Phase Transitions in Factor Graph Models

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

vorgelegt von
Matija Pasch
am 01.03.2023

Erstgutachter: Konstantinos Panagiotou
Zweitgutachter: Will Perkins
Drittgutachter: Charilaos Efthymiou
Tag der mündlichen Prüfung: 06.10.2023

Eidesstattliche Versicherung

(siehe Promotionsordnung vom 12.07.11, §8, Abs. 2 Pkt. 5.)

Ich versichere hiermit an Eidesstatt, dass die vorgelegte Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, 03.02.2024

Matija Pasch

Abstract

Factor graphs provide a unified framework for the discussion of constraint satisfaction problems (CSPs) like boolean satisfiability or graph coloring, communication through noisy channels using e.g. LDPC or LDGM codes, statistical inference tasks like community detection, and spin glasses such as the Potts or the p-spin model. However, their most intriguing feature is that these objects are both versatile and conceptually simple.

Intuitively, a factor graph G is a bipartite graph. The nodes are given by variables and factors, and each factor is equipped with a weight function, whose arity agrees with the factor degree. Values assigned to the variables of G are communicated through the edges of the graph to the factors, which then allows to evaluate the weights. The total weight $\psi_G(\sigma)$ of an assignment σ is then the product of the individual weights over all factors.

Variants of factor graphs were developed and studied independently in mathematics, physics and computer science since the 1960's, up until the discovery of similarities in the model and problem descriptions in the 1990's. Subsequently, the fusion of the available methods, both rigorous and non-rigorous, and of the established approaches led to major breakthroughs in the field.

We study phase transitions in random factor graph models, that is we specify a distribution on (factor) graphs with n variables, such that the (expected) factor to variable ratio has the desired asymptotic behavior as n grows large. Then, we analyze the limiting behavior of quantities derived from the random graph as functions of the factor to variable ratio. For a given ratio, a phase transition occurs whenever the limiting behavior significantly changes at this point (e.g. a discontinuity).

We exclusively consider graphs with weight functions of constant arity, or equivalently k-wise interactions for fixed k, and finite constant variable domains, so all weight functions are of the form $\{1, \dots, q\}^k \to \mathbb{R}_{\geq 0}$. Further, we only discuss the arguably most widespread models, the (binomial or uniform) Erdős–Rényi and the (uniformly) random regular factor graph. The key quantity of our discussion is the partition function $Z(G) = \sum_\sigma \psi_G(\sigma)$, i.e. the sum of the total weights $\psi_G(\sigma)$ over all assignments $\sigma \in \{1, \dots, q\}^n$ to the n variables under the factor graph G.

In the first part of this thesis we consider the Erdős–Rényi model at positive temperature, meaning that all weights are strictly positive. Under mild assumptions on this model, denoted by $\boldsymbol{G}$, we establish the location of the condensation threshold $\alpha_c \in \mathbb{R}_{\geq 0}$, that is for all (limiting) factor to variable ratios $\alpha \in [0, \alpha_c]$ below the threshold, the unruly quenched free entropy (density) coincides with the very accessible annealed free entropy (density), i.e. $\lim_{n\to\infty} \mathbb{E}[\frac{1}{n}\ln(Z(\boldsymbol{G}))] = \lim_{n\to\infty} \frac{1}{n}\ln(\mathbb{E}[Z(\boldsymbol{G})])$, while for ratios $\alpha > \alpha_c$ above the threshold the quenched free entropy is strictly smaller than the annealed counterpart. We establish this result by proving that the quenched free entropy of the planted model converges to the supremum of the corresponding Bethe free entropy, which with some additional effort further allows to establish the limiting mutual information (for graphical channels) and the information-theoretic threshold (for community detection). Due to significantly weaker assumptions and stronger (quantified uniform) results, this part is devoted to a direct generalization and strengthening of the results by Coja-Oghlan, Krzakala, Perkins and Zdeborová in 2018.

The location of the condensation threshold in this generality is still an open problem if the weights may vanish, specifically for CSPs, where the weights only take values in $\{0, 1\}$. In this context, the condensation threshold is conjectured to be responsible for the easy to hard transition, that is below the threshold solutions can be found easily, using local search methods, while finding solutions efficiently above the threshold may not be possible at all or require advanced techniques like survey propagation. In the remainder of this thesis, we turn to the satisfiability thresholds of CSPs, that is below the threshold there exist solutions (with high probability) and hence they can be found in exponential time, while above the threshold the solution space is empty, hence no algorithm finds solutions in this regime. Deriving general results for these thresholds is very ambitious, so we confine ourselves to specific problems. In the second part of the thesis we discuss implications of the satisfiability threshold for perfect matchings in hypergraphs, which has been established by Kahn in 2022. Moreover, Kahn showed that this threshold coincides with the threshold for the existence of isolated vertices in a very strong sense, that is, in the standard hypergraph process formulation the hitting time for the disappearance of the last isolated vertex coincides with the hitting time for the existence of a perfect matching (and the bounds on the hitting time yield the location of both thresholds). Riordan, extended by Heckel in 2021, used an ingenious coupling to derive the threshold for k-clique factors, that is a set of disjoint cliques of size k in the graph such that every vertex appears in exactly one clique, from the perfect matching threshold. We use Kahn's hitting time result to strengthen Riordan's and Heckel's result, by establishing that the hitting time for the k-clique factor coincides with the hitting time for the k-clique cover, i.e. a set of k-cliques in the graph such that every vertex appears in at least one clique. These results are not only of interest in their own right, they also illustrate how closely thresholds may be related and that very involved problems can possibly be reduced to significantly simpler problems, e.g. by checking for necessary conditions.

In the last part of this thesis we turn to an extension of the perfect matching threshold for random regular k-uniform hypergraphs. This threshold has been established by Cooper, Frieze, Molloy and Reed in 1998. Both this problem and the existence of exact covers, studied by Moore in 2016, are equivalent to the 1-in-k occupation problem, i.e. a selection of vertices such that each hyperedge contains exactly one selected vertex. We discuss the 2-in-k occupation problem and show that the first moment bound (the root of the annealed free entropy) is tight for all k. In the proof we exploit the connection of an associated optimization problem regarding the overlap of satisfying assignments to a fixed point problem inspired by belief propagation, a message passing algorithm developed for solving such CSPs.

Zusammenfassung

Faktorgraphen bieten eine einheitliche Modellierungsgrundlage für Constraint Satisfaction Problems (CSPs, auch Bedingungserfüllungsprobleme) wie das Erfüllbarkeitsproblem der Aussagenlogik oder Graphenfärbung, für die Kommunikation über einen gestörten Übertragungskanal, zum Beispiel mittels LDPC oder LDGM Codes, für statistische Inferenz-Probleme wie die Gemeinschaftserkennung und für Spin-Gläser wie das Potts- oder das p-Spin-Modell. Insbesondere zeichnen sich diese Objekte jedoch dadurch aus, dass sie sowohl vielfältig einsetzbar als auch konzeptionell einfach sind.

Unter einem Faktorgraphen G kann man sich einen bipartiten Graphen vorstellen. Die zwei Knotenmengen sind Variablen und Faktoren, und jeder Faktor ist mit einer Gewichtsfunktion ausgestattet, deren Stelligkeit mit dem Faktorgrad übereinstimmt. Den Variablen zugewiesene Werte werden über die Kanten des Graphen den Faktoren kommuniziert, wo dann die zugehörigen Gewichte bestimmt werden. Das Gesamtgewicht $\psi_G(\sigma)$ einer Belegung σ ist schließlich das Produkt der individuellen Gewichte über alle Faktoren.

Varianten von Faktorgraphen wurden seit den 1960er Jahren unabhängig in der Mathematik, der Physik und der Informatik entwickelt und untersucht, bis in den 1990er Jahren Ähnlichkeiten bei den Modellen und Fragestellungen bemerkt wurden. Daraufhin führte die Zusammenführung bekannter Verfahren, ob rigoros oder nicht, und etablierter Ansätze zu bedeutenden Durchbrüchen.

Wir untersuchen Phasenübergänge in zufälligen Faktorgraphmodellen, das heißt wir legen eine Verteilung auf (Faktor-)Graphen mit n Variablen fest, sodass das (erwartete) Verhältnis von Faktoren zu Variablen mit wachsendem n das gewünschte asymptotische Verhalten zeigt. Dann analysieren wir das Grenzverhalten von aus dem Zufallsgraphen abgeleiteten Größen als Funktionen des Faktor-Variablen-Verhältnisses. Für ein gegebenes Verhältnis findet ein Phasenübergang statt, wenn sich das Grenzverhalten an diesem Punkt maßgeblich ändert (zum Beispiel eine Unstetigkeitsstelle).

Wir betrachten ausschließlich Graphen mit fester Stelligkeit, anders ausgedrückt k-weise Interaktionen mit festem k, und endlichem festen Wertebereich für die Variablen, das heißt alle Gewichtsfunktionen sind von der Form $\{1, \ldots, q\}^k \to \mathbb{R}_{\geq 0}$. Außerdem besprechen wir nur die wohl am weitesten verbreiteten Modelle, den (binomialen oder gleichverteilten) Erdős–Rényi- und den (gleichverteilten) zufälligen regulären Faktorgraphen. Unsere besondere Aufmerksamkeit gilt der Zustandssumme $Z(G) = \sum_\sigma \psi_G(\sigma)$, also der Summe der Gesamtgewichte $\psi_G(\sigma)$ über alle Belegungen $\sigma \in \{1, \ldots, q\}^n$ der n Variablen im Faktorgraphen G.

Im ersten Teil der Dissertation widmen wir uns dem Erdős–Rényi-Modell bei positiver Temperatur, was bedeutet, dass alle Gewichte strikt positiv sind. Unter sehr schwachen Modellannahmen bestimmen wir den Kondensationspunkt $\alpha_\mathrm{c} \in \mathbb{R}_{\geq 0}$, das heißt, für alle Faktor-Variablen-Verhältnisse $\alpha \in [0, \alpha_\mathrm{c}]$ unter dem Schwellwert stimmt die eigensinnige Quenched Free Entropy (Density) mit der sehr zugänglichen Annealed Free Entropy des zufälligen Faktorgraphen $\boldsymbol{G}$ überein, also $\lim_{n\to\infty} \mathbb{E}[\frac{1}{n}\ln(Z(\boldsymbol{G}))] = \lim_{n\to\infty} \frac{1}{n}\ln(\mathbb{E}[Z(\boldsymbol{G})])$, während für Verhältnisse $\alpha > \alpha_\mathrm{c}$ über dem Schwellenwert die Quenched Free Entropy strikt kleiner ist als die Annealed Free Entropy. Wir überzeugen uns von diesem Ergebnis, indem wir beweisen, dass die Quenched Free Entropy eines verwandten, gewichteten, Modells gegen das Supremum des entsprechenden Bethe-Funktionals konvergiert, was mit etwas zusätzlichem Aufwand die Bestimmung des Grenzwertes der gegenseitigen Information (für eine Klasse von gestörten Übertragungskanälen) und des informationstheoretischen Schwellenwertes (für Gemeinschaftserkennung) ermöglicht. Wegen wesentlich schwächerer Annahmen und stärkeren (quantifizierten, gleichmäßigen) Ergebnissen, widmet sich der erste Teil also der direkten Verallgemeinerung der Resultate von Coja-Oghlan, Krzakala, Perkins und Zdeborová aus 2018.

Für nichtnegative Gewichte ist der Wert des Kondensationspunktes in dieser Allgemeinheit weit-

erhin unbekannt und damit insbesondere für CSPs, bei denen die Gewichte nur die Werte $\{0, 1\}$ annehmen. In diesem Zusammenhang wird vermutet, dass der Kondensationspunkt für den Übergang von einfachen zu schweren Instanzen verantwortlich ist, das heißt, unterhalb des Schwellenwertes können Lösungen, mithilfe lokaler Verfahren, einfach und effizient bestimmt werden, während die effiziente Berechnung von Lösungen überhalb des Schwellenwertes aussichtslos, oder zumindest auf konzeptionell anspruchsvollere Methoden wie Survey Propagation angewiesen ist. Im weiteren Verlauf dieser Dissertation befassen wir uns mit dem Schwellenwert für Erfüllbarkeit von CSPs. Unter dem Schwellenwert existieren also (mit hoher Wahrscheinlichkeit) Lösungen und können in exponentieller Laufzeit bestimmt werden, während der Lösungsraum über dem Schwellenwert leer ist und dementsprechend kein Algorithmus Lösungen berechnen kann. Allgemeine Ergebnisse für solche kritischen Werte zu beweisen, ist sehr ambitioniert und daher beschränken wir uns auf spezifische Probleme. Im zweiten Teil dieser Dissertation besprechen wir Implikationen der Schwellenwerte für perfekte Matchings in Hypergraphen, die letztes Jahr von Kahn bestimmt wurden. Darüber hinaus zeigte Kahn, dass der Zusammenhang dieser Schwellenwerte mit den Schwellenwerten für die Existenz von isolierten Knoten sehr tiefgehender Natur ist, nämlich dass die entsprechenden Übergangszeiten im gewöhnlichen Hypergraphenprozess übereinstimmen (und die Schranken an die Übergangszeiten die Schwellenwerte implizieren). Riordan, ergänzt durch Heckel im Jahr 2021, benutzte ein brilliantes Coupling, um aus Kahn's Schwellenwert für perfekte Matchings den Schwellenwert für k-Cliquen-Faktoren abzuleiten, also eine Menge von disjunkten Cliquen der Größe k im binomialen Zufallsgraphen, sodass jeder Knoten in genau einer Clique enthalten ist. Wir verwenden Kahn's Ergebnis für die Übergangszeiten, um Riordans und Heckels Ergebnis zu verstärken, indem wir zeigen, dass im gewöhnlichen Graphenprozess die Übergangszeit für k-Cliquen-Faktoren mit der Übergangszeit für die Existenz einer k-Cliquen-Überdeckung übereinstimmt, also einer Menge von k-Cliquen im Zufallsgraphen, so dass jeder Knoten in mindestens einer Clique enthalten ist. Diese Ergebnisse zeigen also insbesondere auf, dass solche kritischen Werte sehr eng miteinander verbunden sein können und dass sehr komplizierte Probleme unter Umständen auf sehr viel einfachere Probleme reduziert werden können, zum Beispiel durch das Überprüfen von notwendigen Bedingungen.

Im letzten Teil dieser Dissertation wenden wir uns der Verallgemeinerung der Schwellenwertergebnisse für perfekte Matchings im zufälligen regulären k-uniformen Hypergraphen zu. Dieser kritische Wert wurde von Cooper, Frieze, Molloy und Reed im Jahr 1998 bestimmt. Sowohl dieses Problem als auch die von Moore im Jahr 2016 untersuchten exakten Überdeckungen sind äquivalent zum 1-in-k Occupation-Problem, also eine Auswahl von Knoten, sodass jede Hyperkante genau einen dieser Knoten enthält. Wir besprechen das 2-in-k Occupation-Problem und zeigen, dass die durch die erwartete Anzahl von Lösungen gegebene Schranke (die Nullstelle der Annealed Free Entropy) scharf ist, für alle k. Unser Beweis erfordert die Lösung eines Optimierungsproblems bezüglich der Überlappung von zwei Lösungen des Occupation-Problems, welches wir auf ein Fixpunktproblem reduzieren, das von Belief Propagation inspiriert ist, ein für die Lösung solcher CSPs entwickelter Message-Passing-Algorithmus.

Acknowledgements

I truly admire Kosta Panagiotou's everlasting patience, his lighthearted optimism and his unconditional love and passion for maths, be it research or teaching. It is this spark in his eyes that piqued my curiosity and lured me into the study of random CSPs. After he gently introduced me to belief propagation in my master thesis, he supported me on my adventure to identify interesting questions in the realm of spin glasses and devise their answers. Rather than dictating the direction, he let me explore the terrain and along the way, he did not only introduce me to strategies, countless results and methods whenever I hit a dead end, he made me appreciate their usefulness and joined me in cherishing elegant arguments. Also beyond the world of rigorous argumentation, he was a great mentor and team lead, and by that, made me understand what this means in the first place. He taught me how to tackle complex problems in general, familiarized me with the publication process, introduced me to the scientific community, provided comfort in times of despair, and was also always up for a relaxed chat. Thank you, Kosta, for this amazing journey, all the beautiful insights and your company!

Another dear colleague who accompanied me on this journey is Noela Müller. Since we met at my first conference, the AofA in 2019, I had the pleasure to work, collaborate, engage in insightful discussions and amusing chats with her. Over the past years, she has helped me with proofs, pointed out events, introduced me to researchers and always offered her support. Thanks so much, Noela!

I feel honored to have worked and collaborated with Annika Heckel. Witnessing her focus, sharp wit and expertise is a fascinating and humbling experience, in our team lunch conversations, but even more so during our project. She patiently introduced me to a new argumentation style, answered my numerous questions and dispelled my doubts. Thank you for your guidance and support, Annika!

I also want to thank Marc Kaufmann for his commitment, his contribution, and our entertaining, fruitful discussions during the project and beyond. I am grateful to Angelika Steger for organizing the amazing and productive Buchboden retreats, which work wonders in networking researchers.

Establishing the condensation threshold for given degree distributions was a very demanding venture, both conceptually and technically. Next to Noela and Kosta, I want to thank Amin Coja-Oghlan, Max Hahn-Klimroth and Philipp Loick for their hospitality, the insightful discussions and their efforts to tame this beast. I am grateful to Markus Heydenreich, Ulrich Mansmann, Thilo Meyer-Brandis, Annika Steibel, Simon Reisser, Vincent Gögl and the fellow contributors for their commitment, the exciting events and discussions surrounding the modeling and analysis of epidemics in complex networks. The funding by the Volkswagen Stiftung for this project is appreciated. I am grateful to the European Research Council for the funding that facilitated this thesis. Further, I want to thank Will Perkins and Charilaos Efthymiou for accepting to review this thesis, and Sabine Jansen for chairing my disputation.

Thank you, Leon, my beloved office mate, for saving my day with your kind-hearted spirit countless times, for the support with the bureaucracy, the travels, the proofreading and the chats about enumerative combinatorics! Thank you, Thomas, my dear colleague, for the fun times, the introduction to the orthant model and the insightful discussions! Also, thank you, Alejandro, Anna, Benedikt, Felix, Jakob, Kilian, Leonid, Nannan, Philipp, Stefan, Steffi, Tamás, Tom and Umberto for the great time during the past few years, we've been an amazing team! I am grateful to the administration of our Institute for the organization, resources and maintenance that facilitated my travels and my work, and in particular to Michaela Platting, who managed to turn bureaucracy into a pleasant experience.

Finally, I want to thank my family and my friends, who bear with me, who have supported me for decades and who have stood by my side along this rollercoaster ride of a life. I am truly lucky to have you, thank you for everything!

Contributing Manuscripts

This thesis is based on the following manuscripts, which were developed by me, the thesis' author Matija Pasch, in collaboration with my PhD supervisor Prof. Dr. Konstantinos Panagiotou, with Annika Heckel, Marc Kaufmann and Noela Müller:

(I) K. Panagiotou, M. Pasch. Satisfiability thresholds for regular occupation problems, 2023 [103]. The preprint is available at `https://arxiv.org/abs/1811.00991`. A weaker version of this result was published in 2019 in the 46th ICALP, volume 132 of LIPIcs [101].

Section 2.3 and Section 5 are based on this contribution.

(II) K. Panagiotou, M. Pasch. Mutual Information, Information-Theoretic Thresholds and the Condensation Phenomenon at Positive Temperature, 2022 [102]. The preprint is available at `https://arxiv.org/abs/2207.11002`.

Section 2.1 and Section 3 are based on this contribution.

(III) A. Heckel, M. Kaufmann, N. Müller, M. Pasch. The hitting time of clique factors, 2023 [64]. The preprint is available at `https://arxiv.org/abs/2302.08340`.

Section 2.2 and Section 4 are based on this contribution.

I stated the problem in [103] and developed the proof strategy based on [36, 96]. I further implemented the proofs, guided by Konstantinos Panagiotou in the initial stages and in particular for the small subgraph conditioning, and we jointly worked on the presentation of the results. Similarly, I formulated the problem in [102], developed the proof strategy based on [33] and implemented the proofs. Then I worked jointly with Konstantinos Panagiotou on the presentation of the results. The problem in [64] was stated by Annika Heckel in the course of the Buchboden retreat in 2021. I significantly contributed to the development of the proof strategy and the implementation of the proofs. I further strengthened the results in [63, 64] for $k = 3$ and proposed the extension to Corollary 4.4 for the discussion in this thesis. Here, Annika Heckel proposed the improvement upon the Poisson coupling in the proof of Lemma 4.17, and the beautiful argument for the proof of Corollary 4.4.

Due to space limitations, the following contributions are not covered in this thesis. In collaboration with Amin Coja-Oghlan, Max Hahn-Klimroth, Philipp Loick, Noela Müller and Konstantinos Panagiotou we published the paper "Inference and mutual information on random factor graphs" in the conference proceedings of the 38th STACS in 2021 [31]. Moreover, I have been involved in an ongoing research project led by Markus Heydenreich, Ulrich Mansmann and Konstantinos Panagiotou dedicated to the analysis of epidemic models on random graphs.

Funding

The research leading to this thesis has received funding from the European Research Council, ERC Grant Agreement 772606–PTRCSP.

Contents

1 Introduction

We motivate and prepare the upcoming formal discussion, and define central notions, including constraint satisfaction problems, phase transitions, satisfiability thresholds and factor graphs. Next to illustrating the broad scope of applicability, the presented examples introduce the relevant models and existing results that are extended in this work.

1.1 Constraint Satisfaction Problems

Right from the definition, it will be obvious that constraint satisfaction problems (CSPs) are both conceptually simple and immensely useful in practice. An instance C of a CSP is given by n variables $[n]=\{1,\dots,n\}$ taking one of q values $[q]$ on the one hand, and m constraints $[m]$ on the other, where each constraint $a\in[m]$ is equipped with a set $\mathcal{V}_a\subseteq[n]$ of involved variables and a set $\mathcal{S}_a\subseteq[q]^{\mathcal{V}_a}$ of satisfying assignments. An assignment $\sigma\in[q]^n$ is a solution of C if for all $a\in[m]$ we have $(\sigma_i)_{i\in\mathcal{V}_a}\in\mathcal{S}_a$, i.e. the assignment σ satisfies each constraint a.

Given C, there is a number of associated problems. First and foremost, we have the problem of identifying a solution, and the closely related decision problem if a solution even exists. On a more granular level, we may attempt to determine the entire solution space, and in particular the number of solutions. Whether solutions exist or not, we may ask for one (or all or the number of) assignments that satisfy the maximum number (or at least a specified number) of constraints. All of these problems, including the decision, the enumeration and the optimization version, are NP-complete in general, meaning that we assume that it will take exponentially long (in terms of n) to solve any of these for some instance.

We illustrate these concepts using a well-known problem from graph theory, namely coloring. To be specific, we consider the decision problem q-COL if a given graph with n vertices and m edges admits a proper coloring with q colors. An instance of q-COL is just a graph G, which is, formally, equipped with an (irrelevant) order $(e_a)_{a\in[m]}$ of the edges, the involved variables being the edge $e_a=\{i_a,j_a\}\subseteq[n]$, and the satisfying assignments being $\mathcal{S}_a=\{\tau\in[q]^{e_a}:\tau_{i_a}\neq\tau_{j_a}\}$. Now, the solutions of G in the sense above are exactly the proper colorings of G. Thus, the decision problem asks if G is q-colorable, the enumeration problem asks for the number of such colorings and the optimization problem asks for a not necessarily proper q-coloring of G with as few monochromatic edges as possible. We know that also the subclass q-COL of all CSPs is NP-complete for $q\geq 3$. Another classic NP-complete example is the exact cover problem in hypergraphs. Recall that the hyperedges of a k-uniform hypergraph $H=([n],\mathcal{E})$ with $|\mathcal{E}|=m$ are k-subsets $\mathcal{E}\subseteq\binom{[n]}{k}$, where $\binom{[n]}{k}=\{E\subseteq[n]:|E|=k\}$. A solution of the exact cover problem is a subset $\mathcal{V}\subseteq[n]$ of the vertices such that each hyperedge $E\in\mathcal{E}$ is incident to exactly one vertex in $\mathcal{V}$. In the CSP version of this problem we choose $q=2$ over the values $\{0,1\}$, i.e. a binary CSP. Then, we associate the constraints with the hyperedges $E\in\mathcal{E}$ as for q-COL, thus the involved variables are E, and the satisfying assignments are $\{\tau\in\{0,1\}^E:\|\tau\|_1=1\}$, where $\|\cdot\|_r$ for $r\in[1,\infty]$ is the r-norm on $\mathbb{R}^d$. Clearly, the solutions $\sigma\in\{0,1\}^n$ of H are exactly the indicators of the solutions $\mathcal{V}=\sigma^{-1}(1)\subseteq[n]$.

However, CSPs and subclasses thereof are not only present in graph theory. A detailed overview of both CSPs and their applications in discrete mathematics, complexity theory, physics, artificial intelligence, image processing, mechanical engineering, transportation, scheduling, natural language processing, robotics, biology and more can be found in [85, 118, 81, 124] and references therein. Since these problems appear in so many disciplines, but are so hard to tackle in both theory and practice, a great amount of research over the last centuries and in particular the last four decades is devoted to separating easy from hard instances, respectively to identifying structural properties that are responsible for this hardness. One approach is to systematically identify subclasses of CSPs and instances for which the decision problem can be solved, say hypergraphs with maximum degree 1 for the exact cover problem, or the enumeration and optimization versions are tractable.

1.2 Random CSPs, Phase Transitions and Satisfiability Thresholds

We follow another line of research, where we equip the instances of (a subclass of) CSPs with probability distributions, usually for fixed n. For example, this allows to determine whether solutions exist for *most* instances in a rigorous manner, or to determine the *average* number of solutions. The certainty that CSPs over a small number n of variables can be eventually dealt with using computational power, combined with the simplifications that arise from the restriction to asymptotics, steers our focus towards large n. Now, a fairly weak version of the decision problem asks if *most* instances are satisfiable, in the sense that the probability that a solution exists tends to 1 as n increases.

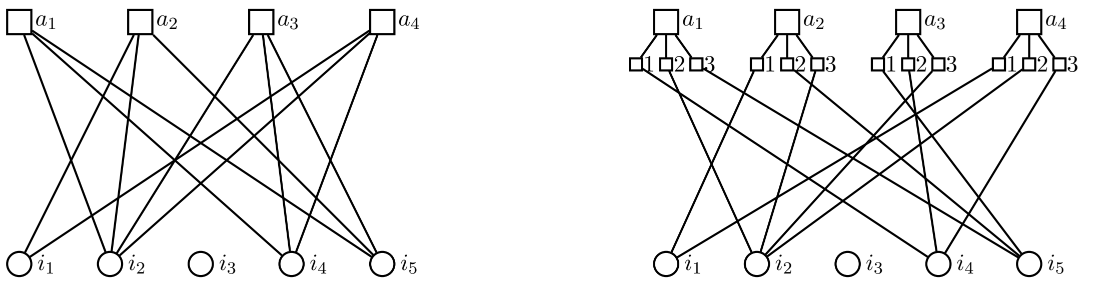
This approach was introduced by Erdős and Rényi in their seminal paper [47], where they discussed the decision problem if a giant component exists. To be specific, consider the (uniform) Erdős-Rényi graph $\boldsymbol{G}_{n,m}$ which is obtained by choosing a set of m edges from $\binom{[n]}{2}$ uniformly at random. For a given average degree $d \in \mathbb{R}_{\geq 0}$ let $P_d(n)$ be the probability that there exists a (connected) component in $\boldsymbol{G}_{n,m}$, where $m = \lfloor dn/2 \rfloor = \max \mathbb{Z}_{\leq dn/2}$, of size at least $\sqrt{n}$. The results of [47, 57] suggest that there exists a *sharp phase transition* at $d = 1$, that is, for $d > 1$ there exists a component of size at least $\sqrt{n}$ *with high probability* (whp), meaning $\lim_{n\to\infty} P_d(n) = 1$, and for $d < 1$ there exists no such component whp, meaning $\lim_{n\to\infty} P_d(n) = 0$.

Ever since, such phase transitions have been established for various problems, a rather recent prominent example being the satisfiability threshold for random k-SAT for large k [40]. Not only the existence, but also the location of the satisfiability threshold for q-COL, that is, a critical average degree $d_{\mathrm{c}} \in \mathbb{R}_{>0}$ such that $\boldsymbol{G}_{n,m}$ is q-colorable whp for $d < d_{\mathrm{c}}$ and $\boldsymbol{G}_{n,m}$ is not q-colorable whp for $d > d_{\mathrm{c}}$, are conjectured for decades, but still outstanding [14].

Moreover, such results have not only been extended to other problems, but to other models as well. Both the uniform and the closely related binomial Erdős-Rényi graph have been discussed since the 50s, and have subsequently been generalized to the uniformly random k-uniform hypergraph, where a set of m hyperedges in $\binom{[n]}{k}$ is drawn uniformly at random, and the binomial k-uniform hypergraph, where each hyperedge in $\binom{[n]}{k}$ is included independently with the same probability. The third standard model is the random d-regular k-uniform hypergraph, which is chosen uniformly at random from the set of all d-regular k-uniform hypergraphs, i.e. hypergraphs with hyperedges in $\binom{[n]}{k}$ such that each vertex is incident to exactly d hyperedges. As opposed to the other two closely related models, this model is far more rigid and usually requires additional or different arguments. However, also for this model, methods have been developed that allowed to rigorously locate satisfiability thresholds, one example being the exact cover [96, 36] introduced above.

1.3 Factor Graphs

Now, we turn to the representation of CSPs using factor graphs. For this purpose, fix the number q of values, the total number n of variables as before, and further let m be the number of factors, which can be thought of as generalized constraints. Each factor $a \in [m]$ is equipped with an arity k_a, an ordered neighborhood $v_a \in [n]^{k_a}$ and a weight function $\psi_a : [q]^{k_a} \to \mathbb{R}_{\geq 0}$. This determines the factor graph $G = (v_a, \psi_a)_a$, where we keep the number n of variables implicit. Depending on the context,



(a) Factor Graph Visualization

(b) Factor Graph With Ordered Neighborhoods

Figure 1: On the left, we see a factor graph for $k = 3$, where the rectangles and circles depict the factors and variables respectively. The weights and the neighborhood orders are kept implicit. The figure on the right shows the same factor graph, with explicitly modeled ordered neighborhoods.

the factor graph G can be thought of as a bipartite graph (cf. Figure 1), with the vertices given by the variables $[n]$ and the factors $[m]$, or as a hypergraph, with variables as vertices and factors as (labeled) hyperedges. The weight of an assignment $\sigma \in [q]^n$ under G is

$$\psi_G(\sigma) = \prod_{a\in[m]} \psi_a\left(\sigma_{v_a}\right), \quad \sigma_v = \left(\sigma_{v(h)}\right)_{h\in[k]}, \, v \in [n]^k.$$

Assuming that for all a the weight function $\psi_a : [q]^{k_a} \to \{0, 1\}$ is an indicator, we recover an instance of a CSP where the constraint $a \in [m]$ involves the variables $\mathcal{V}_a = \{v_a(h) : h \in [k_a]\}$ and is satisfied by $\mathcal{S}_a = \{\tau \in [q]^{\mathcal{V}_a} : \psi_a((\tau_{v_a(h)})_{h\in[k]}) = 1\}$. Usually, here we also consider k_a distinct variables $\mathcal{V}_a$, so the arity $k_a = |\mathcal{V}_a|$ is the number of involved vertices. Notice that the weight $\psi_G(\sigma) \in \{0, 1\}$ of $\sigma \in [q]^n$ indicates if σ is a solution to the CSP, so the solution space is simply $\psi_G^{-1}(1)$.

The values $[q]$ are also referred to as spins or colors, the variables $[n]$ as particles and the assignments $\sigma \in [q]^n$ as (spin) configurations or spins. Another key quantity is the partition function $Z(G) = \sum_{\sigma\in[q]^n} \psi_G(\sigma)$, or the number of solutions in the context of CSPs. If the partition function is positive, the Gibbs measure or Boltzmann distribution $\mu_G : [q]^n \to \mathbb{R}_{\geq 0}$, $\sigma \mapsto \frac{\psi_G(\sigma)}{Z(G)}$, is closely related to structural properties of the solution space in the context of CSPs. Hence, this simple model captures all relevant aspects of CSPs regarding the problems in Section 1.1.

For example, to model q-COL we let $k_a = 2$, further let the involved variables $\mathcal{V}_a \neq \mathcal{V}_b$ be distinct, and let the two variables $v_a = (v_a(1), v_a(2))$ be distinct, to ensure that G reflects a simple graph, where $v_a \in [n]^2$ represents an edge. The appropriate weight $\psi_a(\tau) = \mathbb{1}\{\tau_1 \neq \tau_2\}$ indicates if the colors of the endpoints are distinct, further $\psi_G(\sigma)$ indicates if $\sigma \in [q]^n$ is a proper coloring and $Z(G)$ is the number of proper q-colorings of G viewed as a graph.

We proceed similarly for the exact cover, using $q = 2$, colors $\{0, 1\}$, and $\psi_a(\tau) = \mathbb{1}\{\|\tau\|_1 = 1\}$. However, for this example there are two prominent ways to map the factor graph to a hypergraph. As above, we may map the variables to the vertices and the factors to the hyperedges, enforcing distinct involved variables $\mathcal{V}_a \neq \mathcal{V}_b$ and distinct neighbors $(v_a(1), \ldots, v_a(k))$. Then, a solution $\sigma \in \psi_G^{-1}(1)$ corresponds to an exact vertex cover $\sigma^{-1}(1) \subseteq [n]$, i.e. each hyperedge $\mathcal{V}_a$ is incident to exactly one vertex in $\sigma^{-1}(1)$. On the other hand, we may map the variables to the hyperedges and the factors to the vertices, so here we would enforce distinct neighborhoods $\mathcal{F}_i = \{a \in [m] : i \in v_a([k])\}$ and unique neighbors, i.e. $|v_a^{-1}(i)| = 1$ for $a \in \mathcal{F}_i$. Now, a solution $\sigma \in \psi_G^{-1}(1)$ corresponds to a set $\mathcal{M} = \{\mathcal{F}_i : i \in \sigma^{-1}(1)\}$ of hyperedges, such that each vertex $a \in [m]$ is incident to exactly one hyperedge in $\mathcal{M}$ – thus, the solutions are perfect matchings.

1.4 Factor Graph Models and Applications

Analogously to the discussion in Section 1.2, we introduce some well-known distributions on general factor graphs in Section 1.4.1. The related teacher-student model is introduced in Section 1.4.2, where the random graph prefers a large weight on a specific assignment. In Section 1.4.3 we discuss phase transitions and thresholds, like the satisfiability threshold in Section 1.2, for general factor graph models. Then we introduce several applications of factor graphs, to motivate the generalization. In Section 1.4.4 we introduce spin glass models from physics, and the stochastic block model in Section 1.4.5, which is prominent in statistics, machine learning and network science. In Section 1.4.6 we discuss examples from coding theory.

1.4.1 Random Factor Graph Models. As for (hyper-) graphs and as motivated above, we also consider random factor graphs. The uniform Erdős-Rényi model is given by the following parameters.

- The number $q \geq 1$ of colors.
- The arity $k \geq 1$.
- A random weight function $\boldsymbol{\psi} : [q]^k \to \mathbb{R}_{\geq 0}$ with law p.
- The number $n > 0$ of variables.
- The number $m \geq 0$ of constraints.

The uniformly random factor graph $\boldsymbol{G}_{n,m} = (\boldsymbol{v}_a, \boldsymbol{\psi}_a)_{a \in [m]}$ is obtained by choosing all neighborhoods $\boldsymbol{v}_a \in [n]^k$ uniformly at random, and $\boldsymbol{\psi}_a$ from p, all mutually independent. The closely related binomial factor graph $\boldsymbol{G}_{n,\pi}$ for $\pi \in [0,1]$ is obtained by including each neighborhood $v \in [n]^k$ with probability π and a weight function $\boldsymbol{\psi}_v$ from p for each (included) neighborhood, all mutually independent. For CSPs, we would usually replace $[n]^k$ by the subset of pairwise distinct variables, or even by unordered neighborhoods $\binom{[n]}{k}$, depending on the context. Moreover, as discussed above, we would usually enforce that the neighborhoods are distinct. These restrictions can be achieved by conditioning on the respective events and lead to the desired standard models. However, for the sake of brevity we only consider possibly duplicate neighborhoods $[n]^k$ unless stated otherwise and postpone the extension to the other models. Finally, for a degree $d > 0$, the random regular factor graph $\boldsymbol{G}_{n,d}$ (with possibly duplicate neighborhoods) is obtained from the uniform factor graph $\boldsymbol{G}_{n,m}$ over $\binom{[n]}{k}$ by conditioning on the event that all variable degrees $|\{a \in [m] : i \in \boldsymbol{v}_a\}|$ for $i \in [n]$ are equal to d, or alternatively by choosing a (d,k)-biregular graph uniformly at random and equipping it with weights as above (if this is well-defined).

1.4.2 The Teacher-Student Model. The three factor graph models introduced so far are unbiased, in the sense that there is no preference of certain neighborhoods or weights inherent to the model, relative to the law p. While these are the most prominent distributions for random CSPs, other relevant distributions have emerged over the decades. A particularly prominent family is the *planted model*, where we consider the random factor graph conditional to the event that a specific assignment $\sigma \in [q]^n$ is a solution. In this context, the three unbiased random factor graphs introduced so far are the *null models*, while the planted models generalized from CSPs, where the weights are indicators, to general factor graphs are given as follows.

For an assignment $\sigma \in [q]^n$, the *ground truth*, and the uniform null model $\boldsymbol{G}_{n,m}$, whenever the weight is positive with positive probability $\mathbb{P}(\psi_{\boldsymbol{G}_{n,m}}(\sigma) > 0) > 0$, the teacher-student model, or planted model, $\boldsymbol{G}^*_{n,m}(\sigma)$ is given by the Radon-Nikodym derivative $G \mapsto \psi_G(\sigma)/\mathbb{E}[\psi_{\boldsymbol{G}_{n,m}}(\sigma)]$ with respect to $\boldsymbol{G}_{n,m}$. The planted model for the regular null model $\boldsymbol{G}_{n,d}$ is defined analogously. For the binomial null model $\boldsymbol{G}_{n,\pi}$, we take a slightly different approach to maintain independence. Let $(\boldsymbol{v}, \boldsymbol{\psi})$ be independent, with $\boldsymbol{v}$ uniform from $[n]^k$ and $\boldsymbol{\psi}$ from p. If we have $\mathbb{E}[\boldsymbol{\psi}(\sigma_{\boldsymbol{v}})] > 0$, then the planted model $\boldsymbol{G}^*_{n,\pi}(\sigma)$ is obtained by independently including each neighborhood $v \in [n]^k$ with probability $\frac{\mathbb{E}[\boldsymbol{\psi}(\sigma_v)]}{\mathbb{E}[\boldsymbol{\psi}(\sigma_{\boldsymbol{v}})]}\pi$ and the weight is given by the Radon-Nikodym derivative $\psi \mapsto \psi(\sigma_v)/\mathbb{E}[\boldsymbol{\psi}(\sigma_v)]$ with respect to $\boldsymbol{\psi}$. Notice that the expected number of factors in both the binomial null model and planted model is $n^k\pi$, which is the motivation for the scaling with $\mathbb{E}[\boldsymbol{\psi}(\sigma_{\boldsymbol{v}})]$ in the definition.

Finally, we consider the uniformly random ground truth $\boldsymbol{\sigma}^* \in [q]^n$, independent of anything else, for the same reason that we consider random CSPs, as detailed in Section 1.2. Further, we choose the uniform distribution for $\boldsymbol{\sigma}^*$ because in most assignments all colors appear roughly the same number of times, and because this is consistent with the partition function $Z(G) = \sum_\sigma \psi_G(\sigma) = q^n\mathbb{E}[\psi_G(\boldsymbol{\sigma}^*)]$.

The planted model has already been mentioned in the seminal paper [47] by Erdős and Rényi, and has accompanied the null model ever since. In the context of CSPs, the planted model serves as a tool to focus on (typical) factor graphs that admit a specific solution, to analyze their structural particularities, their solution spaces, to compare them to factor graphs admitting a different solution, and to compare them to factor graphs from the null model. Inspired by the *cavity method* from physics and based on rigorous results for a large class of CSPs [32], for CSPs that are not biased towards specific colors we assume that $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{n,m}(\boldsymbol{\sigma}^*))$ and $(\boldsymbol{\sigma}_{\boldsymbol{G}_{n,m}}, \boldsymbol{G}_{n,m})$, where $\boldsymbol{\sigma}_G$ with probability mass function μ_G are the Gibbs spins, are very similar for sufficiently small constraint densities, mutually contiguous to be precise, that the partition functions $Z(\boldsymbol{G}_{n,m})$ and $Z(\boldsymbol{G}^*_{n,m}(\boldsymbol{\sigma}^*))$ are comparable and tightly concentrated, amongst other properties – in a nutshell, the planted model with the ground truth can be used to analyze the null model and its solution space through the Gibbs spins. For larger constraint densities, the planted model significantly differs from the null model. The reason is that factor graphs with many solutions are preferred in the planted model since any such solution can serve as ground truth, and the impact of this bias is significant for larger densities. The same reasoning holds for general factor graphs [30, 32].

1.4.3 Phase Transitions. Next to the very intuitive satisfiability threshold in Section 1.2, based on the discussion in Section 1.4.2, we assume that another phase transition occurs, in general factor graphs, at the condensation threshold. A formal definition, discussion and rigorous results can be found in [30, 33, 32]. Also further phase transitions are conjectured, for example at the clustering, rigidity and freezing threshold. An excellent overview can be found in [58, 94].

Proving the existence, let alone the location, of phase transitions in factor graph models is highly non-trivial even for fairly simple specific problems. Celebrated contributions establishing such transitions and the behavior of related quantities, include the quenched free entropy (density) for the Sherrington-Kirkpatrick model [123], inspired by the replica method [106], verifications of physics predictions [79, 78, 60] for channel coding, results [3, 43] for the limiting mutual information in the stochastic block model (SBM), the freezing threshold [88] and the condensation threshold [18] for q-COL, for large enough q, as well as the satisfiability thresholds for k-NAESAT [35] and k-SAT [40], for large enough k, and all inspired by the cavity method. Next to such problem specific discussions, also problem independent results were derived. However, a truly general theory is non-existent, certain cases still have to be distinguished. We focus on the (almost) sparse regime, that is, the number of (hyper-)edges is (almost) linear in the number of vertices. Within this regime, we focus on Erdős-Rényi type and regular factor graphs as introduced in Section 1.4.1.

The seminal paper [54] proves the existence of certain phase transitions. The results in [33] establish not only the existence [4], but also the exact limit of the mutual information, the closely related quenched free entropy, moreover the limit of a certain relative entropy and the (standard) quenched free entropy up to the condensation threshold, and also a qualitative description beyond the threshold, all at positive temperature. These results were subsequently extended and improved under partially weaker, partially stronger, and other assumptions in [30], to determine limiting distributions, to compare and characterize thresholds and bounds, and establish mutual contiguity of models, which is highly useful for a number of applications. The contribution [32] further extends these results to zero temperature, which is e.g. crucial for CSPs.

Rigorous results for satisfiability thresholds at zero temperature for CSPs do not exist for larger classes of models, but a few specific thresholds have been established [85]. A rather recent result is the perfect matching threshold for the Erdős-Rényi hypergraph [71]. In fact, the result is even stronger in that it establishes that the hitting time for the existence of a perfect matching coincides with the hitting time for the existence of a hyperedge cover whp, in the corresponding hypergraph process. The location of this threshold was successfully translated to the threshold for the existence of a k-clique factor in the Erdős-Rényi graph using an ingenious coupling [113, 63].

Satisfiability threshold results for the regular hypergraph also include the perfect matching threshold [36], or equivalently the exact cover threshold [96], amongst others [89, 17, 73, 29, 41, 42].

1.4.4 Spin Glasses. A very influential concept that fueled the intuition for a plethora of results and gave rise to countless conjectures, is the cavity method and the closely related replica theory from physics, the foundation of spin glass theory [87, 85, 77]. For simplicity, we focus on spin glasses that correspond to the uniformly random factor graph from Section 1.4.1 for now.

Let $\boldsymbol{e} : [q]^k \to \mathbb{R}$ be the random energy for given spins $\tau \in [q]^k$, then the Hamiltonian $\boldsymbol{E} : [q]^n \to \mathbb{R}$ is the energy $\boldsymbol{E}(\sigma) = \sum_{a\in[m]} \boldsymbol{e}_a(\sigma_{\boldsymbol{v}_a})$ of the spin configuration $\sigma \in [q]^n$, where $\boldsymbol{G}_{n,m} = (\boldsymbol{v}_a, \boldsymbol{e}_a)_{a\in[m]}$ is obtained by independently choosing $\boldsymbol{v}_a$ uniformly from $[n]^k$ and $\boldsymbol{e}_a$ from $\boldsymbol{e}$. For a given inverse temperature $\beta \in \mathbb{R}_{\geq 0}$ the weight of $\sigma \in [q]^n$ is $\psi_{\boldsymbol{G}_{n,m},\beta}(\sigma) = e^{-\beta \boldsymbol{E}(\sigma)}$, the partition function is $Z_{\boldsymbol{G}_{n,m}}(\beta) = \sum_\sigma \psi_{\boldsymbol{G}_{n,m},\beta}(\sigma)$ and $\mu_{\boldsymbol{G}_{n,m},\beta}(\sigma) = \psi_{\boldsymbol{G}_{n,m},\beta}(\sigma)/Z_{\boldsymbol{G}_{n,m}}(\beta)$ defines the Boltzmann distribution. Hence, for given $\boldsymbol{e}$ and β, this model is equivalent to the factor graph model given by $\boldsymbol{\psi}(\tau) = e^{-\beta \boldsymbol{e}(\tau)} \in \mathbb{R}_{>0}$. This defines a diluted mean-field spin glass model, that is, we do not (necessarily) consider all possible combinations $[n]^k$, and all neighborhoods are equally likely to be considered. We further focus on the sparse case as before, where the number of factors is linear in the number of particles, and fix $m = \lfloor dn/k \rfloor$ for given $d \in \mathbb{R}_{\geq 0}$. This model represents microscopic variables (e.g. atoms or electrons) and their interactions, but in this context we are mostly interested in macroscopic properties (e.g. locations of glass transitions). Hence, similarly to CSPs we focus on the thermodynamic limit, that is, the asymptotics in n for fixed β and d. As opposed to CSPs, for the definition of phase transitions we usually keep d fixed and consider a variation of β.

Here, the two limiting cases $\beta = 0$ and $\beta \to \infty$ deserves special attention. The infinite-temperature limit $\beta = 0$ corresponds to total chaos, meaning that there is no interaction between the particles $[n]$, or formally, that the contributions $\beta \boldsymbol{e}_a \equiv 0$ are trivial. On the other hand, the zero temperature limit $\beta \to \infty$ mimics the case that the weight $\boldsymbol{\psi}(\tau) = e^{-\beta \boldsymbol{e}(\tau)}$ may vanish. While this does establish a connection to the previously discussed CSPs, which is used in applications (e.g. simulated annealing), it is highly non-trivial to determine whether the limit $\beta \to \infty$ can be used to analyze the model at zero temperature, i.e. the model for $\beta = \infty$.

To illustrate the phase transitions and the zero-temperature limit in spin glasses, we consider the arguably most studied quantity, the free entropy (density) $\boldsymbol{\phi}_{n,m}(\beta) = \frac{1}{n}\ln(Z_{\boldsymbol{G}_{n,m}}(\beta))$.[1] The conceptually simple example $\boldsymbol{e} \in \{0,1\}$ almost surely is not only typically used to approximate CSPs, but also serves as a reasonable base case, for which we notice that $\boldsymbol{\phi}_{n,m}(\beta) \in [\ln(q) - \frac{\beta m}{n}, \ln(q)]$ almost surely, i.e. this quantity remains bounded in the thermodynamic limit. In general and under weak assumptions, the free entropy is self-averaging, that is, it concentrates around the quenched free entropy $\phi_{\mathrm{q},n,m}(\beta) = \mathbb{E}[\boldsymbol{\phi}_{n,m}(\beta)]$. Here and in the following we restrict to spin glasses for which the discussed quantities are well-defined and finite. In particular, we also assume that this holds for the annealed free entropy $\phi_{\mathrm{a},n,m}(\beta) = \frac{1}{n}\ln(\mathbb{E}[Z_{\boldsymbol{G}_{n,m}}(\beta)])$, as well as the limits $\phi_{\mathrm{q},d}(\beta) = \lim_{n\to\infty}\phi_{\mathrm{q},n,m}(\beta)$ and $\phi_{\mathrm{a},d}(\beta) = \lim_{n\to\infty}\phi_{\mathrm{a},n,m}(\beta)$.[2] Similar to the discussion in Section 1.4.2, for sufficiently high temperatures we expect the partition function to be sharply concentrated (we even have $Z_{\boldsymbol{G}_{n,m}}(0) = q^n$ almost surely), and in particular that $\phi_{\mathrm{q},d}(\beta) = \phi_{\mathrm{a},d}(\beta)$. The latter is very convenient since the annealed free entropy can be easily computed. For sufficiently small average degrees d, say no factors, the quenched and annealed averages even coincide for all temperatures. However, as before, we expect that this ceases to be true for larger densities, and that a phase transition occurs. For this purpose, as indicated in Section 1.4.2, we consider the teacher-student model $\boldsymbol{G}^*_{n,m}(\boldsymbol{\sigma}^*)$ using the established representation of the spin glass as a factor graph, and let $\phi^*_{\mathrm{q},d}(\beta) = \lim_{n\to\infty}\mathbb{E}[\frac{1}{n}\ln(Z_{\boldsymbol{G}^*_{n,m}(\boldsymbol{\sigma}^*)}(\beta))]$. For not too small d we expect the condensation threshold $\beta_{\mathrm{c}}(d) = \inf\{\beta \in \mathbb{R}_{\geq 0} : \phi_{\mathrm{a},d}(\beta) < \phi^*_{\mathrm{q},d}(\beta)\}$ to be in $\mathbb{R}_{>0}$, further $\phi_{\mathrm{q},d}(\beta) = \phi_{\mathrm{a},d}(\beta) = \phi^*_{\mathrm{q},d}(\beta)$ for $\beta \in [0, \beta_{\mathrm{c}}(d)]$ and $\phi_{\mathrm{q},d}(\beta) < \phi_{\mathrm{a},d}(\beta) < \phi^*_{\mathrm{q},d}(\beta)$ for $\beta > \beta_{\mathrm{c}}(d)$. For sufficiently well-behaved spin glasses, we expect the following picture. Consider the threshold $d_{\mathrm{c}} = \inf\{d \in \mathbb{R}_{\geq 0} : \beta_{\mathrm{c}}(d) < \infty\}$ at zero temperature. For $d \leq d_{\mathrm{c}}$ we expect to have $\beta_{\mathrm{c}}(d) = \infty$, and for $d > d_{\mathrm{c}}$ the curve $\beta_{\mathrm{c}}(d) \in \mathbb{R}_{>0}$ is continuous and non-increasing, thereby separating the replica-symmetric regime on the bottom-left from the condensation regime on the top-right. Further, next to the accessible annealed free entropy $\phi_{\mathrm{a},d}(\beta)$, also the limit $\phi^*_{\mathrm{q},d}(\beta)$ has been established for a large class of models [33, 30], which facilitates the computation of $\beta_{\mathrm{c}}(d)$. Usually for $\boldsymbol{e} \in \{0,1\}$ almost surely, we also consider the spin glass at zero temperature, i.e. $\beta = \infty$, which is a CSP. Using minor modifications[3], we may consider the free entropies $\phi_{\mathrm{q},d}(\infty)$, $\phi_{\mathrm{a},d}(\infty)$ and $\phi^*_{\mathrm{q},d}(\infty)$. For a large class of models, it has been established that these are the limits of the positive temperature free entropies and that the threshold d_{c} is the condensation threshold $d_{\mathrm{c}} = \inf\{d : \phi_{\mathrm{a},d}(\infty) < \phi_{\mathrm{q},d}(\infty)\}$[32]. So, next to the importance of the free entropy for spin glasses, where it is e.g. crucial for the analysis of the Boltzmann distribution and the identification of ground states, meaning minimum energy configurations, the free entropy at positive temperature can also be used as an approximation of the free entropy at zero temperature, i.e. a rough estimate for the size of the solution space.

1.4.5 Stochastic Block Model. Community detection in the stochastic block model (SBM) received considerable attention over the last decades [1, 97], in particular since the seminal paper [38]. The problem is stated as follows. For fixed weights $w = (w_0, w_1) \in \mathbb{R}^2_{\geq 0}$ and a fixed ground truth $\sigma \in [q]^n$, we choose a graph $\boldsymbol{G}^*_{n,w}(\sigma)$ on $n \geq \|w\|_\infty$ vertices by including each edge $\{u, v\}$ independently with probability $\frac{w_0}{n}$ if $\sigma_u = \sigma_v$ and probability $\frac{w_1}{n}$ otherwise. Now, a teacher samples a ground truth $\boldsymbol{\sigma}^*$ uniformly from $[q]^n$, then the graph $\boldsymbol{G}^*_{n,w}(\boldsymbol{\sigma}^*)$, and reveals the graph to a student who is tasked to recover (as much information as possible about) the ground truth from the graph (knowing the model parameters). This recovery problem is usually subdivided into the following types: exact

[1] The free energy $-\frac{1}{\beta}\boldsymbol{\phi}_{n,m}(\beta)$ is more common than the free entropy introduced in [85].

[2] Establishing merely the existence of the limiting quenched free entropy is already highly non-trivial [62, 53, 4] and still open in many cases.

[3] The free entropy may not be finite, so we replace $\frac{1}{n}\ln(Z)$ by related quantities like $\frac{1}{n}\ln(Z+1)$ or $Z^{1/n}$ [32], or avoid the expectation and consider $\boldsymbol{\phi}_{m,n}(\infty)$ directly [33].

recovery, almost exact recovery, partial recovery and weak recovery [1]. Another problem which is closely related to weak recovery, is the distinguishability problem. Here, the teacher flips a fair coin $\boldsymbol{c} \in \{0,1\}$ and then chooses a graph from $\boldsymbol{G}^*_{n,w}(\boldsymbol{\sigma}^*)$ for $\boldsymbol{c} = 1$, and the standard binomial graph $\boldsymbol{G}_{n,p}$ for $\boldsymbol{c} = 0$, where each edge is included with probability $p = \frac{w_0}{qn} + \frac{(q-1)w_1}{qn}$. Now, the student has to guess the value of $\boldsymbol{c}$ from the observed graph, with high probability.

This model can be (asymptotically) captured by the binomial factor graph from Section 1.4.1 with arity $k = 2$ over neighborhoods $\binom{[n]}{2}$, where the null model takes the role of the reference distribution $\boldsymbol{G}_{n,\pi}$ and the teacher-student model $\boldsymbol{G}^*_{n,\pi}(\boldsymbol{\sigma}^*)$ the role of the stochastic block model. Recall the discussion in Section 1.4.1. There, we argued that for sufficiently small average degrees d the planted model can be used to approximate the null model. Here, this means that the student cannot distinguish the models. To further deepen the connection using the CSP perspective, corresponding to $w_0 = 0$, for small average degrees the solution space covers almost all of $[q]^n$ in almost all graphs (in both models since they are similar), and any such solution might be the ground truth, equally likely, so we cannot extract any significant information about the ground truth. On the other hand, for large average degrees, as argued before, the null model and the planted model cease to be similar due to the fluctuation of the number of solutions. Further, the smaller solution spaces shatter into not too large (symmetric) clusters, which allows the student to make a guess that performs better than an entirely random guess. Thus, the replica-symmetric regime which is desirable for CSPs and spin glasses, is catastrophic in terms of statistical inference, since in this region both weak recovery and distinguishability are information-theoretically impossible. While this picture is still non-rigorous in general, the distinguishability threshold, the weak recovery threshold and their equivalence to the condensation threshold were rigorously established in [33, 30, 32] for the disassortative case, that is, for the case where $w_1 \geq w_0$.

1.4.6 Coding Theory. First, we focus on the following example from channel coding, namely low-density generator matrix (LDGM) encoded communication through a binary symmetric channel (BSC). A priori, we choose the generator matrix $\boldsymbol{M} \in \{0,1\}^{m\times n}$ by choosing k positions for the 1's uniformly at random from $[n]$, independently for each row. Then, guided by a reasoning similar to Section 1.2 and Section 1.4.2, we consider a stream of independent uniform input bits for the communication, which we partition into blocks of size n for the encoding process, a strategy commonly used for block codes. Let $\boldsymbol{x} \in \{0,1\}^n$ be one such input message, drawn uniformly at random. Encoding the message $\boldsymbol{x}$ amounts to matrix multiplication with $\boldsymbol{M}$, over the field $\mathbb{F}_2$, resulting in the codeword $\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} \in \{0,1\}^m$. Now, the encoded bits $\boldsymbol{y}$ are communicated through the noisy BSC, which means that each bit is flipped independently with probability $\varepsilon \in (0,1/2)$ (given $\boldsymbol{y}$), and results in the scrambled output $\boldsymbol{z} \in \{0,1\}^m$.

Clearly, the purpose of communication is to convey as much information as possible. Here, a widely used and reasonable measure is the conditional mutual information $I(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{M})$ of $\boldsymbol{x}$ and $\boldsymbol{z}$ given the generator matrix $\boldsymbol{M}$. Using the data processing inequality, the chain rule for the conditional mutual information and independence, we obtain the upper bound mc, where $c = 1 - H(\varepsilon)/H(1/2)$ is the capacity of the BSC and $H(\varepsilon) = -\varepsilon\ln(\varepsilon) - (1-\varepsilon)\ln(1-\varepsilon)$ is the entropy. In fact, this upper bound is tight if and only if $\boldsymbol{y} \in \{0,1\}^m$ (given $\boldsymbol{M}$) is uniform. This is indeed the case, whp over $\boldsymbol{M}$, for sequences $m = m(n) = o(\sqrt{n})$. Now, we fix an average degree $d \in \mathbb{R}_{\geq 0}$ and let $m = \lfloor dn/k \rfloor$, as before. In this sparse regime, the bound mc is out of reach due to the arising dependencies. However, we may still hope that the normalized mutual information $\frac{1}{m}I(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{M})$ converges to c for small d. Indeed, there does exist a threshold d_{c} such that $\lim_{n\to\infty}\frac{1}{m}I(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{M}) = c$ for $d \leq d_{\mathrm{c}}$, and $\lim_{n\to\infty}\frac{1}{m}I(\boldsymbol{x},\boldsymbol{z}|\boldsymbol{M}) < c$ for $d > d_{\mathrm{c}}$ [33].

Clearly, the choice of $\boldsymbol{M}$ is reminiscent of the choice of the neighborhoods (in $\binom{[n]}{k}$) for the null model, however, it is not immediate how this model relates to the discussed factor graph models. As it turns out, we can choose the weight $\boldsymbol{\psi}$ such that, with the other parameters unchanged, the mutual information $I(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{M}) = I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{n,m}(\boldsymbol{\sigma}^*))$ coincides with the mutual information of the ground truth and the teacher-student model. Details can e.g. be found in [33] or further below.

While this model is certainly oversimplified[4] from an application perspective, it does allow to focus on the highly non-trivial discussion of the limiting mutual information, without requiring the additional, highly non-trivial treatment of desirable degree sequences – thus, since the former is already involved enough as is, we circumvent the latter. The extension to general degree sequences can be found in [31].

Next to this somewhat indirect example from channel coding, planted models are also proposed as one-way functions in cryptography, and used as benchmark tests for SAT solvers. An example for the former is parity-majority in [32], an example of the latter and the ties to cryptography is discussed in [19]. Another prominent example is $k - \mathrm{LIN}(\eta)$ [33, 10, 12, 51], and both planted XORSAT and SAT models are thoroughly discussed in [51].

2 Main Results

Now, we turn to the new contributions established in this work, respectively in the underlying papers [102, 64, 103]. In Section 2.1, we extend the results in [33] for the uniform ground truth with applications in Section 1.4.4, Section 1.4.5 and Section 1.4.6 to the biased case, which manifests as external fields in spin glasses, as a community size bias in the (generalized) SBM, and as a biased distribution of the input bits for the LDGM-BAC pair, where the binary asymmetric channel (BAC) covers the BSC as a special case. The proofs and further discussion can be found in Section 3.

After this discussion of *general* Erdős-Rényi factor graphs with strictly *positive* weights, the positive temperature case, we turn to a very *specific* example of the *zero* temperature case in Section 2.2, namely perfect matchings in 3-uniform hypergraphs. To be specific, we pinpoint the location of the hitting time for a triangle factor in the standard graph process, i.e. the 2-uniform hypergraph, and relate it to the hitting time for a triangle cover. This result is not only of interest in its own right, the embedding of the model and the problem into the general framework also gives an interesting perspective. In particular, next to the practical requirements of the LDGM-BAC pair, also the embedding of this graph-theoretical model and problem into the general factor graph framework highlights the importance of further generalizations and the discussion of seemingly unintuitive distributions. The proofs and further discussion with respect to this result can be found in Section 4.

In Section 2.3, we leave the realm of Erdős-Rényi hypergraphs entirely and move to regular hypergraphs, where we extend the existing results for the perfect matching satisfiability threshold [36, 96], viewed as 1-in-k occupation problem, to the 2-in-k occupation problem. As opposed to the hitting time result, the embedding of this problem into the general framework is canonical and simultaneously covers two equivalent problems on hypergraphs. On the other hand, as opposed to the Erdős-Rényi hypergraph, the rigid structure of the regular hypergraph requires different arguments, that typically pave the way to the treatment of more complex prescribed degree distributions. Further details and the proofs for the occupation problems are postponed to Section 5.

[4] In the relevant sparse regime a linear number of columns in $\boldsymbol{M}$ is likely to be trivial, so a linear number of input bits are discarded in the encoding.

2.1 Mutual Information, Entropy and Condensation at Positive Temperature

The results in this section have been established in [102] and are based on the cavity method [86] from statistical mechanics, which is closely related to the powerful belief propagation [59, 107] and survey propagation [25, 83] algorithms as well as the Bethe free entropy [21, 126].

In Section 2.1.1, we motivate the results and present extensions of the applications in Section 1.4. Then, we extend the models from Section 1.4 and discuss the main results in Section 2.1.2.

2.1.1 Motivation and Applications. In Section 2.1.1.1, we briefly explain the extension of the unbiased models (without external fields) in Section 1.4 to biased models (with external fields). In Section 2.1.1.2 we propose a significant extension of the SBM from Section 1.4.5 as an application, and in Section 2.1.1.3 we present the LDGM-BAC pair as another application.

2.1.1.1 Biased Ground Truths. As indicated in Section 1.4, a common assumption in essentially all previous related works is that the ground truth $\boldsymbol{\sigma}^*$ is uniform. To build some intuition on why this choice is so convenient, we slightly extend the SBM from Section 1.4.5 for two colors as follows. Now, we choose $\boldsymbol{\sigma}^* = (\boldsymbol{\sigma}^*_i)_i \in \{0,1\}^n$ independently with success probability $p_{\mathrm{c}} \in (0,1)$. Then, edges within community '0' are chosen with probability p_0, within community '1' with p_1, and in between the two with p_{01}. As discussed in Section 1.4.5, we are interested in weak recovery, knowing $p_{\mathrm{c}}, p_0, p_1, p_{01}$. To this end, a simple idea is to compute the expected community sizes n_0, n_1 and the expected vertex degrees d_0, d_1. For $d_0 < d_1$ we may put the n_0 vertices with smaller degrees into community '0', and the remainder into the other. However, if the instance is *balanced*, meaning that $d_0 = d_1$, we might not be able to beat a completely random guess. Certainly, the most simple such case occurs when $p_0 = p_1$ and $p_{\mathrm{c}} = 1/2$, which is exactly the *symmetric* SBM on two colors from Section 1.4.5. However, the instance may be 'difficult', that is, balanced, even if $p_{\mathrm{c}} \neq 1/2$, and thus recovery may be highly non-trivial. On the other hand, the assumption that the two communities are symmetric seems overly restrictive. Similarly, for the LDGM codes from Section 1.4.6, the assumptions that the input bits are uniformly distributed and that the distortion of individual bits is symmetric certainly reflect the most relevant case, but understanding biased inputs and asymmetric distortions is clearly also highly desirable. Such symmetry assumptions are omnipresent and necessary in previous works; the absence of symmetry is not just a technical inconvenience, but introduces fundamentally new phenomena and requires a different analysis.

Our main contribution here is the development of a *general* and *readily applicable* theory for *sparse* and *not necessarily symmetric* factor graph models. Under certain assumptions, we study asymptotically several key observables that enable us to obtain a fine-grained understanding of important phase transitions in these models. Specifically, as in [33], we perform a thorough study and determine formulas for central quantities, like the mutual information, the quenched free entropy and the relative entropy. Additionally, we provide convergence guarantees that are polynomial in the system size, and uniformly so, under a variation of the model parameters. We further locate the condensation threshold and derive strong bounds for the quenched free entropy above the threshold. Including these, our contributions are as follows:

- We cover all reasonable, meaning balanced, pairs of model and ground truth that satisfy our assumptions (which are significantly weaker than in previous work).
- Both verifying and refuting one of the model assumptions is very hard, since it amounts to the solution of an infinite dimensional optimization problem. In the spirit of [105], we present a broader than ever, and in particular *explicit* class of models that satisfies this assumption. To our knowledge, this class covers *all models* from *all contributions* that worked with similar assumptions,

and many, many more. We provide a comprehensive list in Section 3.1.1.
- In previous work, limits have been established. We do not only establish limits, we give convergence guarantees, and the order of convergence is *polynomial.* Moreover, we provide error bounds that are *uniform* over all model parameters, that is, the expected degree, the ground truth and the weight distribution. As mentioned above, we further strengthen comparable results.

Before we proceed with the general exposition in the following section, we demonstrate the applicability of our results by discussing in more detail two fundamental specific models, whose treatment was out of reach with current methods. We keep the discussion of the applications as casual as possible to build intuition for the upcoming formal treatment. Details for the SBM (on hypergraphs) can be found in Section 3.5.9, and in Section 3.5.11 for the LDGM-BAC and more general channels.

2.1.1.2 The Stochastic Block Model. The symmetric SBM in Section 1.4.5 is a special case of the following general SBM, where we replace the weights w by a symmetric matrix $\psi \in \mathbb{R}_{\ge 0}^{q\times q}$ of weights. Further, fix a ground truth distribution $\gamma \in \mathcal{P}([q])$, where $\mathcal{P}([q])$ is the set of all probability measures over $[q]$. For a fixed parameter $d \in \mathbb{R}_{\ge 0}$, a given number $n \in \mathbb{Z}_{\ge d\|\psi\|_\infty}$ of vertices $[n]$ and a *ground truth* $\sigma \in [q]^n$ we obtain the graph $\boldsymbol{G}^*(\sigma)$ by including each edge $\{i,j\} \in \binom{[n]}{2}$ independently with probability $\frac{d}{n}\psi(\sigma_i,\sigma_j)$. The coordinates of the random ground truth $\boldsymbol{\sigma}^* \in [q]^n$ are identically independently distributed (iid), with distribution γ, which we denote by $\boldsymbol{\sigma}^* \sim \gamma^{\otimes n}$. The random graph $\boldsymbol{G}^*(\boldsymbol{\sigma}^*)$ is the *stochastic block model* (SBM).

We can easily check if (ψ,γ) is balanced, meaning, check if all vertices in $\boldsymbol{G}^*(\sigma)$ have the 'same' expected degree for likely σ. Assuming the model is balanced, we focus on two accessible types, the *assortative* and *disassortative* SBMs, defined as follows. First, we introduce a hierarchy on the colors $[q]$, to model how *similar* they are. A hierarchy with $L \in \mathbb{Z}_{>0}$ levels is given by partition refinements $C = (C_\ell)_\ell$, where $C_\ell : [q] \to [q]$ is the color class on level $0 \le \ell \le L$, such that all colors $\sigma \in [q]$ on level 0 are in the parent class $C_0(\sigma) = 1$, and colors on level L are separated, i.e. $C_L(\sigma) = \sigma$. Partition refinement means that for each level $\ell \in [L]$ and class $c \in C_{\ell-1}([q])$ on the lower level, the classes $\mathcal{C} = \{C_\ell(\sigma) : \sigma \in C_{\ell-1}^{-1}(c)\}$ on the higher level refine c, i.e. $\bigcup_{c'\in\mathcal{C}} C_\ell^{-1}(c') = C_{\ell-1}^{-1}(c)$. Now, we assign a weight $w(\ell,c) \in \mathbb{R}$ to each class $c \in [q]$ on each level ℓ, and let

$$\psi(\sigma) = \sum_{\ell=0}^{L} \mathbb{1}\{C_\ell(\sigma_1) = C_\ell(\sigma_2)\} w(\ell, C_\ell(\sigma_1)), \quad \sigma \in [q]^2.$$

This is well-defined if $\psi \ge 0$ (componentwise), and then ψ defines a (hierarchical) SBM. The SBM is assortative if $w(\ell,c) \ge 0$ for all ℓ and $c \in [q]$, and disassortative if $w(\ell,c) \le 0$ for all $\ell \ge 1$ and $c \in [q]$. For both types, it's not only easy to *check if* (ψ,γ) is balanced, it is easy to exactly determine all balanced (ψ,γ) (thanks to nice convexity properties).

The highly non-trivial study of these models received an enormous amount of attention in the last decade. Driven by the conjectures in the seminal paper [38], a long line [84, 99, 4, 16, 7, 8, 9] of research iteratively improved the results. A major success was the localization of the thresholds for the *binary* balanced symmetric two-part SBM in [100], followed by the celebrated localization of the weak recovery threshold for the balanced symmetric disassortative two-part SBM in [33], for *all* $q \in \mathbb{Z}_{>1}$, and the distinguishability threshold in [30]. Both thresholds coincide with the *condensation threshold* d_{cond}, and the crucial step in [33] was to establish d_{cond} using the limit of the mutual information $\frac{1}{n}I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$ of the SBM and the ground truth, a measure of how much information on $\boldsymbol{\sigma}^*$ is captured by $\boldsymbol{G}^*(\boldsymbol{\sigma}^*)$.

Regarding the *general* balanced disassortative SBM, where the number of parts is almost arbitrary, and the ground truth γ is not the uniform distribution, much less is known. This general case can be readily treated with our results, and we provide a detailed picture. In particular, we establish the limit of $\frac{1}{n}I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$ of the mutual information and the condensation threshold $d_{\mathrm{cond}}(\psi)$. We also obtain the limit of the relative entropy $\frac{1}{n}D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)\|\boldsymbol{\sigma}_{\boldsymbol{G}}, \boldsymbol{G})$, where $\boldsymbol{\sigma}_G$ is the posterior given a graph G (our guess what σ is, given G) and $\boldsymbol{G}$ is the corresponding null model, that is the binomial graph with the edge probability chosen such that the average degrees in both models coincide. Finally, we show that the convergence to the limits is polynomially fast, and uniformly over d, γ, ψ.

2.1.1.3 LDGM Codes. We extend the LDGM-BSC pair from Section 1.4.6 as follows. The input message is $\boldsymbol{x} \sim \gamma^{\otimes n}$ for some distribution $\gamma \in \mathcal{P}(\{0,1\})$ and $n \in \mathbb{Z}_{>0}$; we say that the input is binary and memoryless. The parameters for the code are the arity $k \in \mathbb{Z}_{\geq 2}$ and the block length $m \in \mathbb{Z}_{\geq k}$, and the parameters for the noisy channel are the error probabilities $0 < \delta \leq \varepsilon \leq \frac{1}{2}$. Given the parameters, we randomly choose a generator matrix $\boldsymbol{M} \in \{0,1\}^{m\times n}$ by choosing the positions for the 1's in each row independently and uniformly from $\binom{[n]}{k}$ as in Section 1.4.6. Also, as before, we obtain the codeword $\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} \in \{0,1\}^m$ using matrix multiplication over $\mathbb{F}_2$. Finally, we transmit $\boldsymbol{y}$ through the BAC, that is, given $\boldsymbol{y}$, each bit is transmitted independently, flipped with probability δ if it is 0, and otherwise flipped with probability ε. Let $\boldsymbol{z} \in \{0,1\}^m$ be the produced output message.

Similar to Section 1.4.6, for sequences $m = m(n) = o(\sqrt{n})$ and whp over $\boldsymbol{M}$, the bits of the codeword $\boldsymbol{y} = (\boldsymbol{y}_a)_{a\in[m]}$ are conditionally iid given $\boldsymbol{M}$. So, if we let $\boldsymbol{b} = (\boldsymbol{b}_h)_h \in \{0,1\}^k$ be iid with law γ, further let the sum $\boldsymbol{y}_\circ = \sum_{h=1}^k \boldsymbol{b}_h \in \{0,1\}$ over $\mathbb{F}_2$ be the input to the BAC with output $\boldsymbol{z}_\circ$, then whp the law of $(\boldsymbol{y}_a, \boldsymbol{z}_a)_a$ given $\boldsymbol{M}$ is $(\boldsymbol{y}_\circ, \boldsymbol{z}_\circ)^{\otimes m}$. Hence, in this case, the mutual information of $\boldsymbol{x}$ and $\boldsymbol{z}$ given $\boldsymbol{M}$ is exactly $mI(\boldsymbol{y}_\circ, \boldsymbol{z}_\circ)$. Since it's very likely that $\boldsymbol{M}$ is of this form, this also holds for the expectation of the mutual information of $\boldsymbol{x}$ and $\boldsymbol{z}$ given $\boldsymbol{M}$, which is nothing but the conditional mutual information $I(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{M})$.

This reasoning naturally raises the question: Up to which point can we expect this to be true, i.e. up to what block length m are the dependencies in the generation of $\boldsymbol{y}$ so weak that we do have $\frac{1}{m}I(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{M}) \approx I(\boldsymbol{y}_\circ, \boldsymbol{z}_\circ)$? Phrased asymptotically, and in terms of the rate $\frac{n}{m}$ as is common, the question is: What is the value of $R^*(\gamma) = \inf\{R : \lim_{n\to\infty} \frac{1}{m}I(\boldsymbol{x}, \boldsymbol{z}|\boldsymbol{M}) = I(\boldsymbol{y}_\circ, \boldsymbol{z}_\circ)\}$, where $m = m(R, n) > 0$ is such that $\lim_{n\to\infty} \frac{n}{m} = R$?

Motivated by a conjecture in [80] and based on [91], where bounds and concentration were established, this question was studied for uniform γ and the binary symmetric channel, i.e. $\delta = \varepsilon$, in [4], where the existence of the limit was established, and in [33], where the conjecture was settled. The mutual information limit was explicitly determined and it was shown that $R^*(\gamma)$ is the condensation threshold. Crucially, these contributions paved the way for follow-up work [[31] that extends the results to more general degree distributions, since the Poisson ensemble is not suitable for encoding due to the non-vanishing fraction of trivial columns in $\boldsymbol{M}$.

Our results directly yield the mutual information limit and $R^*(\gamma)$ for any γ, be it uniform or not. More importantly, not only for the binary symmetric channel, i.e. for $\delta = \varepsilon$, but for any binary asymmetric channel, i.e. for any $\delta \leq \varepsilon$, for any choice of γ, and for even k the mutual information limit and $R^*(\gamma)$ are also directly implied. A noteworthy special case in the above is when $I(\boldsymbol{y}_\circ, \boldsymbol{z}_\circ)$ is the channel capacity of the BAC. In particular, this is the case for uniform γ and $\delta = \varepsilon$, since the XOR of k uniform bits is uniform, which is the capacity achieving distribution for the BAC with $\delta = \varepsilon$. For the BAC with $\delta < \varepsilon$ and even k we have explicitly computed the two possible choices for the distribution of the input bits such that the normalized mutual information reaches capacity for sufficiently large rates, and thereby completed the picture for the general case.

2.1.2 Free Entropies, Divergence and the Mutual Information. We turn to the main results. First, we extend and adjust the notions from Section 1.4, then we discuss the required assumptions in general and for a more accessible class of models. Subsequently, we present the results for the planted model quenched free entropy, the relative entropy, the condensation threshold and the mutual information. Related results in the literature, extensions of the results in this section, their proofs and their connection to the applications in Section 2.1.1 can be found in Section 3.

2.1.2.1 Factor Graphs. In this section, we introduce an adjusted version of the factor graphs in Section 1.3 and Section 1.4.1. In particular, we manipulate the partition function and the Gibbs measure to take the bias into account. As before, fix a number $q \in \mathbb{Z}_{\geq 1}$ of colors and a factor degree $k \in \mathbb{Z}_{\geq 1}$ throughout. Further, we fix a small lower bound $\psi_\downarrow \in (0,1)$ for the weights and let $\psi_\uparrow = \psi_\downarrow^{-1}$ be the upper bound. We also fix a large upper bound $d_\uparrow \in \mathbb{R}_{>0}$ for the expected variable degree.

Now, for a given distribution $\gamma^* \in \mathcal{P}([q])$, a number $n \in \mathbb{Z}_{>0}$ of variables and a number $m \in \mathbb{Z}_{\geq 0}$ of factors, a factor graph $G = (v_a, \psi_a)_{a\in[m]}$ is given by the ordered neighborhoods $v_a \in [n]^k$ and the weight functions $\psi_a : [q]^k \to [\psi_\downarrow, \psi_\uparrow]$. For an assignment $\sigma \in [q]^n$ and $v \in [n]^k$ we use the shorthand $\sigma_v = (\sigma(v_h))_{h\in[k]} \in [q]^k$. The weight $\psi_G(\sigma)$ of σ and the Gibbs measure $\mu_{\gamma^*,G}(\sigma)$ are given by

$$\psi_G(\sigma) = \prod_{a\in[m]} \psi_a(\sigma_{v_a}), \quad \mu_{\gamma^*,G}(\sigma) = \frac{\gamma^{*\otimes n}(\sigma)\psi_G(\sigma)}{Z_{\gamma^*}(G)}, \quad Z_{\gamma^*}(G) = \sum_{\sigma\in[q]^n} \gamma^{*\otimes n}(\sigma)\psi_G(\sigma). \tag{1}$$

Further, let $\phi_{\gamma^*}(G) = \frac{1}{n}\ln(Z_{\gamma^*}(G))$ be the free entropy. For the random factor graphs, fix a random weight $\boldsymbol{\psi} : [q]^k \to [\psi_\downarrow, \psi_\uparrow]$ with distribution p. Then the null model is the random factor graph $\boldsymbol{G}_{n,m,p} \sim (\mathrm{u}([n]^k) \otimes p)^{\otimes m}$, where $\mathrm{u}([n]^k)$ is the uniform distribution, as before. The teacher-student model $\boldsymbol{G}^*_{n,m,p}(\sigma)$ for a fixed ground truth $\sigma \in [q]^n$, is still given by the Radon-Nikodym (RN) derivative $G \mapsto \psi_G(\sigma)/\mathbb{E}[\psi_{\boldsymbol{G}_{n,m,p}}(\sigma)]$ with respect to $\boldsymbol{G}_{n,m,p}$.

As opposed to the previous models and applications, we work with random factor counts for convenience in the following. Thus, for given $d \in [0, d_\uparrow]$ we let $(\boldsymbol{\sigma}^*_{\gamma^*,n}, \boldsymbol{m}_{d,n}) \sim \gamma^{*\otimes n} \otimes \mathrm{Po}(dn/k)$ be the ground truth and factor count pair, where $\mathrm{Po}(\lambda)$ is the Poisson distribution. Random factor counts other than $\boldsymbol{m}$ are discussed in Section 3.1.3. In the following we suppress dependencies, so e.g. we abbreviate $\boldsymbol{G} = \boldsymbol{G}_{n,m,p}$, $\boldsymbol{G}_{\boldsymbol{m}} = \boldsymbol{G}_{n,\boldsymbol{m},p}$, $\boldsymbol{G}^*(\sigma) = \boldsymbol{G}^*_{n,m,p}(\sigma)$ and $\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*) = \boldsymbol{G}^*_{n,\boldsymbol{m},p}(\boldsymbol{\sigma}^*)$.

2.1.2.2 The Assumptions. Next to the assumption $\psi_\downarrow \leq \boldsymbol{\psi} \leq \psi_\uparrow$, we assume that the following two properties hold for the pair (p, γ^*). Let $\mathcal{P}^2([q]) = \mathcal{P}(\mathcal{P}([q]))$ be the set of probability distributions over the simplex $\mathcal{P}([q]) \subseteq \mathbb{R}^q$, and $\mathcal{P}^2_{*,\gamma^*}([q]) = \{\pi \in \mathcal{P}^2([q]) : \mathbb{E}[\boldsymbol{\gamma}_\pi] = \gamma^*\}$, where $\boldsymbol{\gamma}_\pi \sim \pi$.

BAL: For $\overline{Z}_{\mathrm{f}} : \mathcal{P}([q]) \to \mathbb{R}_{>0}$, $\gamma \mapsto \sum_\sigma \mathbb{E}[\boldsymbol{\psi}(\sigma)] \prod_{h=1}^k \gamma(\sigma_h)$, and $\xi_p = \|\overline{Z}_{\mathrm{f}}\|_\infty$ we have $\overline{Z}_{\mathrm{f}}(\gamma^*) = \xi_p$.

POS: For all $\pi_1, \pi_2 \in \mathcal{P}^2_{*,\gamma^*}([q])$ and using $\Lambda(x) = x\ln(x)$, $(\boldsymbol{\psi}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \sim p \otimes \pi_1^{\otimes k} \otimes \pi_2^{\otimes k}$, we have

$$\mathbb{E}\left[\Lambda(Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_1)) + (k-1)\Lambda(Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_2)) - \sum_{h=1}^k \Lambda(Z_{\mathrm{fm}}(\boldsymbol{\psi}, h, \boldsymbol{\gamma}))\right] \geq 0,$$

where

$$Z_{\mathrm{f}}(\psi, \gamma) = \sum_\sigma \psi(\sigma) \prod_{h=1}^k \gamma_h(\sigma_h), \quad Z_{\mathrm{fm}}(\psi, h, \gamma) = \sum_\sigma \psi(\sigma)\gamma_{1,h}(\sigma_h) \prod_{h'\neq h} \gamma_{2,h'}(\sigma_{h'}).$$

The assumption **BAL** does not only ensure that the model is balanced, it also requires that the expected weight $\mathbb{E}[\psi_{\boldsymbol{G}}(\sigma)]$ is maximized by assignments σ with color frequencies close to γ^*. The assumption **POS** originated as a convexity assumption [4] that ensured subadditivity of the planted model quenched free entropy. Notice that **BAL** requires the solution $\xi = \|\overline{Z}_{\mathrm{f}}\|_\infty$ of a $(q-1)$-dimensional optimization problem. Depending on the problem, this may already be intractable, but **POS** requires to solve an optimization problem over $\mathcal{P}^2_{*,\gamma^*}([q])$. Thus, in order to facilitate the application of the main results, we introduce the following natural and explicit classes of random weights, that turn out to satisfy **POS**.

1. Let $\boldsymbol{a} \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{b}, \boldsymbol{f}_{h,i}(\sigma) \in \mathbb{R}$ for $\sigma \in [q]$, $i \in \mathbb{Z}_{>0}$ and $h \in [k]$. Let $\sum_\sigma \sum_i |\prod_h \boldsymbol{f}_{h,i}(\sigma_h)| < \infty$ almost surely, $\boldsymbol{\Delta}(\sigma) = \sum_\sigma \sum_i \prod_h \boldsymbol{f}_{h,i}(\sigma_h)$ and $|\boldsymbol{b}|\|\boldsymbol{\Delta}\|_\infty \leq 1$ almost surely. Let $(\boldsymbol{a}, \boldsymbol{b})$ and $(\boldsymbol{f}_{h,i}(\sigma))_{h,i,\sigma}$ be independent, and let the k sequences $(\boldsymbol{f}_{1,i}(\sigma))_{i,\sigma}, \dots, (\boldsymbol{f}_{k,i}(\sigma))_{i,\sigma}$ be iid. Let $\mathbb{E}[|\boldsymbol{a}\boldsymbol{b}|^\ell], \mathbb{E}[\|\boldsymbol{\Delta}\|_\infty^\ell] < \infty$ and $\mathbb{E}[\boldsymbol{a}\boldsymbol{b}^\ell] \geq 0$ for all $\ell \in 2\mathbb{Z}_{>0}+1$. Assume that $\boldsymbol{\psi}(\sigma) = \boldsymbol{a}(1 - \boldsymbol{b}\boldsymbol{\Delta}(\sigma))$.
2. Assume that $\boldsymbol{\psi}$ is of Type 1, that $\boldsymbol{f}_{h,i}(\sigma) = 0$ for $i > 1$, $\sigma \in [q]$, and $\mathbb{E}[\boldsymbol{a}\boldsymbol{b}^\ell] = 0$ for $\ell \in 2\mathbb{Z}_{>0}+1$.
3. Assume that $\boldsymbol{\psi}$ is of Type 1 and that $\boldsymbol{f}_{h,i}(\sigma) \geq 0$ for h, i and σ.
4. Assume that $\boldsymbol{\psi}(\sigma) = \prod_h \boldsymbol{f}_h(\sigma_h)$ for $\boldsymbol{f} \in (\mathbb{R}^q_{\geq 0})^k$, possibly dependent.

For even k, let $\mathcal{P}$ be the union of Type 1 and 4. For odd k, let $\mathcal{P}$ be the union of Type 2, 3 and 4. Finally, let $\mathfrak{A}$ be the set of all pairs (p, γ^*) that satisfy **BAL** and **POS**.

Proposition 2.1. *We have* $\{(p, \gamma^*) \in \mathcal{P} \times \mathcal{P}([q]) : \overline{Z}_{\mathrm{f}}(\gamma^*) = \xi_p\} \subseteq \mathfrak{A}$.

We will prove Proposition 2.1 under weaker assumptions. We will further show that the set $\mathfrak{A}$ is closed with respect to dozens of operations, and that it is convex (cf. Section 3.5.5).

2.1.2.3 Uniform Convergence. In the following, all bounds only depend on $\mathfrak{g} = (q, k, \psi_\downarrow, d_\uparrow)$, i.e. they are uniform in p, γ^* and d. For this purpose let $f(n) = \mathcal{O}_{\mathrm{u}}(g(n))$ if there exists $n_\circ(\mathfrak{g}) \in \mathbb{Z}_{>0}$ and $c(\mathfrak{g}) \in \mathbb{R}_{>0}$ such that $|f(n)| \leq cg(n)$ for $n \geq n_\circ$.

2.1.2.4 The Quenched Free Entropy of the Planted Ensemble. Our first main result yields the limit of the teacher-student model quenched free entropy. Recall Λ, Z_{f}, $\mathcal{P}^2([q])$, $\mathcal{P}^2_{*,\gamma^*}([q])$ from the assumption **POS**. For $d' \in \mathbb{Z}_{\geq 0}$ and $(\psi, h, \gamma) \in (\mathbb{R}_{>0}^{[q]^k} \times [k] \times \mathcal{P}([q])^k)^{\mathbb{Z}_{>0}}$ let

$$Z_{\mathrm{v},\gamma^*}(d', \psi, h, \gamma) = \sum_{\sigma \in [q]} \gamma^*(\sigma) \prod_{a \in [d']} \left(\sum_{\tau \in [q]^k} \mathbb{1}\{\tau(h_a) = \sigma\} \psi(\tau) \prod_{h' \neq h_a} \gamma_{a,h'}(\tau(h')) \right).$$

Let $\pi \in \mathcal{P}^2([q])$, $(\boldsymbol{d}, \boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}) \sim \mathrm{Po}(d) \otimes (p \otimes \mathrm{u}([k]) \otimes \pi^{\otimes k})^{\otimes \mathbb{Z}_{>0}}$, $(\boldsymbol{\psi}_\circ, \boldsymbol{\gamma}_\circ) \sim p \otimes \pi^{\otimes k}$, further let

$$B_{p,\gamma^*,d}(\pi) = \mathbb{E}\left[\frac{1}{\xi_p^{\boldsymbol{d}}} \Lambda\left(Z_{\mathrm{v},\gamma^*}(\boldsymbol{d}, \boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma})\right)\right] - \frac{d(k-1)}{k\xi_p} \mathbb{E}\left[\Lambda\left(Z_{\mathrm{f}}(\boldsymbol{\psi}_\circ, \boldsymbol{\gamma}_\circ)\right)\right]$$

be the (limiting) Bethe free entropy (for $\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*)$), and $B_{\uparrow,p,\gamma^*}(d) = \sup_{\pi \in \mathcal{P}^2_{*,\gamma^*}([q])} B_{p,\gamma^*,d}(\pi)$.

Theorem 2.2. *There exists* $\rho(\mathfrak{g}) \in \mathbb{R}_{>0}$ *such that for* $(p, \gamma^*) \in \mathfrak{A}$ *we have*

$$\mathbb{E}\left[\phi_{\gamma^*}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))\right] = \mathbb{E}\left[n^{-1} \ln(Z_{\gamma^*}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*)))\right] = B_\uparrow(d) + \mathcal{O}_{\mathrm{u}}(n^{-\rho}).$$

This result establishes that the Bethe functional is the limit of the planted model quenched free entropy density and thereby rigorously establishes the predictions from physics using the cavity method.

2.1.2.5 The Information-Theoretic Threshold. The second main result addresses the relative entropy of $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))$ with respect to $(\boldsymbol{\sigma}_{\gamma^*,\boldsymbol{G}_{\boldsymbol{m}}}, \boldsymbol{G}_{\boldsymbol{m}})$, where $\boldsymbol{\sigma}_{\gamma^*,G} \sim \mu_{\gamma^*,G}$ are the Gibbs spins from Equation (1). If $\boldsymbol{a}$ has a RN derivative r with respect to $\boldsymbol{b}$, let $D_{\mathrm{KL}}(\boldsymbol{a}\|\boldsymbol{b}) = \mathbb{E}[\ln(r(\boldsymbol{a}))]$ and $D_{\mathrm{KL}}(\boldsymbol{a}\|\boldsymbol{b}) = \infty$ otherwise. Further, let $\phi_{\mathrm{a}}(d) = \phi_{\mathrm{a},p}(d) = \frac{d}{k}\ln(\xi_p)$.

Theorem 2.3. *With ρ from Theorem 2.2 and for $(p,\gamma^*) \in \mathfrak{A}$ we have*

$$\frac{1}{n} D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*) \| \boldsymbol{\sigma}_{\gamma^*,\boldsymbol{G}_{\boldsymbol{m}}}, \boldsymbol{G}_{\boldsymbol{m}}) = B_{\uparrow}(d) - \phi_{\mathrm{a}}(d) + \mathcal{O}_{\mathrm{u}}(n^{-\rho}).$$

Intuitively, Theorem 2.3 states that the teacher-student and the null model are similar in the replica symmetric regime $\mathfrak{P}_{\mathrm{r}} = \{(p,\gamma^*,d) \in \mathfrak{P} : B_{\uparrow}(d) = \phi_{\mathrm{a}}(d)\}$, $\mathfrak{P} = \mathfrak{A} \times [0, d_{\uparrow}]$, while they are significantly distinct in the condensation regime $\mathfrak{P}_{\mathrm{c}} = \mathfrak{P} \setminus \mathfrak{P}_{\mathrm{r}} = \{(p,\gamma^*,d) \in \mathfrak{P} : B_{\uparrow}(d) > \phi_{\mathrm{a}}(d)\}$. Hence this result establishes that for some rough approximations we can sample from $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))$ instead of $(\boldsymbol{\sigma}_{\gamma^*,\boldsymbol{G}_{\boldsymbol{m}}}, \boldsymbol{G}_{\boldsymbol{m}})$ (cf. Section 1.4.2 and Section 1.4.5).

2.1.2.6 The Condensation Threshold. We confirmed that the replica symmetric and the condensation regime govern the behavior of the relative entropy. Next, we ensure that the quenched free entropy for the null model indeed behaves as expected. For this purpose let

$$\phi_{\mathrm{q}\uparrow,p,\gamma^*}(d) = \limsup_{n\to\infty} \mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}_{\boldsymbol{m}})], \quad \phi_{\mathrm{q}\downarrow,p,\gamma^*}(d) = \liminf_{n\to\infty} \mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}_{\boldsymbol{m}})].$$

Further, let $d_{\mathrm{cond}}(p,\gamma^*) = \inf\{d \in \mathbb{R}_{>0} : B_{\uparrow}(d) - \phi_{\mathrm{a}}(d) > 0\}$ be the condensation threshold.

Theorem 2.4. *Recall ρ from Theorem 2.2.*
a) We have $\mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}_{\boldsymbol{m}})] = \phi_{\mathrm{a}}(d) + \mathcal{O}_{\mathrm{u}}(n^{-\rho/2})$ for $(p,\gamma^*,d) \in \mathfrak{P}_{\mathrm{r}}$.
b) There exists $c(\mathfrak{g}) \in \mathbb{R}_{>0}$ such that for $(p,\gamma^*,d) \in \mathfrak{P}$ we have

$$\phi_{\mathrm{a}}(d) - \phi_{\mathrm{q}\uparrow}(d) \geq c \sup_{d' \in [0,d]} (B_{\uparrow}(d') - \phi_{\mathrm{q}\downarrow}(d'))^2.$$

c) We have $\mathfrak{P}_{\mathrm{r}} = [0, d_{\mathrm{cond}}]$ and $\mathfrak{P}_{\mathrm{c}} = (d_{\mathrm{cond},\infty})$.

Theorem 2.4b) allows to establish upper bounds for $\phi_{\mathrm{q}\uparrow}(d)$, the simplest by considering $d' = d$ and solving the quadratic inequality, i.e. (since c is such that $\delta^*(d) \leq 1/(4c)$)

$$\phi_{\mathrm{a}}(d) - \phi_{\mathrm{q}\uparrow}(d) \geq \tilde{\delta}(d) - \sqrt{\tilde{\delta}(d)^2 - \delta^*(d)^2}, \quad \tilde{\delta}(d) = \frac{1}{2c} - \delta^*(d), \quad \delta^*(d) = B_{\uparrow}(d) - \phi_{\mathrm{a}}(d).$$

Again, this result rigorously establishes the behavior predicted by the cavity method, and discussed in Section 1.4.4. Theorem 2.3 provides a finite size interpretation of the condensation threshold, and the combination of these results establishes the asymptotic relationship of the planted model quenched free entropy, the annealed free entropy, the relative entropy, and the null model quenched free entropy (below the threshold).

2.1.2.7 The Mutual Information. We turn to the fourth and last main result, regarding the limit of the mutual information. In general the mutual information is given by $I(\boldsymbol{a},\boldsymbol{b}) = D_{\mathrm{KL}}(\boldsymbol{a},\boldsymbol{b}\|\boldsymbol{a} \otimes \boldsymbol{b})$.

Theorem 2.5. *With ρ from Theorem 2.2 and for $(p,\gamma^*,d) \in \mathfrak{P}$ we have*

$$\frac{1}{n} I\left(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*)\right) = \frac{d}{k\xi_p} \mathbb{E}\left[\Lambda\left(\boldsymbol{\psi}(\boldsymbol{\sigma})\right)\right] - B_{\uparrow}(d) + \mathcal{O}(n^{-\rho}), \quad (\boldsymbol{\psi}, \boldsymbol{\sigma}) \sim p \otimes \gamma^{*\otimes k}.$$

This result establishes the asymptotic relationship of the mutual information with the other key quantities. Moreover, the phase transition for $B_\uparrow(d)$ at d_{cond} corresponds to a phase transition for the mutual information, since we can replace $B_\uparrow$ by ϕ_{a} below d_{cond}. Regarding the applications, we obtain the desired limiting mutual information for the LDGM codes (cf. Section 1.4.6 and Section 2.1.1.3), and the expected phase transition occurs exactly at d_{cond}. Further, this mutual information is also a key quantity in the analysis of SBMs (cf. Section 1.4.5 and Section 2.1.1.2).

2.2 Triangle Factors in the Graph Process

The main result in this section has been established in [64] and is based on the hitting time results for hypergraphs in [71] as well as the translation of the binomial hypergraph threshold to the binomial graph via the ingenious coupling in [113, 63]. Instead of discussing these results in full generality, we focus on triangles in simple graphs, the most demanding case, and only briefly sketch the extension to larger cliques and cliques in hypergraphs.

In Section 2.2.1 we discuss the satisfiability thresholds for hyperedge covers and perfect matchings in the binomial hypergraph, as well as the hitting time version [49, 120, 23, 71]. In Section 2.2.2 we discuss the satisfiability thresholds for k-clique covers and k-clique factors in the binomial graph [68, 113, 63]. Then we turn to the main result, namely the hitting time version, in Section 2.2.3, and in Section 2.2.4 we briefly discuss how the problems can be modeled as random CSPs.

2.2.1 Covers and Matchings. Recently, Jeff Kahn [71] answered a question by Schmidt and Shamir [50, 120] – a well-studied problem that has been open for four decades – and in particular showed that the threshold for the existence of a perfect matching coincides with the threshold for the existence of a hyperedge cover. To be precise, let $k \in \mathbb{Z}_{\geq 2}$ and $\pi \in [0, 1]$. As in Section 1.2, Section 1.4.1, Section 1.4.5 and Section 2.1.1.2 we consider the binomial k-uniform hypergraph $\boldsymbol{H}_k(n, \pi)$ with vertices $[n]$ where each hyperedge is included independently with probability π. A hyperedge cover $\mathcal{C}$ of $\boldsymbol{H}_k(n, p) = ([n], \boldsymbol{\mathcal{H}})$ is a subset $\mathcal{C} \subseteq \boldsymbol{\mathcal{H}}$ of the hyperedges such that $\bigcup_{E \in \mathcal{C}} E = [n]$, and a perfect matching, or 1-factor, is a hyperedge cover composed of vertex disjoint hyperedges. Let

$$\pi_\pm = \frac{\ln(n) \pm g(n)}{\binom{n-1}{k-1}},$$

where $g : \mathbb{Z}_{>0} \to \mathbb{R}_{\geq 0}$ is any sufficiently slowly increasing function with $\lim_{n\to} g(n) = \infty$. Erdős and Rényi [49] established for the special case $k = 2$ that there are no hyperedge covers (and hence no perfect matchings) in $\boldsymbol{H}_k(n, \pi_-)$ whp, and on the other hand that there exists a perfect matching (and hence a cover) in $\boldsymbol{H}_k(n, \pi_+)$ whp. Twenty years later, Schmidt and Shamir [50, 120] asked if a similar bound exists for $k \geq 3$. Forty years later and after extensive research, Jeff Kahn [71] finally extended the result by Erdős and Rényi for $k = 2$ to arbitrary k-uniform hypergraphs.

However, Jeff Kahn established a significantly stronger and more instructive result, namely that the hitting times for the existence of hyperedge covers and perfect matchings in the k-uniform hypergraph process coincide whp, which yields the answer to the question by Schmidt and Shamir as an immediate corollary. To be specific, let $\boldsymbol{H}_{k,n} = (\boldsymbol{H}_{k,n}(S))_S$ be the standard k-uniform hypergraph process, where $\boldsymbol{H}_{k,n}(0) = ([n], \emptyset)$ is empty and we obtain $\boldsymbol{H}_{k,n}(S+1)$ from $\boldsymbol{H}_{k,n}(S)$ by adding one hyperedge, chosen uniformly at random from all hyperedges that are not present in $\boldsymbol{H}_{k,n}(S)$. Let $\boldsymbol{S}_{\mathrm{c}}$ be the minimum step $S \in [\binom{n}{k}]$ such that a hyperedge cover exists in $\boldsymbol{H}_{k,n}(S)$, and let $\boldsymbol{S}_{\mathrm{f}}$ be the minimum step $S \in [\binom{n}{k}]$ such that a perfect matching exists in $\boldsymbol{H}_{k,n}(S)$. Further, let $(\boldsymbol{S}_-, \boldsymbol{S}_+)$ be independent of $\boldsymbol{H}_{k,n}$, where $\boldsymbol{S}_+$ is binomial with parameters $\binom{n}{k}$ and π_+, and $\boldsymbol{S}_-$ given $\boldsymbol{S}_+$ is binomial with parameters $\boldsymbol{S}_+$ and π_-/π_+

(for sufficiently large n). Bollobás and Thomason [23] established for $k=2$ that $\boldsymbol{S}_- \le \boldsymbol{S}_{\mathrm{c}} = \boldsymbol{S}_{\mathrm{f}} \le \boldsymbol{S}_+$ whp, i.e. they showed that the appearance of an edge cover coincides with the appearance of a perfect matching whp. Jeff Kahn [71] extended this result to $k \ge 3$ (using the bounds from [39]).

Theorem 2.6. *For $k \in \mathbb{Z}_{\ge 2}$ we have $\boldsymbol{S}_- \le \boldsymbol{S}_{\mathrm{c}} \le \boldsymbol{S}_+$ whp, and $\boldsymbol{S}_{\mathrm{f}} = \boldsymbol{S}_{\mathrm{c}}$ whp over $n \in k\mathbb{Z}_{>0}$.*

Notice that $\boldsymbol{S}_{\mathrm{f}}$ is finite if and only if $n \in k\mathbb{Z}_{>0}$ since the size of a perfect matching has to be n/k.

2.2.2 Covers and Factors. Oliver Riordan [113] and Annika Heckel [63] devised ingenious couplings to derive k-clique factor thresholds for the binomial graph from Theorem 2.6. To be precise, let $\boldsymbol{G}(n,p) = \boldsymbol{H}_2(n,p)$ for $p \in [0,1]$ be the binomial graph. For $k \in \mathbb{Z}_{\ge 2}$, recall that a clique E of size k, or k-clique, in $\boldsymbol{G}(n,p) = ([n], \boldsymbol{\mathcal{G}})$ is a set $E \in \binom{[n]}{k}$ such that $\binom{E}{2} \subseteq \boldsymbol{\mathcal{G}}$. A k-clique cover $\mathcal{C}$ of $\boldsymbol{G}(n,p)$ is a set $\mathcal{C} \subseteq \binom{[n]}{k}$ of k-cliques in $\boldsymbol{G}(n,p)$ such that $\bigcup_{E \in \mathcal{C}} E = [n]$, and a k-clique factor is a k-clique cover composed of vertex disjoint cliques. Using g from Section 2.2.1, consider the critical window

$$p_\pm = \frac{\ln(n)^{1/\binom{k}{2}} \pm g(n)}{\binom{n-1}{k-1}}.$$

It is well-known [68] that whp there is no k-clique cover in $\boldsymbol{G}(n,p_-)$, and that whp there is a k-clique cover in $\boldsymbol{G}(n,p_+)$. For the k-clique factor threshold, notice that the case $k=2$ follows from Theorem 2.6 since these are perfect matchings. The k-clique factor threshold for $k \ge 4$ was established by Oliver Riordan [113], and by Annika Heckel [63] for $k=3$.

Theorem 2.7. *For $k \in \mathbb{Z}_{\ge 2}$ there exists no k-clique cover (and hence no k-clique factor) in $\boldsymbol{G}(n,p_-)$ whp, and there exists a k-clique factor (and hence a k-clique cover) in $\boldsymbol{G}(n,p_+)$ whp, over $n \in k\mathbb{Z}_{>0}$.*

The result for k-clique covers holds over $n \in \mathbb{Z}_{>0}$, only for k-clique factors we need $n \in k\mathbb{Z}_{>0}$.

2.2.3 Clique Factor Hitting Time. Our contribution [64] is the hitting time version of Theorem 2.7, i.e. the analogue to Theorem 2.6 for k-clique covers and k-clique factors. To be specific, let $\boldsymbol{G}_n = (\boldsymbol{G}_n(s))_s$ be the standard graph process, where $\boldsymbol{G}_n(0) = ([n], \emptyset)$ is empty and we obtain $\boldsymbol{G}_n(s+1)$ from $\boldsymbol{G}_n(s)$ by adding an edge, chosen uniformly at random from all edges that are not present in $\boldsymbol{G}_n(s)$. Let $\boldsymbol{s}_{\mathrm{c}}$ be the minimum step $s \in [\binom{n}{2}]$ such that a k-clique cover exists in $\boldsymbol{G}_n(s)$, and let $\boldsymbol{s}_{\mathrm{f}}$ be the minimum step $s \in [\binom{n}{2}]$ such that a k-clique factor exists in $\boldsymbol{G}_n(S)$. Further, let $(\boldsymbol{s}_-, \boldsymbol{s}_+)$ be independent of $\boldsymbol{G}_n$, where $\boldsymbol{s}_+$ is binomial with parameters $\binom{n}{2}$ and p_+, and $\boldsymbol{s}_-$ given $\boldsymbol{s}_+$ is binomial with parameters $\boldsymbol{s}_+$ and p_-/p_+ (for sufficiently large n).

As before, the case $k=2$ reduces to perfect matchings, i.e. $\boldsymbol{s}_{\mathrm{c}} = \boldsymbol{s}_{\mathrm{f}}$ whp was already established in [23]. In [64], we establish $\boldsymbol{s}_{\mathrm{c}} = \boldsymbol{s}_{\mathrm{f}}$ whp for all $k \ge 3$ and further extend this result to k-clique covers and factors in hypergraphs. Here, we exclusively focus on the case $k=3$.

Theorem 2.8. *For $k=3$ we have $\boldsymbol{s}_- \le \boldsymbol{s}_{\mathrm{c}} \le \boldsymbol{s}_+$ whp, and $\boldsymbol{s}_{\mathrm{f}} = \boldsymbol{s}_{\mathrm{c}}$ whp over $n \in 3\mathbb{Z}_{>0}$.*

The case $k=3$ is a special case which requires special attention, as can be seen in both [113, 63] and [64]. While the treatment of overlaps is slightly simpler (two triangles can only overlap in zero to three vertices), we typically have to deal with about $\ln(n)^3$ copies of a certain subgraph, coined clean 3-cycle, which significantly complicate matters compared to the case $k \ge 4$. Moreover, the discussion of the graph case is more involved compared to the hypergraph case, since in the former we have to deal with non-trivial overlaps. Thus, we choose to thoroughly discuss the most demanding case, and in turn refrain from establishing the general result, for the sake of accessibility. For the general case, we refer the reader to [64].

2.2.4 Matchings, Factors and CSPs. Finally, we address the question how Theorem 2.6, Theorem 2.7 and Theorem 2.8 can be embedded into the framework presented in Section 1. For this purpose we first discuss how perfect matchings in hypergraphs can be modeled as solutions of a CSP. So, given a hypergraph $H = ([n], \mathcal{H})$ let the factor graph $F = (\mathcal{V}, \mathcal{F}, v, \psi)$ be given by the variables $\mathcal{V} = \mathcal{H}$, the factors $\mathcal{F} = [n]$, the unordered neighborhoods $v = (v_a)_{a\in[n]}$ with $v_a = \{E \in \mathcal{H} : a \in E\}$ for $a \in [n]$, and the weights $\psi = (\psi_a)_{a\in[n]}$ with $\psi_a : \{0,1\}^{v_a} \to \{0,1\}$, $\tau \mapsto \mathbb{1}\{\|\tau\|_1 = 1\}$. An assignment $\sigma \in \{0,1\}^{\mathcal{V}}$ to the variables $\mathcal{V} = \mathcal{H}$, i.e. to the hyperedges, is nothing but a selection $\sigma^{-1}(1) \subseteq \mathcal{H}$ of the hyperedges. If this selection is such that $\psi_F(\sigma) = \prod_a \psi_a(\sigma_{v_a}) = 1$, then all constraints $a \in \mathcal{F} = [n]$, i.e. all vertices, are satisfied, meaning that each vertex is incident to exactly one hyperedge in $\sigma^{-1}(1)$ and hence $\sigma^{-1}(1)$ is a perfect matching. For the cover we only have to relax the weights to $\mathbb{1}\{\|\tau\|_1 > 0\}$.

Turning to the binomial hypergraph $\boldsymbol{H}_k(n,\pi) = ([n], \boldsymbol{\mathcal{H}})$, it is obvious that the corresponding factor graph significantly differs from the models in Section 1.4.1. The reason is that we consider a fixed number n of factors, but a random number $|\boldsymbol{\mathcal{H}}|$ of variables, and neighborhoods of varying arity, while the variable degrees are constant and equal to k. So, although we consider the arguably simplest, most widespread hypergraph model $\boldsymbol{H}_k(n,\pi)$, due to the non-standard mapping of vertices to factors, we need a completely different model in terms of factor graphs. Next to the LDGM code example in Section 1.4.6 and Section 2.1.1.3, this application also stresses the necessity of extensions of the factor graph model to more general distributions on both the variable and the factor side.

Modeling k-clique covers and k-clique factors in the binomial graph $\boldsymbol{G}(n,p) = ([n], \boldsymbol{\mathcal{G}})$ further stresses the necessity to consider more general distributions. For a graph $G = ([n], \mathcal{G})$ the associated factor graph $F = (\mathcal{V}, \mathcal{F}, v, \psi)$ is given by the variables $\mathcal{V} = \mathcal{H}$, where $\mathcal{H} = \{E \in \binom{[n]}{k} : \binom{E}{2} \subseteq \mathcal{G}\}$ are the cliques in G, by the factors $\mathcal{F} = [n]$, the neighborhoods $v = (v_a)_a$ with $v_a = \{E \in \mathcal{H} : a \in E\}$, and the weights $\psi = (\psi_a)_a$ with $\psi_a(\tau) = \mathbb{1}\{\|\tau\|_1 = 1\}$ for k-clique factors, respectively $\psi_a(\tau) = \mathbb{1}\{\|\tau\|_1 > 0\}$ for k-clique covers, as before.

Turning to the binomial graph $\boldsymbol{G}(n,p)$, we proceed as above. However, for hypergraphs it would have been sufficient to draw the variables independently as suggested above, while here we have to deal with dependencies for the distribution of the k-cliques and thus for the variables. This advanced example showcases that next to the models from Section 1.4.1 even much more elaborate models [31] are still not even close to cover all conceptually simple applications.

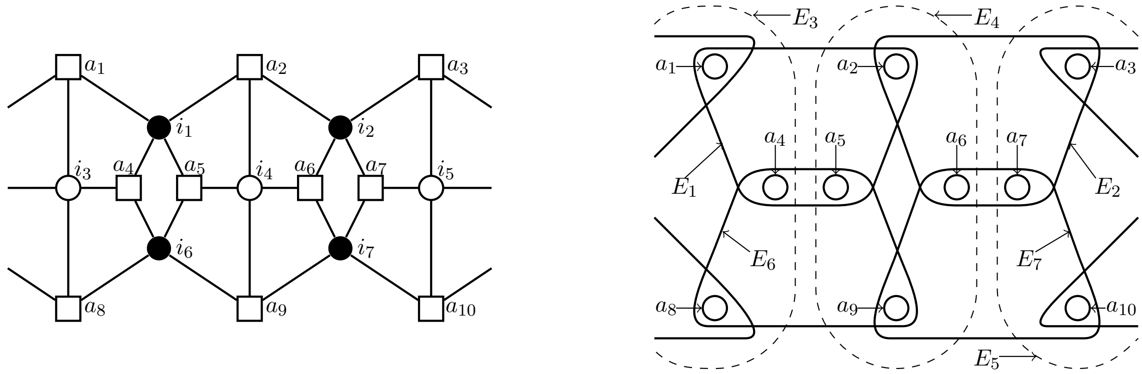
Notice that the standard hypergraph process $\boldsymbol{H}_{k.n}$ can be canonically extended to factor graphs. In this example, with unordered unique neighborhoods, we would choose one new neighborhood $\boldsymbol{v}_a \in \binom{[n]}{k}$ uniformly in each step a, where the step a is nothing but the factor. Independently, we would choose a weight $\boldsymbol{\psi}_a$ from the reference distribution. This process further canonically extends to other versions, e.g. neighborhoods in $[n]^k$ chosen with repetition. However, due to the identification of vertices with factors and hyperedges with variables, this and related processes are not suitable for an embedding of Theorem 2.6 and Theorem 2.8.

Thus, the results in this section demonstrate how far the general theory for factor graph models is from capturing even the simplest models. So, while it is undoubtedly most valuable to further develop the general theory, a case by case analysis of random factor graphs will still be indispensable for decades to come.

2.3 Satisfiability Thresholds for Regular Occupation Problems

In this section, we present the results established in [103], which are based on the results for the one-dimensional problem in [36] and [96].

In Section 2.3.1 we introduce the specific type of occupation problem under consideration. Then,



(a) Solution of the 2-occupation problem

(b) A 2-factor in a hypergraph

Figure 2: On the left we see a solution of the 4-regular 2-in-3 occupation problem on a 4-regular 3-factor graph, where the rectangles and circles depict the constraints (factors) and variables (filled if they take the value one in the solution). The figure on the right shows a 2-factor in a 3-regular 4-uniform hypergraph, where the circles, solid and dashed shapes represent the vertices, hyperedges in the 2-factor and the other hyperedges respectively.

in Section 2.3.2 we discuss the existing results. In Section 2.3.3 we establish the main result of this part.

2.3.1 Occupation Problems. Let k, $d \in \mathbb{Z}_{>1}$ and $r \in [k-1]$ be fixed. Additionally, we are given non-empty sets $[n]$ of variables and constraints $[m]$. An instance G of the d-regular r-in-k occupation problem is given by a sequence $G = (v_a)_{a\in[m]}$ of subsets $v_a \in \binom{[n]}{k}$ such that each of the n variables is contained in d of the subsets. The instance G has a natural interpretation as a (d, k)-biregular graph (or d-regular k-factor graph) with node sets $[n], [m]$ and edges $\{i, a\} \in \mathcal{E}$ if $i \in v_a$. By the handshaking lemma, such objects only exist if $dn = km$, which we assume in the following.

Given an instance G as just described, an assignment $x \in \{0, 1\}^n$ satisfies a constraint $a \in [m]$ if $\sum_{i\in v_a} x_i = r$, otherwise x violates a. If x satisfies all constraints $a \in [m]$, then x is a solution of G. Notice that d times the number of 1's in x matches the total number $rm = rdn/k$ of 1's observed on the factor side, so k has to divide rn, which we also assume in the following. We write $Z(G)$ for the number of solutions of G. Figure 2a shows an example of a 4-regular 2-in-3 occupation problem.

Further, for given m, $n \in \mathbb{Z}_{>0}$ let $\mathcal{G} = \mathcal{G}_{k,d,n,m}$ denote the set of all instances G with variables $[n]$ and constraints $[m]$. If $\mathcal{G}$ is not empty, then $\boldsymbol{G} \sim \mathrm{u}(\mathcal{G})$ is the random d-regular r-in-k occupation problem and $\boldsymbol{Z} = Z(\boldsymbol{G})$ its number of solutions. The random CSP above is given by $\boldsymbol{G}_{n,d}$ from Section 1.4.1 equipped with the weights $\mathbb{1}\{\|x_{v_a}\|_1 = r\}$, where $x_{v_a} = (x_i)_{i\in v_a} \in \{0, 1\}^{v_a}$ for $x \in \{0, 1\}^n$. The graph $\boldsymbol{G}$ from above is just $\boldsymbol{G}_{n,d}$ without the weights, since these are given by the problem definition.

2.3.2 Perfect Matchings. Notice that for $r = 0$ and $r = k$ we always have $\boldsymbol{Z} > 0$ almost surely since the all-0, respectively the all-1, assignment is a solution. Also, notice that the r-in-k and the $(k-r)$-in-k problems are equivalent, by switching 0 and 1, so we may restrict to $k \geq 2r > 0$. For the case $r = 1$, the existence and the location of the satisfiability threshold for the regular r-in-k occupation problem have been rigorously established in [36, 96]. Recall the entropy $H(p) = -p\ln(p) - (1-p)\ln(1-p)$

for $p \in [0,1]$ and let

$$d_r^*(k) = \frac{kH(w_1^*)}{kH(w_1^*) + \ln(w_2^*)} \in \mathbb{R}_{>0}, \quad w_1^* = \frac{r}{k}, \quad w_2^* = \binom{k}{r}^{-1} \tag{2}$$

be the satisfiability threshold estimate obtained from the first moment method. Notice that $kH(w_1^*)$ is the entropy of k iid Bernoulli variables with success probability w_1^* and $-\ln(w_2^*)$ is the entropy of the uniform distribution on $\binom{[k]}{r}$.

Theorem 2.9. *For $r = 1$ and $k \in \mathbb{Z}_{\geq 2}$ we have $\boldsymbol{Z} > 0$ whp if $d < d_r^*(k)$, and $\boldsymbol{Z} = 0$ whp if $d \geq d_r^*(k)$.*

Strictly speaking, Theorem 2.9 was established in [36, 96] for different models than the one considered here and only for the case $k \geq 3$. Perfect matchings were studied in [36] on hypergraphs, as opposed to the biregular graphs from Section 2.3.1. On the other hand, in [96] the result was established for the exact cover, also on hypergraphs. However, using standard arguments that are explained in the following, and the proof in [96], we verified that Theorem 2.9 holds for all models and extends to the case $k = 2$ as pointed out in [96].

2.3.3 Extension to 2-Factors. In the context of (hyper-) graphs (cf. [36]), perfect matchings, or 1-factors, can be generalized to r-factors, a collection of (hyper-) edges such that each vertex is incident to exactly r hyperedges. The exact cover problem (cf.[96]) can be generalized analogously, and both problems on regular uniform hypergraphs are variants of the d-regular r-in-k occupation problem (on biregular graphs) as defined in Section 2.3.1. The discussion in [103] covers the case $r = 2$, which is the main result of this section.

Theorem 2.10. *Let $k \in \mathbb{Z}_{\geq 4}$, $d \in \mathbb{Z}_{\geq 2}$, and let $\boldsymbol{Z}$ be the number of solutions from Section 2.3.1. There exists a sharp satisfiability threshold at d^*, i.e. for any increasing sequence $(n_i)_{i \in \mathbb{Z}_{>0}} \subseteq \mathcal{N} = \{n : dn,\ 2n \in k\mathbb{Z}_{>0}\}$ and $m_i = dn_i/k$ we have*

$$\lim_{i \to \infty} \mathbb{P}(\boldsymbol{Z} > 0) = \begin{cases} 1 & , d < d^* \\ 0 & , d \geq d^* \end{cases}.$$

We provide a self-contained proof for Theorem 2.10 using the first and second moment method with small subgraph conditioning for $\boldsymbol{Z}$. In particular, a main technical contribution in proving Theorem 2.10 is the optimization of a certain multivariate function that appears in the computation of the second moment, which encodes the interplay between the 'similarity' of various assignments and the change in the corresponding probability of being satisfying that they induce. A direct corollary of this optimization step at the threshold d^* is the confirmation of the conjecture by the authors in [101]. Among other things, at the core of our contribution we take a novel and rather different approach to tackle the optimization, inspired by [114] and [126] as well as other works relating the fixed points of belief propagation to the stationary points of the Bethe free entropy, respectively to the computation of the annealed free entropy density; see Section 5.5.6 for details. Finally, we show that d^* is not an integer in Lemma 5.9 below, so as opposed to the case $r = 1$ [96], for $r = 2$ there is no need for a dedicated analysis at criticality.

3 Condensation Threshold

This section is devoted to the proofs of the results in Section 2.1. Moreover, we discuss significant extensions of the results, thoroughly discuss applications and their embedding into the general framework, related work and open problems. A more detailed overview of the contents and the structure is presented in Section 3.1.

3.1 Preliminaries

We start with the discussion of related work to put the results into context, in Section 3.1.1. Our proof strategy is presented in Section 3.1.2, to provide an overview of the required steps to establish the main results on a high level. In Section 3.1.3, we address several extensions of the main results that we establish as well, but chose to omit in the introduction for brevity and simplicity. The terminology and notions with respect to uniform bounds are clarified in Section 3.1.4 Then, we explain the structure of the following discussion in Section 3.1.5. In Section 3.1.6 we introduce some basic notation and results from the literature.

3.1.1 Related Work. Not only the symmetric SBM has received a great deal of attention [1, 97], also the general (asymmetric) SBM is discussed algorithmically [5, 6, 7, 9, 61], and in particular using the cavity method [38, 130, 93, 112]. Partial threshold results where the Kesten-Stigum bound is tight, and general bounds for the binary asymmetric SBM can be found in [26]. The mutual information limit under weak convexity assumptions in the dense regime can be found in [111]. We are not aware of a discussion of the general (dis-) assortative SBM in the literature, however, conceptually similar examples like the $q_1 + q_2$ SBM in [112] have been analyzed. The mutual information of the ground truth and the planted model was established in [33] for the symmetric disassortative case, which we extend here to the general disassortative case. The binary assortative case was recently discussed in [3] and [43].

The existence of the mutual information limit for graphical channels was established in [4], and the limit was determined in [33] under similar assumptions, in both cases for uniform input messages. We extend the class of channels for which it is known that the results are applicable (BISO and SAT-type channels) two-fold via direct extension and via closure properties, and further determine the limit for any binary memoryless source, uniform or not. As a special case, we obtain the mutual information limit for the BAC (and even k). We also extend a generalized version of Gallager's mapping (cf. [90]) to factor graphs (cf. Section 3.5.5.8). The uniform convergence results in this work combined with the continuity of the limit give a theoretical justification for such approximations in the large n limit.

The results in this work rely on a convexity-like assumption, based on Hypothesis H in [4] with respect to channels. Intuitively, such channels have the property that more information can be transmitted reliably using a single code of block length $n_1 + n_1$, rather than a code of block length n_1 and a code of block length n_2. Techniques that do not rely on this assumption were used in [3], the spatial coupling in [60], and proofs based on convex relaxation hierarchies [66]. Bounds on the Chi-squared mutual information and connection to percolation probabilities were obtained in [2]. More general factor graphs have also been discussed [24, 31], under stronger assumptions.

The contributions [30, 32] address more general weights, and stronger results below the threshold are derived, namely for the partition function, and the result for the relative entropy is strengthened to mutual contiguity, next to additional results. However, both contributions rely on other assumptions, in particular on **MIN**. Conversely, previous work [30, 33, 32, 31] benefits from the discussion of **POS** in Section 3.5.5, and in particular from the models in Section 2.1.2.2 (inspired by [105]), which extends the scope both theoretically and application-wise. We did verify that these models cover k-SAT, k-NAESAT, k-XORSAT, the k-spin model, k-COL, the Potts model, and the binary memoryless symmetric channels (covering BISO, BSC and BEC), which were discussed in [33, 33, 32, 31, 4, 3]. We use the weakened assumption **POS** (cf. [30]), which is (conceptually) significantly weaker than Hypothesis H. In particular, this is why we do not only cover all examples, but also the counterexample (the 3-XORSAT) in [4]. The issue with the parity of k is discussed in Remark 3.128. Our discussion of the closure properties also implies that permutation invariance assumptions for weights [30] are not required (cf. [31]). The absence of **SYM** and the uniform prior allows to naturally cover all products $\boldsymbol{\psi} = \prod_h \boldsymbol{\psi}_h$, which is desirable from a theoretical viewpoint.

The strengthened results required an improvement and a generalization of the pinning lemma. This amounts to explicit bounds on the probability for given parameters, and not only for two, but for any number of coordinates. We derived a proof based on [33, 92], a similar result has been established independently in [66]. A general theory is discussed in [13].

We adapt the strategy in [33] to establish the main results. Thus, we combine the pinning lemma with the Aizenman-Sims-Starr scheme and the interpolation method to obtain the limit for the planted model quenched free entropy. As noted in [32] and [31], this strategy compresses the (compressed) 85 pages of proofs in [33] into one sentence. Explaining all modifications, tweaks and boosts is equivalent to explaining the proof, so we present an arbitrary selection. We modify every second argument to avoid using **SYM**, mostly by constructing more subtle couplings. Accommodating general ground truth distributions is canonical at times, on other occasions the model complexity (e.g. interpolation method), the proof complexity (e.g. Gibbs marginal distributions) and generalizing definitions requires a deep understanding of the interplay of the objects involved. Establishing uniform convergence mostly requires explicit bounds and dedicated bookkeeping, the polynomial convergence guarantee on the other hand required fundamental changes and improvements, say, the resolution of the double limit into a single limit, which requires to increase the pinning probabilities and thereby an adjustment of the couplings, or boosting the concentration of the posterior from $o(1)$ to subgaussian concentration. The improved condensation threshold bound and connection to applications is self-evident.

3.1.2 Outline of the Proof. Without further mention, assume that $(p, \gamma^*, d) \in \mathfrak{P}$, i.e. (p, γ^*) satisfy **BAL** and **POS**, and $d \le d_\uparrow$. The bounds in the following results only depend on $\mathfrak{g} = (q, k, \psi_\downarrow, d_\uparrow)$. We obtain Proposition 2.1 by direct verification for the the trivial weights, and otherwise using a Taylor series expansion, similar to **POS** in [33]. Then we use a new twist to show that the resulting contributions are non-negative. The proofs of all main results rely on the properties of the Nishimori ground truth $\hat{\boldsymbol{\sigma}}_{p,\gamma^*,n,m} \in [q]^n$. We present the details in Section 3.1.2.1. Next, we derive two crucial properties of the free entropies in the main results: concentration and Lipschitz continuity of the conditional expectations. Details can be found in Section 3.1.2.2.

The proof of Theorem 2.2 relies on the pinning lemma discussed in Section 3.1.2.3. In Section 3.1.2.4 we explain its application and the steps required to obtain Theorem 2.2. Theorem 2.3 follows from Theorem 2.2 using the properties of $\hat{\boldsymbol{\sigma}}_m$. The result is immediate for $D_{\mathrm{KL}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}) \| \boldsymbol{G}_{\boldsymbol{m}})$, only the discussion of the conditional relative entropy requires some care. The proof of Theorem 2.4 is also rather short, but relies on two clever ideas. Compared to the preceding two results, for the proof of Theorem 2.5 it is rather cumbersome to decompose the mutual information into the ground truth weight and the free entropy, and to derive the asymptotics of the former using $\hat{\boldsymbol{\sigma}}_m$.

3.1.2.1 The Nishimori Ground Truth. As explained in Section 2.1.2.5, we need to consider Gibbs measures $\mu_{\gamma^*,G}$ that are consistent with $\boldsymbol{\sigma}^*$. In order to control both $\mu_{\gamma^*,G}$ and $\boldsymbol{\sigma}^*$ given $\boldsymbol{G}^*_m(\boldsymbol{\sigma}^*)$, we recover the Bayes optimal case and hence ensure that the Nishimori condition holds (Section 1.2.2 in

[129]), by introducing the Nishimori ground truth $\hat{\boldsymbol{\sigma}}_m$, given by the RN derivative

$$\hat{r}_{p,\gamma^*,n,m} : [q]^n \to \mathbb{R}_{>0},\ \sigma \mapsto \frac{\mathbb{E}[\psi_{\boldsymbol{G}(m)}(\sigma)]}{\mathbb{E}[Z_{\gamma^*}(\boldsymbol{G}(m)]}$$

with respect to $\boldsymbol{\sigma}^*$. Let $m_\uparrow = 2d_\uparrow n/k$, and let $\|\boldsymbol{a} - \boldsymbol{b}\|_{\mathrm{tv}} = \sup_{\mathcal{E}} |\mu(\mathcal{E}) - \nu(\mathcal{E})|$ be the total variation distance of $\boldsymbol{a}$, $\boldsymbol{b}$ with laws μ, ν respectively. Let $\gamma_{\mathrm{n},\sigma}(\tau) = \frac{1}{n}|\{i \in [n] : \sigma(i) = \tau\}|$, $\tau \in [q]$, be the color frequencies of $\sigma \in [q]^n$.

Proposition 3.1. *Let $m \in \mathbb{Z}_{\geq 0}$ with $m \leq m_\uparrow$.*
a) There exists $c \in \mathbb{R}_{>0}$ with $\hat{r}_m \leq c$.
b) There exists $c \in \mathbb{R}_{>0}$ such that $\hat{r}_m(\sigma) \geq \exp(-c\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2 n)$.
c) The ground truths $\boldsymbol{\sigma}^*$ given $\gamma_{\mathrm{n},\boldsymbol{\sigma}^*}$ and $\hat{\boldsymbol{\sigma}}_m$ given $\gamma_{\mathrm{n},\hat{\boldsymbol{\sigma}}_m}$ have the same law.
d) We have $(\hat{\boldsymbol{\sigma}}_m, \boldsymbol{G}^*_m(\hat{\boldsymbol{\sigma}}_m)) \sim (\boldsymbol{\sigma}_{\gamma^*,\boldsymbol{G}^*(m,\hat{\boldsymbol{\sigma}}(m))}, \boldsymbol{G}^*_m(\hat{\boldsymbol{\sigma}}_m))$.

3.1.2.2 Concentration and Continuity. Before we turn to the proof of Theorem 2.2, we establish concentration (self-averaging) and continuity of the free entropies.

Proposition 3.2. *Let $m \in \mathbb{Z}_{\geq 0}$, $\sigma \in [q]^n$ and $\boldsymbol{G}^\circ \in \{\boldsymbol{G}_m, \boldsymbol{G}^*_m(\sigma), \boldsymbol{G}^*_m(\boldsymbol{\sigma}^*), \boldsymbol{G}^*_m(\hat{\boldsymbol{\sigma}}_m)\}$.*
a) There exists $c \in \mathbb{R}_{>0}$ such that $|\phi_{\gamma^*}(\boldsymbol{G}^\circ)| \leq ckm/n$ almost surely.
b) There exists $c \in \mathbb{R}^2_{>0}$ such that $\mathbb{P}(|\phi_{\gamma^*}(\boldsymbol{G}^\circ) - \mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}^\circ)]| \geq r) \leq c_2 e^{-c_1 r^2 n}$ for $m \leq m_\uparrow$ and $r \in \mathbb{R}_{\geq 0}$.
c) For $\gamma_{\mathrm{n},\sigma} \geq \psi_\downarrow/2$ and $m \leq m_\uparrow$, $\sigma' \in [q]^n$, $m' \in \mathbb{Z}_{\geq 0}$ and $m^\circ \in \mathbb{Z}^2_{\geq 0}$ we have

$$|\mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}^*_m(\sigma)] - \mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}^*_{m'}(\sigma')]| \leq L\left(\|\gamma_{\mathrm{n},\sigma} - \gamma_{\mathrm{n},\sigma'}\|_{\mathrm{tv}} + \left|\frac{km}{n} - \frac{km'}{n}\right|\right),$$
$$|\mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}(m^\circ_1)] - \mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}(m^\circ_2)]| \leq L\left|\frac{km^\circ_1}{n} - \frac{km^\circ_2}{n}\right|.$$

Proposition 3.2a), with the concentration of $\boldsymbol{m}$ around its expectation dn/k, suggests that we can restrict the expectations accordingly, and Proposition 3.2c) ensures that the conditional expectations asymptotically coincide close to the expectation dn/k. Regarding the ground truths, we recall that $\mathbb{P}(\|\gamma_{\mathrm{n},\boldsymbol{\sigma}^*} - \gamma^*\|_{\mathrm{tv}} \geq r) \leq c' \exp(-cr^2 n)$ for suitable c, c', and hence we can restrict to converging color frequencies (since the free entropies are uniformly bounded for $m \leq m_\uparrow$). Proposition 3.2c) then ensures that the conditional expected free entropies asymptotically coincide.

3.1.2.3 The Pinning Lemma. Proposition 3.1 and Proposition 3.2 establish that the quenched free entropy (densities) for $\boldsymbol{\sigma}^*$, $\hat{\boldsymbol{\sigma}}_m$ asymptotically coincide. The following pinning lemma illustrates *why* working with $\hat{\boldsymbol{\sigma}}_m$ is desirable. Recall that the product measure $\alpha = \bigotimes_{a,h} \gamma_{a,h}$ of the marginals γ is used in the definition of Z_{f}, Z_{v} for the Bethe free entropy. The law α corresponds to the joint distribution of $\mu_{\gamma^*,\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*)}$ on a random number of variables in the finite size case. One of the main obstacles is to show that these joint distributions indeed asymptotically factorize, and this is exactly where the pinning lemma comes into play. For $\boldsymbol{\sigma} \in [q]^n$ with law μ, $\ell \in \mathbb{Z}_{>0}$ and $v \in [n]^\ell$ let $\mu|_v$ be the law of $\boldsymbol{\sigma}_v \in [q]^\ell$. For $\ell = 1$ we use the shorthand $\mu|_{v(1)} = \mu|_v$. Further, let $\iota_\circ(\mu, v) = D_{\mathrm{KL}}(\mu|_v \| \bigotimes_h \mu|_{v(h)})$ and $\iota_\ell(\mu) = \mathbb{E}[\iota_\circ(\mu, \boldsymbol{v})]$ with $\boldsymbol{v} \sim \mathrm{u}([n]^\ell)$. For $\check{\sigma} \in [q]^n$ and $\mathcal{U} \subseteq [n]$ let $[\mu]^\downarrow_{\mathcal{U},\check{\sigma}} \in \mathcal{P}([q]^n)$ be the law of $\boldsymbol{\sigma}|(\boldsymbol{\sigma}(i))_{i\in\mathcal{U}} = (\check{\sigma}(i))_{i\in\mathcal{U}}$, if this is defined. The next result generalizes Lemma 3.5 in [33], corresponding to $\ell = 2$, and states a stronger version that addresses the conditional relative entropy directly.

Lemma 3.3. *For $n \in \mathbb{Z}_{>0}$, $\mu \in \mathcal{P}([q]^n)$, $\ell \in \mathbb{Z}_{>0}$ and $\Theta^{\downarrow} \in (0, n]$ the following holds. Let $\boldsymbol{\theta} \sim \mathrm{u}([0, \Theta^{\downarrow}])$, further iid Bernoulli $\check{\boldsymbol{u}} \in \{0,1\}^n$ with success probability $\boldsymbol{\theta}/n \in [0,1]$, $\boldsymbol{\mathcal{U}} = \check{\boldsymbol{u}}^{-1}(1)$ and $\boldsymbol{\sigma} \sim \mu$ with $(\boldsymbol{\sigma}, \boldsymbol{\mathcal{U}}) \sim \boldsymbol{\sigma} \otimes \boldsymbol{\mathcal{U}}$. Then we have $\mathbb{E}[\iota([\mu]^{\downarrow}_{\boldsymbol{\mathcal{U}},\boldsymbol{\sigma}})] \leq \binom{\ell}{2} \ln(q)/\Theta^{\downarrow}$.*

3.1.2.4 The Planted Model Quenched Free Entropy. We use the interpolation method to obtain the lower bound in Theorem 2.2.

Proposition 3.4. *We have $\mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))] \geq B_{\uparrow}(d) - \mathcal{O}_{\mathrm{u}}(n^{-1/4})$.*

We use the Aizenman-Sims-Starr scheme to obtain the upper bound in Theorem 2.2.

Proposition 3.5. *There exists $\rho \in \mathbb{R}_{>0}$ such that $\mathbb{E}[\phi_{\gamma^*}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))] \leq B_{\uparrow}(d) + \mathcal{O}_{\mathrm{u}}(n^{-\rho})$.*

Both methods require that $\boldsymbol{m} \sim \mathrm{Po}(dn/k)$ is Poisson distributed. Further, we need the pinning lemma in both cases. For this purpose we decorate the graphs with an additional type of factors, say pins, to turn $\mu_{\gamma^*,G}$ into $[\mu_{\gamma^*,G}]^{\downarrow}_{\mathcal{U},\check{\sigma}}$, which then ensures asymptotic independence. A careful calibration of the parameters ensures that the effect of this pinning procedure on the quenched free entropy is asymptotically negligible, while breaking dependencies sufficiently fast.

For the interpolation method we need an additional type of factors, say interpolators, to morph the decoupled model, underlying the Bethe functional, into the graph. The derivative of this transition is (almost) the map in **POS**, so this assumption yields the lower bound in the end.

The Aizenman-Sims-Starr scheme only relies on the standard factors and pins (and external fields). Here, we use that the quenched free entropy *density* can be rewritten as the average difference of the quenched free entropies for $n+1$ and n. Hence, deriving the limit of the difference yields the limit of the quenched free entropy density. This limit *is asymptotically equal to* the expectation of the Bethe free entropy over a certain distribution $\boldsymbol{\pi}_n \in \mathcal{P}^2_{*,\gamma^*}([q])$.

So, in a nutshell, the first part of the proof clarifies how exactly we can utilize the pinning lemma and justifies the application. In the second part we implement the interpolation method using the fully decorated graphs, followed by the implementation of the Aizenman-Sims-Starr scheme with the slightly simpler graphs in the third part.

3.1.3 Extensions and Remarks. For the sake of brevity we did not present our main results in their full generality. In the following, we discuss the actual scope and strength of our results, further implications and related work.

The main results hold for more general $(\boldsymbol{m}^*_n)_n$. To be specific, let $\varepsilon_{\mathrm{m}}, \delta_{\mathrm{m}} : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$ with $\lim_{n\to\infty} \varepsilon_{\mathrm{m}}(n) = \lim_{n\to\infty} \delta_{\mathrm{m}}(n) = 0$ and $d^* = \limsup_{n\to\infty} \mathbb{E}[\boldsymbol{m}^*_n]$. Then the results hold for $\boldsymbol{m}$ replaced by $\boldsymbol{m}^*_n$ and d replaced by d^* if $\mathbb{P}(|\boldsymbol{d}^*_n - d^*| > \delta_{\mathrm{m}}(n)) \leq \varepsilon_{\mathrm{m}}(n)$, $\mathbb{E}[\mathbb{1}\{|\boldsymbol{d}^*_n - d^*| > \delta_{\mathrm{m}}(n)\}\boldsymbol{d}^*_n] \leq \varepsilon_{\mathrm{m}}(n)$ for all $n \in \mathbb{Z}_{>0}$ and $d^* \leq d_{\uparrow}$, where $\boldsymbol{d}^*_n = k\boldsymbol{m}^*_n/n$. Also uniform convergence holds, i.e. the results hold for $\mathfrak{g} = (q, k, p, \gamma^*, d_{\uparrow}, \delta_{\mathrm{m}}, \varepsilon_{\mathrm{m}})$.

The results cover maximizers $\gamma^* \in \mathcal{P}([q])$ with *any support* (cf. Section 3.5.1), but *uniformly* over fixed support. We also establish the results for *all* standard Erdős-Rényi type models in Section 3.5.7. Results for graphs with external fields are discussed in Section 3.5.2. We also establish convergence in probability for the conditional expectations given $\boldsymbol{m}^*$ in Section 3.5.4.

3.1.4 Global Parameters and Uniform Convergence. Let $q, k \in \mathbb{Z}_{>0}$, $\psi_{\downarrow} \in (0, 1/q)$, $\psi_{\uparrow} = 1/\psi_{\downarrow}$ and $d_{\uparrow} \in \mathbb{R}_{>0}$ (cf. Section 2.1.2.1 and Section 2.1.2.3). Further, let $\delta_{\mathrm{m}}, \varepsilon_{\mathrm{m}} : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$ with $\lim_{n\to\infty} \delta_{\mathrm{m}}(n) = 0$, $\lim_{n\to\infty} \varepsilon_{\mathrm{m}}(n) = 0$ be the bounds for $\boldsymbol{m}^*$ from Section 3.1.3. Hence, the global parameters are $\mathfrak{g} = (q, k, \psi_{\downarrow}, d_{\uparrow}, \delta_{\mathrm{m}}, \varepsilon_{\mathrm{m}})$. We keep $\mathfrak{g}$ fixed throughout the remainder and do not track dependencies on $\mathfrak{g}$. Thus, all values depending on $\mathfrak{g}$, meaning all functions of $\mathfrak{g}$, and only such values, are considered to be constant. Still, we may write $c_{\mathfrak{g}}$ to stress that c only depends on $\mathfrak{g}$.

Without loss of generality we may assume that $\psi_{\downarrow}$ is arbitrarily small and that $d_{\uparrow}$ is arbitrarily large since this only increases the set of model parameters. We further assume without loss of generality that ε_{m} and δ_{m} are non-increasing. After we restricted to the δ_{m}-ball around d, the largest average degree to be considered is $d+\delta_{\mathrm{m}}(n)\leq d_{\uparrow}+\delta_{\mathrm{m}}(1)$. Without loss of generality we take $\delta_{\mathrm{m}}(1)=d_{\uparrow}$, so $m_{\uparrow,n}=2d_{\uparrow}n/k$ from Section 3.1.2.1 is the desired maximal factor count.

We use the Bachmann-Landau notation $\mathcal{O}(f(n)), o(f(n)), \Omega(f(n)), \omega(f(n)), \Theta(f(n))$ only with constants, as disussed in Section 2.1.2.3 (but without subscripts).

3.1.5 A Roadmap to the Proofs. Throughout the remainder we assume without loss of generality that $q,k\geq 2$, except for Section 3.5.1, where we justify this assumption. In Section 3.2 we derive basic results. Specifically, in Section 3.2.1 we introduce decorated graphs and establish basic properties. In Section 3.2.2 we discuss basic properties of the Nishimori ground truth $\hat{\boldsymbol{\sigma}}$, including the proof of Proposition 3.1. In Section 3.2.3 we establish boundedness, continuity and concentration for the free entropy, including the proof of Proposition 3.2.

Section 3.3 is devoted to the proof of Theorem 2.2. Specifically, in Section 3.3.1 we discuss the Gibbs measures of decorated graphs, establish the pinning lemma 3.3 and apply it to the graphs, establish a result for reweighted marginal distributions of general measures and apply it to the graphs, and finally discuss projections of $\mathcal{P}^2([q])$ onto $\mathcal{P}^2_*([q])$. In Section 3.3.2 we turn to the interpolation method including the proof of Proposition 3.4. The discussion in Section 3.3.3 addresses the Aizenman-Sims-Starr scheme including the proof of Proposition 3.5, where we also establish Theorem 2.2.

In Section 3.4 we present the proofs of the remaining main results. We establish Theorem 2.3 in Section 3.4.1, followed by the proof of Theorem 2.4 in Section 3.4.2, and conclude the proof of the main results with Theorem 2.5 in Section 3.4.3.

The remainder is discussed in Section 3.5, where we in particular formalize the statements and prove the claims in Section 3.1.3, discuss the assumptions $\mathfrak{A}$ and prove Proposition 2.1, and finally formalize the statements and prove the claims in Section 2.1.1.2 and Section 2.1.1.3.

3.1.6 Notions, Notation and Results from the Literature. We consider a sufficiently rich probability space $(\Omega,\mathcal{F},\mathbb{P})$. Random quantities (variables, vectors) are measurable maps $\boldsymbol{a}:\Omega\to\mathcal{A}$ and denoted in bold. An event $\mathcal{E}\in\mathcal{F}$ holds (almost surely) if $\mathbb{P}(\mathcal{E})=1$. The $(\boldsymbol{a},\boldsymbol{b})$-derivative is the Radon-Nikodym derivative of $\boldsymbol{a}$ with respect to $\boldsymbol{b}$. Further, we use $\mathcal{A}\dot{\cup}\mathcal{B}$ for the disjoint union, $[n]=\mathbb{Z}\cap[1,n]$, $2^{\mathcal{S}}$ for the power set of $\mathcal{S}$, $\binom{\mathcal{A}}{a}\subseteq 2^{\mathcal{S}}$ for the a-subsets $\mathcal{B}\subseteq\mathcal{A}$ with $|\mathcal{B}|=a$, $\mathcal{A}^{\mathcal{B}}$ for the maps $f:\mathcal{B}\to\mathcal{A}$, and $\mathcal{A}^{\underline{\mathcal{B}}}=\{f\in\mathcal{A}^{\mathcal{B}}:\forall a\in\mathcal{A}\,|f^{-1}(a)|\leq 1\}$ for the injections, with $\mathcal{A}^n=\mathcal{A}^{[n]}$, $\mathcal{A}^{\underline{n}}=\mathcal{A}^{\underline{[n]}}$. We consider spaces equipped with their canonical structure unless mentioned otherwise, mostly subspaces of $\mathbb{R}^a$. We use $\leq$ for componentwise inequalities and $\equiv$ for componentwise equality. We use $\sim$ for equality in distribution. A space $\mathcal{A}$ is a copy of a space $\mathcal{B}$ if it carries the same structure under some bijection, in which case we identify $\mathcal{A}$ with $\mathcal{B}$. We identify $\mu\in\mathcal{P}(\mathbb{Z})$ with its probability mass function $\mu:\mathbb{Z}\to[0,1]$. We further use similar identifications to focus on the relevant arguments while avoiding technical routine discussions.

We (partially) suppress dependencies for brevity, e.g. $f_a(x)=f_a=f$. Clearly, this leads to ambiguities, e.g. $\boldsymbol{G}^*$ may refer to $\boldsymbol{G}^*(\boldsymbol{m},\boldsymbol{\sigma}^*)$, $\boldsymbol{G}^*(m,\sigma)$ or any other combination. Hence, when we omit a dependency, the dependency is the same quantity as in the definition and thereby uniquely identified. Further, we keep the notation consistent to earn this degree of flexibility. Finally, we may use $f_{a,x}=f_a(x)=f(a,x)$ interchangeably, for readability or to indicate the distinction between variables and parameters. Similarly, we use mixed notation for random quantities $\boldsymbol{x}$ and their distributions (laws) $\boldsymbol{x} \sim \mu$, e.g. $D_{\mathrm{KL}}(\boldsymbol{x}_1\|\boldsymbol{x}_2) = D_{\mathrm{KL}}(\mu_1\|\mu_2)$.

We extend $(\sigma_i)_{i\in[n]} \in [q]^n$ to maps, i.e. for $v \in [n]^k$ let $\sigma_v = (\sigma_{v(h)})_h$, and let $\mu|_v$ be the law of $\boldsymbol{\sigma}_v$ with $\boldsymbol{\sigma} \sim \mu$, as in Section 3.1.2. If v is the enumeration of $\mathcal{V} \subseteq [n]$, i.e. the unique strictly increasing map $v : [|\mathcal{V}|] \to \mathcal{V}$, we use the shorthands $\sigma_{\mathcal{V}} = \sigma_v$ and $\mu|_{\mathcal{V}} = \mu|_v$ and in particular $\sigma_i = \sigma_{\{i\}}$, $\mu|_i = \mu_{\{i\}}$. Further, let $\mu|_* = \sum_i \frac{1}{n}\mu|_i \in \mathcal{P}([q])$ be the law of $\boldsymbol{\sigma}_{\boldsymbol{i}} \in [q]$, with $(\boldsymbol{\sigma}, \boldsymbol{i}) \sim \mu \otimes \mathrm{u}([n])$. We denote the total variation distance by $\|\mu - \nu\|_{\mathrm{tv}} = \sup_{\mathcal{E}} |\mu(\mathcal{E}) - \nu(\mathcal{E})|$ and let $\Gamma(\mu_1, \mu_2) \subseteq \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ be the couplings of $\mu_1 \in \mathcal{P}(\mathcal{X}_1)$ and $\mu_2 \in \mathcal{P}(\mathcal{X}_2)$, i.e. for $\nu \in \Gamma(\mu_1, \mu_2)$ we have $\nu|_1 = \mu_1$ and $\nu|_2 = \mu_2$.

Observation 3.6. *Notice that the following holds.*

a) For $\mu \in \mathcal{P}([q])^2$ we have $\|\mu_1 - \mu_2\|_{\mathrm{tv}} = \frac{1}{2}\|\mu_1 - \mu_2\|_1$.
b) For $\mu \in (\mathcal{P}([q])^n)^2$ we have $\| \bigotimes_i \mu_{1,i} - \bigotimes_i \mu_{2,i}\|_{\mathrm{tv}} \le \sum_i \|\mu_{1,i} - \mu_{2,i}\|_{\mathrm{tv}}$.
c) For $\boldsymbol{x}, \boldsymbol{x}' \in [q]$ and $\boldsymbol{y}(x) \in [q']$, $x \in [q]$, we have $\|(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x})) - (\boldsymbol{x}', \boldsymbol{y}(\boldsymbol{x}'))\|_{\mathrm{tv}} = \|\boldsymbol{x} - \boldsymbol{x}'\|_{\mathrm{tv}}$.
d) For $\ell \in \mathbb{Z}_{>0}^2$ with $\ell_1 \le \ell_2$, further $v_i \in [n]^{\ell(i)}$, $i \in [2]$, with $v_1 = v_{2,[\ell(1)]}$, and $\mu \in \mathcal{P}([q]^n)^2$ we have $\|\mu_1|_{v_1} - \mu_2|_{v_1}\|_{\mathrm{tv}} \le \|\mu_1|_{v_2} - \mu_2|_{v_2}\|_{\mathrm{tv}}$ and $\|\mu_1|_* - \mu_2|_*\|_{\mathrm{tv}} \le \mathbb{E}[\|\mu_1|_{\boldsymbol{i}} - \mu_2|_{\boldsymbol{i}}\|_{\mathrm{tv}}]$.
e) For a coupling $\boldsymbol{y}$ of $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in [q]$ we have $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathrm{tv}} \le \mathbb{P}(\boldsymbol{y}_1 \ne \boldsymbol{y}_2)$ and there exists $\boldsymbol{y} \in \Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2)$ with $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathrm{tv}} = \mathbb{P}(\boldsymbol{y}_1 \ne \boldsymbol{y}_2)$.
f) For $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in [q]$ we have $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathrm{tv}} \le \sqrt{\frac{1}{2}D_{\mathrm{KL}}(\boldsymbol{x}_1\|\boldsymbol{x}_2)}$.
g) For a measurable set $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ equipped with $\|\cdot\|_{\mathrm{tv}}$ and the Borel algebra, and for $\boldsymbol{p} \in \mathcal{P}(\mathcal{X})$, the pair $(\boldsymbol{p}, \boldsymbol{x}_{\boldsymbol{p}})$ is well-defined, where $\boldsymbol{x}_p \sim p$ for $p \in \mathcal{P}(\mathcal{X})$.

Proof. Part 3.6a) can be found on page 153 in [68], Part 3.6e) on page 10 in [125], Part 3.6f) is Pinsker's inequality, e.g. Equation (2.8) in [33]. For Part 3.6b) we have

$$\begin{aligned} 2\|\mu_1 \otimes \mu_2 - \nu_1 \otimes \nu_2\|_{\mathrm{tv}} &= \sum_x |\mu_1(x_1)\mu_2(x_2) - \nu_1(x_1)\nu_2(x_2)| \\ &\le \sum_x \mu_1(x_1)|\mu_2(x_2) - \nu_2(x_2)| + \sum_x \nu_2(x_2)|\mu_1(x_1) - \nu_1(x_1)|, \end{aligned}$$

so the assertion holds for $n = 2$. The general case follows by induction analogous to the above. Part 3.6c) follows similarly, using Part 3.6a), distributivity and normalization of the conditional laws. For Part 3.6d) notice that

$$\|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\|_{\mathrm{tv}} \le \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathrm{tv}}$$

holds in general and specifically for restrictions. The second part of the assertion follows from Part 3.6a) and the triangle inequality. For Part 3.6g) let Σ be the σ-algebra of $\mathcal{X}$, and $\kappa : \Sigma \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}$, $(\mathcal{E}, p) \mapsto p(\mathcal{E})$. Clearly, $\mathcal{E} \mapsto \kappa(\mathcal{E}, p)$ is $p \in \mathcal{P}(\mathcal{X})$, and $p \mapsto \kappa(\mathcal{E}, p)$ is 1-Lipschitz since $|p(\mathcal{E}) - p'(\mathcal{E})| \le \|p - p'\|_{\mathrm{tv}}$ by the definition of $\|\cdot\|_{\mathrm{tv}}$, for all $\mathcal{E} \in \Sigma$, so κ is a kernel. Thus, for given $\boldsymbol{p}$ we recover the composition $\boldsymbol{p} \otimes \kappa \in \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{X}))$. □

We use the Poisson distribution $\mathrm{Po}(\lambda)$, the binomial distribution $\mathrm{Bin}(n, p)$, the uniform distribution $\mathrm{u}(\mathcal{S})$) and the one-point mass $\mu_{\bullet,\mathcal{S},s} \in \mathcal{P}(\mathcal{S})$ on $s \in \mathcal{S}$.

Observation 3.7. *Let $\overline{m} \in \mathbb{R}_{\ge 0}$ and $\boldsymbol{m} \sim \mathrm{Po}(\overline{m})$.*

a) We have $\mathbb{P}(\boldsymbol{m} = m)m = \overline{m}\mathbb{P}(\boldsymbol{m} = m - 1)$.
b) Let $k \in \mathbb{Z}_{>0}$ and $p \in \mathcal{P}([k])$. Further, let $\boldsymbol{n} \sim \bigotimes_{i\in[k]} \mathrm{Po}(p_k\overline{m})$, and let $\boldsymbol{n}(m) \in \mathbb{Z}_{\ge 0}^k$ be multinomial with parameters m and p, then we have $(\sum_i \boldsymbol{n}_i, \boldsymbol{n}) \sim (\boldsymbol{m}, \boldsymbol{n}(\boldsymbol{m}))$.
c) For $\overline{m}' \in \mathbb{R}_{>0}$, $\boldsymbol{m}' \sim \mathrm{Po}(\overline{m}')$ we have $D_{\mathrm{KL}}(\boldsymbol{m}\|\boldsymbol{m}') = \overline{m}' - \overline{m} + \overline{m}\ln(\overline{m}/\overline{m}')$.

d) There exist $c, c' \in \mathbb{R}_{>0}$ with $\mathbb{P}(|\boldsymbol{m} - \overline{m}| \geq r) \leq c' \exp(-\frac{cr^2}{\overline{m}+r})$ for $r \in \mathbb{R}_{\geq 0}$.
e) Let $p \in \mathcal{P}([K])$, and let $\boldsymbol{x} \sim p^{\otimes N}$ be multinoulli. Let $\boldsymbol{m} \in \mathbb{Z}_{\geq 0}^K$ be multinomial with N and p, and $\boldsymbol{U}_m \sim \mathrm{u}(\{x \in [K]^N : (|x^{-1}(y)|)_y = m\})$. Then we have $\boldsymbol{x} \sim \boldsymbol{U}_{\boldsymbol{m}}$.

Proof. The first three parts can be easily verified directly, the last part follows from Theorem 2.1 with Remark 2.6 in [68] and $c = 1/2$, $c' = 2$, where we notice that for $\overline{m} = r = 0$ the exponent is 0. The last part follows by direct computation. □

For $x \in \mathbb{R}$ let $\lceil x \rceil = \min \mathbb{Z}_{\geq x}$ and $\lfloor x \rfloor = \max \mathbb{Z}_{\leq x}$. For $n, k \in \mathbb{Z}_{\geq 0}$ let $n^{\underline{k}} = \prod_{h=0}^{k-1}(n-h) = |[n]^{\underline{k}}|$.

3.2 Preparations

Decorated factor graphs, the central objects in the proof of Theorem 2.2, are introduced in Section 3.2.1. The Nishimori ground truth is discussed in Section 3.2.2. The we turn to the concentration of the free entropy and the Lipschitz continuity of the conditional expectations in Section 3.2.3.

3.2.1 Decorated Factor Graphs. Let $\mathcal{D}_\Psi = [\psi_\downarrow, \psi_\uparrow]^{[q]^k}$ be the domain of the weights and let $\mathcal{G}_{n,m} = ([n]^k \times \mathcal{D}_\Psi)^m$ be the domain of the graphs without external fields.

3.2.1.1 Random Decorated Graphs. The decorated factor graphs are given by

- a weight function $\boldsymbol{\psi}_\circ \in \mathcal{D}_\Psi$ with law μ_Ψ and expectation $\overline{\psi}_\circ = \mathbb{E}[\boldsymbol{\psi}_\circ]$,
- a ground truth distribution $\gamma^* \in \mathcal{P}([q])$ with $\gamma^* \geq \psi_\downarrow$,
- an average degree $\bar{d} \in [0, d_\uparrow]$ such that $(\mu_\Psi, \gamma^*, \bar{d}) \in \mathfrak{P} = \mathfrak{A} \times [0, d_\uparrow]$,
- a Gibbs marginal distribution $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])$,
- an interpolation time $t^\leftrightarrow \in [0,1]$,
- a pinning bound $\Theta^\downarrow \in \mathbb{R}_{\geq 0}$,
- a number $n \in \mathbb{Z}_{>\Theta^\downarrow}$ of variables,
- a number $m \in \mathbb{Z}_{\geq 0}$ of factors,
- a ground truth $\sigma \in [q]^n$,
- interpolator counts $m^\leftrightarrow \in \mathbb{Z}^n_{\geq 0}$ with $\mathcal{A}^\leftrightarrow_{m^\leftrightarrow} = \{i, h) : i \in [n], h \in [m^\leftrightarrow_i]\}$,
- pins $\mathcal{U} \subseteq [n]$ and
- a pinning assignment $\check{\sigma} \in [q]^n$,

which we will keep fixed throughout the remainder. For $G = (v, \psi) \in \mathcal{G}$ let $[G]^\Gamma_{\gamma^*} = G' = (v'_a, \psi'_a)_{a \in \mathcal{A}}$ be given by $\mathcal{A} = [m] \,\dot\cup\, [n]$, $G'_{[m]} = G$ and $(v'_a, \psi'_a) = (a, \gamma^*)$ for $a \in [n]$, i.e. we attach the unary weight γ^* to each variable. Similarly, for $\psi^\leftrightarrow \in \mathcal{D}_\Psi^{\leftrightarrow \mathcal{A}^\leftrightarrow}$, $\mathcal{D}^\leftrightarrow_\Psi = [\psi_\downarrow, \psi_\uparrow]^q$, let $[G]^\leftrightarrow_{m^\leftrightarrow, \psi^\leftrightarrow} = G' = (v'_a, \psi'_a)_{a \in \mathcal{A}}$ be given by $\mathcal{A} = [m] \,\dot\cup\, \mathcal{A}^\leftrightarrow$, $G'_{[m]} = G$ and $(v'_a, \psi'_a) = (i, \psi^\leftrightarrow_{i,h})$ for $a = (i,h) \in \mathcal{A}^\leftrightarrow$, i.e. to each variable $i \in [n]$ we attach $m^\leftrightarrow_i$ unary interpolation weights $\psi^\leftrightarrow_{i,h}$, $h \in [m^\leftrightarrow_i]$. Finally, the pinned graph is $[G]^\downarrow_{\mathcal{U},\check{\sigma}} = G' = (v'_a, \psi'_a)_{a \in \mathcal{A}}$ given by $\mathcal{A} = [m] \,\dot\cup\, \mathcal{U}$, $G'_{[m]} = G$ and the unary wires-weight pairs $(i, \mu_{\bullet,[q],\check{\sigma}(i)})$ for $i \in \mathcal{U}$.

For $G' = [G]^\Gamma$ we let $[G']^\leftrightarrow$ be the graph obtained from G' by attaching interpolators analogously to the above, and also define other combinations analogously. Further, we define the combined operators analogously, e.g. $[G]^{\Gamma\downarrow}$ attaches external fields and pins.

For the interpolation weight in the null model let $(\boldsymbol{\psi}_\circ, \boldsymbol{h}, \boldsymbol{\gamma}) \sim \mu_\Psi \otimes \mathrm{u}([k]) \otimes \pi^{\otimes k}$ and

$$\boldsymbol{\psi}^\leftrightarrow_{\circ,\mu_\Psi,\gamma^*,\pi} : [q] \to [\psi_\downarrow, \psi_\uparrow],\ \sigma \mapsto \sum_{\tau \in [q]^k} \mathbb{1}\{\tau_{\boldsymbol{h}} = \sigma\} \boldsymbol{\psi}_\circ(\tau) \prod_{h \neq \boldsymbol{h}} \boldsymbol{\gamma}(\tau_h). \tag{3}$$

With $\mathfrak{p}=(\mu_\Psi,\gamma^*,\pi,n,m,m^{\leftrightarrow},\mathcal{U})$ the null model $\boldsymbol{G}_{\mathfrak{p}}=[\boldsymbol{w}]^{\Gamma\leftrightarrow\downarrow}_{\boldsymbol{\psi}^{\leftrightarrow},\check{\boldsymbol{\sigma}}}$ is given by $(\boldsymbol{w},\boldsymbol{\psi}^{\leftrightarrow},\check{\boldsymbol{\sigma}})\sim\boldsymbol{w}\otimes\boldsymbol{\psi}^{\leftrightarrow}\otimes\check{\boldsymbol{\sigma}}$, $\boldsymbol{w}_{\mu_\Psi,n,m}=(\boldsymbol{v}_{n,m},\boldsymbol{\psi}_{\mu_\Psi,m})\sim\boldsymbol{w}_\circ^{\otimes m}$, $\boldsymbol{w}_{\circ,\mu_\Psi,n}=(\boldsymbol{v}_{\circ,n},\boldsymbol{\psi}_{\circ,\mu_\Psi})\sim\mathrm{u}([n]^k)\otimes\mu_\Psi$, $\boldsymbol{\psi}^{\leftrightarrow}_{\mu_\Psi,\gamma^*,\pi,n,m^{\leftrightarrow}}\sim\boldsymbol{\psi}_\circ^{\leftrightarrow\otimes\mathcal{A}^{\leftrightarrow}}$ and $\check{\boldsymbol{\sigma}}_n\sim\mathrm{u}([q]^n)$. The standard weight, Gibbs measure, partition function and free entropy density of $G=(v',\psi')=[(v,\psi)]^{\Gamma\leftrightarrow\downarrow}$ are

$$\psi_{\mathrm{g},G}(\sigma)=\prod_{a\in\mathcal{A}}\psi'_a(\sigma_{v'_a})=\gamma^{*\otimes n}(\sigma)\mathbb{1}\{\sigma_{\mathcal{U}}=\check{\sigma}_{\mathcal{U}}\}\prod_{a\in[m]}\psi_a(\sigma_{v(a)})\prod_{(i,h)\in\mathcal{A}^{\leftrightarrow}}\psi^{\leftrightarrow}_{i,h}(\sigma_i),$$

$$\mu_{\mathrm{g},G}(\sigma)=\frac{\psi_{\mathrm{g},G}(\sigma)}{Z_{\mathrm{g}}(G)},\ \boldsymbol{\sigma}_{\mathrm{g},G}\sim\mu_{\mathrm{g},G},\ Z_{\mathrm{g}}(G)=\sum_{\sigma\in[q]^n}\psi_{\mathrm{g},G}(\sigma),\ \phi_{\mathrm{g}}(G)=\frac{1}{n}\ln(Z_{\mathrm{g}}(G)).$$

Let $\overline{\psi}_{\circ,\mu_\Psi}(\sigma)=\mathbb{E}[\boldsymbol{\psi}_\circ(\sigma)]$, $\overline{\psi}_{\mathrm{m},\mathfrak{p}}(\sigma)=\mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}}(\sigma)]$, $\overline{Z}_{\mathrm{m},\mathfrak{p}}=\mathbb{E}[Z_{\mathrm{g}}(\boldsymbol{G})]$ and $\overline{\phi}_{\mathrm{m},\mathfrak{p}}=\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})]$, and let $\boldsymbol{G}^*_{\mathfrak{p}}(\sigma)$ be the teacher-student model given by the $(\boldsymbol{G}^*(\sigma),\boldsymbol{G})$-derivative $G\mapsto\psi_{\mathrm{g},G}(\sigma)/\overline{\psi}_{\mathrm{m}}(\sigma)$.

Now, let $\boldsymbol{\theta}_{\Theta\downarrow}\sim\mathrm{u}([0,\Theta^{\downarrow}])$, for $\theta\in\mathbb{R}_{\ge0}$ let $\check{\boldsymbol{u}}_{\mathrm{t}\circ,\theta,n}\in\{0,1\}$ be Bernoulli with success probability $\check{p}_{\mathrm{t},\theta,n}=\theta/n\in[0,1]$, $\check{\boldsymbol{u}}_{\mathrm{t},\theta,n}\sim\check{\boldsymbol{u}}_{\mathrm{t}\circ}^{\otimes n}$ and $\boldsymbol{\mathcal{U}}_{\Theta\downarrow,n}=\check{\boldsymbol{u}}^{-1}_{\mathrm{t},\boldsymbol{\theta}}(1)$. Let $\boldsymbol{m}_{\bar{d},t^{\leftrightarrow},n}\sim\mathrm{Po}(t^{\leftrightarrow}\bar{d}n/k)$, $\boldsymbol{m}^{\leftrightarrow}_{\circ,\bar{d},t^{\leftrightarrow}}\sim\mathrm{Po}((1-t^{\leftrightarrow})\bar{d})$, $\boldsymbol{m}^{\leftrightarrow}_{\bar{d},t^{\leftrightarrow},n}\sim\boldsymbol{m}_\circ^{\leftrightarrow\otimes n}$, and let the joint distribution be given by $(\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}})\sim\boldsymbol{m}\otimes\boldsymbol{m}^{\leftrightarrow}\otimes\boldsymbol{\mathcal{U}}$. Let $\bar{\boldsymbol{d}}_{\bar{d},t^{\leftrightarrow},n}=k\boldsymbol{m}/n$ be the average degree (with respect to the standard factors) and $\bar{\boldsymbol{d}}^{\leftrightarrow}_{\bar{d},t^{\leftrightarrow},n}=\|\boldsymbol{m}^{\leftrightarrow}\|_1/n$ be the average degree with respect to the interpolators. Finally, we consider $\boldsymbol{\sigma}^*_{\gamma^*,n}$ to be independent of $(\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}})$.

Remark 3.8. As discussed in Section 3.1.6, the change in notation is required due to the complexity. The expected average degree $\bar{d}$ corresponds to the expected degree d in Section 2.1.2.1 and to the limit d^* for $\boldsymbol{m}^*$ in Section 3.1.3, hence we let $\bar{d}=d=d^*$ in the remainder without further mention. Similarly, the law μ_Ψ is clearly p from Section 2.1.2.1, so we let $p=\mu_\Psi$ in the remainder without further mention.

3.2.1.2 Factor Assignment Distribution. Recall $\overline{Z}_{\mathrm{f},\mu_\Psi}$ and ξ_{μ_Ψ} from Section 2.1.2.2. For $\gamma\in\mathcal{P}([q])$ and $\tau\in[q]^k$ let $\mu_{\mathrm{T}|\Gamma,\mu_\Psi,\gamma}\in\mathcal{P}([q]^k)$ be given by

$$\mu_{\mathrm{T}|\Gamma,\gamma}(\tau)=\frac{1}{\overline{Z}_{\mathrm{f}}(\gamma)}\overline{\psi}_\circ(\tau)\prod_{h\in[k]}\gamma(\tau_h).\tag{4}$$

We will see that $\mu_{\mathrm{T}|\Gamma}$ is the law of the assignment to a factor induced by σ under $\boldsymbol{G}^*(\sigma)$. Further, it is clearly closely related to $\boldsymbol{\psi}^{\leftrightarrow}_\circ$.

Observation 3.9. *Let $\gamma\in\mathcal{P}([q])$ and notice that the following holds.*
a) We have $\psi_\downarrow\le\boldsymbol{\psi}_\circ,\overline{\psi}_\circ,\overline{Z}_{\mathrm{f}},\xi\le\psi_\uparrow$.
b) The map $\overline{Z}_{\mathrm{f}}$ is a q-variate polynomial of degree k on a compact set and hence attains ξ.
c) There exists $L_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $\overline{Z}_{\mathrm{f}}$ is L-Lipschitz.
d) There exists $c_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $\overline{Z}_{\mathrm{f}}(\gamma)\ge\xi-c\|\gamma-\gamma^*\|^2_{\mathrm{tv}}$.
e) There exists $L_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $\mu_{\mathrm{T}|\Gamma}:\mathcal{P}([q])\to\mathcal{P}([q]^k)$ is L-Lipschitz.
f) There exists $c_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $c\le r\le c^{-1}$ for the $(\mu_{\mathrm{T}|\Gamma,\gamma},\gamma^{\otimes k})$-derivative r_γ.
g) There exists $L_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $r:\mathcal{P}([q])\to\mathbb{R}^{[q]^k}_{>0}$, $\gamma\mapsto r_\gamma$, is L-Lipschitz.
h) There exists $c_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $c\le r_*\le c^{-1}$ for the $(\mu_{\mathrm{T}|\Gamma,\gamma}|_*,\gamma)$-derivative $r_{*,\gamma}$.
i) There exists $L_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $r_*:\mathcal{P}([q])\to\mathbb{R}^q_{>0}$, $\gamma\mapsto r_{*,\gamma}$, is L-Lipschitz.
j) We have $\mu_{\mathrm{T}|\Gamma,\gamma^*}|_*=\gamma^*$ and hence $r_{*,\gamma^*}\equiv1$.

Proof. Recall that $\boldsymbol{\psi}_\circ \in \mathcal{D}_\Psi$, $\overline{\psi}_\circ = \mathbb{E}[\boldsymbol{\psi}_\circ]$, $\overline{Z}_\mathrm{f}(\gamma) = \mathbb{E}[\overline{\psi}_\circ(\boldsymbol{\sigma}^*_{\gamma,k})]$ and $\xi = \sup_\gamma \overline{Z}_\mathrm{f}(\gamma)$ for Part 3.9a). For Part 3.9b) notice that $\overline{Z}_\mathrm{f}$ is the restriction of $f : \mathbb{R}^q \to \mathbb{R}$, $x \mapsto \sum_\tau \overline{\psi}_\circ(\tau)\prod_h x_{\tau(h)}$ to $\mathcal{P}([q]) \subseteq \mathbb{R}^q$. Since we need the derivatives anyway, notice that the τ-th partial derivative $f_\tau(x)$ of f for $\tau \in [q]$ at $x \in \mathbb{R}^q$, using the product rule, is given by

$$f_\tau(x) = \sum_h \sum_{\tau'} \overline{\psi}_\circ(\tau')\mathbb{1}\{\tau'_h = \tau\} \prod_{h' \neq h} x_{\tau'(h')}, \tag{5}$$

so $k\psi_\downarrow \le f_\tau(\gamma) \le k\psi_\uparrow$ for $\gamma \in \mathcal{P}([q])$. Now, the (one-dimensional) fundamental theorem of calculus ensures that

$$|\overline{Z}_\mathrm{f}(\gamma_1) - \overline{Z}_\mathrm{f}(\gamma_2)| \le k\psi_\uparrow\|\gamma_1 - \gamma_2\|_1 = 2k\psi_\uparrow\|\gamma_1 - \gamma_2\|_\mathrm{tv}$$

for $\gamma \in \mathcal{P}([q])^2$. For Part 3.9d) we compute the Hessian

$$H_{x,\tau} = \sum_{h \in [k]^2} \sum_{\tau'} \overline{\psi}_\circ(\tau')\mathbb{1}\{\tau'_h = \tau\} \prod_{h' \notin \{h_1,h_2\}} x_{h'},\ \tau \in [q]^2.$$

This yields $k(k-1)\psi_\downarrow \le H_\gamma \le k(k-1)\psi_\uparrow$. So, using that $\overline{Z}_\mathrm{f}(\gamma^*) = \xi$ is the maximum for γ^* in the interior, i.e. the first derivative vanishes, yields the assertion using the first order Taylor approximation with the Lagrange form of the remainder and $c = 2k(k-1)\psi_\uparrow$.

For Part 3.9e) we use the triangle inequality, boundedness and Lipschitz continuity of $\overline{Z}_\mathrm{f}$ and Observation 3.6 to obtain

$$\|\mu_{\mathrm{T}|\Gamma,\gamma_1} - \mu_{\mathrm{T}|\Gamma,\gamma_2}\|_\mathrm{tv} \le \frac{1}{2}\left|1 - \frac{\overline{Z}_\mathrm{f}(\gamma_1)}{\overline{Z}_\mathrm{f}(\gamma_2)}\right| + \psi_\uparrow^2\|\gamma_1^{\otimes k} - \gamma_1^{\otimes k}\|_\mathrm{tv} \le 2k\psi_\uparrow^2\|\gamma_1 - \gamma_2\|_\mathrm{tv}.$$

Part 3.9f) follows from Part 3.9a) with $c = \psi_\uparrow^2$, Part 3.9g) from Part 3.9a) and Part 3.9c) since

$$\|r_{\gamma_2} - r_{\gamma_1}\|_1 = \sum_\tau \frac{\overline{\psi}_\circ(\tau)}{\overline{Z}_\mathrm{f}(\gamma_1)\overline{Z}_\mathrm{f}(\gamma_2)}|\overline{Z}_\mathrm{f}(\gamma_1) - \overline{Z}_\mathrm{f}(\gamma_2)| \le 2kq^k\psi_\uparrow^4\|\gamma_2 - \gamma_1\|_\mathrm{tv},\ \gamma \in \mathcal{P}([q]).$$

Part 3.9h) follows from Part 3.9f) and $\gamma^{\otimes k}|_* = \gamma$. For Part 3.9i) we use the triangle inequality and Part 3.9a) to get

$$\|r_{*,\gamma_2} - r_{*,\gamma_1}\|_1 \le \sum_{\tau,h} \frac{\psi_\uparrow}{k} \sum_{\tau' \in [q]^k} \mathbb{1}\{\tau'_h = \tau\} \left| \frac{\gamma_2^{\otimes [k]\setminus\{h\}}\left(\tau'_{[k]\setminus\{h\}}\right)}{\overline{Z}_\mathrm{f}(\gamma_2)} - \frac{\gamma_1^{\otimes [k]\setminus\{h\}}\left(\tau'_{[k]\setminus\{h\}}\right)}{\overline{Z}_\mathrm{f}(\gamma_1)} \right|.$$

Relabeling, the triangle inequality and Observation 3.6b) yield

$$\|r_{*,\gamma_2} - r_{*,\gamma_1}\|_1 \le q\psi_\uparrow\left(\psi_\uparrow(k-1)\|\gamma_2 - \gamma_1\|_\mathrm{tv} + \psi_\uparrow^2\left|\overline{Z}_\mathrm{f}(\gamma_1) - \overline{Z}_\mathrm{f}(\gamma_2)\right|\right),$$

so Part 3.9c) completes the proof. Finally, for Part 3.9j) we recall the partial derivative $f_\tau(\gamma^*) = k\overline{Z}_\mathrm{f}(\gamma^*)\mu_{\mathrm{T}|\Gamma,\gamma^*}|_*(\tau)/\gamma^*(\tau) = k\overline{Z}_\mathrm{f}(\gamma^*)r_{*,\gamma^*}(\tau)$ from Equation (5) and that γ^* is a maximizer, so the derivatives in the directions $\mu_{\bullet,[q],\tau} - \mu_{\bullet,[q],q}$, $\tau \in [q-1]$, vanish and hence $r_{*,\gamma^*}(\tau) = r_{*,\gamma^*}(q)$ for all τ, which completes the proof. □

Remark 3.10. Let $\gamma \in \mathcal{P}([q])$ be a stationary point of $\overline{Z}_\mathrm{f}$ if and only if γ is a stationary point of the restriction $\overline{Z}_{\mathrm{f}}:\mathcal{P}(\gamma^{-1}(\mathbb{R}_{>0}))\to\mathbb{R}_{>0}$, $\gamma'\mapsto\overline{Z}_{\mathrm{f}}(\gamma')$, or $|\gamma^{-1}(\mathbb{R}_{>0})|=1$. In the proof of Observation 3.9j), we have actually proven that $\mu_{\mathrm{T}|\Gamma,\gamma^*}|_*=\gamma^*$ if and only if γ^* is a stationary point of $\overline{Z}_{\mathrm{f}}$.

3.2.1.3 Expectations and Bounds. We derive naive bounds and compute the expectations, which ensures that the teacher-student model is well-defined since $\overline{\psi}_{\mathrm{m}}>0$ and ϕ_{g}, μ_{g} are well-defined since $Z_{\mathrm{g}}>0$.

Observation 3.11. *Let $M=m+\|m^{\leftrightarrow}\|_1$.*
a) We have $\mathbb{E}[\boldsymbol{\psi}_\circ^{\leftrightarrow}]\equiv\xi$ and $\psi_\downarrow\le\boldsymbol{\psi}_\circ^{\leftrightarrow}\le\psi_\uparrow$.
b) We have $\psi_\downarrow^M\mathbb{1}\{\sigma_{\mathcal{U}}=\check{\sigma}_{\mathcal{U}}\}\gamma^{*\otimes n}(\sigma)\le\psi_{\mathrm{g},G}(\sigma)\le\psi_\uparrow^M\mathbb{1}\{\sigma_{\mathcal{U}}=\check{\sigma}_{\mathcal{U}}\}\gamma^{*\otimes n}(\sigma)$ for $G\in[\mathcal{G}]^{\Gamma\leftrightarrow\downarrow}$.
c) We have $\psi_\downarrow^M\gamma^{*\otimes\mathcal{U}}(\check{\sigma}_{\mathcal{U}})\le Z_{\mathrm{g}}(G)\le\psi_\uparrow^M\gamma^{*\otimes\mathcal{U}}(\check{\sigma}_{\mathcal{U}})$ for $G\in[\mathcal{G}]^{\Gamma\leftrightarrow\downarrow}$.
d) We have $\psi_\downarrow^{2M}\mathbb{P}(\boldsymbol{\sigma}^*=\sigma|\boldsymbol{\sigma}^*_{\mathcal{U}}=\check{\sigma}_{\mathcal{U}})\le\mu_{\mathrm{g},G}(\sigma)\le\psi_\uparrow^{2M}\mathbb{P}(\boldsymbol{\sigma}^*=\sigma|\boldsymbol{\sigma}^*_{\mathcal{U}}=\check{\sigma}_{\mathcal{U}})$ for $G\in[\mathcal{G}]^{\Gamma\leftrightarrow\downarrow}$.
e) We have $\overline{\psi}_{\mathrm{m}}(\sigma)=q^{-|\mathcal{U}|}\xi^{\|m^{\leftrightarrow}\|_1}\gamma^{*\otimes n}(\sigma)\overline{Z}_{\mathrm{f}}(\gamma_{\mathrm{n},\sigma})^m$.
f) We have $\overline{Z}_{\mathrm{m}}=q^{-|\mathcal{U}|}\xi^{\|m^{\leftrightarrow}\|_1}\mathbb{E}[\overline{Z}_{\mathrm{f}}(\gamma_{\mathrm{n},\boldsymbol{\sigma}^*})^m]$.

Proof. For Part 3.11a) we use independence, further $\pi\in\mathcal{P}^2_{*,\gamma^*}([q])$ for the expectations, $\overline{Z}_{\mathrm{f}}(\gamma^*)=\xi$ for the normalization and Observation 3.9j) to obtain

$$\mathbb{E}[\boldsymbol{\psi}_\circ^{\leftrightarrow}(\tau)]=\sum_h\frac{1}{k}\sum_{\tau'}\mathbb{1}\{\tau'_h=\tau\}\overline{\psi}_\circ(\tau')\prod_{h'\neq h}\gamma^*(\tau'_{h'})=\xi r_{*,\gamma^*}(\tau)=\xi,\tau\in[q],$$

while $\boldsymbol{\psi}_\circ^{\leftrightarrow}\in\mathcal{D}_\Psi^{\leftrightarrow}$ is immediate from Observation 3.9a). Part 3.11b) follows with Part 3.11a) and Observation 3.9a), Part 3.11c) follows with Part 3.11b), Part 3.11d) follows with Part 3.11b) and Part 3.11c). Part 3.11e) follows with independence, Part 3.11a) and

$$\mathbb{E}[\psi_{\mathrm{g},\boldsymbol{w}}(\sigma)]=\mathbb{E}[\boldsymbol{\psi}_\circ(\sigma_{\boldsymbol{v}_\circ})]^m=\left(\sum_\tau\overline{\psi}_\circ(\tau)\sum_v\frac{1}{n^k}\mathbb{1}\{\sigma_v=\tau\}\right)^m=\overline{Z}_{\mathrm{f}}(\gamma_{\mathrm{n},\sigma})^m.$$

Part 3.11f) follows from Part 3.11e) by summing over σ. □

Next, we notice that $\boldsymbol{m}^*$ (cf. Section 3.1.3) can be chosen to be $\boldsymbol{m}$, for very restrictive bounds.

Corollary 3.12. *Let $r\in\mathbb{R}_{\ge0}$, $\mathcal{B}^\circ=(t^{\leftrightarrow}\bar{d}-r,t^{\leftrightarrow}\bar{d}+r)$ and $\mathcal{B}^{\leftrightarrow}=((1-t^{\leftrightarrow})\bar{d}-r,(1-t^{\leftrightarrow})\bar{d}+r)$. We have $\|\boldsymbol{m}^{\leftrightarrow}\|_1\sim\mathrm{Po}((1-t^{\leftrightarrow})\bar{d}n)$. Further, there exists $c_{\mathfrak{g}}\in\mathbb{R}^2_{>0}$ such that*

$$\mathbb{P}(\bar{\boldsymbol{d}}\notin\mathcal{B}^\circ),\,\mathbb{P}(\bar{\boldsymbol{d}}^{\leftrightarrow}\notin\mathcal{B}^{\leftrightarrow}),\,\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}\notin\mathcal{B}^\circ\}\bar{\boldsymbol{d}}],\,\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}^{\leftrightarrow}\notin\mathcal{B}^{\leftrightarrow}\}\bar{\boldsymbol{d}}^{\leftrightarrow}]\le c_2\exp\left(-\frac{c_1r^2n}{1+r}\right).$$

Proof. We have $\|\boldsymbol{m}^{\leftrightarrow}\|_1\sim\mathrm{Po}((1-t^{\leftrightarrow})\bar{d}n)$ by Observation 3.7b). Further, we have

$$\mathbb{P}(|\bar{\boldsymbol{d}}-t^{\leftrightarrow}\bar{d}|\ge r)=\mathbb{P}\left(\left|\boldsymbol{m}-\frac{t^{\leftrightarrow}\bar{d}n}{k}\right|\ge\frac{rn}{k}\right)\le c_2\exp\left(-\frac{c_1nr^2}{k(t^{\leftrightarrow}\bar{d}+r)}\right)$$

with $c\in\mathbb{R}^2_{>0}$ from Observation 3.7d), and hence the first part follows with $c_1/(kd_\uparrow)$, using $t^{\leftrightarrow}\le1$ and $d_\uparrow\ge1$. For the third part, fix some large $\rho\in\mathbb{R}_{>0}$. For $r\le\rho/\sqrt{n}$, $c_1>0$ notice that $\exp(-c_1r^2n/(1+r))\ge e^{-c_1\rho^2}$, so for $c_2\ge d_\uparrow e^{c_1\rho^2}$ we have

$$\mathbb{E}[\mathbb{1}\{|\bar{\boldsymbol{d}}-t^{\leftrightarrow}\bar{d}|\ge r\}\bar{\boldsymbol{d}}]\le\mathbb{E}[\bar{\boldsymbol{d}}]=t^{\leftrightarrow}\bar{d}\le d_\uparrow\le c_2\exp\left(-\frac{c_1r^2n}{1+r}\right).$$

So, let $r \geq \rho/\sqrt{n}$. Using Observation 3.7a) and the triangle inequality yields

$$\mathbb{E}[\mathbb{1}\{|\bar{\boldsymbol{d}} - t^{\leftrightarrow}\bar{d}| \geq r\}\bar{\boldsymbol{d}}] = t^{\leftrightarrow}\bar{d}\mathbb{P}\left(\left|\boldsymbol{m} + 1 - \frac{t^{\leftrightarrow}\bar{d}n}{k}\right| \geq \frac{rn}{k}\right) \leq d_{\uparrow}\mathbb{P}\left(\left|\boldsymbol{m} - \frac{t^{\leftrightarrow}\bar{d}n}{k}\right| \geq \frac{rn-k}{k}\right)$$

and $rn - k \geq \frac{1}{2}rn + \frac{1}{2}\rho - k \geq \frac{1}{2}rn$ using $n \geq 1$, so the first part completes the proof, since the results for $\bar{\boldsymbol{d}}^{\leftrightarrow}$ follow analogously. □

3.2.1.4 Independent Factors. Let $\gamma = \gamma_{\mathrm{n},\sigma}$. Further, let the teacher-student model wires-weight pair $\boldsymbol{w}^*_{\circ,\mu_\Psi,n,\sigma}$ be given by the $(\boldsymbol{w}^*_\circ, \boldsymbol{w}_\circ)$-derivative $(v,\psi) \mapsto \psi(\sigma_v)/\overline{Z}_{\mathrm{f}}(\gamma)$. For $\tau \in [q]$ the interpolation weight $\boldsymbol{\psi}^{*\leftrightarrow}_{\circ,\mu_\Psi,\gamma^*,\pi,\tau}$ is given by the $(\boldsymbol{\psi}^{*\leftrightarrow}_\circ, \boldsymbol{\psi}^{\leftrightarrow}_\circ)$-derivative $\psi \mapsto \psi(\tau)/\xi$. Finally, let

$$(\boldsymbol{w}^*_{\mu_\Psi,n,m,\sigma}, \boldsymbol{\psi}^{*\leftrightarrow}_{\mu_\Psi,\gamma^*,\pi,n,m^{\leftrightarrow},\sigma}) \sim \boldsymbol{w}^{*\otimes m}_\circ \otimes \bigotimes_{i\in[n]} \boldsymbol{\psi}^{*\leftrightarrow\otimes m^{\leftrightarrow}_i}_{\circ,\sigma(i)}.$$

Observation 3.13. *We have $\boldsymbol{G}^*(\sigma) \sim [\boldsymbol{w}^*_\sigma]^{\Gamma\leftrightarrow\downarrow}_{\boldsymbol{a}}$ with $\boldsymbol{a} = (\boldsymbol{\psi}^{*\leftrightarrow}_\sigma, \mathcal{U}, \sigma)$.*

Proof. Let $\gamma = \gamma_{\mathrm{n},\sigma}$ and $r(v,\psi) = \psi(\sigma_v)/\overline{Z}_{\mathrm{f}}(\gamma)$, $r^{\leftrightarrow}_\tau(\psi) = \psi(\tau)/\xi$ denote the derivatives. Observation 3.11e) shows that $\gamma^{*\otimes n}(\sigma)$ cancels out in the $(\boldsymbol{G}^*(\sigma), \boldsymbol{G})$-derivative and further

$$\mathbb{P}(\boldsymbol{G}^*(\sigma) \in \mathcal{E}) = \mathbb{E}\left[q^{|\mathcal{U}|}\mathbb{1}\{\sigma_{\mathcal{U}} = \check{\boldsymbol{\sigma}}_{\mathcal{U}}\}\prod_a r(\boldsymbol{w}_a)\prod_{i,h} r^{\leftrightarrow}_{\sigma(i)}(\boldsymbol{\psi}^{\leftrightarrow}_{i,h})\mathbb{1}\left\{[\boldsymbol{w}]^{\Gamma\leftrightarrow\downarrow}_{\boldsymbol{\psi}^{\leftrightarrow},\mathcal{U},\check{\boldsymbol{\sigma}}} \in \mathcal{E}\right\}\right].$$

Recall that $[\cdot]^{\downarrow}_{\mathcal{U},\check{\sigma}}$ only depends on $\check{\sigma}$ through the values $\check{\sigma}_{\mathcal{U}}$ to be pinned, so on the event $\sigma_{\mathcal{U}} = \check{\boldsymbol{\sigma}}_{\mathcal{U}}$ we have $[\boldsymbol{w}]^{\Gamma\leftrightarrow\downarrow}_{\boldsymbol{\psi}^{\leftrightarrow},\mathcal{U},\check{\boldsymbol{\sigma}}} = [\boldsymbol{w}]^{\Gamma\leftrightarrow\downarrow}_{\boldsymbol{\psi}^{\leftrightarrow},\mathcal{U},\sigma}$. After this substitution we can take the expectation over $\check{\boldsymbol{\sigma}}$ due to independence, i.e. $\mathbb{E}[q^{|\mathcal{U}|}\mathbb{1}\{\sigma_{\mathcal{U}} = \check{\boldsymbol{\sigma}}_{\mathcal{U}}\}] = 1$. This completes the proof, due to independence. □

Remark 3.14. Observation 3.13 allows to discuss the standard graph $\boldsymbol{w}^*$, the interpolators $\boldsymbol{\psi}^{*\leftrightarrow}_\sigma$ and the pins separately in most situations. We will make use of this convenient feature to reduce the (notational) complexity and increase the transparency. For example, in the following we will discuss the law of the standard graph, further notions and properties. This discussion directly applies to $\boldsymbol{w}^*$ and in this sense to $\boldsymbol{G}^*(\sigma)$.

3.2.1.5 Factor Side Assignments. For $G = (v,\psi) \in \mathcal{G}$ let $\tau_{\mathrm{g},G,\sigma} = (\sigma_{v(a)})_{a\in[m]}$ be the assignment to the factors induced by σ under G, and $\boldsymbol{\tau}^*_{\mu_\Psi,n,m,\sigma} = \tau_{\mathrm{g},\boldsymbol{G}^*(\sigma),\sigma} = \tau_{\mathrm{g},\boldsymbol{w}^*(\sigma),\sigma}$ the induced ground truth factor assignment. Using $\gamma = \gamma_{\mathrm{n},\sigma}$, let $\mathcal{D}_{\Gamma,\gamma} = \gamma^{-1}(\mathbb{R}_{>0}) \subseteq [q]$ and notice that $\tau_{\mathrm{g},G,\sigma} \in (\mathcal{D}^k_{\Gamma,\gamma})^m$.

On the other hand, let $\boldsymbol{\tau}^*_{\circ,\mu_\Psi,\gamma} \sim \mu_{\mathrm{T}|\Gamma,\gamma}$ with $\mu_{\mathrm{T}|\Gamma,\gamma}$ from Section 3.2.1.2 and notice that the support of $\boldsymbol{\tau}^*_{\circ,\gamma}$ is $\mathcal{D}^k_{\Gamma,\gamma}$ by Observation 3.9f). For $\tau \in \mathcal{D}^k_{\Gamma,\gamma}$ let $\boldsymbol{w}^\star_{\circ,\mu_\Psi,n,\sigma,\tau} = (\boldsymbol{v}^\star_\circ, \boldsymbol{\psi}^\star_\circ) \sim \boldsymbol{v}^\star_\circ \otimes \boldsymbol{\psi}^\star_\circ$, where $\boldsymbol{v}^\star_{\circ,n,\sigma,\tau} \sim \bigotimes_h \mathrm{u}(\sigma^{-1}(\tau_h))$ and $\boldsymbol{\psi}^\star_{\circ,\mu_\Psi,\tau}$ is given by the $(\boldsymbol{\psi}^\star_\circ, \boldsymbol{\psi}_\circ)$-derivative $\psi \mapsto \psi(\tau)/\overline{\psi}_\circ(\tau)$. Finally, for $\tau \in (\mathcal{D}^k_{\Gamma,\gamma})^m$ let $\boldsymbol{w}^\star_{\mu_\Psi,n,m}(\sigma,\tau) \sim \bigotimes_{a\in[m]} \boldsymbol{w}^\star_{\circ,\tau(a)}$.

Observation 3.15. *We have $(\boldsymbol{\tau}^*_\sigma, \boldsymbol{w}^*(\sigma)) \sim (\boldsymbol{\tau}, \boldsymbol{w}^\star(\sigma,\boldsymbol{\tau}))$, $\boldsymbol{\tau} \sim \boldsymbol{\tau}^{*\otimes m}_{\circ,\gamma}$, $\gamma = \gamma_{\mathrm{n},\sigma}$.*

Proof. We restrict to $(\boldsymbol{\tau}^*_\circ, \boldsymbol{w}^\star_{\circ,\boldsymbol{\tau}^*_\circ})$ and $(\sigma_{\boldsymbol{v}^*_\circ}, \boldsymbol{w}^*_\circ)$ using independence. Now, the assertion holds since

$$\mathbb{P}((\boldsymbol{\tau}^*_\circ, \boldsymbol{w}^\star_\circ(\boldsymbol{\tau}^*_\circ)) \in \mathcal{E}) = \sum_{\tau,v}\frac{\mathbb{1}\{\sigma_v = \tau\}}{\overline{Z}_{\mathrm{f}}(\gamma)n^k}\mathbb{E}\left[\boldsymbol{\psi}_\circ(\tau)\mathbb{1}\{(\tau,v,\boldsymbol{\psi}_\circ) \in \mathcal{E}\}\right] = \mathbb{P}((\sigma_{\boldsymbol{v}^*_\circ}, \boldsymbol{w}^*_\circ) \in \mathcal{E}).$$

□

In words, the law of $\boldsymbol{\tau}^*(\sigma)$ factorizes and $\boldsymbol{w}^*(\sigma)$ conditional to $\boldsymbol{\tau}^*(\sigma)=\tau$ is given by $\boldsymbol{w}^\star(\sigma,\tau)$. Let $\boldsymbol{G}^\star(\sigma,\tau)$ be given by the law of $\boldsymbol{G}^*(\sigma)|\boldsymbol{\tau}^*_\sigma=\tau$.

3.2.1.6 Variable Degrees. For $\tau\in([q]^k)^m$ and $\tau'\in[q]^k$ let

$$\alpha_{\mathrm{m},\tau}(\tau')=\frac{|\tau^{-1}(\tau')|}{m}=\frac{1}{m}\left|\{a\in[m]:\tau_a=\tau'\}\right|. \tag{6}$$

Notice that α_{m} is not well-defined for $m=0$, but $m\alpha_{\mathrm{m}}$ is. We reserve the preimage notation $\tau^{-1}(\tau')$ for $\tau'\in[q]^k$ (as opposed to $\tau'\in[q]$). For $G=(v,\psi)\in\mathcal{G}$ let $\mathcal{A}_{\mathrm{v},G}(i)=\{a\in[m]:i\in v_a([k])\}$ be the (factor) neighborhood of $i\in[n]$ and $d_{\mathrm{f},G}(i)=|\mathcal{A}_{\mathrm{v},G}(i)|$ the (factor) degree.

Similarly, let $\mathcal{H}_{\mathrm{v},G}(i)=\{(a,h):v_a(h)=i\}$ be the (wire) neighborhood of i and $d_{\mathrm{w},G}(i)=|\mathcal{H}_{\mathrm{v},G}(i)|$ the (wire) degree. Finally, let $p_{\mathrm{d},\mu_\Psi,n,\sigma}(\sigma_i)=\mathbb{P}(i\in\boldsymbol{v}^*_{\circ,\sigma}([k]))$.

Observation 3.16. *Let $i\in[n]$, $\gamma=\gamma_{\mathrm{n},\sigma}$, $\mu=\mu_{\mathrm{T}|\Gamma,\gamma}$ and $\eta=\mathbb{E}[|\boldsymbol{v}^{*-1}_{\circ,\sigma}(i)|]$.*
a) We have $p_{\mathrm{d}}=1$ if $n=1$ and $p_{\mathrm{d}}\in(0,1)$ otherwise.
b) We have $\eta=\frac{k\mu|_*(\sigma_i)}{n\gamma(\sigma_i)}$.
c) There exists $c_{\mathfrak{g}}\in\mathbb{R}_{>0}$ with $\frac{1}{cn}\le p_{\mathrm{d}}(\sigma_i)\le\eta\le\frac{c}{n}$ and $p_{\mathrm{d}}(\sigma_i)\ge\eta-\frac{c}{n^2}$.

Proof. With Observation 3.15 we obtain p_{d} and η, i.e.

$$p_{\mathrm{d}}=\mathbb{P}(i\in\boldsymbol{v}^\star_{\circ,\boldsymbol{\tau}^*_\circ}([k]))=\mathbb{P}(\exists h\in[k]\,\boldsymbol{\tau}^*_{\circ,h}=\sigma_i,\boldsymbol{v}^\star_{\circ,\boldsymbol{\tau}^*_\circ}(h)=i)=\frac{\mathbb{P}(\boldsymbol{\tau}^*_\circ\notin([q]\setminus\{\sigma_i\})^k)}{n\gamma(\sigma_i)},$$
$$\eta=\sum_h\mathbb{P}(\boldsymbol{v}^*_{\circ,h}=i)=\sum_h\mathbb{P}(\boldsymbol{v}^\star_{\circ,\boldsymbol{\tau}^*_\circ,h}=i)=\frac{k\mu|_*(\sigma_i)}{n\gamma(\sigma_i)}.$$

With the union bound we have $p_{\mathrm{d}}(\sigma_i)\le\eta$ and with c from Observation 3.9h) further $\eta\le kc/n$. For $n=1$ we clearly have $p_{\mathrm{d}}=1$, $\eta=k$ and hence both lower bounds hold for $c\ge k-1$. For $n>1$ we have $n-1\ge n/2$ and hence

$$\begin{aligned}p_{\mathrm{d}}&=\mathbb{E}\left[\frac{\boldsymbol{\psi}_\circ(\sigma_{\boldsymbol{v}_\circ})}{\overline{Z}_{\mathrm{f}}(\gamma)}\mathbb{1}\left\{i\in\boldsymbol{v}_\circ([k])\right\}\right]\ge\psi_\downarrow^2\left(1-\frac{(n-1)^k}{n^k}\right)=\frac{\psi_\downarrow^2}{n^k}\sum_{\ell=0}^{k-1}\binom{k}{\ell}(n-1)^\ell\\&\ge\frac{k\psi_\downarrow^2(n-1)^{k-1}}{n^k}\ge\frac{k\psi_\downarrow^2}{2^{k-1}n}.\end{aligned}$$

This also shows that $p_{\mathrm{d}}\in(0,1)$ for $n>1$. The upper bound on the derivative gives

$$\begin{aligned}\eta-p_{\mathrm{d}}(\sigma_i)&\le\psi_\uparrow^2\mathbb{E}\left[|\boldsymbol{v}_\circ^{-1}(i)|-\mathbb{1}\{i\in\boldsymbol{v}_\circ([k])\}\right]=\frac{\psi_\uparrow^2}{n^k}\left(kn^{k-1}-(n^k-(n-1)^k)\right)\\&=\frac{\psi_\uparrow^2}{n^k}\left(k\sum_{\ell=0}^{k-1}\binom{k-1}{\ell}(n-1)^\ell-\sum_{\ell=0}^{k-1}\binom{k}{\ell}(n-1)^\ell\right)\\&=\frac{\psi_\uparrow^2}{n^k}\sum_{\ell=0}^{k-1}(k-\ell-1)\binom{k}{\ell}(n-1)^\ell\le\frac{\psi_\uparrow^2}{n^2}\sum_{\ell=0}^{k-2}(k-\ell-1)\binom{k}{\ell}.\end{aligned}$$

□

Next, we apply the bounds for the success probabilities to the degrees. Let $\boldsymbol{d}^\star_{\mathrm{w},n,m,\sigma,\tau}(i) \sim \mathrm{Bin}(km\alpha_{\mathrm{m},\tau}|_*(\sigma_i), \frac{1}{n\gamma_{\mathrm{n},\sigma}(\sigma_i)})$ and $\boldsymbol{d}^*_{\mathrm{f},\mu_\Psi,n,m,\sigma}(i) = \mathrm{Bin}(m, p_{\mathrm{d}}(\sigma_i))$. Notice that both degrees only depend on i through σ_i.

Observation 3.17. *Notice that the following holds for $i \in [n]$.*
a) We have $d_{\mathrm{w},\boldsymbol{w}^\star(\sigma,\tau)}(i) \sim \boldsymbol{d}^\star_{\mathrm{w}}(i)$ and $\mathbb{E}[d_{\mathrm{w},\boldsymbol{w}^*(\sigma)}(i)] = m\mathbb{E}[|\boldsymbol{v}^{*-1}_{\circ,\sigma}(i)|]$.
b) We have $d_{\mathrm{f},\boldsymbol{w}^*(\sigma)}(i) \sim \boldsymbol{d}^*_{\mathrm{f}}(i)$ and $\mathbb{E}[d_{\mathrm{f},\boldsymbol{w}^*(\sigma)}(i)] = mp_{\mathrm{d},\sigma}(\sigma_i)$.

Proof. Notice that

$$km\alpha_{\mathrm{m},\tau}|_*(\sigma_i) = \sum_h \sum_{\tau'} \mathbb{1}\{\tau'_h = \sigma_i\} \sum_a \mathbb{1}\{\tau_a = \tau'\} = |\{(a,h) : \tau_{a,h} = \sigma_i\}|.$$

Hence, Part 3.17a) and Part 3.17b) follow from Observation 3.15 and Observation 3.13. □

By an abuse of notation we use the shorthands $\boldsymbol{d}^\star_{\mathrm{w}}(i) = d_{\mathrm{w},\boldsymbol{w}^\star(\sigma,\tau)}(i)$ and $\boldsymbol{d}^*_{\mathrm{f}}(i) = d_{\mathrm{f},\boldsymbol{w}^*(\sigma)}(i)$. Combining these observations does not only yield uniform bounds (also in the choice of σ!), we also obtain the law of the degrees and bounds under the Poisson number of factors, and uniform Lipschitz continuity of the degree in σ.

Corollary 3.18. *Let $i \in [n]$, $\gamma = \gamma_{\mathrm{n},\sigma}$, $\mu = \mu_{\mathrm{T}|\Gamma,\gamma}$, $m\alpha = m\alpha_{\mathrm{m},\tau}$ and $\overline{m} = t^{\leftrightarrow}\bar{d}n/k$.*
a) We have $\boldsymbol{d}^*_{\mathrm{f}} \le \boldsymbol{d}^*_{\mathrm{w}}$. Further, there exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $\frac{km}{cn} \le \mathbb{E}[\boldsymbol{d}^*_{\mathrm{f}}] \le \mathbb{E}[\boldsymbol{d}^*_{\mathrm{w}}] \le \frac{ckm}{n}$ and $\mathbb{E}[\boldsymbol{d}^*_{\mathrm{w}}] - \mathbb{E}[\boldsymbol{d}^*_{\mathrm{f}}] \le \frac{ckm}{n^2}$.
b) We have $(\boldsymbol{d}^*_{\mathrm{f},\boldsymbol{m}}, \boldsymbol{m} - \boldsymbol{d}^*_{\mathrm{f},\boldsymbol{m}}) \sim \mathrm{Po}(p_{\mathrm{d}}\overline{m}, (1-p_{\mathrm{d}})\overline{m})$. Further, there exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $\frac{t^{\leftrightarrow}\bar{d}}{c} \le \mathbb{E}[\boldsymbol{d}^*_{\mathrm{f},\boldsymbol{m}}] \le \mathbb{E}[\boldsymbol{d}^*_{\mathrm{w},\boldsymbol{m}}] \le ct^{\leftrightarrow}\bar{d}$ and $\mathbb{E}[\boldsymbol{d}^*_{\mathrm{w},\boldsymbol{m}}] - \mathbb{E}[\boldsymbol{d}^*_{\mathrm{f},\boldsymbol{m}}] \le \frac{ct^{\leftrightarrow}\bar{d}}{n}$.
c) Using $\delta = \|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}$, there exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that

$$\left|\mathbb{E}[\boldsymbol{d}^*_{\mathrm{w},m}] - \frac{km}{n}\right| \le \frac{ckm}{n}\delta,\ \left|\mathbb{E}[\boldsymbol{d}^*_{\mathrm{w},\boldsymbol{m}}] - t^{\leftrightarrow}\bar{d}\right| \le ct^{\leftrightarrow}\bar{d}\delta,$$
$$\left|\mathbb{E}[\boldsymbol{d}^*_{\mathrm{f},m}] - \frac{km}{n}\right| \le \frac{ckm}{n}\left(\delta + \frac{1}{n}\right),\ \left|\mathbb{E}[\boldsymbol{d}^*_{\mathrm{f},\boldsymbol{m}}] - t^{\leftrightarrow}\bar{d}\right| \le ct^{\leftrightarrow}\bar{d}\left(\delta + \frac{1}{n}\right).$$

Proof. Using $\boldsymbol{w}^* = (\boldsymbol{v}^*, \boldsymbol{\psi}^*)$ and $\eta = \mathbb{E}[|\boldsymbol{v}^{*-1}_\circ(i)|]$, for Part 3.18a) we have

$$\boldsymbol{d}^*_{\mathrm{f}} = \sum_{a\in[m]} \mathbb{1}\{\exists h \in [k]\ \boldsymbol{v}^*_{a,h} = i\} \le \sum_{a\in[m]} \left|\boldsymbol{v}^{*-1}_a(i)\right|, = \boldsymbol{d}^*_{\mathrm{w}},$$

further we have $\mathbb{E}[\boldsymbol{d}^*_{\mathrm{f}}] = mp_{\mathrm{d}}$ and $\mathbb{E}[\boldsymbol{d}^*_{\mathrm{w}}] = m\eta$ from Observation 3.17, hence the bounds follow from Observation 3.16 by rescaling with $k \ge 2$. The law in Part 3.18b) follows from Observation 3.17b) and Observation 3.7b). The bounds are obtained by taking expectations in Part 3.18a). For Part 3.18c) we use Observation 3.16b), Observation 3.9j) and Observation 3.9i) to obtain

$$\left|\mathbb{E}[\boldsymbol{d}^*_{\mathrm{w},m}] - \frac{km}{n}\right| = \frac{km}{n}\left|\frac{\mu_{\mathrm{T}|\Gamma,\gamma_{\mathrm{n},\sigma}}|_*(\sigma_i)}{\gamma_{\mathrm{n},\sigma}(\sigma_i)} - \frac{\mu_{\mathrm{T}|\Gamma,\gamma^*}|_*(\sigma_i)}{\gamma^*(\sigma_i)}\right| \le \frac{ckm}{n}\delta.$$

The remainder is now immediate from Jensen's inequality and Part 3.18a). □

3.2.1.7 Neighborhood Decomposition. Let $\mathcal{D}_- = ([n] \setminus \{i\})^k$ be the factor neighborhoods excluding i and $\mathcal{D}_+ = [n]^k \setminus \mathcal{D}_-$ the neighborhoods covering i. For $n > 1$ let $\boldsymbol{w}_{-\circ,\mu_\Psi,n,i} = (\boldsymbol{v}_{-\circ,n,i}, \boldsymbol{\psi}_{-\circ,\mu_\Psi}) \sim$

$\mathrm{u}(\mathcal{D}_-)\otimes\mu_\Psi$ and let $\boldsymbol{w}^*_{-\circ,\mu_\Psi,n,i,\sigma}$ be given by the $(\boldsymbol{w}^*_{-\circ},\boldsymbol{w}_{-\circ})$-derivative $(v,\psi)\mapsto\psi(\sigma_v)/\mathbb{E}[\boldsymbol{\psi}_{-\circ}(\sigma_{\boldsymbol{v}_{-\circ}})]$. Let $\boldsymbol{w}_{+\circ,\mu_\Psi,n,i}=(\boldsymbol{v}_{+\circ,n,i},\boldsymbol{\psi}_{+\circ,\mu_\Psi})\sim\mathrm{u}(\mathcal{D}_+)\otimes\mu_\Psi$ and let $\boldsymbol{w}^*_{+\circ,\mu_\Psi,n,i,\sigma}$ be given by the $(\boldsymbol{w}^*_{+\circ},\boldsymbol{w}_{+\circ})$-derivative $(v,\psi)\mapsto\psi(\sigma_v)/\mathbb{E}[\boldsymbol{\psi}_{+\circ}(\sigma_{\boldsymbol{v}_{+\circ}})]$. For given $d\in\mathbb{Z}\cap[0,m]$ let

$$(\boldsymbol{w}^*_{-,\mu_\Psi,n,m-d,i,\sigma},\boldsymbol{w}^*_{+,\mu_\Psi,n,d,i,\sigma})\sim\boldsymbol{w}^{*\otimes(m-d)}_{-\circ}\otimes\boldsymbol{w}^{*\otimes d}_{+\circ}.$$

For given $\mathcal{A}\in\binom{[m]}{d}$ let $\alpha_+:[d]\to\mathcal{A}$, $\alpha_-:[m-d]\to[m]\setminus\mathcal{A}$ be the enumerations, and let $\boldsymbol{w}^*_{\mathrm{a},\mathcal{A}}=(\boldsymbol{w}^*_{\mathrm{a},\mathcal{A}}(a))_{a\in[m]}$ be given by $\boldsymbol{w}^*_{\mathrm{a},\mathcal{A}}(a)=\boldsymbol{w}^*_+(\alpha_+^{-1}(a))$ for $a\in\mathcal{A}$ and $\boldsymbol{w}^*_{\mathrm{a},\mathcal{A}}(a)=\boldsymbol{w}^*_-(\alpha_-^{-1}(a))$ for $a\in[m]\setminus\mathcal{A}$. Finally, let $\boldsymbol{\mathcal{A}}_{\mathrm{d},m,d}=\mathrm{u}(\binom{[m]}{d})$ and $\boldsymbol{w}^*_{\mathrm{d},\mu_\Psi,n,m,i,d,\sigma}=\boldsymbol{w}^*_{\mathrm{a},\boldsymbol{\mathcal{A}}_\mathrm{d}}$.

Observation 3.19. *We have $\boldsymbol{w}^*_\mathrm{d}(i,\boldsymbol{d}^*_\mathrm{f}(i),\sigma)\sim\boldsymbol{w}^*(\sigma)$.*

Proof. With $\boldsymbol{w}^*=(\boldsymbol{v}^*,\boldsymbol{\psi}^*)$, $\boldsymbol{\mathcal{A}}^*=\{a\in[m]:i\in\boldsymbol{v}^*_a([k])\}$ and $\boldsymbol{b}^*=(\mathbb{1}\{a\in\boldsymbol{\mathcal{A}}^*\})_a$ we have $\boldsymbol{b}^*\sim\boldsymbol{b}^{*\otimes m}_\circ$, where $\boldsymbol{b}^*_\circ\in\{0,1\}$ is given by the success probability $p_\mathrm{d}(\sigma_i)$. So, for $b\in\{0,1\}^m$ using $p_a=\mathbb{P}(\boldsymbol{b}^*_\circ=b_a)$, $\boldsymbol{w}=(\boldsymbol{v},\boldsymbol{\psi})$ from Section 3.2.1.1 and $\boldsymbol{b}=(\mathbb{1}\{i\in\boldsymbol{v}_a([k])\})_a$ we have

$$\mathbb{P}(\boldsymbol{w}^*\in\mathcal{E}|\boldsymbol{\mathcal{A}}=\mathcal{A})=\mathbb{E}\left[\prod_a\frac{\boldsymbol{\psi}_a(\sigma_{\boldsymbol{v}(a)})\mathbb{1}\{\boldsymbol{b}_a=b_a\}}{\overline{Z}_\mathrm{f}(\gamma_{\mathrm{n},\sigma})p_a}\mathbb{1}\{\boldsymbol{w}\in\mathcal{E}\}\right].$$

Now, the $(\boldsymbol{w}_{-\circ},\boldsymbol{w}_\circ)$-derivative is $(v,\psi)\mapsto\mathbb{1}\{i\notin v([k])\}/\mathbb{P}(i\notin\boldsymbol{v}_\circ([k]))$, so the $(\boldsymbol{w}^*_{-\circ},\boldsymbol{w}_\circ)$-derivative is $(v,\psi)\mapsto\mathbb{1}\{i\notin v([k])\}\psi(\sigma_v)/\mathbb{E}[\mathbb{1}\{i\notin\boldsymbol{v}_\circ([k])\}\boldsymbol{\psi}_\circ(\sigma_{\boldsymbol{v}_\circ})]$. But for any $a\in b^{-1}(0)$ we have

$$\overline{Z}_\mathrm{f}(\gamma_{\mathrm{n},\sigma})p_a=\overline{Z}_\mathrm{f}(\gamma_{\mathrm{n},\sigma})\mathbb{E}\left[\frac{\boldsymbol{\psi}_\circ(\sigma_{\boldsymbol{v}_\circ})}{\overline{Z}_\mathrm{f}(\gamma_{\mathrm{n},\sigma})}\mathbb{1}\{i\notin\boldsymbol{v}_\circ([k])\}\right]=\mathbb{E}[\mathbb{1}\{i\notin\boldsymbol{v}_\circ([k])\}\boldsymbol{\psi}_\circ(\sigma_{\boldsymbol{v}_\circ})].$$

Similarly, the $(\boldsymbol{w}_{+\circ},\boldsymbol{w}_\circ)$-derivative is $(v,\psi)\mapsto\mathbb{1}\{i\in v([k])\}\psi(\sigma_v)/\mathbb{E}[\mathbb{1}\{i\in\boldsymbol{v}_\circ([k])\}\boldsymbol{\psi}_\circ(\sigma_{\boldsymbol{v}_\circ})]$ and $\overline{Z}_\mathrm{f}(\gamma_{\mathrm{n},\sigma})p_a=\mathbb{E}[\mathbb{1}\{i\in\boldsymbol{v}_\circ([k])\}\boldsymbol{\psi}_\circ(\sigma_{\boldsymbol{v}_\circ})]$ for $a\in b^{-1}(1)$. So, $\boldsymbol{w}^*|\boldsymbol{\mathcal{A}}^*=\mathcal{A}$ and $\boldsymbol{w}^*_{\mathrm{a},\mathcal{A}}$ have the same law.

Next, notice that $\boldsymbol{w}^*\sim\boldsymbol{w}^*\circ\alpha$ for any permutation $\alpha\in[m]^{\underline{m}}$ of the factors, which yields $\boldsymbol{\mathcal{A}}^*\sim\alpha(\boldsymbol{\mathcal{A}}^*)$, hence that $\boldsymbol{\mathcal{A}}^*||\boldsymbol{\mathcal{A}}^*|=d$ is uniform and thereby has the same law as $\boldsymbol{\mathcal{A}}_\mathrm{d}$. This shows that $\boldsymbol{w}^*||\boldsymbol{\mathcal{A}}^*|=d$ has the same law as $\boldsymbol{w}^*_\mathrm{d}$ and thereby completes the proof. □

Remark 3.20. Notice that $\boldsymbol{w}^*_{\circ,-\circ}$ does not depend on σ_i due to the definition of $\boldsymbol{v}_{-\circ}$.

3.2.1.8 Standard Graphs. In this section we discuss the relation of the decorated factor graphs from Section 3.2.1.1 and the standard factor graphs from Section 2.1.2.1. For this purpose let $\boldsymbol{G}_\mathrm{d}$, $\boldsymbol{G}^*_\mathrm{d}$ denote the decorated graphs and $\boldsymbol{G}_\mathrm{s}$, $\boldsymbol{G}^*_\mathrm{s}$ the standard graphs.

Observation 3.21. *Let $\Theta^\downarrow=0$ and $t^\leftrightarrow=1$. Let $\boldsymbol{G}_\mathrm{d}=\boldsymbol{G}_{\mathrm{d},\boldsymbol{m},\boldsymbol{m}^\leftrightarrow,\boldsymbol{\mathcal{U}}}$, $\boldsymbol{G}^*_\mathrm{d}=\boldsymbol{G}^*_{\mathrm{d},\boldsymbol{m},\boldsymbol{m}^\leftrightarrow,\boldsymbol{\mathcal{U}}}$, $\boldsymbol{G}_\mathrm{s}=\boldsymbol{G}_{\mathrm{s},\boldsymbol{m}}$ and $\boldsymbol{G}^*_\mathrm{s}=\boldsymbol{G}^*_{\mathrm{s},\boldsymbol{m}}$. Then we have $\boldsymbol{G}_\mathrm{d}\sim[\boldsymbol{G}_\mathrm{s}]^\Gamma$, $\boldsymbol{G}^*_\mathrm{d}(\sigma)\sim[\boldsymbol{G}^*_\mathrm{s}(\sigma)]^\Gamma$, $\psi_{\mathrm{g},\boldsymbol{G}_\mathrm{d}}(\sigma)\sim\gamma^{*\otimes n}(\sigma)\psi_{\boldsymbol{G}_\mathrm{s}}(\sigma)$, $Z_\mathrm{g}(\boldsymbol{G}_\mathrm{d})\sim Z_{\gamma^*}(\boldsymbol{G}_\mathrm{s})$, $Z_\mathrm{g}(\boldsymbol{G}^*_\mathrm{d}(\sigma))\sim Z_{\gamma^*}(\boldsymbol{G}^*_\mathrm{s}(\sigma))$ and $\mu_{\mathrm{g},\boldsymbol{G}_\mathrm{d}}\sim\mu_{\gamma^*,\boldsymbol{G}_\mathrm{s}}$, $\mu_{\mathrm{g},\boldsymbol{G}^*_\mathrm{d}(\sigma)}\sim\mu_{\gamma^*,\boldsymbol{G}^*_\mathrm{s}(\sigma)}$.*

Proof. Notice that $\boldsymbol{m}^\leftrightarrow\equiv0$ and $\boldsymbol{\mathcal{U}}=\emptyset$. Further, we have $\boldsymbol{G}_\mathrm{d}\sim[\boldsymbol{G}_\mathrm{s}]^\Gamma$ by definition, hence $\psi_{\mathrm{g},\boldsymbol{G}_\mathrm{d}}(\sigma)\sim\gamma^{*\otimes n}(\sigma)\psi_{\boldsymbol{G}_\mathrm{s}}(\sigma)$ and $\mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}_\mathrm{d}}(\sigma)|\boldsymbol{m}]=\gamma^{*\otimes n}(\sigma)\mathbb{E}[\psi_{\boldsymbol{G}_\mathrm{s}}(\sigma)|\boldsymbol{m}]$, so as for Observation 3.13 $\gamma^{*\otimes n}(\sigma)$ cancels out in the $(\boldsymbol{G}^*(\sigma),\boldsymbol{G})$-derivative and thereby $\boldsymbol{G}^*_\mathrm{d}(\sigma)\sim[\boldsymbol{G}^*_\mathrm{s}(\sigma)]^\Gamma$. Finally, notice that $Z_{\gamma^*}(G)=\sum_\sigma\gamma^{*\otimes n}(\sigma)\psi_G(\sigma)=\sum_\sigma\psi_{[G]^\Gamma}(\sigma)=Z_\mathrm{g}([G]^\Gamma)$. The remainder follows analogously. □

Notice that the result for ϕ_g, ϕ is implied. Thus, the standard graphs follow as a special case.

3.2.2 The Nishimori Ground Truth. In this section we discuss the Nishimori ground truth $\hat{\boldsymbol{\sigma}}$ from Section 3.1.2.1 and its relation to $\boldsymbol{\sigma}^*$. In Section 3.2.2.1 we show that $\hat{\boldsymbol{\sigma}}$ satisfies the Nishimori condition for the decorated graph and prove Proposition 3.1d). In Section 3.2.2.2 we discuss the color frequencies $\boldsymbol{\gamma}^*_{\gamma^*,n} = \gamma_{\mathrm{n},\boldsymbol{\sigma}^*}$, $\hat{\boldsymbol{\gamma}}_{\mu_\Psi,\gamma^*,n,m} = \gamma_{\mathrm{n},\hat{\boldsymbol{\sigma}}}$ and the conditional laws $\boldsymbol{\sigma}^*|\boldsymbol{\gamma}^*$, $\hat{\boldsymbol{\sigma}}|\hat{\boldsymbol{\gamma}}$, including the proof of Proposition 3.1c), the upper bound in Proposition 3.1a) and the lower bound in Proposition 3.1b). Finally, in Section 3.2.2.3 we bound the total variation distance of Nishimori ground truths for different values of m.

3.2.2.1 Decorated Graphs. Recall the $(\hat{\boldsymbol{\sigma}}, \boldsymbol{\sigma}^*)$-derivative $\hat{r}_{\mu_\Psi,\gamma^*,n,m}$ from Section 3.1.2.1.

Observation 3.22. *Notice that the following holds.*
a) We have $\mathbb{P}(\hat{\boldsymbol{\sigma}} = \sigma) = \overline{\psi}_{\mathrm{m}}(\sigma)/\overline{Z}_{\mathrm{m}}$ *and* $\hat{r}(\sigma) = \overline{Z}_{\mathrm{f}}(\gamma_{\mathrm{n},\sigma})^m/\mathbb{E}[\overline{Z}_{\mathrm{f}}(\gamma_{\mathrm{n},\boldsymbol{\sigma}^*})^m]$.
b) The Radon-Nikodym derivative of $(\hat{\boldsymbol{\sigma}}, \boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}))$ *with respect to* $\boldsymbol{\sigma}^* \otimes \boldsymbol{G}$ *is*

$$(\sigma, G) \mapsto \frac{\psi_{\mathrm{g},G}(\sigma)}{\gamma^{*\otimes n}(\sigma)\overline{Z}_{\mathrm{m}}}.$$

c) The $(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}), \boldsymbol{G})$*-derivative is* $G \mapsto Z_{\mathrm{g}}(G)/\overline{Z}_{\mathrm{m}}$.
d) We have $(\hat{\boldsymbol{\sigma}}, \boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})) \sim (\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})}, \boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}))$.

Proof. With $\boldsymbol{G}_{\mathrm{s}}$ denoting the standard graph and using Observation 3.21 we have

$$\mathbb{P}(\hat{\boldsymbol{\sigma}} = \sigma) = \frac{\gamma^{*\otimes n}(\sigma)\mathbb{E}[\psi_{\boldsymbol{G}_{\mathrm{s}}}(\sigma)]}{\mathbb{E}[Z_{\gamma^*}(\boldsymbol{G}_{\mathrm{s}})]} = \frac{\mathbb{E}[\psi_{\mathrm{g},[\boldsymbol{G}_{\mathrm{s}}]^\Gamma}(\sigma)]}{\mathbb{E}[Z_{\mathrm{g}}([\boldsymbol{G}_{\mathrm{s}}]^\Gamma)]},$$

i.e. the ratio of the expectations for the decorated factor graph without interpolators and pins. But with Observation 3.11e) and Observation 3.11f) this gives

$$\mathbb{P}(\hat{\boldsymbol{\sigma}} = \sigma) = \frac{\gamma^{*\otimes n}(\sigma)\overline{Z}_{\mathrm{f}}(\gamma_{\mathrm{n},\sigma})^m}{\mathbb{E}[\overline{Z}_{\mathrm{f}}(\gamma_{\mathrm{n},\boldsymbol{\sigma}^*})^m]} = \frac{\overline{\psi}_{\mathrm{m}}(\sigma)}{\overline{Z}_{\mathrm{m}}}.$$

Using Part 3.22a), $(\boldsymbol{\sigma}^*, \boldsymbol{G}) \sim \boldsymbol{\sigma}^* \otimes \boldsymbol{G}$ and for an event $\mathcal{E}$ we have

$$\mathbb{P}((\hat{\boldsymbol{\sigma}}, \boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})) \in \mathcal{E}) = \mathbb{E}\left[\frac{\overline{\psi}_{\mathrm{m}}(\boldsymbol{\sigma}^*)}{\gamma^{*\otimes n}(\boldsymbol{\sigma}^*)\overline{Z}_{\mathrm{m}}}\frac{\psi_{\mathrm{g},\boldsymbol{G}}(\boldsymbol{\sigma}^*)}{\overline{\psi}_{\mathrm{m}}(\boldsymbol{\sigma}^*)}\mathbb{1}\{(\boldsymbol{\sigma}^*, \boldsymbol{G}) \in \mathcal{E}\}\right].$$

This shows that the $(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}), \boldsymbol{G})$-derivative is $\sum_\sigma \gamma^{*\otimes n}(\sigma)\frac{\psi_{\mathrm{g},G}(\sigma)}{\gamma^{*\otimes n}(\sigma)\overline{Z}_{\mathrm{m}}} = \frac{Z_{\mathrm{g}}(G)}{\overline{Z}_{\mathrm{m}}}$. This also shows Part 3.22d) since the joint derivative is the product of the individual derivatives. □

Since $\mu_{\mathrm{g},[G]^\Gamma} = \mu_{\gamma^*,G}$ and $G \mapsto [G]^\Gamma$ is a bijection, Proposition 3.1d) follows with Observation 3.21 and Observation 3.22d).

3.2.2.2 Ground Truths. Since both ground truths $\boldsymbol{\sigma}^*$, $\hat{\boldsymbol{\sigma}}$ are invariant to decorations we assume that $m^{\leftrightarrow} \equiv 0$ and $\mathcal{U} = \emptyset$ in this section. Using $\gamma = \gamma_{\mathrm{n},\sigma}$ let $\mathcal{S} = \{\sigma' \in [q]^n : \gamma_{\mathrm{n},\sigma'} = \gamma\}$ and $\boldsymbol{\sigma}_{\Gamma,n,\gamma} \sim \mathrm{u}(\mathcal{S})$.

Observation 3.23. *Notice that the following holds.*
a) We have $(\boldsymbol{\gamma}^*, \boldsymbol{\sigma}^*) \sim (\boldsymbol{\gamma}^*, \boldsymbol{\sigma}_{\Gamma,\boldsymbol{\gamma}^*})$.
b) There exist $c_{\mathfrak{J}} \in \mathbb{R}^2_{>0}$ *with* $\mathbb{P}(\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}} \ge r) \le c_2 e^{-c_1 r^2 n}$ *for* $r \in \mathbb{R}_{\ge 0}$.
c) There exists $c \in \mathbb{R}_{>0}$ *such that* $\mathbb{E}[\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}}] \le c/\sqrt{n}$.
d) There exists $c \in \mathbb{R}_{>0}$ *such that* $\mathbb{E}[\|\boldsymbol{\gamma}^* - \gamma^*\|^2_{\mathrm{tv}}] \le c/n$.

Proof. The first part is clear, further the union bound with Hoeffding's inequality yields

$$\mathbb{P}(\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}} \ge r) \le \mathbb{P}(\exists \tau \in [q] |\boldsymbol{\gamma}^*(\tau) - \gamma^*(\tau)| \ge 2r/q) \le 2q \exp\left(-\frac{8}{q^2} r^2 n\right).$$

For the next part we have $\mathbb{E}[\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}}] \le \int_0^\infty c_2 e^{-c_1 r^2 n} \mathrm{d}r = \frac{c_2\sqrt{\pi}}{2\sqrt{c_1 n}}$. Similarly, for the last part we have $\mathbb{E}[\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}}^2] \le \int_0^\infty c_2 e^{-c_1 r n} \mathrm{d}r = \frac{c_2}{c_1 n}$. □

Next, we establish the bounds for $\overline{\psi}_{\mathrm{m}}$ and $\overline{Z}_{\mathrm{m}}$.

Lemma 3.24. *Notice that the following holds for* $m \le m_\uparrow$.
a) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $c\xi^m \le \overline{Z}_{\mathrm{m}} \le \xi^m$.
b) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *with* $\exp(-c\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2 n)\xi^m \gamma^{*\otimes n}(\sigma) \le \overline{\psi}_{\mathrm{m}}(\sigma) \le \xi^m \gamma^{*\otimes n}(\sigma)$.

Proof. Observation 3.11e) yields $\overline{\psi}_{\mathrm{m}}(\sigma) = \gamma^{*\otimes n}(\sigma)\overline{Z}_{\mathrm{f}}(\gamma)^m \le \gamma^{*\otimes n}(\sigma)\xi^m$, so $\overline{Z}_{\mathrm{m}} \le \xi^m$, where $\gamma = \gamma_{\mathrm{n},\sigma}$. Next, with $\tilde{c}$ from Observation 3.9d) let $c_{\mathfrak{g}} = 8d_\uparrow \tilde{c}\Lambda(\psi_\uparrow)/k$ and $\delta = \|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}$. For $\delta^2 \ge 1/(2\tilde{c}\psi_\uparrow)$ we have

$$\frac{\overline{\psi}_{\mathrm{m}}(\sigma)}{\gamma^{*\otimes n}(\sigma)\xi^m} \ge \left(\frac{\psi_\downarrow}{\psi_\uparrow}\right)^{m_\uparrow} = \psi_\uparrow^{-4d_\uparrow n/k} = \exp\left(-\frac{c}{2\tilde{c}\psi_\uparrow} n\right) \ge e^{-c\delta^2 n}.$$

For $\delta^2 \le 1/(2\tilde{c}\psi_\uparrow)$ we have $\frac{\overline{Z}_{\mathrm{f}}(\gamma)}{\xi} \ge 1 - \frac{\tilde{c}\delta^2}{\psi_\downarrow} \ge \frac{1}{2}$. For $t \in \mathbb{R}_{\ge 1/2}$ we have $\ln(t) \ge (2t-3)(1-t)$ and thereby

$$\ln\left(\frac{\overline{Z}_{\mathrm{f}}(\gamma)}{\xi}\right) \ge -\left(1 + \frac{2\tilde{c}\delta^2}{\psi_\downarrow}\right)\frac{\tilde{c}\delta^2}{\psi_\downarrow} \ge -\frac{2\tilde{c}\delta^2}{\psi_\downarrow},\ \frac{\overline{\psi}_{\mathrm{m}}(\sigma)}{\xi^m \gamma^{*\otimes n}(\sigma)} \ge \exp\left(-\frac{2m_\uparrow \tilde{c}\delta^2}{\psi_\downarrow}\right) \ge e^{-c\delta^2 n}.$$

So, by Jensen's inequality and with $\tilde{c}$ from Observation 3.23d) we have

$$\frac{\overline{Z}_{\mathrm{m}}}{\xi^m} \ge \mathbb{E}\left[\exp\left(-c\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}}^2 n\right)\right] \ge \exp\left(-c\mathbb{E}[\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}}^2] n\right) \ge \exp(-c\tilde{c}).$$

□

Now, we prove the remainder of Proposition 3.1 and translate Observation 3.23 to $\hat{\boldsymbol{\sigma}}$.

Corollary 3.25. *Notice that the following holds for* $m \le m_\uparrow$.
a) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\hat{r} \le c$.
b) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\hat{r}(\sigma) \ge \exp(-c\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2 n)$.
c) We have $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\sigma}}) \sim (\hat{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\Gamma,\hat{\boldsymbol{\gamma}}})$.
d) There exist $c_{\mathfrak{g}} \in \mathbb{R}_{>0}^2$ *with* $\mathbb{P}(\|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}} \ge r) \le c_2 e^{-c_1 r^2 n}$ *for* $r \in \mathbb{R}_{\ge 0}$.
e) There exists $c \in \mathbb{R}_{>0}$ *such that* $\mathbb{E}[\|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}}] \le c/\sqrt{n}$.
f) There exists $c \in \mathbb{R}_{>0}$ *such that* $\mathbb{E}[\|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}}^2] \le c/n$.

Proof. With Observation 3.22a), $\tilde{c}$ from Lemma 3.24a) and Lemma 3.24b) we have $\hat{r} \le \tilde{c}^{-1}$. With c from Lemma 3.24b) and Lemma 3.24a) we have $\hat{r}(\sigma) \ge \exp(-c\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2 n)$. The next part is immediate from the result for $\hat{r}$ in Observation 3.22a), which also shows that $\overline{\psi}_{\mathrm{m}}$ is invariant to permutations of σ. The last parts follow from Part 3.25a) applied to Observation 3.23. □

3.2.2.3 Coupling Nishimori Ground Truths. Since $\hat{\boldsymbol{\sigma}}$ is invariant to decorations we assume that $m^{\leftrightarrow} \equiv 0$ and $\mathcal{U} = \emptyset$. In this section we derive a bound for $\|\hat{\boldsymbol{\sigma}}_{m+1} - \hat{\boldsymbol{\sigma}}_m\|_{\mathrm{tv}}$, which then extends to any $\hat{\boldsymbol{\sigma}}_m$, $\hat{\boldsymbol{\sigma}}_{\tilde{m}}$ using the triangle inequality.

Observation 3.26. *Notice that the following holds for $m \le m_\uparrow$.*
a) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ with $1 - c\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2 \le r(\sigma) \le 1 + \frac{c}{n}$ for the $(\hat{\boldsymbol{\sigma}}_{m+1}, \hat{\boldsymbol{\sigma}}_m)$-derivative r.
b) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $\|\hat{\boldsymbol{\sigma}}_{m+1} - \hat{\boldsymbol{\sigma}}_m\|_{\mathrm{tv}} \le c/n$.

Proof. With Observation 3.11e) and c' from Observation 3.9d) we obtain

$$\frac{\overline{\psi}_{\mathrm{m},m+1}(\sigma)}{\overline{\psi}_{\mathrm{m},m}(\sigma)} = \overline{Z}_{\mathrm{f}}(\gamma_{\mathrm{n},\sigma}) \in [\xi - c'\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2, \xi].$$

This equality and c'' from Corollary 3.25f) further yield

$$\frac{\overline{Z}_{\mathrm{m},m+1}}{\overline{Z}_{\mathrm{m},m}} = \mathbb{E}\left[\overline{Z}_{\mathrm{f}}(\hat{\boldsymbol{\gamma}}_m)\right] \ge \xi - c'\mathbb{E}\left[\|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}}^2\right] \ge \xi - \frac{c'c''}{n}$$

and the upper bound ξ. Hence, we have

$$\frac{\mathbb{P}(\hat{\boldsymbol{\sigma}}_{m+1} = \sigma)}{\mathbb{P}(\hat{\boldsymbol{\sigma}}_m = \sigma)} \ge 1 - \frac{c'}{\xi}\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2 \ge 1 - \frac{c'}{\psi_\uparrow}\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2.$$

For $n \le n_\circ$ with $n_\circ = 2c'c''/\xi \le 2c'c''\psi_\uparrow$ we use Observation 3.22a) to obtain

$$\frac{\mathbb{P}(\hat{\boldsymbol{\sigma}}_{m+1} = \sigma)}{\mathbb{P}(\hat{\boldsymbol{\sigma}}_m = \sigma)} \le \psi_\uparrow^{2m_\uparrow} \le 1 + \frac{n_\circ \psi_\uparrow^{4d_\uparrow n_\circ/k}}{n} \le 1 + \frac{2c'c''\exp(1 + \frac{8}{k}c'c''\Lambda(\psi_\uparrow)d_\uparrow)}{n}.$$

For $n \ge n_\circ$ we use the bounds above and $1/(1-t) \le 1 + 2t$ for $t \in [0, 1/2]$ to obtain

$$\frac{\mathbb{P}(\hat{\boldsymbol{\sigma}}_{m+1} = \sigma)}{\mathbb{P}(\hat{\boldsymbol{\sigma}}_m = \sigma)} \le \frac{1}{1 - \frac{c'c''}{\xi n}} \le 1 + \frac{2c'c''}{\xi n} \le 1 + \frac{2c'c''\psi_\uparrow}{n},$$

which completes the proof of Part 3.26a). Combining this result with Observation 3.6a) gives

$$\|\hat{\boldsymbol{\sigma}}_{m+1} - \hat{\boldsymbol{\sigma}}_m\|_{\mathrm{tv}} = \frac{1}{2}\mathbb{E}\left[|r(\hat{\boldsymbol{\sigma}}_m) - 1|\right] \le \frac{c}{2}\left(\mathbb{E}\left[\|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}}^2\right] + \frac{1}{n}\right),$$

which completes the proof using Corollary 3.25f). □

This completes the discussion of the Nishimori ground truth $\hat{\boldsymbol{\sigma}}$.

3.2.2.4 Ground Truth Given the Graph. In this section we consider arbitrary choices of $m^{\leftrightarrow}$ and $\mathcal{U}$. Due to the Nishimori condition 3.22d) the Nishimori ground truth $\hat{\boldsymbol{\sigma}}$ conditional to $\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})$ has the same distribution as the Gibbs spins $\boldsymbol{\sigma}$. Hence, we only need to discuss the kernel for $\boldsymbol{\sigma}^*$ given $\boldsymbol{G}^*(\boldsymbol{\sigma}^*)$. For this purpose let $r_{\mathrm{g},\sigma}(G) = \psi_{\mathrm{g},G}(\sigma)/\overline{\psi}_{\mathrm{m}}(\sigma)$ be the $(\boldsymbol{G}^*(\sigma), \boldsymbol{G})$-derivative, $r_{\mathrm{g}}^*(G) = \mathbb{E}[r_{\mathrm{g},\boldsymbol{\sigma}^*}(G)]$ and for $G \in [\mathcal{G}]^{\Gamma\leftrightarrow\downarrow}$ let $\boldsymbol{\sigma}_{\mathrm{g},G}^* \in [q]^n$ be given by the $(\boldsymbol{\sigma}_{\mathrm{g},G}^*, \boldsymbol{\sigma}^*)$-derivative $r_{\mathrm{s},G}(\sigma) = r_{\mathrm{g},\sigma}(G)/r_{\mathrm{g}}^*(G)$.

Observation 3.27. *Let $\boldsymbol{G}^* = \boldsymbol{G}^*(\boldsymbol{\sigma}^*)$ and $M = m + \|m^{\leftrightarrow}\|_1$.*
a) The $(\boldsymbol{G}^*, \boldsymbol{G})$-derivative is r_{g}^* with $\psi_\downarrow^{2M}(q\psi_\downarrow)^{|\mathcal{U}|} \le r_{\mathrm{g}}^* \le \psi_\uparrow^{2M} q^{|\mathcal{U}|}$.

b) We have $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*) \sim (\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*}, \boldsymbol{G}^*)$.

Proof. For $G = [(v,\psi)]^{\Gamma\leftrightarrow\downarrow}$ and using Observation 3.11 we have

$$r^*_{\mathrm{g}}(G) = \mathbb{E}\left[\prod_a \frac{\psi_a(\boldsymbol{\sigma}^*_{v(a)})}{\overline{Z}_{\mathrm{f}}(\boldsymbol{\gamma}^*)} \prod_{(i,h)\in\mathcal{A}^{\leftrightarrow}} \frac{\psi^{\leftrightarrow}_{i,h}(\sigma_i)}{\xi} \prod_{i\in\mathcal{U}} \frac{\mathbb{1}\{\boldsymbol{\sigma}^*_i = \check{\sigma}_i\}}{q^{-1}}\right] \ge \psi_{\downarrow}^{2M}(q\psi_{\downarrow})^{|\mathcal{U}|}$$

using $\gamma^{*\otimes\mathcal{U}}(\check{\sigma}_{\mathcal{U}}) \ge \psi_{\downarrow}^{|\mathcal{U}|}$ and the upper bound follows analogously with $\gamma^{*\otimes\mathcal{U}}(\check{\sigma}_{\mathcal{U}}) \le 1$. For the second part with $(\boldsymbol{\sigma}^*, \boldsymbol{G}) \sim \boldsymbol{\sigma}^* \otimes \boldsymbol{G}$ we have

$$\mathbb{P}\left((\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*}, \boldsymbol{G}^*) \in \mathcal{E}\right) = \mathbb{E}\left[r^*_{\mathrm{g}}(\boldsymbol{G}) r_{\mathrm{s},\boldsymbol{G}}(\boldsymbol{\sigma}^*)\mathbb{1}\left\{(\boldsymbol{\sigma}^*, \boldsymbol{G}) \in \mathcal{E}\right\}\right] = \mathbb{P}\left((\boldsymbol{\sigma}^*, \boldsymbol{G}^*) \in \mathcal{E}\right).$$

□

3.2.2.5 Gibbs Spins. In this section we consider arbitrary choices of $m^{\leftrightarrow}$ and $\mathcal{U}$. Due to the Nishimori condition 3.22d) we have $\hat{\boldsymbol{\sigma}} \sim \boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})}$. Hence, we only need to discuss $\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$.

Observation 3.28. *Let* $\boldsymbol{\gamma} = \gamma_{\mathrm{n},\boldsymbol{\sigma}}$ *with* $\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ *and* $m \le m_{\uparrow}$.
a) There exists $c_{\mathfrak{g}} \in \mathbb{R}^2_{>0}$ *such that* $\mathbb{P}(\|\boldsymbol{\gamma} - \gamma^*\|_{\mathrm{tv}} \ge r) \le c_2 e^{-c_1 r^2 n}$.
b) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\mathbb{E}[\|\boldsymbol{\gamma} - \gamma^*\|_{\mathrm{tv}}] \le c/\sqrt{n}$.
c) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\mathbb{E}[\|\boldsymbol{\gamma} - \gamma^*\|^2_{\mathrm{tv}}] \le c/n$.

Proof. Let $c^* \in \mathbb{R}^2_{>0}$ be from Observation 3.23b), $\hat{c} \in \mathbb{R}^2_{>0}$ be from Corollary 3.25d) and $c' \in \mathbb{R}_{>0}$ from Corollary 3.25b). With $r^* = \sqrt{\frac{\hat{c}_1}{2c'}}r$ we have

$$\begin{aligned}\mathbb{P}\left(\|\boldsymbol{\gamma} - \gamma^*\|_{\mathrm{tv}} \ge r\right) &\le e^{c' r^{*2} n}\mathbb{P}\left(\|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}} \ge r, \|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}} < r^*\right) + c_2^* e^{-c_1^* r^{*2} n}\\ &\le \exp\left(\frac{1}{2}\hat{c}_1 r^2 n\right)\hat{c}_2 e^{-\hat{c}_1 r^2 n} + c_2^* \exp\left(-\frac{c_1^*\hat{c}_1}{2c'} r^2 n\right),\end{aligned}$$

which completes the proof with $c_2 = \hat{c}_2 + c_2^*$ and $c_1 = \min(\hat{c}_1/2, c_1^*\hat{c}_1/(2c'))$. The remainder is completely analogous to the proof of Observation 3.23. □

3.2.2.6 Relative Entropies. We compare the various assignments using relative entropies.

Observation 3.29. *Let* $\boldsymbol{G}^* = \boldsymbol{G}^*(\boldsymbol{\sigma}^*)$ *and* $m \le m_{\uparrow}$.
a) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *with* $D_{\mathrm{KL}}(\hat{\boldsymbol{\sigma}}\|\boldsymbol{\sigma}^*) \le c$.
b) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *with* $D_{\mathrm{KL}}(\boldsymbol{\sigma}^*\|\hat{\boldsymbol{\sigma}}) \le c$.
c) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\mathbb{E}[\mathbb{E}[D_{\mathrm{KL}}(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*}\|\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*})|\boldsymbol{G}^*]] \le c$.

Proof. With c' from Corollary 3.25a) we have $D_{\mathrm{KL}}(\hat{\boldsymbol{\sigma}}\|\boldsymbol{\sigma}^*) = \mathbb{E}[\ln(\hat{r}(\hat{\boldsymbol{\sigma}}))] \le \ln(c')$. With c' from Corollary 3.25b) and c'' from Observation 3.23d) we have

$$D_{\mathrm{KL}}(\boldsymbol{\sigma}^*\|\hat{\boldsymbol{\sigma}}) = \mathbb{E}\left[\ln\left(\hat{r}(\boldsymbol{\sigma}^*)^{-1}\right)\right] \le c' n\mathbb{E}\left[\|\boldsymbol{\gamma}^* - \gamma^*\|^2_{\mathrm{tv}}\right] \le c'c''.$$

Using the definitions, the $(\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*}, \boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*})$-derivative $r_{\mathrm{s},G}$ can be composed of

$$r_{\mathrm{s},G}(\sigma) = \frac{\overline{\psi}_{\mathrm{m}}(\sigma) r^*_{\mathrm{g}}(G)}{\psi_{\mathrm{g},G}(\sigma)} \cdot \frac{\psi_{\mathrm{g},G}(\sigma)}{\gamma^{*\otimes n}(\sigma) Z_{\mathrm{g}}(G)} = \mathbb{E}\left[\frac{\hat{r}(\sigma)}{\hat{r}(\boldsymbol{\sigma}_{\mathrm{g},G})}\right].$$

Now, with c' from Corollary 3.25a), c'' from Corollary 3.25b) and $\boldsymbol{\gamma}_{\mathrm{g},G} = \gamma_{\mathrm{n},\boldsymbol{\sigma}_{\mathrm{g},G}}$ we have

$$r_{\mathrm{s},G}(\sigma) \geq \exp\left(-c''\|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}}^2 n\right)/c'.$$

Hence, with the tower property, Observation 3.27 and c''' from Observation 3.23d) we have

$$\mathbb{E}[\mathbb{E}[D_{\mathrm{KL}}(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*}\|\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*})|\boldsymbol{G}^*]] = \mathbb{E}[-\ln(r_{\mathrm{s},\boldsymbol{G}^*}(\boldsymbol{\sigma}^*))] \leq \ln(c') + c''c'''.$$

□

3.2.3 Concentration and Continuity. In this section we prove Proposition 3.2 and related results. In Section 3.2.3.1 we establish boundedness and Lipschitz continuity of the free entropy on the factor graph level for general decorated graphs, yielding Proposition 3.2a).

In Section 3.2.3.2 we establish Lipschitz continuity of $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})]$ for $m^{\leftrightarrow} \equiv 0$ and $\mathcal{U} = \emptyset$. Then we establish Lipschitz continuity for $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^{\star}_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma,\tau))]$ and $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^{*}_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma))]$ in Section 3.2.3.3, yielding Proposition 3.2c). In Section 3.2.3.4 and for $m^{\leftrightarrow} \equiv 0$ and $\mathcal{U} = \emptyset$ we show that $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))] = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))] + o(1)$ which explains why using $\hat{\boldsymbol{\sigma}}$ for Proposition 3.4 is reasonable, and we further show that $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}^*}(\boldsymbol{\sigma}^*))] = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))] + o(1)$, which supports the corresponding claim in Section 3.1.3 regarding $\boldsymbol{m}^*$ and Theorem 2.2. Based on these results we then establish concentration for $m^{\leftrightarrow} \equiv 0$, $\mathcal{U} = \emptyset$ in Section 3.2.3.5, yielding Proposition 3.2b).

3.2.3.1 The Free Entropy. Let $G = [w]^{\Gamma\leftrightarrow\downarrow}_{m^{\leftrightarrow},\psi^{\leftrightarrow},\mathcal{U},\sigma}$ with $w = (v,\psi) \in \mathcal{G}_{n,m}$ and $\tilde{G} = [\tilde{w}]^{\Gamma\leftrightarrow\downarrow}_{\tilde{m}^{\leftrightarrow},\tilde{\psi}^{\leftrightarrow},\tilde{\mathcal{U}},\tilde{\sigma}}$ with $\tilde{w} = (\tilde{v},\tilde{\psi}) \in \mathcal{G}_{n,\tilde{m}}$. Let $\mathcal{V}^{\downarrow}_1 = [n] \setminus (\mathcal{U} \cup \tilde{\mathcal{U}})$ be the unpinned variables, $\mathcal{V}^{\downarrow}_2 = \{i \in \mathcal{U} \cap \tilde{\mathcal{U}} : \sigma_i = \tilde{\sigma}_i\}$ the variables pinned to the same value, and $\mathcal{V}^{\downarrow} = \mathcal{V}^{\downarrow}_1 \cup \mathcal{V}^{\downarrow}_2$. Further, let $m_{\cap} = \min(m,\tilde{m})$, $m^{\leftrightarrow}_{\cap} = (\min(m^{\leftrightarrow}_i,\tilde{m}^{\leftrightarrow}_i))_i$, $\mathcal{A}^{\leftrightarrow}_{\cap} = \{(i,h) : i \in [n], h \in [m_{\cap,i}]\}$ and

$$\mathcal{A}_{=} = \left\{a \in [m_{\cap}] : w_a = \tilde{w}_a, v_a([k]) \subseteq \mathcal{V}^{\downarrow}\right\}, \mathcal{A}^{\leftrightarrow}_{=} = \left\{(i,h) \in \mathcal{A}^{\leftrightarrow}_{\cap} : \psi^{\leftrightarrow}_{i,h} = \tilde{\psi}^{\leftrightarrow}_{i,h}, i \in \mathcal{V}^{\downarrow}\right\}.$$

Now, let $D = m - m_{\cap} + \sum_i(m^{\leftrightarrow}_i - m^{\leftrightarrow}_{\cap,i})$, $\tilde{D} = \tilde{m} - m_{\cap} + \sum_i(\tilde{m}^{\leftrightarrow}_i - m^{\leftrightarrow}_{\cap,i})$ be the excess factors, $D_{\cap} = m_{\cap} - |\mathcal{A}_{=}| + |\mathcal{A}^{\leftrightarrow}_{\cap} \setminus \mathcal{A}^{\leftrightarrow}_{=}|$ the bad factors and let $\mathrm{d}_{\mathrm{g}}(G,\tilde{G}) = D + \tilde{D} + 2D_{\cap} + n - |\mathcal{V}^{\downarrow}|$ be the distance of G and $\tilde{G}$.

Observation 3.30. *There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $|\phi_{\mathrm{g}}(G) - \phi_{\mathrm{g}}(\tilde{G})| \leq \frac{c}{n}\mathrm{d}_{\mathrm{g}}(G,\tilde{G})$ and $|\phi_{\mathrm{g}}(G)| \leq \frac{c}{n}(m + \|m^{\leftrightarrow}\|_1 + |\mathcal{U}|)$.*

Proof. Let G, $\tilde{G}$ be as in the definition of d_{g}. First, we get rid of the excess factor and the bad factors, i.e.

$$Z_{\mathrm{g}}(G) \geq \psi_{\downarrow}^{D+D_{\cap}}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{\sigma}^*_{\mathcal{U}\setminus\mathcal{V}^{\downarrow}_2} = \sigma_{\mathcal{U}\setminus\mathcal{V}^{\downarrow}_2}, \boldsymbol{\sigma}^*_{\mathcal{V}^{\downarrow}_2} = \sigma_{\mathcal{V}^{\downarrow}_2}\}\prod_{a\in\mathcal{A}_{=}}\psi_a(\boldsymbol{\sigma}^*_{v_a})\prod_{(i,h)\in\mathcal{A}^{\leftrightarrow}_{=}}\psi^{\leftrightarrow}_{i,h}(\boldsymbol{\sigma}^*_i)\right].$$

Now, all but the first part of the indicator only depends on $\boldsymbol{\sigma}^*_{\mathcal{V}^{\downarrow}}$, so we can use independence, $\gamma^* \geq \psi_{\downarrow}$ and further transition to $\tilde{G}$, i.e.

$$Z_{\mathrm{g}}(G) \geq \psi_{\downarrow}^{D+D_{\cap}+|\mathcal{U}\setminus\mathcal{V}^{\downarrow}_2|}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{\sigma}^*_{\mathcal{V}^{\downarrow}_2} = \tilde{\sigma}_{\mathcal{V}^{\downarrow}_2}\}\prod_{a\in\mathcal{A}_{=}}\tilde{\psi}_a(\boldsymbol{\sigma}^*_{\tilde{v}_a})\prod_{(i,h)\in\mathcal{A}^{\leftrightarrow}_{=}}\tilde{\psi}^{\leftrightarrow}_{i,h}(\boldsymbol{\sigma}^*_i)\right].$$

This clearly gives $\phi_{\mathrm{g}}(G) \geq \frac{-\ln(\psi_\uparrow)}{n}\mathrm{d}_{\mathrm{g}}(G,\tilde{G}) + \phi_{\mathrm{g}}(\tilde{G})$, the upper bound follows analogously and hence the first part of the assertion holds with $c = \ln(\psi_\uparrow)$. The second part holds due to Observation 3.11c) and $\gamma^* \geq \psi_\downarrow$. □

Proposition 3.2a) follows from Observation 3.21 applied to Observation 3.30.

3.2.3.2 Continuity for the Null Model. In this section we establish Proposition 3.2c) for the null model, implied by the following result for the decorated graph version. Recall $\overline{\phi}_{\mathrm{m}} = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}_m)]$ from Section 3.2.1.1.

Lemma 3.31. *There exists $L_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $|\overline{\phi}_{\mathrm{m}}(m_1) - \overline{\phi}_{\mathrm{m}}(m_2)| \leq L|\frac{km_1}{n} - \frac{km_2}{n}|$ for $m \in \mathbb{Z}^2_{\geq 0}$, $m^{\leftrightarrow} \equiv 0$ and $\mathcal{U} = \emptyset$.*

Proof. For $m^{\leftrightarrow} \equiv 0$ and $\mathcal{U} = \emptyset$ we have $\mathrm{d}_{\mathrm{g}}(G,\tilde{G}) = m + \tilde{m} - 2|\mathcal{A}_=|$ in Section 3.2.3.1. Assume without loss of generality that $m_1 \leq m_2$ and consider the canonical coupling of $\boldsymbol{G}(m_1) = [\boldsymbol{w}_{m_1}]^\Gamma$ and $\boldsymbol{G}(m_2) = [\boldsymbol{w}_{m_2}]^\Gamma$, i.e. $\boldsymbol{w}_{m_1} = \boldsymbol{w}_{m_2,[m_1]}$. Under this coupling we have $\mathcal{A}_= = [m_1]$ and hence $\mathrm{d}_{\mathrm{g}}(\boldsymbol{G}(m_1),\boldsymbol{G}(m_2)) = m_2 - m_1$, so Jensen's inequality and Observation 3.30 yield

$$|\overline{\phi}_{\mathrm{m}}(m_1) - \overline{\phi}_{\mathrm{m}}(m_2)| \leq \mathbb{E}\left[|\phi_{\mathrm{g}}(\boldsymbol{G}(m_1)) - \phi_{\mathrm{g}}(\boldsymbol{G}(m_2))|\right] \leq \frac{c'}{n}(m_2 - m_1) = L\left|\frac{km_1}{n} - \frac{km_2}{n}\right|$$

with $L = c'/k$, and thereby complete the proof. □

Proposition 3.2c) for the null model follows from Observation 3.21 applied to Lemma 3.31.

3.2.3.3 Continuity for the Teacher-Student Model. In this section we establish a version of Proposition 3.2c) for the expected free entropy $\phi^\star(m,\sigma,\tau) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^\star_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma,\tau))]$ over the two-sided planted model and the more general decorated graphs. The result for the teacher-student model then follows as a corollary.

Lemma 3.32. *Let $\gamma_{\mathrm{n},\sigma} \geq \frac{1}{2}\psi_\downarrow$, $m \leq m_\uparrow$, further $\tilde{m} \in \mathbb{Z}_{\geq 0}$, $\tilde{\sigma} \in [q]^n$ and $\tilde{\tau} \in (\mathcal{D}^k_{\Gamma,\tilde{\gamma}})^{\tilde{m}}$. There exists $L_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that*

$$|\phi^\star(m,\sigma,\tau) - \phi^\star(\tilde{m},\tilde{\sigma},\tilde{\tau})| \leq \frac{L}{n}\left(\|n\gamma_{\mathrm{n},\sigma} - n\gamma_{\mathrm{n},\tilde{\sigma}}\|_1 + \|m\alpha_{\mathrm{m},\tau} - \tilde{m}\alpha_{\mathrm{m},\tilde{\tau}}\|_1\right).$$

Proof. Let $\gamma = \gamma_{\mathrm{n},\sigma}$, $\tilde{\gamma} = \gamma_{\mathrm{n},\tilde{\sigma}}$, $\alpha = \alpha_{\mathrm{m},\tau}$ and $\tilde{\alpha} = \alpha_{\mathrm{m},\tilde{\tau}}$. First, we show that $\phi^\star(m,\sigma,\tau) = \phi^\star(\tilde{m},\tilde{\sigma},\tilde{\tau})$ for the special case that $\tilde{m} = m$, $\tilde{\gamma} = \gamma$ and $\tilde{\alpha} = \alpha$, i.e. there exist permutations $\nu \in [n]^{\underline{n}}$ and $\mu \in [m]^{\underline{m}}$ such that $\tilde{\sigma} \circ \nu = \sigma$ and $\tilde{\tau} \circ \mu = \tau$. Similar to the proof of Observation 3.19 we consider a permutation μ of the factors, and moreover a permutation ν of the variables. For given $(v,\psi) \in \mathcal{G}$ let $f(v,\psi) = (\tilde{v},\tilde{\psi}) \in \mathcal{G}$ be given by $\tilde{v}_{\mu(a),h} = \nu(v_{a,h})$ and $\tilde{\psi}_{\mu(a)} = \psi_a$. Notice that $\boldsymbol{w}^\star(\tilde{\sigma},\tilde{\tau}) \sim f(\boldsymbol{w}^\star(\sigma,\tau))$ since f is a simple relabeling of variables and factors. Further, let $f^{\leftrightarrow}(\psi^{\leftrightarrow}) = \tilde{\psi}^{\leftrightarrow}$ with $\tilde{\psi}^{\leftrightarrow}_{\nu(i),h} = \psi^{\leftrightarrow}_{i,h}$ and notice that $\boldsymbol{\psi}^{*\leftrightarrow}_{\tilde{\sigma}} \sim f^{\leftrightarrow}(\boldsymbol{\psi}^{*\leftrightarrow}_{\sigma})$. Finally, using $\tilde{m}^{\leftrightarrow} \circ \nu = m^{\leftrightarrow}$ and $\tilde{\mathcal{U}} = \nu(\mathcal{U})$ we have

$$\phi_{\mathrm{g}}\left([(v,\psi)]^{\Gamma\leftrightarrow\downarrow}_{m^{\leftrightarrow},\psi^{\leftrightarrow},\mathcal{U},\sigma}\right) = \phi_{\mathrm{g}}\left([(\tilde{v},\tilde{\psi})]^{\Gamma\leftrightarrow\downarrow}_{\tilde{m}^{\leftrightarrow},\tilde{\psi}^{\leftrightarrow},\tilde{\mathcal{U}},\tilde{\sigma}}\right),$$

i.e. the free entropy is invariant to a relabeling of factors and variables. This shows that

$$\phi_{\mathrm{g}}\left(\boldsymbol{G}^\star_{m,\tilde{m}^{\leftrightarrow},\tilde{\mathcal{U}}}(\tilde{\sigma},\tilde{\tau})\right) \sim \phi_{\mathrm{g}}\left([f(\boldsymbol{w}^\star(\sigma,\tau))]^{\Gamma\leftrightarrow\downarrow}_{\tilde{m}^{\leftrightarrow},f^{\leftrightarrow}(\boldsymbol{\psi}^{*\leftrightarrow}_{\sigma}),\tilde{\mathcal{U}},\tilde{\sigma}}\right) = \phi_{\mathrm{g}}\left(\boldsymbol{G}^\star_{m,m^{\leftrightarrow},\mathcal{U}}(\sigma,\tau)\right).$$

Since both $\boldsymbol{m}^{\leftrightarrow}$ and $\boldsymbol{\mathcal{U}} = \check{\boldsymbol{u}}_{\mathrm{t},\boldsymbol{\theta}}^{-1}(1)$ are obtained from i.d.d. random variables (given $\boldsymbol{\theta}$), we have $\boldsymbol{G}^{\star}_{m,\boldsymbol{m}^{\leftrightarrow}\circ\nu,\nu(\boldsymbol{\mathcal{U}})}(\tilde\sigma,\tilde\tau) \sim \boldsymbol{G}^{\star}_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\tilde\sigma,\tilde\tau)$ and thereby $\phi^\star(m,\sigma,\tau) = \phi^\star(m,\tilde\sigma,\tilde\tau)$. This completes the proof of the special case and in particular shows that $\phi^\star(m,n\gamma,m\alpha) = \phi^\star(m,\sigma,\tau)$ is well-defined.

Hence, for the general case we assume without loss of generality that $m \le \tilde m$ and that σ, τ, $\tilde\sigma$ and $\tilde\tau$ are ordered as follows. Let $n_{\cap\Gamma} = (\min(n\gamma(\tau'), n\tilde\gamma(\tau')))_{\tau'\in[q]}$ and $n_\cap = \|n_{\cap\Gamma}\|_1$. Analogously, let $m_{\cap\mathrm{A}} = (\min(m\alpha(\tau'), \tilde m\tilde\alpha(\tau')))_{\tau'\in[q]^k}$ and $m_\cap = \|m_{\cap\mathrm{A}}\|_1$. We assume that $\sigma_{[n_\cap]} = \tilde\sigma_{[n_\cap]}$ and $\tau_{[m_\cap]} = \tilde\tau_{[m_\cap]}$.

Next, we consider the following union. Let $n_\cup = n + (n - n_\cap)$ and $\sigma_\cup = (\sigma, \tilde\sigma_{[n]\setminus[n_\cap]})$. Analogously, let $m_\cup = m + (\tilde m - m_\cap)$ and $\tau_\cup = (\tau, \tilde\tau_{[\tilde m]\setminus[m_\cap]})$. Let $\nu_1 : [n] \to [n]$ be the identity, $\nu_2 : [n] \to [n_\cap] \cup ([n_\cup] \setminus [n])$ the enumeration, $\mu_1 : [m] \to [m]$ the identity and $\mu_2 : [\tilde m] \to [m_\cap] \cup ([m_\cup] \setminus [m])$ the enumeration. The union graph $\boldsymbol{G}_\cup = [\boldsymbol{w}_\cup]^{\Gamma\leftrightarrow\downarrow}_{\boldsymbol{m}_\cup^{\leftrightarrow},\boldsymbol{\psi}_\cup^{*\leftrightarrow},\boldsymbol{\mathcal{U}}_\cup,\sigma_\cup}$ is given by

$$(\boldsymbol{w}_\cup, \boldsymbol{m}_\cup^{\leftrightarrow}, \boldsymbol{\psi}_\cup^{*\leftrightarrow}, \boldsymbol{\mathcal{U}}_\cup) \sim \boldsymbol{w}_\cup \otimes (\boldsymbol{m}_\cup^{\leftrightarrow}, \boldsymbol{\psi}_\cup^{*\leftrightarrow}) \otimes \boldsymbol{\mathcal{U}}_\cup,$$

where $\boldsymbol{w}_\cup \sim \boldsymbol{w}^\star_{m_\cup}(\sigma_\cup,\tau_\cup)$ and the remainder is given as follows. The interpolator counts $\boldsymbol{m}_\cup^{\leftrightarrow}$ are given by $\boldsymbol{m}_{\cup,[n]}^{\leftrightarrow} \sim \boldsymbol{m}_n^{\leftrightarrow}$ and $\boldsymbol{m}_\cup^{\leftrightarrow} \circ \nu_2 = \boldsymbol{m}_\cup^{\leftrightarrow} \circ \nu_1$, i.e. we copy the values to the remaining positions. Given $\boldsymbol{m}_\cup^{\leftrightarrow}$ we have $\boldsymbol{\psi}_\cup^{*\leftrightarrow} \sim \boldsymbol{\psi}^*_{n_\cup,\sigma_\cup}$ for the interpolation weights. Similarly, for the pins let $\check{\boldsymbol{u}}_{\mathrm{t}\cup,\theta} \in \{0,1\}^{n_\cup}$ be given by $\check{\boldsymbol{u}}_{\mathrm{t}\cup,\theta,[n]} \sim \check{\boldsymbol{u}}_{\mathrm{t},n,\theta}$ and $\check{\boldsymbol{u}}_{\mathrm{t}\cup,\theta}\circ\nu_2 = \check{\boldsymbol{u}}_{\mathrm{t}\cup,\theta}\circ\nu_1$. Further, let $\boldsymbol{\mathcal{U}} = \check{\boldsymbol{u}}_{\mathrm{t}\cup,\boldsymbol{\theta}_n}^{-1}(1)$ with $\boldsymbol{\theta}_n \sim \mathrm{u}([0,\Theta^\downarrow])$ from Section 3.2.1.1. In words, we obtain $\boldsymbol{m}_\cup^{\leftrightarrow}$ and $\boldsymbol{\mathcal{U}}_\cup$ by choosing the correct distribution on $[n]$ and copying the values to the remainder (yielding the correct distribution there), and then take the law $\boldsymbol{G}^\star(\sigma_\cup,\tau_\cup)$.

Given a graph $G_\cup = [(v_\cup,\psi_\cup)]^{\Gamma\leftrightarrow\downarrow}_{m_\cup^{\leftrightarrow},\psi_\cup^{\leftrightarrow},\mathcal{U}_\cup,\sigma_\cup}$ from $\boldsymbol{G}_\cup$ and $i\in[2]$, let $\boldsymbol{G}_i(G_\cup) = [(\boldsymbol{v},\psi)]^{\Gamma\leftrightarrow\downarrow}_{m^{\leftrightarrow},\psi^{\leftrightarrow},\mathcal{U},\sigma}$ be given by $m^{\leftrightarrow} = m_\cup^{\leftrightarrow}\circ\nu_i$, $\psi^{\leftrightarrow} = \psi_\cup^{\leftrightarrow}\circ\nu_i$, $\mathcal{U} = \nu_i^{-1}(\mathcal{U}_\cup)$, $\psi = \psi_\cup\circ\mu_i$, $\boldsymbol{v}(a,h) = \nu_i^{-1}(v_\cup(\mu(a),h))$ if $v_\cup(\mu(a),h) \in \nu_i([n])$ and otherwise $\boldsymbol{v}(a,h) \sim \mathrm{u}(\mathcal{S}_i)$ independent of everything else, where $\mathcal{S}_i = \nu_i^{-1}(\mathcal{S}_\cup)$ and $\mathcal{S}_\cup = \sigma_\cup^{-1}(\tau_\cup(\mu_i(a),h))$. Notice that $\tau_\cup(\mu_1(a),h) = \tau(a,h)$ and further $\mathcal{S}_1 = \sigma^{-1}(\tau_{a,h})$. Analogously, we obtain $\mathcal{S}_2 = \tilde\sigma^{-1}(\tilde\tau_{a,h})$.

Now, we claim that $\boldsymbol{G}_1(\boldsymbol{G}_\cup) \sim \boldsymbol{G}^\star_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma,\tau)$ and $\boldsymbol{G}_2(\boldsymbol{G}_\cup) \sim \boldsymbol{G}^\star_{\tilde m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\tilde\sigma,\tilde\tau)$. Due to the absence of dependencies and by construction it is straightforward to see that the pinning indicators (sets), the interpolator counts, the interpolation weights and the standard weights have the correct distribution, which leaves us with the (standard) neighborhoods. But using $\boldsymbol{G}_\cup = [(\boldsymbol{v}_\cup,\boldsymbol{\psi}_\cup)]^{\Gamma\leftrightarrow\downarrow}$, $\boldsymbol{G}_1(\boldsymbol{G}_\cup) = [(\boldsymbol{v},\boldsymbol{\psi})]^{\Gamma\leftrightarrow\downarrow}$, for $a\in[m]$, $h\in[k]$ and $i\in\sigma^{-1}(\tau_{a,h})$ we have

$$\begin{aligned}\mathbb{P}(\boldsymbol{v}(a,h)=i) &= \mathbb{P}(\boldsymbol{v}_\cup(a,h)=i) + \mathbb{P}(\boldsymbol{v}_\cup(a,h)\notin\sigma^{-1}(\tau_{a,h}), \boldsymbol{v}(a,h)=i)\\ &= \frac{1}{|\sigma_\cup^{-1}(\tau_{a,h})|} + \frac{|\sigma_\cup^{-1}(\tau_{a,h})| - |\sigma^{-1}(\tau_{a,h})|}{|\sigma_\cup^{-1}(\tau_{a,h})|}\cdot\frac{1}{|\sigma^{-1}(\tau_{a,h})|} = \frac{1}{|\sigma^{-1}(\tau_{a,h})|},\end{aligned}$$

and thereby also $\boldsymbol{v}(a,h) \sim \mathrm{u}(\sigma^{-1}(\tau_{a,h}))$ has the correct distribution. This shows that $\boldsymbol{G}_1(\boldsymbol{G}_\cup) \sim \boldsymbol{G}^\star_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma,\tau)$, and we obtain $\boldsymbol{G}_2(\boldsymbol{G}_\cup) \sim \boldsymbol{G}^\star_{\tilde m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\tilde\sigma,\tilde\tau)$ analogously.

In the next step we want to apply Observation 3.30, hence we have to bound $\mathrm{d_g}(\boldsymbol{G}_1,\boldsymbol{G}_2)$ using $\boldsymbol{G}_i = \boldsymbol{G}_i(\boldsymbol{G}_\cup) = [(\boldsymbol{v}_i,\boldsymbol{\psi}_i)]^{\Gamma\leftrightarrow\downarrow}_{\boldsymbol{m}_i^{\leftrightarrow},\boldsymbol{\psi}_i^*,\boldsymbol{\mathcal{U}}_i,\sigma_\cup\circ\nu_i}$. By construction we have $\boldsymbol{m}^{\leftrightarrow} = \boldsymbol{m}_1^{\leftrightarrow} = \boldsymbol{m}_2^{\leftrightarrow}$ and $\boldsymbol{\mathcal{U}} = \boldsymbol{\mathcal{U}}_1 = \boldsymbol{\mathcal{U}}_2$ (almost surely), so $\mathcal{V}_1^\downarrow = [n]\setminus\boldsymbol{\mathcal{U}}$, $\mathcal{V}_2^\downarrow = \boldsymbol{\mathcal{U}}\cap[n_\cap]$, $\mathcal{V}^\downarrow = ([n]\setminus\boldsymbol{\mathcal{U}})\cup[n_\cap]$, $\min(m,\tilde m) = m$, $m_\cap^{\leftrightarrow} = \boldsymbol{m}^{\leftrightarrow}$, $\mathcal{A}_\cap^{\leftrightarrow} = \mathcal{A}^{\leftrightarrow}_{\boldsymbol{m}^{\leftrightarrow}}$, $D=0$, $\tilde D = \tilde m - m$, $D_\cap = m - |\mathcal{A}_=| + |\mathcal{A}_\cap^{\leftrightarrow}\setminus\mathcal{A}_=^{\leftrightarrow}|$ and $\mathrm{d_g}(\boldsymbol{G}_1,\boldsymbol{G}_2) = \tilde m - m + 2D_\cap + |\boldsymbol{\mathcal{U}}\setminus[n_\cap]|$. Notice that $\boldsymbol{\mathcal{A}}_=^\circ \subseteq \mathcal{A}_=$ with

$$\boldsymbol{\mathcal{A}}_=^\circ = \{a\in[m_\cap] : \boldsymbol{v}_{\cup,a}([k]) \subseteq [n_\cap]\}$$

and $\{(i,h) : i \in [n_\cap], h \in [\boldsymbol{m}_i^{\leftrightarrow}]\} \subseteq \mathcal{A}_=^{\leftrightarrow}$ by construction, so

$$D_\cap \leq m - m_\cap + |\{a \in [m_\cap] : \exists h \in [k]\, \boldsymbol{v}_\cup(a,h) > n_\cap\}| + \sum_{i=n_\cap+1}^{n} \boldsymbol{m}_i^{\leftrightarrow}.$$

Hence, we can upper bound the number of factors by the number of wires, which is then the total degree of the variables $[n_\cup] \setminus [n_\cap]$ with respect to the factors $[m_\cap]$, i.e.

$$\begin{aligned}
D_\cap &\leq m - m_\cap + |\{(a,h) \in [m_\cap] \times [k] : \boldsymbol{v}_\cup(a,h) > n_\cap\}| + \sum_{i=n_\cap+1}^{n} \boldsymbol{m}_i^{\leftrightarrow} \\
&= m - m_\cap + \sum_{i=n_\cap+1}^{n_\cup} \boldsymbol{d}(i) + \sum_{i=n_\cap+1}^{n} \boldsymbol{m}_i^{\leftrightarrow}, \\
\boldsymbol{d}(i) &= |\{(a,h) \in [m_\cap] \times [k] : \boldsymbol{v}_\cup(a,h) = i\}|.
\end{aligned}$$

Notice that $\boldsymbol{d}(i)$ is exactly the (wire) degree of $i \in [n_\cup]$ in $\boldsymbol{w}^\star_{m_\cap}(\sigma_\cup, \tau_{\cup,[m_\cap]})$, so the discussion in Section 3.2.1.6 applies. Further, notice that

$$\mathrm{d_g}(\boldsymbol{G}_1, \boldsymbol{G}_2) \leq \tilde{m} + m - 2m_\cap + \sum_{i=n_\cap+1}^{n_\cup} \boldsymbol{d}(i) + \sum_{i=n_\cap+1}^{n} \boldsymbol{m}_i^{\leftrightarrow}.$$

Now, taking the expectation, using the coupling, Jensen's inequality and c from Observation 3.30 yields

$$|\phi^\star(m,\sigma,\tau) - \phi^\star(\tilde{m},\tilde{\sigma},\tilde{\tau})| \leq \frac{c}{n}\left(\tilde{m} + m - 2m_\cap + \sum_{i=n_\cap+1}^{n_\cup} \mathbb{E}[\boldsymbol{d}(i)] + \sum_{i=n_\cap+1}^{n} \mathbb{E}[\boldsymbol{m}_i^{\leftrightarrow}]\right).$$

By definition we have $\mathbb{E}[\boldsymbol{m}_i^{\leftrightarrow}] = (1-t^{\leftrightarrow})\bar{d} \leq d_\uparrow$, and by Observation 3.17a) we have $\boldsymbol{d}(i) \sim \mathrm{Bin}(|\mathcal{H}|, 1/|\sigma_\cup^{-1}(\sigma')|)$ for $i \in [n_\cup] \setminus [n_\cap]$, with $\sigma' = \sigma_\cup(i)$ and $\mathcal{H} = \{(a,h) \in [m_\cap] \times [k] : \tau_\cup(a,h) = \sigma'\}$ from the proof of Observation 3.17a). This gives

$$\mathbb{E}[\boldsymbol{d}(i)] = \frac{|\mathcal{H}|}{|\sigma_\cup^{-1}(\sigma')|} \leq \frac{km_\cap}{|\sigma^{-1}(\sigma')|} \leq \frac{km_\uparrow}{n\psi_\downarrow/2} = 4d_\uparrow\psi_\uparrow.$$

Using $\|n\gamma - n\tilde{\gamma}\|_1 = 2(n - n_\cap) = n_\cup - n_\cap$ and $\|m\alpha - \tilde{m}\tilde{\alpha}\|_1 = m + \tilde{m} - 2m_\cap$ yields

$$|\phi^\star(m,\sigma,\tau) - \phi^\star(\tilde{m},\tilde{\sigma},\tilde{\tau})| \leq \frac{c}{n}\left(\|m\alpha - \tilde{m}\tilde{\alpha}\|_1 + (4d_\uparrow\psi_\uparrow + d_\uparrow)\|n\gamma - n\tilde{\gamma}\|_1\right),$$

and completes the proof with $L = cd_\uparrow(4\psi_\uparrow + 1)$. □

Remark 3.33. In preparation of the upcoming Aizenman-Sims-Starr scheme, notice that the coupling construction works because we consider fixed $t^{\leftrightarrow}$, $\bar{d}$, $\Theta^{\downarrow}$ and n, i.e. we have the same type of decorations for $\boldsymbol{G}_1(\boldsymbol{G}_\cup)$, $\boldsymbol{G}_2(\boldsymbol{G}_\cup)$.

Now, we obtain the result for $\phi^*(m,\sigma) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma))]$ as a corollary.

Corollary 3.34. *Let $\gamma_{\mathrm{n},\sigma} \geq \frac{1}{2}\psi_{\downarrow}$, $m \leq m_{\uparrow}$, $\tilde{m} \in \mathbb{Z}_{\geq 0}$ and $\tilde{\sigma} \in [q]^n$. For some $L_{\mathfrak{g}} \in \mathbb{R}_{>0}$ we have*

$$|\phi^*(m,\sigma) - \phi^*(\tilde{m},\tilde{\sigma})| \leq L\left(\|\gamma_{\mathrm{n},\sigma} - \gamma_{\mathrm{n},\tilde{\sigma}}\|_1 + \left|\frac{km}{n} - \frac{k\tilde{m}}{n}\right|\right).$$

Proof. Let $\gamma = \gamma_{\mathrm{n},\sigma}$, $\tilde{\gamma} = \gamma_{\mathrm{n}\tilde{\sigma}}$ and assume without loss of generality that $m \leq \tilde{m}$. Using the coupling lemma 3.6e), fix a coupling μ of $\mu_{\mathrm{T}|\Gamma,\gamma}$ and $\mu_{\mathrm{T}|\Gamma\tilde{\gamma}}$ and let $\boldsymbol{\tau} \sim \mu^{\otimes\tilde{m}}$. With Observation 3.15 we have

$$\boldsymbol{G}^*_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma) \sim \boldsymbol{G}^\star_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma,\boldsymbol{\tau}_{1,[m]}),\ \boldsymbol{G}^*_{\tilde{m},\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma) \sim \boldsymbol{G}^\star_{\tilde{m},\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\tilde{\sigma},\boldsymbol{\tau}_2).$$

With the tower property of the expectation, Jensen's inequality and $L^\star$ from Lemma 3.32 we have

$$|\phi^*(m,\sigma) - \phi^*(\tilde{m},\tilde{\sigma})| \leq \frac{L^\star}{n}\mathbb{E}\left[\|n\gamma - n\tilde{\gamma}\|_1 + \|m\boldsymbol{\alpha} - \tilde{m}\tilde{\boldsymbol{\alpha}}\|_1\right]$$

with $\boldsymbol{\alpha} = \alpha_{\mathrm{m},\boldsymbol{\tau}_{1,[m]}}$ and $\tilde{\boldsymbol{\alpha}} = \alpha_{\mathrm{m},\boldsymbol{\tau}_2}$. The triangle inequality gives

$$\|m\boldsymbol{\alpha} - \tilde{m}\tilde{\boldsymbol{\alpha}}\|_1 \leq \sum_{\tau'}\left(\sum_{a\in[m]}|\mathbb{1}\{\boldsymbol{\tau}_{1,a} = \tau'\} - \mathbb{1}\{\boldsymbol{\tau}_{2,a} = \tau'\}| + \sum_{a=m+1}^{\tilde{m}}\mathbb{1}\{\boldsymbol{\tau}_{2,a} = \tau'\}\right)$$

and hence $\mathbb{E}[\|m\boldsymbol{\alpha} - \tilde{m}\tilde{\boldsymbol{\alpha}}\|_1] \leq 2m\mathbb{P}(\boldsymbol{\tau}_{1,1} \neq \boldsymbol{\tau}_{2,1}) + \tilde{m} - m = 2m\|\mu_{\mathrm{T}|\Gamma,\gamma} - \mu_{\mathrm{T}|\Gamma,\tilde{\gamma}}\|_{\mathrm{tv}} + \tilde{m} - m$, so with L' from Observation 3.9e) we have

$$|\phi^*(m,\sigma) - \phi^*(\tilde{m},\tilde{\sigma})| \leq 2L^\star\|\gamma - \tilde{\gamma}\|_{\mathrm{tv}} + \frac{2L'L^\star}{k}\frac{km}{n}\|\gamma - \tilde{\gamma}\|_1 + \frac{L^\star}{k}\left(\frac{k\tilde{m}}{n} - \frac{km}{n}\right),$$

so the assertion holds with $L = \frac{2L^\star}{k}(k + 2L'd_{\uparrow})$. □

Observation 3.21 and Lemma 3.34 yield Proposition 3.2c) for the planted model.

3.2.3.4 Teacher-Student Model Asymptotics. Throughout this section we assume that $m^{\leftrightarrow} \equiv 0$ and $\mathcal{U} = \emptyset$ for convenience. We discuss the behavior of the expected free entropies under random factor counts and random ground truths. For this purpose let $\Gamma^+ = (\lceil n\gamma^*(\tau)\rceil)_\tau$, $\Gamma^- = (\lfloor n\gamma^*(\tau)\rfloor)_\tau$, further let $\Gamma \in \mathbb{Z}^q_{\geq 0}$ be such that $\Gamma^- \leq \Gamma \leq \Gamma^+$ and $\|\Gamma\|_1 = n$, so for $\gamma^\circ = \frac{1}{n}\Gamma$ we have $\gamma^\circ \in \mathcal{P}([q])$ and $\|\gamma^\circ - \gamma^*\|_\infty \leq 1/n$. Let $\sigma^\circ \in [q]^n$ be the non-decreasing assignment with $\gamma_{\mathrm{n},\sigma^\circ} = \gamma^\circ$. Finally, let $m^\circ = \lfloor dn/k \rfloor$ and recall $\boldsymbol{m}^*$, ε_{m}, δ_{m} from the introduction to Section 3.2.

Corollary 3.35. *Let $m \leq m_{\uparrow}$, $m^{\leftrightarrow} \equiv 0$, $\mathcal{U} = \emptyset$ and $\phi^*_m(\sigma) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\sigma))]$.*
a) For some $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ we have $|\mathbb{E}[\phi^*_m(\boldsymbol{\sigma}^*)] - \phi^*_m(\sigma^\circ)| \leq c/\sqrt{n}$, also for $\boldsymbol{\sigma}^*$ replaced by $\hat{\boldsymbol{\sigma}}$.
b) We have $\mathbb{E}[\phi^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*)] = \phi^*_{m^\circ}(\sigma^\circ) + \mathcal{O}(\varepsilon_{\mathrm{m}} + \delta_{\mathrm{m}} + n^{-1/2})$ and the same holds for $\boldsymbol{\sigma}^*$ replaced by $\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}$. Further, this statement also holds for $\boldsymbol{m}$ replaced by $\boldsymbol{m}^*$.

Proof. For $n \geq 2\psi_{\uparrow}$ we have $\gamma^\circ \geq \psi_{\downarrow}/2$. Hence, with Jensen's inequality, L from Corollary 3.34 and c^* from Observation 3.23c) we have

$$|\mathbb{E}[\phi^*_m(\boldsymbol{\sigma}^*)] - \phi^*_m(\sigma^\circ)| \leq L\mathbb{E}[\|\boldsymbol{\gamma}^* - \gamma^\circ\|_{\mathrm{tv}}] \leq c'/\sqrt{n}$$

with $c' = Lc^*$, and the same holds for $\boldsymbol{\sigma}^*$ replaced by $\hat{\boldsymbol{\sigma}}$ and $c' = L\hat{c}$ with $\hat{c}$ from Corollary 3.25e), so Part 3.35a) holds with $c = L\max(c^*,\hat{c}) = L\hat{c}$ for $n \geq 2\psi_{\uparrow}$. For $n \leq 2\psi_{\uparrow}$ we take c' from Observation 3.30 to obtain $|\mathbb{E}[\phi^*_m(\boldsymbol{\sigma}^*)] - \phi^*_m(\sigma^\circ)| \leq 2c'm/n \leq 4c'd_{\uparrow}/k \leq c/\sqrt{n}$ with $c = \sqrt{2\psi_{\uparrow}}4c'd_{\uparrow}/k$.

For δ_{m}, ε_{m} sufficiently large and using Corollary 3.12 we may consider $\boldsymbol{m}$ to be a special case of $\boldsymbol{m}^*$. With c from Observation 3.30 notice that

$$E = |\mathbb{E}[\phi^*_{\boldsymbol{m}^*}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}^*})] - \phi^*_{m^\circ}(\sigma^\circ)| \le \mathbb{E}\left[\frac{c\boldsymbol{m}^*}{n}\right] + \frac{cm^\circ}{n} \le \frac{c}{k}\varepsilon_{\mathrm{m}} + \frac{2cd_\uparrow}{k} + \frac{cd_\uparrow}{k}$$

is uniformly bounded for all n. For $n \ge 2\psi_\uparrow$ recall that $\gamma^\circ \ge \psi_\downarrow/2$ and $m^\circ \le m_\uparrow$. Using Jensen's inequality, L as above, $\hat{c}$ from Corollary 3.25e), $d^\circ = km^\circ/n$ and the triangle inequality we obtain $E \le LE_1 + LE_2$ with

$$\begin{aligned}
E_1 &= \mathbb{E}[\|\hat{\boldsymbol{\gamma}}_{\boldsymbol{m}^*} - \gamma^\circ\|_{\mathrm{tv}}] \le \mathbb{E}[\|\hat{\boldsymbol{\gamma}}_{\boldsymbol{m}^*} - \gamma^*\|_{\mathrm{tv}}] + \frac{q}{2n}\\
&\le \mathbb{E}[\mathbb{1}\{|\boldsymbol{d}^* - \bar{d}| \le \delta_{\mathrm{m}}\}\|\hat{\boldsymbol{\gamma}}_{\boldsymbol{m}^*} - \gamma^*\|_{\mathrm{tv}}] + \varepsilon_{\mathrm{m}} + \frac{q}{2n} \le \frac{\hat{c}}{\sqrt{n}} + \varepsilon_{\mathrm{m}} + \frac{q}{2n},\\
E_2 &= \mathbb{E}[|\boldsymbol{d}^* - d^\circ|] \le \mathbb{E}[|\boldsymbol{d}^* - \bar{d}|] + \frac{k}{n}\\
&\le \delta_{\mathrm{m}} + \mathbb{E}[\mathbb{1}\{|\boldsymbol{d}^* - \bar{d}| > \delta_{\mathrm{m}}\}\boldsymbol{d}^*] + \bar{d}\mathbb{P}(|\boldsymbol{d}^* - \bar{d}| > \delta_{\mathrm{m}}) + \frac{k}{n} \le \delta_{\mathrm{m}} + \varepsilon_{\mathrm{m}} + d_\uparrow\varepsilon_{\mathrm{m}} + \frac{k}{n}.
\end{aligned}$$

The result for $\boldsymbol{\sigma}^*$ follows analogously with $\hat{c}$ replaced by c^* from Observation 3.23c). □

3.2.3.5 Concentration. Throughout this section we assume that $m^{\leftrightarrow} \equiv 0$ and $\mathcal{U} = \emptyset$ for convenience. First, we establish concentration for the models over iid factors.

Lemma 3.36. *Let $m^{\leftrightarrow} \equiv 0$, $\mathcal{U} = \emptyset$ and $m \le m_\uparrow$. There exists $c_{\mathfrak{g}} \in \mathbb{R}^2_{>0}$ such that*

$$\mathbb{P}\left(|\phi_{\mathrm{g}}(\boldsymbol{G}) - \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})]| \ge r\right) \le c_2 e^{-c_1 r^2 n}$$

for $r \in \mathbb{R}_{\ge 0}$ and the same holds for $\boldsymbol{G}$ replaced by $\boldsymbol{G}^\star(\sigma,\tau)$ and $\boldsymbol{G}^*(\sigma)$.

Proof. Recall the proof of Lemma 3.31. For $\tilde{m} = m$ in Section 3.2.3.1 we have

$$\mathrm{d}_{\mathrm{g}}(G, \tilde{G}) = 2m - 2|\mathcal{A}_=| = 2|\{a \in [m] : (v_a, \psi_a) \ne (\tilde{v}_a, \tilde{\psi}_a)\}|.$$

So, for $|\mathcal{A}_=| = m - 1$ and c' from Observation 3.30 we have $|\phi_{\mathrm{g}}(G) - \phi_{\mathrm{g}}(\tilde{G})| \le \frac{2c'}{n}$. Since $\phi_{\mathrm{g}}(\boldsymbol{G}) = \phi_{\mathrm{g}}([\boldsymbol{w}]^\Gamma)$ is a function of m iid pairs McDiarmid's inequality yields

$$\mathbb{P}\left(|\phi_{\mathrm{g}}(\boldsymbol{G}) - \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})]| \ge r\right) \le 2\exp\left(-\frac{2r^2}{m\left(\frac{2c'}{n}\right)^2}\right) \le c_2 e^{-c_1 r^2 n}$$

with $c_2 = 2$ and $c_1 = \frac{k}{4c'^2 d_\uparrow}$. Using Observation 3.13 and Observation 3.15, the proofs for $\boldsymbol{G}^*(\sigma)$ and $\boldsymbol{G}^\star(\sigma,\tau)$ are completely analogous, with the same constants. □

Remark 3.37. This proof extends to any fixed $\mathcal{U}$ (and $\check{\sigma}$) since this determines the pinning weights due to fixed σ, and to not too large $\|m^{\leftrightarrow}\|_1$ analogously to the standard factors.

Next, we establish concentration for random ground truths.

Lemma 3.38. *Let $m^{\leftrightarrow} \equiv 0$, $\mathcal{U} = \emptyset$ and $m \leq m_{\uparrow}$. There exists $c_{\mathfrak{g}} \in \mathbb{R}^2_{>0}$ such that*

$$\mathbb{P}\left(|\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*)) - \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*))]| \geq r\right) \leq c_2 e^{-c_1 r^2 n}$$

for $r \in \mathbb{R}_{\geq 0}$ and the same holds for $\boldsymbol{\sigma}^*$ replaced by $\hat{\boldsymbol{\sigma}}$.

Proof. Let $\phi^*(\sigma) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\sigma))]$ and $\bar{\phi}^* = \mathbb{E}[\phi^*(\boldsymbol{\sigma}^*)]$. With c° from Corollary 3.35 let $\rho = 3c^\circ$, and with L from Corollary 3.34 let $n_\circ = \max(2\psi_{\uparrow}, (3qL/\rho)^2)$. In the following we consider the case $n \geq n_\circ$ and $r \geq \rho/\sqrt{n}$, then the case $n \leq n_\circ$, and finally the case $r \leq \rho/\sqrt{n}$.

For $n \geq n_\circ$ and $r \geq \rho/\sqrt{n}$ the following holds. Using $\|\gamma^\circ - \gamma^*\|_\infty \leq 1/n \leq 1/n^\circ$ we have $\gamma^\circ \geq \psi_{\downarrow}/2$ and hence Corollary 3.34 applies and yields $|\phi^*(\boldsymbol{\sigma}^*) - \phi^*(\sigma^\circ)| \leq L\|\boldsymbol{\gamma}^* - \gamma^\circ\|_{\mathrm{tv}}$. Notice that γ° is also sufficiently close to γ^* relative to r, to be precise we have $\|\gamma^\circ - \gamma^*\|_{\mathrm{tv}} \leq \frac{q}{2n} \leq \frac{q\rho}{6qL\sqrt{n}} \leq \frac{r}{6L}$. The same holds for the expected free entropy, i.e. $|\phi^*(\sigma^\circ) - \bar{\phi}^*| \leq \frac{c^\circ}{\sqrt{n}} = \frac{\rho}{3\sqrt{n}} \leq \frac{1}{3}r$. So, using the triangle inequalities suggested by the above yields

$$|\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*)) - \bar{\phi}^*| \leq |\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*)) - \phi^*(\boldsymbol{\sigma}^*)| + L\left(\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}} + \frac{r}{6L}\right) + \frac{1}{3}r.$$

On $|\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*)) - \bar{\phi}^*| \geq r$ we have $\|\boldsymbol{\gamma}^* - \gamma^*\|_{\mathrm{tv}} \geq r/(6L)$ or $|\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*)) - \phi^*(\boldsymbol{\sigma}^*)| \geq r/3$, so with c^*_Γ from Observation 3.23b) and c_{m} from Lemma 3.36 we have

$$\begin{aligned} P &= \mathbb{P}\left(|\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*)) - \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*))]| \geq r\right) \\ &\leq c_{\Gamma,2}\exp\left(-\frac{c_{\Gamma,1}}{36L^2}r^2 n\right) + c_{\mathrm{m},2}\exp\left(-\frac{c_{\mathrm{m},1}}{9}r^2 n\right) \leq c'_2 e^{-c_1 r^2 n} \end{aligned}$$

with $c'_2 = c_{\Gamma,2} + c_{\mathrm{m},2}$ and $c_1 = \min(\frac{c_{\Gamma,1}}{36L^2}, \frac{c_{\mathrm{m},1}}{9})$. For $n \leq n_\circ$ with $c_{\uparrow}$ from Observation 3.30 we have $|\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*)) - \bar{\phi}^*| \leq c_{\uparrow} m_{\uparrow}/n = r_{\uparrow}$ with $r_{\uparrow} = 2c_{\uparrow} d_{\uparrow}/k$. For $r \leq r_{\uparrow}$ we have

$$P \leq 1 = c''_2 \exp\left(-c_1 r_{\uparrow}^2 n_\circ\right) \leq c''_2 e^{-c_1 r^2 n}$$

with $c''_2 = \exp(c_1 r_{\uparrow}^2 n_\circ)$, but for $r > r_{\uparrow}$ we have $P = 0 \leq c''_2 e^{-c_1 r^2 n}$. For $r \leq \rho/\sqrt{n}$ we have

$$P \leq 1 = e^{c_1 r^2 n} e^{-c_1 r^2 n} \leq c'''_2 e^{-c_1 r^2 n}$$

with $c'''_2 = e^{c_1 \rho^2}$. Choosing $c_2 = \max(c'_2, c''_2, c'''_2)$ completes the proof, since c^*_Γ replaced by $\hat{c}_\Gamma$ from Corollary 3.25d) yields the analogous result for $\hat{\boldsymbol{\sigma}}$. □

Observation 3.21, Lemma 3.36 and Lemma 3.38 yield Proposition 3.2b).

3.3 The Planted Model Quenched Free Entropy

We turn to the proof of Theorem 2.2. In Section 3.3.1 we prove Lemma 3.3, apply it to $\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})$, $\boldsymbol{G}^*(\boldsymbol{\sigma}^*)$ and show that the pinning does not alter the quenched free entropy too much.

In Section 3.3.2 we use the interpolation method to prove Proposition 3.4. Then, we can finally discard the interpolators. In Section 3.3.3 we use the Aizenman-Sims-Starr scheme for the simplified model. Finally, in Section 3.3.3.18 we complete the proof.

3.3.1 Pinned Measures and Their Marginal Distributions. This section is composed of four parts. First, we prove Lemma 3.3, based on [33], [92], in Sections 3.3.1.1 to 3.3.1.4. Then, we discuss the pinning of Gibbs measures in Section 3.3.1.5 and Section 3.3.1.6. In the third part, Section 3.3.1.7, we discuss the marginal distributions of (pinned) measures and prove another proposition for general (pinned) measures. In the last part, Sections 3.3.1.8 and 3.3.1.9, we apply this proposition to decorated graphs and discuss projections onto $\mathcal{P}^2_*([q])$.

In Section 3.3.1.1 we discuss the underlying model, the erasure channel, and the conditional entropy of the assignment. In Section 3.3.1.2 we take the derivative of the conditional entropy with respect to the pinning probability, yielding the crucial connection to the mutual information. In Section 3.3.1.3 we introduce the total correlation, and then establish Lemma 3.3 in Section 3.3.1.4.

In Section 3.3.1.5 we apply Lemma 3.3 to Gibbs measures μ_{g} of decorated graphs. In Section 3.3.1.6 we show that pinning does not alter the quenched free entropy too much for $\Theta^{\downarrow} = \Theta^{\downarrow}(n) = o(n)$.

Next, we introduce empirical marginal distributions in Section 3.3.1.7, further conditional and reweighted versions of the marginal distribution, and show that these asymptotically coincide if the empirical color frequencies concentrate and the measure is ε-symmetric, e.g. most pinned measures.

In Section 3.3.1.8 we show that the empirical color frequencies of the Gibbs spins concentrate and hence in particular the proposition for general measures applies to the Gibbs measure induced by the graph. Finally, in Section 3.3.1.9 we introduce a projection of $\mathcal{P}^2([q])$ onto $\mathcal{P}^2_*([q])$, and then show that the distance of the Gibbs marginal distribution to its projection vanishes.

3.3.1.1 The Erasure Channel, Conditional Entropy and Random Conditioning. For $q \in \mathbb{Z}_{\geq 2}$ and $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in [q]^3$, the cross entropy, the entropy and the relative entropy are

$$H(\boldsymbol{x}\|\boldsymbol{y}) = \sum_x -\mathbb{P}(\boldsymbol{x} = x)\ln(\mathbb{P}(\boldsymbol{y} = x)), H(\boldsymbol{x}) = H(\boldsymbol{x}\|\boldsymbol{x}), D_{\mathrm{KL}}(\boldsymbol{x}\|\boldsymbol{y}) = H(\boldsymbol{x}\|\boldsymbol{y}) - H(\boldsymbol{x})$$

respectively. Notice that the definition of the relative entropy is consistent with the general case from Section 2.1.2.5, and in particular both the cross entropy and the relative entropy are finite if and only if $\boldsymbol{x}$ is absolutely continuous with respect to $\boldsymbol{y}$. The conditional cross entropy, the conditional entropy and the conditional relative entropy are

$$H(\boldsymbol{x}\|\boldsymbol{y}|\boldsymbol{z}) = \mathbb{E}[\mathbb{E}[H(\boldsymbol{x}\|\boldsymbol{y})|\boldsymbol{z}]],\ H(\boldsymbol{x}|\boldsymbol{z}) = H(\boldsymbol{x}\|\boldsymbol{x}|\boldsymbol{z}),\ D_{\mathrm{KL}}(\boldsymbol{x}\|\boldsymbol{y}|\boldsymbol{z}) = \mathbb{E}[\mathbb{E}[D_{\mathrm{KL}}(\boldsymbol{x}\|\boldsymbol{y})|\boldsymbol{z}]].$$

The conditional mutual information is $I(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}) = D_{\mathrm{KL}}(\boldsymbol{x}, \boldsymbol{y}\|\boldsymbol{x} \otimes \boldsymbol{y}|\boldsymbol{z})$. For now, we focus on the following conditional entropy.

Let $n \in \mathbb{Z}_{>0}$, $\mu \in \mathcal{P}([q]^n)$ and $\boldsymbol{x}_\mu \sim \mu$ a random vector of values. Further, let $p \in [0,1]^n$ and let $\boldsymbol{r} \in \{0,1\}^n$ be the revealment given by $\boldsymbol{r} \sim \bigotimes_i \boldsymbol{r}_i$ and Bernoulli variables $\boldsymbol{r}_i$ with success probability p_i. Using the joint distribution $(\boldsymbol{x}, \boldsymbol{r}) \sim \boldsymbol{x} \otimes \boldsymbol{r}$ let $\boldsymbol{\chi} = (\boldsymbol{r}_i \boldsymbol{x}_i)_i \in [q]_\circ^n$ with $[q]_\circ = [q] \cup \{0\}$ be the partial observation. This approach reflects [92].

Fix values $x \in [q]^n$, revealments $r \in \{0,1\}^n$ and let $\chi = (r_i x_i)_i$ be the partial observation, and $\mathcal{R} = r^{-1}(1) = \chi^{-1}([q])$ the revealed coordinates. Further, fix known coordinates $\mathcal{K} \subseteq [n]$, tested coordinates $\mathcal{T} \subseteq [n]$ and selected coordinates $\mathcal{S} \subseteq [n]$. Now, let

$$\begin{aligned} \eta_{n,\mu,p}(\mathcal{S}, x_{\mathcal{K}}, \chi_{\mathcal{T}}) &= H(\boldsymbol{x}_{\mathcal{S}}|\boldsymbol{x}_{\mathcal{K}} = x_{\mathcal{K}}, \boldsymbol{\chi}_{\mathcal{T}} = \chi_{\mathcal{T}}), \\ \overline{\eta}_{n,\mu,p}(\mathcal{S}, \mathcal{K}, \mathcal{T}) &= H(\boldsymbol{x}_{\mathcal{S}}|\boldsymbol{x}_{\mathcal{K}}, \boldsymbol{\chi}_{\mathcal{T}}) = \mathbb{E}\left[\eta(\mathcal{S}, \boldsymbol{x}_{\mathcal{K}}, \boldsymbol{\chi}_{\mathcal{T}})\right] \end{aligned}$$

be the (pointwise) entropy and the conditional entropy respectively. As already indicated by the definition of $\boldsymbol{v}$ in Section 3.1.2.3 we consider selections with repetition. Hence, we establish that the definition above is indeed sufficient for our purposes and further derive a few useful basic properties.

Observation 3.39. *Notice that the following holds.*

a) We have $\eta(\mathcal{S}, x_\emptyset, \chi_\mathcal{T}) = H(\boldsymbol{x}_\mathcal{S}|\boldsymbol{\chi}_\mathcal{T} = \chi_\mathcal{T})$, $\eta(\mathcal{S}, x_\mathcal{K}, \chi_\emptyset) = H(\boldsymbol{x}_\mathcal{S}|\boldsymbol{x}_\mathcal{K} = x_\mathcal{K})$ *and* $\eta(\emptyset, \cdot, \cdot) = 0$.

b) Let $s, k, t \in \mathbb{Z}_{\geq 0}$, $\sigma \in [n]^s$, $\kappa \in [n]^k$ *and* $\tau \in [n]^t$ *such that* $\sigma([s]) = \mathcal{S}$, $\kappa([k]) = \mathcal{K}$ *and* $\tau([t]) = \mathcal{T}$. *Then we have* $H(\boldsymbol{x}_\sigma|\boldsymbol{x}_\kappa = x_\kappa, \boldsymbol{\chi}_\tau = \chi_\tau) = \eta(\mathcal{S}, x_\mathcal{K}, \chi_\mathcal{T})$.

c) We have $\eta(\mathcal{S}, x_\mathcal{K}, \chi_\mathcal{T}) = \eta(\mathcal{S} \setminus \mathcal{K}, x_\mathcal{K}, \chi_{\mathcal{T}\setminus\mathcal{K}}) = \eta(\mathcal{S} \setminus \mathcal{K}^*, x_{\mathcal{K}^*}, \chi_\emptyset)$ *with* $\mathcal{K}^* = \mathcal{K} \cup (\mathcal{T} \cap \mathcal{R})$.

d) For $\mathcal{S} = \mathcal{S}_1 \dot\cup \mathcal{S}_2$ *we have*

$$\overline{\eta}(\mathcal{S}, \mathcal{K}, \mathcal{T}) = \overline{\eta}(\mathcal{S}_1 \setminus \mathcal{K}, \mathcal{K}, \mathcal{T} \setminus \mathcal{K}) + \overline{\eta}(\mathcal{S}_2 \setminus \mathcal{K}, \mathcal{K} \cup \mathcal{S}_1, \mathcal{T} \setminus \mathcal{K}).$$

Proof. Recall well-known properties of the conditional entropy, in particular that $H(\boldsymbol{a}|\boldsymbol{b}) = 0$ if and only if $\boldsymbol{a}$ is determined by $\boldsymbol{b}$, and the chain rule. Further, notice that the conditional entropy is exclusively a function of the laws, and that

$$\mathbb{P}(\boldsymbol{a} = a, \boldsymbol{a} = a, \boldsymbol{b} = b, \boldsymbol{c}_1 = c_1|\boldsymbol{b} = b, \boldsymbol{b} = b, \boldsymbol{c}_2 = c_2) = \mathbb{P}(\boldsymbol{a} = a|\boldsymbol{b} = b)$$

whenever $\boldsymbol{c}_1 = c_1$, $\boldsymbol{c}_2 = c_2$ almost surely. This shows Part 3.39a) and Part 3.39b). Notice that

$$\begin{aligned}\eta(\mathcal{S}, x_\mathcal{K}, \chi_\mathcal{T}) &= H(\boldsymbol{x}_\mathcal{S}|\boldsymbol{x}_\mathcal{K} = x_\mathcal{K}, \boldsymbol{\chi}_\mathcal{T} = \chi_\mathcal{T})\\ &= H(\boldsymbol{x}_\mathcal{S}|\boldsymbol{x}_\mathcal{K} = x_\mathcal{K}, \boldsymbol{x}_{\mathcal{T}\cap\mathcal{R}} = x_{\mathcal{T}\cap\mathcal{R}}, \boldsymbol{r}_\mathcal{T} = r_\mathcal{T}) = H(\boldsymbol{x}_{\mathcal{S}\setminus\mathcal{K}^*}|\boldsymbol{x}_{\mathcal{K}^*} = x_{\mathcal{K}^*})\\ &= \eta(\mathcal{S} \setminus \mathcal{K}^*, x_{\mathcal{K}^*}, \chi_\emptyset)\end{aligned}$$

using $(\boldsymbol{x}, \boldsymbol{r}) = \boldsymbol{x} \otimes \boldsymbol{r}$, so Part 3.39c) holds since this also holds for $\mathcal{S}^\circ = \mathcal{S} \setminus \mathcal{K}$, $\mathcal{T}^\circ = \mathcal{T} \setminus \mathcal{K}$, and $\mathcal{K}^* = \mathcal{K} \cup (\mathcal{T}^\circ \cap \mathcal{R})$ and $\mathcal{S} \setminus \mathcal{K}^* = \mathcal{S}^\circ \setminus \mathcal{K}^*$. With $\mathcal{S}_1^\circ = \mathcal{S}_1 \setminus \mathcal{K}$, $\mathcal{S}_2^\circ = \mathcal{S}_2 \setminus \mathcal{K}$, $\mathcal{S}^\circ = \mathcal{S}_1^\circ \dot\cup \mathcal{S}_2^\circ$, Part 3.39c) and the chain rule for the conditional entropy we have

$$\overline{\eta}(\mathcal{S}, \mathcal{K}, \mathcal{T}) = \overline{\eta}(\mathcal{S}^\circ, \mathcal{K}, \mathcal{T}^\circ) = H(\boldsymbol{x}_{\mathcal{S}_1^\circ}|\boldsymbol{x}_\mathcal{K}, \boldsymbol{\chi}_{\mathcal{T}^\circ}) + H(\boldsymbol{x}_{\mathcal{S}_2^\circ}|\boldsymbol{x}_{\mathcal{K}\cup\mathcal{S}_1^\circ}, \boldsymbol{\chi}_{\mathcal{T}^\circ}),$$

which completes the proof of Part 3.39d). □

Based on Observation 3.39 we assume that $\mathcal{S} \cap \mathcal{K} = \emptyset$ and $\mathcal{T} \cap \mathcal{K} = \emptyset$. Notice that Observation 3.39c) using $\boldsymbol{\mathcal{R}} = \mathcal{T} \cap \boldsymbol{r}^{-1}(1)$ yields the minimal form

$$\overline{\eta}(\mathcal{S}, \mathcal{K}, \mathcal{T}) = \mathbb{E}\left[\eta\left(\mathcal{S} \setminus \boldsymbol{\mathcal{R}}, \boldsymbol{x}_{\mathcal{K}\cup\boldsymbol{\mathcal{R}}}, \boldsymbol{\chi}_\emptyset\right)\right] = \mathbb{E}\left[\mathbb{E}\left[H(\boldsymbol{x}_{\mathcal{S}\setminus\boldsymbol{\mathcal{R}}}|\boldsymbol{x}_{\mathcal{K}\cup\boldsymbol{\mathcal{R}}})\middle|\boldsymbol{\mathcal{R}}\right]\right].$$

This representation reflects the approach in [33].

3.3.1.2 The Conditional Entropy Derivative. Let $\mathcal{S}\cap\mathcal{K} = \emptyset$ and $\mathcal{T}\cap\mathcal{K} = \emptyset$ in this section. Let $\frac{\partial}{\partial x_i} f(x)$ denote the i-th partial derivative of f at x.

Lemma 3.40. *For* $i \in [n]$ *we have* $\frac{\partial}{\partial p_i}\overline{\eta}_p(\mathcal{S}, \mathcal{K}, \mathcal{T}) = -\mathbb{1}\{i \in \mathcal{T}\}I(\boldsymbol{x}_\mathcal{S}, \boldsymbol{x}_i|\boldsymbol{x}_\mathcal{K}, \boldsymbol{\chi}_{\mathcal{T}\setminus\{i\}})$.

Proof. With Observation 3.39 and $\boldsymbol{\mathcal{R}} = \mathcal{T} \cap \boldsymbol{r}^{-1}(1)$ we have

$$\begin{aligned}\overline{\eta}(\mathcal{S}, \mathcal{K}, \mathcal{T}) &= \mathbb{E}\left[\mathbb{E}\left[H(\boldsymbol{x}_{\mathcal{S}\setminus\boldsymbol{\mathcal{R}}}|\boldsymbol{x}_{\mathcal{K}\cup\boldsymbol{\mathcal{R}}})\middle|\boldsymbol{\mathcal{R}}\right]\right]\\ &= \sum_{r\in\{0,1\}^\mathcal{T}} \prod_{i\in\mathcal{T}} \mathbb{P}(\boldsymbol{r}_i = r_i) H(\boldsymbol{x}_{\mathcal{S}\setminus r^{-1}(1)}|\boldsymbol{x}_{\mathcal{K}\cup r^{-1}(1)}).\end{aligned}$$

This shows that $\frac{\partial}{\partial p_i}\overline{\eta}(\mathcal{S},\mathcal{K},\mathcal{T})=0$ for $i\in[n]\setminus\mathcal{T}$. For $i\in\mathcal{T}$ let $\mathcal{T}^\circ=\mathcal{T}\setminus\{i\}$. Then we have

$$\begin{aligned}\frac{\partial}{\partial p_i}\overline{\eta}(\mathcal{S},\mathcal{K},\mathcal{T})&=\sum_{r\in\{0,1\}^{\mathcal{T}}}\mathbb{P}(\boldsymbol{r}_{\mathcal{T}^\circ}=r_{\mathcal{T}^\circ})H(\boldsymbol{x}_{\mathcal{S}\setminus r^{-1}(1)}|\boldsymbol{x}_{\mathcal{K}\cup r^{-1}(1)})(r_i-(1-r_i))\\&=\overline{\eta}(\mathcal{S}\setminus\{i\},\mathcal{K}\cup\{i\},\mathcal{T}^\circ)-\overline{\eta}(\mathcal{S},\mathcal{K},\mathcal{T}^\circ)\\&=\overline{\eta}(\mathcal{S},\mathcal{K}\cup\{i\},\mathcal{T}^\circ)-\overline{\eta}(\mathcal{S},\mathcal{K},\mathcal{T}^\circ)=-I(\boldsymbol{x}_{\mathcal{S}},\boldsymbol{x}_i|\boldsymbol{x}_{\mathcal{K}},\boldsymbol{\chi}_{\mathcal{T}^\circ})\end{aligned}$$

since $I(\boldsymbol{a},\boldsymbol{b}|\boldsymbol{c})=H(\boldsymbol{b}|\boldsymbol{c})-H(\boldsymbol{b}|\boldsymbol{a},\boldsymbol{c})$. □

3.3.1.3 Mutual Information, Relative Entropy and the Product of the Marginals. The last sections were dedicated to the conditional entropy. Now, we turn to the following relative entropy. The (generalized conditional) mutual information (total correlation) for $(\boldsymbol{a},\boldsymbol{b})=((\boldsymbol{a}_h)_{h\in\mathcal{H}},\boldsymbol{b})$ is

$$I(\boldsymbol{a}|\boldsymbol{b})=D_{\mathrm{KL}}\left(\boldsymbol{a}\,\middle\|\,\bigotimes_{h\in\mathcal{H}}\boldsymbol{a}_h\,\middle|\,\boldsymbol{b}\right)=H\left(\boldsymbol{a}\,\middle\|\,\bigotimes_{h\in\mathcal{H}}\boldsymbol{a}_h\,\middle|\,\boldsymbol{b}\right)-H(\boldsymbol{a}|\boldsymbol{b})=\sum_{h\in\mathcal{H}}H(\boldsymbol{a}_h|\boldsymbol{b})-H(\boldsymbol{a}|\boldsymbol{b}).$$

For $\boldsymbol{a}=(\boldsymbol{a}_1,\boldsymbol{a}_2)$, this notion of $I(\boldsymbol{a}|\boldsymbol{b})$ coincides with $I(\boldsymbol{a}_1,\boldsymbol{a}_2|\boldsymbol{b})$ from Section 3.3.1.1.

Observation 3.41. *Let $(\boldsymbol{a},\boldsymbol{b})\in\mathcal{A}^m\times\mathcal{B}^n$ with $m,n\in\mathbb{Z}_{\geq0}$ and $\mathcal{A},\mathcal{B}\neq\emptyset$. Further, let $k,\ell\in\mathbb{Z}_{\geq0}$, $v\in[m]^k$, $w\in[n]^\ell$, $\mathcal{V}=v([k])$ and $\mathcal{W}=w([\ell])$.*
a) For $\ell'\in\mathbb{Z}_{\geq0}$, $w'\in[m]^{\ell'}$ with $w'([\ell'])\subseteq\mathcal{W}$ we have

$$I(\boldsymbol{a}_v,\boldsymbol{b}_{w'}|\boldsymbol{b}_w)=I(\boldsymbol{a}_v|\boldsymbol{b}_{\mathcal{W}})=\sum_{i\in[m]}|v^{-1}(i)|H(\boldsymbol{a}_i|\boldsymbol{b}_{\mathcal{W}})-H(\boldsymbol{a}_{\mathcal{V}}|\boldsymbol{b}_{\mathcal{W}}).$$

b) For $j\in\mathbb{Z}_{\geq0}$ and $\dot{\bigcup}_{i\in[j]}\mathcal{K}_i=[k]$ with $v'(i)=v_{\mathcal{K}_i}$ we have

$$I(\boldsymbol{a}_v|\boldsymbol{b}_w)=\sum_{i\in[j]}I(\boldsymbol{a}_{v'(i)}|\boldsymbol{b}_{\mathcal{W}})+I((\boldsymbol{a}_{v'(i)})_{i\in[j]}|\boldsymbol{b}_{\mathcal{W}}).$$

Proof. Part 3.41a) is immediate from the properties of the conditional entropy since

$$\begin{aligned}I(\boldsymbol{a}_v,\boldsymbol{b}_{w'}|\boldsymbol{b}_w)&=\sum_{h\in[k]}H(\boldsymbol{a}_{v(h)}|\boldsymbol{b}_w)+\sum_{h\in[\ell']}H(\boldsymbol{b}_{w'(h)}|\boldsymbol{b}_w)-H(\boldsymbol{a}_v,\boldsymbol{b}_{w'}|\boldsymbol{b}_w)\\&=\sum_{i\in[m]}|v^{-1}(i)|H(\boldsymbol{a}_i|\boldsymbol{b}_{\mathcal{W}})-H(\boldsymbol{a}_{\mathcal{V}}|\boldsymbol{b}_{\mathcal{W}}).\end{aligned}$$

The second part is also immediate from the conditional entropy representation since

$$\begin{aligned}I(\boldsymbol{a}_v|\boldsymbol{b}_w)&=\sum_{h\in[k]}H(\boldsymbol{a}_{v(h)}|\boldsymbol{b}_{\mathcal{W}})-H(\boldsymbol{a}_v|\boldsymbol{b}_{\mathcal{W}})\\&=\sum_{i\in[j]}I(\boldsymbol{a}_{v'(i)}|\boldsymbol{b}_{\mathcal{W}})+\sum_{i\in[j]}H(\boldsymbol{a}_{v'(i)}|\boldsymbol{b}_{\mathcal{W}})-H(\boldsymbol{a}_{\mathcal{V}}|\boldsymbol{b}_{\mathcal{W}})\\&=\sum_{i\in[j]}I(\boldsymbol{a}_{v'(i)}|\boldsymbol{b}_{\mathcal{W}})+I((\boldsymbol{a}_{v'(i)})_{i\in[j]}|\boldsymbol{b}_{\mathcal{W}}).\end{aligned}$$

□

As for the conditional entropy, Observation 3.41a) yields a normalized form, and Observation 3.41b) is a partitioning property of the mutual information.

Fix known coordinates $\mathcal{K} \subseteq [n]$, tested coordinates $\mathcal{T} \subseteq [n]$, further $s \in \mathbb{Z}_{\geq 0}$, a selection $\sigma \in [n]^s$ and let $\mathcal{S} = \sigma([s])$. In the following we discuss the mutual information given by

$$\iota_{n,\mu,p}(\sigma, x_{\mathcal{K}}, \chi_{\mathcal{T}}) = I(\boldsymbol{x}_\sigma | \boldsymbol{x}_{\mathcal{K}} = x_{\mathcal{K}}, \boldsymbol{\chi}_{\mathcal{T}} = \chi_{\mathcal{T}}),$$
$$\bar{\iota}_{n,\mu,p}(\sigma, \mathcal{K}, \mathcal{T}) = I(\boldsymbol{x}_\sigma | \boldsymbol{x}_{\mathcal{K}}, \boldsymbol{\chi}_{\mathcal{T}}) = \mathbb{E}[\iota(\sigma, \boldsymbol{x}_{\mathcal{K}}, \boldsymbol{\chi}_{\mathcal{T}})].$$

Observation 3.41a) ensures that it is sufficient to consider sets $\mathcal{K}$, $\mathcal{T}$. Next, we establish basic properties and build the connection to the conditional entropy.

Observation 3.42. *Let $\mathcal{D} = \sigma^{-1}(\mathcal{S} \setminus \mathcal{K})$, $\sigma^\circ = \sigma_{\mathcal{D}}$ and $\mathcal{T}^\circ = \mathcal{T} \setminus \mathcal{K}$.*
a) Let $\mathcal{K}^* = \mathcal{K} \cup (\mathcal{T} \cap \mathcal{R})$, $\mathcal{D}^* = \sigma^{-1}(\mathcal{S} \setminus \mathcal{K}^*)$ and $\sigma^* = \sigma_{\mathcal{D}^*}$. Then we have

$$\iota(\sigma, x_{\mathcal{K}}, \chi_{\mathcal{T}}) = \iota(\sigma^\circ, x_{\mathcal{K}}, \chi_{\mathcal{T}^\circ}) = \iota(\sigma^\circ, x_{\mathcal{K}^*}, \chi_\emptyset) = \iota(\sigma^*, x_{\mathcal{K}^*}, \chi_\emptyset).$$

Further, we have $\iota(\sigma, x_{\mathcal{K}}, \chi_{\mathcal{T}}) = \sum_h \eta(\{\sigma(h)\}, x_{\mathcal{K}}, \chi_{\mathcal{T}}) - \eta(\mathcal{S}, x_{\mathcal{K}}, \chi_{\mathcal{T}})$.
b) We have $\bar{\iota}(\sigma, \mathcal{K}, \mathcal{T}) = \bar{\iota}(\sigma^\circ, \mathcal{K}, \mathcal{T}^\circ)$ and $\bar{\iota}(\sigma, \mathcal{K}, \mathcal{T}) = \sum_h \bar{\eta}(\{\sigma(h)\}, \mathcal{K}, \mathcal{T}) - \bar{\eta}(\mathcal{S}, \mathcal{K}, \mathcal{T})$.

Proof. Observation 3.41a) implies $\iota(\sigma, x_{\mathcal{K}}, \chi_{\mathcal{T}}) = \sum_h \eta(\{\sigma(h)\}, x_{\mathcal{K}}, \chi_{\mathcal{T}}) - \eta(\mathcal{S}, x_{\mathcal{K}}, \chi_{\mathcal{T}})$, and further Observation 3.39c) yields

$$\begin{aligned}\iota(\sigma, x_{\mathcal{K}}, \chi_{\mathcal{T}}) &= \sum_{i \in \mathcal{S} \setminus \mathcal{K}} |\sigma^{\circ -1}(i)| \eta(\{i\}, x_{\mathcal{K}}, \chi_{\mathcal{T}^\circ}) - \eta(\mathcal{S} \setminus \mathcal{K}, x_{\mathcal{K}}, \chi_{\mathcal{T}^\circ}) \\ &= \sum_{i \in \mathcal{S} \setminus \mathcal{K}^*} |\sigma^{*-1}(i)| \eta(\{i\}, x_{\mathcal{K}^*}, \chi_\emptyset) - \eta(\mathcal{S} \setminus \mathcal{K}^*, x_{\mathcal{K}^*}, \chi_\emptyset),\end{aligned}$$

which establishes the remainder of Part 3.42a), and Part 3.42b) follows by taking expectations. □

Based on Observation 3.42 we assume that $\mathcal{S} \cap \mathcal{K} = \emptyset$ and $\mathcal{T} \cap \mathcal{K} = \emptyset$. Notice that Observation 3.42a) using $\boldsymbol{\mathcal{R}} = \mathcal{T} \cap \boldsymbol{r}^{-1}(1)$ yields the minimal form

$$\bar{\iota}(\sigma, \mathcal{K}, \mathcal{T}) = \mathbb{E}\left[\iota\left(\sigma, \boldsymbol{x}_{\mathcal{K} \cup \boldsymbol{\mathcal{R}}}, \boldsymbol{\chi}_\emptyset\right)\right] = \mathbb{E}\left[\iota\left(\boldsymbol{\sigma}^*, \boldsymbol{x}_{\mathcal{K} \cup \boldsymbol{\mathcal{R}}}, \boldsymbol{\chi}_\emptyset\right)\right] = \mathbb{E}\left[\mathbb{E}\left[I(\boldsymbol{x}_{\boldsymbol{\sigma}^*} | \boldsymbol{x}_{\mathcal{K} \cup \boldsymbol{\mathcal{R}}}) | \boldsymbol{\mathcal{R}}\right]\right],$$

where $\boldsymbol{\sigma}^* = \sigma_{\boldsymbol{\mathcal{D}}}$ with $\boldsymbol{\mathcal{D}} = \sigma^{-1}(\mathcal{S} \setminus \boldsymbol{\mathcal{R}})$. This representation reflects the approach in [33].

Remark 3.43. In the proof of Lemma 3.40 we have already seen a recursive structure. Further, Observation 3.42b) yields a representation of $\bar{\iota}$ as a linear combination of $\bar{\eta}$-terms, while Observation 3.41b) applied to Lemma 3.40 yields a representation of the derivative as a linear combination of $\bar{\iota}$-terms (and hence $\bar{\eta}$-terms). This is one way to obtain all higher derivatives of $\bar{\eta}$ and $\bar{\iota}$.

3.3.1.4 The Pinning Lemma. In this section we prove Lemma 3.3. Recall the pinning operation $[\mu]^{\downarrow}_{\mathcal{U},\check{\sigma}}$ from Section 3.1.2.3 and let $\boldsymbol{\mu}_{n,\mu,p} = [\mu]^{\downarrow}_{\boldsymbol{r}^{-1}(1),\boldsymbol{x}}$.

Lemma 3.44. *We have $\bar{\iota}(\sigma, \emptyset, [n]) = \mathbb{E}[\mathbb{E}[I(\boldsymbol{x}_{\boldsymbol{\mu},\sigma}) | \boldsymbol{\mu}]]$.*

Proof. Notice that $\iota(\sigma, x_{\mathcal{K}}, \chi_\emptyset) = I(\boldsymbol{x}_{\mu,\sigma} | \boldsymbol{x}_{\mathcal{K}} = x_{\mathcal{K}}) = I(\boldsymbol{x}_{[\mu]^{\downarrow}_{\mathcal{K},x},\sigma})$, since by definition $[\mu]^{\downarrow}_{\mathcal{K},x}$ is the law of $\boldsymbol{x}_\mu | \boldsymbol{x}_{\mu,\mathcal{K}} = x_{\mathcal{K}}$. Now, the assertion follows from Observation 3.42a) since $\bar{\iota}(\sigma, \emptyset, [n]) = \mathbb{E}[\iota(\sigma, \boldsymbol{x}_{\boldsymbol{\mathcal{R}}}, \boldsymbol{\chi}_\emptyset)] = \mathbb{E}[\mathbb{E}[I(\boldsymbol{x}_{\boldsymbol{\mu},\sigma}) | \boldsymbol{\mu}]]$. □

Fix $\ell \in \mathbb{Z}_{>0}$ and $p \in [0,1]$. Let $p_{/} = (p)_{i\in[n]}$, $\boldsymbol{v}_{n,\ell} \sim \mathrm{u}([n]^{\ell})$, $\eta^*_{n,\mu,\ell}(p) = \mathbb{E}[\overline{\eta}_{p_/}(\boldsymbol{v}([\ell]), \emptyset, [n])]$, $\iota^*_{n,\mu,\ell}(p) = \mathbb{E}[\overline{\iota}_{p_/}(\boldsymbol{v}, \emptyset, [n])]$ and $\delta^*_{n,\mu,\ell}(p) = \iota^*_{n,\mu,\ell+1}(p) - \iota^*_{n,\mu,\ell}(p)$.

Lemma 3.45. *Let $\ell \in \mathbb{Z}_{>0}$ and $\boldsymbol{v} = \boldsymbol{v}_{n,\ell+1} \in [n]^{\ell+1}$.*
a) We have $\delta^*(p) = I(\boldsymbol{x}_{\boldsymbol{v}_{[\ell]}}, \boldsymbol{x}_{\mu,\boldsymbol{v}_{\ell+1}} | \boldsymbol{\chi}_{p_/}, \boldsymbol{v})$.
b) We have $\frac{\partial}{\partial p}\eta^*(p) = -\frac{n}{1-p}\delta^*(p)$.

Proof. Notice that $\boldsymbol{v}_{[\ell]} \sim \boldsymbol{v}_\circ$ with $\boldsymbol{v}_\circ = \boldsymbol{v}_{n,\ell} \in [n]^\ell$, let $\boldsymbol{\chi} = \boldsymbol{\chi}_{p_/}$ and $\boldsymbol{a} = (\boldsymbol{\chi}_{p_/}, \boldsymbol{v})$. We have

$$\delta^*(p) = I(\boldsymbol{x}_{\boldsymbol{v}}|\boldsymbol{a}) - I(\boldsymbol{x}_{\boldsymbol{v}_{[\ell]}}|\boldsymbol{a}) = I(\boldsymbol{x}_{\boldsymbol{v}}|\boldsymbol{a}) - I(\boldsymbol{x}_{\boldsymbol{v}_{[\ell]}}|\boldsymbol{a}) - I(\boldsymbol{x}_{\boldsymbol{v}_{\ell+1}}|\boldsymbol{a}) = I(\boldsymbol{x}_{\boldsymbol{v}_{[\ell]}}, \boldsymbol{x}_{\mu,\boldsymbol{v}_{\ell+1}}|\boldsymbol{a}),$$

with Observation 3.41b), since $I(\boldsymbol{b}|\boldsymbol{a}) = D_{\mathrm{KL}}(\boldsymbol{b}\|\boldsymbol{b}|\boldsymbol{a}) = 0$ for $\boldsymbol{b} \in \mathcal{B}^1$. For Part 3.45b) we have

$$\begin{aligned}\frac{\partial}{\partial p}\eta^*(p) &= \sum_{i\in[n]} \mathbb{E}\left[\frac{\partial}{\partial p_i}\overline{\eta}_{p_/}(\boldsymbol{v}([\ell]), \emptyset, [n])\right] = -\sum_{i\in[n]} I\left(\boldsymbol{x}_{\boldsymbol{v}([\ell])}, \boldsymbol{x}_i | \boldsymbol{\chi}_{[n]\setminus\{i\}}, \boldsymbol{v}\right)\\ &= -nI\left(\boldsymbol{x}_{\boldsymbol{v}([\ell])}, \boldsymbol{x}_{\boldsymbol{v}_{\ell+1}} | \boldsymbol{\chi}_{[n]\setminus\{\boldsymbol{v}_{\ell+1}\}}, \boldsymbol{v}\right) = -nI\left(\boldsymbol{x}_{\boldsymbol{v}_{[\ell]}}, \boldsymbol{x}_{\boldsymbol{v}_{\ell+1}} | \boldsymbol{\chi}_{[n]\setminus\{\boldsymbol{v}_{\ell+1}\}}, \boldsymbol{v}\right),\end{aligned}$$

using Lemma 3.40 and the chain rule. For given v and χ with $i = v_{\ell+1}$ we have

$$\begin{aligned}\iota_v(\chi) &= I(\boldsymbol{x}_{v_{[\ell]}}, \boldsymbol{x}_i | \boldsymbol{\chi} = \chi) = (1 - r_i) I(\boldsymbol{x}_{v_{[\ell]}}, \boldsymbol{x}_i | \boldsymbol{x}_{\mathcal{R}\setminus\{i\}} = x_{\mathcal{R}\setminus\{i\}})\\ &= (1 - r_i) I(\boldsymbol{x}_{v_{[\ell]}}, \boldsymbol{x}_i | \boldsymbol{\chi}_{[n]\setminus\{i\}} = \chi_{[n]\setminus\{i\}}).\end{aligned}$$

Due to independence this gives $I(\boldsymbol{x}_{v_{[\ell]}}, \boldsymbol{x}_i|\boldsymbol{\chi}) = \mathbb{E}[\iota_v(\boldsymbol{\chi})] = (1-p)I(\boldsymbol{x}_{v_{[\ell]}}, \boldsymbol{x}_i|\boldsymbol{\chi}_{[n]\setminus\{i\}})$. This completes the proof by taking the expectation over $\boldsymbol{v}$ and using the first part. □

An immediate consequence of Lemma 3.45 is a uniform bound for the integral over ι^*.

Corollary 3.46. *For $\ell \in \mathbb{Z}_{>0}$ we have $\int_0^1 \iota^*(p)\mathrm{d}p \le \binom{\ell}{2}\ln(q)/n$.*

Proof. With $\iota^* \equiv 0$ for $\ell = 1$, a telescoping sum and Lemma 3.45b) we have

$$\begin{aligned}\int_0^1 \iota^*(p)\mathrm{d}p &= \int_0^1 \sum_{\ell'=1}^{\ell-1} \delta^*(p)\mathrm{d}p \le \int_0^1 \sum_{\ell'=1}^{\ell-1} \frac{\delta^*(p)}{1-p}\mathrm{d}p = \frac{1}{n}\sum_{\ell'=1}^{\ell-1}(\eta^*_{\ell'}(0) - \eta^*_{\ell'}(1))\\ &= \frac{1}{n}\sum_{\ell'=1}^{\ell-1} H(\boldsymbol{x}_{\boldsymbol{v}_{\ell'}}|\boldsymbol{v}_{\ell'}) \le \frac{1}{n}\sum_{\ell'=1}^{\ell-1}\ln(q^{\ell'}) = \frac{\ln(q)}{n}\binom{\ell}{2}.\end{aligned}$$

□

With $\boldsymbol{\mu}^* = [\mu]^{\downarrow}_{\boldsymbol{\mathcal{U}},\boldsymbol{\sigma}}$ as defined in Lemma 3.3 we have $\boldsymbol{\mu}^* \sim \boldsymbol{\mu}_{\boldsymbol{p}_/}$ with $\boldsymbol{p}_/ = (\boldsymbol{p})_{i\in[n]}$ and $\boldsymbol{p} \sim \mathrm{u}([0,P])$, $P = \Theta^{\downarrow}/n$, so with Lemma 3.44, Corollary 3.46 and $(\boldsymbol{\mu}^*, \boldsymbol{v}_\ell) \sim \boldsymbol{\mu}^* \otimes \boldsymbol{v}_\ell$ we get

$$\begin{aligned}\mathbb{E}[\mathbb{E}[I(\boldsymbol{x}_{\boldsymbol{\mu}^*,\boldsymbol{v}_\ell})|\boldsymbol{\mu}^*, \boldsymbol{v}_\ell]] &= \int_0^P \frac{n}{\Theta^{\downarrow}}\mathbb{E}[\mathbb{E}[I(\boldsymbol{x}_{\boldsymbol{\mu},\boldsymbol{v}_\ell})|\boldsymbol{\mu}_{p_/}, \boldsymbol{v}_\ell]]\mathrm{d}p = \frac{n}{\Theta^{\downarrow}}\int_0^P \mathbb{E}[\overline{\iota}_{p_/}(\boldsymbol{v}, \emptyset, [n])]\mathrm{d}p\\ &= \frac{n}{\Theta^{\downarrow}}\int_0^P \iota^*(p)\mathrm{d}p \le \frac{n}{\Theta^{\downarrow}}\int_0^1 \iota^*(p)\mathrm{d}p \le \frac{\binom{\ell}{2}\ln(q)}{\Theta^{\downarrow}}.\end{aligned}$$

3.3.1.5 Asymptotic Independence of Gibbs Spins. We start with a few basic results.

Observation 3.47. *Let $\Theta^{\downarrow} \le n$ and $\theta \le \Theta^{\downarrow}$. We have $|\check{\boldsymbol{u}}_{\mathrm{t}}^{-1}(1)| \sim \mathrm{Bin}(n, \theta/n)$, $\mathbb{E}[\boldsymbol{\theta}] = \Theta^{\downarrow}/2$, $\mathbb{E}[\check{\boldsymbol{u}}_{\mathrm{t}\circ}] = \Theta^{\downarrow}/(2n)$ and $\mathbb{E}[|\boldsymbol{\mathcal{U}}|] = \Theta^{\downarrow}/2$.*

Proof. The proof is left as an exercise to the reader. □

The following result is one of the main reasons to work with the Nishimori ground truth. Recall the notions from Section 3.1.2.3, in particular $\iota_\circ$, ι and $\boldsymbol{v}$ for $\ell \in \mathbb{Z}_{\ge 0}$. Notice that $\iota_\circ \equiv 0$ for $\ell \le 1$.

Proposition 3.48. *Let $\boldsymbol{G}^*(\sigma) = \boldsymbol{G}^*_{\boldsymbol{\mathcal{U}}}(\sigma)$, $\hat{\boldsymbol{\mu}} = \mu_{\mathrm{g},\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})}$, $\boldsymbol{\mu}^* = \mu_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ and $\ell \in \mathbb{Z}_{\ge 0}$.*
a) We have $\mathbb{E}[\iota(\hat{\boldsymbol{\mu}})], \mathbb{E}[\iota(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow}})] \le \binom{\ell}{2}\ln(q)/\Theta^{\downarrow}$.
b) There exists $C_{\mathfrak{g}} \in (0,1) \times \mathbb{R}_{>0}$ such that for $c \in (0, C_1]$ and $m \le m_{\uparrow}$ we have

$$\mathbb{E}[\iota(\boldsymbol{\mu}^*)], \mathbb{E}[\iota(\boldsymbol{\mu}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow}})] \le C_2(\ell-1)\left(\frac{\ell}{\Theta^{\downarrow}}\right)^c.$$

Proof. Notice that for any G we have $\mu_{\mathrm{g},[G]^{\downarrow}_{\mathcal{U},\sigma}} = [\mu_{\mathrm{g},G}]^{\downarrow}_{\mathcal{U},\sigma}$ and that $\boldsymbol{\mathcal{U}}$ as defined in Section 3.2.1.1 coincides with $\boldsymbol{\mathcal{U}}$ from Lemma 3.3, so with $(\boldsymbol{\mathcal{U}}, \boldsymbol{\sigma}_{\mathrm{g},G}) \sim \boldsymbol{\mathcal{U}} \otimes \boldsymbol{\sigma}_{\mathrm{g},G}$, $\boldsymbol{G}(G) = [G]^{\downarrow}_{\boldsymbol{\mathcal{U}},\boldsymbol{\sigma}_{\mathrm{g},G}}$ we have $\mu_{\mathrm{g},\boldsymbol{G}(G)} = [\mu_{\mathrm{g},G}]^{\downarrow}_{\boldsymbol{\mathcal{U}},\boldsymbol{\sigma}_{\mathrm{g},G}}$ and hence Lemma 3.3 yields $\mathbb{E}[\iota(\mu_{\mathrm{g},\boldsymbol{G}(G)})] \le \binom{\ell}{2}\ln(q)/\Theta^{\downarrow}$. Since this holds for any G, the expectation for the unpinned graph $\boldsymbol{G}^\circ(\sigma) = [\boldsymbol{w}^*(\sigma)]^{\Gamma\leftrightarrow}_{m^{\leftrightarrow},\psi^{*\leftrightarrow}}$ is also bounded by $\mathbb{E}[\iota(\mu_{\mathrm{g},\boldsymbol{G}(\boldsymbol{G}^\circ(\sigma))})] \le \binom{\ell}{2}\ln(q)/\Theta^{\downarrow}$. Notice that by Observation 3.13 the graphs $\boldsymbol{G}^*(\sigma)$ and $\boldsymbol{G}(\boldsymbol{G}^\circ(\sigma))$ differ exactly in the choice of the pinning assignment, i.e. σ for the former and $\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^\circ(\sigma)}$ for the latter. Since this bound holds for any σ, it holds for $\hat{\boldsymbol{\sigma}}$. But with Observation 3.22d) (for $\mathcal{U} = \emptyset$) we have $(\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^\circ(\hat{\boldsymbol{\sigma}})}, \boldsymbol{G}^\circ(\hat{\boldsymbol{\sigma}})) \sim (\hat{\boldsymbol{\sigma}}, \boldsymbol{G}^\circ(\hat{\boldsymbol{\sigma}}))$ and using Observation 3.13 further

$$\boldsymbol{G}(\boldsymbol{G}^\circ(\hat{\boldsymbol{\sigma}})) = [\boldsymbol{G}^\circ(\hat{\boldsymbol{\sigma}})]^{\downarrow}_{\boldsymbol{\mathcal{U}},\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^\circ(\hat{\boldsymbol{\sigma}})}} \sim [\boldsymbol{G}^\circ(\hat{\boldsymbol{\sigma}})]^{\downarrow}_{\boldsymbol{\mathcal{U}},\hat{\boldsymbol{\sigma}}} \sim \boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}).$$

Thus, we have $\mathbb{E}[\iota(\hat{\boldsymbol{\mu}})] \le \binom{\ell}{2}\ln(q)/\Theta^{\downarrow}$, and $\mathbb{E}[\iota(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow}})] \le \binom{\ell}{2}\ln(q)/\Theta^{\downarrow}$ by taking expectations.

Now, let $r \in \mathbb{R}_{>0}$, $c_{\mathrm{m}} \in \mathbb{R}^2_{>0}$ from Corollary 3.12, $c^* \in \mathbb{R}^2_{>0}$ from Observation 3.23b) and $\hat{c} \in \mathbb{R}_{>0}$ from Corollary 3.25b). Then for $\ell > 0$ we have

$$\iota_\circ(\mu, v) = \sum_h H(\mu|_{v(h)}) - H(\mu|v) = \sum_{h>1} H(\mu|_{v(h)}) - H(\mu|_v|\mu|_{v(1)}) \le (\ell-1)\ln(q),$$

so we have $\iota(\mu) \le (\ell-1)\ln(q)$. Hence, for $E_1 = \mathbb{E}[\iota(\boldsymbol{\mu}^*)]$, $E_2 = \mathbb{E}[\iota(\boldsymbol{\mu}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow}})]$ and $\ell \ge \Theta^{\downarrow}$ we have $E_1, E_2 \le c_2(\ell-1)(\ell/\Theta^{\downarrow})^{c_1}$ for $c \in \mathbb{R}_{>0} \times \mathbb{R}_{\ge \ln(q)}$. Otherwise, we have

$$E_1 \le e^{\hat{c}r^2}\mathbb{E}\left[\mathbb{1}\{\|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}} < r/\sqrt{n}\}\iota(\hat{\boldsymbol{\mu}})\right] + \delta,$$
$$E_2 \le e^{\hat{c}r^2}\mathbb{E}\left[\mathbb{1}\{\|\hat{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}} < r/\sqrt{n}, \boldsymbol{m} \le m_{\uparrow}\}\iota(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow}})\right] + \delta,$$

with $\delta = (\ell-1)\ln(q)(c^*_2 e^{-c^*_1 r^2} + c_{\mathrm{m},2}e^{-\tilde{c}n})$ and $\tilde{c} = c_{\mathrm{m},1}d^2_{\uparrow}/(1+d_{\uparrow})$ obtained by choosing $r = d_{\uparrow}$ in Corollary 3.12 to enforce $\bar{\boldsymbol{d}} \le t^{\leftrightarrow}\bar{d} + d_{\uparrow} \le 2d_{\uparrow}$. With Part 3.48a) and $i \in [2]$ we have

$$E_i \le e^{\hat{c}r^2}\frac{\ln(q)}{\Theta^{\downarrow}}\binom{\ell}{2} + \delta = (\ell-1)\ln(q)\left(\frac{\ell e^{\hat{c}r^2}}{2\Theta^{\downarrow}} + c^*_2 e^{-c^*_1 r^2} + c_{\mathrm{m},2}e^{-\tilde{c}n}\right).$$

In order to compensate the last contribution notice that $\ell/\Theta^{\downarrow} \geq 1/n$, $e^{\hat{c}r^2} \geq 1$ and hence

$$E_i \leq (\ell-1)\ln(q)\left(c'\frac{\ell e^{\hat{c}r^2}}{\Theta^{\downarrow}} + c_2^* e^{-c_1^* r^2}\right),\ c' = \frac{1}{2} + c_{\mathrm{m},2}\max_{n>0} ne^{-\tilde{c}n}.$$

So, with $C_2 = 2\ln(q)\max(c', c_2^*)$, $r = \sqrt{\frac{1}{\hat{c}+c_1^*}\ln(\frac{\Theta^{\downarrow}}{\ell})} > 0$, $C_1 = \frac{c_1^*}{\hat{c}+c_1^*}$ and for $\ell > 0$, $c \in (0, C_1]$ we have $E_i \leq c_2(\ell-1)(\ell/\Theta^{\downarrow})^c$, and Part 3.48b) holds since the assertion is trivial for $\ell = 0$. □

Remark 3.49. The relative entropy in Section 3.1.2.3 can be extended to general f-divergences, in particular to the total variation. Thus, let $\nu_\circ(\mu, v) = \|\mu|_v - \bigotimes_h \mu|_{v(h)}\|_{\mathrm{tv}}$ and $\nu_\ell(\mu) = \mathbb{E}[\nu_\circ(\mu, \boldsymbol{v})]$. For $\varepsilon \in \mathbb{R}_{\geq 0}$ let $\mathcal{E}_{\mathrm{s},n,\varepsilon,\ell} = \{\mu \in \mathcal{P}([q]^n) : \nu(\mu) \leq \varepsilon\}$ be the (ε,ℓ)-symmetric measures. Pinsker's inequality 3.6f) yields $\nu_\circ(\mu, v) \leq \sqrt{\iota_\circ(\mu, v)/2}$ and hence $\nu(\mu) \leq \sqrt{\iota(\mu)/2}$ using Jensen's inequality.

3.3.1.6 Pinning Impact on the Quenched Free Entropy. We bound the distance of the pinned and the unpinned quenched free entropy.

Proposition 3.50. *There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that*

$$0 \leq \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow},\emptyset}(\boldsymbol{\sigma}^*))] - \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*))] \leq \frac{c\Theta^{\downarrow}}{n}$$

and the same result holds for $\boldsymbol{\sigma}^*$ replaced by $\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}$.

Proof. Let $\boldsymbol{G}^* = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow},\emptyset}(\boldsymbol{\sigma}^*)$ and $\boldsymbol{G}^{\downarrow} = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*)$. Recall that $\boldsymbol{G}^{\downarrow} \sim [\boldsymbol{G}^*]^{\downarrow}_{\boldsymbol{\mathcal{U}},\boldsymbol{\sigma}^*}$ from Observation 3.13, so given $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*)$ we can obtain $\boldsymbol{G}^{\downarrow}$ by choosing $\boldsymbol{\mathcal{U}}$, which means that $\boldsymbol{G}^{\downarrow}$ and $\boldsymbol{G}^*$ then exactly differ in the pins. Hence, using this coupling we have $\phi_{\mathrm{g}}(\boldsymbol{G}^{\downarrow}) \leq \phi_{\mathrm{g}}(\boldsymbol{G}^*)$ and using the notions from Section 3.2.3.1 with $\boldsymbol{G}^* = (\boldsymbol{v}^*, \boldsymbol{\psi}^*)$ further

$$\mathrm{d_g}(\boldsymbol{G}^*, \boldsymbol{G}^{\downarrow}) = 2\left(|\{a \in [\boldsymbol{m}] : \boldsymbol{v}^*_a([k]) \cap \boldsymbol{\mathcal{U}} \neq \emptyset\}| + \sum_{i \in \boldsymbol{\mathcal{U}}} \boldsymbol{m}^{\leftrightarrow}_i\right).$$

Bounding $\mathrm{d_g}$, using the sum over the (factor) degrees of $i \in \boldsymbol{\mathcal{U}}$ and c' from Corollary 3.18a), yields

$$\mathbb{E}\left[\mathrm{d_g}(\boldsymbol{G}^*, \boldsymbol{G}^{\downarrow})\right] \leq 2\mathbb{E}\left[\left(\frac{c'k\boldsymbol{m}}{n} + (1-t^{\leftrightarrow})\bar{d}\right)|\boldsymbol{\mathcal{U}}|\right] = \left(t^{\leftrightarrow}c' + 1 - t^{\leftrightarrow}\right)\bar{d}\Theta^{\downarrow},$$

using Observation 3.47. With c'' from Observation 3.30 we have $|\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*)] - \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^{\downarrow})]| \leq \frac{c\Theta^{\downarrow}}{n}$ for $c = c''(c'+1)d_{\uparrow}$, and the result for $\hat{\boldsymbol{\sigma}}$ follows analogously. □

3.3.1.7 Reweighted Marginal Distributions. Recall the set $\mathcal{P}^2([q])$ from Section 2.1.2.2. In the following we use Observation 3.6 implicitly when working with random measures, if required. The empirical marginal distribution $\pi_\mu \in \mathcal{P}^2([q])$ of $\mu \in \mathcal{P}([q]^n)$ is given by

$$\pi(\mathcal{E}) = \frac{1}{n}|\{i \in [n] : \mu|_i \in \mathcal{E}\}|$$

for an event $\mathcal{E} \subseteq \mathcal{P}([q])$. Let $\boldsymbol{\sigma}_\mu \sim \mu$, $\boldsymbol{\gamma}_\pi \sim \pi$ and $\overline{\gamma}_\pi = \mathbb{E}[\boldsymbol{\gamma}]$. For $\tau \in [q]$ we have

$$\overline{\gamma}(\tau) = \sum_{i\in[n]} \frac{1}{n}\mu|_i(\tau) = \mu|_*(\tau) = \mathbb{E}\left[\gamma_{\mathrm{n},\boldsymbol{\sigma}}(\tau)\right]. \tag{7}$$

For $\sigma \in [q]^n$ and $\tau \in \sigma([n])$ let $\check{\pi}_{\mu,\sigma,\tau} \in \mathcal{P}^2([q])$ be given by

$$\check{\pi}(\mathcal{E}) = \frac{1}{|\sigma^{-1}(\tau)|}|\{i \in [n] : \sigma_i = \tau, \mu|_i \in \mathcal{E}\}.$$

For $\tau \in \overline{\gamma}^{-1}(\mathbb{R}_{>0})$ let $\hat{\pi}_{\mu,\tau} \in \mathcal{P}^2([q])$ be given by the $(\hat{\pi},\pi)$-derivative $\gamma \mapsto \gamma(\tau)/\overline{\gamma}(\tau)$. Recall the couplings $\Gamma(\check{\pi},\hat{\pi})$ from Section 3.1.6 and for $\sigma \in [q]^n$, $\tau \in \sigma([n]) \cap \overline{\gamma}^{-1}(\mathbb{R}_{>0})$ let

$$\mathrm{d_w}(\check{\pi},\hat{\pi}) = \inf_{\rho\in\Gamma(\check{\pi},\hat{\pi})} \mathbb{E}[\|\boldsymbol{\gamma}_{\rho,1} - \boldsymbol{\gamma}_{\rho,2}\|_{\mathrm{tv}}],\ \boldsymbol{\gamma}_\rho \sim \rho,$$

be the Wasserstein distance of $\check{\pi}$ and $\hat{\pi}$. Recall $\mathcal{E}_{\mathrm{s},\varepsilon,\ell}$ from Remark 3.49.

Proposition 3.51. *Let $\delta, \varepsilon, \varepsilon_{\mathrm{s}} \in \mathbb{R}_{>0}$ and $\mu \in \mathcal{P}([q]^n)$ be such that $\overline{\gamma} \geq \psi_\downarrow/2$, $\mu \in \mathcal{E}_{\mathrm{s},\varepsilon_{\mathrm{s}},2}$ and $\mathbb{P}(\|\gamma_{\mathrm{n},\boldsymbol{\sigma}} - \overline{\gamma}\|_{\mathrm{tv}} > \delta) \leq \varepsilon$. Then there exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that*

$$\mathbb{E}\left[\sum_{\tau\in\boldsymbol{\sigma}([n])} \mathrm{d_w}(\check{\pi}_{\boldsymbol{\sigma},\tau},\hat{\pi}_\tau)\right] \leq c(\delta + \varepsilon + \varepsilon_{\mathrm{s}}^{1/(2q+1)}).$$

Proof. If $\max(\delta,\varepsilon,\varepsilon_{\mathrm{s}}) \geq 1$, then the assertion holds with $c \geq q$ since the left hand side is at most q, so let $\delta,\varepsilon,\varepsilon_{\mathrm{s}} \in (0,1]$. Notice that $\overline{\gamma}^{-1}(\mathbb{R}_{>0}) = \bigcup_{\sigma\in\mu^{-1}(\mathbb{R}_{>0})} \sigma([n])$, let $\sigma \in \mu^{-1}(\mathbb{R}_{>0})$ and $\tau \in \sigma([n])$. Let $a \in \mathbb{Z}_{>0}$, $\mathcal{B}^* = (a^{-1}\mathbb{Z})^{q-1}$, $\mathcal{Q}_\circ = [-1/(2a), 1/(2a))^{q-1}$ and $\mathcal{Q}_b = b + \mathcal{Q}_\circ$ for $b \in \mathcal{B}^*$, then $(\mathcal{Q}_b)_{b\in\mathcal{B}^*}$ is a partition of $\mathbb{R}^{q-1}$. This induces a partition $(\mathcal{Q}_b)_{b\in\mathcal{B}}$ of $\mathcal{P}([q])$, where

$$\mathcal{B} = \{(b^*, 1 - \|b^*\|_1) : b^* \in \mathcal{B}^*\} \cap \mathcal{P}([q]),$$
$$\mathcal{Q}_b = \{(b^*, 1 - \|b^*\|_1) : b^* \in \mathcal{Q}^*_{b_{[q-1]}}\},\ b \in \mathcal{B}.$$

Notice that $|\mathcal{B}| \leq (a+1)^{q-1} = |\mathcal{B}^* \cap [0,1]^{q-1}|$. For $\gamma \in \mathcal{Q}_b$ we have $\tilde{\gamma} \in \mathcal{Q}_{\tilde{b}}$, where $\tilde{\gamma} = \gamma_{[q-1]}$ and $\tilde{b} = b_{[q-1]}$, hence

$$\|\gamma - b\|_{\mathrm{tv}} = \frac{1}{2}\|\tilde{\gamma} - \tilde{b}\|_1 + \frac{1}{2}|1 - \|\tilde{\gamma}\|_1 - (1 - \|\tilde{b}\|_1)| \leq \|\tilde{\gamma} - \tilde{b}\|_1 \leq \frac{q-1}{2a}.$$

Next, let $\mathcal{I}_b = \{i \in [n] : \mu|_i \in \mathcal{Q}_b\}$ and $\check{\mathcal{I}}_{\sigma,\tau,b} = \{i \in [n] : \sigma_i = \tau, \mu|_i \in \mathcal{Q}_b\}$. Then we have

$$\pi(\mathcal{Q}_b) = \frac{|\mathcal{I}_b|}{n},\ \check{\pi}(\mathcal{Q}_b) = \frac{|\check{\mathcal{I}}_{\sigma,\tau,b}|}{|\sigma^{-1}(\tau)|}.$$

The expectation $\bar{I}_{\tau,b} = \mathbb{E}[|\check{\mathcal{I}}_{\boldsymbol{\sigma},\tau,b}|]$ is given by

$$\bar{I}_{\tau,b} = \sum_i \mathbb{1}\{\mu|_i \in \mathcal{Q}_b\}\mathbb{P}(\boldsymbol{\sigma}_i = \tau) = \sum_i \mathbb{1}\{\mu|_i \in \mathcal{Q}_b\}\mu|_i(\tau) = n\overline{\gamma}(\tau)\hat{\pi}(\mathcal{Q}_b).$$

The variances, using $\mu \in \mathcal{E}_{\mathrm{s},n,\varepsilon_{\mathrm{s}},2}$, are given by

$$\begin{aligned}
V &= \sum_{b,\tau} \mathrm{Var}(|\check{\mathcal{I}}_{\boldsymbol{\sigma},\tau,b}|) = \sum_{b,\tau} \left(\mathbb{E}\left[|\check{\mathcal{I}}_{\boldsymbol{\sigma},\tau,b}|^2\right] - \mathbb{E}\left[|\check{\mathcal{I}}_{\boldsymbol{\sigma},\tau,b}|\right]^2 \right) \\
&= \sum_{b,\tau} \sum_{v\in[n]^2} \mathbb{1}\left\{\mu|_{v(1)}, \mu|_{v(2)} \in \mathcal{Q}_b\right\} \left(\mathbb{P}\left(\boldsymbol{\sigma}|_v = (\tau,\tau)\right) - \mathbb{P}\left(\boldsymbol{\sigma}|_{v(1)} = \tau\right) \mathbb{P}\left(\boldsymbol{\sigma}|_{v(2)} = \tau\right)\right) \\
&\le \sum_{v\in[n]^2} \sum_{\tau\in[q]^2} \left|\mathbb{P}\left(\boldsymbol{\sigma}|_v = \tau\right) - \mathbb{P}\left(\boldsymbol{\sigma}|_{v(1)} = \tau_1\right) \mathbb{P}\left(\boldsymbol{\sigma}|_{v(2)} = \tau_2\right)\right| \\
&= 2 \sum_{v\in[n]^2} \left\|\mu|_v - \mu|_{v(1)} \otimes \mu|_{v(2)}\right\|_{\mathrm{tv}} \le 2n^2\varepsilon_{\mathrm{s}},
\end{aligned}$$

where in the extension of the summation region b is determined by $\mu|_{v(1)}$, i.e. the unique point b with $\mu|_{v(1)} \in \mathcal{Q}_b$, while we drop the restriction $\mu|_{v(2)} \in \mathcal{Q}_b$. With the union bound and Chebyshev's inequality we have

$$\mathbb{P}\left(\sum_{b,\tau} \left||\check{\mathcal{I}}_{\boldsymbol{\sigma},\tau,b}| - \bar{I}_{\tau,b}\right| \ge rn\right) \le \sum_{b,\tau} \mathbb{P}\left(\left||\check{\mathcal{I}}_{\boldsymbol{\sigma},\tau,b}| - \bar{I}_{\tau,b}\right| \ge \frac{rn}{q|\mathcal{B}|}\right) \le \frac{Vq^2|\mathcal{B}|^2}{r^2n^2} \le \frac{2q^2|\mathcal{B}|^2\varepsilon_{\mathrm{s}}}{r^2}$$

for $r = \varepsilon_{\mathrm{s}}^{1/(1+2q)}$. Next, we recall that the color frequencies concentrate and let

$$\mathcal{S} = \left\{\sigma \in [q]^n : \|\gamma_{\mathrm{n},\sigma} - \overline{\gamma}\|_{\mathrm{tv}} \le \delta, \sum_{b,\tau} \left||\check{\mathcal{I}}_{\sigma,\tau,b}| - \bar{I}_{\tau,b}\right| < rn\right\}.$$

Further, notice that $\mathrm{d_w}(\check{\pi}, \hat{\pi}) \in [0,1]$ since $\|\cdot\|_{\mathrm{tv}} \in [0,1]$, so

$$\mathbb{E}\left[\sum_{\tau\in\boldsymbol{\sigma}([n])} \mathrm{d_w}\left(\check{\pi}_{\boldsymbol{\sigma},\tau}, \hat{\pi}_\tau\right)\right] \le \mathbb{E}\left[\mathbb{1}\{\boldsymbol{\sigma} \in \mathcal{S}\} \sum_{\tau\in\boldsymbol{\sigma}([n])} \mathrm{d_w}\left(\check{\pi}_{\boldsymbol{\sigma},\tau}, \hat{\pi}_\tau\right)\right] + q\varepsilon + \frac{2q^3|\mathcal{B}|^2\varepsilon_{\mathrm{s}}}{r^2}.$$

For $\sigma \in \mathcal{S}$, $\tau \in \sigma([n])$ and $b \in \mathcal{B}$ we have

$$\begin{aligned}
\Delta_{\sigma,\tau}(b) = |\check{\pi}(\mathcal{Q}_b) - \hat{\pi}(\mathcal{Q}_b)| &\le \left|\frac{|\check{\mathcal{I}}|}{|\sigma^{-1}(\tau)|} - \frac{|\check{\mathcal{I}}|}{n\overline{\gamma}(\tau)}\right| + \left|\frac{|\check{\mathcal{I}}|}{n\overline{\gamma}(\tau)} - \frac{\bar{I}}{n\overline{\gamma}(\tau)}\right| \\
&= \frac{1}{\overline{\gamma}(\tau)}\left(\frac{|\check{\mathcal{I}}|}{|\sigma^{-1}(\tau)|}|\overline{\gamma}(\tau) - \gamma_{\mathrm{n},\sigma}(\tau)| + \frac{1}{n}\left||\check{\mathcal{I}}| - \bar{I}\right|\right).
\end{aligned}$$

With $\sum_b \check{\mathcal{I}}_b = |\sigma^{-1}(\tau)|$, $\overline{\gamma} \ge \psi_\downarrow/2$ and $\|\gamma_{\mathrm{n},\sigma} - \overline{\gamma}\|_1 = 2\|\gamma_{\mathrm{n},\sigma} - \overline{\gamma}\|_{\mathrm{tv}} \le 2\delta$ we have

$$\sum_{b,\tau} \Delta_{\sigma,\tau}(b) < \frac{4\delta + 2r}{\psi_\downarrow}.$$

Now, let $\check{\boldsymbol{\gamma}} \sim \check{\pi}$ and $\hat{\boldsymbol{\gamma}} \sim \hat{\pi}$. For $\gamma \in \mathcal{P}([q])$ let $b(\gamma) \in \mathcal{B}$ be the unique index with $\gamma \in \mathcal{Q}_b$. Then we have $\sum_b \Delta_{\sigma,\tau}(b) = 2\|b(\check{\boldsymbol{\gamma}}) - b(\hat{\boldsymbol{\gamma}})\|_{\mathrm{tv}}$. With the coupling lemma we obtain a coupling of $b(\check{\boldsymbol{\gamma}})$ and $b(\hat{\boldsymbol{\gamma}})$ that extends to a coupling of $\check{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ via $(\check{\boldsymbol{\gamma}}|b(\check{\boldsymbol{\gamma}}) = \check{b}) \otimes (\hat{\boldsymbol{\gamma}}|b(\hat{\boldsymbol{\gamma}}) = \hat{b})$ given $(b(\check{\boldsymbol{\gamma}}), b(\hat{\boldsymbol{\gamma}})) = (\check{b}, \hat{b})$ (by an abuse of notation in that we use the same notation for the coupling). The triangle inequality yields $\|\check{\boldsymbol{\gamma}}-\hat{\boldsymbol{\gamma}}\|_{\mathrm{tv}} \leq \frac{q-1}{a}+\|b(\check{\boldsymbol{\gamma}})-b(\hat{\boldsymbol{\gamma}})\|_{\mathrm{tv}}$ and hence

$$\mathbb{P}\left(\|\check{\boldsymbol{\gamma}}-\hat{\boldsymbol{\gamma}}\|_{\mathrm{tv}}>\frac{q-1}{a}\right) \leq \mathbb{P}(b(\check{\boldsymbol{\gamma}}) \neq b(\hat{\boldsymbol{\gamma}}))=\|b(\check{\boldsymbol{\gamma}})-b(\hat{\boldsymbol{\gamma}})\|_{\mathrm{tv}}=\frac{1}{2} \sum_{b} \Delta_{\sigma, \tau}(b)<\frac{2 \delta+r}{\psi_{\downarrow}}.$$

Using that $\|\cdot\|_{\mathrm{tv}} \leq 1$ this gives

$$\mathrm{d}_{\mathrm{w}}(\check{\pi}, \hat{\pi}) \leq \mathbb{E}\left[\|\check{\boldsymbol{\gamma}}-\hat{\boldsymbol{\gamma}}\|_{\mathrm{tv}}\right] \leq \frac{q-1}{a}+\frac{2 \delta+r}{\psi_{\downarrow}}.$$

Hence, combining the results for $\boldsymbol{\sigma} \in \mathcal{S}$ and $\boldsymbol{\sigma} \notin \mathcal{S}$ yields

$$\mathbb{E}\left[\sum_{\tau \in \boldsymbol{\sigma}([n])} \mathrm{d}_{\mathrm{w}}\left(\check{\pi}_{\boldsymbol{\sigma}, \tau}, \hat{\pi}_{\tau}\right)\right] \leq q\left(\frac{q-1}{a}+\frac{2 \delta+r}{\psi_{\downarrow}}\right)+q \varepsilon+\frac{2 q^{3}(a+1)^{2(q-1)} \varepsilon_{\mathrm{s}}}{r^{2}}.$$

Now, let $a=\lfloor r^{-1}\rfloor$. With $r=\varepsilon_{\mathrm{s}}^{1/(2q+1)} \leq 1$ we have $a \in \mathbb{Z}_{\geq 1}$, hence $a+1 \leq 2a$, $a \geq \frac{1}{2}(a+1) \geq \frac{1}{2r}$ and further

$$\mathbb{E}\left[\sum_{\tau \in \boldsymbol{\sigma}([n])} \mathrm{d}_{\mathrm{w}}\left(\check{\pi}_{\boldsymbol{\sigma}, \tau}, \hat{\pi}_{\tau}\right)\right] \leq q\left(2(q-1)+\frac{1}{\psi_{\downarrow}}+2^{2q-1} q^{2}\right) r+\frac{2 q}{\psi_{\downarrow}} \delta+q \varepsilon.$$

□

3.3.1.8 Reweighted Gibbs Marginal Distribution. Using the notions from Section 3.3.1.7 and for a decorated factor graph G with Gibbs measure $\mu=\mu_{\mathrm{g},G}$ let $\pi_{\mathrm{g},G}=\pi_{\mu}$, $\overline{\gamma}_{\mathrm{g},G}=\overline{\gamma}_{\mu}$, $\check{\pi}_{\mathrm{g},G,\sigma,\tau}=\check{\pi}_{\mu,\sigma,\tau}$ and $\hat{\pi}_{\mathrm{g},G,\tau}=\hat{\pi}_{\mu,\tau}$. First, we focus on the expected Gibbs marginal $\overline{\gamma}_{\mathrm{g}}$ under various versions of the teacher-student model, and start with $\boldsymbol{G}^{*}(\hat{\boldsymbol{\sigma}})$.

Lemma 3.52. *Let $m \leq m_{\uparrow}$, $\overline{\boldsymbol{\gamma}}=\overline{\gamma}_{\boldsymbol{G}^{*}}$, $\boldsymbol{\gamma}=\gamma_{\mathrm{n},\boldsymbol{\sigma}_{\mathrm{g}}}$ with $\boldsymbol{\sigma}_{\mathrm{g}}=\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^{*}}$, $\boldsymbol{G}^{*}=\boldsymbol{G}^{*}(\boldsymbol{\sigma})$, $\boldsymbol{\sigma}=\boldsymbol{\sigma}^{*}$ or $\boldsymbol{\sigma}=\hat{\boldsymbol{\sigma}}$.*
a) There exists $c \in \mathbb{R}_{>0}^{2}$ such that $\mathbb{P}(\|\overline{\boldsymbol{\gamma}}-\gamma^{*}\|_{\mathrm{tv}} \geq r) \leq c_{2} e^{-c_{1} r^{2} n}$.
b) There exists $c \in \mathbb{R}_{>0}^{2}$ such that $\mathbb{P}(\|\boldsymbol{\gamma}-\overline{\boldsymbol{\gamma}}\|_{\mathrm{tv}} \geq r) \leq c_{2} e^{-c_{1} r^{2} n}$.
c) There exists $c \in \mathbb{R}_{>0}$ such that $\mathbb{E}[\|\overline{\boldsymbol{\gamma}}-\gamma^{*}\|_{\mathrm{tv}}] \leq c/\sqrt{n}$.
d) There exists $c \in \mathbb{R}_{>0}$ such that $\mathbb{E}[\|\overline{\boldsymbol{\gamma}}-\gamma^{*}\|_{\mathrm{tv}}^{2}] \leq c/n$.

Proof. Let $\overline{D}(G)=\|\overline{\gamma}_{\mathrm{g},G}-\gamma^{*}\|_{\mathrm{tv}}$ and $\boldsymbol{D}(G)=\|\gamma_{\mathrm{n},\boldsymbol{\sigma}_{\mathrm{g},G}}-\gamma^{*}\|_{\mathrm{tv}}$. With Equation (7) and Jensen's inequality we have

$$\overline{D}(G)^{x}=\|\mathbb{E}[\gamma_{\mathrm{n},\boldsymbol{\sigma}_{\mathrm{g},G}}]-\gamma^{*}\|_{\mathrm{tv}}^{x} \leq \mathbb{E}[\boldsymbol{D}(G)^{x}], e^{y\overline{D}(G)^{2}} \leq \mathbb{E}\left[e^{y\boldsymbol{D}(G)^{2}}\right]$$

for $x \in \mathbb{R}_{\geq 1}$ and $y \in \mathbb{R}_{\geq 0}$, so $\mathbb{E}[\overline{D}(\boldsymbol{G}^{*}))^{x}] \leq \mathbb{E}[\boldsymbol{D}(\boldsymbol{G}^{*})^{x}]$ and $\mathbb{E}[e^{y\overline{D}(\boldsymbol{G}^{*})^{2}}] \leq \mathbb{E}[e^{y\boldsymbol{D}(\boldsymbol{G}^{*})^{2}}]$. With Observation 3.22d) we have $\boldsymbol{D}(\boldsymbol{G}^{*}(\hat{\boldsymbol{\sigma}})) \sim \|\hat{\boldsymbol{\gamma}}-\gamma^{*}\|_{\mathrm{tv}}$, so Corollary 3.25 applies, and we have $\boldsymbol{D}(\boldsymbol{G}^{*}(\boldsymbol{\sigma}^{*}))=\|\boldsymbol{\gamma}-\gamma^{*}\|_{\mathrm{tv}}$ with $\boldsymbol{\gamma}$ from Observation 3.28. So, for some $c \in \mathbb{R}_{>0}^{2}$ we have

$$\mathbb{P}(\boldsymbol{D}(\boldsymbol{G}^{*}) \geq r) \leq c_{2} e^{-c_{1} r^{2} n}, \mathbb{E}[\boldsymbol{D}(\boldsymbol{G}^{*})] \leq c_{2}/\sqrt{n}, \mathbb{E}[\boldsymbol{D}(\boldsymbol{G}^{*})^{2}] \leq c_{2}/n,$$

in both cases. Notice that $\overline{D}(\boldsymbol{G}^{*})=\|\overline{\boldsymbol{\gamma}}-\gamma^{*}\|_{\mathrm{tv}}$ and recall that $\mathbb{E}[\overline{D}(\boldsymbol{G}^{*})^{x}] \leq \mathbb{E}[\boldsymbol{D}(\boldsymbol{G}^{*})^{x}]$, so Part

3.52c) follows for $x=1$ and Part 3.52d) follows for $x=2$. Further, notice that

$$\mathbb{E}\left[\exp\left(\frac{c_1}{2}\boldsymbol{D}(\boldsymbol{G}^*)^2n\right)\right]=\int_0^\infty\mathbb{P}\left(\exp\left(\frac{c_1}{2}\boldsymbol{D}(\boldsymbol{G}^*)^2n\right)>r\right)\mathrm{d}r$$
$$=1+\int_1^\infty\mathbb{P}\left(\boldsymbol{D}(\boldsymbol{G}^*)>\sqrt{\frac{2\ln(r)}{c_1n}}\right)\mathrm{d}r$$
$$\le 1+c_2\int_1^\infty e^{-2\ln(r)}\mathrm{d}r=1+c_2\int_1^\infty\frac{1}{r^2}\mathrm{d}r=1+c_2.$$

So, the bound above for $y=c_1n/2$ and Markov's inequality yield

$$\mathbb{P}\left(\overline{D}(\boldsymbol{G}^*)\ge r\right)=\mathbb{P}\left(\exp\left(\frac{c_1}{2}\overline{D}(\boldsymbol{G}^*)^2n\right)\ge\exp\left(\frac{c_1}{2}r^2n\right)\right)\le(1+c_2)\exp\left(-\frac{c_1}{2}r^2n\right).$$

This establishes Part 3.52a). Let $c'\in\mathbb{R}^2_{>0}$ be the constants from Part 3.52a). Then the triangle inequality and the union bound yields

$$\mathbb{P}(\|\boldsymbol{\gamma}-\overline{\boldsymbol{\gamma}}\|_{\mathrm{tv}}\ge r)\le\mathbb{P}(\boldsymbol{D}(\boldsymbol{G}^*)\ge r/2)+\mathbb{P}(\overline{D}(\boldsymbol{G}^*)\ge r/2)\le c_2e^{-c_1r^2n/4}+c_2'e^{-c_1'r^2n/4},$$

so Part 3.52b) holds with $c_1'/4$ and c_2+c_2'. □

Remark 3.53. Similar to $\iota(\mu)$ and $\nu(\mu)$ in Remark 3.49, quantifying the ℓ-wise dependencies of μ, we may consider $P_\mu:[0,1]\to[0,1]$, $r\mapsto\mathbb{P}(\|\gamma_{\mathrm{n},\boldsymbol{\sigma}_\mu}-\overline{\gamma}_\mu\|_{\mathrm{tv}}\ge r)$ to quantify the concentration of the color frequencies. As for Proposition 3.48, Lemma 3.52b) suggests that $\mathbb{E}[P(\boldsymbol{\mu},r)]\le c_2e^{-c_1r^2n}$ for both $\boldsymbol{\mu}=\boldsymbol{\mu}^*$ and $\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}$.

Lemma 3.52 facilitates the application of Proposition 3.51. For this purpose let

$$D(\sigma,\mu)=\sum_{\tau\in\sigma([n])}\mathrm{d_w}(\check{\pi}_{\mu,\sigma,\tau},\hat{\pi}_{\mu,\tau})$$

for $\mu\in\mathcal{P}([q]^n)$ using the notions from Proposition 3.51.

Corollary 3.54. *Let $m\le m_\uparrow$, $\hat{\boldsymbol{\mu}}=\mu_{\mathrm{g},\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})}$, $\boldsymbol{\mu}^*=\mu_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ and $\boldsymbol{\sigma}_\mu\sim\mu$. Denote the constants from Proposition 3.48b) by $C_{\mathfrak{g}}\in(0,1)\times\mathbb{R}_{>0}$. There exists $c_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that the following holds.*
a) We have $\mathbb{E}[D(\hat{\boldsymbol{\sigma}},\hat{\boldsymbol{\mu}})],\mathbb{E}[D(\boldsymbol{\sigma}_{\hat{\boldsymbol{\mu}}},\hat{\boldsymbol{\mu}})]\le c/\Theta^{\downarrow}$.
b) We have $\mathbb{E}[D(\boldsymbol{\sigma}^*,\boldsymbol{\mu}^*)]\le c/\Theta^{\downarrow c'}$ for $c'\in(0,C_1]$, and $\mathbb{E}[D(\boldsymbol{\sigma}_{\boldsymbol{\mu}^*},\boldsymbol{\mu}^*)]\le c/\Theta^{\downarrow c'}$ for $c'\in(0,C_1/3]$.

Proof. For $\Theta^\downarrow\le 1$ and $c\in\mathbb{R}_{\ge0}\times\mathbb{R}_{\ge q}$ we have $\mathbb{E}[D(\cdot)]\le q\le c_2\Theta^{\downarrow-c_1}$. Otherwise, notice that $n>1$, let $\delta=\varepsilon=\ln(n)/\sqrt{n}$, $\varepsilon_{\mathrm{s}}\in\mathbb{R}_{>0}$ and

$$\mathcal{M}=\left\{\mu\in\mathcal{P}([q]^n):\|\overline{\gamma}_\mu-\gamma^*\|_{\mathrm{tv}}\le\frac{\psi_\downarrow}{4},\iota_2(\mu)\le\varepsilon_{\mathrm{s}},P(\mu,\delta)\le\varepsilon\right\}$$

with P from Remark 3.53. For $\mu\in\mathcal{M}$ we have $\|\overline{\gamma}_\mu-\gamma^*\|_\infty\le\psi_\downarrow/2$, so $\overline{\gamma}_\mu\ge\psi_\downarrow/2$, and $\mathbb{E}[D(\boldsymbol{\sigma}_\mu,\mu)]\le c_{\mathrm{d}}(\delta+\varepsilon+\sqrt{\varepsilon_{\mathrm{s}}/2})$ with $c_{\mathrm{d}}\in\mathbb{R}_{>0}$ from Proposition 3.51, using Remark 3.49. With $c_{\mathrm{c}}\in\mathbb{R}^2_{>0}$ for both

Lemma 3.52a) and Lemma 3.52b), Proposition 3.48a) and Markov's inequality we have

$$\mathbb{P}(\hat{\boldsymbol{\mu}} \notin \mathcal{M}) \le c_{\mathrm{c},2} \exp\left(-\frac{c_{\mathrm{c},1}\psi_{\downarrow}^2}{16} n\right) + \frac{\ln(q)}{\varepsilon_{\mathrm{s}}\Theta^{\downarrow}} + \frac{c_{\mathrm{c},2}}{\varepsilon} e^{-c_{\mathrm{c},1}\delta^2 n}$$
$$\le \tilde{c}e^{-n/\tilde{c}} + \frac{\tilde{c}}{\Theta^{\downarrow 1/3}} + \frac{\tilde{c}\sqrt{n}}{\ln(n)} e^{-\ln(n)^2/\tilde{c}} \le \frac{c'}{\Theta^{\downarrow 1/3}},$$

where $\tilde{c}$ is the implied maximum, $\varepsilon_{\mathrm{s}} = \Theta^{\downarrow -2/3}$, $c' = \tilde{c}c_1'' + \tilde{c} + \tilde{c}c_2''$, $c_1'' = \max_{n>1} n^{1/3}e^{-n/\tilde{c}}$ and $c_2'' = \max_{n>1} n^{5/6}e^{-\ln(n)^2/\tilde{c}}/\ln(n)$ using $\Theta^{\downarrow} \le n$. Using $D(\cdot) \le q$ gives

$$\mathbb{E}\left[D\left(\boldsymbol{\sigma}_{\hat{\boldsymbol{\mu}}}, \hat{\boldsymbol{\mu}}\right)\right] \le \frac{2c_{\mathrm{d}}\ln(n)}{\sqrt{n}} + \frac{c_{\mathrm{d}}}{\sqrt{2}\Theta^{\downarrow 1/3}} + \frac{qc'}{\Theta^{\downarrow 1/3}} \le \frac{c}{\Theta^{\downarrow 1/3}}$$

with $c = \max(c_{\mathrm{d}} + qc' + 2c_{\mathrm{d}} \max_{n>1} \ln(n)/n^{1/6}, q)$, so $\mathbb{E}[D(\boldsymbol{\sigma}_{\hat{\boldsymbol{\mu}}}, \hat{\boldsymbol{\mu}})] \le c/\Theta^{\downarrow 1/3}$ holds for all $\Theta^{\downarrow} \ge 0$. The Nishimori condition 3.22d) completes the proof of Part 3.54a). With C from Proposition 3.48b), branching off in the discussion above yields

$$\mathbb{P}(\boldsymbol{\mu}^* \notin \mathcal{M}) \le c_{\mathrm{c},2} \exp\left(-\frac{c_{\mathrm{c},1}\psi_{\downarrow}^2}{16} n\right) + \frac{C_2 2^{C_1}}{\varepsilon_{\mathrm{s}}\Theta^{\downarrow C_1}} + \frac{c_{\mathrm{c},2}}{\varepsilon} e^{-c_{\mathrm{c},1}\delta^2 n} \le \frac{c'}{\Theta^{\downarrow C_1/3}}$$

with $\varepsilon_{\mathrm{s}} = \Theta^{\downarrow -2C_1/3}$ and c' obtained analogously to the above. Repeating the remaining steps yields $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ with $\mathbb{E}[D(\boldsymbol{\sigma}_{\boldsymbol{\mu}^*}, \boldsymbol{\mu}^*)] \le c/\Theta^{\downarrow C_1/3}$. For the second part of Corollary 3.54b) and $\Theta^{\downarrow} > 1$ let $\hat{c}$ from Corollary 3.25b), c^* from Observation 3.23b), c' from Part 3.54a) and $r = \sqrt{\ln(\Theta^{\downarrow})/((c_1^* + \hat{c})n)}$, then we have

$$\mathbb{E}[D(\boldsymbol{\sigma}^*, \boldsymbol{\mu}^*)] \le e^{\hat{c}r^2 n}\frac{c'}{\Theta^{\downarrow}} + qc_2^* e^{-c_1^* r^2 n} = \frac{c}{\Theta^{\downarrow\rho}}$$

with $c = c' + qc_2^*$ and $\rho = c_1^*/(c_1^* + \hat{c})$. Recall from the proof of Proposition 3.48 that $\rho = C_1$. □

3.3.1.9 Gibbs Marginal Distribution Projection. Let $\boldsymbol{G}^* = \boldsymbol{G}^*(\boldsymbol{\sigma})$ with $\boldsymbol{\sigma} = \boldsymbol{\sigma}^*$ or $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}$. Recall that $\overline{\gamma}_{\mathrm{g},\boldsymbol{G}^*}$ is defined as the expected law under the empirical marginal distribution $\boldsymbol{\pi} = \pi_{\mathrm{g},\boldsymbol{G}^*}$, given $\boldsymbol{\pi}$. Lemma 3.52a) ensures that $\overline{\gamma}_{\boldsymbol{G}^*}$ is close to γ^*, but this is not sufficient for **POS** and $B_{\uparrow}$, being extremal only on $\mathcal{P}_*^2([q])$, meaning that the expectation has to be exactly γ^*.

Hence, we map $\boldsymbol{\pi} \in \mathcal{P}^2([q])$ to some $\boldsymbol{\pi}^{\circ} \in \mathcal{P}_*^2([q])$ such that the Wasserstein distance $\mathrm{d}_{\mathrm{w}}(\boldsymbol{\pi}, \boldsymbol{\pi}^{\circ})$ vanishes, which is sufficient because both the map in **POS** and B will turn out to be Lipschitz continuous. First, we identify a suitable counterweight to $\overline{\gamma}$.

Let $\alpha_{\mathfrak{g}} : \mathcal{P}([q]) \to [0, 1]$ and $f_{\mathfrak{g}} : \mathcal{P}([q]) \to \mathcal{P}([q])$, $\gamma \mapsto [\gamma]_{\mathrm{c}}$, be given as follows. Let $\ell_{\circ} = \ell_{\mathrm{c}}^{-1}(\psi_{\downarrow})$ with $\ell_{\mathrm{c}} : [0, 1) \to \mathbb{R}_{\ge 0}$, $\ell \mapsto -(1 + \Lambda(\ell))/\ln(\ell)$ and $\ell(\gamma) = \|\gamma - \gamma^*\|_2$ for $\gamma \in \mathcal{P}([q])$. For $\ell(\gamma) = 0$ let $[\gamma]_{\mathrm{c}} = \gamma^*$ and $\alpha(\gamma) = 0$. For $\ell(\gamma) \in (0, \ell_{\circ}]$ let $[\gamma]_{\mathrm{c}} = \gamma^* + \frac{\ell_{\mathrm{c}}(\ell(\gamma))}{\ell(\gamma)}(\gamma^* - \gamma)$ and $\alpha(\gamma) = -\Lambda(\ell(\gamma))$. For $\ell(\gamma) \ge \ell_{\circ}$ let $[\gamma]_{\mathrm{c}} = \gamma^* + \frac{\psi_{\downarrow}}{\ell(\gamma)}(\gamma^* - \gamma)$ and $\alpha(\gamma) = \ell(\gamma)/(\ell(\gamma) + \psi_{\downarrow})$.

Observation 3.55. *The maps α and f are continuous with $\alpha(\gamma)[\gamma]_{\mathrm{c}} + (1-\alpha(\gamma))\gamma = \gamma^*$ for $\gamma \in \mathcal{P}([q])$. With $c = (e-1)/e$ we have $\ell_{\circ} \in [e^{-\psi_{\uparrow}}, e^{-c\psi_{\uparrow}}]$ and α is increasing in $\ell(\gamma)$.*

Proof. Clearly, the maps ℓ and ℓ_{c} are continuous. Further, ℓ_{c} is strictly increasing with $\ell_{\mathrm{c}}(0) = 0$ and $\ell_{\mathrm{c}}(1) = \infty$, so $\ell_{\circ} \in (0, 1)$ is well-defined. Hence, we have $[\gamma]_{\mathrm{c}} = \gamma^* + s(\gamma)(\gamma^* - \gamma)$ with $s(\gamma) =$

$\min(\psi_\downarrow, \ell_c(\ell(\gamma)))/\ell(\gamma)$ and thereby

$$\|[\gamma]_c - \gamma^*\|_\infty \le \|[\gamma]_c - \gamma^*\|_2 = \min(\psi_\downarrow, \ell_c(\ell(\gamma))) \le \psi_\downarrow,$$

so $[\gamma]_c \ge 0$ and thereby $[\gamma]_c \in \mathcal{P}([q])$. The map s is clearly continuous for $\gamma \ne \gamma^*$. For $\ell(\gamma) \le \ell_\circ$ we further have $\|[\gamma]_c - \gamma^*\|_2 = \ell_c(\ell(\gamma))$ and hence f is continuous. Notice that $\ell_c(\ell_\circ) = \psi_\downarrow$ implies $-\Lambda(\ell_\circ) = \ell_\circ/(\ell_\circ+\psi_\downarrow)$ and hence α is clearly continuous for $\gamma \ne \gamma^*$, while continuity for $\gamma = \gamma^*$ follows from $-\Lambda(\ell(\gamma^*)) = 0$. We have $\alpha(\gamma)[\gamma]_c + (1-\alpha(\gamma))\gamma = \gamma^*$ by construction. With $\ell_c(\ell) \le -1/\ln(\ell)$ we have $\ell_\circ \ge e^{-\psi_\uparrow}$, while the upper bound follows with $\ell_c(\ell) \ge -c/\ln(\ell)$. With $\psi_\downarrow \le 1/q \le 1/2$ we have $c\psi_\uparrow > 1$, so α is increasing since Λ takes its unique minimum at e^{-1}. □

So, with the notation from Section 3.3.1.7 for the general case let

$$\pi^\circ_\mu = (1-\alpha_\mu)\pi_\mu + \alpha_\mu \pi_\bullet$$

with $\pi_\bullet = \mu_{\bullet,\mathcal{P}([q]),\gamma}$, $\gamma = [\overline{\gamma}_\mu]_c$, and $\alpha_\mu = \alpha(\|\overline{\gamma}_\mu - \gamma^*\|_2)$. For a decorated graph G let $\pi^\circ_{g,G} = \pi^\circ_{\mu_{g,G}}$ be the projection of $\pi_{g,G}$ onto $\mathcal{P}^2_*([q])$.

Lemma 3.56. *Let $m \le m_\uparrow$ and $\boldsymbol{G}^* = \boldsymbol{G}^*(\boldsymbol{\sigma})$ with $\boldsymbol{\sigma} = \boldsymbol{\sigma}^*$ or $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}$.*
a) For $\mu \in \mathcal{P}([q]^n)$ we have $\pi^\circ_\mu \in \mathcal{P}^2_*([q])$ and $d_w(\pi_\mu, \pi^\circ_\mu) \le \alpha_\mu$.
b) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $\mathbb{E}[d_w(\pi_{g,\boldsymbol{G}^*}, \pi^\circ_{g,\boldsymbol{G}^*})] \le c\sqrt{\ln(n+1)^3/n}$.

Proof. With $(\boldsymbol{b}, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \sim \mathrm{Bin}(1, \alpha_\mu) \otimes \pi_\mu \otimes \pi_\bullet$ we have $\boldsymbol{\gamma}_{\boldsymbol{b}} \sim \pi^\circ_\mu$ and

$$\mathbb{E}[\boldsymbol{\gamma}_{\boldsymbol{b}}] = \mathbb{E}[\mathbb{1}\{\boldsymbol{b}=0\}\boldsymbol{\gamma}_0] + \mathbb{E}[\mathbb{1}\{\boldsymbol{b}=1\}\boldsymbol{\gamma}_1] = (1-\alpha_\mu)\overline{\gamma}_\mu + \alpha_\mu[\overline{\gamma}_\mu]_c = \gamma^*,$$

so $\pi^\circ_\mu \in \mathcal{P}^2_*([q])$. Further, we have

$$d_w(\pi_\mu, \pi^\circ_\mu) \le \mathbb{E}[\|\boldsymbol{\gamma}_{\boldsymbol{b}} - \boldsymbol{\gamma}_0\|_{tv}] = \alpha_\mu \mathbb{E}[\|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_0\|_{tv}] \le \alpha_\mu.$$

With $c' \in (0,1] \times \mathbb{R}_{\ge 1}$ from Lemma 3.52a) let $r = \sqrt{\ln(n)/(2c'_1 n)}$ and let $n_{\circ,\mathfrak{g}} \in \mathbb{Z}_{\ge 3}$ be such that $2r \le \ell_\circ$ if $n \ge n_\circ$. For $n \le n_0$ we have

$$\mathbb{E}[d_w(\pi_{g,\boldsymbol{G}^*}, \pi^\circ_{g,\boldsymbol{G}^*})] \le q \le \sqrt{\frac{q^2 n_\circ}{\ln(2)^3}}\sqrt{\frac{\ln(n+1)^3}{n}}.$$

Otherwise, with $d_w(\cdot) \le 1$, the first part and $\overline{\boldsymbol{\gamma}} = \overline{\gamma}_{g,\boldsymbol{G}^*}$ we have

$$\mathbb{E}[d_w(\pi_{g,\boldsymbol{G}^*}, \pi^\circ_{g,\boldsymbol{G}^*})] \le \mathbb{E}\left[\mathbb{1}\{\|\overline{\boldsymbol{\gamma}} - \gamma^*\|_2 < 2r\}\alpha(\|\overline{\boldsymbol{\gamma}} - \gamma^*\|_2)\right] + \mathbb{P}(\|\overline{\boldsymbol{\gamma}} - \gamma^*\|_2 \ge 2r).$$

With $\|\overline{\boldsymbol{\gamma}} - \gamma^*\|_2 \le 2\|\overline{\boldsymbol{\gamma}} - \gamma^*\|_{tv}$ and Observation 3.55 we have

$$\mathbb{E}[d_w(\pi_{g,\boldsymbol{G}^*}, \pi^\circ_{g,\boldsymbol{G}^*})] \le \alpha(2r) + c'_2 e^{-c'_1 r^2 n} = -\Lambda(2r) + \frac{c'_2}{\sqrt{n}}.$$

Hence, for any sufficiently large $\tilde{c}$ we have

$$\mathbb{E}[d_w(\pi_{g,\boldsymbol{G}^*}, \pi^\circ_{g,\boldsymbol{G}^*})] \le \tilde{c}\sqrt{\frac{\ln(n)}{n}}\ln\left(\frac{n}{\ln(n)}\right) + \tilde{c}\sqrt{\frac{\ln(n+1)^3}{n}} \le 2\tilde{c}\sqrt{\frac{\ln(n+1)^3}{n}}.$$

□

3.3.2 The Interpolation Method. Let $\Theta^{\downarrow}_{\mathfrak{g}} : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$ with $\Theta^{\downarrow} = \omega(1)$ and $\Theta^{\downarrow} \le n$, so dependencies on $\Theta^{\downarrow}$ change to dependencies on n. Assume that $\bar{d} > 0$. Recall the map in **POS**, i.e.

$$\nabla : \mathcal{P}^2([q]) \to \mathbb{R},\ \pi \mapsto \mathbb{E}\left[\Lambda(Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_1)) + (k-1)\Lambda(Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_2)) - \sum_{h=1}^{k} \Lambda(Z_{\mathrm{fm}}(\boldsymbol{\psi}, h, \boldsymbol{\gamma}))\right].$$

3.3.2.1 Overview. Recall the Bethe free entropy in Section 2.1.2.4 and the related notions, in particular $(\boldsymbol{\psi}_\circ, \boldsymbol{\gamma}_\circ)$. The interpolation method relies on the derivative of the function

$$\phi^{\leftrightarrow}_{\mu_\Psi, \gamma^*, \bar{d}, \pi, n}(t^{\leftrightarrow}) = \mathbb{E}\left[\phi_{\mathrm{g}}(\hat{\boldsymbol{G}})\right] + t^{\leftrightarrow}\phi_\circ,\ \phi_\circ = \frac{\bar{d}(k-1)}{\xi k}\mathbb{E}\left[\Lambda\left(Z_{\mathrm{f}}(\boldsymbol{\psi}_\circ, \boldsymbol{\gamma}_\circ)\right)\right],$$

using the shorthand $\hat{\boldsymbol{G}} = \boldsymbol{G}^*_{\boldsymbol{m}, \boldsymbol{m}^{\leftrightarrow}, \boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}})$. If the derivative is (asymptotically) non-negative, then we have $\phi^{\leftrightarrow}(0) \le \phi^{\leftrightarrow}(1)$, and realignment yields Proposition 3.4. Hence, we determine the asymptotics of the derivative. Recall $\pi_{\mathrm{g},G}$ from Section 3.3.1.8 and its projection $\pi^\circ_{\mathrm{g},G} \in \mathcal{P}^2_*([q])$ from Section 3.3.1.9.

Proposition 3.57. *We have* $\frac{\partial}{\partial t^{\leftrightarrow}}\phi^{\leftrightarrow}(t^{\leftrightarrow}) = \frac{\bar{d}}{\xi k}\mathbb{E}[\nabla(\pi^\circ_{\mathrm{g},\hat{\boldsymbol{G}}}, \pi)] + \mathcal{O}(\Theta^{\downarrow -1/3})$.

In order to establish Proposition 3.57, we first compute the derivative of $\phi^{\leftrightarrow}$.

Lemma 3.58. *We have* $\frac{\partial}{\partial t^{\leftrightarrow}}\phi^{\leftrightarrow}(t^{\leftrightarrow}) = \frac{\bar{d}}{k}\Delta^\circ + \phi_\circ - \bar{d}\Delta^{\leftrightarrow}$*, where*

$$\Delta^\circ = \mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}+1, \boldsymbol{m}^{\leftrightarrow}, \boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}+1}))] - \mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}, \boldsymbol{m}^{\leftrightarrow}, \boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))],$$
$$\Delta^{\leftrightarrow} = \sum_{i \in [n]} \frac{1}{n}\left(\mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}, \boldsymbol{m}^{\leftrightarrow} + \mu_{\bullet,[n],i}, \boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))] - \mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}, \boldsymbol{m}^{\leftrightarrow}, \boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))]\right).$$

Proof. We consider $n+1$ independent Poisson variables $\boldsymbol{m}$, $\boldsymbol{m}^{\leftrightarrow}$ depending on $t^{\leftrightarrow}$, while the remainder does not. Hence, we use the product rule, which amounts to taking each derivative individually using Observation 3.30 and Corollary 3.12. But for $\boldsymbol{x} \sim \mathrm{Po}(at + b)$ we have

$$\frac{\partial}{\partial t}\mathbb{E}[f(\boldsymbol{x})] = \sum_x f(x)\frac{\partial}{\partial t}\mathbb{P}(\boldsymbol{x} = x) = -a\mathbb{E}[f(\boldsymbol{x})] + a\mathbb{E}[f(\boldsymbol{x}+1)].$$

□

The second contribution $\phi_\circ$ in Lemma 3.58 is exactly what we need. For the other contributions recall $(\boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}_\pi)$ from the definition of ∇ and let $\boldsymbol{\pi} = (\pi_{\mathrm{g},\hat{\boldsymbol{G}}}, \pi)$.

Lemma 3.59. *We have* $\xi\Delta^{\leftrightarrow} = \mathbb{E}[\Lambda(Z_{\mathrm{fm}}(\boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}_{\boldsymbol{\pi}}))]$.

The proof of Lemma 3.59 is presented in Section 3.3.2.2. The first contribution is demanding, since the joint Gibbs law is not a product measure. This is where Proposition 3.48 comes into play.

Lemma 3.60. *We have* $\xi\Delta^\circ = \mathbb{E}[\Lambda(Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_{\boldsymbol{\pi},1}))] + \mathcal{O}(\Theta^{\downarrow -1/3})$.

The proof of Lemma 3.60 is presented in Section 3.3.2.3. Proposition 3.57 now follows by establishing Lipschitz continuity of ∇ and thereby justifying the transition to the projection $\pi^\circ_{\mathrm{g},\hat{\boldsymbol{G}}}$. The proof is presented in Section 3.3.2.4. The proof of Proposition 3.4 and the respective version for graphs with external fields over random factor counts $\boldsymbol{m}^*$ is presented in Section 3.3.2.5.

3.3.2.2 Adding an Interpolator. Fix the variable $i \in [n]$ with the additional interpolator and let

$$\Delta_{\mathrm{v}}^{\leftrightarrow}(i) = \mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow}+\mu_{\bullet,[n],i},\boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))] - \mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))].$$

Lemma 3.61. *With* $(\boldsymbol{G}', \boldsymbol{\psi}^{\leftrightarrow}) \sim \hat{\boldsymbol{G}} \otimes \boldsymbol{\psi}_{\circ}^{\leftrightarrow}$ *we have*

$$\Delta_{\mathrm{v}}^{\leftrightarrow}(i) = \frac{1}{\xi}\mathbb{E}\left[\Lambda\left(\sum_{\tau\in[q]} \mu_{\mathrm{g},\boldsymbol{G}'}|_i(\tau)\boldsymbol{\psi}^{\leftrightarrow}(\tau)\right)\right].$$

Proof. Using independence due to Observation 3.13, we have a coupling $(\boldsymbol{G}_-, \boldsymbol{G}_+)$ of $\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}})$ and $\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{m}^{\leftrightarrow}+\mu_{\bullet,[n],i},\boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}})$, i.e. given $\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}$ and $\boldsymbol{G}_-$ we attach the factor $(m_i^{\leftrightarrow} + 1)$ to i equipped with a weight given by $\boldsymbol{\psi}^{*\leftrightarrow}_{\circ,\sigma(i)}$ to obtain $\boldsymbol{G}_+$. Explicitly introducing the conditional expectation gives

$$\Delta_{\mathrm{v}}^{\leftrightarrow}(i) = \mathbb{E}\left[\Delta^{\leftrightarrow}_{\mathrm{vms},i}(\boldsymbol{m}, \boldsymbol{m}^{\leftrightarrow}, \boldsymbol{\mathcal{U}}, \hat{\boldsymbol{\sigma}}_{\boldsymbol{m}})\right],\ \Delta^{\leftrightarrow}_{\mathrm{vms},i}(m, m^{\leftrightarrow}, \mathcal{U}, \sigma) = \mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}_+) - n\phi_{\mathrm{g}}(\boldsymbol{G}_-)].$$

With $\delta(G, G') = n\phi_{\mathrm{g}}(G') - n\phi_{\mathrm{g}}(G)$ we have $\delta(G, G') = \ln(Z_{\mathrm{g}}(G')/Z_{\mathrm{g}}(G))$, and further $Z_{\mathrm{g}}(G') = \sum_\sigma \psi_{\mathrm{g},G'}(\sigma)$. If G' is an extension of G as above, i.e. obtained by adding factors $\mathcal{A}_+$ with wire-weight pairs $w = (v_a, \psi_a)_{a\in\mathcal{A}_+}$, then we have $\psi_{\mathrm{g},G'}(\sigma) = \psi_{\mathrm{g},G}(\sigma)\prod_{a\in\mathcal{A}_+}\psi_a(\sigma_{v_a})$. This gives $\delta(G, G') = \ln(\overline{\psi}_{\mathrm{w|g},G}(w))$ with

$$\overline{\psi}_{\mathrm{w|g},G}(w) = \sum_\sigma \mu_{\mathrm{g},G}(\sigma) \prod_{a\in\mathcal{A}_+} \psi_a(\sigma_{v(a)}) = \mathbb{E}\left[\prod_{a\in\mathcal{A}_+} \psi_a\left(\boldsymbol{\sigma}_{\mathrm{g},G,v(a)}\right)\right], \tag{8}$$

so the difference of the free entropies is the logarithm of the expected weight of the additional factors of G', under the Gibbs measure of the (smaller) base graph G. So, using $(\boldsymbol{G}''(\sigma), \boldsymbol{\psi}^{*\leftrightarrow}, \boldsymbol{\psi}^{\leftrightarrow}) \sim \boldsymbol{G}^*(\sigma) \otimes \boldsymbol{\psi}^{*\leftrightarrow}_{\circ,\sigma(i)} \otimes \boldsymbol{\psi}^{\leftrightarrow}_{\circ}$ for brevity, we have

$$\Delta^{\leftrightarrow}_{\mathrm{vms}} = \mathbb{E}\left[\ln\left(\overline{\psi}_{\mathrm{w|g},\boldsymbol{G}''(\sigma)}\left(i, \boldsymbol{\psi}^{*\leftrightarrow}\right)\right)\right] = \mathbb{E}\left[\frac{\boldsymbol{\psi}^{\leftrightarrow}(\sigma_i)}{\xi}\ln\left(\overline{\psi}_{\mathrm{w|g},\boldsymbol{G}''(\sigma)}(i, \boldsymbol{\psi}^{\leftrightarrow})\right)\right].$$

Taking the expectation over $\hat{\boldsymbol{\sigma}}$, using $\boldsymbol{G} = \boldsymbol{G}''(\hat{\boldsymbol{\sigma}})$ and the Nishimori condition 3.22d) yields

$$\mathbb{E}[\Delta^{\leftrightarrow}_{\mathrm{vms}}(\hat{\boldsymbol{\sigma}})] = \mathbb{E}\left[\frac{\boldsymbol{\psi}^{\leftrightarrow}(\hat{\boldsymbol{\sigma}}_i)}{\xi}\ln\left(\overline{\psi}_{\mathrm{w|g},\boldsymbol{G}}(i, \boldsymbol{\psi}^{\leftrightarrow})\right)\right] = \mathbb{E}\left[\frac{\boldsymbol{\psi}^{\leftrightarrow}(\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G},i})}{\xi}\ln\left(\overline{\psi}_{\mathrm{w|g},\boldsymbol{G}}(i, \boldsymbol{\psi}^{\leftrightarrow})\right)\right].$$

For the leading coefficient we take the conditional expectation given $\boldsymbol{G}$ and $\boldsymbol{\psi}^{\leftrightarrow}$, i.e. the expectation over the Gibbs spins $\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}$ only, which exactly matches the definition of $\overline{\psi}_{\mathrm{w|g}}$ and hence

$$\mathbb{E}[\Delta^{\leftrightarrow}_{\mathrm{vms}}(\hat{\boldsymbol{\sigma}})] = \frac{1}{\xi}\mathbb{E}\left[\Lambda\left(\overline{\psi}_{\mathrm{w|g},\boldsymbol{G}}(i, \boldsymbol{\psi}^{\leftrightarrow})\right)\right],\ \overline{\psi}_{\mathrm{w|g},\boldsymbol{G}}(i, \boldsymbol{\psi}^{\leftrightarrow}) = \sum_{\tau\in[q]} \mu_{\mathrm{g},\boldsymbol{G}}|_i(\tau)\boldsymbol{\psi}^{\leftrightarrow}(\tau).$$

□

Now, recall that we have $\Delta^{\leftrightarrow} = \mathbb{E}[\Delta_{\mathrm{v}}^{\leftrightarrow}(\boldsymbol{i})]$ for $\boldsymbol{i} \sim \mathrm{u}([n])$ and that $\mu_{\mathrm{g},G}|_{\boldsymbol{i}} \sim \pi_{\mathrm{g},G}$. Further, recall

$(\boldsymbol{\psi}_\circ, \boldsymbol{h}, \boldsymbol{\gamma})$ and the definition of $\boldsymbol{\psi}^{\leftrightarrow} \sim \boldsymbol{\psi}_\circ^{\leftrightarrow}$ from Equation (3), which gives

$$\sum_{\tau \in [q]} \mu_{\mathrm{g},G}|_{\boldsymbol{i}}(\tau)\boldsymbol{\psi}^{\leftrightarrow}(\tau) \sim \sum_{\tau \in [q]} \mu_{\mathrm{g},G}|_{\boldsymbol{i}}(\tau) \sum_{\tau'} \mathbb{1}\{\tau'_{\boldsymbol{h}} = \tau\}\boldsymbol{\psi}_\circ(\tau') \prod_{h \neq \boldsymbol{h}} \boldsymbol{\gamma}_h(\tau'_h) \sim Z_{\mathrm{fm}}(\boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}_{\pi'})$$

with $\pi' = (\pi_{\mathrm{g},G}, \pi)$, $(\boldsymbol{i}, \boldsymbol{\psi}^{\leftrightarrow}) \sim \boldsymbol{i} \otimes \boldsymbol{\psi}^{\leftrightarrow}$ and $(\boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}_{\pi'})$ from the assertion in Lemma 3.59. This completes the proof by considering the conditional expectation given $\boldsymbol{G}'$ in $\Delta^{\leftrightarrow}$.

3.3.2.3 Adding a Factor. Using the shorthand $\phi^*_m(\sigma) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{m,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\sigma))]$ we may rewrite $\Delta^\circ = \mathbb{E}\left[n\phi^*_{\boldsymbol{m}+1}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}+1})\right] - \mathbb{E}[n\phi^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}})]$. In the first step we align the ground truths, i.e. we replace $\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}+1}$ by $\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}$, and introduce the following typical event. With $r(n) = \ln(n)/\sqrt{n}$ and $\mathcal{B}^\circ$ from Corollary 3.12 for r, let $\mathcal{B}^\Gamma = \{\sigma \in [q]^n : \|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}} < r\}$ and further $\mathcal{E} = \{\bar{\boldsymbol{d}} \in \mathcal{B}^\circ, \hat{\boldsymbol{\sigma}}_{\boldsymbol{m}} \in \mathcal{B}^\Gamma\}$.

Lemma 3.62. *We have* $\Delta^\circ = n\mathbb{E}\left[\mathbb{1}\mathcal{E}(\phi^*_{\boldsymbol{m}+1}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}) - \phi^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))\right] + \mathcal{O}(r(n))$.

Proof. Let $(\hat{\boldsymbol{\sigma}}^-_m, \hat{\boldsymbol{\sigma}}^+_m)$ be a coupling of $\hat{\boldsymbol{\sigma}}_m$, $\hat{\boldsymbol{\sigma}}_{m+1}$ from the coupling lemma 3.6e) and further $\mathcal{E}' = \{\bar{\boldsymbol{d}} \in \mathcal{B}^\circ, \hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}} \in \mathcal{B}^\Gamma, \hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}} \in \mathcal{B}^\Gamma\}$. With $\boldsymbol{\Phi}^+ = n\phi^*_{\boldsymbol{m}+1}(\hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}})$, $\boldsymbol{\Phi}^- = n\phi^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}})$, c_Φ from Observation 3.30, c° from Corollary 3.12, Observation 3.47, $\ln(n)/\sqrt{n} \le 2/e$, $n \ge 1$ and $\Theta^\downarrow \le n$ we obtain

$$|\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}} \notin \mathcal{B}^\circ\}\boldsymbol{\Phi}^+]| \le \mathbb{E}\left[\mathbb{1}\left\{\bar{\boldsymbol{d}} \notin \mathcal{B}^\circ\right\} c_\Phi \left(\frac{\bar{\boldsymbol{d}}n}{k} + 1 + (1 - t^{\leftrightarrow})\bar{d}n + \frac{\Theta^\downarrow}{2}\right)\right] \le c'_2 n e^{-c'_1 \ln(n)^2},$$
$$c'_1 = \frac{c^\circ_1}{2},\, c'_2 = c_\Phi\left(\frac{1}{k} + 1 + d_\uparrow + \frac{1}{2}\right)c^\circ_2,$$

and we obtain the same bound for $\boldsymbol{\Phi}^-$. On the event $\bar{\boldsymbol{d}} \in \mathcal{B}^\circ$ we have $\boldsymbol{m} + 1 \le m_\uparrow$ since $d_\uparrow$ is large. But then with $\hat{c}$ from Corollary 3.25d) we obtain

$$|\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}} \in \mathcal{B}^\circ, \hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}} \notin \mathcal{B}^\Gamma\}\boldsymbol{\Phi}^+]| \le c_\Phi \hat{c}_2 \left(\frac{d_\uparrow n}{k} + 1 + d_\uparrow n + \frac{\Theta^\downarrow}{2}\right) e^{-\hat{c}_1 \ln(n)^2} \le c''_2 n e^{-c''_1 \ln(n)^2},$$
$$c''_1 = \hat{c}_1,\, c''_2 = c_\Phi \hat{c}_2\left(\frac{d_\uparrow}{k} + 1 + d_\uparrow + \frac{1}{2}\right).$$

The bound for $\hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}}$ is the same, and the same bounds also follow for $\boldsymbol{\Phi}^-$. This shows that

$$\Delta^\circ = n\mathbb{E}\left[\mathbb{1}\mathcal{E}'(\phi^*_{\boldsymbol{m}+1}(\hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}}) - \phi^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}}))\right] + \mathcal{O}\left(ne^{-\tilde{c}\ln(n)^2}\right)$$

for $\tilde{c} = \min(c'_1, c''_1)$. Since $(\hat{\boldsymbol{\sigma}}^-, \hat{\boldsymbol{\sigma}}^+)$ is a coupling from the coupling lemma 3.6e), we can use c from Corollary 3.26b) on $\mathcal{E}'$. With $n_{\circ,\mathfrak{g}}$ such that $\ln(n)/\sqrt{n} \le \psi_\downarrow/4$ for $n \ge n_\circ$, we have $\gamma_{\mathrm{n},\hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}}} \ge \psi_\downarrow/2$ on $\mathcal{E}'$ if $n \ge n_\circ$, so using c' from Corollary 3.34 and $\|\gamma_{\mathrm{n},\hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}}} - \gamma_{\mathrm{n},\hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}}}\|_{\mathrm{tv}} \le 2r(n)$ on $\mathcal{E}'$ we obtain

$$|n\mathbb{E}\left[\mathbb{1}\mathcal{E}'\mathbb{1}\{\hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}} \neq \hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}}\}(\phi^*_{\boldsymbol{m}+1}(\hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}}) - \phi^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}}))\right]| \le cc'\left(2r(n) + \frac{k}{n}\right) = \mathcal{O}\left(\frac{\ln(n)}{\sqrt{n}}\right).$$

Now, we substitute $\hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}}$ and then drop $\mathbb{1}\{\hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}} = \hat{\boldsymbol{\sigma}}^-_{\boldsymbol{m}}, \hat{\boldsymbol{\sigma}}^+_{\boldsymbol{m}} \in \mathcal{B}^\Gamma\}$ at expense $\mathcal{O}(1/n)$. □

With Lemma 3.62 we obtain $\boldsymbol{G}^*_{\boldsymbol{m}+1,\boldsymbol{m}^{\leftrightarrow},\boldsymbol{\mathcal{U}}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}})$ from $\hat{\boldsymbol{G}}$ given $(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}, \hat{\boldsymbol{G}})$ by attaching a single additional standard factor $\boldsymbol{m}+1$, since the ground truths coincide, as do the decorations. So, we consider $\hat{\boldsymbol{G}} \otimes \boldsymbol{w}^*_{\circ,\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}}$, follow the steps in Section 3.3.2.2 to reduce this to $(\hat{\boldsymbol{G}}, \boldsymbol{v}, \boldsymbol{\psi}) \sim \hat{\boldsymbol{G}} \otimes \boldsymbol{w}_\circ$ and thereby, using

$\hat{\gamma}_m = \gamma_{\mathrm{n},\hat{\sigma}}$, obtain

$$\Delta^\circ = \mathbb{E}\left[\frac{\mathbb{1}\mathcal{E}}{\overline{Z}_{\mathrm{f}}(\hat{\gamma}_{\boldsymbol{m}})}\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m},\boldsymbol{v}})\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\hat{\boldsymbol{G}}}(\boldsymbol{v},\boldsymbol{\psi})\right)\right] + \mathcal{O}(r(n)).$$

Now, recall that $\mathcal{E}$ covers $\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}} \in \mathcal{B}^\Gamma$, so with c from Observation 3.9d) and Observation 3.9a) we have $1 - \psi_\uparrow cr(n)^2 \le \overline{Z}_{\mathrm{f}}(\hat{\gamma}_{\boldsymbol{m}})/\xi \le 1$, which yields $\overline{Z}_{\mathrm{f}}(\hat{\gamma}_{\boldsymbol{m}})/\xi = 1 + \mathcal{O}(r(n)^2)$ and

$$\Delta^\circ = (1 + \mathcal{O}(r(n)^2))\mathbb{E}\left[\frac{\mathbb{1}\mathcal{E}}{\xi}\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m},\boldsymbol{v}})\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\hat{\boldsymbol{G}}}(\boldsymbol{v},\boldsymbol{\psi})\right)\right] + \mathcal{O}(r(n)).$$

With $|\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m},\boldsymbol{v}})\ln(\overline{\psi}_{\mathrm{w}|\mathrm{g},\hat{\boldsymbol{G}}}(\boldsymbol{v},\boldsymbol{\psi}))/\xi| \le \psi_\uparrow^2\ln(\psi_\uparrow)$ and the Nishimori condition 3.22d), analogously to Section 3.3.2.2, we obtain

$$\Delta^\circ = \mathbb{E}\left[\frac{\mathbb{1}\mathcal{E}}{\xi}\Lambda\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\hat{\boldsymbol{G}}}(\boldsymbol{v},\boldsymbol{\psi})\right)\right] + \mathcal{O}(r(n)).$$

We can drop the restriction to $\mathcal{E}$ due to the uniform bound $\psi_\uparrow^2\ln(\psi_\uparrow)$ on the argument of the expectation at expense $\mathcal{O}(e^{-\tilde{c}\ln(n)^2})$ with $\tilde{c}$ from the proof of Lemma 3.62. Finally, we turn to the application of Proposition 3.48. With $\iota_\circ$ from Section 3.3.1.5 and $\hat{\boldsymbol{\mu}}$ from Proposition 3.48 notice that $\mathbb{E}[\iota_\circ(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^\leftrightarrow},\boldsymbol{v})] = \mathbb{E}[\iota_\ell(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^\leftrightarrow})]$ for $\ell = k$ and hence we can use Proposition 3.48a), Markov's inequality and the bound $\psi_\uparrow^2\ln(\psi_\uparrow)$ on the argument of the expectation with $\delta = \Theta^{\downarrow-2/3}$ to obtain

$$\Delta^\circ = \mathbb{E}\left[\frac{\mathbb{1}\mathcal{E}}{\xi}\Lambda\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\hat{\boldsymbol{G}}}(\boldsymbol{v},\boldsymbol{\psi})\right)\right] + \mathcal{O}\left(r(n) + \Theta^{\downarrow-1/3}\right),\ \mathcal{E} = \{\iota_\circ(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^\leftrightarrow},\boldsymbol{v}) < \delta\}.$$

With $\nu_\circ$ from Remark 3.49 and by Observation 3.6e) there exists a coupling $(\boldsymbol{\tau},\boldsymbol{\tau}')$ of $\mu|_v$ and $\bigotimes_h \mu|_{v(h)}$ such that $\mathbb{P}(\boldsymbol{\tau} \neq \boldsymbol{\tau}^*) \le \nu_\circ(\mu, v)$ and hence $|\mathbb{E}[\psi(\boldsymbol{\tau}) - \psi(\boldsymbol{\tau}')]| \le 2\psi_\uparrow\nu_\circ(\mu, v)$, $\psi \in \mathcal{D}_\Psi$. So, with $\boldsymbol{\zeta} = \sum_\tau \boldsymbol{\psi}(\tau)\prod_h \mu_{\mathrm{g},\hat{\boldsymbol{G}}}|_{\boldsymbol{v}(h)}(\tau_h)$ we have $|\overline{\psi}_{\mathrm{w}|\mathrm{g},\hat{\boldsymbol{G}}}(\boldsymbol{v},\boldsymbol{\psi}) - \boldsymbol{\zeta}| \le 2\psi_\uparrow\nu_\circ(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^\leftrightarrow},\boldsymbol{v})$. Now, we can use Lipschitz continuity of Λ on $[\psi_\downarrow,\psi_\uparrow]$ since both arguments live in this interval, i.e. we obtain $L_{\mathfrak{g}}$ such that $|\Lambda(\overline{\psi}_{\mathrm{w}|\mathrm{g},\hat{\boldsymbol{G}}}(\boldsymbol{v},\boldsymbol{\psi})) - \Lambda(\boldsymbol{\zeta})| \le 2L\psi_\uparrow\nu_\circ(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^\leftrightarrow},\boldsymbol{v})$. With Remark 3.49 we have $|\Lambda(\overline{\psi}_{\mathrm{w}|\mathrm{g},\hat{\boldsymbol{G}}}(\boldsymbol{v},\boldsymbol{\psi})) - \Lambda(\boldsymbol{\zeta})| \le \sqrt{2}L\psi_\uparrow\sqrt{\iota_\circ(\hat{\boldsymbol{\mu}}_{\boldsymbol{m},\boldsymbol{m}^\leftrightarrow},\boldsymbol{v})} \le \sqrt{2}L\psi_\uparrow\sqrt{\delta}$ on $\mathcal{E}$. Then we drop the restriction to $\mathcal{E}$ and notice that $(\mu_{\mathrm{g},\hat{\boldsymbol{G}}}|_{\boldsymbol{v}(h)})_h \sim \pi_{\mathrm{g},\hat{\boldsymbol{G}}}^{\otimes k}$ since $\boldsymbol{v}$ is uniform, so

$$\xi\Delta^\circ = \mathbb{E}[\Lambda(Z_{\mathrm{f}}(\boldsymbol{\psi},\boldsymbol{\gamma}_{\boldsymbol{\pi},1}))] + \mathcal{O}\left(r(n) + \Theta^{\downarrow-1/3}\right).$$

This completes the proof since $\Theta^\downarrow \le n$ and hence $\Theta^{\downarrow-1/3} = \omega(r(n))$.

3.3.2.4 Proof of Proposition 3.57. Combining Lemma 3.58, Lemma 3.59 and Lemma 3.60 gives

$$\begin{aligned}\frac{\partial}{\partial t^\leftrightarrow}\phi^\leftrightarrow(t^\leftrightarrow) &= \frac{\bar{d}}{k\xi}\mathbb{E}[\Lambda(Z_{\mathrm{f}}(\boldsymbol{\psi},\boldsymbol{\gamma}_{\boldsymbol{\pi},1}))] + \mathcal{O}\left(\Theta^{\downarrow-1/3}\right) + \phi_\circ - \frac{\bar{d}}{\xi}\mathbb{E}[\Lambda(Z_{\mathrm{fm}}(\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}_{\boldsymbol{\pi}}))]\\ &= \frac{\bar{d}}{\xi k}\mathbb{E}[\nabla(\boldsymbol{\pi})] + \mathcal{O}\left(\Theta^{\downarrow-1/3}\right)\end{aligned}$$

using $\bar{d} \in [0, d_\uparrow]$ and $\psi_\downarrow \le \xi \le \psi_\uparrow$ from Observation 3.9a).

Lemma 3.63. *For some $L_{\mathfrak{g}} \in \mathbb{R}_{>0}$ we have $|\nabla(\pi_1,\pi_3) - \nabla(\pi_2,\pi_3)| \le L\mathrm{d}_{\mathrm{w}}(\pi_1,\pi_2)$, $\pi \in \mathcal{P}^2([q])^3$.*

Proof. Let $\ell_{\mathfrak{g}} \in \mathbb{R}_{>0}$ be such that $|\Lambda(t_1) - \Lambda(t_2)| \le \ell|t_1 - t_2|$ for $t \in [\psi_\downarrow, \psi_\uparrow]^2$. For a coupling $\rho \in \Gamma(\pi_1, \pi_2)$ let $(\boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}) \sim \mu_\Psi \otimes \mathrm{u}([k]) \otimes (\rho^{\otimes k} \otimes \pi_3^{\otimes k})$ with $\boldsymbol{\gamma} \in (\mathcal{P}([q])^k)^3$, so with Jensen's inequality and the triangle inequality we have

$$\begin{aligned}\delta &= |\nabla(\pi_1, \pi_3) - \nabla(\pi_2, \pi_3)| \\ &\le \ell\mathbb{E}[|Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_1) - Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_2)|] + k\ell\mathbb{E}[|Z_{\mathrm{fm}}(\boldsymbol{\psi}, \boldsymbol{h}, (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_3)) - Z_{\mathrm{fm}}(\boldsymbol{\psi}, \boldsymbol{h}, (\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3))|].\end{aligned}$$

Expanding the definitions, using the triangle inequality, $\boldsymbol{\psi} \le \psi_\uparrow$ and Observation 3.6b)yields

$$\delta \le \ell\psi_\uparrow\mathbb{E}\left[\left\|\bigotimes_h \boldsymbol{\gamma}_{1,h} - \bigotimes_h \boldsymbol{\gamma}_{1,h}\right\|_{\mathrm{tv}}\right] + k\psi_\uparrow\mathbb{E}\left[\|\boldsymbol{\gamma}_{1,\boldsymbol{h}} - \boldsymbol{\gamma}_{2,\boldsymbol{h}}\|_{\mathrm{tv}}\right] \le L\mathbb{E}\left[\|\boldsymbol{\gamma}_{1,1} - \boldsymbol{\gamma}_{2,1}\|_{\mathrm{tv}}\right]$$

with $L = 2k\ell\psi_\uparrow$, which completes the proof since this holds uniformly for all couplings ρ. ☐

Now, Proposition 3.57 follows from Lemma 3.63 with Lemma 3.56b), $|\nabla| \le 2k\Lambda(\psi_\uparrow)$, further Corollary 3.12 and $1/\Theta^\downarrow = \Omega(1/n)$.

3.3.2.5 Proof of Proposition 3.4. First, we derive the result for graphs with pins and external fields, but without interpolators.

Lemma 3.64. *For $t^\leftrightarrow = 1$ we have $\mathbb{E}[\phi_{\mathrm{g}}(\hat{\boldsymbol{G}})] \ge B_\uparrow + \mathcal{O}(\frac{1}{\Theta^{\downarrow 1/3}} + \frac{\Theta^\downarrow}{n})$.*

Proof. From Proposition 3.57 we obtain $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that

$$\frac{\partial}{\partial t^\leftrightarrow}\phi^\leftrightarrow(t^\leftrightarrow) \ge \nabla_\downarrow - \frac{c}{\Theta^{\downarrow 1/3}} \ge -\frac{c}{\Theta^{\downarrow 1/3}}$$

since $\nabla_\downarrow \ge 0$ by assumption, so integration yields $\phi^\leftrightarrow(1) - \phi^\leftrightarrow(0) \ge -c/\Theta^{\downarrow 1/3}$. But for $t^\leftrightarrow = 0$ with $\check{\boldsymbol{\psi}}_i$ denoting the pin, i.e. $\check{\boldsymbol{\psi}}_i \equiv 1$ for $i \notin \boldsymbol{\mathcal{U}}$ and $\check{\boldsymbol{\psi}}_i(\tau) = \mathbb{1}\{\tau = \hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}\}$ otherwise, and using the notions from Observation 3.13 we have

$$\begin{aligned}\phi^\leftrightarrow(0) &= \frac{1}{n}\mathbb{E}\left[\ln\left(\prod_i \sum_\tau \gamma^*(\tau)\check{\boldsymbol{\psi}}_i(\tau) \prod_{h \in [\boldsymbol{m}_i^\leftrightarrow]} \boldsymbol{\psi}_{i,h}^{*\leftrightarrow}(\tau)\right)\right] \\ &= \sum_i \frac{1}{n}\mathbb{E}\left[\ln\left(\sum_\tau \gamma^*(\tau)\check{\boldsymbol{\psi}}_i(\tau) \prod_{h \in [\boldsymbol{m}_i^\leftrightarrow]} \boldsymbol{\psi}_{i,h}^{*\leftrightarrow}(\tau)\right)\right]\end{aligned}$$

Notice that $\hat{\boldsymbol{\sigma}}_0 \sim \boldsymbol{\sigma}^*$ and $\boldsymbol{m}_i^\leftrightarrow \sim \mathrm{Po}(\bar{d})$, so

$$\phi^\leftrightarrow(0) = \mathbb{E}\left[\ln\left(\sum_\tau \gamma^*(\tau)\check{\boldsymbol{\psi}}_1(\tau) \prod_{h \in [\boldsymbol{m}_1^\leftrightarrow]} \boldsymbol{\psi}_{1,h}^{*\leftrightarrow}(\tau)\right)\right].$$

Notice that the argument of the logarithm is in $[\psi_\downarrow^{\boldsymbol{m}_1^\leftrightarrow + 1}, \psi_\uparrow^{\boldsymbol{m}^\leftrightarrow}]$ and the probability of $\check{\boldsymbol{\psi}}_1 \equiv 1$ is

$1-\mathbb{E}[\boldsymbol{\theta}/n]=1-\Theta^{\downarrow}/(2n)$, so we have

$$\phi^{\leftrightarrow}(0)=\left(1-\frac{\Theta^{\downarrow}}{2n}\right)\mathbb{E}\left[\ln\left(\sum_{\tau}\gamma^*(\tau)\prod_{h\in[\boldsymbol{m}_1^{\leftrightarrow}]}\boldsymbol{\psi}^{*\leftrightarrow}_{1,h}(\tau)\right)\right]+(\bar{d}+1)\ln(\psi_{\uparrow})\mathcal{O}\left(\frac{\Theta^{\downarrow}}{n}\right)$$
$$=\mathbb{E}\left[\ln\left(\sum_{\tau}\gamma^*(\tau)\prod_{h\in[\boldsymbol{m}_1^{\leftrightarrow}]}\boldsymbol{\psi}^{*\leftrightarrow}_{1,h}(\tau)\right)\right]+\mathcal{O}\left(\frac{\Theta^{\downarrow}}{n}\right).$$

We recover the first contribution to the Bethe functional with the $(\boldsymbol{\psi}^{*\leftrightarrow}_{1,h},\boldsymbol{\psi}^{\leftrightarrow}_{\circ})$-derivatives, so

$$\mathbb{E}\left[\phi_{\mathrm{g}}(\hat{\boldsymbol{G}})\right]\geq B(\pi)+\mathcal{O}\left(\frac{1}{\Theta^{\downarrow 1/3}}+\frac{\Theta^{\downarrow}}{n}\right).$$

□

With Lemma 3.64 we restrict to $t^{\leftrightarrow}=1$, $m^{\leftrightarrow}\equiv 0$ in the remainder, where we also discuss all $\bar{d}\in[0,d_{\uparrow}]$. First, we turn to graphs with external fields. Recall m° and σ° from Proposition c).

Proposition 3.65. *Let $m^{\leftrightarrow}\equiv 0$ and $\mathcal{U}=\emptyset$.*
a) We have $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))]\geq B_{\uparrow}+\mathcal{O}(n^{-1/4})$.
b) For $d=km/n\leq d_{\uparrow}$ we have $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*))]\geq B_{\uparrow}(d)+\mathcal{O}(n^{-1/4})$.
c) We have $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}^*}(\boldsymbol{\sigma}^*))]\geq B_{\uparrow}+\mathcal{O}(n^{-1/4}+\delta_{\mathrm{m}}+\varepsilon_{\mathrm{m}})$.

Proof. With Lemma 3.64 and Proposition 3.50 we have

$$\mathbb{E}\left[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))\right]\geq B_{\uparrow}+\mathcal{O}\left(\frac{1}{\Theta^{\downarrow 1/3}}+\frac{\Theta^{\downarrow}}{n}\right).$$

This yields $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))]\geq B_{\uparrow}+\mathcal{O}(n^{-1/4})$ for $\Theta^{\downarrow}(n)=n^{3/4}\in(0,n]$ and $\bar{d}\in(0,d_{\uparrow}]$. For $\bar{d}=0$ notice that $\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))=0$ and $B\equiv 0$. Without loss of generality let $\delta_{\mathrm{m}}\geq\delta^{\circ}_{\mathrm{m}}$ with $\delta^{\circ}_{\mathrm{m}}=\ln(n)/\sqrt{n}\leq 1$ and $\varepsilon_{\mathrm{m}}\geq\varepsilon^{\circ}_{\mathrm{m}}$ with $\varepsilon^{\circ}_{\mathrm{m}}=c_2e^{-c_1\ln(n)^2/2}$ and c, c_2 large, from Corollary 3.12 since this does not affect the assertions. Hence, we may take $\boldsymbol{m}^*=\boldsymbol{m}$, and then Corollary 3.35b) applied to $\boldsymbol{m}^*$ and to $\boldsymbol{m}$ yields

$$\begin{aligned}\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}^*}(\boldsymbol{\sigma}^*))]&=\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{m^{\circ}}(\sigma^{\circ}))]+\mathcal{O}(\varepsilon_{\mathrm{m}}+\delta_{\mathrm{m}}+n^{-1/2})\\&=\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}}))]+\mathcal{O}(\varepsilon_{\mathrm{m}}+\delta_{\mathrm{m}}+n^{-1/2})\\&\geq B_{\uparrow}+\mathcal{O}(\varepsilon_{\mathrm{m}}+\delta_{\mathrm{m}}+n^{-1/4}),\end{aligned}$$

which establishes Part 3.65c). Now, consider the special case $\delta_{\mathrm{m}}=\delta^{\circ}_{\mathrm{m}}=o(n^{-1/4})$ and $\varepsilon_{\mathrm{m}}=\varepsilon^{\circ}_{\mathrm{m}}=o(n^{-1/4})$, then Part 3.65a) follows as a special case from Part 3.65c). But also Part 3.65b) now follows as a special case from Part 3.65c) by further considering $\bar{d}=d$ and $\boldsymbol{m}^*_{n'}=m^{\circ}_{n'}$ for $n'\in\mathbb{Z}_{>0}$, which in particular gives $\boldsymbol{m}^*=m$. □

Observation 3.21 yields the results for graphs without external fields, so Proposition 3.4 follows.

3.3.3 The Aizenman-Sims-Starr Scheme. This section is dedicated to the proof of Proposition 3.5, and hence Theorem 2.2. Fix $t^{\leftrightarrow}=1$, $m^{\leftrightarrow}\equiv 0$ in the remainder, which resolves dependencies on π,

$\psi^{\leftrightarrow}$ and $\boldsymbol{\psi}^{\leftrightarrow}$. With $C_1 \in (0,1)$ from Proposition 3.48b) let $c = C_1/3$, $\rho = c/(1+c)$ and $\Theta^{\downarrow}(n) = n^{1-\rho}$. Notice that $\rho \in (0, 1/4)$ and $\Theta^{\downarrow} \in [0, n]$. Let $\bar{d} > 0$ for now.

3.3.3.1 Overview. We avoided the introduction of the projected Gibbs marginal distribution $\pi^{\circ}_{\mathrm{g},G}$ from Section 3.3.1.9 in Section 3.1.2.4, but now we can state the stronger version.

Proposition 3.66. *We have* $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*)] = \mathbb{E}[B(\pi^{\circ}_{\mathrm{g},\boldsymbol{G}^*})] + \mathcal{O}(n^{-\rho})$ *with* $\boldsymbol{G}^* = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*)$.

We use the Aizenman-Sims-Starr scheme for Proposition 3.66, where the quenched free entropy density is understood as the average change of the quenched free entropies, meaning

$$\mathbb{E}\left[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{n,\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*))\right] = \sum_{n'=0}^{n-1} \frac{1}{n}\Phi_{\Delta,n'},\ \Phi_{\Delta,n'} = \mathbb{E}\left[(n'+1)\phi_{\mathrm{g}}\left(\boldsymbol{G}_{+,n'}\right)\right] - \mathbb{E}\left[n'\phi_{\mathrm{g}}\left(\boldsymbol{G}_{-,n'}\right)\right],$$
$$\boldsymbol{G}_{+,n} = \boldsymbol{G}^*_{n+1,\boldsymbol{m}_{n+1},\boldsymbol{\mathcal{U}}_{n+1}}(\boldsymbol{\sigma}^*_{n+1}),\ \boldsymbol{G}_{-,n} = \boldsymbol{G}^*_{n,\boldsymbol{m}_n,\boldsymbol{\mathcal{U}}_n}(\boldsymbol{\sigma}^*_n),$$

using $\mathbb{E}[n'\phi_{\mathrm{g}}(\boldsymbol{G}_{-,n'})] = 0$ for $n' = 0$. Intuitively, we observe that if $\Phi_{\Delta,n}$ converges, so does the quenched free entropy density, with the same limit. Hence, the main focus of this section is to establish the following result.

Lemma 3.67. *We have* $\Phi_{\Delta,n} = \mathbb{E}[B(\pi^{\circ}_{\mathrm{g},\boldsymbol{G}^*})] + \mathcal{O}(n^{-\rho})$ *with* $\boldsymbol{G}^* = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*)$.

Similar to Section 3.3.2 and Section 3.2.3 we will control the difference Φ_{Δ} of the expectations by introducing a coupling of $\boldsymbol{G}_-$ and $\boldsymbol{G}_+$, say $(\boldsymbol{G}^-, \boldsymbol{G}^+)$. However, as opposed to the previous sections we now have to deal with an additional variable. Since the average degree is $\bar{d}$, i.e. we expect the new variable to wire to $\bar{d}$ factors, but the expected difference in the number of factors is only $\bar{d}/k$, we will have to rewire factors – like in Section 3.2.3. But as opposed to Section 3.2.3 we cannot afford rough estimates, and have to control the behavior on a very granular level instead.

We can partially recover the convenient situation in Section 3.3.2 by taking the intersection graph, or base graph, as a starting point and then enrich this graph to obtain $\boldsymbol{G}^-$ and $\boldsymbol{G}^+$ each, say a triplet $(\boldsymbol{G}_{\cap}, \boldsymbol{G}^-, \boldsymbol{G}^+)$. The expectations give $\bar{d}n/k$ factors for $\boldsymbol{G}^-$, $\bar{d}(n+1)/k$ factors for $\boldsymbol{G}^+$, with roughly $\bar{d}$ wired to $i = n+1$. So, we can hope for $\frac{\bar{d}(n+1)}{k} - \bar{d} = \frac{\bar{d}n}{k} - \frac{\bar{d}(k+1)}{k}$ factors in $\boldsymbol{G}_{\cap}$ and attaching the remaining factors to obtain $\boldsymbol{G}^-$ and $\boldsymbol{G}^+$ respectively. This coupling allows to rewrite

$$\Phi_{\Delta}(n) = \mathbb{E}\left[\ln\left(\frac{Z_{\mathrm{g}}(\boldsymbol{G}^+)}{Z_{\mathrm{g}}(\boldsymbol{G}_{\cap})}\right)\right] - \mathbb{E}\left[\ln\left(\frac{Z_{\mathrm{g}}(\boldsymbol{G}^-)}{Z_{\mathrm{g}}(\boldsymbol{G}_{\cap})}\right)\right].$$

Since the coupling is fairly involved, we present it in three parts. In Section 3.3.3.2 we use the discussion in Section 3.2.1.7 to couple the standard factor graphs. Then we couple the factor counts using Observation 3.7. Finally, we turn to the pins in Section 3.3.3.4 and combine the three parts. In Section 3.3.3.5 we show that the law of $\boldsymbol{G}_{\cap}$ is close to $\boldsymbol{G}^-$, which allows to recycle our results for the teacher-student model. Then, in Section 3.3.3.6 we show that our rough estimates for the expectations are asymptotically correct.

Next, we discuss the asymptotics of the two contributions to $\Phi_{\Delta}(n)$ separately. We start with the easier $\boldsymbol{G}^-$-contribution, since only factors are added, which is covered by Sections 3.3.3.7 to 3.3.3.11. Then we discuss the $\boldsymbol{G}^+$-contribution in Sections 3.3.3.12 to 3.3.3.16.

While the discussion of the $\boldsymbol{G}^-$-contribution is conceptually similar to the discussion in Section 3.3.2.3, there are several additional obstacles. In Section 3.3.3.7 we discuss the restriction to typical instances. In Section 3.3.3.8 we introduce an approximation of the joint distribution to resolve dependencies. In Section 3.3.3.9 we use Proposition 3.48 to transition to independent Gibbs marginals.

In Section 3.3.3.10 we use Proposition 3.54 to resolve the dependencies of the Gibbs marginals on the ground truth. Finally, in Section 3.3.3.11 we discuss the remaining asymptotics, followed by the Lipschitz continuity of the factor contribution to the Bethe free entropy, which allows to transition to the projected Gibbs marginal distributions from Section 3.3.1.9.

Sections 3.3.3.12 to 3.3.3.16 are devoted to the respective steps for the $\boldsymbol{G}^+$-contribution, approaching the variable contribution to the Bethe free entropy. Finally, in Section 3.3.3.17 we establish Lemma 3.67, Proposition 3.66, Proposition 3.5 and the respective version for graphs with (normalized) external fields over general factor counts $\boldsymbol{m}^*$. In Section 3.3.3.18 we derive Theorem 2.2 for graphs with (normalized) external fields over general factor counts $\boldsymbol{m}^*$, and thereby complete the proof of Theorem 2.2.

3.3.3.2 Coupling Standard Graphs. For the sake of readability we suppress dependencies in the following sections unless required. Fix a ground truth $\sigma^+ \in [q]^{n+1}$ with $\sigma^- = \sigma^+_{[n]}$ and $m_{\mathrm{a}} = (m_\cap, m^-_\Delta, m^+_{\Delta-}, m^+_{\Delta+}) \in \mathbb{Z}^4_{\ge 0}$, meaning $m_\cap$ factors in the base graph with variables $[n]$, additional m^-_Δ factors in $\boldsymbol{G}^-$, additional $m^+_{\Delta-}$ factors in $\boldsymbol{G}^+$ that do not wire to the variable $i = n+1$ and $m^+_{\Delta+}$ factors that do wire to i. From these atoms we obtain the derived factor counts, namely

$$m^- = m_\cap + m^-_\Delta,\ m^+_- = m_\cap + m^+_{\Delta-},\ m^+_\Delta = m^+_{\Delta-} + m^+_{\Delta+}, m^+ = m_\cap + m^+_\Delta.$$

Recall the discussion in Section 3.2.1.7. We consider the wires-weight pairs

$$(\boldsymbol{w}_\cap, \boldsymbol{w}^-_\Delta, \boldsymbol{w}^+_{\Delta-}, \boldsymbol{w}^+_{\Delta+}) \sim \boldsymbol{w}^{*\otimes m_\cap}_{-\circ,n+1,i,\sigma^-} \otimes \boldsymbol{w}^{*\otimes m^-_\Delta}_{-\circ,n+1,i,\sigma^-} \otimes \boldsymbol{w}^{*\otimes m^+_{\Delta-}}_{-\circ,n+1,i,\sigma^-} \otimes \boldsymbol{w}^{*\otimes m^+_{\Delta+}}_{+\circ,n+1,i,\sigma^+}.$$

This yields the graph $\boldsymbol{w}^- = (\boldsymbol{w}_\cap, \boldsymbol{w}^-_\Delta)$ on n variables. Mimicking Section 3.2.1.7 for $n+1$ variables, m^+ factors, $i = n+1$ and $d = m^+_\Delta$, let $\boldsymbol{w}^+_- = (\boldsymbol{w}_\cap, \boldsymbol{w}^+_{\Delta-})$ be the pairs not connected to i and $\boldsymbol{w}^+_+ = \boldsymbol{w}^+_{\Delta+}$ the pairs connected to i. For $\mathcal{A} \in \binom{[m^+]}{d}$ let $\boldsymbol{w}^+_{\mathrm{a},\mathcal{A}}$ be the corresponding relabeling of $(\boldsymbol{w}^+_-, \boldsymbol{w}^+_+)$ and let $\boldsymbol{w}^+_{\mathrm{d}} = \boldsymbol{w}^+_{\mathrm{a},\boldsymbol{\mathcal{A}}}$ with $\boldsymbol{\mathcal{A}} \sim \mathrm{u}(\binom{[m^+]}{d})$. Further, recall the degree $\boldsymbol{d}^+_{m^+} \sim \mathrm{Bin}(m^+, p^+_{\mathrm{d}})$ with success probability $p^+_{\mathrm{d}} = p_{\mathrm{d},n+1,\sigma^+}(\sigma^+_i)$ from Section 3.2.1.6, $\boldsymbol{w}^*$ from Observation 3.13 and $\boldsymbol{w}^*_{\mathrm{d}}$ from Observation 3.19.

Lemma 3.68. *Let $i = n+1$ and $d = m^+_{\Delta+}$.*
a) We have $\boldsymbol{w}_{\cap,n,m_\cap}(\sigma^-) \sim \boldsymbol{w}^*_{n,m_\cap}(\sigma^-)$ and $\boldsymbol{w}^-_{n,m^-}(\sigma^-) \sim \boldsymbol{w}^*_{n,m^-}(\sigma^-)$.
b) We have $\boldsymbol{w}^+_{\mathrm{d},n,m^+,d}(\sigma^+) \sim \boldsymbol{w}^*_{\mathrm{d},n+1,m^+,i,d}(\sigma^+)$.
c) Let $(\boldsymbol{m}_\cap, \boldsymbol{m}^-_{\Delta-}, \boldsymbol{m}^+_{\Delta-}, \boldsymbol{d}) \in \mathbb{Z}^4_{\ge 0}$, $\boldsymbol{m}^+ = \boldsymbol{m}_\cap + \boldsymbol{m}^+_{\Delta-} + \boldsymbol{d}$ and $\boldsymbol{w}^+_n(\sigma^+) = \boldsymbol{w}^+_{\mathrm{d},n,\boldsymbol{m}^+,\boldsymbol{d}}(\sigma^+)$. Then we have $\boldsymbol{w}^+_n(\sigma^+) \sim \boldsymbol{w}^*_{n+1,\boldsymbol{m}^+}(\sigma^+)$ if $(\boldsymbol{m}^+, \boldsymbol{d}) \sim (\boldsymbol{m}^+, \boldsymbol{d}^+_{\boldsymbol{m}^+})$.

Proof. Recall from the proof of Observation 3.19 that the $(\boldsymbol{w}^*_{-\circ}, \boldsymbol{w}_\circ)$-derivative is given by $(v, \psi) \mapsto \mathbb{1}\{i \notin v([k])\}\psi(\sigma_v)/\mathbb{E}[\mathbb{1}\{i \notin \boldsymbol{v}_\circ([k])\}\boldsymbol{\psi}_\circ(\sigma_{\boldsymbol{v}_\circ})]$, which yields $\boldsymbol{w}^*_{-\circ,n+1,i,\sigma^-} \sim \boldsymbol{w}^*_{\circ,n,\sigma^-}$ since clearly $\boldsymbol{v}_{\circ,n+1}|i \notin \boldsymbol{v}_{\circ,n+1}([k]) \sim \boldsymbol{v}_{\circ,n}$, and thereby completes the proof of Part 3.68a). Part 3.68b) holds by construction since we explicitly mimicked the construction in Section 3.2.1.7. Part 3.68c) follows directly from Part 3.68b) and Observation 3.19. □

Remark 3.69. Notice that the coupling of the graphs does not require Poisson counts, but they are very convenient to avoid case distinctions, as mentioned below. In particular, we could take $\boldsymbol{m}^+ = m^+$, $\boldsymbol{m}^- = m^-$ and let $\boldsymbol{m}_\cap = \min(m^-, m^+ - \boldsymbol{d}^+_{m^+})$.

3.3.3.3 Coupling Factor Counts. In this section we introduce a coupling of $\boldsymbol{m}_n$ and $\boldsymbol{m}_{n+1}$ that meets the requirements of Lemma 3.68c). For this purpose recall the Poisson parameter $\overline{m}_n = \bar{d}n/k$ of $\boldsymbol{m}_n$, let $\overline{m}_n^- = \overline{m}_n$ and $\overline{m}_n^+ = \overline{m}_{n+1}$. Guided by Lemma 3.68c) and Observation 3.7b) let $\overline{m}_{\Delta+}^+ = p_{\mathrm{d}}^+\overline{m}^+$ and $\overline{m}_-^+ = (1-p_{\mathrm{d}}^+)\overline{m}^+$. Inspired by Remark 3.69, let $\overline{m}_\cap = \min(\overline{m}^-, \overline{m}_-^+)$, and denote the gaps by $\overline{m}_\Delta^- = \overline{m}^- - \overline{m}_\cap$ and $\overline{m}_{\Delta-}^+ = \overline{m}_-^+ - \overline{m}_\cap$. So, the basic Poisson counts are

$$(\boldsymbol{m}_\cap, \boldsymbol{m}_\Delta^-, \boldsymbol{m}_{\Delta-}^+, \boldsymbol{m}_{\Delta+}^+) \sim \mathrm{Po}(\overline{m}_\cap) \otimes \mathrm{Po}(\overline{m}_\Delta^-) \otimes \mathrm{Po}(\overline{m}_{\Delta-}^+) \otimes \mathrm{Po}(\overline{m}_{\Delta+}^+).$$

Let $\boldsymbol{m}^- = \boldsymbol{m}_\cap + \boldsymbol{m}_\Delta^-$, $\boldsymbol{m}^+ = \boldsymbol{m}_\cap + \boldsymbol{m}_{\Delta-}^+ + \boldsymbol{m}_{\Delta+}^+$ and $\boldsymbol{w}_\cap$, $\boldsymbol{w}^-$, $\boldsymbol{w}^+$ from Lemma 3.68.

Lemma 3.70. *We have* $\boldsymbol{w}_{\cap,n,\boldsymbol{m}_\cap(\sigma^+)}(\sigma^-) \sim \boldsymbol{w}^*_{n,\boldsymbol{m}_\cap(\sigma^+)}(\sigma^-)$, $\boldsymbol{w}^-_{n,\boldsymbol{m}^-}(\sigma^-) \sim \boldsymbol{w}^*_{n,\boldsymbol{m}^-}(\sigma^-)$, $\boldsymbol{w}_n^+(\sigma^+) \sim \boldsymbol{w}^*_{n+1,\boldsymbol{m}^+}(\sigma^+)$, *further* $\boldsymbol{m}^- \sim \boldsymbol{m}_n$ *and* $\boldsymbol{m}^+ \sim \boldsymbol{m}_{n+1}$.

Proof. With Observation 3.7b) we have $\boldsymbol{m}^- \sim \mathrm{Po}(\overline{m}^-)$ since $\overline{m}^- = \overline{m}_\cap + \overline{m}_\Delta^-$, further $\boldsymbol{m}_-^+ = \boldsymbol{m}_\cap + \boldsymbol{m}_{\Delta-}^+ \sim \mathrm{Po}(\overline{m}_-^+)$ since $\overline{m}_-^+ = \overline{m}_\cap + \overline{m}_{\Delta-}^+$ and $\boldsymbol{m}^+ \sim \mathrm{Po}(\overline{m}^+)$ since $\overline{m}^+ = \overline{m}_-^+ + \overline{m}_{\Delta+}^+$. Hence, Observation 3.7b) further yields that $(\boldsymbol{m}^+, \boldsymbol{m}_{\Delta+}^+) \sim (\boldsymbol{m}^+, \boldsymbol{d}_{\boldsymbol{m}^+}^+)$, so Lemma 3.68c) applies, and completes the proof with Lemma 3.68a). Notice that $\boldsymbol{m}_-^+$ depends on σ^+ through p_{d}^+ and hence $\boldsymbol{m}_\cap = \boldsymbol{m}_\cap(\sigma^+)$ depends on σ^+. □

Remark 3.71. The notation for $\boldsymbol{w}_\cap$, $\boldsymbol{w}^-$ and $\boldsymbol{w}^+$ is inconsistent, so we change it as follows. Let $\boldsymbol{W}_{\mathrm{p}}(\sigma^+) = (\boldsymbol{w}_{\cap,\boldsymbol{m}_\cap(\sigma^+)}(\sigma^-), \boldsymbol{w}^-_{\boldsymbol{m}^-}(\sigma^-), \boldsymbol{w}^+(\sigma^+))$ and $\boldsymbol{W}_{\mathrm{m},M}(\sigma^-) = (\boldsymbol{W}_{\mathrm{p}}|(\boldsymbol{m}_\cap, \boldsymbol{m}^-, \boldsymbol{m}^+) = M)$ be the pairs for random and given counts respectively. Let $(\boldsymbol{w}_{\cap,m_\cap}(\sigma^-), \boldsymbol{w}^-_{m^-}(\sigma^-), \boldsymbol{w}^+_{m^+}(\sigma^+)) = \boldsymbol{W}_{\mathrm{m},(m_\cap,m^-,m^+)}(\sigma^+)$ and $\boldsymbol{W}_{\mathrm{p}}(\sigma^+) \sim (\boldsymbol{w}_{\cap,\boldsymbol{m}_\cap(\sigma^+)}(\sigma^-), \boldsymbol{w}^-_{\boldsymbol{m}^-}(\sigma^-), \boldsymbol{w}^+_{\boldsymbol{m}^+}(\sigma^+))$.

Also in the new notation we let $\boldsymbol{m}_{\Delta+}^+ \sim \mathrm{Po}(\overline{m}_{\Delta+}^+)$ be the factor degree of $i = n+1$ in $\boldsymbol{w}^+_{\boldsymbol{m}^+}(\sigma^+)$, so $\boldsymbol{m}_-^+ = \boldsymbol{m}^+ - \boldsymbol{m}_{\Delta-}^+ \sim \mathrm{Po}(\overline{m}_-^+)$ factors are not wired to i in $\boldsymbol{w}^+_{\boldsymbol{m}^+}(\sigma^+)$. Similarly, we e.g. still have $\boldsymbol{w}^-_{\boldsymbol{m}^-,[\boldsymbol{m}_\cap(\sigma^+)]}(\sigma^-) = \boldsymbol{w}_{\cap,\boldsymbol{m}_\cap(\sigma^+)}(\sigma^-)$.

3.3.3.4 Coupling Pins. Notice that $\Theta^\downarrow_{n+1} - \Theta^\downarrow_n > 0$. Let $\boldsymbol{\theta}_n^+ \sim \mathrm{u}([0, \Theta^\downarrow_{n+1}])$ and let $\boldsymbol{\theta}_n^- = \boldsymbol{\theta}^-_{n,\boldsymbol{\theta}^+}$ be given by $\boldsymbol{\theta}_\theta^- = \theta$ for $\theta \in [0, \Theta^\downarrow_n]$ and $\boldsymbol{\theta}_\theta^- \sim \mathrm{u}([0, \Theta^\downarrow_n])$ otherwise. Recall $\boldsymbol{\theta}$ from Section 3.2.1.1.

Lemma 3.72. *We have* $\boldsymbol{\theta}^- \le \boldsymbol{\theta}^+$, $\boldsymbol{\theta}_n^- \sim \boldsymbol{\theta}_n$ *and* $\boldsymbol{\theta}_n^+ \sim \boldsymbol{\theta}_{n+1}$.

Proof. By construction we have $\boldsymbol{\theta}_n^+ \sim \boldsymbol{\theta}_{n+1}$ and $\boldsymbol{\theta}^- \le \boldsymbol{\theta}^+$. Further, for an event $\mathcal{E} \subseteq [0, \Theta^\downarrow_n]$ and with $(\boldsymbol{u}, \boldsymbol{\theta}^+) \sim \boldsymbol{u} \otimes \boldsymbol{\theta}^+$, $\boldsymbol{u} \sim \mathrm{u}([0, \Theta^\downarrow_n])$, $I = \int \mathbb{1}\{t \in \mathcal{E}\}\mathrm{d}t$ we have

$$\mathbb{P}(\boldsymbol{\theta}^- \in \mathcal{E}) = \mathbb{P}(\boldsymbol{\theta}^+ \in \mathcal{E}) + \mathbb{P}(\boldsymbol{\theta}^+ > \Theta^\downarrow_n, , \boldsymbol{u} \in \mathcal{E}) = \frac{I}{\Theta^\downarrow_{n+1}} + \frac{(\Theta^\downarrow_{n+1} - \Theta^\downarrow_n)I}{\Theta^\downarrow_{n+1}\Theta^\downarrow_n} = \mathbb{P}(\boldsymbol{\theta}_n \in \mathcal{E}).$$

□

For given $\theta^- \in [0, n]$, $\theta^+ \in [0, n+1]$ with $\theta^- \le \theta^+$ we consider the success probabilities

$$p_\cap = \frac{\theta^-}{n+1},\ p_\Delta^- = \frac{\theta^-}{n(n+1) - n\theta^-},\ p_\Delta^+ = \frac{\theta^+ - \theta^-}{n+1-\theta^-}.$$

Let $\check{\boldsymbol{u}}_{\cap\circ}, \check{\boldsymbol{u}}^-_{\Delta\circ}, \check{\boldsymbol{u}}^+_{\Delta\circ} \in \{0,1\}$ be given by the success probabilities $p_\cap, p^-_\Delta, p^+_\Delta$ respectively, $\check{\boldsymbol{u}}_{\times\mathrm{t},\theta^-,\theta^+} = (\check{\boldsymbol{u}}_{\cap\circ} \otimes \check{\boldsymbol{u}}^-_{\Delta\circ} \otimes \check{\boldsymbol{u}}^+_{\Delta\circ})^{\otimes(n+1)}$, $\check{\boldsymbol{u}}_{\times,n} = (\check{\boldsymbol{u}}_\cap, \check{\boldsymbol{u}}^-_\Delta, \check{\boldsymbol{u}}^+_\Delta) = \check{\boldsymbol{u}}_{\times\mathrm{t},\boldsymbol{\theta}^-,\boldsymbol{\theta}^+}$ and

$$\boldsymbol{\mathcal{U}}_\cap = \{i \in [n] : \check{\boldsymbol{u}}_\cap(i) > 0\}, \boldsymbol{\mathcal{U}}^- = \{i \in [n] : \check{\boldsymbol{u}}_\cap(i) + \check{\boldsymbol{u}}^-_\Delta(i) > 0\},$$
$$\boldsymbol{\mathcal{U}}^+ = \{i \in [n+1] : \check{\boldsymbol{u}}_\cap(i) + \check{\boldsymbol{u}}^+_\Delta(i) > 0\}.$$

Recall the pinning set $\boldsymbol{\mathcal{U}}_n$ from Section 3.2.1.1.

Lemma 3.73. *We have* $\boldsymbol{\mathcal{U}}^- \sim \boldsymbol{\mathcal{U}}_n$ *and* $\boldsymbol{\mathcal{U}}^+ \sim \boldsymbol{\mathcal{U}}_{n+1}$.

Proof. Recall $\check{\boldsymbol{u}}_{\mathrm{t}\circ}$ with success probability θ/n from Section 3.2.1.1. For given $\theta^- \in [0,n]$, $\theta^+ \in [0,n+1]$ with $\theta^- \le \theta^+$ let $(\check{\boldsymbol{u}}_{\cap\circ}, \check{\boldsymbol{u}}^-_{\Delta\circ}, \check{\boldsymbol{u}}^+_{\Delta\circ}) = \check{\boldsymbol{u}}_{\cap\circ} \otimes \check{\boldsymbol{u}}^-_{\Delta\circ} \otimes \check{\boldsymbol{u}}^+_{\Delta\circ}$ and notice that

$$\mathbb{P}(\check{\boldsymbol{u}}_{\cap\circ} + \check{\boldsymbol{u}}^-_{\Delta\circ} > 0) = p_\cap + (1-p_\cap)p^-_\Delta = \frac{\theta^-}{n} = \mathbb{P}(\check{\boldsymbol{u}}_{\mathrm{t}\circ,\theta^-,n} = 1),$$
$$\mathbb{P}(\check{\boldsymbol{u}}_{\cap\circ} + \check{\boldsymbol{u}}^+_{\Delta\circ} > 0) = p_\cap + (1-p_\cap)p^+_\Delta = \frac{\theta^+}{n+1} = \mathbb{P}(\check{\boldsymbol{u}}_{\mathrm{t}\circ,\theta^+,n+1} = 1).$$

Now, let $n^- = n$, $n^+ = n+1$ and $\check{\boldsymbol{u}}_{\times\mathrm{t}} = (\check{\boldsymbol{u}}_{\cap\mathrm{t}}, \check{\boldsymbol{u}}^-_{\Delta\mathrm{t}}, \check{\boldsymbol{u}}^+_{\Delta\mathrm{t}})$. Then the above shows that $\check{\boldsymbol{u}}^\pm_{\theta^\pm} = (\min(1, \check{\boldsymbol{u}}_{\cap\mathrm{t},i} + \check{\boldsymbol{u}}^\pm_{\Delta\mathrm{t},i}))_{i\in[n^\pm]} \sim \check{\boldsymbol{u}}^{\otimes n^\pm}_{\mathrm{t}\circ,\theta^\pm,n^\pm}$, so Lemma 3.72 yields $\boldsymbol{\mathcal{U}}^\pm = \check{\boldsymbol{u}}^{\pm-1}_{\boldsymbol{\theta}^\pm}(1) \sim \boldsymbol{\mathcal{U}}_{n^\pm}$. □

Finally, we complete the coupling with $\boldsymbol{\sigma}^+ \sim \boldsymbol{\sigma}^*_{n+1}$ and $\boldsymbol{\sigma}^- = \boldsymbol{\sigma}^+_{[n]}$. In order to clarify the dependency structure recall $\boldsymbol{W}_{\mathrm{p},n}(\sigma^+)$ from Remark 3.71 and that it determines the factor counts. On the other hand we have $\boldsymbol{U}_n = (\boldsymbol{\theta}^-, \boldsymbol{\theta}^+, \check{\boldsymbol{u}}_\times)$ which determines $\boldsymbol{\mathcal{U}}_\cap, \boldsymbol{\mathcal{U}}^-, \boldsymbol{\mathcal{U}}^+$. The joint distribution is now given by $(\boldsymbol{\sigma}^+, \boldsymbol{W}_\mathrm{p}(\boldsymbol{\sigma}^+), \boldsymbol{U}) \sim (\boldsymbol{\sigma}^+, \boldsymbol{W}_\mathrm{p}(\boldsymbol{\sigma}^+)) \otimes \boldsymbol{U}$. Now, the graphs are $\boldsymbol{G}^\pm(\boldsymbol{\sigma}^\pm) = [\boldsymbol{w}^\pm_{n,\boldsymbol{m}^\pm}(\boldsymbol{\sigma}^\pm_n)]^{\Gamma\downarrow}_{\boldsymbol{\mathcal{U}}^\pm_n,\sigma^\pm_n}$ and $\boldsymbol{G}_\cap = [\boldsymbol{w}_{\cap,\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)}(\boldsymbol{\sigma}^-)]^{\Gamma\downarrow}_{\boldsymbol{\mathcal{U}}_\cap,\boldsymbol{\sigma}^-}$.

Proposition 3.74. *We have* $\boldsymbol{G}_\cap \sim \boldsymbol{G}^*_{\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+),\boldsymbol{\mathcal{U}}_\cap}(\boldsymbol{\sigma}^-)$*, further* $(\boldsymbol{\sigma}^-, \boldsymbol{G}^-(\boldsymbol{\sigma}^-)) \sim (\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*))$ *and* $(\boldsymbol{\sigma}^+, \boldsymbol{G}^+(\boldsymbol{\sigma}^+)) \sim (\boldsymbol{\sigma}^*_{n+1}, \boldsymbol{G}^*_{n+1,\boldsymbol{m}_{n+1},\boldsymbol{\mathcal{U}}_{n+1}}(\boldsymbol{\sigma}^*_{n+1}))$.

Proof. The result is immediate from Lemma 3.70, Lemma 3.73 and Observation 3.13. □

Notice that $\boldsymbol{G}^-(\boldsymbol{\sigma}^-)$ and $\boldsymbol{G}^+(\boldsymbol{\sigma}^+)$ are conditionally independent given $(\boldsymbol{\theta}^-, \boldsymbol{\theta}^+, \boldsymbol{\sigma}^+, \boldsymbol{G}_\cap)$ and obtained as follows. For $\boldsymbol{G}^-(\boldsymbol{\sigma}^-)$ we choose $\boldsymbol{m}^-_\Delta(\boldsymbol{\sigma}^+)$ and the additional standard factors iid from $\boldsymbol{w}^*_{\circ,n,\boldsymbol{\sigma}^-}$. Further, we perform a second sweep of pinning with probability $p^-_{\Delta,\boldsymbol{\theta}^-}$, i.e. pinning each (unpinned) variable $i \in [n]$ to $\boldsymbol{\sigma}^-(i)$ independently with probability $p^-_{\Delta,\boldsymbol{\theta}^-}$.

For $\boldsymbol{G}^+(\boldsymbol{\sigma}^+)$ we choose $\boldsymbol{m}^+_{\Delta-}(\boldsymbol{\sigma}^+)$ and the corresponding additional standard factors iid from $\boldsymbol{w}^*_{\circ,n,\boldsymbol{\sigma}^-}$. Further, we choose $\boldsymbol{m}^+_{\Delta+}(\boldsymbol{\sigma}^+)$ and the corresponding additional standard factors independently from $\boldsymbol{w}^*_{+\circ,n+1,i,\boldsymbol{\sigma}^+}$. Formally, we also have to randomly relabel all factors. Finally, we perform a second sweep of pinning with probability $p^+_{\Delta,\boldsymbol{\theta}^-,\boldsymbol{\theta}^+}$ for the variables $[n]$ and pin $i = n+1$ with probability $\boldsymbol{\theta}^+/(n+1)$.

Since both $\boldsymbol{G}^-(\boldsymbol{\sigma}^-)$ and $\boldsymbol{G}^+(\boldsymbol{\sigma}^+)$ are obtained from $\boldsymbol{G}_\cap$ exclusively by adding factors, the ratios in $\Phi_\Delta(n) = \Phi_\mathrm{v}(n) - \Phi_\mathrm{f}(n)$ with

$$\Phi_\mathrm{v}(n) = \mathbb{E}\left[\ln\left(\frac{Z_\mathrm{g}(\boldsymbol{G}^+(\boldsymbol{\sigma}^+))}{Z_\mathrm{g}(\boldsymbol{G}_\cap)}\right)\right], \Phi_\mathrm{f}(n) = \mathbb{E}\left[\ln\left(\frac{Z_\mathrm{g}(\boldsymbol{G}^-(\boldsymbol{\sigma}^+))}{Z_\mathrm{g}(\boldsymbol{G}_\cap)}\right)\right], \tag{9}$$

can be understood as the expected additional weight caused by the new factors under the Gibbs spins $\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}_\cap}$, as in Section 3.3.2.2 and Section 3.3.2.3.

3.3.3.5 The Base Graph. We define a coupling for the pairs $(\boldsymbol{\sigma}^-,\boldsymbol{G}_\cap)$ and $(\boldsymbol{\sigma}^*_n,\boldsymbol{G}^*_{n,\boldsymbol{m}_n,\boldsymbol{\mathcal{U}}_n}(\boldsymbol{\sigma}^*_n))$. For this purpose we start with a coupling of $(\boldsymbol{\sigma}^-,\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+),\boldsymbol{\mathcal{U}}_\cap)$ and $(\boldsymbol{\sigma}^*_n,\boldsymbol{m}_n,\boldsymbol{\mathcal{U}}_n)$. For $\sigma\in[q]^{n+1}$ let $(\boldsymbol{m}'_{\cap,\sigma},\boldsymbol{m}'_\sigma)$ be a coupling of $\boldsymbol{m}_\cap(\sigma)$ and $\boldsymbol{m}^-_n$ from the coupling lemma 3.6e). This conditional law and $(\boldsymbol{\sigma}'_+,\boldsymbol{\mathcal{U}}'_\cap,\boldsymbol{\mathcal{U}}')\sim(\boldsymbol{\sigma}^+,\boldsymbol{\mathcal{U}}_\cap,\boldsymbol{\mathcal{U}}^-)$ induce $\boldsymbol{a}=(\boldsymbol{\sigma}'_+,\boldsymbol{m}'_\cap(\boldsymbol{\sigma}'_+),\boldsymbol{m}'(\boldsymbol{\sigma}'_+),\boldsymbol{\mathcal{U}}'_\cap,\boldsymbol{\mathcal{U}}')$, which further determines $\boldsymbol{\sigma}'_\cap=\boldsymbol{\sigma}'=\boldsymbol{\sigma}'_{+,[n]}$. For given $a=(\sigma_+,m_\cap,m,\mathcal{U}_\cap,\mathcal{U})$ with $\sigma=\sigma_{+,[n]}$ we obtain the graphs as follows. For $m_\cap=m$ and $\mathcal{U}_\cap=\mathcal{U}$ let $\boldsymbol{G}'_\cap(a)=\boldsymbol{G}'(a)\sim\boldsymbol{G}^*_{m,\mathcal{U}}(\sigma)$, otherwise let $(\boldsymbol{G}'_\cap(a),\boldsymbol{G}'(a))\sim\boldsymbol{G}^*_{m_\cap,\mathcal{U}_\cap}(\sigma)\otimes\boldsymbol{G}^*_{m,\mathcal{U}}(\sigma)$.

Lemma 3.75. *We have $(\boldsymbol{\sigma}'_\cap,\boldsymbol{G}'_\cap(\boldsymbol{a}))\sim(\boldsymbol{\sigma}^-,\boldsymbol{G}_\cap)$ and $(\boldsymbol{\sigma}',\boldsymbol{G}'(\boldsymbol{a}))\sim(\boldsymbol{\sigma}^*,\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*))$. Further, we have $\mathbb{P}((\boldsymbol{\sigma}'_\cap,\boldsymbol{G}'_\cap(\boldsymbol{a}))\neq(\boldsymbol{\sigma}',\boldsymbol{G}'(\boldsymbol{a})))=\mathcal{O}(n^{-\rho})$.*

Proof. We have $(\boldsymbol{\sigma}'_+,\boldsymbol{\sigma}'_\cap,\boldsymbol{\mathcal{U}}'_\cap,\boldsymbol{G}'_\cap(\boldsymbol{a}))\sim(\boldsymbol{\sigma}^+,\boldsymbol{\sigma}^-,\boldsymbol{\mathcal{U}}_\cap,\boldsymbol{G}^*_{\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+),\boldsymbol{\mathcal{U}}_\cap}(\boldsymbol{\sigma}^-))$ by definition, and further $(\boldsymbol{\sigma}',\boldsymbol{\mathcal{U}}',\boldsymbol{G}'(\boldsymbol{a}))\sim(\boldsymbol{\sigma}^-,\boldsymbol{\mathcal{U}}^-,\boldsymbol{G}^*_{\boldsymbol{m}^-,\boldsymbol{\mathcal{U}}^-}(\boldsymbol{\sigma}^-))$, so the first two assertions hold by Proposition 3.74. Since the graphs coincide if the counts do, we have $\mathbb{P}((\boldsymbol{\sigma}'_\cap,\boldsymbol{G}'_\cap(\boldsymbol{a}))\neq(\boldsymbol{\sigma}',\boldsymbol{G}'(\boldsymbol{a})))\leq\mathbb{P}((\boldsymbol{m}'_{\cap,\boldsymbol{\sigma}^+},\boldsymbol{\mathcal{U}}'_\cap)\neq(\boldsymbol{m}'_{\boldsymbol{\sigma}^+},\boldsymbol{\mathcal{U}}'))\leq\mathbb{P}(\boldsymbol{m}'_{\cap,\boldsymbol{\sigma}^+}\neq\boldsymbol{m}'_{\boldsymbol{\sigma}^+})+\mathbb{P}(\boldsymbol{\mathcal{U}}'_\cap\neq\boldsymbol{\mathcal{U}}')$. For the latter we have $\mathbb{P}(\boldsymbol{\mathcal{U}}'_\cap\neq\boldsymbol{\mathcal{U}}')\leq\mathbb{P}(|\check{\boldsymbol{u}}^{-\,-1}_\Delta(1)|>0)$, and further $\mathbb{P}(|\check{\boldsymbol{u}}^{-\,-1}_\Delta(1)|>0)\leq\mathbb{E}[|\check{\boldsymbol{u}}^{-\,-1}_\Delta(1)|]$ by Markov's inequality, where

$$\mathbb{E}[|\check{\boldsymbol{u}}^{-\,-1}_\Delta(1)|]=n\mathbb{E}[p^-_\Delta(\boldsymbol{\theta}^-)]\leq\frac{n\Theta^{\downarrow}_n}{n(n+1)-n\Theta^{\downarrow}_n}=(1+o(1))\frac{\Theta^{\downarrow}_n}{n}. \tag{10}$$

For the factor counts we use the definition, i.e. the coupling lemma 3.6e), Pinsker's inequality 3.6f) and Observation 3.7c) to obtain

$$\begin{aligned}\mathbb{P}(\boldsymbol{m}'_{\cap,\boldsymbol{\sigma}^+}\neq\boldsymbol{m}'_{\boldsymbol{\sigma}^+})&=\mathbb{E}\left[\|\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)-\boldsymbol{m}_n\|_{\mathrm{tv}}\right]\leq\mathbb{E}\left[\mathbb{E}\left[\sqrt{\frac{1}{2}D_{\mathrm{KL}}\left(\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)\|\boldsymbol{m}_n\right)}\,\middle|\,\boldsymbol{\sigma}^+\right]\right]\\&=\frac{1}{\sqrt{2}}\mathbb{E}\left[\sqrt{\overline{m}_n-\overline{m}_\cap(\boldsymbol{\sigma}^+)+\overline{m}_\cap(\boldsymbol{\sigma}^+)\ln\left(\frac{\overline{m}_\cap(\boldsymbol{\sigma}^+)}{\overline{m}_n}\right)}\right].\end{aligned}$$

The argument of the expectation vanishes for $\overline{m}_\cap(\boldsymbol{\sigma}^+)=\overline{m}^-=\overline{m}_n$. Otherwise, we have $\overline{m}_\cap(\boldsymbol{\sigma}^+)=\overline{m}^+-\overline{m}^+_{\Delta+}(\boldsymbol{\sigma}^+)<\overline{m}^-$, or $\boldsymbol{\Delta}>0$ with $\boldsymbol{\Delta}=\overline{m}^+_{\Delta+}(\boldsymbol{\sigma}^+)-\bar{d}/k$, and using $\ln(1-t)\leq-t$ further

$$\begin{aligned}\mathbb{P}(\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)\neq\boldsymbol{m}_n)&\leq\frac{1}{\sqrt{2}}\mathbb{E}\left[\mathbb{1}\left\{\boldsymbol{\Delta}>0\right\}\sqrt{\boldsymbol{\Delta}-\overline{m}_\cap(\boldsymbol{\sigma}^+)\frac{\boldsymbol{\Delta}}{\overline{m}_n}}\right]\\&=\frac{1}{\sqrt{2}}\mathbb{E}\left[\mathbb{1}\left\{\boldsymbol{\Delta}>0\right\}\sqrt{\boldsymbol{\Delta}-\left(1-\frac{\boldsymbol{\Delta}}{\overline{m}_n}\right)\boldsymbol{\Delta}}\right]=\frac{\mathbb{E}[\mathbb{1}\{\boldsymbol{\Delta}>0\}\boldsymbol{\Delta}]}{\sqrt{2\overline{m}_n}}.\end{aligned}$$

With $\tilde{c}$ from Corollary 3.18b) we have $\boldsymbol{\Delta}\leq\tilde{c}\bar{d}-\frac{\bar{d}}{k}$, hence

$$\mathbb{P}(\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)\neq\boldsymbol{m}_n)\leq\left(\tilde{c}-\frac{1}{k}\right)\frac{\sqrt{k\bar{d}}}{\sqrt{2\bar{d}n}}\leq\left(\tilde{c}-\frac{1}{k}\right)\frac{\sqrt{kd_\uparrow}}{\sqrt{2n}}.$$

This completes the proof since $\Theta^{\downarrow}/n=n^{-\rho}=\omega(n^{-1/4})$. □

3.3.3.6 Factor Count Asymptotics. Let $c_{\mathrm{r},\mathfrak{g}} \in \mathbb{R}_{>0}$ be large, $r(n) = c_{\mathrm{r}}\sqrt{\ln(n)/n}$ and $\mathcal{B}_+^{\Gamma} = \{\sigma^+ \in [q]^{n+1} : \|\gamma_{\mathrm{n},\sigma^+} - \gamma^*\|_{\mathrm{tv}} \le r(n)\}$. In this section we show that for typical spins $\mathcal{B}_+^{\Gamma}$ and for sufficiently large n the coupling of the graphs simplifies.

Lemma 3.76. *Let $\sigma \in \mathcal{B}_+^{\Gamma}$. There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $\overline{m}_{\Delta}^{\pm}(\sigma) \le c$. Further, there exists $n_{\circ,\mathfrak{g}} \in \mathbb{Z}_{>0}$ such that for $n \ge n_\circ$ we have $\overline{m}_{\Delta-}^{+}(\sigma) = 0$ and*

$$\left|\overline{m}_{\Delta+}^{+}(\sigma) - \bar{d}\right|, \left|\overline{m}_{\Delta}^{-}(\sigma) - \frac{\bar{d}(k-1)}{k}\right| \le cr(n).$$

Proof. With $\tilde{c}$ from Corollary 3.18c) and with $\boldsymbol{d}^*_{\mathrm{f},\boldsymbol{m}} \sim \mathrm{Po}(p_{\mathrm{d}}\overline{m})$ from Corollary 3.18b) we have $|\overline{m}_{\Delta+}^{+} - \bar{d}| \le \tilde{c}\bar{d}(\|\gamma_{\mathrm{n},\sigma^+} - \gamma^*\|_{\mathrm{tv}} + n^{-1}) \le \tilde{c}\bar{d}(r(n) + n^{-1})$. Using $k \ge 2$ fix

$$n_\circ = \min\left\{n_0 \in \mathbb{Z}_{\ge 3} : \sup_{n \ge n_0} \tilde{c}\left(r(n) + \frac{1}{n}\right) < \frac{k-1}{k}\right\}.$$

For $n \le n^\circ$ we have $\overline{m}_{\Delta}^{\pm} \le \overline{m}^{+} \le d_{\uparrow}(n_\circ + 1)/k$. For $n \ge n_\circ$ we have $|\overline{m}_{\Delta+}^{+} - \bar{d}| < \bar{d}(k-1)/k$ and $|\overline{m}_{\Delta+}^{+} - \bar{d}| \le 2\tilde{c}d_{\uparrow}r(n) \le 2\tilde{c}d_{\uparrow}c_{\mathrm{r}}/\sqrt{e}$. The former yields $\overline{m}_{-}^{+} = \overline{m}^{+} - \overline{m}_{\Delta+}^{+} < \overline{m}^{+} - \frac{\bar{d}}{k} = \overline{m}^{-}$, so $\overline{m}_{\Delta-}^{+} = 0$. Finally, notice that

$$\left|\overline{m}_{\Delta}^{-} - \frac{\bar{d}(k-1)}{k}\right| = \left|\overline{m}^{-} - \overline{m}_{-}^{+} - \frac{\bar{d}(k-1)}{k}\right| = \left|\overline{m}_{\Delta+}^{+} - \bar{d}\right|.$$

□

3.3.3.7 Typical Events for the Factor Contribution. Analogously to the ball $\mathcal{B}_+^{\Gamma}$ on $(n+1)$ variables, let $\mathcal{B}_-^{\Gamma} = \{\sigma \in [q]^n : \|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}} \le r(n)\}$, and $\mathcal{B}^\circ = (\bar{d} - r(n), \bar{d} + r(n))$ for the average degree. Further, let $\Phi_{\mathrm{f}}(n) = \mathbb{E}[\boldsymbol{\Phi}]$ with $\boldsymbol{\Phi} = \ln(Z_{\mathrm{g}}(\boldsymbol{G}^-(\boldsymbol{\sigma}^-))/Z_{\mathrm{g}}(\boldsymbol{G}_\cap))$ and $\bar{\boldsymbol{d}}^- = k\boldsymbol{m}^-/n$.

Lemma 3.77. *We have $\Phi_{\mathrm{f}}(n) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \in \mathcal{B}_+^{\Gamma}, \boldsymbol{\sigma}^- \in \mathcal{B}_-^{\Gamma}, \bar{\boldsymbol{d}}^- \in \mathcal{B}^\circ\}\boldsymbol{\Phi}] + o(n^{-1})$.*

Proof. With $\mathcal{E} = \{\boldsymbol{\sigma}^+ \in \mathcal{B}_+^{\Gamma}, \boldsymbol{\sigma}^- \in \mathcal{B}_-^{\Gamma}, \bar{\boldsymbol{d}}^- \in \mathcal{B}^\circ\}$ we have $\Phi_{\mathrm{f}}(n) = \mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}] + \varepsilon$ with

$$\varepsilon(n) = \mathbb{E}\left[\mathbb{1}\neg\mathcal{E}\overline{\boldsymbol{\Phi}}\right], \ \overline{\boldsymbol{\Phi}} = \mathbb{E}\left[\boldsymbol{\Phi}\middle|\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+), \boldsymbol{\sigma}^+, \boldsymbol{\sigma}^-, \bar{\boldsymbol{d}}^-\right].$$

With Jensen's inequality we can consider the atypical events separately, i.e.

$$|\varepsilon(n)| \le \mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}^- \notin \mathcal{B}^\circ\}|\overline{\boldsymbol{\Phi}}|] + \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \notin \mathcal{B}_+^{\Gamma}\}|\overline{\boldsymbol{\Phi}}|] + \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^- \notin \mathcal{B}_-^{\Gamma}\}|\overline{\boldsymbol{\Phi}}|].$$

With $\boldsymbol{\Phi} = n\phi_{\mathrm{g}}(Z_{\mathrm{g}}(\boldsymbol{G}^-(\boldsymbol{\sigma}^-))) - n\phi_{\mathrm{g}}(Z_{\mathrm{g}}(\boldsymbol{G}_\cap))$, Jensen's inequality, the triangle inequality, $\tilde{c}$ from Observation 3.30, Lemma 3.73, Lemma 3.72 and Observation 3.47 we have

$$|\overline{\boldsymbol{\Phi}}| \le \tilde{c}\left(\boldsymbol{m}^- + \frac{\Theta_n^{\downarrow}}{2} + \boldsymbol{m}_\cap(\boldsymbol{\sigma}^+) + \frac{n\Theta_n^{\downarrow}}{(n+1)2}\right) \le \tilde{c}(2\boldsymbol{m}^- + n).$$

So, with Lemma 3.70 and c from Corollary 3.12 we have

$$\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}^- \notin \mathcal{B}^\circ\}|\overline{\boldsymbol{\Phi}}|] \le \frac{2\tilde{c}n}{k}\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}} \notin \mathcal{B}^\circ\}\bar{\boldsymbol{d}}] + \tilde{c}n\mathbb{P}(\bar{\boldsymbol{d}} \notin \mathcal{B}^\circ) \le c'n\exp\left(-\frac{c_1 r^2 n}{1+r}\right)$$

with $c' = \tilde{c}c_2(2+k)/k$. With $r = o(1)$ and $c_{\mathrm{r}}^2 > 2/c_1$ we get

$$\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}^- \notin \mathcal{B}^\circ\}|\overline{\boldsymbol{\Phi}}|] \le c'n\exp\left(-(1+o(1))c_1c_{\mathrm{r}}^2\ln(n)\right) = c'n^{-(1+o(1))c_1c_{\mathrm{r}}^2+1} = o(n^{-1}).$$

With c from Observation 3.23b), independence and $c_{\mathrm{r}}^2 > 2/c_1$ we have

$$\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \notin \mathcal{B}_+^\Gamma\}|\overline{\boldsymbol{\Phi}}|] \le \tilde{c}\left(\frac{2d_\uparrow}{k}+1\right)c_2ne^{-c_1r^2(n+1)} = \Theta\left(n^{1-c_1c_{\mathrm{r}}^2}\right) = o(n^{-1}),$$

and $\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^- \notin \mathcal{B}_-^\Gamma\}|\overline{\boldsymbol{\Phi}}|] = o(n^{-1})$ follows analogously. □

The following result further restricts the very typical event in Lemma 3.77 to the typical event that no variables are pinned in the second sweep.

Lemma 3.78. *We have*

$$\Phi_{\mathrm{f}}(n) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \in \mathcal{B}_+^\Gamma, \boldsymbol{\sigma}^- \in \mathcal{B}_-^\Gamma, \bar{\boldsymbol{d}}^- \in \mathcal{B}^\circ, \boldsymbol{\mathcal{U}}^- = \boldsymbol{\mathcal{U}}_\cap\}\boldsymbol{\Phi}] + \mathcal{O}(n^{-\rho}).$$

Proof. With Lemma 3.77 it is sufficient to consider $\mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\Phi}|]$, where

$$\mathcal{E} = \left\{\boldsymbol{\sigma}^+ \in \mathcal{B}_+^\Gamma, \boldsymbol{\sigma}^- \in \mathcal{B}_-^\Gamma, \bar{\boldsymbol{d}}^- \in \mathcal{B}^\circ, \boldsymbol{\mathcal{U}}^- \ne \boldsymbol{\mathcal{U}}_\cap\right\}.$$

With $\tilde{c}$ from Observation 3.30 we have $\mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\Phi}|] \le \tilde{c}\mathbb{E}[\mathbb{1}\mathcal{E}\mathrm{d_g}(\boldsymbol{G}_\cap, \boldsymbol{G}^-(\boldsymbol{\sigma}^-))]$. With the notions in Section 3.2.3.1 we have $\mathcal{V}_1^\downarrow = [n]\setminus\boldsymbol{\mathcal{U}}^-$, $\mathcal{V}_2^\downarrow = \boldsymbol{\mathcal{U}}_\cap$, $\mathcal{V}^\downarrow = [n]\setminus\boldsymbol{\mathcal{U}}_\Delta$ with $\boldsymbol{\mathcal{U}}_\Delta = \boldsymbol{\mathcal{U}}^-\setminus\boldsymbol{\mathcal{U}}_\cap$, further $m_\cap = \boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)$, $\mathcal{A}_= = [\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)]\setminus\boldsymbol{\mathcal{A}}_{\ne}$,

$$\boldsymbol{\mathcal{A}}_{\ne} = \{a \in [\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)] : \boldsymbol{v}_{\cap,a}([k]) \cap \boldsymbol{\mathcal{U}}_\Delta \ne \emptyset\},$$

where $\boldsymbol{v}_\cap$ are the neighborhoods of $\boldsymbol{G}_\cap$, so $D = 0$, $\tilde{D} = \boldsymbol{m}^- - \boldsymbol{m}_\cap(\boldsymbol{\sigma}^+) = \boldsymbol{m}_\Delta^-(\boldsymbol{\sigma}^+)$, $D_\cap = |\boldsymbol{\mathcal{A}}_{\ne}|$ and hence $\mathrm{d_g}(\boldsymbol{G}_\cap, \boldsymbol{G}^-(\boldsymbol{\sigma}^-)) = \boldsymbol{m}_\Delta^-(\boldsymbol{\sigma}^+) + 2|\boldsymbol{\mathcal{A}}_{\ne}| + |\boldsymbol{\mathcal{U}}_\Delta|$. Recall that

$$|\boldsymbol{\mathcal{A}}_{\ne}| \le \sum_{i\in\boldsymbol{\mathcal{U}}_\Delta} d_{\mathrm{f},\boldsymbol{w}_\cap}(i) \le \sum_{i\in\boldsymbol{\mathcal{U}}_\Delta} d_{\mathrm{f},\boldsymbol{w}^-}(i)$$

with $\boldsymbol{w}_\cap = \boldsymbol{w}_{\cap,\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+)}(\boldsymbol{\sigma}^-)$, $\boldsymbol{w}^- = \boldsymbol{w}_{\boldsymbol{m}^-}^-(\boldsymbol{\sigma}^-)$ and $d_{\mathrm{f},G}(i)$ from Section 3.2.1.6. With c from Corollary 3.18a) this gives

$$\mathbb{E}\left[\mathrm{d_g}(\boldsymbol{G}_\cap, \boldsymbol{G}^-(\boldsymbol{\sigma}^-))\middle|\boldsymbol{\sigma}^+, \boldsymbol{m}_\cap(\boldsymbol{\sigma}^+), \boldsymbol{m}^-, \boldsymbol{\mathcal{U}}_\cap, \boldsymbol{\mathcal{U}}^-\right] \le \boldsymbol{m}_\Delta^-(\boldsymbol{\sigma}^+) + 2|\boldsymbol{\mathcal{U}}_\Delta|c\bar{\boldsymbol{d}}^- + |\boldsymbol{\mathcal{U}}_\Delta|.$$

On $\mathcal{E}$ we further have $\bar{\boldsymbol{d}} \le d_\uparrow + r(n)$. With this bound, standard bounds, and taking conditional expectations we obtain $\mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\Phi}|] \le \tilde{c}E_1 + \tilde{c}(2c(d_\uparrow + r(n)) + 1)E_2$, where

$$E_1 = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \in \mathcal{B}_+^\Gamma, \boldsymbol{\mathcal{U}}^- \ne \boldsymbol{\mathcal{U}}_\cap\}\overline{m}_\Delta^-(\boldsymbol{\sigma}^+)],\ E_2 = \mathbb{E}[|\boldsymbol{\mathcal{U}}_\Delta|].$$

With c' from Lemma 3.76 we have $\overline{m}_\Delta^-(\boldsymbol{\sigma}^+) \le c'$, so as in the proof of Lemma 3.75 we have

$$E_1 \le c'\mathbb{P}(\boldsymbol{\mathcal{U}}^- \ne \boldsymbol{\mathcal{U}}_\cap) \le (1+o(1))c'\frac{\Theta_n^\downarrow}{n},\ E_2 \le (1+o(1))\frac{\Theta_n^\downarrow}{n}.$$

This completes the proof since $\Theta^{\downarrow}/n = n^{-\rho} = \omega(n^{-1})$. □

3.3.3.8 Normalization Step for the Factor Contribution. In Section 3.3.3.7 we restricted the expectation over the coupled graphs to typical events, now we change the underlying law. In particular, we replace $\boldsymbol{G}_\cap$ by $\boldsymbol{G}^*(\sigma) = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\sigma)$ and $\boldsymbol{m}^-_\Delta(\boldsymbol{\sigma}^+)$ by a Poisson variable $\boldsymbol{m}_\Delta \sim \mathrm{Po}(\bar{d}(k-1)/k)$. Clearly, we obtain the additional wires-weight pairs given $\boldsymbol{\sigma}^*$ from Observation 3.13, i.e. we consider

$$(\boldsymbol{G}^*(\sigma), \boldsymbol{m}_\Delta, \boldsymbol{w}^*(\sigma)) \sim \boldsymbol{G}^*(\sigma) \otimes \boldsymbol{m}_\Delta \otimes \boldsymbol{w}^{*\otimes \mathbb{Z}_{>0}}_{\circ,\sigma}. \tag{11}$$

Further, let $\boldsymbol{\Phi} = \ln(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}(\boldsymbol{w}^*_{\boldsymbol{\sigma}^*,[\boldsymbol{m}_\Delta]}))$ with $\overline{\psi}_{\mathrm{w}|\mathrm{g}}$ from Equation (8).

Lemma 3.79. *We have* $\Phi_{\mathrm{f}}(n) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}^\Gamma_-\}\boldsymbol{\Phi}] + \mathcal{O}(n^{-\rho})$.

Proof. Let $\boldsymbol{\Phi'} = \ln(Z_{\mathrm{g}}(\boldsymbol{G}^-(\boldsymbol{\sigma}^-))/Z_{\mathrm{g}}(\boldsymbol{G}_\cap))$, and let $(\boldsymbol{G}_\cap, \boldsymbol{m}_\Delta, \boldsymbol{w}^*(\boldsymbol{\sigma}^-))$ be conditionally independent given $\boldsymbol{\sigma}^+$. As explained in Section 3.3.3.4 and analogously to Section 3.3.2.3 on

$$\mathcal{E} = \left\{\boldsymbol{\sigma}^+ \in \mathcal{B}^\Gamma_+, \boldsymbol{\sigma}^- \in \mathcal{B}^\Gamma_-, \bar{\boldsymbol{d}}^- \in \mathcal{B}^\circ, \boldsymbol{\mathcal{U}}^- = \boldsymbol{\mathcal{U}}_\cap\right\}$$

we have $\boldsymbol{\Phi'} \sim \ln(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}_\cap}(\boldsymbol{w}^*_{\boldsymbol{\sigma}^-,[\boldsymbol{m}^-_\Delta(\boldsymbol{\sigma}^+)]}))$, i.e. there are no additional pins, the additional factors are independent of the remainder and iid from the teacher-student model for the given ground truth. For given $\sigma^+ \in [q]^{n+1}$ let $\boldsymbol{\delta}(\sigma^+) \sim \mathrm{Po}(|\overline{m}^-_\Delta(\sigma^+) - \overline{m}_\Delta|)$ with $\overline{m}_\Delta = \frac{\bar{d}(k-1)}{k}$. For $\overline{m}^-_\Delta(\sigma^+) \geq \overline{m}_\Delta$ and using Observation 3.7b) we consider the coupling $\boldsymbol{m}^-_\Delta(\sigma^+) = \boldsymbol{m}_\Delta + \boldsymbol{\delta}(\sigma^+)$, and $\boldsymbol{m}_\Delta = \boldsymbol{m}^-_\Delta(\sigma^+) + \boldsymbol{\delta}(\sigma^+)$ otherwise. This gives

$$\left|\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}_\cap}\left(\boldsymbol{w}^*_{\boldsymbol{\sigma}^-,[\boldsymbol{m}^-_\Delta(\boldsymbol{\sigma}^+)]}\right)\right) - \ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}_\cap}\left(\boldsymbol{w}^*_{\boldsymbol{\sigma}^-,[\boldsymbol{m}_\Delta]}\right)\right)\right| \leq \boldsymbol{\delta}(\boldsymbol{\sigma}^+)\ln(\psi_\uparrow).$$

With Lemma 3.76 we can bound $\mathbb{E}[\boldsymbol{\delta}(\sigma^+)]$ on $\mathcal{E}$, so with Lemma 3.78 we have

$$\Phi_{\mathrm{f}}(n) = \mathbb{E}\left[\mathbb{1}\mathcal{E}\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}_\cap}\left(\boldsymbol{w}^*_{\boldsymbol{\sigma}^-,[\boldsymbol{m}_\Delta]}\right)\right)\right] + \mathcal{O}\left(r(n) + n^{-\rho}\right).$$

Due to the independence of $\boldsymbol{m}_\Delta$ from the remainder we can use the upper bound $\boldsymbol{m}_\Delta \ln(\psi)$ on the argument of the expectation and then take the expectation with respect to $\boldsymbol{m}_\Delta$ to obtain the upper bound $c = \overline{m}_\Delta \ln(\psi_\uparrow) \leq d_\uparrow \ln(\psi_\uparrow)(k-1)/k$ given the rest. This shows that reducing $\mathcal{E}$ to $\{\boldsymbol{\sigma}^- \in \mathcal{B}^\Gamma_-\}$ causes an error of $\mathcal{O}(n^{-\rho})$. Also, with the coupling from Lemma 3.75 we then get

$$\Phi_{\mathrm{f}}(n) = \mathbb{E}\left[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}^\Gamma_-\}\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}\left(\boldsymbol{w}^*_{\boldsymbol{\sigma}^*,[\boldsymbol{m}_\Delta]}\right)\right)\right] + \mathcal{O}\left(n^{-\rho}\right)$$

since $n^{-\rho} = \omega(r(n))$, which completes the proof. □

3.3.3.9 Gibbs Marginal Product for the Factor Contribution. Now, it is time to apply Proposition 3.48. Using the distribution (11) let

$$\boldsymbol{\Phi} = \ln\left(\sum_\tau\left(\bigotimes_{(a,h)\in[\boldsymbol{m}_\Delta]\times[k]}\gamma_{a,h}\right)(\tau)\prod_{a\in[\boldsymbol{m}_\Delta]}\boldsymbol{\psi}^*_a(\tau_a)\right) = \sum_{a\in[\boldsymbol{m}_\Delta]}\ln\left(Z_{\mathrm{f}}(\boldsymbol{\psi}^*_a, \boldsymbol{\gamma}_a)\right),$$

where $\boldsymbol{\gamma} = (\mu_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}|_{\boldsymbol{v}^*(a,h)})_{a,h}$, $(\boldsymbol{v}^*, \boldsymbol{\psi}^*) = \boldsymbol{w}^*(\boldsymbol{\sigma}^*)$, and Z_{f} from the Bethe functional.

Lemma 3.80. *We have* $\Phi_{\mathrm{f}}(n) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}^\Gamma_-\}\boldsymbol{\Phi}] + \mathcal{O}(n^{-\rho})$.

Proof. Resolving the Radon-Nikodym derivative of the additional pairs in Lemma 3.79 yields $\Phi_{\mathrm{f}}(n) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}_-^\Gamma\}\boldsymbol{\Phi}^*] + \mathcal{O}(n^{-\rho})$ with $\boldsymbol{w} = (\boldsymbol{v}, \boldsymbol{\psi}) \sim (\mathrm{u}([n]^k) \otimes \mu_\Psi)^{\otimes \mathbb{Z}_{>0}}$ and

$$\boldsymbol{\Phi}^* = \prod_{a \in [\boldsymbol{m}_\Delta]} \frac{\boldsymbol{\psi}_a(\boldsymbol{\sigma}^*_{\boldsymbol{v}(a)})}{\overline{Z}_{\mathrm{f}}(\boldsymbol{\gamma}^*)} \ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}(\boldsymbol{w}_{[\boldsymbol{m}_\Delta]})\right).$$

Recall that for $\boldsymbol{v}' = \boldsymbol{v}_{[\boldsymbol{m}_\Delta]}$ and $\boldsymbol{\alpha}^* = \boldsymbol{\mu}^*|_{\boldsymbol{v}'}$ with $\boldsymbol{\mu}^* = \mu_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ we have

$$\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}(\boldsymbol{w}_{[\boldsymbol{m}_\Delta]}) = \sum_{\tau \in ([q]^k)^{\boldsymbol{m}_\Delta}} \boldsymbol{\alpha}^*(\tau) \prod_{a \in [\boldsymbol{m}_\Delta]} \boldsymbol{\psi}_a(\tau_a).$$

Let C from Proposition 3.48b), $\varepsilon = C_1/3$ and $\delta = \Theta^{\downarrow -2\varepsilon}$. Using $\iota_\circ$ from Section 3.3.1.5 let $\mathcal{E} = \{\boldsymbol{\sigma}^* \in \mathcal{B}_-^\Gamma, \iota_\circ(\boldsymbol{\mu}^*, \boldsymbol{v}') \le \delta\}$ and notice that $\iota_\circ(\boldsymbol{\mu}^*, \boldsymbol{v}') = 0 < \delta$ on $\boldsymbol{m}_\Delta = 0$. Hence, the bound $|\boldsymbol{\Phi}^*| \le \ln(\psi_\uparrow^{\boldsymbol{m}_\Delta})\psi_\uparrow^{2\boldsymbol{m}_\Delta} \le \psi_\uparrow^{3\boldsymbol{m}_\Delta}$ and Markov's inequality conditional to $\boldsymbol{m}_\Delta$ give $\Delta \le \varepsilon' + \mathcal{O}(n^{-\rho})$, where $\Delta = |\Phi_{\mathrm{f}} - \mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}^*]|$ and

$$\varepsilon' = \mathbb{E}\left[\mathbb{1}\{\boldsymbol{m}_\Delta > 0\}\frac{C_2(k\boldsymbol{m}_\Delta - 1)}{\delta}\left(\frac{k\boldsymbol{m}_\Delta}{\Theta^\downarrow}\right)^{C_1}\psi_\uparrow^{3\boldsymbol{m}_\Delta}\right].$$

Standard bounds imply $\varepsilon' \le \tilde{c}\mathbb{E}[\exp(\tilde{c}\boldsymbol{m}_\Delta)]/\Theta^{\downarrow\varepsilon}$ for some $\tilde{c} \in \mathbb{R}_{>0}$. The canonical coupling of $\boldsymbol{m}_\Delta \sim \mathrm{Po}(\bar{d}(k-1)/k)$ and $\boldsymbol{m}_{\Delta\uparrow} \sim \mathrm{Po}(d_\uparrow(k-1)/k)$ from Observation 3.7b) gives $\boldsymbol{m}_\Delta \le \boldsymbol{m}_{\Delta\uparrow}$ and hence $\varepsilon' \le \tilde{c}\mathbb{E}[\exp(\tilde{c}\boldsymbol{m}_{\Delta\uparrow})]/\Theta^{\downarrow\varepsilon}$. Finally, the moment generating function of the Poisson distribution gives $\varepsilon' = \mathcal{O}(\Theta^{\downarrow -\varepsilon})$. But $\rho = \varepsilon/(1+\varepsilon)$ and $\Theta^\downarrow = n^{1-\rho}$ yields $\Theta^{\downarrow\varepsilon} = n^\rho$, thereby $\delta = n^{-2\rho}$ and $\varepsilon', \Delta = \mathcal{O}(n^{-\rho})$. Now, with $\boldsymbol{\alpha} = \bigotimes_{(a,h) \in [\boldsymbol{m}_\Delta] \times [k]} \boldsymbol{\mu}^*|_{\boldsymbol{v}'(a,h)}$ and

$$\boldsymbol{\Phi} = \prod_{a \in [\boldsymbol{m}_\Delta]} \frac{\boldsymbol{\psi}_a(\boldsymbol{\sigma}^*_{\boldsymbol{v}(a)})}{\overline{Z}_{\mathrm{f}}(\boldsymbol{\gamma}^*)} \ln\left(\sum_{\tau \in ([q]^k)^{\boldsymbol{m}_\Delta}} \boldsymbol{\alpha}(\tau) \prod_{a \in [\boldsymbol{m}_\Delta]} \boldsymbol{\psi}_a(\tau_a)\right),$$

notice that the arguments of the logarithm for both $\boldsymbol{\Phi}^*$ and $\boldsymbol{\Phi}$ are in $[\psi_\downarrow^{\boldsymbol{m}_\Delta}, \psi_\uparrow^{\boldsymbol{m}_\Delta}]$ and that the logarithm is $\psi_\uparrow^{\boldsymbol{m}_\Delta}$-Lipschitz on this domain, so

$$|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}| \le \psi_\uparrow^{3\boldsymbol{m}_\Delta}\left|\sum_{\tau \in ([q]^k)^{\boldsymbol{m}_\Delta}} \boldsymbol{\alpha}^*(\tau) \prod_{a \in [\boldsymbol{m}_\Delta]} \boldsymbol{\psi}_a(\tau_a) - \sum_{\tau \in ([q]^k)^{\boldsymbol{m}_\Delta}} \boldsymbol{\alpha}(\tau) \prod_{a \in [\boldsymbol{m}_\Delta]} \boldsymbol{\psi}_a(\tau_a)\right|.$$

This yields $|\boldsymbol{\Phi}^* - \boldsymbol{\Phi}| \le 2\psi_\uparrow^{4\boldsymbol{m}_\Delta}\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}\|_{\mathrm{tv}} = 2\psi_\uparrow^{4\boldsymbol{m}_\Delta}\nu_\circ(\boldsymbol{\mu}^*, \boldsymbol{v}') \le \sqrt{2}\psi_\uparrow^{4\boldsymbol{m}_\Delta}\sqrt{\iota_\circ(\boldsymbol{\mu}^*, \boldsymbol{v}')}$ with standard bounds and Remark 3.49. Since we have the same bound $|\boldsymbol{\Phi}| \le \psi_\uparrow^{3\boldsymbol{m}_\Delta}$, we can spare another ε' from above to obtain

$$\Phi_{\mathrm{f}}(n) = \mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}] + \mathcal{O}\left(\sqrt{\delta} + n^{-\rho}\right) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}_-^\Gamma\}\boldsymbol{\Phi}] + \mathcal{O}\left(n^{-\rho}\right).$$

The assertion follows by reintroducing $\boldsymbol{w}^*$ using the Radon-Nikodym derivative in $\boldsymbol{\Phi}$. □

With $(\boldsymbol{\psi}^*_a, \boldsymbol{\gamma}_a)_a$ being iid given $\boldsymbol{\sigma}^*$ and $\mathbb{E}[\overline{m}_\Delta] = \bar{d}(k-1)/k$, Lemma 3.80 yields

$$\Phi_{\mathrm{f}}(n) = \frac{\bar{d}(k-1)}{k}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}^{\Gamma}_{-}\}\ln(Z_{\mathrm{f}}(\boldsymbol{\psi}^*, \boldsymbol{\gamma}))\right] + \mathcal{O}\left(n^{-\rho}\right),$$

with $\boldsymbol{w}^* = (\boldsymbol{v}^*, \boldsymbol{\psi}^*) \sim \boldsymbol{w}^*_{\circ,\boldsymbol{\sigma}^*}$, $\boldsymbol{\gamma} = (\mu_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}|_{\boldsymbol{v}^*(h)})_{h\in[k]}$ and $(\boldsymbol{w}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$ conditionally independent given $\boldsymbol{\sigma}^*$.

3.3.3.10 Marginal Distribution for the Factor Contribution. Now, we work towards the discussion in Section 3.3.1.8. Let $\gamma = \gamma_{\mathrm{n},\sigma}$, $\boldsymbol{G}^*(\sigma) = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\sigma)$ and $\boldsymbol{w}^\star_\circ$, $\boldsymbol{\tau}^*_\circ$, $\mathcal{D}_\Gamma$ from Section 3.2.1.5. Let $(\boldsymbol{v}^\star_{\sigma,\tau}, \boldsymbol{\psi}^\star_\tau, \boldsymbol{\gamma}_{\sigma,\tau}) \sim \boldsymbol{w}^\star_{\circ,\sigma,\tau} \otimes \bigotimes_{h\in[k]} \check{\pi}_{\mathrm{g},\boldsymbol{G}^*(\sigma),\sigma,\tau(h)}$ for $\tau \in \mathcal{D}^k_{\Gamma,\gamma}$, further $(\boldsymbol{\tau}^*(\sigma), \boldsymbol{G}^*(\sigma)) \sim \boldsymbol{\tau}^*_{\circ,\sigma} \otimes \boldsymbol{G}^*(\sigma)$ and $\boldsymbol{\tau}^* = \boldsymbol{\tau}^*(\boldsymbol{\sigma}^*)$.

Lemma 3.81. *We have* $\Phi_{\mathrm{f}}(n) = \frac{\bar{d}(k-1)}{k}\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}^{\Gamma}_{-}\}\ln(Z_{\mathrm{f}}(\boldsymbol{\psi}^\star_{\boldsymbol{\tau}^*}, \boldsymbol{\gamma}_{\boldsymbol{\sigma}^*,\boldsymbol{\tau}^*}))] + \mathcal{O}(n^{-\rho})$.

Proof. With Lemma 3.80, Observation 3.15, independence and $\boldsymbol{\mu}^*_\sigma = \mu_{\mathrm{g},\boldsymbol{G}^*(\sigma)}$ we have

$$\Phi_{\mathrm{f}}(n) = \frac{\bar{d}(k-1)}{k}\mathbb{E}\left[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}^{\Gamma}_{-}\}E(\boldsymbol{\sigma}^*, \boldsymbol{\tau}^*)\right] + \mathcal{O}(n^{-\rho}),$$
$$E(\sigma, \tau) = \mathbb{E}\left[\ln\left(Z_{\mathrm{f}}\left(\boldsymbol{\psi}^\star_\tau, \left(\boldsymbol{\mu}^*_\sigma|_{\boldsymbol{v}^\star_{\sigma,\tau}(h)}\right)_{h\in[k]}\right)\right)\right], \tau \in \mathcal{D}^k_{\Gamma,\gamma}, \gamma = \gamma_{\mathrm{n},\sigma}.$$

Next, we use independence, expand the definition of $\boldsymbol{v}^\star_{\sigma,\tau}$ and obtain

$$E(\sigma, \tau) = \mathbb{E}\left[\sum_v \prod_{h\in[k]} \frac{\mathbb{1}\{\sigma_{v(h)} = \tau_h\}}{|\sigma^{-1}(\tau_h)|}\ln\left(Z_{\mathrm{f}}\left(\boldsymbol{\psi}^\star_\tau, \left(\boldsymbol{\mu}^*_\sigma|_{v(h)}\right)_{h\in[k]}\right)\right)\right].$$

The definition of $\check{\pi}_{\mathrm{g},G,\sigma,\tau(h)}$ completes the proof. □

Now, we can combine Lemma 3.81 with Corollary 3.54. Hence, we introduce the reweighted marginals $(\boldsymbol{v}^\star_{\sigma,\tau}, \boldsymbol{\psi}^\star_\tau, \hat{\boldsymbol{\gamma}}_{\sigma,\tau}) \sim \boldsymbol{w}^\star_{\circ,\sigma,\tau} \otimes \bigotimes_{h\in[k]} \hat{\pi}_{\mathrm{g},\boldsymbol{G}^*(\sigma),\tau(h)}$.

Lemma 3.82. *We have* $\Phi_{\mathrm{f}}(n) = \frac{\bar{d}(k-1)}{k}\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}^{\Gamma}_{-}\}\ln(Z_{\mathrm{f}}(\boldsymbol{\psi}^\star_{\boldsymbol{\tau}^*}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{\sigma}^*,\boldsymbol{\tau}^*}))] + \mathcal{O}(n^{-\rho})$.

Proof. Fix σ, τ, $G = [w]^{\Gamma\downarrow}_{\mathcal{U},\sigma}$ with $w \in \mathcal{G}$ and $\psi \in \mathcal{D}_\Psi$. Let $\check{E}(\sigma, \tau, G, \psi) = \mathbb{E}[\ln(Z_{\mathrm{f}}(\psi, \boldsymbol{\gamma}))]$, where $\boldsymbol{\gamma} \sim \bigotimes_h \check{\pi}_{\mathrm{g},G,\sigma,\tau(h)}$, and $\hat{E}(\sigma, \tau, G, \psi) = \mathbb{E}[\ln(Z_{\mathrm{f}}(\psi, \boldsymbol{\gamma}))]$, where $\boldsymbol{\gamma} \sim \bigotimes_h \hat{\pi}_{\mathrm{g},G,\tau(h)}$. Let $\pi_h \in \Gamma(\check{\pi}_{\mathrm{g},G,\sigma,\tau(h)}, \hat{\pi}_{\mathrm{g},G,\tau(h)})$ be a coupling for $h \in [k]$ and $(\check{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}) \sim \bigotimes_h \pi_h$ with $\check{\gamma}, \hat{\gamma} \in \mathcal{P}([q])^k$. We have $Z_{\mathrm{f}}(\psi, \check{\gamma}), Z_{\mathrm{f}}(\psi, \hat{\gamma}) \in [\psi_\downarrow, \psi_\uparrow]$, so the logarithm is $\psi_\uparrow$-Lipschitz on this domain, and thereby using Observation 3.6b) we obtain

$$\Delta(\sigma, \tau, G, \psi) = \left|\check{E} - \hat{E}\right| \le \psi_\uparrow\mathbb{E}\left[\sum_{\tau'}\psi(\tau')\left|\prod_h \check{\boldsymbol{\gamma}}_h(\tau'_h) - \prod_h \hat{\boldsymbol{\gamma}}_h(\tau'_h)\right|\right]$$
$$\le 2\psi^2_\uparrow\sum_h \|\check{\boldsymbol{\gamma}}_h - \hat{\boldsymbol{\gamma}}_h\|_{\mathrm{tv}}.$$

Since this holds for any choice of coupling we have $\Delta \le 2\psi^2_\uparrow\sum_h \mathrm{d}_{\mathrm{w}}(\check{\pi}_{\mathrm{g},G,\sigma,\tau(h)}, \hat{\pi}_{\mathrm{g},G,\tau(h)})$. With $\tau \in \sigma([n])^k$ and $D(\sigma, \mu)$ from Corollary 3.54 this yields $\Delta \le 2k\psi^2_\uparrow D(\sigma, \mu_{\mathrm{g},G})$. Hence, taking the expectation and using c, C_1 from Corollary 3.54b) with $\boldsymbol{\gamma}_{\sigma,\tau}$ from Lemma 3.81 gives

$$\begin{aligned}\Delta &= \left|\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}_-^\Gamma\} \ln(Z_\mathrm{f}(\boldsymbol{\psi}^\star_{\boldsymbol{\tau}^*}, \boldsymbol{\gamma}_{\boldsymbol{\sigma}^*,\boldsymbol{\tau}^*}))] - \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^* \in \mathcal{B}_-^\Gamma\} \ln(Z_\mathrm{f}(\boldsymbol{\psi}^\star_{\boldsymbol{\tau}^*}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{\sigma}^*,\boldsymbol{\tau}^*}))]\right| \\ &\le 2k\psi_\uparrow^2 \mathbb{E}\left[D(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))\right] \le 2k\psi_\uparrow\left(\frac{c}{\Theta^{\downarrow C_1}} + q\mathbb{P}(\boldsymbol{m} > m_\uparrow)\right).\end{aligned}$$

Recall that $\mathbb{P}(\boldsymbol{m} > m_\uparrow) = o(1/n)$, $\Theta^\downarrow = n^{1-\rho}$ with $\rho = c/(1+c)$, $c = C_1/3$, and notice that $(1-\rho)C_1 = 3\rho > \rho$, so $\Delta = o(n^{-\rho})$ and hence the assertion holds with Lemma 3.81. □

3.3.3.11 The Factor Contribution. Now, we complete the discussion of Φ_f. First, we resolve the reweighting, and then project onto $\mathcal{P}^2_*([q])$. Let $(\boldsymbol{\psi}, \boldsymbol{\gamma}) \sim \mu_\Psi \otimes \pi^{\otimes k}_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ with $\boldsymbol{G}^*(\boldsymbol{\sigma}^*) = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*)$.

Lemma 3.83. *We have* $\Phi_\mathrm{f}(n) = \frac{\bar{d}(k-1)}{k\xi}\mathbb{E}[\Lambda(Z_\mathrm{f}(\boldsymbol{\psi}, \boldsymbol{\gamma}))] + \mathcal{O}(n^{-\rho})$.

Proof. Let $\boldsymbol{\gamma}^* = \gamma_{\mathrm{n},\boldsymbol{\sigma}^*}$ and $\overline{\boldsymbol{\gamma}} = \overline{\gamma}_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$. With c from Lemma 3.52a) we have

$$\mathbb{P}(\|\overline{\boldsymbol{\gamma}} - \gamma^*\|_\mathrm{tv} \ge r) \le c_2 e^{-c_1 c_\mathrm{r} \ln(n)} + \mathbb{P}(\boldsymbol{m} > m_\uparrow) = o(1/n)$$

since $c_\mathrm{r} > 1/c_1$ is large. Lemma 3.82, using that the argument to the logarithm is in $[\psi_\downarrow, \psi_\uparrow]$ (and the leading coefficient in $[0, d_\uparrow]$), with $\mathcal{E} = \{\boldsymbol{\sigma}^* \in \mathcal{B}_-^\Gamma, \|\overline{\boldsymbol{\gamma}} - \gamma^*\|_\mathrm{tv} \le r\}$ yields

$$\Phi_\mathrm{f}(n) = \frac{\bar{d}(k-1)}{k}\mathbb{E}[\mathbb{1}\mathcal{E}\ln(Z_\mathrm{f}(\boldsymbol{\psi}^\star_{\boldsymbol{\tau}^*}, \hat{\boldsymbol{\gamma}}_{\boldsymbol{\sigma}^*,\boldsymbol{\tau}^*}))] + \mathcal{O}(n^{-\rho}).$$

Resolving the Radon-Nikodym derivatives gives $\Phi_\mathrm{f}(n) = \frac{\bar{d}(k-1)}{k}\mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}] + \mathcal{O}(n^{-\rho})$ with

$$\begin{aligned}\boldsymbol{\Phi} &= \sum_\tau \frac{\overline{\psi}_\circ(\tau)\prod_h \boldsymbol{\gamma}^*(\tau_h)}{\overline{Z}_\mathrm{f}(\boldsymbol{\gamma}^*)} \cdot \frac{\boldsymbol{\psi}(\tau)}{\overline{\psi}_\circ(\tau)} \cdot \prod_h \frac{\boldsymbol{\gamma}_h(\tau_h)}{\overline{\boldsymbol{\gamma}}(\tau_h)} \ln(Z_\mathrm{f}(\boldsymbol{\psi}, \boldsymbol{\gamma})) \\ &= \sum_\tau \frac{\boldsymbol{\psi}(\tau)\prod_h \boldsymbol{\gamma}_h(\tau_h)\prod_h \boldsymbol{\gamma}^*(\tau_h)}{\overline{Z}_\mathrm{f}(\boldsymbol{\gamma}^*)\prod_h \overline{\boldsymbol{\gamma}}(\tau_h)} \ln(Z_\mathrm{f}(\boldsymbol{\psi}, \boldsymbol{\gamma})).\end{aligned}$$

On $\mathcal{E}$ we have $\boldsymbol{\gamma}^*(\tau_h)/\gamma^*(\tau_h) \le 1 + \psi_\uparrow\|\boldsymbol{\gamma}^* - \gamma^*\|_\infty$, which with the corresponding lower bound yields $\boldsymbol{\gamma}^*(\tau_h)/\gamma^*(\tau_h) = 1 + \mathcal{O}(r)$. With Observation 3.9d) and Observation 3.9a) we further have $\overline{Z}_\mathrm{f}(\boldsymbol{\gamma}^*)/\xi = 1 + \mathcal{O}(r^2)$. Analogously to $\boldsymbol{\gamma}^*$ we get $\overline{\boldsymbol{\gamma}}(\tau_h)/\gamma^*(\tau_h) = 1 + \mathcal{O}(r)$, so

$$\boldsymbol{\Phi} = (1 + \mathcal{O}(r))\sum_\tau \frac{\boldsymbol{\psi}(\tau)\prod_h \boldsymbol{\gamma}_h(\tau_h)\prod_h \gamma^*(\tau_h)}{\xi\prod_h \gamma^*(\tau_h)} \ln(Z_\mathrm{f}(\boldsymbol{\psi}, \boldsymbol{\gamma})) = (1 + \mathcal{O}(r))\frac{\Lambda(Z_\mathrm{f}(\boldsymbol{\psi}, \boldsymbol{\gamma}))}{\xi}.$$

With $|\boldsymbol{\Phi}| \le \psi_\uparrow^2 \ln(\psi_\uparrow)$ and $r = o(n^{-\rho})$ we have $\Phi_\mathrm{f}(n) = \frac{\bar{d}(k-1)}{k\xi}\mathbb{E}[\mathbb{1}\mathcal{E}\Lambda(Z_\mathrm{f}(\boldsymbol{\psi}, \boldsymbol{\gamma}))] + \mathcal{O}(n^{-\rho})$. Since the argument is still uniformly bounded, resolving $\mathcal{E}$ comes at a cost $o(1/n)$. □

Next, we show that we can replace $\pi_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ by its projection $\pi^\circ_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ by using Lemma 3.56. For this purpose we show that the factor contribution

$$B_\mathrm{f} : \mathcal{P}^2([q]) \to \mathbb{R},\ \pi \mapsto \frac{\bar{d}(k-1)}{k\xi}\mathbb{E}\left[\Lambda\left(Z_\mathrm{f}(\boldsymbol{\psi}, \boldsymbol{\gamma}_\pi)\right)\right],\ (\boldsymbol{\psi}, \boldsymbol{\gamma}_\pi) \sim \mu_\Psi \otimes \pi^{\otimes k},$$

to the Bethe functional is Lipschitz in π with respect to d_w.

Lemma 3.84. *There exists $L_{\mathfrak{g}}$ such that B_{f} is L-Lipschitz.*

Proof. Let $\pi_\circ \in \Gamma(\pi_1, \pi_2)$ be a coupling of $\pi \in \mathcal{P}^2([q])^2$ and $(\boldsymbol{\psi}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \sim \mu_\Psi \otimes \pi_\circ^{\otimes k}$ with $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{P}([q])^k$. Using that Λ is L'-Lipschitz on $[\psi_\downarrow, \psi_\uparrow]$ with $L' = \ln(\psi_\uparrow) + 1$, we have

$$\Delta = |B_{\mathrm{f}}(\pi_1) - B_{\mathrm{f}}(\pi_2)| \le \frac{d_\uparrow L'(k-1)}{k\psi_\downarrow} \mathbb{E}\left[|Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_1) - Z_{\mathrm{f}}(\boldsymbol{\psi}, \boldsymbol{\gamma}_2)|\right].$$

With the triangle inequality, $\boldsymbol{\psi} \le \psi_\uparrow$ and Observation 3.6b) this gives

$$\Delta \le 2d_\uparrow L'\psi_\uparrow^2(k-1)\sum_h \frac{1}{k}\mathbb{E}[\|\boldsymbol{\gamma}_{1,h} - \boldsymbol{\gamma}_{2,h}\|_{\mathrm{tv}}] = 2d_\uparrow L'\psi_\uparrow^2(k-1)\mathbb{E}[\|\boldsymbol{\gamma}_{1,1} - \boldsymbol{\gamma}_{2,1}\|_{\mathrm{tv}}].$$

This completes the proof since $\pi_\circ \in \Gamma(\pi_1, \pi_2)$ was arbitrary. □

Now, we are finally ready to establish the easier part of Lemma 3.67.

Lemma 3.85. *We have $\Phi_{\mathrm{f}}(n) = \mathbb{E}[B_{\mathrm{f}}(\pi^\circ_{\mathrm{g},\boldsymbol{G}^*})] + \mathcal{O}(n^{-\rho})$ with $\boldsymbol{G}^* = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*)$.*

Proof. With Lemma 3.83 we have $\Phi_{\mathrm{f}}(n) = \mathbb{E}[B_{\mathrm{f}}(\pi_{\mathrm{g},\boldsymbol{G}^*})] + \mathcal{O}(n^{-\rho})$. Lemma 3.84 and Lemma 3.56b) complete the proof, since $\mathrm{d_w} \le q$ and $\mathbb{P}(\boldsymbol{m} > m_\uparrow) = o(1/n)$. □

3.3.3.12 Typical Events for the Variable Contribution. For the variable contribution to the Bethe functional, let $\boldsymbol{\Phi} = \ln(Z_{\mathrm{g}}(\boldsymbol{G}^+(\boldsymbol{\sigma}^+))/Z_{\mathrm{g}}(\boldsymbol{G}_\cap))$ for $\Phi_{\mathrm{v}}(n) = \mathbb{E}[\boldsymbol{\Phi}]$ from Equation (9). Recall $r(n)$, $\mathcal{B}^\Gamma_+$ from Section 3.3.3.6, $\mathcal{B}^\circ$ from Section 3.3.3.7 and let $\bar{\boldsymbol{d}}^+ = k\boldsymbol{m}^+/n$.

Lemma 3.86. *We have $\Phi_{\mathrm{v}}(n) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \in \mathcal{B}^\Gamma_+, \boldsymbol{\sigma}^- \in \mathcal{B}^\Gamma_-, \bar{\boldsymbol{d}}^+ \in \mathcal{B}^\circ\}\boldsymbol{\Phi}] + o(n^{-1})$.*

Proof. With $\mathcal{E} = \{\boldsymbol{\sigma}^+ \in \mathcal{B}^\Gamma_+, \boldsymbol{\sigma}^- \in \mathcal{B}^\Gamma_-, \bar{\boldsymbol{d}}^+ \in \mathcal{B}^\circ\}$ we have $\Phi_{\mathrm{v}}(n) = \mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}] + \varepsilon$ with

$$\varepsilon(n) = \mathbb{E}\left[\mathbb{1}\neg\mathcal{E}\overline{\boldsymbol{\Phi}}\right], \ \overline{\boldsymbol{\Phi}} = \mathbb{E}\left[\boldsymbol{\Phi}\middle|\boldsymbol{m}_\cap(\boldsymbol{\sigma}^+), \boldsymbol{\sigma}^+, \boldsymbol{\sigma}^-, \bar{\boldsymbol{d}}^+\right].$$

With Jensen's inequality we can consider the atypical events separately, i.e.

$$|\varepsilon(n)| \le \mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}^+ \notin \mathcal{B}^\circ\}|\overline{\boldsymbol{\Phi}}|] + \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \notin \mathcal{B}^\Gamma_+\}|\overline{\boldsymbol{\Phi}}|] + \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^- \notin \mathcal{B}^\Gamma_-\}|\overline{\boldsymbol{\Phi}}|].$$

With $\boldsymbol{\Phi} = (n+1)\phi_{\mathrm{g}}(Z_{\mathrm{g}}(\boldsymbol{G}^+(\boldsymbol{\sigma}^+))) - n\phi_{\mathrm{g}}(Z_{\mathrm{g}}(\boldsymbol{G}_\cap))$, Jensen's inequality, the triangle inequality, $\tilde{c}$ from Observation 3.30, Lemma 3.73, Lemma 3.72 and Observation 3.47 we have

$$|\overline{\boldsymbol{\Phi}}| \le \tilde{c}\left(\boldsymbol{m}^+ + \frac{\Theta^\downarrow_{n+1}}{2} + \boldsymbol{m}_\cap(\boldsymbol{\sigma}^+) + \frac{n\Theta^\downarrow_n}{(n+1)2}\right) \le \tilde{c}(2\boldsymbol{m}^+ + n + 1).$$

So, with Lemma 3.70, c from Corollary 3.12 and $n' = n+1$ we have

$$\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}^+ \notin \mathcal{B}^\circ\}|\overline{\boldsymbol{\Phi}}|] \le \frac{2\tilde{c}n'}{k}\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}_{n'} \notin \mathcal{B}^\circ\}\bar{\boldsymbol{d}}_{n'}] + \tilde{c}n'\mathbb{P}(\bar{\boldsymbol{d}}_{n'} \notin \mathcal{B}^\circ) \le c'n'\exp\left(-\frac{c_1 r^2 n'}{1+r}\right)$$

with $c' = \tilde{c}c_2(2+k)/k$, so $\mathbb{E}[\mathbb{1}\{\bar{\boldsymbol{d}}^+ \notin \mathcal{B}^\circ\}|\overline{\boldsymbol{\Phi}}|] = o(1/n)$. With c from Observation 3.23b), independence and $c_{\mathrm{r}}^2 > 2/c_1$ we have

$$\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \notin \mathcal{B}^\Gamma_+\}|\overline{\boldsymbol{\Phi}}|] \le \tilde{c}\left(\frac{2d_\uparrow}{k} + 1\right)c_2 n' e^{-c_1 r^2 n'} = o(n^{-1}),$$

and $\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^- \notin \mathcal{B}_-^\Gamma\}|\overline{\boldsymbol{\Phi}}|] = o(n^{-1})$ follows analogously. □

The following result further restricts the very typical event in Lemma 3.86 to the typical event that no variables are pinned in the second sweep.

Lemma 3.87. *We have*

$$\Phi_{\mathrm{v}}(n) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^+ \in \mathcal{B}_+^\Gamma, \boldsymbol{\sigma}^- \in \mathcal{B}_-^\Gamma, \bar{\boldsymbol{d}}^+ \in \mathcal{B}^\circ, \boldsymbol{\mathcal{U}}^+ = \boldsymbol{\mathcal{U}}_\cap\}\boldsymbol{\Phi}] + \mathcal{O}(n^{-\rho}).$$

Proof. With Lemma 3.86 it is sufficient to consider $\mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\Phi}|]$, where

$$\mathcal{E} = \left\{\boldsymbol{\sigma}^+ \in \mathcal{B}_+^\Gamma, \boldsymbol{\sigma}^- \in \mathcal{B}_-^\Gamma, \bar{\boldsymbol{d}}^+ \in \mathcal{B}^\circ, \boldsymbol{\mathcal{U}}^+ \neq \boldsymbol{\mathcal{U}}_\cap\right\}.$$

As opposed to the proof of Lemma 3.78 we cannot use Observation 3.30 since now the numbers of variables do not coincide. Let $\boldsymbol{\mathcal{A}}_\cap \dot\cup \boldsymbol{\mathcal{A}}_+ = [\boldsymbol{m}^+]$ be the partition of the standard factors of $\boldsymbol{G}^+(\boldsymbol{\sigma}^+)$ such that $\boldsymbol{\mathcal{A}}_\cap$ is the relabeling of the standard factors $[\boldsymbol{m}_\cap]$ in $\boldsymbol{G}_\cap$. Further, let $\boldsymbol{\mathcal{V}}^\downarrow = \boldsymbol{\mathcal{U}}^+ \setminus \boldsymbol{\mathcal{U}}_\cap$ be the additional pins and $\boldsymbol{\mathcal{A}}^\downarrow = \{a \in \boldsymbol{\mathcal{A}}_\cap : \boldsymbol{v}_a^+([k]) \cap \boldsymbol{\mathcal{V}}^\downarrow \neq \emptyset\}$ with $\boldsymbol{v}^+$ being the neighborhoods in $\boldsymbol{G}^+(\boldsymbol{\sigma}^+)$. The bounds from the proof of Observation 3.30 and normalization of the external field for the last variable give

$$Z_{\mathrm{g}}(\boldsymbol{G}^+(\boldsymbol{\sigma}^+)) \le \psi_\uparrow^{|\boldsymbol{\mathcal{A}}_+|} Z_{\mathrm{g}}(\boldsymbol{G}_\cap),\ Z_{\mathrm{g}}(\boldsymbol{G}^+(\boldsymbol{\sigma}^+)) \ge \psi_\downarrow^{|\boldsymbol{\mathcal{A}}_+|+2|\boldsymbol{\mathcal{A}}^\downarrow|+|\boldsymbol{\mathcal{V}}^\downarrow|} Z_{\mathrm{g}}(\boldsymbol{G}_\cap).$$

This shows that $|\boldsymbol{\Phi}| \le \ln(\psi_\uparrow)(\boldsymbol{m}_\Delta^+ + 2|\boldsymbol{\mathcal{A}}^\downarrow| + |\boldsymbol{\mathcal{V}}^\downarrow|)$. Bounding $|\boldsymbol{\mathcal{A}}^\downarrow|$ by the sum of the degrees of $i \in \boldsymbol{\mathcal{V}}^\downarrow \cap [n]$ in $\boldsymbol{G}_\cap$ and taking the conditional expectation as in the proof of Lemma 3.78 gives the bound $\ln(\psi_\uparrow)(\boldsymbol{m}_\Delta^+ + 2c\frac{k\boldsymbol{m}_\cap}{n}|\boldsymbol{\mathcal{V}}^\downarrow \cap [n]| + |\boldsymbol{\mathcal{V}}^\downarrow|)$ with c from Corollary 3.18a). With c' from Lemma 3.76 we get the bound $\ln(\psi_\uparrow)(c' + 2c\frac{n+1}{n}(\bar{d}+r)|\boldsymbol{\mathcal{V}}^\downarrow \cap [n]| + |\boldsymbol{\mathcal{V}}^\downarrow|)$ on $\mathcal{E}$, so for some $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ we have

$$\mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\Phi}|] \le c\mathbb{P}(\boldsymbol{\mathcal{U}}^+ \neq \boldsymbol{\mathcal{U}}) + c\mathbb{E}[\mathbb{1}\{\boldsymbol{\mathcal{U}}^+ \neq \boldsymbol{\mathcal{U}}\}|\boldsymbol{\mathcal{U}}^+ \setminus \boldsymbol{\mathcal{U}}|] \le 2c\mathbb{E}[|\boldsymbol{\mathcal{U}}^+ \setminus \boldsymbol{\mathcal{U}}|].$$

As in the proof of Lemma 3.78, we trace this back to the indicators $\check{\boldsymbol{u}}_\Delta^+$ for the variables $[n]$, and pinning probability $\boldsymbol{\theta}_{n+1}/(n+1)$ for $i = n+1$. This yields

$$\mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\Phi}|] \le 2c\mathbb{E}\left[np_\Delta^+(\boldsymbol{\theta}^-, \boldsymbol{\theta}^+) + \frac{\boldsymbol{\theta}^+}{n+1}\right] \le c\left(\frac{n(\Theta_{n+1}^\downarrow - \Theta_n^\downarrow)}{n+1-\Theta_n^\downarrow} + \frac{\Theta_{n+1}^\downarrow}{n+1}\right).$$

With $\Theta^\downarrow(n) = n^{1-\rho}$ we have $\Theta_{n+1}^\downarrow - \Theta_n^\downarrow = (1-\rho)\int_n^{n+1} t^{-\rho}\mathrm{d}t \le n^{-\rho}$, and hence the assertion follows since $\mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\Phi}|] = \mathcal{O}(n^{-\rho})$. □

3.3.3.13 Normalization Step for the Variable Contribution. Now, we simplify the underlying law using the typical behavior. As before, we replace $\boldsymbol{G}_\cap$ by $\boldsymbol{G}^*(\sigma^-) = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\sigma^-)$, and $\boldsymbol{m}_\Delta^+(\boldsymbol{\sigma}^+)$ by a Poisson variable $\boldsymbol{d} \sim \mathrm{Po}(\bar{d})$ reflecting the degree of $i = n+1$. Recalling Lemma 3.76, we obtain the additional wires-weight pairs given $\boldsymbol{\sigma}^+$ from Observation 3.19, i.e. we consider

$$(\boldsymbol{G}^*(\sigma^-), \boldsymbol{d}, \boldsymbol{w}^*(\sigma^+)) \sim \boldsymbol{G}^*(\sigma^-) \otimes \boldsymbol{d} \otimes \boldsymbol{w}^{*\otimes\mathbb{Z}_{>0}}_{+\circ,n+1,i,\sigma^+}.$$

Since we have an additional variable, we have to adjust the definition

$$\overline{\psi}_{\mathrm{w}|\mathrm{g},G}(v,\psi)=\mathbb{E}\left[\prod_{a\in[d]}\psi_a(\boldsymbol{\sigma}_{v(a)})\right],\ \boldsymbol{\sigma}\sim\mu_{\mathrm{g},G}\otimes\gamma^*,\ (v,\psi)\in([n+1]^k\times\mathcal{D}_\Psi)^d,$$

from Equation (8), where G is still a decorated graph on n variables.

Lemma 3.88. *We have* $\Phi_{\mathrm{v}}(n)=\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-\}\ln(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^-)}(\boldsymbol{w}^*_{\boldsymbol{\sigma}^+,[\boldsymbol{d}]}))]+\mathcal{O}(n^{-\rho})$.

Proof. Let $\boldsymbol{d}$ and $\boldsymbol{w}^*(\sigma^+)$ be independent of anything else and

$$\mathcal{E}=\left\{\boldsymbol{\sigma}^+\in\mathcal{B}^\Gamma_+,\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-,\bar{\boldsymbol{d}}^+\in\mathcal{B}^\circ,\boldsymbol{\mathcal{U}}^+=\boldsymbol{\mathcal{U}}_\cap\right\}.$$

For $n\ge n_\circ$ with $n_\circ$ from Lemma 3.76 and on $\mathcal{E}$, as explained in Section 3.3.3.4 and conditional to $\boldsymbol{\sigma}^+$, $\boldsymbol{G}_\cap$, $\boldsymbol{\mathcal{U}}^+$ and $\boldsymbol{d}^+=\boldsymbol{m}^+_{\Delta+}(\boldsymbol{\sigma}^+)$, we obtain $\boldsymbol{G}^+(\boldsymbol{\sigma}^+)$ from $\boldsymbol{G}_\cap$ by adding the variable $i=n+1$ with external field γ^* and $\boldsymbol{d}^+$ standard factors with wires-weight pairs from $(\boldsymbol{v}^*,\boldsymbol{\psi}^*)=\boldsymbol{w}^*=\boldsymbol{w}^*_{\boldsymbol{\sigma}^+,[\boldsymbol{d}^+]}$. Hence, on $\mathcal{E}$ we have

$$\boldsymbol{r}=\frac{Z_{\mathrm{g}}(\boldsymbol{G}^+(\boldsymbol{\sigma}^+))}{\boldsymbol{G}_\cap}\sim\sum_{\sigma^+}\frac{\psi_{\mathrm{g},\boldsymbol{G}_\cap}(\sigma^+_{[n]})}{Z_{\mathrm{g}}(\boldsymbol{G}_\cap)}\gamma^*(\sigma^+_i)\prod_{a\in[\boldsymbol{d}^+]}\boldsymbol{\psi}^*_a(\sigma^+_{\boldsymbol{v}^*(a)})=\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}_\cap}(\boldsymbol{w}^*).$$

We couple $\boldsymbol{d}^+$ and $\boldsymbol{d}$ using $\boldsymbol{\delta}(\boldsymbol{\sigma}^+)\sim\mathrm{Po}(|\overline{m}^+_{\Delta+}(\boldsymbol{\sigma}^+)-\bar{d}|)$ as in the proof of Lemma 3.79 to obtain

$$\left|\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}_\cap}\left(\boldsymbol{w}^*_{\boldsymbol{\sigma}^+,[\boldsymbol{d}^+]}\right)\right)-\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}_\cap}\left(\boldsymbol{w}^*_{\boldsymbol{\sigma}^+,[\boldsymbol{d}]}\right)\right)\right|\le\boldsymbol{\delta}(\boldsymbol{\sigma}^+)\ln(\psi_\uparrow).$$

With Lemma 3.76 we can bound $\mathbb{E}[\boldsymbol{\delta}(\sigma^+)]$ on $\mathcal{E}$, so with Lemma 3.87 we have

$$\Phi_{\mathrm{v}}(n)=\mathbb{E}\left[\mathbb{1}\mathcal{E}\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}_\cap}\left(\boldsymbol{w}^*_{\boldsymbol{\sigma}^+,[\boldsymbol{d}]}\right)\right)\right]+\mathcal{O}\left(r(n)+n^{-\rho}\right),$$

since the expectations can be bounded by $c'\ln(\psi_\uparrow)$ with c' from Lemma 3.76 and $\ln(\psi_\uparrow)\bar{d}$ respectively for $n\le n_\circ$. This also shows that reducing $\mathcal{E}$ to $\{\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-\}$ causes an error of $\mathcal{O}(n^{-\rho})$, and that with the coupling from Lemma 3.75 we get

$$\Phi_{\mathrm{v}}(n)=\mathbb{E}\left[\mathbb{1}\{\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-\}\ln\left(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^-)}\left(\boldsymbol{w}^*_{\boldsymbol{\sigma}^+,[\boldsymbol{d}]}\right)\right)\right]+\mathcal{O}\left(n^{-\rho}\right).$$

□

In a second normalization step we simplify $\boldsymbol{w}^*_{\boldsymbol{\sigma}^+,[\boldsymbol{d}]}$ by establishing that $i=n+1$ typically does not wire more than once to the same factor. As seen in Section 3.3.3.10, it is reasonable to explicitly control the factor assignments. With $\sigma^+\in[q]^{n+1}$, $\sigma^-=\sigma^+_{[n]}$, $\gamma^-=\gamma_{\mathrm{n},\sigma^-}$ and $\sigma^\circ=\sigma^+_i$ let $(\boldsymbol{\tau}^+_{\circ,n,\sigma^+},\boldsymbol{h}^+_{\circ,n,\sigma^+})\in[q]^k\times[k]$ be given by

$$\mathbb{P}(\boldsymbol{\tau}^+_\circ=\tau,\boldsymbol{h}^+_\circ=h)=\frac{W(\tau,h)}{\overline{Z}^+_{\mathrm{f}}(\sigma^\circ,\gamma^-)},\ \overline{Z}^+_{\mathrm{f}}(\sigma^\circ,\gamma^-)=\sum_{\tau,h}W(\tau,h),$$
$$W(\tau,h)=\mathbb{1}\{\tau_h=\sigma^\circ\}\frac{1}{k}\overline{\psi}_\circ(\tau)\prod_{h'\in[k]\setminus\{h\}}\gamma^-(\tau_{h'}).$$

So, with $\mathcal{T}^+_{\sigma^+}=\{(\tau,h)\in[q]^k\times[k]:\tau_h=\sigma^\circ,\tau([k]\setminus\{h\})\subseteq\sigma^-([n])\}$ we have $(\boldsymbol{\tau}^+_\circ,\boldsymbol{h}^+_\circ)\in\mathcal{T}^+$ almost surely. Further, for $\sigma^\circ\in\sigma^-([n])$ we have $\overline{Z}^+_{\mathrm{f}}(\sigma^\circ,\gamma^-)=\overline{Z}_{\mathrm{f}}(\gamma^-)\mu|_*(\sigma^\circ)/\gamma^-(\sigma^\circ)$ with $\mu=\mu_{\mathrm{T}|\Gamma,\gamma^-}$. For $(\tau,h)\in\mathcal{T}^+$ let $(\boldsymbol{v}^+_{\circ,\sigma^-,\tau,h},\boldsymbol{\psi}^+_{\circ,\tau})\sim\mathrm{u}(\mathcal{V}^+)\otimes\boldsymbol{\psi}^\star_{\circ,\tau}$ with $\boldsymbol{\psi}^\star_\circ$ from Section 3.2.1.5 and $\mathcal{V}^+_{\sigma^-,\tau,h}=\{v\in[n+1]^k:v_h=n+1,\forall h'\in[k]\setminus\{h\}\,v(h')\in\sigma^{--1}(\tau_{h'})\}$.

For $d\in\mathbb{Z}_{\ge0}$ and $(\tau,h)\in\mathcal{T}^d_{\sigma^+}$ let $\boldsymbol{w}^+_{\sigma^-,d,\tau,h}\sim\bigotimes_{a\in[d]}(\boldsymbol{v}^+_{\circ,\sigma^-,\tau(a),h(a)},\boldsymbol{\psi}^+_{\circ,\tau(a)})$ and

$$\left(\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\sigma^-),\boldsymbol{d},\boldsymbol{\tau}^+_{\sigma^+},\boldsymbol{h}^+_{\sigma^+}\right)\sim\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\sigma^-)\otimes\boldsymbol{d}\otimes(\boldsymbol{\tau}^+_{\circ,\sigma^+},\boldsymbol{h}^+_{\circ,\sigma^+})^{\otimes\mathbb{Z}_{>0}}. \tag{12}$$

Finally, let $\boldsymbol{G}^*(\sigma^-)=\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\sigma^-)$, $\boldsymbol{\tau}^+=\boldsymbol{\tau}^+_{\boldsymbol{\sigma}^+,[\boldsymbol{d}]}$, $\boldsymbol{h}^+=\boldsymbol{h}^+_{\boldsymbol{\sigma}^+,[\boldsymbol{d}]}$ and $\boldsymbol{w}^+=\boldsymbol{w}^+_{\boldsymbol{\sigma}^-,\boldsymbol{d},\boldsymbol{\tau}^+,\boldsymbol{h}^+}$.

Lemma 3.89. *We have* $\Phi_{\mathrm{v}}(n)=\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-\}\ln(\overline{\psi}_{\mathrm{w|g},\boldsymbol{G}^*(\boldsymbol{\sigma}^-)}(\boldsymbol{w}^+))]+\mathcal{O}(n^{-\rho})$.

Proof. Let $n_+=n+1$ and $i=n+1$. Further, let $\mathcal{V}_0=\mathcal{V}_1\,\dot\cup\,\mathcal{V}_2$ with

$$\mathcal{V}_1=\{v\in[n_+]^k:|v^{-1}(i)|=1\},\ \mathcal{V}_2=\{v\in[n_+]^k:|v^{-1}(i)|>1\}.$$

For $a\in\{0,1\}$ let $(\boldsymbol{v}_{\circ,a},\boldsymbol{\psi}_{\circ,a})\sim\mathrm{u}(\mathcal{V}_a)\otimes\mu_\Psi$ and let $(\boldsymbol{v}^*_{\circ,a},\boldsymbol{\psi}^*_{\circ,a})$ be given by the Radon-Nikodym derivative $r_a(v,\psi)=\psi(\sigma^+_v)/z_a(\sigma^+)$ with $z_a(\sigma^+)=\mathbb{E}[\boldsymbol{\psi}_{\circ,a}(\sigma^+_{\boldsymbol{v}_{\circ,a}})]$. With $(\boldsymbol{v},\boldsymbol{\psi})=(\boldsymbol{v}_{\circ,0},\boldsymbol{\psi}_{\circ,0})$, $P_\circ=\mathbb{P}(\boldsymbol{v}\in\mathcal{V}_2)$, $P=\mathbb{P}(\boldsymbol{v}^*_{\circ,0}\in\mathcal{V}_2)$, $\boldsymbol{v}_{\mathrm{u}}\sim\mathrm{u}([n_+]^k)$ and using $r_a\in[\psi^2_\downarrow,\psi^2_\uparrow]$ we have

$$P\le\psi^2_\uparrow P_\circ=\frac{\psi^2_\uparrow\mathbb{P}(\boldsymbol{v}_{\mathrm{u}}\in\mathcal{V}_2)}{\mathbb{P}(\boldsymbol{v}_{\mathrm{u}}\in\mathcal{V}_0)}\le\frac{\psi^2_\uparrow\mathbb{P}(\boldsymbol{v}_{\mathrm{u}}\in\mathcal{V}_2)}{\mathbb{P}(\boldsymbol{v}_{\mathrm{u}}\in\mathcal{V}_1)}=\frac{\psi^2_\uparrow(n^k_+-n^k-kn^{k-1})}{kn^{k-1}}\le\frac{c}{n},\ c=\frac{\psi^2_\uparrow2^k}{k},$$

as in the proof of Observation 3.16c). Since the $(\boldsymbol{v}_{\circ,1},\boldsymbol{v})$-derivative is $r_{\mathrm{v}}(v)=\mathbb{1}\{v\in\mathcal{V}_1\}/\mathbb{P}(\boldsymbol{v}\in\mathcal{V}_1)$, the Radon-Nikodym derivative of $(\boldsymbol{v}^*_{\circ,1},\boldsymbol{\psi}^*_{\circ,1})$ with respect to $(\boldsymbol{v},\boldsymbol{\psi})$ is $r(v,\psi)=r_1(v,\psi)r_{\mathrm{v}}(v)=\mathbb{1}\{v\in\mathcal{V}_1\}\psi(\sigma^+_v)/z^\circ_1(\sigma^+)$ with $z^\circ_1(\sigma^+)=\mathbb{E}[\mathbb{1}\{\boldsymbol{v}\in\mathcal{V}_1\}\boldsymbol{\psi}(\sigma^+_{\boldsymbol{v}})]$. Clearly, we have $z^\circ_1(\sigma^+)\le z_0(\sigma^+)$, and on the other hand

$$R(\sigma^+)=\frac{z_0(\sigma^+)}{z^\circ_1(\sigma^+)}=1+\frac{\mathbb{E}[\mathbb{1}\{\boldsymbol{v}\in\mathcal{V}_2\}\boldsymbol{\psi}_0(\sigma^+_{\boldsymbol{v}})]}{z^\circ_1(\sigma^+)}\le1+\frac{\psi_\uparrow P_\circ}{\psi_\downarrow\mathbb{P}(\boldsymbol{v}\in\mathcal{V}_1)}=1+\frac{\psi^2_\uparrow\mathbb{P}(\boldsymbol{v}_{\mathrm{u}}\in\mathcal{V}_2)}{\mathbb{P}(\boldsymbol{v}_{\mathrm{u}}\in\mathcal{V}_1)},$$

so the bound for P from above yields $1\le R(\sigma^+)\le1+\frac{c}{n}$.

Now, we turn back to Φ_{v}. With Lemma 3.88 we have $\Phi_{\mathrm{v}}(n)=\mathbb{E}[\mathbb{1}\mathcal{E}^\circ\boldsymbol{\Phi}]+\mathcal{O}(n^{-\delta})$, where

$$\boldsymbol{\Phi}=f_{\boldsymbol{G}}(\boldsymbol{w}^*),\ f_G(w)=\ln(\overline{\psi}_{\mathrm{w|g},G}(w)),\ \boldsymbol{G}=\boldsymbol{G}^*(\boldsymbol{\sigma}^-),\ \boldsymbol{w}^*=\boldsymbol{w}^*_{\boldsymbol{\sigma}^+,[\boldsymbol{d}]}.$$

Notice that $|\boldsymbol{\Phi}|\le\boldsymbol{d}\ln(\psi_\uparrow)$, and that $\boldsymbol{w}^*$ given $\boldsymbol{\sigma}^+$, $\boldsymbol{d}$ are $\boldsymbol{d}$ iid copies of $(\boldsymbol{v}^*_{\circ,0},\boldsymbol{\psi}^*_{\circ,0})$ from above. Hence, the bound on $\boldsymbol{\Phi}$ with the union bound yield $\Delta=|\mathbb{E}[\mathbb{1}\mathcal{E}^\circ\mathbb{1}\neg\mathcal{E}\boldsymbol{\Phi}]|\le\ln(\psi_\uparrow)\mathbb{E}[\boldsymbol{d}^2]P$. Recall that $\boldsymbol{d}\sim\mathrm{Po}(\bar d)$, hence $\mathbb{E}[\boldsymbol{d}^2]=\bar d+\bar d^2\le d_\uparrow(d_\uparrow+1)$, and that $P\le c/n$, so $\Delta=\mathcal{O}(1/n)$ and $\Phi_{\mathrm{v}}(n)=\Phi^\circ_{\mathrm{v}}(n)+\mathcal{O}(n^{-\delta})$ with $\Phi^\circ_{\mathrm{v}}(n)=\mathbb{E}[\mathbb{1}\mathcal{E}^\circ\mathbb{1}\mathcal{E}\boldsymbol{\Phi}]$. Now, let $\boldsymbol{w}=(\boldsymbol{v},\boldsymbol{\psi})\sim(\boldsymbol{v}_{\circ,0},\boldsymbol{\psi}_{\circ,0})^{\otimes\mathbb{Z}_{>0}}$ and $\boldsymbol{w}^*_{\sigma^+}\sim(\boldsymbol{v}^*_{\circ,1},\boldsymbol{\psi}^*_{\circ,1})^{\otimes\mathbb{Z}_{>0}}$ be independent of anything else. With the shorthand $\boldsymbol{w}^*=\boldsymbol{w}^*_{\boldsymbol{\sigma}^+}$ we have

$$\Phi^\circ_{\mathrm{v}}(n)=\mathbb{E}\left[\mathbb{1}\mathcal{E}^\circ\prod_{a\in[\boldsymbol{d}]}(\mathbb{1}\{\boldsymbol{v}_a\in\mathcal{V}_1\}r_0(\boldsymbol{w}_a))f_{\boldsymbol{G}}(\boldsymbol{w}_{[\boldsymbol{d}]})\right]=\mathbb{E}\left[\mathbb{1}\mathcal{E}^\circ f_{\boldsymbol{G}}(\boldsymbol{w}^*_{[\boldsymbol{d}]})R(\boldsymbol{\sigma}^+)^{-\boldsymbol{d}}\right].$$

Using $|f_{\boldsymbol{G}}(\boldsymbol{w}^*_{[\boldsymbol{d}]})| \le \boldsymbol{d}c'$, $c' = \ln(\psi_\uparrow)$, and $1 \le \boldsymbol{R} \le 1 + \frac{c}{n}$ with $\boldsymbol{R} = R(\boldsymbol{\sigma}^+)$ further gives

$$\Delta = \left|\Phi^\circ_{\mathrm{v}}(n) - \mathbb{E}\left[\mathbb{1}\mathcal{E}^\circ f_{\boldsymbol{G}}(\boldsymbol{w}^*_{[\boldsymbol{d}]})\right]\right| \le c'\mathbb{E}\left[\boldsymbol{d}\left(1 - \boldsymbol{R}^{-\boldsymbol{d}}\right)\right] \le c'\mathbb{E}\left[\boldsymbol{d}\left(\left(1 + \frac{c}{n}\right)^{\boldsymbol{d}} - 1\right)\right].$$

With $\boldsymbol{d}_\uparrow \sim \mathrm{Po}(d_\uparrow)$, the standard coupling of $\boldsymbol{d}$ and $\boldsymbol{d}_\uparrow$, Lipschitz continuity (for $\boldsymbol{d}_\uparrow > 0$) and the moment generating function of $\mathrm{Po}(d_\uparrow)$ we have

$$\Delta \le \frac{c'c}{n}\mathbb{E}\left[\boldsymbol{d}_\uparrow^2\left(1 + \frac{c}{n}\right)^{\boldsymbol{d}_\uparrow - 1}\right] \le \frac{c'c}{n}\mathbb{E}\left[e^{\boldsymbol{d}_\uparrow \lambda}\right] = \mathcal{O}(n^{-1}),\ \lambda = 2 + \ln\left(1 + \frac{c}{n}\right) \le 2 + \frac{c}{n}.$$

This shows that $\Phi_{\mathrm{v}}(n) = \mathbb{E}[\mathbb{1}\mathcal{E}^\circ f_{\boldsymbol{G}}(\boldsymbol{w}^*_{[\boldsymbol{d}]})] + \mathcal{O}(n^{-\rho})$. Now, due to the conditional independence given $\boldsymbol{\sigma}^+$ it suffices to show that $\boldsymbol{w}^+$ and $\boldsymbol{w}^*_{[\boldsymbol{d}]}$ given $\boldsymbol{\sigma}^+$, $\boldsymbol{d}$ have the same law. Hence, for fixed σ^+ we have to show that $(\boldsymbol{v}^+, \boldsymbol{\psi}^+) \sim (\boldsymbol{v}^*, \boldsymbol{\psi}^*)$, where

$$(\boldsymbol{v}^+, \boldsymbol{\psi}^+) = \left(\boldsymbol{v}^+_{\circ,\boldsymbol{\tau}^+_\circ,\boldsymbol{h}^+_\circ}, \boldsymbol{\psi}^+_{\circ,\boldsymbol{\tau}^+_\circ}\right),\ (\boldsymbol{v}^*, \boldsymbol{\psi}^*) = \left(\boldsymbol{v}^*_{\circ,1}, \boldsymbol{\psi}^*_{\circ,1}\right).$$

First, notice that the normalization constants coincide, i.e.

$$z_1 = \sum_v \frac{\mathbb{1}\{v \in \mathcal{V}_1\}}{kn^{k-1}}\overline{\psi}_\circ(\sigma^+_v) = \sum_{(\tau,h)\in\mathcal{T}^+} \frac{1}{k}\overline{\psi}_\circ(\tau)\sum_v \frac{\mathbb{1}\{v \in \mathcal{V}^+\}}{n^{k-1}} = \overline{Z}^+_{\mathrm{f}},$$

similar to the discussion in Section 3.2.1.5. For $v \in \mathcal{V}_1$ let $\tau(v) = \sigma^+_v$ and $h(v) \in [k]$ uniquely determined by $v(h(v)) = i$. Notice that we have $\boldsymbol{\tau}^+_\circ = \tau(\boldsymbol{v}^+_\circ)$ and $\boldsymbol{h}^+_\circ = h(\boldsymbol{v}^+_\circ)$ by definition, and $\mathcal{T}^+ = \{(\tau(v), h(v)) : v \in \mathcal{V}_1\}$. So, for an event $\mathcal{E}$ and with $\boldsymbol{\psi} \sim \mu_\Psi$ we have

$$\begin{aligned}\mathbb{P}((\boldsymbol{v}^+, \boldsymbol{\psi}^+) \in \mathcal{E}) &= \mathbb{E}\left[\sum_{(\tau,h)\in\mathcal{T}^+}\sum_{v\in\mathcal{V}^+_{\tau,h}} \frac{\overline{\psi}_\circ(\tau)\prod_{h'\ne h}\gamma^-(\tau_{h'})\boldsymbol{\psi}(\tau)}{k\overline{Z}^+_{\mathrm{f}}\prod_{h'\ne h}(n\gamma^-(\tau_{h'}))\overline{\psi}_\circ(\tau)}\mathbb{1}\{(v,\boldsymbol{\psi}) \in \mathcal{E}\}\right]\\ &= \mathbb{E}\left[\sum_{(\tau,h)\in\mathcal{T}^+}\sum_{v\in\mathcal{V}^+_{\tau,h}} \frac{\boldsymbol{\psi}(\tau)}{kn^{k-1}z_1}\mathbb{1}\{(v,\boldsymbol{\psi}) \in \mathcal{E}\}\right]\\ &= \mathbb{E}\left[r_1(\boldsymbol{v}_{\circ,1}, \boldsymbol{\psi}_{\circ,1})\mathbb{1}\{(\boldsymbol{v}_{\circ,1}, \boldsymbol{\psi}_{\circ,1}) \in \mathcal{E}\}\right] = \mathbb{P}((\boldsymbol{v}^*, \boldsymbol{\psi}^*) \in \mathcal{E}).\end{aligned}$$

□

3.3.3.14 Gibbs Marginal Product for the Variable Contribution. We are ready to apply Proposition 3.48. Using the distribution (12) and the corresponding shorthands let $(\boldsymbol{v}^+, \boldsymbol{\psi}^+) = \boldsymbol{w}^+$, $\boldsymbol{\gamma}^+ = (\mu_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^-)}|_{\boldsymbol{v}^+(a,h)})_{a\in[\boldsymbol{d}],h\ne\boldsymbol{h}^+(a)}$, recall Z_{v} from Section 2.1.2.4 and let $\boldsymbol{\Phi} = \ln\left(Z_{\mathrm{v}}\left(\boldsymbol{d}, \boldsymbol{\psi}^+, \boldsymbol{h}^+, \boldsymbol{\gamma}^+\right)\right)$, where we dropped the redundant dependencies on $\gamma_{a,h(a)}$ in the definition of Z_{v}.

Lemma 3.90. *We have* $\Phi_{\mathrm{v}}(n) = \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^- \in \mathcal{B}^\Gamma_-\}\boldsymbol{\Phi}] + \mathcal{O}(n^{-\rho})$.

Proof. Recall n_+, i, $\mathcal{V}_1$, $(\boldsymbol{v}_{\circ,1}, \boldsymbol{\psi}_{\circ,1})$, $(\boldsymbol{v}^*_{\circ,1}, \boldsymbol{\psi}^*_{\circ,1})$, r_1, $\mathcal{E}^\circ$ from the proof of Lemma 3.89, and that $(\boldsymbol{v}^+_{\circ,\boldsymbol{\tau}^+_\circ,\boldsymbol{h}^+_\circ}, \boldsymbol{\psi}^+_{\circ,\boldsymbol{\tau}^+_\circ}) \sim (\boldsymbol{v}^*_{\circ,1}, \boldsymbol{\psi}^*_{\circ,1})$, all for given σ^+. First, we resolve the reweighting, i.e. we consider $(\boldsymbol{\sigma}^+, \boldsymbol{G}^*(\boldsymbol{\sigma}^-), \boldsymbol{d}, \boldsymbol{w}) \sim (\boldsymbol{\sigma}^+, \boldsymbol{G}^*(\boldsymbol{\sigma}^-)) \otimes \boldsymbol{d} \otimes (\boldsymbol{v}_{\circ,1}, \boldsymbol{\psi}_{\circ,1})^{\otimes\mathbb{Z}_{>0}}$, and Lemma 3.89 yields $\Phi_{\mathrm{v}} = \Phi^\circ_{\mathrm{v}} + \mathcal{O}(n^{-\rho})$

with $\Phi^\circ_{\mathrm{v}} = \mathbb{E}[\mathbb{1}\mathcal{E}^\circ \prod_{a\in[\boldsymbol{d}]} r_1(\boldsymbol{w}_a)\ln(\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^-)}(\boldsymbol{w}_{[\boldsymbol{d}]}))]$. Next, notice that $V:[k]\times[n]^{k-1}\to\mathcal{V}_1$ is a bijection, where $v'=V(h,v)$ is given by $v'_h=n+1$ and $v'\circ\eta=v$, with $\eta:[k-1]\to[k]\setminus\{h\}$ denoting the enumeration. Notice that $\boldsymbol{v}_{\circ,1}\sim V(\boldsymbol{h},\boldsymbol{v})$ with $(\boldsymbol{h},\boldsymbol{v})\sim\mathrm{u}([k])\otimes\mathrm{u}([n]^{k-1})$. So, with

$$\left(\boldsymbol{\sigma}^+,\boldsymbol{G}^*(\boldsymbol{\sigma}^-),\boldsymbol{d},\boldsymbol{h},\boldsymbol{v},\boldsymbol{\psi}\right)\sim(\boldsymbol{\sigma}^+,\boldsymbol{G}^*(\boldsymbol{\sigma}^-))\otimes\boldsymbol{d}\otimes\left(\mathrm{u}([k])\otimes\mathrm{u}([n]^{k-1})\otimes\mu_\Psi\right)^{\otimes\mathbb{Z}_{>0}}$$

we have $\Phi^\circ_{\mathrm{v}}=\mathbb{E}[\mathbb{1}\mathcal{E}^\circ\boldsymbol{\Phi}^*]$, where $\boldsymbol{\Phi}^*=\prod_{a\in[\boldsymbol{d}]}r_1(\boldsymbol{w}_a)\ln(\boldsymbol{Z}^*)$, $\boldsymbol{w}=(V(\boldsymbol{h}_a,\boldsymbol{v}_a),\boldsymbol{\psi}_a)_a$,

$$\boldsymbol{Z}^*=\overline{\psi}_{\mathrm{w}|\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^-)}(\boldsymbol{w}_{[\boldsymbol{d}]})=\sum_{\sigma^\circ}\gamma^*(\sigma^\circ)\sum_\tau\boldsymbol{\alpha}^*(\tau)\prod_{a\in[\boldsymbol{d}]}\boldsymbol{\psi}_a(t(\tau_a,\sigma^\circ,\boldsymbol{h}_a)),$$

further $\boldsymbol{\alpha}^*=\boldsymbol{\mu}^*|_{\boldsymbol{v}'}$, $\boldsymbol{v}'=\boldsymbol{v}_{[\boldsymbol{d}]}\in([n]^{k-1})^{\boldsymbol{d}}$, $\boldsymbol{\mu}^*=\mu_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^-)}$, and $\tau'=t(\tau,\sigma^\circ,h)\in[q]^k$ given by $\tau'_h=\sigma^\circ$ and $\tau'\circ\eta=\tau$ using the enumeration $\eta:[k-1]\to[k]\setminus\{h\}$. Now, regarding $\boldsymbol{\alpha}^*$, the situation is very similar to the proof of Lemma 3.80, in particular given $\boldsymbol{d}$ we have $(\boldsymbol{\mu}^*,\boldsymbol{v}')\sim\boldsymbol{\mu}^*\otimes\mathrm{u}([n])^{\otimes(k-1)\boldsymbol{d}}$. Hence, let C from Proposition 3.48b), $\varepsilon=C_1/3$ and $\delta=\Theta^{\downarrow-2\varepsilon}$. Using $\iota_\circ$ from Section 3.3.1.5 let $\mathcal{E}=\{\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-,\iota_\circ(\boldsymbol{\mu}^*,\boldsymbol{v}')\le\delta\}$. Hence, the bound $|\boldsymbol{\Phi}^*|\le\ln(\psi_\uparrow^{\boldsymbol{d}})\psi_\uparrow^{2\boldsymbol{d}}\le\psi_\uparrow^{3\boldsymbol{d}}$ and Markov's inequality conditional to $\boldsymbol{d}$ give $\Delta\le\varepsilon'+\mathcal{O}(n^{-\rho})$, where $\Delta=|\Phi_{\mathrm{v}}-\mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}]|$ and

$$\varepsilon'=\mathbb{E}\left[\mathbb{1}\{\boldsymbol{d}>0\}\frac{C_2((k-1)\boldsymbol{d}-1)}{\delta}\left(\frac{(k-1)\boldsymbol{d}}{\Theta^\downarrow}\right)^{C_1}\psi_\uparrow^{3\boldsymbol{d}}\right].$$

Standard bounds imply $\varepsilon'\le\tilde{c}\mathbb{E}[\exp(\tilde{c}\boldsymbol{d})]/\Theta^{\downarrow\varepsilon}$ for some $\tilde{c}\in\mathbb{R}_{>0}$. The canonical coupling of $\boldsymbol{d}\sim\mathrm{Po}(\bar{d})$ and $\mathrm{Po}(d_\uparrow)$ gives $\varepsilon'=\mathcal{O}(\Theta^{\downarrow-\varepsilon})$. Recall that $\delta=n^{-2\rho}$ and $\varepsilon',\Delta=\mathcal{O}(n^{-\rho})$ as in the proof of Lemma 3.80. Now, with $\boldsymbol{\alpha}=\bigotimes_{(a,h)\in[\boldsymbol{d}]\times[k-1]}\boldsymbol{\mu}^*|_{\boldsymbol{v}'(a,h)}$, further

$$\boldsymbol{Z}=\sum_{\sigma^\circ}\gamma^*(\sigma^\circ)\sum_\tau\boldsymbol{\alpha}(\tau)\prod_{a\in[\boldsymbol{d}]}\boldsymbol{\psi}_a(t(\tau_a,\sigma^\circ,\boldsymbol{h}_a)),$$

and $\boldsymbol{\Phi}=\prod_{a\in[\boldsymbol{d}]}r_1(\boldsymbol{w}_a)\ln(\boldsymbol{Z})$, notice that $\boldsymbol{Z}^*,\boldsymbol{Z}\in[\psi_\downarrow^{\boldsymbol{d}},\psi_\uparrow^{\boldsymbol{d}}]$, so Lipschitz continuity of the logarithm gives $|\boldsymbol{\Phi}^*-\boldsymbol{\Phi}|\le\psi_\uparrow^{3\boldsymbol{d}}|\boldsymbol{Z}^*-\boldsymbol{Z}|\le2\psi_\uparrow^{4\boldsymbol{d}}\|\boldsymbol{\alpha}^*-\boldsymbol{\alpha}\|_{\mathrm{tv}}$. Remark 3.49 yields $|\boldsymbol{\Phi}^*-\boldsymbol{\Phi}|\le\sqrt{2}\psi_\uparrow^{4\boldsymbol{d}}\sqrt{\iota_\circ(\boldsymbol{\mu}^*,\boldsymbol{v}')}\le\sqrt{2}\psi_\uparrow^{4\boldsymbol{d}}\sqrt{\delta}$ on $\mathcal{E}$ and hence

$$\Phi_{\mathrm{v}}=\mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}]+\mathcal{O}\left(\sqrt{\delta}+n^{-\rho}\right)=\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-\}\boldsymbol{\Phi}]+\mathcal{O}\left(n^{-\rho}\right).$$

The assertion follows by reintroducing $\boldsymbol{w}^+$ using the Radon-Nikodym derivative in $\boldsymbol{\Phi}$. □

3.3.3.15 Marginal Distribution for the Variable Contribution. Now, we work towards the discussion in Section 3.3.1.8. Using the distribution (12) and for σ^+, d, τ, h let $(\boldsymbol{w}^+,\boldsymbol{\gamma}_{\sigma^-,d,\tau,h})\sim\boldsymbol{w}^+\otimes\bigotimes_{a\in[d],h'\in[k]\setminus\{h(a)\}}\check{\pi}_{\mathrm{g},\boldsymbol{G}^*(\sigma^-),\sigma^-,\tau(a,h')}$, let $\boldsymbol{\gamma}=\boldsymbol{\gamma}_{\boldsymbol{\sigma}^-,\boldsymbol{d},\boldsymbol{\tau}^+,\boldsymbol{h}^+}$ and $(\boldsymbol{v}^+,\boldsymbol{\psi}^+)=\boldsymbol{w}^+$.

Lemma 3.91. *We have* $\Phi_{\mathrm{v}}(n)=\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-\}\ln(Z_{\mathrm{v}}(\boldsymbol{d},\boldsymbol{\psi}^+,\boldsymbol{h}^+,\boldsymbol{\gamma}))]+\mathcal{O}(n^{-\rho})$.

Proof. As for Lemma 3.81, the assertion is immediate using the definition of $\boldsymbol{v}^+$ and $\check{\pi}_{\mathrm{g}}$. □

Let $(\boldsymbol{w}^+,\hat{\gamma}_{\sigma^-,d,\tau,h})\sim\boldsymbol{w}^+\otimes\bigotimes_{a\in[d],h'\in[k]\setminus\{h(a)\}}\hat{\pi}_{\mathrm{g},\boldsymbol{G}^*(\sigma^-),\tau(a,h')}$, and $\hat{\boldsymbol{\gamma}}=\hat{\boldsymbol{\gamma}}_{\boldsymbol{\sigma}^-,\boldsymbol{d},\boldsymbol{\tau}^+,\boldsymbol{h}^+}$.

Lemma 3.92. *We have* $\Phi_{\mathrm{v}}(n)=\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^-\in\mathcal{B}^\Gamma_-\}\ln(Z_{\mathrm{v}}(\boldsymbol{d},\boldsymbol{\psi}^+,\boldsymbol{h}^+,\hat{\boldsymbol{\gamma}}))]+\mathcal{O}(n^{-\rho})$.

Proof. As for Lemma 3.82, let $\Delta = |\check{E} - \hat{E}|$ be the difference of $\check{E} = \mathbb{E}[\ln(Z_{\mathrm{v}}(d, \psi, h, \boldsymbol{\gamma}))]$, where $\boldsymbol{\gamma} \sim \bigotimes_{a,h' \neq h(a)} \check{\pi}_{\mathrm{g},G,\sigma^-,\tau(a,h')}$, and $\hat{E} = \mathbb{E}[\ln(Z_{\mathrm{v}}(d, \psi, h, \boldsymbol{\gamma}))]$, where $\boldsymbol{\gamma} \sim \bigotimes_{a,h' \neq h(a)} \hat{\pi}_{\mathrm{g},G,\tau(a,h')}$. For couplings $\pi_{a,h'} \in \Gamma(\check{\pi}_{\mathrm{g},G,\sigma^-,\tau(a,h')}, \hat{\pi}_{\mathrm{g},G,\tau(a,h')})$ with $h' \neq h(a)$ we define $(\check{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}) \sim \bigotimes_{a,h' \neq h(a)} \pi_{a,h'}$ analogously. With $\psi_{\downarrow}^d \leq Z_{\mathrm{v}}(d, \psi, h, \cdot) \leq \psi_{\uparrow}^d$ and Observation 3.6b) we get

$$\begin{aligned}
\Delta &\leq \psi_{\uparrow}^{2d} \sum_{\sigma^\circ} \gamma^*(\sigma^\circ) \mathbb{E}\left[\sum_{\tau} \left| \prod_{a,h' \neq h(a)} \check{\boldsymbol{\gamma}}_{a,h'}(\tau_{a,h'}) - \prod_{a,h' \neq h(a)} \hat{\boldsymbol{\gamma}}_{a,h'}(\tau_{a,h'}) \right|\right] \\
&\leq 2\psi_{\uparrow}^{2d} \sum_{a,h' \neq h(a)} \mathbb{E}\left[\|\check{\boldsymbol{\gamma}}_{a,h'} - \hat{\boldsymbol{\gamma}}_{a,h'}\|_{\mathrm{tv}}\right].
\end{aligned}$$

Hence, we have $\Delta \leq 2\psi_{\uparrow}^{2d} \sum_{a,h' \neq h(a)} \mathrm{d_w}(\check{\pi}_{\mathrm{g},G,\sigma^-,\tau(a,h')}, \hat{\pi}_{\mathrm{g},G,\tau(a,h')})$, so with $D(\sigma, \mu)$ from Corollary 3.54 this yields $\Delta \leq 2d(k-1)\psi_{\uparrow}^{2d} D(\sigma^-, \mu_{\mathrm{g},G})$. Taking the expectation and using Corollary 3.54b) with $\boldsymbol{\gamma}$ from Lemma 3.91 gives

$$\begin{aligned}
\Delta &= \left|\mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^- \in \mathcal{B}_-^{\Gamma}\} \ln(Z_{\mathrm{v}}(\boldsymbol{d}, \boldsymbol{\psi}^+, \boldsymbol{h}^+, \boldsymbol{\gamma}))] - \mathbb{E}[\mathbb{1}\{\boldsymbol{\sigma}^- \in \mathcal{B}_-^{\Gamma}\} \ln(Z_{\mathrm{v}}(\boldsymbol{d}, \boldsymbol{\psi}^+, \boldsymbol{h}^+, \hat{\boldsymbol{\gamma}}))]\right| \\
&\leq \mathbb{E}[2\boldsymbol{d}(k-1)\psi_{\uparrow}^{2\boldsymbol{d}}] \mathbb{E}\left[D(\boldsymbol{\sigma}^-, \boldsymbol{G}^*(\boldsymbol{\sigma}^-))\right] = o(n^{-\rho})
\end{aligned}$$

as in the proof of Lemma 3.82 by coupling $\boldsymbol{d}$, $\mathrm{Po}(\bar{d})$, so Lemma 3.91 yields the assertion. □

3.3.3.16 The Variable Contribution. In this section we complete the discussion of Φ_{v}. First, we resolve the reweighting, then we turn to the projection onto $\mathcal{P}_*^2([q])$. Similar to Section 2.1.2.4 let

$$(\boldsymbol{d}, \boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}) \sim \mathrm{Po}(\bar{d}) \otimes (\mu_\Psi \otimes \mathrm{u}([k]) \otimes \pi_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}^{\otimes k})^{\otimes \mathbb{Z}_{>0}}$$

with $\boldsymbol{G}^*(\boldsymbol{\sigma}^*) = \boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*)$, $\boldsymbol{\psi} = \boldsymbol{\psi}_{[\boldsymbol{d}]}$, $\boldsymbol{h} = \boldsymbol{h}_{[\boldsymbol{d}]}$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}_{[\boldsymbol{d}]}$ by an abuse of notation.

Lemma 3.93. *We have* $\Phi_{\mathrm{v}}(n) = \mathbb{E}[\xi^{-\boldsymbol{d}} \Lambda(Z_{\mathrm{v}}(\boldsymbol{d}, \boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}))] + \mathcal{O}(n^{-\rho})$.

Proof. Let $i = n+1$, $\boldsymbol{\gamma}^- = \gamma_{\mathrm{n},\boldsymbol{\sigma}^-}$, $\boldsymbol{\sigma}^\circ = \boldsymbol{\sigma}_i^+$ and $\overline{\boldsymbol{\gamma}} = \overline{\gamma}_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^-)}$. With Lemma 3.52a) we have $\mathbb{P}(\|\overline{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}} \geq r) = o(1/n)$. Lemma 3.92 with $\mathcal{E} = \{\boldsymbol{\sigma}^- \in \mathcal{B}_-^{\Gamma}, \|\overline{\boldsymbol{\gamma}} - \gamma^*\|_{\mathrm{tv}} \leq r\}$ yields

$$\Phi_{\mathrm{v}}(n) = \mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}] + \mathcal{O}(n^{-\rho}),\ \boldsymbol{\Phi} = \ln(Z_{\mathrm{v}}(\boldsymbol{d}, \boldsymbol{\psi}^+, \boldsymbol{h}^+, \hat{\boldsymbol{\gamma}})),$$

using $|\boldsymbol{\Phi}| \leq \boldsymbol{d}\ln(\psi_{\uparrow})$ and independence. Using $(\boldsymbol{d}, \boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma})$ with $\boldsymbol{\sigma}^*$ replaced by $\boldsymbol{\sigma}^-$, resolving the Radon-Nikodym derivatives and reusing the terms $1/k$ to introduce $\boldsymbol{h}$ gives $\Phi_{\mathrm{v}}(n) = \mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}] + \mathcal{O}(n^{-\rho})$, where $\boldsymbol{\Phi} = \boldsymbol{r}\ln(Z_{\mathrm{v}}(\boldsymbol{d}, \boldsymbol{\psi}, \boldsymbol{h}, \boldsymbol{\gamma}))$ and

$$\begin{aligned}
\boldsymbol{r} &= \prod_{a \in [\boldsymbol{d}]} \sum_{\tau} \frac{\mathbb{1}\{\tau_{\boldsymbol{h}(a)} = \boldsymbol{\sigma}^\circ\} \overline{\psi}_\circ(\tau) \prod_{h \neq \boldsymbol{h}(a)} \boldsymbol{\gamma}^-(\tau_h)}{\overline{Z}_{\mathrm{f}}^+(\boldsymbol{\sigma}^\circ, \boldsymbol{\gamma}^-)} \cdot \frac{\boldsymbol{\psi}_a(\tau)}{\overline{\psi}_\circ(\tau)} \cdot \prod_{h \neq \boldsymbol{h}(a)} \frac{\boldsymbol{\gamma}_{a,h}(\tau_h)}{\overline{\boldsymbol{\gamma}}(\tau_h)} \\
&= \prod_{a \in [\boldsymbol{d}]} \sum_{\tau} \frac{\mathbb{1}\{\tau_{\boldsymbol{h}(a)} = \boldsymbol{\sigma}^\circ\} \boldsymbol{\psi}_a(\tau) \prod_{h \neq \boldsymbol{h}(a)} \boldsymbol{\gamma}_{a,h}(\tau_h) \prod_{h \neq \boldsymbol{h}(a)} \boldsymbol{\gamma}^-(\tau_h)}{\overline{Z}_{\mathrm{f}}^+(\boldsymbol{\sigma}^\circ, \boldsymbol{\gamma}^-) \prod_{h \neq \boldsymbol{h}(a)} \overline{\boldsymbol{\gamma}}(\tau_h)}.
\end{aligned}$$

As in the proof of Lemma 3.83 we have $\boldsymbol{\gamma}^-(\tau_h)/\gamma^*(\tau_h) = 1 + \mathcal{O}(r)$ and $\overline{\boldsymbol{\gamma}}(\tau_h)/\gamma^*(\tau_h) = 1 + \mathcal{O}(r)$ on $\mathcal{E}$. For $r(n) < \psi_{\downarrow}/2$ we have $\boldsymbol{\gamma}^- > 0$ and hence $\overline{Z}_{\mathrm{f}}^+(\boldsymbol{\sigma}^\circ, \boldsymbol{\gamma}^-) = \overline{Z}_{\mathrm{f}}(\boldsymbol{\gamma}^-)\boldsymbol{\mu}|_*(\boldsymbol{\sigma}^\circ)/\boldsymbol{\gamma}^-(\boldsymbol{\sigma}^\circ)$ with $\boldsymbol{\mu} = \mu_{\mathrm{T}|\Gamma,\boldsymbol{\gamma}^-}$ as pointed out after the definition of $\overline{Z}_{\mathrm{f}}^+$, above Equation (12). As in the proof of Lemma 3.83 with

Observation 3.9i) and Observation 3.9j) this yields $\overline{Z}_{\mathrm{f}}^{+}(\boldsymbol{\sigma}^\circ,\boldsymbol{\gamma}^-)/\xi=(1+\mathcal{O}(r^2))(1+\mathcal{O}(r))=1+\mathcal{O}(r)$. Hence, there exists $c_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $\boldsymbol{r}\le(1+cr)^{\boldsymbol{d}}\boldsymbol{r}^\circ$ and $\boldsymbol{r}^\circ\le(1+cr)^{\boldsymbol{d}}\boldsymbol{r}$, where

$$\boldsymbol{r}^\circ=\prod_{a\in[\boldsymbol{d}]}\sum_{\tau}\frac{\mathbb{1}\{\tau_{\boldsymbol{h}(a)}=\boldsymbol{\sigma}^\circ\}\boldsymbol{\psi}_a(\tau)\prod_{h\neq\boldsymbol{h}(a)}\boldsymbol{\gamma}_{a,h}(\tau_h)}{\xi}.$$

Now, with $\boldsymbol{\Phi}^\circ=\boldsymbol{r}^\circ\ln(Z_{\mathrm{v}}(\boldsymbol{d},\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}))$ and $|\boldsymbol{\Phi}^\circ|\le\ln(\psi_\uparrow)\boldsymbol{d}\psi_\uparrow^{2\boldsymbol{d}}$ we get

$$|\mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}]-\mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{\Phi}^\circ]|\le\mathbb{E}\left[\ln(\psi_\uparrow)\boldsymbol{d}\psi_\uparrow^{2\boldsymbol{d}}\left((1+cr)^{\boldsymbol{d}}-(1+cr)^{-\boldsymbol{d}}\right)\right]=\mathcal{O}(r).$$

Since $\ln(Z_{\mathrm{v}}(\boldsymbol{d},\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}))$ does not depend on $\boldsymbol{\sigma}^\circ$, the sum over $\boldsymbol{\sigma}^\circ$ with $\boldsymbol{\sigma}^-\sim\boldsymbol{\sigma}^*$ yields the assertion. □

Next, we show that we can replace $\pi_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ by its projection $\pi^\circ_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}$ by using Lemma 3.56. For this purpose we show that the variable contribution

$$B_{\mathrm{v}}:\mathcal{P}^2([q])\to\mathbb{R},\,\pi\mapsto\mathbb{E}\left[\xi^{-\boldsymbol{d}}\Lambda\left(Z_{\mathrm{v}}(\boldsymbol{d},\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}_\pi)\right)\right],$$
$$(\boldsymbol{d},\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}_\pi)\sim\mathrm{Po}(\bar{d})\otimes(\mu_\Psi\otimes\mathrm{u}([k])\otimes\pi^{\otimes k})^{\otimes\mathbb{Z}_{>0}},$$

to the Bethe functional is Lipschitz in π with respect to d_{w}.

Lemma 3.94. *There exists $L_{\mathfrak{g}}$ such that B_{v} is L-Lipschitz.*

Proof. Let $\pi_\circ\in\Gamma(\pi_1,\pi_2)$ be a coupling of $\pi\in\mathcal{P}^2([q])^2$. Further, let $(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2)\sim(\pi_\circ^{\otimes k})^{\otimes\mathbb{Z}_{>0}}$ with $\gamma_1,\gamma_2\in(\mathcal{P}([q])^k)^{\mathbb{Z}_{>0}}$. With $(\boldsymbol{d},\boldsymbol{\psi},\boldsymbol{h})$ from the definition of B_{v} let $(\boldsymbol{d},\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2)\sim\boldsymbol{d}\otimes\boldsymbol{\psi}\otimes\boldsymbol{h}\otimes(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2)$. With the Lipschitz continuity of Λ yields

$$\Delta=|B_{\mathrm{v}}(\pi_1)-B_{\mathrm{v}}(\pi_2)|\le\mathbb{E}\left[\psi_\uparrow^{\boldsymbol{d}}(\boldsymbol{d}\ln(\psi_\uparrow)+1)\,|Z_{\mathrm{v}}(\boldsymbol{d},\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}_1)-Z_{\mathrm{v}}(\boldsymbol{d},\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}_2)|\right].$$

With the triangle inequality, $\boldsymbol{\psi}\le\psi_\uparrow$ and Observation 3.6b) this gives

$$\begin{aligned}\Delta&\le\mathbb{E}\left[\psi_\uparrow^{2\boldsymbol{d}}(\boldsymbol{d}\ln(\psi_\uparrow)+1)\sum_{a\in[\boldsymbol{d}]}\sum_{h'\neq\boldsymbol{h}(a)}\|\boldsymbol{\gamma}_{1,a,h'}-\boldsymbol{\gamma}_{2,a,h'}\|_{\mathrm{tv}}\right]\\&=\mathbb{E}\left[\psi_\uparrow^{2\boldsymbol{d}}(\boldsymbol{d}\ln(\psi_\uparrow)+1)\boldsymbol{d}(k-1)\right]\mathbb{E}[\|\boldsymbol{\gamma}_{1,1,1}-\boldsymbol{\gamma}_{2,1,1}\|_{\mathrm{tv}}].\end{aligned}$$

This completes the proof since $\pi_\circ\in\Gamma(\pi_1,\pi_2)$ was arbitrary. □

Now, we finally obtain the asymptotics of Φ_{v}.

Lemma 3.95. *We have $\Phi_{\mathrm{v}}(n)=\mathbb{E}[B_{\mathrm{v}}(\pi^\circ_{\mathrm{g},\boldsymbol{G}^*})]+\mathcal{O}(n^{-\rho})$ with $\boldsymbol{G}^*=\boldsymbol{G}^*_{\boldsymbol{m},\boldsymbol{\mathcal{U}}}(\boldsymbol{\sigma}^*)$.*

Proof. With Lemma 3.93 we have $\Phi_{\mathrm{v}}(n)=\mathbb{E}[B_{\mathrm{v}}(\pi_{\mathrm{g},\boldsymbol{G}^*})]+\mathcal{O}(n^{-\rho})$. Lemma 3.94 and Lemma 3.56b) complete the proof, since $\mathrm{d}_{\mathrm{w}}\le q$ and $\mathbb{P}(\boldsymbol{m}>m_\uparrow)=o(1/n)$. □

3.3.3.17 Proof of Proposition 3.5. First, we establish Lemma 3.67 and Proposition 3.66. Then, we establish a stronger version of Proposition 3.5 for graphs with external fields.

Proof of Lemma 3.67. Lemma 3.67 follows from Lemma 3.85 and Lemma 3.95 with Equation (9). □

Proof of Proposition 3.66. With $\boldsymbol{G}^* = \boldsymbol{G}^*_{\boldsymbol{m},\mathcal{U}}(\boldsymbol{\sigma}^*)$ let $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ be such that $|\Phi_{\Delta,n} - \mathbb{E}[B(\pi^\circ_{\mathrm{g},\boldsymbol{G}^*})]| \le cn^{-\rho}$. Recall that $|\mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}^*_n)]| \le c'(\frac{d_\uparrow n}{k} + \frac{1}{2}n^{1-\rho})$ with c' from Observation 3.30 using Observation 3.47, and notice that $\Phi_{\Delta,0} = \mathbb{E}[n\phi_{\mathrm{g}}(\boldsymbol{G}^*_n)]$ for $n=1$. With $|B| \le \mathbb{E}[\psi_\uparrow^{\boldsymbol{d}}\Lambda(\psi_\uparrow^{\boldsymbol{d}})] + \bar{d}\psi_\uparrow\Lambda(\psi_\uparrow) \le c'$ for some $c'_{\mathfrak{g}} \in \mathbb{R}_{>0}$ the telescoping sum with the triangle inequality yields

$$\begin{aligned}|\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*)] - \mathbb{E}[B(\pi^\circ_{\mathrm{g},\boldsymbol{G}^*})]| &\le \mathcal{O}(n^{-1}) + \frac{c}{n}\sum_{n'=2}^{n-1} n'^{-\rho} = \mathcal{O}(n^{-1}) + \frac{c}{n}\int_1^{n-1}\lceil t^{-\rho}\rceil \mathrm{d}t \\ &\le \mathcal{O}(n^{-1}) + \frac{c}{n}\int_1^{n-1} t^{-\rho}\mathrm{d}t = \mathcal{O}(n^{-1}) + \frac{c((n-1)^{1-\rho}-1)}{(1-\rho)n},\end{aligned}$$

which shows that $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*)] = \mathbb{E}[B(\pi^\circ_{\mathrm{g},\boldsymbol{G}^*})] + \mathcal{O}(n^{-\rho})$ and thereby completes the proof. □

In the remainder we let $\boldsymbol{\pi} = \pi^\circ_{\mathrm{g},\boldsymbol{G}^*_{\boldsymbol{m},\mathcal{U}}(\boldsymbol{\sigma}^*)} \in \mathcal{P}^2_*([q])$ be the projected marginal distributions *including pins*. On the other hand, we let $\mathcal{U} = \emptyset$ in the remainder, where we also cover the case $\bar{d} = 0$. Now, we turn to Proposition 3.5 for graphs with external fields.

Proposition 3.96. *Notice that the following holds.*
a) We have $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))] = \mathbb{E}[B(\boldsymbol{\pi})] + \mathcal{O}(n^{-\rho})$.
b) For $d = km/n \le d_\uparrow$ *we have* $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_m(\boldsymbol{\sigma}^*))] = \mathbb{E}[B_d(\boldsymbol{\pi})] + \mathcal{O}(n^{-\rho})$.
c) We have $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}^*}(\boldsymbol{\sigma}^*))] = \mathbb{E}[B(\boldsymbol{\pi})] + \mathcal{O}(\delta_{\mathrm{m}} + \varepsilon_{\mathrm{m}} + n^{-\rho})$.

Proof. Proposition 3.66, Proposition 3.50 and $\Theta^\downarrow = n^{1-\rho}$ yield $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))] = \mathbb{E}[B(\boldsymbol{\pi})] + \mathcal{O}(n^{-\rho})$ for $\bar{d} > 0$. Recall from the proof of Proposition 3.65 that $\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*)) = \phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\hat{\boldsymbol{\sigma}}_{\boldsymbol{m}})) = 0$ and $B \equiv 0$ for $\bar{d} = 0$, so Part 3.96a) holds. The remainder follows similar to the proof of Proposition 3.65, but easier since the transition from $\hat{\boldsymbol{\sigma}}$ to $\boldsymbol{\sigma}^*$ is not required. □

Observation 3.21 yields the corresponding results for graphs without external fields and thereby completes the proof of Proposition 3.5.

3.3.3.18 Proof of Theorem 2.2. The following result for graphs with external fields implies Theorem 2.2. Recall ρ from Section 3.3.3 and that $m^\leftrightarrow \equiv 0$, $\mathcal{U} = \emptyset$, $t^\leftrightarrow = 1$ and $\Theta^\downarrow = 0$, i.e. we consider standard graphs with external fields only.

Theorem 3.97. *Notice that the following holds.*
a) We have $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}}(\boldsymbol{\sigma}^*))] = B_\uparrow(\bar{d}) + \mathcal{O}(n^{-\rho})$.
b) For $d = km/n \le d_\uparrow$ *we have* $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_m(\boldsymbol{\sigma}^*))] = B_\uparrow(d) + \mathcal{O}(n^{-\rho})$.
c) We have $\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_{\boldsymbol{m}^*}(\boldsymbol{\sigma}^*))] = B_\uparrow(\bar{d}) + \mathcal{O}(\delta_{\mathrm{m}} + \varepsilon_{\mathrm{m}} + n^{-\rho})$.

Proof. The assertion follows from Proposition 3.65, Proposition 3.96, $B_d \le B_\uparrow(d)$, and $\rho \in (0, 1/4)$ as discussed in the introduction of Section 3.3.3. □

Observation 3.21 yields the corresponding results for graphs without external fields and thereby completes the proof of Theorem 2.2.

3.4 Relative Entropy, Condensation and Mutual Information

In this section we derive Theorem 2.3, Theorem 2.4 and Theorem 2.5 from Theorem 3.97, for both graphs with and without external fields over more general factor counts $\boldsymbol{m}^*$. We also establish Lipschitz continuity in the average degree for all key quantities, i.e. the corresponding versions of Proposition 3.2c). Let $m^\leftrightarrow \equiv 0$, $\mathcal{U} = \emptyset$ and ρ from Section 3.3.3.

3.4.1 The Relative Entropy. We first discuss the annealed free entropy in Section 3.4.1.1, and then turn to the proof of Theorem 2.3 in Section 3.4.1.2.

3.4.1.1 The Annealed Free Entropy. In this section we briefly discuss the properties of the annealed free entropy. For this purpose recall $\phi_{\mathrm{a}}(d) = \frac{d}{k}\ln(\xi)$ from Section 2.1.2.5.

Observation 3.98. *Let* $\phi(m) = \frac{1}{n}\ln(\mathbb{E}[Z_{\mathrm{g}}(\boldsymbol{G})])$.
a) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* ϕ *is* kc/n*-Lipschitz and* $|\phi(m)| \le ckm/n$.
b) We have $\phi(m) = \phi_{\mathrm{a}}(km/n) + \mathcal{O}(1/n)$ *for* $m \le m_{\uparrow}$.
c) We have $\mathbb{E}[\phi(\boldsymbol{m}^*)] = \phi_{\mathrm{a}}(\bar{d}) + \mathcal{O}(\varepsilon_{\mathrm{m}} + \delta_{\mathrm{m}} + n^{-1})$*, so* $\mathbb{E}[\phi(\boldsymbol{m})] = \phi_{\mathrm{a}}(\bar{d}) + \mathcal{O}(\sqrt{\ln(n)/n})$.

Proof. With the proofs of Observation 3.30 and Lemma 3.31 we get $|\phi(m)| \le \frac{cm}{n}$ and $|\phi(m'_1) - \phi(m'_2)| \le \frac{c}{n}|m'_1 - m'_2|$ for $m' \in \mathbb{Z}^2_{\ge 0}$ and $c = \ln(\psi_{\uparrow})$. With Lemma 3.24a) we have $\phi(m) = \phi_{\mathrm{a}}(km/n) + \mathcal{O}(1/n)$ for $m \le m_{\uparrow}$. With Part 3.98a) and the expectation bound we have $\mathbb{E}[\phi(\boldsymbol{m}^*)] = \mathbb{E}[\mathbb{1}\{|\bar{\boldsymbol{d}}^* - \bar{d}| \le \delta_{\mathrm{m}}\}\phi(\boldsymbol{m}^*)] + \mathcal{O}(\varepsilon_{\mathrm{m}})$, so e.g. with Part 3.98b) and the probability bound we get $\mathbb{E}[\phi(\boldsymbol{m}^*)] = \phi_{\mathrm{a}}(\bar{d}) + \mathcal{O}(\varepsilon_{\mathrm{m}} + \delta_{\mathrm{m}} + n^{-1})$. The result for $\boldsymbol{m}$ then follows with Corollary 3.12 and $r = c'\sqrt{\ln(n)/n}$ for large c'. □

3.4.1.2 Proof of Theorem 2.3. The Nishimori ground truth establishes a finite size connection between the quenched free entropies, the annealed free entropy and the relative entropies.

Observation 3.99. *With* $\phi(m) = \frac{1}{n}\ln(\mathbb{E}[Z_{\mathrm{g}}(\boldsymbol{G})])$ *we have*

$$\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}))] = \phi(m) + D_{\mathrm{KL}}(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})\|\boldsymbol{G}) \ge \phi(m) - D_{\mathrm{KL}}(\boldsymbol{G}\|\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})].$$

Proof. Notice that Observation 3.22c) yields both $D_{\mathrm{KL}}(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})\|\boldsymbol{G}) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}))] - \phi(m)$ and $D_{\mathrm{KL}}(\boldsymbol{G}\|\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})) = \phi(m) - \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})]$. □

The asymptotics from Theorem 3.97 using Corollary 3.35 and from Observation 3.98 with the first equality in Observation 3.99 yield the asymptotics of $D_{\mathrm{KL}}(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})\|\boldsymbol{G})$. Now, we obtain Theorem 2.3 for graphs with external fields using the results of Section 3.2.2.

Theorem 3.100. *Let* $\delta(m) = \frac{1}{n}D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)\|\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}, \boldsymbol{G})$ *and* $\delta^*(d) = B_{\uparrow}(d) - \phi_{\mathrm{a}}(d)$.
a) We have $\delta(m) = \delta^*(km/n) + \mathcal{O}(n^{-\rho})$ *for* $km/n \le d_{\uparrow}$.
b) We have $\mathbb{E}[\delta(\boldsymbol{m}^*)] = \delta^*(\bar{d}) + \mathcal{O}(\varepsilon_{\mathrm{m}} + \delta_{\mathrm{m}} + n^{-\rho})$*, so* $\mathbb{E}[\delta(\boldsymbol{m})] = \delta^*(\bar{d}) + \mathcal{O}(n^{-\rho})$.

Proof. The Radon-Nikodym derivative of $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$ with respect to $(\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}, \boldsymbol{G})$ is

$$(\sigma, G) \mapsto \frac{\gamma^{*\otimes n}(\sigma)\psi_{\mathrm{g},G}(\sigma)Z_{\mathrm{g}}(G)}{\overline{\psi}_{\mathrm{m}}(\sigma)\psi_{\mathrm{g},G}(\sigma)} = \frac{Z_{\mathrm{g}}(G)}{\hat{r}(\sigma)\overline{Z}_{\mathrm{m}}},$$

and thereby $\delta(m) = \phi^*(m) - \frac{m}{n}\mathbb{E}[\ln(\overline{Z}_{\mathrm{f}}(\boldsymbol{\gamma}^*))] = \phi^*(m) - \phi(m) + \delta'(m)$ using Observation 3.11e) and with $\phi^*(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*))]$, $\phi(m) = \frac{1}{n}\ln(\overline{Z}_{\mathrm{m}})$ and $\delta'(m) = \frac{1}{n}D_{\mathrm{KL}}(\boldsymbol{\sigma}^*\|\hat{\boldsymbol{\sigma}})$. For Part 3.100a) we combine Theorem 3.97b) with Observation 3.98b) and Observation 3.29b). For Part 3.100b) we use Observation 3.29b), Observation 3.11e) and 3.11f) to obtain

$$0 \le \mathbb{E}[\delta'(\boldsymbol{m}^*)] \le \frac{c}{n} + \mathbb{E}\left[\mathbb{1}\{\boldsymbol{m}^* > m_{\uparrow}\}\frac{2\ln(\psi_{\uparrow})\boldsymbol{m}^*}{n}\right] = \mathcal{O}\left(\frac{1}{n} + \varepsilon_{\mathrm{m}}\right).$$

Now, the assertion follows with Theorem 3.97c), Observation 3.98c) and Corollary 3.12. □

Let $\boldsymbol{G}_{\circ,m}$, $\boldsymbol{G}^*_{\circ,m}(\boldsymbol{\sigma}^*) \in \mathcal{G}$ be the graphs without external fields from Section 2.1.2.1. We use $\boldsymbol{G} = \boldsymbol{G}_{\boldsymbol{m}^*}$, $\boldsymbol{G}_\circ = \boldsymbol{G}_{\circ,\boldsymbol{m}^*}$, $\boldsymbol{G}^*(\boldsymbol{\sigma}^*) = \boldsymbol{G}^*_{\boldsymbol{m}^*}(\boldsymbol{\sigma}^*)$ and $\boldsymbol{G}^*_\circ(\boldsymbol{\sigma}^*) = \boldsymbol{G}^*_{\circ,\boldsymbol{m}^*}(\boldsymbol{\sigma}^*)$. Recall $Z_{\gamma^*}(G)$ from Section 2.1.2.1 and $\boldsymbol{\sigma}_{\gamma^*,G}$ from Section 2.1.2.5 for $G \in \mathcal{G}$. The expectation in Theorem 3.100b) recovers

$$n\mathbb{E}[\delta(\boldsymbol{m}^*)] = D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*) \| \boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}, \boldsymbol{G}|\boldsymbol{m}^*) = D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*) \| \boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}, \boldsymbol{G}).$$

Let $r(\sigma, [G]^\Gamma) = Z_{\mathrm{g}}(G)/(\hat{r}(\sigma)\overline{Z}_{\mathrm{m}})$ be the Radon-Nikodym derivative of $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$ with respect to $(\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}, \boldsymbol{G})$ from the proof of Theorem 3.100. Further, let

$$r_\circ(\sigma, G) = \frac{\gamma^{*\otimes n}(\sigma)\psi_{\mathrm{g},G}(\sigma)Z_{\gamma^*}(G)}{\mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}_\circ(m)}(\sigma)]\gamma^{*\otimes n}(\sigma)\psi_{\mathrm{g},G}(\sigma)} = \frac{\gamma^{*\otimes n}(\sigma)Z_{\mathrm{g}}([G]^\Gamma)}{\overline{\psi}_{\mathrm{m}}(\sigma)} = r(\sigma, [G]^\Gamma)$$

be the Radon-Nikodym derivative of $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_\circ(\boldsymbol{\sigma}^*))$ with respect to $(\boldsymbol{\sigma}_{\gamma^*,\boldsymbol{G}_\circ}, \boldsymbol{G}_\circ)$. Combining this with Observation 3.21 completes the proof of Theorem 2.3 since

$$\begin{aligned} D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*) \| \boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}, \boldsymbol{G}) &= \mathbb{E}\left[\ln\left(r\left(\boldsymbol{\sigma}^*, [\boldsymbol{G}^*_\circ(\boldsymbol{\sigma}^*)]^\Gamma\right)\right)\right] = \mathbb{E}\left[\ln\left(r_\circ\left(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_\circ(\boldsymbol{\sigma}^*)\right)\right)\right] \\ &= D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_\circ(\boldsymbol{\sigma}^*) \| \boldsymbol{\sigma}_{\gamma^*,\boldsymbol{G}_\circ}, \boldsymbol{G}_\circ). \end{aligned}$$

3.4.2 The Condensation Threshold. In this section we establish Theorem 2.4. First, we show Theorem 2.4a) in Section 3.4.2.1, followed by the proof of Theorem 2.4b) in Section 3.4.2.2.

3.4.2.1 The Replica Symmetric Regime. Recall that $(\mu_\Psi, \gamma^*, \bar{d}) \in \mathfrak{P}_{\mathrm{r}}$ means that $B_\uparrow(\bar{d}) = \phi_{\mathrm{a}}(\bar{d})$.

Lemma 3.101. *Assume that $B_\uparrow(\bar{d}) = \phi_{\mathrm{a}}(\bar{d})$ and let $\phi(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})]$.*
a) We have $\phi(m) = \phi_{\mathrm{a}}(\bar{d}) + \mathcal{O}(n^{-\rho/2})$ if $\bar{d} = km/n$.
b) We have $\mathbb{E}[\phi(\boldsymbol{m}^*)] = \phi_{\mathrm{a}}(\bar{d}) + \mathcal{O}(\delta_{\mathrm{m}} + \varepsilon_{\mathrm{m}} + n^{-\rho/2})$, so $\mathbb{E}[\phi(\boldsymbol{m})] = \phi_{\mathrm{a}}(\bar{d}) + \mathcal{O}(n^{-\rho/2})$.

Proof. Using Theorem 3.97b), Corollary 3.35a) and $B_\uparrow(\bar{d}) = \phi_{\mathrm{a}}(\bar{d})$ let $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ be such that $|\hat{\phi}(m) - \phi_{\mathrm{a}}(\bar{d})| \le r$, where $\hat{\phi}(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}))]$ and $r = cn^{-\rho}$. With $\hat{c}$ from Lemma 3.38 and $\hat{\mathcal{E}} = \{|\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}) - \hat{\phi}(m)| < r\}$ we have $\mathbb{P}(\neg\hat{\mathcal{E}}) \le \hat{c}_2 \exp(-\hat{c}_1 n^{1-2\rho})$. Further, with

$$n_{\circ,\mathfrak{g}} = \left(\frac{\ln(2\hat{c}_2)}{\hat{c}_1}\right)^{1/(1-2\rho)}$$

we have $\mathbb{P}(\hat{\mathcal{E}}) \ge 1/2$ for $n \ge n_\circ$ (for $n \le n_\circ$ we use $|\phi(m) - \phi_{\mathrm{a}}(\bar{d})| \le \frac{1}{k}\ln(\psi_\uparrow)d_\uparrow n_\circ^{\rho/2} n^{-\rho/2}$). Notice that $\hat{\mathcal{E}}_{\mathrm{a}} = \{|\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}}) - \phi_{\mathrm{a}}(\bar{d})| < 2r\}$ holds on $\hat{\mathcal{E}}$ by the triangle inequality, so with $\mathcal{E}_{\mathrm{a}} = \{|\phi_{\mathrm{g}}(\boldsymbol{G}) - \phi_{\mathrm{a}}(\bar{d})| < 2r\}$, $\boldsymbol{Z} = Z_{\mathrm{g}}(\boldsymbol{G})\mathbb{1}\mathcal{E}_{\mathrm{a}}$ and $\overline{Z} = \mathbb{E}[\boldsymbol{Z}]$ Observation 3.22c) yields

$$\overline{Z} = \overline{Z}_{\mathrm{m}}\mathbb{E}\left[\frac{Z_{\mathrm{g}}(\boldsymbol{G})}{\overline{Z}_{\mathrm{m}}}\mathbb{1}\mathcal{E}_{\mathrm{a}}\right] = \overline{Z}_{\mathrm{m}}\mathbb{P}(\hat{\mathcal{E}}_{\mathrm{a}}) \ge \frac{1}{2}\overline{Z}_{\mathrm{m}}.$$

Further, we have $\boldsymbol{Z}^2 = \exp(2n\phi_{\mathrm{g}}(\boldsymbol{G}))\mathbb{1}\mathcal{E}_{\mathrm{a}} \le \exp(2n\phi_{\mathrm{a}}(\bar{d}) + 2rn) = e^{2rn}\overline{Z}_{\mathrm{m}}^2$, using $\mathcal{E}_{\mathrm{a}}$ and the definition of ϕ_{a}. Now, the Paley-Zygmund inequality yields

$$\mathbb{P}\left(\boldsymbol{Z} \ge \frac{1}{2}\overline{Z}\right) \ge \frac{\overline{Z}^2}{4\mathbb{E}[\boldsymbol{Z}^2]} \ge \frac{\overline{Z}_{\mathrm{m}}^2}{16\overline{Z}_{\mathrm{m}}^2}e^{-2rn} > 0.$$

Using $\boldsymbol{Z} \le Z_{\mathrm{g}}(\boldsymbol{G})$ and $\overline{Z} \ge \frac{1}{2}\overline{Z}_{\mathrm{m}}$ gives $Z_{\mathrm{g}}(\boldsymbol{G}) \ge \frac{1}{4}\overline{Z}_{\mathrm{m}}$ on $\boldsymbol{Z} \ge \frac{1}{2}\overline{Z}$, so

$$P = \mathbb{P}\left(\phi_{\mathrm{g}}(\boldsymbol{G}) \ge \phi_{\mathrm{a}}(\bar{d}) - \frac{\ln(4)}{n}\right) = \mathbb{P}\left(Z_{\mathrm{g}}(\boldsymbol{G}) \ge \frac{1}{4}\overline{Z}_{\mathrm{m}}\right) \ge \frac{\overline{Z}_{\mathrm{m}}^2}{16\overline{Z}_{\mathrm{m}}^2}e^{-2rn}\frac{1}{16}e^{-2rn} > 0.$$

Now, with c° from Lemma 3.36 and $r_\circ = \sqrt{\frac{1}{c_1^\circ n}\ln(\frac{2c_2^\circ}{P})}$ we have

$$\mathbb{P}\left(\phi_{\mathrm{g}}(\boldsymbol{G}) \ge \phi_{\mathrm{a}}(\bar{d}) - \frac{\ln(4)}{n}, |\phi_{\mathrm{g}}(\boldsymbol{G}) - \phi(m)| < r_\circ\right) \ge P - c_2^\circ e^{-c_1^\circ r_\circ^2 n} = \frac{1}{2}P > 0.$$

On this event we have $\phi(m) \ge \phi_{\mathrm{a}}(\bar{d}) - r_\circ - \frac{\ln(4)}{n}$, which establishes Part 3.101a) since $\phi_{\mathrm{a}}(\bar{d}) \ge \phi(m)$ by Observation 3.99 and Lemma 3.24a), and further

$$r_\circ + \frac{\ln(4)}{n} = \sqrt{\frac{2}{c_1^\circ}r + \frac{\ln(32c_2^\circ)}{c_1^\circ n}} + \frac{\ln(4)}{n} \le c'n^{-\rho/2},\ c' = \sqrt{\frac{2c + \ln(32c_2^\circ)}{c_1^\circ}} + \ln(4).$$

Observation 3.30 and Lemma 3.31 give $\mathbb{E}[\phi(\boldsymbol{m}^*)] = \phi(\lfloor \bar{d}n/k \rfloor) + \mathcal{O}(\delta_{\mathrm{m}} + \varepsilon_{\mathrm{m}} + n^{-1})$, so Part 3.101a) completes the proof. □

Observation 3.21 establishes Theorem 2.4a).

3.4.2.2 The Condensation Regime. As opposed to all other results, Theorem 2.4b) does not address asymptotics, only limits, so we don't discuss finite size approximations like Lemma 3.101a). Let

$$\phi_{\mathrm{q}\uparrow}(\bar{d}) = \limsup_{n\to\infty}\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}_{\boldsymbol{m}^*})],\ \phi_{\mathrm{q}\downarrow}(\bar{d}) = \liminf_{n\to\infty}\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}_{\boldsymbol{m}^*})].$$

Lemma 3.102. *There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that for $d \in [0, d_\uparrow]$ we have*

$$\phi_{\mathrm{a}}(d) - \phi_{\mathrm{q}\uparrow}(d) \ge c \sup_{d'\in[0,d]}(B_\uparrow(d') - \phi_{\mathrm{q}\downarrow}(d'))^2.$$

Proof. Observation 3.99, Observation 3.98c), Theorem 3.97c) and Corollary 3.35b) yield

$$B_\uparrow(d) \ge \phi_{\mathrm{a}}(d) \ge \phi_{\mathrm{q}\uparrow}(d) \ge \phi_{\mathrm{q}\downarrow}(d).$$

For $d' \in [0, d]$ with $\delta^*(d') = 0$, where $\delta^*(d) = B_\uparrow(d) - \phi_{\mathrm{a}}(d)$, we have $\phi_{\mathrm{q}\downarrow}(d') = \phi_{\mathrm{q}\uparrow}(d') = \phi_{\mathrm{a}}(d') = B_\uparrow(d')$ by Lemma 3.101b), and hence $\phi_{\mathrm{a}}(d) - \phi_{\mathrm{q}\uparrow}(d) \ge c(B_\uparrow(d') - \phi_{\mathrm{q}\downarrow}(d'))^2$ for all $c \in \mathbb{R}$. Hence, assume that $\delta^*(d') > 0$, let $m'_n = \lfloor d'n/k \rfloor$ and $m_n = \lfloor dn/k \rfloor$. Notice that $m'_n \le m_n \le d_\uparrow n/k$. Fix $\varepsilon \in (0, 1)$ with $\varepsilon < \delta^*(d)/2$ and let $\delta'(n) = \phi^*(m') - \phi(m')$ with $\phi^*(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*))]$ and $\phi(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})]$. With $\tilde{c}_{\mathfrak{g}}$ satisfying both Theorem 3.97c) and Observation 3.98c) for any small $\delta_{\mathrm{m}} \ge k/n$, $\varepsilon_{\mathrm{m}} \ge 0$ and using $n^{-\rho}$, let $n_\circ(\varepsilon) = (\varepsilon/\tilde{c})^{-\rho}$, so for $n \ge n_\circ(\varepsilon)$ we have $|\phi^*(m') - B_\uparrow(d')| \le \varepsilon$ and $|\bar{\phi}(m') - \phi_{\mathrm{a}}(d')| \le \varepsilon$, where $\bar{\phi}(m) = \frac{1}{n}\ln(\overline{Z}_{\mathrm{m}})$. This yields $\delta'(n) > 0$ since

$$\phi^*(m') \ge B_\uparrow(d') - \varepsilon > \phi_{\mathrm{a}}(d') + \varepsilon \ge \bar{\phi}(m') \ge \phi(m').$$

With c° from Lemma 3.36, c^* from Lemma 3.38, $\hat{c}$ from Corollary 3.25a) and the canonical coupling

$(\boldsymbol{G},\boldsymbol{G}')$ of $\boldsymbol{G}_m$ and $\boldsymbol{G}_{m'}$, meaning $\boldsymbol{G}'=R(\boldsymbol{G})$ with $R([w]^\Gamma)=[w_{[m']}]^\Gamma$, we have

$$\begin{aligned}P(n)&=\mathbb{P}(\phi_{\mathrm{g}}(\boldsymbol{G}')\le\phi(m')+\varepsilon\delta')\\&\le c_2^\circ e^{-c_1^\circ\varepsilon^2n}+\mathbb{P}\left(\phi_{\mathrm{g}}(\boldsymbol{G}')\le\phi(m')+\varepsilon\delta',|\phi_{\mathrm{g}}(\boldsymbol{G})-\phi(m)|<\varepsilon\right)\\&\le c_2^\circ e^{-c_1^\circ\varepsilon^2n}+\mathbb{E}\left[\frac{Z_{\mathrm{g}}(\boldsymbol{G})}{\exp(n(\phi(m)-\varepsilon))}\mathbb{1}\{\phi_{\mathrm{g}}(\boldsymbol{G}')\le\phi(m')+\varepsilon\delta'\}\right]\\&=c_2^\circ e^{-c_1^\circ\varepsilon^2n}+e^{n(\bar\phi(m)-\phi(m)+\varepsilon)}\mathbb{P}\left(\phi_{\mathrm{g}}(R(\boldsymbol{G}^*(\hat{\boldsymbol{\sigma}})))\le\phi(m')+\varepsilon\delta'\right),\end{aligned}$$

where we used Observation 3.22c). Observation 3.13 yields $R(\boldsymbol{G}^*(\boldsymbol{\sigma}^*))\sim\boldsymbol{G}^*_{m'}(\boldsymbol{\sigma}^*)$, so Corollary 3.25a) with Lemma 3.38 yields

$$\begin{aligned}P&\le c_2^\circ e^{-c_1^\circ\varepsilon^2n}+\hat ce^{n(\bar\phi(m)-\phi(m)+\varepsilon)}\mathbb{P}\left(\phi_{\mathrm{g}}(\boldsymbol{G}^*_{m'}(\boldsymbol{\sigma}^*)))\le\phi(m')+\varepsilon\delta'\right)\\&\le c_2^\circ e^{-c_1^\circ\varepsilon^2n}+\hat cc_2^*\exp(n\beta_\varepsilon(n)),\ \beta_\varepsilon(n)=\bar\phi(m)-\phi(m)+\varepsilon-c_1^*(1-\varepsilon)^2\delta'^2,\end{aligned}$$

where we used that $\phi(m')+\varepsilon\delta'=\phi^*(m')-(1-\varepsilon)\delta'$. For $\beta(\varepsilon)=\liminf_{n\to\infty}\beta_\varepsilon(n)$ taking the limits yields $\beta(\varepsilon)=\phi_{\mathrm{a}}(d)-\phi_{\mathrm{q}\uparrow}(d)-c_1^*(1-\varepsilon)^2(B_\uparrow(d')-\phi_{\mathrm{q}\downarrow}(d'))^2+\varepsilon$. On the other hand, Lemma 3.36 yields $P\ge1-c_2^\circ\exp(-n\beta'_\varepsilon(n))$ with $\beta'_\varepsilon(n)=c_1^\circ\varepsilon^2\delta'^2$. Since we assume $\delta^*(d)>0$, we have $\beta'(\varepsilon)=\liminf_{n\to\infty}\beta'_\varepsilon(n)=c_1^\circ\varepsilon^2(B_\uparrow(d')-\phi_{\mathrm{q}\uparrow}(d'))^2>0$. This shows that $\lim_{n\to\infty}P(n)=1$, which in turn yields $\beta(\varepsilon)\ge0$. Since β is a quadratic polynomial in ε, and in particular continuous, we have $\beta(0)\ge0$, so the assertion holds with c_1^*. □

Observation 3.21 establishes Theorem 2.4b).

3.4.2.3 The Condensation Threshold. Recall from Theorem 2.3 that $\delta^*(d)=B_\uparrow(d)-\phi_{\mathrm{a}}(d)\ge0$. Theorem 2.4b) implies that $\delta_\uparrow(d)=\phi_{\mathrm{a}}(d)-\phi_{\mathrm{q}\uparrow}(d)\ge0$, so in particular $\delta_\downarrow(d)=\phi_{\mathrm{a}}(d)-\phi_{\mathrm{q}\downarrow}(d)\ge\delta_\uparrow(d)\ge0$. Now, looking at Theorem 2.4b) through the eyes of the annealed free entropy gives

$$\delta_\uparrow(d)\ge c\sup_{d'\in[0,d]}(\delta^*(d')+\delta_\downarrow(d'))^2\ge c\sup_{d'\in[0,d]}(\delta^*(d')+\delta_\uparrow(d'))^2.$$

With Theorem 2.4a) suggests that $\phi_{\mathrm{q}\downarrow}(d)=\phi_{\mathrm{q}\uparrow}(d)=\phi_{\mathrm{a}}(d)=B_\uparrow(d)$ for $d\in[0,d_{\mathrm{cond}})$, since $\delta^*(d)=0$, and in particular $(p,\gamma^*,d)\in\mathfrak{P}_{\mathrm{r}}$. For $d\in(d_{\mathrm{cond}},\infty)$ there exists $d'\in[d_{\mathrm{cond}},d)$ such that $\delta^*(d')>0$, so Theorem 2.4b) suggests that $\delta_\uparrow(d)>0$. But then Theorem 2.4a) requires that $\delta^*(d)>0$ and thus $(p,\gamma^*,d)\in\mathfrak{P}_{\mathrm{c}}$. Since $\delta^*(0)=0$ and δ^* is Lipschitz, we have $(p,\gamma^*,d_{\mathrm{cond}})\in\mathfrak{P}_{\mathrm{r}}$.

3.4.3 The Mutual Information. We turn to the proof of the last main result. We show that the mutual information for graphs with external fields converges to $\iota^*(d)=\frac{d}{k\xi}\mathbb{E}[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}))]-B_\uparrow(d)$ from Theorem 2.5, and then obtain Theorem 2.5 as a corollary.

Theorem 3.103. *Let $\iota(m)=\frac{1}{n}I(\boldsymbol{\sigma}^*,\boldsymbol{G}^*(\boldsymbol{\sigma}^*))$.*
a) We have $\iota(m)=\iota^*(km/n)+\mathcal{O}(n^{-\rho})$ for $km/n\le d_\uparrow$.
b) We have $\mathbb{E}[\iota(\boldsymbol{m}^*)]=\iota^*(\bar d)+\mathcal{O}(\varepsilon_{\mathrm{m}}+\delta_{\mathrm{m}}+n^{-\rho})$, so $\mathbb{E}[\iota(\boldsymbol{m})]=\iota^*(\bar d)+\mathcal{O}(n^{-\rho})$.

We prove Theorem 3.103 in three parts. For this purpose recall the notions from Section 3.3.1.1 and $\boldsymbol{\sigma}^*_{\mathrm{g}}$ from Section 3.2.2.4. First, we split ι into three contributions, the ground truth entropy $H(\gamma^*)$, the conditional cross entropy $\overline{\eta}(m)=\mathbb{E}[\mathbb{E}[H(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}\|\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)})|\boldsymbol{G}^*(\boldsymbol{\sigma}^*)]]$ and the conditional relative entropy $\overline{\delta}(m)=\mathbb{E}[\mathbb{E}[D_{\mathrm{KL}}(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)}\|\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*(\boldsymbol{\sigma}^*)})|\boldsymbol{G}^*(\boldsymbol{\sigma}^*)]]$.

Lemma 3.104. *We have $\iota(m) = H(\gamma^*) - \overline{\eta}(m) + \overline{\delta}(m)$.*

The proof is presented in Section 3.4.3.1. Then we determine the limit of $\overline{\eta}$.

Lemma 3.105. *Notice that the following holds.*
a) We have $\overline{\eta}(m) = H(\gamma^*) - \iota^*(km/n) + \mathcal{O}(n^{-\rho})$ for $km/n \le d_\uparrow$.
b) We have $\mathbb{E}[\overline{\eta}(\boldsymbol{m}^*)] = H(\gamma^*) - \iota^*(\bar{d}) + \mathcal{O}(\varepsilon_{\mathrm{m}} + \delta_{\mathrm{m}} + n^{-\rho})$.

The proof is presented in Section 3.4.3.2. Finally, we complete the proof of Theorem 3.103 in Section 3.4.3.3, where we also establish Theorem 2.5.

3.4.3.1 The Entropy Decomposition. Using $\boldsymbol{G}^* = \boldsymbol{G}^*(\boldsymbol{\sigma}^*)$, recall that $(\boldsymbol{\sigma}^*, \boldsymbol{G}^*) \sim (\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*}, \boldsymbol{G}^*)$ from Observation 3.27b), so we have $n\iota(m) = D_{\mathrm{KL}}(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*} \| \boldsymbol{\sigma} | \boldsymbol{G}^*)$, using $(\boldsymbol{\sigma}, \boldsymbol{G}^*) \sim \boldsymbol{\sigma}^* \otimes \boldsymbol{G}^*$, by the chain rule of the relative entropy. The decomposition into the (conditional) cross entropy and the entropy gives $n\iota(m) = H(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*} \| \boldsymbol{\sigma} | \boldsymbol{G}^*) - H(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*} | \boldsymbol{G}^*)$. Using linearity of the cross entropy in the first component and independence, we can take the expectation over $\boldsymbol{G}^*$ to obtain $H(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*} \| \boldsymbol{\sigma} | \boldsymbol{G}^*) = H(\boldsymbol{\sigma}^*)$ since $\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*} \sim \boldsymbol{\sigma}^*$. We split the latter entropy into the cross entropy and the relative entropy with respect to $\boldsymbol{\sigma}_{\mathrm{g}}$, yielding $H(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*} | \boldsymbol{G}^*) = H(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*} \| \boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*} | \boldsymbol{G}^*) - D_{\mathrm{KL}}(\boldsymbol{\sigma}^*_{\mathrm{g},\boldsymbol{G}^*} \| \boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*} | \boldsymbol{G}^*)$, and hence $\iota(m) = H(\gamma^*) - \overline{\eta}(m) + \overline{\delta}(m)$.

3.4.3.2 The Cross Entropy Contribution. By the definition of the cross entropy and $\boldsymbol{\sigma}_{\mathrm{g},G}$ we have $H(\boldsymbol{\sigma}^*_{\mathrm{g},G} \| \boldsymbol{\sigma}_{\mathrm{g},G}) = \mathbb{E}[-\ln(\psi_{\mathrm{g},G}(\boldsymbol{\sigma}^*_{\mathrm{g},G})/Z_{\mathrm{g}}(G))]$, so Observation 3.27b) yields

$$\overline{\eta}(m) = \mathbb{E}\left[\phi_{\mathrm{g}}(\boldsymbol{G}^*)\right] - \mathbb{E}\left[\frac{1}{n}\ln\left(\psi_{\mathrm{g},\boldsymbol{G}^*}(\boldsymbol{\sigma}^*)\right)\right].$$

Unlike the partition function Z_{g}, the weight $\psi_{\mathrm{g},\boldsymbol{G}^*(\sigma)}(\sigma) \sim \gamma^{*\otimes n}(\sigma)\prod_{a\in[m]}\boldsymbol{\psi}^*_a(\sigma_{\boldsymbol{v}^*(a)})$ factorizes, where $(\boldsymbol{v}^*, \boldsymbol{\psi}^*) \sim \boldsymbol{w}^{*\otimes m}_{\circ,\sigma}$, and hence $\overline{\eta}(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*)] + H(\gamma^*) - \frac{m}{n}\mathbb{E}[\ln(\boldsymbol{\psi}^*(\boldsymbol{\sigma}^*_{\boldsymbol{v}^*}))]$, where $(\boldsymbol{v}^*, \boldsymbol{\psi}^*) \sim \boldsymbol{w}^*_{\circ,\boldsymbol{\sigma}^*}$. Resolving the Radon-Nikodym derivative yields

$$\overline{\eta}(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*)] + H(\gamma^*) - \frac{m}{n}\mathbb{E}\left[\frac{\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}^*_{\boldsymbol{v}}))}{\overline{Z}_{\mathrm{f}}(\gamma^*)}\right]$$

with $(\boldsymbol{\sigma}^*, \boldsymbol{v}, \boldsymbol{\psi}) \sim \gamma^{*\otimes n} \otimes \mathrm{u}([n]^k) \otimes \mu_\Psi$. Hence, with Observation 3.9d), Observation 3.23b) and Theorem 3.97b) we obtain Part 3.105a), since $\boldsymbol{\sigma}^*_{\boldsymbol{v}} \sim \boldsymbol{\sigma}$. For Part 3.105b) we notice that $|\mathbb{E}[\bar{\boldsymbol{d}}^*] - \bar{d}| \le \delta_{\mathrm{m}} + d_\uparrow \varepsilon_{\mathrm{m}}$, hence Observation 3.9d), Observation 3.23b) and Theorem 3.97c) complete the proof.

3.4.3.3 Proof of Theorem 2.5. Part 3.103a) is immediate from Lemma 3.104, Lemma 3.105a) and Observation 3.29c). Part 3.103b) follows from Lemma 3.104, Lemma 3.105b), Observation 3.29c) and the expectation bound for the relative entropy and $\boldsymbol{m}^* > m_\uparrow$, since the proof of Observation 3.29 reveals that the $(\boldsymbol{\sigma}_{\mathrm{g},G}, \boldsymbol{\sigma}^*_{\mathrm{g},G})$-derivative is $r_G(\sigma) = \psi_{\mathrm{g},G}(\sigma)/(\gamma^{*\otimes n}(\sigma) Z_{\mathrm{g}}(G))$, thereby establishing $|\ln(r_G(\sigma))| \le 2m\ln(\psi_\uparrow)$ and further $D_{\mathrm{KL}}(\boldsymbol{\sigma}^*_{\mathrm{g},G} \| \boldsymbol{\sigma}_{\mathrm{g},G}) \le 2m\ln(\psi_\uparrow)$ for $G \in \mathcal{G}$. This completes the proof of Theorem 3.103. Theorem 2.5 follows with Observation 3.21 and analogously to the derivation of Theorem 2.3 from Theorem 3.100.

3.5 Additional Discussion

In Section 3.5.1 we discuss the remaining cases $k = 1$ and $q = 1$, as well the treatment of ground truths that are not fully supported. In Section 3.5.2 we translate the main results to graphs with external fields. Different representations of B and ∇ (cf. Section 3.3.2) are discussed in Section 3.5.3 that are both instructive and of formal importance. Lipschitz continuity, boundedness and convergence in probability for the target quantities is discussed in Section 3.5.4. The closure properties for **POS** are discussed in Section 3.5.5, followed by the proof of Proposition 2.1 in Section 3.5.6. In Section 3.5.7, we translate the main results to all related standard models, that is, various versions of the uniform and the binomial model. There, we also discuss (un-) balanced problems for the general case. The SBM is then discussed in Section 3.5.9, the spin glass version in Section 3.5.10, and graphical channels are treated in Section 3.5.11. Finally, we discuss some open problems in Section 3.5.12.

3.5.1 Embeddings and Projections. In this section we argue that $k, q \in \mathbb{Z}_{>0}$ can be chosen arbitrarily large, which in particular covers the case $q = 1$, and the case $k = 1$.

Observation 3.106. *We may choose k and q arbitrarily large and assume that γ^* is fully supported without loss of generality.*

Proof. These transformations with respect to the assumptions will be discussed in Section 3.5.1 in greater detail. Details on why it is sufficient to consider the weights ψ_{g} for the finite size models can be found in Section 3.5.7. For a given model (μ_Ψ, γ^*) such that the support of γ^* is $[q']$ with $q' \leq q$ consider the model given by the restriction $\boldsymbol{\psi}' : [q']^k \to \mathbb{R}_{>0}$, $\sigma \mapsto \boldsymbol{\psi}(\sigma)$, with law μ'_Ψ, suggesting that the restriction $\gamma^* \in \mathcal{P}([q'])$ is now fully supported. Notice that the partition functions and Gibbs measures of both the null model and teacher-student model are not affected because $[q]^n \setminus [q']^n$ is a null set under $\gamma^{*\otimes n}$, hence the left hand sides in the results are not affected. We observe that the limiting quantities, i.e. the Bethe functional, the annealed free entropy density and the limiting mutual information are also invariant to this restriction.

For $k' \geq k$, $q' \geq q$ let the projection $f : [q'] \to [q]$ be given by $f(\sigma) = \sigma$ for $\sigma \in [q]$ and $f(\sigma) = q$ otherwise. Further, let $\boldsymbol{\psi}' : [q']^{k'} \to \mathbb{R}_{>0}$, $\sigma \mapsto \boldsymbol{\psi}((f(\sigma_h))_{h \in [k]})$, so $\boldsymbol{\psi}'$ does not depend on the additional coordinates, and only depends on the color class $\mathcal{C} = \{\sigma :\in [q'] : \sigma \geq q\}$. Let $\gamma' \in \mathcal{P}([q'])$ be given by $\gamma'(\sigma) = \gamma^*(\sigma)$ for $\sigma < q$ and by $\gamma'(\sigma) = \gamma^*(q)/(q' - q + 1)$ for $\sigma \geq q$. It is sufficient to observe that the weights do not depend on the additional coordinates, thus its distribution is the same under both null models, hence the reweighting for the teacher-student model is not affected, and thereby the teacher-student model weight function is not affected. The coefficient for the Poisson factor count $\boldsymbol{m}$ changes, which results in the consideration of the average degree $k\bar{d}/k'$. This exactly corresponds to the rescaling of the limiting quantities. □

3.5.2 External Fields. In this section we follow up on the discussion of graphs with external fields in Section 3.1.3. For $\eta : [q] \to \mathbb{R}_{>0}$ and $G = (v, \psi) \in \mathcal{G}$ let $[G]^\Gamma_\eta = G' = (v'_a, \psi'_a)_{a \in \mathcal{A}}$ with $\mathcal{A} = [m] \,\dot\cup\, [n]$ be given by $G'_{[m]} = G$, $G'_a = (a, \eta)$ for $a \in [n]$, i.e. a graph with fixed external fields. Let $\boldsymbol{G}_{\mathrm{e}} = [\boldsymbol{w}]^\Gamma_\eta$ with $\boldsymbol{w}$ from Section 3.2.1.1 and let $\boldsymbol{G}^*_{\mathrm{e}}(\sigma)$ be given by the $(\boldsymbol{G}^*_{\mathrm{e}}(\sigma), \boldsymbol{G}_{\mathrm{e}})$-derivative $G \mapsto \psi_{\mathrm{g},G}(\sigma)/\mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}_{\mathrm{e}}}(\sigma)]$. Let $\mathfrak{P}_{\mathrm{e}} = \{(\mu_\Psi, \gamma^*, d, c\gamma^*) : (\mu_\Psi, \gamma^*, d) \in \mathfrak{P}, c \in \mathbb{R}_{>0}\}$ be the parameters including external fields. We reduce the general case to normalized external fields via $\phi^\circ_{\mathrm{g}}([G]^\Gamma_\eta) = \phi_{\mathrm{g}}([G]^\Gamma_\eta) - \ln(\|\eta\|_1)$.

Corollary 3.107. *Theorem 3.97, Theorem 3.100, Lemma 3.101, Lemma 3.102 and Theorem 3.103 hold for $\boldsymbol{G}$, $\boldsymbol{G}^*$, ϕ_{g} replaced by $\boldsymbol{G}_{\mathrm{e}}$, $\boldsymbol{G}^*_{\mathrm{e}}$, ϕ°_{g} and $(\mu_\Psi, \gamma^*, d, \eta) \in \mathfrak{P}_{\mathrm{e}}$.*

Proof. We have $\|c\gamma^*\|_1 = c$ and hence $\phi^\circ_{\mathrm{g}}([G]^\Gamma_{c\gamma^*}) = \frac{1}{n}\ln(c^n Z_{\mathrm{g}}([G]^\Gamma_{\gamma^*})) - \ln(c) = \phi_{\mathrm{g}}([G]^\Gamma_{\gamma^*})$. Similarly, notice that $\psi_{\mathrm{g},[G]^\Gamma_{c\gamma^*}}(\sigma) = c^n \psi_{\mathrm{g},[G]^\Gamma_{\gamma^*}}(\sigma)$, so the Radon-Nikodym derivatives coincide in this sense and analogously to Observation 3.21, and thereby $\boldsymbol{G}^*_{\mathrm{e}}(\sigma) \sim [\boldsymbol{w}^*(\sigma)]^\Gamma_{c\gamma^*}$. The remainder is analogous to the translation of the results to graphs without external fields. Notice that the results are uniform over $c \in \mathbb{R}_{>0}$. □

3.5.3 Reweighting and Relative Entropies. In this section we build some context for ∇ from Section 2.1.2.2 and B from Section 2.1.2.4. As opposed to the proofs, for the theory in this section we exclusively consider the restrictions to $\mathcal{P}^2_*([q])$ with $\overline{Z}_{\mathrm{f}}(\gamma^*)=\xi$ for $\gamma^*\in\mathcal{P}([q])$, i.e. we require γ^* to be a maximizer of $\overline{Z}_{\mathrm{f}}$.

Let $r_{\mathrm{f}}:\mathcal{D}_\Psi\times\mathcal{P}([q])^k\to\mathbb{R}_{>0}$, $(\psi,\gamma)\mapsto Z_{\mathrm{f}}(\psi,\gamma)/\xi$. Further, for $\sigma\in[q]$ and $\pi\in\mathcal{P}^2_*([q])$ with $\mathcal{R}=\{(\psi,h,\gamma):\psi\in\mathcal{D}_\Psi,h\in[k],\gamma\in\mathcal{P}([q])^{[k]\setminus\{h\}}\}$ let

$$r_{\mathrm{v},\sigma}:\mathcal{R}\to\mathbb{R}_{>0},\ (\psi,h,\gamma)\mapsto\frac{1}{\xi}\sum_\tau\mathbb{1}\{\tau_h=\sigma\}\psi(\tau)\prod_{h'\neq h}\gamma_{h'}(\tau_{h'}).$$

For $\pi\in\mathcal{P}^2_*([q])^2$ let $\boldsymbol{x}_i\sim\mu_\Psi\otimes\pi_i^{\otimes k}$, $i\in[2]$, further $\boldsymbol{h}\sim\mathrm{u}([k])$ and for $h\in[k]$ let $\boldsymbol{x}_{3,h}\sim\mu_\Psi\otimes\bigotimes_{h'\in[k]}\pi_{3,h'}$ with $\pi_{3,h}=\pi_1$ and $\pi_{3,h'}=\pi_2$ for $h'\in[k]\setminus\{h\}$. Let $\boldsymbol{x}^*_1$, $\boldsymbol{x}^*_2$, $\boldsymbol{x}^*_{3,h}$ be given by the Radon-Nikodym derivative r_{f} with respect to $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_{3,h}$ respectively, and

$$\nabla_2(\pi_1,\pi_2)=\xi(\mathbb{E}[\ln(Z_{\mathrm{f}}(\boldsymbol{x}^*_1))]+(k-1)\mathbb{E}[\ln(Z_{\mathrm{f}}(\boldsymbol{x}^*_2))]-k\mathbb{E}[\ln(Z_{\mathrm{f}}(\boldsymbol{x}^*_{3,\boldsymbol{h}}))]),$$
$$\nabla_3(\pi_1,\pi_2)=\xi(D_{\mathrm{KL}}(\boldsymbol{x}^*_1\|\boldsymbol{x}_1)+(k-1)D_{\mathrm{KL}}(\boldsymbol{x}^*_2\|\boldsymbol{x}_2)-kD_{\mathrm{KL}}(\boldsymbol{x}^*_{3,\boldsymbol{h}}\|\boldsymbol{x}_{3,\boldsymbol{h}}|\boldsymbol{h})).$$

For $\pi\in\mathcal{P}^2_*([q])$ let $\boldsymbol{x}_{\mathrm{f}}\sim\mu_\Psi\otimes\pi^{\otimes k}$, $\boldsymbol{x}_{\mathrm{v},\circ}=(\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}_{[k]\setminus\{\boldsymbol{h}\}})$, where $(\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma})\sim\mu_\Psi\otimes\mathrm{u}([k])\otimes\pi^{\otimes k}$, let $\boldsymbol{x}^*_{\mathrm{f}}$ be given by the Radon-Nikodym derivative r_{f}, and let $\boldsymbol{x}^*_{\mathrm{v},\circ,\sigma}$ be given by the Radon-Nikodym derivative $r_{\mathrm{v},\sigma}$ for $\sigma\in[q]$. Further, let $(\boldsymbol{d},\boldsymbol{x}_{\mathrm{v}})\sim\mathrm{Po}(d)\otimes\boldsymbol{x}_{\mathrm{v},\circ}^{\otimes\mathbb{Z}_{>0}}$, $\boldsymbol{x}^*_{\mathrm{v},\sigma}\sim\boldsymbol{x}^{*\otimes\mathbb{Z}_{>0}}_{\mathrm{v},\circ,\sigma}$ and $\boldsymbol{\sigma}^*\sim\gamma^*$ with $(\boldsymbol{d},\boldsymbol{\sigma}^*,\boldsymbol{x}^*_{\mathrm{v},\boldsymbol{\sigma}^*})\sim\boldsymbol{d}\otimes(\boldsymbol{\sigma}^*,\boldsymbol{x}^*_{\mathrm{v},\boldsymbol{\sigma}^*})$. Finally, let $\boldsymbol{X}_{\mathrm{v}}=\boldsymbol{x}_{\mathrm{v},[\boldsymbol{d}]}$, $\boldsymbol{X}^*_{\mathrm{v}}=\boldsymbol{x}^*_{\mathrm{v},\boldsymbol{\sigma}^*,[\boldsymbol{d}]}$, $Z_{\mathrm{v}}(\psi_{[d]},h_{[d]},(\gamma_{a,h'})_{a\in[d],h'\neq h(a)})=Z_{\mathrm{v}}(d,\psi,h,\gamma)$ and

$$B_{2,d}(\pi)=\mathbb{E}\left[\ln\left(Z_{\mathrm{v}}(\boldsymbol{X}^*_{\mathrm{v}})\right)\right]-\frac{d(k-1)}{k}\mathbb{E}[[\ln\left(Z_{\mathrm{f}}(\boldsymbol{x}^*_{\mathrm{f}})\right)],$$
$$B_{3,d}(\pi)=\phi_{\mathrm{a}}(d)+D_{\mathrm{KL}}(\boldsymbol{X}^*_{\mathrm{v}}\|\boldsymbol{X}_{\mathrm{v}})-\frac{d(k-1)}{k}D_{\mathrm{KL}}(\boldsymbol{x}^*_{\mathrm{f}}\|\boldsymbol{x}_{\mathrm{f}}).$$

Lemma 3.108. *We have $\nabla=\nabla_2=\nabla_3$ and $B=B_2=B_3$.*

Proof. For $\pi\in\mathcal{P}^2_*([q])^k$ and $(\boldsymbol{\psi},\boldsymbol{\gamma})\sim\mu_\Psi\otimes\bigotimes_h\pi_h$ we have $\mathbb{E}[Z_{\mathrm{f}}(\boldsymbol{\psi},\boldsymbol{\gamma})]=\overline{Z}_{\mathrm{f}}(\gamma^*)=\xi$, which shows that $\boldsymbol{x}^*_1$, $\boldsymbol{x}^*_2$, $\boldsymbol{x}^*_{3,h}$ for ∇ and $\boldsymbol{x}^*_{\mathrm{f}}$ for B are well-defined. Let $\mathcal{C}=\gamma^{*-1}(\mathbb{R}_{>0})$ be the support of γ^* and $\mu=\mu_{\mathrm{T}|\Gamma,\gamma^*}$ from Section 3.2.1.2. For $|\mathcal{C}|=1$ we have $\mu|_*=\gamma^*$ since both are necessarily one-point masses on the only element of $\mathcal{C}$, otherwise we have $\mu|_*=\gamma^*$ by Observation 3.9j) (since γ^* is a fully supported stationary point of $\overline{Z}_{\mathrm{f}}$ on $\mathcal{P}(\mathcal{C})$). Hence, for $\pi\in\mathcal{P}^2_*([q])^k$ and $\sigma\in[q]$ with $(\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma})\sim\mu_\Psi\otimes\mathrm{u}([k])\otimes\bigotimes_h\pi_h$ we have

$$\mathbb{E}\left[r_{\mathrm{v},\sigma}\left(\boldsymbol{\psi},\boldsymbol{h},\boldsymbol{\gamma}_{[k]\setminus\{\boldsymbol{h}\}}\right)\right]=\frac{1}{\xi}\sum_h\frac{1}{k}\sum_\tau\mathbb{1}\{\tau_h=\sigma\}\overline{\psi}_\circ(\tau)\prod_{h'\neq h}\gamma^*(\tau_{h'})=\frac{\mu|_*(\sigma)}{\gamma^*(\sigma)}=1.$$

This shows that $\boldsymbol{X}^*_{\mathrm{v}}$ is well-defined, and hence the assertion clearly holds. □

3.5.4 Lipschitz Continuity and Boundedness. In this section we stress the relevant properties that allow to extend the main results to $\boldsymbol{m}^*$ and the equivalence of various modes of convergence.

Let $\bar{\phi}_n(m)=\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}_m)]$, $\phi^*_{\sigma,n}(m)=\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_m(\sigma))]$ and $\bar{\phi}^*_n(m)=\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*_m(\boldsymbol{\sigma}^*))]$.

Lemma 3.109. *Notice that the following holds.*
a) There exists $c_{\mathfrak{g}}\in\mathbb{R}_{>0}$ such that $|\bar{\phi}(m)|\leq c\frac{km}{n}$. The same holds for $\bar{\phi}$ replaced by $\phi^*_\sigma,\bar{\phi}^*$.

b) There exists $L_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $|\bar{\phi}(m_1) - \bar{\phi}(m_2)| \le L\left|\frac{km_1}{n} - \frac{km_2}{n}\right|$ *for* $m \in \mathbb{Z}_{\ge 0}^2$*. The same holds for* $\bar{\phi}$ *replaced by* $\phi^*_\sigma, \bar{\phi}^*$*.*

Proof. Part 3.109a) follows from Observation 3.30 and Observation 3.47. For Part 3.109b) assume that $m_1 \le m_2$ and let $\boldsymbol{G}^\circ_m$ be $\boldsymbol{G}_m$ or $\boldsymbol{G}^*_m(\sigma)$. Under the canonical coupling (using Observation 3.13) we obtain $\boldsymbol{G}^\circ(m_2)$ from $\boldsymbol{G}^\circ(m_1)$ given $\boldsymbol{G}^\circ(m_1)$ by adding $m_2 - m_1$ factors with pairs drawn iid from the underlying wires-weight pair distribution, then with c from Observation 3.30 we have

$$\left|\mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^\circ_{m_2})] - \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^\circ_{m_1})]\right| \le L\,|m_2 - m_1|\,,\; L = kc.$$

The result for $\bar{\phi}^*$ now follows from $\bar{\phi}^*(m) = \mathbb{E}[\phi^*_{\boldsymbol{\sigma}^*}(m)]$ and Jensen's inequality. □

On the finite size side let $\bar{d}_n(m) = km/n$. Lemma 3.109 suggests that e.g. $|\bar{\phi}(m)| \le c\bar{d}(m)$ and $|\bar{\phi}(m_1) - \bar{\phi}(m_2)| \le L|\bar{d}(m_1) - \bar{d}(m_2)|$. Next, we show that in general under these two properties convergence in probability, convergence of the expectation and pointwise convergence with respect to $m^\circ_n = \lfloor \bar{d}n/k \rfloor$ coincide for $\boldsymbol{m}^*$. Let $\mathcal{F}_\circ = \mathbb{R}^{\mathbb{Z}_{>0}\times\mathbb{Z}_{\ge 0}} = \{f : \mathbb{Z}_{>0} \times \mathbb{Z}_{\ge 0} \to \mathbb{R}\}$, and for $c \in \mathbb{R}_{>0}$ let $\mathcal{F}_c = \mathcal{F}_{\mathrm{b},c} \cap \mathcal{F}_{\mathrm{l},c}$ with

$$\begin{aligned}\mathcal{F}_{\mathrm{b},c} &= \left\{f \in \mathcal{F}_\circ : \forall n \forall m\, |f(n,m)| \le c\bar{d}_n(m)\right\},\\ \mathcal{F}_{\mathrm{l},c} &= \left\{f \in \mathcal{F}_\circ : \forall n \forall m \le m_\uparrow\, |f(n,m_1) - f(n,m_2)| \le c|\bar{d}_n(m_1) - \bar{d}_n(m_2)|\right\}.\end{aligned}$$

Lemma 3.110. *For* $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$*,* $f \in \mathcal{F}_c$*,* $f^* : \mathbb{R}_{\ge 0} \to \mathbb{R}$ *and* $g_{\mathfrak{g}} : \mathbb{Z}_{>0} \to \mathbb{R}_{\ge 0}$ *with* $g(n) = o(1)$ *the following statements are equivalent. Let* $\Delta_{\mathfrak{g}}(n) = g(n) + \delta_{\mathrm{m}}(n) + \varepsilon_{\mathrm{m}}(n) + n^{-1}$*.*
a) We have $|f_n(m^\circ) - f^*(\bar{d})| = \mathcal{O}(\Delta(n))$*.*
b) There exists $C_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\mathbb{P}(|f_n(\boldsymbol{m}^*) - f^*(\bar{d})| > C\Delta(n)) \le \varepsilon_{\mathrm{m}}(n)$*.*
c) We have $|\mathbb{E}[f_n(\boldsymbol{m}^*)] - f^*(\bar{d})| = \mathcal{O}(\Delta(n))$*.*

Proof. Using $E(n) = \mathbb{E}[f_n(\boldsymbol{m}^*)]$ we have

$$\begin{aligned}E(n) &\le \mathbb{E}[\mathbb{1}\{|\bar{\boldsymbol{d}}^* - \bar{d}| \le \delta_{\mathrm{m}}\}f(\boldsymbol{m}^*)] + c\varepsilon_{\mathrm{m}} \le \mathbb{E}[\mathbb{1}\{|\bar{\boldsymbol{d}}^* - \bar{d}| \le \delta_{\mathrm{m}}\}f(m^\circ)] + c\delta_{\mathrm{m}} + c\varepsilon_{\mathrm{m}}\\ &\le f(m^\circ) + c\bar{d}(m^\circ)\varepsilon_{\mathrm{m}} + c\delta_{\mathrm{m}} + c\varepsilon_{\mathrm{m}} \le f(m^\circ) + 2cd_\uparrow\Delta\end{aligned}$$

and analogously $E(n) \ge f(m^\circ) - 2cd_\uparrow\Delta$, which shows that the statements 3.110a) and 3.110c) are equivalent. Now, assume that 3.110a) holds and let $C'_{\mathfrak{g}}$ be such that $|f(m^\circ) - f^*(\bar{d})| \le C'\Delta$. By the triangle inequality we have $|f(m) - f^*(\bar{d})| \le c'|\bar{d}(m) - \bar{d}(m^\circ)| + C'\Delta \le (ck + C')\Delta$ if $|\bar{d}(m) - \bar{d}| \le \delta_{\mathrm{m}}$, since then $|\bar{d}(m) - \bar{d}(m^\circ)| \le \delta_{\mathrm{m}} + \frac{k}{n} \le k\Delta$ by the triangle inequality, so 3.110b) holds with $C = ck + C'$. Conversely, let C' be the constant from 3.110b) and $n_{\circ,\mathfrak{g}}$ such that $\varepsilon_{\mathrm{m}} < 1/2$ for all $n \ge n_\circ$. In this case we have

$$\mathbb{P}(|f_n(\boldsymbol{m}^*) - f^*(\bar{d})| \le C'\Delta(n), |\bar{\boldsymbol{d}}^* - \bar{d}| \le \delta_{\mathrm{m}}) \ge 1 - 2\varepsilon_{\mathrm{m}} > 0,$$

so there exists m with $|f(m) - f^*(\bar{d})| \le C'\Delta(n)$ and $|\bar{d}(m) - \bar{d}| \le \delta_{\mathrm{m}}$. As above, the triangle inequality and Lipschitz continuity give $|f(m^\circ) - f^*(\bar{d})| \le C'\Delta + ck\Delta$. By taking the limit this shows that $|f^*(\bar{d})| \le c\bar{d}$, so for $n \le n_\circ$ we have $|f(m^\circ) - f^*(\bar{d})| \le 2c\bar{d} \le 2cd_\uparrow n_\circ\Delta$ and thereby Part 3.110a) holds with $C = \max(C' + ck, 2cd_\uparrow n_\circ)$. □

We only verify that Lemma 3.110 holds for the target functions $\bar{\phi}$, $\bar{\phi}^*$, further $\bar{\phi}_{\mathrm{a}}(m) = \frac{1}{n}\ln(\overline{Z}_{\mathrm{m}})$, $\iota(m) = \frac{1}{n}I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$ and $\delta(m) = \frac{1}{n}D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)\|\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}, \boldsymbol{G})$.

Lemma 3.111. *There exists* $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\bar{\phi}, \bar{\phi}^*, \bar{\phi}_{\mathrm{a}}, \iota, \delta \in \mathcal{F}_c$.

Proof. The assertion for $\bar{\phi}, \bar{\phi}^*$ follows from Lemma 3.109. The assertion for $\bar{\phi}_{\mathrm{a}}$ is Observation 3.98a). For ι we recall that $\iota = \frac{1}{n} D_{\mathrm{KL}}(\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}^*} \| \boldsymbol{\sigma} | \boldsymbol{G}^*)$ from Section 3.4.3.1 with $(\boldsymbol{\sigma}, \boldsymbol{G}^*) \sim \boldsymbol{\sigma}^* \otimes \boldsymbol{G}^*(\boldsymbol{\sigma}^*)$ and $\boldsymbol{\sigma}_{\mathrm{g}}$ from Section 3.2.2.4 given by the $(\boldsymbol{\sigma}_{\mathrm{g},G}, \boldsymbol{\sigma})$-derivative $r_{\mathrm{s},G}$, so with Observation 3.27a) we have $\psi_{\downarrow}^{4m} \le r_{\mathrm{s},G} \le \psi_{\uparrow}^{4m}$ for $G \in \mathcal{G}$ and thereby $|\iota| \le 4\ln(\psi_{\uparrow}) m/n$. Further, for $G \in \mathcal{G}_m$ and an extension $G' \in \mathcal{G}_{m+1}$ we have $\psi_{\downarrow}^4 r_{\mathrm{s},G} \le r_{\mathrm{s},G'} \le \psi_{\uparrow}^4 r_{\mathrm{s},G}$ so $|\iota(m_1) - \iota(m_2)| \le 4k\ln(\psi_{\uparrow})|\bar{d}(m_1) - \bar{d}(m_2)|$ using the canonical coupling of $\boldsymbol{G}^*_{m_1}(\boldsymbol{\sigma}^*)$ and $\boldsymbol{G}^*_{m_2}(\boldsymbol{\sigma}^*)$ from the proof of Lemma 3.109. For δ we recall the derivative $r(G) = Z_{\mathrm{g}}(G)/(\hat{r}(\sigma)\overline{Z}_{\mathrm{m}}) = \gamma^{*\otimes n}(\sigma) Z_{\mathrm{g}}(G)/\overline{\psi}_{\mathrm{m}}(\sigma)$ from the proof of Theorem 3.100, and notice that $\psi_{\downarrow}^2 r(G) \le r(G') \le \psi_{\uparrow}^2 r(G)$ for an extension $G' \in \mathcal{G}_{m+1}$ of $G \in \mathcal{G}_m$, so $|\delta(m_1) - \delta(m_2)| \le 2k\ln(\psi_{\uparrow})|\bar{d}(m_1) - \bar{d}(m_2)|$, and $|\delta(m)| \le 2k\ln(\psi_{\uparrow})\bar{d}(m)$. □

Remark 3.112. The combination of Lemma 3.111 and Lemma 3.110 yields all main results for graphs with and without external fields for the modes of convergence in Lemma 3.110.

For the limiting quantities, assuming $\overline{Z}_{\mathrm{f}}(\gamma^*) = \xi$, let

$$\mathcal{F}_c = \{f : \mathbb{R}_{\ge 0} \to \mathbb{R} : \forall d |f(d_1) - f(d_2)| \le c|d_1 - d_2|, f(0) = 0\},$$

further $\phi_{\mathrm{a}}(d) = d\ln(\xi)/k$, $\iota^*(d)$ from Section 3.4.3 and $\delta^*(d) = B_{\uparrow}(d) - \phi_{\mathrm{a}}(d)$.

Lemma 3.113. *There exists* $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $B_\pi, B_{\uparrow}, \phi_{\mathrm{a}}, \iota^*, \delta^* \in \mathcal{F}_c$.

Proof. With Lemma 3.108 we have

$$|B_d(\pi)| = |B_{2,d}(\pi)| \le \mathbb{E}[\bar{\boldsymbol{d}} \ln(\psi_{\uparrow})] + \frac{d(k-1)}{k}\ln(\psi_{\uparrow}) = d,\ c = \frac{(2k-1)\ln(\psi_{\uparrow})}{k}.$$

For $d_2 \ge d_1$ we use the canoncial coupling of $\bar{\boldsymbol{d}}_i \sim \mathrm{Po}(d_i)$, $i \in [2]$, to obtain

$$|B_{2,d_1}(\pi) - B_{2,d_2}(\pi)| \le (d_2 - d_1)\ln(\psi_{\uparrow}) + (d_2 - d_1)\frac{k-1}{k}\ln(\psi_{\uparrow}) = c|d_2 - d_1|.$$

This also yields $|B_{\uparrow}(d)| \le cd$ and $B_{\uparrow}(d_1) - B_{\uparrow}(d_2)| \le c|d_1 - d_2|$, where the former is obvious and the latter follows by considering maximizing sequences $(\pi_{1,n})_n$, $(\pi_{2,n})_n$ to obtain

$$B_{\uparrow}(d_1) = \lim_{n\to\infty} B_{d_1}(\pi_{1,n}) \le \lim_{n\to\infty} B_{d_2}(\pi_{1,n}) + c|d_2 - d_1| \le B_{\uparrow}(d_2) + c|d_2 - d_1|$$

and the analogous result by switching 1 and 2 in the above. The result for ϕ_{a} is immediate, which directly implies the result for δ^*. The result for ι^* follows from the result for $B_{\uparrow}$ and the immediate result $d\mathbb{E}[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}))]/(k\xi) \in \mathcal{F}_c$ for $c = \ln(\psi_{\uparrow})/k$ using that $(\sigma, \psi) \mapsto \psi(\sigma)/\xi$ is a Radon-Nikodym derivative for $(\boldsymbol{\sigma}, \boldsymbol{\psi})$ since $\mathbb{E}[\boldsymbol{\psi}(\boldsymbol{\sigma})] = \overline{Z}_{\mathrm{f}}(\gamma^*) = \xi$. □

Specifically for the Bethe functional we also recall the Lipschitz continuity in π.

Lemma 3.114. *There exists* $L_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $B_d : \mathcal{P}^2([q]) \to \mathbb{R}$ *is L-Lipschitz if* $d \le d_{\uparrow}$. *Hence, there exists* $\pi \in \mathcal{P}^2_*([q])$ *such that* $B_d(\pi) = B_{\uparrow}(d)$.

Proof. Lipschitz continuity follows from Lemma 3.94 and Lemma 3.84. Recall from [33] that $\mathcal{P}^2([q])$ is a compact Polish space (Corollary 2.2.5, Theorem 2.2.7 and Proposition 2.2.8 in [104]), notice that $\mathcal{P}^2_*([q]) \subseteq \mathcal{P}^2([q])$ is closed and hence compact, so by the extreme value theorem the maximum is attained. □

3.5.5 The Assumption POS. We discuss the difficult assumption **POS** in detail. Many of the properties discussed below directly translate to other parts, say, from **POS** to **BAL**, from ∇ (cf. Section 3.3.2) to B, or even to all quantities (cf. Observation 3.106).

3.5.5.1 Extended Definition. Equip $\mathcal{P}([q])$ with $\|\cdot\|_{\mathrm{tv}}$, $\mathcal{D}_\Psi = \mathbb{R}_{\ge 0}^{[q]^k}$ with $\|\cdot\|_\infty$, and both with the Borel algebra, which defines $\mathcal{P}^2([q])$ and $\mathcal{P}(\mathcal{D}_\Psi)$. Fix $\gamma^* \in \mathcal{P}([q])$ throughout the remainder, recall $Z_{\mathrm{fm}}(\psi, h, \gamma) = \sum_y \psi(y)\gamma_{1,h}(y_h)\prod_{h'\neq h}\gamma_{2,h'}(y_{h'})$ and $Z_{\mathrm{f}}(\psi, \gamma_1) = Z_{\mathrm{fm}}(\psi, h, (\gamma_1, \gamma_1))$ for $\psi \in \mathcal{D}_\Psi$, $h \in [k]$ and $\gamma \in \mathcal{P}([q])^{2\times k}$. Notice that

$$\nabla_\circ : D_\Psi \times \mathcal{P}([q])^{2\times k} \to \mathbb{R},\ (\psi, \gamma) \mapsto \Lambda(Z_{\mathrm{f}}(\psi, \gamma_1)) + (k-1)\Lambda(Z_{\mathrm{f}}(\psi, \gamma_2)) - \sum_{h=1}^k \Lambda(Z_{\mathrm{fm}}(\psi, h, \gamma))$$

is well-defined and continuous. For fixed $p \in \mathcal{P}(D_\Psi)$, $\pi \in \mathcal{P}^2([q])$ let $(\boldsymbol{\psi}, \boldsymbol{\gamma}) \sim p \otimes \pi_1^{\otimes k} \otimes \pi_2^{\otimes k}$. Notice that $\min_y \psi(y) \le Z_{\mathrm{fm}}(\psi, \cdot, \cdot) \le \|\psi\|_\infty$ and that $\Lambda \ge -1/e$, hence it is both sufficient and convenient to consider

$$\mathcal{D}_{\mathrm{p}} = \{p \in \mathcal{P}(\mathcal{D}_\Psi) : \mathbb{E}\left[\Lambda(\|\boldsymbol{\psi}\|_\infty)\right] < \infty\}.$$

This does not only ensure that $\nabla : \mathcal{D}_{\mathrm{p}} \times \mathcal{P}^2([q])^2 \to \mathbb{R}$, $(p, \pi) \mapsto \mathbb{E}[\nabla_\circ(\boldsymbol{\psi}, \boldsymbol{\gamma})]$, is well-defined, but also that the expectations of the contributions are. Using the tower property and independence we have $\nabla(p, \pi) = \mathbb{E}[\nabla_\bullet(\boldsymbol{\psi}, \pi)]$ with $\nabla_\bullet : \mathcal{D}_\Psi \times \mathcal{P}^2([q])^2 \to \mathbb{R}$, $(\psi, \pi) \mapsto \mathbb{E}[\nabla_\circ(\psi, \boldsymbol{\gamma})]$. Finally, let $\nabla_\downarrow : \mathcal{D}_{\mathrm{p}} \times \mathcal{P}([q]) \to \mathbb{R}$, $(p, \gamma) \mapsto \inf_{\pi \in \mathcal{P}^2_{*,\gamma}([q])^2} \nabla(p, \pi)$, be our target function, and let $\mathfrak{A} = \nabla_\downarrow^{-1}(\mathbb{R}_{\ge 0})$ be the pairs satisfying **POS**. Let $\mathfrak{P}(\gamma^*) = \{p : (p, \gamma^*) \in \mathfrak{A}\} = \{p \in \mathcal{D}_{\mathrm{p}} : \nabla_\downarrow(p, \gamma^*) \ge 0\}$ be the weights for γ^*. Throughout the remainder we fix $p \in \mathcal{D}_{\mathrm{p}}$, $\gamma^* \in \mathcal{P}([q])$ and $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])^2$ unless mentioned otherwise.

3.5.5.2 Basic Observations. In this section we draw easy conclusions that help to build some intuition, to identify special cases and to get rid of pathological cases. Thus, let $p_0 \in \mathcal{D}_{\mathrm{p}}$ be the one-point mass on the trivial weight, i.e. $p_0(\mathcal{E}) = \mathbb{1}\{(0)_\sigma \in \mathcal{E}\}$. For $p \in \mathcal{D}_{\mathrm{p}} \setminus \{p_0\}$ let $p^\circ \in \mathcal{P}(\mathcal{D}_\Psi)$ be given by $p^\circ(\mathcal{E}) = p(\mathcal{E} \setminus \{(0)_\sigma\})/(1 - p(\{(0)_\sigma\}))$, i.e. the law of $\boldsymbol{\psi}|\boldsymbol{\psi} \neq (0)_\sigma$. Let $\mathcal{D}_{\mathrm{p}}^\circ = \{p \in \mathcal{D}_{\mathrm{p}} : p(\{(0)_\sigma\}) = 0\}$ and $\mathfrak{A}^\circ = \{(p, \gamma^*) : p \in \mathcal{D}_{\mathrm{p}}^\circ, \nabla_\downarrow(p, \gamma^*) \ge 0\}$.

Observation 3.115. *Notice that the following holds.*

a) We have $(p_0, \gamma^*) \in \mathfrak{A}$, and for $p \in \mathcal{D}_{\mathrm{p}} \setminus \{p_0\}$ we have $P = \mathbb{P}(\|\boldsymbol{\psi}\|_\infty > 0) > 0$. Further, we have $\mathbb{E}[\Lambda(\boldsymbol{\psi})] = P\mathbb{E}[\Lambda(\boldsymbol{\psi})|\|\boldsymbol{\psi}\|_\infty > 0]$ and $\nabla(p, \pi) = P\nabla(p^\circ, \pi)$.

b) For $\psi \in \mathcal{D}_\Psi$, $\gamma \in \mathcal{P}([q])^k$ we have $\nabla_\circ(\psi, (\gamma_i)_i) = 0$, hence $\nabla_\bullet(\psi, (\pi, \pi)) = 0$ for all $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])$ and thus $\nabla_\downarrow \le 0$.

Proof. For $\psi = (0)_\sigma$ all Z_{fm} vanish, thus $\nabla_\circ$ does, so $\nabla_\downarrow(p_0, \gamma^*) = 0$ and hence $(p_0, \gamma^*) \in \mathfrak{A}$. For $p \neq p_0$ we have $\mathbb{E}[\Lambda(\boldsymbol{\psi})] = \mathbb{E}[\mathbb{1}\{\|\boldsymbol{\psi}\|_\infty > 0\}\Lambda(\boldsymbol{\psi})] = \mathbb{P}(\|\boldsymbol{\psi}\|_\infty > 0)\mathbb{E}[\Lambda(\boldsymbol{\psi})|\|\boldsymbol{\psi}\|_\infty > 0]$, and of course $\mathbb{P}(\|\boldsymbol{\psi}\|_\infty > 0) > 0$, so the left hand side is finite iff the right hand side is. The same argumentation for ∇ yields $\nabla(p, \pi) = \mathbb{P}(\|\boldsymbol{\psi}\|_\infty > 0)\nabla(p^\circ, \pi)$, which completes the proof of the first part. For the second part, notice that $Z_{\mathrm{fm}}(\psi, h, \gamma) = Z_{\mathrm{f}}(\psi, \gamma_2)$ if $\gamma_{1,h} = \gamma_{2,h}$, and the rest follows. □

Now, we may restrict to $\mathfrak{A}^\circ$ and $\mathcal{D}_{\mathrm{p}}^\circ$, and know that $\mathfrak{A} = \nabla_\downarrow^{-1}(\{0\})$.

3.5.5.3 POS without POS and Products. In this section we establish the special role of product weights, and provide pairs for which ∇ is trivially non-negative. Fix $\psi \in \mathcal{D}_\Psi$, $\gamma \in \mathcal{P}([q])^k$ and let $\alpha = \bigotimes_h \gamma_h$. If we have $Z = Z_{\mathrm{f}}(\psi, \gamma) > 0$, let $p_{\mathrm{s}} \in \mathcal{P}([q]^k)$ be given by $p_{\mathrm{s}}(\sigma) = \alpha(\sigma)\psi(\sigma)/Z$, and let $\boldsymbol{\sigma} \sim p_{\mathrm{s}}$. Let $\pi^{\circ}_{\gamma^*} \in \mathcal{P}^2_{*,\gamma^*}([q])$ be the measure supported on the one-point masses, i.e. given by $\pi^{\circ}_{\gamma^*}(\mathcal{E}) = \sum_\tau \gamma^*(\tau)\mathbb{1}\{\delta_\tau \in \mathcal{E}\}$, where $\delta_\tau \in \mathcal{P}([q])$ is given by $\delta_\tau(\tau) = 1$. Further, let $\pi^{\bullet}_{\gamma^*} \in \mathcal{P}^2_{*,\gamma^*}([q])$ be the one-point mass on γ^*, i.e. $\pi^{\bullet}_{\gamma^*}(\mathcal{E}) = \mathbb{1}\{\gamma^* \in \mathcal{E}\}$.

Observation 3.116. *Let $\pi' \in \mathcal{P}^2_{*,\gamma^*}([q])$, $\pi = (\pi^{\circ}_{\gamma^*}, \pi')$ and $\gamma \in \mathcal{P}([q])^k$.*
a) We have $\mathbb{E}[\nabla_\circ(\psi, \boldsymbol{\gamma}_1, \gamma)] = Z(\sum_h H(\boldsymbol{\sigma}_h) - H(\boldsymbol{\sigma})) \geq 0$ with equality if and only if $\boldsymbol{\sigma} \sim \bigotimes_h \boldsymbol{\sigma}_h$.
b) Let $\pi' = \pi^{\bullet}_{\gamma^*}$. Then we have $\nabla(p, \pi) = 0$ if and only if $\boldsymbol{\psi}$ is a product almost surely.
c) For all $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])$ we have $\nabla_\bullet(\psi, \pi) = 0$ if ψ is a product.

Proof. For the first part we notice that

$$\mathbb{E}[\nabla_\circ(\psi, \boldsymbol{\gamma}_1, \gamma)] = \sum_\sigma \alpha(\sigma)\Lambda(\psi(\sigma)) + (k-1)\Lambda(Z) - \sum_h \sum_\tau \gamma^*(\tau)\Lambda(\overline{\psi}_h(\tau)),$$

where $\overline{\psi}_h : [q] \to \mathbb{R}_{\geq 0}$, $\tau \mapsto \mathbb{E}[\psi(\boldsymbol{\sigma}^*)|\boldsymbol{\sigma}^*_h = \tau]$, and $\boldsymbol{\sigma}^* \sim \alpha$. Now, notice that if $Z = 0$, then $\alpha(\sigma)\psi(\sigma) = 0$ for all σ and hence the expectation is 0, which coincides with the right hand side. For $Z > 0$ we expand the definition of Λ, the leading Z, $\overline{\psi}_h$, and rearrange the sums to obtain

$$\mathbb{E}[\nabla_\circ(\psi, \boldsymbol{\gamma}_1, \gamma)] = \sum_\sigma \alpha(\sigma)\psi(\sigma)\left[\ln(\psi(\sigma)) + (k-1)\ln(Z) - \sum_h \ln(\overline{\psi}_h(\sigma_h))\right].$$

Notice that we can restrict to σ such that both α and ψ are non-trivial. Next, we distribute Z and introduce appropriate masses to obtain

$$\begin{aligned}\mathbb{E}[\nabla_\circ(\psi, \boldsymbol{\gamma}_1, \gamma)] &= Z\left[D_{\mathrm{KL}}(\boldsymbol{\sigma}\|\boldsymbol{\sigma}^*) - \sum_h D_{\mathrm{KL}}(\boldsymbol{\sigma}_h\|\boldsymbol{\sigma}^*_h)\right] \\ &= Z\left[H(\boldsymbol{\sigma}\|\boldsymbol{\sigma}^*) - \sum_h H(\boldsymbol{\sigma}_h\|\boldsymbol{\sigma}^*_h) + \sum_h H(\boldsymbol{\sigma}_h) - H(\boldsymbol{\sigma})\right] = Z\left[\sum_h H(\boldsymbol{\sigma}_h) - H(\boldsymbol{\sigma})\right].\end{aligned}$$

Since the sum of the coordinate entropies is at least the entropy, with equality iff the coordinates are independent, the result follows. For the second part we have $\nabla_\bullet(\psi, \pi) = \mathbb{E}[\nabla_\circ(\psi, \boldsymbol{\gamma}_1, \gamma^*)]$. For $f \in (\mathbb{R}^q_{\geq 0})^k$ let $[f] : [q]^k \to \mathbb{R}_{\geq 0}$, $\sigma \mapsto \prod_h f_h(\sigma_h)$ be the product. Since γ^* is fully supported we have $\psi \equiv 0$ if and only if $Z = 0$. For $Z > 0$ there exists f with $\psi = [f]$ if and only if $\boldsymbol{\sigma} \sim \bigotimes_h \boldsymbol{\sigma}_h$, for the following reasons. If there exists f, then we have $Z = \prod_h Z_h$, $Z_h = \sum_\tau f_h(\tau)\gamma^*(\tau)$, and thus $p_{\mathrm{s}}(\sigma) = \prod_h \frac{\gamma^*(\sigma_h) f_h(\sigma_h)}{Z_h}$. Conversely, if the law factorizes into some $(p_h)_h$, then we have $\frac{\psi(\sigma)}{Z}\prod_h \gamma^*(\sigma_h) = \prod_h p_h(\sigma_h)$, so $\psi(\sigma) = \prod_h \frac{Z^{1/k} p_h(\sigma_h)}{\gamma^*(\sigma_h)}$ factorizes, using that γ^* is fully supported. But we have $\boldsymbol{\sigma} \sim \bigotimes_h \boldsymbol{\sigma}_h$ if and only if $\sum_h H(\boldsymbol{\sigma}_h) - H(\boldsymbol{\sigma}) = 0$, so we can rewrite $\{\exists f \boldsymbol{\psi} = [f]\} = \{\nabla_\bullet(\boldsymbol{\psi}, \pi) = 0\}$. Since continuity is not abundantly clear, we may replace the entropy condition by the total variation distance (half the 1-norm), in which case continuity in ψ essentially follows from the definition. This shows that $\{\exists f\, \boldsymbol{\psi} = [f]\}$ is indeed an event, the event on which $\nabla_\bullet$ vanishes and $\boldsymbol{\psi}$ factorizes, while otherwise $\nabla_\bullet$ is positive and $\boldsymbol{\psi}$ does not factorize. So, $\nabla(p, \pi)$ vanishes if and only if $\mathbb{P}(\exists f\, \boldsymbol{\psi} = [f]) = 1$.

For $\psi(\sigma) = \prod_h \psi_h(\sigma_h)$, we have $Z_{\mathrm{fm}}(\psi, h, \gamma) = z(\psi_h, \gamma_{1,h})\prod_{h' \neq h} z(\psi_{h'}, \gamma_{2,h'})$, where $z(\psi', \gamma') =$

$\sum_\tau \gamma'(\tau)\psi'(\tau)$. Using $s_{i,h} = \mathbb{E}[\Lambda(\psi, h, \boldsymbol{\gamma}_{i,1})]$, $z_h = z(\psi_h, \gamma^*)$ this yields

$$\nabla_\bullet(\psi,\pi) = \sum_h s_{1,h} \prod_{h'\neq h} z_{h'} + (k-1)\sum_h s_{2,h} \prod_{h'\neq h} z_{h'} - \sum_h \left(s_{1,h} \prod_{h'\neq h} z_{h'} + \sum_{h'\neq h} s_{2,h'} z_h \prod_{h''\notin\{h,h'\}} z_{h''} \right)$$

and thereby $\nabla_\bullet(\psi,\pi) = 0$. □

This result suggests that $\nabla(p,\pi) \leq 0$ for all $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])^2$ is equivalent to $\nabla(p,\pi) = 0$ for all π (because of $\pi = (\pi^\circ, \pi^\bullet)$), which is also equivalent to $\boldsymbol{\psi}$ being a product.

Also, the conditional expectation vanishes on products, so $\nabla(p,\pi) = \mathbb{P}(\mathcal{E})\nabla(p^\circ,\pi)$, where $\mathcal{E}$ is the event that $\boldsymbol{\psi}$ is not a product and $p^\circ(\mathcal{F}) = p(\mathcal{F}\cap\mathcal{E})/p(\mathcal{E})$ is the law of $\boldsymbol{\psi}|\mathcal{E}$. Notice that $p^\circ \in \mathcal{D}_\mathrm{p}$ whenever p is, so as for the trivial weight, we may now assume $\mathcal{E}$ almost surely. Finally, notice that all weights for $k=1$ are products.

Remark 3.117. Observation 3.116 can be extended. For $\boldsymbol{\gamma} \in \mathcal{P}([q])$ with $\mathbb{E}[\boldsymbol{\gamma}] = \gamma^*$ and a σ-algebra $\mathcal{F}' \subseteq \mathcal{F}$ (cf. Section 3.1.6) let $\boldsymbol{\gamma}' = \mathbb{E}[\boldsymbol{\gamma}|\mathcal{F}']$ and notice that $\mathbb{E}[\boldsymbol{\gamma}'] = \gamma^*$. It can be shown that $\nabla(p,\pi,\pi') \geq 0$, where $\boldsymbol{\gamma} \sim \pi$ and $\boldsymbol{\gamma}' \sim \pi'$.

3.5.5.4 Scaling Invariance. In this section we discuss $\boldsymbol{a}\boldsymbol{\psi}$ for given $(\boldsymbol{a},\boldsymbol{\psi})$. Let

$$\Delta_\circ : D_\Psi \times \mathcal{P}([q])^{2\times k} \to \mathbb{R},\ (\psi,\gamma) \mapsto Z_\mathrm{f}(\psi,\gamma_1) + (k-1)Z_\mathrm{f}(\psi,\gamma_2) - \sum_{h=1}^{k} Z_\mathrm{fm}(\psi,h,\gamma).$$

Observation 3.118. *Let $a \in \mathbb{R}_{\geq 0}$, $(\boldsymbol{a},\boldsymbol{\psi}) \in \mathbb{R}_{\geq 0}\times\mathcal{D}_\Psi$, $\boldsymbol{\psi} \sim p \in \mathcal{D}_\mathrm{p}$, $\boldsymbol{a}\boldsymbol{\psi} \sim p' \in \mathcal{D}_\mathrm{p}$ and $\mathbb{E}[\boldsymbol{a}] < \infty$.*
a) We have $\nabla_\circ(a\psi,\gamma) = a\nabla_\circ(\psi,\gamma) + \Delta_\circ(\psi,\gamma)\Lambda(a)$ and $\nabla_\bullet(a\psi,\pi) = a\nabla_\bullet(\psi,\pi)$.
b) We have $\nabla(p',\pi) = \mathbb{E}[\boldsymbol{a}]\nabla(p'',\pi)$, where $p'' \in \mathcal{D}_\mathrm{p}$ is given by the p''-p derivative $\mathbb{E}[\boldsymbol{a}|\boldsymbol{\psi}]/\mathbb{E}[\boldsymbol{a}]$.

Proof. Notice that $\Lambda(az) = a\Lambda(z) + z\Lambda(a)$, $z \in \mathbb{R}_{\geq 0}$, and $Z_\mathrm{fm}(a\psi,h,\gamma') = aZ_\mathrm{fm}(\psi,h,\gamma')$, so the first part follows. Linearity of the expectation yields $\nabla_\bullet(a\psi,\pi) = a\nabla_\bullet(\psi,\pi) + \Delta_\bullet(\psi,\pi)\Lambda(a)$ with $\Delta_\bullet(\psi,\pi) = \mathbb{E}[\Delta_\circ(\psi,\boldsymbol{\gamma})]$. This gives $\Delta_\bullet(\psi,\pi) = \Delta_\circ(\psi,(\mathbb{E}[\boldsymbol{\gamma}_{i,h}])_{i,h}) = \Delta_\circ(\psi,(\mathbb{E}[\boldsymbol{\gamma}_{1,1}])_{i,h}) = 0$ using linearity, the definition of $\boldsymbol{\gamma}$ and $\pi \in \mathcal{P}^2_*([q])^2$. Thus, taking the expectation gives $\nabla(p',\pi) = \mathbb{E}[\boldsymbol{a}\nabla_\bullet(\boldsymbol{\psi},\pi)] = \mathbb{E}[\mathbb{E}[\boldsymbol{a}|\boldsymbol{\psi}]\nabla_\bullet(\boldsymbol{\psi},\pi)]$, which establishes that $\nabla(p',\pi) = \mathbb{E}[\boldsymbol{a}]\nabla(p'',\pi)$ since for $(\boldsymbol{a},\boldsymbol{\psi}) \sim \boldsymbol{a}\otimes\boldsymbol{\psi}$ we have $p'' = p$, and otherwise for $\boldsymbol{\psi}'' \sim p''$ we have $p'' \in \mathcal{D}_\mathrm{p}$ since

$$\mathbb{E}[\boldsymbol{a}]\mathbb{E}[\Lambda(\|\boldsymbol{\psi}''\|_\infty)] = \mathbb{E}[\boldsymbol{a}\Lambda(\|\boldsymbol{\psi}\|_\infty)] = \mathbb{E}[\Lambda(\|\boldsymbol{a}\boldsymbol{\psi}\|_\infty) - \|\boldsymbol{\psi}\|_\infty\Lambda(\boldsymbol{a})] < \infty.$$

□

3.5.5.5 Continuity. We equip $\mathcal{P}(\mathcal{D}_\Psi)$ and $\mathcal{P}^2([q])$ with the Wasserstein distance $\mathrm{d_w}$ (cf. Section 3.3.1.7), and $\mathcal{P}(\mathcal{D}_\Psi)\times\mathcal{P}^2([q])^2$ as well as $\mathcal{P}(\mathcal{D}_\Psi)\times\mathcal{P}([q])$ with the 1-product metric. For $\beta \in \mathbb{R}_{>0}$ let

$$\mathcal{D}^\circ_\mathrm{p}(\beta) = \{p \in \mathcal{P}(\mathcal{D}_\Psi) : \mathbb{E}[\|\boldsymbol{\psi}\|_\infty^{1+\beta}] \leq \psi_\uparrow\},$$

Observation 3.119. *The map ∇ is continuous on $\mathcal{D}^\circ_\mathrm{p}(\beta)\times\mathcal{P}^2([q])^2$, $\beta \in \mathbb{R}_{>0}$. For all $p \in \mathcal{D}^\circ_\mathrm{p}(\beta)$ there exists $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])$ such that $\nabla_\downarrow(p,\gamma^*) = \nabla(p,\pi)$, and $\nabla_\downarrow$ is continuous on $\mathcal{D}^\circ_\mathrm{p}(\beta)\times\mathcal{P}([q])$.*

Proof. Let $\mathcal{D}^\circ_\nabla(\beta) = \mathcal{D}^\circ_\mathrm{p}(\beta)\times\mathcal{P}^2([q])^2$ and $\boldsymbol{L}_\uparrow = \Lambda(\|\boldsymbol{\psi}\|_\infty)$. For $\beta' \in [0,\beta)$ there exists $b \in \mathbb{R}_{>0}$ such that $\mathbb{E}[\boldsymbol{L}_\uparrow^{1+\beta'}] \leq \psi_\uparrow + b$ for all $(p,\pi) \in \mathcal{D}^\circ_\nabla(\beta)$. Let $\mathcal{D}_\mathrm{z} = \mathcal{D}_\Psi\times\mathcal{P}([q])^{2\times k} \subseteq \mathbb{R}^d$, $d = q^k + (q-1)^{2k}$, equipped with $\|\cdot\|_\infty$ and the Borel algebra. For $Z_{\mathrm{fm},h}:\mathcal{D}_{\mathrm{z}}\to\mathbb{R}_{\ge 0}$, $(\psi,\gamma)\mapsto Z_{\mathrm{fm}}(\psi,h,\gamma)$, we have

$$|Z_{\mathrm{fm},h}(\psi,\gamma)-Z_{\mathrm{fm},h}(\psi',\gamma')|\le\|\psi-\psi'\|_\infty+2\|\psi'\|_\infty\left(\|\gamma_{1,h}-\gamma'_{1,h}\|_{\mathrm{tv}}+\sum_{h'\ne h}\|\gamma_{2,h'}-\gamma'_{2,h'}\|_{\mathrm{tv}}\right),$$

using Observation 3.6, so $Z_{\mathrm{fm},h}$ is continuous, and thus $\Lambda_{\mathrm{z},h}=\Lambda\circ Z_{\mathrm{fm},h}$ is continuous. We equip $\mathcal{D}^*=\{P\in\mathcal{P}(\mathcal{D}_{\mathrm{z}}):\mathbb{E}[\|\boldsymbol{\psi}\|_\infty^{1+\beta}]\le\psi_\uparrow\}\subseteq\mathcal{P}(\mathbb{R}^d)$ with $\mathrm{d_w}$, the topology of weak convergence, and let

$$\Lambda^*_{\mathrm{z},h}:\mathcal{D}^*\to\mathbb{R},\ P\mapsto\mathbb{E}[\Lambda_{\mathrm{z},h}(\boldsymbol{\psi},\boldsymbol{\gamma})],$$

using $(\boldsymbol{\psi},\boldsymbol{\gamma})\sim P$ for $P\in\mathcal{P}(\mathcal{D}_{\mathrm{z}})$, further $\boldsymbol{Z}=Z_{\mathrm{fm},h}(\boldsymbol{\psi},\boldsymbol{\gamma})$ and $\boldsymbol{L}=\Lambda_{\mathrm{z},h}(\boldsymbol{\psi},\boldsymbol{\gamma})$. We turn to the continuity. For this purpose let $P\in\mathcal{D}^*$ and $\varepsilon\in(0,1)$ small. Fix $L\in\mathbb{R}_{>0}$ large, then we have

$$T(L,P)=\mathbb{E}[\mathbb{1}\{\boldsymbol{L}_\uparrow>L\}\boldsymbol{L}]\le L\mathbb{P}(\boldsymbol{L}_\uparrow>L)+\int_L^\infty\mathbb{P}(\boldsymbol{L}_\uparrow>t)\mathrm{d}t\le\frac{\psi_\uparrow+b}{L^{\beta'}}+\frac{\psi_\uparrow+b}{\beta'L^{\beta'}}.$$

Consider the law P^* of $(\boldsymbol{\psi},\boldsymbol{\gamma})|\boldsymbol{L}_\uparrow\le L$ and the following coupling. For $(\boldsymbol{\psi}_2,\boldsymbol{\gamma}_2)\sim P^*$ independent of $(\boldsymbol{\psi},\boldsymbol{\gamma})$, let $(\boldsymbol{\psi}^*,\boldsymbol{\gamma}^*)=(\boldsymbol{\psi},\boldsymbol{\gamma})$ on $\{\boldsymbol{L}_\uparrow\le L\}$ and $(\boldsymbol{\psi}^*,\boldsymbol{\gamma}^*)=(\boldsymbol{\psi}_2,\boldsymbol{\gamma}_2)$ on $\{\boldsymbol{L}_\uparrow>L\}$. This yields

$$\begin{aligned}\mathrm{d_w}(P^*,P)&=\mathbb{E}[\mathbb{1}\mathcal{E}\|(\boldsymbol{\psi}_2,\boldsymbol{\gamma}_2)-(\boldsymbol{\psi},\boldsymbol{\gamma})\|_\infty]\le\mathbb{P}(\mathcal{E})(\mathbb{E}[\|\boldsymbol{\psi}_2\|_\infty]+2)+\mathbb{E}[\mathbb{1}\mathcal{E}\|\boldsymbol{\psi}\|_\infty]\\&\le\mathbb{P}(\mathcal{E})\left(\frac{\psi_\uparrow}{\mathbb{P}(\boldsymbol{L}_\uparrow\le L)}+3\right)+\mathbb{E}[\mathbb{1}\mathcal{E}\boldsymbol{L}_\uparrow]\le\frac{8\psi_\uparrow^2}{L^{1+\beta'}}+\frac{(1+\beta')(\psi_\uparrow+b)}{\beta'L^{\beta'}},\quad\mathcal{E}=\{\boldsymbol{L}_\uparrow>L\},\end{aligned}$$

for $\psi_\uparrow$, L sufficiently large. Fix some small $\delta,\varepsilon'\in(0,1)$ and let $\hat{P}\in\mathcal{D}^*$ with $\mathrm{d_w}(P,\hat{P})<\delta$. Adapting the notation above for $\hat{P}$, we have $T(L,P),T(L,\hat{P}),\mathrm{d_w}(P^*,P),\mathrm{d_w}(\hat{P}^*,\hat{P})<\varepsilon'$ for L sufficiently large. This gives $\mathrm{d_w}(P^*,\hat{P}^*)<\delta+2\varepsilon'$ and further

$$|\mathbb{E}[\boldsymbol{L}]-\mathbb{E}[\hat{\boldsymbol{L}}]|\le2\varepsilon'+|\mathbb{E}[\mathbb{1}\{\boldsymbol{L}_\uparrow\le L\}\boldsymbol{L}]-\mathbb{E}[\mathbb{1}\{\hat{\boldsymbol{L}}_\uparrow\le L\}\hat{\boldsymbol{L}}]|.$$

Moving to the conditional expectation, we let $\boldsymbol{L}^*=(\boldsymbol{L}|\boldsymbol{L}_\uparrow\le L)$, adopt this for $\hat{P}$, and derive

$$|\mathbb{E}[\boldsymbol{L}]-\mathbb{E}[\hat{\boldsymbol{L}}]|\le2\varepsilon'+L\mathbb{P}(\boldsymbol{L}_\uparrow>L)+L\mathbb{P}(\hat{\boldsymbol{L}}_\uparrow>L)+|\mathbb{E}[\boldsymbol{L}^*]-\mathbb{E}[\hat{\boldsymbol{L}}^*]|.$$

From the bound for $T(L,P)$ we obtain $|\mathbb{E}[\boldsymbol{L}]-\mathbb{E}[\hat{\boldsymbol{L}}]|\le4\varepsilon'+|\mathbb{E}[\boldsymbol{L}^*]-\mathbb{E}[\hat{\boldsymbol{L}}^*]|$. Now, let $\mathcal{D}^\circ=\Lambda_{\mathrm{z},h}^{-1}([-1/e,L])$ and notice that $\Lambda_{\mathrm{z},h}:\mathcal{D}^\circ\to[-1/e,L]$ is continuous and bounded. Thus, the Portmanteau Theorem applies and offers δ' such that if $\mathrm{d_w}(P^*,\hat{P}^*)<\delta'$, then $|\mathbb{E}[\boldsymbol{L}^*]-\mathbb{E}[\hat{\boldsymbol{L}}^*]|<\varepsilon/2$. Thus, choosing δ and ε' sufficiently small yields $|\mathbb{E}[\boldsymbol{L}]-\mathbb{E}[\hat{\boldsymbol{L}}]|<\varepsilon$. This shows that $\Lambda^*_{\mathrm{z},h}$ is continuous. Next, notice that the map $f:\mathcal{P}(\mathcal{D}_\Psi)\times\mathcal{P}^2([q])^2\to\mathcal{P}(\mathcal{D}_{\mathrm{z}})$, $(p,\pi)\mapsto p\otimes\pi_1^{\otimes k}\otimes\pi_2^{\otimes k}$, is continuous, since for couplings $(\boldsymbol{\psi},\boldsymbol{\psi}')$, $(\boldsymbol{\gamma}_{\circ,i},\boldsymbol{\gamma}'_{\circ,i})$, $i\in[2]$, we take the canonical coupling in the image, in particular $(\boldsymbol{\gamma}_i,\boldsymbol{\gamma}'_i)\sim\bigotimes_h(\boldsymbol{\gamma}_{\circ,i},\boldsymbol{\gamma}'_{\circ,i})$. This yields $\mathbb{E}[\|(\boldsymbol{\psi},\boldsymbol{\gamma})-(\boldsymbol{\psi}',\boldsymbol{\gamma}')\|_\infty]\le\mathbb{E}[\|\boldsymbol{\psi}-\boldsymbol{\psi}'\|\infty]+2\mathbb{E}[\|\boldsymbol{\gamma}-\boldsymbol{\gamma}'\|_{\mathrm{tv}}]$, which shows that the map is $2k$-Lipschitz using Observation 3.6. Hence, the map $L_h:\mathcal{D}^\circ_{\mathrm{p}}(\beta)\times\mathcal{P}^2([q])^2\to\mathbb{R}$, $(p,\pi)\mapsto\mathbb{E}[\Lambda(Z_{\mathrm{fm}}(\boldsymbol{\psi},h,\boldsymbol{\gamma}))]$, is continuous. This completes the proof, because $\nabla(p,\pi)=L_1(p,\pi_1,\pi_1)+(k-1)L_1(p,\pi_2,\pi_2)-\sum_h L_h(p,\pi)$. Recall from the proof of Lemma 3.114 that $\mathcal{P}^2_{*,\gamma^*}([q])$ is compact, so the infimum of ∇ over $\mathcal{P}^2_{*,\gamma^*}([q])$ is attained, yielding $\pi\in\mathcal{P}^2_{*,\gamma^*}([q])$ such that $\nabla(p,\pi)=\nabla_\downarrow(p,\gamma^*)$. Continuity of $\nabla_\downarrow$ is left to the reader. □

The fact that $\nabla_\downarrow$ (over $\mathcal{D}^\circ_{\nabla\downarrow}(\beta)$) is continuous is very useful, since it suggests that the set $\mathfrak{A}=$

$\nabla_{\downarrow}^{-1}(0)$ of pairs satisfying **POS** is closed, and thereby also limits of such pairs satisfy **POS**.

3.5.5.6 Convexity. Equip $\mathcal{P}(\mathcal{D}_\Psi)$ with the Borel algebra for $\|\cdot\|_{\mathrm{tv}}$ (so the metric $\mathrm{d_w}$ for $\mathcal{P}(\mathcal{D}_\Psi)$ and the topology inducing the Borel algebra are not consistent, but that's not a problem, just something to be aware of), which defines $\mathcal{P}(\mathcal{P}(\mathcal{D}_\Psi))$, and let $p \mapsto \boldsymbol{\psi}_p$ be the kernel from Observation 3.6. Further, let $\mathcal{D}_{\mathrm{p}\uparrow} = \{p \in \mathcal{D}_\mathrm{p} : \mathbb{E}[\Lambda(\|\boldsymbol{\psi}\|_\infty)] \le \psi_\uparrow\}$ and $\mathfrak{P}_\uparrow(\gamma^*) = \mathfrak{P}(\gamma^*) \cap \mathcal{D}_{\mathrm{p}\uparrow} = \{p \in \mathcal{D}_{\mathrm{p}\uparrow} : \nabla_\downarrow(p, \gamma^*) \ge 0\}$.

Observation 3.120. *Let $\boldsymbol{p} \in \mathcal{P}(\mathcal{D}_\Psi)$, fix an event $\mathcal{E} \subseteq \mathcal{P}(\mathcal{D}_\Psi)$ with $\boldsymbol{p} \in \mathcal{E}$ almost surely, and let p' be given by $\boldsymbol{\psi}_{\boldsymbol{p}} \sim p'$. Notice that the following holds*
a) For $|\mathcal{E}| < \infty$ we have $p' \in \mathcal{D}_\mathrm{p}$ if $\mathcal{E} \subseteq \{\boldsymbol{p} \in \mathcal{D}_\mathrm{p}\}$, and $p' \in \mathfrak{P}(\gamma^*)$ if $\mathcal{E} \subseteq \{\boldsymbol{p} \in \mathfrak{P}(\gamma^*)\}$.
b) We have $p' \in \mathcal{D}_{\mathrm{p}\uparrow}$ if $\mathcal{E} \subseteq \{\boldsymbol{p} \in \mathcal{D}_{\mathrm{p}\uparrow}\}$, and $p' \in \mathfrak{P}_\uparrow(\gamma^*)$ if $\mathcal{E} \subseteq \{\boldsymbol{p} \in \mathfrak{P}_\uparrow(\gamma^*)\}$.

Proof. For $\mathcal{E} \subseteq \{\boldsymbol{p} \in \mathcal{D}_{\mathrm{p}\uparrow}\}$ we have $\mathbb{E}[\Lambda(\|\boldsymbol{\psi}_{\boldsymbol{p}}\|_\infty)] = \mathbb{E}[\mathbb{1}\mathcal{E}\Lambda(\|\boldsymbol{\psi}_{\boldsymbol{p}}\|_\infty)] \le \psi_\uparrow$ and hence $p' \in \mathcal{D}_{\mathrm{p}\uparrow}$. For $\mathcal{E} \subseteq \{\boldsymbol{p} \in \mathfrak{P}(\gamma^*)\}$ we have $p' \in \mathcal{D}_{\mathrm{p}\uparrow}$ by the above, and further $\nabla(p', \pi) = \mathbb{E}[\mathbb{1}\mathcal{E}\nabla(\boldsymbol{p}, \pi)] \ge 0$, $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])$, using the tower property and independence. This shows that $\nabla_\downarrow(p', \pi) \ge 0$ and hence $p' \in \mathfrak{P}(\gamma^*)$. The first part is shown analogously. □

3.5.5.7 Permutation Invariance. For $\psi \in \mathcal{D}_\Psi$ and $\iota \in [k]!$ let $[\psi]_{\mathrm{p},\iota} \in \mathcal{D}_\Psi$ be given by $[\psi]_{\mathrm{p},\iota}(\sigma) = \psi(\sigma \circ \iota)$. Further, let $[\psi] = \{[\psi]_{\mathrm{p},\iota} : \iota \in [k]!\}$ be the equivalence class and $[\mathcal{D}_\Psi] = \{[\psi] : \psi \in \mathcal{D}_\Psi\}$ the quotient space, i.e. $\eta : \mathcal{D}_\Psi \to [\mathcal{D}_\Psi]$, $\psi \mapsto [\psi]$, is a quotient map, and equip $[\mathcal{D}_\Psi]$ with the Borel algebra, so η is measurable. Let $\eta_\mathrm{p} : \mathcal{P}(\mathcal{D}_\Psi) \to \mathcal{P}([\mathcal{D}_\Psi])$ be given by $\eta(\boldsymbol{\psi}) \sim \eta_\mathrm{p}(p)$ for $p \in \mathcal{P}(\mathcal{D}_\Psi)$, using $\boldsymbol{\psi} \sim p$. Finally, let $[p] = \eta_\mathrm{p}^{-1}(\eta_\mathrm{p}(p)) \subseteq \mathcal{P}(\mathcal{D}_\Psi)$ be the equivalence class of p.

Fact 3.121. *Notice that the following holds.*
a) For $\iota \in [k]!$ we have $\nabla_\bullet([\psi]_{\mathrm{p},\iota}, \pi) = \nabla_\bullet(\psi, \pi)$.
b) For $p \in \mathfrak{P}(\gamma^*)$ we have $[p] \subseteq \mathfrak{P}(\gamma^*)$.

Proof. Using $\gamma'_i = \gamma_i \circ \iota$, $\gamma' = (\gamma'_1, \gamma'_2)$, and $h' = \iota^{-1}(h)$ we notice that

$$Z_{\mathrm{fm}}(\psi', h, \gamma) = \sum_\sigma \psi'(\sigma)\gamma_{1,h}(\sigma_h) \prod_{j \ne h} \gamma_{2,j}(\sigma_j) = \sum_\sigma \psi(\sigma)\gamma'_{1,h'}(\sigma_{h'}) \prod_{j \ne h'} \gamma'_{2,j}(\sigma_j) = Z_{\mathrm{fm}}(\psi, h', \gamma').$$

Since h is only relevant for the negative contribution to $\nabla_\circ$ where we sum over h, we obtain $\nabla_\circ(\psi', \gamma) = \nabla_\circ(\psi, \gamma')$. Finally, since $\boldsymbol{\gamma}_i$ is exchangeable, we have $(\boldsymbol{\gamma}_1 \circ \iota, \boldsymbol{\gamma}_2 \circ \iota) \sim \boldsymbol{\gamma}$ and thereby $\nabla_\bullet(\psi', \pi) = \nabla_\bullet(\psi, \pi)$. This shows that $\nabla_\bullet([\psi], \pi) = \nabla_\bullet(\psi, \pi)$ is well-defined, which shows that $\nabla_\downarrow([p], \gamma^*) = \nabla_\downarrow(p, \gamma^*)$ is well-defined and thereby completes the proof. □

Thus, it doesn't matter if we consider p, the symmetrized version $[\boldsymbol{\psi}]_{\mathrm{p},\boldsymbol{\iota}}$ with $(\boldsymbol{\psi}, \boldsymbol{\iota}) \sim p \otimes \mathrm{u}([k]!)$ or versions with minimal support. Of course, we may also permute colors using $\iota \in [q]!$, but obviously for $\boldsymbol{\psi}$ and γ^* consistently.

3.5.5.8 Embeddings and Projections. In this section we consider the product space embedding for an increasing number of coordinates and the color projection for a decreasing number of colors. For the first type fix $\psi \in \mathbb{R}_{\ge 0}^{[q]^k} \times \mathbb{R}_{\ge 0}^{[q]^{k'}}$ and let $\psi_1\psi_2 : [q]^{k+k'} \to \mathbb{R}_{\ge 0}$, $\sigma \mapsto \psi_1(\sigma_{[k]})\psi_2(\sigma_{[k']})$, where we use $[k + k'] = [k] \,\dot\cup\, [k']$. For $p \in \mathcal{P}(\mathbb{R}_{\ge 0}^{[q]^k}) \times \mathcal{P}(\mathbb{R}_{\ge 0}^{[q]^{k'}})$ this yields the embedding $p_1p_2 \in \mathcal{P}(\mathbb{R}_{\ge 0}^{[q]^{k+k'}})$, given by $\boldsymbol{\psi}_1\boldsymbol{\psi}_2 \sim p_1p_2$ with $(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \sim p_1 \otimes p_2$. On the other hand, let $[\psi]_\mathrm{q} : [q-1]^k \to \mathbb{R}_{\ge 0}$, $\sigma \mapsto \psi(\sigma)$ be the restriction of ψ. For $p \in \mathcal{P}(\mathcal{D}_\Psi)$ this yields the color restriction $[p]_\mathrm{q} \in \mathcal{P}(\mathbb{R}_{\ge 0}^{[q-1]^k})$ of p, given by $[\boldsymbol{\psi}]_\mathrm{q} \sim [p]_\mathrm{q}$. Let $\mathfrak{A} = \mathfrak{A}(k, q)$.

Observation 3.122. *Notice that the following holds.*

a) For $p \in \mathcal{P}(\mathbb{R}_{\geq 0}^{[q]^k}) \times \mathcal{P}(\mathbb{R}_{\geq 0}^{[q]^{k'}})$ with $p_1, p_2 \in \mathcal{D}_{\mathrm{p}}$ we have $(p_1p_2, \gamma^*) \in \mathfrak{A}(k+k', q)$ if and only if $(p_1, \gamma^*) \in \mathfrak{A}(k, q)$ and $(p_2, \gamma^*) \in \mathfrak{A}(k', q)$. In particular we have $p^* \in \mathfrak{A}(k+k', q)$ if and only if $p \in \mathfrak{A}(k, q)$, where p^* is the law of $\boldsymbol{\psi}\psi$ and $\psi \equiv 1$.
b) For $\gamma^*(q) = 0$ we have $(p, \gamma^*) \in \mathfrak{A}(k, q)$ if and only if $([p]_{\mathrm{q}}, \gamma^*_{[q-1]}) \in \mathfrak{A}(k, q-1)$.

Proof. For the first part, recall the proof of Observation 3.116 to obtain $\nabla(p_1p_2, \pi) = \nabla_1 + \nabla_2$,

$$\nabla_1 = \mathbb{E}\left[\boldsymbol{Z}_{21}\Lambda(\boldsymbol{Z}_{11}) + (k+k'-1)\boldsymbol{Z}_{22}\Lambda(\boldsymbol{Z}_{12}) - \sum_{h=1}^{k+k'} \boldsymbol{Z}_{23,h}\Lambda(\boldsymbol{Z}_{13,h})\right] = Z_2\nabla(p_1, \pi),$$

similarly $\nabla_2 = Z_2\nabla(p_2, \pi)$ and where $\boldsymbol{Z}_{ij} = Z_{\mathrm{f}}(\boldsymbol{\psi}_i, \boldsymbol{\gamma}_{j,[k_i]})$, $\boldsymbol{Z}_{i3,h} = Z_{\mathrm{fm}}(\boldsymbol{\psi}_i, h, (\boldsymbol{\gamma}_{1,[k_i]}, \boldsymbol{\gamma}_{2,[k_i]}))$ for $h \in [k_i]$ and $\boldsymbol{Z}_{i3,h} = Z_{\mathrm{f}}(\boldsymbol{\psi}_i, \boldsymbol{\gamma}_{2,[k_i]})$ otherwise, $Z_i = Z_{\mathrm{f}}(\mathbb{E}[\boldsymbol{\psi}_i], (\gamma^*)_{h\in[k_i]})$, $k_1 = k$ and $k_2 = k'$.

For $\gamma^*(q) = 0$ and $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])^2$ we have $\boldsymbol{\gamma}_{i,h}(q) = 0$ almost surely, since the $\boldsymbol{\gamma} \geq 0$ and $\mathbb{E}[\boldsymbol{\gamma}_{i,h}(q)] = 0$. Hence, we have $Z_{\mathrm{fm}}(\psi, h, \boldsymbol{\gamma}) = Z_{\mathrm{fm}}([\psi]_{\mathrm{q}}, h, (\boldsymbol{\gamma}_{i,h,[q-1]})_{i,h})$ almost surely. □

The corresponding result for **BAL** is immediate, completing the pending part in the proof of Observation 3.106. Next, we consider the following color embedding (inspired by Gallager's mapping, cf. [90]). Let $q' \in \mathbb{Z}_{\geq q}$ and let $\dot{\bigcup}_\tau \mathcal{C}_\tau = [q']$ be a partition into color classes $\mathcal{C} = (\mathcal{C}_\tau)_{\tau\in[q]}$ such that $\tau \in \mathcal{C}_\tau$ for all $\tau \in [q]$. For $\tau \in [q']$ let $[\tau]_{\mathcal{C}} \in [q]$ be the unique representant with $\tau \in \mathcal{C}_{[\tau]_{\mathcal{C}}}$. For $\psi \in \mathbb{R}_{\geq 0}^{[q]^k}$ let $[\psi]_{\mathcal{C}} : [q']^k \to \mathbb{R}_{\geq 0}$, $\sigma \mapsto \psi(([\sigma_h]_{\mathcal{C}})_h)$. For $p \in \mathcal{P}(\mathbb{R}_{\geq 0}^{[q]^k})$ let $[p]_{\mathcal{C}} \in \mathcal{P}(\mathbb{R}_{\geq 0}^{[q']^k})$ be given by $[\boldsymbol{\psi}]_{\mathcal{C}} \sim [p]_{\mathcal{C}}$. For $\gamma \in \mathcal{P}([q'])$ let $[\gamma]_{\mathcal{C}} \in \mathcal{P}([q])$ be given by $[\gamma]_{\mathcal{C}}(\tau) = \gamma(\mathcal{C}_\tau)$.

Observation 3.123. *For $p \in \mathcal{D}_{\mathrm{p}}$ and $\gamma^* \in \mathcal{P}([q'])$ we have $([p]_{\mathcal{C}}, \gamma^*) \in \mathfrak{A}(k, q')$ if and only if $(p, [\gamma^*]_{\mathcal{C}}) \in \mathfrak{A}(k, q)$.*

Proof. First, notice that $Z_{\mathrm{fm}}([\psi]_{\mathcal{C}}, h, \gamma) = Z_{\mathrm{fm}}(\psi, h, ([\gamma_{i,h}]_{\mathcal{C}})_{i,h})$. Thus, for $\pi \in \mathcal{P}^2_{*,\gamma^*}([q'])$ with $\boldsymbol{\gamma} \sim \pi$ and $[\pi]_{\mathcal{C}} \in \mathcal{P}^2_{*,[\gamma^*]_{\mathcal{C}}}([q])$ given by $[\boldsymbol{\gamma}]_{\mathcal{C}} \sim [\pi]_{\mathcal{C}}$ we have $\nabla_\bullet([\psi]_{\mathcal{C}}, \pi) = \nabla_\bullet(\psi, [\pi]_{\mathcal{C}})$ and thereby $\nabla([p]_{\mathcal{C}}, \pi) = \nabla(p, [\pi]_{\mathcal{C}})$. Since $\mathcal{P}^2_{*,\gamma^*}([q']) \to \mathcal{P}^2_{*,[\gamma^*]_{\mathcal{C}}}([q])$, $\pi \mapsto [\pi]_{\mathcal{C}}$, is surjective, the result follows. □

Next, we turn to a reweighting scheme in the flavor of Observation 3.118 that heavily generalizes Gallager's mapping. Recall q' and $\mathcal{C}$ from above. Using Observation 3.122, let $r : [q'] \to \mathbb{R}_{\geq 0}$ and $\beta^* \in \mathcal{P}([q'])$ be such that $\sum_{\rho\in\mathcal{C}_\tau} \beta^*(\rho)r(\rho) = \beta^*(\mathcal{C}_\tau) > 0$. For $\rho \in [q']$ let $[\rho] \in [q]$ be the unique element with $\rho \in \mathcal{C}_{[\rho]}$ and let $[\sigma] = ([\sigma_h])_h$ for $\sigma \in [q']^k$. For $\psi \in \mathbb{R}_{\geq 0}^{[q]^k}$ let $[\psi] \in \mathbb{R}_{\geq 0}^{[q']^k}$ be given by $[\psi](\sigma) = \prod_h r(\sigma_h)\psi([\sigma])$, and let $[p] \in \mathcal{P}(\mathbb{R}_{\geq 0}^{[q']^k})$ be given by $[\boldsymbol{\psi}] \sim [p]$. For $\beta \in \mathcal{P}([q'])$ let $[\beta] \in \mathbb{R}^q_{\geq 0}$ be given by $[\beta](\tau) = \sum_{\rho\in\mathcal{C}_\tau} r(\rho)\beta(\rho)$.

Observation 3.124. *We have $(p, [\beta^*]) \in \mathfrak{A}(k, q)$ if we have $([p], \beta^*) \in \mathfrak{A}(k, q')$.*

Proof. For $\tau \in [q]$ let $\beta_\tau \in \mathcal{P}([q'])$ be given by $\beta_\tau(\rho) = \beta^*(\rho)/\beta^*(\mathcal{C}_\tau)$, $\rho \in \mathcal{C}_\tau$. Notice that the inverse of the map $\beta^\circ : \mathcal{P}([q]) \to \mathcal{P}([q'])$, $\gamma \mapsto \sum_\tau \gamma(\tau)\beta_\tau$ is $\gamma : \mathcal{P}([q']) \to \mathbb{R}^q_{\geq 0}$, $\beta \mapsto [\beta]$, meaning that $\beta^\circ : \mathcal{P}([q]) \to \beta^\circ(\mathcal{P}([q]))$ is a bijection, and that $\beta^\circ(\gamma^*) = \beta^*$, where $\gamma^* = [\beta^*]$. For $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])$ let $[\pi] \in \mathcal{P}^2_{*,\beta^*}([q'])$ be given by $\beta^\circ(\boldsymbol{\gamma}) \sim [\pi]$, using $\boldsymbol{\gamma} \sim \pi$.

Next, notice that $Z_{\mathrm{fm}}([\psi], h, \gamma) = Z_{\mathrm{fm}}(\psi, h, ([\gamma_{i,h}])_{i,h})$ which yields that $\nabla_\bullet([\psi], [\pi]) = \nabla_\bullet(\psi, \pi)$ for all $\pi \in \mathcal{P}^2_{*,\gamma^*}([q])$, which suggests that $\nabla([p], [\pi]) = \nabla(p, \pi)$ and hence $\nabla_\downarrow([p], \beta^*) \leq \nabla_\downarrow(p, \gamma^*)$. □

3.5.6 Valid Models. So far, we have discussed a plentitude of closure properties, but all models so far are trivial in some sense. Let $f:\mathbb{R}^2\to\mathbb{R}$, $(x,y)\mapsto x^k+(k-1)y^k-kxy^{k-1}$ and $\mathcal{S}=f^{-1}(\mathbb{R}_{\ge 0})$.

Observation 3.125. *We have $\mathbb{R}_{\ge 0}^2\subseteq\mathcal{S}$. For $k\in 2\mathbb{Z}_{>0}$ we have $\mathcal{S}=\mathbb{R}^2$.*

Proof. The first part holds because $f(x,y)=(x-y)^2\sum_{\ell=0}^{k-2}(\ell+1)x^{k-2-\ell}y^\ell$. For the second part we distinguish the following cases. First notice that $f(x,0)=x^k\ge 0$ since $k\in 2\mathbb{Z}_{>0}$. For $y\neq 0$ let $r=x/y$ and notice that $f(x,y)=g(r)y^k$ with $g(r)=r^k+(k-1)-kr$. Using $g'(r)=kr^{k-1}-k$ and $g''(r)=k^2r^{k-2}$ we notice that g is convex since $k\in 2\mathbb{Z}_{>0}$ and thus $g''\ge 0$. Hence, the only stationary point at $r=1$ is the unique minimum, with $g(1)=0$. □

As simple as it seems, many proofs in the past twenty years relied on this result (cf. Remark 3.128) For $p\in\mathcal{P}(\mathcal{T})$, $\mathcal{T}=\mathbb{R}_{\ge 0}\times\mathbb{R}\times\mathbb{R}^{[q]^k}$, let $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{\Delta})\sim p$. The base class of functions is

$$\mathfrak{P}_{\mathrm{b}}=\{p\in\mathcal{P}(\mathcal{T}):\mathbb{E}[\Lambda(\boldsymbol{a})]<\infty,\,\mathbb{P}(|\boldsymbol{b}\boldsymbol{\Delta}|\le 1)=1\}\,.$$

For $p\in\mathfrak{P}_{\mathrm{b}}$ let $\boldsymbol{\psi}:[q]^k\to\mathbb{R}$, $\sigma\mapsto\boldsymbol{a}(1-\boldsymbol{b}\boldsymbol{\Delta}(\sigma))$, and let $[p]$ be given by $\boldsymbol{\psi}\sim[p]$.

Observation 3.126. *For $p\in\mathfrak{P}_{\mathrm{b}}$ we have $\nabla([p],\pi)=\sum_{\ell=2}^{\infty}\frac{1}{\ell(\ell-1)}\mathbb{E}[\boldsymbol{a}\boldsymbol{b}^\ell\boldsymbol{\nabla}^\circ(\ell)]$,*

$$\boldsymbol{\nabla}^\circ(\ell)=Z_{\mathrm{f}}(\boldsymbol{\Delta},\boldsymbol{\gamma}_1)^\ell+(k-1)Z_{\mathrm{f}}(\boldsymbol{\Delta},\boldsymbol{\gamma}_2)^\ell-\sum_h Z_{\mathrm{fm}}(\boldsymbol{\Delta},h,\boldsymbol{\gamma})^\ell.$$

Proof. Notice that $\|\boldsymbol{\psi}\|_\infty\le\boldsymbol{a}(1+\|\boldsymbol{b}\boldsymbol{\Delta}\|_\infty)\le 2\boldsymbol{a}$ almost surely, so $[p]\in\mathcal{D}_{\mathrm{p}}$ and thus $\nabla([p],\pi)$ is well-defined. Using Observation 3.118 we obtain $\nabla([p],\pi)=\mathbb{E}[\boldsymbol{a}\nabla_\bullet(\boldsymbol{\psi}^\circ,\pi)]$ with $\boldsymbol{\psi}^\circ:[q]^k\to\mathbb{R}$, $\sigma\mapsto 1-\boldsymbol{b}\boldsymbol{\Delta}(\sigma)$. Hence, we have $\nabla([p],\pi)=\mathbb{E}[\boldsymbol{a}\nabla_\circ(\boldsymbol{\psi}^\circ,\boldsymbol{\gamma})]$. Next, we use that $\Lambda(1-t)+t=\sum_{\ell=2}^\infty\frac{1}{\ell(\ell-1)}t^\ell$ for $|t|\le 1$. For this purpose we rewrite $\nabla([p],\pi)=\mathbb{E}[\boldsymbol{a}(\nabla_\circ(\boldsymbol{\psi}^\circ,\boldsymbol{\gamma})+\boldsymbol{L})-\boldsymbol{a}\boldsymbol{L}]$ with

$$\boldsymbol{L}=Z_{\mathrm{f}}(\boldsymbol{b}\boldsymbol{\Delta},\boldsymbol{\gamma}_1)+(k-1)Z_{\mathrm{f}}(\boldsymbol{b}\boldsymbol{\Delta},\boldsymbol{\gamma}_2)-\sum_h Z_{\mathrm{f}}(\boldsymbol{b}\boldsymbol{\Delta},h,\boldsymbol{\gamma})\in[-k,k].$$

Using linearity we obtain $\nabla([p],\pi)=\mathbb{E}[\boldsymbol{a}(\nabla_\circ(\boldsymbol{\psi}^\circ,\boldsymbol{\gamma})+\boldsymbol{L})]$ since $\mathbb{E}[\boldsymbol{a}\boldsymbol{L}]=0$, using that all color distributions have the same expectation. Using the power series we obtain

$$\nabla([p],\pi)=\mathbb{E}\left[\sum_{\ell=2}^\infty\frac{\boldsymbol{a}\boldsymbol{b}^\ell\boldsymbol{\nabla}^\circ(\ell)}{\ell(\ell-1)}\right]=\sum_{\ell=2}^\infty\mathbb{E}\left[\frac{\boldsymbol{a}\boldsymbol{b}^\ell\boldsymbol{\nabla}^\circ(\ell)}{\ell(\ell-1)}\right].$$

□

Now, we formally introduce the non-trivial models from Section 2.1.2.2. Recall that these models are defined on $\mathcal{T}=\mathbb{R}_{\ge 0}\times\mathbb{R}\times((\mathbb{R}^q)^{\mathbb{Z}_{>0}})^k$, ultimately still $\mathbb{R}^{\mathbb{Z}}$, and thereby all of these spaces are Borel (Definition 8.35 in [75]). Thus, we can work with Markov kernels without loss of generality. Further, we can modify $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{f})$ as long as it does not affect $\boldsymbol{\psi}$, since this consequently does not affect the assertion in Proposition 2.1.

Let $\mathcal{F}$ be the σ-algebra from Section 3.1.6, Σ_{c} the σ-algebra of $\mathbb{R}_{\ge 0}\times\mathbb{R}$, Σ_{f} the σ-algebra of $(\mathbb{R}^q)^{\mathbb{Z}_{>0}}$, and $\Sigma_\times=\Sigma_{\mathrm{c}}\otimes\Sigma_{\mathrm{f}}^{\otimes k}$. The following triplets $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{f})\in\mathcal{T}$ are of Type 1.

- There exists a sub-σ-algebra $\Sigma\subseteq\mathcal{F}$ and Markov kernels $\kappa_{\mathrm{c}}:\Omega\times\Sigma_{\mathrm{c}}\to[0,1]$, $\kappa_{\mathrm{f}}:\Omega\times\Sigma_{\mathrm{f}}\to[0,1]$ with the following property. The product kernel $\kappa:\Omega\times\Sigma_\times\to[0,1]$, i.e. $\kappa_\omega=\kappa_{\mathrm{c},\omega}\otimes\kappa_{\mathrm{f},\omega}^{\otimes k}$ for $\omega\in\Omega$, is a regular conditional distribution for $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{f})$ given Σ.

- We have $\sum_{\sigma\in[q]^k}\sum_{i=1}^{\infty}\left|\prod_{h=1}^{k}\boldsymbol{f}_{h,i}(\sigma_h)\right|<\infty$ almost surely. Redefine $\boldsymbol{f}$ by setting it 0 on the event that the series diverges. Let $\boldsymbol{\Delta}:[q]^k\to\mathbb{R}$, $\sigma\mapsto\sum_i\prod_h\boldsymbol{f}_{h,i}(\sigma_h)$.
- We have $|\boldsymbol{b}|\|\boldsymbol{\Delta}\|_\infty\le 1$ almost surely.
- We have $\mathbb{E}[\Lambda(\boldsymbol{a})]<\infty$.
- We have $\mathbb{E}[|\boldsymbol{a}\boldsymbol{b}^\ell|],\mathbb{E}[\|\boldsymbol{\Delta}\|_\infty^\ell]<\infty$ for all $\ell\in\mathbb{Z}_{>0}$.
- We have $\mathbb{E}[\boldsymbol{a}\boldsymbol{b}^\ell|\Sigma]\ge 0$ almost surely for $\ell\in 2\mathbb{Z}_{>0}+1$.

Notice that redefining $\boldsymbol{f}$ on a null event does not change the distribution p of $\boldsymbol{\psi}$. A triplet is of Type 2 if it is of Type 1, further $\boldsymbol{f}_{1,i}\equiv 0$, $i\in\mathbb{Z}_{>1}$, and $\mathbb{E}[\boldsymbol{a}\boldsymbol{b}^\ell|\Sigma]=0$, $\ell\in 2\mathbb{Z}_{>0}+1$, almost surely. A triplet is of Type 3 if it is of Type 1 and $\boldsymbol{f}_1\ge 0$ almost surely. For $i\in[3]$ the distribution of $\sigma\mapsto\boldsymbol{a}(1-\boldsymbol{b}\boldsymbol{\Delta}(\sigma))$ is of Type i if $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{f})$ is of Type i. Let $\mathfrak{P}_i\subseteq\mathcal{D}_{\mathrm{p}}$ be the set of all distributions of Type $i\in[3]$. Let $\mathfrak{P}_4$ be the set of all distributions p of $\sigma\mapsto\prod_h\boldsymbol{f}_h(\sigma_h)$ for $\boldsymbol{f}\in(\mathbb{R}_{\ge 0}^q)^k$ with $p\in\mathcal{D}_{\mathrm{p}}$. For $k\in 2\mathbb{Z}$ let $\mathfrak{P}=\mathfrak{P}_1\cup\mathfrak{P}_4$, otherwise let $\mathfrak{P}=\mathfrak{P}_2\cup\mathfrak{P}_3\cup\mathfrak{P}_4$.

Proposition 3.127. *We have $\mathfrak{P}\times\mathcal{P}([q])\subseteq\mathfrak{A}$.*

Proof. The set $\mathfrak{P}_4$ was discussed in Observation 3.116. For $p\in\mathfrak{P}_1$, Observation 3.126 applies, and there exists a σ-algebra $\Sigma\subseteq\mathcal{F}$ that yields conditional independence. Then we have

$$\nabla(p,\pi)=\sum_{\ell=2}^{\infty}\frac{1}{\ell(\ell-1)}\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{a}\boldsymbol{b}^\ell\middle|\Sigma\right]\mathbb{E}\left[\boldsymbol{\nabla}^\circ(\ell)|\Sigma\right]\right].$$

Now, we turn to the various cases. First, assume that $k\in 2\mathbb{Z}$. Recall that we have $\mathbb{E}[\boldsymbol{a}\boldsymbol{b}^\ell|\Sigma]\ge 0$ almost surely for the leading factors, by assumption for $\ell\in 2\mathbb{Z}_{>0}+1$, and for $\ell\in 2\mathbb{Z}_{>0}$ because $\boldsymbol{a}\ge 0$ almost surely and $\boldsymbol{b}^\ell\ge 0$ almost surely. Hence, we focus on

$$\boldsymbol{\nabla}^\circ(\ell)=\sum_{M=1}^{\infty}\sum_{i\in\mathcal{I}_M}\left(\prod_{m=1}^{\ell}\prod_{h=1}^{k}\boldsymbol{s}_{1,i_m,h}+(k-1)\prod_{m=1}^{\ell}\prod_{h=1}^{k}\boldsymbol{s}_{2,i_m,h}-\sum_h\prod_{m=1}^{\ell}\left(\boldsymbol{s}_{1,i_m,h}\prod_{h'\ne h}\boldsymbol{s}_{2,i_m,h'}\right)\right)$$

with $\mathcal{I}_M=\{i\in\mathbb{Z}_{>0}^\ell:\sum_m i_m=M\}$, $\boldsymbol{s}_{j,i,h}=\sum_\sigma\boldsymbol{\gamma}_{j,h}(\sigma)\boldsymbol{f}_{h,i}(\sigma)$, using that the series underlying $\boldsymbol{\Delta}$ are absolutely convergent almost surely. Now, taking the conditional expectation yields

$$\mathbb{E}[\boldsymbol{\nabla}^\circ(\ell)|\Sigma]=\sum_{M=1}^{\infty}\sum_{i\in\mathcal{I}_M}\left(\boldsymbol{S}_{i,1}^k+(k-1)\boldsymbol{S}_{i,2}^k-k\boldsymbol{S}_{i,1}\boldsymbol{S}_{i,2}^{k-1}\right)$$

with $\boldsymbol{S}_{i,j}=\mathbb{E}[\prod_m\boldsymbol{s}_{j,i_m,1}|\Sigma]$ using conditional independence and identical laws along the coordinates. With $k\in 2\mathbb{Z}_{>0}$ and Observation 3.125 we're done. For $p\in\mathfrak{P}_3$ we argue exactly as above, only in the very last step we use that $\boldsymbol{S}_{i,j}\ge 0$ almost surely. For the remaining case we use that $\mathbb{E}[\boldsymbol{a}\boldsymbol{b}^\ell|\Sigma]=0$ almost surely for odd ℓ, to restrict to $\ell\in 2\mathbb{Z}_{>0}$. Since we only have a single contribution in this case, the set $\mathcal{I}_M$ is empty for $M\ne k$, thus we only have to evaluate $i=(1)_m$ and thereby $\boldsymbol{S}_{i,j}=\mathbb{E}[\boldsymbol{s}_{j,1,1}^\ell|\Sigma]\ge 0$ almost surely, so Observation 3.125 applies as above. □

Proposition 2.1 is a corollary to Proposition 3.127, for bounded weights. Recall that all previously established closure properties apply on top of Proposition 3.127.

Remark 3.128. The issue with the parity of k has a long history. Maneva [95] noticed that the restriction to even k in [53, 105] was not necessary, regarding the Type 2 models. Then, it was pointed out in [33] that the restriction to even k in [4] was not necessary, again, regarding the Type

2 models. The restriction to even k in this work stems from the Type 1 model, which is new, to our knowledge.

3.5.7 Related Models. In this section we discuss closely related models, given by the following parameters. The individual variable neighborhoods are $\mathcal{U}_1 = [n]^k$ or $\mathcal{U}_2 = [n]^{\underline{k}} \subseteq \mathcal{U}_1$. A weight $\psi : [q]^k \to \mathbb{R}_{\geq 0}$ is permutation invariant if $\psi(\tau \circ \iota) = \psi(\tau)$ for all $\tau \in [q]^k$ and $\iota \in [k]!$. If $\boldsymbol{\psi}_\circ$ is permutation invariant almost surely, let $\mathcal{U}_3 = \{u \in \mathcal{U}_2 : \forall h < k\, u_h < u_{h+1}\}$ (which is $\binom{[n]}{k}$), and let $\mathcal{T}_3 = \{\tau \in [q]^k : \forall h < k\, \tau_h \leq \tau_{h+1}\}$.

Let $\mathcal{V}_{r,1} = \mathcal{U}_r^m$, $\mathcal{V}_{r,2} = \mathcal{U}_r^{\underline{m}} \subseteq \mathcal{V}_{r,1}$ and $\mathcal{V}_{r,3} = \{v \in \mathcal{V}_{r,2} : \forall a < m\, v_a < v_{a+1}\}$ (using the lexicographical order, this is $\binom{\mathcal{U}_r}{m}$) be the joint variable neighborhoods. Null model and teacher-student model pairs are uniform or binomial. Let $n_{\circ,\mathfrak{g}}$ be sufficiently large, $m_{\circ,\mathfrak{g},n} = \Theta(n^k)$ sufficiently small, $n \geq n_\circ$ and $m \leq m_\circ$.

3.5.7.1 Model Definitions. Let $\boldsymbol{w}_{r,s,m} = (\boldsymbol{v}_{r,s}, \boldsymbol{\psi}_{r,s}) \sim \mathrm{u}(\mathcal{V}_{r,s}) \otimes \mu_\psi^{\otimes m}$, and let $\boldsymbol{G}_{r,s,m} = [\boldsymbol{w}]^\Gamma_{\gamma^*}$ be the graph with external fields. Let $\psi_{\mathrm{g},\boldsymbol{w}}(\sigma) = \prod_a \boldsymbol{\psi}_a(\sigma_{\boldsymbol{v}_a})$ and $\psi_{\mathrm{g},\boldsymbol{G}}(\sigma) = \gamma^{*\otimes n}(\sigma)\psi_{\mathrm{g},\boldsymbol{w}}(\sigma)$ be the weights, and $\boldsymbol{G}^*_{r,s,m}(\sigma)$ given by $G \mapsto \psi_{\mathrm{g},G}(\sigma)/\overline{\psi}(\sigma)$, where $\overline{\psi}_{r,s,m} = \mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}}(\sigma)]$.

For the binomial null model let $\overline{\psi}_\circ = \mathbb{E}[\boldsymbol{\psi}_\circ]$, $\gamma_\sigma = \gamma_{\mathrm{n},\sigma}$, $\alpha_\tau = (|\tau^{-1}(\rho)|)_\rho$ for $\tau \in [q]^k$, and

$$\overline{Z}_{\mathrm{f},1}(\gamma_\sigma) = \sum_\tau \prod_h \gamma_\sigma(\tau_h)\overline{\psi}_\circ(\tau) = \sum_\tau \frac{1}{n^k} \prod_\rho (n\gamma_\sigma(\rho))^{\alpha_\tau(\rho)}\overline{\psi}_\circ(\tau),$$

$$\overline{Z}_{\mathrm{f},2}(\gamma_\sigma) = \sum_\tau \frac{1}{n^{\underline{k}}} \prod_\rho (n\gamma_\sigma(\rho))^{\underline{\alpha_\tau(\rho)}}\overline{\psi}_\circ(\tau) = \sum_\tau \frac{\prod_\rho \binom{n\gamma_\sigma(\rho)}{\alpha_\tau(\rho)}}{\binom{n}{k}} \frac{1}{\binom{k}{\alpha_\tau}} \overline{\psi}_\circ(\tau),$$

$$\overline{Z}_{\mathrm{f},3}(\gamma_\sigma) = \sum_{\tau \in \mathcal{T}_3} \frac{\prod_\rho \binom{n\gamma_\sigma(\rho)}{\alpha_\tau(\rho)}}{\binom{n}{k}} \overline{\psi}_\circ(\tau) = \overline{Z}_{\mathrm{f},2}(\gamma_\sigma).$$

Notice that each version is an expectation over $\overline{\psi}_\circ$, with respect to a compatible law. Recall the expected number $\overline{m} = \bar{d}n/k$ of factors. Let $p_{\mathrm{b},r,\bar{d},n} = \frac{\overline{m}}{|\mathcal{U}_r|}$ be the success probability. For the case[5] $s = 1$ let $\boldsymbol{v}_\mathrm{b} \sim \mathrm{Po}(p_\mathrm{b})^{\otimes \mathcal{U}_r}$, otherwise let $\boldsymbol{v}_\mathrm{b} \sim \mathrm{Bin}(1, p_\mathrm{b})^{\otimes \mathcal{U}_r}$. Further, let $\boldsymbol{\psi}_\mathrm{b} \sim \mu_\psi^{\otimes \mathcal{U}_r \times \mathbb{Z}_{>0}}$, $\boldsymbol{w}_{\mathrm{b},r,s,\bar{d}} = (\boldsymbol{v}_\mathrm{b}, \boldsymbol{\psi}_\mathrm{b})$ and $\boldsymbol{G}_{\mathrm{b},r,s,\bar{d}} = [\boldsymbol{w}_\mathrm{b}]^\Gamma_{\gamma^*}$. Let $p^*_{\mathrm{b},r,s,\bar{d},\sigma}(\tau) = \frac{\overline{\psi}_\circ(\tau)}{\overline{Z}_{\mathrm{f},r}(\gamma_\sigma)} p_\mathrm{b}$, further $\boldsymbol{v}^*_{\mathrm{b},r,s,\bar{d},\sigma} \sim \bigotimes_{u \in \mathcal{U}_r} \mathrm{Po}(p^*_\mathrm{b}(\sigma_u))$ for $s = 1$ and $\boldsymbol{v}^*_\mathrm{b} \sim \bigotimes_{u \in \mathcal{U}_r} \mathrm{Bin}(1, p^*_\mathrm{b}(\sigma_u))$ otherwise. Let $\boldsymbol{\psi}^*_{\mathrm{b},\sigma} \sim \bigotimes_{u \in \mathcal{U}_r} \boldsymbol{\psi}^{*\otimes \mathbb{Z}_{>0}}_{\circ,\sigma_u}$, where $\boldsymbol{\psi}^*_{\circ,\tau}$ is given by the Radon-Nikodym derivative $\psi \mapsto \psi(\tau)/\overline{\psi}_\circ(\tau)$, and let $\boldsymbol{w}^*_{\mathrm{b},r,s,\bar{d},\sigma} = (\boldsymbol{v}^*_\mathrm{b}, \boldsymbol{\psi}^*_\mathrm{b}) \sim \boldsymbol{v}^*_\mathrm{b} \otimes \boldsymbol{\psi}^*_\mathrm{b}$, $\boldsymbol{G}^*_{\mathrm{b},r,s,\bar{d}}(\sigma) = [\boldsymbol{w}^*]^\Gamma_{\gamma^*}$. For $w = (v, \psi)$ from $\boldsymbol{w}_\mathrm{b}$ let $\psi_{\mathrm{g},w}(\sigma) = \prod_{u \in \mathcal{U}_r} \prod_{a=1}^{v(u)} \psi_{u,a}(\sigma_u)$ and $\psi_{\mathrm{g},[w]^\Gamma_{\gamma^*}}(\sigma) = \gamma^{*\otimes n}(\sigma)\psi_{\mathrm{g},w}(\sigma)$. We identify $w = (v, (\psi_{u,a})_{a \in [v(u)]})$ with its restriction, and also $(\psi_{u,a})_{a \in [1]} = \psi_u$.

Remark 3.129. We defined the 18 model pairs as they appear in the literature, apart from minor modifications (like the normalization constant for unconditional 'binomial' teacher-student model, discussed later in more detail). The combination $r = s = 1$ is used in proofs, e.g. in this contribution, and in software applications that tolerate repetitions, e.g. for performance reasons. The case $r = 2$ is used to prevent multiple occurrences of variables for weight functions that are not permutation invariant (e.g. k-SAT). The case $r = 3$ is used for permutation invariant weights, especially on bipartite or hypergraphs (e.g. occupation problems). The case $s = 2$ is used to enforce neighborhood uniqueness with order (e.g. bipartite graphs, with $r = 3$). The case $s = 3$ is used to enforce neighborhood

[5] Notice that the name Poisson model would be more appropriate in this case.

uniqueness without order (e.g. hypergraphs with $r = 3$). The binomial versions are used in proofs, e.g. this contribution, and discussions of binomial bipartite or hypergraphs.

In the following we use *main results* to refer to Theorem 3.97, Theorem 3.100, Lemma 3.101, Lemma 3.102 and Theorem 3.103, further to Observation 3.98 for completeness, and to Lemma 3.110, Lemma 3.111 for convergence in probability. With *target functions* we refer to the maps $\bar\phi_{\mathrm{a}}(m) = \frac{1}{n}\ln(\mathbb{E}[Z_{\mathrm{g}}(\boldsymbol{G})])$, $\bar\phi(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G})]$, $\bar\phi^*(m) = \mathbb{E}[\phi_{\mathrm{g}}(\boldsymbol{G}^*(\boldsymbol{\sigma}^*))]$, $\bar\phi_{\mathrm{a}}(m) = \frac{1}{n}\ln(\mathbb{E}[Z_{\mathrm{g}}(\boldsymbol{G})])$, $\iota(m) = \frac{1}{n}I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$ and $\delta(m) = \frac{1}{n}D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)\|\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}}, \boldsymbol{G})$ (cf. Section 3.5.4). These definitions directly extend to the uniform model pair, for any r, s.

This covers convergence for $\boldsymbol{m} \sim \mathrm{Po}(\bar dn/k)$, pointwise for $m \le m_\uparrow$, and for $\boldsymbol{m}^*$, for $\bar\phi_{\mathrm{a}}$, $\bar\phi^*$, $\delta(m)$, $\iota(m)$, and $\bar\phi(m)$ if $B_\uparrow = \phi_{\mathrm{a}}(\bar d)$, with respective order of convergence (recall $\rho \in (0, 1/4)$. This also covers bounds for the limits $\phi_{\mathrm{q}\uparrow}(\bar d), \phi_{\mathrm{q}\downarrow}(\bar d)$, boundedness and Lipschitz continuity of the target functions and convergence in probability.

The binomial models are defined for $\bar d$, not m, thus the situation is slightly different. The definitions $\bar\phi_{\mathrm{b}}(\bar d) = \mathbb{E}[\frac{1}{n}\ln(Z_{\mathrm{g}}(\boldsymbol{G}_{\mathrm{b}}))]$, $\bar\phi^*_{\mathrm{b}}(\bar d) = \mathbb{E}[\frac{1}{n}\ln(Z_{\mathrm{g}}(\boldsymbol{G}^*_{\mathrm{b}}(\boldsymbol{\sigma})))]$, $\delta_{\mathrm{b}}(\bar d) = \frac{1}{n}D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{\mathrm{b}}(\boldsymbol{\sigma}^*)\|\boldsymbol{\sigma}_{\mathrm{g},\boldsymbol{G}_{\mathrm{b}}}, \boldsymbol{G}_{\mathrm{b}})$ and $\iota_{\mathrm{b}}(\bar d) = \frac{1}{n}I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*_{\mathrm{b}}(\boldsymbol{\sigma}^*))$ are consistent with the uniform models over random factor counts. The annealed free entropy density for the binomial (null) model is $\bar\phi_{\mathrm{ba}}(\bar d) = \mathbb{E}[\frac{1}{n}\ln(\mathbb{E}[Z_{\mathrm{g}}(\boldsymbol{G}_{\mathrm{b}})|\|\boldsymbol{v}_{\mathrm{b}}\|_1])]$. Let $\boldsymbol{m}_{n,\bar d} \sim \mathrm{Po}(\overline{m})$ for $s = 1$, and $\boldsymbol{m} \sim \mathrm{Bin}(|\mathcal{U}|, p_{\mathrm{s}})$ otherwise.

Observation 3.130. *Notice that the following holds.*
a) The main results hold for $r = s = 1$, and are invariant to factors labels.
b) We have $\boldsymbol{G}_{\boldsymbol{m}} \sim \boldsymbol{G}_{\mathrm{b}}$ for all r, s (up to factor labels).
c) For $s = 1$ and the uniform model, the case $r = 2$ covers $r = 3$.
d) The case $s = 2$ covers $s = 3$.

Proof. For Part 3.130a) we notice that $r = s = 1$ is indeed exactly the case that we discussed. For $w = (v, \psi)$ from $\boldsymbol{w}$ let $\mathcal{A}_w = \{(v_a, \psi_a) : a \in [m]\}$, and $\mathcal{A}_w = \{(u, \psi_{u,a}) : u \in \mathcal{U}, a \in [v(u)]\}$ for $w = (v, \psi)$ from $\boldsymbol{w}_{\mathrm{b}}$. Then for any w from $\boldsymbol{w}$ or $\boldsymbol{w}_{\mathrm{b}}$ we have $\psi_{\mathrm{g},w}(\sigma) = \prod_{(u,\psi)\in\mathcal{A}_w}\psi(\sigma_u)$. The corresponding result for $[w]^\Gamma_{\gamma^*}$ is immediate. Next, we show that the target functions only depend on the graphs through the weights ψ_{g}. This holds for $Z_{\mathrm{g}}([w]^\Gamma) = \sum_\sigma \psi_{\mathrm{g},[w]^\Gamma}(\sigma)$ and hence for $\bar\phi_{\mathrm{a}}$, $\bar\phi$ and $\bar\phi^*$. The Radon-Nikodym derivative for δ is

$$r(\sigma, G) = \frac{\gamma^{*\otimes n}(\sigma)Z_{\mathrm{g}}(G)}{\mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}}(\sigma)]}.$$

Hence, this also holds for $\delta(m) = \frac{1}{n}\mathbb{E}[\ln(r(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)))]$. The Radon-Nikodym derivative for $\iota(m) = \frac{1}{n}D_{\mathrm{KL}}(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)\|\boldsymbol{\sigma}^* \otimes \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$ is

$$r(\sigma, G) = \frac{\psi_{\mathrm{g},G}(\sigma)/\mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}}(\sigma)]}{\sum_{\sigma^*}\gamma^{*\otimes n}(\sigma^*)\psi_{\mathrm{g},G}(\sigma^*)/\mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}}(\sigma^*)]},$$

so also $\iota(m) = \frac{1}{n}\mathbb{E}[\ln(r(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)))]$ only depends on ψ_{g}. Thus, Part 3.130a) holds.

For Part 3.130b) let $s = 1$. With $\boldsymbol{v}_{\mathrm{b}} \sim \mathrm{Po}(p_{\mathrm{b}})^{\otimes\mathcal{U}}$, with $\boldsymbol{m} \sim \mathrm{Po}(\overline{m})$, and with $\boldsymbol{v}_m \in \mathcal{U}$ being iid multinoulli with law $\mathrm{u}(\mathcal{U})$, due to $s = 1$, the frequencies $\boldsymbol{v}'_m = (|\boldsymbol{v}_m^{-1}(u)|)_u$ are multinomial with parameters m and $\mathrm{u}(\mathcal{U})$, so the claim follows with Observation 3.7b). For $s > 1$ the neighborhoods $\boldsymbol{v}_{\mathrm{b}} \sim \mathrm{Bin}(1, p_{\mathrm{b}})^{\otimes\mathcal{U}}$ are iid Bernoulli, $\boldsymbol{m} \sim \mathrm{Bin}(|\mathcal{U}|, p_{\mathrm{b}})$ is binomial, and $\boldsymbol{v}_m \sim \mathrm{u}(\mathcal{V})$, so the result follows with Observation 3.7e).

For Part 3.130c) we recall that $\boldsymbol{\psi}_\circ$ is permutation invariant almost surely for $r=3$, and from Part 3.130a) that $r=3$ is covered by $r=2$ if $\psi_{\mathrm{g},\boldsymbol{G}_{2,1}} \sim \psi_{\mathrm{g},\boldsymbol{G}_{3,1}}$ and $(\boldsymbol{\sigma}^*, \psi_{\mathrm{g},\boldsymbol{G}^*_{2,1}(\boldsymbol{\sigma}^*)}) \sim (\boldsymbol{\sigma}^*, \psi_{\mathrm{g},\boldsymbol{G}^*_{3,1}(\boldsymbol{\sigma}^*)})$. For $\boldsymbol{v}_\circ \sim \mathrm{u}(\mathcal{U}_2)$ we have $\boldsymbol{v}^\circ_\circ \sim \mathrm{u}(\mathcal{U}_3)$, where $u^\circ \in \mathcal{U}_3$ is given by $u^\circ([k]) = u([k])$, and thereby

$$\psi_{\mathrm{g},\boldsymbol{w}_2}(\sigma) = \prod_a \boldsymbol{\psi}_a(\sigma_{\boldsymbol{v}^\circ_a}) \sim \psi_{\mathrm{g},\boldsymbol{w}_3}(\sigma),$$

which shows that $\boldsymbol{G}^*_3(\sigma)$ is obtained from $\boldsymbol{G}^*_2(\sigma)$ via $u \mapsto u([k])$ and thus $\psi_{\mathrm{g},\boldsymbol{G}^*_2(\sigma)} \sim \psi_{\mathrm{g},\boldsymbol{G}^*_3(\sigma)}$. For Part 3.130d) and $w \in \mathcal{V}_{r,2} \times \mathcal{D}^m_\Psi$ we let $w^\circ = (v^\circ, \psi^\circ)$ be given by the unique $v^\circ \in \mathcal{V}_{r,3}$ with $v^\circ([m]) = v([m])$, and $\psi^\circ = (\psi_{v^{-1}(v^\circ(a))})_a$, and proceed analogously. □

Since $s=3$ is covered, we only work with labeled factors from here on.

3.5.7.2 Managing Expectations. In this section we focus on the expected weights $\overline{\psi}^\circ_{r,s}(\sigma) = \mathbb{E}[\psi_{\mathrm{g},\boldsymbol{w}}(\sigma)]$ and $\overline{\psi}_{r,s}(\sigma) = \mathbb{E}[\psi_{\mathrm{g},\boldsymbol{G}}(\sigma)] = \gamma^{*\otimes n}\overline{\psi}^\circ(\sigma)$. Let $\boldsymbol{w}^*_{r,1} = (\boldsymbol{v}^*, \boldsymbol{\psi}^*) \sim \boldsymbol{w}^{*\otimes m}_\circ$, where $\boldsymbol{w}^*_{\circ,r} = (\boldsymbol{v}^*_\circ, \boldsymbol{\psi}^*_\circ)$ is given by the derivative $(u,\psi) \mapsto \psi(\sigma_u)/Z_\mathrm{f}(\gamma_\sigma)$ with respect to $\boldsymbol{w}_{\circ,r} = (\boldsymbol{v}_\circ, \boldsymbol{\psi}_\circ) \sim \mathrm{u}(\mathcal{U}) \otimes \mu_\Psi$. Let $\boldsymbol{w}^*_{r,2} \sim (\boldsymbol{w}^*_{r,1} | \boldsymbol{v}^*_{r,1} \in \mathcal{V}_2)$.

Observation 3.131. *Notice that the following holds.*
a) We have $\overline{\psi}^\circ_{r,1}(\sigma) = \overline{Z}_\mathrm{f}(\gamma_\sigma)^m$.
b) We have $\overline{Z}_{\mathrm{f},r}(\gamma_\sigma) = \mathbb{E}[\overline{\psi}_\circ(\sigma_{\boldsymbol{v}_{\circ,r}})]$, *so* $\overline{Z}_{\mathrm{f},r}(\gamma_\sigma) = \frac{\mathbb{P}(\boldsymbol{v}^*_{\circ,1} \in \mathcal{U}_r)}{\mathbb{P}(\boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)} \overline{Z}_{\mathrm{f},1}(\gamma_\sigma)$.
c) We have $\overline{\psi}^\circ_{r,2}(\sigma) = \frac{\mathbb{P}(\boldsymbol{v}^*_{r,1} \in \mathcal{V}_2)}{\mathbb{P}(\boldsymbol{v}_{r,1} \in \mathcal{V}_2)} \overline{\psi}^\circ_{r,1}(\sigma)$.
d) We have $\boldsymbol{w}_{\circ,r} \sim (\boldsymbol{w}_{\circ,1} | \boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)$ *and* $\boldsymbol{w}^*_{\circ,r} \sim (\boldsymbol{w}^*_{\circ,1} | \boldsymbol{v}^*_{\circ,1} \in \mathcal{U}_r)$.
e) We have $\boldsymbol{w}_{r,1} \sim (\boldsymbol{w}_{\circ,1} | \boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)^{\otimes m}$, $\boldsymbol{G}^*_{r,1}(\sigma) \sim [\boldsymbol{w}^*_{r,1}]^\Gamma$ *and* $\boldsymbol{w}^*_{r,1} \sim (\boldsymbol{w}^*_{\circ,1} | \boldsymbol{v}^*_{\circ,1} \in \mathcal{U}_r)^{\otimes m}$.
f) We have $\boldsymbol{w}_{r,2} \sim (\boldsymbol{w}_{r,1} | \boldsymbol{v}_{r,1} \in \mathcal{V}_2)$ *and* $\boldsymbol{G}^*_{r,2}(\sigma) \sim [\boldsymbol{w}^*_{r,2}]^\Gamma$.
g) We have $\boldsymbol{G}^*_\mathrm{b}(\sigma) \sim \boldsymbol{G}^*_{\boldsymbol{m}}(\sigma)$ *for* $s=1$ *(up to relabeling of factors).*

Proof. The properties of the uniform distribution yield $\boldsymbol{v}_{r,1} \sim \mathrm{u}(\mathcal{U}_r)^{\otimes m}$, $\boldsymbol{v}_{\circ,r} \sim (\boldsymbol{v}_{\circ,1} | \boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)$ and $\boldsymbol{v}_{r,2} \sim (\boldsymbol{v}_{r,1} | \boldsymbol{v}_{r,1} \in \mathcal{V}_2)$. Thus, independence gives $\boldsymbol{w}_{\circ,r} \sim (\boldsymbol{w}_{\circ,1} | \boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)$, $\boldsymbol{w}_{r,1} \sim \boldsymbol{w}^{\otimes m}_{\circ,r} \sim (\boldsymbol{w}_{\circ,1} | \boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)^{\otimes m}$ and $\boldsymbol{w}_{r,2} \sim (\boldsymbol{w}_{r,1} | \boldsymbol{v}_{r,1} \in \mathcal{V}_2)$. We verify $\overline{Z}_{\mathrm{f},2} = \mathbb{E}[\overline{\psi}_\circ(\sigma_{\boldsymbol{v}_{\circ,2}})]$ analogous to Observation 3.11, which implies the result for $r=3$, analogous to the proof of Observation 3.130. This establishes Part 3.131a) and Part 3.131b) since

$$\overline{Z}_{\mathrm{f},r}(\gamma_\sigma) = \mathbb{E}\left[\overline{\psi}_\circ(\sigma_{\boldsymbol{v}_{\circ,1}}) \middle| \boldsymbol{v}_{\circ,1} \in \mathcal{U}_r\right] = \frac{\overline{Z}_{\mathrm{f},1}(\gamma_\sigma)\mathbb{E}\left[\frac{\overline{\psi}_\circ(\sigma_{\boldsymbol{v}_{\circ,1}})}{\overline{Z}_{\mathrm{f},1}(\gamma_\sigma)} \mathbb{1}\{\boldsymbol{v}_{\circ,1} \in \mathcal{U}_r\}\right]}{\mathbb{P}(\boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)} = \frac{\overline{Z}_{\mathrm{f},1}(\gamma_\sigma)\mathbb{P}(\boldsymbol{v}^*_{\circ,1} \in \mathcal{U}_r)}{\mathbb{P}(\boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)}.$$

Part 3.131c) is obtained analogously. This also yields $\frac{\psi(\sigma_v)\mathbb{1}\{v \in \mathcal{U}_r\}}{\overline{Z}_{\mathrm{f},r}(\gamma_\sigma)\mathbb{P}(\boldsymbol{v}_{\circ,1} \in \mathcal{U}_r)} = \frac{\psi(\sigma_v)\mathbb{1}\{v \in \mathcal{U}_r\}}{\overline{Z}_{\mathrm{f},1}(\gamma_\sigma)\mathbb{P}(\boldsymbol{v}^*_{\circ,1} \in \mathcal{U}_r)}$ and thereby Part 3.131d). With Part 3.131a) and analogously to Observation 3.13 we obtain $\boldsymbol{G}^*_{r,1}(\sigma) \sim [\boldsymbol{w}^*_{r,1}]^\Gamma$, and thereby Part 3.131e). Part 3.131f) follows analogously to Part 3.131d).

Part 3.131g) holds due to Observation 3.7b), since $\boldsymbol{v}^*_\mathrm{b} \sim \bigotimes_u \mathrm{Po}(p^*_\mathrm{b}(\sigma_u))$, thus $\|\boldsymbol{v}^*_\mathrm{b}\|_1 \sim \mathrm{Po}(\overline{m})$ since $\sum_u p^*_\mathrm{b}(\sigma_u) = \overline{m}\mathbb{E}[\overline{\psi}_\circ(\sigma_{\boldsymbol{v}_{\circ,r}})]/\overline{Z}_\mathrm{f}(\gamma_\sigma) = \overline{m}$, further since $\boldsymbol{m} \sim \mathrm{Po}(\overline{m})$, and since $\boldsymbol{v}^*_m \in \mathcal{U}$ is multinoulli with parameters m and $\boldsymbol{v}_\circ$, i.e. probabilities $\frac{\overline{\psi}_\circ(\sigma_u)}{\overline{Z}_\mathrm{f}(\gamma_\sigma)|\mathcal{U}|}$, thus $(|\boldsymbol{v}^{*-1}_m(u)|)_u$ is multinomial with parameters m and $\boldsymbol{v}_\circ$ and thereby $(|\boldsymbol{v}^{*-1}_m(u)|)_u \sim \boldsymbol{v}_\mathrm{b}$. □

Combining Observation 3.130 and Observation 3.131 yields the following. The main results hold for both model pairs for $r=s=1$, where the binomial model is exactly the model in Section 1. Thus, for $s = 1$ we only have to establish the results for the uniform model and $r = 2$, which covers the binomial model for $r = 2$ as well as the uniform model for $r = 3$, which in turn covers the binomial model for $r = 3$. Then, for each r and both model pairs, we are only left to derive $s = 2$ from $s = 1$, since $s = 3$ is covered by $s = 2$.

Next, we establish bounds for the error terms given by Observation 3.131.

Observation 3.132. *There exists* $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that the following holds.*

a) We have $\mathbb{P}(\boldsymbol{v}_{\circ,1} \notin \mathcal{U}_2), \mathbb{P}(\boldsymbol{v}^*_{\circ,1} \notin \mathcal{U}_2), |\ln(\frac{\overline{Z}_{\mathrm{f},2}(\gamma_\sigma)}{\overline{Z}_{\mathrm{f},1}(\gamma_\sigma)})| \le c/n$.

b) We have $\exp(-c\binom{m}{2}/|\mathcal{U}|) \le \mathbb{P}(\boldsymbol{v}_{r,1} \in \mathcal{U}^{\underline{m}}), \mathbb{P}(\boldsymbol{v}^*_{r,1} \in \mathcal{U}^{\underline{m}}) \le \exp(-\binom{m}{2}/(c|\mathcal{U}|))$.

Proof. The first claim holds since the $(\boldsymbol{v}^*_\circ, \boldsymbol{v}_\circ)$-derivative is uniformly bounded. For the second claim let $P = \mathbb{P}(\boldsymbol{v} \in \mathcal{U}_2)$ and $P^* = \mathbb{P}(\boldsymbol{v}^* \in \mathcal{U}_2)$. Using that the $(\boldsymbol{v}^*, \boldsymbol{v})$-derivative is uniformly bounded and induction over m, we obtain $\prod_{a=0}^{m-1}(1 - \psi_\uparrow^2 \frac{a}{|\mathcal{U}|}) \le P, P^* \le \prod_{a=0}^{m-1}(1 - \psi_\downarrow^2 \frac{a}{|\mathcal{U}|})$. Now, the assertion follows with standard arguments. □

3.5.7.3 Joint Distributions. In this section we establish joint distributions per model type over all parameters. We start with the coupling of $r = 1$ and $r = 2$ for $s = 1$, which then completes the discussion of $s = 1$. Then, we will couple $s = 1$ with $s = 2$ for each r and model type. For $w = (v, \psi) \in \mathcal{G}_{n,m}$, $w' = (v', \psi') \in \mathcal{G}_{n,m'}$ we recall the Hamming distance

$$\mathrm{d}(w, w') = \sum_{a=1}^{\min(m,m')} \mathbb{1}\{(v_a, \psi_a) \ne (v'_a, \psi'_a)\} + |m - m'|.$$

Observation 3.133. *There exist joint distributions* $(\boldsymbol{w}_{1,1}, \boldsymbol{w}_{1,2}, \boldsymbol{B})$, $(\boldsymbol{w}^*_{1,1}, \boldsymbol{w}^*_{1,2}, \boldsymbol{B})$ *and* $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\mathrm{d}(\boldsymbol{G}_{1,1}, \boldsymbol{G}_{1,2}) \le \boldsymbol{B}$ *almost surely,* $\mathrm{d}(\boldsymbol{G}^*_{1,1}(\sigma), \boldsymbol{G}^*_{1,2}(\sigma)) \le \boldsymbol{B}$ *almost surely, and* $\boldsymbol{B} \sim \mathrm{Bin}(m, c/n)$.

Proof. Let $\boldsymbol{w}' = (\boldsymbol{v}', \boldsymbol{\psi}')$ and $\boldsymbol{w}'_{\mathrm{r}} = (\boldsymbol{v}'_{\mathrm{r}}, \boldsymbol{\psi}'_{\mathrm{r}})$ be the null models for $r = 2$ and $r = 1$, or the corresponding teacher-student models. We couple $\boldsymbol{w}'$ and $\boldsymbol{w}'_{\mathrm{r}}$ using rejection sampling, i.e. let $(\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{\psi}}) \sim (\boldsymbol{v}'_{\mathrm{r},\circ}, \boldsymbol{\psi}'_{\mathrm{r},\circ})^{\otimes([m]\times\mathbb{Z}_{>0})}$, then we have $(\boldsymbol{v}'_{\mathrm{r},a}, \boldsymbol{\psi}'_{\mathrm{r},a})_a \sim (\tilde{\boldsymbol{v}}_{a,1}, \tilde{\boldsymbol{\psi}}_{a,1})_a$ and $(\boldsymbol{v}'_a, \boldsymbol{\psi}'_a)_a \sim (\tilde{\boldsymbol{v}}_{a,\boldsymbol{b}(a)}, \tilde{\boldsymbol{\psi}}_{a,\boldsymbol{b}(a)})_a$, where $\boldsymbol{b}(a) = \inf\{b \in \mathbb{Z}_{>0} : \tilde{\boldsymbol{v}}_{a,b} \in [n]^{\underline{k}}\}$. We use this coupling to define the joint distribution of $\boldsymbol{w}'$ and $\boldsymbol{w}'_{\mathrm{r}}$. Let $\boldsymbol{s} = (\mathbb{1}\{\boldsymbol{b}(a) = 1\})_a$ be the indicator of where the coupling succeeded and notice that $\boldsymbol{s} \sim \mathrm{Bin}(m, P')$, where $P' = \mathbb{P}(\boldsymbol{v}_{\circ,1} \in \mathcal{U}_2)$ for the null models and $P' = \mathbb{P}(\boldsymbol{v}^*_{\circ,1} \in \mathcal{U}_2)$ for the teacher-student models. This gives $\mathrm{d}(\boldsymbol{G}', \boldsymbol{G}'_{\mathrm{r}}) \le \boldsymbol{B}^\circ$ with $\boldsymbol{B}^\circ = (m - \boldsymbol{s}(a))_a \sim \mathrm{Bin}(m, 1 - P')$. Thus, the maximal coupling of $\boldsymbol{b}^\circ \sim \mathrm{Bin}(1, 1 - P')^{\otimes m}$ and $\boldsymbol{b} \sim \mathrm{Bin}(1, c/n)^{\otimes m}$, using Observation 3.132, completes the proof. □

This defines a joint distribution for all null models (covering both uniform and binomial), and a joint distribution for all planted models, for $s = 1$. Observation 3.7 shows that $\boldsymbol{B} \sim \mathrm{Po}(c\overline{m}/n)$ is Poisson with $c\overline{m}/n = c\overline{d}/k$ for the binomial models. The next result provides joint distributions for $s = 1$ and $s = 2$, for given r and model type.

Observation 3.134. *There exist* $(\boldsymbol{w}_{r,1}, \boldsymbol{w}_{r,2}, \boldsymbol{B})$, $(\boldsymbol{w}^*_{r,1}, \boldsymbol{w}^*_{r,2}, \boldsymbol{B})$, $(\boldsymbol{w}^*_{\mathrm{b},r,1}, \boldsymbol{w}^*_{\mathrm{b},r,2}, \boldsymbol{B}')$ *and* $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *such that* $\mathrm{d}(\boldsymbol{w}_{r,1}, \boldsymbol{w}_{r,2}) \le \boldsymbol{B}$, $\mathrm{d}(\boldsymbol{w}^*_{r,1}, \boldsymbol{w}^*_{r,2}) \le \boldsymbol{B}$, $\mathrm{d}(\boldsymbol{w}^*_{\mathrm{b},r,1}, \boldsymbol{w}^*_{\mathrm{b},r,2}) \le \boldsymbol{B}'$ *almost surely, where* $\boldsymbol{B} \sim \mathrm{Bin}(m, cm/|\mathcal{U}|)$ *and* $\boldsymbol{B}' \sim \mathrm{Bin}(|\mathcal{U}|, c\overline{m}^2/|\mathcal{U}|^2)$.

Proof. We follow the proof of Observation 3.133. Hence, we introduce the rejection sampling $(\tilde{\boldsymbol{v}}, \tilde{\boldsymbol{\psi}}) \sim (\boldsymbol{v}'_{\mathrm{r},\circ}, \boldsymbol{\psi}'_{\mathrm{r},\circ})^{\otimes([m]\times\mathbb{Z}_{>0})}$, notice that $(\boldsymbol{v}'_{\mathrm{r}}, \boldsymbol{\psi}'_{\mathrm{r}}) \sim (\tilde{\boldsymbol{v}}_{a,1}, \tilde{\boldsymbol{\psi}}_{a,1})_a$ and that $(\boldsymbol{v}', \boldsymbol{\psi}') \sim (\tilde{\boldsymbol{v}}_{a,\boldsymbol{b}(a)}, \tilde{\boldsymbol{\psi}}_{a,\boldsymbol{b}(a)})_a$, where

$\boldsymbol{b}(a) = \inf\{b : \forall a' \in [a-1]\, \tilde{\boldsymbol{v}}_{a,b} \neq \tilde{\boldsymbol{v}}_{a',\boldsymbol{b}(a')}\}$. For $\boldsymbol{B}^\circ = (\mathbb{1}\{\boldsymbol{b}(a) > 1\})_a$ we obtain the bound $\mathbb{P}(\boldsymbol{B}^\circ(a) = 1) \le \psi_\uparrow^2 m/|\mathcal{U}|$, thus we have a joint distribution $(\boldsymbol{G}', \boldsymbol{G}'_{\mathrm{r}}, \boldsymbol{B})$ with $\boldsymbol{B} \sim \mathrm{Bin}(m, \psi_\uparrow^2 m/|\mathcal{U}|)$.

For $\boldsymbol{G}^*_{\mathrm{b},1,2}(\sigma) = (\boldsymbol{v}', \boldsymbol{\psi}')$ and $\boldsymbol{G}^*_{\mathrm{b},1,1}(\sigma) = (\boldsymbol{v}'_{\mathrm{r}}, \boldsymbol{\psi}'_{\mathrm{r}})$ we have $\mathbb{P}(\boldsymbol{v}'_{\mathrm{r}}(u) > 0) = 1 - e^{-p^*_{\mathrm{b}}} \le p^*_{\mathrm{b}} = \mathbb{P}(\boldsymbol{v}'(u) > 0)$ and thus the maximal coupling yields $\boldsymbol{B}^\circ \sim \bigotimes_u \mathrm{Bin}(1, p^*_{\mathrm{b}}(\sigma_u) - (1 - e^{p^*_{\mathrm{b}}(\sigma_u)}))$. Using $e^x - 1 - x \le (e-2)x^2$ on $[0,1]$ we obtain the bound $(e-2)\psi_\uparrow^4 \overline{m}^2/|\mathcal{U}|^2$. □

3.5.7.4 Main Results. We extend the main results to all related models.

Corollary 3.135. *The main results apply for all $r, s \in [3]$ and the uniform models.*

Proof. For $w \in \mathcal{G}_{n,m}$ and $w' \in \mathcal{G}_{n,m'}$ we recall the bounds

$$\left|\ln\left(\frac{\psi_{\mathrm{g},w}(\sigma)}{\psi_{\mathrm{g},w'}(\sigma)}\right)\right|, \left|\ln\left(\frac{Z_{\mathrm{g}}([w]^\Gamma)}{Z_{\mathrm{g}}([w']^\Gamma)}\right)\right| \le 2\ln(\psi_\uparrow)\mathrm{d}(w,w'),$$
$$|\ln(\psi_{\mathrm{g},w}(\sigma))|, \left|\ln\left(Z_{\mathrm{g}}([w]^\Gamma)\right)\right| \le \ln(\psi_\uparrow)m.$$

The latter also yield $|\ln(\overline{\psi}^\circ_{r,s}(\sigma)|, |\ln(\overline{Z}_{r,s})| \le \ln(\psi_\uparrow)m$, using the shorthand $\overline{Z} = \mathbb{E}[Z_{\mathrm{g}}(\boldsymbol{G})]$, and thereby $|\bar{\phi}_{\mathrm{a}}(m)|, |\bar{\phi}(m)|, |\bar{\phi}^*(m)| \le \ln(\psi_\uparrow)m/n$. Using the proof of Observation 3.130, also the bounds $|\delta(m)| \le 2\ln(\psi_\uparrow)m/n$, $|\iota(m)| \le 4\ln(\psi_\uparrow)m/n$ follow. This justifies the restriction to $m \le m_\uparrow$. Observation 3.131 and Observation 3.132 give

$$\left|\ln\left(\frac{\overline{\psi}^\circ_{r,s}(\sigma)}{\overline{\psi}^\circ_{1,1}(\sigma)}\right)\right|, \left|\ln\left(\frac{\overline{Z}_{r,s}}{\overline{Z}_{1,1}}\right)\right| \le c\left(\frac{m^2}{|\mathcal{U}|} + \frac{m}{n}\right) \le c\left(\frac{m_\uparrow^2}{\binom{n}{k}} + \frac{m_\uparrow}{n}\right) \le C$$

for a sufficiently large constant $C_{\mathfrak{g}} > 0$. This directly yields $|\bar{\phi}_{\mathrm{a}}(m) - \bar{\phi}_{\mathrm{a},1,1}(m)| \le C/n$, which completes the discussion for $\bar{\phi}_{\mathrm{a}}$. With Observation 3.133 and Observation 3.134 we further get

$$\left|\bar{\phi}(m) - \bar{\phi}_{1,1}(m)\right|, \left|\bar{\phi}^*(m) - \bar{\phi}^*_{1,1}(m)\right| \le \frac{2\ln(\psi_\uparrow)}{n}\left(\frac{cm}{n} + \frac{cm^2}{|\mathcal{U}|}\right) \le \frac{C}{n},$$

which completes the discussion for the free entropies. The bound $|\delta(m) - \delta_{1,1}(m)| \le C/n$ follows analogous to the above. The nominator for $\iota(m)$ follows analogously, for the denominator we use that the bounds above for both the expected and the unexpected values are uniform, to recover $|\iota(m) - \iota_{r,s}(m)| \le C/n$. □

Before we establish the main results for the binomial models, we derive some basic results for the binomial teacher-student model for $s = 2$. Let $\boldsymbol{m}_{\mathrm{b}} = \|\boldsymbol{v}_{\mathrm{b}}\|_1$, $\boldsymbol{m}^*_{\mathrm{b}}(\sigma) = \|\boldsymbol{v}^*_{\mathrm{b}}\|_1$, $\bar{\boldsymbol{d}}_{\mathrm{b}} = k\boldsymbol{m}_{\mathrm{b}}/n$ and $\bar{\boldsymbol{d}}^*_{\mathrm{b}}(\sigma) = k\boldsymbol{m}^*(\sigma)/n$. For $i \in [n]$ let $\boldsymbol{d}^*_{\mathrm{b},\sigma}(i) = |\{u \in \boldsymbol{v}^{*-1}_{\mathrm{b}}(1) : i \in u([k])\}|$.

Observation 3.136. *Notice that the following holds.*
a) We have $\mathbb{E}[\boldsymbol{m}_{\mathrm{b}}] = \mathbb{E}[\boldsymbol{m}^*_{\mathrm{b}}(\sigma)] = \overline{m}$ and $\mathbb{E}[\bar{\boldsymbol{d}}_{\mathrm{b}}] = \mathbb{E}[\bar{\boldsymbol{d}}^*_{\mathrm{b}}(\sigma)] = \bar{d}$.
b) There exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $\mathbb{P}(|\bar{\boldsymbol{d}}_{\mathrm{b}} - \bar{d}| \ge r), \mathbb{P}(|\bar{\boldsymbol{d}}^*_{\mathrm{b}}(\sigma) - \bar{d}| \ge r) \le c_1 \exp\left(-\frac{c_2 r^2 n}{1+r}\right)$.
c) We have $\mathbb{E}[\boldsymbol{d}^*_{\mathrm{b},\boldsymbol{\sigma}^*}(i)] = (1 + \mathcal{O}(1/\sqrt{n}))\bar{d}$ and $\mathbb{P}(|\bar{\boldsymbol{d}}^*_{\mathrm{b},\boldsymbol{\sigma}^*}(i) - \bar{d}| \ge r) \le c_1 \exp\left(-\frac{c_2 r^2 n}{1+r}\right)$.

Proof. For $s = 1$ we have $\boldsymbol{m}_{\mathrm{b}}(\sigma) \sim \boldsymbol{m}_{\mathrm{b}} \sim \boldsymbol{m}$. For $s = 2$, using Observation 3.131 notice that $\mathbb{E}[\boldsymbol{m}^*_{\mathrm{b}}(\sigma)] = \sum_u p^*_{\mathrm{b}}(\sigma_u) = \overline{m}$. Hence, with the Chernoff-Hoeffding Theorem we recover the bounds in

Corollary 3.12 for both $\boldsymbol{m}_{\mathrm{b}}$ and $\boldsymbol{m}^*_{\mathrm{b}}(\sigma)$ and thus $\bar{\boldsymbol{d}}_{\mathrm{b}}$, $\bar{\boldsymbol{d}}^*_{\mathrm{b}}(\sigma)$. For the second part, we have

$$\mathbb{E}\left[\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)\right] = \bar{d}\sum_{\tau}\frac{1}{k}\frac{\overline{\psi}_{\circ}(\tau)}{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}\frac{n|\{u\in\mathcal{U}: i\in u([k]),\sigma_u=\tau\}|}{|\mathcal{U}|}.$$

in all cases. The case $r=3$ reduces to $r=2$, and we reduce $r=1$ to $r=2$ at the cost of the relative error $\mathcal{O}(1/n)$. For $r=2$ there exists a unique position with $u_h=i$, hence we have

$$\lambda_\sigma(i) = \mathbb{E}\left[\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)\right] = (1+\mathcal{O}(1/n))\bar{d}\sum_{h}\frac{1}{k}\sum_{\tau}\frac{\overline{\psi}_{\circ}(\tau)}{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}\frac{n|\{u\in\mathcal{U}: u_h=i,\sigma_u=\tau\}|}{|\mathcal{U}|}.$$

Earning another $\mathcal{O}(1/n)$, we move back to $r=1$ to obtain $\lambda_\sigma(i) = (1+\mathcal{O}(1/n))\bar{d}\frac{\mu_{\gamma_\sigma}|_*(\sigma_i)}{\gamma_\sigma(\sigma_i)}$, with μ_{γ_σ} from Observation 3.9. Lipschitz continuity and Observation 3.23 yield the expectation result. For $s=1$ we have $\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)\sim\mathrm{Po}(\lambda_\sigma(i))$, while for $s=2$ the degree $\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)$ is a sum of independent Bernoullis with average success probability $\lambda_\sigma(i)/|\mathcal{U}(i)|$, where $\mathcal{U}(i)=\{u\in\mathcal{U}: i\in u([k])\}$. Bernstein's inequality gives $\mathbb{P}(|\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)-\lambda_\sigma(i)|\geq r)\leq c_1\exp(-c'\frac{r^2}{1+r)})$ for both cases, and $c_{\mathfrak{g}}\in\mathbb{R}^2_{>0}$. The result follows with Lipschitz continuity and standard methods. □

This result implies that for *any* type of ground truth distribution, the average number of factors and the average degree do not reveal any information about the chosen ground truth. It also shows that for γ^* the individual degrees do not reveal any information, in particular not on the color of the vertex. Looking into the proof, the result is significantly stronger. We showed that the variable degrees do not reveal any information if and only if γ^* is a stationary point of $\overline{Z}_{\mathrm{f}}$.

Corollary 3.137. *All main results apply for the binomial model for $s=1$. For $s=2$ all main results for $\bar{\phi}_{\mathrm{a}}$, $\bar{\phi}$, $\bar{\phi}^*$ hold. Further, we have $\iota_{r,2}(\bar{d})=\iota_{r,1}(\bar{d})+\mathcal{O}(1/n^{k-1})$ and $\delta_{r,2}(\bar{d})=\delta_{r,1}(\bar{d})+\mathcal{O}(1/n^{k-1})$, thus all results hold except for the pointwise asymptotics.*

Proof. Since we entirely focus on the binomial model in this proof, we discard the subscript b. Corollary 3.135 with Observation 3.130 and Observation 3.131 establishes the main results for the binomial model for $s=1$, where we notice that all target functions for the binomial model are expectations over $\|\boldsymbol{v}\|_1\sim\boldsymbol{m}$, as for the case $r=s=1$ from Section 1.

The case $s=2$ differs from all previous cases because $\|\boldsymbol{v}^*\|_1$ and $\|\boldsymbol{v}\|_1$ do not have the same law. Still, Observation 3.130 shows that $\bar{\phi}$ and $\bar{\phi}_{\mathrm{a}}$ are special cases of Corollary 3.135. Also, Observation 3.136 shows that $\boldsymbol{m}^*$ is of the same quality as the Poisson distribution, so the pointwise results from Corollary 3.135 imply the result for $\bar{\phi}^*$.

Thus, we turn to $\iota(\bar{d})=\mathbb{E}[\frac{1}{n}\ln(r_{\mathrm{i},\boldsymbol{\sigma}^*}(\boldsymbol{w}^*_{\boldsymbol{\sigma}^*}))]$, where $(\sigma,w)\mapsto r_{\mathrm{i},\sigma}(w)$ is the $(\boldsymbol{\sigma}^*,\boldsymbol{w}^*_{r,2,\boldsymbol{\sigma}^*})$ to $\boldsymbol{\sigma}^*\otimes\boldsymbol{w}^*_{r,2,\boldsymbol{\sigma}^*}$ derivative. We approximate the binomial with the Poisson distribution to reduce the case $s=2$ to $s=1$. Hence, we write $r_{\mathrm{i},\sigma}(w)=r_\sigma(w)r^{\circ}_\sigma(w)/\bar{r}(w)$ in terms of the $(\boldsymbol{\sigma}^*,\boldsymbol{w}^*_{r,2,\sigma})$ to $(\boldsymbol{\sigma}^*,\boldsymbol{w}^*_{r,1,\sigma})$ derivative $r_\sigma(w)$, the $(\boldsymbol{\sigma}^*,\boldsymbol{w}^*_{r,1,\sigma})$ to $\boldsymbol{\sigma}^*\otimes\boldsymbol{w}^*_{r,1,\sigma}$ derivative $r^{\circ}_\sigma(w)$, and the $\boldsymbol{\sigma}^*\otimes\boldsymbol{w}^*_{r,2,\sigma}$ to $\boldsymbol{\sigma}^*\otimes\boldsymbol{w}^*_{r,1,\sigma}$ derivative $\bar{r}(w)$. With $p^*_\sigma(\tau)=\frac{\overline{\psi}_{\circ}(\tau)}{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}p$ and $\rho_\sigma(\tau)=p_\sigma(\tau)\frac{\psi_u(\tau)}{\overline{\psi}_{\circ}(\tau)}$, we have

$$r_\sigma(v,\psi) = \frac{\prod_{u\in v^{-1}(1)}\rho_\sigma(\sigma_u)\prod_{u\in v^{-1}(0)}(1-p^*_\sigma(\sigma_u))}{\prod_{u\in v^{-1}(1)}(e^{-p^*_\sigma(\sigma_u)}\rho_\sigma(\sigma_u))\prod_{u\in v^{-1}(0)}e^{-p^*_\sigma(\sigma_u)}} = \exp\left(\mathcal{O}(p\overline{m})+\mathcal{O}(p)\|v\|_1\right),$$

$$\bar{r}(v,\psi) = \frac{\mathbb{E}\left[\prod_{u\in v^{-1}(1)}\rho_{\boldsymbol{\sigma}^*}(\boldsymbol{\sigma}^*_u)\prod_{u\in v^{-1}(0)}(1-p^*_{\boldsymbol{\sigma}^*}(\boldsymbol{\sigma}^*_u))\right]}{\mathbb{E}\left[\prod_{u\in v^{-1}(1)}(e^{-p^*_{\boldsymbol{\sigma}^*}(\boldsymbol{\sigma}^*_u)}\rho_{\boldsymbol{\sigma}^*}(\boldsymbol{\sigma}^*_u))\prod_{u\in v^{-1}(0)}e^{-p^*_{\boldsymbol{\sigma}^*}(\boldsymbol{\sigma}^*_u)}\right]} = \exp\left(\mathcal{O}(p\overline{m})+\mathcal{O}(p)\|v\|_1\right).$$

Recall that we take the logarithm and the expectation to obtain ι, so Observation 3.136 yields

$$\iota_{r,2}(\bar{d}) = \mathbb{E}\left[\frac{1}{n}\ln\left(r^{\circ}_{\boldsymbol{\sigma}^*}\left(\boldsymbol{w}^*_{r,2,\boldsymbol{\sigma}^*}\right)\right)\right] + \mathcal{O}\left(\frac{p\overline{m}}{n}\right) = \mathbb{E}\left[\frac{1}{n}\ln\left(r^{\circ}_{\boldsymbol{\sigma}^*}\left(\boldsymbol{w}^*_{r,2,\boldsymbol{\sigma}^*}\right)\right)\right] + \mathcal{O}\left(\frac{1}{n^{k-1}}\right).$$

Now, we recovered the the target function for $s = 1$, that is, the log-density

$$\iota^{\circ}_{\sigma}(v,\psi) = \frac{1}{n}\ln\left(r^{\circ}_{\sigma}(v,\psi)\right) = \frac{1}{n}\ln\left(\frac{\prod_u\prod_{a=1}^{v(u)}\frac{\psi_{u,a}(\sigma_u)}{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}}{\mathbb{E}\left[\prod_u\prod_{a=1}^{v(u)}\frac{\psi_{u,a}(\boldsymbol{\sigma}^*_u)}{\overline{Z}_{\mathrm{f}}(\gamma_{\boldsymbol{\sigma}^*})}\right]}\right).$$

Recall the distance from Section 3.5.7.3 and notice that $|\iota^{\circ}_{\sigma}(w)-\iota^{\circ}_{\sigma}(w')| \le \frac{c}{n}\mathrm{d}(w,w')$ for some $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$, same as for the log-densities of all preceding derivatives. Hence, we couple $\boldsymbol{w}^*_{r,2,\sigma}$ and $\boldsymbol{w}^*_{r,1,\sigma}$, which amounts to coupling $\mathrm{Bin}(p^*_{\sigma}(\tau))$ and $\mathrm{Po}(p^*_{\sigma}(\tau))$, due to independence and since the weights have the same law. We take $\mathbb{P}(\boldsymbol{x}_{\circ,\sigma,\tau} = x) = p(x)$ with

$$p(0,0) = 1-p^*,\ p(0,1) = e^{-p^*} - (1-p^*),\ p(a,1) = e^{-p^*}\frac{p^{*a}}{a!},\ a\in\mathbb{Z}_{>0},\ p^* = p^*_{\sigma}(\tau),$$

which gives $\boldsymbol{x}_{\circ}(1) \sim \mathrm{Po}(p^*)$ and $\boldsymbol{x}_{\circ}(2) \sim \mathrm{Bin}(1,p^*)$. This induces the coupling $(\boldsymbol{x}_{\sigma},\boldsymbol{\psi}^*) \sim \bigotimes_u(\boldsymbol{x}_{\circ,\sigma,\sigma_u}\otimes \boldsymbol{\psi}^*_{\circ,\sigma_u})$, which bounds the distance with $\boldsymbol{D}_{\sigma} = \sum_u|\boldsymbol{x}_{\sigma,u}(2)-\boldsymbol{x}_{\sigma,u}(1)|$, and thereby

$$\varepsilon(\sigma) = \left|\mathbb{E}\left[\iota^{\circ}_{\sigma}(\boldsymbol{w}^*_{r,2,\sigma})\right] - \mathbb{E}\left[\iota^{\circ}_{\sigma}(\boldsymbol{w}^*_{r,1,\sigma})\right]\right| \le \frac{c}{n}\mathbb{E}[\boldsymbol{D}_{\sigma}] = \frac{c|\mathcal{U}|}{n}\mathbb{E}[\boldsymbol{D}_{\circ}],$$

where for $\boldsymbol{D}_{\circ} = |\boldsymbol{x}_{\circ}(2)-\boldsymbol{x}_{\circ}(1)|$ we have $\mathbb{P}(\boldsymbol{D}_{\circ} = D) = p(D)$ with $p(0) = 1-p^*+e^{-p^*}p^*$, $p(1) = e^{-p^*}-(1-p)^*+\frac{1}{2}e^{-p^*}p^{*2}$ and $p(D) = \frac{1}{(D+1)!}e^{-p^*}p^{*(D+1)}$ for $D\in\mathbb{Z}_{\ge 2}$. Notice that with $D \le D+1$ we obtain $Dp(D) \le p^*p(D-1) = p^*\mathbb{P}(\boldsymbol{x}_{\circ}(1) = D)$ for $D \ge 2$, and thereby

$$\mathbb{E}[\boldsymbol{D}_{\circ}] = p(1) + p^*\mathbb{P}(\boldsymbol{x}_{\circ}(1)\ge 2) = e^{-p^*}-(1-p^*)-p^*(e^{-p^*}-1+\frac{1}{2}e^{-p^*}p^*) \le cp^2 = \frac{c\overline{m}^2}{|\mathcal{U}|^2}$$

for some $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$. This gives $\varepsilon(\sigma) = \mathcal{O}(1/n^{k-1})$, uniformly in σ, so taking the expectation over $\boldsymbol{\sigma}^*$ and Jensen's inequality yield $|\iota_{r,2}(\bar{d})-\iota_{r,1}(\bar{d})| = \mathcal{O}(1/n^{k-1})$.

The proof for $\delta(\bar{d}) = \mathbb{E}[\frac{1}{n}\ln(r_{\mathrm{d},\boldsymbol{\sigma}^*}(\boldsymbol{w}^*_{\boldsymbol{\sigma}^*}))]$ over the $(\boldsymbol{\sigma}^*,\boldsymbol{w}^*_{\boldsymbol{\sigma}^*})$ to $(\boldsymbol{\sigma}_{\mathrm{g},[\boldsymbol{w}]^{\Gamma}},\boldsymbol{w})$ derivative $r_{\mathrm{d},\sigma}(w)$ is similar. With $r_{\sigma}(v,\psi) = \mathcal{O}(p\overline{m}) + \mathcal{O}(p)\|v\|_1$ from above, further with the $(\boldsymbol{w}_{r,2},\boldsymbol{w}_{r,1})$-derivative

$$r'(v,\psi) = \frac{p^{\|v\|_1}(1-p)^{|\mathcal{U}|-\|v\|_1}}{e^{-p\|v\|_1}p^{\|v\|_1}e^{-p(|\mathcal{U}|-\|v\|_1)}} = \exp(\mathcal{O}(p\overline{m})+\mathcal{O}(p)\|v\|_1),$$

and with the $(\boldsymbol{\sigma}^*,\boldsymbol{w}^*_1(\boldsymbol{\sigma}^*))$ to $(\boldsymbol{\sigma}_{\mathrm{g},[\boldsymbol{w}_1]^{\Gamma}},\boldsymbol{w}_1)$ derivative r°_{σ} we obtain $\delta_{r,2}(\bar{d}) = \mathbb{E}[\delta^{\circ}_{\sigma}(\boldsymbol{\sigma}^*,\boldsymbol{w}^*_{r,2}(\boldsymbol{\sigma}^*))] + \mathcal{O}(1/n^{k+1})$, where

$$\delta^{\circ}_{\sigma}(v,\psi) = \frac{1}{n}\ln(r^{\circ}_{\sigma}(v,\psi)) = \frac{1}{n}\ln\left(\frac{\mathbb{E}\left[\prod_u\prod_{a=1}^{v(u)}\psi_{u,a}(\boldsymbol{\sigma}^*_u)\right]}{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)^{\|v\|_1}}\right).$$

Clearly, this map is Lipschitz, so we can reuse our coupling to establish the claim. □

Finally, we notice that replacing $\overline{Z}_{\mathrm{f}}(\gamma_\sigma)$ by ξ in the binomial teacher-student model has no impact on the results.

Observation 3.138. *Corollary 3.137 also holds with $\overline{Z}_{\mathrm{f}}(\gamma_\sigma)$ replaced by ξ.*

Proof. We suppress the subscripts b for the binomial model. Notice that these models do not depend on r, and that the results for $\bar{\phi}_{\mathrm{a}}$ and $\bar{\phi}$ are not affected. Hence, let $r=1$, $s\in[2]$, and let $f(\bar{d})=\mathbb{E}[f^\circ_{\boldsymbol{\sigma}^*}(\boldsymbol{w}^*)]$ be one of the remaining target functions from Corollary 3.137, i.e. $f=\bar{\phi}^*$ with $f^\circ_\sigma(w)=\phi_{\mathrm{g}}([w]^\Gamma)$, $f=\iota$ with $f^\circ_\sigma(w)=\frac{1}{n}\ln(r_{\mathrm{i}}(\sigma,w))$, or $f=\delta$ with $f^\circ_\sigma(w)=\frac{1}{n}\ln(r_{\mathrm{d}}(\sigma,w))$. Let g, g° denote the counterparts with $\overline{Z}_{\mathrm{f}}(\gamma_\sigma)$ replaced by ξ. In the following we bound $|f^\circ-g^\circ|$ and introduce a coupling of the two corresponding planted models. For the latter, notice that the weights have the law given by $\psi(\sigma_u)/\overline{\psi}_\circ(\sigma_u)$ with respect to μ_Ψ in both models, so it is sufficient to couple the neighborhoods.

First, notice that f°,g° are bounded and Lipschitz, in the sense that $|f^\circ_\sigma(v,\psi)|,|g^\circ_\sigma(v,\psi)|\le\frac{c\|v\|_1}{n}$ and $|f^\circ_\sigma(w)-f^\circ_\sigma(w')|,|g^\circ_\sigma(w)-g^\circ_\sigma(w')|\le\frac{c}{n}\mathrm{d}(w,w')$ for a constant $c_{\mathfrak{g}}>0$. Recall from Observation 3.131 that $|\ln(\overline{Z}_{\mathrm{f}}(\gamma_\sigma)/\overline{Z}_{\mathrm{f},1}(\gamma_\sigma))|\le c/n$. Recall that $\overline{Z}_{\mathrm{f},1}\le\xi$, and from Observation 3.9 that $\overline{Z}_{\mathrm{f},1}(\gamma)/\xi\ge\max(1-c\|\gamma-\gamma^*\|^2_{\mathrm{tv}},\psi^2_\downarrow)$ for some other constant $c_{\mathfrak{g}}$.

Let $p^*_\sigma(\tau)=\frac{\overline{\psi}_\circ(\tau)}{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}p$, $p^\bullet(\tau)=\frac{\overline{\psi}_\circ(\tau)}{\xi}p\le p^*_\sigma(\tau)$ be success probabilities, let $\boldsymbol{v}^*_{1,\sigma}\sim\bigotimes_u\mathrm{Po}(p^*_\sigma(\sigma_u))$, $\boldsymbol{v}^\bullet_{1,\sigma}\sim\bigotimes_u\mathrm{Po}(p^\bullet_\sigma(\sigma_u))$ be neighborhoods for $s=1$, and $\boldsymbol{v}^*_{2,\sigma}\sim\bigotimes_u\mathrm{Bin}(p^*_\sigma(\sigma_u))$, $\boldsymbol{v}^\bullet_{2,\sigma}\sim\bigotimes_u\mathrm{Bin}(p^\bullet_\sigma(\sigma_u))$ neighborhoods for $s=2$. Notice that $\boldsymbol{m}^*_2(\sigma)=\sum_\tau\boldsymbol{m}^\star_\sigma(\tau)$ is a sum of independent binomials $\boldsymbol{m}^\star\sim\bigotimes_\tau\mathrm{Bin}(|\mathcal{U}^*_\sigma(\tau)|,p^*_\sigma(\tau))$, where $\mathcal{U}^*_\sigma(\tau)=\{u\in\mathcal{U}:\sigma_u=\tau\}$. This does not only suggest that $\boldsymbol{m}^*_2$ is subpoissonian, i.e. $\mathbb{E}[e^{\boldsymbol{m}^*_2(\sigma)t}]\le\mathbb{E}[e^{\boldsymbol{m}t}]$, it also suggests that for the reweighted counts $\hat{\boldsymbol{m}}_2$, we have $\hat{\boldsymbol{m}}_2\sim\boldsymbol{m}^*_2(\sigma)+1$ and $\hat{\boldsymbol{m}}_1\sim\boldsymbol{m}+1$, which holds for both sums of independent binomials, and for Poisson distributions. Thus, this also holds for $\boldsymbol{m}'_i(\sigma)=\|\boldsymbol{v}^\bullet_{i,\sigma}\|_1$ and their reweighted counterparts $\hat{\boldsymbol{m}}'_i$. Notice that due to the replacement we obtain $\mathbb{E}[\boldsymbol{m}'_i(\sigma)]=\frac{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}{\xi}\overline{m}\le\overline{m}=\mathbb{E}[\boldsymbol{m}]=\mathbb{E}[\boldsymbol{m}^*(\sigma)]$ and $\mathbb{E}[\hat{\boldsymbol{m}}'_i]=\frac{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}{\xi}\overline{m}+1\le\overline{m}+1=\mathbb{E}[\hat{\boldsymbol{m}}_i]$. On the other hand, recall that $\overline{Z}_{\mathrm{f}}(\gamma_\sigma)/\xi\ge\psi^2_\downarrow$, thus the ratio of the expectations is uniformly bounded. Using that f°, $g^\circ\le c\|v\|_1/n$ are bounded, the tail expectation $\mathbb{E}[\mathbb{1}\{\boldsymbol{m}_*\ge m_\uparrow\}\boldsymbol{m}_*]=\mathbb{E}[\boldsymbol{m}_*]\mathbb{P}(\hat{\boldsymbol{m}}_*\ge m_\uparrow)=o(n^{-k})$ is negligible for any choice of factor count $\boldsymbol{m}_*$, and for $\boldsymbol{m}^*<m_\uparrow$ the tartget functions are $f^\circ,g^\circ\le cm_\uparrow/n\le cd_\uparrow/k$ are uniformly bounded. Let $B_{\mathfrak{g}}=C\sqrt{\ln(n)/n}$ for some large $C_{\mathfrak{g}}\in\mathbb{R}_{>0}$, so that the probability bound in Observation 3.23 is of order $\mathcal{O}(1/n^k)$. Then we have

$$f(\bar{d})=\mathbb{E}\left[\mathbb{1}\left\{\|\boldsymbol{v}^*\|_1<m_\uparrow,\|\gamma_{\boldsymbol{\sigma}^*}-\gamma^*\|_{\mathrm{tv}}<B\right\}f^\circ_{\boldsymbol{\sigma}^*}(\boldsymbol{w}^*)\right]+\mathcal{O}\left(1/n^k\right)$$

and the corresponding result for g. Using $p^\bullet\le p^*$, we obtain the coupling $\boldsymbol{v}^*_1(u)=\boldsymbol{v}^\bullet_1(u)+\boldsymbol{\varepsilon}_1(u)$ with $\boldsymbol{\varepsilon}_1\sim\bigotimes_u\mathrm{Po}(p^*(\sigma_u)-p^\bullet(\sigma_u))$ and the coupling $\boldsymbol{v}^*_2(u)=\boldsymbol{v}^\bullet_2(u)+\boldsymbol{\varepsilon}_2(u)$, where $\boldsymbol{\varepsilon}_2\sim\bigotimes_u\boldsymbol{\varepsilon}_2(u)\in\{0,1\}^{\mathcal{U}}$ is given by $\boldsymbol{\varepsilon}_2(u)=0$ on $\boldsymbol{v}^\bullet_2(u)=1$ and $(\boldsymbol{\varepsilon}_2(u)|\boldsymbol{v}^\bullet_2(u)=0)\sim\mathrm{Bin}(1,(p^*(\sigma_u)-p^\bullet(\sigma_u))/(1-p^\bullet(\sigma_u)))$. From the factor count discussion, we obtain $\mathbb{E}[\|\boldsymbol{\varepsilon}_i\|_1]=\left(1-\frac{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}{\xi}\right)\overline{m}$. With B we further obtain the upper bound $\mathbb{E}[\|\boldsymbol{\varepsilon}_i\|_1]\le c\ln(n)$ for some $c_{\mathfrak{g}}>0$. Now, Lipschitz continuity yields

$$f(\bar{d})=\mathbb{E}\left[\mathbb{1}\left\{\|\boldsymbol{v}^\bullet\|_1<m_\uparrow,\|\gamma_{\boldsymbol{\sigma}^*}-\gamma^*\|_{\mathrm{tv}}<B\right\}f^\circ_{\boldsymbol{\sigma}^*}(\boldsymbol{w}^\bullet)\right]+\mathcal{O}\left(\ln(n)/n\right),$$

where $(\boldsymbol{\sigma}^*,\boldsymbol{w}^\bullet)$ is the distribution of the weight-adjusted model, in particular $\boldsymbol{w}^\bullet=(\boldsymbol{v}^\bullet,\boldsymbol{\psi}^\bullet)$ with $\boldsymbol{\psi}^\bullet\sim\boldsymbol{\psi}^*$. This completes the proof for $\bar{\phi}^*$, since in this case $f^\circ=g^\circ$. For the remaining four functions ι°_i, δ°_i we track the required replacements to replace f° by g°. We only discuss ι, since the discussion for δ is analogous, but simpler. For the binomial version $s=2$ we have to take care of $(\overline{Z}_{\mathrm{f}}(\gamma_\sigma)/\xi)^{\|v\|_1}$ and $((1-p^*)/(1-p^\bullet))^{|\mathcal{U}|-\|v\|_1}$, in the nominator and denominator each. The log-density of the first error is $\mathcal{O}(1)\frac{\overline{m}}{n}\ln(1+\mathcal{O}(\ln(n)/n)) = \mathcal{O}(\ln(n)/n)$, due to the discussion of the factor count expectations above, and using B. The base of the second error can be rewritten as $1-\frac{\xi-\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}{(1-p^\bullet)\overline{Z}_{\mathrm{f}}(\gamma_\sigma)\xi}\overline{\psi}_\circ(\sigma_u)p = 1-\mathcal{O}(\ln(n)p/n)$, thus the log-density is once more of order $\mathcal{O}(\ln(n)/n)$. For the case $s=1$ we also have to account for $(\overline{Z}_{\mathrm{f}}(\gamma_\sigma)/\xi)^{\|v\|_1}$, analogous to the above. However, instead of the quotient we now have to introduce $\exp(\frac{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}{\xi}\overline{m})$ to the nominator and the denominator (the cooresponding terms $e^{\overline{m}}$ for f° cancel out), where the term in the denominator sits inside the expectation. However, by rewriting the term as $e^{\overline{m}}\exp((1-\frac{\overline{Z}_{\mathrm{f}}(\gamma_\sigma)}{\xi})\overline{m})$, the first part cancels out in the nominator and the denominator, while the log-density of the latter is, again, of order $\mathcal{O}(\ln(n)/n)$. □

3.5.8 Degrees and Balance. In this section we discuss the special role of the stationary points from Remark 3.10, in particular why inference problems can be easily solved whenever γ is not stationary. We only present the discussion for the uniform model with $r=s=1$ in Section 3.5.7.

3.5.8.1 Degree Distributions. In this section we prepare the discussion of unbalanced problems in the next section by establishing convergence of the empirical degree distributions. Thus, we assume neither **POS** nor **BAL**. First, we show convergence of the empirical distribution for iid discrete integrable random variables. Then, we show convergence of the empirical degree distribution when edges are drawn uniformly. This immediately yields the result for the null model. For the planted model, we establish concentration of the ground truth and factor assignment frequencies, which then allows to control the degree distribution also in this case.

For the first part let $\mu_{\mathrm{n},n,x} = (\frac{1}{n}|x^{-1}(y)|)_{y\in\mathbb{Z}}$ be the empirical distribution of $x\in(\mathbb{Z}^d)^n$.

Observation 3.139. *For $d\in\mathbb{Z}_{>0}$, $c\in\mathbb{R}_{>0}$, $\delta:\mathbb{Z}_{>0}\to\mathbb{R}_{>0}$, $\delta=\omega(n^{-1/(3d+2)})$, there exists $\varepsilon_{d,c,\delta}=o(1)$ such that the following holds. For $\boldsymbol{x}_\circ\in\mathbb{Z}^d$ with $\mathbb{E}[\|\boldsymbol{x}_\circ\|_\infty]\le c$ we have $\mathbb{P}(\|\mu_{\mathrm{n},\boldsymbol{x}}-\mu\|_{\mathrm{tv}}\ge\delta)\le\varepsilon$, where $\boldsymbol{x}_\circ\sim\mu$ and $\boldsymbol{x}_n\sim\mu^{\otimes n}$.*

Proof. Let $Y:\mathbb{Z}_{>0}\to\mathbb{R}_{>0}$ with $Y=o(n^{1/\beta})$, $\beta=3d+2$ be sufficiently large, $\eta=1/Y^{d+1}$, and $\mathcal{Y}=\{y\in\mathbb{Z}^d:\|y\|_\infty\le Y\}$. Let $\boldsymbol{\mu}=\mu_{\mathrm{n},\boldsymbol{x}}$ and notice that $n\boldsymbol{\mu}(y)=\sum_i\mathbb{1}\{\boldsymbol{x}_i=y\}\sim\mathrm{Bin}(n,\mu(y))$ for all $y\in\mathbb{Z}^d$. For $y\in\mathcal{Y}$, Chebyshev's inequality yields

$$\mathbb{P}\left(|n\boldsymbol{\mu}(y)-n\mu(y)|\ge n\eta\right)\le\frac{n\mu(y)(1-\mu(y))}{n^2\eta^2}<\frac{1}{n\eta^2}.$$

This yields $\mathbb{P}(\boldsymbol{x}\notin\mathcal{X})\le\varepsilon$ for $\mathcal{X}=\{x\in\mathbb{Z}^n:\forall y\in\mathcal{Y}\,|\mu_{\mathrm{n},x}(y)-\mu(y)|<\eta\}$ with the union bound, where $\varepsilon=\frac{|\mathcal{Y}|}{n\eta^2}$. For $x\in\mathcal{X}$ we have $\mu_{\mathrm{n},x}(\mathcal{Y})>\mu(\mathcal{Y})-|\mathcal{Y}|\eta$ and thereby $\mu_{\mathrm{n},x}(\mathbb{Z}^d\setminus\mathcal{Y})<\mu(\mathbb{Z}^d\setminus\mathcal{Y})+|\mathcal{Y}|\eta$, which gives $\|\mu_{\mathrm{n},x}-\mu\|_1<|\mathcal{Y}|\eta+\mu_{\mathrm{n},x}(\mathbb{Z}\setminus\mathcal{Y})+\mu(\mathbb{Z}\setminus\mathcal{Y})<2\delta(n)$, where $\delta(n)=|\mathcal{Y}|\eta+\frac{c}{Y}$, using Markov's inequality. Finally, notice that $|\mathcal{Y}|=\Theta(Y^d)$, so $\varepsilon=\Theta(Y^\beta)=o(1)$ and $\delta=\Theta(n^{-1/\beta})$. □

Better bounds can be derived from [20], in particular for the Poisson distribution. Now, we turn to the empirical degree distribution. For a given total degree $D\in\mathbb{Z}_{\ge0}$ let $\boldsymbol{V}_{n,D}\sim\mathrm{u}([n])^{\otimes D}$. For given $v\in[n]^D$ let $d_{\mathrm{D},v}=(|v^{-1}(i)|)_i$ be the degree sequence and $\mu_{\mathrm{D},v}=\mu_{\mathrm{n},d_{\mathrm{D},v}}$ the empirical degree distribution. Finally, let $\boldsymbol{\mu}_{\mathrm{D},n,D}=\mu_{\mathrm{D},\boldsymbol{V}_{n,D}}$.

Observation 3.140. *Let $\alpha\in\mathbb{R}_{>0}$ and $\delta_{\mathrm{D}},\varepsilon_{\mathrm{D}}:\mathbb{Z}_{>0}\to\mathbb{R}_{>0}$ with $\delta_{\mathrm{D}},\varepsilon_{\mathrm{D}}=o(1)$. There exists $\delta=\delta(\alpha,\delta_{\mathrm{D}},\varepsilon_{\mathrm{D}})$ and $\varepsilon=\varepsilon(\alpha,\delta_{\mathrm{D}},\varepsilon_{\mathrm{D}})$ with $\delta,\varepsilon=o(1)$ such that the following holds. For $D\in[0,\alpha n]$ and $\boldsymbol{D}\in\mathbb{Z}_{\ge0}$ such that $\mathbb{P}(|\boldsymbol{D}-D|>\delta_{\mathrm{D}}n)\le\varepsilon_{\mathrm{D}}$ we have $\mathbb{P}(\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}}-\mathrm{Po}(\frac{D}{n})\|_{\mathrm{tv}}>\delta)\le\varepsilon$.*

Proof. Fix some $\eta = o(1)$, $D^* \le \alpha n$, let $\boldsymbol{D}^* \sim \mathrm{Po}(D^*)$ and assume that $\delta_{\mathrm{D}}, \varepsilon_{\mathrm{D}} = o(1)$ are sufficiently slowly decreasing, so $\boldsymbol{D}^*$ is a choice for $\boldsymbol{D}$. Let $\mathcal{D} = \{D \in \mathbb{Z}_{\ge 0} : |D - D^*| \le \delta_{\mathrm{D}} n\}$, then we have

$$\mathbb{P}\left(\left\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}} - \mathrm{Po}\left(\frac{D^*}{n}\right)\right\|_{\mathrm{tv}} > \eta\right) \le \mathbb{P}\left(\left\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}} - \mathrm{Po}\left(\frac{D^*}{n}\right)\right\|_{\mathrm{tv}} > \eta, \boldsymbol{D} \in \mathcal{D}\right) + \varepsilon_{\mathrm{D}}.$$

Now, let $\boldsymbol{U} \sim \mathrm{u}([n])^{\otimes \mathbb{Z}_{>0}}$ and redefine $\boldsymbol{\mu}_{\mathrm{D},D} = \mu_{\mathrm{D},\boldsymbol{U}_{[D]}}$. Notice that adding an edge changes exactly one degree, so we have $\|\boldsymbol{\mu}_{\mathrm{D},D_1} - \boldsymbol{\mu}_{\mathrm{D},D_2}\|_{\mathrm{tv}} \le \frac{1}{n}|D_1 - D_2|$ for all $D \in \mathbb{Z}_{\ge 0}^2$. With the triangle inequality we have $\|\boldsymbol{\mu}_{\mathrm{D},D_1} - \boldsymbol{\mu}_{\mathrm{D},D_2}\|_{\mathrm{tv}} \le 2\delta_{\mathrm{D}}$ for all $D \in \mathcal{D}^2$, so for $(\boldsymbol{D}, \boldsymbol{D}^*) \sim \boldsymbol{D} \otimes \boldsymbol{D}^*$ we have

$$\mathbb{P}\left(\left\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}} - \mathrm{Po}\left(\frac{D^*}{n}\right)\right\|_{\mathrm{tv}} > \eta\right) \le \mathbb{P}\left(\left\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*} - \mathrm{Po}\left(\frac{D^*}{n}\right)\right\|_{\mathrm{tv}} > \eta - 2\delta_{\mathrm{D}}\right) + 2\varepsilon_{\mathrm{D}}.$$

Thus, we consider $\eta = \omega(\delta_{\mathrm{D}})$ with $\eta > 2\delta_{\mathrm{D}}$. Now, notice that the degree sequence $\boldsymbol{d}_D = d_{\boldsymbol{V}_D}$ is multinomial with D samples over the distribution $\mathrm{u}([n])$, so Observation 3.7 yields that $\boldsymbol{d}_{\boldsymbol{D}^*} \sim \mathrm{Po}(\frac{D^*}{n})^{\otimes n}$. Now, restricting η to $\eta = \omega(n^{-1/5})$, we summon Observation 3.139 with $d = 1$, $c = \alpha$ and $\delta = \eta - 2\delta_{\mathrm{D}} = (1 + o(1))\eta = \omega(n^{-1/5})$ to obtain $\varepsilon' = \varepsilon_{c,\delta} = o(1)$. But for $\boldsymbol{x}_\circ \sim \mathrm{Po}(\frac{D^*}{n})$ we have $\mathbb{E}[\boldsymbol{x}_\circ] = D^*/n \le \alpha$, and for $\boldsymbol{x} = \boldsymbol{d}_{\boldsymbol{D}^*} \sim \boldsymbol{x}_\circ^{\otimes n}$ we have $\mu_{\mathrm{n},\boldsymbol{x}} = \boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*}$, since $\mu_{\mathrm{D},v} = \mu_{\mathrm{n},d_{\mathrm{D},v}}$, so

$$\mathbb{P}\left(\left\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*} - \mathrm{Po}\left(\frac{D^*}{n}\right)\right\|_{\mathrm{tv}} > \eta - 2\delta_{\mathrm{D}}\right) \le \varepsilon'.$$

The choice $\delta = \eta$ and $\varepsilon = \varepsilon' + 2\varepsilon_{\mathrm{D}}$ completes the proof. □

Now, we immediately obtain the asymptotics for the null model $\boldsymbol{G} = (\boldsymbol{v}, \boldsymbol{\psi})$. The variable degrees are $\boldsymbol{d}_{\mathrm{m},m} = (|\boldsymbol{v}^{-1}(i)|)_i$ and their empirical distribution is $\boldsymbol{\delta}_m = \mu_{\mathrm{n},\boldsymbol{d}_{\mathrm{m}}}$.

Observation 3.141. *There exist $\delta_{\mathfrak{g}}, \varepsilon_{\mathfrak{g}} = o(1)$ such that $\mathbb{P}(\|\boldsymbol{\delta}_{\boldsymbol{m}^*} - \mathrm{Po}(\bar{d})\|_{\mathrm{tv}} > \delta) < \varepsilon$.*

Proof. Let $\alpha = d_\uparrow$, $\delta_{\mathrm{D}} = \delta_{\mathrm{m}}$ and $\varepsilon_{\mathrm{D}} = \varepsilon_{\mathrm{m}}$, and summon Observation 3.140 to obtain $\delta_{\mathfrak{g}}, \varepsilon_{\mathfrak{g}} = o(1)$. For $D = \bar{d}n$, $\boldsymbol{D} = k\boldsymbol{m}^*$ we have $\mathbb{P}(|\boldsymbol{D} - D| > \delta_{\mathrm{D}} n) = \mathbb{P}(|\frac{k\boldsymbol{m}^*}{n} - \bar{d}| > \delta_{\mathrm{m}}) \le \varepsilon_{\mathrm{m}} = \varepsilon_{\mathrm{D}}$, $D = \bar{d}n \le d_\uparrow n = \alpha n$ and $\boldsymbol{v}_{\boldsymbol{m}^*} \sim \boldsymbol{V}_{\boldsymbol{D}}$, so $\boldsymbol{\delta}_{\boldsymbol{m}^*} \sim \boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}}$, which directly yields $\mathbb{P}(\|\boldsymbol{\delta}_{\boldsymbol{m}^*} - \mathrm{Po}(\bar{d})\|_{\mathrm{tv}} > \delta) < \varepsilon$. □

The situation for the planted model is far more involved. The strategy is to recover uniformly distributed edges using Observation 3.15, for fixed σ, m and τ. However, for this purpose we still need concentration of $\boldsymbol{\tau}^*_{m,\sigma}$. Clearly, we can restrict to

$$\mathcal{M}_{\bar{d},n} = \left\{m \in \mathbb{Z}_{\ge 0} : \left|\frac{km}{n} - \bar{d}\right| \le \delta_{\mathrm{m}}\right\}, \quad \mathcal{S}_{n,\gamma^*} = \left\{\sigma \in [q]^n : \|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}} \le \frac{\ln(n)}{\sqrt{n}}\right\}.$$

Let $\mu_{\gamma_{\mathrm{n},\sigma}} = \mu_{\mathrm{T}|\Gamma,\gamma_{\mathrm{n},\sigma}}$ and recall that $\boldsymbol{\tau}^*_{m,\sigma} \sim \mu_{\gamma_{\mathrm{n},\sigma}}^{\otimes m}$. Let $\boldsymbol{\mu}_{\mathrm{T},m,\sigma} = \mu_{\mathrm{n},m,\boldsymbol{\tau}^*_{m,\sigma}}$ for $m > 0$.

Observation 3.142. *There exists $\delta_{\mathfrak{g}}, \varepsilon_{\mathfrak{g}} : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$ with $\delta, \varepsilon = o(1)$ such that the following holds. For $\bar{d} \ge \sqrt{\delta_{\mathrm{m}}} + 1/\ln(n)$, $m \in \mathcal{M}$ and $\sigma \in \mathcal{S}$ we have $\mathbb{P}(\|\boldsymbol{\mu}_{\mathrm{T},m,\sigma} - \mu_{\gamma_{\mathrm{n},\sigma}}\|_{\mathrm{tv}} > \delta) \le \varepsilon$.*

Proof. Observation 3.139 does not help in this case, thus we notice that $\boldsymbol{\mu}_{\mathrm{T}}(\tau) \sim \mathrm{Bin}(m, \mu_{\gamma_{\mathrm{n},\sigma}}(\tau))$. But now, Hoeffding's inequality completes the proof, since it implies that $\mathbb{P}(|\boldsymbol{\mu}_{\mathrm{T}}(\tau) - \mu_{\gamma_{\mathrm{n},\sigma}}(\tau)| \ge \delta) \le 2\exp(-2\delta^2 m) \le 2\exp(-2\delta^2(\sqrt{\delta_{\mathrm{m}}} + 1/\ln(n) - \delta_{\mathrm{m}})n/k)$. □

We are finitely ready to derive the asymptotics of the planted model. For this purpose let $\boldsymbol{d}^*(\tau) \sim \mathrm{Po}(\bar{d}\frac{\mu_{\gamma^*}|_*(\tau)}{\gamma^*(\tau)})$, $\boldsymbol{\tau}^* \sim \gamma^*$, and let $\mu^* \in \mathcal{P}(\mathbb{Z})$ be given by $\boldsymbol{d}^*(\boldsymbol{\tau}^*) \sim \mu^*$. Further, for the planted model $\boldsymbol{G}^*_{n,m}(\sigma) = (\boldsymbol{v}^*_{m,\sigma}, \boldsymbol{\psi}^*_{m,\sigma})$ let $\boldsymbol{d}^*_{\mathrm{m},m,\sigma} = (|\boldsymbol{v}^{*-1}(i)|)_i$ and $\boldsymbol{\delta}^*_{m,\sigma} = \mu_{\mathrm{n},\boldsymbol{d}^*_{\mathrm{m}}}$.

Observation 3.143. *There exists $\delta_{\mathfrak{g}}, \varepsilon_{\mathfrak{g}} = o(1)$ such that $\mathbb{P}(\|\boldsymbol{\delta}^*_{\boldsymbol{m}^*,\boldsymbol{\sigma}^*} - \mu^*\|_{\mathrm{tv}} > \delta) < \varepsilon$.*

Proof. We will mostly consider the colors separately. Thus, recall from Observation 3.9 that $\rho(\tau') = \mu_{\gamma^*}|_*(\tau')/\gamma^*(\tau')$ is uniformly bounded and let $\boldsymbol{d}^*(\tau') \sim \mu^*_{\tau'} = \mathrm{Po}(\bar{d}\rho(\tau'))$. On the other hand, let $\mathcal{I}_\sigma(\tau') = \{i \in [n] : \sigma_i = \tau'\}$ and $\mathcal{A}_\tau(\tau') = \{(a,h) : \tau_{a,h} = \tau'\}$, for Observation 3.15. Let $\boldsymbol{w}^\star_m(\sigma,\tau) = (\boldsymbol{v}^\star_{\sigma,\tau}, \boldsymbol{\psi}^\star_{\sigma,\tau})$ from Observation 3.15. Further, let $\boldsymbol{v}^\star_{\sigma,\tau} = (\boldsymbol{v}^\star_{\sigma,\tau,\tau'})_{\tau'}$ be the decomposition into the independent components $\boldsymbol{v}^\star_{\sigma,\tau,\tau'} \sim \mathrm{u}(\mathcal{I}(\tau'))^{\otimes\mathcal{A}(\tau')}$. This induces the independent degree sequences $\boldsymbol{d}^\star_{\sigma,\tau,\tau'} = (|\boldsymbol{v}^{\star-1}(i)|)_{i\in\mathcal{I}(\tau')}$. Let $N_\sigma(\tau') = |\mathcal{I}_\sigma(\tau')|$ and $D_\tau(\tau') = |\mathcal{A}_\tau(\tau')|$. In the following we use Observation 3.23 to restrict to $\sigma \in \mathcal{S}$ and choose n sufficiently large such that $\gamma_{\mathrm{n},\sigma} \ge \psi_\downarrow/2$, yielding $\mathcal{I}_\sigma(\tau') = \Theta(n)$ and in particular $N(\tau') > 0$. Thus, we have $\boldsymbol{v}^\star_{\tau'} \sim \boldsymbol{V}_{N(\tau'),D(\tau')}$, where $\boldsymbol{V}_{N,D} = \boldsymbol{U}_{N,[D]}$ and $\boldsymbol{U}_N \sim \mathrm{u}([N])^{\otimes\mathbb{Z}_{>0}}$, using the identifications $\mathcal{I}(\tau') = [N(\tau')]$ and $\mathcal{A}(\tau') = [D(\tau')]$, suggesting that Observation 3.140 is applicable. Let $\boldsymbol{d}_{N,D} = d_{\mathrm{D},\boldsymbol{V}_{N,D}}$ and $\boldsymbol{\mu}_{N,D} = \mu_{\mathrm{D},N,\boldsymbol{V}_{N,D}}$. Let $\boldsymbol{D}_{m,\sigma}(\tau') = D_{\boldsymbol{\tau}^*_{m,\sigma}}(\tau')$ and $\boldsymbol{D}^*_\sigma(\tau') = \boldsymbol{D}_{\boldsymbol{m}^*}(\tau')$.

First, consider $\bar{d} \le d_-$ with $d_- = \sqrt{\delta_{\mathrm{m}}} + 1/\ln(n) = o(1)$. Let $\mu_\circ \in \mathcal{P}(\mathbb{Z})$ be given by $\mu_\circ(0) = 1$. Then we have $\|\mu^*_{\tau'} - \mu_\circ\|_{\mathrm{tv}} = \mathcal{O}(d_-)$ and thus $\|\mu^* - \mu_\circ\|_{\mathrm{tv}} = \mathcal{O}(d_-)$. On the other hand, we use $D(\tau') \le km$ to obtain $\boldsymbol{d}_{N(\tau'),D(\tau')} \le \boldsymbol{d}_{N(\tau'),km}$ and thereby $\|\boldsymbol{\mu}_{N(\tau'),D(\tau')} - \mu_\circ\|_{\mathrm{tv}} \le \|\boldsymbol{\mu}_{N(\tau'),km} - \mu_\circ\|_{\mathrm{tv}}$. Now, consider Observation 3.140 with $\alpha = 1$, $\delta_{\mathrm{D}} = d_- + \delta_{\mathrm{m}} = o(1)$ and $\varepsilon_{\mathrm{D}} = \varepsilon_{\mathrm{m}}$, to obtain δ', ε'. For $D = 0$ we have

$$\mathbb{P}(|k\boldsymbol{m}^* - D| > \delta_{\mathrm{D}} n) = \mathbb{P}(k\boldsymbol{m}^* > d_- n + \delta_{\mathrm{m}} n) \le \mathbb{P}(k\boldsymbol{m}^* > \bar{d}n + \delta_{\mathrm{m}} n) \le \varepsilon_{\mathrm{m}} = \varepsilon_{\mathrm{D}}.$$

This shows that $\mathbb{P}(\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*(\tau')} - \mu_\circ\|_{\mathrm{tv}} > \delta') \le \mathbb{P}(\|\boldsymbol{\mu}_{\mathrm{D},k\boldsymbol{m}^*} - \mu_\circ\|_{\mathrm{tv}} > \delta') \le \varepsilon'$.

Now, we turn to $\bar{d} \ge d_-$. We use δ'', ε'' from Observation 3.142 to restrict to $m \in \mathcal{M}$ and $\tau \in ([q]^k)^m$ such that $\|\mu_{\mathrm{n},m,\tau} - \mu_{\gamma_{\mathrm{n},\sigma}}\|_{\mathrm{tv}} \le \delta''$, whp with failure probability $\varepsilon_{\mathrm{m}} + \varepsilon''$. Using Lipschitz continuity of μ_γ from Observation 3.9 and $\sigma \in \mathcal{S}$, we obtain $\|\mu_{\mathrm{n},m,\tau} - \mu_{\gamma^*}\|_{\mathrm{tv}} \le \tilde{\delta}$ for some $\tilde{\delta}_{\mathfrak{g}} = o(1)$. Thus, we have $\|\mu_{\mathrm{n},m,\tau}|_* - \mu_{\gamma^*}|_*\|_{\mathrm{tv}} \le \tilde{\delta}$ and $\|\mu_{\mathrm{n},m,\tau}|_* - \mu_{\gamma^*}|_*\|_\infty \le 2\tilde{\delta}$. This shows that $|D(\tau') - km\mu_{\gamma^*}|_*(\tau')| \le 2\tilde{\delta}km$. With Observation 3.9 we have $\mu_{\gamma^*}|_*(\tau') \ge c\gamma^*(\tau') \ge c\psi_\downarrow$, so we have $D(\tau') = (1 + \mathcal{O}(\tilde{\delta}))km\mu_{\gamma^*}|_*(\tau')$. With $\bar{d} \ge d_- = \omega(\delta_{\mathrm{m}})$ we further have $D(\tau') = (1 + o(1))\bar{d}\mu_{\gamma^*}|_*(\tau')n$. On the other side, we clearly have $N(\tau') = (1 + o(1))\gamma^*(\tau')n$. Now, for sufficiently large n, with $\bar{d} \le d_\uparrow$ and Observation 3.9 we obtain $\alpha_{\mathfrak{g}}$ such that $D(\tau')/N(\tau') \le \alpha$. Now, we summon Observation 3.140 with α and any $\delta_{\mathrm{D}}, \varepsilon_{\mathrm{D}}$ to obtain $\delta''', \varepsilon'''$. Then, for $\boldsymbol{D} = D = D(\tau')$ and n replaced by $N(\tau') = \Theta(n)$, we have $\mathbb{P}(\|\boldsymbol{\mu}_{\mathrm{D},D(\tau')} - \mathrm{Po}(\frac{D(\tau')}{N(\tau')})\|_{\mathrm{tv}} > \delta'''(N(\tau'))) \le \varepsilon'''(N(\tau'))$. Since we have $N_\sigma(\tau') = \Theta(n)$ uniformly for all $\sigma \in \mathcal{S}$ and τ', this yields bounds in n. Finally, since we have uniform relative errors for $\frac{D(\tau')}{N(\tau')} = (1 + o(1))\bar{d}\rho(\tau')$, we obtain $\|\mathrm{Po}(\frac{D(\tau')}{N(\tau')}) - \mu^*_{\tau'}\|_{\mathrm{tv}} = o(1)$. This yields bounds $\delta'', \varepsilon'' = o(1)$ with $\mathbb{P}(\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*(\tau')} - \mu^*_{\tau'}\|_{\mathrm{tv}} > \delta'') \le \varepsilon''$. For $\delta^\circ = \max(\delta', \delta'')$ and $\varepsilon^\circ = \max(\varepsilon', \varepsilon'')$ we have $\mathbb{P}(\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*(\tau')} - \mu^*_{\tau'}\|_{\mathrm{tv}} > \delta^\circ) \le \varepsilon^\circ$ in any case. Thus, we have $\|\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*(\tau')} - \mu^*_{\tau'}\|_{\mathrm{tv}} \le \delta^\circ$ for all τ' jointly with probability at least $1 - q\varepsilon^\circ$ using the union bound. On this event we have $\|\sum_{\tau'} \gamma^*(\tau')\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*(\tau')} - \mu^*\|_{\mathrm{tv}} \le \delta^\circ$, so we are left to bound the total variation distance $\|\sum_{\tau'} \gamma^*(\tau')\boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*(\tau')} - \boldsymbol{\delta}^*_{\boldsymbol{m}^*,\sigma}\|_{\mathrm{tv}} = \|\sum_{\tau'} \boldsymbol{\mu}_{\mathrm{D},\boldsymbol{D}^*(\tau')}(\gamma^*(\tau') - \frac{1}{n}N(\tau'))\|_{\mathrm{tv}} \le \|\gamma^* - \gamma_{\mathrm{n},\sigma}\|_{\mathrm{tv}} = o(1)$. Thus, with the triangle inequality we obtain the desired bound. □

For a finite index set $\mathcal{I} \ne \emptyset$, a finite edge set $\mathcal{A}$ and a graph $v \in \mathcal{I}^{\mathcal{A}}$ let $d_v = (|v^{-1}(i)|)_i$ be the degree sequence of v, in particular $d_v \equiv 0$ for $\mathcal{A} = \emptyset$, and let $\delta_v = (\frac{1}{n}|d_v^{-1}(d)|)_d \in \mathcal{P}(\mathbb{Z}_{\geq 0})$ be the degree distribution of v.

One of two crucial ingredients in the study of the degree distribution asymptotics is the empirical distribution of iid Poisson variables.

Observation 3.144. *Let $\varepsilon_{\mathfrak{g}}, D_{\mathfrak{g}} \in \mathbb{R}_{>0}$, $d^* \in [0, D]$ and $\mathcal{I} \subseteq [n]$ be such that $|\mathcal{I}| \geq \varepsilon n$. For $\boldsymbol{d} \sim \mathrm{Po}(d^*)^{\otimes \mathcal{I}}$ let $\boldsymbol{\delta} = (\frac{1}{|\mathcal{I}|}|\boldsymbol{d}^{-1}(d)|)_d$, then there exist $r_{\mathfrak{g}}, \varepsilon'_{\mathfrak{g}} = o(1)$ with $\mathbb{P}(\|\boldsymbol{\delta} - \mathrm{Po}(d^*)\|_{\mathrm{tv}} \geq r) \leq \varepsilon'$.*

Proof. Without loss of generality we assume that $\mathcal{I} = [N]$. Fix $D_+ = \omega(1)$ slowly increasing. Notice that $N\boldsymbol{\delta}(d) = \sum_i \mathbb{1}\{\boldsymbol{d}(i) = d\} \sim \mathrm{Bin}(N, P(d))$ for $d \in \mathbb{Z}_{\geq 0}$, where $P(d) = \mathbb{P}(\boldsymbol{d}(1) = d)$. Hence, Chebyshev's inequality yields $\mathbb{P}(|\boldsymbol{\delta}(d) - P(d)| \geq N^{-1/4}) \leq N^{-1/2}$, which yields the assertion combined with the union bound for $d < D_+$, with

$$r(n) = \frac{D_+}{(\varepsilon n)^{1/4}} + \mathbb{P}(\boldsymbol{d}_1 \geq D_+),\ \varepsilon'(n) = \frac{D_+}{(\varepsilon n)^{1/2}}.$$

□

Now, we are ready to prove the main result of this section

Observation 3.145.

Proof. For the null model, recall $\boldsymbol{v}_m \sim \mathrm{u}(([n]^k)^m)$. Let $\boldsymbol{U}_n \sim \mathrm{u}([n])^{\otimes \mathbb{Z}_{>0}}$ and $\boldsymbol{V}_{n,M} = \boldsymbol{U}_{[M]} \sim \mathrm{u}([n]^M)$, then we have $\boldsymbol{v}_m \sim \boldsymbol{V}_{km}$. Now, let $\boldsymbol{d} \sim \mathrm{Po}(\bar{d})^{\otimes n}$ and $\boldsymbol{D} = \|\boldsymbol{d}\|_1$. Recall from Observation 3.7 that $\boldsymbol{d} \sim d_{\boldsymbol{v}_{\boldsymbol{D}}}$. So, Observation 3.144 applies to $\delta_{\boldsymbol{v}_{\boldsymbol{D}}}$. Notice that for M, M' we have $\|\delta_{\boldsymbol{V}_M} - \delta_{\boldsymbol{V}_{M'}}\|_{\mathrm{tv}} \leq \frac{1}{n}|M - M'|$. Thus, for any $\boldsymbol{D}'$ that concentrates around $\bar{d}n$ we recover Observation 3.144, in particular for $\boldsymbol{D}' = k\boldsymbol{m}^*$. This completes the proof for the null model using that $\boldsymbol{\delta} \sim \delta_{\boldsymbol{V}_{\boldsymbol{D}'}}$.

For the planted model recall Observation 3.15. So, given m, σ and $\tau \in ([q]^k)^m$ we consider $\boldsymbol{v}_{\tau'} \sim \mathrm{u}(\mathcal{I}_\sigma(\tau')^{\mathcal{A}_\tau(\tau')})$ with $\mathcal{I}_\sigma(\tau') = \sigma^{-1}(\tau')$ and $\mathcal{A}_\tau(\tau') = \{(a, h) : \tau_{a,h} = \tau'\}$. Using $\mathcal{I}_\sigma(\tau') = [|\sigma^{-1}(\tau')|] = [n\gamma_{\mathrm{n},\sigma}(\tau')]$ and $\mathcal{A}_\tau(\tau') = [\alpha_\tau(\tau')]$, where $\alpha_\tau(\tau') = |\mathcal{A}_\tau(\tau')|$, for simplified notation, we thus have $\boldsymbol{v}_{\tau'} \sim \boldsymbol{V}_{n\gamma_{\mathrm{n},\sigma}(\tau'),\alpha_\tau(\tau')}$. For $D_{-,\mathfrak{g}} = o(1)$ slowly decreasing, in particular $D_- \geq 2\delta_{\mathrm{m}}$, and $\bar{d} \leq D_-$ we can restrict to $m \leq (\delta_{\mathrm{m}} + D_-)n/k$, so $\alpha_\tau(\tau') \leq km$ with the Lipschitz continuity above yields that the total variation of $\delta_{\boldsymbol{v}_{\tau'}}$ and the one-point mass on 0 is greater than $\delta_{\mathrm{m}} + D_-$ with probability at most ε_{m}. For $\bar{d} \geq D_-$ we have $\boldsymbol{m}^* \geq (D_- - \delta_{\mathrm{m}})n/k = \Theta(D_- n)$ with probability at least $1 - \varepsilon_{\mathrm{m}}$. With $\mu_{\gamma_{\mathrm{n},\sigma}} = \mu_{\mathrm{T}|\Gamma,\gamma_{\mathrm{n},\sigma}}$ from Section 3.2.1.2, the assignment frequencies $\boldsymbol{\beta}_{m,\sigma} = (|\boldsymbol{\tau}^{*-1}_{m,\sigma}(\tau)|)_\tau$ for $\boldsymbol{\tau}^*$ from Observation 3.15 are multinomial with m samples over the distribution $\mu_{\gamma_{\mathrm{n},\sigma}}$. So, with the bounds on m the Bretagnolle–Huber–Carol inequality yields $\mathbb{P}(\|\frac{1}{m}\boldsymbol{\beta}_{\sigma,m} - \mu_{\gamma_{\mathrm{n},\sigma}}\|_{\mathrm{tv}} \geq \varepsilon) \leq 2^{q^k} e^{-2m\varepsilon^2} \leq 2^{q^k} e^{-cD_- n\varepsilon^2}$. for some $c_{\mathfrak{g}} > 0$ and $n \geq n_{\circ,\mathfrak{g}}$. Notice that this bound is uniform in σ. Finally, we also restrict σ using Observation 3.23. Using that D_- is sufficiently slowly decreasing, we let $r(n) = \ln(n)/\sqrt{n}$, further $\beta_\tau = (\frac{1}{m}|\tau^{-1}(\tau')|)_{\tau'}$, $m > 0$, and restrict to the whp event

$$\mathcal{T} = \left\{(m, \sigma, \tau) : \left|m - \frac{\bar{d}n}{k}\right| \leq \frac{\delta_{\mathrm{m}} n}{k}, \|\gamma_{\mathrm{n},\sigma} - \gamma^*\|_{\mathrm{tv}} \leq r(n), \|\beta_\tau - \mu_{\gamma_{\mathrm{n},\sigma}}\|_{\mathrm{tv}} \leq r(n)\right\}.$$

Now, we can use Lipschitz continuity of $\mu_{\gamma_{\mathrm{n},\sigma}}$ from Observation 3.9 to obtain $\|\beta_\tau - \mu_{\gamma^*}\|_{\mathrm{tv}} \leq (L+1)r(n)$. This shows that $\alpha_\tau(\tau') = (1 + o(1))\bar{d}n\mu_{\gamma^*}|_*(\tau')$ for $\tau \in \mathcal{T}$, using $D_- = \omega(D_-)$. On the other hand we have $n\gamma_{\mathrm{n},\sigma}(\tau') = (1 + o(1))n\gamma^*(\tau')$. Now, consider the degrees $\boldsymbol{d}_{\tau'} \sim \mathrm{Po}(\frac{\alpha_\tau(\tau')}{n\gamma_{\mathrm{n},\sigma}(\tau')})^{\otimes n\gamma_{\mathrm{n},\sigma}(\tau')}$. □

3.5.8.2 Balanced Problems. In this section, we extend the discussion of balance in Section 2.1.1.2 to general factor graphs. Thus, we discuss the expected degree of a variable in the planted model for a given ground truth. We do not explicitly distinguish the binomial model in Corollary 3.137 and the model in Observation 3.138. Recall the degrees $\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)$ from Observation 3.136. For the uniform model, let $\boldsymbol{G}^*(\sigma) = (\boldsymbol{v}^*, \boldsymbol{\psi}^*)$ and $\boldsymbol{d}^*_{m,\sigma}(i) = \{a \in [m] : \exists h \in [k] \boldsymbol{v}^*_{a,h} = i\}$. Let $d^*_{\mathrm{b},\sigma}(i) = \mathbb{E}[\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)]$ and $d^*_{m,\sigma}(i) = \mathbb{E}[\boldsymbol{d}^*_{m,\sigma}(i)]$. Further, let $\Delta(m,\sigma) = (d^*_{m,\sigma}(i) - \bar{d})_i$ and $\Delta_{\mathrm{b}}(\sigma) = (d^*_{\mathrm{b},\sigma}(i) - \bar{d})_i$. Recall Observation 3.9, let $\mu_\gamma = \mu_{\mathrm{T}|\Gamma,\gamma}$ and $\mathcal{P}(\mu_\Psi) = \{\gamma \in \mathcal{P}([q]) : \mu_\gamma|_* = \gamma\}$.

Observation 3.146. *We have* $\mathbb{E}[|\Delta(\boldsymbol{m}^*, \boldsymbol{\sigma}^*)_1|] = o(1)$ *if and only if* $\bar{d} = 0$ *or* $\gamma^* \in \mathcal{P}(\mu_\Psi)$*. In this case we have* $\mathbb{E}[\|\Delta(\boldsymbol{m}^*, \boldsymbol{\sigma}^*)\|_\infty] = \mathcal{O}(\varepsilon_{\mathrm{m}} + \delta_{\mathrm{m}} + 1/\sqrt{n})$*, and* $\|\Delta(\boldsymbol{m}^*, \boldsymbol{\sigma}^*)\|_\infty \le c\delta_{\mathrm{m}} + \varepsilon$ *whp for some* $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ *and all* $\varepsilon = \omega(1/\sqrt{n})$*. The same holds for* $\boldsymbol{d}^*$ *replaced by* $\boldsymbol{d}^*_{\mathrm{b}}$*.*

Proof. Observation 3.130d) applies, so let $s \in [2]$. Next, notice that there exists $c_{\mathfrak{g}} \in \mathbb{R}_{>0}$ such that $d^*_{m,\sigma}(i) = m\mathbb{P}(\exists h \in [k]\, \boldsymbol{v}^*_{r,s}(1,h) = i) \le cm/n$, using the union bound. This justifies the restriction to $m \le m_\uparrow$. For $s = 1$, Observation 3.131e), Observation 3.132a) and the union bound yield

$$\begin{aligned} d^*_{m,\sigma}(i) &= m\mathbb{P}(\exists h \in [k]\, \boldsymbol{v}^*_{\circ,1}(h) = i | \boldsymbol{v}^*_{\circ,1} \in \mathcal{U}_r) = (1 + \mathcal{O}(1/n))\frac{km}{n}\sum_h \frac{1}{k} n\mathbb{P}(\boldsymbol{v}^*_{\circ,1}(h) = i) + \mathcal{O}(1/n) \\ &= (1 + \mathcal{O}(1/n))\frac{km}{n}\frac{\mu_{\gamma_\sigma}|_*(\sigma_i)}{\gamma_\sigma(\sigma_i)} + \mathcal{O}(1/n) = \frac{km}{n}\frac{\mu_{\gamma_\sigma}|_*(\sigma_i)}{\gamma_\sigma(\sigma_i)} + \mathcal{O}(1/n). \end{aligned}$$

Now, we obtain $\mathbb{E}[|\Delta(\boldsymbol{m}^*, \boldsymbol{\sigma}^*)_1|] = 2\bar{d}\|\mu_{\gamma^*}|_* - \gamma^*\|_{\mathrm{tv}} + \mathcal{O}(\varepsilon_{\mathrm{m}} + \delta_{\mathrm{m}} + 1/\sqrt{n})$ using concentration of $\boldsymbol{m}^*$, Observation 3.23 and Observation 3.9. This establishes the first assertion, and the second part follows analogously. For $s = 2$ we use the rejection sampling in the proof of Observation 3.134. The difference of the degrees of i in the two models is bounded by the number of positions where we rejected twice plus the positions where we rejected once and either the first or the second factor is adjacent to i. Thus, the expected difference is $\mathcal{O}(n^3/n^{2k} + n/n^k + n^2/n^{k+1})$. So, we exactly recover the bound above. For the binomial model, Observation 3.131 takes care of the case $s = 1$. The case $s = 2$ was discussed in the proof of Observation 3.136, and we showed that $\mathbb{E}[\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)] = \lambda_\sigma(i) = (1 + \mathcal{O}(1/n))\bar{d}\frac{\mu_{\gamma_\sigma}|_*(\sigma_i)}{\gamma_\sigma(\sigma_i)}$, thus this case is simpler. The case for the constant normalization constant in Observation 3.138 follows with Observation 3.9 and Observation 3.23, we leave the details to the reader. □

Since we focus on balanced problems, not on unbalanced problems, a brief, superficial discussion of weak recovery and distinguishability in this case has to suffice. As we have seen in the proof, the expected degrees $d^*_{m,\sigma}(i)$, $d^*_{\mathrm{b},\sigma}(i)$ given σ are essentially reweighted by $\mu_{\gamma^*}|_*(\sigma_i)/\gamma^*(\sigma_i)$. Since we also roughly know $|\boldsymbol{\sigma}^{*-1}(\tau)|$ for $\tau \in [q]$ due to Observation 3.23, we use the order given by $\tau \mapsto \mu_{\gamma^*}|_*(\tau)/\gamma^*(\tau)$, and the order given by, say, $(\boldsymbol{d}^*_{\mathrm{b},\sigma}(i))_i$ to construct the assignment to the variables, starting with a minimizer τ of $\mu_{\gamma^*}|_*(\tau)/\gamma^*(\tau)$, and the $n\gamma^*(\tau)$ variables with smallest degrees $\boldsymbol{d}^*_{\mathrm{b},\sigma}(i)$. For unbalanced models, this algorithm should efficiently solve weak recovery.

For distinguishability, we look at the empirical degree distributions. This empirical measure converges to $\mathrm{Po}(\bar{d})$ in the null model. For the planted model, the proof of Observation 3.146 suggests that the empirical measure converges to a mixed Poisson with parameter $\lambda_{\boldsymbol{\tau}}$, where $\lambda_\tau = \frac{\mu_{\gamma^*}|_*(\tau)}{\gamma^*(\tau)}$ and $\boldsymbol{\tau} \sim \gamma^*$. Since these do not coincide in the unbalanced case, the degree distributions converge to different limits. This should show why unbalanced models are not contiguous.

3.5.9 The Stochastic Block Model. In this section we discuss a generalized and normalized version of the hierarchical SBMs in Section 2.1.1.2. We start with the definition of the weight function in

Section 3.5.9.1, followed by the model definition in Section 3.5.9.2 and the results in Section 3.5.9.3

3.5.9.1 Weight Definition. In this section we rigorously discuss the SBM, on k-uniform hypergraphs for any $k \in \mathbb{Z}_{\geq 2}$. The definition in Section 2.1.1.2 suffers from redundancies, e.g. since $C_{\ell}^{-1}(c) = C_{\ell-1}^{-1}(c')$ is not only possible, but enforced, or $w(\ell, c)$ is redundant for $c \in [q] \setminus C_\ell([q])$. A clean definition of the class hierarchy on $[q]$ is given by the following family of subsets. For a set $\mathcal{S} \subseteq \mathbb{Z}$ let $\mathfrak{P}(\mathcal{S}) = \{\mathfrak{P} \subseteq 2^{\mathcal{S}} \setminus \{\emptyset\} : \forall \mathcal{P} \in \mathfrak{P}^2 \mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset, \bigcup_{\mathcal{P} \in \mathfrak{P}} \mathcal{P} = \mathcal{S}\}$ be the set of partitions of $\mathcal{S}$. The interior of $\mathfrak{S} \subseteq 2^{\mathcal{S}} \setminus \{\emptyset\}$ is $\mathfrak{I}(\mathfrak{S}) = \{\mathcal{T} \in \mathfrak{S} : \exists \mathcal{T}' \in \mathfrak{S}\, \mathcal{T}' \subsetneq \mathcal{T}\}$, and $\mathfrak{L}(\mathcal{S}) = \mathfrak{S} \setminus \mathfrak{I}(\mathcal{S})$ are the leaves. Let $\mathfrak{R}_0 = \{\{[q]\}\}$, for $n \in \mathbb{Z}_{>0}$ let $\mathfrak{R}_n = \{\mathfrak{S} \cup \mathfrak{P} : \mathfrak{S} \in \mathfrak{R}_{n-1}, \mathcal{S} \in \mathfrak{L}(\mathfrak{S}), \mathfrak{P} \in \mathfrak{P}(\mathcal{S})\}$, and finally let $\mathfrak{R} = \bigcup_n \mathfrak{R}_n = \lim_{n\to\infty} \mathfrak{R}_n$ be the refinement schemes on $[q]$. For a refinement scheme $\mathcal{R} \in \mathfrak{R}$ and $S \in 2^{[q]} \setminus \{\emptyset\}$ let $C_{\mathcal{R}}(S) = \{S' \in \mathcal{R} : S \subseteq S'\}$ be the chain for S in $\mathcal{R}$. Further, for $S \in \mathcal{R}$ let $\mathcal{K}_{\mathcal{R}}(S) = \{S' \in \mathcal{R} : C_{\mathcal{R}}(S') = C_{\mathcal{R}}(S) \cup \{S'\}\}$ be the children of S.

Let $\mathfrak{W}(\mathcal{R}) = \{w \in \mathbb{R}^{\mathcal{R}} : \forall S \in \mathfrak{I}(\mathcal{R})\, \exists K \in \mathcal{K}_{\mathcal{R}}(S)\, w(K) \neq 0\}$ be the set of weights for a refinement scheme, where the normalization property ensures that do not have redundant splits, i.e. the entire partition of a subset is weighted with 0. For $w \in \mathfrak{W}(\mathcal{R})$ let $[w] : 2^{[q]} \setminus \{\emptyset\} \to \mathbb{R}$, $S \mapsto \sum_{S' \in C_{\mathcal{R}}(S)} w(S')$, be the accumulated weights. Let $\mathfrak{W} = \{(\mathcal{R}, w) : \mathcal{R} \in \mathfrak{R}, w \in \mathfrak{W}(\mathcal{R})\}$ be the weighted refinement schemes. Further, we restrict to non-negative and strictly positive accumulated weights, i.e. $\mathfrak{W}_{\geq 0} = \{(\mathcal{R}, w) \in \mathfrak{W} : [w] \geq 0\}$ and $\mathfrak{W}_{>0} = \{(\mathcal{R}, w) \in \mathfrak{W} : [w] > 0\}$.

For $(\mathcal{R}, w) \in \mathfrak{W}_{\geq 0}$ let $\psi_{\mathcal{R},w} : [q]^k \to \mathbb{R}_{\geq 0}$, $\sigma \mapsto [w](\sigma([k]))$, be the accumulated weight of the image of σ. The hierarchical SBMs on k-uniform hypergraphs are $\{\psi_{\mathcal{R},w} : (\mathcal{R}, w) \in \mathfrak{W}_{\geq 0}\}$, and $\{\psi_x : x \in \mathfrak{W}_{>0}\}$ for soft constraints. This recovers the hierarchical SBMs in Section 2.1.1.2 for $k = 2$. In this sense, we also refer to $\mathfrak{W}_{\geq 0}$ as hierarchical SBMs. In the following, we focus on $\mathfrak{W}_{\geq 0}$, the restrictions to soft constraints will be obvious. The SBMs $\mathfrak{W}^+_{\geq 0} = \{(\mathcal{R}, w) \in \mathfrak{W}_{\geq 0} : w \geq 0\}$ are assortative, and the SBMs $\mathfrak{W}^-_{\geq 0} = \{(\mathcal{R}, w) \in \mathfrak{W}_{\geq 0} : w \leq 0\}$ are disassortative.

3.5.9.2 Model Definition. Using the weight ψ from Section 3.5.9.1, we can define the SBM on any of the models from Section 3.5.7. The most widespread model is arguably the binomial model for $r = 3$, $s = 2$ with fixed normalization constant ξ, i.e. from Observation 3.138. As thoroughly discussed in Section 3.5.7, all the models discussed therein are equivalent for our purposes. We prefer to work with the binomial model using the flexible normalization with $\overline{Z}_{\mathrm{f}}(\gamma_\sigma)$ from Corollary 3.137, for its conceptual beauty and, as explained in Observation 3.136, since it hides the ground truth from the expected factor count, for all $r, s \in [3]$, thus it is even harder then the more popular choice.

3.5.9.3 Results. Weak recovery and distinguishability for unbalanced problems have been discussed in Section 3.5.8.2 for the general case. Recall from Remark 3.10 and from Observation 3.146 that a model is balanced if and only if γ is a stationary point of $\overline{Z}_{\mathrm{f}}$, or equivalently $\mu_{\mathrm{T}|\Gamma,\gamma^*}|_* = \gamma^*$ using the notions from Section 3.2.1.2. Although this is very easy to verify, *identifying all* stationary points of $\overline{Z}_{\mathrm{f}}$ is usually not. In this section, we establish all stationary points *explicitly* for the assortative and disassortative SBM. Based on Section 3.5.1, it is sufficient to consider fully supported $\gamma \in \mathcal{P}([q])$. Since we have $S \in C_{\mathcal{R}}(\sigma([k]))$ if and only if $\sigma \in S^k$, exchanging the sums yields

$$\begin{aligned}\overline{Z}_{\mathrm{f}}(\gamma) &= \sum_{\sigma} [w](\sigma([k])) \prod_h \gamma(\sigma_h) = \sum_{S,\sigma} \mathbb{1}\{S \in C(\sigma([k]))\} w(S) \prod_h \gamma(\sigma_h) \\ &= w([q]) + (-1)^t \sum_{S \in \mathcal{R} \setminus \{[q]\}} |w(S)| \gamma(S)^k,\end{aligned}$$

where $t \in \{0, 1\}$ is 1 for the disassortative case. This shows that $\overline{Z}_{\mathrm{f}}$ is (concave) convex if the model is (dis-) assortative. The (maximizers) minimizers are the minimizers of $\sum_{S \in \mathcal{R} \setminus \{[q]\}} |w(S)| \gamma(S)^k$, for both $t \in \{0,1\}$, and can be computed recursively as follows. Let $\mathcal{R}_\circ = \mathcal{R} \cup \binom{[q]}{1}$ be the completion of $\mathcal{R}$, in case some minimal element in $\mathcal{R}$ does not have size 1. Let $w_\circ : \mathcal{R}_\circ \to\to \mathbb{R}$ be given by $w_\circ(S) = w(S)$ for $S \in \mathcal{R}$ and $w_\circ(S) = 0$ otherwise, so in the exceptional case that all children are singletons, they may all have weight 0. Let $\mathcal{R} = \mathcal{R}_\circ$ and $w = w_\circ$ for brevity. We consider the canonical projection $\mathcal{P} \to \mathcal{P}([q])$, where

$$\mathcal{P} = \prod_{S \in \mathfrak{I}(\mathcal{R})} \mathcal{P}(\mathcal{K}(S))$$

are the conditional distributions. In detail, let $A_{\mathcal{R},\sigma} \in \mathcal{R}^{\mathcal{I}(\sigma)}$, $\mathcal{I}_\sigma = \mathbb{Z} \cap [0, |C(\sigma)| - 1]$, be the enumeration of $C(\sigma)$, i.e. $A_\sigma(\mathcal{I}_\sigma) = C(\sigma)$ and $A_\sigma(i) \subseteq A_\sigma(i-1)$ for $i \in \mathcal{I}_\sigma \setminus \{0\}$. Then we consider the projection given by $\gamma(\sigma) = \prod_{i=0}^{|\mathcal{I}_\sigma|-1} \gamma_{A_\sigma(i)}(A_\sigma(i+1))$.

For $S \in \mathcal{R}$ let $\mathcal{R}_{\mathcal{S}} = \mathcal{R}(S) = \{S' \in \mathcal{R} : S' \subseteq S\}$ be the subtree rooted at S, further let $S_{\mathcal{R}(S)} = S$ be its root, and let $\mathcal{K}(\mathcal{R}(S)) = \mathcal{K}_{\mathcal{R}(S)}(S) = \mathcal{K}_{\mathcal{R}}(S)$ be its children. Then, for $\gamma \in \mathcal{P}$ we have $\overline{Z}_{\mathrm{f}}(\gamma) = w([q]) + (-1)^t f_{\mathcal{R}}(\gamma)$, where $f_{\mathcal{R}'}(\gamma') = \sum_{S \in \mathcal{K}(\mathcal{R}')}(|w(S)| + f_{\mathcal{R}'(S)}(\gamma'_{\mathfrak{I}(\mathcal{R}'(S))}))\gamma'_{S_{\mathcal{R}'}}(S)^k$. This recursive description allows to compute the minimizers using the following, recursive algorithm.

- For $S \in \mathcal{K}(\mathcal{R}')$, let $\mathcal{M}_S \subseteq \prod_{S' \in \mathfrak{I}(\mathcal{R}'(S))} \mathcal{P}(\mathcal{K}(S'))$ be the minimizers and $f_S \geq 0$ the minimum.
- Let $\mathcal{N} = \{S' \in \mathcal{K}(\mathcal{R}') : |w(S')| + f_{S'} = 0\}$.
  - For $\mathcal{N} \neq \emptyset$ let $\mathcal{M} = \{\gamma \in \prod_{S \in \mathfrak{I}(\mathcal{R}')} \mathcal{P}(\mathcal{K}(S)) : \gamma_{S_{\mathcal{R}'}} \in \mathcal{P}(\mathcal{N}), \forall S \in \gamma^{-1}_{S_{\mathcal{R}'}}(\mathbb{R}_{>0})\, \gamma_{\mathfrak{I}(\mathcal{R}'(S))} \in \mathcal{M}_S\}$ be the minimizers and $f = 0$ the minimum.
  - For $\mathcal{N} = \emptyset$ let $\mathcal{M} = \{\gamma \in \prod_{S \in \mathfrak{I}(\mathcal{R}')} \mathcal{P}(\mathcal{K}(S)) : \gamma_{S_{\mathcal{R}'}} = \gamma^*, \forall S \in \mathcal{K}(\mathcal{R}')\, \gamma_{\mathfrak{I}(\mathcal{R}'(S))} \in \mathcal{M}_S\}$ be the minimizers and $f = \sum_{S \in \mathcal{K}(\mathcal{R}')}(|w(S)| + f_S)\gamma^*(S)^k$ the minimum, where $\gamma^* \in \mathcal{P}(\mathcal{K}(\mathcal{R}'))$ is given by $\gamma^*(S) = (|w(S)| + f_S)^{-1/(k-1)}/Z$ for $S \in \mathcal{K}(\mathcal{R}')$ and $Z = \sum_S (|w(S)| + f_S)^{-1/(k-1)}$.
- Output the minimizers $\mathcal{M}$ and the minimum f.

This determines all (maximizers) minimizers of the (dis-) assortative SBM, thus *all balanced (dis-) assortative SBMs*. The main results apply to the strictly positive balanced disassortative SBMs, i.e. to $\mathfrak{W}^-_{>0}$ with a maximizer γ^*, since this model is of type 3 in Section 2.1.2.2, where we can choose $w([q]) = 1$ without loss of generality, thus $\boldsymbol{a} = 1$, further $\boldsymbol{b} = 1$ and $\boldsymbol{f}_{h,S}(\sigma) = |w(S)|^{1/k}\mathbb{1}\{\sigma \in S\}$ for $S \in \mathcal{R}$, $h \in [k]$, $\sigma \in [q]$ (and 0 otherwise).

3.5.10 Potts Models. In this section, we discuss the spin glass version the hierarchical SBM.

3.5.10.1 Model Definition. Recall the notions from Section 3.5.9.1 and let $(\mathcal{R}, w) \in \mathfrak{W}$. Following the conventions in the context of spin glasses, we consider the energy $E_\circ(\sigma) = [w](\sigma([k]))$ for $\sigma \in [q]^k$, which induces the weight $\psi(\sigma) = \exp(-E_\circ(\sigma))$. Further, while this model may also be defined on any of the versions in Section 3.5.7, it is typically defined on the binomial model, thus a diluted mean-field model in physics jargon. We also make use of the external fields η from Section 3.5.2. Then, for a fixed graph $v \in \{0,1\}^{\binom{[n]}{k}}$, the Boltzmann distribution μ_v, as defined in Section 2.1.2, is given by the weights $\prod_{i \in [n]} \eta(\sigma_i) \prod_u \psi(\sigma_u)^{v(u)} = \exp(-E_v(\sigma))$, where

$$E_v(\sigma) = \sum_u v(u) E_\circ(\sigma_u) + \sum_{i \in [n]} \eta_\circ(\sigma_i), \quad \eta_\circ(\tau) = -\ln(\eta(\tau)),$$

is the *Hamiltonian.* In this context, we usually take $r = 3$, $s = 2$ in Section 3.5.7 (using permutation invariance of ψ), thus $v(u) \in \{0,1\}$, yielding the model on the binomial hypergraph. The model is assortative/ferromagnetic if $w_S \leq 0$ for $S \neq [q]$, and it is disassortative/antiferromagnetic if $w_S \geq 0$ for $S \neq [q]$. The sign change is due to the convention for the sign in the exponent of ψ, and $w_{[q]}$ can be chosen freely as before, since it only determines the scaling of the weight.

3.5.10.2 Discussion. First, we notice that ψ can be written in the additive form in Section 3.5.9.1, and the definitions of (dis-) assortative models are consistent. Notice that we do not impose any restrictions on the sign of the external field $\eta_\circ$, in particular since we can normalize the model and use $\eta = \gamma^*$ as outlined in Section 3.5.2. Letting $\mathcal{E} = v^{-1}(1)$ denoting the selected edges and expanding the definitions for the Hamiltonian yields

$$E_v(\sigma) = \sum_{u\in\mathcal{E}}\sum_{S\in\mathcal{R}} \mathbb{1}\{\sigma_u \subseteq S\}w(S) + \sum_i \eta_\circ(\sigma_i).$$

Since this form is fairly abstract, notice that for $k = 2$, $w([q]) = 0$, and for a single split, i.e. a partition $\mathcal{P}$ of $[q] = \bigcup_{\mathcal{T}\in\mathcal{P}}\mathcal{T}$ with $\mathcal{R} = \{[q]\}\cup\mathcal{P}$, there exists a unique *type* $\mathcal{T}(\sigma)$ for all $\sigma\in[q]$, i.e. $\sigma\in\mathcal{T}(\sigma)$, and we obtain $E_v(\sigma) = \sum_{\{i,j\}}\mathbb{1}\{\mathcal{T}(\sigma_i) = \mathcal{T}(\sigma_j)\}w(\mathcal{T}(\sigma_i)) + \sum_i \eta_\circ(\sigma_i)$, which is still a generalization of the Potts model, obtained for $\mathcal{T}(\sigma) = \{\sigma\}$, $w(\sigma) = c \in \mathbb{R}_{\geq 0}$ and $\eta_\circ \equiv 0$.

3.5.11 Graphical Channels. We start with a discussion of the channels in [4], then we look at specific channels. In Section 3.5.11.1 we define graphical channels, in Section 3.5.11.2 we discuss the connection to factor graphs, and complete the discussion of general channels with the results in Section 3.5.11.3. We continue with a discussion of LDGM codes in Section 3.5.11.4 and the discussion of the BAC in Section 3.5.11.5 After a discussion of general properties, we focus on LDGM codes. Then we further consider the BAC specifically, and the BISO specifically.

3.5.11.1 Definition. Recall $\mathcal{U}_r \subseteq [q]^k$ and $\mathcal{V}_{r,s} \subseteq \mathcal{U}_r^m$ from Section 3.5.7. A graphical channel is given by $q' \in \mathbb{Z}_{>0}$ and $\nu = (\nu_y) \in \mathcal{P}([q'])^{\mathcal{U}}$. Without loss of generality we assume that there are no redundant outputs, i.e. for all $z \in [q']$ there exists $y \in \mathcal{U}$ such that $\nu_y(z) > 0$.

Fix the message length $n \in \mathbb{Z}_{>0}$ and the block length $m \in \mathbb{Z}_{\geq 0}$. For given $v \in \mathcal{V}$, we use the block code[6] $y_v : [q]^n \to \mathcal{U}_r^m$, $x \mapsto (x_{v_a})_{a\in[m]}$. For given $x \in [q]^n$, let $\boldsymbol{z}(v, x) \sim \bigotimes_{a\in[m]} \nu_{y_{v,x}(a)}$ be the output for the transmitted codeword $y_{v,x}$. We choose the code uniformly at random, independently from the discrete memoryless source given by $\gamma^* \in \mathcal{P}([q])$, i.e. we consider $(\boldsymbol{v}, \boldsymbol{x}) \sim \mathrm{u}(\mathcal{V}) \otimes \gamma^{*\otimes n}$. Recall the conditional mutual information, e.g. from Section 3.3.1.1. In this section we focus on the asymptotics of $I(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x})|\boldsymbol{v})$.

3.5.11.2 Induced Factor Graphs. First, we discuss how the model in Section 3.5.11.1 is mapped onto a factor graph model. For fully supported $p^* \in \mathcal{P}([q'])$ let $\boldsymbol{z}^* \sim p^*$, for $z \in [q']$ let $\psi_z : [q]^k \to \mathbb{R}_{\geq 0}$, $\sigma \mapsto \nu_\sigma(z)/p^*(z)$, and finally let $\boldsymbol{\psi}_\circ = \psi_{\boldsymbol{z}^*}$.

Observation 3.147. *We have* $\mathbb{E}[\boldsymbol{\psi}_\circ] \equiv 1$, $\overline{Z}_{\mathrm{f}} \equiv 1$, $\xi = 1$, *and* $I(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x})|\boldsymbol{v}) = I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*))$.

Proof. Unfortunately, we have to deal with a technical obstacle. Let $\Psi = \{\psi_z : z \in [q']\}$, and for $\psi \in \Psi$ let $\mathcal{Z}(\psi) = \{z \in [q'] : \psi_z = \psi\}$. Notice that $\mu_\Psi(\psi) = \sum_{z\in\mathcal{Z}(\psi)} p^*(z)$. Thus, we can only distinguish the equivalence classes $\mathcal{Z}(\psi)$. So, we introduce the auxiliary channel $\nu^\circ_\sigma(\psi) = \nu_\sigma(\mathcal{Z}(\psi)) = \sum_{z\in\mathcal{Z}(\psi)} \nu_\sigma(z)$ and notice that $\mu_\Psi(\psi)\psi(\sigma) = \nu^\circ_\sigma(\psi)$. This directly yields $\mathbb{E}[\boldsymbol{\psi}_\circ(\sigma)] = \sum_\psi \nu^\circ_\sigma(\psi) = 1$ and $\overline{Z}_{\mathrm{f}} \equiv 1$. Using $\boldsymbol{z}^\circ(v, x)$ to denote the output for ν°, we even have

$$\begin{aligned}\mathbb{P}(\boldsymbol{\sigma}^* = \sigma, \boldsymbol{G}^*(\sigma) = (v, \psi)) &= \frac{\gamma^{*\otimes n}(\sigma)}{(n^k)^m}\prod_a (\mu_\Psi(\psi_a)\psi_a(\sigma_{v_a})) = \frac{\gamma^{*\otimes n}(\sigma)}{(n^k)^m}\prod_a \nu^\circ_\sigma(\psi_a)\\ &= \mathbb{P}\left(\boldsymbol{x} = \sigma, \boldsymbol{v} = v, \boldsymbol{z}^\circ(\boldsymbol{v}, \boldsymbol{x}) = \psi\right).\end{aligned}$$

[6]This code is *not* injective, but it does serve the purpose of a code, as we will see shortly.

This yields $I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)) = I(\boldsymbol{x}, (\boldsymbol{v}, \boldsymbol{z}^\circ(\boldsymbol{v}, \boldsymbol{x})))$ and builds the bridge from channel to factor graph.

In the next step, we resolve the technical obstacle and reduce the mutual information to the conditional mutual information. First, we notice that $(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{z}^\circ(\boldsymbol{v}, \boldsymbol{x})) \sim (\boldsymbol{x}, \boldsymbol{v}, \psi_{\boldsymbol{z}(\boldsymbol{v},\boldsymbol{x})})$, and that

$$\mathbb{P}(\boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x}) = z | \boldsymbol{v} = v, \boldsymbol{x} = x, \psi_{\boldsymbol{z}(\boldsymbol{v},\boldsymbol{x})} = \psi) = \frac{\prod_a \nu_{x_{v_a}}(z_a)}{\prod_a \nu_{x_{v_a}}(\mathcal{Z}(\psi_a))},$$

for v, x, ψ, z such that the probability is well-defined and positive, thus in particular $z_a \in \mathcal{Z}(\psi_a)$. But now, choose a fixed representant $z(\psi) \in \mathcal{Z}(\psi)$ per $\psi \in \Psi$ and notice that

$$\nu_\sigma(z) = p^*(z)\psi_z(\sigma) = p^*(z)\psi_{z(\psi)}(\sigma) = \frac{p^*(z)}{p^*(z(\psi))}\nu_\sigma(z(\psi)), \quad z \in \mathcal{Z}(\psi).$$

This suggests that $\nu_\sigma(\mathcal{Z}(\psi)) = \frac{\mu_\Psi(\psi)}{p^*(z(\psi))}\nu_\sigma(z(\psi))$, hence the conditional probability simplifies to

$$\mathbb{P}(\boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x}) = z | \boldsymbol{v} = v, \boldsymbol{x} = x, \psi_{\boldsymbol{z}(\boldsymbol{v},\boldsymbol{x})} = \psi) = \prod_a \frac{p^*(z_a)}{\mu_\Psi(\psi_a)},$$

which does not depend on v and x, yielding $I(\boldsymbol{x}, (\boldsymbol{v}, \boldsymbol{z}^\circ(\boldsymbol{v}, \boldsymbol{x}))) = I(\boldsymbol{x}, (\boldsymbol{v}, \boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x})))$. Finally, we have $I(\boldsymbol{x}, (\boldsymbol{v}, \boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x}))) = I(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x}) | \boldsymbol{v})$ since $\boldsymbol{v} \sim \mathrm{u}(\mathcal{V})$ is independent of $\boldsymbol{x}$. □

Observation 3.147 explains why we can consider the planted model and further shows that **BAL** holds for all $\gamma^* \in \mathcal{P}([q])$, which is the precise reason why we can compute the mutual information limit for all discrete memoryless sources – well, which brings us to **POS**. Unfortunately, the assumption is hard to check, but on the upside, we have the following result.

Observation 3.148. *Fix $\gamma^* \in \mathcal{P}([q])$. If* **POS** *holds for some fully supported $p^* \in \mathcal{P}([q])$, then* **POS** *holds for all fully supported $p^* \in \mathcal{P}([q])$.*

Proof. Fix a fully supported measure $p^* \in \mathcal{P}([q'])$ with $|\Psi^*| = q'$, where $\Psi^* = \{\psi^*_z : z \in [q']\}$ and $\psi^*_z = \psi_{p^*,z}$. Let $p \in \mathcal{P}([q'])$ be fully supported, $\psi_z = \psi_{p,z}$ and $\Psi = \{\psi_z : z \in [q']\}$. Let $\boldsymbol{z}^* \sim p^*$, $\boldsymbol{z} \sim p$, $\boldsymbol{\psi}_* = \psi^*_{\boldsymbol{z}^*}$ and $\boldsymbol{\psi}_\circ = \psi_{\boldsymbol{z}}$. Finally, let $\boldsymbol{\psi}_* \sim \mu^*_\Psi$, $\boldsymbol{\psi}_\circ \sim \mu_\Psi$ and assume that **POS** holds for μ_Ψ.

Using that $z \mapsto \psi^*_z$ is a bijection, let $\mu'_\Psi(\psi^*_z) = p(z)$ and $\boldsymbol{\psi}' \sim \mu'_\Psi$. Further, let $a(\psi^*_z) = p^*(z)/p(z)$ and notice that a is the (μ^*_Ψ, μ'_Ψ)-derivative. On the other hand, we have $\psi_z = a(\psi^*_z)\psi^*_z$, so we have $\boldsymbol{\psi} \sim a(\boldsymbol{\psi}')\boldsymbol{\psi}'$. Observation 3.118 yields $\nabla(\mu_\Psi, \pi) = \nabla(\mu^*_\Psi, \pi)$ since $\mathbb{E}[a(\boldsymbol{\psi}')] = 1$. so **POS** holds for p^* if and only if it holds for p, which completes the proof using transitivity. □

3.5.11.3 Results. Since we only built the bridge to the conditional mutual information[7], in Observation 3.147, we focus on Theorem 2.5. Using ϕ_a from Theorem 2.3 and that a phase transition exists for $B_\uparrow(d) = \phi_\mathrm{a}(d)$, a conceptually reasonable way to rewrite the mutual information limit is

$$\lim_{n\to} \frac{1}{n} I(\boldsymbol{\sigma}^*, \boldsymbol{G}^*(\boldsymbol{\sigma}^*)) = \frac{d}{k\xi}\mathbb{E}[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}))] - \phi_\mathrm{a}(d) - (B_\uparrow(d) - \phi_\mathrm{a}(d)) = \frac{d}{k}\left(\mathbb{E}[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}))] - \frac{k}{d}B_\uparrow(d)\right),$$

for $d > 0$, where we pulled out d/k since the normalization with m is reasonable for codes, as discussed in Section 2.1.1.3, used that $\phi_\mathrm{a}(d) = \frac{d}{k}\Lambda(\xi)$, and that $\xi = 1$ from Observation 3.147. For transparency,

[7] Recall from the proof of Observation 3.147 that the joint laws coincide, so this bridge is very solid.

let $\boldsymbol{m}^* = m$ be the sequence from Section 2.1.1.3, then for any channel $\nu > 0$ and $\gamma^* \in \mathcal{P}([q])$ satisfying **POS** (cf. Observation 3.148) we have $\lim_{n\to\infty} \frac{1}{m} I(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x})|\boldsymbol{v}) = \iota(d)$,

$$\iota(d) = \mathbb{E}[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}))] - \frac{k}{d}(B_\uparrow(d) - \phi_{\mathrm{a}}(d)) = \mathbb{E}[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}))] - \frac{k}{d} B_\uparrow(d).$$

Of course, we still have to resolve the auxiliary construction $(\boldsymbol{\sigma}, \boldsymbol{\psi}) \sim \gamma^{*\otimes k} \otimes \mu_\Psi$. Notice that

$$\mathbb{E}[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}))] = \sum_{\sigma, z} \gamma^{*\otimes k}(\sigma) p^*(z) \frac{\nu_\sigma(z)}{p^*(z)} \ln\left(\frac{\nu_\sigma(z)}{p^*(z)}\right) = D_{\mathrm{KL}}(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) \| \boldsymbol{z}^* | \boldsymbol{y}_\circ),$$

which is confusing, to say the least, because p^* from the *auxiliary* construction is still present in the limit. Formally, this is confusing because $I(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x})|\boldsymbol{v})$ does no depend on p^*, so neither does the limit. The only possible explanation is that $B_\uparrow(d)$ depends on p^*.

Before we continue this thread, we rewrite the relative entropy as

$$\begin{aligned}\mathbb{E}[\Lambda(\boldsymbol{\psi}(\boldsymbol{\sigma}))] &= H(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) \| \boldsymbol{z}^*) - H(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) | \boldsymbol{y}_\circ) = H(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ)) + D_{\mathrm{KL}}(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) \| \boldsymbol{z}^*) - H(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) | \boldsymbol{y}_\circ) \\ &= I(\boldsymbol{y}_\circ, \boldsymbol{z}_\circ(\boldsymbol{y}_\circ)) + D_{\mathrm{KL}}(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) \| \boldsymbol{z}^*),\end{aligned}$$

using linearity of the cross entropy. So, if we choose p^* to be $p^* = p^\circ$, where $\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) \sim p^\circ$, which is fully supported by assumption, then we have $\iota(d) = I(\boldsymbol{y}_\circ, \boldsymbol{z}_\circ(\boldsymbol{y}_\circ)) - \frac{k}{d} B_\uparrow(d)$. Continuing the thread above, we *know* that $\iota(d)$ does not depend on p^*, meaning that $B_\circ(d) = \frac{k}{d} B_\uparrow(d) - D_{\mathrm{KL}}(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) \| \boldsymbol{z}^*)$ does not depend on p^*, which means that $B_\uparrow(d) = B_\uparrow(d, p^*) = \frac{d}{k} B_\circ(d) + \frac{d}{k} D_{\mathrm{KL}}(\boldsymbol{z}_\circ(\boldsymbol{y}_\circ) \| \boldsymbol{z}^*)$ is convex in p^* with unique minimizer p°. For $0 < d \le d_{\mathrm{cond}}$ we have $B_\uparrow(d) = \phi_{\mathrm{a}}(d) = 0$, since $\xi = 1$, and on the other hand $B_\uparrow(d) > 0$ for $p^* \neq p^\circ$, so $d_{\mathrm{cond}} = 0$ for $p^* \neq p^\circ$. Let $p^* = p^\circ$ in the remainder.

Now, we have $\lim_{n\to\infty} \frac{1}{m} I(\boldsymbol{x}, \boldsymbol{z}(\boldsymbol{v}, \boldsymbol{x})|\boldsymbol{v}) = I(\boldsymbol{y}_\circ, \boldsymbol{z}_\circ(\boldsymbol{y}_\circ)) - B_\circ(d)$ with $B_\circ(d) = 0$ for $d \le d_{\mathrm{cond}}$ by the above, for (ν, γ^*) satisfying **POS** and $\nu > 0$. The translation of degree into rate is given by $R = k/d$, in particular $R^*(\gamma^*) = k/d_{\mathrm{cond}}(\gamma^*)$. Of course, if **POS** can be verified for γ^* such that $\gamma^{*\otimes k}$ is a capacity achieving distribution for ν, then we can replace the mutual information by the capacity in the above.

3.5.11.4 LDGM Codes. In this section we focus on LDGM codes. First, we discuss the connection to graphical channels, and thus factor graphs. Then, we discuss the idealized mapping of the discrete memoryless source onto the input distribution to the noisy channel.

The definition of LDGM codes in Section 2.1.1.3 requires $r = 3$ and $s = 1$ in Section 3.5.7. Let $\iota : \{-1, 1\} \to \{0, 1\}$, $x \mapsto \mathbb{1}\{x = -1\}$, then for $y \in \{-1, 1\}^k$ we have $\iota(\prod_h y_h) = \bigoplus_h \iota(y_h)$. In the following, we work with the product representation, i.e. with $q = 2$ over $\{-1, 1\}$. For a given channel $\nu^\circ \in \mathcal{P}([q'])^{\{-1,1\}}$ let $\nu \in \mathcal{P}([q'])^{\{-1,1\}^k}$ be given by $\nu_y(z) = \nu^\circ_{b(y)}(z)$, where $b(y) = \prod_h y_h$. So, LDGM encoded communication through a noisy channel ν° can be modeled as a graphical channel, and thus as a planted model.

As explained in Section 2.1.1.3, the code used for the graphical channels in Section 3.5.11.1 preserves[8] the input distribution. So, for a success probability $x \in [0, 1]$ let $\gamma_x \in \mathcal{P}(\{-1, 1\})$ be given by $\gamma_x(1) = x$, and further let $\boldsymbol{y}_x \sim \gamma_x^{\otimes k}$ be our idealized input to the channel ν. For the analysis of the second step, the XOR, let $s : [0, 1] \mapsto [0, 1]$, $x \mapsto \mathbb{P}(b(\boldsymbol{y}_x) = 1)$.

Observation 3.149. *We have $s(x) = \frac{1}{2}(1 + (2x - 1)^k)$. The map s is an increasing bijection for $k \notin 2\mathbb{Z}$. For $k \in 2\mathbb{Z}$ we have $s(0) = 1$, $s(x) = s(1 - x)$ and the unique minimum $\frac{1}{2}$ at $x = \frac{1}{2}$.*

[8]This is obviously not true, but in the replica symmetric regime it is sufficiently true (cf. Section 3.5.11.3)

Proof. We only show the first part, the remainder is obvious. But the binomial theorem yields

$$(x-(1-x))^k=\sum_{\ell=0}^{k}\binom{k}{\ell}[-(1-x)]^\ell x^{k-\ell}=s(x)-(1-s(x)).$$

□

Thus, for $k\in 2\mathbb{Z}$ the input distribution at ν° is necessarily biased towards 1 (meaning $0\in\{0,1\}$). Moreover, for any non-uniform input at ν° biased towards 1, there exist exactly *two* choices for the discrete memoryless source.

3.5.11.5 The Binary Asymmetric Channel. We further restrict to $q'=2$, and thus arrive at the LDGM-BAC pair from Section 2.1.1.3. So, let $\nu^\circ_1(-1)=\delta\in[0,\varepsilon]$ and $\nu^\circ_{-1}(1)=\varepsilon\in[0,1/2]$. Let $c^\circ(\delta,\varepsilon)=\max_{p\in\mathcal{P}(\{-1,1\})}I(\boldsymbol{x}_p,\boldsymbol{y}(\boldsymbol{x}_p))$ be the channel capacity, where $\boldsymbol{x}_p\sim p$ and $\boldsymbol{y}(x)\sim\nu^\circ_x$. Notice that $c^\circ(1/2,1/2)=0$ is trivial, so let $(\delta,\varepsilon)\neq(1/2,1/2)$ in the remainder.

Observation 3.150. *Notice that the following holds.*
a) The unique distributions $p_{\mathrm{c}},q_{\mathrm{c}}\in\mathcal{P}(\{-1,1\})$ *with* $\boldsymbol{y}(\boldsymbol{x}_{p_{\mathrm{x}}})\sim q_{\mathrm{c}}$ *and* $c^\circ(\delta,\varepsilon)=I(p_{\mathrm{c}},q_{\mathrm{c}})$ *are*

$$p_{\mathrm{c}}(1)=\frac{q_{\mathrm{c}}(1)-\varepsilon}{1-\varepsilon-\delta},\ q_{\mathrm{c}}(1)=\frac{a}{1+a},\ a=\exp\left(\frac{H(\varepsilon)-H(\delta)}{1-\varepsilon-\delta}\right).$$

b) We have $p_{\mathrm{c}}(1)\geq 1/2$.

Proof. For $p\in[0,1]$ let $p_{\mathrm{c}},q_{\mathrm{c}}\in\mathcal{P}(\{-1,1\})$ be given by $\boldsymbol{x}_p\sim p_{\mathrm{c}}$ and $\boldsymbol{y}(\boldsymbol{x}_p)\sim q_{\mathrm{c}}$, i.e. $p_{\mathrm{c}}(1)=p$ and $q_{\mathrm{c}}=\mathbb{E}[\nu^\circ_{\boldsymbol{x}_p}]$. Let $\iota:[0,1]\to\mathbb{R}_{\geq 0}$, $p\mapsto I(\boldsymbol{x}_p,\boldsymbol{y}(\boldsymbol{x}_p))$ and notice that

$$\begin{aligned}\iota(p)&=\sum_{x,y\in\{-1,1\}}p_{\mathrm{c}}(x)\nu^\circ_x(y)\ln\left(\frac{p_{\mathrm{c}}(x)\nu^\circ_x(y)}{p_{\mathrm{c}}(x)q_{\mathrm{c}}(y)}\right)=\sum_{x,y\in\{-1,1\}}p_{\mathrm{c}}(x)\nu^\circ_x(y)\ln\left(\frac{\nu^\circ_x(y)}{q_{\mathrm{c}}(y)}\right)\\&=H(q_{\mathrm{c}})-\sum_{x\in\{-1,1\}}p_{\mathrm{c}}(x)H(\nu^\circ_x).\end{aligned}$$

Using the chain rule with $\Lambda'(t)=\ln(t)+1$, $q_{\mathrm{c}}(y)'=\nu^\circ_1(y)-\nu^\circ_{-1}(y)$ and linearity yields

$$\begin{aligned}\iota'(p)&=-\sum_y\ln(q_{\mathrm{c}}(y))(\nu^\circ_1(y)-\nu^\circ_{-1}(y))-(H(\nu^\circ_1)-H(\nu^\circ_{-1}))\\&=H(\nu^\circ_1\|q_{\mathrm{c}})-H(\nu^\circ_{-1}\|q_{\mathrm{c}})-(H(\nu^\circ_1)-H(\nu^\circ_{-1}))=D_{\mathrm{KL}}(\nu^\circ_1\|q_{\mathrm{c}})-D_{\mathrm{KL}}(\nu^\circ_{-1}\|q_{\mathrm{c}}).\end{aligned}$$

The former relative entropy is strictly decreasing, while the latter is strictly increasing in p, so ι' is strictly decreasing with $\iota'(0)=D_{\mathrm{KL}}(\nu^\circ_1\|\nu^\circ_{-1})$ and $\iota'(1)=-D_{\mathrm{KL}}(\nu^\circ_{-1}\|\nu^\circ_1)$. Hence, the unique maximizer $p\in(0,1)$ satisfies $\iota'(p)=0$ and thus

$$q_{\mathrm{c}}(1)^{\nu^\circ_1(1)-\nu^\circ_{-1}(1)}q_{\mathrm{c}}(-1)^{\nu^\circ_1(-1)-\nu^\circ_{-1}(-1)}=e^{H(\nu^\circ_1)-H(\nu^\circ_{-1})}=a^{1-\varepsilon-\delta}.$$

We further obtain $q_{\mathrm{c}}(1)\in(0,1)$ since $\nu_1\neq\nu_{-1}$, so using $\nu^\circ_1(1)-\nu^\circ_{-1}(1)=1-\varepsilon-\delta$ we have $q_{\mathrm{c}}(1)/q_{\mathrm{c}}(-1)=a$ and thus $q_{\mathrm{c}}(1)=\frac{a}{1+a}$. Solving $(1-\delta)p+\varepsilon(1-p)=q_{\mathrm{c}}(1)$ for p completes the proof of the first part.

For the second part we notice that $p_{\mathrm{c}}(1)\geq 1/2$ is equivalent to $q_{\mathrm{c}}(0)\leq m$, where $m=\frac{1}{2}(\ell+u)$ with $u=1-\varepsilon$ and $\ell=\delta$. This is equivalent to $a\geq\frac{1-m}{m}$, and by taking the logarithm on both sides we notice that this is equivalent to $H(u) - H(\ell) \geq (u-\ell)H'(m)$. Now, we have

$$H'(u) - H'(\ell) = (u-\ell)H'(m) - \int_\ell^u H'(m) - H'(p)\mathrm{d}p = (u-\ell)H'(m) - \int_\ell^u \int_p^m H''(q)\mathrm{d}q\mathrm{d}p$$
$$= (u-\ell)H'(m) - \int_\ell^m \int_p^m H''(q)\mathrm{d}q\mathrm{d}p + \int_m^u \int_m^p H''(q)\mathrm{d}q\mathrm{d}p.$$

Notice that for $q \in (\ell, m)$ and $r = 2m - q$ we have $\frac{1}{2} - q = r - \frac{1}{2} + 1 - 2m$, which shows that $|\frac{1}{2} - q| > |\frac{1}{2} - r|$. So, the symmetry of $H''(q) = \frac{-1}{q(1-q)}$ gives

$$H'(u) - H'(\ell) \geq (u-\ell)H'(m) - \int_\ell^m \int_p^m H''(2m-q)\mathrm{d}q\mathrm{d}p + \int_m^u \int_m^p H''(q)\mathrm{d}q\mathrm{d}p = (u-\ell)H'(m).$$

□

Observation 3.150 shows that we can hope to reach capacity for $k \in 2\mathbb{Z}$, as discussed in the last section. It also shows that there are two capacity achieving distributions $\gamma^* \in \mathcal{P}([q])$ for $\delta \neq \varepsilon$, and both are fully supported. Also, recall from Observation 3.147 that all $\gamma^* \in \mathcal{P}([q])$ satisfy **BAL**.

For **POS**, we use Observation 3.148 and uniformly distributed $\psi_z(y) = 2\nu_y(z) = 2\nu^\circ_{b(y)}(z)$, i.e.

$$\psi_z(y) = (\nu^\circ_1(z) + \nu^\circ_{-1}(z)) + (\nu^\circ_1(z) - \nu^\circ_{-1}(z))\prod_h y_h,\ z \in \{-1, 1\}.$$

With $a_z = \nu^\circ_1(z) + \nu^\circ_{-1}(z)$, $b_z = -(\nu^\circ_1(z) - \nu^\circ_{-1}(z))/a_z$ we have $\psi_z(y) = a_z(1 - b_z \prod_h y_h)$, where

$$a_1 = 1 - \delta + \varepsilon,\ b_1 = -\frac{1-\delta-\varepsilon}{1-\delta+\varepsilon},\ a_{-1} = 1 - \varepsilon + \delta,\ b_{-1} = \frac{1-\delta-\varepsilon}{1-\varepsilon+\delta}.$$

For $\varepsilon = \delta$ we recover the binary symmetric channel, which has already been thoroughly discussed. Verifying that this weight is of Type 1 in Section 2.1.2.2, amounts to the evaluation of

$$\mathbb{E}[a_{\boldsymbol{z}} b_{\boldsymbol{z}}^\ell] = \frac{1}{2}(a_1 b_1^\ell + a_{-1} b_{-1}^\ell) = \frac{1}{2}(1-\delta-\varepsilon)^\ell (a_{-1}^{-(\ell-1)} - a_1^{-(\ell-1)}), \quad \boldsymbol{z} \sim \mathrm{u}(\{-1,1\}),\ \ell \in 2\mathbb{Z}_{>0} + 1.$$

This shows that $\mathbb{E}[a_{\boldsymbol{z}} b_{\boldsymbol{z}}^\ell] > 0$ since $\delta < \varepsilon$ and thereby $a_{-1} < 1 < a_1$, so **BAL** and **POS** holds for all choices of $0 \leq \delta < \varepsilon \leq \frac{1}{2}$, $\gamma^* \in \mathcal{P}(\{-1,1\})$, in particular the two capacity achieving distributions.

3.5.11.6 Convex Combinations of Channels. We consider the BISO channels in [4], defined as follows. Let $q_2 \in \mathbb{Z}_{\geq 0}$, $q_1 \in \mathbb{Z}_{\geq 0}$ be such that $q' = q_2 + q_1 > 0$, and let the output alphabet $\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_2$ be given by $\mathcal{Z}_2 = [q_2] \times \{-1, 1\}$ and $\mathcal{Z}_1 = [q_1] \times \{1\}$. Let $\nu^\circ = (\nu^\circ_1, \nu^\circ_{-1})$ be such that for each $i \in [q_2]$ we have $\nu^\circ_1((i,-1)) = \nu^\circ_{-1}((i,1))$ and $\nu^\circ_1((i,1)) = \nu^\circ_{-1}((i,-1))$, while for $i \in [q_1]$ we have $\nu^\circ_1((i,1)) = \nu^\circ_{-1}((i,1))$.

For $i \in [q_2]$ let $\alpha(i) = \nu^\circ_1((i,1)) + \nu^\circ_1((i,-1)) = \nu^\circ_{-1}((i,1)) + \nu^\circ_{-1}((i,-1))$. Notice that we have $\alpha(i) > 0$, otherwise (i,s) would both be isolated. Let $\nu^\circ_{i,s}((i,s)) = \nu^\circ_s((i,s))/\alpha(i)$, then ν° is a binary symmetric channel (modulo the output symbols). Also for $i \in [q_1]$ we have $\alpha(i) = \nu^\circ_1((i,1)) = \nu^\circ_{-1}((i,1)) > 0$, and we let $\nu^\circ_{i,s}((i,1)) = 1$. This completes the channel decomposition.

Let $\nu = (\nu^\circ_{b(y)})_y$ and $\nu_{i,y} = \nu^\circ_{i,b(y)}$ be the associated channels with LDGM frontend. Let $\psi_{(i,s)}(y) = 2\nu_{i,y}((i,s))$ for $(i,s) \in \mathcal{Z}_2$, $\psi_{(i,s)}(y) = \nu_{i,y}((i,s))$ for $(i,s) \in \mathcal{Z}_1$, $p^* \in \mathcal{P}(\mathcal{Z})$ given by $p^*(i,s) = \frac{1}{2}\alpha(i)$ for $(i,s) \in \mathcal{Z}_2$ and $p^*(i,s) = \alpha(i)$ for $(i,s) \in \mathcal{Z}_1$, and let $\boldsymbol{\psi} \sim \psi_{\boldsymbol{z}^*}$ for $\boldsymbol{z}^* \sim p^*$ be the associated weight. Let $\boldsymbol{b} \sim \mathrm{u}(\{-1,1\})$, further $\boldsymbol{\psi}_i \sim \psi_{i,\boldsymbol{b}}$ for $i \in [q_2]$ and $\boldsymbol{\psi}_i = \psi_{(i,1)}$ for $i \in [q_1]$. Notice that the former are (scaled) binary symmetric channels and the latter deterministic constant functions (from the product class), hence all p_i, given by $\boldsymbol{\psi}_i \sim p_i$, satisfy $\{p_i\} \times \mathcal{P}([q]) \subseteq \mathfrak{A}$. On the other hand, the law p given by $\boldsymbol{\psi} \sim p$ is just the convex combination $p = \sum_i \alpha(i) p_i$ and hence $\{p\} \times \mathcal{P}([q]) \subseteq \mathfrak{A}$, for all $k \in \mathbb{Z}_{>0}$.

Using the probability space over the weight functions from Observation 3.120, this result directly extends to the binary memoryless symmetric channels from Definition 1 in Appendix A of [3]. However, notice that Observation 3.147 was only proven for finite output alphabets.

3.5.12 Open Problems. The extensions of [30, 32, 31] to other ground truths are highly desirable, the first to e.g. establish weak recovery and distinguishability threshold results for the general disassortative SBM, the second for the extension to CSPs and the last for the analysis of SBMs over given degree sequences, and the efficiency analysis of LDGM codes that do not suffer from isolated input bits. We further believe that the results in this work can be extended to arbitrary balanced models subject to **POS**, i.e. ground truths that are stationary but not necessarily maximizers. This should be achievable using a (possibly soft) truncation argument to keep the Nishimori model close to the planted model. Since the Aizenman-Sims-Starr scheme does not rely on **POS**, this should be sufficient to recover the bounds in [43].

The assumption **POS** is still poorly understood. To our knowledge, there is no proof of a violation of **POS**, however, there are claims in [30, 31].Combined with the above, verifying **POS** for a single binary assortative SBM would close the gap in [3]. Moreover, next to product weights, all models known to satisfy **POS** are essentially sums of products of conditionally independent factors. An extension of this class to other models is highly desirable, e.g. for cryptography [19].

Finally, for graphical channels the replica symmetric regime trivializes except for one specific weight distribution. While we have extended this reweighting of weights without changing the mutual information, the analysis and a deeper understanding of the impact on the Bethe functional and in particular the condensation threshold remain open.

4 Triangle Factors in the Graph Process

This section is dedicated to the proof of Theorem 2.8. We continue to use the convention to suppress dependencies for the sake of brevity and hence adjust the notation, e.g. to distinguish the binomial graph and the graph process. In Section 4.1 we look at the relevant contributions, formally introduce the notions using the adjusted notation, present a stronger result which yields Theorem 2.8 as a corollary, and discuss the proof steps on a high level. In Section 4.2, we establish a few basics that are required for the main steps of the proof. Then, we derive a stronger version of the main result in [63] in Section 4.3. The core coupling of the hypergraph process and the graph process is executed in Section 4.4. Finally, we conclude the proof in Section 4.5, where we couple two copies of the hypergraph process to deal with some exceptionally persistent hyperedges, and then harvest the insights of the preceding parts to derive the main results.

4.1 Introduction

The research on hyperedge covers, perfect matchings, k-clique covers and k-clique factors is reviewed in Section 4.1.1. The new notation for random hypergraphs is introduced in Section 4.1.2, while triangle hypergraphs and the hitting times are defined in Section 4.1.3. Then we explicitly state the relevant result from [71] in Section 4.1.4, followed by the main result of this part in Section 4.1.5.

Then, we briefly discuss open problems in Section 4.1.6, before providing a high-level overview of the proof in Section 4.1.7.

4.1.1 Related Work. Here and in the following, it is convenient to identify k-uniform hypergraphs H with their indicator function $H \in \{0,1\}^{\binom{[n]}{k}}$, keeping the vertex set $[n]$ implicit. Recall that a hyperedge cover $\mathcal{C} \subseteq H^{-1}(1)$ is a set of hyperedges that covers the vertices, i.e. $\bigcup_{E\in\mathcal{C}} E = [n]$, and that $\mathcal{C}$ is a perfect matching, or 1-factor, if the hyperedges are pairwise disjoint. For a graph $G \in \{0,1\}^{\binom{[n]}{2}}$, recall that a k-clique $E \in \binom{[n]}{k}$ is a vertex set such that the complete graph on E is contained in G, i.e. $\binom{E}{2} \subseteq G^{-1}(1)$, that a k-clique cover $\mathcal{C} \subseteq \binom{[n]}{k}$ is a set of k-cliques $E \in \mathcal{C}$ that covers the vertices, i.e. $\bigcup_{E\in\mathcal{C}} E = [n]$, and that $\mathcal{C}$ is a k-clique factor if the k-cliques are disjoint [68].

Since 2-cliques are edges, the notions of edge cover and 2-clique cover, as well as perfect matching and 2-clique factor coincide for $k = 2$. The study of this case in general graphs dates back at least to 1891 [108], and also in the context of random graphs the locations of the sharp thresholds for the existence of edge covers (connectivity, component sizes) and perfect matchings have already been established by Erdős and Rényi in the 1960's [46, 47, 48, 49]. Finally, the hitting time result Theorem 2.8 for $k = 2$ was established two decades later by Bollobás and Thomason in 1985 [23].

The extension towards perfect matchings (and hyperedge covers) in k-uniform hypergraphs was initiated by Schmidt and Shamir in 1979 [50], who derived the first bounds in 1983 [120]. These results were subsequently improved in [55, 74, 69], until not only the threshold, but also the stronger hitting time result was established by Jeff Kahn [70, 71]. The first time that this threshold was conjectured was seemingly in [36], where the perfect matching threshold for regular uniform hypergraphs was established.

As stated in [68, 69], the extension towards k-clique factors (and covers) was seemingly initiated by Ruciński. First results were obtained in [121, 117, 11], yielding the cover threshold [68], which were further improved in [76, 74, 69], until the location of the k-clique factor threshold was established by Oliver Riordan [113] and Annika Heckel [63] using ingenious couplings.

The missing hitting time result for k-clique covers and k-clique factors, a problem stated by Erdős and Spencer for $k = 3$ [28], was recently established in [64]. In the following, we will slightly strengthen the result in [63] and present the result in [64] for the special case $k = 3$, i.e. for triangle covers and triangle factors. In particular, we will obtain the cover threshold as a byproduct along the lines.

4.1.2 Random Graphs. We keep the number $n \in \mathbb{Z}_{\geq 12}$ of vertices arbitrary but fixed throughout. For given $\pi \in [0,1]$ let $\boldsymbol{H}_{\mathrm{b},n,\pi} \sim \mathrm{Bin}(1,\pi)^{\otimes\binom{[n]}{3}}$ be the binomial hypergraph, further $\boldsymbol{E}_{\mathrm{p},n} \sim \mathrm{u}(\binom{[n]}{3}!)$ the hyperedge process with associated hypergraph process $\boldsymbol{H}_{\mathrm{p},n} = (\boldsymbol{H}_{\mathrm{p},n,S})_S \in (\{0,1\}^{\binom{[n]}{3}})^{\binom{n}{3}}$ given by $\boldsymbol{H}_{\mathrm{p},S}^{-1}(1) = \boldsymbol{E}_{\mathrm{p}}([S])$, and for $p \in [0,1]$ let $\boldsymbol{G}_{\mathrm{b},n,p} \sim \mathrm{Bin}(1,p)^{\otimes\binom{[n]}{2}}$, $\boldsymbol{e}_{\mathrm{p},n} \sim \mathrm{u}(\binom{[n]}{2}!)$, and $\boldsymbol{G}_{\mathrm{p},n} = (\boldsymbol{G}_{\mathrm{p},n,s})_s \in (\{0,1\}^{\binom{[n]}{2}})^{\binom{n}{2}}$ given by $\boldsymbol{G}_{\mathrm{p},s}^{-1}(1) = \boldsymbol{e}_{\mathrm{p}}([s])$, be the corresponding notions for graphs.

4.1.3 Triangle Hypergraphs. Let $G \in \{0,1\}^{\binom{[n]}{2}}$ be a graph, and let $H_{\mathrm{t},G} \in \{0,1\}^{\binom{[n]}{3}}$ given by $H_{\mathrm{t},G}^{-1}(1) = \{E \in \binom{[n]}{3} : \binom{E}{2} \subseteq G^{-1}(1)\}$ be the triangle hypergraph induced by G. There is a clear one to one correspondence between the triangle covers (triangle factors) in G and the hyperedge covers (1-factors) in $H_{\mathrm{t},G}$. Let $\boldsymbol{H}_{\mathrm{t},n,p} = H_{\mathrm{t},\boldsymbol{G}_{\mathrm{b}}}$ be the binomial triangle hypergraph, and let $\boldsymbol{H}_{\mathrm{tp},n} = (\boldsymbol{H}_{\mathrm{tp},n,s})_s = (H_{\mathrm{t},\boldsymbol{G}_{\mathrm{p},s}})_s$ be the triangle hypergraph process. For the hitting times let $\overline{\mathcal{F}}_n = \{F \subseteq \binom{[n]}{3} : |F| = n/3, \bigcup_{E\in F} E = [n]\}$ be the set of all 1-factors (perfect matchings). Then the four hitting times are given by

$$\boldsymbol{S}_{\mathrm{c}}=\inf\left\{S\in\left[\binom{n}{3}\right]:\bigcup_{E\in\boldsymbol{H}_{\mathrm{p},S}^{-1}(1)}E=[n]\right\},\quad \boldsymbol{S}_{\mathrm{f}}=\inf\left\{S\in\left[\binom{n}{3}\right]:\exists F\in\overline{\mathcal{F}}\,F\subseteq\boldsymbol{H}_{\mathrm{p},S}^{-1}(1)\right\},$$

$$\boldsymbol{s}_{\mathrm{c}}=\inf\left\{s\in\left[\binom{n}{2}\right]:\bigcup_{E\in\boldsymbol{H}_{\mathrm{tp},s}^{-1}(1)}E=[n]\right\},\quad \boldsymbol{s}_{\mathrm{f}}=\inf\left\{s\in\left[\binom{n}{2}\right]:\exists F\in\overline{\mathcal{F}}\,F\subseteq\boldsymbol{H}_{\mathrm{tp},s}^{-1}(1)\right\}.$$

For $n\in 3\mathbb{Z}$ we have $\frac{1}{3}n\le\boldsymbol{S}_{\mathrm{c}}\le\boldsymbol{S}_{\mathrm{f}}\le\binom{n}{3}$ almost surely, and $n\le\boldsymbol{s}_{\mathrm{c}}\le\boldsymbol{s}_{\mathrm{f}}\le\binom{n}{2}$ almost surely. Now, $\boldsymbol{S}_{\mathrm{c}}$ is the hitting time for a hyperedge cover of the hypergraph process, and $\boldsymbol{S}_{\mathrm{f}}$ is the hitting time for a 1-factor in the hypergraph process. Analogously, $\boldsymbol{s}_{\mathrm{c}}$ is the hitting time for a triangle cover in the graph process, and $\boldsymbol{s}_{\mathrm{f}}$ is the hitting time for a triangle factor in the graph process.

4.1.4 Perfect Matchings. Theorem 1.3 in [71] states that $\boldsymbol{S}_{\mathrm{f}}$ and $\boldsymbol{S}_{\mathrm{c}}$ coincide whp. The location of $\boldsymbol{S}_{\mathrm{c}}$ is also well-known, and is for example given by Lemma 5.1 in [39]. We combine these results into a single theorem that summarizes the properties of these hitting times for the hypergraph process. For this purpose fix a function $g:\mathbb{Z}_{>0}\to\mathbb{R}_{\ge 0}$ with $g(n)=\omega(1)$, $g(n)\le\ln(n)$ and $g(n)=o(\frac{\ln(n)}{\ln(\ln(n))})$ throughout the remainder. Further, let $\pi_{\pm}=\frac{\ln(n)\pm g(n)}{\binom{n-1}{2}}$, and let $(\boldsymbol{S}_-,\boldsymbol{S}_+)$ be given by $\boldsymbol{S}_+\sim\mathrm{Bin}(\binom{n}{3},\pi_+)$ and $(\boldsymbol{S}_-|\boldsymbol{S}_+=S_+)\sim\mathrm{Bin}(S_+,\pi_-/\pi_+)$, independent of $\boldsymbol{E}_{\mathrm{p}}$.

Theorem 4.1. *We have $\boldsymbol{S}_-\le\boldsymbol{S}_{\mathrm{c}}\le\boldsymbol{S}_+$ whp. Further, we have $\boldsymbol{S}_{\mathrm{f}}=\boldsymbol{S}_{\mathrm{c}}$ whp over $n\in 3\mathbb{Z}_{>0}$.*

The location of the triangle cover hitting time $\boldsymbol{s}_{\mathrm{c}}$ is also well-known, and for example given by Theorem 3.22 in [68]. Notice that we have $\boldsymbol{s}_{\mathrm{c}}\le\boldsymbol{s}_{\mathrm{f}}$ since a factor is a cover. The ingenious coupling underlying Theorem 2 in [63], which is based on [113], embeds a binomial hypergraph into the triangle hypergraph. Theorem 4.5 below is a refined version of this result. This embedding ensures the existence of a triangle factor, and thereby establishes a tight upper bound for $\boldsymbol{s}_{\mathrm{f}}$.

4.1.5 Main Result. The main result states that the random process hypergraph at the hitting time can be embedded into the random process triangle hypergraph at the cover hitting time, and hence the triangle cover and the triangle factor hitting time coincide whp.

Theorem 4.2. *There exists a coupling $(\tilde{\boldsymbol{e}}_{\mathrm{p}},\tilde{\boldsymbol{E}}_{\mathrm{p}})$ of $\boldsymbol{e}_{\mathrm{p}}$ and $\boldsymbol{E}_{\mathrm{p}}$ such that $\tilde{\boldsymbol{H}}_{\mathrm{p},\tilde{\boldsymbol{S}}_{\mathrm{c}}}\le\tilde{\boldsymbol{H}}_{\mathrm{tp},\tilde{\boldsymbol{s}}_{\mathrm{c}}}$ whp, using the notions for $(\tilde{\boldsymbol{e}}_{\mathrm{p}},\tilde{\boldsymbol{E}}_{\mathrm{p}})$ corresponding to $\boldsymbol{e}_{\mathrm{p}}$ and $\boldsymbol{E}_{\mathrm{p}}$. In particular, this gives $\boldsymbol{s}_{\mathrm{f}}=\boldsymbol{s}_{\mathrm{c}}$ whp, over $n\in 3\mathbb{Z}_{>0}$.*

For completeness, we combine all hitting time results for triangles into the following corollary. Let $p_{\pm}=\pi_{\pm}^{1/3}$, and let $(\boldsymbol{s}_-,\boldsymbol{s}_+)$ be given by $\boldsymbol{s}_+\sim\mathrm{Bin}(\binom{n}{2},p_+)$ and $(\boldsymbol{s}_-|\boldsymbol{s}_+=s_+)\sim\mathrm{Bin}(s_+,p_-/p_+)$, independent of $\boldsymbol{e}_{\mathrm{p}}$. Notice that this is Theorem 2.8.

Corollary 4.3. *We have $\boldsymbol{s}_-\le\boldsymbol{s}_{\mathrm{c}}\le\boldsymbol{s}_+$ whp. Further, we have $\boldsymbol{s}_{\mathrm{f}}=\boldsymbol{s}_{\mathrm{c}}$ whp over $n\in 3\mathbb{Z}_{>0}$.*

Finally, we consider the following extension of Theorem 2.6 and Corollary 4.3. For this purpose let $\overline{\mathcal{F}}^*_n=\{F\subseteq\binom{[n]}{3}:|F|=\lceil n/r\rceil,\bigcup_{E\in F}E=[n]\}$ be the covers with the minimal number of hyperedges, and notice that $\overline{\mathcal{F}}^*=\overline{\mathcal{F}}$ for $n\in 3\mathbb{Z}_{>0}$. The corresponding hitting times are

$$\boldsymbol{S}^*_{\mathrm{f}}=\inf\left\{S\in\left[\binom{n}{3}\right]:\exists F\in\overline{\mathcal{F}}\,F\subseteq\boldsymbol{H}_{\mathrm{p},S}^{-1}(1)\right\},\quad \boldsymbol{s}^*_{\mathrm{f}}=\inf\left\{s\in\left[\binom{n}{2}\right]:\exists F\in\overline{\mathcal{F}}\,F\subseteq\boldsymbol{H}_{\mathrm{tp},s}^{-1}(1)\right\}.$$

Corollary 4.4. *Notice that the following holds.*
a) We have $\boldsymbol{S}_- < \boldsymbol{S}_{\mathrm{c}} = \boldsymbol{S}_{\mathrm{f}}^* \leq \boldsymbol{S}_+$ *whp.*
b) We have $\boldsymbol{s}_- < \boldsymbol{s}_{\mathrm{c}} = \boldsymbol{s}_{\mathrm{f}}^* \leq \boldsymbol{s}_+$ *whp.*

4.1.6 Open Problems. Theorem 2.6 is an extension of the corresponding result in [23], and the hitting time results for the connectivity thresholds were extended to hypergraphs in [110], thus we know that the hitting times for minimum degree 1, connectivity and perfect matchings coincide. We further know that the thresholds for minimum degree 2 and 2-connectivity coincide. Although there was significant progress towards the threshold for loose Hamilton cycles [44, 45, 52, 109], both its exact location and the hitting time version extending the corresponding result in [23] are still outstanding.

One may also consider 2-factors, i.e. 2-regular spanning subgraphs, and in particular connected 2-factors. As opposed to loose Hamilton cycles, where all but two 2-degree vertices in each hyperedge have degree 1, all vertices in connected 2-factors have degree 2. Notice that this problem is also still open for regular hypergraphs, as discussed in Section 2.3, respectively Section 5 and [101]. There, we also establish that the exact cover is essentially equivalent, however the problems are different for the binomial model, and locating the threshold for this problem is also still open [72]. For open problems regarding the thresholds for more general F-factors, we refer the reader to [113], in particular the strengthening of Theorem 9 therein for 1-balanced hypergraphs. Therein, Riordan also asks for the constants for $k \geq 4$, which we have determined in Theorem 4.5 for $k = 3$.

4.1.7 Outline of the Proof. We split the proof of Theorem 4.2 into the following three parts. First, we prove a refined version of Theorem 2 in [63].

Theorem 4.5. *For any* $p_\uparrow = \Omega(\ln(n)^{1/3}/n^{2/3})$, $p_\uparrow = o(n^{-20/31})$ *there exists* $c \in \mathbb{R}_{>0}$ *and* $\varepsilon = o(1)$ *such that the following holds. For all* $p \leq p_\uparrow$ *and* $\pi \leq \max(0, (1 - cn^7 p_\uparrow^{11})p^3)$ *there exists a coupling* $(\tilde{\boldsymbol{G}}_{\mathrm{b}}, \tilde{\boldsymbol{H}}_{\mathrm{b}})$ *of* $\boldsymbol{G}_{\mathrm{b}}$ *and* $\boldsymbol{H}_{\mathrm{b}}$ *such that* $\tilde{\boldsymbol{H}}_{\mathrm{b}} \leq H_{\mathrm{t},\tilde{\boldsymbol{G}}_{\mathrm{b}}}$ *with probability at least* $1 - \varepsilon$.

We prove Theorem 4.5 in Section 4.3. Next, we use Theorem 4.5 to couple the hypergraph process $\boldsymbol{H}_{\mathrm{p}}$ and the graph process $\boldsymbol{G}_{\mathrm{p}}$. This coupling fails to completely embed the random process hypergraph in the random process triangle hypergraph, but *almost* all hyperedges are embedded (all but less than $\ln(n)^2$) and in particular a hyperedge cover can be embedded.

Proposition 4.6. *There exists a coupling* $(\tilde{\boldsymbol{e}}_{\mathrm{p}}, \tilde{\boldsymbol{E}}_{\mathrm{p}})$ *of* $\boldsymbol{e}_{\mathrm{p}}$ *and* $\boldsymbol{E}_{\mathrm{p}}$ *such that* $\bigcup_{E \in \boldsymbol{\mathcal{H}}} E = [n]$ *whp, where* $\boldsymbol{\mathcal{H}} = \tilde{\boldsymbol{H}}^{-1}_{\mathrm{p},\tilde{\boldsymbol{S}}_{\mathrm{c}}}(1) \cap \tilde{\boldsymbol{H}}^{-1}_{\mathrm{tp},\tilde{\boldsymbol{s}}_{\mathrm{c}}}(1)$*, and using the notions for* $(\tilde{\boldsymbol{e}}_{\mathrm{p}}, \tilde{\boldsymbol{E}}_{\mathrm{p}})$ *corresponding to* $\boldsymbol{e}_{\mathrm{p}}$ *and* $\boldsymbol{E}_{\mathrm{p}}$.

We prove Proposition 4.6 in Section 4.4. In the third step we extend the coupling in Proposition 4.6 to deal with the excess hyperedges $\boldsymbol{\mathcal{H}} = \tilde{\boldsymbol{H}}^{-1}_{\mathrm{p},\tilde{\boldsymbol{S}}_{\mathrm{c}}}(1) \setminus \tilde{\boldsymbol{H}}^{-1}_{\mathrm{tp},\tilde{\boldsymbol{s}}_{\mathrm{c}}}(1)$. For this purpose, we couple $\tilde{\boldsymbol{E}}_{\mathrm{p}}$ from Proposition 4.6 with $\boldsymbol{E}_{\mathrm{p}}$, i.e. itself, such that the (standard) hypergraph at the cover hitting time of the latter process is entirely contained in the hitting time hypergraph of the former process with the excess hyperedges removed, i.e. in $\tilde{\boldsymbol{H}}^{-1}_{\mathrm{p},\tilde{\boldsymbol{S}}_{\mathrm{c}}}(1) \setminus \boldsymbol{\mathcal{H}}$. This extension of the coupling is discussed in Section 4.5, where we also complete the proof of Theorem 4.2, and establish Corollary 4.3 as well as Corollary 4.4.

4.2 Preparations

We keep the number $n \in \mathbb{Z}_{\geq 12}$ of vertices arbitrary but fixed throughout. For given $p, \pi \in [0, 1]$, we recall a few basics for the binomial hypergraph $\boldsymbol{H}_{\mathrm{b},n,\pi} \sim \mathrm{Bin}(1, \pi)^{\otimes \binom{[n]}{3}}$, further for the hyperedge process $\boldsymbol{E}_{\mathrm{p},n} \sim \mathrm{u}(\binom{[n]}{3}!)$ with associated hypergraph process $\boldsymbol{H}_{\mathrm{p},n} = (\boldsymbol{H}_{\mathrm{p},n,S})_S \in (\{0,1\}^{\binom{[n]}{3}})^{\binom{n}{3}}$ given by $\boldsymbol{H}_{\mathrm{p},S}^{-1}(1) = \boldsymbol{E}_{\mathrm{p}}([S])$, and the corresponding notions $\boldsymbol{G}_{\mathrm{b},n,p} \sim \mathrm{Bin}(1,p)^{\otimes\binom{[n]}{2}}$, $\boldsymbol{e}_{\mathrm{p},n} \sim \mathrm{u}(\binom{[n]}{2}!)$, and $\boldsymbol{G}_{\mathrm{p},n} = (\boldsymbol{G}_{\mathrm{p},n,s})_s \in (\{0,1\}^{\binom{[n]}{2}})^{\binom{n}{2}}$ given by $\boldsymbol{G}_{\mathrm{p},s}^{-1}(1) = \boldsymbol{e}_{\mathrm{p}}([s])$, for graphs.

4.2.1 Uniform Bounds. In order to derive bounds that do not depend on specific choices of sequences $p(n), \pi(n)$ of (hyper-) edge probabilities (which is very restrictive and not required), we fix $\pi_\uparrow, p_\uparrow : \mathbb{Z}_{>0} \to [0,1]$, exclusively consider $\pi \in [0,\pi_\uparrow]$, $p \in [0,p_\uparrow]$, and use the Landau notation as follows. We write $f(n) = \mathcal{O}(f^*(n))$ if there exists $c_{p_\uparrow,\pi_\uparrow}, n_{\circ,p_\uparrow,\pi_\uparrow} \in \mathbb{R}_{>0}$ such that for all $n \in \mathbb{Z}_{\geq n_\circ}$ we have $|f(n)| \leq cf^*(n)$, i.e. the constants hidden in the notation only depend on $p_\uparrow, \pi_\uparrow$. We define $o(f^*(n)), \Omega(f^*(n)), \omega(f^*(n))$ and $\Theta(f^*(n))$ analogously. Clearly, when we consider bounds for $\pi_\uparrow, p_\uparrow$ say $\pi_\uparrow = \omega(n^{-2})$, the constants hidden in the Landau notation are absolute. This also affects whp statements, in that events $\mathcal{E}_{n,p,\pi}$ hold whp if and only if $\mathbb{P}(\mathcal{E}) = 1 + o(1)$, i.e. whp uniformly in $p \in [0,p_\uparrow]$ and $\pi \in [0,\pi_\uparrow]$.

4.2.2 Critical Window Coupling. In this section we discuss the basic coupling underlying the proof of the main result. For this purpose fix the start $\pi_- \in [0,\pi]$ of the critical window. We consider the joint distribution $(\boldsymbol{E}_{\mathrm{p}}, \boldsymbol{S}_+, \boldsymbol{S}_-) \sim \boldsymbol{E}_{\mathrm{p}} \otimes (\boldsymbol{S}_+, \boldsymbol{S}_-)$, where $\boldsymbol{S}_+ \sim \mathrm{Bin}(\binom{n}{3}, \pi)$ and $(\boldsymbol{S}_-|\boldsymbol{S}_+) \sim \mathrm{Bin}(\boldsymbol{S}_+, \pi_-/\pi)$, using $\frac{0}{0} = 1$. On the other hand, let $(\boldsymbol{H}_{\mathrm{b}}, \boldsymbol{E})$ be given by $(\boldsymbol{E}|\boldsymbol{H}_{\mathrm{b}}) \sim \mathrm{u}(\boldsymbol{H}_{\mathrm{b}}^{-1}(1)!)$, and let $(\boldsymbol{H}_{\mathrm{b}}, \boldsymbol{H}_-)$ be given by $(\boldsymbol{H}_-|\boldsymbol{H}_{\mathrm{b}}) \sim \bigotimes_E \mathrm{Bin}(1, \boldsymbol{H}_{\mathrm{b}}(E)\pi_-/\pi)$, using $\frac{0}{0} = 1$, and where the product is over $E \in \binom{[n]}{3}$. Hence, we obtain $\boldsymbol{H}_-$ by thinning out $\boldsymbol{H}_{\mathrm{b}}$, keeping each hyperedge with probability π_-/π. We define $(\boldsymbol{G}_{\mathrm{p}}, \boldsymbol{s}_+, \boldsymbol{s}_-)$, $(\boldsymbol{G}_{\mathrm{b}}, \boldsymbol{e})$, $p_- \in [0,p]$ and $(\boldsymbol{G}_{\mathrm{b}}, \boldsymbol{G}_-)$ analogously.

Observation 4.7. *Notice that the following holds.*
a) We have $(\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_+}, \boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_+]}) \sim (\boldsymbol{H}_{\mathrm{b}}, \boldsymbol{E})$, $(\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_+}, \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_-}) \sim (\boldsymbol{H}_{\mathrm{b}}, \boldsymbol{H}_-)$ *and* $\boldsymbol{H}_- \sim \boldsymbol{H}_{\mathrm{b},\pi_-}$.
b) We have $(\boldsymbol{G}_{\mathrm{p},\boldsymbol{s}_+}, \boldsymbol{e}_{\mathrm{p},[\boldsymbol{s}_+]}) \sim (\boldsymbol{G}_{\mathrm{b}}, \boldsymbol{e})$, $(\boldsymbol{G}_{\mathrm{p},\boldsymbol{s}_+}, \boldsymbol{G}_{\mathrm{p},\boldsymbol{s}_-}) \sim (\boldsymbol{G}_{\mathrm{b}}, \boldsymbol{G}_-)$ *and* $\boldsymbol{G}_- \sim \boldsymbol{G}_{\mathrm{b},p_-}$.

Proof. Let $\overline{\mathcal{H}} = \binom{[n]}{3}$, $\boldsymbol{S} = \boldsymbol{S}_+$, $\boldsymbol{\mathcal{H}} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}}^{-1}(1)$, and notice that $|\boldsymbol{\mathcal{H}}| = \boldsymbol{S}$. Let $\boldsymbol{\mathcal{H}}' = \boldsymbol{H}_{\mathrm{b}}^{-1}(1)$ and $\boldsymbol{S}' = |\boldsymbol{\mathcal{H}}'|$. Notice that $\boldsymbol{S}' \sim \boldsymbol{S}$, further that $(\boldsymbol{\mathcal{H}}'|\boldsymbol{S}') \sim \mathrm{u}(\overline{\mathcal{H}}^{\underline{\boldsymbol{S}'}})$ and that $(\boldsymbol{\mathcal{H}}|\boldsymbol{S}) \sim \mathrm{u}(\overline{\mathcal{H}}^{\underline{\boldsymbol{S}}})$, which shows that $\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}} \sim \boldsymbol{H}_{\mathrm{b}}$. Using another symmetry argument, we have $(\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}]}|\boldsymbol{\mathcal{H}}) \sim \mathrm{u}(\boldsymbol{\mathcal{H}}!)$, and using $(\boldsymbol{E}|\boldsymbol{\mathcal{H}}') \sim \mathrm{u}(\boldsymbol{\mathcal{H}}'!)$ thereby $(\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}}, \boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}]}) \sim (\boldsymbol{H}_{\mathrm{b}}, \boldsymbol{E})$.

For the second part, let $\boldsymbol{S}'_- = |\boldsymbol{H}_-^{-1}(1)|$, notice that $|\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_-}^{-1}(1)| = \boldsymbol{S}_-$, further that $(\boldsymbol{S}'_-|\boldsymbol{H}_{\mathrm{b}}) \sim \mathrm{Bin}(\boldsymbol{S}', \pi_-/\pi)$, and that $(\boldsymbol{S}_-|\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}}) \sim \mathrm{Bin}(\boldsymbol{S}, \pi_-/\pi)$. Let $\boldsymbol{\mathcal{H}}_- = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_-}^{-1}(1)$ and $\boldsymbol{\mathcal{H}}'_- = \boldsymbol{H}_-^{-1}(1)$. Notice that $(\boldsymbol{\mathcal{H}}_-|\boldsymbol{S}_-, \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}}) \sim \mathrm{u}(\binom{\boldsymbol{\mathcal{H}}}{\boldsymbol{S}_-})$ and that $(\boldsymbol{\mathcal{H}}'_-|\boldsymbol{S}'_-, \boldsymbol{H}_{\mathrm{b}}) \sim \mathrm{u}(\binom{\boldsymbol{\mathcal{H}}'}{\boldsymbol{S}'_-})$ using a symmetry argument, so $(\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}}, \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_-}) \sim (\boldsymbol{H}_{\mathrm{b}}, \boldsymbol{H}_-)$. Finally, we have $\boldsymbol{H}_- \sim \boldsymbol{H}_{\mathrm{b},\pi_-}$ since independence is preserved and each hyperedge is present with probability π_-. The results for graphs follow analogously. □

In more detail, we may start with $\boldsymbol{H}_{\mathrm{b}}$, obtain $\boldsymbol{H}_-$ by thinning out, order the hyperedges of $\boldsymbol{H}_-$ to obtain the process up to the start of the critical window, and then order the remaining hyperedges $\boldsymbol{H}_{\mathrm{b}}^{-1}(1) \setminus \boldsymbol{H}_-^{-1}(1)$ to obtain the process up to the end of the critical window.

4.2.3 Hypergraph Degrees. We discuss the vertex degrees for $\boldsymbol{H}_{\mathrm{b}}$. For $H \in \{0,1\}^{\binom{[n]}{3}}$ and $v \in [n]$ let $D_H(v) = |\{E \in \binom{[n]}{3} : v \in E\}|$ be the degree of v in H and let $\Delta_{\mathrm{h}}(H) = \max_v D_H(v)$ be the maximum degree. Define $d_G(v) = |\{e \in \binom{[n]}{2} : v \in e\}|$ and $\Delta_{\mathrm{g}}(G) = \max_v d_G(v)$ for $G \in \{0,1\}^{\binom{[n]}{2}}$ analogously. Further, let $\boldsymbol{\Delta}_{\mathrm{h},n,\pi} = \Delta_{\mathrm{h}}(\boldsymbol{H}_{\mathrm{b}})$ and $\boldsymbol{\Delta}_{\mathrm{g},n,p} = \Delta_{\mathrm{g}}(\boldsymbol{G}_{\mathrm{b}})$. Let $\bar{D}_{n,\pi} = \mathbb{E}[D_{\boldsymbol{H}_{\mathrm{b}}}(v)]$ and $\bar{d}_{n,p} = \mathbb{E}[d_{\boldsymbol{G}_{\mathrm{b}}}(v)]$ denote the expected degrees.

Observation 4.8. *Let $v \in [n]$, $\Delta_{\mathrm{h}\uparrow,n,\pi} = \overline{D}_\pi + \max(\overline{D}_\pi, 4\ln(n))$, $\Delta_{\mathrm{g}\uparrow,n,p} = \bar{d}_p + \max(\bar{d}_p, 4\ln(n))$.*
a) We have $D_{\boldsymbol{H}_\mathrm{b}}(v) \sim \mathrm{Bin}(\binom{n-1}{2}, \pi)$ and $d_{\boldsymbol{G}_\mathrm{b}}(v) \sim \mathrm{Bin}(n-1, p)$.
b) We have $\mathbb{P}(\boldsymbol{\Delta}_\mathrm{h} \geq \Delta_{\mathrm{h}\uparrow}) \leq n^{-1/3}$ and $\mathbb{P}(\boldsymbol{\Delta}_\mathrm{g} \geq \Delta_{\mathrm{g}\uparrow}) \leq n^{-1/3}$.

Proof. Since $D_{\boldsymbol{H}_\mathrm{b}}(v) = \sum_{E\in\mathcal{H}} \boldsymbol{H}_\mathrm{b}(E)$, $\mathcal{H} = \{E \in \binom{[n]}{3} : v \in E\}$, is a sum of iid Bernoulli variables, we have $D_{\boldsymbol{H}_\mathrm{b}}(v) \sim \mathrm{Bin}(\binom{n-1}{2}, \pi)$. Thus, combining Observation 4.7 with the union bound and the Chernoff bound, using $\delta = \Delta_{\uparrow\mathrm{h}} - \overline{D}_\pi = \max(\overline{D}_\pi, 4\ln(n))$, yields

$$\mathbb{P}(\boldsymbol{\Delta}_\mathrm{h} \geq \Delta_{\mathrm{h}\uparrow}) \leq n \exp\left(-\frac{\delta^2}{2\overline{D}_\pi + \delta}\right) \leq \frac{1}{n^{1/3}}.$$

The claims for the graph follow analogously. □

4.2.4 Avoidable Configurations. In the following, we will frequently discuss copies of some hypergraph in the random hypergraphs. Whenever no such copies exist whp, we can use this to significantly simplify the discussion. In this section, we present this argument, which is closely related to the avoidable configurations in [113] and [63]. Fix a set $\mathcal{H} \subseteq \binom{\mathbb{Z}}{3}$ of hyperedges. Let $\mathcal{V} = \mathcal{V}(\mathcal{H}) = \bigcup_{E\in\mathcal{H}} E$ be the vertices, and for $H \in \{0,1\}^{\binom{[n]}{3}}$ let

$$N_{\mathrm{hc},H}(\mathcal{H}) = \left|\left\{\{\iota(E) : E \in \mathcal{H}\} \subseteq H^{-1}(1) : \iota \in [n]^{\underline{\mathcal{V}}}\right\}\right|$$

be the number of copies of $\mathcal{H}$ in H. Similarly, for $\mathcal{G} \subseteq \binom{\mathbb{Z}}{2}$, $\mathcal{V} = \mathcal{V}(\mathcal{G})$, and $G \in \{0,1\}^{\binom{[n]}{2}}$ let $N_{\mathrm{gc},G}(\mathcal{G}) = |\{\{\iota(e) : e \in \mathcal{G}\} \subseteq G^{-1}(1) : \iota \in [n]^{\underline{\mathcal{V}}}\}|$ be the number of copies of $\mathcal{G}$ in G.

Observation 4.9. *Notice that the following holds.*
a) Let $\mathcal{H} \subseteq \binom{[n]}{3}$ and $V = |\mathcal{V}(\mathcal{H})|$. Then we have $\mathbb{E}[N_{\mathrm{hc},\boldsymbol{H}_\mathrm{b}}(\mathcal{H})] \leq n^{\underline{V}}\pi^{|\mathcal{H}|} \leq n^V \pi_\uparrow^{|\mathcal{H}|}$.
b) Let $\mathcal{G} \subseteq \binom{[n]}{2}$ and $V' = |\mathcal{V}(\mathcal{G})|$. Then we have $\mathbb{E}[N_{\mathrm{gc},\boldsymbol{G}_\mathrm{b}}(\mathcal{G})] \leq n^{\underline{V'}}p^{|\mathcal{G}|} \leq n^{V'} p_\uparrow^{|\mathcal{G}|}$.

Proof. Let $H^* \in \{0,1\}^{\binom{[n]}{3}}$, $H^* \equiv 1$, be the complete hypergraph and $V = |\mathcal{V}(\mathcal{H})|$. Then we have $\mathbb{E}[N_{\mathrm{hc},\boldsymbol{H}_\mathrm{b}}(\mathcal{H})] = N_{\mathrm{hc},H^*}(\mathcal{H})\pi^{|\mathcal{H}|} \leq n^{\underline{V}}\pi^{|\mathcal{H}|}$. The proof for graphs is analogous. □

4.2.5 Induced Hyperedges. We describe the triangle hypergraphs $\mathcal{T}_\mathrm{h} = \{H_{\mathrm{t},G} : G \in \{0,1\}^{\binom{[n]}{2}}\}$, in terms of triangle graphs and 3-cycles, given as follows. For $H \in \{0,1\}^{\binom{[n]}{3}}$ let $G_{\mathrm{t},H} \in \{0,1\}^{\binom{[n]}{3}}$ given by $G_{\mathrm{t},H}^{-1}(1) = \bigcup_{E\in H^{-1}(1)} \binom{E}{2}$ be the triangle graph induced by H. This gives rise to the set $\mathcal{T}_\mathrm{g} = \{G_{\mathrm{t},H} : H \in \{0,1\}^{\binom{[n]}{3}}\}$ of all triangle graphs.

The 3-cycle $C_{E,v} \subseteq \binom{[n]}{3}$ with interior $E \in \binom{[n]}{3}$ and outer corners $v \in ([n] \setminus E)^{\binom{E}{2}}$ is given by $C_{E,v} = \{e \cup \{v_e\} : e \in \binom{E}{2}\}$. Let $\overline{\mathcal{C}}_n = \{C_{E,v} : E \in \binom{[n]}{3}, v \in ([n] \setminus E)^{\binom{E}{2}}\}$ be the set of all 3-cycles. For a hypergraph $H \in \{0,1\}^{\binom{[n]}{3}}$ let $\mathcal{C}_\mathrm{h}(H) = \{C \in \overline{\mathcal{C}} : C \subseteq H^{-1}(1)\}$ be the 3-cycles in H.

Observation 4.10. *Let $H \in \{0,1\}^{\binom{[n]}{3}}$, $G' = G_{\mathrm{t},H}$, further $G \in \{0,1\}^{\binom{[n]}{2}}$ and $H' = H_{\mathrm{t},G}$.*
a) We have $G_{\mathrm{t},H'} \leq G$ and $H_{\mathrm{t},G'}^{-1}(1) = H^{-1}(1) \cup \{E \in \binom{[n]}{3} : v \in ([n] \setminus E)^{\binom{E}{2}}, C_{E,v} \in \mathcal{C}_\mathrm{h}(H)\}$.
b) The map $\mathcal{T}_\mathrm{g} \to \mathcal{T}_\mathrm{h}$, $G \mapsto H_{\mathrm{t},G}$, is a bijection with inverse $H \mapsto G_{\mathrm{t},H}$.
c) We have $\mathcal{T}_\mathrm{h} = \{H \in \{0,1\}^{\binom{[n]}{3}} : H_{\mathrm{t},G_{\mathrm{t},H}} = H\}$.

Proof. We have $G_{\mathrm{t},H'} \leq G$ because $\tilde{G} = G_{\mathrm{t},H'}$ are exactly the triangles in G, that is, since we have $\tilde{G}^{-1}(1) = \{e \in \binom{E}{2} : E \in \binom{[n]}{3}, \binom{E}{2} \subseteq G^{-1}(1)\}$. Since $\tilde{G}$ are the triangles in G, we have $H_{\mathrm{t},\tilde{G}} = H'$, so $\mathcal{T}_{\mathrm{h}} = \{H_{\mathrm{t},G} : G \in \mathcal{T}_{\mathrm{g}}\}$. On the other hand, we have $H \leq H_{\mathrm{t},G'}$ because $\tilde{H} = H_{\mathrm{t},G'}$ is induced by the triangles G' given by H, i.e.

$$\tilde{H}^{-1}(1) = \left\{E \in \binom{[n]}{3} : E' \in (H^{-1}(1))^{\binom{E}{2}}, \forall e \in \binom{E}{2}\, e \subseteq E'_e\right\},$$

which gives $\tilde{H}(E) = 1$ for $E \in H^{-1}(1)$, by choosing $E' \equiv E$. But we may also choose $E' \in C_{\mathrm{h}}(H)$ with interior E, which gives $\mathcal{I}(H) \subseteq \tilde{H}^{-1}(1)$ for the induced hyperedges

$$\mathcal{I}(H) = \left\{E \in \binom{[n]}{3} : v \in ([n] \setminus E)^{\binom{E}{2}}, C_{E,v} \in C_{\mathrm{h}}(H)\right\}.$$

Conversely, for any $E \in \tilde{H}^{-1}(1) \setminus H^{-1}(1)$ there exists $E' \in (H^{-1}(1))^{\binom{E}{2}}$ such that $e \subseteq E'_e$ for all $e \in \binom{E}{2}$, i.e. $E'_e = e \cup \{v_e\}$ for some $v_e \in [n] \setminus e$. With $E \in H^{-1}(0)$ we have $E'_e \neq E$, hence $v = (v_e)_e \in ([n] \setminus E)^{\binom{E}{2}}$, thereby $C_{E,v} \in C_{\mathrm{h}}(H)$ and thus $E \in \mathcal{I}(H)$. This shows that $\tilde{H}^{-1}(1) = H^{-1}(1) \cup \mathcal{I}(H)$. As before, since $\tilde{H}$ are the (hyperedges corresponding to) triangles in G', we have $G_{\mathrm{t},\tilde{H}} = G'$, and hence $\mathcal{T}_{\mathrm{g}} \to \mathcal{T}_{\mathrm{h}}$, $G \mapsto H_{\mathrm{t},G}$, is a bijection with inverse $H \mapsto G_{\mathrm{t},H}$. The last part is now immediate. □

An implication of Observation 4.10 is that $H \in \mathcal{T}_{\mathrm{h}}$ if and only if all hyperedges E induced by 3-cycles $C_{E,v} \in C_{\mathrm{h}}(H)$ in H are included in H, that is $H(E) = 1$.

We are particularly interested in the clean 3-cycles $\overline{\mathcal{C}}^{\circ}_n = \{C_{E,v} : E \in \binom{[n]}{3}, v \in ([n] \setminus E)^{\underline{\binom{E}{2}}}\}$, i.e. 3-cycles with distinct outer corners, and the clean 3-cycles $\mathcal{C}^{\circ}(H) = \overline{\mathcal{C}}^{\circ} \cap \mathcal{C}_{\mathrm{h}}(H)$ in $H \in \{0,1\}^{\binom{[n]}{3}}$. Notice that $\{(E,v) : E \in \binom{[n]}{3}, v \in ([n] \setminus E)^{\underline{\binom{E}{2}}}\} \to \overline{\mathcal{C}}^{\circ}$, $(E,v) \mapsto C_{E,v}$, is a bijection. Hence, for $C^{\circ} \in \overline{\mathcal{C}}^{\circ}$ there exists a unique induced hyperedge $E_{\mathrm{ci}}(C) \in \binom{[n]}{3}$. Among the clean 3-cycles, we are particularly interested in the vertex disjoint clean 3-cycles. So, for a family $\mathcal{S} = \{\mathcal{H}_i : i \in [N]\}$ of $N \in \mathbb{Z}_{\geq 0}$ sets $\mathcal{H}_i \subseteq \binom{[n]}{3}$ of hyperedges let $\mathrm{vdj}(\mathcal{S}) = \mathbb{1}\{\forall \mathcal{H} \in \mathcal{S}^{\underline{2}}\, \mathcal{V}(\mathcal{H}_1) \cap \mathcal{V}(\mathcal{H}_2) = \emptyset\}$ indicate if the hyperedge sets $\mathcal{S}$ are pairwise vertex disjoint, where $\mathcal{V}(\mathcal{H}) = \bigcup_{E \in \mathcal{H}} E$.

On the next level, not only triangles can be induced (by 3-cycles), but also 3-cycles can be induced (by inducing one of their hyperedges using 3-cycles). We are particularly interested in induced clean 3-cycles, induced by vertex disjoint clean 3-cycles. In this constellation, two of the hyperedges of the induced clean 3-cycle have to already be present, thus the pair of induced and inducing clean 3-cycle is given by a total of five hyperedges. In this sense, the set of all induced clean 3-cycles is

$$\overline{\mathcal{C}}^{\circ}_{\mathrm{i},n} = \left\{(C_1 \cup C_2) \setminus \{E_{\mathrm{ci}}(C_2)\} : C \in \overline{\mathcal{C}}^{\circ 2}, E_{\mathrm{ci}}(C_2) \in C_1\right\}.$$

Finally, let $\mathcal{C}_{\mathrm{i}}(H) = \{C^{\circ}_{\mathrm{i}} \in \overline{\mathcal{C}}^{\circ}_{\mathrm{i}} : C^{\circ}_{\mathrm{i}} \subseteq H^{-1}(1)\}$ be the induced clean 3-cycles in $H \in \{0,1\}^{\binom{[n]}{3}}$. For the random instances we consider the 3-cycles $\boldsymbol{\mathcal{C}}_{\mathrm{h},n,\pi} = \mathcal{C}(\boldsymbol{H}_{\mathrm{b}})$, the clean 3-cycles $\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h},n,\pi} = \mathcal{C}^{\circ}(\boldsymbol{H}_{\mathrm{b}})$, $\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{t},n,p} = \mathcal{C}^{\circ}(\boldsymbol{H}_{\mathrm{t}})$, and the induced clean 3-cycles $\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{hi}} = \mathcal{C}^{\circ}_{\mathrm{i}}(\boldsymbol{H}_{\mathrm{b}})$.

Observation 4.11. *Notice that the following holds.*

a) We have $\mathbb{P}(\boldsymbol{\mathcal{C}}_{\mathrm{h}} \neq \boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}), \mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{i}} \neq \emptyset), \mathbb{P}(\mathrm{vdj}(\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}) = 0) = o(1)$ *for* $\pi_{\uparrow} = o(n^{-11/6})$.

b) We have $\mathbb{P}(\mathrm{vdj}(\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{t}}) = 0) = o(1)$ *for* $p_{\uparrow} = o(n^{-7/11})$.

c) We have $\mathbb{E}[|\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}|] = 120\binom{n}{6}\pi^3$ *and* $\mathbb{E}[|\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{t}}|] = 120\binom{n}{6}p^9$.

d) We have $\mathrm{Var}(|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|) \leq (1+o(1))\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|]$ *for* $\pi_\uparrow = o(n^{-5/3})$.
e) We have $\mathrm{Var}(|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}|) \leq (1+o(1))\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}|]$ *for* $p_\uparrow = o(n^{-5/9})$.

Proof. For Part 4.11a) we use Observation 4.9. The set $\overline{\mathcal{C}} \setminus \overline{\mathcal{C}}^\circ$ consists of copies of $\mathcal{H}_4 = \{E_1, E_2, E_3\}$ with $E_1 = \{1,2,4\}$, $E_2 = \{1,3,4\}$, $E_3 = \{2,3,4\}$, and of copies of $\mathcal{H}_5 = \{E_1, E_2, E_3\}$ with $E_1 = \{1,2,4\}$, $E_2 = \{1,3,4\}$, $E_3 = \{2,3,5\}$, so the union bound yields

$$\mathbb{P}(\boldsymbol{\mathcal{C}}_{\mathrm{h}} \neq \boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}) \leq \mathbb{P}(N_{\mathrm{c},\boldsymbol{H}_{\mathrm{b}}}(\mathcal{H}_4) > 0) + \mathbb{P}(N_{\mathrm{c},\boldsymbol{H}_{\mathrm{b}}}(\mathcal{H}_5) > 0) \leq n^4\pi_\uparrow^3 + n^5\pi_\uparrow^3 \leq 2n^5\pi_\uparrow^3 = o(1).$$

We proceed similarly for $\mathbb{P}(\boldsymbol{\mathcal{C}}^\circ_{\mathrm{i}} \neq \emptyset)$. For $C^\circ_{\mathrm{i}} \in \overline{\mathcal{C}}^\circ_{\mathrm{i}}$ we have $|C^\circ_{\mathrm{i}}| = 5$ and $|\mathcal{V}(C^\circ_{\mathrm{i}})| \leq 9$, yielding $\mathbb{P}(\boldsymbol{\mathcal{C}}^\circ_{\mathrm{i}} \neq \emptyset) \leq \binom{9}{3}^{\underline{5}} n^9\pi_\uparrow^5 = o(1)$ since there are $\binom{9}{3}^{\underline{5}}$ sets $\mathcal{H} \subseteq \binom{[9]}{3}$ of hyperedges of size $|\mathcal{H}| = 5$ in total. Since the computations for vertex disjoint clean 3-cycles are lengthy and closely related to the variance, we postpone the proof of the remainder.

For Part 4.11c), we reconcile the leading coefficient 120. Given six vertices, choose three for the first triangle. For the second triangle, choose one of the vertices in the first triangle, and two of the remaining three vertices. For the third triangle, choose one of the remaining two vertices in the first triangle, and one of the remaining two in the second triangle. In total, this gives $\binom{6}{3}\binom{3}{1}\binom{3}{2}\binom{2}{1}^2 = 6!$, and dividing by 6 to drop the order yields $5! = 120$.

For the remainder, i.e. the remainder of Part 4.11a), for Part 4.11b), Part 4.11d) and Part 4.11e), we consider pairs of clean 3-cycles. For $r \in \mathbb{Z}_{\geq 3}$ and $s \in \{0,1\}$ let

$$\overline{\mathcal{P}}_{\mathrm{h},r,s} = \left\{C \in (\overline{\mathcal{C}}^\circ)^2 : |C_1 \cup C_2| = r,\, \mathrm{vdj}(\{C_1, C_2\}) = s\right\}, \overline{\mathcal{U}}_{\mathrm{h},r,s} = \left\{C_1 \cup C_2 : C \in \overline{\mathcal{P}}_{\mathrm{h},r,s}\right\}.$$

Let $\boldsymbol{\mathcal{U}}_{\mathrm{h},n,\pi,r,s} = \{U \in \overline{\mathcal{U}}_{\mathrm{h},r,s} : U \subseteq \boldsymbol{H}_{\mathrm{b}}^{-1}(1)\}$ and $\boldsymbol{\mathcal{P}}_{\mathrm{h},n,\pi,r,s} = \{C \in \overline{\mathcal{P}}_{\mathrm{h},r,s} : C_1 \cup C_2 \subseteq \boldsymbol{H}_{\mathrm{b}}^{-1}(1)\}$. Notice that the sets are empty unless $3 \leq r \leq 5$, $s = 0$, or $r = 6$, and that $|\boldsymbol{\mathcal{U}}_{\mathrm{h}}| \leq |\boldsymbol{\mathcal{P}}_{\mathrm{h}}| \leq \binom{r}{3}^2 |\boldsymbol{\mathcal{U}}_{\mathrm{h}}|$, by selecting any pair of three hyperedges each. Let $P_{\mathrm{h},n,\pi,r,s} = \mathbb{E}[|\boldsymbol{\mathcal{P}}_{\mathrm{h}}|]$, $U \in \overline{\mathcal{U}}_{\mathrm{h}}$ and $V = \bigcup_{E \in U} E$.

Fix $s = 0$ for now. For $r = 3$ we have $P_{\mathrm{h}} = \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|]$. For $r = 4$ we have $V \leq 7$ and hence $P_{\mathrm{h}} \leq \binom{4}{3}^2\binom{7}{3}^{\underline{4}} n^7\pi^4$ by Observation 4.9, analogous to the above. For $r = 5$ we have $V \leq 9$ and thus $P_{\mathrm{h}} \leq \binom{5}{3}^2\binom{9}{3}^{\underline{5}} n^9\pi^5$. For $r = 6$ we have $V \leq 11$ since $s = 0$ and thereby $P_{\mathrm{h}} \leq \binom{6}{3}^2\binom{11}{3}^{\underline{6}} n^{11}\pi^6$. For $r = 6$ and $s = 1$ we have $P_{\mathrm{h}} = 120^2\binom{n}{6,6,n-12}\pi^6 \leq \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|]^2$ using the proof of Part 4.11c), so

$$P_{\mathrm{h},3} = \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|],\ P_{\mathrm{h},4} \leq cn^7\pi^4,\ P_{\mathrm{h},5} \leq cn^9\pi^5,\ P_{\mathrm{h},6,0} \leq cn^{11}\pi^6,\ P_{\mathrm{h},6,1} \leq \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|]^2 \tag{13}$$

for some $c \in \mathbb{R}_{>0}$. This gives $\mathbb{P}\left(\mathrm{vdj}(\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}) = 0\right) \leq P_{\mathrm{h},4} + P_{\mathrm{h},5} + P_{\mathrm{h},6,0} = o(1)$, and further

$$\begin{aligned}\mathrm{Var}(|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|) &= \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|^2] - \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|]^2 \leq \sum_{r=3}^{6} \mathcal{P}_{\mathrm{h},r,0} \leq \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|] + c(n^7\pi^4 + n^9\pi^5 + n^{11}\pi^6)\\ &= \left(1 + (1+\mathcal{O}(n^{-1}))6c(n\pi + n^3\pi^2 + n^5\pi^3)\right)\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|].\end{aligned}$$

Since we have $\mathrm{Var}(|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|) = \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|] = 0$ for $\pi = 0$, this shows that $\mathrm{Var}(|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|) = (1+o(1))\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|]$. For the triangle hypergraph we proceed analogously, but we have to count the edges in the union, not the hyperedges. Hence, for $\mathcal{H} \subseteq \binom{[n]}{3}$ let $\mathcal{E}(\mathcal{H}) = \bigcup_{E \in \mathcal{H}} \binom{E}{2}$ be the edges and further let

$$\overline{\mathcal{P}}_{\mathrm{t},r,s} = \left\{C \in (\overline{\mathcal{C}}^\circ)^2 : |\mathcal{E}(C_1) \cup \mathcal{E}(C_2)| = r,\, \mathrm{vdj}(\{C_1, C_2\}) = s\right\}, \overline{\mathcal{U}}_{\mathrm{t},r,s} = \left\{C_1 \cup C_2 : C \in \overline{\mathcal{P}}_{\mathrm{t},r,s}\right\}.$$

Notice that the sets are empty unless $9 \leq r \leq 17$ and $s = 0$, or $r = 18$. Adjusting the definitions and following the argumentation above gives $c \in \mathbb{R}_{>0}$ such that

$$P_{\mathrm{t},9} = \mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|],\ P_{\mathrm{t},10} = 0,\ P_{\mathrm{t},r,0} \leq cn^{v_r}p^r, 11 \leq r \leq 18,\ P_{\mathrm{t},18,1} \leq \mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|]^2, \tag{14}$$

where $v_r = \lfloor \frac{1}{2}(r+3) \rfloor$ for $11 \leq r \leq 17$ and $v_{18} = 11$. The remainder follows as above. □

Taking into account $\pi_{\uparrow} \leq p_{\uparrow}^3$, we obtain $p_{\uparrow} = o(n^{-7/11})$ and $\pi_{\uparrow} = p_{\uparrow}^3 = o(n^{-21/11})$.

4.2.6 Diamonds. Let $\overline{\mathcal{D}} = \{\{E, E'\} \subseteq \binom{[n]}{3} : |E \cap E'| = 2\}$ be the set of hyperedge pairs, or diamond graphs (in the appropriate interpretation). For $H \in \{0,1\}^{\binom{[n]}{3}}$ let $\mathcal{D}_{\mathrm{h}}(H) = \{D \in \overline{\mathcal{D}} : D \subseteq H^{-1}(1)\}$, and let $\boldsymbol{\mathcal{D}}_{\mathrm{h},n,\pi} = \mathcal{D}_{\mathrm{h}}(\boldsymbol{H}_{\mathrm{b}})$ be the set of diamonds in $\boldsymbol{H}_{\mathrm{b}}$.

Observation 4.12. *Notice that the following holds.*
a) We have $\mathbb{E}[|\boldsymbol{\mathcal{D}}_{\mathrm{h}}|] = 6\binom{n}{4}\pi^2$.
b) We have $\mathbb{P}(\mathrm{vdj}(\boldsymbol{\mathcal{D}}_{\mathrm{h}} \cup \boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}) = 0) = o(1)$ *for* $\pi_{\uparrow} = o(n^{-11/6})$.
c) We have $\mathrm{Var}(|\boldsymbol{\mathcal{D}}_{\mathrm{h}}|) = (1 + \mathcal{O}(n^3\pi_{\uparrow}^2))\mathbb{E}[|\boldsymbol{\mathcal{D}}_{\mathrm{h}}|] + o(1)$ *for* $\pi_{\uparrow} = o(n^{-7/4})$.

Proof. The coefficient 6 is obtained by choosing two of the four vertices for the overlap edge. For the next part we consider the split $\mathcal{E} = \mathcal{E}_{\mathrm{d}} \cup \mathcal{E}_{\mathrm{c}} \cup \mathcal{E}_{\mathrm{r}}$, where $\mathcal{E} = \{\mathrm{vdj}(\boldsymbol{\mathcal{D}}_{\mathrm{h}} \cup \boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}) = 0\}$, $\mathcal{E}_{\mathrm{d}} = \{\mathrm{vdj}(\boldsymbol{\mathcal{D}}_{\mathrm{h}}) = 0\}$ are overlapping diamonds, $\mathcal{E}_{\mathrm{c}} = \{\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}) = 0\}$ are overlapping clean 3-cycles, and $\mathcal{E}_{\mathrm{r}} = \mathcal{E} \setminus (\mathcal{E}_{\mathrm{d}} \cup \mathcal{E}_{\mathrm{c}})$ is the rest, suggesting that there exists a diamond overlapping with a clean 3-cycle. Analogous to the proof of Observation 4.11, for $\mathcal{E}_{\mathrm{d}}$ and the variance we consider a diamond overlap of 1 hyperedge on at most 5 vertices, and the split for an overlap of 0 hyperedges with at most 7 vertices if the diamonds are not vertex disjoint, and expectation $(1 - \delta(n))\mathbb{E}[|\boldsymbol{\mathcal{D}}_{\mathrm{h}}|]^2$ otherwise, where $\delta(n) = 1 - \binom{n-4}{4}/\binom{n}{4}$. This establishes the asymptotics for the variance and takes care of $\mathcal{E}_{\mathrm{d}}$, since the contribution by diamonds with vertex overlap is $\mathcal{O}(n^5\pi^3 + n^7\pi^4) = o(1)$, while $\mathcal{E}_{\mathrm{c}}$ is taken care of by Observation 4.11. For the 3-cycle diamond pairs, covering $\mathcal{E}_{\mathrm{r}}$, we may have 4 hyperedges with up to 7 vertices or 5 hyperedges with up to 9 vertices since they are not disjoint, yielding the bound $\mathcal{O}(n^7\pi_{\uparrow}^4 + n^9\pi_{\uparrow}^5) = o(1)$. □

4.3 Binomial Graph Coupling

Now, we couple $\boldsymbol{G}_{\mathrm{b},n,p} \sim \mathrm{Bin}(1,p)^{\otimes\binom{[n]}{2}}$ with $\boldsymbol{H}_{\mathrm{b},n,\pi} \sim \mathrm{Bin}(1,\pi)^{\otimes\binom{[n]}{3}}$ and establish Theorem 4.5. We continue to use the convention for the Landau notation from Section 4.2.1.

4.3.1 Overview. We establish Theorem 4.5 using the explicit coupling described in Figure 3, which relies on the notions in Section 4.2.5 and on coupling orders $E_{n,\mathcal{C}}^* \in \binom{[n]}{3}!$ for $\mathcal{C} \subseteq \overline{\mathcal{C}}^{\circ}$ with $\mathrm{vdj}(\mathcal{C}) = 1$, such that the hyperedges in the clean 3-cycles come first, i.e. $E_{\mathcal{C}}^*([|\mathcal{S}|]) = \mathcal{S}$ for $\mathcal{S} = \{E \in C : C \in \mathcal{C}\}$, followed by the induced hyperedges, i.e. $E_{\mathcal{C}}^*([|\mathcal{S} \cup \mathcal{S}'|] \setminus [|\mathcal{S}|]) = \mathcal{S}'$ for $\mathcal{S}' = \{E_{\mathrm{ci}}(C) : C \in \mathcal{C}\}$. It further relies on the conditional hyperedge inclusion probabilities

$$\pi_{\mathrm{t}}(E, \mathcal{C}, \mathcal{Y}, \mathcal{N}) = \mathbb{P}\left(\boldsymbol{H}_{\mathrm{t}}(E) = 1 | \boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ} = \mathcal{C}, \mathcal{Y} \subseteq \boldsymbol{H}_{\mathrm{t}}^{-1}(1), \mathcal{N} \subseteq \boldsymbol{H}_{\mathrm{t}}^{-1}(0)\right),$$
$$\pi_{\mathrm{h}}(E, \mathcal{C}, \mathcal{Y}, \mathcal{N}) = \mathbb{P}\left(\boldsymbol{H}_{\mathrm{b}}(E) = 1 | \boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ} = \mathcal{C}, \mathcal{Y} \subseteq \boldsymbol{H}_{\mathrm{b}}^{-1}(1), \mathcal{N} \subseteq \boldsymbol{H}_{\mathrm{b}}^{-1}(0)\right).$$

In order to establish Theorem 4.5, we show that the coupling is successful, i.e. that the event

$$\mathcal{E} = \left\{\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{t}}^{\circ} = \tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ}, \tilde{\boldsymbol{b}}_{\mathrm{h}} = 1, \tilde{\boldsymbol{H}}_{\mathrm{b}} \leq \tilde{\boldsymbol{H}}_{\mathrm{t}}\right\}, \tilde{\boldsymbol{H}}_{\mathrm{t}} = H_{\mathrm{t},\tilde{\boldsymbol{G}}_{\mathrm{b}}}, \tag{15}$$

1. We use the following coupling $(\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{t}}^{\circ},\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ})$ of the clean 3-cycles $\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}$ and $\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}$.
   (a) Use a maximal coupling $(\tilde{\boldsymbol{n}}_{\mathrm{t}},\tilde{\boldsymbol{n}}_{\mathrm{h}})$ to couple $|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|$ and $|\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|$.
   (b) Given $\mathcal{E}=\{(\tilde{\boldsymbol{n}}_{\mathrm{t}},\tilde{\boldsymbol{n}}_{\mathrm{h}})=(n_{\mathrm{t}},n_{\mathrm{h}})\}$, let $(\tilde{\boldsymbol{b}}_{\mathrm{t}},\tilde{\boldsymbol{b}}_{\mathrm{h}})\in\{0,1\}^2$ be conditionally independent with $(\tilde{\boldsymbol{b}}_{\mathrm{t}}|\mathcal{E})\sim(\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ})||\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|=n_{\mathrm{t}})$ and $(\tilde{\boldsymbol{b}}_{\mathrm{h}}|\mathcal{E})\sim(\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ})||\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|=n_{\mathrm{h}})$.
   (c) We define $(\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{t}}^{\circ},\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ})$ conditional to $\mathcal{E}=\{(\tilde{\boldsymbol{n}}_{\mathrm{t}},\tilde{\boldsymbol{n}}_{\mathrm{h}},\tilde{\boldsymbol{b}}_{\mathrm{t}},\tilde{\boldsymbol{b}}_{\mathrm{h}})=(n_{\mathrm{t}},n_{\mathrm{h}},b_{\mathrm{t}},b_{\mathrm{h}})\}$.
      i. Let $(\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ}|\mathcal{E})\sim(\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|\,\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ})=b_{\mathrm{h}},|\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|=n_{\mathrm{h}})$.
      ii. If the coupling succeeded, i.e. $n_{\mathrm{t}}=n_{\mathrm{h}}$ and $b_{\mathrm{t}}=b_{\mathrm{h}}=1$, set $\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{t}}^{\circ}=\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ}$ almost surely.
      iii. If not, let $(\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{t}}^{\circ},\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ})$ be conditionally independent with $(\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{t}}^{\circ}|\mathcal{E})\sim(\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|\,\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ})=b_{\mathrm{t}},|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|=n_{\mathrm{t}})$.
2. We define $(\tilde{\boldsymbol{G}}_{\mathrm{b}},\tilde{\boldsymbol{H}}_{\mathrm{b}})$ conditional to $\mathcal{E}=\{(\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{t}}^{\circ},\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ},\tilde{\boldsymbol{b}}_{\mathrm{t}},\tilde{\boldsymbol{b}}_{\mathrm{h}})=(\mathcal{C}_{\mathrm{t}},\mathcal{C}_{\mathrm{h}},b_{\mathrm{t}},b_{\mathrm{h}})\}$, by recursively defining yes and no sets $\boldsymbol{x}=(\boldsymbol{\mathcal{Y}}_{\mathrm{t},i},\boldsymbol{\mathcal{Y}}_{\mathrm{h},i},\boldsymbol{\mathcal{N}}_{\mathrm{t},i},\boldsymbol{\mathcal{N}}_{\mathrm{h},i})_i$ over $i\in\mathbb{Z}\cap[0,\binom{n}{3}]$, where 'yes' suggests inclusion.
   (a) If $\mathcal{C}_{\mathrm{t}}\neq\mathcal{C}_{\mathrm{h}}$ or $b_{\mathrm{h}}=0$, set $\boldsymbol{x}_i\equiv\emptyset$ for all i.
   (b) Otherwise, we have $\mathcal{C}_{\mathrm{t}}=\mathcal{C}_{\mathrm{h}}$, $b_{\mathrm{t}}=b_{\mathrm{h}}=1$, and set $\boldsymbol{x}_0\equiv\emptyset$. For $i\in[\binom{n}{3}]$, conditional to previous choices, $\boldsymbol{x}_i$ only depends on $\mathcal{E}$ and $\boldsymbol{x}_{i-1}$. Hence, it is sufficient to define $\boldsymbol{x}_i$ given $\mathcal{E}\cap\{\boldsymbol{x}_{i-1}=x\}$, with $x=(\mathcal{Y}_{\mathrm{t}},\mathcal{Y}_{\mathrm{h}},\mathcal{N}_{\mathrm{t}},\mathcal{N}_{\mathrm{h}})$. Let $\pi_i=\pi_{\mathrm{t}}(E_i^*,\mathcal{C}_{\mathrm{t}},\mathcal{Y}_{\mathrm{t}},\mathcal{N}_{\mathrm{t}})$ and $\pi_i'=\pi_{\mathrm{h}}(E_i^*,\mathcal{C}_{\mathrm{h}},\mathcal{Y}_{\mathrm{h}},\mathcal{N}_{\mathrm{h}})$.
      i. In the case $\mathcal{Y}_{\mathrm{h}}\cap\mathcal{N}_{\mathrm{t}}\neq\emptyset$ let $\boldsymbol{x}_i=x$, otherwise proceed as follows.
      ii. For $\pi_i=\pi_i'=0$ let $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i}=\mathcal{Y}_{\mathrm{t}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{t},i}=\mathcal{N}_{\mathrm{t}}\cup\{E_i^*\}$, $\boldsymbol{\mathcal{Y}}_{\mathrm{h},i}=\mathcal{Y}_{\mathrm{h}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{h},i}=\mathcal{N}_{\mathrm{h}}\cup\{E_i^*\}$.
      iii. Otherwise, for $\pi_i\geq\pi_i'$ let $\boldsymbol{b}\sim\mathrm{Bin}(1,\pi_i'/\pi_i)$.
         A. For $\boldsymbol{b}=0$ let $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i}=\mathcal{Y}_{\mathrm{t}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{t},i}=\mathcal{N}_{\mathrm{t}}$ and $\boldsymbol{\mathcal{Y}}_{\mathrm{h},i}=\mathcal{Y}_{\mathrm{h}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{h},i}=\mathcal{N}_{\mathrm{h}}\cup\{E_i^*\}$.
         B. Given $\boldsymbol{b}=1$ let $\boldsymbol{b}'\sim\mathrm{Bin}(1,\pi_i)$. For $\boldsymbol{b}'=1$ let $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i}=\mathcal{Y}_{\mathrm{t}}\cup\{E_i^*\}$, $\boldsymbol{\mathcal{N}}_{\mathrm{t},i}=\mathcal{N}_{\mathrm{t}}$ and $\boldsymbol{\mathcal{Y}}_{\mathrm{h},i}=\mathcal{Y}_{\mathrm{h}}\cup\{E_i^*\}$, $\boldsymbol{\mathcal{N}}_{\mathrm{h},i}=\mathcal{N}_{\mathrm{h}}$. For $\boldsymbol{b}'=0$ let $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i}=\mathcal{Y}_{\mathrm{t}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{t},i}=\mathcal{N}_{\mathrm{t}}\cup\{E_i^*\}$ and $\boldsymbol{\mathcal{Y}}_{\mathrm{h},i}=\mathcal{Y}_{\mathrm{h}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{h},i}=\mathcal{N}_{\mathrm{h}}\cup\{E_i^*\}$.
      iv. For $\pi_i<\pi_i'$ let $\boldsymbol{m}\in\{0,1,2\}$ with $\mathbb{P}(\boldsymbol{m}=m)=P_m$, $P_2=\pi_i$, $P_1=\pi_i'-\pi_i$, $P_0=1-\pi_i'$.
         A. For $\boldsymbol{m}=2$ let $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i}=\mathcal{Y}_{\mathrm{t}}\cup\{E_i^*\}$, $\boldsymbol{\mathcal{N}}_{\mathrm{t},i}=\mathcal{N}_{\mathrm{t}}$ and $\boldsymbol{\mathcal{Y}}_{\mathrm{h},i}=\mathcal{Y}_{\mathrm{h}}\cup\{E_i^*\}$, $\boldsymbol{\mathcal{N}}_{\mathrm{h},i}=\mathcal{N}_{\mathrm{h}}$.
         B. For $\boldsymbol{m}=1$ let $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i}=\mathcal{Y}_{\mathrm{t}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{t},i}=\mathcal{N}_{\mathrm{t}}\cup\{E_i^*\}$ and $\boldsymbol{\mathcal{Y}}_{\mathrm{h},i}=\mathcal{Y}_{\mathrm{h}}\cup\{E_i^*\}$, $\boldsymbol{\mathcal{N}}_{\mathrm{h},i}=\mathcal{N}_{\mathrm{h}}$.
         C. For $\boldsymbol{m}=0$ let $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i}=\mathcal{Y}_{\mathrm{t}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{t},i}=\mathcal{N}_{\mathrm{t}}\cup\{E_i^*\}$ and $\boldsymbol{\mathcal{Y}}_{\mathrm{h},i}=\mathcal{Y}_{\mathrm{h}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{h},i}=\mathcal{N}_{\mathrm{h}}\cup\{E_i^*\}$.
   (c) Conditional to a particular outcome of the experiment above, $\tilde{\boldsymbol{G}}_{\mathrm{b}}$ and $\tilde{\boldsymbol{H}}_{\mathrm{b}}$ are conditionally independent and only depend on $\mathcal{E}$ and $\boldsymbol{x}_{\binom{n}{3}}$. Hence, using $\mathcal{E}^*=\mathcal{E}\cap\{\boldsymbol{x}_{\binom{n}{3}}=(\mathcal{Y}_{\mathrm{t}},\mathcal{Y}_{\mathrm{h}},\mathcal{N}_{\mathrm{t}},\mathcal{N}_{\mathrm{h}})\}$, their distribution is determined by $(\tilde{\boldsymbol{G}}_{\mathrm{b}}|\mathcal{E}^*)\sim(\boldsymbol{G}_{\mathrm{b}}|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}=\mathcal{C}_{\mathrm{t}},\mathcal{Y}_{\mathrm{t}}\subseteq\boldsymbol{H}_{\mathrm{t}}^{-1}(1),\mathcal{N}_{\mathrm{t}}\subseteq\boldsymbol{H}_{\mathrm{t}}^{-1}(0))$ and $(\tilde{\boldsymbol{H}}_{\mathrm{b}}|\mathcal{E}^*)\sim(\boldsymbol{H}_{\mathrm{b}}|\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}=\mathcal{C}_{\mathrm{h}},\mathcal{Y}_{\mathrm{h}}\subseteq\boldsymbol{H}_{\mathrm{b}}^{-1}(1),\mathcal{N}_{\mathrm{h}}\subseteq\boldsymbol{H}_{\mathrm{b}}^{-1}(0))$.

Figure 3: The two main steps in the coupling of $\boldsymbol{G}_{\mathrm{b}}$ and $\boldsymbol{H}_{\mathrm{b}}$ are the coupling of their clean 3-cycles $\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}$ and $\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}$, followed by the coupling of $\boldsymbol{G}_{\mathrm{b}}$ and $\boldsymbol{H}_{\mathrm{b}}$ given $\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}$ and $\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}$. In the first step we couple the numbers of clean 3-cycles, then determine independently if they are pairwise vertex disjoint, and if the coupling succeeds, i.e. if the numbers of cycles are equal and both are disjoint, then we can choose the same cycles for both hypergraphs, otherwise we choose independently.
In the second step, we first iterate through all hyperedges, where $\boldsymbol{x}_i$ depends on all past decisions only through $\mathcal{E}$ and $\boldsymbol{x}_{i-1}$. Subject to the conditional inclusion probabilities π_i,π_i' at hand, we use different coupling strategies. When the coupling fails, i.e. the case $\mathcal{Y}_{\mathrm{h}}\cap\mathcal{N}_{\mathrm{t}}\neq\emptyset$ where we marked a hyperedge for inclusion in $\tilde{\boldsymbol{H}}_{\mathrm{b}}$ and for exclusion in $H_{\mathrm{t},\tilde{\boldsymbol{G}}_{\mathrm{b}}}$, we stop marking hyperedges. For $\pi_i\geq\pi_i'$ and $\boldsymbol{b}=0$, we do not make a decision for $H_{\mathrm{t},\tilde{\boldsymbol{G}}_{\mathrm{b}}}$ and exclude E_i^* for $\tilde{\boldsymbol{H}}_{\mathrm{b}}$, while for $\boldsymbol{b}=1$ we include or exclude E_i^* in both hypergraphs depending on $\boldsymbol{b}'$. For $\pi_i<\pi_i'$ we use a maximal coupling. Here, the coupling fails if $\boldsymbol{m}=1$, and otherwise we include or exclude E_i^* in both hypergraphs depending on $\boldsymbol{m}$.
In the last step, based on the marked hyperedges, we complete the graph $\tilde{\boldsymbol{G}}_{\mathrm{b}}$ by considering the corresponding triangles, and the hypergraph $\tilde{\boldsymbol{H}}_{\mathrm{b}}$ independently.

holds whp. We show this in three steps. First, we bound the probability for the first part.

Proposition 4.13. *Let $\pi_\uparrow = o(n^{-11/6})$ and $p_\uparrow = o(n^{-7/11})$. Then we have* $\mathbb{P}(\tilde{\boldsymbol{\mathcal{C}}}^\circ_{\mathrm{t}} = \tilde{\boldsymbol{\mathcal{C}}}^\circ_{\mathrm{h}}, \tilde{\boldsymbol{b}}_{\mathrm{h}} = 1) \geq 1 - \min(1, \frac{1}{\sqrt{2\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|]}}, \frac{1}{\sqrt{2\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}|]}})|\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}|] - \mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|]| + o(1)$.

We turn to the conditional probabilities π_{t} and π_{h}. The domain for π_{t} is

$$\mathcal{D}_{\mathrm{t}} = \left\{(E, \mathcal{C}, \mathcal{Y}, \mathcal{N}) : E \in \binom{[n]}{3}, \mathbb{P}\left(\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}} = \mathcal{C}, \mathcal{Y} \subseteq \boldsymbol{H}_{\mathrm{t}}^{-1}(1), \mathcal{N} \subseteq \boldsymbol{H}_{\mathrm{t}}^{-1}(1)\right) > 0\right\},$$

and $\mathcal{D}_{\mathrm{h}}$ for π_{h} is defined analogously using $\boldsymbol{H}_{\mathrm{b}}$. For $X = (E, \mathcal{C}, \mathcal{Y}, \mathcal{N}) \in \mathcal{D}_{\mathrm{h}}$ let $\overline{\mathcal{Y}} = \mathcal{Y} \cup \bigcup_{C \in \mathcal{C}} C$, let $H \in \{0,1\}^{\binom{[n]}{3}}$ be given by $H^{-1}(1) = \overline{\mathcal{Y}} \cup \{E\}$ and let $\Delta_{\mathrm{c}}(X) = \Delta_{\mathrm{h}}(H)$ be the minimal maximal degree for a hypergraph subject to X. Aiming towards Observation 4.8, we consider the restriction $\mathcal{D}^\circ_{\mathrm{t},n} = \{X \in \mathcal{D}_{\mathrm{t}} : \Delta_{\mathrm{c}}(X) < \binom{n-1}{2}p_\uparrow^3 + \max(\binom{n-1}{2}p_\uparrow^3, 4\ln(n))\}$ be the restriction to bounded degrees. Define $\mathcal{D}^\circ_{\mathrm{h}}$ analogously. The first result addresses triangle hypergraphs.

Proposition 4.14. *Let $X = (E, \mathcal{C}, \mathcal{Y}, \mathcal{N}) \in \mathcal{D}_{\mathrm{t}}$ and $\overline{\mathcal{Y}} = \mathcal{Y} \cup \bigcup_{C \in \mathcal{C}} C$.*

a) For $p_\uparrow = \Omega(\ln(n)^{1/3}/n^{2/3})$, $p_\uparrow = o(n^{-7/11})$ there exists $c \in \mathbb{R}_{>0}$ such that the following holds. For all $X \in \mathcal{D}^\circ_{\mathrm{t}}$ with $\pi_{\mathrm{t}}(X) > 0$ we have $\pi_{\mathrm{t}}(X) \geq (1 - cn^7 p_\uparrow^{11}))p^3$.

b) We have $\pi_{\mathrm{t}}(X) = 0$ if $p = 0$, $E \in \mathcal{N}$ or $C \subseteq \overline{\mathcal{Y}} \cup \{E\}$ for some $C \in \overline{\mathcal{C}}^\circ \setminus \mathcal{C}$.

c) Assume that $\mathrm{vdj}(\mathcal{C}) = 1$, that $\pi_{\mathrm{t}}(X) = 0$ and that the assumptions in Part 4.14b) do not hold. Then we have $\mathcal{C}_{\mathrm{h}}(H) \setminus \mathcal{C}^\circ_{\mathrm{h}}(H) \neq \emptyset$ or $\mathcal{C}^\circ_{\mathrm{i}}(H) \neq \emptyset$.

Now, we turn to the corresponding results for the binomial hypergraph.

Proposition 4.15. *Let $X = (E, \mathcal{C}, \mathcal{Y}, \mathcal{N}) \in \mathcal{D}_{\mathrm{h}}$ and $\overline{\mathcal{Y}} = \mathcal{Y} \cup \bigcup_{C \in \mathcal{C}} C$.*

a) For $E \notin \overline{\mathcal{Y}}$ we have $\pi_{\mathrm{h}}(X) \leq \pi$.

b) For $\pi_\uparrow = \Omega(\ln(n)/n^2)$, $\pi_\uparrow = o(n^{-3/2})$ there exists $c \in \mathbb{R}_{>0}$ such that the following holds. For all $X \in \mathcal{D}^\circ_{\mathrm{h}}$ with $\pi_{\mathrm{h}}(X) > 0$ we have $\pi_{\mathrm{h}}(X) \geq (1 - cn^3\pi_\uparrow^2)\pi$.

c) We have $\pi_{\mathrm{h}}(X) = 0$ if and only if $\pi = 0$, $E \in \mathcal{N}$ or $C \subseteq \overline{\mathcal{Y}} \cup \{E\}$ for some $C \in \overline{\mathcal{C}}^\circ \setminus \mathcal{C}$.

Notice that there is no reasonable upper bound for $\pi_{\mathrm{t}}(X)$, in particular since the induced hyperedges given by $\mathcal{C}$ are definitely included. Also, notice that $\pi_{\mathrm{h}}(X) = 0$ implies that $\pi_{\mathrm{t}}(X) = 0$ (or $\pi = 0$). Using these two results, it is immediate that the coupling is successful whp.

Corollary 4.16. *Let $p_\uparrow = \Omega(\ln(n)^{1/3}/n^{2/3})$, $p_\uparrow = o(n^{-20/31})$ and c from Proposition 4.14a). Then we have $\mathbb{P}(\mathcal{E}) = 1 + o(1)$ uniformly for all $p \in [0, p_\uparrow]$ and $\pi = \max(0, (1 - cn^7 p_\uparrow^{11})p^3)$.*

Notice that as opposed to Theorem 4.5, Corollary 4.16 relies on a very specific choice of π. We show Proposition 4.13 in Section 4.3.2, Proposition 4.14 in Section 4.3.3, Proposition 4.15 in Section 4.3.4, Corollary 4.16 in Section 4.3.5 and Theorem 4.5 in Section 4.3.7.

4.3.2 Proof of Proposition 4.13. We follow the definition of the coupling and start with the analysis of $(\tilde{\boldsymbol{n}}_{\mathrm{t}}, \tilde{\boldsymbol{n}}_{\mathrm{h}})$, a maximal coupling of $|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}|$ and $|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|$. Thus, the coupling lemma suggests that $\mathbb{P}(\tilde{\boldsymbol{n}}_{\mathrm{t}} \neq \tilde{\boldsymbol{n}}_{\mathrm{h}}) = \||\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}| - |\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|\|_{\mathrm{tv}}$. We bound the total variation distance using the limiting distributions of $|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}|$ and $|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|$. For this purpose let $\boldsymbol{n}_{\mathrm{t}} = \mathrm{Po}(\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}|])$ and $\boldsymbol{n}_{\mathrm{h}} = \mathrm{Po}(\mathbb{E}[|\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}|])$, then we have

$$\mathbb{P}(\tilde{\boldsymbol{n}}_{\mathrm{t}} \neq \tilde{\boldsymbol{n}}_{\mathrm{h}}) \leq \||\boldsymbol{\mathcal{C}}^\circ_{\mathrm{t}}| - \boldsymbol{n}_{\mathrm{t}}\|_{\mathrm{tv}} + \|\boldsymbol{n}_{\mathrm{t}} - \boldsymbol{n}_{\mathrm{h}}\|_{\mathrm{tv}} + \||\boldsymbol{\mathcal{C}}^\circ_{\mathrm{h}}| - \boldsymbol{n}_{\mathrm{h}}\|_{\mathrm{tv}}.$$

Our first result will be that $\boldsymbol{n}_{\mathrm{t}}$, $\boldsymbol{n}_{\mathrm{h}}$ are indeed the limiting distributions and that they asymptotically coincide for our choices of p and π.

Lemma 4.17. *Notice that the following holds.*
a) We have $\||\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|-\boldsymbol{n}_{\mathrm{t}}\|_{\mathrm{tv}}=o(1)$ *for* $p_{\uparrow}=o(n^{-1/2})$.
b) We have $\||\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|-\boldsymbol{n}_{\mathrm{h}}\|_{\mathrm{tv}}=o(1)$ *for* $\pi_{\uparrow}=o(n^{-3/2})$.
c) We have $\|\boldsymbol{n}_{\mathrm{t}}-\boldsymbol{n}_{\mathrm{h}}\|_{\mathrm{tv}}\le\min(1,\frac{1}{\sqrt{2\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|]}},\frac{1}{\sqrt{2\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|]}})|\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|]-\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|]|$.

Proof. For the first two parts we use Theorem 4.7 in [116]. For this purpose we notice that the cases $\pi=0$, $p=0$ are trivial, and otherwise the bound yields

$$\||\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|-\boldsymbol{n}_{\mathrm{h}}\|_{\mathrm{tv}}\le\min\left(1,P_{\mathrm{h},3}^{-1}\right)(\pi^3P_{\mathrm{h},3}+(1+\pi^2)P_{\mathrm{h},4}+(1+\pi)P_{\mathrm{h},5})=o(1),$$

using Equation (13) in the proof of Observation 4.11. The result for the triangle hypergraph follows with Equation (14). For the last part and $\pi\le p^3$, we couple $\boldsymbol{n}_{\mathrm{t}}=\boldsymbol{n}_{\mathrm{h}}+\boldsymbol{\Delta}$ using $(\boldsymbol{n}_{\mathrm{h}},\boldsymbol{\Delta})\sim\mathrm{Po}(\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|])\otimes\mathrm{Po}(\Delta)$ and $\Delta=|\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|]-\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|]|$, and the corresponding coupling for $\pi>p^3$ with the coupling lemma gives $\|\boldsymbol{n}_{\mathrm{t}}-\boldsymbol{n}_{\mathrm{h}}\|_{\mathrm{tv}}=\mathbb{P}(\boldsymbol{n}_{\mathrm{t}}\ne\boldsymbol{n}_{\mathrm{h}})=\mathbb{P}(\boldsymbol{\Delta}>0)=1-e^{-\Delta}\le\Delta$. For $\pi>0$, $p\in[0,1]$ and using $\ln(1+x)\le x$, we have

$$D_{\mathrm{KL}}(\boldsymbol{n}_{\mathrm{t}}\|\boldsymbol{n}_{\mathrm{h}})=\mathbb{E}[\boldsymbol{n}_{\mathrm{h}}]-\mathbb{E}[\boldsymbol{n}_{\mathrm{t}}]+\mathbb{E}[\boldsymbol{n}_{\mathrm{t}}]\ln\left(\frac{\mathbb{E}[\boldsymbol{n}_{\mathrm{t}}]}{\mathbb{E}[\boldsymbol{n}_{\mathrm{h}}]}\right)\le\frac{(\mathbb{E}[\boldsymbol{n}_{\mathrm{h}}]-\mathbb{E}[\boldsymbol{n}_{\mathrm{t}}])^2}{\mathbb{E}[\boldsymbol{n}_{\mathrm{h}}]},$$

so Pinsker's inequality yields $\|\boldsymbol{n}_{\mathrm{t}}-\boldsymbol{n}_{\mathrm{h}}\|_{\mathrm{tv}}\le\frac{\Delta}{\sqrt{2\mathbb{E}[\boldsymbol{n}_{\mathrm{h}}]}}$. □

Now, notice that $\mathbb{P}(\tilde{\boldsymbol{b}}_{\mathrm{t}}=\tilde{\boldsymbol{b}}_{\mathrm{h}}=1)=1+o(1)$ is an immediate consequence of Observation 4.11 since $\tilde{\boldsymbol{b}}_{\mathrm{t}}\sim\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ})$, $\tilde{\boldsymbol{b}}_{\mathrm{h}}\sim\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ})$, $p_{\uparrow}=o(n^{-7/11})$ and $\pi_{\uparrow}=o(n^{-11/6})$. Finally, a symmetry argument shows that $(\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|\,\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ})=1,|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|=N)\sim\mathrm{u}(\mathcal{S})$ and $(\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|\,\mathrm{vdj}(\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ})=1,|\boldsymbol{\mathcal{C}}_{\mathrm{h}}^{\circ}|=N)\sim\mathrm{u}(\mathcal{S})$, where $\mathcal{S}=\{\mathcal{C}\subseteq\overline{\mathcal{C}}^{\circ}:|\mathcal{C}|=N,\mathrm{vdj}(\mathcal{C})=1\}$.

4.3.3 Proof of Proposition 4.14. First, we briefly discuss the cases $p=0$, $p=1$ and $p\in(0,1)$. For $p=0$ we have $\mathcal{D}_{\mathrm{t}}=\{(E,\emptyset,\emptyset,\mathcal{N}):E\in\binom{[n]}{3},\mathcal{N}\subseteq\binom{[n]}{3}\}$ and $\pi_{\mathrm{t}}\equiv0$ since $\boldsymbol{H}_{\mathrm{t}}\equiv0$ almost surely. For $p=1$ we have $\boldsymbol{H}_{\mathrm{t}}\equiv1$ almost surely, hence $\mathcal{D}_{\mathrm{t}}=\{(E,\overline{\mathcal{C}}^{\circ},\mathcal{Y},\emptyset):E\in\binom{[n]}{3},\mathcal{Y}\subseteq\binom{[n]}{3}\}$ and $\pi_{\mathrm{t}}\equiv1$. For $p\in(0,1)$ and using $\mathcal{T}_{\mathrm{h}}=\{H_{\mathrm{t}}(G):G\in\{0,1\}^{\binom{[n]}{2}}\}$ we have

$$\mathcal{D}_{\mathrm{t}}=\left\{(E,\mathcal{C}_{\mathrm{h}}(H),\mathcal{Y},\mathcal{N}):E\in\binom{[n]}{3},H\in\mathcal{T}_{\mathrm{h}},\mathcal{Y}\subseteq H^{-1}(1),\mathcal{N}\subseteq H^{-1}(0)\right\}.\tag{16}$$

We start with Part 4.14a) and let $X=(E^*,\mathcal{C},\mathcal{Y},\mathcal{N})\in\mathcal{D}_{\mathrm{t}}$. Our first result will establish a refined lower bound. For this purpose, let $\overline{\mathcal{Y}}=\mathcal{Y}\cup\bigcup_{C\in\mathcal{C}}C$ be all included hyperedges, let $H_{\pm}\in\{0,1\}^{\binom{[n]}{3}}$ be given by $H_-^{-1}(1)=\overline{\mathcal{Y}}$, $H_+^{-1}(1)=\overline{\mathcal{Y}}\cup\{E^*\}$, and let $G_{\pm}=G_{\mathrm{t}}(H_{\pm})$ be the triangle graphs. For $E\in\binom{[n]}{3}$ let $\mathcal{M}_{\pm}(E)=\binom{E}{2}\setminus G_{\pm}^{-1}(1)$ be the missing edges, and let

$$\mathcal{N}_{\mathrm{e}}=\{E\in\mathcal{N}:\mathcal{M}_-(E^*)\cap\mathcal{M}_-(E)\ne\emptyset\},\,\mathcal{N}_{\mathrm{c}}=\left\{C\in\overline{\mathcal{C}}^{\circ}\setminus\mathcal{C}:\mathcal{M}_-(E^*)\cap\bigcup_{E\in C}\mathcal{M}_-(E)\ne\emptyset\right\}$$

be the excluded hyperedges and cycles that profit from the inclusion of E^*. For $E\in\mathcal{N}_{\mathrm{e}}$ let $m(E)=|\mathcal{M}_+(E)|$, for $C\in\mathcal{N}_{\mathrm{c}}$ let $m(C)=|\bigcup_{E\in C}\mathcal{M}_+(E)|$, and let $Q=\sum_{E\in\mathcal{N}_{\mathrm{e}}}p^{m(E)}+\sum_{C\in\mathcal{N}_{\mathrm{c}}}p^{m(C)}$.

Lemma 4.18. *We have* $\pi_{\mathrm{t}}(X)\ge(1-Q)p^3$ *for all* $p\in[0,1]$ *and* $X\in\mathcal{D}_{\mathrm{t}}$.

Proof. Let $\mathcal{Y}_e = \bigcup_{E \in \overline{\mathcal{Y}}} \binom{E}{2}$ be the edges, $\boldsymbol{G}' \sim \mathrm{Bin}(1,1)^{\otimes \mathcal{Y}_e} \otimes \mathrm{Bin}(1,p)^{\otimes \binom{[n]}{2} \setminus \mathcal{Y}_e}$ the graph $\boldsymbol{G}_b$ given that $\mathcal{Y}_e$ is included, and let $\boldsymbol{H}'_t = H_t(\boldsymbol{G}')$ be its triangle hypergraph. Then we have $\pi_t(X) = \mathbb{P}(\mathcal{U}|\mathcal{D}_\cap) = \frac{\mathbb{P}(\mathcal{U} \cap \mathcal{D}_\cap)}{\mathbb{P}(\mathcal{D}_\cap)}$, with $\mathcal{U} = \{\boldsymbol{H}'_t(E^*) = 1\}$,

$$\mathcal{D}_\cap = \left\{\mathcal{C}^\circ_h(\boldsymbol{H}') \subseteq \mathcal{C}, \mathcal{N} \subseteq \boldsymbol{H}'^{-1}_t(0)\right\} = \left\{\mathcal{C}^\circ_h(\boldsymbol{H}') \cap (\overline{\mathcal{C}}^\circ \setminus \mathcal{C}) = \emptyset, \boldsymbol{H}'^{-1}_t(1) \cap \mathcal{N} = \emptyset\right\}.$$

Now, let $\mathcal{N}^c_e = \mathcal{N} \setminus \mathcal{N}_e$ and $\mathcal{N}^c_c = (\overline{\mathcal{C}}^\circ \setminus \mathcal{C}) \setminus \mathcal{N}_c$. Then we have $\mathcal{D}_\cap = \mathcal{D}_0 \cap \mathcal{D}_1$ with the decomposition

$$\mathcal{D}_0 = \left\{\mathcal{C}^\circ_h(\boldsymbol{H}') \cap \mathcal{N}^c_c = \emptyset, \boldsymbol{H}'^{-1}_t(1) \cap \mathcal{N}^c_e = \emptyset\right\}, \mathcal{D}_1 = \left\{\mathcal{C}^\circ_h(\boldsymbol{H}') \cap \mathcal{N}_c = \emptyset, \boldsymbol{H}'^{-1}_t(1) \cap \mathcal{N}_e = \emptyset\right\}.$$

This gives $\pi_t(X) \geq \frac{\mathbb{P}(\mathcal{U} \cap \mathcal{D}_\cap)}{\mathbb{P}(\mathcal{D}_0)} = \mathbb{P}(\mathcal{U} \cap \mathcal{D}_1|\mathcal{D}_0) = \mathbb{P}(\mathcal{U}) - \mathbb{P}(\mathcal{U} \cap \mathcal{D}^c_1|\mathcal{D}_0)$, using that $\mathcal{U}$ is independent of $\mathcal{D}_0$, and $\mathcal{E}^c$ to denote the complement of an event $\mathcal{E}$. Since $\mathcal{U} \cap \mathcal{D}^c_1$ is an up-set and $\mathcal{D}_0$ is a down-set, the Harris inequality gives $\pi_t(X) \geq \mathbb{P}(\mathcal{U}) - \mathbb{P}(\mathcal{U} \cap \mathcal{D}^c_1) = (1 - \mathbb{P}(\mathcal{D}^c_1|\mathcal{U}))\mathbb{P}(\mathcal{U})$ for $p > 0$, which we may assume. With $\mathbb{P}(\mathcal{U}) = p^{|\mathcal{M}_-(E^*)|} \geq p^3$ we have $\pi_t(X) \geq (1 - \mathbb{P}(\mathcal{D}^c_1|\mathcal{U}))p^3$. The union bound with $\mathcal{Y}_+ = \mathcal{Y}_e \cup \binom{E^*}{2}$, $\boldsymbol{G}'' \sim \mathrm{Bin}(1,1)^{\otimes \mathcal{Y}_+} \otimes \mathrm{Bin}(1,p)^{\otimes \binom{[n]}{2} \setminus \mathcal{Y}_+}$ and $\boldsymbol{H}'' = H_t(\boldsymbol{G}'')$ yields

$$\mathbb{P}(\mathcal{D}^c_1|\mathcal{U}) = \mathbb{P}\left((\mathcal{C}^\circ_h(\boldsymbol{H}'') \cap \mathcal{N}_c) \cup (\boldsymbol{H}''^{-1}(1) \cap \mathcal{N}_e) \neq \emptyset\right) \leq Q,$$

since $\mathbb{P}(C \in \mathcal{C}^\circ_h(\boldsymbol{H}'')) = p^{m(C)}$ for $C \in \mathcal{N}_c$ and $\mathbb{P}(\boldsymbol{H}''(E) = 1) = p^{m(E)}$ for $E \in \mathcal{N}_e$. Substituting this bound above yields $\pi_t(X) \geq (1 - Q)p^3$ and thereby completes the proof. □

Lemma 4.18 yields a lower bound which is valid in any case. Next, we turn to the cases for which we have $\pi_t(X) = 0$. Let $\mathcal{N}^\circ_e = \{E \in \mathcal{N}_e : m(E) = 0\}$ and $\mathcal{N}^\circ_c = \{C \in \mathcal{N}_c : m(C) = 0\}$.

Lemma 4.19. *For all $p \in [0,1]$ and $X \in \mathcal{D}_t$ the following holds. We have $\pi_t(X) = 0$ if and only if one of the following holds.*
- *We have $p = 0$.*
- *We have $\mathcal{N}^\circ_e \neq \emptyset$.*
- *We have $\mathcal{N}^\circ_c \neq \emptyset$.*

Proof. We may assume $p \in (0,1)$. Assume that $\pi_t(X) > 0$, i.e. there exists a graph G, $H = H_t(G)$, with $H(E^*) = 1$ that induces X in $\mathcal{D}_t$ as given by Equation (16). This gives $G_+ \leq G$. Thus, for $E \in \mathcal{N}_e$ we have $m(E) > 0$, and $m(C) > 0$ for all $C \in \mathcal{N}_c$. This shows that $\pi_t(X) = 0$ if one of the three conditions holds. For the other direction, let X be such that $\mathcal{N}^\circ_e = \emptyset$ and $\mathcal{N}^\circ_c = \emptyset$. Let G, $H = H_{t,G}$ be a graph that induces X, then we have $G_- \leq G$, $H'_- = H_{t,G_-} \leq H$, and G_- also induces X. So, for $E \in \mathcal{N}$ we have $\mathcal{M}_-(E) \neq \emptyset$ since $H'_-(E) = 0$. For $E \notin \mathcal{N}_e$ we have $m(E) = |\mathcal{M}_-(E)| > 0$ since $G^{-1}_+(1) \setminus G^{-1}_-(1) \subseteq \binom{E^*}{2}$, and thereby $H'_+(E) = 0$, where $H'_+ = H_{t,G_+}$. For $E \in \mathcal{N}_e$ we have $H'_+(E) = 0$ because $\mathcal{N}^\circ_e = \emptyset$, and thereby $\mathcal{N} \subseteq H'^{-1}_+(0)$. Similarly, let $C \in \overline{\mathcal{C}}^\circ \setminus \mathcal{C}$. For $C \notin \mathcal{N}_c$ we have $C \notin \mathcal{C}^\circ_h(H'_+)$ because the edges missing in C under G_- are also missing under G_+. For $C \in \mathcal{N}_c$ we have $C \notin \mathcal{C}^\circ_h(H'_+)$ because there are still edges missing since $\mathcal{N}^\circ_c = \emptyset$. This yields $\mathcal{C}^\circ_h(H'_+) \subseteq \mathcal{C}$. Clearly, we have $\overline{\mathcal{Y}} \cup \{E^*\} \subseteq H'^{-1}_+(1)$ and thereby G_+ induces X and covers $H'_+(E^*) = 1$, so $\pi_t(X) > 0$. □

Intuitively, Lemma 4.19 just states that $\pi_t(X) = 0$ if an excluded hyperedge or clean 3-cycle is covered by including E^*, thus violating the condition. This shows that we have $\mathcal{N}^\circ_e, \mathcal{N}^\circ_c = \emptyset$ whenever $\pi_t(X) > 0$. Hence, combining this with the next result establishes Proposition 4.14a).

Lemma 4.20. *For all $p \in [0,1]$, $X \in \mathcal{D}_\mathrm{t}$ and an absolute constant $c \in \mathbb{R}_{>0}$ we have*

$$0 \le Q \le |\mathcal{N}_\mathrm{e}^\circ| + |\mathcal{N}_\mathrm{c}^\circ| + c\sum_{m=1}^{8} \Delta_\mathrm{g}(G_+)^{4-v_m} n^{v_m} p^m, \quad v_m = \left\lfloor \frac{m}{2} \right\rfloor.$$

Proof. We consider the split $\mathcal{N}_\mathrm{e} = \bigcup_m \mathcal{N}_\mathrm{m}(m)$, $\mathcal{N}_\mathrm{m}(m) = \{E \in \mathcal{N}_\mathrm{e} : m(E) = m\}$, according to the exponent, notice that $\mathcal{N}_\mathrm{m}(m) = \emptyset$ unless $m \in \mathbb{Z} \cap [0,2]$ and that $\mathcal{N}_\mathrm{m}(0) = \mathcal{N}_\mathrm{e}^\circ$. Hence, let $m \in [2]$ and notice that $\mathcal{N}_\mathrm{m}(m) = \{E \in \mathcal{N}_\mathrm{m}^*(m) : E \in \mathcal{N}\}$, where

$$\mathcal{N}_\mathrm{m}^*(m) = \left\{ E \in \binom{[n]}{3} : E \cap E^* \in G_+^{-1}(1), \left| \binom{E}{2} \setminus G_+^{-1}(1) \right| = m \right\}.$$

We obtain an upper bound on $|\mathcal{N}_\mathrm{m}^*(1)|$ by choosing two distinct vertices $u, v \in E^*$, yielding a factor 6, and a neighbor w of u, yielding a factor $\Delta = \Delta_\mathrm{g}(G_+)$, to obtain $E^* = \{u, v, w\}$. This gives the bound $|\mathcal{N}_\mathrm{m}^*(1)| \le 6\Delta$. We obtain $|\mathcal{N}_\mathrm{m}^*(2)| \le 3n$ by choosing a pair in $\binom{E^*}{2}$ and a third vertex. This gives the bound $\sum_{E \in \mathcal{N}_\mathrm{e}} p^{m(E)} \le |\mathcal{N}_\mathrm{e}| + 6\Delta p + 3np^2$ for the first contribution to Q.

Let $\mathcal{N}_\mathrm{c} = \bigcup_m \mathcal{N}_\mathrm{m}(m)$, $\mathcal{N}_\mathrm{m}(m) = \{C \in \mathcal{N}_\mathrm{c} : m(C) = m\}$ and notice that $\mathcal{N}_\mathrm{m}(m) = \emptyset$ unless $m \in \mathbb{Z} \cap [0,8]$. We proceed as above, and notice that we now always have at least two vertices in E^*. For $m = 8$ we choose the remaining 4 vertices freely, say $|\mathcal{N}_\mathrm{m}(8)| \le 3\binom{6}{3}^3 n^4$. For $m = 7$, we take two vertices in E^*, choose 3 vertices freely and a neighbor to one of the chosen vertices to obtain $|\mathcal{N}_\mathrm{m}(7)| \le 3 \cdot 5 \cdot \binom{6}{3}^3 \Delta n^3$. For $m = 6$ we can close a triangle, thus we may still have 3 free vertices. However, in this case the choice of the third edge is determined, which gives $|\mathcal{N}_\mathrm{m}(6)| \le 3 \cdot 4 \cdot 5\binom{6}{3}^3 \Delta^2 n^2 + 3 \cdot 2\binom{6}{3}^3 \Delta n^3 \le c\Delta n^3$ using $\Delta \le n$ and the implied constant c. Thus, we only need the maximum number of isolated vertices in a subgraph of a clean 3-cycle with $9 - m$ edges. This gives $|\mathcal{N}_\mathrm{m}(m)| \le c\Delta^{4-v_m} n^{v_m}$ with $v_m = \lfloor \frac{1}{2} m \rfloor$ and for some large enough absolute constant c. Using $2 \le \Delta$ we combine the bounds above to obtain an absolute constant $c \in \mathbb{R}_{>0}$ and

$$Q \le |\mathcal{N}_\mathrm{e}^\circ| + |\mathcal{N}_\mathrm{c}^\circ| + c\sum_{m=1}^{8} \Delta^{4-v_m} n^{v_m} p^m.$$

□

We summarize the last three results in a corollary before restricting the choices of p and X.

Corollary 4.21. *For $p \in [0,1]$, $X \in \mathcal{D}_\mathrm{t}$ and an absolute constant $c \in \mathbb{R}_{>0}$ we have*

$$\pi_\mathrm{t}(X) \ge \left(1 - |\mathcal{N}_\mathrm{e}^\circ| - |\mathcal{N}_\mathrm{c}^\circ| - c\sum_{m=1}^{8} \Delta_\mathrm{c}(X)^{4-v_m} n^{v_m} p^m \right) p^3, \quad v_m = \left\lfloor \frac{m}{2} \right\rfloor.$$

Proof. The assertion follows from Lemma 4.18, Lemma 4.20 and $\Delta_\mathrm{g}(G_+) \le 2\Delta_\mathrm{h}(H_+) = 2\Delta_\mathrm{c}(X)$. □

We turn to the proof of Proposition 4.14a). Hence, let $p_\uparrow = \Omega(\ln(n)^{1/3}/n^{2/3})$, $p_\uparrow = o(n^{-7/11})$, further $p \le p_\uparrow$ and $X \in \mathcal{D}_\mathrm{t}^\circ$ with $\pi_\mathrm{t}(X) > 0$. Let $n_\circ$ be sufficiently large, $c_\circ$ sufficiently small and in particular such that $p_\uparrow(n) \ge c_\circ \frac{\ln(n)^{1/3}}{n^{2/3}}$ for $n \ge n_\circ$, yielding

$$\Delta_\mathrm{c}(X) \le \binom{n-1}{2} p_\uparrow^3 + \max\left(\binom{n-1}{2} p_\uparrow^3, 4\ln(n) \right) \le c' n^2 p_\uparrow^3,\ c' = \frac{1}{2} + \frac{4}{c_\circ^3},$$

since $X \in \mathcal{D}_{\mathrm{t}}^{\circ}$. With $\tilde{c} \in \mathbb{R}_{>0}$ from Corollary 4.21 and Lemma 4.19 we obtain $\pi_{\mathrm{t}}(X) \geq (1-\varepsilon)p^3$, where $\varepsilon(n) = \tilde{c}c'^4 \sum_{m=1}^{8} (n^2 p_{\uparrow}^3)^{4-v_m} n^{v_m} p_{\uparrow}^m \leq 2\tilde{c}c'^4 (n^8 p_{\uparrow}^{13} + \sum_{a=4}^{7} n^a p_{\uparrow}^{a+4})$, since the contribution for $m = 2\ell + 1$, $\ell \in [3]$, is bounded by the contribution for 2ℓ. This further yields $\varepsilon \leq 10\tilde{c}c'^4 n^7 p_{\uparrow}^{11}$. Choosing $c > 10\tilde{c}c'^4$ sufficiently large such that $cn^7 p_{\uparrow}^{11} > 1$ for $n \leq n_\circ$ completes the proof.

We turn to Part 4.14b) and Part 4.14c), which are immediate from the following, general result.

Lemma 4.22. *Let $p \in [0,1]$ and $X \in \mathcal{D}_{\mathrm{t}}$. Then the following holds.*
a) We have $\pi_{\mathrm{t}}(X) = 0$ if $p = 0$, $E^* \in \mathcal{N}$ or $C \subseteq \overline{\mathcal{Y}} \cup \{E^*\}$ for some $C \in \overline{\mathcal{C}}^{\circ} \setminus \mathcal{C}$.
b) Assume that $\pi_{\mathrm{t}}(X) = 0$, $\mathrm{vdj}(\mathcal{C}) = 1$, $p > 0$, $E^* \notin \mathcal{N}$, and $C \not\subseteq \overline{\mathcal{Y}} \cup \{E^*\}$ for all $C \in \overline{\mathcal{C}}^{\circ} \setminus \mathcal{C}$. Then we have $\mathcal{C}_{\mathrm{h}}(H_+) \setminus \mathcal{C}_{\mathrm{h}}^{\circ}(H_+) \neq \emptyset$ or $\mathcal{C}_{\mathrm{i}}(H_+) \neq \emptyset$.

Proof. Using Lemma 4.19 we can restrict to $p \in (0,1)$. The remaining two items in the first part yield $\pi_{\mathrm{t}}(X) = 0$, since these directly violate the given event in π_{t}. For the second part, we have $\mathcal{C}_{\mathrm{h}}^{\circ}(H_+) = \mathcal{C}$. Using Lemma 4.19, we have $\mathcal{N}_{\mathrm{e}}^{\circ} \neq \emptyset$ or $\mathcal{N}_{\mathrm{c}}^{\circ} \neq \emptyset$. As in the proof of Lemma 4.19, for any G inducing X in $\mathcal{D}_{\mathrm{t}}$ we have $G_- \leq G$, and G_- induces X. Also, for any G, $H = H_{\mathrm{t}}(G)$, with $H(E^*) = 1$ inducing X in $\mathcal{D}_{\mathrm{t}}$ we have $G_+ \leq G$, and G_+ induces X. Let $H'_{\pm} = H_{\mathrm{t},G_{\pm}}$ be the closures. Now, assume that $\mathcal{N}_{\mathrm{e}}^{\circ} \neq \emptyset$ and let $E \in \mathcal{N}_{\mathrm{e}}^{\circ}$. Then $E \in \mathcal{N}$ has to be induced by some $C \in \mathcal{C}_{\mathrm{h}}(H_+)$ because $H_-(E) = H'_-(E) = H_+(E) = 0$ and $H'_+(E) = 1$. This shows that $\mathcal{C}_{\mathrm{h}}(H_+) \setminus \mathcal{C}_{\mathrm{h}}^{\circ}(H_+) \neq \emptyset$.

Otherwise, we have $\mathcal{N}_{\mathrm{c}}^{\circ} \neq \emptyset$, so let $C \in \mathcal{N}_{\mathrm{c}}^{\circ}$ and assume that also $\mathcal{C}_{\mathrm{h}}(H_+) \setminus \mathcal{C}_{\mathrm{h}}^{\circ}(H_+) = \emptyset$, meaning that $\mathcal{C}_{\mathrm{h}}(H_+) = \mathcal{C}$. Now, we have $C \notin \mathcal{C}$, $C \not\subseteq \overline{\mathcal{Y}} \cup \{E^*\}$, but we do have $C \in \mathcal{C}_{\mathrm{h}}^{\circ}(H'_+)$. Hence, there exists $E \in C$ with $H'_-(E) = 0$. Assume that we have $H_+(E) = 0$, then there exists $C' \in \mathcal{C}_{\mathrm{h}}(H_+) = \mathcal{C}$ that induces E, but $C' \in \mathcal{C} = \mathcal{C}_{\mathrm{h}}^{\circ}(H'_-)$ then suggests that $H'_-(E) = 1$, a contradiction. Hence, we have $H_+(E) = 1$ and thereby $E = E^*$. This shows that $E^* \in C$ and that $H'_-(E) = 1$ for $E \in C \setminus \{E^*\}$. This gives $E^* \notin \overline{\mathcal{Y}}$ and $H_+(E) = 0$ for some $E \in C$. But due to $H'_+(E) = 1$ there exists $C' \in \mathcal{C}_{\mathrm{h}}(H_+) = \mathcal{C}$ with $E = E^{\circ}(C')$. For the remaining hyperedge E', i.e. $C = \{E^*, E, E'\}$, we cannot have $H_+(E') = 0$ since $\mathrm{vdj}(\mathcal{C}) = 1$, so we have $E' \in \overline{\mathcal{Y}}$. This gives $C' \cup \{E^*, E'\} \in \mathcal{C}_{\mathrm{i}}(H_+)$. ☐

4.3.4 Proof of Proposition 4.15. For the binomial hypergraph, we start with the discussion of the support for $\pi = 0$, $\pi = 1$ and $\pi \in (0,1)$. For $\pi = 0$ we have $\mathcal{D}_{\mathrm{h}} = \{(E, \emptyset, \emptyset, \mathcal{N}) : E \in \binom{[n]}{3}, \mathcal{N} \subseteq \binom{[n]}{3}\}$ and $\pi_{\mathrm{h}} \equiv 0$ since $\boldsymbol{H}_{\mathrm{b}} \equiv 0$ almost surely. For $\pi = 1$ we have $\mathcal{D}_{\mathrm{h}} = \{(E, \overline{\mathcal{C}}^{\circ}, \mathcal{Y}, \emptyset) : E \in \binom{[n]}{3}, \mathcal{Y} \subseteq \binom{[n]}{3}\}$ and $\pi_{\mathrm{h}} \equiv 1$ since $\boldsymbol{H}_{\mathrm{b}} \equiv 1$ almost surely. For $\pi \in (0,1)$ we have

$$\mathcal{D}_{\mathrm{h}} = \left\{(E, \mathcal{C}_{\mathrm{h}}(H), \mathcal{Y}, \mathcal{N}) : E \in \binom{[n]}{3}, H \in \{0,1\}^{\binom{[n]}{3}}, \mathcal{Y} \subseteq H^{-1}(1), \mathcal{N} \subseteq H^{-1}(0)\right\}. \qquad (17)$$

We turn to Part 4.15a) and let $X = (E^*, \mathcal{C}, \mathcal{Y}, \mathcal{N}) \in \mathcal{D}_{\mathrm{h}}$. As before, let $\overline{\mathcal{Y}} = \mathcal{Y} \cup \bigcup_{C \in \mathcal{C}} C$ and let $H_{\pm} \in \{0,1\}^{\binom{[n]}{3}}$ be given by $H_-^{-1}(1) = \overline{\mathcal{Y}}$, $H_+^{-1}(1) = \overline{\mathcal{Y}} \cup \{E^*\}$.

Lemma 4.23. *For all $\pi \in [0,1]$ and $X \in \mathcal{D}_{\mathrm{h}}$ with $E^* \notin \overline{\mathcal{Y}}$ we have $\pi_{\mathrm{h}}(X) \leq \pi$.*

Proof. We may restrict to $\pi \in (0,1)$. Consider $\boldsymbol{H}' \sim \mathrm{Bin}(1,1)^{\otimes \overline{\mathcal{Y}}} \otimes \mathrm{Bin}(1,\pi)^{\otimes \binom{[n]}{3} \setminus \overline{\mathcal{Y}}}$ and notice that $\pi_{\mathrm{h}}(X_{\mathrm{h}}) = \mathbb{P}(\boldsymbol{H}'(E^*) = 1 | \mathcal{C}_{\mathrm{h}}^{\circ}(\boldsymbol{H}') \subseteq \mathcal{C}, \mathcal{N} \subseteq \boldsymbol{H}'^{-1}(0)) \leq \mathbb{P}(\boldsymbol{H}'(E^*) = 1)$ by the Harris inequality. ☐

Before we establish the lower bound, we discuss when the conditional probability vanishes.

Lemma 4.24. *Let $\pi \in [0,1]$ and $X \in \mathcal{D}_{\mathrm{h}}$. We have $\pi_{\mathrm{h}}(X) = 0$ if and only if $\pi = 0$, $E^* \in \mathcal{N}$ or $C \subseteq \overline{\mathcal{Y}} \cup \{E^*\}$ for some $C \in \overline{\mathcal{C}}^{\circ} \setminus \mathcal{C}$.*

Proof. Since $\pi_{\mathrm{h}}(X)$ is well-defined, the definition of conditional probability gives $\pi_{\mathrm{h}}(X)=0$ if and only if $\mathbb{P}(\boldsymbol{H}_{\mathrm{b}}(E^*)=1,\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}\subseteq\mathcal{C},\overline{\mathcal{Y}}\subseteq\boldsymbol{H}_{\mathrm{b}}^{-1}(1),\mathcal{N}\subseteq\boldsymbol{H}_{\mathrm{b}}^{-1}(0))=0$. Hence, we have $\pi_{\mathrm{h}}(X)=0$ if $\pi=0$, $E^*\in\mathcal{N}$ or $C\subseteq\overline{\mathcal{Y}}\cup\{E^*\}$ for some $C\in\overline{\mathcal{C}}^{\circ}\setminus\mathcal{C}$. Conversely, for $\pi>0$, $E^*\notin\mathcal{N}$ and $C\not\subseteq\overline{\mathcal{Y}}\cup\{E^*\}$ for all $C\in\overline{\mathcal{C}}^{\circ}\setminus\mathcal{C}$, we have $\mathcal{C}^{\circ}_{\mathrm{h}}(H_+)\subseteq\mathcal{C}$ as well as $\mathcal{N}\subseteq H_+^{-1}(0)$, thus $\pi_{\mathrm{h}}(X)>0$. □

This establishes Part 4.15c). Now, we turn to the remaining Part 4.15b).

Lemma 4.25. *For $\pi\in[0,1]$ and $X\in\mathcal{D}_{\mathrm{h}}$ such that $\pi_{\mathrm{h}}(X)>0$ we have*

$$\pi_{\mathrm{h}}(X)\geq(1-12(\Delta_{\mathrm{c}}(X)n\pi+n^3\pi^2))\pi.$$

Proof. Using Lemma 4.24 we may assume that $\pi>0$, $E^*\notin\mathcal{N}$ and $\mathcal{C}^{\circ}_{\mathrm{h}}(H_+)=\mathcal{C}$. We may further assume that $\pi_{\mathrm{h}}(X)<1$ and in particular that $\pi<1$, $E^*\notin\overline{\mathcal{Y}}$. Using the two shorthands $\mathcal{E}_-=\{\overline{\mathcal{Y}}\subseteq\boldsymbol{H}_{\mathrm{b}}^{-1}(1),\mathcal{N}\subseteq\boldsymbol{H}_{\mathrm{b}}^{-1}(0)\}$ and $\mathcal{E}_+=\mathcal{E}_-\cap\{\boldsymbol{H}_{\mathrm{b}}(E^*)=1\}$ we have $\pi_{\mathrm{h}}(X)=\frac{\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}\subseteq\mathcal{C}|\mathcal{E}_+)}{\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}\subseteq\mathcal{C}|\mathcal{E}_-)}\pi$. Let $\overline{\mathcal{C}}^{\circ}_1=\{C\in\overline{\mathcal{C}}^{\circ}\setminus\mathcal{C}:E^*\in C\}$, $\overline{\mathcal{C}}^{\circ}_0=(\overline{\mathcal{C}}^{\circ}\setminus\mathcal{C})\setminus\overline{\mathcal{C}}^{\circ}_1$ and $\boldsymbol{\mathcal{C}}^{\circ}_s=\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}\cap\overline{\mathcal{C}}^{\circ}_s$ for $s\in\{0,1\}$. Then we have $\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}\subseteq\mathcal{C}|\mathcal{E}_\pm)=\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_1=\emptyset|\mathcal{E}'_\pm)\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_0=\emptyset|\mathcal{E}_\pm)$, where $\mathcal{E}'_\pm=\{\boldsymbol{\mathcal{C}}^{\circ}_0=\emptyset\}\cap\mathcal{E}_\pm$. But since $\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_0=\emptyset|\mathcal{E}_\pm)=\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_0=\emptyset|\mathcal{E}_-)$, the probabilities in the numerator and denominator cancel out, yielding $\pi_{\mathrm{h}}(X)=\frac{\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_1=\emptyset|\mathcal{E}'_+)}{\mathbb{P}(\boldsymbol{\mathcal{C}}^{\circ}_1=\emptyset|\mathcal{E}'_-)}\pi\geq(1-\sum_{C\in\overline{\mathcal{C}}^{\circ}_1}\mathbb{P}(C\in\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}|\mathcal{E}'_+))\pi$. By assumption we have $C\not\subseteq\overline{\mathcal{Y}}\cup\{E^*\}$ for all $C\in\overline{\mathcal{C}}^{\circ}_1$ and hence $\overline{\mathcal{C}}^{\circ}_1=\mathcal{X}_0\cup\mathcal{X}_1$ with $\mathcal{X}_s=\{C\in\overline{\mathcal{C}}^{\circ}_1:|C\cap\overline{\mathcal{Y}}|=s\}$. In order to obtain $C=\{E^*,Y,E\}\in\mathcal{X}_1$ with $Y\in\overline{\mathcal{Y}}$, we first choose $E^*\cap Y=\{u\}$, then one of the remaining neighbors of u in H_+ to obtain Y, one of the remaining vertices $E^*\cap E=\{v\}$ in $E^*\setminus\{u\}$ and $Y\cap E=\{w\}$ in $Y\setminus\{u\}$, and finally any vertex to complete E, yielding $|\mathcal{X}_1|\leq 3(\Delta_{\mathrm{h}}(H_+)-1)4n\leq 12\Delta_{\mathrm{c}}(X)n$. Looking at the auxiliary hypergraph with $\overline{\mathcal{Y}}\cup\{E^*\}$ present and applying the Harris inequality yields $\mathbb{P}(C\in\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}|\mathcal{E}'_+)\leq\pi$. For $C\in\mathcal{X}_0$ we choose three vertices freely, yielding $|\mathcal{X}_0|\leq 12n^3$, and on the other hand $\mathbb{P}(C\in\boldsymbol{\mathcal{C}}^{\circ}_{\mathrm{h}}|\mathcal{E}'_+)\leq\pi^2$ since none of the two remaining hyperedges is known to be present. □

Now, Part 4.15b) is immediate.

4.3.5 Proof of Corollary 4.16. Though cumbersome, it is straightforward that the joint distribution $(\tilde{\boldsymbol{G}}_{\mathrm{b}},\tilde{\boldsymbol{H}}_{\mathrm{b}})$ defined in Figure 3 is well-defined and indeed a coupling of $\boldsymbol{G}_{\mathrm{b}}$ and $\boldsymbol{H}_{\mathrm{b}}$, for any choice of $p,\pi\in[0,1]$. In the following, we thoroughly discuss the ideas underlying the coupling and thus establish stronger results, which immediately imply Corollary 4.16.

For $p_\uparrow=\Omega(\ln(n)^{1/3}/n^{2/3})$, $p_\uparrow=o(n^{-20/31})$, further $\pi_\uparrow=p_\uparrow^3$, let c be from Proposition 4.14a), and let $n_\circ\in\mathbb{Z}_{\geq 12}$ be the minimum n such that $p_\uparrow(n)\in(0,1)$ and $cn^7p_\uparrow(n)^{11}<1$ for all $n\geq n_\circ$. Aiming towards Observation 4.8, Observation 4.11 and Observation 4.12, we consider the following typical hypergraphs. Let $\overline{D}=\binom{n-1}{2}p_\uparrow^3$, $\Delta=\overline{D}+\max\left(\overline{D},4\ln(n)\right)$ and

$$\mathcal{T}=\left\{H\in\{0,1\}^{\binom{[n]}{3}}:\mathcal{C}_{\mathrm{h}}(H)=\mathcal{C}^{\circ}_{\mathrm{h}}(H),\,\mathrm{vdj}(\mathcal{D}_{\mathrm{h}}(H)\cup\mathcal{C}^{\circ}_{\mathrm{h}}(H))=1,\mathcal{C}_{\mathrm{i}}(H)=\emptyset,\,\Delta_{\mathrm{h}}(H)<\Delta\right\}.\quad(18)$$

Thus, we restrict the maximal degree, we enforce that all 3-cycles are clean and no clean 3-cycles can be induced, we further require that the clean 3-cycles are vertex disjoint, which we upgrade to $\mathrm{vdj}(\mathcal{D}_{\mathrm{h}}(H)\cup\mathcal{C}^{\circ}_{\mathrm{h}}(H))=1$ for convenience. The upgrade only ensures that $H(E_{\mathrm{ci}}(C))=0$ for all $C\in\mathcal{C}^{\circ}_{\mathrm{h}}(H)$, i.e. that the induced hyperedges $E_{\mathrm{ci}}(C)$ from Section 4.2.5 are not present in H.

Let $\mathcal{E}_1=\{\tilde{\boldsymbol{\mathcal{C}}}^{\circ}_{\mathrm{t}}=\tilde{\boldsymbol{\mathcal{C}}}^{\circ}_{\mathrm{h}},\tilde{\boldsymbol{b}}_{\mathrm{h}}=1\}$ be the event that the first step is successful, and recall the event $\mathcal{E}$ from Equation (15) that both steps are successful. Further, let $\mathcal{E}_{\mathrm{h}}(H)=\{\tilde{\boldsymbol{\mathcal{C}}}^{\circ}_{\mathrm{t}}=\mathcal{C}^{\circ}_{\mathrm{h}}(H),\tilde{\boldsymbol{H}}_{\mathrm{b}}=H\}$ for $H\in\mathcal{T}$ and $\mathcal{E}_{\mathrm{t}}=\{\tilde{\boldsymbol{\mathcal{C}}}^{\circ}_{\mathrm{t}}=\tilde{\boldsymbol{\mathcal{C}}}^{\circ}_{\mathrm{h}},\tilde{\boldsymbol{H}}_{\mathrm{b}}\in\mathcal{T}\}=\bigcup_{H\in\mathcal{T}}\mathcal{E}_{\mathrm{h}}(H)$.

Let $M = \binom{n}{3}$ be the number of steps. For a typical hypergraph $H \in \mathcal{T}$ and using $\mathcal{C} = \mathcal{C}_{\mathrm{h}}^{\circ}(H)$, let $E^*_{\mathrm{h},H} = E^*_{\mathcal{C}}$ be the coupling order. For $i \in [M] \cup \{0\}$ let $\mathcal{Y}_{\mathrm{h},H,i} = \{E^*_{\mathrm{h},j} \in H^{-1}(1) : j \in [i]\}$ and $\mathcal{N}_{\mathrm{h},H,i} = \{E^*_{\mathrm{h},j} \in H^{-1}(0) : j \in [i]\}$. Let $\mathcal{I}_{\mathrm{c},H} = [3|\mathcal{C}|]$ be the first steps dedicated to $\mathcal{C}$, and $\mathcal{I}_{\mathrm{i},H} = [4|\mathcal{C}|] \setminus [3|\mathcal{C}|]$ the steps dedicated to the induced hyperedges. We denote the trivial steps by $\mathcal{I}_{\mathrm{t},H} = \{i \in [M] \setminus [4|\mathcal{C}|] : \exists C \in \overline{\mathcal{C}}^{\circ} \setminus \mathcal{C}\, C \subseteq \mathcal{Y}_{\mathrm{h},i-1} \cup \{E^*_{\mathrm{h},i}\}\}$. Further, the inclusion steps are given by $\mathcal{I}_{1,H} = \{i \in [M] \setminus [4|\mathcal{C}|] : E^*_{\mathrm{h},i} \in H^{-1}(1)\}$. The remaining steps are given by $\mathcal{I}_{*,H} = \{i \in [M] \setminus [4|\mathcal{C}|] : E^*_{\mathrm{h},i} \in H^{-1}(0)\} \setminus \mathcal{I}_{\mathrm{t}}$. For $i \in [M] \cup \{0\}$ let

$$\mathcal{X}_{\mathrm{h},H,i} = \{(\mathcal{Y}_{\mathrm{h},i}, \mathcal{Y}_{\mathrm{h},i}, \{E^*_{\mathrm{h},j} : j \in \mathcal{I} \cap [i]\}, \mathcal{N}_{\mathrm{h},i}) : \mathcal{I}_{\mathrm{t}} \subseteq \mathcal{I} \subseteq \mathcal{I}_{\mathrm{t}} \cup \mathcal{I}_*\}.$$

On $\mathcal{E}_1$ let $\boldsymbol{E}^* = E^*_{\tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ}}$, let $\boldsymbol{\pi}_i = \pi_{\mathrm{t}}(\boldsymbol{E}^*_i, \tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{t}}^{\circ}, \boldsymbol{\mathcal{Y}}_{\mathrm{t},i-1}, \boldsymbol{\mathcal{N}}_{\mathrm{t},i-1})$, and let $\boldsymbol{\pi}'_i = \pi_{\mathrm{h}}(\boldsymbol{E}^*_i, \tilde{\boldsymbol{\mathcal{C}}}_{\mathrm{h}}^{\circ}, \boldsymbol{\mathcal{Y}}_{\mathrm{h},i-1}, \boldsymbol{\mathcal{N}}_{\mathrm{h},i-1})$ for $i \in [M] \cup \{0\}$ be the conditional inclusion probabilities. Further, for $i \in [M]$ let $\boldsymbol{b}_i$ be the variable in Step 2.2.3, $\boldsymbol{b}'_i$ the variable in Step 2.2.3.2, and $\boldsymbol{m}_i$ the variable in Step 2.2.4.

Lemma 4.26. *Let $p_\uparrow = \Omega(\ln(n)^{1/3}/n^{2/3})$, $p_\uparrow = o(n^{-20/31})$, $n \geq n_\circ$, $H \in \mathcal{T}$ and $p \in [0, p_\uparrow]$.*
a) We have $\pi = (1 - cn^7 p_\uparrow^{11})p^3$. Further, we have $\pi = 0$ if and only if $p = 0$.
b) We have $\mathbb{P}(\mathcal{E}_1) \geq 1 - \min(\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|], \sqrt{\frac{1}{2}\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|]})3cn^7 p_\uparrow^{11} + o(1)$ and thus $\mathcal{E}_1$ holds whp.
c) For $p = 0$ we have $\tilde{\boldsymbol{G}}_{\mathrm{b}} \equiv 0$, $\tilde{\boldsymbol{H}}_{\mathrm{b}} \equiv 0$ and $\boldsymbol{x}_M = (\emptyset, \emptyset, \binom{[n]}{3}, \binom{[n]}{3})$ almost surely.
d) For $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{Y}}_{\mathrm{h},M} \cap \boldsymbol{\mathcal{N}}_{\mathrm{t},M}$ we have $\boldsymbol{\mathcal{Y}}_{\mathrm{h},M} = \boldsymbol{\mathcal{Y}}_{\mathrm{t},M} \cup \boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{N}}_{\mathrm{t},M} \subseteq \boldsymbol{\mathcal{N}}_{\mathrm{h},M} \cup \boldsymbol{\mathcal{F}}$ and $|\boldsymbol{\mathcal{F}}| \leq 1$.
e) We have $\mathcal{E} = \mathcal{E}_1 \cap \{\boldsymbol{\mathcal{F}} = \emptyset\} = \{\boldsymbol{\mathcal{Y}}_{\mathrm{t},M} = \boldsymbol{\mathcal{Y}}_{\mathrm{h},M} = \tilde{\boldsymbol{H}}_{\mathrm{b}}^{-1}(1), \boldsymbol{\mathcal{N}}_{\mathrm{t},N} \subseteq \boldsymbol{\mathcal{N}}_{\mathrm{h},N} = \tilde{\boldsymbol{H}}_{\mathrm{b}}^{-1}(0)\}$.
f) On $|\boldsymbol{\mathcal{F}}| = 1$ the event $\mathcal{E}_1$ holds, and using $\boldsymbol{\mathcal{F}} = \{\boldsymbol{E}^*_{\boldsymbol{i}}\}$ we have $\boldsymbol{\pi}_{\boldsymbol{i}} < \boldsymbol{\pi}'_{\boldsymbol{i}} \leq \pi$, $\boldsymbol{m}_{\boldsymbol{i}} = 1$.
g) We have $\mathcal{E}_{\mathrm{t}} \subseteq \mathcal{E}$. Further, we have $\mathcal{E}_{\mathrm{t}}$ whp.
h) The sets $\mathcal{I}_{\mathrm{c}}$, $\mathcal{I}_{\mathrm{i}}$, $\mathcal{I}_{\mathrm{t}}$, $\mathcal{I}_1$ and $\mathcal{I}_*$ are a partition of $[M]$.
i) For $p > 0$, $i \in [M] \cup \{0\}$ we have $\{x : \mathbb{P}(\boldsymbol{x}_i = x | \mathcal{E}_{\mathrm{h}}(H)) > 0\} \subseteq \mathcal{X}_{\mathrm{h},i}$.
j) For $p > 0$, $i \in \mathcal{I}_{\mathrm{c}}$ and on $\mathcal{E}_{\mathrm{h}}(H)$ we have $\boldsymbol{\pi}_i = \boldsymbol{\pi}'_i = 1$, $\boldsymbol{b}_i = 1$ and $\boldsymbol{b}'_i = 1$.
k) For $p > 0$, $i \in \mathcal{I}_{\mathrm{i}}$ and on $\mathcal{E}_{\mathrm{h}}(H)$ we have $\boldsymbol{\pi}_i = 1 > \boldsymbol{\pi}'_i > 0$ and $\boldsymbol{b}_i = 0$.
l) For $p > 0$, $i \in \mathcal{I}_{\mathrm{t}}$ and on $\mathcal{E}_{\mathrm{h}}(H)$ we have $\boldsymbol{\pi}_i = \boldsymbol{\pi}'_i = 0$.
m) For $p > 0$, $i \in \mathcal{I}_1$ and on $\mathcal{E}_{\mathrm{h}}(H)$ we have $\boldsymbol{\pi}_i \geq \pi \geq \boldsymbol{\pi}'_i > 0$, $\boldsymbol{b}_i = 1$ and $\boldsymbol{b}'_i = 1$.
n) For $p > 0$, $i \in \mathcal{I}_*$ and on $\mathcal{E}_{\mathrm{h}}(H)$ one of the following holds.

- *We have $\boldsymbol{\pi}_i \geq \pi \geq \boldsymbol{\pi}'_i > 0$ and $\boldsymbol{b}_i = 0$.*
- *We have $\boldsymbol{\pi}_i \geq \pi \geq \boldsymbol{\pi}'_i > 0$ and $\boldsymbol{b}_i = 1$, $\boldsymbol{b}'_i = 0$.*
- *We have $\boldsymbol{\pi}_i = 0 < \boldsymbol{\pi}'_i \leq \pi$ and $\boldsymbol{m}_i = 0$.*

Proof. Part 4.26a) follows from $n \geq n_\circ$. Proposition 4.13 with $1 - (1-q)^3 \leq 3q$ for $q \in [0,1]$ and $n \geq n_\circ$ yields $\mathbb{P}(\mathcal{E}_1) \geq 1 - \min(\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|], \sqrt{\frac{1}{2}\mathbb{E}[|\boldsymbol{\mathcal{C}}_{\mathrm{t}}^{\circ}|]})3cn^7 p_\uparrow^{11} + o(1)$. For $p \leq p_{\mathrm{c}} = (\frac{1}{240\binom{n}{6}})^{1/9}$ we have $\mathbb{P}(\mathcal{E}_1) \geq 1 - \frac{3}{2}n^7 p_\uparrow^{11} + o(1) = 1 + o(1)$ and for $p \in [p_{\mathrm{c}}, p_\uparrow]$ we have $\mathbb{P}(\mathcal{E}_1) \geq 1 - \mathcal{O}(n^{10} p_\uparrow^{31/2}) + o(1) = 1 + o(1)$. Part 4.26c) is immediate. For Part 4.26d), notice that if the first step fails, we have $\boldsymbol{x}_M \equiv \emptyset$, and notice that $\boldsymbol{\mathcal{F}} \neq \emptyset$ if and only if we entered Step 2.2.4.2 in Figure 3, and that then Step 2.2.1 ensures that $|\boldsymbol{\mathcal{F}}| = 1$. The remainder is now immediate. For Part 4.26e), we clearly have $\mathcal{E} \subseteq \mathcal{E}_1 \cap \{\boldsymbol{\mathcal{F}} = \emptyset\} = \mathcal{E}' \subseteq \mathcal{E}$ and thus equality, where

$$\mathcal{E}' = \{\boldsymbol{\mathcal{Y}}_{\mathrm{t},M} = \boldsymbol{\mathcal{Y}}_{\mathrm{h},M} = \tilde{\boldsymbol{H}}_{\mathrm{b}}^{-1}(1), \boldsymbol{\mathcal{N}}_{\mathrm{t},N} \subseteq \boldsymbol{\mathcal{N}}_{\mathrm{h},N} = \tilde{\boldsymbol{H}}_{\mathrm{b}}^{-1}(0)\}.$$

For Part 4.26f), on $|\boldsymbol{\mathcal{F}}| = 1$ the event $\mathcal{E}_1$ holds since otherwise $\boldsymbol{x}_M \equiv \emptyset$, so $\boldsymbol{E}^*$ is well-defined. Part 4.26d) suggests that there exists a unique $\boldsymbol{i} \in [M]$ with $\boldsymbol{\mathcal{F}} = \{\boldsymbol{E}^*_{\boldsymbol{i}}\}$, and clearly we have $\boldsymbol{\pi}_{\boldsymbol{i}} < \boldsymbol{\pi}'_{\boldsymbol{i}}$, $\boldsymbol{m}_{\boldsymbol{i}} = 1$. This gives $p > 0$, $\pi > 0$ by Part 4.26a), further $\boldsymbol{\pi}_{\boldsymbol{i}} < 1$ and hence Proposition 4.15 shows that $\boldsymbol{\pi}'_{\boldsymbol{i}} \leq \pi$. For Part 4.26g) let $p > 0$, yielding $\pi > 0$. Part 4.26b), Observation 4.8, Observation 4.11, and Observation 4.12 yield that $\mathcal{E}_{\mathrm{t}}$ holds whp. For $\mathcal{E}_{\mathrm{t}} \subseteq \mathcal{E}$, we use Part 4.26f) and show that $\mathcal{B} \cap \mathcal{E}_{\mathrm{t}} = \emptyset$, where $\mathcal{B} = \mathcal{E}_1 \cap \{\boldsymbol{\mathcal{F}} \neq \emptyset\}$. On $\mathcal{B}$ we know that $\mathcal{E}_1$ holds, $\boldsymbol{\pi}_{\boldsymbol{i}} < \boldsymbol{\pi}'_{\boldsymbol{i}} \leq \pi$ and $\boldsymbol{m}_{\boldsymbol{i}} = 1$. By the choice of π, Proposition 4.14 suggests that $\boldsymbol{\pi}_i = 0$ and $\mathcal{C}_{\mathrm{h}}(\boldsymbol{H}) \setminus \mathcal{C}^\circ_{\mathrm{h}}(\boldsymbol{H}) \neq \emptyset$ or $\mathcal{C}^\circ_{\mathrm{i}}(\boldsymbol{H}) \neq \emptyset$ unless $\boldsymbol{X} = (\boldsymbol{E}^*_{\boldsymbol{i}}, \tilde{\boldsymbol{\mathcal{C}}}^\circ_{\mathrm{t}}, \boldsymbol{\mathcal{Y}}_{\mathrm{t},\boldsymbol{i}-1}, \boldsymbol{\mathcal{N}}_{\mathrm{t},\boldsymbol{i}-1}) \notin \mathcal{D}^\circ_{\mathrm{t}}$, where $\boldsymbol{H}$ is given by $\boldsymbol{H}^{-1}(1) = \boldsymbol{\mathcal{Y}}_{\mathrm{t},\boldsymbol{i}-1} \cup \{\boldsymbol{E}^*_{\boldsymbol{i}}\}$. However, in this step we necessarily have $\boldsymbol{\mathcal{Y}}_{\mathrm{t},\boldsymbol{i}-1} = \boldsymbol{\mathcal{Y}}_{\mathrm{h},\boldsymbol{i}-1}$, $\tilde{\boldsymbol{\mathcal{C}}}^\circ_{\mathrm{t}} = \tilde{\boldsymbol{\mathcal{C}}}^\circ_{\mathrm{h}}$ and $\boldsymbol{\mathcal{N}}_{\mathrm{t},\boldsymbol{i}-1} \subseteq \boldsymbol{\mathcal{N}}_{\mathrm{h},\boldsymbol{i}-1}$, so $\boldsymbol{H}$ is also the hypergraph for $(\boldsymbol{E}^*_{\boldsymbol{i}}, \tilde{\boldsymbol{\mathcal{C}}}^\circ_{\mathrm{h}}, \boldsymbol{\mathcal{Y}}_{\mathrm{h},\boldsymbol{i}-1}, \boldsymbol{\mathcal{N}}_{\mathrm{h},\boldsymbol{i}-1})$. Since $\tilde{\boldsymbol{H}}_{\mathrm{b}}$ is chosen conditional to $\tilde{\boldsymbol{\mathcal{C}}}^\circ_{\mathrm{h}}$ and $\boldsymbol{\mathcal{Y}}_{\mathrm{h},\boldsymbol{i}-1} \subseteq \tilde{\boldsymbol{H}}^{-1}_{\mathrm{b}}(1)$, this gives $\mathcal{B} \subseteq \{\tilde{\boldsymbol{H}}_{\mathrm{b}} \notin \mathcal{T}\}$, and hence $\mathcal{E}_{\mathrm{t}} \subseteq \mathcal{E}$.

Now, let $\mathcal{C} = \mathcal{C}^\circ_{\mathrm{h}}(H)$ and $E^* = E^*_{\mathrm{h},H}$. Part 4.26h) follows with $H^{-1}(1) = E^*(\mathcal{I}_{\mathrm{c}} \cup \mathcal{I}_1)$ and $H^{-1}(0) = E^*(\mathcal{I}_{\mathrm{i}} \cup \mathcal{I}_{\mathrm{t}} \cup \mathcal{I}_0 \cup \mathcal{I}_*)$. For the remainder, let $p \in (0, p_\uparrow]$, and notice that $\mathbb{P}(\mathcal{E}_{\mathrm{h}}(H)) > 0$. Let $x = (\mathcal{Y}_{\mathrm{t}}, \mathcal{Y}_{\mathrm{h}}, \mathcal{N}_{\mathrm{t}}, \mathcal{N}_{\mathrm{h}})$ be such that $\mathbb{P}(\boldsymbol{x}_M = x | \mathcal{E}_{\mathrm{h}}(H)) > 0$. By Part 4.26g) we have $\mathcal{E}_{\mathrm{h}}(H) \subseteq \mathcal{E}_{\mathrm{t}} \subseteq \mathcal{E}$, so $\mathcal{Y}_{\mathrm{t}} = \mathcal{Y}_{\mathrm{h}} = H^{-1}(1)$ and $\mathcal{N}_{\mathrm{t}} \subseteq \mathcal{N}_{\mathrm{h}} = H^{-1}(0) = E^*(\mathcal{I}_{\mathrm{i}} \cup \mathcal{I}_{\mathrm{t}} \cup \mathcal{I}_0 \cup \mathcal{I}_*)$ by Part 4.26e). This shows that $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i} = \boldsymbol{\mathcal{Y}}_{\mathrm{h},i} = \mathcal{Y}_{\mathrm{h},i}$ and $\boldsymbol{\mathcal{N}}_{\mathrm{t},i} \subseteq \boldsymbol{\mathcal{N}}_{\mathrm{h},i} = \mathcal{N}_{\mathrm{h},i}$ on $\mathcal{E}_{\mathrm{h}}(H)$.

Part 4.26j) is immediate. For Part 4.26k) we have $\boldsymbol{\pi}_i = 1 \geq \boldsymbol{\pi}'_i$, and $\boldsymbol{b} = 0$ since we have $H(E^*_i) = 0$, but on $\boldsymbol{b} = 1$ we would habe $\boldsymbol{b}' = 1$, which also yields $\boldsymbol{\pi}'_i < 1$. Part 4.26l) is trivial, since on $\mathcal{E}_{\mathrm{h}}(H)$ and for $i \in \mathcal{I}_{\mathrm{t}}$ there exist $E, E' \in \mathcal{Y}_{\mathrm{h},i-1} = \boldsymbol{\mathcal{Y}}_{\mathrm{h},i-1} = \boldsymbol{\mathcal{Y}}_{\mathrm{t},i-1}$ with $\{E, E', E^*_i\} \in \overline{\mathcal{C}}^\circ \setminus \mathcal{C}$, and thus $\boldsymbol{\pi}_i = \boldsymbol{\pi}'_i = 0$ by Proposition 4.14 and Proposition 4.15. For Part 4.26m), using that on $\mathcal{E}_{\mathrm{h}}(H)$ we have $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i} = \boldsymbol{\mathcal{Y}}_{\mathrm{h},i} = \mathcal{Y}_{\mathrm{h},i}$ with $E^*_i \in \mathcal{Y}_{\mathrm{h},i}$, so using $H \in \mathcal{T}$ we have $\boldsymbol{\pi}_i > 0$ and further $\boldsymbol{\pi}_i \geq \pi \geq \boldsymbol{\pi}'_i > 0$. Since we have $H(E^*_i) = 1$, this gives $\boldsymbol{b}_i = \boldsymbol{b}'_i = 1$. For Part 4.26n), we start with $\boldsymbol{\pi}_i > 0$, which yields $\boldsymbol{\pi}_i \geq \pi \geq \boldsymbol{\pi}'_i > 0$ as before. Further, we have $\boldsymbol{b}_i = 0$ or $\boldsymbol{b}_i = 1$ and $\boldsymbol{b}'_i = 0$ since $H(E^*_i) = 0$. For $\boldsymbol{\pi}_i = 0$ we have $\boldsymbol{\pi}'_i > 0$ since $i \notin \mathcal{I}_{\mathrm{t}}$ and $\boldsymbol{m}_i = 0$ since $H(E^*_i) = 0$.

Now, Part 4.26k) implies that $\mathcal{N}_{\mathrm{t}} \cap E^*(\mathcal{I}_{\mathrm{i}}) = \emptyset$, Part 4.26l) yields $E^*(\mathcal{I}_{\mathrm{t}}) \subseteq \mathcal{N}_{\mathrm{t}}$, and hence we have $x \in \mathcal{X}_{\mathrm{h},M}$. This completes the proof of the remaining Part 4.26i). □

Corollary 4.16 follows from Lemma 4.26g). Moreover, Lemma 4.26 provides information about the coupling that can be extracted from the observed hypergraph $H \in \mathcal{T}$ (if we also know that the first step was successful). Finally, notice that in Part 4.26n) the case distinction for $\boldsymbol{\pi}_i = 0$ depends on the specific choice of $\boldsymbol{\mathcal{N}}_{\mathrm{t},i-1}$, in particular if we chose to ignore or to exclude hyperedges for the case $\boldsymbol{\pi}_j > 0$ in previous steps.

4.3.6 Additional Triangles. Under the assumptions of Corollary 4.16 and working towards Section 4.4, we discuss the additional triangles $\boldsymbol{\mathcal{H}}_{\mathrm{t},n,p,\pi} = \tilde{\boldsymbol{H}}^{-1}_{\mathrm{t}}(1) \setminus \tilde{\boldsymbol{H}}^{-1}_{\mathrm{b}}(1)$. Let $\boldsymbol{\mathcal{H}}_{\mathrm{c},n,p,\pi} = \bigcup_{E \in \tilde{\boldsymbol{\mathcal{C}}}^\circ_{\mathrm{h}}} E$ be the hyperedges in the clean 3-cycles. Now, we show that vertex sets that are typically not too large and not incident to clean 3-cycles (in $\tilde{\boldsymbol{H}}_{\mathrm{b}}$) are typically also not incident to additional triangles.

Lemma 4.27. *Let $p_\uparrow = \Omega(\ln(n)^{1/3}/n^{2/3})$, $p_\uparrow = o(n^{-20/31})$ and c from Proposition 4.14a). For $\delta(n), V(n) : \mathbb{Z}_{>0} \to \mathbb{R}_{\geq 0}$ with $\delta(n) = o(1)$ and $V(n) = o(n^{-9}p_\uparrow^{-14})$ there exists $\varepsilon(n) = o(1)$ such that the following holds. For $p \in [0, p_\uparrow]$, $\pi = \max(0, (1 - cn^7 p_\uparrow^{11})p)$ and $\mathcal{V}(H) \subseteq [n]$, $H \in \{0,1\}^{\binom{[n]}{3}}$, with $\mathbb{P}(|\mathcal{V}(\tilde{\boldsymbol{H}}_{\mathrm{b}})| \leq V, \mathcal{V}(\tilde{\boldsymbol{H}}_{\mathrm{b}}) \cap \bigcup_{E \in \boldsymbol{\mathcal{H}}_{\mathrm{c}}} E = \emptyset) \geq 1 - \delta$ we have $\mathbb{P}(\mathcal{V}(\tilde{\boldsymbol{H}}_{\mathrm{b}}) \cap \bigcup_{E \in \boldsymbol{\mathcal{H}}_{\mathrm{t}}} E \neq \emptyset) \leq \varepsilon$.*

Proof. Fix $p_\uparrow$, $\pi_\uparrow = p_\uparrow^3$, δ and V. As in Section 4.3.5, let $n_\circ \in \mathbb{Z}_{\geq 12}$ be sufficiently large such that $p_\uparrow \in (0,1)$, $cn^7 p_\uparrow^{11} < 1$ for $n \geq n_\circ$. For $n < n_\circ$ let $\varepsilon(n) = 1$, and thus let $n \geq n_\circ$ in the remainder, so in particular $\pi = (1 - cn^7 p_\uparrow^{11})p$. For $p = 0$ we have $\pi = 0$, $\boldsymbol{\mathcal{H}}_{\mathrm{t}} = \emptyset$ and thus the assertion is trivial, so let $p > 0$ and thus $\pi > 0$ in the remainder.

Let $\mathcal{T}$ be the set from Equation (18), and $\mathcal{E}_{\mathrm{t}}$, $\mathcal{E}_{\mathrm{h}}(H)$, $\varepsilon_1 = o(1)$ from Lemma 4.26 such that

$\mathbb{P}(\mathcal{E}_\mathrm{t}) \geq 1 - \varepsilon_1$. Then we have $\mathbb{P}(\mathcal{E}_\mathrm{v}) \geq 1 - \delta - \varepsilon_1$, where $\mathcal{E}_\mathrm{v} = \{\tilde{\boldsymbol{\mathcal{C}}}^\circ_\mathrm{t} = \tilde{\boldsymbol{\mathcal{C}}}^\circ_\mathrm{h}, \tilde{\boldsymbol{H}}_\mathrm{b} \in \mathcal{T}_\mathrm{v}\}$ and

$$\mathcal{T}_{\mathrm{v},V,\mathcal{V}} = \left\{ H \in \mathcal{T} : |\mathcal{V}(H)| \leq V, \mathcal{V}(H) \cap \bigcup_{E \in \mathcal{H}_{\mathrm{c},H}} E = \emptyset \right\}, \mathcal{H}_{\mathrm{c},H} = \bigcup_{C \in \mathcal{C}^\circ_\mathrm{h}(H)} C.$$

Let $\boldsymbol{\mathcal{V}} = \mathcal{V}(\tilde{\boldsymbol{H}}_\mathrm{b})$, and $\boldsymbol{N} = |\{(v, E) \in \boldsymbol{\mathcal{V}} \times \boldsymbol{\mathcal{H}}_\mathrm{t} : v \in E\}|$ be the number of pairs of target vertices v and additional triangles E that contain $v \in E$. Then we have

$$\mathbb{P}\left(\boldsymbol{\mathcal{V}} \cap \bigcup_{E \in \boldsymbol{\mathcal{H}}_\mathrm{t}} E \neq \emptyset \right) = \mathbb{P}(\boldsymbol{N} > 0) \leq \mathbb{E}\left[\mathbb{1}\mathcal{E}_\mathrm{v} \mathbb{E}[\boldsymbol{N} | \mathcal{E}_\mathrm{h}(\tilde{\boldsymbol{H}}_\mathrm{b})] \right] + \delta + \varepsilon_1.$$

Let $H \in \mathcal{T}_\mathrm{v}$ and $\mathcal{C} = \mathcal{C}^\circ_\mathrm{h}(H)$. With $\mathcal{P} = \{(v, i) : v \in \mathcal{V}(H) \cap E^*_{\mathrm{h},i}, i \in \mathcal{I}_\mathrm{i} \cup \mathcal{I}_*\}$, Lemma 4.26 yields

$$\mathbb{E}[\boldsymbol{N} | \mathcal{E}_\mathrm{h}(H)] = \sum_{(v,i) \in \mathcal{P}} \mathbb{P}(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 | \mathcal{E}_\mathrm{h}(H)).$$

Notice that $\mathcal{P} = \{(v, i) : v \in \mathcal{V}(H) \cap E^*_{\mathrm{h},i}, i \in \mathcal{I}_*\}$ by the definition of $\mathcal{T}_\mathrm{v}$. Thus, let $i \in \mathcal{I}_*$ with $v \in E^*_{\mathrm{h},i}$. Lemma 4.26 yields $\mathbb{P}(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 | \mathcal{E}_\mathrm{h}(H)) = \mathbb{E}[\mathbb{P}(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 | \boldsymbol{x}_M) | \mathcal{E}_\mathrm{h}(H)]$ with $M = \binom{n}{3}$, and thereby $\mathbb{E}[\boldsymbol{N} | \mathcal{E}_\mathrm{h}(H)] = \sum_{v,i} \mathbb{E}[\mathbb{1}\{\boldsymbol{x}_M \in \mathcal{X}\} \mathbb{P}(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 | \boldsymbol{x}_M) | \mathcal{E}_\mathrm{h}(H)]$, where $\mathcal{X} = \{x : \mathbb{P}(\boldsymbol{x}_M = x | \mathcal{E}_\mathrm{h}(H)) > 0, \mathbb{P}(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 | \boldsymbol{x}_M = x) > 0\}$. Next, let

$$\pi^*_{H,i} = \mathbb{P}\left(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 \middle| H^{-1}(1) \subseteq \tilde{\boldsymbol{H}}_\mathrm{t}^{-1}(1) \right) \geq p^3. \tag{19}$$

Then, for $x = (\mathcal{Y}_\mathrm{t}, \mathcal{Y}_\mathrm{h}, \mathcal{N}_\mathrm{t}, \mathcal{N}_\mathrm{h}) \in \mathcal{X}$ the Harris inequality for $\tilde{\boldsymbol{G}}_\mathrm{b}$ with the edges for $\mathcal{Y}_\mathrm{t}$ given yields

$$\mathbb{P}\left(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 | \boldsymbol{x}_M = x \right) = \mathbb{P}\left(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 | \mathcal{C}^\circ_\mathrm{h}(\tilde{\boldsymbol{H}}_\mathrm{t}) \subseteq \mathcal{C}, \mathcal{Y}_\mathrm{t} \subseteq \tilde{\boldsymbol{H}}_\mathrm{t}^{-1}(1), \mathcal{N}_\mathrm{t} \subseteq \tilde{\boldsymbol{H}}_\mathrm{t}^{-1}(0) \right) \leq \pi^*,$$

using Lemma 4.26. Using that $\mathbb{P}(\tilde{\boldsymbol{H}}_\mathrm{t}(E^*_{\mathrm{h},i}) = 1 | \boldsymbol{x}_M = x) > 0$ for $x \in \mathcal{X}$, Lemma 4.26 further gives $\mathbb{E}[\boldsymbol{N} | \mathcal{E}_\mathrm{h}(H)] \leq \sum_{v,i} \mathbb{P}(\boldsymbol{x}_M \in \mathcal{X} | \mathcal{E}_\mathrm{h}(H)) \pi^* \leq \sum_{v,i} \mathbb{P}(\boldsymbol{\pi}_i > \boldsymbol{\pi}'_i, \boldsymbol{b}_i = 0 | \mathcal{E}_\mathrm{h}(H)) \pi^*$. We also have

$$\begin{aligned} \mathbb{P}(\boldsymbol{\pi}_i > \boldsymbol{\pi}'_i, \boldsymbol{b}_i = 0 | \mathcal{E}_\mathrm{h}(H)) &= \mathbb{E}\left[\mathbb{1}\{\boldsymbol{\pi}_i > \boldsymbol{\pi}'_i\} \mathbb{P}(\boldsymbol{b}_i = 0 | \boldsymbol{x}_{i-1}, \mathcal{E}_\mathrm{h}(H)) | \mathcal{E}_\mathrm{h}(H) \right] \\ &= \mathbb{E}\left[\mathbb{1}\{\boldsymbol{\pi}_i > \boldsymbol{\pi}'_i\} \frac{1 - \frac{\boldsymbol{\pi}'_i}{\boldsymbol{\pi}_i}}{1 - \boldsymbol{\pi}'_i} \middle| \mathcal{E}_\mathrm{h}(H) \right] \end{aligned}$$

since $\boldsymbol{b}_i$ does not depend on the future decisions. Proposition 4.15 with c' therein and Proposition 4.14 yield the bounds for $(1 - c' n^3 \pi_\uparrow^2)\pi \leq \boldsymbol{\pi}'_i \leq \pi$ since $H \in \mathcal{T}_\mathrm{v}$ and $\boldsymbol{\pi}_i > \boldsymbol{\pi}'_i$. For $\boldsymbol{\pi}_i$ we use the Harris inequality for $\tilde{\boldsymbol{G}}_\mathrm{b}$ given that the edges induced by $\boldsymbol{\mathcal{Y}}_{\mathrm{t},i-1} = \mathcal{Y}_{\mathrm{h},H,i-1}$ (from Lemma 4.26) are present to eliminate the down-sets, and a second time to introduce the up-set that the edges given by $H^{-1}(1) \setminus \mathcal{Y}_{\mathrm{h},H,i-1}$ are present, yielding $\boldsymbol{\pi}_i \leq \pi^*$ from Equation (19). Combining these bounds yields $\mathbb{P}(\boldsymbol{\pi}_i > \boldsymbol{\pi}'_i, \boldsymbol{b}_i = 0 | \mathcal{E}_\mathrm{h}(H)) \leq \frac{\pi^* - (1 - c' n^3 \pi_\uparrow)\pi}{\pi^*(1-\pi)}$ and $\mathbb{E}[\boldsymbol{N} | \mathcal{E}_\mathrm{h}(H)] \leq \frac{1}{1 - p_\uparrow^3}(S + S')$, where

$$S = \sum_{(v,i) \in \mathcal{P}} \left(\pi^*_i - p^3 \right), \ S' = \sum_{(v,i) \in \mathcal{P}} \left(p^3 - (1 - c' n^3 \pi_\uparrow^2)\pi \right).$$

Notice that $|\mathcal{P}| \leq V\binom{n-1}{2}$, so $\pi = (1 - cn^7 p_\uparrow^{11})p^3$ yields $S' \leq (c'n^3\pi_\uparrow^2 + cn^7 p_\uparrow^{11})p^3|\mathcal{P}| = \mathcal{O}(n^9 p_\uparrow^{14} V)$. For S, we notice that $\pi_i^* = p^r$ for some $r \in \mathbb{Z} \cap [0, 3]$. For $r = 3$ the contributions vanish, and $r = 0$ implies that all edges of $E_{\mathrm{h},i}^*$ are already present in H, which is not possible since $H \in \mathcal{T}_\mathrm{v}$. This gives $S = \sum_{r=1}^2 N_r(p^r - p^3)$, where $N_r = |\{(v, i) \in \mathcal{P} : \pi_i^* = p^r\}| \leq \sum_{v \in \mathcal{V}(H)} |\mathcal{H}(v)|$ is the number of valid pairs and $\mathcal{H}_r(v) = \{E \in \binom{[n]}{3} : v \in E, |\binom{E}{2} \setminus G^{-1}(1)| = r\}$, $G = G_{\mathrm{t},H}$, are the hyperedges adjacent to v with r edges missing. Notice that $\Delta = \Theta(n^2 p_\uparrow^3)$ in the definition of $\mathcal{T}_\mathrm{v}$ and thus the maximum degree of G is $\Delta_\mathrm{g}(G) = \mathcal{O}(n^2 p_\uparrow^3)$, yielding $|\mathcal{H}_r(v)| = \mathcal{O}(p_\uparrow^{9-3r} n^{5-r})$, thereby $S = \mathcal{O}(n^4 p_\uparrow^7 V)$ and $\mathbb{E}[\boldsymbol{N}|\mathcal{E}_\mathrm{h}(H)] = \mathcal{O}(n^9 p_\uparrow^{14} V)$. Hence, we have $\mathbb{E}[\boldsymbol{N}|\mathcal{E}_\mathrm{h}(H)] \leq \varepsilon_2$ for some $\varepsilon_2 = o(1)$ and thus $\mathbb{P}(\boldsymbol{N} > 0) \leq \varepsilon$ for $\varepsilon = \delta + \varepsilon_1 + \varepsilon_2 = o(1)$. □

Thus, all vertex sets that are not too large do not meet additional triangles. Notice that $n^{-9} p_\uparrow^{-14} = \omega(n^{1/31})$ for $p_\uparrow = o(n^{-20/31})$, so in any case V can be polynomial in n.

4.3.7 Proof of Theorem 4.5. The only difference between Theorem 4.5 and Corollary 4.16 is that the equality in Corollary 4.16 is replaced by the inequality $\pi \leq \pi_+$ with $\pi_+ = \max(0, (1 - cn^7 p_\uparrow^{11})p^3)$. Thus, combining the coupling of $\boldsymbol{G}_\mathrm{b}$ and $\boldsymbol{H}_{\mathrm{b},\pi_+}$ from Corollary 4.16 with the coupling of $\boldsymbol{H}_{\mathrm{b},\pi_+}$ and $\boldsymbol{H}_{\mathrm{b},\pi}$ from Observation 4.7 completes the proof.

4.4 The Cover Coupling

We turn to the proof of Proposition 4.6. As opposed to the previous sections, we use the Landau notation in this section in the traditional sense, since we will only consider fixed choices of p and π. Recall $g(n)$, $\pi_\pm$ from Section 4.1.4. Let $\pi_\uparrow = 2\ln(n)/\binom{n-1}{2}$, $p_\uparrow = \pi_\uparrow^{1/3}$ and c from Theorem 4.5. Let $n_\circ \in \mathbb{Z}_{\geq 12}$ be such that $0 < \pi_- < \pi_+ < \pi_\uparrow < p_\uparrow < 1$, $1 - cn^7 p_\uparrow^{11} > 0$ and $p_+ = (\frac{\pi_+}{1 - cn^7 p_\uparrow^{11}})^{1/3} \in (0, p_\uparrow)$ for all $n \geq n_\circ$. Further, let $p_- = \pi_-^{1/3} \in (0, p_+)$, $\boldsymbol{H}_\pm = \boldsymbol{H}_{\mathrm{b},\pi_\pm}$ and $\boldsymbol{G}_\pm = \boldsymbol{G}_{\mathrm{b},p_\pm}$.

Section 4.4.1 is devoted to the discussion of the coupling underlying Proposition 4.6. The remainder of the proof is discussed in Section 4.4.2.

4.4.1 The Union Process. We establish Proposition 4.6 using the explicit coupling described in Figure 4. For the remainder of this section we let the joint distribution $(\boldsymbol{G}_+, \boldsymbol{H}_+, \boldsymbol{e}_\mathrm{p}, \boldsymbol{E}_\mathrm{p})$ be given by this coupling, for the sake of brevity, i.e. we do neither distinguish $(\tilde{\boldsymbol{e}}_\mathrm{p}, \tilde{\boldsymbol{E}}_\mathrm{p})$ and $(\boldsymbol{e}_\mathrm{p}, \boldsymbol{E}_\mathrm{p})$, nor $(\tilde{\boldsymbol{G}}_+, \tilde{\boldsymbol{H}}_+)$ and $(\boldsymbol{G}_+, \boldsymbol{H}_+)$. Notice that all notions in Figure 3 and Figure 4 are determined by $\boldsymbol{J}' = (\boldsymbol{x}_{\binom{n}{3}}, \boldsymbol{G}_+, \boldsymbol{H}_+, \boldsymbol{e}_\cup)$, where $\boldsymbol{x}_{\binom{n}{3}}$ are the final yes and no decisions in the coupling $(\boldsymbol{G}_+, \boldsymbol{H}_+)$ in Figure 3. In particular, $\boldsymbol{J}'$ determines $\boldsymbol{e}_\mathrm{p}$, $\boldsymbol{E}_\mathrm{p}$, $\boldsymbol{G}_\mathrm{p}$, $\boldsymbol{H}_\mathrm{tp}$, $\boldsymbol{H}_\mathrm{p}$, $\boldsymbol{s}_\mathrm{c}$, $\boldsymbol{s}_\mathrm{f}$, $\boldsymbol{S}_\mathrm{c}$ and $\boldsymbol{S}_\mathrm{f}$. Also, if we have $\mathrm{vdj}(\boldsymbol{\mathcal{C}}_+^\circ) = 1$ for the clean 3-cycles $\boldsymbol{\mathcal{C}}_+^\circ = \mathcal{C}_\mathrm{h}^\circ(\boldsymbol{H}_+)$, then these also determine the coupling order $\boldsymbol{E}^* = E_{\boldsymbol{\mathcal{C}}_+^\circ}^*$ used when the first step in Figure 3 is successful. Let $\boldsymbol{s}_+ = \|\boldsymbol{G}_+\|_1$, $\boldsymbol{S}_+ = \|\boldsymbol{H}_+\|_1$ be the ends of the critical windows, and notice that $\boldsymbol{G}_+ = \boldsymbol{G}_{\mathrm{p},\boldsymbol{s}_+}$ and $\boldsymbol{H}_+ = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_+}$, i.e. this further recovers the corresponding joint distribution in Observation 4.7. Let $\boldsymbol{J} = (\boldsymbol{J}', \boldsymbol{S}_-)$ with $\boldsymbol{S}_- \in \mathbb{Z} \cap [0, \boldsymbol{S}_+]$ depending on $\boldsymbol{J}'$ only through $\boldsymbol{S}_+$, given by $(\boldsymbol{S}_-|\boldsymbol{S}_+ = S_+) \sim \mathrm{Bin}(S_+, \pi_-/\pi_+)$. Thus, this completely defines the critical window in the hypergraph process. Let $\boldsymbol{H}_- = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_-}$.

With the joint distribution $\boldsymbol{J}$ in place, we discuss the coupling in Figure 4 in detail and introduce further derived notions. Let $\boldsymbol{H}_{\mathrm{t}+} = H_{\mathrm{t},\boldsymbol{G}_+}$ be the triangle hypergraph for $\boldsymbol{G}_+$, let $\boldsymbol{G}_{\mathrm{t}+} = G_{\mathrm{t},\boldsymbol{H}_+}$ be the triangle graph for $\boldsymbol{H}_+$ from Section 4.2.5, and let $\boldsymbol{\mathcal{D}}_\pm = \mathcal{D}_\mathrm{h}(\boldsymbol{H}_\pm)$ be the diamonds from Observation 4.12. Let $\boldsymbol{\mathcal{H}}_\mathrm{p} = \{E : \{E, E'\} \in \boldsymbol{\mathcal{D}}_+\}$ be the partner hyperedges and $\boldsymbol{\mathcal{H}}_\mathrm{s} = \boldsymbol{H}_+^{-1}(1) \setminus \boldsymbol{\mathcal{H}}_\mathrm{p}$ the single

1. Use the coupling $(\tilde{\boldsymbol{G}}_+, \tilde{\boldsymbol{H}}_+)$ of $\boldsymbol{G}_+$ and $\boldsymbol{H}_+$ from Theorem 4.5.
   (a) The first step is successful if the coupling in Theorem 4.5 is successful and $\mathrm{vdj}(\mathcal{D}_{\mathrm{h}}(\tilde{\boldsymbol{H}}_+)) = 1$.
   (b) Otherwise, the first step failed.
2. Next, the union edge set $\boldsymbol{\mathcal{G}}_\cup = \boldsymbol{\mathcal{G}}_{\mathrm{g}} \cup \boldsymbol{\mathcal{G}}_{\mathrm{h}}$ and the hyperedges $\boldsymbol{E}_{\mathrm{e}} : \boldsymbol{\mathcal{G}}_{\mathrm{h}} \to \tilde{\boldsymbol{H}}_+^{-1}(1)$ are determined conditional to $\mathcal{E} = \{\tilde{\boldsymbol{G}}_+ = G_+, \tilde{\boldsymbol{H}}_+ = H_+\}$, and given as follows. The union edge process $(\boldsymbol{e}_\cup|\mathcal{E}) \sim \mathrm{u}(\boldsymbol{\mathcal{G}}_\cup!)$ is uniform given $\mathcal{E}$.
   (a) If the first step failed, let $\boldsymbol{\mathcal{G}}_{\mathrm{g}} = G_+^{-1}(1)$, $\boldsymbol{\mathcal{G}}_{\mathrm{h}} = H_+^{-1}(1) \times [3]$ and $\boldsymbol{E}_{\mathrm{e}}(E, i) = E$.
   (b) If the first step was successful, let $G = G_{\mathrm{t},H_+}$. Consider the partition $G_+^{-1}(1) = \mathcal{G}_* \cup \mathcal{G}_1 \cup \mathcal{G}_2$ into $\mathcal{G}_* = G_+^{-1}(1) \setminus G^{-1}(1)$, $\mathcal{G}_2 = \{E_1 \cap E_2 : E \in H_+^{-1}(1)^2, |E_1 \cap E_2| = 2\}$ and $\mathcal{G}_1 = G^{-1}(1) \setminus \mathcal{G}_2$. Let $\mathcal{G}_2'$ be a copy of $\mathcal{G}_2$. For $e \in \mathcal{G}_1$ there exists a unique hyperedge $\boldsymbol{E}_{\mathrm{e}}(e) \in H_+^{-1}(1)$ with $e \in \binom{\boldsymbol{E}_{\mathrm{e}}(e)}{2}$, while for $e \in \mathcal{G}_2$ there exist two such hyperedges E, E', without loss of generality $E = E_i^*$ and $E' = E_j^*$ with $i < j$. Let $\boldsymbol{E}_{\mathrm{e}}(e) = E$ for $e \in \mathcal{G}_2$ and $\boldsymbol{E}_{\mathrm{e}}(e) = E'$ for $e \in \mathcal{G}_2'$. Finally, let $\boldsymbol{\mathcal{G}}_{\mathrm{g}} = G_+^{-1}(1) = \mathcal{G}_* \,\dot\cup\, \mathcal{G}_1 \,\dot\cup\, \mathcal{G}_2$ and $\boldsymbol{\mathcal{G}}_{\mathrm{h}} = \mathcal{G}_1 \,\dot\cup\, \mathcal{G}_2 \,\dot\cup\, \mathcal{G}_2'$.
3. Let $\mathcal{E} = \{\boldsymbol{\mathcal{G}}_{\mathrm{g}} = \mathcal{G}_{\mathrm{g}}, \boldsymbol{\mathcal{G}}_{\mathrm{h}} = \mathcal{G}_{\mathrm{h}}, \boldsymbol{E}_{\mathrm{e}} = E_{\mathrm{e}}, \boldsymbol{e}_\cup = e_\cup\}$. The processes $\tilde{\boldsymbol{e}}_{\mathrm{p}}$ and $\tilde{\boldsymbol{E}}_{\mathrm{p}}$ are conditionally independent given $\mathcal{E}$ and obtained as follows. Let $s_+ = |\mathcal{G}_{\mathrm{g}}|$ and $S_+ = |E_{\mathrm{e}}(\mathcal{G}_{\mathrm{h}})|$.
   (a) Let $e \in \mathcal{G}_{\mathrm{g}}!$ be the subsequence of $e_\cup$ obtained by the restriction to $\mathcal{G}_{\mathrm{g}}$. Then the law of $\tilde{\boldsymbol{e}}_{\mathrm{p}}|\mathcal{E}$ is the same as the law of $\boldsymbol{e}_{\mathrm{p}}|\boldsymbol{e}_{\mathrm{p},[s_+]} = e$.
   (b) Let $E \in E_{\mathrm{e}}(\mathcal{G}_{\mathrm{h}})!$ be the enumeration induced by $s_{\cup\mathrm{h}}(E) = \max\{e_\cup^{-1}(e) : e \in E_{\mathrm{e}}^{-1}(E)\}$. Then the law of $\tilde{\boldsymbol{E}}_{\mathrm{p}}|\mathcal{E}$ is the same as the law of $\boldsymbol{E}_{\mathrm{p}}|\boldsymbol{E}_{\mathrm{p},[S_+]} = E$.

Figure 4: The three main steps in the coupling $(\tilde{\boldsymbol{e}}_{\mathrm{p}}, \tilde{\boldsymbol{E}}_{\mathrm{p}})$ of $\boldsymbol{e}_{\mathrm{p}}$ and $\boldsymbol{E}_{\mathrm{p}}$ are the coupling of the binomial hypergraphs at the end of the critical window, followed by the coupling of the orders of the hyperedges and the edges, and finalized by the completions of the processes.

For the first step, recall that the coupling in Theorem 4.5 is successful if the event $\mathcal{E}$ from Equation (15) holds. Further, recall the diamonds $\mathcal{D}_{\mathrm{h}}(H)$ from Observation 4.12. The first step is successful if the coupling of the binomial graph and hypergraph is successful, and the diamonds in the hypergraph are vertex disjoint.

For the second step, recall the coupling order E^* used in the coupling in Figure 3. If the first step failed, we take the disjoint union of the edges in G_+ and three edges per hypergedge in H_+, and impose these with the uniform order. If the first step was successful, we take the edges of G_+ and clone the overlap edges in the diamonds of H_+. This gives the partition into edges $\mathcal{G}_*$ that only appear in G_+, the edges $\mathcal{G}_1$ in the triangle graph G that belong to a unique hyperedge, the edges $\mathcal{G}_2$ of G_+ that are the shared edges of diamonds in H_+ and belong to the hyperedge with the lower index in the coupling order by definition, and the clones $\mathcal{G}_2'$ that belong to the hyperedge with the larger index. Again, we choose the order uniformly.

In the third step, we extract the subsequence e relevant for the graph process, and complete the process by adding the remaining edges uniformly at random. For the hyperedge process, we extract the subsequence on $\mathcal{G}_{\mathrm{h}}$ and notice that each hyperedge is represented by its own three edges, with the uniform order imposed on the entire edge set. Intuitively, we consider the uniform order on the edges of S_+ vertex disjoint triangles. The index $s_{\cup\mathrm{h}}(E)$ is the step in which the triangle corresponding to E is closed. Thus, the induced enumeration on $E_{\mathrm{e}}(\mathcal{G}_{\mathrm{h}})$ is uniform. Finally, we complete the process by adding the remaining hyperedges uniformly at random.

hyperedges. Assuming the first step in Figure 4 is successful, we can further distinguish the good partners $\boldsymbol{\mathcal{H}}_{\mathrm{g}} = \{\boldsymbol{E}^*_i : i < j, \{\boldsymbol{E}^*_i, \boldsymbol{E}^*_j\} \in \boldsymbol{\mathcal{D}}_+\}$ and the bad partners $\boldsymbol{\mathcal{H}}_{\mathrm{b}} = \boldsymbol{\mathcal{H}}_{\mathrm{p}} \setminus \boldsymbol{\mathcal{H}}_{\mathrm{g}}$.

For now, we stick to the case that the first step is successful. Let $\boldsymbol{\mathcal{G}}_2 = \{E_1 \cap E_2 : \{E_1, E_2\} \in \boldsymbol{\mathcal{D}}_+\}$ be the shared edges, and let $\boldsymbol{\mathcal{G}}_1 = \boldsymbol{G}^{-1}_{\mathrm{t}+}(1) \setminus \boldsymbol{\mathcal{G}}_2$ be the exclusive edges. By the definition of $\boldsymbol{G}_{\mathrm{t}+}$, for each $e \in \boldsymbol{G}^{-1}_{\mathrm{t}+}(1)$ there exists a hyperedge $E \in \boldsymbol{H}^{-1}_+(1)$ that e belongs to, i.e. $e \in \binom{E}{2}$. Since the first step was successful, an edge cannot be shared by more than two hyperedges, otherwise we would get at least three overlapping diamonds by choosing two out of the at least three hyperedges. This gives the partition $\boldsymbol{G}^{-1}_{\mathrm{t}+}(1) = \boldsymbol{\mathcal{G}}_1 \cup \boldsymbol{\mathcal{G}}_2$ into edges $e \in \boldsymbol{\mathcal{G}}_1$ that belong to exactly one hyperedge $\boldsymbol{E}_{\mathrm{e}}(e) \in \boldsymbol{H}^{-1}_+(1)$, and edges $e \in \boldsymbol{\mathcal{G}}_2$ that belong to two hyperedges, one good partner $\boldsymbol{E}_{\mathrm{e}}(e) \in \boldsymbol{\mathcal{H}}_{\mathrm{g}}$ and one bad partner $E \in \boldsymbol{\mathcal{H}}_{\mathrm{b}}$. To account for this missing edge due to sharing, we introduce clones $\boldsymbol{\mathcal{G}}'_2$ of $\boldsymbol{\mathcal{G}}_2$, and choose $\boldsymbol{E}_{\mathrm{e}}(e') = E \in \boldsymbol{\mathcal{H}}_{\mathrm{b}}$ for the clone $e' \in \boldsymbol{\mathcal{G}}'_2$ of $e \in \boldsymbol{\mathcal{G}}_2$. Now, each hyperedge $E \in \boldsymbol{H}^{-1}_+(1)$ is assigned three edges, i.e. $|\boldsymbol{E}^{-1}_{\mathrm{e}}(E)| = 3$. To be specific, since the single hyperedges $E \in \boldsymbol{\mathcal{H}}_{\mathrm{s}}$ do not share edges, all their edges are exclusive, i.e. $\boldsymbol{E}^{-1}_{\mathrm{e}}(E) \subseteq \boldsymbol{\mathcal{G}}_1$. The good partners $E \in \boldsymbol{\mathcal{H}}_{\mathrm{g}}$ still have two exclusive edges, i.e. $|\boldsymbol{E}^{-1}_{\mathrm{e}}(E) \cap \boldsymbol{\mathcal{G}}_1| = 2$, and further one shared edge in $|\boldsymbol{E}^{-1}_{\mathrm{e}}(E) \cap \boldsymbol{\mathcal{G}}_2| = 1$. The bad partners $E \in \boldsymbol{\mathcal{H}}_{\mathrm{b}}$ also have two exclusive edges, i.e. $|\boldsymbol{E}^{-1}_{\mathrm{e}}(E) \cap \boldsymbol{\mathcal{G}}_1| = 2$, and have to make do with the clone of the shared edge in $|\boldsymbol{E}^{-1}_{\mathrm{e}}(E) \cap \boldsymbol{\mathcal{G}}'_2| = 1$.

Since the first step is successful, we have $\boldsymbol{\mathcal{G}}_{\mathrm{g}} = \boldsymbol{G}^{-1}_+(1) = \boldsymbol{\mathcal{G}}_1 \cup \boldsymbol{\mathcal{G}}_2 \cup \boldsymbol{\mathcal{G}}_*$ with $\boldsymbol{\mathcal{G}}_* = \boldsymbol{G}^{-1}_+(1) \setminus \boldsymbol{G}^{-1}_{\mathrm{t}+}(1)$. Hence, in this case we have a non-trivial intersection of $\boldsymbol{\mathcal{G}}_{\mathrm{g}}$ and $\boldsymbol{\mathcal{G}}_{\mathrm{h}} = \boldsymbol{\mathcal{G}}_1 \cup \boldsymbol{\mathcal{G}}_2 \cup \boldsymbol{\mathcal{G}}'_2$ given by $\boldsymbol{\mathcal{G}}_{\mathrm{g}} \cap \boldsymbol{\mathcal{G}}_{\mathrm{h}} = \boldsymbol{\mathcal{G}}_1 \cup \boldsymbol{\mathcal{G}}_2$. On the other hand, if the coupling failed, we have $\boldsymbol{\mathcal{G}}_{\mathrm{g}} \cap \boldsymbol{\mathcal{G}}_{\mathrm{h}} = \emptyset$ for the edge sets $\boldsymbol{\mathcal{G}}_{\mathrm{g}} = \boldsymbol{G}^{-1}_+(1)$ and $\boldsymbol{\mathcal{G}}_{\mathrm{h}} = \boldsymbol{H}^{-1}_+(1) \times [3]$. However, in any case each edge in $\boldsymbol{G}_+$ is, trivially, represented by exactly one edge in $\boldsymbol{\mathcal{G}}_{\mathrm{g}}$, and each hyperedge in $\boldsymbol{H}_+$ is represented by its own exclusive three edges in $\boldsymbol{\mathcal{G}}_{\mathrm{h}}$, where $\boldsymbol{E}_{\mathrm{e}} : \boldsymbol{\mathcal{G}}_{\mathrm{h}} \to \boldsymbol{H}^{-1}_+(1)$ models the 'belongs to' relation. The fact that $\boldsymbol{e}_\cup \in \boldsymbol{\mathcal{G}}_\cup!$ is uniform given $\boldsymbol{\mathcal{G}}_\cup$, suggests that so are the restrictions $\boldsymbol{e}_{\cup,\boldsymbol{\mathcal{G}}_{\mathrm{g}}}$ and $\boldsymbol{e}_{\cup,\boldsymbol{\mathcal{G}}_{\mathrm{h}}}$. Thus, the subsequence given by the former exactly corresponds to $\boldsymbol{e}_{\mathrm{p},\boldsymbol{s}_+}$ as discussed in Observation 4.7. For the latter, we take a viewpoint similar to the triangle hypergraph process $\boldsymbol{H}_{\mathrm{tp}}$. That is, if we consider m triangles represented by edges $[m] \times [3]$ that we order uniformly, i.e. $\boldsymbol{x} \sim \mathrm{u}([m] \times [3])$, and we extract the steps when triangles were closed, i.e. $\boldsymbol{i}(t) = \max\{\boldsymbol{x}^{-1}(t, e) : e \in [3]\} \in [3m]$ for $t \in [m]$, then the enumeration $\boldsymbol{t} \in [m]!$ induced by $\boldsymbol{i}$ is uniform. This explains why the joint distribution $(\boldsymbol{e}_{\mathrm{p}}, \boldsymbol{E}_{\mathrm{p}})$ defined in Figure 4 is indeed a coupling of $\boldsymbol{e}_{\mathrm{p}}$ and $\boldsymbol{E}_{\mathrm{p}}$.

Now, we introduce the graph and the hypergraph process to the shared timeline. Let $\boldsymbol{G}_{\mathrm{g}} = (\boldsymbol{G}_{\mathrm{g},s})_s$ with $\boldsymbol{G}_{\mathrm{g},s} \in \{0,1\}^{\binom{[n]}{2}}$ for $s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|]$ be given by $\boldsymbol{G}^{-1}_{\mathrm{g},s}(1) = \boldsymbol{e}_\cup([s]) \cap \boldsymbol{\mathcal{G}}_{\mathrm{g}}$. Clearly, the graph process $\boldsymbol{G}_{\mathrm{g}}$ is just a stretched version of the graph process $\boldsymbol{G}_{\mathrm{p}}$ up to the end $\boldsymbol{s}_+$ of the critical window. Formally, the steps $\boldsymbol{e}^{-1}_\cup(\boldsymbol{\mathcal{G}}_{\mathrm{g}}) \cup \{0\}$ in the union process are in one to one correspondence with the steps $\mathbb{Z} \cap [0, \tilde{\boldsymbol{s}}_+]$ in the graph process, i.e. apart from the initially empty graph, these are exactly the steps where an edge is added to the graph, and the edges as well as their order are the same in both processes. Of course, this also gives rise to the triangle hypergraph $\boldsymbol{H}_{\mathrm{g}} = (\boldsymbol{H}_{\mathrm{g},s})_s$ in the union process, with $\boldsymbol{H}_{\mathrm{g},s} = H_{\mathrm{t},\boldsymbol{G}_{\mathrm{g},s}}$. For the hypergraph process let $\boldsymbol{G}_{\mathrm{h}} = (\boldsymbol{G}_{\mathrm{h},s})_s$ with $\boldsymbol{G}_{\mathrm{h},s} \in \{0,1\}^{\binom{[n]}{2}}$ for $s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|]$ be given by $\boldsymbol{G}^{-1}_{\mathrm{h},s}(1) = \boldsymbol{e}_\cup([s]) \cap \boldsymbol{\mathcal{G}}_{\mathrm{h}}$, i.e. we consider the graph composed of the edges associated with the hyperedges. Now, the hypergraph process is $\boldsymbol{H}_{\mathrm{h}} = (\boldsymbol{H}_{\mathrm{h},s})_s$ given by $\boldsymbol{H}^{-1}_{\mathrm{h},s}(1) = \{E \in \binom{[n]}{3} : \boldsymbol{E}^{-1}_{\mathrm{e}}(E) \subseteq \boldsymbol{G}_{\mathrm{h},s}\}$, i.e. all edges belonging to E under $\boldsymbol{G}_{\mathrm{h}}$ are present in $\boldsymbol{G}_{\mathrm{h},s}$. Now, there is a one to one correspondence between the steps $\{0\} \cup \{s \in [|\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{h},s} \neq \boldsymbol{H}_{\mathrm{h},s-1}\}$ in which a hyperedge is added and the steps $\mathbb{Z} \cap [0, \boldsymbol{S}_+]$ in the hypergraph process, i.e. apart from the initially empty hypergraph, these are exactly the steps where a hyperedge is added to the hypergraph, and the hyperedges as well as their order are the same in both processes.

4.4.2 Overview. For the proof of Proposition 4.6, we first ensure that the coupling in Figure 4 does not fail whp. We further state a key property of the coupling if the first step is successful. For this purpose let $\boldsymbol{s}_{\mathrm{g}}(E) = \min\{s \in [|\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{g},s}(E) = 1\}$ be the step where $E \in \boldsymbol{H}_{\mathrm{t}+}^{-1}(1)$ is added, and let $\boldsymbol{s}_{\mathrm{h}}(E) = \min\{s \in [|\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{h},s}(E) = 1\}$ be the step where $E \in \boldsymbol{H}_{+}^{-1}(1)$ is added.

Lemma 4.28. *Notice that the following holds.*
a) The first step in Figure 4 is successful whp.
b) Given that the first step is successful, we have $\boldsymbol{s}_{\mathrm{g}}(E) = \boldsymbol{s}_{\mathrm{h}}(E)$ *for all* $E \in \boldsymbol{\mathcal{H}}_{\mathrm{s}} \cup \boldsymbol{\mathcal{H}}_{\mathrm{g}}$.
c) Given that the first step is successful, we have $\boldsymbol{s}_{\mathrm{g}}(E) \le \max(\boldsymbol{s}_{\mathrm{h}}(E), \boldsymbol{s}_{\mathrm{h}}(E'))$ *for all* $E \in \boldsymbol{\mathcal{H}}_{\mathrm{b}}$*, where* $E' \in \boldsymbol{\mathcal{H}}_{\mathrm{g}}$ *is such that* $\{E, E'\} \in \boldsymbol{\mathcal{D}}_{+}$.

This result suggests that we can assume that the first step is successful. It further ensures that all hyperedges but the bad partners are added in the same step in both processes. Finally, it also ensures that all diamonds that are present under the hypergraph process are also present under the graph process. Next, we discuss isolated vertices in the critical window $(\boldsymbol{E}_{\mathrm{p}}, \boldsymbol{S}_{-}, \boldsymbol{S}_{+})$, which will yield the whp bounds $\boldsymbol{S}_{-} < \boldsymbol{S}_{\mathrm{c}} \le \boldsymbol{S}_{+}$ in Theorem 2.6 as a byproduct. For this purpose let $\boldsymbol{D}_{\pm} = D_{\boldsymbol{H}_{\pm}}$ be the vertex degrees from Observation 4.8. Further, let $\boldsymbol{\mathcal{I}}_{\pm} = \{v \in [n] : \boldsymbol{D}_{\pm}(v) = 0\}$ be the isolated vertices and $\boldsymbol{\mathcal{L}}_{\pm} = \{v \in [n] : \boldsymbol{D}_{\pm}(v) \le 6g(n)\}$ the low-degree vertices.

Lemma 4.29. *Notice that the following holds.*
a) We have $\boldsymbol{\mathcal{I}}_{-} \neq \emptyset$ *and* $\boldsymbol{\mathcal{I}}_{+} = \emptyset$ *whp, and thereby* $\boldsymbol{S}_{-} < \boldsymbol{S}_{\mathrm{c}} \le \boldsymbol{S}_{+}$ *whp.*
b) We have $\mathbb{E}[|\boldsymbol{\mathcal{I}}_{-}|] \le e^{g(n)}$, $|\boldsymbol{\mathcal{L}}_{+}| < e^{7g(n)\ln(\ln(n))} = n^{o(1)}$ *whp and* $\boldsymbol{\mathcal{I}}_{-} \subseteq \boldsymbol{\mathcal{L}}_{+}$ *whp.*
c) We have $\boldsymbol{\mathcal{L}}_{+} \cap \bigcup_{E \in \boldsymbol{\mathcal{H}}_{\mathrm{p}} \cup \boldsymbol{\mathcal{H}}} E = \emptyset$ *whp, where* $\boldsymbol{\mathcal{H}} = \bigcup_{C \in \boldsymbol{\mathcal{C}}_{+}^{\circ}} C$.
d) We have $[n] \setminus \boldsymbol{\mathcal{I}}_{-} \subseteq \bigcup_{E \in \boldsymbol{\mathcal{H}}} E$ *whp, where* $\boldsymbol{\mathcal{H}} = \boldsymbol{H}_{-}^{-1}(1) \setminus \{E : \{E, E'\} \in \boldsymbol{\mathcal{D}}_{-}\}$.

Lemma 4.29 does not only show that our choice for the critical window is reasonable, it also establishes typical properties of the binomial hypergraph regarding isolated (low-degree) vertices for π_{-} (π_{+}). The next crucial observation is that also the additional triangles $\boldsymbol{\mathcal{H}}_{\mathrm{t}} = \boldsymbol{H}_{\mathrm{t}+}^{-1}(1) \setminus \boldsymbol{H}_{+}^{-1}(1)$ do not interfere with the property of being isolated in the critical window.

Lemma 4.30. *We have* $\boldsymbol{\mathcal{L}}_{+} \cap \bigcup_{E \in \boldsymbol{\mathcal{H}}_{\mathrm{t}}} E = \emptyset$ *whp.*

Proof. The result follows from Lemma 4.29 and Lemma 4.27 for p_{+} and the vertex set $\boldsymbol{\mathcal{L}}_{+}$. □

Lemma 4.29 and Lemma 4.30 give sufficient control over the isolated vertices in the critical window. Now, we translate Lemma 4.29 and Theorem 2.6 to the union process. For this purpose, let $\boldsymbol{s}_{\mathrm{h}-} = \min\{s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{h},s} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{-}}\}$ be the start and $\boldsymbol{s}_{\mathrm{h}+} = \min\{s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{h},s} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{+}}\}$ the end of the critical window. Of course, we also introduce the hitting times

$$\boldsymbol{s}_{\mathrm{hc}} = \inf\{s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{h},s} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{\mathrm{c}}}\},\ \boldsymbol{s}_{\mathrm{hf}} = \inf\{s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{h},s} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{\mathrm{f}}}\}$$

and notice that these are not necessarily finite since $\boldsymbol{S}_{\mathrm{c}}, \boldsymbol{S}_{\mathrm{f}} > \boldsymbol{S}_{+}$ is possible. For the graph, let

$$\boldsymbol{s}_{\mathrm{gc}} = \inf\{s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{g},s} = \boldsymbol{H}_{\mathrm{tp},\boldsymbol{s}_{\mathrm{c}}}\},\ \boldsymbol{s}_{\mathrm{gf}} = \inf\{s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{g},s} = \boldsymbol{H}_{\mathrm{tp},\boldsymbol{s}_{\mathrm{f}}}\}.$$

As mentioned below Theorem 2.6, the location of $\boldsymbol{s}_{\mathrm{c}}$ is well-known. However, since the lower bound is also almost immediate from our results, we let $(\boldsymbol{J}, \boldsymbol{s}_{-})$ be such that $\boldsymbol{s}_{-}$ depends on $\boldsymbol{J}$ only through $\boldsymbol{s}_{+}$ and $(\boldsymbol{s}_{-}|\boldsymbol{s}_{+} = s_{+}) \sim \mathrm{Bin}(s_{+}, p_{-}/p_{+})$. As for the hypergraph process, we also introduce $\boldsymbol{s}_{\mathrm{g}-} = \min\{s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{g},s} = \boldsymbol{H}_{\mathrm{tp},\boldsymbol{s}_{-}}\}$ and $\boldsymbol{s}_{\mathrm{g}+} = \min\{s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|] : \boldsymbol{H}_{\mathrm{g},s} = \boldsymbol{H}_{\mathrm{tp},\boldsymbol{s}_{+}}\}$.

Lemma 4.31. *Notice that the following holds.*
a) We have $\boldsymbol{S}_- < \boldsymbol{S}_\mathrm{c} \le \boldsymbol{S}_+$ *whp, and* $\boldsymbol{S}_\mathrm{c} = \boldsymbol{S}_\mathrm{f}$ *whp over* $n \in 3\mathbb{Z}_{>0}$.
b) We have $\boldsymbol{s}_{\mathrm{h}-} < \boldsymbol{s}_\mathrm{hc} = \boldsymbol{s}_\mathrm{gc} \le \boldsymbol{s}_{\mathrm{h}+}$ *whp, and* $\boldsymbol{s}_\mathrm{hc} = \boldsymbol{s}_\mathrm{hf}$ *whp over* $n \in 3\mathbb{Z}_{>0}$.
c) We have $\boldsymbol{s}_{\mathrm{g}-} < \boldsymbol{s}_\mathrm{hc} = \boldsymbol{s}_\mathrm{gc} \le \boldsymbol{s}_{\mathrm{g}+}$ *whp, and* $\boldsymbol{s}_\mathrm{hc} = \boldsymbol{s}_\mathrm{hf}$ *whp over* $n \in 3\mathbb{Z}_{>0}$.
d) For $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_\mathrm{s} \cap \boldsymbol{H}^{-1}_{\mathrm{h},\boldsymbol{s}_\mathrm{hc}}(1)$ *we have* $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_\mathrm{s} \cap \boldsymbol{H}^{-1}_{\mathrm{g},\boldsymbol{s}_\mathrm{gc}}(1)$ *and* $\bigcup_{E\in\boldsymbol{\mathcal{H}}} E = [n]$ *whp.*
e) We have $\boldsymbol{s}_- < \boldsymbol{s}_\mathrm{c} \le \boldsymbol{s}_+$ *whp.*

Proposition 4.6 is now immediate from Lemma 4.31, using $\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_\mathrm{c}} = \boldsymbol{H}_{\mathrm{h},\boldsymbol{s}_\mathrm{hc}}$ on $\boldsymbol{S}_\mathrm{c} \le \boldsymbol{S}_+$ and $\boldsymbol{H}_{\mathrm{tp},\boldsymbol{s}_\mathrm{c}} = \boldsymbol{H}_{\mathrm{g},\boldsymbol{s}_\mathrm{gc}}$ on $\boldsymbol{s}_\mathrm{c} \le \boldsymbol{s}_+$. Lemma 4.28 is proved in Section 4.4.3, Lemma 4.29 is proved in Section 4.4.4, and Lemma 4.31 is proved in Section 4.4.5.

4.4.3 Proof of Lemma 4.28. Part 4.28a) is an immediate consequence of Theorem 4.5 and Observation 4.12. Part 4.28b) is immediate from the definition and the discussion in Section 4.4.1, since these hyperedges are associated with the same underlying edges, thus they appear in the same step. Part 4.28c) is also immediate from the definition and the discussion in Section 4.4.1, since the two exclusive edges $\binom{E}{2}\cap\boldsymbol{\mathcal{G}}_1$ belong to E under both $\boldsymbol{G}_\mathrm{g}$ and $\boldsymbol{G}_\mathrm{h}$, and the third, shared edge $\binom{E}{2}\cap\boldsymbol{\mathcal{G}}_2 = \{e\}$ belonging to E under $\boldsymbol{G}_\mathrm{g}$, belongs to the good partner E' under $\boldsymbol{G}_\mathrm{h}$, so at the step $\max(\boldsymbol{s}_\mathrm{h}(E), \boldsymbol{s}_\mathrm{h}(E'))$ where both partners are present, all of their edges are, thus all edges of E under $\boldsymbol{G}_\mathrm{g}$ are present and thereby $\boldsymbol{H}_\mathrm{g}(E) = 1$.

4.4.4 Proof of Lemma 4.29. Using Observation 4.8 we have $\mathbb{E}[|\boldsymbol{\mathcal{I}}_\pm|] = n(1-\pi_\pm)^{\binom{n-1}{2}}$, which gives $0 < \mathbb{E}[|\boldsymbol{\mathcal{I}}_\pm|] \le e^{\mp g(n)}$ and $\mathbb{E}[|\boldsymbol{\mathcal{I}}_\pm|] = (1+\mathcal{O}(\ln(n)^2/n^2))e^{\mp g(n)}$. The second factorial moment is $\mathbb{E}[|\boldsymbol{\mathcal{I}}_\pm|^{\underline{2}}] = n^{\underline{2}}(1-\pi_\pm)^{\binom{n-1}{2}+\binom{n-2}{2}}$, and hence

$$\varepsilon_\pm = \frac{\mathrm{Var}(|\boldsymbol{\mathcal{I}}_\pm|)}{\mathbb{E}[|\boldsymbol{\mathcal{I}}_\pm|]} - 1 = \beta n(1-\pi_\pm)^{\binom{n-2}{2}},\ \beta = 1 - (1-\pi_\pm)^{n-2} - \frac{1}{n}.$$

For β we have $-\frac{1}{n} \le \beta \le (n-2)\pi_\pm = \frac{2}{n-1}(\ln(n) \pm g(n)) \le \frac{4\ln(n)}{n-1}$ using $g(n) \le \ln(n)$, and on the other hand we have $0 \le n(1-\pi_\pm)^{\binom{n-2}{2}} \le n\exp(-(\ln(n) \pm g(n))) \le e^{g(n)}$. Combining these bounds gives $|\varepsilon_\pm| \le \frac{4\ln(n)}{n-1}e^{g(n)} = n^{-1+o(1)}$ using $g(n) = o(\ln(n))$. This shows that $\mathrm{Var}(|\boldsymbol{\mathcal{I}}_\pm|) = (1+o(1))\mathbb{E}[|\boldsymbol{\mathcal{I}}_\pm|]$. Markov's inequality gives $\mathbb{P}(\boldsymbol{\mathcal{I}}_+ \ne \emptyset) \le e^{-g(n)} = o(1)$, and Chebyshev's inequality gives $\mathbb{P}(\boldsymbol{\mathcal{I}}_- = \emptyset) \le \frac{1+o(1)}{\mathbb{E}[|\boldsymbol{\mathcal{I}}_-|]} = (1+o(1))e^{-g(n)}$. This completes the proof of Part 4.29a) and the first part of Part 4.29b).

For the last part of Part 4.29b) let $\boldsymbol{N} = |\boldsymbol{\mathcal{I}}_- \setminus \boldsymbol{\mathcal{L}}_+|$ be the number of vertices that are isolated in $\boldsymbol{H}_-$ and not of low degree in $\boldsymbol{H}_+$. Using Observation 4.7, for a vertex known to be isolated in $\boldsymbol{H}_-$, each adjacent hyperedge is present in $\boldsymbol{H}_+$ independently with probability $\pi = \frac{\pi_+-\pi_-}{1-\pi_-}$, so with $\boldsymbol{b} \sim \mathrm{Bin}(\binom{n-1}{2}, \pi)$ we have $\mathbb{E}[\boldsymbol{N}] = n(1-\pi_-)^{\binom{n-1}{2}}\mathbb{P}(\boldsymbol{b} > 6g(n)) = \mathbb{E}[|\boldsymbol{\mathcal{I}}_-|]\mathbb{P}(\boldsymbol{b} > 6g(n))$. With $\mathbb{E}[|\boldsymbol{\mathcal{I}}_-|] \le e^{g(n)}$ and the Chernoff bound we obtain $\mathbb{E}[\boldsymbol{N}] \le \exp(g(n) - (1+\mathcal{O}(\pi_-))2g(n)) = o(1)$, which shows that $\boldsymbol{\mathcal{I}}_- \subseteq \boldsymbol{\mathcal{L}}_+$ whp. Notice that $\mathbb{E}[|\boldsymbol{\mathcal{L}}_+|] = n\mathbb{P}(\boldsymbol{b} \le 6g(n))$ for $\boldsymbol{b} \sim \mathrm{Bin}(\binom{n-1}{2}, \pi_+)$, so the Chernoff bound gives $\mathbb{E}[|\boldsymbol{\mathcal{L}}_+|] \le \exp(f(n))$, where using $\pi = 6g(n)/\binom{n-1}{2}$ we have

$$f(n) = \ln(n) - \binom{n-1}{2} D_\mathrm{KL}(\pi\|\pi_+) = (1+o(1))6g(n)\ln(\ln(n)).$$

Markov's inequality yields $|\boldsymbol{\mathcal{L}}_+| < e^{7g(n)\ln(\ln(n))}$ whp, and completes the proof of Part 4.29b).

For Part 4.29c), let $\boldsymbol{N} = |\{(v,E,E') : v \in \boldsymbol{\mathcal{L}}_+ \cap (E \setminus E'), \{E,E'\} \in \boldsymbol{\mathcal{D}}_+\}|$ be the number of triplets where $\{E,E'\}$ is a diamond and the outer corner v of E is of low degree. Analogously, let

$\boldsymbol{N}' = |\{(v,E,E') : v \in \boldsymbol{\mathcal{L}}_+ \cap (E \cap E'), \{E,E'\} \in \boldsymbol{\mathcal{D}}_+\}|$ be the number of triplets where v is one of the endpoints of the shared edge. The union bound and Markov's inequality yield

$$\mathbb{P}\left(\boldsymbol{\mathcal{L}}_+ \cap \bigcup_{E\in\boldsymbol{\mathcal{H}}_\mathrm{p}} E \neq \emptyset\right) \leq \mathbb{E}[\boldsymbol{N}] + \mathbb{E}[\boldsymbol{N}'].$$

We have $\mathbb{E}[\boldsymbol{N}] = \binom{n}{3}\binom{3}{2}\binom{n-3}{1}\pi_+^2\mathbb{P}(\boldsymbol{b} \leq 6g(n)-1)$, where $\boldsymbol{b} \sim \mathrm{Bin}(\binom{n-1}{2}-1, \pi_+)$. Since the leading coefficient is only of order $\mathcal{O}(\ln(n)^2)$, proceeding analogous to the above yields $\mathbb{E}[\boldsymbol{N}] = n^{-1+o(1)}$, and further $\mathbb{E}[\boldsymbol{N}'] = \binom{n}{3}\binom{3}{2}\binom{n-3}{1}\pi_+^2\binom{2}{1}\mathbb{P}(\boldsymbol{b} \leq 6g(n)-2) = n^{-1+o(1)}$ for $\boldsymbol{b} \sim \mathrm{Bin}(\binom{n-1}{2}-2, \pi_+)$. This shows that $\boldsymbol{\mathcal{L}}_+ \cap \bigcup_{E\in\boldsymbol{\mathcal{H}}_\mathrm{p}} E = \emptyset$ whp.

To complete Part 4.29c), let $\boldsymbol{N} = |\{(v,E) \in \boldsymbol{\mathcal{L}}_+ \times \binom{[n]}{3}^3 : \{E_1,E_2,E_3\} \in \boldsymbol{\mathcal{C}}_+^\circ, v \in E_1 \setminus (E_2 \cup E_3)\}|$ and $\boldsymbol{N}' = |\{(v,E) \in \boldsymbol{\mathcal{L}}_+ \times \binom{[n]}{3}^3 : \{E_1,E_2,E_3\} \in \boldsymbol{\mathcal{C}}_+^\circ, E_1 \cap E_2 = \{v\}\}|$ be the number of pairs of clean 3-cycles and corners, for outer and inner corners respectively. Then we have

$$\mathbb{P}\left(\boldsymbol{\mathcal{L}}_+ \cap \bigcup_{E\in\boldsymbol{\mathcal{H}}} E \neq \emptyset\right) \leq \mathbb{E}[\boldsymbol{N}] + \mathbb{E}[\boldsymbol{N}'].$$

We have $\mathbb{E}[\boldsymbol{N}] = \binom{n}{3}\binom{3}{1}\binom{n-3}{2}\binom{2}{1}^2\binom{n-5}{1}\pi_+^3\mathbb{P}(\boldsymbol{b} \leq 6g(n)-1)$ for $\boldsymbol{b} \sim \mathrm{Bin}(\binom{n-1}{2}-1, \pi_+)$, yielding $\mathbb{E}[\boldsymbol{N}] = n^{-1+o(1)}$, and analogously $\mathbb{E}[\boldsymbol{N}'] = n^{-1+o(1)}$.

For Part 4.29d), let $\boldsymbol{N} = |\{(v,E,E') : v \in E \setminus E', \boldsymbol{D}_-(v) = 2, \{E,E'\} \in \boldsymbol{\mathcal{D}}_-\}|$ the number of triplets of a diamond with an otherwise isolated outer corner, and $\boldsymbol{N}' = |\{(v,E,E') : v \in E \cap E', \boldsymbol{D}_-(v) = 1, \{E,E'\} \in \boldsymbol{\mathcal{D}}_-\}|$ the number of triplets of a diamond with an otherwise isolated inner corner. The union bound and Markov's inequality yield

$$\mathbb{P}\left(([n] \setminus \boldsymbol{\mathcal{I}}_-) \setminus \bigcup_{E\in\boldsymbol{\mathcal{H}}} E \neq \emptyset\right) \leq \mathbb{P}\left(\mathrm{vdj}(\boldsymbol{\mathcal{D}}_-) = 0\right) + \mathbb{E}[\boldsymbol{N}] + \mathbb{E}[\boldsymbol{N}']$$

Observation 4.12 yields $\mathbb{P}(\mathrm{vdj}(\boldsymbol{\mathcal{D}}_-) = 0) = o(1)$, we have $\mathbb{E}[\boldsymbol{N}] = \binom{n}{3}\binom{3}{2}\binom{n-3}{1}\pi_-^2(1-\pi_-)^{\binom{n-1}{2}-1} = n^{-1+o(1)}$ and we have $\mathbb{E}[\boldsymbol{N}'] = \binom{n}{3}\binom{3}{2}\binom{n-3}{1}\pi_-^2\binom{2}{1}(1-\pi_-)^{\binom{n-1}{2}-2} = n^{-1+o(1)}$.

4.4.5 Proof of Lemma 4.31. Part 4.31a) is Theorem 2.6, a combination of Lemma 4.29 and Theorem 1.3 in [71]. For Part 4.31b), notice that we have $\{\boldsymbol{s}_{\mathrm{h}-} < \boldsymbol{s}_{\mathrm{hc}} \leq \boldsymbol{s}_{\mathrm{h}+}\} = \{\boldsymbol{S}_- < \boldsymbol{S}_\mathrm{c} \leq \boldsymbol{S}_+\}$ and $\{\boldsymbol{s}_{\mathrm{hf}} = \boldsymbol{s}_{\mathrm{hc}} \leq \boldsymbol{s}_{\mathrm{h}+}\} = \{\boldsymbol{S}_\mathrm{f} = \boldsymbol{S}_\mathrm{c} \leq \boldsymbol{S}_+\}$, because $\boldsymbol{H}_{\mathrm{p},[\boldsymbol{S}_+]}$ is a subprocess of $\boldsymbol{H}_\mathrm{h}$. Lemma 4.28 yields that $\boldsymbol{\mathcal{H}}_\mathrm{s} \cap \boldsymbol{H}_{\mathrm{g},s}^{-1}(1) = \boldsymbol{\mathcal{H}}_\mathrm{s} \cap \boldsymbol{H}_{\mathrm{h},s}^{-1}(1)$ for all $s \in \mathbb{Z} \cap [0, |\boldsymbol{\mathcal{G}}_\cup|]$ on $\mathcal{E}_2 = \mathcal{E} \cap \{\mathrm{vdj}(\boldsymbol{\mathcal{D}}_+) = 1\}$, with $\mathcal{E}$ from Equation (15). Further, with Lemma 4.29 and the above it yields that whp $\mathcal{E}_2$ and $\boldsymbol{s}_{\mathrm{h}-} < \boldsymbol{s}_{\mathrm{hc}} \leq \boldsymbol{s}_{\mathrm{h}+}$ hold, and that for all $v \in \boldsymbol{\mathcal{I}}_- \subseteq \boldsymbol{\mathcal{L}}_+$ and $s \in \mathbb{Z} \cap [\boldsymbol{s}_{\mathrm{h}-}, \boldsymbol{s}_{\mathrm{h}+}]$ we have $\boldsymbol{\mathcal{N}}_\mathrm{h}(v,s) \subseteq \boldsymbol{\mathcal{H}}_\mathrm{s}$, where $\boldsymbol{\mathcal{N}}_\mathrm{h}(v,s) = \{E \in \boldsymbol{H}_{\mathrm{h},s}^{-1}(1) : v \in E\}$. This further gives $\boldsymbol{\mathcal{N}}_\mathrm{h}(v,s) \subseteq \boldsymbol{\mathcal{N}}_\mathrm{g}(v,s) \cap \boldsymbol{\mathcal{H}}_\mathrm{s}$, where $\boldsymbol{\mathcal{N}}_\mathrm{g}(v,s) = \{E \in \boldsymbol{H}_{\mathrm{g},s}^{-1}(1) : v \in E\}$. We also obtain that whp for all $v \in [n] \setminus \boldsymbol{\mathcal{I}}_-$ we have $\boldsymbol{\mathcal{N}}_\mathrm{h}(v,s) \cap \boldsymbol{\mathcal{H}}_\mathrm{s} = \boldsymbol{\mathcal{N}}_\mathrm{g}(v,s) \cap \boldsymbol{\mathcal{H}}_\mathrm{s} \neq \emptyset$. This shows that $\boldsymbol{s}_{\mathrm{gc}} \leq \boldsymbol{s}_{\mathrm{hc}}$ whp. Now, Lemma 4.30 yields $\boldsymbol{\mathcal{N}}_\mathrm{h}(v,s) = \boldsymbol{\mathcal{N}}_\mathrm{g}(v,s) \subseteq \boldsymbol{\mathcal{H}}_\mathrm{s}$ for all $v \in \boldsymbol{\mathcal{I}}_-$, $s \in \mathbb{Z} \cap [\boldsymbol{s}_{\mathrm{h}-}, \boldsymbol{s}_{\mathrm{h}+}]$ and thus $\boldsymbol{s}_{\mathrm{gc}} = \boldsymbol{s}_{\mathrm{hc}}$ whp. For Part 4.31d), we have $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_\mathrm{s} \cap \boldsymbol{H}_{\mathrm{h},\boldsymbol{s}_{\mathrm{gc}}}^{-1}(1)$ and $\bigcup_{E\in\boldsymbol{\mathcal{H}}} E = [n]$ by the above. For Part 4.31c), we have $\boldsymbol{s}_{\mathrm{gc}} \leq \boldsymbol{s}_{\mathrm{g}+}$ on $\boldsymbol{s}_{\mathrm{gc}} < \infty$ and hence $\boldsymbol{s}_{\mathrm{gc}} = \boldsymbol{s}_{\mathrm{hc}} \leq \boldsymbol{s}_{\mathrm{g}+}$ whp. For the lower bound, we use Theorem 4.5 to couple $\boldsymbol{G}_-$ with $\boldsymbol{H}'_- = \boldsymbol{H}_{\mathrm{b},\pi'_-}$, $\pi'_- = (1 - cn^7p_\uparrow^{11})p_-^3 > 0$. Notice that $\pi'_- = \frac{\ln(n)-g'(n)}{\binom{n-1}{2}}$ with $g'(n) = (1+o(1))g(n)$

given by the above, so the results for $\boldsymbol{H}_-$ also apply for $\boldsymbol{H}'_-$. Let $\boldsymbol{\mathcal{I}}'_- = \{v \in [n] : D_{\boldsymbol{H}'_-}(v) = 0\}$ be the isolated vertices in $\boldsymbol{H}'_-$, and let $\boldsymbol{\mathcal{H}}'_\mathrm{t} = \boldsymbol{H}_{\mathrm{t}-}^{-1}(1) \setminus \boldsymbol{H}'^{-1}_-(1)$ be the extra cliques in $\boldsymbol{H}_{\mathrm{t}-} = H_{\mathrm{t},\boldsymbol{G}_-}$. Notice that by Lemma 4.29, Lemma 4.27 applies for $\boldsymbol{\mathcal{I}}'_-$ and $\boldsymbol{\mathcal{H}}'_\mathrm{t}$ and hence $\boldsymbol{\mathcal{I}}'_- \cap \bigcup_{E\in\boldsymbol{\mathcal{H}}'_\mathrm{t}} E = \emptyset$ whp. This yields $\{v : D_{\boldsymbol{H}_{\mathrm{t}-}}(v) = 0\} = \boldsymbol{\mathcal{I}}'_- \neq \emptyset$ whp, so $\boldsymbol{s}_\mathrm{c} > \boldsymbol{s}_-$ whp. This completes the proof of Part 4.31c), and of Part 4.31e) analogously to the above.

4.5 Existence of a Perfect Matching

In this section we complete the proof of Theorem 4.2 based on the coupling from Proposition 4.6. As in Section 4.4, we use the Landau notation in the traditional sense. As in Section 4.4, we continue to define couplings of two (unrelated) random objects by defining a joint distribution for the pair.

4.5.1 Overview. Recall the notions from Section 4.4.1 and Section 4.4.2. With $\mathcal{E}$ from Equation (15), let $\mathcal{E}_2 = \mathcal{E} \cap \{\mathrm{vdj}(\boldsymbol{\mathcal{D}}_+) = 1\}$ be the event that the first step in Figure 4 was successful. Further, let $\mathcal{E}_3 = \mathcal{E}_2 \cap \{\boldsymbol{s}_\mathrm{gc} = \boldsymbol{s}_\mathrm{hc} = \boldsymbol{s}_\mathrm{hf} \leq \boldsymbol{s}_{\mathrm{h}+}\}$ be the whp event that the hitting times in the union process coincide. On $\mathcal{E}_3$, let $\boldsymbol{\mathcal{H}}_\circ = \boldsymbol{H}^{-1}_{\mathrm{h},\boldsymbol{s}_\mathrm{hc}}(1) \setminus \boldsymbol{H}^{-1}_{\mathrm{g},\boldsymbol{s}_\mathrm{hc}}(1)$ be the excess hyperedges in $\boldsymbol{H}_{\mathrm{h},\boldsymbol{s}_\mathrm{hc}}$ and notice that $\boldsymbol{\mathcal{H}}_\circ = \boldsymbol{H}^{-1}_{\mathrm{p},\boldsymbol{S}_\mathrm{c}}(1) \setminus \boldsymbol{H}^{-1}_{\mathrm{tp},\boldsymbol{s}_\mathrm{c}}(1)$. Theorem 4.2 is immediate from the following result. Let $\tilde{\boldsymbol{E}}_\mathrm{p}$ be a copy of $\boldsymbol{E}_\mathrm{p}$, together with all derived notions, e.g. $\tilde{\boldsymbol{H}}_\mathrm{p}$.

Proposition 4.32. *There exists a coupling $(\boldsymbol{J}, \tilde{\boldsymbol{E}}_\mathrm{p})$ such that $\tilde{\boldsymbol{H}}^{-1}_{\mathrm{p},\tilde{\boldsymbol{S}}_\mathrm{c}}(1) \subseteq \boldsymbol{H}^{-1}_{\mathrm{p},\boldsymbol{S}_\mathrm{c}}(1) \setminus \boldsymbol{\mathcal{H}}_\circ$ whp.*

The first obstacle in the proof of Proposition 4.32 is that $\boldsymbol{\mathcal{H}}_\circ$ heavily relies on the coupling in Figure 4. But this can be easily remedied by considering the set

$$\boldsymbol{\mathcal{H}} = \left\{E \in \boldsymbol{H}^{-1}_{\mathrm{p},\boldsymbol{S}_\mathrm{c}}(1) : \{E, E'\} \in \boldsymbol{\mathcal{D}}_+, E' \in \boldsymbol{H}^{-1}_{\mathrm{p},\boldsymbol{S}_\mathrm{c}}(0)\right\}.$$

Notice that $\boldsymbol{\mathcal{H}}$ is determined by $(\boldsymbol{E}_\mathrm{p}, \boldsymbol{S}_+)$, and thus does not involve our previous constructions.

Lemma 4.33. *We have $\boldsymbol{\mathcal{H}}_\circ \subseteq \boldsymbol{\mathcal{H}}$ whp.*

Thus, it is sufficient to establish a coupling $(\boldsymbol{J}, \tilde{\boldsymbol{E}}_\mathrm{p})$ such that $\tilde{\boldsymbol{H}}^{-1}_{\mathrm{p},\tilde{\boldsymbol{S}}_\mathrm{c}}(1) \subseteq \boldsymbol{H}^{-1}_{\mathrm{p},\boldsymbol{S}_\mathrm{c}}(1) \setminus \boldsymbol{\mathcal{H}}$ whp by Lemma 4.33, to obtain Proposition 4.32. But this only amounts to a coupling $(\boldsymbol{E}_\mathrm{p}, \boldsymbol{S}_+, \tilde{\boldsymbol{E}}_\mathrm{p})$, i.e. in the remainder we can forget about Section 4.3 and Section 4.4 entirely, and only discuss a coupling of the standard hyperedge process with itself. For this purpose, let $(\hat{\boldsymbol{E}}_\mathrm{p}, \boldsymbol{r}) \sim \mathrm{u}((\binom{[n]}{3})!) \otimes \mathrm{Bin}(1, \pi_\mathrm{r})^{\binom{n}{3}}$ and $\hat{\boldsymbol{\mathcal{H}}} = \{\hat{\boldsymbol{E}}_\mathrm{p}(S) : S \in [\hat{\boldsymbol{S}}_\mathrm{c}] \cap \boldsymbol{r}^{-1}(1)\}$, using the derived notions for the copy $\hat{\boldsymbol{E}}_\mathrm{p}$ of $\boldsymbol{E}_\mathrm{p}$ and $\pi_\mathrm{r} = \frac{18g(n)}{n-1}$.

Lemma 4.34. *There exists a coupling $(\boldsymbol{E}_\mathrm{p}, \boldsymbol{S}_+, \hat{\boldsymbol{E}}_\mathrm{p}, \boldsymbol{r})$ such that $\boldsymbol{\mathcal{H}} \subseteq \hat{\boldsymbol{\mathcal{H}}}$ whp.*

In the last step, we establish that the subprocess obtained from $\hat{\boldsymbol{E}}_\mathrm{p}$ by restricting to $\boldsymbol{r}^{-1}(0)$ is nothing but a standard hyperedge process, i.e. $\tilde{\boldsymbol{E}}_{\mathrm{p},[|\boldsymbol{r}^{-1}(0)|]}$, and that whp the hyperedge removal $\boldsymbol{r}$ does not affect the hitting time.

Lemma 4.35. *We have $\hat{\boldsymbol{S}}_\mathrm{c} = \inf\{S : \bigcup_{E\in\hat{\boldsymbol{E}}_\mathrm{p}([S]\cap\boldsymbol{r}^{-1}(0))} E = [n]\}$ whp. Further, there exists a coupling $(\hat{\boldsymbol{E}}_\mathrm{p}, \boldsymbol{r}, \tilde{\boldsymbol{E}}_\mathrm{p})$ with $\tilde{\boldsymbol{H}}^{-1}_{\mathrm{p},\tilde{\boldsymbol{S}}_\mathrm{c}}(1) \subseteq \hat{\boldsymbol{H}}^{-1}_{\mathrm{p},\hat{\boldsymbol{S}}_\mathrm{c}}(1) \setminus \hat{\boldsymbol{\mathcal{H}}}$.*

At this point, Proposition 4.32 and Theorem 4.2 are almost immediate. We establish Lemma 4.33 in Section 4.5.2, Lemma 4.34 in Section 4.5.3, Lemma 4.35 in Section 4.5.4, Proposition 4.32 in Section 4.5.5, Theorem 4.2 in Section 4.5.6, Corollary 4.3 in Section 4.5.7 and Corollary 4.4 in Section 4.5.8.

4.5.2 Proof of Lemma 4.33. Lemma 4.28 and Lemma 4.29 show that $\mathcal{E}_3$ is a whp event. On $\mathcal{E}_3$ we have $\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{\mathrm{c}}} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{s}_{\mathrm{hc}}}$ and $\boldsymbol{H}_{+} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{+}} = \boldsymbol{H}_{\mathrm{h},\boldsymbol{s}_{\mathrm{h}+}}$, so Lemma 4.28 shows that

$$\boldsymbol{\mathcal{H}}_{\circ} \subseteq \left\{E \in \boldsymbol{\mathcal{H}}_{\mathrm{b}} \cap \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{\mathrm{c}}}^{-1}(1) : \{E, E'\} \in \boldsymbol{\mathcal{D}}_{+}, E' \in \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{\mathrm{c}}}^{-1}(0)\right\} \subseteq \boldsymbol{\mathcal{H}}.$$

4.5.3 Proof of Lemma 4.34. We consider two additional hitting times. For this purpose, let

$$\pi_{\mathrm{w}} = \frac{3g(n)}{\binom{n-1}{2}}, \quad \pi_{\mathrm{u}} = \frac{\ln(n) + 5g(n)}{\binom{n-1}{2}}$$

be the hyperedge inclusion probabilities for the partner window and for the upper bound. Consider the joint distribution $(\boldsymbol{E}_{\mathrm{p}}, \boldsymbol{S}_{+}, \boldsymbol{\Delta}_{\mathrm{w}}, \boldsymbol{S}_{\mathrm{u}}) \sim (\boldsymbol{E}_{\mathrm{p}}, \boldsymbol{S}_{+}, \boldsymbol{\Delta}_{\mathrm{w}}) \otimes \boldsymbol{S}_{\mathrm{u}}$, where $\boldsymbol{\Delta}_{\mathrm{w}}$ depends on $(\boldsymbol{E}_{\mathrm{p}}, \boldsymbol{S}_{+})$ only through $\boldsymbol{S}_{\mathrm{c}}$ and is thus determined by $(\boldsymbol{\Delta}_{\mathrm{w}}|\boldsymbol{S}_{\mathrm{c}} = S) \sim \mathrm{Bin}(\binom{n}{3} - S, \pi_{\mathrm{w}})$, and $\boldsymbol{S}_{\mathrm{u}}$ is determined by $\boldsymbol{S}_{\mathrm{u}} \sim \mathrm{Bin}(\binom{n}{3}, \pi_{\mathrm{u}})$. Let $\boldsymbol{S}_{\mathrm{w}} = \boldsymbol{S}_{\mathrm{c}} + \boldsymbol{\Delta}_{\mathrm{w}}$ be the end of the partner window, and further let $\boldsymbol{H}_{\mathrm{c}} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{\mathrm{c}}}$, $\boldsymbol{H}_{\mathrm{w}} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{\mathrm{w}}}$, $\boldsymbol{H}_{\mathrm{u}} = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_{\mathrm{u}}}$ be the hypergraphs corresponding to the hitting times. Let $\boldsymbol{\mathcal{D}}_{\mathrm{c}} = \mathcal{D}_{\mathrm{h}}(\boldsymbol{H}_{\mathrm{c}})$, $\boldsymbol{\mathcal{D}}_{\mathrm{w}} = \mathcal{D}_{\mathrm{h}}(\boldsymbol{H}_{\mathrm{w}})$ and $\boldsymbol{\mathcal{D}}_{\mathrm{u}} = \mathcal{D}_{\mathrm{h}}(\boldsymbol{H}_{\mathrm{u}})$ be the diamonds. Analogously to the hyperedges $\boldsymbol{\mathcal{H}}$ with partners in $(\boldsymbol{S}_{\mathrm{c}}, \boldsymbol{S}_{+}]$, let $\boldsymbol{\mathcal{H}}_{\mathrm{w}} = \{E \in \boldsymbol{H}_{\mathrm{c}}^{-1}(1) : \{E, E'\} \in \boldsymbol{\mathcal{D}}_{\mathrm{w}}, E' \in \boldsymbol{H}_{\mathrm{c}}^{-1}(0)\}$ be the hyperedges with partners in $(\boldsymbol{S}_{\mathrm{c}}, \boldsymbol{S}_{\mathrm{w}}]$.

Lemma 4.36. *Let $f : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$ with $f = \omega(1)$.*
a) We have $|\boldsymbol{S}_{\pm} - \frac{n}{3}(\ln(n) \pm g(n))| \leq \sqrt{f(n)n\ln(n)}$ whp.
b) We have $\frac{n}{3}(\ln(n) - g(n)) - \sqrt{f(n)n\ln(n)} \leq \boldsymbol{S}_{\mathrm{c}} \leq \frac{n}{3}(\ln(n) + g(n)) + \sqrt{f(n)n\ln(n)}$ whp.
c) We have $\frac{n}{3}(\ln(n) + 2g(n)) - \sqrt{f(n)n\ln(n)} \leq \boldsymbol{S}_{\mathrm{w}} \leq \frac{n}{3}(\ln(n) + 4g(n)) + \sqrt{f(n)n\ln(n)}$ and further $|\boldsymbol{\Delta}_{\mathrm{w}} - ng(n)| \leq \sqrt{f(n)ng(n)}$ whp.
d) We have $|\boldsymbol{S}_{\mathrm{u}} - \frac{n}{3}(\ln(n) + 5g(n))| \leq \sqrt{f(n)n\ln(n)}$ whp.
e) We have $\boldsymbol{S}_{-} < \boldsymbol{S}_{\mathrm{c}} \leq \boldsymbol{S}_{+} \leq \boldsymbol{S}_{\mathrm{w}} \leq \boldsymbol{S}_{\mathrm{u}}$ whp, $\boldsymbol{\mathcal{D}}_{\mathrm{c}} \subseteq \boldsymbol{\mathcal{D}}_{+} \subseteq \boldsymbol{\mathcal{D}}_{\mathrm{w}} \subseteq \boldsymbol{\mathcal{D}}_{\mathrm{u}}$ and $\boldsymbol{\mathcal{H}} \subseteq \boldsymbol{\mathcal{H}}_{\mathrm{w}}$ whp.
f) We have $\mathrm{vdj}(\boldsymbol{\mathcal{D}}_{\mathrm{c}}) = \mathrm{vdj}(\boldsymbol{\mathcal{D}}_{+}) = \mathrm{vdj}(\boldsymbol{\mathcal{D}}_{\mathrm{w}}) = \mathrm{vdj}(\boldsymbol{\mathcal{D}}_{\mathrm{u}}) = 1$ whp.

Proof. Let $B(n) = \sqrt{f(n)n\ln(n)}$ and $b(n) = \sqrt{f(n)ng(n)}$ be the bounds. For Part 4.36a) we have $\mathrm{Var}(\boldsymbol{S}_{\pm}) \leq \mathbb{E}[\boldsymbol{S}_{\pm}] = \frac{n}{3}(\ln(n) \pm g(n)) \leq \frac{2}{3}n\ln(n)$, $\mathbb{P}(|\boldsymbol{S}_{\pm} - \mathbb{E}[\boldsymbol{S}_{\pm}]| \geq B(n)) \leq \frac{2}{3f(n)}$ by Chebyshev's inequality. Part 4.36d) follows analogously. Lemma 4.29 gives $\boldsymbol{S}_{-} < \boldsymbol{S}_{\mathrm{c}} \leq \boldsymbol{S}_{+}$ whp in Part 4.36e), which yields Part 4.36b) using Part 4.36a). For Part 4.36c), let $S \in \mathbb{Z} \cap [\mathbb{E}[\boldsymbol{S}_{-}] - B(n), \mathbb{E}[\boldsymbol{S}_{+}] + B(n)]$ and $\boldsymbol{\Delta} \sim \mathrm{Bin}(\binom{n}{3} - S, \pi_{\mathrm{w}})$, i.e. $(\boldsymbol{\Delta}_{\mathrm{w}}|\boldsymbol{S}_{\mathrm{c}} = S) \sim \boldsymbol{\Delta}$. Using $\mathrm{Var}(\boldsymbol{\Delta}) \leq \mathbb{E}[\boldsymbol{\Delta}] = ng(n) - S\pi_{\mathrm{w}}$, Chebyshev's inequality yields $|\boldsymbol{\Delta} - \mathbb{E}[\boldsymbol{\Delta}]| \leq b'(n)$ whp for any $b'(n) = \omega(\sqrt{ng(n)})$ and hence $|\boldsymbol{\Delta} - ng(n)| < b(n)$ whp. This yields $|\boldsymbol{\Delta}_{\mathrm{w}} - ng(n)| \leq \sqrt{f(n)ng(n)}$ whp using Part 4.36b), and the remainder of Part 4.36c) follows with $\boldsymbol{S}_{-} + \boldsymbol{\Delta}_{\mathrm{w}} < \boldsymbol{S}_{\mathrm{w}} \leq \boldsymbol{S}_{+} + \boldsymbol{\Delta}_{\mathrm{w}}$ whp. The remainder of Part 4.36e) is now immediate, and Part 4.36f) then follows with Observation 4.12. □

For $S \in [\boldsymbol{S}_{\mathrm{c}}]$ let $\overline{\boldsymbol{\mathcal{P}}}(S) = \{E' \in \boldsymbol{H}_{\mathrm{c}}^{-1}(0) : \{\boldsymbol{E}_{\mathrm{p},S}, E'\} \in \overline{\mathcal{D}}, \forall E'' \in \boldsymbol{H}_{\mathrm{c}}^{-1}(1) \setminus \{\boldsymbol{E}_{\mathrm{p},S}\} \, |E' \cap E''| \leq 1\}$ be the potential new exclusive partners of $\boldsymbol{E}_{\mathrm{p},S}$, let $\boldsymbol{\mathcal{P}}(S) = \overline{\boldsymbol{\mathcal{P}}}(S) \cap \boldsymbol{H}_{\mathrm{w}}^{-1}(1)$ be the new exclusive partners in $\boldsymbol{H}_{\mathrm{w}}$, and let $\boldsymbol{R} = (\mathbb{1}\{|\boldsymbol{\mathcal{P}}(S)| = 1\})_{S \in [\boldsymbol{S}_{\mathrm{c}}]}$ be the indicator for meeting exactly one new exclusive partner. Let $\boldsymbol{P} = (|\overline{\boldsymbol{\mathcal{P}}}(S)|\pi_{\mathrm{w}}(1 - \pi_{\mathrm{w}})^{|\overline{\boldsymbol{\mathcal{P}}}(S)| - 1})_{S \in [\boldsymbol{S}_{\mathrm{c}}]}$ be the corresponding success probability.

Lemma 4.37. *We have $(\boldsymbol{R}|\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}) \sim \bigotimes_{s \in [\boldsymbol{S}_{\mathrm{c}}]} \mathrm{Bin}(1, \boldsymbol{P}(s))$.*

Proof. Notice that $\overline{\boldsymbol{\mathcal{P}}}(S) \subseteq \boldsymbol{H}_{\mathrm{c}}^{-1}(0)$, $S \in [\boldsymbol{S}_{\mathrm{c}}]$, is determined by $\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}$ and $\overline{\boldsymbol{\mathcal{P}}}(S_1) \cap \overline{\boldsymbol{\mathcal{P}}}(S_2) = \emptyset$ for $S \in [\boldsymbol{S}_{\mathrm{c}}]^{\underline{2}}$. Also notice that $\boldsymbol{H}_{\mathrm{w}}$ given $\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}$ has the law $\mathrm{Bin}(1, 1)^{\otimes \boldsymbol{H}_{\mathrm{c}}^{-1}(1)} \otimes \mathrm{Bin}(1, \pi_{\mathrm{w}})^{\otimes \boldsymbol{H}_{\mathrm{c}}^{-1}(0)}$ and hence $(\boldsymbol{N}|\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}) \sim \bigotimes_{S \in [\boldsymbol{S}_{\mathrm{c}}]} \mathrm{Bin}(|\overline{\boldsymbol{\mathcal{P}}}(S)|, \pi_{\mathrm{w}})$, where $\boldsymbol{N} = (|\boldsymbol{\mathcal{P}}(S)|)_S$. This completes the proof. □

Now, let $(\boldsymbol{E}_{\mathrm{p}}, \hat{\boldsymbol{E}}_{\mathrm{p}}, \boldsymbol{r}, \boldsymbol{S}_+) \sim (\boldsymbol{E}_{\mathrm{p}}, \hat{\boldsymbol{E}}_{\mathrm{p}}, \boldsymbol{r}) \otimes \boldsymbol{S}_+$ be given as follows. First, choose $\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}$ and set $\hat{\boldsymbol{E}}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]} = \boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}$. Next, using Lemma 4.37 let $(\boldsymbol{R}, \boldsymbol{r}_{[\boldsymbol{S}_{\mathrm{c}}]})|\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}$ be given by the componentwise maximal coupling, i.e. the pairs $(\boldsymbol{R}(s), \boldsymbol{r}(s))_s$ are independent with $\mathbb{P}(\boldsymbol{R}(s) \neq \boldsymbol{r}(s)|\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}) = |\boldsymbol{P}(s) - \pi_{\mathrm{r}}|$. Finally, let $(\boldsymbol{E}_{\mathrm{p}}, \hat{\boldsymbol{E}}_{\mathrm{p}}, \boldsymbol{r})$ be conditionally independent given $(\boldsymbol{E}_{\mathrm{p},[\boldsymbol{S}_{\mathrm{c}}]}, \boldsymbol{R}, \boldsymbol{r}_{[\boldsymbol{S}_{\mathrm{c}}]})$.

Lemma 4.38. *We have* $\boldsymbol{\mathcal{H}} \subseteq \boldsymbol{\mathcal{H}}_{\mathrm{w}} = \boldsymbol{E}_{\mathrm{p}}(\boldsymbol{R}^{-1}(1)) \subseteq \hat{\boldsymbol{E}}_{\mathrm{p}}(\boldsymbol{r}^{-1}(1))$ *whp.*

Proof. Lemma 4.36 yields $\boldsymbol{\mathcal{H}} \subseteq \boldsymbol{\mathcal{H}}_{\mathrm{w}} = \boldsymbol{E}_{\mathrm{p}}(\boldsymbol{R}^{-1}(1))$ whp. With $|\overline{\boldsymbol{\mathcal{P}}}(s)| \leq 3(n-3)$ we have $\boldsymbol{P} \leq \pi_{\mathrm{r}}$. □

Using $\boldsymbol{\mathcal{H}} \subseteq \boldsymbol{H}_{\mathrm{c}}^{-1}(1)$ and $\boldsymbol{S}_{\mathrm{c}} = \hat{\boldsymbol{S}}_{\mathrm{c}}$ gives $\boldsymbol{\mathcal{H}} \subseteq \hat{\boldsymbol{\mathcal{H}}}$ whp.

4.5.4 Proof of Lemma 4.35. Consider $(\hat{\boldsymbol{E}}_{\mathrm{p}}, \boldsymbol{r}, \hat{\boldsymbol{S}}_+, \hat{\boldsymbol{S}}_-) \sim (\hat{\boldsymbol{E}}_{\mathrm{p}}, \boldsymbol{r}) \otimes (\boldsymbol{S}_+, \boldsymbol{S}_-)$, $\hat{\boldsymbol{H}}_{\pm} = \boldsymbol{H}_{\mathrm{p},\hat{\boldsymbol{S}}_{\pm}}$, the removed hyperedges $\boldsymbol{\mathcal{R}}_+ = \{\hat{\boldsymbol{E}}_{\mathrm{p},S} : S \in \boldsymbol{r}^{-1}(1) \cap [\hat{\boldsymbol{S}}_+]\}$, the isolated vertices $\hat{\boldsymbol{\mathcal{I}}}_{\pm}$ and the low-degree vertices $\hat{\boldsymbol{\mathcal{L}}}_{\pm}$ from Lemma 4.29 for $\hat{\boldsymbol{H}}_{\pm}$.

Lemma 4.39. *Notice that the following holds.*
a) We have $\forall E \in \boldsymbol{\mathcal{R}}_+^2\, E_1 \cap E_2 = \emptyset$ *whp.*
b) We have $\hat{\boldsymbol{\mathcal{L}}}_+ \cap \bigcup_{E \in \boldsymbol{\mathcal{R}}_+} E = \emptyset$ *whp.*
c) We have $[n] \setminus \hat{\boldsymbol{\mathcal{I}}}_- \subseteq \bigcup_{E \in \boldsymbol{\mathcal{H}}} E$ *whp, where* $\boldsymbol{\mathcal{H}} = \hat{\boldsymbol{H}}_-^{-1}(1) \setminus \boldsymbol{\mathcal{R}}_+$.
d) We have $\hat{\boldsymbol{S}}_{\mathrm{c}} = \inf\{S : \bigcup_{E \in \hat{\boldsymbol{E}}_{\mathrm{p}}([S] \cap \boldsymbol{r}^{-1}(0))} E = [n]\}$ *whp.*

Proof. For Part 4.39a), we consider $\boldsymbol{N}_r = |\{E \in \boldsymbol{\mathcal{R}}_+^2 : |E_1 \cap E_2| = r\}|$ for $r \in [2]$ and notice that $\mathbb{P}(\exists E \in \boldsymbol{\mathcal{R}}_+^2\, E_1 \cap E_2 \neq \emptyset) \leq \mathbb{E}[\boldsymbol{N}_1] + \mathbb{E}[\boldsymbol{N}_2] = \mathcal{O}(\ln(n)^2 g(n)^2/n)$. For Part 4.39b), we have $\mathbb{P}(\hat{\boldsymbol{\mathcal{L}}}_+ \cap \bigcup_{E \in \boldsymbol{\mathcal{R}}_+} E \neq \emptyset) \leq \mathbb{E}[\boldsymbol{N}] = 3\binom{n}{3}\pi_+\pi_{\mathrm{r}}\mathbb{P}(\boldsymbol{b} < 6g(n))$, where $\boldsymbol{b} \sim \mathrm{Bin}(\binom{n-1}{2} - 1, \pi_+)$ and $\boldsymbol{N} = |\{(v, E) \in \hat{\boldsymbol{\mathcal{L}}}_+ \times \boldsymbol{\mathcal{R}}_+ : v \in E\}|$. As in the proof of Lemma 4.29 we obtain $\mathbb{P}(\boldsymbol{b} < 6g(n)) \leq n^{-1+o(1)}$ and thereby $\mathbb{E}[\boldsymbol{N}] \leq n^{-1+o(1)}$. For Part 4.39c), let $\boldsymbol{\mathcal{V}} = ([n] \setminus \hat{\boldsymbol{\mathcal{I}}}_-) \setminus (\bigcup_{E \in \boldsymbol{\mathcal{H}}} E)$ and notice that by Part 4.39a) it holds whp that for all $v \in \boldsymbol{\mathcal{V}}$ there exists exactly one $E \in \hat{\boldsymbol{H}}_-^{-1}(1)$ with $v \in E$, for which we have $E \in \boldsymbol{\mathcal{R}}_-$. So, with $\boldsymbol{N} = |\{(v, E) \in [n] \times \boldsymbol{\mathcal{R}}_- : v \in E, D_{\hat{\boldsymbol{H}}_-}(v) = 1\}|$ and Part 4.39a) we have $\mathbb{P}(\boldsymbol{\mathcal{V}} \neq \emptyset) \leq \mathbb{E}[\boldsymbol{N}] + o(1) = o(1)$. Now, Part 4.39d) follows with Lemma 4.29, since whp for all $S \in \mathbb{Z} \cap [\boldsymbol{S}_-, \boldsymbol{S}_+]$ we have

$$\bigcup_{E \in \hat{\boldsymbol{H}}_{\mathrm{p},S}^{-1}(1)} E = \bigcup_{E \in \hat{\boldsymbol{H}}_{\mathrm{p},S}^{-1}(1) \setminus \boldsymbol{\mathcal{R}}_+} E.$$

□

Now, given $\boldsymbol{J} = (\hat{\boldsymbol{E}}_{\mathrm{p}}, \boldsymbol{r}, \hat{\boldsymbol{S}}_+, \hat{\boldsymbol{S}}_-)$ let $\boldsymbol{\eta}$ be the enumeration of $\boldsymbol{r}^{-1}(0)$, further let $\tilde{\boldsymbol{E}}_{\mathrm{p},[|\boldsymbol{r}^{-1}(0)|]} = (\hat{\boldsymbol{E}}_{\mathrm{p},\boldsymbol{\eta}(S)})_S$ be the subprocess obtained from the restriction to $\boldsymbol{r}^{-1}(0)$ and let $\tilde{\boldsymbol{E}}_{\mathrm{p}}$ depend on $\boldsymbol{J}$ only through $\tilde{\boldsymbol{E}}_{\mathrm{p},[|\boldsymbol{r}^{-1}(0)|]}$. Lemma 4.39 suggests that $\boldsymbol{\eta}(\tilde{\boldsymbol{S}}_{\mathrm{c}}) = \hat{\boldsymbol{S}}_{\mathrm{c}}$ whp, which completes the proof.

4.5.5 Proof of Proposition 4.32. With Lemma 4.33, Lemma 4.34 and Lemma 4.35 we have $\boldsymbol{\mathcal{H}}_\circ \subseteq \boldsymbol{\mathcal{H}} \subseteq \hat{\boldsymbol{\mathcal{H}}}$ and $\tilde{\boldsymbol{H}}_{\mathrm{p},\tilde{\boldsymbol{S}}_{\mathrm{c}}}^{-1}(1) \subseteq \hat{\boldsymbol{H}}_{\mathrm{p},\hat{\boldsymbol{S}}_{\mathrm{c}}}^{-1}(1) \setminus \hat{\boldsymbol{\mathcal{H}}} \subseteq \hat{\boldsymbol{H}}_{\mathrm{p},\hat{\boldsymbol{S}}_{\mathrm{c}}}^{-1}(1) \setminus \boldsymbol{\mathcal{H}}_\circ$ whp, by gluing the couplings.

4.5.6 Proof of Theorem 4.2. Theorem 4.2 is immediate from Lemma 4.31, Proposition 4.32 and Theorem 2.6.

4.5.7 Proof of Corollary 4.3. This result follows from Theorem 4.2 with Lemma 4.31.

4.5.8 Proof of Corollary 4.4. We obtain Part 4.4a) by reducing it to the case $n\in 3\mathbb{Z}_{>0}$ as follows. Let $n\in\mathbb{Z}_{\geq 12}$, $n^\circ=3\lfloor n\rfloor$ and $\Delta=n-n^\circ$. let $\Delta\in\{1,2\}$. Further, let $(\boldsymbol{H}_+,\boldsymbol{E}_*)\sim\boldsymbol{H}_{\mathrm{b},\pi_+}\otimes\mathrm{u}(\binom{[n]}{3}!)$ and let $\boldsymbol{H}_*\in\{0,1\}^{\binom{[n]}{3}}$ be given by $\boldsymbol{H}_*^{-1}(1)=\boldsymbol{H}_+^{-1}(1)\cup\{\boldsymbol{E}_*\}$.

Lemma 4.40. *We have* $\|\boldsymbol{H}_+-\boldsymbol{H}_*\|_{\mathrm{tv}}=o(1)$.

Proof. We have $\|\boldsymbol{H}_+-\boldsymbol{H}_*\|_{\mathrm{tv}}=\frac{1}{2}\sum_H|\pi_+^{|H^{-1}(1)|}(1-\pi_+)^{|H^{-1}(0)|}-\frac{|H^{-1}(1)|}{\binom{n}{3}}\pi_+^{|H^{-1}(1)|-1}(1-\pi_+)^{|H^{-1}(0)|}|$ and using Jensen's inequality with $\boldsymbol{M}=|\boldsymbol{H}_+^{-1}(1)|\sim\mathrm{Bin}(\binom{n}{3},\pi_+)$ further

$$\|\boldsymbol{H}_+-\boldsymbol{H}_*\|_{\mathrm{tv}}=\frac{\mathbb{E}[|\mathbb{E}[\boldsymbol{M}]-\boldsymbol{M}|]}{2\mathbb{E}[\boldsymbol{M}]}\leq\frac{\sqrt{\mathrm{Var}(\boldsymbol{M})}}{2\mathbb{E}[\boldsymbol{M}]}\leq\frac{1}{2\sqrt{\mathbb{E}[\boldsymbol{M}]}}=\mathcal{O}((n\ln(n))^{-1/2}).$$

□

For notational transparency let $\boldsymbol{E}_{\mathrm{p}}$ be the process on the hypergraph with $\boldsymbol{E}_*$ planted, given by $(\boldsymbol{E}_{\mathrm{p}}|\boldsymbol{H}_+,\boldsymbol{E}_*)\sim\mathrm{u}(\boldsymbol{H}_*^{-1}(1)!)$ and let $\boldsymbol{S}_*=|\boldsymbol{H}_*^{-1}(1)|$. Further, let $\boldsymbol{\mathcal{V}}_\Delta$ be given by $(\boldsymbol{\mathcal{V}}_\Delta|\boldsymbol{H}_+,\boldsymbol{E}_*,\boldsymbol{E}_{\mathrm{p}})\sim\mathrm{u}(\binom{\boldsymbol{E}_*}{\Delta})$. Let $\boldsymbol{\mathcal{V}}^\circ=[n]\setminus\boldsymbol{\mathcal{V}}_\Delta$, enumerated by $\boldsymbol{\eta}_{\mathrm{v}}^\circ$, and let $\boldsymbol{H}_\circ\in\{0,1\}^{\binom{[n^\circ]}{3}}$ be given by

$$\boldsymbol{H}_\circ^{-1}(1)=\left\{E\in\binom{[n^\circ]}{3}:\boldsymbol{H}_*(\boldsymbol{\eta}_{\mathrm{v}}^\circ(E))=1\right\},$$

that is, we obtain $\boldsymbol{H}_\circ$ from $\boldsymbol{H}_*$ by removing the Δ vertices $\boldsymbol{\mathcal{V}}_\Delta\subseteq\boldsymbol{E}_*$ (and all incident hyperedges) and relabeling the remaining vertices. Notice that $\boldsymbol{H}_{*,\binom{\boldsymbol{\mathcal{V}}^\circ}{3}}\sim\mathrm{Bin}(1,\pi_+)^{\otimes\binom{\boldsymbol{\mathcal{V}}^\circ}{3}}$ and hence $\boldsymbol{H}_\circ\sim\boldsymbol{H}_{\mathrm{b},n^\circ,\pi_{+,n}}$. Further, notice that $\pi_{+,n}=\frac{\ln(n^\circ)+g'(n^\circ)}{\binom{n^\circ-1}{2}}$ for some suitable g'. Let $\boldsymbol{\eta}_{\mathrm{p}}^\circ$ be the enumeration of $\boldsymbol{E}_{\mathrm{p}}^{-1}(\boldsymbol{H}_\circ^{-1}(1))$ and $\boldsymbol{E}_{\mathrm{p}}^\circ=\boldsymbol{E}_{\mathrm{p}}\circ\boldsymbol{\eta}_{\mathrm{p}}^\circ$, i.e. we consider the subprocess on the remaining hyperedges. Notice that $\boldsymbol{E}_{\mathrm{p}}^\circ\sim\tilde{\boldsymbol{E}}_{\mathrm{p},[\tilde{\boldsymbol{S}}]}$ with $(\tilde{\boldsymbol{E}}_{\mathrm{p}},\tilde{\boldsymbol{S}})\sim\mathrm{u}(\binom{[n_\circ]}{3}!)\otimes\mathrm{Bin}(\binom{n^\circ}{3},\pi_{+,n})$ is a standard stopped process. Thus, Theorem 2.6 yields that $\boldsymbol{S}_{\mathrm{c}}^\circ=\boldsymbol{S}_{\mathrm{f}}^\circ\leq\boldsymbol{S}_+^\circ$ whp, where $\boldsymbol{S}_+^\circ=|\boldsymbol{H}_\circ^{-1}(1)|$ and using the notions for $\boldsymbol{E}_{\mathrm{p}}^\circ$ corresponding to $\boldsymbol{E}_{\mathrm{p}}$, e.g. the hitting times $\boldsymbol{S}_{\mathrm{c}}^\circ$, $\boldsymbol{S}_{\mathrm{f}}^\circ$. Further, we clearly have $\boldsymbol{H}_{\mathrm{p},S}^\circ\leq\boldsymbol{H}_{\mathrm{p},\boldsymbol{\eta}_{\mathrm{p}}^\circ(S)}$ and hence $\boldsymbol{S}_{\mathrm{c}}\leq\boldsymbol{\eta}_{\mathrm{p}}^\circ(\boldsymbol{S}_{\mathrm{c}})$. Now, let $\boldsymbol{\mathcal{L}}_*=\{v\in[n]:D_{\boldsymbol{H}_*}(v)\leq 6g(n)\}$ be the low-degree vertices and let $\boldsymbol{\mathcal{H}}=\{E\in\boldsymbol{H}_*^{-1}(1):E\cap\boldsymbol{\mathcal{V}}_\Delta\neq\emptyset\}$ be the hyperedges incident to $\boldsymbol{\mathcal{V}}_\Delta$.

Lemma 4.41. *Notice that the following holds.*
a) We have $\boldsymbol{\mathcal{L}}_*\cap\bigcup_{E\in\boldsymbol{\mathcal{H}}}E=\emptyset$ *whp.*
b) We have $\boldsymbol{S}_{\mathrm{c}}=\boldsymbol{S}_{\mathrm{f}}^*=\boldsymbol{\eta}_{\mathrm{p}}^\circ(\boldsymbol{S}_{\mathrm{c}})$ *whp.*

Proof. For Part 4.41a), we consider $\boldsymbol{N}=|\{(v,E)\in\boldsymbol{\mathcal{L}}_*\times\boldsymbol{\mathcal{H}}:v\in E\setminus\boldsymbol{E}^*\}|$, then the union bound yields $\mathbb{P}(\boldsymbol{\mathcal{L}}_*\cap\bigcup_{E\in\boldsymbol{\mathcal{H}}}E\neq\emptyset)\leq\mathbb{E}[|\boldsymbol{\mathcal{L}}_*\cap\boldsymbol{E}_*|]+\mathbb{E}[\boldsymbol{N}]$. We have $\mathbb{E}[|\boldsymbol{\mathcal{L}}_*\cap\boldsymbol{E}_*|]=3\mathbb{P}(\boldsymbol{b}<6g(n))=n^{-1+o(1)}$ with $\boldsymbol{b}\sim\mathrm{Bin}(\binom{n-1}{2}-1,\pi_+)$, as in the proof of Lemma 4.29, Further, we have

$$\mathbb{E}[\boldsymbol{N}]\leq\sum_{r=1}^{\Delta}\binom{\Delta}{r}\binom{n-\Delta}{3-r}\pi_+\binom{3}{1}\mathbb{P}(\boldsymbol{b}<6g(n))\leq n^{-1+o(1)}$$

and thereby Part 4.41a) holds. For Part 4.41b), let $\boldsymbol{S}_+=|\boldsymbol{H}_+^{-1}(1)|$ and let $\boldsymbol{S}_-$ be given by $(\boldsymbol{S}_-|\boldsymbol{H}_+,\boldsymbol{E}_*,\boldsymbol{E}_{\mathrm{p}},\boldsymbol{\mathcal{V}}_\Delta)\sim\mathrm{Bin}(\boldsymbol{S}_+,\pi_-/\pi_+)$. Further, let $\boldsymbol{H}_-\in\{0,1\}^{\binom{[n]}{3}}$ be given by $\boldsymbol{H}_-^{-1}(1)=\boldsymbol{E}_{\mathrm{p}}([\boldsymbol{S}_-+1])\setminus\{\boldsymbol{E}_*\}$ on the event $\{\boldsymbol{H}_+(\boldsymbol{E}_*)=0,\boldsymbol{E}_{\mathrm{p}}^{-1}(\boldsymbol{E}_*)\leq\boldsymbol{S}_-\}$, and $\boldsymbol{H}_-^{-1}(1)=\boldsymbol{E}_{\mathrm{p}}([\boldsymbol{S}_-])$ otherwise. Finally, let $\boldsymbol{\mathcal{I}}_-=\{v\in[n]:D_{\boldsymbol{H}_-}(v)=0\}$ be the isolated vertices. Notice that $\boldsymbol{H}_-\sim\boldsymbol{H}_{\mathrm{b},\pi_-}$, and thus the proof of Lemma 4.29 gives $\mathbb{E}[|\boldsymbol{\mathcal{I}}_-|] = (1+o(1))e^{g(n)}$ and $\mathrm{Var}(|\boldsymbol{\mathcal{I}}_-|) = (1+o(1))\mathbb{E}[|\boldsymbol{\mathcal{I}}_-|]$. Hence, Chebyshev's inequality yields $|\boldsymbol{\mathcal{I}}_-| > \frac{1}{2}e^{g(n)} = \omega(1)$ whp, which shows that $\boldsymbol{\mathcal{I}}' \neq \emptyset$ whp, where $\boldsymbol{\mathcal{I}}' = \{v \in [n] : D_{\boldsymbol{H}'}(v) = 0\} = \boldsymbol{\mathcal{I}}_- \setminus \boldsymbol{E}_*$ and $\boldsymbol{H}' \in \{0,1\}^{\binom{[n]}{3}}$ is given by $\boldsymbol{H}'^{-1}(1) = \boldsymbol{H}_-^{-1}(1) \cup \{\boldsymbol{E}_*\}$. Notice that $\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_-} \leq \boldsymbol{H}'$ and hence there exist isolated vertices in $\boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_-}$ whp. Next, we show that $\mathcal{E} = \{\boldsymbol{H}_+(\boldsymbol{E}_*) = 0, \boldsymbol{E}_{\mathrm{p}}^{-1}(\boldsymbol{E}_*) \leq \boldsymbol{S}_-\}$ whp. We have $\mathbb{P}(\boldsymbol{H}_+(\boldsymbol{E}_*) = 1) = \pi_+$ and using Lemma 4.36 further

$$\mathbb{P}(\boldsymbol{H}_+(\boldsymbol{E}_*) = 0, \boldsymbol{E}_{\mathrm{p}}^{-1}(\boldsymbol{E}_*) \leq \boldsymbol{S}_-) = \mathbb{E}\left[\frac{\mathbb{1}\{\boldsymbol{H}_+(\boldsymbol{E}_*) = 0\}\boldsymbol{S}_-}{\boldsymbol{S}_+ + 1}\right] = \mathbb{E}\left[\frac{\mathbb{1}\{\boldsymbol{H}_+(\boldsymbol{E}_*) = 0\}\pi_-\boldsymbol{S}_+}{\pi_+(\boldsymbol{S}_+ + 1)}\right]$$
$$\geq \frac{\pi_-}{\pi_+} - \pi_+ - \mathbb{E}\left[\frac{1}{\boldsymbol{S}_+ + 1}\right] = 1 + o(1).$$

Thus, we know that $\boldsymbol{H}' = \boldsymbol{H}_{\mathrm{p},\boldsymbol{S}_-+1}$ whp. Analogously to Lemma 4.29 we obtain that $\boldsymbol{\mathcal{I}}' \subseteq \boldsymbol{\mathcal{L}}_*$ whp, using $\mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\mathcal{L}}_* \setminus \boldsymbol{\mathcal{I}}'|] = \mathbb{E}[\mathbb{1}\mathcal{E}|\boldsymbol{\mathcal{I}}'|]\mathbb{P}(\boldsymbol{b} > 6g(n)) \leq \mathbb{E}[|\boldsymbol{\mathcal{I}}_-|]\mathbb{P}(\boldsymbol{b} > 6g(n))$ with $\boldsymbol{b} \sim \mathrm{Bin}(\binom{n-1}{2}, \frac{\pi_+ - \pi_-}{1-\pi_-})$. With Part 4.41a) the last isolated vertices $\boldsymbol{\mathcal{I}}'$ are not incident to $\boldsymbol{\mathcal{H}}$ and hence $\boldsymbol{S}_{\mathrm{c}} = \boldsymbol{\eta}_{\mathrm{p}}^{\circ}(\boldsymbol{S}_{\mathrm{c}}^{\circ})$ whp. Finally, with $\boldsymbol{S}_{\mathrm{f}}^{\circ} = \boldsymbol{S}_{\mathrm{c}}^{\circ}$ whp we have a perfect matching $F_\circ$ in $\boldsymbol{H}_\circ$, which induces a matching $F_* \subseteq \boldsymbol{H}_*^{-1}(1)$ with $\bigcup_{E \in F_*} E = \boldsymbol{\mathcal{V}}^{\circ}$, so for $F = F_* \cup \{\boldsymbol{E}_*\}$ we have $|F| = \lceil n/3 \rceil$ and $\bigcup_{E \in F} E = [n]$. □

Part 4.4a) follows from Lemma 4.40 combined with Lemma 4.41. This extends Theorem 2.6. Part 4.4b) is now immediate from Theorem 4.2, since we embed the hitting time hypergraph and thus obtain the desired cover.

5 Occupation Problems on Regular Graphs

In this section we provide further context and the proofs for the results presented in Section 2.3. First, we present applications of occupation problems, other contributions and open problems in Section 5.1, before we turn to the proofs. In Section 5.2 we give a detailed overview of our procedure. Before we turn to the proof, we clarify some preliminaries, prepare arguments and discuss the translation to different models in Section 5.3. Then, the first step is the first moment method, which we implement in Section 5.4. The second step is the second moment method, which we implement in Section 5.5. Crucially, this is the part where we establish the conjecture from [101]. In Section 5.6, we boost the probability obtained using the second moment to 1 using small subgraph conditioning, based on the proof of Lemma 5.8 in Section 5.7.

5.1 Applications, Results and Problems

We take a closer look at the examples from [36, 96] in Section 5.1.1, then we discuss related work in Section 5.1.2 and open problems in Section 5.1.3.

5.1.1 Examples and Related Problems. A problem that is closely related and can be reduced to the d-regular r-in-k occupation problem is the d-regular positive r-in-k SAT problem, a variant of k-SAT introduced above. In this case, we consider a boolean formula

$$f = \bigwedge_{a \in \mathcal{F}} c_a, \quad c_a = \bigvee_{i \in v(a)} i,\, a \in [m],$$

in conjunctive normal form with m clauses over n variables $i \in [n]$, such that no literal appears negated (hence *positive* r-in-k SAT), and where each clause c_a is the disjunction of k literals and each variable appears in exactly d clauses (hence d-regular). The decision problem is to determine if there exists an assignment x such that exactly r literals in each clause evaluate to true (hence r-in-k SAT). In [96] the satisfiability threshold for this problem was determined for $r = 1$, i.e. the case where exactly one literal in each clause evaluates to true. Our Theorem 2.10 solves this problem for $r = 2$ and $k \in \mathbb{Z}_{\geq 4}$.

Our second example deals with a prominent problem related to graph theory. A k-regular d-uniform hypergraph H is a pair $H = ([m], \mathcal{E})$ with vertices $[m]$ and $n = |\mathcal{E}|$ (hyper-)edges such that each edge contains d vertices and the degree of each vertex is k. An r-factor $\mathcal{F} \subseteq \mathcal{E}$ is a subset of the hyperedges such that each vertex $a \in [m]$ is incident to r hyperedges $e_i \in \mathcal{F}$. In this case the problem is to determine if H has an r-factor. For example, the case $r = 1$ is the well-known perfect matching problem and the threshold was determined in [36]. An example of a 2-factor in a hypergraph is shown in Figure 2b. Theorem 2.10 solves also this problem for $r = 2$ and $k \in \mathbb{Z}_{\geq 4}$.

There are several other problems in complexity and graph theory that are closely related to the examples above. The satisfiability threshold in Theorem 2.10 also applies to a variant of the vertex cover problem (or hitting set problem from set theory perspective), where we choose a subset of the vertices (variables with value one) in a d-regular k-uniform hypergraph such that each hyperedge is incident to exactly two vertices in the subset. Analogously, Theorem 2.10 also establishes the threshold for a variant of the set cover problem in set theory corresponding to 2-factors in hypergraphs, i.e. given a family of d-subsets (hyperedges) and a universe (vertices) with each element contained in k subsets, the problem is to find a subfamily of the subsets such that each element of the universe is contained in exactly two subsets of the subfamily. Further, Theorem 2.10 can e.g. also be used to give sufficient conditions for the (asymptotic) existence of Euler families in regular uniform hypergraphs as discussed in [15].

5.1.2 Related Work. The regular version of the random 1-in-k occupation problem (and related problems) has been studied in [36, 96] using the first and second moment method with small subgraph conditioning. The paper [114] shows that $\lim_{i\to\infty} \mathbb{P}(\boldsymbol{Z} > 0) = 1$ for $d = 2$ and $k \in \mathbb{Z}_{\geq 2}$ in the d-regular 2-in-k occupation problem, i.e. the existence of 2-factors in k-regular simple graphs. A recent discussion of 2-factors (and the related Euler families) that does not rely on the probabilistic method is presented in [15]. Further, randomized polynomial time algorithms for the generation and approximate counting of 2-factors in random regular graphs have been developed in [56].

The study of Erdős Rényi (hyper-)graphs was initiated by the groundbreaking paper [47] in 1960 and turned into a fruitful field of research with many applications, including early results on 1-factors in simple graphs [49]. On the contrary, results for the random d-regular k-uniform (hyper-)graph ensemble were rare before the introduction of the configuration (or pairing) model by Bollobás [22] and the development of the small subgraph conditioning method [67, 68]. While the proof scheme facilitated rigorous arguments to establish the existence and location of satisfiability thresholds of random regular CSPs [89, 17, 73, 29, 34, 41, 42], the problems are treated on a case by case basis, while results on entire classes of random regular CSPs are still outstanding.

One of the main reasons responsible for the complexity of a rigorous analysis of random (regular) CSPs seems to be a conjectured structural change of the solution space for increasing densities. This hypothesis has been put forward by physicists, verified in parts and mostly for ER ensembles, further led to new rigorous proof techniques [40, 34, 30] and to randomized algorithms [25, 83] for NP-hard problems that are not only of great value in practice, but can also be employed for precise numerical (though non-rigorous) estimates of satisfiability thresholds. An excellent introduction to this replica theory can be found in [85, 77, 122]. Specifically, numerical results indicating the satisfiability thresholds for d-regular r-in-k occupation problems (more general variants, and for ER type hypergraphs) based on this conjecture were discussed in various publications [27, 37, 119, 58, 65, 128, 127], where occupation problems were introduced for the first time in [98].

Another fundamental obstacle in the rigorous analysis is of a very technical nature and directly related to the second moment method as discussed in detail in our current work. In the case of regular 2-in-k occupation problems (amongst others) this optimization problem can be solved by exploiting a connection to the fixed points of belief propagation. This well-studied message passing algorithm is thoroughly discussed in [85].

5.1.3 Open Problems. In this work we rigorously establish the threshold for $r = 2$ and $k \in \mathbb{Z}_{\geq 4}$ for the random regular r-in-k occupation problem. A rigorous proof for general r (and k) seems to be involved, but further assumptions may significantly simplify the analysis. For example, as an extension of the current work one may focus on r-in-$2r$ occupation problems, where the constraints are symmetric in the colors. As can be seen from our proof, this yields useful symmetry properties. Further, as suggested by the literature [30, 33] such balanced problems [127, 128] are usually more accessible to a rigorous study. On the other hand, the optimization usually also significantly simplifies if only carried out for $k \geq k_0(r)$ for some (large) $k_0(r)$.

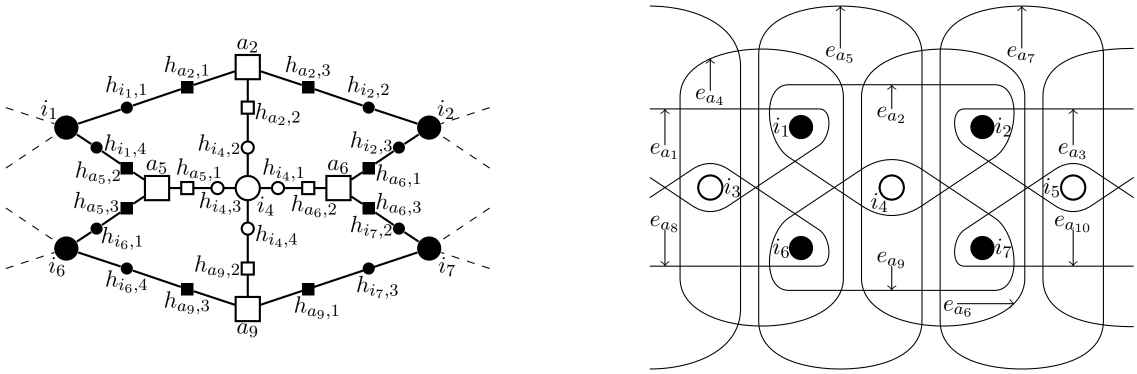
Apart from the generalizations discussed above, results for the general r-in-k occupation problems are also still outstanding for Erdős-Rényi type CSPs, the only exception being the satisfiability threshold for perfect matchings which was recently established by Kahn [71]. Further, there only exist bounds for the exact cover problem [72] on 3-uniform hypergraphs, i.e. $r = 1$ and $k = 3$.

Outline of the Proofs. In Section 5.2 we present the proof strategy on a high level. Then, we turn to the notation and do some groundwork, in particular the analysis of $d^*(k)$, in Section 5.3. The easy part of the main result is established in Section 5.4 using the first moment method. The remainder is devoted to the proof that solutions exist below the threshold with probability tending to one, starting with the second moment method in Section 5.5. Most of the twenty pages in this section are devoted to the solution of the optimization problem and related conjecture from [101] using a belief propagation inspired approach. Finally, we complete the small subgraph conditioning method in Section 5.6, using the proof of Lemma 5.8 in Section 5.7 as a blueprint.

5.2 Proof Techniques

In this section we give a high-level overview of our proof. We make heavy use of the so-called configuration model for the generation of random instances in the form used by Moore [96].

5.2.1 The Configuration Model. Working with the uniform distribution on d-regular k-uniform hypergraphs directly is challenging. Instead, we show Theorem 2.10 for occupation problems on so-called configurations. A d-regular k-configuration is a bijection $G : [n] \times [d] \to [m] \times [k]$, where the v-edges $(i, h') \in [n] \times [d]$ represent pairs of variables $i \in [n]$ and so-called i-edges, i.e. half-edge indices $h' \in [d]$. The image $(a, h) = G(i, h')$ is an f-edge, i.e. a pair of a constraint (factor) $a \in [m]$ and an a-edge (or half-edge) $h \in [k]$, indicating that the i-edge h' of the variable i is wired to the a-edge h of a and thereby suggesting that i is connected to a in the corresponding d-regular k-factor graph. Notice that we can represent G by an equivalent, four-partite, graph with (disjoint) vertex sets given by the variables $[n]$, constraints (factors) $[m]$, v-edges $\mathcal{H}' = [n] \times [d]$ and f-edges $\mathcal{H} = [m] \times [k]$, where



(a) occupation problem on configurations (b) 4-regular 2-in-3 vertex cover

Figure 5: The figure on the left shows the solution on a configuration corresponding to the solution in Figure 2. We only denoted a-edges (small boxes, filled if they the a-edge takes the value one) and i-edges (small circles, filled if the i-edge takes the value one) instead of f-edges and v-edges for brevity (e.g. $h_{a_1,1}$ instead of $(a_1, h_{a_1,1})$). The figure on the right illustrates the corresponding 2-in-3 vertex cover (given by the filled circles).

each variable $i \in [n]$ connects to all its v-edges $(i, h') \in \mathcal{H}'$, each constraint $a \in [m]$ to all its f-edges $(a, h) \in \mathcal{H}$ and a v-edge (i, h') connects to an f-edge (a, h) if $G(i, h') = (a, h)$.

Let $\mathcal{G}_\mathrm{c} = \mathcal{G}_{\mathrm{c},k,d,n,m}$ be the set of all d-regular k-configurations on n variables, and notice that $|\mathcal{G}_\mathrm{c}| = \emptyset$ iff $dn \neq km$ and $|\mathcal{G}_\mathrm{c}| = (dn)! = (km)!$ for $m = dn/k \in \mathbb{Z}$, which we assume from here on. Further, the occupation problem on factor graphs directly translates to configurations, i.e. an assignment $x \in \{0,1\}^n$ is a solution of $G \in \mathcal{G}_\mathrm{c}$ if for each constraint $a \in [m]$ there exist exactly two distinct a-edges h, $h' \in [k]$ such that $x_{i(a,h)} = x_{i(a,h')} = 1$, where $i(a,h) = (G^{-1}(a,h))_1$ denotes the h-th neighbor of a. Say, the occupation problem on a configuration corresponding to the example in Figure 2a is shown in Figure 5a.

Let $Z(G)$ be the number of solutions of $G \in \mathcal{G}_\mathrm{c}$, let $\boldsymbol{G}_\mathrm{c} \sim \mathrm{u}(\mathcal{G}_\mathrm{c})$ be the uniformly random configuration and $\boldsymbol{Z}_\mathrm{c} = Z(\boldsymbol{G}_\mathrm{c})$. As before, $\boldsymbol{Z}_\mathrm{c} = 0$ almost surely unless $2n \in k\mathbb{Z}$. Theorem 2.10 will be a straightforward consequence of the following result.

Theorem 5.1. *Theorem 2.10 also holds for* $\boldsymbol{Z}_\mathrm{c}$.

5.2.2 The First Moment Method. In the first step we apply the first moment method to the occupation problem on configurations, yielding the following result.

Lemma 5.2. *Let* $k \in \mathbb{Z}_{\geq 4}$, $d \in \mathbb{Z}_{\geq 2}$. *For* $n \in \mathcal{N}$ *tending to infinity*

$$\mathbb{E}[\boldsymbol{Z}_\mathrm{c}] = (1 + o(1))\sqrt{d}e^{n\phi_1}, \quad \textit{where} \quad \phi_1 = \frac{d}{k}\ln\binom{k}{2} - (d-1)H\left(\frac{2}{k}\right).$$

In particular, $\mathbb{E}[\boldsymbol{Z}_\mathrm{c}] \to \infty$ for $d < d^*$ and $\mathbb{E}[\boldsymbol{Z}_\mathrm{c}] \to 0$ for $d > d^*$ with d^* as in (2). So, Markov's inequality implies $\mathbb{P}(\boldsymbol{Z}_\mathrm{c} > 0) \to 0$ for $d > d^*$. The map ϕ_1 is known as annealed free entropy density.

5.2.3 The Second Moment Method. Let $k \in \mathbb{Z}_{\geq 4}$ and $d \in \mathbb{Z}_{\geq 2}$. We denote the set of distributions on a finite set $\mathcal{S}$ by $\mathcal{P}(\mathcal{S})$ and identify $p \in \mathcal{P}(\mathcal{S})$ with its probability mass function, meaning $\mathcal{P}(\mathcal{S}) =$

$\{p \in [0,1]^{\mathcal{S}} : \sum_{x\in\mathcal{S}} p(s) = 1\}$. Further, let $\mathcal{P}_\ell(\mathcal{S}) = \{p \in \mathcal{P}(\mathcal{S}) : \ell p \in \mathbb{Z}^{\mathcal{S}}\}$ be the empirical distributions over $\ell \in \mathbb{Z}_{>0}$ trials.

In order to apply the second moment method we will consider a (new) CSP with m factors on n variables with the larger domain $\{0,1\}^2$, and where the constraint $a \in [m]$ is satisfied by an assignment $x \in (\{0,1\}^2)^n$ if $\sum_{i\in v(a)} x_{i,1} = \sum_{i\in v(a)} x_{i,2} = 2$. Here, there are qualitatively three types of satisfying assignments for the constraints, namely with 0, 1 or 2 overlapping ones. We will analyze the empirical overlap distributions $p \in \mathcal{P}_m(\{0,1,2\})$ of assignments satisfying all constraints, which determine the empirical distributions $p_{\mathrm{e}} \in \mathcal{P}_{km}(\{0,1\}^2)$ of the values $\{0,1\}^2$ over the km edges, given by

$$p_{\mathrm{e}}(11) = \frac{1}{k}p(1) + \frac{2}{k}p(2) \quad \text{and} \quad p_{\mathrm{e}}(10) = p_{\mathrm{e}}(01) = \frac{1}{k}p(1) + \frac{2}{k}p(0).$$

So, if $p \in \mathcal{P}_m(\{0,1,2\})$ is an *achievable* empirical overlap distribution on the m factors, then p_{e} is necessarily an empirical distribution on the n variables; thus the achievable overlap distributions are contained in $\mathcal{P}_n = \{p \in \mathcal{P}_m(\{0,1,2\}) : p_{\mathrm{e}} \in \mathcal{P}_n(\{0,1\}^2)\}$.

In the first – combinatorial – part we establish that the second moment for $n \in \mathcal{N}$ can be written as a sum of contributions for fixed overlap distributions.

Lemma 5.3. *For any $n \in \mathcal{N}$ we have*

$$\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^2] = \sum_{p\in\mathcal{P}_n} E(p), \ \textit{where}\ E(p) = \binom{m}{mp} \prod_{s\in\{0,1,2\}} \binom{k}{s, 2-s, 2-s, k-4+s}^{mp(s)} \binom{n}{np_{\mathrm{e}}} \binom{dn}{dnp_{\mathrm{e}}}^{-1}.$$

Here, we use the notation $\binom{m}{mp}$, $p \in \mathcal{P}_m(\{0,1,2\})$, for multinomial coefficients.

To study further the second moment in Lemma 5.3, we identify the maximal contributions. For this purpose, let $p^* \in \mathcal{P}(\{0,1,2\})$ be the hypergeometric distribution with

$$p^*(s) = \frac{\binom{2}{s}\binom{k-2}{2-s}}{\binom{k}{2}} \ \text{ for } \ s \in \{0,1,2\}, \quad \text{and} \quad p^*_{\mathrm{e}}(1,1) = \frac{4}{k^2},\ p^*_{\mathrm{e}}(1,0) = \frac{2(k-2)}{k^2}. \tag{20}$$

The overlap distribution p^* is a natural candidate for maximizing $E(p)$. Indeed, we obtain p^* when we consider two independent uniformly random assignments in $\{0,1\}^k$ with 2 ones each, and p^*_{e} is exactly the marginal probability if we jointly consider two independent uniformly random assignments in $\{0,1\}^n$ to the variables with $2n/k$ ones each. In the next step, we derive the limits of the log-densities $\frac{1}{n}\ln(E(p))$. Recall that the K(ullback)-L(eibler) divergence $D_{\mathrm{KL}}(p\|q)$ of two distributions $p, q \in \mathcal{P}(\mathcal{S})$, such that p is absolutely continuous with respect to q, is

$$D_{\mathrm{KL}}(p\|q) = \sum_{x\in\mathcal{S}} p(x) \ln\left(\frac{p(x)}{q(x)}\right).$$

Lemma 5.4. *For a fully supported $p \in \mathcal{P}(\{0,1,2\})$ and a sequence $(p_n)_{n\in\mathcal{N}} \subseteq \mathcal{P}_n$ with $\lim_{n\to\infty} p_n = p$ we have $\lim_{n\to\infty} \frac{1}{n}\ln(E(p_n)) = \phi_2(p)$, where*

$$\phi_2(p) = 2\phi_1 - \frac{d}{k}\Delta_d(p) \quad \textit{and} \quad \Delta_d(p) = D_{\mathrm{KL}}(p\|p^*) - \frac{(d-1)k}{d} D_{\mathrm{KL}}(p_{\mathrm{e}}\|p^*_{\mathrm{e}}).$$

The following proposition is the main contribution of this work.

Proposition 5.5. *For $k = 4$ the global minimizers of $\Delta_{d^*(4)}$ are p^*, $p^{(0)}$ given by $p^{(0)}(0) = 1$ and $p^{(2)}$ given by $p^{(2)}(2) = 1$. For $k \in \mathbb{Z}_{\geq 5}$ the global minimizers of $\Delta_{d^*(k)}$ are p^* and $p^{(2)}$.*

With Proposition 5.5, we easily verify that p^* is the unique minimizer of Δ_d for any $d < d^*(k)$, since the KL divergence is minimized by its unique root and $(d-1)k/d$ is increasing in d. This conclusion then allows to compute the limit of the scaled second moment using Laplace's method for sums. Moreover, we confirm the authors' conjecture in [101] as an immediate corollary, and further obtain the Hirschfeld-Gebelein-Rényi maximal correlation for the same noisy channel with input p^*.

Proposition 5.6. *For any $k \in \mathbb{Z}_{\geq 4}$ and $d < d^*(k)$*

$$\frac{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^2]}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]^2} = (1 + o(1))\sqrt{\frac{k-1}{k-d}}, \quad \text{as } n \in \mathcal{N} \text{ tends to infinity.}$$

Proposition 5.6 and the Paley-Zygmund inequality yield $\liminf_{n\to\infty} \mathbb{P}(\boldsymbol{Z}_{\mathrm{c}} > 0) \geq \sqrt{\frac{k-d}{k-1}}$. While this bound suggests that a threshold exists, we need to show that the threshold at d^* is *sharp*.

5.2.4 Small Subgraph Conditioning. We complete the proof of Theorem 5.1 using the small subgraph conditioning method. For this purpose let $a^{\underline{b}} = \prod_{c=0}^{b-1}(a - c)$ denote the falling factorial.

Theorem 5.7 (Small Subgraph Conditioning, [96, Theorem 2])**.** *Let $\boldsymbol{Z}_n$ and $\boldsymbol{X}_{n,1}, \boldsymbol{X}_{n,2}, \ldots$ be non-negative integer-valued random variables. Suppose that $\mathbb{E}[\boldsymbol{Z}_n] > 0$ and that for each $\ell \in \mathbb{Z}_{>0}$ there are $\lambda_\ell \in \mathbb{R}_{>0}$, $\delta_\ell \in \mathbb{R}_{>-1}$ such that for any $L \in \mathbb{Z}_{>0}$*
a) the variables $(\boldsymbol{X}_{n,\ell})_{\ell\in[L]}$ are asymptotically independent and Poisson with $\mathbb{E}[\boldsymbol{X}_{n,\ell}] = (1 + o(1))\lambda_\ell$,
b) for any sequence $r_1, \ldots, r_L$ of non-negative integers,

$$\frac{\mathbb{E}\left[\boldsymbol{Z}_n \prod_{\ell=1}^{L} \boldsymbol{X}_{n,\ell}^{\underline{r_\ell}}\right]}{\mathbb{E}[\boldsymbol{Z}_n]} = (1 + o(1)) \prod_{\ell=1}^{L} [\lambda_\ell(1 + \delta_\ell)]^{r_\ell},$$

c) we explain the variance, i.e.

$$\frac{\mathbb{E}[\boldsymbol{Z}_n^2]}{\mathbb{E}[\boldsymbol{Z}_n]^2} = (1 + o(1)) \exp\left(\sum_{\ell\geq 1} \lambda_\ell \delta_\ell^2\right) \quad \text{and} \quad \sum_{\ell\geq 1} \lambda_\ell \delta_\ell^2 < \infty.$$

Then $\lim_{n\to\infty} \mathbb{P}(\boldsymbol{Z}_n > 0) = 1$.

We will apply Theorem 5.7 to the number $\boldsymbol{Z}_{\mathrm{c}}$ of solutions from Section 5.2.1 and the numbers $\boldsymbol{X}_\ell$ of small cycles in the configuration $\boldsymbol{G}_{\mathrm{c}}$. In order to understand what a cycle in a configuration is, we recall the representation of a configuration G as a four-partite graph from Section 5.2.1. Since we are mostly interested in the factor graph associated with a configuration we divide lengths of paths by three, e.g. what we call a cycle of length four in the bijection, is actually a cycle of length twelve in its equivalent four-partite graph representation. Figures 2a and 5a show an example of a factor graph and the corresponding configuration in its graph representation. Showing the following statement, which establishes Assumption 5.7a), is rather routine.

Lemma 5.8. *For $\ell \in \mathbb{Z}_{>0}$ let $\boldsymbol{X}_\ell$ be the number of 2ℓ-cycles in $\boldsymbol{G}_{\mathrm{c}}$, and set*

$$\lambda_\ell = \frac{[(k-1)(d-1)]^\ell}{2\ell}.$$

Then $(\boldsymbol{X}_\ell)_{\ell\in[L]}$ *are asymptotically independent and Poisson with* $\mathbb{E}[\boldsymbol{X}_\ell]=(1+o(1))\lambda_\ell$ *for all* $L\in\mathbb{Z}_{>0}$.

We give a self-contained proof of Lemma 5.8 in Section 5.7, which we build upon to argue that Assumption 5.7b) in Theorem 5.7 holds. With Lemma 5.8 in place, we consider the base case in Assumption 5.7b), i.e. for $\ell\in\mathbb{Z}_{>0}$ we let $r_\ell=1$ and $r_{\ell'}=0$ otherwise, to determine $\delta_\ell=(1-k)^{-\ell}$. We easily verify that $\sum_{\ell\geq 1}\lambda_\ell\delta_\ell^2=\frac{1}{2}\ln(\frac{k-1}{k-d})$ and thereby establish Assumption 5.7c) using Proposition 5.6. Finally, we follow the proof of Lemma 5.8 to complete the verification of Assumption 5.7b) and thereby complete the proof of Theorem 2.10.

5.2.5 Translation of the Results. Before the proof of the main result, we translate the results for configurations to factor graphs using Lemma 5.8, to obtain contiguity of the factor graph model with respect to the configuration model. For completeness we then also provide self-contained proofs to establish the applicability to hypergraphs (where the constraints may be attached to either the vertices or to the hyperedges). This establishes our claims in Sections 5.1.1 and 2.3.3.

5.3 Preliminaries and Notation

After introducing notation in Section 5.3.1, we establish a few basic facts in Section 5.3.2. In Section 5.3.3 we derive Theorem 2.10 from the configuration version. Finally, in Section 5.3.4 we derive the thresholds for the positive 2-in-k SAT and 2-factors from Theorem 2.10. Hence, this section primarily addresses readers new to this field, treating established concepts, results and the corresponding subtle technical difficulties.

5.3.1 Notation. We use the notation $[n]=\{1,\dots,n\}$ and $[n]_0=\{0\}\cup[n]$ for $n\in\mathbb{Z}_{>0}$, denote the falling factorial (or k-factorial) with $n^{\underline{k}}$ for $n,k\in\mathbb{Z}_{\geq 0}$, $k\leq n$, and multinomial coefficients with $\binom{n}{k}$ for $n\in\mathbb{Z}_{\geq 0}$ and $k\in\mathbb{Z}_{\geq 0}^d$, $d\in\mathbb{Z}_{>1}$, such that $\sum_{i\in[d]}k_i=n$. Recall Stirling's formula from [115], i.e.

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}}\leq n!\leq\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}},\quad n\in\mathbb{Z}_{>0},$$

and in particular $n!=(1+o(1))\sqrt{2\pi n}(\frac{n}{e})^n$. If a random variable $\boldsymbol{X}$ has law p we write $\boldsymbol{X}\sim p$ and use $\mathrm{Po}(\lambda)$ to denote the Poisson distribution with parameter λ. Distributions $p\in\mathcal{P}(\mathcal{S})$ in the convex polytope $\mathcal{P}(\mathcal{S})$ of distributions with finite support $\mathcal{S}$ are identified with their probability mass functions $p\in[0,1]^{\mathcal{S}}$. Further, $\mathcal{P}_n(\mathcal{S})=\{p\in\mathcal{P}(\mathcal{S}):np\in[n]_0\}$ denotes the set of empirical distributions obtained from $n\in\mathbb{Z}_{\geq 1}$ trials. Let v^{t} denote the transpose of a vector v. Finally, we use 'iff' for 'if and only if' and 'whp' for 'with high probability', i.e. $\lim_{n\to\infty}P(\mathcal{E}_n)=1$ for events $\mathcal{E}_n$.

5.3.2 Basic Observations. We briefly establish the claims in Section 2.3.1 for the configuration version, and the claim that d^* is not an integer.

Lemma 5.9. *The set* $\mathcal{G}_{\mathrm{c}}$ *is empty iff* $dn\neq km$*, so let* $dn=km$*. Then, we have* $\boldsymbol{Z}_{\mathrm{c}}=0$ *almost surely if* $n_1=2n/k\notin\mathbb{Z}$*. Finally,* $d^*\in(1,\infty)\setminus\mathbb{Z}$.

Proof. Since $G\in\mathcal{G}_{\mathrm{c}}$ is a bijection $G:[n]\times[d]\to[m]\times[k]$, the set $\mathcal{G}_{\mathrm{c}}$ is empty for $dn\neq km$ and $|\mathcal{G}_{\mathrm{c}}|=(dn)!=(km)!$ otherwise, which proves the first assertion. Next, we fix a solution $x\in\{0,1\}^n$ of $g\in\mathcal{G}_{\mathrm{c}}$ with n_1' ones. Then two a-edges h have to take the value one, i.e. $x_{i(a,h)}=1$, for each $a\in[m]$ and hence $2m$ f-edges $(a,h)\in[m]\times[k]$ in total. On the other hand, there are dn_1' v-edges $(i,h)\in[n]\times[d]$ that take the value one. Since G is a bijection, $dn_1'=2m$, so $n_1=n_1'\in\mathbb{Z}$.

For the last assertion, we first focus on the denominator of d^*, i.e.

$$kH\left(\frac{2}{k}\right)-\ln\binom{k}{2}=-\ln\left(\binom{k}{2}\left(\frac{2}{k}\right)^2\left(\frac{k-2}{k}\right)^{k-2}\right)>0,$$

so $d^*>0$ for $k\in\mathbb{Z}_{\geq 3}$. Next, notice that d^* is a solution of $f(d)=1$ with

$$f(d)=e^{(d-1)(kH(2/k)-\ln\binom{k}{2})-\ln\binom{k}{2}}=\frac{2}{k(k-1)}\left(\frac{k^{k-1}}{2(k-2)^{k-2}(k-1)}\right)^{d-1},$$

which directly implies that $d^*>1$ and further, since $\gcd(k,k-1)=1$, that $d^*\in(1,\infty)\setminus\mathbb{Z}$. □

5.3.3 From Configurations to Factor Graphs. In this section, we derive Theorem 2.10 from Theorem 5.1. First, we notice that there are $(k!)^m(d!)^n$ configurations $G_\mathrm{c}\in\mathcal{G}_\mathrm{c}$ corresponding to each occupation problem $G\in\mathcal{G}$ due to the labeling of the half-edges for each variable and constraint (where we consider G and G_c in their graph representations as illustrated in Figures 2a and 5a). Further, a configuration G_c corresponding to an occupation problem G obviously cannot contain two-cycles, so let $\mathcal{G}_{\mathrm{c},1}\subseteq\mathcal{G}_\mathrm{c}$ denote the set of configurations without two-cycles. Then the uniformly random $\boldsymbol{G}_\mathrm{c}\in\mathcal{G}_\mathrm{c}$ conditional to $\boldsymbol{G}_\mathrm{c}\in\mathcal{G}_{\mathrm{c},1}$ is uniform on $\mathcal{G}_{\mathrm{c},1}$ and further Lemma 5.8 directly implies that $\mathbb{P}(\boldsymbol{G}_\mathrm{c}\in\mathcal{G}_{\mathrm{c},1})\to\mathbb{P}(\boldsymbol{N}=0)>0$ with $\boldsymbol{N}\sim\mathrm{Po}(\lambda_1)$, so the uniform distribution on $\mathcal{G}_{\mathrm{c},1}$ is contiguous with respect to the uniform distribution on $\mathcal{G}_\mathrm{c}$, i.e. for any sequence of events $(\mathcal{E}_n)_n$ we have that $\mathbb{P}(\boldsymbol{G}_\mathrm{c}\in\mathcal{E}_n)\to 0$ implies $\mathbb{P}(\boldsymbol{G}_\mathrm{c}\in\mathcal{E}_n|\boldsymbol{G}_\mathrm{c}\in\mathcal{G}_{\mathrm{c},1})\to 0$ as n tends to infinity. As explained above, the uniform distribution on $\mathcal{G}$ is the pushforward of the uniform distribution on $\mathcal{G}_{\mathrm{c},1}$, so since the number $Z(\boldsymbol{G}_\mathrm{c})$ of solutions in $\boldsymbol{G}_\mathrm{c}\in\mathcal{G}_{\mathrm{c},1}$ properly translates to $Z(\boldsymbol{G})$, we can translate Theorem 5.1 to Theorem 2.10 using the contiguity result discussed above (while the translation of Lemma 5.9 is obvious). Further, we notice that we can also derive a version of Lemma 5.8 for (d,k)-biregular graphs (respectively d-regular k-factor graphs) using a similar argumentation.

5.3.4 Variants of the Occupation Problem. In this section we establish the thresholds for the positive 2-in-k SAT and the 2-factors from Section 5.1.1. In the former example, uniqueness of the clauses is guaranteed by the CNF representation, so instances of this problem are given by d-regular k-uniform hypergraphs. For any such hypergraph $H=([n],\mathcal{E})$ there are $m!$ factor graphs $G\in\mathcal{G}$, that map onto H via $\mathcal{E}=\{v_a:a\in[m]\}$. Further, a factor graph G corresponding to H can obviously not have redundant constraints, i.e. pairs $\{a,b\}\subseteq[m]$ of two distinct constraints a,b such that $v_a=v_b$.

Lemma 5.10. *The number R of redundant constraints in $\boldsymbol{G}$ is zero whp.*

Proof. We will apply the first moment method. For the number $\boldsymbol{S}$ of redundant constraints in $\boldsymbol{G}_\mathrm{c}$, i.e. pairs $\{a,b\}$ of distinct constraints $a,b\in[m]$ such that $v_a=v_b$ and $|v_a|=k$ (where v_a is the set of adjacent variables $i(a,h)$ for $h\in[k]$), the expectation is given by

$$\mathbb{E}[\boldsymbol{S}]=\frac{|\mathcal{E}|}{|\mathcal{G}_\mathrm{c}|}=\frac{1}{(dn)!}\binom{m}{2}n^{\underline{k}}k!(d(d-1))^k(dn-2k)!,$$

where $\mathcal{E}$ is the set of pairs $(G,\{a,b\})$ such that $\{a,b\}$ is a pair of redundant constraints in $G\in\mathcal{G}_\mathrm{c}$, and the terms on the right hand side are derived as follows. We choose two constraints a and b, the k variables they connect to in the order that a connects to them, then choose the order in which b connects to the k variables, further choose the edge that connects to a and to b respectively for each of the k variables and wire the rest. Computing the asymptotics yields $\mathbb{E}[\boldsymbol{S}] = o(1)$. Then, Markov's inequality implies that $\mathbb{P}(\boldsymbol{S} > 0) = \mathbb{P}(\boldsymbol{S} \geq 1) \leq \mathbb{E}[\boldsymbol{S}] \to 0$ for $n \to \infty$. Using contiguity this yields that we do not have redundant constraints in $\boldsymbol{G}_{\mathrm{c}} | \boldsymbol{G}_{\mathrm{c}} \in \mathcal{G}_{\mathrm{c},1}$ whp and further no redundant constraints in $\boldsymbol{G}$ whp because analogously to the previous examples $\boldsymbol{S}$ properly translates to $\boldsymbol{R}$. □

A direct consequence of this lemma is that the uniform distribution on $\mathcal{G}$ and the uniform distribution on the set $\mathcal{G}_1$ of factor graphs with no redundant constraints are mutually contiguous. The remaining steps to translate Theorem 2.10 to d-regular k-uniform hypergraphs with constraints on the hyperedges are completely analogous to the translation of Theorem 5.1. In particular, this establishes the satisfiability threshold for the d-regular 2-in-k SAT.

The proof that the number of redundant variables in $\boldsymbol{G}$, i.e. pairs of distinct variables whose neighborhoods coincide, equals zero whp is completely analogous for $d > 2$, while for the graph case $d = 2$ the result follows analogously to Section 5.3.3 using Lemma 5.8 for 4-cycles. This shows that Theorem 2.10 also applies to the existence of 2-factors in k-regular d-uniform hypergraphs, where the variables are now mapped to the hyperedges, while the constraints are mapped to the vertices.

5.4 The First Moment Method – Proof of Lemma 5.2

This short section is dedicated to the proof of Lemma 5.2. We write the expectation in terms of the number $|\mathcal{E}|$ of pairs $(G, x) \in \mathcal{E}$ such that $x \in \{0,1\}^n$ satisfies $G \in \mathcal{G}_{\mathrm{c}}$, i.e.

$$\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}] = \frac{|\mathcal{E}|}{|\mathcal{G}_{\mathrm{c}}|} = \frac{1}{(dn)!}\binom{n}{n_1}\binom{k}{2}^m (2m)!(dn-2m)!,$$

with $n_1 = 2n/k$ and for the following reasons. First, we choose the n_1 variables with value one in x, then we choose the two a-edges for each constraint $a \in [m]$ with value one, wire the v-edges and f-edges with value one and finally wire the edges with value zero. In particular, this implies that $\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}] > 0$ for all $n \in \mathcal{N}$. Stirling's formula yields after some straightforward but tedious manipulations $\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}] = (1 + o(1))\sqrt{d} e^{n\phi_1}$, as claimed.

5.5 The Second Moment Method

In this section we consider the case $d < d^*$. We prove Lemma 5.3, Lemma 5.4, Proposition 5.5 and Proposition 5.6, the main contribution of this work.

5.5.1 How to Square a Constraint Satisfaction Problem. In order to facilitate the presentation we introduce the *squared* d-regular 2-in-k occupation problem. As before, an instance of this problem is given by a bijection $G : [n] \times [d] \to [m] \times [k]$. Now, for an assignment $x \in (\{0,1\}^2)^n$ let $y_{G,x} = (x_{i(a,h)})_{a \in [m], h \in [k]}$ be the corresponding f-edge assignment under G, where we recall from Section 5.2.1 that $i(a,h) = (G^{-1}(a,h))_1 \in [n]$ is the variable $i(a,h)$ wired to the f-edge (a,h) under G. A constraint $a \in [m]$ is satisfied by a constraint assignment $x \in (\{0,1\}^2)^k$ iff $x \in \mathcal{S}^{(2)}$, where

$$\mathcal{S}^{(2)} = \left\{ x \in (\{0,1\}^2)^k : \sum_{h \in [k]} x_{h,1} = \sum_{h \in [k]} x_{h,2} = 2 \right\}.$$

An f-edge assignment $x \in (\{0,1\}^2)^{m \times k}$ is satisfying if $x_a = (x_{a,h})_{h \in [k]}$ satisfies a for all $a \in [m]$. Finally, an assignment $x \in (\{0,1\}^2)^n$ is a solution of G if $y_{G,x}$ is satisfying. Notice that the pairs of solutions $x,\ x' \in \{0,1\}^n$ of the standard problem on G are in one to one correspondence with the solutions $y \in (\{0,1\}^2)^n$ of the squared problem on G via $y = (x_i, x'_i)_{i\in[n]}$. So, $Z^{(2)}(G) = Z(G)^2$ for the number $Z^{(2)}(G)$ of solutions of the squared problem, hence $\boldsymbol{Z}_{\mathrm{c}}^{(2)} = \boldsymbol{Z}_{\mathrm{c}}^2$ for $\boldsymbol{Z}_{\mathrm{c}}^{(2)} = Z^{(2)}(\boldsymbol{G}_{\mathrm{c}})$ and in particular $\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^{(2)}] = \mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^2]$. This equivalence allows us to entirely focus on the squared problem.

5.5.2 Proof of Lemma 5.3. As before, we can write $\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^{(2)}] = \frac{1}{(dn)!}|\mathcal{E}|$, where $|\mathcal{E}|$ is the number of pairs $(G, x) \in \mathcal{E}$ such that $x \in (\{0,1\}^2)^n$ solves G. Set

$$\mathcal{Y} = \left\{ y \in (\{0,1\}^2)^{m\times k} : y_a \in \mathcal{S}^{(2)} \text{ for all } a \in [m] \right\}.$$

For $y \in \mathcal{Y}$ let the overlap distribution $p_y \in \mathcal{P}_m(\{0,1,2\})$ be given by

$$p_y(s) = \frac{1}{m}|\{a \in [m] : |y_a^{-1}(1,1)| = s\}|, \quad s \in \{0,1,2\}.$$

Further, let the edge distribution $q_y \in \mathcal{P}_{km}(\{0,1\}^2)$ be given by

$$q_y(x) = \frac{1}{km}|\{(a,h) \in [m]\times[k] : y_{a,h} = x\}| = \frac{1}{km}|y^{-1}(x)|, \quad x \in \{0,1\}^2.$$

Using that $|y_a^{-1}(1,0)| = |y_a^{-1}(0,1)| = 2 - |y_a^{-1}(1,1)|$ and hence $|y^{-1}(0,0)| = k - 4 + |y^{-1}(1,1)|$ we directly get

$$\begin{aligned} q_y(1,1) &= \frac{1}{km}\sum_{a\in[m]}|y_a^{-1}(1,1)| = \frac{1}{km}\sum_{s\in\{0,1,2\}} s|\{a\in[m] : |y_a^{-1}(1,1)| = s\}| = \sum_{s\in\{0,1,2\}}\frac{s}{k}p_y(s),\\ q_y(1,0) = q_y(0,1) &= \frac{1}{km}\sum_{s\in\{0,1,2\}}(2-s)|\{a\in[m] : |y_a^{-1}(1,1)| = s\}| = \sum_{s\in\{0,1,2\}}\frac{2-s}{k}p_y(s),\\ q_y(0,0) &= \sum_{s\in\{0,1,2\}}\frac{k-4+s}{k}p_y(s). \end{aligned}$$

Hence, let $p_{\mathrm{e}} = Wp \in \mathcal{P}(\{0,1\}^2)$ denote the edge distribution of any (not necessarily empirical) overlap distribution $p \in \mathcal{P}(\{0,1,2\})$, where $W \in [0,1]^{\{0,1\}^2\times\{0,1,2\}}$ is given by

$$W_{11,s} = \frac{s}{k}, \quad W_{10,s} = W_{01,s} = \frac{2-s}{k} \text{ and } W_{00,s} = \frac{k-4+s}{k}, \quad s \in \{0,1,2\}. \tag{21}$$

Now, notice that for any $(G,x) \in \mathcal{E}$ we have $y_{G,x,a,h} = x_{i(a,h)}$ for all $a \in [m]$, $h \in [k]$, hence $G(x^{-1}(z)\times[d]) = y_{G,x}^{-1}(z)$ and by that

$$q_{y_{G,x}}(z) = \frac{|y_{G,x}^{-1}(z)|}{km} = \frac{d|x^{-1}(z)|}{km} = \frac{1}{n}|x^{-1}(z)| = q_x(z) \text{ for } z \in \{0,1\}^2,$$

i.e. the relative frequencies of the values in the f-edge assignment $y_{G,x}$ coincide with the relative frequencies $q_x \in \mathcal{P}_n(\{0,1\}^2)$ of the values in the variable assignment x. In particular, this shows that a satisfying f-edge assignment $y \in \mathcal{Y}$ is only *attainable* if $q_y \in \mathcal{P}_n(\{0,1\}^2)$, and thereby

$$\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^{(2)}] = \frac{1}{(dn)!} \sum_{p \in \mathcal{P}_n} |\{(G,x) \in \mathcal{E} : p_{y_{G,x}} = p\}|.$$

Now, fix an attainable satisfying f-edge assignment $y \in \mathcal{Y}$ and an assignment $x \in (\{0,1\}^2)^n$ with $q_y = q_x$, i.e. $|x^{-1}(z) \times [d]| = |y^{-1}(z)|$ for all $z \in \{0,1\}^2$. Any bijection G with $y = y_{G,x}$ needs to respect $G(x^{-1}(z) \times [d]) = y_{G,x}^{-1}(z)$ for $z \in \{0,1\}^2$ and can hence be uniquely decomposed into its restrictions $G_z : x^{-1}(z) \times [d] \to y_{G,x}^{-1}(z)$. On the other hand, any choice of such restrictions G_z gives a bijection G with $y = y_{G,x}$, and so

$$|\mathcal{E}_{x,y}| = \prod_{z \in \{0,1\}^2} (dnq_x(z))! = \prod_{z \in \{0,1\}^2} (dnp_{y,\mathrm{e}}(z))! \ \text{, where } \ \mathcal{E}_{x,y} = \{(G,x) \in \mathcal{E} : y_{G,x} = y\}.$$

Notice that $\mathcal{E}_{x,y} \cap \mathcal{E}_{x',y'} = \emptyset$ for any $(x,y) \neq (x',y')$, which is obvious for $x \neq x'$, and also for $y \neq y'$, since $y_{G,x} = y \neq y' = y_{G',x}$ implies that $G \neq G'$. But since $|\mathcal{E}_{x,y}|$ only depends on p_y (actually only on $p_{y,\mathrm{e}}$) this completes the proof, because for any fixed attainable overlap distribution $p \in \mathcal{P}_n$, we can now independently choose the satisfying f-edge assignment y and variable assignment x, subject to $q_x = p_{\mathrm{e}}$ and $p_y = p$ (which implies $q_y = q_x$). So we have $\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^{(2)}] = \sum_p E(p)$ with $p \in \mathcal{P}_n$ and

$$E(p) = \frac{1}{(dn)!} \binom{n}{np_{\mathrm{e}}} \binom{m}{mp} \prod_{s \in \{0,1,2\}} \binom{k}{s, 2-s, 2-s, k-4+s}^{mp(s)} \prod_{x \in \{0,1\}^2} (dnp_{\mathrm{e}}(z))!,$$

where we choose a variable assignment x with $q_x = p_{\mathrm{e}}$, an f-edge assignment y with $p_y = p$ by first choosing one of the $\binom{m}{mp}$ options for $(|y_a^{-1}(1,1)|)_{a \in [m]}$ and then independently one of the $\binom{k}{s,2-s,2-s,k-4+s}$ satisfying constraint assignments for each of the $mp(s)$ constraints with overlap $s \in \{0,1,2\}$, and finally choosing a bijection G with $(G,x) \in \mathcal{E}_{x,y}$.

5.5.3 Empirical Overlap Distributions. This section is dedicated to deriving properties of the set $\mathcal{P}_n$ for $n \in \mathcal{N}$. In the following we will use the canonical ascending order on $\{0,1,2\}$ to denote points in $\mathbb{R}^{\{0,1,2\}}$ and the ascending lexicographical order on $\{0,1\}^2$ to denote points in $\mathbb{R}^{\{0,1\}^2}$. Recall that $p^{(s)} \in \mathcal{P}(\{0,1,2\})$ given by $p^{(s)}(s) = 1$ for $s \in \{0,1,2\}$ denote the corners of the convex polytope $\mathcal{P}(\{0,1,2\})$ and further consider the vectors in $\mathbb{R}^{\{0,1,2\}}$

$$b_1 = (-d, d, 0)^{\mathrm{t}}, \quad b_2 = (1, -2, 1)^{\mathrm{t}}. \tag{22}$$

Finally, let

$$\mathcal{X} = \left\{ x \in \mathbb{R}^2 : (b_1, b_2)x \geq -p^{(0)} \right\}, \quad \mathcal{X}_n = \mathcal{X} \cap (m^{-1}\mathbb{Z})^2.$$

Lemma 5.11. *The map $\iota_n : \mathcal{X}_n \to \mathcal{P}_n$, $x \mapsto p^{(0)} + (b_1, b_2)x$ is a bijection.*

Proof. We use the shorthands $1_{\{0,1,2\}} = (1)_{s \in \{0,1,2\}}$ and

$$1_{\{0,1,2\}}^{\perp} = \left\{ x \in \mathbb{R}^{\{0,1,2\}} : 1_{\{0,1,2\}}^{\mathrm{t}} x = 0 \right\} = \left\{ x \in \mathbb{R}^{\{0,1,2\}} : \sum_{s \in \{0,1,2\}} x_s = 0 \right\}.$$

Note that $\mathcal{P}(\{0,1,2\}) \subseteq p^{(0)} + 1^{\perp}_{\{0,1,2\}} = \{p^{(0)} + x : x \in 1^{\perp}_{\{0,1,2\}}\}$. On the other hand, (b_1, b_2) is a basis of $1^{\perp}_{\{0,1,2\}}$, and hence

$$\iota : \mathbb{R}^2 \to p^{(0)} + 1^{\perp}_{\{0,1,2\}}, \quad x \mapsto p^{(0)} + (b_1, b_2)x, \tag{23}$$

is bijective. This gives that $\iota(\mathcal{X}) = \mathcal{P}(\{0,1,2\})$ and that ι_n is the restriction of ι to $\mathcal{X}_n$, so ι_n is a bijection from $\mathcal{X}_n$ to $\mathcal{P}(\{0,1,2\}) \cap \iota\left((m^{-1}\mathbb{Z})^2\right)$. Consequently, it remains to show that $\mathcal{P}_n = \mathcal{P}(\{0,1,2\}) \cap \iota\left((m^{-1}\mathbb{Z})^2\right)$, where

$$\iota\left((m^{-1}\mathbb{Z})^2\right) = \left\{p^{(0)} + \frac{i_1}{m}b_1 + \frac{i_2}{m}b_2 : i \in \mathbb{Z}^2\right\}$$

is a grid anchored at $p^{(0)}$ and spanned by $m^{-1}b_1$ and $m^{-1}b_2$. Note that

$$p^{(0)}_{\mathrm{e}} = \left(\frac{k-4}{k}, \frac{2}{k}, \frac{2}{k}, 0\right)^{\mathrm{t}},$$

so $np^{(0)}_{\mathrm{e}} \in \mathbb{Z}^{\{0,1\}^2}$ since $n \in \mathcal{N}$, and hence $p^{(0)} \in \mathcal{P}_n$ by the definition of $\mathcal{P}_n$. Next, we show that $\mathcal{P}_n$ is on the grid, i.e. $\mathcal{P}_n \subseteq \iota\left((m^{-1}\mathbb{Z})^2\right)$. For this purpose fix $p \in \mathcal{P}_n$ and let $x = \iota^{-1}(p)$, i.e. $mp \in \mathbb{Z}^{\{0,1,2\}}$, $n(Wp) \in \mathbb{Z}^{\{0,1\}^2}$ and $p = p^{(0)} + x_1 b_1 + x_2 b_2$. This directly gives $mx_2 = mp(2) \in \mathbb{Z}$. Further, we notice that b_2 is in the kernel of W from Equation (21), i.e. $Wb_2 = 0_{\{0,1\}^2}$, and $Wb_1 = \frac{d}{k}w$ with $w = (1,-1,-1,1)^{\mathrm{t}}$. This directly gives $p_{\mathrm{e}}(1,1) = 0 + \frac{d}{k}x_1 + 0$ and hence $mx_1 = np_{\mathrm{e}}(1,1) \in \mathbb{Z}$, i.e. $x \in (m^{-1}\mathbb{Z})^2$ and hence $p = \iota(x) \in \iota\left((m^{-1}\mathbb{Z})^2\right)$. Conversely, for any $x \in \mathcal{X}_n$ and with $p = \iota(x)$ we have $p \in \mathcal{P}(\{0,1,2\})$ since $x \in \mathcal{X}$, further $mp = mp^{(0)} + (b_1, b_2)(mx) \in \mathbb{Z}^{\{0,1,2\}}$ since $mx \in \mathbb{Z}^2$ and the other terms on the right-hand side are integer valued by definition, and finally $np_{\mathrm{e}} = np^{(0)}_{\mathrm{e}} + mx_1 w \in \mathbb{Z}^{\{0,1\}^2}$. □

Using Lemma 5.11 we have $\mathbb{E}[\boldsymbol{Z}^{(2)}_{\mathrm{c}}] = \sum_{x \in \mathcal{X}_n} E(\iota_n(x))$, where $\mathcal{X}_n \subseteq \mathbb{R}^2$ may be considered as a normalization of the grid $\mathcal{P}_n \subseteq p^{(0)} + 1^{\mathrm{t}}_{\{0,1,2\}}$. In order to prepare the upcoming asymptotics of the second moment, we give a complete characterization of the convex polytope $\mathcal{X}$ and the image of $\mathcal{X}$ under $W(b_1, b_2)$, i.e. the image $p_{\mathrm{e}} = Wp$ of $p \in \mathcal{P}(\{0,1,2\})$ under W from Equation (21). Let $w = (1,-1,-1,1)^{\mathrm{t}}$ from the proof of Lemma 5.11, and set

$$\mathcal{W} = \left\{p^{(0)}_{\mathrm{e}} + yw : y \in [0, 2/k]\right\} \subseteq \mathcal{P}(\{0,1\}^2), \quad \mathcal{X}_p = \left\{x \in \mathcal{X} : x_1 = \frac{k}{d}p(1,1)\right\} \text{ for } p \in \mathcal{W}.$$

Moreover, recall the definition of p^* from (20) and the bijection ι from (23), and let

$$x^{(0)} = \begin{pmatrix}0\\0\end{pmatrix}, x^{(1)} = d^{-1}\begin{pmatrix}1\\0\end{pmatrix}, x^{(2)} = d^{-1}\begin{pmatrix}2\\d\end{pmatrix} \in \mathbb{R}^2 \text{ and } x^* = \iota^{-1}(p^*).$$

Lemma 5.12. *The set $\mathcal{X}$ is a two-dimensional convex polytope with corners $x^{(0)}, x^{(1)}, x^{(2)}$, and x^* is in the interior of $\mathcal{X}$. The image of $\mathcal{X}$ under $W(b_1, b_2)$ is the one-dimensional convex polytope $\mathcal{W}$ with corners $p^{(0)}_{\mathrm{e}}$ and $p^{(2)}_{\mathrm{e}}$. Further, the preimage of $p \in \mathcal{W}$ under $W(b_1, b_2)$ is $\mathcal{X}_p$, where $\mathcal{X}_{p^{(s)}_{\mathrm{e}}} = \{p^{(s)}\}$ for $s \in \{0,2\}$ and the intersection of $\mathcal{X}_p$ with the interior of $\mathcal{X}$ is non-empty otherwise.*

Proof. Notice that $\iota(x^{(s)}) = p^{(s)}$ for $s \in \{0,1,2\}$, so since $\mathcal{P}(\{0,1,2\})$ is the convex hull of its corners $p^{(s)}$, $s \in \{0,1,2\}$, we have that $\mathcal{X}$ is the convex hull of $x^{(s)}$, $s \in \{0,1,2\}$, i.e. a two-dimensional convex polytope with corners $x^{(s)}$, since ι is an affine transformation. In particular this also directly yields that x^* is in the interior of $\mathcal{X}$. Further, this shows that for any $x \in \mathcal{X}$ we have $x_1 \geq 0$ with equality iff $x = x^{(0)}$ and further $x_1 \leq \frac{2}{d}$ with equality iff $x = x^{(2)}$. Using $Wb_2 = 0_{\{0,1\}^2}$ and $Wb_1 = \frac{d}{k}w$ from the proof of Lemma 5.11 we directly get that

$$W(b_1, b_2)x = p_{\mathrm{e}}^{(0)} + \frac{d}{k}x_1 w \text{ with } \frac{d}{k}x_1 \in [0, 2/k],$$

hence the image of $\mathcal{X}$ under $W(b_1, b_2)$ is a subset of $\mathcal{W}$. Conversely, for $y \in [0, 2/k]$ and $x = \frac{k}{2}yx^{(2)} \in \mathcal{X}$ we have $W(b_1, b_2)x = p_{\mathrm{e}}^{(0)} + yw$, which shows that $\mathcal{W}$ is the image of $\mathcal{X}$ under $W(b_1, b_2)$. This also shows that $\mathcal{X}_p$ is the preimage of $p \in \mathcal{W}$, since for $y \in [0, 2/k]$ and $p = p_{\mathrm{e}}^{(0)} + yw$ we have $p(1, 1) = y$. This in turn directly yields that $\mathcal{X}_{p_{\mathrm{e}}^{(s)}} = \{p^{(s)}\}$ for $s \in \{0, 2\}$. To see that $\mathcal{X}_p$ contains interior points of $\mathcal{X}$ otherwise, we can consider non-trivial convex combinations of x^* and $x^{(0)}$ for $\frac{k}{d}p(1, 1) < x_1^*$ and non-trivial convex combinations of x^* and $x^{(2)}$ for $\frac{k}{d}p(1, 1) > x_1^*$, which are points in the interior of $\mathcal{X}$. □

Notice that in the two-dimensional case at hand, the proof of Lemma 5.12 is overly formal. The set $\mathcal{X}$ is simply (the convex hull of) the triangle given by $x^{(s)}$, $s \in \{0, 1, 2\}$, with $\mathcal{X}_p$ given by the vertical lines in $\mathcal{X}$ with $x_1 = \frac{d}{k}p(1, 1)$. Further, the set $\mathcal{X}_n$ is a canonical discretization of $\mathcal{X}$ in that it is given by the points of the grid $(m^{-1}\mathbb{Z})^2$ contained in the triangle $\mathcal{X}$.

5.5.4 Proof of Lemma 5.4. We derive Lemma 5.4 from the following stronger assertion.

Lemma 5.13. *Let $\mathcal{U} \subseteq \mathcal{P}(\{0, 1, 2\})$ be a subset with non-empty interior and such that the closure of $\mathcal{U}$ is contained in the interior of $\mathcal{P}(\{0, 1, 2\})$. Then there exists a constant $c = c(\mathcal{U}) \in \mathbb{R}_{>0}$ such that for all $n \in \mathcal{N}$ and all $p \in \mathcal{P}_n \cap \mathcal{U}$ we have $\tilde{E}(p)e^{-c/n} \leq E(p) \leq \tilde{E}(p)e^{c/n}$, where*

$$\tilde{E}(p) = \sqrt{\frac{d^3}{(2\pi)^2 m^2 \prod_s p(s)}}e^{n\phi_2(p)}.$$

Proof. Let $\mathcal{C}$ denote the closure of $\mathcal{U}$ and $\pi_s : \mathcal{C} \to [0, 1]$, $p \mapsto p(s)$ the projection for $s \in \{0, 1, 2\}$. Since $\mathcal{C}$ is compact, the continuous map π_s attains its maximum $p_+(s)$ and its minimum $p_-(s)$, which directly gives $0 < p_-(s) < p_+(s) < 1$ since all $p \in \mathcal{C}$ are fully supported and the interior of $\mathcal{C}$ is non-empty (that gives the second inequality). Using Lemma 5.12, the continuous map $\pi : \mathcal{C} \to [0, 2/k]$, $p \mapsto p_{\mathrm{e}}(1, 1)$, and the same reasoning as above we obtain the maximum $p_{\mathrm{e},+}(1, 1)$ and minimum $p_{\mathrm{e},-}(1, 1)$ of π with $0 < p_{\mathrm{e},-}(1, 1) < p_{\mathrm{e},+}(1, 1) < 2/k$, which directly give the bounds $p_{\mathrm{e},-}(x), p_{\mathrm{e},+}(x) > 0$ for $x \in \{0, 1\}^2$ as functions of $p_{\mathrm{e},+}(1, 1)$ and $p_{\mathrm{e},-}(1, 1)$. Now, we can use these bounds with the Stirling bound to obtain a constant $c \in \mathbb{R}_{>0}$ such that for all $n \in \mathcal{N}$ and $p \in \mathcal{P}_n \cap \mathcal{C}$ we have $E'(p)e^{-c/n} \leq E(p) \leq E'(p)e^{c/n}$, where

$$\begin{aligned} E'(p) &= \sqrt{\frac{2\pi m d^3}{\prod_s (2\pi m p(s))}} \prod_{s \in \{0,1,2\}} \binom{k}{s, 2-s, 2-s, k-4+s}^{mp(s)} e^{mH(p)-(d-1)nH(p_{\mathrm{e}})} \\ &= \sqrt{\frac{d^3}{(2\pi)^2 m^2 \prod_s p(s)}} e^{2m \ln \binom{k}{2} - mD_{\mathrm{KL}}(p\|p^*) - (d-1)nH(p_{\mathrm{e}})}. \end{aligned}$$

To see that $E'(p) = \tilde{E}(p)$, we observe that $D_{\mathrm{KL}}(p_{\mathrm{e}}\|p_{\mathrm{e}}^*) = 2H(2/k) - H(p_{\mathrm{e}})$, since $H(p_{\mathrm{e}}^*) = 2H(2/k)$, $p_{\mathrm{e}}(1, 0) = p_{\mathrm{e}}(0, 1)$, and $p_{\mathrm{e}}(1, 1) + p_{\mathrm{e}}(1, 0) = 2/k$ for any $p \in \mathcal{P}(\{0, 1, 2\})$. □

Now, Lemma 5.4 is an immediate corollary. To see this, fix a fully supported overlap distribution $p \in \mathcal{P}(\{0,1,2\})$ and a sequence $(p_n)_{n\in\mathcal{N}} \subseteq \mathcal{P}_n$ converging to p, e.g. $p_n = \iota(m^{-1}\lfloor mx_1\rfloor, m^{-1}\lfloor mx_2\rfloor)$ with $\iota(x) = p$ and n sufficiently large. Further, fix a neighborhood $\mathcal{U}$ of p as described in Lemma 5.13, which is possible since p is fully supported. Then we have $p_n \in \mathcal{U}$ for sufficiently large n, hence with the continuity of ϕ_2 we have

$$\lim_{n\to\infty}\frac{1}{n}\ln(E(p_n)) = \lim_{n\to\infty}\frac{1}{n}\ln(\tilde{E}(p_n)) = \lim_{n\to\infty}\phi_2(p_n) = \phi_2(p).$$

5.5.5 Proof of Proposition 5.6. We postpone the proof of Proposition 5.5 and continue with Laplace's method for sums using the result. We obtain that

$$\frac{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^2]}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]^2} = \sum_p e(p), \quad e(p) = \frac{E(p)}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]^2} = \frac{\binom{2n/k}{np_{\mathrm{e}}(1,1)}\binom{(k-2)n/k}{np_{\mathrm{e}}(0,1)}\binom{dn}{2dn/k}\binom{m}{mp}\prod_s p^*(s)^{mp(s)}}{\binom{2dn/k}{dnp_{\mathrm{e}}(1,1)}\binom{(k-2)dn/k}{dnp_{\mathrm{e}}(0,1)}\binom{n}{2n/k}},$$

where the sum is over $p \in \mathcal{P}_n$. First, we use Proposition 5.5 to show that Laplace's method of sums is applicable. While we have already established that Δ_{d^*} is non-negative, we still need to ensure that p^* is the unique minimizer of Δ_d for $d < d^*$ and that the Hessian at p^* is positive definite. We will need the second order Taylor approximation of the KL divergence. To be specific, let μ^* have finite non-trivial support $\mathcal{S}$ and let $f : \mathcal{P}(\mathcal{S}) \to \mathbb{R}_{\ge 0}$, $\mu \mapsto D_{\mathrm{KL}}(\mu\|\mu^*)$, be the corresponding KL divergence. Then

$$f^{(2)} : \mathcal{P}(\mathcal{S}) \to \mathbb{R}_{\ge 0}, \quad \mu \mapsto \frac{1}{2}D_{\chi^2}(\mu\|\mu^*) = \frac{1}{2}\sum_s \frac{(\mu(s)-\mu^*(s))^2}{\mu^*(s)} = \frac{1}{2}(\mu-\mu^*)^{\mathrm{t}}D_{\mu^*}^{-1}(\mu-\mu^*),$$

is the second order Taylor approximation of f at μ^*, where $D_{\chi^2}(\mu\|\mu^*)$ denotes Pearson's χ^2 divergence, $D_{\mu^*} = (\delta_{i,j}\mu^*(i))_{i,j\in\mathcal{S}}$ the matrix with μ^* on the diagonal, and $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise; this can be easily seen by considering the extension of f to $\mathbb{R}^{\mathcal{S}}_{\ge 0}$. On the other hand, we would like to consider Δ_d as a function over the suitable domain $\mathcal{X}$ from Section 5.5.3, however relative to the base point p^*. Hence, let $\mathcal{X}^* = \{x - x^* : x \in \mathcal{X}\}$ be the triangle $\mathcal{X}$ centered at x^* instead of $x^{(0)}$, and $\iota^* : \mathcal{X}^* \to \mathcal{P}(\{0,1,2\})$ the bijection given by

$$\iota^*(x) = \iota(x + x^*) = p^{(0)} + (b_1, b_2)x + (b_1, b_2)x^* = \iota(x^*) + (b_1, b_2)x = p^* + (b_1, b_2)x$$

for $x \in \mathcal{X}^*$, with b_1, b_2 from Equation (22). Now, let $\gamma_d : \mathcal{X}^* \to \mathbb{R}_{\ge 0}$, $x \mapsto \Delta_d(\iota^*(x))$, denote the corresponding parametrization of Δ_d. Then, using the chain rule for multivariate calculus as indicated above for both (b_1, b_2) and W from (21), we derive the Hessian

$$H_d = (b_1, b_2)^{\mathrm{t}}\left(D_{p^*}^{-1} - \frac{(d-1)k}{d}W^{\mathrm{t}}D_{p^*_{\mathrm{e}}}^{-1}W\right)(b_1, b_2) \tag{24}$$

of γ_d at $0_{[2]} \in \mathbb{R}^2$, using the shorthand $D_{\mu^*} = (\delta_{i,j}\mu^*(i))_{i,j}$. The properties of the KL divergence imply that $\gamma_d(0_{[2]}) = 0$ and γ_d has a stationary point at $0_{[2]}$. Now, the second order Taylor approximation $\gamma_d^{(2)} : \mathcal{X}^* \to \mathbb{R}$, $x \mapsto \frac{1}{2}x^{\mathrm{t}}H_d x$, of γ_d at $0_{[2]}$ can be written as $\gamma_d^{(2)} = \Delta_d^{(2)} \circ \iota^*$ with

$$\Delta_d^{(2)}(p) = \frac{1}{2}\left[D_{\chi^2}(p\|p^*) - \frac{(d-1)k}{d}D_{\chi^2}(p_{\mathrm{e}}\|p^*_{\mathrm{e}})\right]. \tag{25}$$

Further, for any neighborhood $\mathcal{U}$ of $0_{[2]}$ such that the closure of $\mathcal{U}$ is contained in the interior of $\mathcal{X}^*$, Taylor's theorem yields a constant $c \in \mathbb{R}_{>0}$ such that

$$\gamma_d^{(2)}(x) - c\|x\|_2^3 \le \gamma_d(x) \le \gamma_d^{(2)}(x) + c\|x\|_2^3 \tag{26}$$

for all $x \in \mathcal{U}$. Since H_d is symmetric, let $\lambda_1,\ \lambda_2 \in \mathbb{R}$ with $\lambda_1 \le \lambda_2$ denote its eigenvalues and fix a corresponding orthonormal basis of eigenvectors $v_1,\ v_2 \in \mathbb{R}^2$, i.e. $H_d v_1 = \lambda_1 v_1$ and $H_d v_2 = \lambda_2 v_2$.

Formally and analogously to the KL divergence we will take the liberty to identify Δ_d and $\Delta_d^{(2)}$ with their extensions to the maximal domains $\mathcal{D} \subseteq \mathbb{R}^{\{0,1,2\}}$ and $\mathcal{D}^{(2)} = \mathbb{R}^{\{0,1,2\}}$ respectively. In particular, Lemma 5.12 shows that for any fully supported $p \in \mathcal{P}(\{0,1,2\})$ the edge distribution p_e also has full support, hence we can use the Lipschitz continuity of W on $\mathbb{R}^{\{0,1,2\}}$ to find $\varepsilon \in \mathbb{R}_{>0}$ such that both $p' > 0$ and $Wp' > 0$ for any $p' \in \mathcal{B}_\varepsilon(p) \subseteq \mathbb{R}_{>0}^{\{0,1,2\}}$ and thereby Δ_d is well-defined and smooth on $\mathcal{B}_\varepsilon(p)$.

Lemma 5.14. *Let $k \in \mathbb{Z}_{\ge 4}$ and $d \in (0, d^*)$. Then the unique minimizer of γ_d is $0_{[2]}$ and H_d is positive definite.*

Proof. Using Proposition 5.5 we know that H_{d^*} is positive semidefinite since $0_{[2]}$ is a global minimum of γ_{d^*}. This in turn yields that $\gamma_{d^*}^{(2)} \ge 0$ or equivalently $\Delta_{d^*}^{(2)} \ge 0$. Now, for any $d < d^*$ the unique minimizer of Δ_d is p^* since $\Delta_d(p^*) = 0$, further $\Delta_d(p) > 0$ for any $p \ne p^*$ with $p_\mathrm{e} = p_\mathrm{e}^*$ and

$$\Delta_d(p) = D_{\mathrm{KL}}(p\|p^*) - \left(1 - \frac{1}{d}\right) k D_{\mathrm{KL}}(p_\mathrm{e}\|p_\mathrm{e}^*) > \Delta_{d^*}(p) \ge 0$$

for any p with $p_\mathrm{e} \ne p_\mathrm{e}^*$. But the same argumentation shows that p^* is the unique minimizer of $\Delta_d^{(2)}$, since $D_{\chi^2}(\mu\|\mu^*)$ is also minimal with value 0 iff $\mu = \mu^*$. This in turn shows that $\gamma_d^{(2)}$ is uniquely minimized at $0_{[2]}$ and hence H_d is positive definite. □

Let $\eta_{\mathrm{KL}} = \sup_{p \ne p^*} \frac{D_{\mathrm{KL}}(p_\mathrm{e}\|p_\mathrm{e}^*)}{D_{\mathrm{KL}}(p\|p^*)}$ denote the contraction coefficient with respect to the KL divergence. Notice that by Proposition 5.5 we have $\frac{d^*}{(d^*-1)k} \ge \frac{D_{\mathrm{KL}}(p_\mathrm{e}\|p_\mathrm{e}^*)}{D_{\mathrm{KL}}(p\|p^*)}$ for all $p \ne p^*$ with equality for $p = p^{(2)}$, hence $\eta_{\mathrm{KL}} = \frac{d^*}{(d^*-1)k}$ (so Proposition 5.5 indeed confirms the conjecture by the authors in [101]). Further, let $\eta_{\chi^2} = \sup_{p \ne p^*} \frac{D_{\chi^2}(p_\mathrm{e}\|p_\mathrm{e}^*)}{D_{\chi^2}(p\|p^*)}$ denote the contraction coefficient with respect to the χ^2 divergence. The proof of Lemma 5.14 suggests that $\eta_{\chi^2} \le \eta_{\mathrm{KL}}$, a result known from literature.

In the rest of this section we discuss the straightforward (but cumbersome) application of Laplace's method for sums. For convenience, we first show that the boundaries can be neglected and derive the asymptotics of the sum on the interior using the uniform convergence established in Lemma 5.13.

Lemma 5.15. *Let $d \in (0, d^*)$ and let $\mathcal{U}$ be a neighborhood of p^* such that its closure is contained in the interior of $\mathcal{P}(\{0,1,2\})$. Then*

$$\frac{\mathbb{E}[\boldsymbol{Z}_\mathrm{c}^2]}{\mathbb{E}[\boldsymbol{Z}_\mathrm{c}]^2} = \sum_{p \in \mathcal{P}_n} e(p) = (1 + o(1)) \sum_{p \in \mathcal{P}_n \cap \mathcal{U}} \sqrt{\frac{d}{(2\pi)^2 m^2 \prod_s p(s)}} e^{-m\Delta_d(p)}.$$

Proof. Let $\Delta_{\min} > 0$ denote the global minimum of Δ on $\mathcal{P}(\{0,1,2\}) \setminus \mathcal{U}$. Now, we can use the well-known bounds $\frac{1}{a+1}\exp(aH(\frac{b}{a})) \le \binom{a}{b} \le \exp(aH(\frac{b}{a}))$ for binomial coefficients and the corresponding upper bound for multinomial coefficients (using the entropy of the distribution determined by the weights $\frac{b_i}{a}$) to derive

$$\sum_{p\notin\mathcal{U}} e(p) \leq \rho(n)\sum_{p\notin\mathcal{U}} e^{-m\Delta_d(p)} \leq \rho(n)e^{-m\Delta_{\min}}|\mathcal{P}_m(\{0,1,2\})| = \rho(n)\binom{m+1}{2}e^{-m\Delta_{\min}}, \text{ where}$$
$$\rho(n) = (n+1)\left(\frac{2dn}{k}+1\right)\left(\frac{(k-2)dn}{k}+1\right).$$

Here, we used the form of $e(p)$ introduced at the beginning of this section and further notice that the bounds used are tight for the log-densities, i.e. the exponent is $\Delta_d(p)$ by the computations in Section 5.5.4. The right hand side vanishes for n tending to infinity, hence we have

$$\frac{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^2]}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]^2} = (1+o(1))\sum_{p\in\mathcal{U}} e(p).$$

Now, the result directly follows using Lemma 5.13 and Lemma 5.2. □

Lemma 5.15 shows that the overlap distributions p with material contributions $e(p)$ to the second moment are concentrated around p^*. Hence, instead of considering a fixed neighborhood $\mathcal{U}$ of p^* we consider a sequence $(\mathcal{U}_n)_{n\in\mathcal{N}}$ of decreasing neighborhoods. First, we choose a scaling that improves the assertion of Lemma 5.15 and further allows to simplify the asymptotics of the right hand side, in the sense that the leading factor collapses to a constant and $\gamma_d(x) = \Delta_d(\iota^*(x))$ can be replaced by its second order Taylor approximation $\gamma_d^{(2)}(x) = \Delta_d^{(2)}(\iota^*(x)) = \frac{1}{2}x^{\mathrm{t}}H_dx$ from above (25). For this purpose let $\mathcal{U}^* \subseteq \mathcal{X}^*$ be a sufficiently small neighborhood of $0_{[2]}$ (in particular bounded away from the boundary of $\mathcal{X}^*$), further

$$\mathcal{U}_n^* = \left\{x\in\mathcal{X}^* : \|x\|_2 < \frac{\ln(m)}{\sqrt{m}}\right\} \text{ and } \mathcal{X}_n^* = \{x - x^* : x\in\mathcal{X}_n\}\cap\mathcal{U}_n^* \text{ for } n\in\mathcal{N}.$$

In the following we restrict to $n \geq n_0$ where $n_0\in\mathcal{N}$ is such that $\mathcal{U}_{n_0}^* \subseteq \mathcal{U}^*$.

Lemma 5.16. *For $d\in(0,d^*)$ we have*

$$\frac{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^2]}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]^2} = (1+o(1))\sqrt{\frac{d}{(2\pi)^2m^2\prod_s p^*(s)}}\sum_{x\in\mathcal{X}_n^*} e^{-\frac{m}{2}x^{\mathrm{t}}H_dx}.$$

Proof. First, notice that we can apply Lemma 5.15 to $\iota^*(\mathcal{U}^*)$. So, we need to show that the sum over $\mathcal{U}^*\setminus\mathcal{U}_n^*$ is negligible. Then we proceed to derive the asymptotics of the sum over $\mathcal{U}_n^*$. Obviously, we have $\gamma_{\min}(n)\to 0$ for $n\to\infty$ with $\gamma_{\min}(n) = \min_{x\notin\mathcal{U}_n^*}\gamma_d(x) > 0$, since $\gamma_d(x) = \Delta_d(\iota^*(x))$ is continuous and $\gamma_d(0_{[2]}) = 0$. The main objective of the proof is to show that $\gamma_{\min}(n)$ converges to zero sufficiently slow. But with $\gamma_d^{(2)}(x) = \frac{1}{2}x^{\mathrm{t}}H_dx$ from above (25) and for any $x\in\mathbb{R}^2$ we have

$$\gamma_d^{(2)}((v_1,v_2)x) = \frac{1}{2}(\lambda_1x_1^2+\lambda_2x_2^2) \geq \frac{\lambda_1}{2}\|x\|_2^2 = \frac{\lambda_1}{2}\|(v_1,v_2)x\|_2^2$$

since (v_1,v_2) is an orthonormal basis, so $\gamma_d^{(2)}(x) \geq \frac{\lambda_1}{2}\|x\|_2^2$ for all $x\in\mathbb{R}^2$. Now, for any sufficiently small $\varepsilon\in(0,1)$ let $c\in\mathbb{R}_{>0}$ be the constant for $\mathcal{B}_\varepsilon(0_{[2]})$ from Taylor's theorem applied to γ_d at $0_{[2]}$, then for any $x \in \mathcal{U}^* = \mathcal{B}_\varepsilon(0_{[2]}) \cap \mathcal{B}_\delta(0_{[2]})$, with $\delta = \frac{\varepsilon\lambda_1}{2c}$, we have $\gamma_d(x) \geq (1-\varepsilon)\gamma_d^{(2)}(x)$ since

$$\gamma_d(x) - (1-\varepsilon)\gamma_d^{(2)}(x) \geq \varepsilon\gamma_d^{(2)}(x) - c\|x\|_2^3 \geq \left(\frac{\varepsilon\lambda_1}{2} - c\delta\right)\|x\|_2^2 = 0.$$

In combination we have $\gamma_d(x) \geq \frac{(1-\varepsilon)\lambda_1}{2}\|x\|_2^2$ and using $p = \iota^*(x)$ hence

$$\lim_{n\to\infty} \sum_{x\notin\mathcal{U}_n^*} e(p) = \lim_{n\to\infty} \sum_{x\in\mathcal{U}^*\setminus\mathcal{U}_n^*} \sqrt{\frac{d}{(2\pi)^2 m^2 \prod_s p(s)}} e^{-m\gamma_d(x)} \leq \lim_{n\to\infty} Cme^{-\frac{(1-\varepsilon)\lambda_1}{2}\ln(m)^2} = 0,$$

by using (the proof of) Lemma 5.15 and some sufficiently large constant C. With this we have

$$\frac{\mathbb{E}[\boldsymbol{Z}_\mathrm{c}^2]}{\mathbb{E}[\boldsymbol{Z}_\mathrm{c}]^2} = (1+o(1)) \sum_{x\in\mathcal{X}_n^*} e(\iota^*(x)) = (1+o(1))\sqrt{\frac{d}{(2\pi)^2 m^2 \prod_s p^*(s)}} \sum_{x\in\mathcal{X}_n^*} e^{-m\gamma_d^{(2)}(x)},$$

where the last equivalence follows from the fact that the leading factor converges to the respective constant uniformly on $\mathcal{U}_n^*$ and by (26) on $\mathcal{U}^*$. □

Lemma 5.16 completes the analytical part of the proof. For the last, measure theoretic, part we recall the bijection ι_n from Lemma 5.11. The translation of the sum on the right-hand side of Lemma 5.16 into a Riemann sum and further into the integral $\int g_\infty(x)\mathrm{d}x$, where

$$g_\infty : \mathbb{R}^2 \to \mathbb{R}_{>0}, \quad y \mapsto \sqrt{\frac{d}{(2\pi)^2 \prod_s p^*(s)}} \exp\left(-\frac{1}{2}y^\mathrm{t} H_d y\right),$$

is essentially given by the grid $\mathcal{X}_n \subseteq (m^{-1}\mathbb{Z})^2 \subseteq \mathbb{R}^2$. We make this rigorous in the following.

Lemma 5.17. *We have*

$$\frac{\mathbb{E}[\boldsymbol{Z}_\mathrm{c}^2]}{\mathbb{E}[\boldsymbol{Z}_\mathrm{c}]^2} = (1+o(1)) \int g_\infty(x)\mathrm{d}x.$$

Proof. We start with the partition of $\mathbb{R}^2$ into the squares

$$\mathcal{Q}_{n,x} = \left\{x + \alpha_1 \begin{pmatrix}1\\0\end{pmatrix} + \alpha_2 \begin{pmatrix}0\\1\end{pmatrix} : \alpha \in \left[-\frac{1}{2m}, \frac{1}{2m}\right)^2\right\},\ x \in (m^{-1}\mathbb{Z})^2.$$

Next, we need a suitable selection of squares to cover the disc

$$x^* + \mathcal{U}_n^* = \left\{x^* + x : x \in \mathbb{R}^2, \|x\|_2 < \frac{\ln(m)}{\sqrt{m}}\right\} \subseteq \mathbb{R}^2$$

corresponding to the disc $\mathcal{U}_n^*$. For this purpose let $x_{\min}$, $x_{\max} \in (m^{-1}\mathbb{Z})^2$ be given by

$$x_{\min,1} = m^{-1}\left\lfloor m\left(x_1^* - \frac{\ln(m)}{\sqrt{m}}\right)\right\rfloor,\ x_{\min,2} = m^{-1}\left\lfloor m\left(x_2^* - \frac{\ln(m)}{\sqrt{m}}\right)\right\rfloor,$$
$$x_{\max,1} = m^{-1}\left\lceil m\left(x_1^* + \frac{\ln(m)}{\sqrt{m}}\right)\right\rceil,\ x_{\max,2} = m^{-1}\left\lceil m\left(x_2^* + \frac{\ln(m)}{\sqrt{m}}\right)\right\rceil.$$

Further, let $\mathcal{G}_n = (m^{-1}\mathbb{Z})^2 \cap ([x_{\min,1}, x_{\max,1}] \times [x_{\min,2}, x_{\max,2}])$. By the definition of $x_{\min}$ and $x_{\max}$ the points on the boundary are not in $x^* + \mathcal{U}_n^*$, which ensures that $x^* + \mathcal{U}_n^* \subseteq \mathcal{Q}_n$ with $\mathcal{Q}_n = \bigcup_{x\in\mathcal{G}_n} \mathcal{Q}_{n,x}$. Further, we have $\mathcal{Q}_- \subseteq \mathcal{Q}_n \subseteq \mathcal{Q}_+$ with

$$\mathcal{Q}_- = \left\{x \in \mathbb{R}^2 : \|x - x^*\|_\infty \le \frac{\ln(m)}{\sqrt{m}}\right\},\ \mathcal{Q}_+ = \left\{x \in \mathbb{R}^2 : \|x - x^*\|_\infty \le \frac{\ln(m)}{\sqrt{m}} + \frac{3}{2m}\right\},$$

which ensures that $\mathcal{Q}_n \subseteq \mathcal{X}$ for $n \in \mathcal{N}$ sufficiently large. Now, we translate the notions back to $\mathcal{X}^*$ using the bijection $\tau : \mathbb{R}^2 \to \mathbb{R}^2$, $x \mapsto x - x^*$, i.e. let $\mathcal{G}_n^* = \tau(\mathcal{G}_n)$, $\mathcal{Q}_{n,x}^* = \tau(\mathcal{Q}_{n,\tau^{-1}(x)})$ for $x \in \mathcal{G}_n^*$, $\mathcal{Q}_n^* = \tau(\mathcal{Q}_n)$, $\mathcal{Q}_-^* = \tau(\mathcal{Q}_-)$ and $\mathcal{Q}_+^* = \tau(\mathcal{Q}_+)$. This directly gives

$$\mathcal{Q}_{n,x}^* = \left\{x + \alpha_1 \begin{pmatrix}1\\0\end{pmatrix} + \alpha_2 \begin{pmatrix}0\\1\end{pmatrix} : \alpha \in \left[-\frac{1}{2m}, \frac{1}{2m}\right)^2\right\},\ x \in \mathcal{G}_n^*,\ \mathcal{Q}_n^* = \bigcup_{x\in\mathcal{G}_n^*} \mathcal{Q}_{n,x}^*,$$
$$\mathcal{Q}_-^* = \left\{x \in \mathbb{R}^2 : \|x\|_\infty \le \frac{\ln(m)}{\sqrt{m}}\right\},\ \mathcal{Q}_+^* = \left\{x \in \mathbb{R}^2 : \|x\|_\infty \le \frac{\ln(m)}{\sqrt{m}} + \frac{3}{2m}\right\},$$

and $\mathcal{U}_n^* \subseteq \mathcal{Q}_-^* \subseteq \mathcal{Q}_n^* \subseteq \mathcal{Q}_+^* \subseteq \mathcal{X}^*$ for $n \in \mathcal{N}$ sufficiently large. Further, with Lemma 5.16 and the definition of $\gamma_d^{(2)}$ we now have

$$\frac{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^2]}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]^2} = (1 + o(1)) \sum_{x\in\mathcal{G}_n^*} \sqrt{\frac{d}{(2\pi)^2 m^2 \prod_s p^*(s)}} \exp\left(-\frac{1}{2} m x^{\mathrm{t}} H_d x\right).$$

Finally, we need to adjust the scaling to turn the sum on the right hand side into a Riemann sum. For this purpose let $\sigma : \mathbb{R}^2 \to \mathbb{R}^2$, $x \mapsto \sqrt{m}x$, further $\mathcal{G}_n' = \sigma(\mathcal{G}_n^*)$, $\mathcal{Q}_{n,x}' = \sigma(\mathcal{Q}_{n,\sigma^{-1}(x)}^*)$ for $x \in \mathcal{G}_n'$, $\mathcal{Q}_n' = \sigma(\mathcal{Q}_n^*)$, $\mathcal{Q}_-' = \sigma(\mathcal{Q}_-^*)$ and $\mathcal{Q}_+' = \sigma(\mathcal{Q}_+^*)$. This directly gives

$$\mathcal{Q}_{n,x}' = \left\{x + \alpha_1 \begin{pmatrix}1\\0\end{pmatrix} + \alpha_2 \begin{pmatrix}0\\1\end{pmatrix} : \alpha \in \left[-\frac{1}{2\sqrt{m}}, \frac{1}{2\sqrt{m}}\right)^2\right\},\ x \in \mathcal{G}_n',\ \mathcal{Q}_n' = \bigcup_{x\in\mathcal{G}_n'} \mathcal{Q}_{n,x}',$$
$$\mathcal{Q}_-' = \left\{x \in \mathbb{R}^2 : \|x\|_\infty \le \ln(m)\right\},\ \mathcal{Q}_+' = \left\{x \in \mathbb{R}^2 : \|x\|_\infty \le \ln(m) + \frac{3}{2\sqrt{m}}\right\},$$

and $\mathcal{Q}_-' \subseteq \mathcal{Q}_n' \subseteq \mathcal{Q}_+'$. Using that $m x^{\mathrm{t}} H_d x = \sigma(x)^{\mathrm{t}} H_d \sigma(x)$ for all $x \in \mathcal{G}_n^*$ and further that the area of $\mathcal{Q}_{n,x}'$ is m^{-1} for all $x \in \mathcal{G}_n'$ we have

$$\frac{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}^2]}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]^2} = (1 + o(1)) \sum_{x\in\mathcal{G}_n'} \sqrt{\frac{d}{(2\pi)^2 m^2 \prod_s p^*(s)}} \exp\left(-\frac{1}{2} x^{\mathrm{t}} H_d x\right) = (1 + o(1)) \int g_n(y) \mathrm{d}y,$$
$$g_n(y) = \sum_{x\in\mathcal{G}_n'} \mathbb{1}\{y \in \mathcal{Q}_{n,x}'\} \sqrt{\frac{d}{(2\pi)^2 \prod_s p^*(s)}} \exp\left(-\frac{1}{2} x^{\mathrm{t}} H_d x\right),\ y \in \mathbb{R}^2.$$

In order to show that $\int g_n(y)\mathrm{d}y$ converges to $\int g_\infty(y)\mathrm{d}y$ we recall from Lemma 5.14 that H_d is positive definite, which ensures that $\int g_\infty(y)\mathrm{d}y$ exists and is finite. Now, using Taylor's theorem with order 0 and the Lagrange form of the first order remainder with the fact that the absolutes of the first derivatives of g_∞ are bounded from above yields a constant $c \in \mathbb{R}_{>0}$ such that for all $n \in \mathcal{N}$ and all

$y \in \mathcal{Q}'_n$, with $x \in \mathcal{G}'_n$ such that $y \in \mathcal{Q}'_{n,x}$, we have

$$\|g_\infty(y) - g_n(y)\|_\infty = \|g_\infty(y) - g_\infty(x)\|_\infty \leq \frac{c}{\sqrt{m}}.$$

This bound directly suggests that

$$\left|\int \mathbb{1}\{y \in \mathcal{Q}'_{n,x}\} g_\infty(y)\mathrm{d}y - \int \mathbb{1}\{y \in \mathcal{Q}'_{n,x}\} g_n(y)\mathrm{d}y\right| \leq cm^{-\frac{3}{2}},$$
$$\left|\int \mathbb{1}\{y \in \mathcal{Q}'_n\} g_\infty(y)\mathrm{d}y - \int \mathbb{1}\{y \in \mathcal{Q}'_n\} g_n(y)\mathrm{d}y\right| \leq \frac{c}{\sqrt{m}}\left(2\ln(m) + \frac{3}{\sqrt{m}}\right)^2 \text{ and}$$
$$\left|\int g_\infty(y)\mathrm{d}y - \int g_n(y)\mathrm{d}y\right| \leq \frac{c}{\sqrt{m}}\left(2\ln(m) + \frac{3}{\sqrt{m}}\right)^2 + \int \mathbb{1}\{y \notin \mathcal{Q}'_n\} g_\infty(y)\mathrm{d}y.$$

In particular the last bound suggests that $\int g_n(y)\mathrm{d}y \to \int g_\infty(y)\mathrm{d}y$ since the error on the right-hand side tends to zero as n tends to infinity. □

The only remaining part of the proof is to compute $\int g_\infty(x)\mathrm{d}x$. However, instead of computing the main quantity $\det(H_d)$, i.e. the determinant of a two by two matrix, directly, we present an arguably more insightful and in particular generalizable argument.

Lemma 5.18. *We have* $\int g_\infty(x)\mathrm{d}x = \sqrt{\frac{k-1}{k-d}}$.

Proof. The Gaussian integral gives

$$\int g_\infty(x)\mathrm{d}x = \sqrt{\frac{d}{(2\pi)^2\prod_s p^*(s)}}\sqrt{\frac{(2\pi)^2}{\det(H_d)}} = \sqrt{\frac{d}{\det(H_d)\prod_s p^*(s)}}.$$

In order to compute $\det(H_d)$ we want to extract $\det(D_{p^*}^{-1})$ which in turn requires to split away (b_1, b_2). Hence, we are ultimately interested in the decomposition $H_d = (b_1, b_2)^{\mathrm{t}} D_{p^*}^{-1} M (b_1, b_2)$ with $M = I_{\{0,1,2\}} - \frac{(d-1)k}{d} D_{p^*} W^{\mathrm{t}} D_{p^*_{\mathrm{e}}}^{-1} W$ and $I_{\{0,1,2\}} = D_{1_{\{0,1,2\}}}$ denoting the identity matrix. First, in order to extract (b_1, b_2) we add $b_0 = p^*$. Recall that b_1, b_2 span $1_{\{0,1,2\}}^{\perp}$, hence $B = (b_0, b_1, b_2) \in \mathbb{R}^{\{0,1,2\}\times\{0,1,2\}}$ is a basis of $\mathbb{R}^{\{0,1,2\}}$. Further, since $p^*_{\mathrm{e}} = Wp^*$ and W is column stochastic, hence W^{t} is row stochastic, we have $Mp^* = \lambda_0 p^*$ with $\lambda_0 = 1 - \frac{(d-1)k}{d}$ and further $D_{p^*}^{-1} M p^* = \lambda_0 1_{\{0,1,2\}}$, so $b_0^{\mathrm{t}} D_{p^*}^{-1} M b_0 = \lambda_0$ since p^* is normalized, and $b_s^{\mathrm{t}} D_{p^*}^{-1} M b_0 = b_0^{\mathrm{t}} D_{p^*}^{-1} M b_s = 0$ since $b_s \in 1_{\{0,1,2\}}^{\perp}$ for $s \in \{1,2\}$ and $D_{p^*}^{-1} M$ is symmetric. Combined, this gives

$$B^{\mathrm{t}} D_{p^*}^{-1} M B = \begin{pmatrix} \lambda_0 & 0_{\{0\}\times[2]} \\ 0_{[2]\times\{0\}} & H_d \end{pmatrix},$$

so $\det(B^{\mathrm{t}} D_{p^*}^{-1} M B) = \lambda_0 \det(H_d)$. Notice that for $d \neq \frac{k}{k-1}$ (which is the case for all $d \in \mathbb{Z}_{>0}$) we have $\lambda_0 \neq 0$. Next, we compute $\det(B)$ using Gaussian elimination. Stepwise, we let $\tilde{b}_1 = d^{-1} b_1$ by right multiplication of $G_1 \in \mathbb{R}^{\{0,1,2\}\times\{0,1,2\}}$, then $\tilde{b}_2 = b_2 + 2\tilde{b}_1$ by right multiplication of $G_2 \in \mathbb{R}^{\{0,1,2\}\times\{0,1,2\}}$ and finally $\tilde{b}_0 = b_0 - p^*(1)\tilde{b}_1 - p^*(2)\tilde{b}_2$ by right multiplication with $G_0 \in \mathbb{R}^{\{0,1,2\}\times\{0,1,2\}}$. Combined, we have $\tilde{B} = (\tilde{b}_0, \tilde{b}_1, \tilde{b}_2) = BG_1G_2G_0$, where $\tilde{B}$ is an upper triangular matrix with ones on the diagonal due to the scaling of b_1 and normalization of p^* and hence $\det(\tilde{B}) = 1$, further G_1 is a diagonal matrix with $\det(G_1) = d^{-1}$, G_2 is an upper triangular matrix with $\det(G_2) = 1$ and G_0 is a lower triangular matrix with $\det(G_0) = 1$, so $\det(B) = d$. This directly yields

$$\det(H_d) = \frac{\det(M)d^2}{\lambda_0 \prod_s p^*(s)}.$$

Recall that $b_0 = p^*$ is an eigenvector of M with eigenvalue λ_0, and further b_2 is an eigenvector of M with eigenvalue $\lambda_2 = 1$ since $Wb_2 = 0$. The remaining eigenvector is given by Lemma 5.12 as follows. First, notice that $W' = D_{p^*}W^{\mathrm{t}}D_{p^*_{\mathrm{e}}}^{-1} \in \mathbb{R}^{\{0,1,2\}\times\{0,1\}^2}$ is given by $W'_{s,x} = \frac{W_{x,s}p^*(s)}{p^*_{\mathrm{e}}(x)}$ for $s \in \{0,1,2\}$ and $x \in \{0,1\}^2$, so in particular W' is a column stochastic transition probability matrix.

Aside, the transition probabilities W and W' have a very intuitive interpretation. For this purpose let $\boldsymbol{X}^*$ be a uniformly random satisfying constraint assignment for the squared constraint satisfaction problem introduced in Section 5.5.1, i.e. $\boldsymbol{X}^* \in (\{0,1\}^2)^k$ with $\sum_i \boldsymbol{X}^*_{i,1} = \sum_i \boldsymbol{X}^*_{i,2} = 2$. Further, let $\boldsymbol{S}^*$ be the overlap of $\boldsymbol{X}^*$, i.e. $\boldsymbol{S}^* = |\boldsymbol{X}^{*-1}(1,1)| \in \{0,1,2\}$, and let $\boldsymbol{I} \in [k]$ be the uniformly random coordinate. Then $\mathbb{P}(\boldsymbol{S}^* = s, \boldsymbol{X}^*_{\boldsymbol{I}} = x) = W_{x,s}p^*(s)$ for $s \in \{0,1,2\}$ and $x \in \{0,1\}^2$, $\boldsymbol{S}^* \sim p^*$, $\boldsymbol{X}^*_{\boldsymbol{I}} \sim p^*_{\mathrm{e}}$, further $\mathbb{P}(\boldsymbol{X}^*_{\boldsymbol{I}} = x|\boldsymbol{S}^* = s) = W_{x,s}$ and $\mathbb{P}(\boldsymbol{S}^* = s|\boldsymbol{X}^*_{\boldsymbol{I}} = x) = W'_{s,x}$. In particular, the columns of W' are hypergeometric distributions up to a shift of one for $x = (1,1)$.

Now, we easily verify that the Markov chain induced by $W'W$ has the stationary distribution p^* and the Markov chain induced by WW' has the stationary distribution p^*_{e}. For any distribution $p \in \mathcal{P}(\{0,1\}^2)$ we have $W'p \in \mathcal{P}(\{0,1,2\})$ and hence $WW'p = p^*_{\mathrm{e}} + \alpha w$ for some $\alpha \in \mathbb{R}$ by Lemma 5.12, so $WW'(p - p^*_{\mathrm{e}}) = \alpha w$. Choosing $p \in \mathcal{W}$ directly yields that $(p - p^*_{\mathrm{e}})$ is in the span of w and hence w needs to be an eigenvector of WW'. The corresponding eigenvalue $\lambda'_1 = \frac{1}{k-1}$, determined by $WW'w = \lambda'_1 w$, can be thought of as the rate of convergence to the stationary distribution p^*_{e}. From this we directly get the eigenvector $v_1 = W'w$ for $W'W$ with eigenvalue λ'_1 and by that the eigenvector v_1 for M with eigenvalue $\lambda_1 = 1 - \frac{(d-1)k}{d(k-1)} = \frac{k-d}{d(k-1)}$. Knowing all eigenvalues of M this gives

$$\det(H_d) = \frac{\lambda_1 d^2}{\prod_s p^*(s)} \text{ and hence } \int g_\infty(x)\mathrm{d}x = \sqrt{\frac{1}{d\lambda_1}} = \sqrt{\frac{k-1}{k-d}}.$$

□

Notice that a direct corollary of Lemma 5.18 is that $\eta_{\chi^2} = \frac{1}{k-1}$ or equivalently H_d is positive definite for $d < k$, e.g. since $\det(H_d) > 0$ for $d < k$ and H_d is positive definite for $d = 1$, hence no eigenvalue can change sign using the continuity of H_d with respect to d. For $d = k$ the determinant is zero (so an eigenvalue is 0 and $\Delta^{(2)}_d \equiv 0$ along the direction of the eigenvector), hence $\eta_{\chi^2} = \frac{k}{(k-1)k} = \frac{1}{k-1}$, which is the squared Hirschfeld-Gebelein-Rényi maximal correlation [82]. Finally, combining Lemma 5.17 with Lemma 5.18 completes the proof of Proposition 5.6.

5.5.6 Proof of Proposition 5.5. We start with a characterization of the stationary points of Δ_d for any $d \in \mathbb{R}_{>0}$. In order to determine these, we first determine the stationary points of the restriction of Δ_d to overlap distributions with the same fixed edge distribution. For this purpose, recall the line $\mathcal{W} \subseteq \mathcal{P}(\{0,1\}^2)$ of attainable edge distributions and the lines $\mathcal{P}_q = \iota(\mathcal{X}_q) = \{p \in \mathcal{P}(\{0,1,2\}) : p_{\mathrm{e}} = q\}$ of overlap distributions with fixed edge distribution $q \in \mathcal{W}$ from Lemma 5.12. Further, let $\Delta_{d,q} : \mathcal{P}_q \to \mathbb{R}$ denote the restriction of Δ_d to $\mathcal{P}_q$. For $x \in \mathbb{R}_{>0}$ let $p_x \in \mathcal{P}(\{0,1,2\})$ be given by $p_x(s) = p^*(s)x^s / \sum_s p^*(s)x^s$, $s \in \{0,1,2\}$, further let $p_0 = p^{(0)}$, $p_\infty = p^{(2)}$, and $\mathcal{P}_{\min} = \{p_x : x \in [0,\infty]\}$. Finally, let $\iota_{\mathrm{rp}} : [0,\infty] \to \mathcal{P}_{\min}$, $x \mapsto p_x$, denote the induced map and $\iota_{\mathrm{pe}} : \mathcal{P}_{\min} \to \mathcal{W}$, $p \mapsto p_{\mathrm{e}}$, the corresponding edge distributions.

Lemma 5.19. *For all $q \in \mathcal{W} \setminus \{p_{\mathrm{e}}^{(0)}, p_{\mathrm{e}}^{(2)}\}$ the map $\Delta_{d,q}$ has a unique stationary point $p_q \in \mathcal{P}_q$ that is a global minimum. The unique global minimizer of $\Delta_{d,p_{\mathrm{e}}^{(s)}}$ is $p_{p_{\mathrm{e}}^{(s)}} = p^{(s)}$ for $s \in \{0,2\}$. Further, we have $\mathcal{P}_{\min} = \{p_q : q \in \mathcal{W}\}$ and the maps ι_{rp}, ι_{pe} are bijections.*

Proof. Recall from Lemma 5.12 that $\mathcal{P}_q$ is one-dimensional for $q \in \mathcal{W} \setminus \{p_{\mathrm{e}}^{(0)}, p_{\mathrm{e}}^{(2)}\}$. Further, the map $\Delta_{d,q}$ is strictly convex since the KL divergence $D_{\mathrm{KL}}(p\|p^*)$ (respectively $x\ln(x)$) is and further $D_{\mathrm{KL}}(p_{\mathrm{e}}\|p_{\mathrm{e}}^*) = D_{\mathrm{KL}}(q\|p_{\mathrm{e}}^*)$ is constant. Now, fix an interior point $p_\circ \in \mathcal{P}_q$ and let a boundary point $p_{\mathrm{b}} \in \mathcal{P}_q$ be given. Then p_{b} is not fully supported since it is on the boundary of $\mathcal{P}(\{0,1,2\}$ and hence the derivative of $D_{\mathrm{KL}}(\alpha p_\circ + (1-\alpha)p_{\mathrm{b}}\|p^*)$ tends to $-\infty$ as α tends to 0, which shows that $\Delta_{d,q}$ is not minimized on the boundary. Hence, we know that there exists exactly one stationary point $p_q \in \mathcal{P}_q$ and that $\Delta_{d,q}(p)$ is minimal iff $p = p_q$. As discussed in Lemma 5.12 we have $\mathcal{P}_q = \{p^{(s)}\}$ for $q = p_{\mathrm{e}}^{(s)}$ and $s \in \{0,2\}$, so $p_q = p^{(s)}$ is obviously the unique global minimizer of $\Delta_{d,q}$ in this case and further $\Delta_{d,q}$ has no stationary points (since $\mathcal{P}_q$ has empty interior). This shows that the map $q \mapsto p_q$ for $q \in \mathcal{W}$ is a bijection.

Further, for q in the interior of $\mathcal{W}$ the stationary point p_q is fully supported and the unique root of the first derivative of $\Delta_{d,q}$ in the direction b_2 from (22), i.e.

$$\ln\left(\frac{p_q(0)}{p^*(0)}\right) + \ln\left(\frac{p_q(2)}{p^*(2)}\right) = 2\ln\left(\frac{p_q(1)}{p^*(1)}\right) \text{ or equivalently } \frac{p_q(2)/p^*(2)}{p_q(1)/p^*(1)} = \frac{p_q(1)/p^*(1)}{p_q(0)/p^*(0)}.$$

Let $\mathcal{P}'_{\min}$ denote the set of all fully supported $p \in \mathcal{P}(\{0,1,2\})$ satisfying $\frac{p(2)/p^*(2)}{p(1)/p^*(1)} = \frac{p(1)/p^*(1)}{p(0)/p^*(0)}$, i.e. our set of candidates for stationary points. Now, for $p \in \mathcal{P}'_{\min}$ let $q = p_{\mathrm{e}}$, then we obviously have $p \in \mathcal{P}_q$ and p is a root of the first derivative of $\Delta_{d,q}$ in the direction b_2, so p is the unique root and $p = p_q$. Hence, the map $\iota'_{\mathrm{pe}} : \mathcal{P}'_{\min} \to \mathcal{W}$, $p \mapsto p_{\mathrm{e}}$, is a bijection (up to the corners of $\mathcal{W}$) with inverse $q \mapsto p_q$. Now, let $\iota_{\mathrm{pr}} : \mathcal{P}'_{\min} \to \mathbb{R}_{>0}$, $p \mapsto x_p$, with $x_p = \frac{p(1)p^*(0)}{p^*(1)p(0)}$. Notice that ι_{pr} is surjective since for any $x \in \mathbb{R}_{>0}$ we have

$$\frac{p_x(2)/p^*(2)}{p_x(1)/p^*(1)} = \frac{x^2}{x} = x = \frac{p_x(1)/p^*(1)}{p_x(0)/p^*(0)}$$

and hence $p_x \in \mathcal{P}'_{\min}$. To show that ι_{pr} is injective let $p \in \mathcal{P}'_{\min}$ and $x = x_p$. Using the definition of x_p and the defining property of $\mathcal{P}'_{\min}$ we get

$$p(0) = p^*(0)\frac{p(0)}{p^*(0)},\ p(1) = p^*(1)x\frac{p(0)}{p^*(0)},\ p(2) = p^*(2)x\frac{p(1)}{p^*(1)} = p^*(2)x^2\frac{p(0)}{p^*(0)}, \text{ so}$$

$$p(s) = \frac{p(s)}{p(0)+p(1)+p(2)} = \frac{p^*(s)x^s}{\sum_s p^*(s)x^s} = p_x(s),\ s \in \{0,1,2\}.$$

This shows that $\mathcal{P}_{\min} = \mathcal{P}'_{\min} \cup \{p_0, p_\infty\}$, that ι_{rp} is a bijection with inverse ι_{pr} (canonically extended to the endpoints), and finally that $\iota_{\mathrm{pe}} = \iota'_{\mathrm{pe}}$ is a bijection as well. □

Lemma 5.19 has a few immediate consequences. For one, the only minimizers of Δ_d in the direction b_2, from (22), on the boundary are $p^{(0)}$ and $p^{(2)}$, while all other boundary points are maximizers in the direction b_2, hence if p is a global minimizer of Δ_d on the boundary, we have $p \in \{p^{(0)}, p^{(2)}\}$. Further, all stationary points of Δ_d are either local minima or saddle points. Finally, we have $p \in \mathcal{P}_{\min}$ for any stationary point $p \in \mathcal{P}(\{0,1,2\})$ of Δ_d since then also the derivative in the direction of b_2 vanishes.

For the upcoming characterization of the stationary points of Δ_d let

$$\iota_{\mathrm{rr}}:\mathbb{R}_{>0}\to\mathbb{R}_{>0},\quad \iota_{\mathrm{rr}}(x)=\iota^*_{\mathrm{rr}}(x)^{\frac{d-1}{d}},\ \iota^*_{\mathrm{rr}}(x)=\frac{p_{x,\mathrm{e}}(1,1)p_{x,\mathrm{e}}(0,0)}{p_{x,\mathrm{e}}(1,0)p_{x,\mathrm{e}}(0,1)}.$$

Notice that $\iota_{\mathrm{rr}}(x)\in\mathbb{R}_{>0}$ for $x\in\mathbb{R}_{>0}$ since then p_x is fully supported and hence $p_{x,\mathrm{e}}$ is fully supported by Lemma 5.12. Finally, let $\mathcal{X}_{\mathrm{st}}=\{x\in\mathbb{R}_{>0}:\iota_{\mathrm{rr}}(x)=x\}$ denote the fixed points of ι_{rr} and $\mathcal{P}_{\mathrm{st}}=\{p_x:x\in\mathcal{X}_{\mathrm{st}}\}$ the corresponding distributions. Notice that $p_x=p^*$ for $x=1$ and further $\iota^*_{\mathrm{rr}}(1)=1$, i.e. $\iota_{\mathrm{rr}}(1)=1$ for all $d\in\mathbb{R}_{>0}$, hence $1\in\mathcal{X}_{\mathrm{st}}$ and $p^*\in\mathcal{P}_{\mathrm{st}}$ for all $d\in\mathbb{R}_{>0}$.

Lemma 5.20. *The stationary points of Δ_d are given by $\mathcal{P}_{\mathrm{st}}$.*

Proof. Using Lemma 5.19, a fully supported distribution $p\in\mathcal{P}(\{0,1,2\})$ is a stationary point of Δ_d iff there exists $x\in\mathbb{R}_{>0}$ such that $p=p_x$ and the derivative of Δ_d at p_x in the direction b_1 vanishes, i.e. p_x is a solution of

$$0=\left(\left(\ln\left(\frac{p_x(s)}{p^*(s)}\right)\right)^{\mathrm{t}}_{s\in\{0,1,2\}}-\frac{(d-1)k}{d}\left(\ln\left(\frac{p_{x,\mathrm{e}}(y)}{p^*_{\mathrm{e}}(y)}\right)\right)^{\mathrm{t}}_{y\in\{0,1\}^2}W\right)b_1,$$

where we used the chain rule for multivariate calculus, that W is column stochastic and that $b_1\in 1^{\perp}_{\{0,1,2\}}$. Recall from Section 5.5.3, e.g. from the proof of Lemma 5.11, that $Wb_1=\frac{d}{k}w$, hence computing the dot product with b_1 gives

$$0=d\ln(x)-(d-1)\ln\left(\frac{p_{x,\mathrm{e}}(1,1)p_{x,\mathrm{e}}(0,0)}{p_{x,\mathrm{e}}(1,0)p_{x,\mathrm{e}}(0,1)}\right).$$

Obviously, equality holds if and only if $x\in\mathcal{X}_{\mathrm{st}}$, hence p is a stationary point of Δ_d iff $p\in\mathcal{P}_{\mathrm{st}}$. □

Lemma 5.20 does not only allow to translate the stationary points of Δ_d into fixed points of ι_{rr}, it also allows to translate the types as follows.

Lemma 5.21. *Fix $x\in\mathbb{R}_{>0}$. We have $\iota_{\mathrm{rr}}(x)<x$ iff $(\Delta_d\circ\iota_{\mathrm{rp}})'(x)>0$, $\iota_{\mathrm{rr}}(x)>x$ iff $(\Delta_d\circ\iota_{\mathrm{rp}})'(x)<0$, and $\iota_{\mathrm{rr}}(x)=x$ iff $(\Delta_d\circ\iota_{\mathrm{rp}})'(x)=0$.*

Proof. Fix $x\in\mathbb{R}_{>0}$. The proof of Lemma 5.20 directly suggests that the first derivative of Δ_d at p_x in the direction b_1 is strictly positive iff

$$0<\ln(x)-\frac{d-1}{d}\ln(\iota^*_{\mathrm{rr}}(x)),$$

which holds iff $\iota_{\mathrm{rr}}(x)<x$. We're left to establish that the direction of ι_{rp} is consistent with b_1. Intuitively, using Lemma 5.12 and Lemma 5.19 we can argue that $x\mapsto p_{x,\mathrm{e}}(1,1)$ is a bijection and hence either increasing or decreasing. Taking the limits $x\to 0$ and $x\to\infty$ suggests that it is increasing, hence with $c\in\mathbb{R}^2$ given by $\iota'_{\mathrm{rp}}(x)=(b_1,b_2)c$, we know that $c_1\geq 0$.

Formally, we quantify the direction of ι_{rp}. For this purpose we compute the derivative of ι_{rp} at $x\in\mathbb{R}_{>0}$, given by

$$\iota'_{\mathrm{rp}}(x)=\left(\frac{sp^*(s)x^{s-1}\sum_{s'\in\{0,1,2\}}p^*(s')x^{s'}-p^*(s)x^s\sum_{s'\in\{0,1,2\}}s'p^*(s')x^{s'-1}}{\left(\sum_{s'\in\{0,1,2\}}p^*(s')x^{s'}\right)^2}\right)_{s\in\{0,1,2\}}.$$

Notice that $v = \iota'_{\mathrm{rp}}(x) \in 1^{\perp}_{\{0,1,2\}}$, since $\iota_{\mathrm{rp}}(\mathbb{R}_{>0}) \subseteq \mathcal{P}(\{0,1,2\})$ or by computing $\sum_s v_s = 0$ directly. Now, let $c \in \mathbb{R}^2$ be given by $v = (b_1, b_2)c$. This directly gives $c_2 = v_2$ and hence $c_1 = d^{-1}(v_1 + 2v_2) = d^{-1}\sum_s sv_s$. Now, notice that

$$\begin{aligned} S = dxc_1 &= \sum_{s,s' \in \{0,1,2\}} p_x(s)p_x(s')s(s-s') \\ &= \sum_{s>s'} p_x(s)p_x(s')s(s-s') - \sum_{s>s'} p_x(s)p_x(s')s'(s-s') = \sum_{s>s'} p_x(s)p_x(s')(s-s')^2 > 0, \end{aligned}$$

which directly gives $c_1 = \frac{S}{dx} \in \mathbb{R}_{>0}$. Now, with $\nabla = \left(\frac{\partial \Delta_d}{\partial p(s)}(p_x)\right)_{s\in\{0,1,2\}} \in \mathbb{R}^{\{0,1,2\}}$ denoting the partial derivatives of Δ_d at p_x and using the chain rule we have

$$(\Delta_d \circ \iota_{\mathrm{rp}})'(x) = \nabla^{\mathrm{t}} \iota'_{\mathrm{rp}}(x) = c_1 \nabla^{\mathrm{t}} b_1 + c_2 \nabla^{\mathrm{t}} b_2 = c_1 \nabla^{\mathrm{t}} b_1,$$

since the derivative $\nabla^{\mathrm{t}} b_2$ of Δ_d at p_x in the direction b_2 is zero, hence we have $(\Delta_d \circ \iota_{\mathrm{rp}})'(x) > 0$ iff the derivative $\nabla^{\mathrm{t}} b_1$ of Δ_d at p_x in the direction b_1 is strictly positive, which is the case iff $\iota_{\mathrm{rr}}(x) < x$. □

Lemma 5.21 with Lemma 5.20 shows that control over ι_{rr} gives complete control over the location and characterization of the stationary points of Δ_d. However, instead of solving the fixed point equation given by ι_{rr} directly, we use a slight modification inspired by the belief propagation algorithm applied to the constraint satisfaction discussed in Section 5.5.1 and initialized with uniform messages.

For this purpose let $N \in \mathbb{Z}_{\geq 0}$, further N_1, $N_2 \in [N]_0$ and the hypergeometric distribution $p_{N,N_1,N_2} \in \mathcal{P}(\mathbb{Z})$ be given by

$$p_{N,N_1,N_2}(s) = \frac{\binom{N_1}{s}\binom{N-N_1}{N_2-s}}{\binom{N}{N_2}} = \frac{\binom{N}{N-N_1-N_2+s,N_1-s,N_2-s,s}}{\binom{N}{N_1}\binom{N}{N_2}} \text{ for } s \in \mathbb{Z}.$$

The latter form directly shows that $p_{N,N_1,N_2} = p_{N,N_2,N_1}$. Now, for $y \in \{0,1\}^2$ let $p^*_y \in \mathcal{P}(\{0,1,2\})$ be given by $p^*_{(1,1)}(s) = p_{k-1,1,1}(s-1)$, $p^*_{(1,0)}(s) = p_{k-1,1,2}(s)$, $p^*_{(0,1)}(s) = p_{k-1,2,1}(s)$, $p^*_{(0,0)}(s) = p_{k-1,2,2}(s)$ for $s \in \{0,1,2\}$. Hence, $p^*_y(s)$ gives the probability of seeing a certain overlap $s \in \{0,1,2\}$ when drawing two satisfying constraint assignments for the standard problem uniformly and independently, but knowing the pair y of values of one (fixed or random) coordinate. In particular, this explains why $p^*_{(1,1)}(0) = 0$. Further, notice that $p^*_{(1,0)} = p^*_{(0,1)}$ and finally that the matrix W' in the proof of Lemma 5.18 is given by $W' = (p^*_y)_{y\in\{0,1\}^2}$.

On the other hand, for $y \in \{0,1\}^2$ let $p'_y \in \mathcal{P}(\{0,1,2\})$ be given by $p'_{(1,1)}(s) = p_{k-1,1,1}(s)$ for $s \in \{0,1,2\}$ and further $p'_y = p^*_y$ for $y \in \{0,1\}^2 \setminus \{(1,1)\}$. Hence, knowing the value y of one (fixed or random) coordinate, $p'_y(s)$ gives the probability of seeing a certain overlap *on the remaining coordinates*. This explains both why $p'_{(1,1)}(s) = p^*_{(1,1)}(s-1)$ and $p'_y(s) = p^*_y(s)$ for $y \neq (1,1)$. Finally, for a distribution $p \in \mathcal{P}(\{0,1,2\})$ let $f_p : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $x \mapsto \sum_{s\in\{0,1,2\}} p(s)x^s$, be its probability generating function, and further $\iota_{\mathrm{BP}} : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ given by

$$\iota_{\mathrm{BP}}(x) = \iota^*_{\mathrm{BP}}(x)^{d-1},\ \iota^*_{\mathrm{BP}}(x) = \frac{f_{p'_{(1,1)}}(x) f_{p'_{(0,0)}}(x)}{f_{p'_{(1,0)}}(x) f_{p'_{(0,1)}}(x)}, \text{ for } x \in \mathbb{R}_{>0}.$$

Lemma 5.22. *Fix $x \in \mathbb{R}_{>0}$. Then we have $\iota_{\mathrm{rr}}(x) < x$ iff $\iota_{\mathrm{BP}}(x) < x$, $\iota_{\mathrm{rr}}(x) > x$ iff $\iota_{\mathrm{BP}}(x) > x$, and $\iota_{\mathrm{rr}}(x) = x$ iff $\iota_{\mathrm{BP}}(x) = x$.*

Proof. First, notice that the normalization constant of p_x cancels out in ι^*_{rr}, as does the normalization constant $\binom{k}{2}^2$ of p^*. Further, with $v = (k-4+s, 2-s, 2-s, s)^{\mathrm{t}} \in \mathbb{R}^{\{0,1\}^2}$ we have $W_{y,s}\binom{k}{v} = \frac{v_y}{k}\binom{k}{v} = \binom{k-1}{v-(\delta_{y,z})_z}$ for $y \in \{0,1\}^2$, $s \in \{0,1,2\}$, and thereby

$$\iota^*_{\mathrm{rr}}(x) = \frac{\left(\sum_s \binom{k-1}{k-4+s,2-s,2-s,s-1} x^s\right)\left(\sum_s \binom{k-1}{k-5+s,2-s,2-s,s} x^s\right)}{\left(\sum_s \binom{k-1}{k-4+s,2-s,1-s,s} x^s\right)\left(\sum_s \binom{k-1}{k-4+s,1-s,2-s,s} x^s\right)} \text{ for } x \in \mathbb{R}_{>0}.$$

Now, since the normalization constants cancel out in total, this directly gives

$$\iota^*_{\mathrm{rr}}(x) = \frac{f_{p^*_{(1,1)}}(x) f_{p^*_{(0,0)}}(x)}{f_{p^*_{(1,0)}}(x) f_{p^*_{(0,1)}}(x)} = x\iota^*_{\mathrm{BP}}(x) \text{ for } x \in \mathbb{R}_{>0},$$

using that $p^*_{(1,1)}(s) = p'_{(1,1)}(s-1)$ for $s \in \{0,1,2\}$, hence $f_{p^*_{(1,1)}}(x) = x f_{p'_{(1,1)}}(x)$, and $p^*_y(s) = p'_y(s)$ for $y \neq (1,1)$. Now, we have $x = \iota_{\mathrm{rr}}(x)$ iff $x = x^{\frac{d-1}{d}} \iota_{\mathrm{BP}}(x)^{\frac{1}{d}}$, which holds iff $x^{\frac{1}{d}} = \iota_{\mathrm{BP}}(x)^{\frac{1}{d}}$, which then again is equivalent to $x = \iota_{\mathrm{BP}}(x)$. Equivalence of the inequalities follows analogously. □

The following part is dedicated to the identification of the fixed points of ι_{BP}, and the only part where we actually require $r = 2$ with the occupation number r as defined in Section 2.3.1. We start with a discussion of ι^*_{BP}. For this purpose let g^*_1, $g^*_k : \mathbb{R}_{\geq 1} \to \mathbb{R}_{\geq 1}$ be given by

$$g^*_1(x) = \frac{1}{k-1}(x-1) + 1 \text{ and } g^*_k(x) = \frac{13k-12}{27(k-1)(k-2)}(x-k) + \frac{2(7k-12)}{9(k-2)},\ x \in \mathbb{R}_{\geq 1}.$$

Lemma 5.23. *For any $k \in \mathbb{Z}_{\geq 4}$ we have*

$$\lim_{x \to 0} \iota^*_{\mathrm{BP}}(x) = \frac{k-4}{k-3},\ \iota^*_{\mathrm{BP}}(1) = 1 = g^*_1(1) \text{ and } \iota^*_{\mathrm{BP}}(k) = g^*_k(k).$$

For the first derivative ${\iota^*}'_{\mathrm{BP}}$ we have

$${\iota^*}'_{\mathrm{BP}}(1) = {g^*}'_1(1), \quad {\iota^*}'_{\mathrm{BP}}(k) = {g^*}'_k(k) \text{ and } \lim_{x \to \infty} {\iota^*}'_{\mathrm{BP}}(x) = \frac{1}{2(k-2)}.$$

Moreover, for the second derivative we have

$${\iota^*}''_{\mathrm{BP}}(x) < 0 \text{ for } x \in (0,k), \quad {\iota^*}''_{\mathrm{BP}}(k) = 0, \quad {\iota^*}''_{\mathrm{BP}}(x) > 0 \text{ for } x \in \mathbb{R}_{>k}.$$

For $k = 4$ and $x \in \mathbb{R}_{>0}$ we have $\iota^*_{\mathrm{BP}}(x^{-1}) = (\iota^*_{\mathrm{BP}}(x))^{-1}$.

Proof. Using $f_p(1) = 1$ for the moment generating function of any finitely supported law p, we have $\iota^*_{\mathrm{BP}}(1) = 1$. Further, using that the first moment of a hypergeometric distribution p_{N,N_1,N_2} is $\frac{N_1 N_2}{N}$ and that $f'_p(1)$ is the first moment of p, we have ${\iota^*}'_{\mathrm{BP}}(1) = \frac{1}{k-1} + \frac{4}{k-1} - 2 \cdot \frac{2}{k-1} = \frac{1}{k-1}$. The symmetry of ι^*_{BP} for the special case $k = 4$ can be seen as follows. First, recall that $p_{N,N_1,N_2}(s) = p_{N,N-N_1,N_2}(N_2 - s)$ for any hypergeometric distribution p_{N,N_1,N_2}. For $s \in \{0,1,2\}$ this gives

$$\begin{aligned} p'_{(0,0)}(s) &= p_{3,2,2}(s) = p_{3,1,2}(2-s) = p'_{(1,0)}(2-s) \\ &= p'_{(0,1)}(2-s) = p_{3,2,1}(2-s) = p_{3,1,1}(s-1) = p'_{(1,1)}(s-1). \end{aligned}$$

These relations can be directly translated to the moment generating functions, i.e.

$$f_{p'_{(0,0)}}(x) = x^2 f_{p'_{(1,0)}}(x^{-1}) = x^2 f_{p'_{(0,1)}}(x^{-1}) = x f_{p'_{(1,1)}}(x)$$

for $x \in \mathbb{R}_{>0}$. Using these transformations we have

$$\iota^*_{\mathrm{BP}}(x^{-1}) = \frac{f_{p'_{(1,1)}}(x^{-1}) f_{p'_{(0,0)}}(x^{-1})}{f_{p'_{(1,0)}}(x^{-1}) f_{p'_{(0,1)}}(x^{-1})} = \frac{x^{-1} f_{p'_{(1,0)}}(x) x^{-2} f_{p'_{(0,1)}}(x)}{x^{-1} f_{p'_{(1,1)}}(x) x^{-2} f_{p'_{(0,0)}}(x)} = (\iota^*_{\mathrm{BP}}(x))^{-1}\,.$$

For $k \in \mathbb{Z}_{\geq 4}$ and $x \in \mathbb{R}_{>0}$ direct computation gives

$$\begin{aligned}
\iota^*_{\mathrm{BP}}(x) &= \frac{1}{2(k-2)}x + \frac{2k-5}{2(k-2)} + \frac{(k-1)(k-3)(x-1)}{2(k-2)(2x+k-3)^2},\\
{\iota^*}'_{\mathrm{BP}}(x) &= \frac{1}{2(k-2)} + \frac{(k-1)(k-3)(-2x+k+1)}{2(k-2)(2x+k-3)^3},\\
{\iota^*}''_{\mathrm{BP}}(x) &= \frac{4(k-1)(k-3)(x-k)}{(k-2)(2x+k-3)^4}.
\end{aligned}$$

The remaining assertions follow immediately or with routine computations. □

Lemma 5.23 has the following immediate consequences.

Corollary 5.24. *For any $d \in (0,2]$ we have $\iota^*_{\mathrm{BP}}(x) \in (x,1)$ for $x \in (0,1)$ and $\iota^*_{\mathrm{BP}}(x) \in (1,x)$ for $x \in \mathbb{R}_{>1}$. In particular, p^* is the unique minimizer of Δ_d.*

Proof. Using Lemma 5.23 we notice that ${\iota^*}'_{\mathrm{BP}}(x) \in [{\iota^*}'_{\mathrm{BP}}(k), \frac{1}{k-1}] \subset (0,1)$ for $x \in \mathbb{R}_{\geq 1}$ and $\iota^*_{\mathrm{BP}}(x) = x$ for $x = 1$, hence we have $\iota^*_{\mathrm{BP}}(x) \in (1,x)$ for $x \in \mathbb{R}_{>1}$. For $k = 4$ this gives $\iota^*_{\mathrm{BP}}(x) \in (x,1)$ using the symmetry result. For $k \in \mathbb{Z}_{>4}$ we have $\lim_{x\to 0} \iota^*_{\mathrm{BP}}(x) > 0$, which gives $x^* = \inf\{x \in \mathbb{R}_{>0} : \iota^*_{\mathrm{BP}}(x) \leq x\} \in (0,1]$ using $\iota^*_{\mathrm{BP}}(1) = 1$. Assume that $x^* < 1$, then using the continuity of $x - \iota^*_{\mathrm{BP}}(x)$ we directly get $\iota^*_{\mathrm{BP}}(x^*) = x^*$, and further ${\iota^*}'_{\mathrm{BP}}(x^*) \leq 1$ since $\iota^*_{\mathrm{BP}}(x) > x$ for $x \in (0,x^*)$. But then, since ${\iota^*}''_{\mathrm{BP}}(x) < 0$ for $x \in (0,k)$, this implies that ${\iota^*}'_{\mathrm{BP}}(x) < 1$ for $x \in (x^*,1]$ which yields that ${\iota^*}_{\mathrm{BP}}(1) < 1$ and hence a contradiction. This shows that $\iota^*_{\mathrm{BP}}(x) \in (x,1)$ for $x \in (0,1)$. Now, for any $d \in (0,2]$ we have $\iota_{\mathrm{BP}}(x) \geq \iota^*_{\mathrm{BP}}(x) \in (x,1)$ for $x \in (0,1)$ and $\iota_{\mathrm{BP}}(x) \leq \iota^*_{\mathrm{BP}}(x) \in (1,x)$ for $x \in \mathbb{R}_{>1}$. Hence, using Lemma 5.21 and Lemma 5.22 we immediately get that $p^* = p_1$ is the unique minimizer of Δ_d. □

Corollary 5.24 covers the case of simple graphs that was discussed in [114]. On the other hand, Corollary 5.24 suggests that Proposition 5.5 can only hold if $d^* > 2$.

Corollary 5.25. *For all $k \in \mathbb{Z}_{\geq 4}$ we have $d^* \in \mathbb{R}_{>2}$.*

Proof. For $d \in \mathbb{R}_{>0}$ let $f(d) = (\Delta_d \circ \iota_{\mathrm{rp}})(\infty)$ and notice that $f(d) = \frac{k}{d}\phi_1$. Corollary 5.24 shows that $f(d) > 0$ for all $d \in (0,2]$. On the other hand, as derived in Section 5.4 we know that d^* is the unique root of f, which shows that $d^* > 2$. □

Based on Corollary 5.24 we can restrict to $d \in \mathbb{R}_{>2}$, while Corollary 5.25 motivates the discussion of this interval. Further, the restriction $d \in \mathbb{R}_{>2}$ ensures that x^{d-1} is increasing and convex on $\mathbb{R}_{>0}$. In the following we will consider the intervals $\mathbb{R}_{\geq k}$, $[\bar{x},k]$, $[1,\bar{x}]$ and $(0,1]$ independently, where $\bar{x} = \frac{1}{7}(k+6) \in (1,k)$ is the intersection of g_1 and g_k with $g_1(\bar{x}) = g_k(\bar{x}) = \frac{8}{7}$.

Lemma 5.26. *For $d \in (2, d_k)$, with $d_k = \frac{\ln(k)}{\ln(\iota^*_{\mathrm{BP}}(k))} + 1$, there exists $x_{\max} \in \mathbb{R}_{>k}$ such that $\iota_{\mathrm{BP}}(x) < x$ for $x \in [k, x_{\max})$, $\iota_{\mathrm{BP}}(x_{\max}) = x_{\max}$ and $\iota_{\mathrm{BP}}(x) > x$ for $x \in \mathbb{R}_{>x_{\max}}$.*

Proof. Let $f(d) = (\iota^*_{\mathrm{BP}}(k))^{d-1}$ for $d \in \mathbb{R}_{\geq 2}$, i.e. $f(d) = \iota_{\mathrm{BP}}(k)$ is the value of ι_{BP} at k under a variation of d. We know from Lemma 5.23 that $\iota^*_{\mathrm{BP}}(k) \in \mathbb{R}_{>1}$, hence $f(d)$ is strictly increasing, and further direct computation gives $f(d_k) = k$, so we have $\iota_{\mathrm{BP}}(k) < k$ for any $d \in (2, d_k)$. Now, since ι^*_{BP} is strictly increasing and convex on $\mathbb{R}_{>k}$ by Lemma 5.23 and further the function x^{d-1} is increasing and convex for $d \in (2, d_k)$, we know that ι_{BP} is convex and increasing on $\mathbb{R}_{>k}$, or formally

$$\iota'_{\mathrm{BP}}(x) = (d-1)\iota^*_{\mathrm{BP}}(x)^{d-2}{\iota^*}'_{\mathrm{BP}}(x) = (d-1)\iota_{\mathrm{BP}}(x)\frac{{\iota^*}'_{\mathrm{BP}}(x)}{\iota^*_{\mathrm{BP}}(x)} > 0,$$

$$\iota''_{\mathrm{BP}}(x) = (d-1)\iota_{\mathrm{BP}}(x)\left[(d-2)\left(\frac{{\iota^*}'_{\mathrm{BP}}(x)}{\iota^*_{\mathrm{BP}}(x)}\right)^2 + \frac{{\iota^*}''_{\mathrm{BP}}(x)}{\iota^*_{\mathrm{BP}}(x)}\right] > 0.$$

Using Lemma 5.23 and for $x \to \infty$ we have $\iota^*_{\mathrm{BP}}(x) \to \infty$ since ${\iota^*}'_{\mathrm{BP}}(x) \to \frac{1}{2(k-2)}$ and hence $\iota'_{\mathrm{BP}}(x) \to \infty$ by the above, i.e. $\iota_{\mathrm{BP}}(x) - x \to \infty$, which suggests the existence of $x_{\max} \in \mathbb{R}_{>k}$ with $\iota_{\mathrm{BP}}(x_{\max}) = x_{\max}$ since $\iota_{\mathrm{BP}}(k) < k$. Now, let $x_+ = \inf\{x \in \mathbb{R}_{\geq k} : \iota_{\mathrm{BP}}(x) \geq x\}$, then we have $x_+ \in (k, x_{\max}]$. Since $\iota_{\mathrm{BP}}(x) < x$ for $x \in [k, x_+)$ we need $\iota'_{\mathrm{BP}}(x_+) \geq 1$, which gives $\iota'_{\mathrm{BP}}(x) > 1$ for $x > x_+$ since $\iota''_{\mathrm{BP}}(x) > 0$, hence $\iota_{\mathrm{BP}}(x) > x$, thereby $x_+ = x_{\max}$, and in summary $\iota_{\mathrm{BP}}(x) < x$ for $x \in [k, x_{\max})$, $\iota_{\mathrm{BP}}(x_{\max}) = x_{\max}$ and $\iota_{\mathrm{BP}}(x) > x$ for $x \in \mathbb{R}_{>x_{\max}}$. □

The proof of Lemma 5.26 serves as a blueprint for the next two cases, where we do not consider ι_{BP} directly since ${\iota^*}''_{\mathrm{BP}}(x) < 0$ on $(1, k)$, but work with $g_k(x) = g^*_k(x)^{d-1}$ and $g_1(x) = g^*_1(x)^{d-1}$ instead, which are convex, increasing and upper bounds for ι_{BP} on $[1, k]$ since ${\iota^*}''_{\mathrm{BP}}(x) < 0$ on $(1, k)$. In the spirit of Lemma 5.26 we continue to consider the maximal domain for $d \in \mathbb{R}_{>2}$. Let

$$d_{\bar{x}} = \frac{\ln(\bar{x})}{\ln(g_1(\bar{x}))} + 1 \text{ and } d_{\max} = \min(d_{\bar{x}}, d_k).$$

We postpone the proof that $d^* \leq d_{\max}$, instead we focus on the interval $(1, k)$.

Lemma 5.27. *For any $d \in (2, d_{\max}) \subseteq (2, k)$ and all $x \in (1, k]$ we have $\iota_{\mathrm{BP}}(x) < x$.*

Proof. Fix $d \in (2, d_{\max})$. Since ${\iota^*}''_{\mathrm{BP}}(x) < 0$ for $x \in [1, k)$, we know that $\iota^*_{\mathrm{BP}}(x) \leq g^*_k(x)$ for $x \in [\bar{x}, k]$ and $\iota^*_{\mathrm{BP}}(x) \leq g^*_1(x)$ for $x \in [1, \bar{x}]$, so using that x^{d-1} is increasing we have that $\iota_{\mathrm{BP}}(x) \leq g_k(x)$ for $x \in [\bar{x}, k]$ and $\iota_{\mathrm{BP}}(x) \leq g_1(x)$ for $x \in [1, \bar{x}]$. Analogous to Lemma 5.26 we notice that $g_k(k) = \iota_{\mathrm{BP}}(k) < k$ since $d < d_k$, that $g_k(\bar{x}) = g_1(\bar{x}) < \bar{x}$ since $d < d_{\bar{x}}$ and that $g_1(1) = 1$. Further, since g^*_1, g^*_k are increasing and convex, using that x^{d-1} is increasing and convex yields that g_1, g_k are increasing and convex. In particular, we can upper bound g_k with the line $l_k : [\bar{x}, k] \to [g_k(\bar{x}), g_k(k)]$ connecting $(\bar{x}, g_k(\bar{x}))$ and $(k, g_k(k))$, which is entirely and strictly under the diagonal. Analogously, we can upper bound g_1 with the line $l_1 : [1, \bar{x}] \to [1, g_1(\bar{x})]$ connecting $(1, 1)$ and $(\bar{x}, g_1(\bar{x}))$, which is also entirely and strictly under the diagonal except for $(1, 1)$ where the two lines intersect. In total, $\iota_{\mathrm{BP}}(x) \leq \min(g_1(x), g_k(x)) \leq \min(l_1(x), l_k(x)) < x$ for all $x \in (1, k]$. Finally, another implication is that $\frac{d-1}{k-1} = g'_1(1) \leq l'_1(1) < 1$ since l_1 is below the diagonal, which suggests that $d_{\max} \leq k$. □

Combining Lemma 5.26 and Lemma 5.27 shows for any $d \in (2, d_{\max})$ that $\Delta_d \circ \iota_{\mathrm{rp}}$ has exactly one stationary point $x_{\max}$ on $\mathbb{R}_{>1}$ which is the unique maximizer of $\Delta_d \circ \iota_{\mathrm{rp}}$ on this interval. Further, since $d_{\max} \leq k$ and using $\iota'_{\mathrm{BP}}(1) = \frac{d-1}{k-1}$ we also know that $x = 1$ is an isolated minimizer of $\Delta_d \circ \iota_{\mathrm{rp}}$. Aside, notice that this argumentation can be used to show that H_d as defined in Section 5.5.5 is positive semi-definite for all $d < k$, hence with the arguments from the proof of Lemma 5.14 we see that H_d is positive definite for all $d < k$ and finally with Lemma 5.18 that $\eta_{\chi^2} = \frac{1}{k-1}$.

For the low overlap region $x \in (0,1)$ we need a significantly different approach, since ι^*_{BP} is increasing and concave, but $\iota_{\mathrm{BP}}(x) < \iota^*_{\mathrm{BP}}(x)$ and we need to show that $\iota_{\mathrm{BP}}(x) > x$. This means that first order approximations as used for $(1,k)$ are useless since they are upper bounds to ι^*_{BP} and there are no immediate implications for ι''_{BP} as was the case for $\mathbb{R}_{>k}$. However, the symmetric case $k = 4$ can be discussed easily.

Corollary 5.28. *For $k = 4$ and $d \in (2, d_{\max})$ we have $\iota_{\mathrm{BP}}(x) < x$ for $x \in (0, x_{\max}^{-1})$, $\iota_{\mathrm{BP}}(x_{\max}^{-1}) = x_{\max}^{-1}$, and $\iota_{\mathrm{BP}}(x) > x$ for $x \in (x_{\max}^{-1}, 1)$.*

Proof. Combining Lemma 5.27 and Lemma 5.26 we have $\iota_{\mathrm{BP}}(x) < x$ for $x \in (1, x_{\max})$ and $\iota_{\mathrm{BP}}(x) > x$ for $x \in (x_{\max}, \infty)$, hence using the symmetry from Lemma 5.23 directly gives the result. □

Corollary 5.28 allows to restrict to $k \in \mathbb{Z}_{>4}$ in the remainder. Now, we basically reverse the method used for the interval $(1,k)$, i.e. instead of using tangents g^*_1, g^*_k to ι^*_{BP} and scaling them with $(d-1)$, we scale ι^*_{BP} such that the diagonal is a tangent, meaning we consider $\iota_k = \iota_{\mathrm{BP}}$ for $d = k$ since $\iota'_{\mathrm{BP}}(1) = \frac{d-1}{k-1}$, and show that ι_k is sufficiently convex to ensure $\iota_k(x) > x$ for $x \in (0,1)$. The next lemma ensures that this approach is applicable for all $k \geq 5$.

Lemma 5.29. *For any $k \in \mathbb{Z}_{\geq 5}$, $d \in (2,k]$ and all $x \in (0,1)$ we have $\iota_{\mathrm{BP}}(x) > x$.*

Proof. Let $k \in \mathbb{Z}_{\geq 5}$. As derived in the proof of Lemma 5.27 we have $\iota_k(1) = 1$ and $\iota'_k(1) = 1$, i.e. the diagonal is a tangent to ι_k at $x = 1$. Further, as discussed in the proof of Lemma 5.26,

$$\iota''_k(x) = (k-1)\frac{\iota_k(x)}{\iota^*_{\mathrm{BP}}(x)^2}\left[(k-2){\iota^*}'_{\mathrm{BP}}(x)^2 + \iota^*_{\mathrm{BP}}(x){\iota^*}''_{\mathrm{BP}}(x)\right] \text{ for } x \in (0,1).$$

Since the leading factor is clearly strictly positive, we may focus on the term in the square brackets. Further, using that moment generating functions are strictly positive for strictly positive real numbers we can extract the strictly positive denominator of the term and normalize to get

$$\begin{aligned}
f(x) &= \frac{\iota^*_{\mathrm{BP}}(x)^2\left[(k-1)f_{p'_{(1,0)}}(x)\right]^6 (k-2)^2\iota''_k(x)}{(k-1)\iota_k(x)} = \sum_{i\in[6]_0} a_i x^i, \text{ where}\\
a_6 &= 16(k-2),\\
a_5 &= 48(k-2)(k-3),\\
a_4 &= 4(k-3)\left[9(k-2)(k-3) + 4(k-2)(k-4) + 2(k-1)\right],\\
a_3 &= 8(k-3)\left[3(k-2)^3 + (k-2)(k^2-3k+4) + 2(k-1)(k-4)\right],\\
a_2 &= 4(k-3)b_2,\ b_2 = (k-2)^2\left[(k-4)^2 + 3(k^2-3k+4)\right] - (k-1)(k^2+11k-36),\\
a_1 &= 4(k-3)^2 b_1,\ b_1 = (k-2)(k-4)(k^2-3k+4) - 2(k-1)(2k^2-3k-4),\\
a_0 &= (k-2)(k-3)^2 b_0,\ b_0 = (k^2-3k+4)^2 - 4k(k-1)(k-4).
\end{aligned}$$

We can easily verify that $a_i > 0$ for $i \in [6] \setminus [2]$ using that $x^2 - 3x + 4 > 0$ for all $x \in \mathbb{R}$. Viewing the coefficients b_i, $i \in \{0,1,2\}$, as polynomials $b_i(x)$, $x \in \mathbb{R}$, of degree 4 and evaluated at $x = k$, we have

$$b_2''(x) = 48x^2 - 228x + 254,\ b_2(5) = 82,\ b_2'(5) = 225,$$

hence $b_2''(x) > 0$ for $x > x_2$ with $x_2 = \frac{19}{8} + \frac{\sqrt{201}}{24} < 3$ and by that $b_2(x) > 0$ for all $x \in \mathbb{R}_{\geq 5}$, so in particular $b_2 = b_2(k) > 0$ since $k \in \mathbb{Z}_{\geq 5}$ and thereby $a_2 > 0$. Using the same technique for the degree four polynomials $b_1(x)$ we obtain that $b_1''(x) > 0$ for $x > x_1$ with $x_1 = \frac{13}{4} + \frac{\sqrt{561}}{12} < 6$, $b_1(10) = 564$, $b_1'(10) = 854$, and hence that $b_1 > 0$ if $k \geq 10$. For $b_0(x)$ we have $b_0''(x) > 0$ for $x > x_0$ with $x_0 = \frac{5}{2} + \frac{\sqrt{3}}{6} < 3$, $b_0'(3) = 20$, $b_0(3) = 40$, so $b_0 > 0$ for all $k \in \mathbb{Z}_{\geq 5}$.

Hence, for $k \in \mathbb{Z}_{\geq 10}$ we know that $a_i > 0$ for all $i \in [6]_0$, which directly implies that $f(x) > 0$ for all $x \in \mathbb{R}_{>0}$, so $\iota_k''(x) > 0$, further $\iota_k'(x) < 1$ for $x \in (0,1)$, $\iota_k'(x) > 1$ for $x \in \mathbb{R}_{>1}$ and thereby $\iota_k(x) > x$ for $x \in \mathbb{R}_{>0} \setminus \{1\}$.

For $5 \leq k \leq 9$ we still have $a_i > 0$ for $i \in [6]_0 \setminus \{1\}$. For $6 \leq k \leq 9$ we consider the quadratic function $g_k(x) = \sum_{i \in \{0,1,2\}} a_i x^i$ explicitly, given by

$$g_6(x) = 6120x^2 - 11664x + 8784,\ g_7(x) = 24960x^2 - 25344x + 41600,$$
$$g_8(x) = 72560x^2 - 34400x + 156000,\ g_9(x) = 172944x^2 - 9504x + 484848.$$

It turns out that $g_k(x) > 0$ for all $x \in \mathbb{R}$ and $6 \leq k \leq 9$, which in particular yields $f(x) > 0$ for all $x \in \mathbb{R}_{>0}$. Using the same argumentation as for $k \in \mathbb{Z}_{\geq 10}$ shows that $\iota_k(x) > x$ for all $x \in \mathbb{R}_{>0} \setminus \{1\}$ and $k \in \mathbb{Z}_{\geq 6}$.

As opposed to the previous cases the function ι_k is not convex for $k = 5$, and while this slightly complicates the computation, we will show that this does not affect the overall picture. Now, we consider the complete sixth order polynomial

$$f(x) = 48x^6 + 288x^5 + 592x^4 + 1088x^3 + 656x^2 - 3296x + 1392.$$

We notice that $f''(x) > 0$ for all $x \in \mathbb{R}_{\geq 0}$ since $a_i > 0$ for $i \in [6] \setminus \{1\}$, hence $f'(x)$ is strictly increasing for $x \in \mathbb{R}_{\geq 0}$, which shows the existence of a unique root $x_{\min} \in (0,1)$, using $f'(0) < 0$ and $f'(1) > 0$, i.e. $x_{\min}$ is the unique minimizer of f on $\mathbb{R}_{\geq 0}$. Computing $f(x_{1-}) > 0$, $f(x_{1+}) < 0$ and $f(1) > 0$ with $x_{1-} = 0.581$ and $x_{1+} = 0.582$ ensures the existence of exactly two roots $x_1 \in (x_{1-}, x_{1+})$ and $x_2 \in (x_{1+}, 1)$ of f, hence ι_k is convex on $(0, x_1)$, concave on (x_1, x_2) and convex on $\mathbb{R}_{>x_2}$. Let $g : \mathbb{R} \to \mathbb{R}$, $x \mapsto \iota_k'(x_{1-})(x - x_{1-}) + \iota_k(x_{1-})$ denote the tangent of ι_k at x_{1-}. Since we can write both ι_k and ι_k' as the ratio of polynomials with integer coefficients, we can compute $\iota_k(x_{1-}) \in (0.584, 0.585)$, $\iota_k'(x_{1-}) \in (0.99, 0.991)$ and $g(x_{1+}) \in (0.585, 0.586)$ exactly. Using the convexity of ι_k on $(0, x_1)$, $g' = \iota_k'(x_{1-}) < 1$ and $g(x_{1+}) > x_{1+}$ immediately gives that $\iota_k(x) \geq g(x) > x$ for $x \in (0, x_1]$. The fact that $\iota_k(x) > x$ for $x \in [x_2, 1)$ immediately follows from the convexity of ι_k on $(x_2, 1)$ and that the diagonal is a tangent to ι_k at $x = 1$. But now, since $(x_1, \iota_k(x_1))$ and $(x_2, \iota_k(x_2))$ are above the diagonal, so is the line connecting the two, which is a lower bound to ι_k on (x_1, x_2) since ι_k is concave on this interval. By that we have finally showed that the overall picture is also the same for $k = 5$, i.e. $\iota_k(x) > x$ for $x \in \mathbb{R}_{>0} \setminus \{1\}$.

The fact that $\iota_k(x) > x$, i.e. $\iota_{\mathrm{BP}}(x) > x$ with $d = k$, for $x \in \mathbb{R}_{>0} \setminus \{1\}$ shows that the only stationary point of Δ_k is a saddle at p^* (respectively a maximum of $\Delta_k \circ \iota_{\mathrm{rp}}$ at $x = 1$). But more importantly, since x^{d-1} is decreasing in d for $x \in (0,1)$ and $\iota_{\mathrm{BP}}^*(x) \in (0,1)$, we have $\iota_{\mathrm{BP}}(x) \geq \iota_k(x) > x$ for all $d \in (2, k]$. □

The combination of Lemma 5.26, Lemma 5.27 and Lemma 5.28 shows that for $k = 4$ and all $d \in (2, d_{\max})$ there exist exactly three fixed points $x_- < x_0 < x_+$ of ι_{BP}, with $x_0 = 1$, $x_+ \in (k, \infty)$ and $x_- = x_+^{-1}$, hence Lemma 5.22 and Lemma 5.21 suggest that x_0 is a minimizer of $\Delta_d \circ \iota_{\mathrm{rp}}$ while x_- and x_+ are maximizers. In particular, we have the three minimizers $\{0, 1, \infty\}$ of $\Delta_d \circ \iota_{\mathrm{rp}}$ in total.

The combination of Lemma 5.26, Lemma 5.27 and Lemma 5.29 shows that for $k \in \mathbb{Z}_{\geq 5}$ and all $d \in (2, d_{\max}) \subseteq (0, k)$ there exist exactly two fixed points $x_0 < x_+$ of ι_{BP}, with $x_0 = 1$ and $x_+ \in (k, \infty)$, hence Lemma 5.22 and Lemma 5.21 suggest that x_0 is a minimizer of $\Delta_d \circ \iota_{\mathrm{rp}}$ while x_+ is a maximizer. In particular, we have the two minimizers $\{1, \infty\}$ of $\Delta_d \circ \iota_{\mathrm{rp}}$ in total, while $x = 0$ is a maximizer in these cases.

The last step is to show that we have $d^* \in (2, d_{\max})$, which then directly establishes that the unique minimizers of $\Delta_{d^*} \circ \iota_{\mathrm{rp}}$ are given by $\{0, 1, \infty\}$ for $k = 4$ and $\{1, \infty\}$ for $k \in \mathbb{Z}_{\geq 5}$ as required. On the other hand, direct computation as in the proof of Corollary 5.25 shows that all minimizers are roots of $\Delta_{d^*} \circ \iota_{\mathrm{rp}}$, i.e. all minimizers are global minimizers and the global minimum of $\Delta_{d^*} \circ \iota_{\mathrm{rp}}$ is 0. Lemma 5.19 directly suggests that the global minimizers of $\Delta_{d^*} \circ \iota_{\mathrm{rp}}$ are in one to one correspondence with the global minimizers of Δ_{d^*} via $x \mapsto p_x$, which then completes the proof of Proposition 5.5.

Lemma 5.30. *For all $k \in \mathbb{Z}_{\geq 4}$ we have $2 < d^* < d_k < d_{\bar{x}}$.*

Proof. Recall from Corollary 5.25 that $d^* > 2$. For convenience, we consider the extensions of $d^* - 1$, $d_{\bar{x}} - 1$ and $d_k - 1$ to the real line, i.e. for $x \in \mathbb{R}_{\geq 3}$ let

$$f_0(x) = \frac{\ln\left(\frac{1}{2}x(x-1)\right)}{xH(2/x) - \ln\left(\frac{1}{2}x(x-1)\right)},\ f_1(x) = \frac{\ln\left(\frac{1}{7}(x+6)\right)}{\ln(8/7)},\ f_2(x) = \frac{\ln(x)}{\ln\left(\frac{2(7x-12)}{9(x-2)}\right)},$$

i.e. $d^* = f_0(k) + 1$, $d_{\bar{x}} = f_1(k) + 1$ and $d_k = f_2(k) + 1$ for all $k \in \mathbb{Z}_{\geq 4}$. We start with the asymptotic comparison of f_1 and f_2. The corresponding rearrangement gives

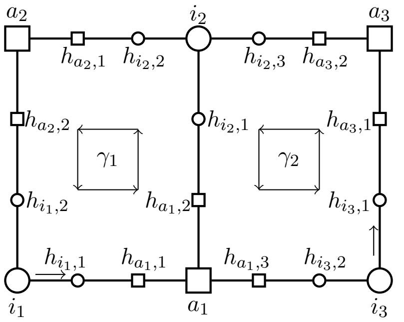
$$f_1(x) = m_1 \ln(x) + t_1(x),\ m_1 = \frac{1}{\ln(8/7)},\ t_1(x) = -\frac{\ln(7)}{\ln(8/7)} + \frac{\ln\left(1 + \frac{6}{x}\right)}{\ln(8/7)},$$
$$f_2(x) = m_2(x)\ln(x),\ m_2(x) = \frac{1}{\ln(14/9) + \ln\left(1 + \frac{2}{7(x-2)}\right)}.$$

Notice that $\ln(x) > 0$ since $x \geq 3$, further $t_1(x)$ is decreasing while $m_2(x)$ is increasing, and thereby we have $f_1(x) \geq f_{1\infty}(x)$ and $f_2(x) \leq f_{2\infty}(x)$ with

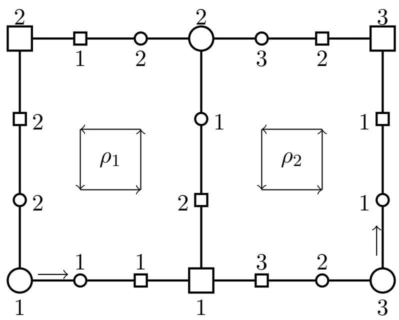
$$f_{1\infty}(x) = m_{1\infty}\ln(x) + t_{1\infty},\ m_{1\infty} = m_1,\ t_{1\infty} = -\frac{\ln(7)}{\ln(8/7)},$$
$$f_{2\infty}(x) = m_{2\infty}\ln(x),\ m_{2\infty} = \frac{1}{\ln(14/9)}.$$

We have $m_{1\infty} > m_{2\infty}$, $t_{1\infty} < 0$ and further $f_{1\infty}(x) > f_{2\infty}(x)$ iff $x > x_{12}$ with $x_{12} = \exp\left(\frac{-t_{1\infty}}{m_{1\infty} - m_{2\infty}}\right) \in (16, 17)$, so $f_1(x) > f_2(x)$ for all $x \in \mathbb{R}_{\geq 17}$ and hence $d_k < d_{\bar{x}}$ for $k \in \mathbb{Z}_{\geq 17}$. We check by hand that $d_k < d_{\bar{x}}$ also holds for $4 \leq k \leq 16$.

Hence, we're left to show that $1 < f_0(x) < f_2(x)$ for $x \in \mathbb{Z}_{\geq 4}$. Again, we start with the asymptotic



(a) pair (γ_1, γ_2) of intersecting 4-cycles



(b) relative positions (ρ_1, ρ_2) for (γ_1, γ_2)

Figure 6: The left figure shows a sequence $\gamma = (\gamma_1, \gamma_2)$ of two directed (intersecting) four-cycles with base variables i_1 and i_3 and directions indicated by the arrows respectively. Analogously to Figure 5a we only denoted the i-edges and a-edges instead of the v-edges and f-edges. The relative positions $\rho = (\rho_1, \rho_2)$ corresponding to γ are depicted in the right figure. Here, the variables, constraints, i-edges and a-edges are labeled according to the order of first traversal (where γ_1 is traversed before γ_2). The numbers $n(\rho) = 3$, $m(\rho) = 3$, $e(\rho) = 7$ of variables, constraints and edges in ρ are equal to the corresponding numbers in γ, further the degree $d_j(\rho)$ of the variable $j \in [3]$ equals the degree of i_j in γ, and analogously for the degrees $k_b(\rho)$ of the constraints $b \in [3]$ in ρ. The absolute values $\alpha = (\alpha_v, \alpha_f, (\alpha_{v,j})_{j\in[3]}, (\alpha_{f,b})_{b\in[3]})$ are given by $\alpha_v = (i_j)_{j\in[3]}$, $\alpha_f = (a_b)_{b\in[3]}$, $\alpha_{v,j} = (h_{i_j,e})_{e\in[d_j(\rho)]}$, $j \in [3]$, and $\alpha_{f,b} = (h_{a_b,e})_{e\in[k_b(\rho)]}$, $b \in [3]$, i.e. they store the (initial) labels of γ corresponding to the labels of ρ.

comparison, where the corresponding rearrangement $f_0(x) = m_0(x)\ln(x) + t_0(x)$ is given by

$$m_0(x) = \frac{2}{n_0(x)},\ \ t_0(x) = \frac{\ln\left(1+\frac{1}{x}\right) - \ln(2)}{n_0(x)},\ \ n_0(x) = -(x-2)\ln\left(1-\frac{2}{x}\right) - \ln\left(1-\frac{1}{x}\right) - \ln(2).$$

Recall that for given $c \in \mathbb{R}$ we have $\ln(1+\frac{c}{x}) = (1+o(1))\frac{c}{x}$ and $\ln(1+\frac{c}{x}) \leq \frac{c}{x}$ for all $x \in \mathbb{R}_{>|c|}$, since $\ln(x)$ is concave and the tangent at 1 is $x-1$. Hence, for all $x \in \mathbb{R}_{\geq 3}$ we have $n_0(x) \geq n_+(x) > 0$ since $x > 2$ and $x > x_1$, where

$$n_+(x) = \ln\left(\frac{e^2}{2}\right) - \frac{3}{x} \text{ and } x_1 = \frac{3}{\ln\left(\frac{e^2}{2}\right)} \in (2,3).$$

Since $t_0(x) < 0$ and $m_0(x) \leq m_+(x)$ we have $f_0(x) \leq f_+(x) = m_+(x)\ln(x)$ with

$$m_+(x) = \frac{2}{\ln\left(\frac{e^2}{2}\right) - \frac{3}{x}}.$$

Now, since $m_+(x)$ is decreasing in x and $m_2(x)$ is increasing in x, we numerically determine $x^* \in \mathbb{R}_{>0}$ such that $m_+(x^*) = m_2(x^*)$ and find that $x^* \in (8,9)$. In particular, we have $f_0(x) \leq f_+(x) < f_2(x)$ for $x \in \mathbb{R}_{\geq 9}$, and check that $d^* < d_k$ for $4 \leq k \leq 8$ by hand. □

Lemma 5.30 concludes the proof of Proposition 5.5 as discussed before.

5.6 Small Subgraph Conditioning

In this section we prove the remaining parts of Theorem 5.7, thereby establishing Theorem 5.1. The first part of the proof heavily relies on Section 5.7 and illustrates the correspondences. We start with the derivation of δ_ℓ by computing $\mathbb{E}[\boldsymbol{Z}_\mathrm{c}\boldsymbol{X}_\ell]$. For this purpose we fix $\ell \in \mathbb{Z}_{>0}$, $n \in \mathcal{N}$ sufficiently large, and let $\bar{c}_\ell$ denote the canonical 2ℓ-cycle, i.e. the cycle with variables i, constraints a in $[\ell]$ and i-edges, a-edges in $\{1,2\}$ with labels ordered by first traversal, see e.g. the left cycle in Figure 6b. Analogous to the previous sections we rewrite the expectation and count the number $|\mathcal{E}|$ of triplets $(G,c,x) \in \mathcal{E}$ such that c is a 2ℓ-cycle and x a solution in G, i.e.

$$\begin{aligned}
\mathbb{E}[\boldsymbol{Z}_\mathrm{c}\boldsymbol{X}_\ell] &= \frac{|\mathcal{E}|}{|\mathcal{G}_\mathrm{c}|} = \sum_{y\in\{0,1\}^\ell} \frac{e_1e_2e_3}{2\ell(dn)!}, \text{ where}\\
e_1 = e_1(y) &= \binom{n}{n_1} n_1^{\underline{r_1}}(n-n_1)^{\underline{\ell-r_1}}(d(d-1))^\ell,\\
e_2 = e_2(y) &= \binom{k}{2}^m m^{\underline{\ell}} 2^{r_2}(2(k-2))^{2(r_1-r_2)}((k-2)(k-3))^{\ell-2r_1+r_2},\\
e_3 = e_3(y) &= (dn_1-2r_1)!(d(n-n_1)-2(\ell-r_1))!,
\end{aligned}$$

and $r = r(y) = (r_1, r_2)$ is defined as follows. For $y \in \{0,1\}^\ell$ we let $r_1 = r_1(y)$ denote the number of ones in y. Further, let $r_2 = r_2(y)$ denote the number of constraints $b \in [\ell]$ in $\bar{c}_\ell$ such that both b-edges take the value one under the assignment y of the variables $j \in [\ell]$ in $\bar{c}_\ell$. With y fixed we can compute the number of suitable triplets (G,c,x) as follows. The denominator in the first line reflects $|\mathcal{G}_\mathrm{c}|^{-1}$ and the compensation 2ℓ as we will count directed cycles γ in G. The sum over $y \in \{0,1\}^\ell$ implements the choice of the assignment of the variables visited by γ such that the variables $i_1,\dots,i_\ell$ traversed by γ correspond to the variables $1,\dots,\ell$ in $\bar{c}_\ell$ in this order, i.e. $x_{i_1} = y_1, \dots, x_{i_\ell} = y_\ell$. The first term in e_1 chooses the variables that take the value one under the solution x. Then we choose the r_1 variables out of the n_1-variables that participate in the directed cycle γ and take the value one consistent with y (hence an ordered choice). Analogously, we then choose the variables in γ taking zero under x. Finally, we choose the two i-edges traversed by γ for each of the ℓ variables i in the cycle.

The first term in e_2 is the usual choice of the two a-edges taking one under x for each $a \in [m]$. Then we choose the constraints visited by γ. The remaining terms account for the ordered choice of the two a-edges that are traversed by γ and that is consistent with the assignments y and x in the following sense. The (already chosen) variables $i_1,\dots,i_\ell$ and constraints $a_1,\dots,a_\ell$ traversed by γ correspond to the variables $1,\dots,\ell$ and constraints $1,\dots,\ell$ in $\bar{c}_\ell$ in this order respectively. Further, the assignment of these variables is already fixed by y and the a-edges taking the value one for each of these constraints are also fixed by our previous choice. Hence, if $y_1 = y_2 = 1$, then we have only two choices for the a_1-edge connecting to i_1, while the a_1-edge connecting to i_2 is fixed afterwards. For $y_1 = 1$ and $y_2 = 0$ we have two choices for the a_1-edge connecting to i_1 and $(k-2)$ choices for the a_1-edge connecting to i_2. The case $y_1 = 0$, $y_2 = 1$ is symmetric and we see that we have $(k-2)$ and $(k-3)$ choices for the remaining case $y_1 = y_2 = 0$ analogously. To derive the number of constraints for each of the cases above we recall that we have $r_1(y)$ ones in total and $r_2(y)$ ones whose successor is one (i.e. the constraint a between the two ones takes the value one on both a-edges, and where the successor of y_ℓ is y_1). But then $(r_1 - r_2)$ ones in y do not have the successor one, i.e. they have the successor zero. Complementarily we see that since r_2 ones are succeeded by a one there are r_2 ones that are preceded by a one, hence there are $(r_1 - r_2)$ ones that are preceded by zero. Then again, this means that there are $(r_1 - r_2)$ zeros that are succeeded by a one, hence the remaining $(\ell - 2r_1 + r_2)$ zeros out of the $(\ell - r_1)$ zeros are succeeded by a zero. This fixes γ, so in particular $2r_1$ v-edges that take the value one and $2(\ell - r_1)$ v-edges that take the value zero. The two terms in e_3 then wire the remaining edges.

We divide by $\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]$ to match the left hand side of Theorem 5.7b), i.e.

$$\frac{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}\boldsymbol{X}_\ell]}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]} = \sum_{y\in\{0,1\}^\ell} \frac{e_1e_2e_3}{2\ell(2m)!(dn-2m)!}, \text{ where}$$
$$e_1 = e_1(y) = n_1^{r_1}(n-n_1)^{\underline{\ell-r_1}}(d(d-1))^\ell,$$
$$e_2 = e_2(y) = m^{\underline{\ell}}2^{r_2}(2(k-2))^{2(r_1-r_2)}((k-2)(k-3))^{\ell-2r_1+r_2},$$
$$e_3 = e_3(y) = (dn_1 - 2r_1)!(d(n-n_1) - 2(\ell - r_1))!,$$

and using Stirling's formula we easily derive that

$$\frac{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}\boldsymbol{X}_\ell]}{\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]} = (1+o(1))\lambda_\ell \sum_{y\in\{0,1\}^\ell} M_{11}^{r_2}M_{01}^{r_1-r_2}M_{10}^{r_1-r_2}M_{00}^{\ell-2r_1+r_2} = (1+o(1))\lambda_\ell(1+\delta_\ell),$$
$$M = \begin{pmatrix} 1-\frac{2}{k-1} & 1-\frac{1}{k-1} \\ \frac{2}{k-1} & \frac{1}{k-1} \end{pmatrix}.$$

The matrix M has a nice interpretation as a (column stochastic) transition probability matrix in a two state Markov process, with

$$1+\delta_\ell = \sum_{\substack{y\in\{0,1\}^\ell \\ y_1=0}} M_{11}^{r_2}M_{01}^{r_1-r_2}M_{10}^{r_1-r_2}M_{00}^{\ell-2r_1+r_2} + \sum_{\substack{y\in\{0,1\}^\ell \\ y_1=1}} M_{11}^{r_2}M_{01}^{r_1-r_2}M_{10}^{r_1-r_2}M_{00}^{\ell-2r_1+r_2}$$

reflecting the probabilities that we return to the starting point given that the starting point is zero and one respectively. Let us consider the first partial sum restricted to sequences y (of Markov states) such that $y_1 = 0$, i.e. we start in the state zero. Then M_{0y_2} reflects the probability that we move from the initial state zero to the state y_2 given that we are in state zero (which is the case because we know that $y_1 = 0$). As discussed above we will move from a one to a one in y exactly r_2 times, from a one to a zero $(r_1 - r_2)$ times, from a zero to a one $(r_1 - r_2)$ times and from a zero to a zero $(\ell - 2r_1 + r_2)$ times. Hence the contribution to the first partial sum for given y exactly reflects the probability that we start in the state zero and (with this given) return to the state zero after ℓ steps (since the successor of y_ℓ is $y_1 = 0$). Since we sum over all such sequences y the first sum reflects the probability that we reach state zero after ℓ steps given that we start in the state zero. The discussion of the second sum is completely analogous. This directly yields

$$1+\delta_\ell = (M^\ell)_{00} + (M^\ell)_{11} = \mathrm{Tr}(M^\ell) = \lambda_1' + \lambda_2' = \lambda_1^\ell + \lambda_2^\ell, \quad \lambda_1 = 1,\, \lambda_2 = -\frac{1}{k-1},$$

where we used the Kolmogorov-Chapman equalities in the first step, i.e. that the ℓ-step transition probability matrix is the ℓ-th power of the one step transition probability matrix, which allow to translate the first sum into the transition probability $(M^\ell)_{00}$ that we reach the state zero after ℓ steps given that we start in the state zero and analogously for the second sum. In the second step we use the definition of the trace, while in the third step we use that the trace is the sum of the eigenvalues λ'_1, λ'_2 of M^ℓ. In the next step we use that the eigenvalues λ'_1 λ'_2 of the ℓ-th power M^ℓ of the matrix M are the ℓ-th powers of the eigenvalues λ_1, λ_2 of M. In particular this also yields that $\delta_\ell > -1$ for all $k > 3$ and establishes $\delta_\ell = (1-k)^{-\ell}$.

Following the strategy of Section 5.7 we turn to the case of disjoint cycles. Similarly, the present case is a canonical extension of the single cycle case discussed above. We fix $L \in \mathbb{Z}_{>0}$, $r \in \mathbb{Z}_{\geq 0}^L$ and $n \in \mathcal{N}$ sufficiently large. Further, as in the previous sections we rewrite the expectation and count the number $|\mathcal{E}|$ of triplets $(G, c, x) \in \mathcal{E}$ such that $c = (c_s)_{s\in[\bar{r}]}$ is a sequence of $\bar{r} = \sum_{\ell\in[L]} r_\ell$ distinct $2\ell_s$-cycles c_s in the configuration g sorted by their length ℓ_s in ascending order (as described in Section 5.7) and x is a solution of g. This yields

$$\mathbb{E}\left[\boldsymbol{Z}_{\mathrm{c}} \prod_{\ell\in[L]} (\boldsymbol{X}_\ell)_{r_\ell}\right] = \frac{|\mathcal{E}|}{|\mathcal{G}_{\mathrm{c}}|} = \frac{|\mathcal{E}_0|}{|\mathcal{G}_{\mathrm{c}}|} + \frac{|\mathcal{E}_1|}{|\mathcal{G}_{\mathrm{c}}|},$$

where $\mathcal{E}_0 \subseteq \mathcal{E}$ is the set over all triplets $(g, c, x) \in \mathcal{E}$ involving sequences c of disjoint cycles and $\mathcal{E}_1 = \mathcal{E} \setminus \mathcal{E}_0$. We begin with the first contribution, which can be regarded as a combination of the discussion of disjoint cycles in Section 5.7 and the single cycle case above, i.e.

$$\begin{aligned}
\frac{|\mathcal{E}_0|}{|\mathcal{G}_{\mathrm{c}}|} &= \sum_{y\in\{0,1\}^{\mathfrak{l}}} \frac{e_1 e_2 e_3}{(dn)! \prod_{s\in[\bar{r}]}(2\ell_s)},\\
e_1 = e_1(y) &= \binom{n}{n_1} n_1^{\underline{\mathfrak{r}_1}} (n-n_1)^{\underline{\mathfrak{l}-\mathfrak{r}_1}} (d(d-1))^{\mathfrak{l}},\\
e_2 = e_2(y) &= \binom{k}{2}^m m^{\mathfrak{l}} 2^{\mathfrak{r}_2} (2(k-2))^{2(\mathfrak{r}_1-\mathfrak{r}_2)} ((k-2)(k-3))^{\mathfrak{l}-2\mathfrak{r}_1+\mathfrak{r}_2},\\
e_3 = e_3(y) &= (dn_1 - 2\mathfrak{r}_1)!(d(n-n_1) - 2(\mathfrak{l}-\mathfrak{r}_1))!,\\
\mathfrak{l} &= \sum_{s\in[\bar{r}]} \ell_s,\ \mathfrak{r}_i = \sum_{s\in[\bar{r}]} r_i(y_s),\ i\in[2],
\end{aligned}$$

where $y = (y_s)_{s\in[\bar{r}]}$ is the subdivision of y corresponding to the definition of c, and r_1, r_2 are the notions defined above. The combinatorial arguments are now fairly self-explanatory, e.g. we make an ordered choice of the $r_1(y_1)$ variables taking one for γ_1, then an ordered choice of $r_1(y_2)$ variables taking one for γ_2 out of the remaining $n_1 - r_1(y_1)$ variables taking one and so on.

The asymptotics are also completely analogous to the single cycle case and Section 5.7. First, we notice that the sum is still bounded, i.e. we can also use the asymptotic equivalences for the corresponding ratio here. Then, the sum can be decomposed into the product of the $\bar{r}$ factors that correspond to the single cycle case above, analogously to Section 5.7, which yields

$$\frac{|\mathcal{E}_0|}{|\mathcal{G}_{\mathrm{c}}|\mathbb{E}[\boldsymbol{Z}_{\mathrm{c}}]} = (1+o(1)) \prod_{\ell\in[L]} \lambda_\ell^{r_\ell}(1+\delta_\ell)^{r_\ell}.$$

Now we turn to the proof that the second contribution involving $\mathcal{E}_1$ is negligible, which is a combination of the above and the discussion of intersecting cycles in Section 5.7. We let

$$\begin{aligned}
&\mathcal{E}_2 = \{(G,\gamma,x) : (G,c(\gamma),x) \in \mathcal{E}_1\},\ \mathcal{R} = \{\rho(\gamma) : (G,\gamma,x) \in \mathcal{E}_2\} \text{ and}\\
&\mathcal{E}_\rho = \{(G,\gamma,x) \in \mathcal{E}_2 : \rho(\gamma) = \rho\} \text{ for } \rho \in \mathcal{R}
\end{aligned}$$

denote the sets that match the corresponding sets in Section 5.7. For relative positions $\rho \in \mathcal{R}$ we consider an assignment $y \in \{0,1\}^{n(\rho)}$ of the variables $V = [n(\rho)]$ in the corresponding union of cycles $c = c(\rho)$ and let

$$\begin{aligned}
\mathfrak{r}_1 = \mathfrak{r}_1(\rho,y) &= |\{j \in V : y_j = 1\}|,\\
\mathfrak{o}(b) = \mathfrak{o}_{\rho,y}(b) &= |\{h \in [k_b(\rho)] : y_{i_c(b,h)} = 1\}| \text{ for } b \in [m(\rho)] \text{ and}\\
\mathfrak{o} = \mathfrak{o}(\rho,y) &= \sum_{b \in [m(\rho)]} \mathfrak{o}(b)
\end{aligned}$$

denote the number of variables $j \in V$ in c that take the value one under y, the number of b-edges for a constraint $b \in [m(\rho)]$ in c that take the value one under y and the number of f-edges in c that take the value one under y respectively. Since c is a configuration the number of v-edges in c that take the value one under y is also $\mathfrak{o}$. We are particularly interested in the assignments

$$y \in \mathcal{Y} = \mathcal{Y}(\rho) = \{z \in \{0,1\}^{n(\rho)} : \forall b \in [m(\rho)] \mathfrak{o}(b) \in [2 + k_b - k, 2]\}$$

that do not directly violate a constraint $b \in [m(\rho)]$ in $c(\rho)$ in the sense that $\mathfrak{o}(b) \le 2$ and also do not indirectly violate b in that $2 - \mathfrak{o}(b) \le k - k_b$, i.e. there are sufficiently many b-edges left to take the remaining $(2 - \mathfrak{o}(b))$ ones. With this slight extension of our machinery we can derive

$$\begin{aligned}
\frac{|\mathcal{E}_1|}{|\mathcal{G}_\mathrm{c}|} &= \sum_{\rho \in \mathcal{R}} \frac{|\mathcal{E}_\rho|}{(dn)! \prod_{s \in [\mathfrak{r}]} (2\ell_s)}, \ |\mathcal{E}_\rho| = \sum_{y \in \mathcal{Y}} e_1 e_2 e_3,\\
e_1 = e_1(\rho,y) &= \binom{n}{n_1} n_1^{\underline{\mathfrak{r}_1}} (n-n_1)^{\underline{n(\rho)-\mathfrak{r}_1}} \prod_{j \in [n(\rho)]} d^{\underline{d_j(\rho)}},\\
e_2 = e_2(\rho,y) &= \binom{k}{2}^m m^{\underline{m(\rho)}} \prod_{b \in [m(\rho)]} \left(2^{\underline{\mathfrak{o}(b)}} (k-2)^{\underline{k_b(\rho)-\mathfrak{o}(b)}}\right),\\
e_3 = e_3(\rho,y) &= (dn_1 - \mathfrak{o})!(d(n-n_1) - (e(\rho) - \mathfrak{o}))!,
\end{aligned}$$

for the following reasons. With $\rho \in \mathcal{R}$ and $y \in \mathcal{Y}(\rho)$ fixed we choose the n_1 variables out of the n variables in the configuration G that should take the value one under x. Out of these n_1 variables we choose the $\mathfrak{r}_1$ variables (ordered by first traversal) that take the value one in the directed cycles γ under x, corresponding to the $\mathfrak{r}_1$ variables in ρ that take one under y (more precisely we choose the values $i \in [n]$ of the absolute values α_v for the $\mathfrak{r}_1$ variables $j \in [n(\rho)]$ in ρ that take the value one under y) and analogously for the variables that take zero. Then, for each variable $j \in [n(\rho)]$ in ρ and corresponding variable $i = \alpha_v(j)$ in γ we choose the i-edges that participate in γ (meaning that we choose $\alpha_{v,j}$). On the constraint side we first choose the two a-edges that take the value one under x in G for each $a \in [m]$. Then we select the $m(\rho)$ constraints that participate in γ (i.e. we fix α_f). Further, for each constraint $b \in [m(\rho)]$ in ρ and its corresponding constraint $a = \alpha_f(b)$ in γ we choose the $\mathfrak{o}(b)$ a-edges that take the value one in γ under x consistent with ρ and y out of the two a-edges that take the value one in G under x and analogously for the a-edges that take the value zero (which means that we fix $\alpha_{f,b}$ for $b \in [m(\rho)]$ consistent with the choice of y and the choice of the two a-edges that take the value one for each $a \in [m]$). This fixes the sequence of the directed cycles (i.e. the isomorphism α and further γ). The remaining terms wire the $(dn_1 - \mathfrak{o})$ remaining v-edges that take the value one and the v-edges taking zero respectively.

As opposed to the rather demanding combinatorial part the asymptotics are still easy to derive since both sums are bounded, so the procedure analogous to Section 5.7 yields

$$\frac{|\mathcal{E}_1|}{|\mathcal{G}_\mathrm{c}|\mathbb{E}[\boldsymbol{Z}_\mathrm{c}]} = (1+o(1)) \sum_{\rho\in\mathcal{R}} \sum_{y\in\mathcal{Y}} c_1(\rho, y) n^{n(\rho)+m(\rho)-e(\rho)},$$

where $c_1(\rho, y)$ is a constant compensating the bounded terms. The right hand side tends to zero by the argumentation in Section 5.7, so this contribution is indeed negligible. This shows that $\frac{|\mathcal{E}|}{|\mathcal{G}_\mathrm{c}|} = (1+o(1))\frac{|\mathcal{E}_0|}{|\mathcal{G}_\mathrm{c}|}$ and thereby establishes Theorem 5.7 (b)).

With $d \in [1, d^*) \subseteq [1, k)$ as discussed in Lemma 5.27 and Lemma 5.30, λ_ℓ as derived in Lemma 5.8, $\delta_\ell = (1-k)^{-\ell}$, the asymptotics of the second moment discussed in Lemma 5.6 and the Taylor series $\ln(1-x) = -\sum_{\ell\geq 1} x^\ell/\ell, x \in (0,1)$, we establish Theorem 5.7 (c)) by applying our results to the sum

$$\sum_{\ell\geq 1} \lambda_\ell \delta_\ell^2 = \sum_{\ell\geq 1} \frac{1}{2\ell}\left(\frac{d-1}{k-1}\right)^\ell = -\frac{1}{2}\ln\left(1 - \frac{d-1}{k-1}\right) = \ln\left(\sqrt{\frac{k-1}{k-d}}\right).$$

This concludes the proof of Theorem 5.7 and further the proof of Theorem 2.10.

5.7 Proof of Lemma 5.8

We present the proof of Lemma 5.8 in detail so as to facilitate the presentation of the small subgraph conditioning method in Section 5.6. Lemma 5.8 can be shown by a direct application of the method of moments, which is discussed, for example, in [68] (Theorem 6.10).

Theorem 5.31 (Method of Moments)**.** *Let $L \in \mathbb{Z}_{>0}$ and $((\boldsymbol{X}_{\ell,i})_{\ell\in[L]})_{i\in\mathbb{Z}_{>0}}$ be a sequence of a vector of random variables. If $\lambda \in \mathbb{R}_{\geq 0}^L$ is such that, as $i \to \infty$,*

$$\mathbb{E}\left[\prod_{\ell=1}^{L} (\boldsymbol{X}_{\ell,i})_{r_\ell}\right] \to \prod_{\ell=1}^{L} \lambda_\ell^{r_\ell}$$

for every $r \in \mathbb{Z}_{\geq 0}^L$, then $(\boldsymbol{X}_{\ell,i})_{\ell\in[L]}$ converges in distribution to $(\boldsymbol{Z}_\ell)_{\ell\in[L]}$, where the $\boldsymbol{Z}_\ell \sim \mathrm{Po}(\lambda_\ell)$ are independent Poisson distributed random variables.

First, we notice that $\boldsymbol{G}_\mathrm{c}$ and further $\boldsymbol{X}_\ell = X_\ell(\boldsymbol{G}_\mathrm{c})$ is only defined for $m = dn/k \in \mathbb{Z}$ as stated in Lemma 5.9, hence Lemma 5.8 only applies to such sequences of configurations.

Fix k, $d \in \mathbb{Z}_{>1}$. Before we turn to the general case we consider the $\mathbb{E}[\boldsymbol{X}_\ell]$ for $\ell \in \mathbb{Z}_{>0}$. For this purpose let n and $m(n)$ be sufficiently large. Let $\mathcal{C}_{\ell,G}$ be the set of all 2ℓ-cycles in $G \in \mathcal{G}_\mathrm{c}$. Then

$$\mathbb{E}[\boldsymbol{X}_\ell] = \sum_{G\in\mathcal{G}_\mathrm{c}} \frac{X_\ell(G)}{|\mathcal{G}_\mathrm{c}|} = |\mathcal{G}_\mathrm{c}|^{-1} \sum_{G\in\mathcal{G}_\mathrm{c}} |\mathcal{C}_{\ell,G}| = \frac{|\mathcal{E}|}{|\mathcal{G}_\mathrm{c}|}, \text{ where } \mathcal{E} = \{(G,c) : G \in \mathcal{G}_\mathrm{c}, c \in \mathcal{C}_{\ell,G}\}.$$

With this at hand we obtain that

$$\mathbb{E}[\boldsymbol{X}_\ell] = \frac{1}{2\ell(dn)!} n^{\underline{\ell}} m^{\underline{\ell}} (d(d-1))^\ell (k(k-1))^\ell (dn-2\ell)!$$

using the following combinatorial arguments. Instead of counting pairs (G, c) of configurations G and 2ℓ-cycles $c \in \mathcal{C}_{\ell,G}$ we count pairs (G, γ) of configurations G and *directed* 2ℓ-cycles γ (based at a variable node) in G. There are exactly 2ℓ directed cycles γ corresponding to each (undirected) cycle c of length 2ℓ since we can choose the base from the ℓ variables in c and γ is then determined by one of the two possible directions. The denominator reflects the compensation for this counting next to the probability $|\mathcal{G}_\mathrm{c}|^{-1}$. Further, the term $n^{\underline{\ell}}$ reflects the ordered choice of the variables for the directed cycle, as does $m^{\underline{\ell}}$ for the constraints. The next two terms account for the choice of the two i-edges and a-edges traversed by the cycle for each of the ℓ variables i and constraints a. This fixes the directed cycle γ and further the corresponding undirected cycle $c(\gamma)$. In particular, the 2ℓ edges of the cycle c in G are fixed, i.e. the corresponding restriction of G to c. This leaves us with $(dn-2\ell)$ half-edges in $[n] \times [d]$ and $(km-2\ell)$ half-edges in $[m] \times [k]$ that have not been wired yet. The last term gives the number of such wirings.

Next, we turn to asymptotics. Extracting λ_ℓ and expanding the falling factorials yields

$$\mathbb{E}[\boldsymbol{X}_\ell] = \lambda_\ell d^\ell k^\ell \frac{n!m!(dn-2\ell)!}{(dn)!(n-\ell)!(m-\ell)!}.$$

Using Stirling's formula we readily obtain that

$$\mathbb{E}[\boldsymbol{X}_\ell] = (1+o(1))\lambda_\ell d^\ell k^\ell \sqrt{\frac{nm(dn-2\ell)}{dn(n-\ell)(m-\ell)}} \frac{n^n m^m (dn-2\ell)^{dn-2\ell}}{(dn)^{dn}(n-\ell)^{n-\ell}(m-\ell)^{m-\ell}},$$

and so

$$\mathbb{E}[\boldsymbol{X}_\ell] = (1+o(1))\lambda_\ell d^\ell k^\ell \sqrt{\frac{(1-\frac{2\ell}{dn})}{(1-\frac{\ell}{n})(1-\frac{\ell}{m})}} \frac{n^\ell m^\ell (1-\frac{2\ell}{dn})^{dn-2\ell}}{(dn)^{2\ell}(1-\frac{\ell}{n})^{n-\ell}(1-\frac{\ell}{m})^{m-\ell}} = (1+o(1))\lambda_\ell d^\ell k^\ell \frac{n^\ell m^\ell}{(dn)^{2\ell}}.$$

Using that $dn = km$ leads to

$$\mathbb{E}[\boldsymbol{X}_\ell] = (1+o(1))\lambda_\ell d^\ell k^\ell \frac{n^\ell (dk^{-1}n)^\ell}{(dn)^{2\ell}} = \lambda_\ell,$$

as claimed. We turn to the general case. For this purpose let $L \in \mathbb{Z}_{>0}$, $r \in \mathbb{Z}_{\geq 0}^L$ and let n and m be sufficiently large. Then

$$X_\ell(G)^{\underline{r_\ell}} = \prod_{s=0}^{r_\ell - 1} (|\mathcal{C}_{\ell,G}| - s) = |\mathcal{C}_{\ell,r_\ell,G}| \text{ , where } \mathcal{C}_{\ell,r_\ell,G} = \{c \in \mathcal{C}_{\ell,G}^{r_\ell} : \forall s \in [r_\ell] \forall s' \in [s-1]\, c_s \neq c_{s'}\}$$

for $G \in \mathcal{G}_\mathrm{c}$, since this corresponds to an ordered choice of 2ℓ-cycles in G without repetition. The product can then be directly written as

$$\prod_{\ell=1}^{L} X_\ell(G)^{\underline{r_\ell}} = |\mathcal{C}_{r,G}| \text{ , where } \mathcal{C}_{r,G} = \prod_{\ell=1}^{L} \mathcal{C}_{\ell,r_\ell,G}.$$

To avoid double indexed sequences we use the equivalent representation $c = (c_s)_{s\in[\mathfrak{r}]} \in \mathcal{C}_{r,G}$ where $\mathfrak{r} = \sum_{1\le\ell\le L} r_\ell$. From the above we see that the cycles c_s are ordered by their length ℓ_s in ascending order and are pairwise distinct. We obtain that

$$\mathbb{E}\left[\prod_{\ell=1}^{L} \boldsymbol{X}_\ell^{r_\ell}\right] = \frac{|\mathcal{E}|}{|\mathcal{G}_{\mathrm{c}}|} \ \text{, where} \quad \mathcal{E} = \{(G,c) : G \in \mathcal{G}_{\mathrm{c}}, c \in \mathcal{C}_{r,G}\}.$$

Since we have ℓ_s distinct variables and constraints in each cycle c_s respectively, we can have at most $\mathfrak{l} = \sum_{s\in[\mathfrak{r}]} \ell_s$ distinct variables and constraints in c. Specifically, we only have $|V(c)| = \mathfrak{l}$ variables and $|F(c)| = \mathfrak{l}$ constraints iff all cycles c_s are disjoint. So, let

$$\mathcal{E}_0 = \{(G,c) \in \mathcal{E} : |V(c)| = |F(c)| = \mathfrak{l}\}$$

denote the set of pairs $(g,c) \in \mathcal{E}$ with disjoint cycles and further $\mathcal{E}_1 = \mathcal{E} \setminus \mathcal{E}_0$ the remaining pairs. Then we have

$$\frac{|\mathcal{E}_0|}{|\mathcal{G}_{\mathrm{c}}|} = \frac{1}{(dn)! \prod_{s=1}^{\mathfrak{r}} (2\ell_s)} n^{\underline{\mathfrak{l}}} m^{\underline{\mathfrak{l}}} (d(d-1)^{\mathfrak{l}} (k(k-1))^{\mathfrak{l}} (dn-2\mathfrak{l})!$$

for the following reasons. For each cycle c_s in c counting the $2\ell_s$ directed cycles facilitates the computation, hence we find the corresponding product in the denominator. Since the variables within each directed cycle and the cycles in the sequence are ordered we have an ordered choice of all variables. Further, since the ℓ_s variables within each cycle are distinct and the cycles are pairwise disjoint we choose all variables without repetition. This explains the first falling factorial. The next term for the constraints follows analogously. But since variables and constraints are disjoint the edges are too, hence we choose two edges for each of the $\mathfrak{l}$ variables and constraints respectively. Then we wire the remaining edges.
The asymptotics are derived analogously to the base case, i.e.

$$\frac{|\mathcal{E}_0|}{|\mathcal{G}_{\mathrm{c}}|} = (1+o(1)) \frac{(d-1)^{\mathfrak{l}}(k-1)^{\mathfrak{l}}}{\prod_{s=1}^{\mathfrak{r}} (2\ell_s)} = (1+o(1)) \prod_{s=1}^{\mathfrak{r}} \lambda_{\ell_s} = (1+o(1)) \prod_{\ell=1}^{L} \lambda_\ell^{r_\ell},$$

using the definition of $c = (c_s)_{s\in[\mathfrak{r}]}$ in the last step. Since the contribution of the disjoint cycles already yields the desired result, we want to show that the contribution of intersecting cycles is negligible. As before, we count directed cycles γ_s and adjust the result accordingly, so let

$$\mathcal{E}_2 = \{(G,\gamma) : (G, c(\gamma)) \in \mathcal{E}_1\}, \text{ i.e. } |\mathcal{E}_2| = |\mathcal{E}_1| \prod_{s\in[\mathfrak{r}]} (2\ell_s).$$

In the next step we consider the relative position representations (α, ρ) of sequences γ of directed cycles. Instead of a formal introduction we illustrate this concept in Figure 6. The corresponding decomposition of the contributions to the expectation according to ρ is

$$\frac{|\mathcal{E}_1|}{|\mathcal{G}_{\mathrm{c}}|} = \sum_{\rho\in\mathcal{R}} \frac{|\mathcal{E}_\rho|}{|\mathcal{G}_{\mathrm{c}}| \prod_{s\in[\mathfrak{r}]} (2\ell_s)}, \ \mathcal{E}_\rho = \{(G,\gamma) \in \mathcal{E}_2 : \rho(\gamma) = \rho\}, \ \mathcal{R} = \{\rho(\gamma) : (G,\gamma) \in \mathcal{E}_2\}.$$

For the following reasons we can then derive

$$|\mathcal{E}_\rho| = n^{\underline{n(\rho)}} m^{\underline{m(\rho)}} \prod_{j\in[n(\rho)]} d^{\underline{d_j(\rho)}} \prod_{b\in[m(\rho)]} k^{\underline{k_b(\rho)}} (dn - e(\rho))!.$$

Since ρ is fixed, we have to fix the absolute values α, thereby the directed cycle γ, and wire the remaining edges. But the first four terms exactly correspond to the number of choices for the index vectors in α. This fixes γ, further the union $c(\gamma)$ of cycles and in particular $e(\rho)$ edges. The remaining term counts the number of choices to wire the remaining edges.

For the asymptotics we notice that $n(\rho)$, $m(\rho) \leq \mathfrak{l}$ and that also the two products are bounded in both the multiplication region and values. But this further implies that $|\mathcal{R}|$ is bounded, i.e. the summation region is also finite in the limit and hence we can consider the asymptotics of each term separately, which yields

$$\begin{aligned}
\frac{|\mathcal{E}_1|}{|\mathcal{G}_{\mathrm{c}}|} &= \sum_{\rho\in\mathcal{R}} \frac{\prod_{i\in[n(\rho)]} d^{\underline{d_i(\rho)}} \prod_{a\in[m(\rho)]} k^{\underline{k_a(\rho)}}}{\prod_{s\in[\mathfrak{r}]}(2\ell_s)} \frac{n^{\underline{n(\rho)}} m^{\underline{m(\rho)}}(dn - e(\rho))!}{(dn)!} \\
&= \sum_{\rho\in\mathcal{R}} c_1(\rho) \frac{n^{\underline{n(\rho)}} m^{\underline{m(\rho)}}(dn - e(\rho))!}{(dn)!} \\
&= (1+o(1)) \sum_{\rho\in\mathcal{R}} c_1(\rho) \left(\frac{1}{e}\right)^{n(\rho)} \left(\frac{d}{ke}\right)^{m(\rho)} \left(\frac{e}{d}\right)^{e(\rho)} n^{n(\rho)+m(\rho)-e(\rho)} \\
&= (1+o(1)) \sum_{\rho\in\mathcal{R}} c_2(\rho) n^{n(\rho)+m(\rho)-e(\rho)},
\end{aligned}$$

where we summarized the terms that only depend on ρ into constants. Now, let $\rho \in \mathcal{R}$ and let $c = c(\rho)$ be the graph of ρ as introduced in Section 5.2.4. Since ρ is a sequence of directed cycles that are not all disjoint, its graph c is the union of the corresponding (undirected) cycles that are not all disjoint. But then c has more edges than vertices, i.e. $3e(\rho) > n(\rho) + m(\rho) + 2e(\rho)$, and hence

$$\frac{|\mathcal{E}_1|}{|\mathcal{G}_{\mathrm{c}}|} = (1+o(1)) \sum_{\rho\in\mathcal{R}} c_2(\rho) n^{n(\rho)+m(\rho)-e(\rho)} \leq (1+o(1)) n^{-1} \sum_{\rho\in\mathcal{R}} c_2(\rho) = (1+o(1)) c_3 n^{-1},$$

which shows that this contribution is negligible. This establishes the asymptotic equivalence

$$\mathbb{E}\left[\prod_{\ell\in[L]} \boldsymbol{X}_\ell^{\underline{r_\ell}}\right] = (1+o(1)) \prod_{\ell\in[L]} \lambda_\ell^{r_\ell}$$

and allows to apply the method of moments, which directly yields Lemma 5.8.

6 Bibliography

[1] E. Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18:1–86, 2018.

[2] E. Abbe and E. Boix-Adserà. An information-percolation bound for spin synchronization on general graphs. *Ann. Appl. Probab.*, 30(3):1066–1090, 2020.

[3] E. Abbe, E. Cornacchia, Y. Gu, and Y. Polyanskiy. Stochastic block model entropy and broadcasting on trees with survey. In *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134, pages 1–25. PMLR, 15–19 Aug 2021.

[4] E. Abbe and A. Montanari. Conditional random fields, planted constraint satisfaction, and entropy concentration. *Theory Comput.*, 11:413–443, 2015.

[5] E. Abbe and C. Sandon. Community detection in general stochastic block models: fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015*, pages 670–688. IEEE Computer Soc., Los Alamitos, CA, 2015.

[6] E. Abbe and C. Sandon. Recovering communities in the general stochastic block model without knowing the parameters. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[7] E. Abbe and C. Sandon. Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[8] E. Abbe and C. Sandon. Crossing the KS threshold in the stochastic block model with information theory. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 840–844, 2016.

[9] E. Abbe and C. Sandon. Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.*, 71(7):1334–1406, 2018.

[10] M. Alekhnovich. More on average case vs approximation complexity. *Comput. Complexity*, 20(4):755–786, 2011.

[11] N. Alon and R. Yuster. Threshold functions for H-factors. *Combin. Probab. Comput.*, 2(2):137–144, 1993.

[12] B. Applebaum, B. Barak, and A. Wigderson. Public-key cryptography from different assumptions. In *STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 171–180. ACM, New York, 2010.

[13] T. Austin. Multi-variate correlation and mixtures of product measures. *Kybernetika*, 56(3):459–499, 2020.

[14] P. Ayre, A. Coja-Oghlan, and C. Greenhill. Lower Bounds on the Chromatic Number of Random Graphs. *Combinatorica*, 42(5):617–658, 2022.

[15] M. A. Bahmanian and M. Šajna. Quasi-Eulerian hypergraphs. *Electron. J. Combin.*, 24(3):Paper 3.30, 12, 2017.

[16] J. Banks, C. Moore, J. Neeman, and P. Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 383–416, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.

[17] V. Bapst, A. Coja-Oghlan, and C. Efthymiou. Planting colourings silently. *Combin. Probab. Comput.*, 26(3):338–366, 2017.

[18] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Raßmann, and D. Vilenchik. The condensation phase transition in random graph coloring. *Comm. Math. Phys.*, 341(2):543–606, 2016.

[19] W. Barthel, A. K. Hartmann, M. Leone, F. Ricci-Tersenghi, M. Weigt, and R. Zecchina. Hiding solutions in random satisfiability problems: A statistical mechanics approach. *Phys. Rev. Lett.*, 88:188701, Apr 2002.

[20] D. Berend and A. Kontorovich. On the convergence of the empirical distribution. *arXiv preprint arXiv:1205.6711v2*, 2012.

[21] H. A. Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 150(871):552–575, 1935.

[22] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.

[23] B. Bollobás and A. Thomason. Random graphs of small order. In *Random graphs '83 (Poznań, 1983)*, volume 118 of *North-Holland Math. Stud.*, pages 47–97. North-Holland, Amsterdam, 1985.

[24] J. Brand and N. Jaafari. The mutual information of LDGM codes. *arXiv preprint arXiv:1707.04413v2*, 2017.

[25] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: an algorithm for satisfiability. *Random Structures Algorithms*, 27(2):201–226, 2005.

[26] F. Caltagirone, M. Lelarge, and L. Miolane. Recovering asymmetric communities in the stochastic block model. *IEEE Transactions on Network Science and Engineering*, 5(3):237–246, 2018.

[27] T. Castellani, V. Napolano, F. Ricci-Tersenghi, and R. Zecchina. Bicolouring random hypergraphs. *J. Phys. A*, 36(43):11037–11053, 2003.

[28] F. Chung and R. Graham. *Erdős on graphs: His legacy of unsolved problems.* AK Peters/CRC Press, 1998.

[29] A. Coja-Oghlan. Constraint satisfaction: random regular k-SAT. In *Statistical physics, optimization, inference, and message-passing algorithms*, pages 231–251. Oxford Univ. Press, Oxford, 2016.

[30] A. Coja-Oghlan, C. Efthymiou, N. Jaafari, M. Kang, and T. Kapetanopoulos. Charting the replica symmetric phase. *Comm. Math. Phys.*, 359(2):603–698, 2018.

[31] A. Coja-Oghlan, M. Hahn-Klimroth, P. Loick, N. Müller, K. Panagiotou, and M. Pasch. Inference and mutual information on random factor graphs. In *38th International Symposium on Theoretical Aspects of Computer Science*, volume 187 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 24, 15. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021.

[32] A. Coja-Oghlan, T. Kapetanopoulos, and N. Müller. The replica symmetric phase of random constraint satisfaction problems. *Combin. Probab. Comput.*, 29(3):346–422, 2020.

[33] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová. Information-theoretic thresholds from the cavity method. *Adv. Math.*, 333:694–795, 2018.

[34] A. Coja-Oghlan and K. Panagiotou. The asymptotic k-SAT threshold. *Adv. Math.*, 288:985–1068, 2016.

[35] A. Coja-Oglan and K. Panagiotou. Catching the K-NAESAT threshold. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12, page 899–908. Association for Computing Machinery, 2012.

[36] C. Cooper, A. Frieze, M. Molloy, and B. Reed. Perfect matchings in random r-regular, s-uniform hypergraphs. *Combin. Probab. Comput.*, 5(1):1–14, 1996.

[37] L. Dall'Asta, A. Ramezanpour, and R. Zecchina. Entropy landscape and non-Gibbs solutions in constraint satisfaction problems. *Phys. Rev. E (3)*, 77(3):031118, 16, 2008.

[38] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, Dec 2011.

[39] P. Devlin and J. Kahn. Perfect fractional matchings in k-out hypergraphs. *Electron. J. Combin.*, 24(3):Paper No. 3.60, 12, 2017.

[40] J. Ding, A. Sly, and N. Sun. Proof of the satisfiability conjecture for large k. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 59–68. ACM, New York, 2015.

[41] J. Ding, A. Sly, and N. Sun. Maximum independent sets on random regular graphs. *Acta Math.*, 217(2):263–340, 2016.

[42] J. Ding, A. Sly, and N. Sun. Satisfiability threshold for random regular NAE-SAT. *Comm. Math. Phys.*, 341(2):435–489, 2016.

[43] T. Dominguez and J. Mourrat. Mutual information for the sparse stochastic block model. *arXiv preprint arXiv:2209.04513v2*, 2022.

[44] A. Dudek and A. Frieze. Loose Hamilton cycles in random uniform hypergraphs. *Electron. J. Combin.*, 18(1):Paper 48, 14, 2011.

[45] A. Dudek, A. Frieze, P. Loh, and S. Speiss. Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs. *Electron. J. Combin.*, 19(4):Paper 44, 17, 2012.

[46] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[47] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[48] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12:261–267, 1961.

[49] P. Erdős and A. Rényi. On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hungar.*, 17:359–368, 1966.

[50] P. Erdős. On the combinatorial problems which I would most like to see solved. *Combinatorica*, 1(1):25–42, 1981.

[51] V. Feldman, W. Perkins, and S. Vempala. On the complexity of random satisfiability problems with planted solutions. *SIAM J. Comput.*, 47(4):1294–1338, 2018.

[52] A. Ferber. Closing gaps in problems related to Hamilton cycles in random graphs and hypergraphs. *Electron. J. Combin.*, 22(1):Paper 1.61, 7, 2015.

[53] S. Franz and M. Leone. Replica bounds for optimization problems and diluted spin systems. *J. Statist. Phys.*, 111(3-4):535–564, 2003.

[54] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain.

[55] A. Frieze and S. Janson. Perfect matchings in random s-uniform hypergraphs. *Random Structures Algorithms*, 7(1):41–57, 1995.

[56] A. Frieze, M. Jerrum, M. Molloy, R. W. Robinson, and N. C. Wormald. Generating and counting Hamilton cycles in random regular graphs. *J. Algorithms*, 21(1):176–198, 1996.

[57] A. Frieze and M. Karoński. *Introduction to random graphs.* Cambridge University Press, Cambridge, 2016.

[58] M. Gabrié, V. Dani, G. Semerjian, and L. Zdeborová. Phase transitions in the q-coloring of random hypergraphs. *J. Phys. A*, 50(50):505002, 44, 2017.

[59] R. G. Gallager. Low-density parity-check codes. *IRE Trans.*, IT-8:21–28, 1962.

[60] A. Giurgiu, N. Macris, and R. Urbanke. Spatial coupling as a proof technique and three applications. *IEEE Trans. Inform. Theory*, 62(10):5281–5295, 2016.

[61] O. Guédon and R. Vershynin. Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields*, 165(3-4):1025–1049, 2016.

[62] F. Guerra and F. Toninelli. The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.*, 230(1):71–79, 2002.

[63] A. Heckel. Random triangles in random graphs. *Random Structures Algorithms*, 59(4):616–621, 2021.

[64] A. Heckel, M. Kaufmann, N. Müller, and M. Pasch. The hitting time of clique factors. *arXiv preprint arXiv:2302.08340*, 2023.

[65] S. Higuchi and M. Mézard. Correlation-based decimation in constraint satisfaction problems. *Journal of Physics: Conference Series*, 233(1):012003, 2010.

[66] V. Jain, F. Koehler, and A. Risteski. Mean-field approximation, convex hierarchies, and the optimality of correlation rounding: A unified perspective. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 1226–1236. Association for Computing Machinery, 2019.

[67] S. Janson. Random regular graphs: asymptotic distributions and contiguity. *Combin. Probab. Comput.*, 4(4):369–405, 1995.

[68] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[69] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures Algorithms*, 33(1):1–28, 2008.

[70] J. Kahn. Asymptotics for shamir's problem. *arXiv preprint arXiv:1909.06834*, 2019.

[71] J. Kahn. Hitting times for Shamir's problem. *Trans. Amer. Math. Soc.*, 375(1):627–668, 2022.

[72] V. Kalapala and C. Moore. The phase transition in exact cover. *Chic. J. Theoret. Comput. Sci.*, pages Article 5, 9, 2008.

[73] G. Kemkes, X. Pérez-Giménez, and N. C. Wormald. On the chromatic number of random d-regular graphs. *Adv. Math.*, 223(1):300–328, 2010.

[74] J. Kim. Perfect matchings in random uniform hypergraphs. *Random Structures Algorithms*, 23(2):111–132, 2003.

[75] A. Klenke. *Probability theory.* Universitext. Springer, London, second edition, 2014. A comprehensive course.

[76] M. Krivelevich. Triangle factors in random graphs. *Combin. Probab. Comput.*, 6(3):337–347, 1997.

[77] F. Krzakala, F. Ricci-Tersenghi, L. Zdeborová, R. Zecchina, E. W. Tramel, and L. F. Cugliandolo, editors. *Statistical physics, optimization, inference, and message-passing algorithms.* Oxford University Press, Oxford, 2016.

[78] S. Kudekar, S. Kumar, M. Mondelli, H. Pfister, E. Şaşoğlu, and R. Urbanke. Reed-muller codes achieve capacity on erasure channels. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 658–669. Association for Computing Machinery, 2016.

[79] S. Kudekar, T. Richardson, and R. Urbanke. Spatially coupled ensembles universally achieve capacity under belief propagation. *IEEE Trans. Inform. Theory*, 59:7761–7813, 2013.

[80] K. Kumar, P. Pakzad, A. Salavati, and A. Shokrollahi. Phase transitions for mutual information. In *2010 6th International Symposium on Turbo Codes & Iterative Information Processing*, pages 137–141, 2010.

[81] C. Lecoutre. *Constraint Networks: Targeting Simplicity for Techniques and Algorithms.* ISTE. Wiley, 2013.

[82] A. Makur and L. Chzhèn. Comparison of contraction coefficients for f-divergences. *Problemy Peredachi Informatsii*, 56(2):3–63, 2020.

[83] E. Maneva, E. Mossel, and M. J. Wainwright. A new look at survey propagation and its generalizations. *J. ACM*, 54(4):Art. 17, 41, 2007.

[84] L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing*, pages 694–703. ACM, New York, 2014.

[85] M. Mézard and A. Montanari. *Information, physics, and computation.* Oxford Graduate Texts. Oxford University Press, Oxford, 2009.

[86] M. Mézard, G. Parisi, and M. A. Virasoro. SK model: The replica solution without replicas. *EPL (Europhysics Letters)*, 1(2):77, 1986.

[87] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin glass theory and beyond*, volume 9 of *World Scientific Lecture Notes in Physics.* World Scientific Publishing Co., Inc., Teaneck, NJ, 1987.

[88] M. Molloy. The freezing threshold for k-colourings of a random graph. *J. ACM*, 65(2):Art. 7, 62, 2018.

[89] M. Molloy, H. Robalewska, R. Robinson, and N. Wormald. 1-factorizations of random regular graphs. *Random Structures Algorithms*, 10(3):305–321, 1997.

[90] M. Mondelli, R. Urbanke, and S. Hassani. How to achieve the capacity of asymmetric channels. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 789–796, 2014.

[91] A. Montanari. Tight bounds for ldpc and ldgm codes under map decoding. *IEEE Transactions on Information Theory*, 51(9):3221–3246, 2005.

[92] A. Montanari. Estimating random variables from random sparse observations. *European Transactions on Telecommunications*, 19(4):385–403, 2008.

[93] A. Montanari. Finding one community in a sparse graph. *J. Stat. Phys.*, 161(2):273–299, 2015.

[94] A. Montanari, R. Restrepo, and P. Tetali. Reconstruction and clustering in random constraint satisfaction problems. *SIAM J. Discrete Math.*, 25(2):771–808, 2011.

[95] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Clusters of solutions and replica symmetry breaking in randomk-satisfiability. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(04):P04004, apr 2008.

[96] C. Moore. The phase transition in random regular exact cover. *Ann. Inst. Henri Poincaré D*, 3(3):349–362, 2016.

[97] C. Moore. The computer science and physics of community detection: landscapes, phase transitions, and hardness. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, pages 26–61, 2017.

[98] T. Mora. *Geometry and Inference in Optimization and in Information Theory.* Theses, Université Paris Sud - Paris XI, September 2007.

[99] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields*, 162(3-4):431–461, 2015.

[100] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

[101] K. Panagiotou and M. Pasch. Satisfiability thresholds for regular occupation problems. In *46th International Colloquium on Automata, Languages, and Programming*, volume 132 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 90, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019.

[102] K. Panagiotou and M. Pasch. Mutual information, information-theoretic thresholds and the condensation phenomenon at positive temperature. *arXiv preprint arXiv:2207.11002*, 2022.

[103] K. Panagiotou and M. Pasch. Satisfiability thresholds for regular occupation problems. *arXiv preprint arXiv:1811.00991v3*, 2023.

[104] V. Panaretos and Y. Zemel. *An invitation to statistics in Wasserstein space.* SpringerBriefs in Probability and Mathematical Statistics. Springer, Cham, 2020.

[105] D. Panchenko and M. Talagrand. Bounds for diluted mean-fields spin glass models. *Probab. Theory Related Fields*, 130(3):319–336, 2004.

[106] G Parisi. A sequence of approximated solutions to the S-K model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, apr 1980.

[107] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[108] J. Petersen. Die theorie der regulären graphs. *Acta Math.*, 15:193–220, 1891.

[109] K. Petrova and M. Trujić. Transference for loose hamilton cycles in random 3-uniform hypergraphs. *arXiv preprint arXiv:2205.11421*, 2022.

[110] D. Poole. On the strength of connectedness of a random hypergraph. *Electron. J. Combin.*, 22(1):Paper 1.69, 16, 2015.

[111] G. Reeves, V. Mayya, and A. Volfovsky. The geometry of community detection via the mmse matrix. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 400–404, 2019.

[112] F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová. Typology of phase transitions in bayesian inference problems. *Phys. Rev. E*, 99:042109, Apr 2019.

[113] O. Riordan. Random cliques in random graphs and sharp thresholds for F-factors. *Random Structures & Algorithms*, 61(4):619–637, 2022.

[114] H. Robalewska. 2-factors in random regular graphs. *J. Graph Theory*, 23(3):215–224, 1996.

[115] H. Robbins. A remark on Stirling's formula. *Amer. Math. Monthly*, 62:26–29, 1955.

[116] N. Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.

[117] A. Ruciński. Matching and covering the vertices of a random graph by copies of a given graph. *Discrete Math.*, 105(1-3):185–197, 1992.

[118] Z. Ruttkay. Constraint satisfaction—a survey. *CWI Quarterly*, 11(2-3):123–161, 1998.

[119] C. Schmidt, N. Guenther, and L. Zdeborová. Circular coloring of random graphs: statistical physics investigation. *J. Stat. Mech. Theory Exp.*, pages 083303, 28, 2016.

[120] J. Schmidt and E. Shamir. A threshold for perfect matchings in random d-pure hypergraphs. *Discrete Math.*, 45(2-3):287–295, 1983.

[121] J. Spencer. Threshold functions for extension statements. *J. Combin. Theory Ser. A*, 53(2):286–305, 1990.

[122] M. Talagrand. *Spin glasses: a challenge for mathematicians*, volume 46 of *Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics*. Springer-Verlag, Berlin, 2003.

[123] M. Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.

[124] E. Tsang. *Foundations of Constraint Satisfaction.* Computation in cognitive science. Academic Press, 1993.

[125] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[126] J. Yedidia, W. Freeman, and Y. Weiss. Bethe free energy, kikuchi approximations, and belief propagation algorithms. Technical Report TR2001-16, MERL - Mitsubishi Electric Research Laboratories, Cambridge, MA 02139, May 2001.

[127] L. Zdeborová and F. Krzakala. Quiet planting in the locked constraint satisfaction problems. *SIAM J. Discrete Math.*, 25(2):750–770, 2011.

[128] L. Zdeborová and M. Mézard. Constraint satisfaction problems with isolated solutions are hard. *J. Stat. Mech. Theory Exp.*, 2008(12):P12004, 2008.

[129] L. Zdeborová and F. Krzakala. Statistical physics of inference: thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

[130] P. Zhang, C. Moore, and M. Newman. Community detection in networks with unequal groups. *Phys. Rev. E*, 93(1):012303, 12, 2016.