

## THE UNIVERSITY of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.


# Algorithms and complexity for approximately counting hypergraph colourings and related problems 

Fiaheng Wang

Doctor of Philosophy<br>Laboratory for Foundations of Computer Science<br>School of Informatics<br>University of Edinburgh

2023

## Abstract

The past decade has witnessed advancements in designing efficient algorithms for approximating the number of solutions to constraint satisfaction problems (CSPs), especially in the local lemma regime. However, the phase transition for the computational tractability is not known. This thesis is dedicated to the prototypical problem of this kind of CSPs, the hypergraph colouring. Parameterised by the number of colours $q$, the arity of each hyperedge $k$, and the vertex maximum degree $\Delta$, this problem falls into the regime of Lovász local lemma when $\Delta \lesssim q^{k}$. In prior, however, fast approximate counting algorithms exist when $\Delta \lesssim q^{k / 3}$, and there is no known inapproximability result. In pursuit of this, our contribution is two-folded, stated as follows.

- When $q, k \geq 4$ are evens and $\Delta \geq 5 \cdot q^{k / 2}$, approximating the number of hypergraph colourings is NP-hard.
- When the input hypergraph is linear and $\Delta \lesssim q^{k / 2}$, a fast approximate counting algorithm does exist.


## Acknowledgements

I cannot be more grateful to my parents, but I still would like to venture into expressing this in my own words. I was not born in a family of scholars, neither a place considered by most people where knowledge is inherited. No one in the village prior to me had ever received college-level education. To gain money, my parents moved to urban area, where my mother kept a market booth, and my father worked as a repairman, electrician and delivery man for our booth. Nothing did sound like a proper place for a five-year-old kid to receive education, but my affection for mathematics started there. Under the guidance of my parents, I had a lot of fun reading outdated newspapers, playing with calculators and cards. I also knew what voltage means, and the nominal voltage of many common drycell batteries by playing my father's voltohmmeter. The childhood is a period of exploration for everyone, and my parents guided me into science, despite not holding any degree in any kind of science. Later on, they made an important decision, and took me to several best primary schools in that city, trying to persuade the principles to get me enrolled. That was not a hard task: I amazed one of them by showing off what I had learnt, including how to properly do $2 \times 2 \times 2$ and form 24 using $1,5,5,5$ and three arithmetic operations (the solution is $5 \times(5-1 \div 5)=24)$. Since then, they focused overwhelmingly on my performance at school, as if they were the ones attending school. Every afternoon they queried my class' teachers about my performance when waiting for the queue to pick me up. As I progressed to higher years, they were no longer able to do any teaching at home, but they still kept urging my teachers on a weekly basis to ensure the quality of education till I gained self-discipline in high school. Things went on quite well, and it was very fortunate for me to be admitted to Peking University for undergraduate study, where I was obsessed in randomised algorithms and complexity: I will talk about this later. In all, I do not think I could achieve the same if I were not born in a family where changing life by education is a belief. They are also in full support of my pursuit of a PhD oversea, and had tried very hard persuading my grandparents regarding this.

I would also want to express my gratitude towards my supervisor Heng Guo. We met for the first time in the summer of 2019 when I was looking for summer research internship on algorithms and complexity. That was also the first time I had been outside China, with Edinburgh being the first destination. I had a rather great time during the stay, and decided to get back to do a PhD. Definitely, doing research itself is a path of thorns and agony, and I must be honest that the journey towards
knowledge is not pleasant, but I managed to acquire some achievements that I take pride in together with Heng. Our relation is more like coauthorship, rather than supervision, and I have learnt a lot from him beyond academic topics. We shared some common hobbies, yet a lot of other differences, including, most importantly, the way of thinking. Though we need to adapt to each other especially when writing papers, it helps me quite much to reconcile my foolhardiness. For this reason, I would also want to thank him for forgiveness - my personality would definitely become a troublemaker for others if it were not him!

There are many faculty members that I would like to express gratitude to, for all kinds of reasons, including Mary Cryan, Radu Curticapean, Xiaotie Deng, Charilaos Efthymiou, Kousha Etessami, Aris Filos-Ratsikas, Andreas Galanis, Raul GarciaPatron, Leslie Ann Goldberg, Kun He, Tyler Helmuth, Mark Jerrum, John Longley, Pinyan Lu, Shuai Shao, Daniel Štefankovič, Xiaoming Sun, Eric Vigoda, Kuan Yang, Yitong Yin, Chihao Zhang (sorted in alphabetical order), and many others from the cohort of PhD or postdoc, including Konrad Anand, Weiming Feng, Graham Freifeld, Yao Fu, Wenzhi Fu, Shangmin Guo, Ben Jourdan, Steinar Laenen, Chao Liao, Mingmou Liu, Muyang Liu, Peter Macgregor, Isja Mannens, Giorgos Mousa, Vishvajeet N, Guillem Ramirez, Guoliang Qiu, Chunyang Wang, Kewen Wu, and a lot of names (sorted in alphabetical order, again). Among these names I wish to express a special thank to Weiming Feng, a frequent and talented coauthor, and a powerful rival on the other side of the boardgame table in front of Heng's bookshelf of boardgames. Wish him the best luck of job-hunting! I would also want to spend a few words acknowledging Giorgos Mousa, for the exchange of thoughts built upon our shared interest in metal music and all the existential discussion.

Finally, my PhD study is financially supported by an Informatics Global PhD Scholarship at The University of Edinburgh, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 947778).

## Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.
(Fiaheng Wang)

## Table of Contents

1 Introduction ..... 1
2 Preliminaries ..... 9
2.1 Graph theory ..... 9
$2.2 \quad f$-divergences ..... 10
2.3 Markov chain and mixing time ..... 12
2.4 The local lemma ..... 13
2.5 Approximate counting and sampling algorithms ..... 14
3 Computing the total variation distance ..... 17
3.1 The algorithm ..... 19
4 Inapproximability ..... 25
4.1 Hardness for searching ..... 27
4.1.1 Proof of Lemma 4.5 ..... 32
4.2 Reducing from spin systems ..... 35
4.3 Analysis of the dominant phases ..... 38
4.3.1 Restricting to three values ..... 41
4.3.2 2,3-maximal triples ..... 46
4.3.3 Stability of maximal $(q / 2, q / 2,0)$ fixpoints ..... 49
4.3.4 (In)stability of $(q, 0,0)$ fixpoints ..... 51
4.3.5 $(q, 0,0)$ fixpoint is not maximal ..... 53
4.4 Remaining proofs of this chapter ..... 60
4.4.1 Proof of Lemma 4.29 and Lemma 4.23 ..... 60
4.4.2 Proof of Lemma 4.21 ..... 61
4.4.3 Proof of Lemma 4.33 and Lemma 4.39 ..... 62
4.4.4 Proof of Lemma 4.35 and Lemma 4.36 ..... 64
4.4.5 Proof of Lemma 4.45 ..... 65
5 FPRAS for linear hypergraphs ..... 67
5.1 Preliminaries of this chapter ..... 69
5.1.1 List hypergraph colouring and local uniformity ..... 69
5.1.2 Projection scheme, projected distribution and conditional dis- tribution ..... 70
5.2 Algorithm ..... 71
5.2.1 The sampling algorithm ..... 71
5.2.2 Proof of the main theorem ..... 73
5.3 Analysis of the Sample subroutine ..... 78
5.3.1 Proof of running time and correctness ..... 78
5.3.2 Bound the probability of $\mathcal{B}_{\text {rej }}(t)$ ..... 79
5.4 Analysis of connected components ..... 80
5.4.1 Proof of Lemma 5.10 ..... 81
5.4.2 Properties of the 2-block-tree generator ..... 90
5.4.3 Property of random configurations ..... 92
5.4.4 Properties of 2-block-trees ..... 93
5.5 Mixing of systematic scan ..... 96
5.5.1 Information percolation analysis ..... 97
5.5.2 Extended hyperedges and the extended hypergraph ..... 99
5.5.3 Proof of Lemma 5.21 ..... 102
5.5.4 Proof of Lemma 5.32 ..... 106
6 Ferromagnetic Ising model ..... 113
6.1 Preliminaries of this chapter ..... 119
6.1.1 The models and their equivalences ..... 119
6.1.2 Markov chains and down-up walks ..... 121
6.1.3 Canonical paths and variance decay ..... 124
6.1.4 Spectral independence and entropy decay ..... 125
6.1.5 Holographic transformation ..... 126
6.2 The grand model and a generalised Grimmett-Janson coupling ..... 127
6.2.1 The grand model ..... 127
6.2.2 Coupling via holographic transformation ..... 129
6.3 Variance decay of Glauber dynamics on the grand model ..... 130
6.3.1 Construction of the canonical path ..... 131
6.3.2 Total congestion and rapid mixing ..... 134
6.4 Entropy decay of Glauber dynamics on the grand model ..... 135
6.5 Rapid mixing of Glauber dynamics on the random cluster model ..... 138
6.5.1 Comparing the decay rates of down walks ..... 141
6.5.2 Faster mixing via perturbed chains ..... 142
6.6 Rapid mixing of Swendsen-Wang dynamics ..... 145
6.6.1 FKES distribution and single-bond dynamics ..... 147
6.6.2 Comparing Glauber dynamics to single-bond dynamics ..... 151
6.7 Perfect sampling via coupling from the past ..... 155
6.7.1 Perfect ferromagnetic Ising sampler ..... 156
6.7.2 CFTP for weighted random cluster models ..... 156
6.7.3 Proof of monotonicity ..... 159
6.8 Remaining proofs of this chapter ..... 160
6.8.1 Proof of the equivalence result ..... 160
6.8.2 Proof of the adjointness ..... 162
6.8.3 Proof of analytic lemmata ..... 163
7 Conclusion and open problems ..... 167
7.1 Towards sampling local lemma ..... 167
7.2 Parity issue ..... 170
7.3 Random cluster model ..... 170
7.4 Fine-grained complexity ..... 173
A A proof of \#P-hardness ..... 175
Bibliography ..... 179

## Chapter 1

## Introduction

Background. The constraint satisfaction problem (CSP) is probably one of the most important subjects to study in the theory of computing. In fact, many problems can be cast as CSPs, e.g., Boolean satisfiability problems (SATs), proper colourings of graphs and hypergraphs, and independent sets, to name a few. In general, deciding if a CSP instance can be satisfied or not is NP-hard. However, efficient algorithms become possible when the number of appearances of each variable (usually referred to as the degree) is not too high. For these instances, the Lovász local lemma [EL75] provides a fundamental criterion to guarantee the existence of a solution. Although the original local lemma does not provide an efficient algorithm, after two decades of effort [Bec91, Alo91, MR98, CS00, Sri08, Mos09], the celebrated work of Moser and Tardos [MT10] provides an efficient algorithm matching the same conditions as the Lovász local lemma. One remarkable aspect of this algorithm in the case of the boundeddegree $k$-SAT problem is that it gives the location of the algorithmic threshold for finding solutions as the degree varies [KST93, GST16], up to lower-order terms. In other words, the bounded-degree $k$-SAT problem exhibits a computational phase transition as the degree varies whose threshold is captured by the Lovász local lemma.

Two related computational problems that have been intensively studied recently is to efficiently sample (almost uniformly) from the solution space, and approximately count the number of solutions. These two problems are closely related to each other, and even computationally equivalent for many computational tasks of interest. It is tempting to think about using the Moser-Tardos algorithm for the sampling problem too, but unfortunately, its output distribution is far from being uniform, and does not suit the need of either counting or sampling tasks. Such deficiency is fundamental, as sampling can be computationally harder than searching in the local lemma
regime. For example, for the $k$-SAT problem where each variable appears in at most $\Delta$ clauses ${ }^{1}$, if $\Delta \leq 2^{k} /(\mathrm{e} k)$, then there must be a satisfying assignment by applying Lovász local lemma, and it can be efficiently found. Yet if $\Delta \geq 5 \cdot 2^{k / 2}$, there is no algorithm to sample or approximately count satisfying assignments unless $\mathbf{N P}=\mathbf{R P}$ [ $\mathrm{BGG}^{+}$19], even when no literal contains any negation (also known as the monotone $k$-SAT problem, or (weak) hypergraph independent sets). ${ }^{2}$ This leaves open the problem whether there is a local-lemma-type threshold for the computational phase transition of the sampling and approximate counting problem.

The main focus of this thesis is a prototypical problem which was also the original setting where the local lemma was developed, the hypergraph colouring problem. We adopt the standard definitions for hypergraphs; refer to Section 2.1 for details. A (proper) $q$-colouring assigns each vertex with a colour out of $q$ possible choices such that no hyperedge is monochromatic. The hypergraph colouring problem is defined formally as follows.

## Name $\operatorname{HypergraphColouring}(q, k, \Delta)$

Instance A $k$-uniform hypergraph $H$ with maximum degree at most $\Delta$.
Output A $q$-colouring of $H$, or $\perp$ if there is no such colouring.

## Name \#HypergraphColouring $(q, k, \Delta)$

Instance A $k$-uniform hypergraph $H$ with maximum degree at most $\Delta$.
Output The number of $q$-colourings of $H$.
In pursuit of efficient algorithms, a common idea is to incorporate the Markov chain Monte Carlo (MCMC) method. This approach works if (1) the stationary distribution of the Markov chain is correct, (2) the implementation of the chain is efficient, and (3) the chain converges fast (i.e., having a low mixing time). It is proved to be successful for uniformly sampling hypergraph independent sets when $\Delta \leq c 2^{k / 2}$ [HSZ19] by using the most straightforward Markov chain called the Glauber dynamics. Note that this matches the NP-hardness result $\left[\mathrm{BGG}^{+} 19\right]$ up to constant factors, and hence we consider it as a sharp computational phase transition for hypergraph

[^0]independent sets. Indeed, the vanilla Glauber dynamics also succeeds sampling hypergraph colourings when $\Delta<q-1$ and $k>4$ [BDK08]. However, unlike hypergraph independent sets, one starts to encounter the so-called disconnectivity barrier as the degree is going beyond this condition. It is well-known that the state space of the vanilla Markov chain is no longer connected even when $\Delta=c q$ for some constant $c$ [FM11], let alone getting into the regime of the local lemma where $\Delta \sim\left(q^{k}\right)^{1 / C}$ for some constant $C \geq 1$ as $q$ and $k$ grows.

This connectivity barrier has been bypassed recently by some exciting developments. These include, the partial rejection sampling method [GJL19], the LP-based marking/unmarking paradigm [Moi19, GLLZ19, GGGY21, JPV21b], the Markov chain projection approach [FGYZ21a, FHY21, JPV21a, HSW21], and the lazy sampler [HWY22, HWY23b, HWY23a]. Specialised to hypergraph colourings, the up-to-date regime one can achieve is roughly $\Delta \lesssim q^{k / 3}$ where the symbol $\lesssim$ hides lower order factors. We list below these algorithms. Note that the threshold for Lovász local lemma is $\Delta \lesssim q^{k}$.

| Reference | Algorithm | Bound | Method |
| :---: | :---: | :---: | :---: |
| [GLLZ19] | FPTAS | $\Delta \lesssim q^{k / 14}$ | Linear programming |
| [FHY21] | FPRAS | $\Delta \lesssim q^{k / 9}$ | Projection |
| [JPV21b] | FPTAS | $\Delta \lesssim q^{k / 7}$ | Linear programming |
| [HWY23a] | FPTAS | $\Delta \lesssim q^{k / 5}$ | Lazy sampler |
| [JPV21a] | FPRAS | $\Delta \lesssim q^{k / 3}$ | Projection |
| [HSW21] | Perfect sampler | $\Delta \lesssim q^{k / 3}$ | Projection |
| [FGW ${ }^{+}$23a] | FPTAS | $\Delta \lesssim q^{k / 3}$ | History backtracking |

Table 1.1: Algorithms for hypergraph colourings

The terms FPRAS and FPTAS in the above table are defined in Section 2.5, though we avoid relying on these definitions when stating our main theorems. In a nutshell, both FPRAS and FPTAS approximate the count in fully polynomial time, and an FPRAS allows randomness while an FPTAS is a deterministic algorithm.

On the other hand, before the recent wave of local lemma inspired sampling algorithms, randomly sampling colourings in linear $k$-uniform hypergraphs has already been studied [FM11, FA17]. A hypergraph is called linear, if any two hyperedges intersect in at most one vertex. Linear hypergraphs are also known as simple hypergraphs based on the context. In particular, Frieze and Anastos [FA17] gave an efficient sam-
pling algorithm when the number of colours satisfies $q \geq \max \left\{C_{k} \log n, 500 k^{3} \Delta^{\frac{1}{k-1}}\right\}$, where $n$ is the number of vertices and $C_{k}$ depends only on $k$. Their algorithm is the standard Glauber dynamics with a random initial (not necessarily proper) colouring. The logarithmic lower bound on the number of colours is crucial to their analysis, as it guarantees that there is a giant connected component in the state space so that connectivity is not an issue.

However, and somewhat surprisingly, it was unclear about the computational hardness of hypergraph colourings in the local lemma settings prior to this work, even for the searching problem. The computational phase transition for the hypergraph colouring problem is therefore open, unlike the case for the hypergraph independent set problem.

Results of this thesis. This thesis strives to narrow the gap for the counting hypergraph colouring problem. The main contribution is therefore two fold.

From the complexity side, we first show that it is NP-hard to find a proper hypergraph colouring if $q \geq 2, k \geq 2$ (but not $q=k=2$ ), and $\Delta \gtrsim k q^{k}$ (see Theorem 4.4), and to approximately count if $q \geq 2, k \geq 4$, and $\Delta \gtrsim k q^{k-1}$ (see Theorem 4.6). These bounds almost match the Lovász local lemma threshold. When restricting to linear hypergraphs, these two bounds still hold. We remark that the aforementioned algorithmic result for counting hypergraph independent sets [HSZ19] further improves to $\Delta \lesssim \frac{2^{k}}{k^{2}}$ when restricted to linear hypergraphs. This upper bound on $\Delta$ almost matches that for the searching algorithm [MT10]. In view of this, it seems reasonable to conjecture that, for linear hypergraphs, such a match also occurs in the colouring problem, and that the sharp hardness threshold (for both approximate counting and searching) is $\Delta \gtrsim q^{k-1}$, up to some polynomial factors in $k$. Our hardness result almost matches it.

Our second and main contribution for hardness is to obtain a far more refined bound for the counting problem that goes well beyond the hardness of finding a colouring and which we conjecture is asymptotically tight (up to constant factors). We show in particular that for all even $q \geq 4$ it is NP-hard to approximate the number of colourings when $\Delta \gtrsim q^{k / 2}$ (see Theorem 4.1). The exponent here is in line with that of the hardness threshold for counting hypergraph independent sets, and confirms that the "sampling-is-computationally-harder" phenomenon manifests into local-lemma-type hypergraph problems with non-Boolean domain and which are not necessarily monotone.

From the algorithmic side, we provide an algorithm with a better regime on linear
hypergraphs. Roughly speaking, the regime we obtain is $\Delta \lesssim q^{k /(2+\delta)}$ for any fixed $\delta>$ 0 and $k \geq k(\delta)$. (see Theorem 5.1 for the formal parameterisation). This improves the bound of [JPV21a, HSW21] for general hypergraphs, and does not require unbounded number of colours, unlike in [FM11, FA17]. For the sampling task, the running time of our algorithm can be made arbitrarily close to linear in the number of vertices. This, with a standard counting to sampling reduction, induces an FPRAS for the count running in quadratic time. Moreover, our algorithm can be modified into a perfect sampler by applying the bounding chain method [Hub98] based on coupling from the past (CFTP) [PW96a], following the same lines of [HSW21].

The exponent (roughly $k / 2$ ) is unlikely to be tight, although it appears to be the limit of current techniques. In fact, we conjecture that the computational transition for sampling $q$-colourings in linear hypergraphs happens around the same threshold of the local lemma (namely, the exponent should be roughly $k / 1$ ). This conjecture is supported by the aforementioned hardness result for general $q$ on linear hypergraphs ${ }^{3}$, and by the algorithm of Frieze and Anastos [FA17] for $q=\Omega(\log n)$. Note that for a linear $k$-uniform hypergraph with maximum degree $\Delta$, Frieze and Mubayi [FM13] showed that the chromatic number $\chi(H) \leq C_{k}\left(\frac{\Delta}{\log \Delta}\right)^{\frac{1}{k-1}}$ where $C_{k}$ depends only on $k$. Their bound is asymptotically better than the bound given by the local lemma. Thus, there may still be a gap between the searching threshold and the sampling threshold.

Other counting problems. This thesis also includes some related counting problems:

- As a warmup, we provide an algorithm that approximates the total variation distance between two product distributions in polynomial time after the section of preliminaries, while the exact computation is $\mathbf{\# P}$-complete $\left[\mathrm{BGM}^{+} 23\right]$.
- At the end of this thesis, we consider the sampling problem for the ferromagnetic Ising model with consistent external fields. It can be unified with the aforementioned hypergraph colouring problem under the paradigm of graphical models, and similarly, exhibits some certain phase transitions. In particular, the edge-flipping dynamics on a related random cluster model, and the Swendsen-Wang dynamics [SW87] on this Ising model, are studied. Under the

[^1]condition that the maximum degree of the input graph is bounded and external fields are non-trivial, we show near-linear time mixing of both dynamics.

Detailed introductions of both are provided at the beginning of the two chapters respectively.

Additional results beyond this thesis. This thesis collects part of the author's work during the study of the degree in pursuit. Aside from the aforementioned work, the following additional results are also shown, details of which are not included in this thesis.

First, the computational hardness of approximate counting and sampling in the local lemma regime can be refined with the parameterisation of the size of overlaps. As one such example, recall that approximately counting hypergraph independent sets is NP-hard when $\Delta \geq 5 \cdot 2^{k / 2}\left[\mathrm{BGG}^{+} 19\right]$. If we further take into account the size of the overlap between any pair of hyperedges, and restrict such size being no more than $b$ where $1 \leq b \leq k / 2$, then the problem is NP-hard if $\Delta \geq 5 \cdot 2^{k-b}$ [QW23]. This not only recovers the hardness result without the overlap restriction by taking $b=k / 2$, but also implies the hardness for linear hypergraphs for $\Delta \geq 2.5 \cdot 2^{k}$. As a consequence, it confirms that the regimes on the maximum degree where the state-of-the-art algorithms work are tight, up to some small factors. These algorithms include the FPRAS requiring $\Delta \leq c 2^{k} / k^{2}$ [HSZ19], the perfect sampler requiring $\Delta \leq c 2^{k} / k^{2}$ [QWZ22], and the FPTAS requiring $\Delta \lesssim 2^{(1-o(1)) k}\left[\mathrm{FGW}^{+} 23 \mathrm{a}\right]$.

Second, in nearly the same regime where the FPRAS [JPV21a] and perfect sampler [HSW21] work for hypergraph colourings, a very recent paper [FGW ${ }^{+}$23a] shows how one can derandomise the Markov chain Monte Carlo algorithm and obtain an FPTAS. This is achieved by backtracking the running history of the Markov chain in a lazy way. Such an argument also applies to counting hypergraph independent sets and acquires a nearly tight regime $\Delta \lesssim 2^{(1-o(1)) k}$.

Organisation. Each chapter begins with a separate introduction, with a detailed elaboration of the main result and a brief overview of the techniques.

In Chapter 2, the necessary definitions and preliminary results are provided. In Chapter 3, we look into the problem of approximating the total variation distance between two product distributions, serving as a warm-up. For completeness, a known hardness result in this chapter is provided in Appendix A. This part features a published article [FGJW23]. In Chapter 4, we kick off the technical part of this thesis and
prove the hardness of the hypergraph colouring problem. This part features a published article [GGW22]. In Chapter 5, we move to the other side of the hypergraph colouring counting problem, and give an algorithm that counts linear hypergraph colourings. This part features a published article [FGW22]. In Chapter 6, the ferromagnetic Ising model is studied. This part features a published article [FGW23b]. Finally, we conclude this thesis by providing several open problems in Chapter 7.

## Chapter 2

## Preliminaries

This section introduces some fundamental concepts and definitions that are essential for the subsequent analysis.

## Nomenclature

Sets. We sometimes do not distinguish the element $u$ and the singleton set $\{u\}$ in sub- or sup-scripts. Given a set $S$, the notation $S+u$ is a shorthand for $S \cup\{u\}$, and $S-u$ for $S \backslash\{u\}$. Note that $u$ does not necessarily belong to $S$.

Integers. We take the convention that natural numbers start with 0 , i.e., $\mathbb{N}:=\{0,1, \cdots\}$. We also use $[n]$ as a shorthand for $\{1,2, \cdots, n\}$, and $[n]$ for $[n] \cup\{0\}$.

Logarithms. Denote by e $\approx 2.718$ the base of the natural logarithm. Throughout, $\log$ and $\ln$ stands for the logarithm in base 2 and e respectively.

Indicator function. Denote by $\mathbb{I}[A]$ the indicator function that takes value 1 when the condition $A$ holds, and 0 otherwise.

### 2.1 Graph theory

Without specification, all graphs are undirected. We adopt all the standard definitions for graphs, which, specifically, include:

- $G[A]$ : the induced subgraph of $G$ on the vertex subset $A \subseteq V$.
- $\operatorname{dist}_{G}(A, B)$ : the distance between two vertex sets $A \subseteq V$ and $B \subseteq V$ on $G$, which is defined by $\operatorname{dist}_{G}(A, B):=\min _{u \in A, v \in B} \operatorname{dist}_{G}(u, v)$ where $\operatorname{dist}_{G}(u, v)$ is the length of the shortest path between $u$ and $v$ in $G$.
- $\Gamma_{G}^{i}(A)$ : the set of vertices $u$ such that $\operatorname{dist}_{G}(A, u)=i$. Specifically, when $i=1$, this notation represents the neighbourhood of the given set $A \subseteq V$, and is also denoted by $\Gamma_{G}(A)$.

For the sake of convenience, we may drop the subscript $G$ when the underlying graph is clear from the context.

Definition 2.1 (Graph power). The $i$-th power of $G$, denoted by $G^{i}$, is another graph that has the same vertex set as $G$, and $\{u, v\}$ is an edge in $G^{i}$ if and only if $1 \leq$ $\operatorname{dist}_{G}(u, v) \leq i$.

A hypergraph $H=(V, \mathcal{E})$ consists of a set of vertices $V$ and a set of hyperedges $\mathcal{E} \subseteq 2^{V}$. It is said to be $k$-uniform, if every hyperedge $e \in \mathcal{E}$ is of size $k$. The degree of a vertex is the number of hyperedges that it appears, and the degree of the hypergraph is the maximum one of all the vertices. The hypergraph is said to have overlap $b$, if for any distinct pair of hyperedges $e \neq e^{\prime}$, they share at most $b$ vertices, i.e., $\left|e \cap e^{\prime}\right| \leq b$. Specifically, the hypergraph is called linear, if it has overlap 1.

Definition 2.2 (Line graph). Let $H=(V, \mathcal{E})$ be a hypergraph. Its line $\operatorname{graph} \operatorname{Lin}(H):=$ $\left(V_{L}, E_{L}\right)$ is given by $V_{L}=\mathcal{E}$, and $\left\{e, e^{\prime}\right\} \in E_{L}$ iff $e \cap e^{\prime} \neq \varnothing$.

## $2.2 f$-divergences

This section is used by Chapter 3 and Chapter 6.
Let $\Omega$ be a finite state space.
A widely-used quantity for measuring the difference between two distributions is the $f$-divergence. Let $f: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ be a convex function satisfying $f(1)=0$. Let $v$ and $\mu$ be two distributions over $\Omega$. The $f$-divergence between $v$ and $\mu$ is defined by

$$
D_{f}(v \| \mu):=\mathbf{E}_{X \sim \mu}\left[f\left(\frac{v(X)}{\mu(X)}\right)\right] .
$$

We remark that it is not necessarily symmetric over the change of roles between $\mu$ and $v$. Three important $f$-divergences are used in this thesis: the total variation distance, the $\chi^{2}$-divergence, and the Kullback-Leibler divergence (KL divergence, also known as the relative entropy).

Total variation distance. The total variation distance, denoted by $d_{\mathrm{TV}}(\cdot, \cdot)$, is the $f$-divergence with $f(x)=\frac{1}{2}|x-1|$. Alternatively,

$$
d_{\mathrm{TV}}(\mu, v)=d_{\mathrm{TV}}(v, \mu):=\frac{1}{2} \sum_{\omega \in \Omega}|\mu(\omega)-v(\omega)|=\sum_{\omega \in \Omega} \max \{0, \mu(\omega)-v(\omega)\}
$$

We then collect some important properties of the total variation distance. The random variable $(X, Y) \in \Omega \times \Omega$ is called a coupling between $\mu$ and $v$ if the marginal distributions satisfy $X \sim \mu$ and $Y \sim v$. From this point of view, the total variation distance not only characterises the difference between two distributions, but also captures the minimum discrepancy of any coupling. This is summarised by the following lemma, usually referred to as the coupling lemma; see e.g., [LP17, Proposition 4.7]

Lemma 2.3 (Coupling lemma). For any coupling $(X, Y)$ between $\mu$ and $v$,

$$
d_{\mathrm{TV}}(\mu, v) \leq \operatorname{Pr}[X \neq Y] .
$$

Moreover, there exists an optimal coupling under which the equality holds.
Optimal couplings are not necessarily unique. However, for any optimal coupling $O$, it holds that

$$
\begin{equation*}
\forall \omega \in \Omega, \quad \operatorname{Pr}_{O}[X=Y=\omega]=\min \{\mu(\omega), \nu(\omega)\} . \tag{2.1}
\end{equation*}
$$

The above equation holds because (1) for any valid coupling $C$, it holds that $\operatorname{Pr}_{C}[X=$ $Y=\omega] \leq \min \{\mu(\omega), \nu(\omega)\} ;$ (2) to achieve the optimal coupling, every $\omega$ must achieve the equality. We then have

$$
\begin{equation*}
\operatorname{Pr}_{O}[X=\omega \wedge Y \neq X]=\operatorname{Pr}_{O}[X=\omega]-\operatorname{Pr}_{O}[X=Y=\omega]=\max \{0, \mu(\omega)-v(\omega)\} \tag{2.2}
\end{equation*}
$$

$\chi^{2}$-divergence. The $\chi^{2}$ divergence, denoted by $D_{\chi^{2}}(\cdot \| \cdot)$, is the $f$-divergence with $f(x)=x^{2}-1$. Equivalently,

$$
D_{\chi^{2}}(v \| \mu):=\sum_{\omega \in \Omega} \frac{v^{2}(\omega)}{\mu(\omega)}-1 .
$$

A standard inequality states

$$
\begin{equation*}
d_{\mathrm{TV}}(v, \mu) \leq \sqrt{D_{\chi^{2}}(v \| \mu)} . \tag{2.3}
\end{equation*}
$$

The $\chi^{2}$-divergence can be alternatively defined from a functional viewpoint. For any function $g: \Omega \rightarrow \mathbb{R}_{\geq 0}$, its relative variance over $\mu$ is given by

$$
\operatorname{Var}_{\mu}(g)=\mathbf{E}_{\mu}\left[g^{2}\right]-\mathbf{E}_{\mu}^{2}[g]=\sum_{\omega \in \Omega} \mu(\omega) g^{2}(\omega)-\left(\sum_{x \in \Omega} \mu(\omega) g(\omega)\right)^{2}
$$

Clearly, if we take $g(\omega)=\frac{v(\omega)}{\mu(\omega)}$, then $\operatorname{Var}_{\mu}(g)=D_{\chi^{2}}(v \| \mu)$.

KL divergence. The Kullback-Leibler divergence, denoted by $D_{\mathrm{KL}}(\cdot \| \cdot)$, is the $f$ divergence with $f(x)=x \log x$. Equivalently,

$$
D_{\mathrm{KL}}(v \| \mu):=\sum_{x \in \Omega} v(x) \log \left(\frac{v(x)}{\mu(x)}\right) .
$$

The following Pinkser's inequality is well known.

$$
\begin{equation*}
d_{\mathrm{TV}}(v, \mu) \leq \sqrt{\frac{D_{\mathrm{KL}}(v \| \mu)}{2}} \tag{2.4}
\end{equation*}
$$

Similarly to the $\chi^{2}$-divergence, we also have a functional interpretation of the KL divergence. For any function $g: \Omega \rightarrow \mathbb{R}_{\geq 0}$, its relative entropy over $\mu$ is given by

$$
\begin{aligned}
\operatorname{Ent}_{\mu}(g) & :=\mathbf{E}_{\mu}[g \log g]-\mathbf{E}_{\mu}[g] \log \mathbf{E}_{\mu}[g] \\
& =\sum_{\omega \in \Omega} \mu(\omega) g(\omega) \log g(\omega)-\left(\sum_{\omega \in \Omega} \mu(\omega) g(\omega)\right) \log \left(\sum_{x \in \Omega} \mu(\omega) g(\omega)\right),
\end{aligned}
$$

where the convention is that $0 \log 0=0$. Clearly, if we take $g(x)=\frac{\nu(x)}{\mu(x)}$, then $\operatorname{Ent}_{\mu}(g)=$ $D_{\mathrm{KL}}(v \| \mu)$.

Data processing inequality. Finally, the following data processing inequality is useful if we are to apply the above functional interpretation to the analysis of Markov chains. Let $P$ be a stochastic matrix over the state space $\Omega$. For any $f$-divergence and pair of distributions $v, \mu$ over $\Omega$, it holds that

$$
D_{f}(v P \| \mu P) \leq D_{f}(v \| \mu)
$$

### 2.3 Markov chain and mixing time

This section is used in Chapter 5 and Chapter 6.
Let $\Omega$ be a finite state space, and $P$ be a $|\Omega| \times|\Omega|$ stochastic matrix, namely a non-negative matrix such that the sum of each row is 1 . We call a sequence of random variables $\left(X_{t}\right)_{t \geq 0}$ a Markov chain with state space $\Omega$ and transition matrix $P$, if $\operatorname{Pr}\left[X_{t+1}=y \mid X_{t}=x\right]=P(x, y)$. We say $P$ is

- irreducible if for any $x, y \in \Omega$, there exists $t>0$ such that $P^{t}(x, y)>0$;
- aperiodic if $\operatorname{gcd}\left\{t \mid P^{t}(x, x)>0\right\}=1$ for all $x \in \Omega$;
- reversible with respect to $\mu$ if the following detailed balance equation holds

$$
\forall x, y \in \Omega, \quad \mu(x) P(x, y)=\mu(y) P(y, x)
$$

A distribution $\mu$ is a stationary distribution of $P$ if $\mu P=\mu$. If $P$ is reversible with respect to $\mu$, then $\mu$ is a stationary distribution of $P$. If $P$ is both irreducible and aperiodic, then $P$ has a unique stationary distribution. The mixing time of $P$ is defined by

$$
\forall \epsilon>0, \quad t_{\mathrm{mix}}(P, \epsilon):=\max _{x \in \Omega} \min \left\{t \mid d_{\mathrm{TV}}\left(P^{t}(x, \cdot), \mu\right) \leq \epsilon\right\}
$$

The joint process $\left(X_{t}, Y_{t}\right)_{t \geq 0}$ is a coupling of Markov chain $\mathbf{P}$ if $\left(X_{t}\right)_{t \geq 0}$ and $\left(Y_{t}\right)_{t \geq 0}$ individually follow the transition rule of $\mathbf{P}$, and if $X_{i}=Y_{i}$ then $X_{j}=Y_{j}$ for all $j \geq i$. By the coupling lemma, for any coupling $\left(X_{t}, Y_{t}\right)_{t \geq 0}$ of $\mathbf{P}$, it holds that

$$
d_{\mathrm{TV}}\left(P^{t}\left(X_{0}, \cdot\right), P^{t}\left(Y_{0}, \cdot\right)\right) \leq \operatorname{Pr}\left[X_{t} \neq Y_{t}\right] .
$$

Hence, the mixing time of $\mathbf{P}$ can be bounded by

$$
\begin{equation*}
t_{\text {mix }}(\boldsymbol{P}, \epsilon) \leq \max _{X_{0}, Y_{0} \in \Omega} \min \left\{t \mid \operatorname{Pr}\left[X_{t} \neq Y_{t}\right] \leq \epsilon\right\} . \tag{2.5}
\end{equation*}
$$

### 2.4 The local lemma

Let $\mathcal{R}=\left\{R_{1}, \cdots, R_{n}\right\}$ be a set of mutually independent random variables. Given an event $A$, denote the set of variables that determines $A$ by $\operatorname{vbl}(A) \subseteq \mathcal{R}$. Let $\mathcal{B}=$ $\left\{B_{1}, \cdots, B_{n}\right\}$ be a collection of "bad" events. For any event $A$ (not necessarily in $\mathcal{B}$ ), let $\Gamma(A):=\{B \in \mathcal{B} \mid B \neq A, \operatorname{vbl}(B) \cap \operatorname{vbl}(A) \neq \varnothing\}$. We will use the following version of Lovász Local Lemma from [HSS11].

Theorem 2.4 ([EL75, HSS11]). If there exists a function $x: \mathcal{B} \rightarrow(0,1)$ such that for any bad event $B \in \mathcal{B}$,

$$
\begin{equation*}
\operatorname{Pr}[B] \leq x(B) \prod_{B^{\prime} \in \Gamma(B)}\left(1-x\left(B^{\prime}\right)\right), \tag{2.6}
\end{equation*}
$$

then it holds that

$$
\operatorname{Pr}\left[\bigwedge_{B \in \mathcal{B}} \bar{B}\right] \geq \prod_{B \in \mathcal{B}}(1-x(B))>0 .
$$

Moreover, for any event $A$,

$$
\begin{equation*}
\operatorname{Pr}\left[A \mid \bigwedge_{B \in \mathcal{B}} \bar{B}\right] \leq \operatorname{Pr}[A] \prod_{B \in \Gamma(A)}(1-x(B))^{-1} . \tag{2.7}
\end{equation*}
$$

### 2.5 Approximate counting and sampling algorithms

Formal definitions of the algorithms that suit our need are given in this section. For the approximate counting tasks, the desired algorithms are called fully polynomialtime randomised approximation schemes (FPRAS), formally defined as below.

Definition 2.5 (FPRAS). An algorithm is called a randomised approximation scheme (RAS) for a function $Z: \Sigma^{*} \rightarrow \mathbb{N}$, if, taking an input pair $x \in \Sigma^{*}, \varepsilon \in(0,1)$, the random output $\hat{Z}$ satisfies $\operatorname{Pr}[(1-\varepsilon) Z \leq \hat{Z} \leq(1+\varepsilon) Z] \geq 2 / 3$. It is further said to be fully polynomial-time (FPRAS), if it runs in time poly $(|x|, 1 / \varepsilon)$.

Remark 2.6 (FPTAS). The probability in the above definition only depends on the randomness of the algorithm; in other words, the error bound should be guaranteed for arbitrary valid inputs. Due to a standard augmentation argument, the success probability $2 / 3$ can be replaced by $1-\delta$ where $\delta \in(0,1)$, while scaling the running time by a factor of $\log 1 / \delta$. Furthermore, if the success probability is replaced by 1 and no randomness is allowed in the algorithm, then such an algorithm is called a fully polynomial-time deterministic approximation scheme (FPTAS).

Defined below are the samplers of interest in this thesis.

Definition 2.7 (FPAUS). An algorithm is called a fully polynomial-time almost uniform sampler (FPAUS) for a distribution $\mu=\mu(x)$ specified by an instance $x$, if the output distribution $\mu^{\prime}$ satisfies $d_{\mathrm{TV}}\left(\mu, \mu^{\prime}\right) \leq \delta$ in time poly $(|x|, \log 1 / \delta)$.

Remark 2.8 (Perfect sampler). A sampler is called perfect, or exact, if we further restrict the output distribution $\mu^{\prime}$ to satisfy $d_{\mathrm{TV}}\left(\mu, \mu^{\prime}\right)=0$, but allow the algorithm to be Las Vegas. That is, the running time could be unbounded, but in expectation it is bounded by poly $(|x|)$.

For many interesting problems, standard reductions [JVV86, ŠVV09] between approximate counting and almost uniform sampling are well-known. We brief explain how the counting to sampling reduction is done for hypergraph colourings as an example. A detailed argument could be found in $\left[\mathrm{FGW}^{+} 23 \mathrm{a}\right.$, Section 2.3] for instance.

Let $H=(V, \mathcal{E})$ be a hypergraph with $m$ edges. Suppose $\mathcal{E}=\left\{e_{1}, \cdots, e_{m}\right\}$. An edge decomposition of $H$ is a sequence of hypergraphs $H_{0}, H_{1}, \cdots H_{m}$, where $H_{i}=$ $\left(V,\left\{e_{1}, \cdots, e_{m}\right\}\right)$. In other words, starting from $H=H_{m}$, we gradually remove a hyperedge until the hypergraph consists solely indepedent vertices. Let $Z_{i}$ be the
number of hypergraph colourings in $H_{i}$, and $\mu_{i}$ be the uniform distribution over all colourings of $H_{i}$. Then we decompose $Z=Z_{m}$ as a telescoping product

$$
Z=Z_{0} \prod_{i=1}^{m} \frac{Z_{i}}{Z_{i-1}}
$$

Let $\mu_{i-1, e_{i}}$ be the distribution $\mu_{i-1}$ projected on the hyperedge $e_{i}$. Then each term in the above product is exactly

$$
\frac{Z_{i}}{Z_{i-1}}=\operatorname{Pr}_{\omega \sim \mu_{i-1, e_{i}}}[\omega \text { is not monochromatic }] .
$$

Therefore, it suffices to approximate the above probability. In the local lemma regime that we consider, the above probability has a constant lower bound, which means we do not need a lot of sample drawing to obtain a good approximation.

The following diagram characterises the connection between these algorithms.


These terms might appear in brief discussion, but when formally stating main results, we avoid using them.

## Chapter 3

## Computing the total variation <br> distance

Let $\Omega=[q]^{n}$ be the state space, where $[q]=\{1, \ldots, q\}$ is a finite set, and $P_{1}, \cdots, P_{n}$, $Q_{1}, \cdots, Q_{n}$ be independent distributions over [q]. Let $P=P_{1} \otimes P_{2} \otimes \cdots \otimes P_{n}$ and $Q=Q_{1} \otimes Q_{2} \otimes \cdots \otimes Q_{n}$ be two product distributions. Namely, for any $\omega_{1} \omega_{2} \cdots \omega_{n}$, it holds that

$$
P\left(\omega_{1} \omega_{2} \cdots \omega_{n}\right)=P_{1}\left(\omega_{1}\right) P_{2}\left(\omega_{2}\right) \cdots P_{n}\left(\omega_{n}\right),
$$

and analogously for $Q$.
In this chapter, we are interested in computing the total variation distance (recall Section 2.2) between two product distributions. This might appear as a simple problem owing to the mutual independency of all the distributions $P_{i}$ and $Q_{i}$. In fact, for many other quantities for similar uses, such as the relative entropy and the $\chi^{2}$-divergence, simple tensorisation formulæ do exist:

$$
\begin{aligned}
D_{\mathrm{KL}}(P \| Q) & =\sum_{i=1}^{n} D_{\mathrm{KL}}\left(P_{i} \| Q_{i}\right), \text { and } \\
1+D_{\chi^{2}}(P \| Q) & =\prod_{i=1}^{n}\left(1+D_{\chi^{2}}\left(P_{i} \| Q_{i}\right)\right) .
\end{aligned}
$$

This means computing these quantities for product distributions are easy if each distribution is described explicitly in the input. Unfortunately, the total variation distance does not tensorise over product distributions, and it was discovered recently that, somewhat surprisingly, exact computation of the total variation distance, even between product distributions over the Boolean domain, is \#P-hard [ $\left.\mathrm{BGM}^{+} 23\right] .{ }^{1}$

[^2]This leaves open the question of approximation complexity of the total variation distance. In $\left[\mathrm{BGM}^{+} 23\right]$, the authors give polynomial-time randomised approximation algorithms in two special cases over the Boolean domain, when one of the distribution has marginals over $1 / 2$ and dominates the other, or when one of the distribution has a constant number of distinct marginals. Their method is based on Dyer's dynamic programming algorithm for approximating the number of knapsack solutions [Dye03].

Our contribution is a simple polynomial-time approximation algorithm for the total variation distance between two product distributions. Our algorithm is based on the Monte Carlo method and does not have further restrictions.

Theorem 3.1. Let $[q]=\{1,2, \ldots, q\}$ be a finite set. There exists an algorithm such that given two product distributions $P, Q$ over $[q]^{n}$ and parameters $\epsilon>0$ and $0<\delta<1$, it outputs a random value $\widehat{d}$ in time $O_{q}\left(\frac{n^{2}}{\epsilon^{2}} \log \frac{1}{\delta}\right)$ such that $(1-\epsilon) d_{\mathrm{TV}}(P, Q) \leq \widehat{d} \leq$ $(1+\epsilon) d_{\mathrm{TV}}(P, Q)$ holds with probability at least $1-\delta$.

Our algorithm can also handle the case where each coordinate has a different domain size without any change. In Theorem 3.1, the input product distributions are given by the marginal probability for each coordinate and each $c \in[q]$ in binary. The stated running time assumes that all arithmetic operations can be done in $O(1)$ time. ${ }^{2}$

To approximate the total distance, the naïve Monte Carlo algorithm works well when the two distributions are sufficiently far away. However, when the total distance is exponentially small, naïve Monte Carlo may require exponentially many samples to return an accurate estimate. Our idea is to consider a distribution that can be efficiently sampled from and yet boosts the probability that the two distributions are different. Ideally, we would want to use the optimal coupling, but that is difficult to compute. We use instead the coordinate-wise greedy coupling where each coordinate is coupled optimally independently. Though such greedy coupling is usually not optimal, ${ }^{3}$ it serves as a proxy in the construction of our estimator. We further condition on the (potentially very unlikely) event that the two samples are different. Normally, conditioning on an unlikely event is a bad move since computational tasks would have become hard. However, here they are still easy thanks to the independence of the coordinates under the coupling. With this conditional distribution, our estimator

[^3]is the ratio between the probabilities of the assignment in the optimal coupling and in the greedy coupling. We show that this estimator is always bounded from above by 1 and its expectation is at least $1 / n$. This means that the standard Monte Carlo method will succeed with high probability using only polynomially many samples.

### 3.1 The algorithm

Let $O$ be an (arbitrary) optimal coupling between $P$ and $Q$. Let $C$ be the coordinatewise greedy coupling. Namely, for each coordinate $i$ and $c \in[q], \operatorname{Pr}_{C}\left[X_{i}=Y_{i}=c\right]=$ $\min \left\{P_{i}(c), Q_{i}(c)\right\}$, and the remaining probability can be assigned arbitrarily as long as $C$ is a valid coupling (but each coordinate is independent). In other words, for each $i \in[n], C$ couples $P_{i}$ and $Q_{i}$ optimally and independently. Note that

$$
\begin{equation*}
\operatorname{Pr}_{C}[X \neq Y]=1-\operatorname{Pr}_{C}[X=Y]=1-\prod_{i=1}^{n}\left(1-d_{\mathrm{TV}}\left(P_{i}, Q_{i}\right)\right) \tag{3.1}
\end{equation*}
$$

can be computed exactly.
Consider the distribution $\pi$ such that

$$
\begin{equation*}
\pi(\omega):=\operatorname{Pr}_{C}[X=\omega \mid X \neq Y] . \tag{3.2}
\end{equation*}
$$

We may assume $P$ and $Q$ are not identical, as otherwise the algorithm just outputs 0 . This makes sure that the distribution $\pi$ is well-defined. The following lemma shows that we can draw random samples from $\pi$ efficiently.

Lemma 3.2. We can sample from the distribution $\pi$ in $O(n)$ time.
Proof. We draw a random sample $\omega \in[q]^{n}$ from $\pi$ index by index. In the $k$-th step, where $1 \leq k \leq n$, we sample $\omega_{k} \in[q]$ from $\pi_{k}\left(\cdot \mid \omega_{1}, \omega_{2}, \ldots, \omega_{k-1}\right)$, which is the marginal distribution on the $k$-th variable conditional on the values of the first $k-1$ variables being $\omega_{1}, \omega_{2}, \ldots, \omega_{k-1}$. By definition,

$$
\pi_{k}\left(\omega_{k} \mid \omega_{1}, \omega_{2}, \ldots, \omega_{k-1}\right)=\frac{\operatorname{Pr}_{X \sim \pi}\left[\forall 1 \leq i \leq k, X_{i}=\omega_{i}\right]}{\operatorname{Pr}_{X \sim \pi}\left[\forall 1 \leq i \leq k-1, X_{i}=\omega_{i}\right]}
$$

As $\omega_{1}, \ldots, \omega_{k-1}$ are sampled from the marginal distribution of $\pi$, the denominator is positive. We show how to compute the numerator next, and the denominator can be computed similarly. By definition

$$
\begin{aligned}
& \operatorname{Pr}_{X \sim \pi}\left[\forall 1 \leq i \leq k, X_{i}=\omega_{i}\right]=\operatorname{Pr}_{(X, Y) \sim \mathcal{C}}\left[\forall 1 \leq i \leq k, X_{i}=\omega_{i} \mid X \neq Y\right] \\
= & \left(1-\operatorname{Pr}_{(X, Y) \sim \mathcal{C}}\left[X=Y \mid \forall 1 \leq i \leq k, X_{i}=\omega_{i}\right]\right) \cdot \frac{\prod_{i=1}^{k} P_{i}\left(\omega_{i}\right)}{1-\prod_{i=1}^{n}\left(1-d_{\mathrm{TV}}\left(P_{i}, Q_{i}\right)\right)}
\end{aligned}
$$

where the last line is due to Bayes' Law.
In the coupling $C$, every pair of $X_{i}$ and $Y_{i}$ is coupled optimally and independently. We have

$$
\begin{align*}
& \operatorname{Pr}_{(X, Y) \sim C}\left[X=Y \mid \forall 1 \leq i \leq k, X_{i}=\omega_{i}\right] \\
= & \prod_{i=1}^{k} \frac{\operatorname{Pr}_{C}\left[X_{i}=Y_{i}=\omega_{i}\right]}{\operatorname{Pr}_{C}\left[X_{i}=\omega_{i}\right]} \prod_{i=k+1}^{n} \operatorname{Pr}_{C}\left[X_{i}=Y_{i}\right] \\
= & \prod_{i=1}^{k} \frac{\min \left\{P_{i}\left(\omega_{i}\right), Q_{i}\left(\omega_{i}\right)\right\}}{P_{i}\left(\omega_{i}\right)} \prod_{i=k+1}^{n}\left(1-d_{\mathrm{TV}}\left(P_{i}, Q_{i}\right)\right) . \tag{3.3}
\end{align*}
$$

where the last line is due to (2.1).
Combining the two equations, we can compute $\operatorname{Pr}_{X \sim \pi}\left[\forall 1 \leq i \leq k, X_{i}=\omega_{i}\right]$, and thus we can compute and sample from $\pi_{k}\left(\cdot \mid \omega_{1}, \omega_{2}, \ldots, \omega_{k-1}\right)$. When sampling from the distribution $\pi$, we pre-process $\prod_{i=k+1}^{n}\left(1-d_{\mathrm{TV}}\left(P_{i}, Q_{i}\right)\right)$ for all $k$, and maintain the prefix products $\prod_{i=1}^{k} \min \left\{P_{i}\left(\omega_{i}\right), Q_{i}\left(\omega_{i}\right)\right\}$ and $\prod_{i=1}^{k} P_{i}\left(\omega_{i}\right)$. This way, each conditional marginal distribution can be computed with $O_{q}(1)$ incremental cost. Hence, the total running time is $O_{q}(n)$, where $O_{q}(\cdot)$ hides a factor linear in $q$.

Let $\omega$ be a random sample from $\pi$. Now consider the following estimator:

$$
\begin{equation*}
f(\omega):=\frac{\operatorname{Pr}_{O}[X=\omega \wedge X \neq Y]}{\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]}=\frac{\max \{0, P(\omega)-Q(\omega)\}}{\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]} \tag{3.4}
\end{equation*}
$$

where the second equality is due to (2.2). This estimator $f$ is well-defined, because when $\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]=0, \pi(\omega)=0$ as well and $\omega$ will not be drawn.

In fact, if $\pi(\omega)=0$, or equivalently $\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]=0$, it must be that $\max \{0, P(\omega)-Q(\omega)\}=0$. This is because $\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]=0$ implies that either $\operatorname{Pr}_{C}[X=\omega]=P(\omega)=0$ or $\operatorname{Pr}_{C}[X \neq Y \mid X=\omega]=0$. In the first case, $\max \{0, P(\omega)-$ $Q(\omega)\}=0$. In the second case $\operatorname{Pr}_{C}[Y=\omega \mid X=\omega]=1$, which implies that $Q(\omega) \geq$ $P(\omega)$, and $\max \{0, P(\omega)-Q(\omega)\}=0$ as well.

Lemma 3.3. For any $\omega \in \Omega$ with $\pi(\omega)>0, f(\omega)$ can be computed in $O(n)$ time.
Proof. Note that

$$
\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]=P(\omega) \operatorname{Pr}_{C}[X \neq Y \mid X=\omega]=P(\omega)\left(1-\operatorname{Pr}_{C}[X=Y \mid X=\omega]\right)
$$

Since $\pi(\omega)>0$, it holds that $P(\omega)>0$. Using (3.3), we have

$$
f(\omega)=\max \left\{0, \frac{1-\frac{Q(\omega)}{P(\omega)}}{\frac{1}{P(\omega)} \operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]}\right\}=\max \left\{0, \frac{1-\prod_{i=1}^{n} \frac{Q_{i}\left(\omega_{i}\right)}{P_{i}\left(\omega_{i}\right)}}{1-\prod_{i=1}^{n} \frac{\min \left\{P_{i}\left(\omega_{i}\right), Q_{i}\left(\omega_{i}\right)\right\}}{P_{i}(\omega)}}\right\}
$$

which can be computed in $O(n)$ time.

Lemma 3.4. We have the following:

$$
\begin{align*}
\mathbf{E}_{\pi} f & =\frac{\operatorname{Pr}_{O}[X \neq Y]}{\operatorname{Pr}_{C}[X \neq Y]}  \tag{3.5}\\
\frac{1}{n} & \leq \mathbf{E}_{\pi} f \leq 1 \tag{3.6}
\end{align*}
$$

Moreover, for any $\omega \in \Omega$ with $\pi(\omega)>0$,

$$
\begin{equation*}
0 \leq f(\omega) \leq 1 \tag{3.7}
\end{equation*}
$$

and it holds that

$$
\begin{equation*}
\mathbf{V a r}_{\pi} f \leq \mathbf{E}_{\pi} f \tag{3.8}
\end{equation*}
$$

Proof. For (3.5), Let $\Omega_{+}=\{\omega \in \Omega \mid \pi(\omega)>0\}$. Then,

$$
\begin{aligned}
\mathbf{E}_{\pi} f & =\sum_{\omega \in \Omega_{+}} \pi(\omega) \times \frac{\operatorname{Pr}_{O}[X=\omega \wedge X \neq Y]}{\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]} \\
& =\sum_{\omega \in \Omega_{+}} \frac{\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]}{\operatorname{Pr}_{C}[X \neq Y]} \times \frac{\operatorname{Pr}_{O}[X=\omega \wedge X \neq Y]}{\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]} \\
& =\frac{\sum_{\omega \in \Omega_{+}} \operatorname{Pr}_{O}[X=\omega \wedge X \neq Y]}{\operatorname{Pr}_{C}[X \neq Y]}=\frac{\operatorname{Pr}_{O}[X \neq Y]}{\operatorname{Pr}_{C}[X \neq Y]},
\end{aligned}
$$

where in the last equation we used the aforementioned fact that $\pi(\omega)=0$ implies $\max \{0, P(\omega)-Q(\omega)\}=0$.

For (3.6), as $O$ is the optimal coupling, $\operatorname{Pr}_{O}[X \neq Y] \leq \operatorname{Pr}_{C}[X \neq Y]$. For the other direction, notice that $O$ projected to coordinate $i$, denoted $O_{i}$, is a coupling between $P_{i}$ and $Q_{i}$. Thus,

$$
\operatorname{Pr}_{O}[X \neq Y] \geq \max _{1 \leq i \leq n} \operatorname{Pr}_{O_{i}}\left[X_{i} \neq Y_{i}\right] \geq \max _{1 \leq i \leq n} d_{\mathrm{TV}}\left(P_{i}, Q_{i}\right)
$$

On the other hand, by the union bound,

$$
\operatorname{Pr}_{C}[X \neq Y] \leq \sum_{i=1}^{n} \operatorname{Pr}_{C_{i}}\left[X_{i} \neq Y_{i}\right]=\sum_{i=1}^{n} d_{\mathrm{TV}}\left(P_{i}, Q_{i}\right) \leq n \max _{1 \leq i \leq n} d_{\mathrm{TV}}\left(P_{i}, Q_{i}\right)
$$

For (3.7), the lower bound is trivial. For the upper bound, we only need to consider $\omega \in \Omega_{+}$such that $P(\omega)>Q(\omega)$. In this case

$$
\begin{aligned}
f(\omega) & =\frac{\max \{0, P(\omega)-Q(\omega)\}}{\operatorname{Pr}_{C}[X=\omega \wedge X \neq Y]}=\frac{P(\omega)-Q(\omega)}{\operatorname{Pr}_{C}[X=\omega] \operatorname{Pr}_{C}[X \neq Y \mid X=\omega]} \\
& =\frac{P(\omega)-Q(\omega)}{P(\omega)\left(1-\operatorname{Pr}_{C}[X=Y \mid X=\omega]\right)}=\frac{1-\frac{Q(\omega)}{P(\omega)}}{1-\operatorname{Pr}_{C}[X=Y \mid X=\omega]}
\end{aligned}
$$

Since $C$ couples each coordinate independently,

$$
\operatorname{Pr}_{C}[X=Y \mid X=\omega]=\prod_{i=1}^{n} \frac{\min \left\{P_{i}\left(\omega_{i}\right), Q_{i}\left(\omega_{i}\right)\right\}}{P_{i}\left(\omega_{i}\right)} \leq \prod_{i=1}^{n} \frac{Q_{i}\left(\omega_{i}\right)}{P_{i}\left(\omega_{i}\right)}=\frac{Q(\omega)}{P(\omega)}
$$

This finishes the proof of (3.7).
For (3.8), since $0 \leq f(\omega) \leq 1$ for all $\Omega \in \Omega_{+}, f(\omega)^{2} \leq f(\omega)$ and thus $\mathbf{E}_{\pi} f^{2} \leq \mathbf{E}_{\pi} f$. We have

$$
\operatorname{Var}_{\pi} f=\mathbf{E}_{\pi} f^{2}-\left(\mathbf{E}_{\pi} f\right)^{2} \leq \mathbf{E}_{\pi} f^{2} \leq \mathbf{E}_{\pi} f
$$

Lemma 3.4 implies that standard Monte Carlo method can be used to accurately estimate $\mathbf{E}_{\pi} f=\frac{\operatorname{Pr}_{O}[X \neq Y]}{\operatorname{Pr}_{C}[X \neq Y]}$. To implement the Monte Carlo algorithm, we use Lemma 3.2 and Lemma 3.3.

To be more specific, our approximate algorithm is to compute the median of means. The input contains the descriptions of $2 n$ distributions $P_{1}, P_{2}, \ldots, P_{n}$ and $Q_{1}, Q_{2}, \ldots, Q_{n}$ together with two parameters $\epsilon>0$ and $0<\delta<1$. The algorithm proceeds as follows:

- for each $i$ from 1 to $m=\left\lceil\frac{10 n}{\epsilon^{2}}\right\rceil$, independently sample $\omega_{i} \sim \pi$ and let

$$
F=\frac{1}{m} \sum_{i=1}^{m} f\left(\omega_{i}\right)
$$

- use independent samples to compute $F$ for $s=10\left\lceil\log \frac{1}{\delta}\right\rceil$ times to get $F_{1}, F_{2}, \ldots, F_{s}$ and let

$$
\widehat{F}=\operatorname{Median}\left\{F_{1}, F_{2}, \ldots, F_{s}\right\}
$$

- output the value $\widehat{d}=\left(1-\prod_{i=1}^{n}\left(1-d_{\mathrm{TV}}\left(P_{i}, Q_{i}\right)\right)\right) \widehat{F}$.

We claim that

$$
\begin{equation*}
\operatorname{Pr}\left[\left|F-\mathbf{E}_{\pi} f\right| \geq \epsilon \mathbf{E}_{\pi} f\right] \leq \frac{1}{10} \tag{3.9}
\end{equation*}
$$

Assuming that (3.9) holds, by the Chernoff bound, it holds that

$$
\operatorname{Pr}\left[\left|\widehat{F}-\mathbf{E}_{\pi} f\right| \geq \epsilon \mathbf{E}_{\pi} f\right] \leq \delta
$$

Using (3.5) in Lemma 3.4 and (3.1), we have

$$
\operatorname{Pr}\left[\left|\widehat{d}-d_{\mathrm{TV}}(P, Q)\right| \geq \epsilon d_{\mathrm{TV}}(P, Q)\right]=\operatorname{Pr}\left[\left|\widehat{F}-\mathbf{E}_{\pi} f\right| \geq \epsilon \mathbf{E}_{\pi} f\right] \leq \delta
$$

By Lemma 3.2 and Lemma 3.3, the total running time is $O(n m s)=O\left(\frac{n^{2}}{\epsilon^{2}} \log \frac{1}{\delta}\right)$. This proves Theorem 3.1.

Finally, we prove the claim (3.9). Note that the expectation and the variance of the random variable $F$ satisfy that $\mathbf{E} F=\mathbf{E}_{\pi} f$ and $\operatorname{Var} F=\frac{1}{m} \operatorname{Var}_{\pi} f$. By Chebyshev's inequality,

$$
\begin{aligned}
\operatorname{Pr}\left[\left|F-\mathbf{E}_{\pi} f\right| \geq \epsilon \mathbf{E}_{\pi} f\right] & =\operatorname{Pr}[|F-\mathbf{E} F| \geq \epsilon \mathbf{E} F] \leq \frac{\operatorname{Var} F}{\epsilon^{2}(\mathbf{E} F)^{2}}=\frac{\operatorname{Var}_{\pi} f}{m \epsilon^{2}\left(\mathbf{E}_{\pi} f\right)^{2}} \\
& \leq \frac{1}{m \epsilon^{2} \mathbf{E}_{\pi} f} \leq \frac{n}{m \epsilon^{2}} \leq \frac{1}{10} . \quad\left(\text { by }(3.8),(3.6), \text { and } m=\left\lceil\frac{10 n}{\epsilon^{2}}\right\rceil\right)
\end{aligned}
$$

## Chapter 4

## Inapproximability

The computational hardness of hypergraph colourings, especially the approximate counting problem, is concerned in this chapter. Our goal is to prove the following main result.

Theorem 4.1. Let $q \geq 4$ be even, $k \geq 4$ be even, and $\Delta \geq 5 q^{k / 2}$. It is NP-hard to approximate the number of proper $q$-colourings in n-vertex $k$-uniform hypergraphs of maximum degree at most $\Delta$, even within a factor of $2^{c n}$ for some constant $c(q, k)>0$.

The requirements on parity of $q$ and $k$ appear bizarre at first glance. Indeed, they are artifacts introduced by the technique. We elaborate this a bit below.

First, our result applies to only even $k$ for $k$-uniform hypergraphs. This is due to a particular halving construction we use in the reduction. The hardness results for (monotone) $k$-SATs $\left[\mathrm{BGG}^{+} 19\right]$ allow hyperedges with sizes at least $k$. This is a stronger assumption and our hardness bound would still apply without changing the asymptotical order. In fact, we expect a slight variant of our construction to work for odd $k$ to achieve a threshold of the same order. (See Remark 4.11.) However, as we explain soon, the details for even $k$ are already very complicated for hypergraph colourings, so for clarity and simplicity we did not pursue the odd $k$ case.

Secondly, our result applies to an even number of colours $q$, which is analogous to hardness results for counting in the graph colouring setting [GŠV15]. It was left as an open problem in [GŠV15] to handle odd $q$ (see also the recent work [CGSV22]), and we met the same difficulty in our setting as well. Our hardness proof for counting builds on ideas from [GŠV15], and we focus on the challenges needed to refine them in the hypergraph setting (rather than addressing the parity of $q$ ). We expect that substantial new ideas are required to resolve the parity of $q$, even in the graph setting.

Technique overview. In order to show Theorem 4.1, we first reduce from an auxiliary weighted binary CSP, namely a "spin system" in graphs.

Definition 4.2 (Spin system). Let $G$ be a simple undirected graph, $q \geq 2$ be an integer and $\boldsymbol{B}$ be a $q$-by- $q$ non-negative matrix (called interaction matrix). A $q$-spin system is specified by the tuple $(G ; \boldsymbol{B})$ as follows. A configuration $\sigma: V \rightarrow[q]$ is an assignment of the $q$ spins to the vertices of $G$. The weight of a configuration is defined as

$$
\operatorname{wt}(\sigma):=\prod_{(u, v) \in E} B_{\sigma(u) \sigma(v)} .
$$

The partition function is the sum of weights over all assignments:

$$
\begin{equation*}
Z_{\boldsymbol{B}}(G):=\sum_{\sigma: V \rightarrow\{1, \ldots, q\}} \mathrm{wt}(\sigma) . \tag{4.1}
\end{equation*}
$$

In particular, the $q$-state antiferromagnetic Potts model corresponds to the case where $\boldsymbol{B}$ is the matrix whose off-diagonal entries are equal to 1 , whereas the diagonal entries equal to some parameter $0 \leq B<1$ (note, $B=0$ corresponds to $q$-colourings).

Basically, we replace each vertex of the graph by a cluster of $k / 2$ vertices in the hypergraph, and an edge by a hyperedge of size $k$. This construction is identical to the one in $\left[\mathrm{BGG}^{+} 19\right]$, via which one reduces from weighted independent set in graphs to hypergraph independent sets. However, in order to reduce to the hypergraph $q$ colouring problem, the variables of the weighted binary CSP take $q+1$ possible values. The interactions among these $q+1$ states are dictated by the hypergraph colouring problem, with $q$ states that correspond to "pure" colours, and one special value that corresponds to a "mixed" state. The mixed state behaves very differently from the pure colours; roughly, the pure colours behave symmetrically (as in the graph case) but the mixed state causes asymmetry.

Our next and main step is to show the desired hardness result for this spin system. We follow an established route of establishing inapproximability for spin systems [DFJ02, MWW09, Sly10, CCGL12, SS14, GŠV16], and in particular [GŠV15], where the key is to understand the system on random regular bipartite graphs which are used as gadgets in the reduction. More precisely, we need to analyze what are the most likely configurations of the system on random regular bipartite graphs, the socalled dominant phases (given by the normalised counts of the colours on each side of the graph). It was shown in [GŠV15] that these are captured by a certain matrix norm of the interaction matrix. These norms are in general very hard to penetrate
analytically and it was already a major difficulty in the perfectly symmetric setting of [GŠV15]. For us, the presence of a special spin together with $q$ symmetric spins makes our spin system very different from all of the spin systems analyzed before and the mixture of symmetry and asymmetry make the analysis substantially harder. For example, in [GŠV15], to show that the two parts of the graph are unbalanced, a simple Hessian calculation suffices, whereas in our setting, there are stable balanced phases due to the presence of this special spin (that can be favoured against the others). Also, being stable means that this phase is locally maximal, making perturbation arguments hard to carry out. What we do instead is to directly compare this phase with the dominant phase via a careful interpolation path and a sequence of delicate estimates. This reflects the main difference between our work and previous works, namely that our estimates and perturbation arguments are significantly more delicate in order to rule out the local-maxima.

Outline of this chapter. This chapter begins with a (de)tour that confirms the optimality of the Moser-Tardos algorithm for the searching problem in Section 4.1, after which we shall focus completely on the counting problem in the sub-LLL regime. In Section 4.2, we carry out the first step of reduction, namely reducing from a certain spin system. And then in Section 4.3, we analyse in detail the dominant phases of the spin system and hence its inapproximability, which leads to the main theorem eventually.

### 4.1 Hardness for searching

In this section we show hardness results for finding hypergraph colourings for parameters beyond the local lemma condition. The key is to find instances that do not have proper colourings.

We will use a configuration model to construct random regular hypergraphs. With constant probability, the resulting hypergraph is linear [CFMR96, PP19]. Frieze and Mubayi [FM13] showed that if $q>c\left(\frac{\Delta}{\log \Delta}\right)^{\frac{1}{k-1}}$ for some constant $c=c(k)$ that only depends on $k$, then any linear $k$-uniform hypergraph with maximum degree $\Delta$ is $q$ colourable. In particular, their condition holds if $\Delta \leq c k q^{k-1} \ln q$ for some constant $c=c(k)$. Our next lemma complements their result by showing as an intermediate result that if $\Delta>k q^{k-1} \ln q+1$, we can find a $k$-uniform hypergraph with maximum degree $\Delta$ which is not $q$-colourable. For our reductions, we use such hypergraphs to
obtain a "disequality" gadget, as detailed in the lemma below.
Lemma 4.3. Let $q, k \geq 2$ be integers. Then, for all integers $\Delta>k q^{k-1} \ln q+1$, there exists a q-colourable $k$-uniform linear hypergraph $H$ with maximum degree $\Delta$ and two distinct vertices $u, v$ such that the degree of $u$ is 1 , the degree of $v$ is at most $\Delta$, and for every $q$-colouring $\sigma$ of $H$ it holds that $\sigma(u) \neq \sigma(v)$.

Proof. We first argue that for all $\Delta>k q^{k-1} \ln q$ there is a $\Delta$-regular hypergraph $H_{0}$ such that $Z_{\text {col }}\left(H_{0}\right)=0$, where $Z_{\text {col }}(H)$ denotes the number of $q$-colourings in $H$.

Let $n$ be such that $m=n \Delta / k$ is an integer. We sample a $k$-uniform $\Delta$-regular hypergraph $H$ according to the following pairing model (see [PP19]). Start with a bipartite graph with the points $[n] \times[\Delta]$ on the left and the points $[m] \times[k]$ on the right, and pair the two sides using a uniformly random perfect matching; the vertex set of the final hypergraph $H$ is obtained in the natural way by projecting the set $[n] \times[\Delta]$ onto $[n]$. Note that it will be convenient to view the hyperedges of $H$ for now as ordered tuples rather than sets; this does not make any difference when considering colourings of $H$ due to the symmetry among possible ordering of the colours within the hyperedge. It is a well-known fact, see for example [CFMR96, Lemma 2] or [PP19, Theorem 2.4 \& Appendix A.4], that the probability that $H$ is linear is bounded away from zero for all sufficiently large $n .{ }^{1}$

For a colouring $\sigma:[n] \rightarrow[q]$, a colour $i \in[q]$ and a $k$-tuple of colours $\boldsymbol{i}=$ $\left(i_{1}, \ldots, i_{k}\right) \in[q]^{k}$, let $n \alpha_{i}$ be the number of vertices with colour $i$, and $m \beta_{i}$ be the number of hyperedges whose vertices are coloured according to $\boldsymbol{i}$ (i.e., the $j$-th vertex of the hyperedge takes the colour $i_{j}$ ). Let $\alpha=\left\{\alpha_{i}\right\}_{i \in[q]}$ and $\beta=\left\{\beta_{i}\right\}_{i \in[q]^{k}}$, and note that $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in S_{q}$, where $S_{q}$ is the space of all pairs of vectors in $\mathbb{R}^{q} \times \mathbb{R}^{q^{k}}$ satisfying

$$
\begin{gather*}
\sum_{i \in[q]} \alpha_{i}=1, \quad \sum_{i \in[q]} t_{i, i} \beta_{i}=k \alpha_{i} \text { for } i \in[q] \\
\alpha_{i} \geq 0 \text { for } i \in[q], \quad \beta_{i} \geq 0 \text { for } \boldsymbol{i} \in[q]^{k}, \quad \beta_{(i, i, \ldots, i)}=0 \text { for } i \in[q], \tag{4.2}
\end{gather*}
$$

where for $i \in[q]$ and $\boldsymbol{i} \in[q]^{k}$ we denote by $t_{i, i}$ the number of occurrences of colour $i$ in the tuple $\boldsymbol{i}$. Then, we have

$$
\mathbf{E}\left[Z_{c o l}(H)\right]=\frac{1}{(\Delta n)!} \sum_{\substack{(\alpha, \beta) \in S_{q}: \\ n \alpha \in \mathbb{Z}^{q}, m \boldsymbol{\beta} \in \mathbb{Z}^{q^{k}}}}\binom{n}{\alpha_{1} n, \ldots, \alpha_{q} n}\binom{m}{\beta_{1} m, \ldots, \beta_{q^{k}} m} \prod_{i \in[q]}\left(\Delta \alpha_{i} n\right)!,
$$

[^4]since a term in the sum corresponding to ( $\boldsymbol{\alpha}, \boldsymbol{\beta}$ ) accounts for the number of ways to choose $\sigma$ and $H$ with vertex-colour frequencies given by the vector $\alpha$ and edge-colour frequencies given by the vector $\boldsymbol{\beta}$. Using Stirling's approximation $(2 \pi c)^{1 / 2}(c / \mathrm{e})^{c} \leq$ $c!\leq \mathrm{e} c^{1 / 2}(c / \mathrm{e})^{c}$ that holds for all integers $c \geq 1$, we obtain by expanding the terms inside the sum (note that there are at most $n^{q^{k}+q}$ of them) that
\[

$$
\begin{equation*}
\mathbf{E}\left[Z_{c o l}(H)\right] \leq n^{o(1)} \exp \left(n \max _{(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in S_{q}} F(\boldsymbol{\alpha}, \boldsymbol{\beta})\right), \tag{4.3}
\end{equation*}
$$

\]

where $F(\boldsymbol{\alpha}, \boldsymbol{\beta})=-(\Delta-1) h(\boldsymbol{\alpha})+\frac{\Delta}{k} h(\boldsymbol{\beta})$ and $h(\cdot)$ is the entropy function (here, we adopt the usual convention that $0 \ln 0=0$ which makes $h$ and $F$ continuous and therefore the maximum in (4.3) well-defined).

For $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in S_{q}$, we have that $\alpha_{i}=\frac{1}{k} \sum_{i \in[q]^{k}} t_{i, i} \beta_{i}$ for $i \in[q]$, and hence

$$
F(\boldsymbol{\alpha}, \boldsymbol{\beta})=h(\boldsymbol{\alpha})+\frac{\Delta}{k} G(\boldsymbol{\alpha}, \boldsymbol{\beta}) \text { where } G(\boldsymbol{\alpha}, \boldsymbol{\beta})=h(\boldsymbol{\beta})-\sum_{i \in[q]} \ln \left(\alpha_{i}\right) \sum_{i \in[q]} t_{i, i} \beta_{i} .
$$

Note that for a fixed vector $\alpha$, the function $G_{\alpha}(\beta):=G(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is concave and the method of Lagrange multipliers yields that the maximum of $G_{\boldsymbol{\alpha}}$ happens at $\boldsymbol{\beta}^{*}=$ $\left\{\beta_{i}^{*}\right\}_{i \in[q]^{k}}$ that satisfies

$$
\beta_{i}^{*}=\frac{\prod_{i \in[q]}\left(\alpha_{i}\right)^{t_{i, i}} \prod_{i \in[q]} \mathbf{1}_{\boldsymbol{i}^{*} \neq(i, i, \ldots, i)}}{1-\|\boldsymbol{\alpha}\|_{k}^{k}} \text { for } \boldsymbol{i} \in[q]^{k}, \quad G_{\alpha}\left(\boldsymbol{\beta}^{*}\right)=\ln \left(1-\|\boldsymbol{\alpha}\|_{k}^{k}\right) .
$$

It follows that

$$
\begin{equation*}
F(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq h(\boldsymbol{\alpha})+\frac{\Delta}{k} \ln \left(1-\|\boldsymbol{\alpha}\|_{k}^{k}\right) \leq \ln \left(q\left(1-\frac{1}{q^{k-1}}\right)^{\Delta / k}\right), \tag{4.4}
\end{equation*}
$$

where the last inequality follows from $h(\boldsymbol{\alpha}) \leq \ln q$ and $\|\boldsymbol{\alpha}\|_{k}^{k} \geq 1 / q^{k-1}$, both of which are simple applications of Jensen's inequality. For $\Delta>k q^{k-1} \ln q$, the r.h.s. of (4.4) is negative and therefore $\max _{(\alpha, \boldsymbol{\beta}) \in S_{q}} F(\boldsymbol{\alpha}, \boldsymbol{\beta})<0$. From (4.3), we conclude that $Z_{\text {col }}(H)=$ 0 with probability $1-\exp (-\Omega(n))$. By a union bound, we obtain a linear $\Delta$-regular hypergraph $H_{0}$ with $Z_{\text {col }}\left(H_{0}\right)=0$, as claimed.

To obtain the final hypergraph $H$ with the desired property, we invoke an argument in [GG16, Lemma 28] (which in turn was inspired by [KST93]). We give the details here for completeness. Given the above $H_{0}=(V, \mathcal{E})$ that $Z_{\text {col }}\left(H_{0}\right)=0$, we can remove hyperedges from $\mathcal{E}$ one by one until removing any more hyperedge makes $Z_{\text {col }}(H)>0$. Call the resulting hypergraph $H_{0}^{\prime}=\left(V, \mathcal{E}_{0}^{\prime}\right)$ where $\mathcal{E}_{0}^{\prime} \subsetneq \mathcal{E}$. Clearly $H_{0}^{\prime}$ is linear and has at least one hyperedge. In other words, $H_{0}^{\prime}$ is the minimal subgraph of $H_{0}$ that cannot be $q$-coloured.

Choose an arbitrary hyperedge $e \in \mathcal{E}_{0}^{\prime}$. Let $S \subseteq e$ be the set of vertices with nonzero degree in $H_{0}^{\prime}-e$. If $S=\varnothing$, then $e$ is disconnected from the rest of the graph. Thus


Figure 4.1: An example construction of $H$ from $H_{0}^{\prime}$. Here, $e=\left\{v_{1}, \cdots, v_{5}\right\}$ and $S=$ $\left\{v_{1}, v_{2}\right\}$.
as $H_{0}^{\prime}$ is not $q$-colourable, removing $e$ would not make the hypergraph $q$-colourable. This contradicts to the minimality of $H_{0}^{\prime}$ and thus $S \neq \varnothing$. Denote the vertices in $S$ by $v_{1}, \ldots, v_{i}$, and the vertices in $e \backslash S$ by $v_{i+1}, \ldots, v_{k}$. We construct $i$ linear hypergraphs $H_{1}, \ldots, H_{i}$ where for $1 \leq j \leq i$, in $H_{j}$ we introduce new vertices $u_{1}, \ldots, u_{j}$ and replace the hyperedge $e$ by $e_{j}:=\left\{u_{1}, \ldots, u_{j}, v_{j+1}, \ldots, v_{k}\right\}$. By minimality of $H_{0}^{\prime}$ again, $Z_{c o l}\left(H_{i}\right)>0$ as $e_{i}$ is disconnected from the rest of $H_{i}$. Thus we can find the smallest $j \geq 1$ such that $Z_{\text {col }}\left(H_{j}\right)>0$ and $Z_{\text {col }}\left(H_{j-1}\right)=0\left(\right.$ or $Z_{\text {col }}\left(H_{0}^{\prime}\right)=0$ if $\left.j=1\right)$. See Figure 4.1 for an example.

For any proper colouring $\sigma$ of $H_{j}$, if $\sigma\left(u_{j}\right)=\sigma\left(v_{j}\right), \sigma$ would be a proper colouring of $H_{j-1}$, contradicting to the above. Thus it must hold that for any colouring $\sigma$ of $H_{j}$, $\sigma\left(u_{j}\right) \neq \sigma\left(v_{j}\right)$. This is the hypergraph required by the lemma, with $u=u_{j}$ and $v=v_{j}$. Moreover, the degree of $u_{j}$ is 1 , and the degree of $v_{j}$ is at most $\Delta$.

Lemma 4.3 leads to the following hardness result, where we lose a factor $2 q$ in the degree bound due to the reduction. We note that for $q=2, k=3, \Delta=4$, and linear hypergraphs, NP-hardness is known [DD20]. However, the main point of the next theorem is that there is a degree bound that scales roughly as $q^{k}$ and makes the problem NP-hard.

Theorem 4.4. Let $q, k \geq 2$ be integers with $(q, k) \neq(2,2)$. Then, it is NP-hard to find a $q$-colouring on a $k$-uniform linear hypergraph of maximum degree at most $\Delta$, when $\Delta \geq 2 k q^{k} \ln q+2 q$.

Proof. For $q>2$, we reduce from the problem of finding $q$-colourings in graphs whose degrees are bounded by $2 q$. The latter problem is shown to be NP-hard by [EHK98]. Given a graph $G$, we replace each edge $(u, v)$ of $G$ by the hypergraph in Lemma 4.3, where $u$ and $v$ are identified with the special vertices in the hypergraph. Then each
such hypergraph is effectively a disequality for the colours of $u$ and $v$. Call the resulting hypergraph $H$. Thus $G$ is $q$-colourable if and only if $H$ is $q$-colourable. The maximum degree of $H$ is $2 q\left(k q^{k-1} \ln q+1\right)$.

For $q=2$ and $k>2$, using two copies of the hypergraph from Lemma 4.3, we build an "equality" gadget, i.e., a linear hypergraph $H$ of maximum degree $\Delta \leq 2\left(k q^{k-1} \ln q+\right.$ 1) $=k 2^{k} \ln 2+2$ with distinct vertices $u, v$ which both have degree 1 such that for every $q$-colouring $\sigma$ it holds that $\sigma(u)=\sigma(v)$. It is well-known that finding 2 -colourings of $k$-uniform linear hypergraphs is NP-hard (or we can use for example [DD20]), and using the equality gadget $H$, for any $k$-uniform linear hypergraph $F$, we can construct a $k$-uniform linear hypergraph $F^{\prime}$ of maximum degree $\Delta$ such that $F$ is 2-colourable if and only if $F^{\prime}$ is 2 -colourable. One possible way to do so is replacing each degree- $d$ vertex $w$ of $F$ with a cycle of length $d$ and then replacing each edge $e$ of the cycle with a distinct copy of the hypergraph $H$ using $u, v$ for the endpoints of the edge $e$; then, for each hyperedge of $F$ that uses $w$, in $F^{\prime}$ we use instead one of the $d$ vertices of the cycle.

Note that the result of Frieze and Mubayi [FM13] is also algorithmic. Thus Theorem 4.4 is sharp for linear hypergraphs up to a factor $c q$ where $c=c(k)$ is a constant depending only on $k$. For general hypergraphs, the algorithm of Moser and Tardos [MT10] applies in this setting when $\Delta \leq \frac{q^{k-1}}{\mathrm{e}(k-1)}$, in which case Theorem 4.4 almost matches the algorithmic result, up to a factor of $c k^{2} q \ln q$ where $c$ is a constant.

For approximate counting, we can avoid the loss of the factor $q$ when $q \geq 2$ and $k \geq 4$. We will use the following hardness result about the Potts model.

Lemma 4.5. There is a constant $C_{1}>5$ such that, for any integers $q \geq 2, \Delta \geq 2 C_{1} q \ln q$, and $B<1-\frac{C_{1} q \ln q}{\Delta}$, there is no FPRAS to approximate the $q$-state antiferromagnetic Potts partition function $Z_{B}$ in graphs with bounded degree $\Delta$, unless $\mathbf{N P}=\mathbf{R P}$.

The proof of Lemma 4.5 is quite a detour from the problems we focus on, so we postpone it to Section 4.1.1. We note that Lemma 4.5 is weaker than the inapproximability result in [GŠV15, Theorem 1.2], which achieves $B<1-\frac{q}{\Delta}$ but only holds for even $q$. We want to deal with general $q$, and thus settle with this weaker version.

Theorem 4.6. There is a constant $C_{1}>5$ such that, for any integers $q \geq 2, k \geq 4$, and $\Delta \geq C_{1} k q^{k-1} \ln q$, unless $\mathbf{N P}=\mathbf{R P}$, there is no FPRAS for the number of $q$-colourings in $k$-uniform linear hypergraphs of maximum degree at most $\Delta$.

Proof. We reduce the partition function of the $q$-state antiferromagnetic Potts model with $B=1-\frac{1}{q^{2}-3 q+3}$ in graphs with bounded degree $\Delta$ to the problem of counting $q$-colourings in $k$-uniform linear hypergraphs of maximum degree at most $\Delta$. Note that if $k \geq 4$ and $\Delta \geq C_{1} k q^{k-1} \ln q$, where $C_{1}$ is from Lemma 4.5 , then $B<1-\frac{C_{1} q \ln q}{\Delta}$. Thus the reduction implies the theorem via Lemma 4.5.

The reduction goes as follows. For each edge $(u, v)$ in a $\Delta$-regular graph $G=$ $(V, E)$, we replace it by a gadget using the hypergraph $H$ in Lemma 4.3, whose degree bound is $\Delta_{0}=k q^{k-1} \ln q+1$. To be more specific, we introduce new vertices $w_{1}$ and $w_{2}$. We add three copies of the hypergraph $H$ with special vertices $\left(u, w_{1}\right),\left(w_{2}, w_{1}\right)$, and ( $v, w_{2}$ ), respectively. Do this for all edges in $G$. Then, the degrees of $u$ and $v$ are still $\Delta$, the degrees of $w_{1}$ 's are at most $2 \Delta_{0}<\Delta$, and the degrees of $w_{2}$ 's are at most $\Delta_{0}+1<\Delta$. All other newly introduced vertices have degrees at most $\Delta_{0}<\Delta$. Thus, the degree requirement is met. Call the resulting hypergraph $H_{G}$.

To finish the reduction, we claim that

$$
Z_{c o l}\left(H_{G}\right)=C^{|E|} Z_{B}(G),
$$

where $C$ is a constant depending only on $H$. First notice that for any pair of colours $i$ and $j$, the number of colourings $\sigma$ of $H$ such that $\sigma(u)=i$ and $\sigma(v)=j$ is a constant, due to the symmetry among colours. Denote this constant by $C_{0}$. Thus, in the gadget above, when the two endpoints $u$ and $v$ have different colours, the number of possible colourings for the gadget is $\left((q-2)^{2}+(q-1)\right) C_{0}^{3}$; when the two endpoints $u$ and $v$ have the same colour, the number of possible colourings for the gadget is $(q-1)(q-$ 2) $C_{0}^{3}$. The claim holds with $C=\left((q-2)^{2}+(q-1)\right) C_{0}^{3}$.

In Theorem 4.6, we could avoid the large constant $C_{1}$ in the degree bound by using [GŠV15, Theorem 1.2] as the starting point of our reduction, but doing so will restrict the result to even $q$ only.

For $k$-SATs on linear hypergraphs, Hermon, Sly, and Zhang [HSZ19] showed an efficient approximate counting and sampling algorithm if $\Delta \leq \frac{c 2^{k}}{k^{2}}$, where $c$ is a constant. In view of their result, Theorem 4.4 and Theorem 4.6 are potentially sharp for linear hypergraphs, up to some polynomial factor in $k$.

### 4.1.1 Proof of Lemma 4.5

We will consider the following computational problem. Given a graph $G=(V, E)$, for a $q$-colouring $\sigma: V \rightarrow\{1, \ldots, q\}$, let $\operatorname{Mono}(G, \sigma)$ be the number of monochromatic edges under $\sigma$.

Name Max-q-Cut
Instance A undirected graph $G=(V, E)$
Output $\max _{\sigma: V \rightarrow\{1, \ldots, q\}}\{|E|-\operatorname{Mono}(G, \sigma)\}$
An approximation for this problem with relative error $\delta$ requires an output that is at least $(1-\delta)$ times the optimal value. Let Max-Cut be the $q=2$ version of Max-$q$-Cut. Alimonti and Kann [AK00] showed the following.

Proposition 4.7. There is a constant $\delta_{0}>0$ such that, there is no randomized polynomialtime approximation algorithm for MAx-CUT in cubic graphs with relative error $\delta_{0}$ unless $\mathbf{N P}=\mathbf{R P}$.

Furthermore, Kann, Khanna, Lagergren, and Panconesi [KKLP97] showed the following reduction.

Proposition 4.8. For any $0 \leq \delta \leq 1$, if MAX- $q$-CUT in $\left(\frac{\Delta(q+1)}{2}+\frac{q-1}{2}\right)$-regular graphs can be approximated within relative error $\frac{\delta}{2(q+1)}$ in polynomial-time, then Max-Cut can be approximated within $\delta$ in polynomial-time for $\Delta$-regular graphs.

The original reduction in [KKLP97, Theorem 1] works for only even $q$ and gives relative error lower bound $\frac{\delta}{2(q-1)}$ instead. For odd $q$ they used a different reduction to achieve the same lower bound but it does not keep the degrees bounded. Here we briefly describe how to modify the reduction in [KKLP97, Theorem 1] such that it works for odd $q$ as well, albeit with a slightly worse relative error lower bound $\frac{\delta}{2(q+1)}$. For odd $q$, given an instance $G=(V, E)$ for Max-Cut, we replace each vertex $v \in V$ by a clique $C_{v}$ of size $\frac{q+1}{2}$ (instead of $\frac{q}{2}$ in the original reduction), and replace each edge $(u, v) \in E$ by a bipartite complete graph between $C_{v}$ and $C_{u}$. Moreover, give weight $\frac{q+1}{q-1} d_{G}(v)$ for edges inside $C_{v}$ (instead of $\left.d_{G}(v)\right)$ and keep weight 1 for all other edges. It can be verified straightforwardly that the proof still works, except that the parameters $\alpha$ and $\beta$ changed from $\left(\frac{q(q-1)}{2}, \frac{2}{q}\right)$ to $\left(\frac{(q+1)^{2}}{2}, \frac{2}{q+1}\right)$, which leads to the worse lower bound $\frac{\delta}{2(q+1)}$. Finally, Crescenzi, Silvestri, and Trevisan [CST01] showed that for a general class of combinatorial optimization problem, including Max- $q$-CuT, the weighted and unweighted versions have the same approximation complexity.

Lemma 4.9. There is a constant $0<C_{0}<1$ such that for any $q \geq 2$, there is no FPRAS for the $q$-state Potts model with weights $B<q^{-1 / C_{0}}$ in $(2 q+1)$-regular graphs unless $\mathbf{N P}=\mathbf{R} \mathbf{P}$.

Proof. Let $C_{0}:=\frac{5 \delta_{0}}{24}$, where $\delta_{0}$ is from Proposition 4.7. We claim that for any $q \geq 2$, an FPRAS for $Z_{B}(G)$ with weight $B<q^{-1 / C_{0}}$ in graphs with degree bound $2 q+1$ implies an efficient approximation of MAX- $q$-CuT within relative error $\varepsilon_{0}:=\frac{\delta_{0}}{2(q+1)}$ in graphs with the same degree bound. Then Proposition 4.7 and Proposition 4.8 (with $\Delta=3$ ) imply the lemma.

Given an instance $G=(V, E)$ to MAx- $q$-Cut, assume the maximum value of $q$-cut is Opt. Let $n:=|V|$ and $m:=|E|$. Then $m=\frac{(2 q+1) n}{2}$. If we had an FPRAS for the $q$-state Potts model, then we can efficiently sample a colouring proportional to its weight. (In the local lemma setting, one such reduction is given in [JPV21b].) The probability that the cut value of the colouring is less than $\left(1-\varepsilon_{0}\right)$ Opt is at most

$$
\frac{q^{n} B^{m-\left(1-\varepsilon_{0}\right) \mathrm{Opt}}}{B^{m-\mathrm{Opt}}+q^{n} B^{m-\left(1-\varepsilon_{0}\right) \mathrm{Opt}}}
$$

In particular, this probability is at most $1 / 2$ if

$$
B^{m-\mathrm{Opt}} \geq q^{n} B^{m-\left(1-\varepsilon_{0}\right) \mathrm{Opt}}
$$

which is equivalent to $B^{-\varepsilon_{0} \mathrm{Opt}} \geq q^{n}$. On the other hand, notice that a uniformly at random colouring achieves cut value $\left(1-\frac{1}{q}\right) m$ in expectation. Thus, $\mathrm{Opt} \geq\left(1-\frac{1}{q}\right) m=$ $\frac{(2 q+1)(q-1) n}{2 q}$. Consequently, for any $q \geq 2, B^{-\varepsilon_{0} \mathrm{Opt}} \geq B^{-\delta_{0} n \frac{(2 q+1)(q-1)}{4 q(q+1)}} \geq B^{-C_{0} n}$, since $C_{0}=\frac{5 \delta_{0}}{24} \leq \frac{(2 q+1)(q-1) \delta_{0}}{4 q(q+1)}$ for $q \geq 2$. Thus if $B<q^{-1 / C_{0}}, B^{-\varepsilon_{0} \mathrm{Opt}} \geq q^{n}$ as desired. Standard methods can boost the success probability from $1 / 2$ to arbitrarily close to 1 .

Now we are ready to show Lemma 4.5.

Proof of Lemma 4.5. Given a $(2 q+1)$-regular graph $G=(V, E)$, we replace each edge by $s:=\left\lfloor\frac{\Delta}{2 q+1}\right\rfloor$ parallel edges to get a new graph $G^{\prime}$ whose degree is at most $(2 q+1) s \leq$ $\Delta$. As $q \geq 2, C_{1} \geq 5$, and $\Delta \geq 2 C_{1} q \ln q, s \geq \frac{\Delta}{2 q+1}-1>0.63 \frac{\Delta}{2 q+1}$.

If we have a Potts model with edge weight $B$ on $G^{\prime}$, then effectively, this is a Potts model on $G$ with $B^{\prime}=B^{s}$. Thus if $B<1-\frac{C_{1} q \ln q}{\Delta}$ for $C_{1}=5 / C_{0}$, where $C_{0}$ is from Lemma 4.9, then

$$
B^{-s C_{0}}>\left(1+\frac{C_{1} q \ln q}{\Delta}\right)^{s C_{0}} \geq e^{\frac{0.8 s C_{0} C_{1} q \ln q}{\Delta}} \geq e^{\frac{0.8 * 0.63 C_{0} C_{1} q \ln q}{2 q+1}}>e^{\frac{C_{0} C_{1} q \ln q}{2(2 q+1)}} \geq q^{0.2 C_{0} C_{1}} \geq q,
$$

where in the first line we used $1+x \geq e^{0.8 x}$ for $x \leq 0.5$. Thus this parallel construction can reduce from the Potts model satisfying the conditions of Lemma 4.9, which is NP-hard to approximate.

### 4.2 Reducing from spin systems

In this section we show our main theorem, Theorem 4.1, namely a refined inapproximability result for counting. As mentioned earlier, we will do this by first relating it to a multi-spin system on graphs with "antiferromagnetic" interaction matrix $\boldsymbol{B}$, and then establishing inapproximability results. It is tempting to pursue a strategy similar to that of Lemma 4.5 to show hardness for the spin system defined by $\boldsymbol{B}$. However, that strategy relies on hardness of finding the maximum weight configuration, and somewhat surprisingly, as we shall see soon, that problem for $\boldsymbol{B}$ is trivial. Instead, we need sharper tools from [GŠV15].

To define the spin system on graphs we will be interested in, we only need to specify its interaction matrix $\boldsymbol{B}$ (recall (4.1)). We use $[q]$ to denote $\{1, \ldots, q\}$ and $[\bar{q}]$ to denote $\{0,1, \ldots, q\}$. Let $t:=\left(q^{k^{\prime}}-q\right)^{1 / \Delta}$, where $k^{\prime}:=k / 2$, and $\boldsymbol{B}=\left\{B_{i j}\right\}_{i, j \in[\bar{q}]}$ be the matrix with block form

$$
\boldsymbol{B}=\left[\begin{array}{cc}
t^{2} & t \mathbf{1}^{\mathrm{T}}  \tag{4.5}\\
& \\
t \mathbf{1} & \boldsymbol{J}
\end{array}\right]
$$

where $\boldsymbol{J}$ is the $q \times q$ matrix with 0 s on the diagonal and 1 s elsewhere, and $\mathbf{1}$ is the $q \times 1$ vector with all ones. In the language of [GŠV15], the matrix $\boldsymbol{B}$ is antiferromagnetic and ergodic. ${ }^{2}$

Let $H$ be a $k$-uniform hypergraph, where $k=2 k^{\prime}$ is even, and recall that we use $Z_{\text {col }}(H)$ to denote the number of proper $q$-colourings of $H$. For any given $\Delta$ regular graph $G=(V, E)$, let $H_{G}$ be the hypergraph where every vertex $v \in V$ is replaced by $k^{\prime}$ new vertices $v_{1}, \ldots, v_{k^{\prime}}$, and each edge ( $u, v$ ) is replaced by a hyperedge $\left\{u_{1}, \ldots, u_{k^{\prime}}, v_{1}, \ldots, v_{k^{\prime}}\right\}$ of size $2 k^{\prime}$. Then $H_{G}$ is $2 k^{\prime}$-uniform and $\Delta$-regular. This construction has been used in $\left[\mathrm{BGG}^{+} 19\right]$, and yields the following lemma in our case.

Lemma 4.10. Let $G=(V, E)$ be a $\Delta$-regular graph, and $H_{G}=\left(V^{\prime}, E^{\prime}\right)$ be the $2 k^{\prime}$ uniform hypergraph constructed as above. Then, $Z_{\boldsymbol{B}}(G)=Z_{\text {col }}\left(H_{G}\right)$.

Proof. Let $\Omega_{\boldsymbol{B}}$ be the set of all assignments $\sigma$ of $G$ whose weights are non-zero. Let $\Omega_{\text {col }}$ be the set of all proper $q$-colourings $\tau$ of $H$. We will construct a surjective mapping $\varphi$ between $\Omega_{\text {col }}$ and $\Omega_{B}$, such that for any $\sigma,\left|\varphi^{-1}(\sigma)\right|=\operatorname{wt}(\sigma)$. This implies the lemma.

[^5]

Figure 4.2: The mapping $\varphi$. The "pure" colours are red, green and blue, while black stands for the "mixed" colour.

The mapping $\varphi$ is as follows. Given $\tau: V^{\prime} \rightarrow\{1,2, \ldots, q\}$, let

$$
\varphi(\tau)(v):= \begin{cases}i & \text { if } \tau\left(v_{1}\right)=\tau\left(v_{2}\right)=\cdots=\tau\left(v_{k^{\prime}}\right)=i \text { for some } 1 \leq i \leq q, \\ 0 & \text { otherwise } .\end{cases}
$$

See Figure 4.2 for an illustration.
We first show that $\varphi$ is surjective. Let $\sigma \in \Omega_{B}$ and we construct $\tau \in \Omega_{\text {col }}$ such that $\varphi(\tau)=\sigma$. For any $v$ such that $\sigma(v) \neq 0, \tau\left(v_{i}\right)=\sigma(v)$ for any $1 \leq i \leq k^{\prime}$. If $\sigma(v)=0$, then let $\tau\left(v_{i}\right)=1$ for any $1 \leq i \leq k^{\prime}-1$, and $\tau\left(v_{k}^{\prime}\right)=2$. It is easy to verify that $\tau$ is a proper $q$-colouring and $\varphi(\tau)=\sigma$ for this construction.

Next we calculate $\left|\varphi^{-1}(\sigma)\right|$. Let $n_{0}(\sigma)$ be the number of vertices assigned 0 under $\sigma$. Then

$$
\left|\varphi^{-1}(\sigma)\right|=\left(q^{k^{\prime}}-q\right)^{n_{0}(\sigma)}
$$

On the other hand, since $G$ is $\Delta$-regular,

$$
\operatorname{wt}(\sigma)=t^{\Delta n_{0}(\sigma)}=\left(q^{k^{\prime}}-q\right)^{n_{0}(\sigma)}=\left|\varphi^{-1}(\sigma)\right|
$$

which verifies the properties of $\varphi$.
Remark 4.11. For odd $k$, we may consider a similar construction, but in addition to clustering half of each hyperedge as a single vertex, we leave one vertex in the middle which appears only in this single hyperedge. The resulting spin system would have a different matrix $\boldsymbol{B}^{\prime}$, but the difference between $\boldsymbol{B}^{\prime}$ and the current $\boldsymbol{B}$ is not too much in the sense that the zeros would be replaced by small constants. We expect that we may obtain a hardness result for $\boldsymbol{B}^{\prime}$ for $\Delta$ of a similar order. However, since the details are already getting very complicated, we will only handle $\boldsymbol{B}$ in the rest of this paper.

Given Lemma 4.10, all we need to show is that the spin system with interaction matrix $\boldsymbol{B}$ is hard to approximate on $\Delta$-regular graphs, with $\Delta$ in the desired range. For this, we will use a result by Galanis, Vigoda, and Štefankovič [GŠV15, Theorem 1.5] which gives a sufficient condition in terms of studying a certain function (that can be formulated in terms of an induced norm of $\boldsymbol{B}$ ). Note that since $t>1$ the corresponding optimization problem related to $\boldsymbol{B}$ is trivial. Thus, we cannot use a strategy similar to that of Lemma 4.5 to show hardness for the spin system defined by $\boldsymbol{B}$.

The main construction in the gadget to show the hardness is the bipartite random $\Delta$-regular graph. Let $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ be a pair of vectors such that for $i \in[\bar{q}], \alpha_{i}$ and $\beta_{i}$ denotes the fraction of vertices with colour $i$ on the left and right sides of the bipartite random regular graph. If we draw a sample $\sigma$ proportional to its weight $\mathrm{wt}(\sigma)$, then with high probability over the choice of the random graph, the fraction of colours $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ will be from one of the dominant phases, for all but an exponentially small probability. Analyzing these dominant phases lies in the heart of [GŠV15, Theorem 1.5].

Let $\mathcal{G}_{n}$ denote the family of $\Delta$-regular bipartite graphs with $n$ vertices on each side. For a bipartite graph $G$ uniformly drawn from $\mathcal{G}_{n}$ and probability vectors $\alpha=$ $\left\{\alpha_{i}\right\}_{i \in[q]}, \boldsymbol{\beta}=\left\{\beta_{i}\right\}_{i \in[q]}$, we use $Z_{\boldsymbol{\beta}}^{\alpha, \boldsymbol{\beta}}(G)$ to denote the total weights of assignments whose fractions of colours on the two sides are given by $\alpha, \boldsymbol{\beta}$ respectively. Consider the function $\Psi_{1}$ that captures the exponential growth of the expectation of $Z_{\boldsymbol{B}}^{\alpha, \beta}(G)$, i.e.,

$$
\begin{equation*}
\Psi_{1}(\boldsymbol{\alpha}, \boldsymbol{\beta}):=\lim _{n \rightarrow \infty} \frac{1}{n} \log \mathbf{E}_{\mathcal{G}_{n}}\left[Z_{\boldsymbol{B}}^{\alpha, \boldsymbol{\beta}}(G)\right] . \tag{4.6}
\end{equation*}
$$

The function $\Psi_{1}$ has a relatively explicit form (see [GŠV15, Section 2]) using entropystyle functions though the exact details are not going to be important and we will in fact use a surrogate function later on (see Section 4.3).

Before stating the main result of [GŠV15], we need some further terminology. A dominant phase $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is a maximizer of the function $\Psi_{1}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and captures the most likely configurations for the spin system with interaction matrix on a random $\Delta$-regular graph. A dominant phase is called Hessian dominant if the Hessian of $\Psi_{1}$ is negative definite. Finally, two dominant phases $\left(\boldsymbol{\alpha}_{1}, \boldsymbol{\beta}_{1}\right)$ and ( $\boldsymbol{\alpha}_{2}, \boldsymbol{\beta}_{2}$ ) are permutation symmetric if there is a permutation matrix $\boldsymbol{P}$ such that $\boldsymbol{B}=\boldsymbol{P} \boldsymbol{B} \boldsymbol{P}^{\mathrm{T}}$ and $\left(\alpha_{1}, \boldsymbol{\beta}_{1}\right)=$ $\left(\boldsymbol{P} \boldsymbol{\alpha}_{2}, \boldsymbol{P} \boldsymbol{\beta}_{2}\right)$ or $\left(\boldsymbol{\alpha}_{1}, \boldsymbol{\beta}_{1}\right)=\left(\boldsymbol{P} \boldsymbol{\beta}_{2}, \boldsymbol{P} \boldsymbol{\alpha}_{2}\right)$. Now we can state [GŠV15, Theorem 1.5]. ${ }^{3}$

Proposition 4.12 ([GŠV15]). Let $\Delta \geq 3$ be an integer, and suppose that $\boldsymbol{B}$ is an ergodic interaction matrix of an antiferromagnetic spin system. Suppose further that the dom-

[^6]inant phases $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ satisfy $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$, are permutation symmetric and Hessian dominant. Then, it is $\mathbf{N P}$-hard to approximate the partition function $Z_{\boldsymbol{B}}(G)$ on $n$-vertex trianglefree $\Delta$-regular graphs $G$, even within a factor of $2^{c n}$ for a constant $c(\boldsymbol{B}, \Delta)>0$.

The key ingredient in the conditions of Proposition 4.12 is the condition that $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$; this enables a reduction in [GŠV15] to the MAx-Cut problem; the Hessian dominance and the permutation symmetry condition are more on the technical side, but is one of the main reasons that complicates the overall arguments (this was already prevalent in [GŠV15]).

The main challenge to show our inapproximability results is to establish the conditions of Proposition 4.12 for $\boldsymbol{B}$ and the relevant range for $\Delta$, which is the scope of the following lemma.

Lemma 4.13. Let $q \geq 4$ be even, $k^{\prime} \geq 2$, and $\Delta=5 q^{k^{\prime}}+1$. Then the dominant phases of the spin system with interaction matrix $\boldsymbol{B}$ (defined in (4.5)) satisfy the conditions of Proposition 4.12.

Theorem 4.1 follows from Lemma 4.10, Proposition 4.12, and Lemma 4.13. It remains to analyse the dominant phases of $\boldsymbol{B}$ and establish Lemma 4.13, which is the focus of Section 4.3.

### 4.3 Analysis of the dominant phases

In this section we analyze the dominant phase. We will state the main lemmata in this section and in Section 4.3.1. However, because the calculations are often very heavy, many of the lemmata are not immediately proved. The sections in which their proofs appear can be found at the end of Section 4.3.1.

Let $q, \Delta \geq 3$ be integers. To prove Lemma 4.13, we need to analyse the function $\Psi_{1}$ from (4.6). The function $\Psi_{1}$ turns out to be inconvenient to work with, but there is a simpler surrogate function $\Phi$ from [GŠV15] that we can use. For vectors $\boldsymbol{r}=\left\{R_{i}\right\}_{i \in[\bar{q}]}$ and $\boldsymbol{c}=\left\{C_{i}\right\}_{i \in[\bar{q}]}$ with nonnegative entries, let

$$
\begin{equation*}
\Phi(\boldsymbol{r}, \boldsymbol{c}):=\Delta \ln \frac{\boldsymbol{r}^{\mathrm{T}} \boldsymbol{B} \boldsymbol{c}}{\|\boldsymbol{r}\|_{p}\|\boldsymbol{c}\|_{p}}, \text { where } p=\Delta /(\Delta-1) . \tag{4.7}
\end{equation*}
$$

It is not hard to see that for the matrix $\boldsymbol{B}$ defined in (4.5), the critical points of $\Phi$
satisfy the following equations: ${ }^{4}$

$$
\begin{array}{lll}
R_{0} \propto t^{d}\left(t C_{0}+\sum_{j \in[q] ; j \neq i} C_{i}\right)^{d}, & R_{i} \propto\left(t C_{0}+\sum_{j \in[q] ; j \neq i} C_{i}\right)^{d} \quad \text { for } i \in[q] ; \\
C_{0} \propto t^{d}\left(t R_{0}+\sum_{i \in[q] ; i \neq j} R_{i}\right)^{d}, & C_{j} \propto\left(t R_{0}+\sum_{i \in[q] ; i \neq j} R_{i}\right)^{d} \text { for } j \in[q], \tag{4.8}
\end{array}
$$

where $t=\left(q^{k^{\prime}}-q\right)^{1 / \Delta}$ and $d:=\Delta-1$. The equations in (4.8) are often called the "tree recursions", because they are the same as the recursion for marginal probabilities on an infinite $d$-ary tree. Note that $1 \leq t \leq 1.0312$ for any $q \geq 4, k^{\prime} \geq 2$ and $d \geq 5 q^{k^{\prime}}$. The connection between the functions $\Psi_{1}$ and $\Phi$ is detailed in the following result from [GŠV15], applied to our setting.

Proposition 4.14 ([GŠV15, Theorem 4.1]). Let $q, \Delta \geq 3$ be integers, and let $p=\Delta /(\Delta-$ 1). Then, the local maxima of $\Phi$ and $\Psi_{1}$ happen at critical points, i.e., there are no local maxima on the boundary. The transformation $(\boldsymbol{r}, \boldsymbol{c}) \mapsto(\boldsymbol{\alpha}, \boldsymbol{\beta})$ given by $\alpha_{i}=R_{i}^{p} /\|\boldsymbol{r}\|_{p}^{p}$ and $\beta_{i}=C_{j}^{p} /\|\boldsymbol{c}\|_{p}^{p}$ for $i \in[\bar{q}]$ yields a one-to-one correspondence between the critical points of $\Phi$ and $\Psi_{1}$. Moreover, for the corresponding critical points $(\boldsymbol{r}, \boldsymbol{c})$ and $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ it holds that $\Psi_{1}(\boldsymbol{\alpha}, \boldsymbol{\beta})=\Phi(\boldsymbol{r}, \boldsymbol{c})$.

The function $\Phi$ is still multi-dimensional ( $2 q$ ), but fortunately we can reduce its dimensions significantly down to 11 by studying the structure of fixpoints to the system (4.8). A first observation is that $R_{i}<R_{j}$ implies $C_{i}>C_{j}$, and $R_{i}=R_{j}$ implies $C_{i}=C_{j}$, where $i, j \neq 0$. The next lemma is similar to [GŠV15, Lemma 7.6].

Lemma 4.15. Let $\left(R_{0}, R_{1}, \cdots, R_{q}, C_{0}, C_{1}, \cdots, C_{q}\right)$ be a positive fixpoint of (4.8). Then the number of distinct values in $\left\{R_{i}\right\}_{1 \leq i \leq q}$ and $\left\{C_{i}\right\}_{1 \leq i \leq q}$ is at most 3 .

Proof. Let $R:=\sum_{i=1}^{q} R_{i}$ and $C:=\sum_{i=1}^{q} C_{i}$. Suppose all variables are normalized so that $R_{0}+R=C_{0}+C=1$. Then for any $i \in[q]$, we have that

$$
\begin{aligned}
\frac{R_{i}}{R_{0}} & =t^{-d}\left(\frac{(t-1) C_{0}+1-C_{i}}{(t-1) C_{0}+1}\right)^{d} \\
& =t^{-d}\left(\frac{(t-1)+C_{0}^{-1}-C_{i} / C_{0}}{(t-1)+C_{0}^{-1}}\right)^{d}=t^{-d}\left(1-\frac{1}{t^{d} C^{\prime}}\left(1-\frac{R_{i}}{R_{0} R^{2}}\right)^{d}\right)^{d},
\end{aligned}
$$

where $C^{\prime}=(t-1)+C_{0}^{-1}$ and $R^{\prime}=(t-1)+R_{0}^{-1}$. Let $x=\left(R_{i} / R_{0}\right)^{1 / d}$ and note that $x \in[0,1]$. Then the above equation becomes $f(x)=0$, where $f(x):=t^{-1}\left(1-\frac{1}{t^{d} C^{\prime}}(1-\right.$

[^7]$\left.\left.\frac{x^{d}}{R^{\prime}}\right)^{d}\right)-x$. We have that
$$
f^{\prime}(x):=(g(x))^{d-1}-1, \text { where } g(x):=\left(\frac{d^{2}}{t^{d+1} R^{\prime} C^{\prime}}\right)^{1 /(d-1)}\left(1-\frac{x^{d}}{R^{\prime}}\right) x .
$$

Note that $g(x)>0$ on the interval $[0,1]$ because $x^{d}=\frac{R_{i}}{R_{0}}<(t-1)+R_{0}^{-1}=R^{\prime}$. Using that $(g(x))^{d-1}-1=(g(x)-1)\left(g(x)^{d-2}+\ldots+1\right)$, we therefore obtain that the roots of $f^{\prime}(x)=0$ can only come from the roots of $g(x)-1$, which has at most two roots by the Descartes' rule of signs. Hence $f^{\prime}(x)$ changes its sign at most twice in the interval of $[0,1]$ and $f(x)$ has at most 3 roots over $[0,1]$, showing that the $R_{i}$ 's for $i \in[q]$ can only be supported on three different values. The statement for the $C_{i}$ 's follows by an analogous argument.

The above lemma motivates the following definition.
Definition 4.16. Let ( $R_{0}, R_{1}, \cdots, R_{q}, C_{0}, C_{1}, \cdots, C_{q}$ ) be a positive fixpoint. We call the fixpoint $m$-supported, if the number of distinct values in $\left\{R_{i}\right\}_{1 \leq i \leq q}$ is $m$, where $m \in\{1,2,3\}$. We call the fixpoint is of type $\left(q_{1}, q_{2}, q_{3}\right)$ where $q_{1}+q_{2}+q_{3}=q$, if the multiplicities of different numbers in $\left\{R_{i}\right\}_{1 \leq i \leq q}$ are $q_{1}, q_{2}, q_{3}$ respectively. ${ }^{5}$ In case that the fixpoint is 2 or 1 -supported, let one or two of $q_{i}$ 's take zero respectively.

From now on we may also abuse the notation $R_{i}$ (also $C_{i}, i=1,2,3$ ) by absorbing all the same values, and hence $R_{1}$ stands for the value that $q_{1}$ of $R$ 's (except $R_{0}$ ) take, rather than the value of $R$ on the first index in the fixpoint.

The main lemma of this section can be stated as follows.
Lemma 4.17. Suppose $q \geq 4$ is even, $k^{\prime} \geq 2$ and $d=5 q^{k^{\prime}}$. The maximum of $\Psi_{1}$ over $\left(q_{1}, q_{2}, q_{3}\right)$-type fixpoints is attained uniquely, when $\left(q_{1}, q_{2}, q_{3}\right)=(q / 2, q / 2,0)$.

We also need to prove that 2-maximal triples ( $q / 2, q / 2,0$ ) yield unique $\mathbf{r}$ and c (up to scaling and permutation), and that the corresponding maxima are Hessian dominant.

Lemma 4.18. Suppose $q \geq 4$ is even, $k^{\prime} \geq 2$ and $d \geq 5 q^{k^{\prime}}$. The fixpoints of type $(q / 2, q / 2,0)$ are unique up to scaling and permutation symmetric. In addition, they are Hessian dominant maxima of $\Psi_{1}$.

Lemma 4.13 follows immediately by combining Lemmata 4.17 and 4.18.
To illustrate how fixpoints look like, we provide an example below for references.

[^8]Example 4.19. Consider the case $q=6, k^{\prime}=3$ and $d=5 q^{k^{\prime}}=1080$. The following three pairs are fixpoints of the system Equation (4.8) under this parameter setting.
(a) $\left\{\begin{array}{l}\left(R_{0}, \cdots, R_{6}\right)=(0.9863,0.0045,0.0045,0.0045,0.0001,0.0001,0.0001) ; \\ \left(C_{0}, \cdots, C_{6}\right)=(0.9863,0.0001,0.0001,0.0001,0.0045,0.0045,0.0045) .\end{array}\right.$
(b)

$$
\begin{aligned}
& \text { (b) } \quad\left\{\begin{array}{l}
\left(R_{0}, \cdots, R_{6}\right)=(0.993,0.001,0.001,0.001,0.001,0.001,0.001) ; \\
\left(C_{0}, \cdots, C_{6}\right)=(0.993,0.001,0.001,0.001,0.001,0.001,0.001)
\end{array}\right. \\
& \text { (c) } \quad\left\{\begin{array}{l}
\left(R_{0}, \cdots, R_{6}\right)=(0.9997,0.0001,0.0001,0.0001,0.0001,0.0001,0.0001) ; \\
\left(C_{0}, \cdots, C_{6}\right)=(0.9732,0.0045,0.0045,0.0045,0.0045,0.0045,0.0045) .
\end{array}\right.
\end{aligned}
$$

By definition, (a) is of type ( $3,3,0$ ), while (b)(c) are of type $(6,0,0)$.

### 4.3.1 Restricting to three values

In order to prove Lemma 4.17, we need to determine which type of fixpoints maximizes $\Psi_{1}$. By using Proposition 4.14, the value of $\Psi_{1}$ corresponding to such a fixpoint in (4.8) can be given by the matrix norm (4.7), which can be seen to be equal to

$$
\begin{align*}
& \overline{\Phi^{S}}(\mathbf{q}, \mathbf{r}, \mathbf{c}):= \\
& \qquad \begin{array}{l}
(d+1) \ln \left(R_{0} C_{0} t^{2}+\left(\sum_{i=1}^{3} q_{i} R_{i}\right) C_{0} t+\left(\sum_{i=1}^{3} q_{i} C_{i}\right) R_{0} t\right. \\
\left.\quad+\left(\sum_{i=1}^{3} q_{i} R_{i}\right)\left(\sum_{i=1}^{3} q_{i} C_{i}\right)-\left(\sum_{i=1}^{3} q_{i} R_{i} C_{i}\right)\right) \\
-d \ln \left(R_{0}^{(d+1) / d}+\sum_{i=1}^{3} q_{i} R_{i}^{(d+1) / d}\right)-d \ln \left(C_{0}^{(d+1) / d}+\sum_{i=1}^{3} q_{i} C_{i}^{(d+1) / d}\right) .
\end{array}
\end{align*}
$$

Where we define the vector $\mathbf{r}=\left(R_{0}, R_{1}, R_{2}, R_{3}\right)$ and $\mathbf{c}=\left(C_{0}, C_{1}, C_{2}, C_{3}\right)$. For instance, Example 4.19(a) can be written as $\mathbf{q}=(3,3,0), \mathbf{r}=(0.9863,0.0045,0.0001$, arbitrary $)$, and $\mathbf{c}=\left(0.9863,0.0001,0.0045\right.$, arbitrary). ${ }^{6}$ It is worth noting that this function is scale-free with respect to $\mathbf{r}$ and $\mathbf{c}$, as this property will be used intensively in our later proofs.

The discrete optimization of (4.9) over all fixpoints of the tree recursion (4.8) is difficult to cope with. Instead, we then try to maximize (4.9) over all nonnegative $\mathbf{q}$ and $\sum_{i=1}^{3} q_{i}=q$, wishing the maximum to be taken at integer $\mathbf{q}$. This is the main reason such approach can only deal with even $q$.

[^9]For all $\mathbf{q}=\left(q_{1}, q_{2}, q_{3}\right)$ with $q_{1}+q_{2}+q_{3}=q, q_{i} \geq 0$, define

$$
\begin{equation*}
\bar{\Phi}(\mathbf{q}):=\max _{\mathbf{r}, \mathbf{c}} \overline{\Phi^{S}}(\mathbf{q}, \mathbf{r}, \mathbf{c}) \tag{4.10}
\end{equation*}
$$

where the maximum is taken over $\mathbf{r}=\left(R_{0}, R_{1}, R_{2}, R_{3}\right), \mathbf{c}=\left(C_{0}, C_{1}, C_{2}, C_{3}\right)$ satisfying

$$
\begin{gather*}
R_{0} C_{0} t^{2}+\left(\sum_{i=1}^{3} q_{i} R_{i}\right) C_{0} t+\left(\sum_{i=1}^{3} q_{i} C_{i}\right) R_{0} t+\left(\sum_{i=1}^{3} q_{i} R_{i}\right)\left(\sum_{i=1}^{3} q_{i} C_{i}\right)-\left(\sum_{i=1}^{3} q_{i} R_{i} C_{i}\right)>0, \\
R_{i}, C_{i} \geq 0, i=0,1,2,3 . \tag{4.11}
\end{gather*}
$$

Our first step is to verify the maximum in (4.10) is well defined, and moreover, the maximum in $\max _{\mathbf{q}} \bar{\Phi}(\mathbf{q})$ can also be taken. This is formalized by the next lemma.

Lemma 4.20 ([GŠV15, Lemma 7.10]). The maximum in (4.10) is well-defined. In addition, $\max _{\mathbf{q}} \bar{\Phi}(\mathbf{q})$ can be attained in the region where $q_{1}+q_{2}+q_{3}=q, q_{i} \geq 0$.

Proof. The argument is verbatim the same as in [GŠV15], the only difference is that the function has slightly different form, but still accounts for the relevant parameters $q_{1}, q_{2}, q_{3}$.

The next trouble we may encounter later is that we are now dealing with all possible $\mathbf{r}, \mathbf{c}$ conditioned on (4.11), instead of just fixpoints of (4.8). The good news is that, in contrast to [GŠV15], we can rule out fairly easily that the maximizer in (4.10) is at the boundary.

Lemma 4.21. For all triples $\boldsymbol{q}=\left(q_{1}, q_{2}, q_{3}\right)$, any maximizer in(4.10) satisfies (a) $R_{0}, C_{0}>$ 0 , (b) for anyi such that $q_{i}>0$, it holds that $R_{i}, C_{i}>0$, and (c) for distinct $i, j$ such that $q_{i}, q_{j}>0$, it holds that $R_{i}=R_{j}$ if and only if $C_{i}=C_{j}$.

The problem in [GŠV15] that also appears in our setting is that it might be that $q_{i}, q_{j}>0$, but $R_{i}=R_{j}$ and $C_{i}=C_{j}$. For example, imagine we are now strengthening the restriction (4.11) by adding $R_{1}=R_{2}$ and $C_{1}=C_{2}$. Then $\bar{\Phi}\left(q_{1}+q_{2}, 0, q_{3}\right) \leq$ $\bar{\Phi}\left(q_{1}, q_{2}, q_{3}\right)$. Such degenerate case makes it difficult to compare between different $\mathbf{q}$ triples because the equality can be taken. This motivates the next definition.

Definition 4.22. Let $m=2,3$. A triple $\mathbf{q}$ is called $m$-maximal, if exactly $m q_{i}$ 's in $\mathbf{q}$ are non-zero, and there exists $\mathbf{r}, \mathbf{c}$ maximizing (4.10) such that, $q_{i}, q_{j}>0$ and $i \neq j$ imply that $R_{i} \neq R_{j}$ and $C_{i} \neq C_{j}$. We also call $\mathbf{q}$ maximal if it is either 2- or 3-maximal.

Now we connect $m$-maximal triples with fixpoints in (4.8).

Lemma 4.23. Suppose a triple $\mathbf{q}$ is m-maximal. Then there exists $\mathbf{r}, \mathbf{c}$ achieving the maximum in (4.10) and specifying an m-supported fixpoint of tree recursion (4.8) of type $\mathbf{q}$.

For 2 and 3-maximal triples, the key is the next lemma.
Lemma 4.24. Suppose $q \geq 4$ is even. Then the following statements hold:
(a) There does not exist any 3-maximal triple that maximizes (4.10).
(b) The only possibility of a 2-maximal triple to maximize (4.10) is $(q / 2, q / 2,0)$ or its permutations, with $R_{i} / R_{j}=C_{j} / C_{i}$, where $i \neq j$ are the two indices such that $q_{i}, q_{j}=q / 2$.

The above lemma is not yet enough to finish the proof of Lemma 4.17 because we have to rule out degenerate cases of all triples, i.e., the triple ( $q, 0,0$ ). This is the main difference with the colour-symmetric setting of [GŠV15]. Instead, we have the special colour corresponding to ( $R_{0}, C_{0}$ ), which makes the system behave like a 2 -spin system when all "pure" colours take the same fraction. What is worse is that, it is possible for the 2 -spin system to have three fixpoints (two of them being symmetric), when the tree recursion lies in the so-called "non-uniqueness" region (see Section 4.3.4). Therefore, we need to discuss such fixpoints by two different cases.

Before continuing the discussion, let us state another useful result from [GŠV15]. A fixpoint $x$ of a mapping $f$ is facobian stable if the Jacobian of $f$ at $x$ has spectral radius less than 1 .

Proposition 4.25 ([GŠV15, Theorem 4.2]). A fixpoint of the tree recursion (4.8) is 7acobian stable if and only if it corresponds to a Hessian dominant local maximum of $\Psi_{1}$.

The first kind of fixpoints satisfy $R_{0} / R_{1} \neq C_{0} / C_{1}$. The fixpoint Example 4.19(c) is one such. As stated in the next lemma, such a fixpoint is Jacobian stable, and hence it is a possible candidate to be the maximizer in $\max _{\mathbf{q}} \bar{\Phi}(\mathbf{q})$. Though the proof of stability is not necessary for our main theorem, we still leave it here for future references.

Lemma 4.26. Suppose $d \geq 5 q^{k^{\prime}}$. The fixpoint corresponding to triple $(q, 0,0)$ and $R_{0} / R_{1} \neq C_{0} / C_{1}$ is unique up to scaling and swapping $R$ and $C$. Moreover, it is facobian stable.

For the reason above, we can only go through a very detailed calculation to rule out this case. The equality in $d=5 q^{k^{\prime}}$ from the next lemma is for the sake of simplification in calculation.

Lemma 4.27. Suppose $d=5 q^{k^{\prime}}$. Any fixpoint corresponding to triple $(q, 0,0)$ and $R_{0} / R_{1} \neq C_{0} / C_{1}$ does not maximize (4.10).

On the other hand, when $R_{0} / R_{1}=C_{0} / C_{1}$, things become easier as such fixpoints are not Jacobian stable. Thus, by Proposition 4.25, these fixpoints do not correspond to local maxima of $\Psi_{1}$.

Lemma 4.28. Suppose $d \geq 5 q^{k^{\prime}}$. Any fixpoint corresponding to triple $(q, 0,0)$ and $R_{0} / R_{1}=C_{0} / C_{1}$ is facobian unstable.

The fixpoint Example 4.19(b) is one such example.
Now we are ready to prove Lemma 4.17, which given the above ingredients can be done by following closely a related argument in [GŠV15]. The main complicacy in the proof is that when we find a maximizer $\mathbf{q}$ of $\bar{\Phi}(\mathbf{q})$, the corresponding $\mathbf{r}$ (or $\mathbf{c}$ ) values are not necessarily distinct. We need to carefully rule out these degenerate cases.

Proof of Lemma 4.17. Denote MAX $:=\max _{\mathbf{q}} \bar{\Phi}(\mathbf{q})$. We first claim that $M A X$ is attained at $\hat{\mathbf{q}}=(q / 2, q / 2,0)$, and $\hat{\mathbf{q}}$ is maximal. Assuming the claim, Lemma 4.23 yields that there exist $\hat{\mathbf{r}}, \hat{\mathbf{c}}$ with $\bar{\Phi}(\hat{\mathbf{q}})=\overline{\Phi^{S}}(\hat{\mathbf{q}}, \hat{\mathbf{r}}, \hat{\mathbf{c}})$, specifying a ( $q / 2, q / 2,0$ )-type fixpoint of the tree recursion (4.8). Hence $M A X=\max \Psi_{1}$. To show that $\hat{\mathbf{q}}$ is the unique type of fixpoint achieving the maximum of $\Psi_{1}$, consider an arbitrary $\mathbf{q}^{*}$-type fixpoint achieving the maximum of $\Psi_{1}$, say $\left(\mathbf{r}^{*}, \mathbf{c}^{*}\right)$. Then $\mathbf{q}^{*}$ must also achieve the maximum in $\max _{\mathbf{q}} \bar{\Phi}(\mathbf{q})$. By Lemma 4.27, $\mathbf{q}^{*} \neq(q, 0,0)$ and hence it is maximal according to Definition 4.22 (using $\left(\mathbf{r}^{*}, \mathbf{c}^{*}\right)$ as the maximizers; Recall Definition 4.16 that $R_{i} \neq R_{j}, C_{i} \neq C_{j}$ for $i \neq j$ and $\left.q_{i}, q_{j}>0\right)$. Therefore we can apply Lemma 4.24 and obtain that $\mathbf{q}^{*}=\hat{\mathbf{q}}$.

It remains to prove the claim above. Let $\mathbf{q}^{*}$ be any maximizer of $\max _{\mathbf{q}} \bar{\Phi}(\mathbf{q})$.
(1) $\mathbf{q}^{*}$ has at least two positive entries. This is a consequence of Lemmata 4.27 and 4.28 (after using Proposition 4.25).
(2) In case $\mathbf{q}^{*}$ has exactly two positive entries, then $\mathbf{q}^{*}$ must be maximal. Otherwise, suppose $\mathbf{q}^{*}=\left(q_{1}, q_{2}, 0\right)$ and the maximizer in (4.10) is achieved at $\mathbf{r}^{*}$, $\mathbf{c}^{*}$ where $R_{1}=R_{2}$ or $C_{1}=C_{2}$. By Lemma 4.21 (c), both equalities are true and hence $\bar{\Phi}\left(\mathbf{q}^{*}\right)=\bar{\Phi}((q, 0,0))$, contradicting item (1).
(3) In case $\mathbf{q}^{*}$ has exactly two positive entries, it must holds that $\mathbf{q}^{*}=\hat{\mathbf{q}}$. This is from item (2), and Lemma 4.24 (b).
(4) If $\mathbf{q}^{*}$ has all positive entries, then it cannot be 3-maximal. This is from Lemma 4.24 (a).
(5) If $\mathbf{q}^{*}$ has all positive entries, then $\bar{\Phi}\left(\mathbf{q}^{*}\right)=\bar{\Phi}(\hat{\mathbf{q}})$. This can be proved by the following argument. Let $\mathbf{r}^{*}, \mathbf{c}^{*}$ be the maximizer corresponding to $\mathbf{q}^{*}$. By item (4), $\mathbf{q}^{*}$ is not 3-maximal, and using the argument of item (2), there exist distinct $i, j \geq 1$ such that $R_{i}=R_{j}$ and $C_{i}=C_{j}$ in $\mathbf{r}^{*}, \mathbf{c}^{*}$. Let $k \geq 1$ be the remaining index.

- If $R_{i}=R_{j}=R_{k}$, then by Lemma 4.21 (c), $C_{i}=C_{j}=C_{k}$, and hence $\bar{\Phi}\left(\mathbf{q}^{*}\right)=$ $\bar{\Phi}(q, 0,0)$, contradicting item (1).
- If $C_{i}=C_{j}=C_{k}$, then by Lemma 4.21 (c), $R_{i}=R_{j}=R_{k}$, and hence $\bar{\Phi}\left(\mathbf{q}^{*}\right)=$ $\bar{\Phi}(q, 0,0)$, contradicting item (1).
- If $R_{i} \neq R_{k}$ and $C_{i} \neq C_{k}$, we can "merge" the indices $i, j$ to get a new triple $\mathbf{q}^{\prime}:=\left(q_{i}+q_{j}, q_{k}, 0\right)$. Let $\boldsymbol{r}^{\prime}:=\left(R_{0}, R_{i}, R_{k}, 0\right), \boldsymbol{c}^{\prime}:=\left(R_{0}, C_{i}, C_{k}, 0\right)$. Then

$$
\bar{\Phi}\left(\mathbf{q}^{*}\right)=\overline{\Phi^{S}}\left(\mathbf{q}^{*}, \mathbf{r}^{*}, \mathbf{c}^{*}\right)=\overline{\Phi^{S}}\left(\mathbf{q}^{\prime}, \mathbf{r}^{\prime}, \mathbf{c}^{\prime}\right) \leq \bar{\Phi}\left(\mathbf{q}^{\prime}\right)
$$

This means that $\mathbf{q}^{\prime}$ is also a maximizer of $\max _{\mathbf{q}} \bar{\Phi}(\mathbf{q})$ since $\mathbf{q}^{*}$ is a maximizer. However, $\mathbf{q}^{\prime}$ has exactly two positive entries. Hence by item (3), $\bar{\Phi}\left(\mathbf{q}^{*}\right)=\bar{\Phi}\left(\mathbf{q}^{\prime}\right)=\bar{\Phi}(\hat{\mathbf{q}})$.

The above arguments imply that for any maximizer $\mathbf{q}^{*}$, it holds that $\bar{\Phi}\left(\mathbf{q}^{*}\right)=\bar{\Phi}(\hat{\mathbf{q}})$, which means that $\hat{\mathbf{q}}$ is indeed a maximizer. This also indicates all items above apply to $\mathbf{q}^{*}=\hat{\mathbf{q}}$, and from item (3), we obtain that $\hat{\mathbf{q}}$ is 2-maximal. This concludes the proof.

Before diving into the proofs of all the lemmata above, we want to mention the following observation. The partial derivatives $\partial \overline{\Phi^{S}} / \partial q_{i}$, conditioned on $\mathbf{r}$ and $\mathbf{c}$ achieving the maximum in (4.10), can be written as follows. (Note that it applies to all triples $\mathbf{q}$, including non-maximal ones.) Based on these partial derivatives, we can argue the non-optimality by perturbing $q_{i}$ 's.

Lemma 4.29. Suppose $\mathbf{r}, \mathbf{c}$ achieve the maximum in (4.10). Then for any $i \in\{1,2,3\}$ such that $q_{i}>0$, it holds that

$$
\frac{\partial \overline{\Phi^{S}}}{\partial q_{i}}=\frac{R_{i} C_{0} t+R_{0} C_{i} t+(d-1) R_{i} C_{i}+R_{i}\left(\sum_{j=1}^{3} C_{j} q_{j}\right)+C_{i}\left(\sum_{j=1}^{3} R_{j} q_{j}\right)}{R_{0} C_{0} t^{2}+\left(\sum_{j=1}^{3} C_{j} q_{j}\right) R_{0} t+\left(\sum_{j=1}^{3} R_{j} q_{j}\right) C_{0} t+\left(\sum_{j=1}^{3} R_{j} q_{j}\right)\left(\sum_{j=1}^{3} C_{j} q_{j}\right)-\left(\sum_{j=1}^{3} R_{j} C_{j} q_{j}\right)} .
$$

Moreover, if there exists $i, j$ such that $q_{i}, q_{j}>0$ and $i \neq j$ and satisfies $\partial \overline{\Phi^{S}} / \partial q_{i}-$ $\partial \overline{\Phi^{S}} / \partial q_{j} \neq 0$, then the maximum in (4.10) is not achieved.

Unproved propositions and lemmata in this subsection can be found later:

- Lemma 4.24 is proved in Section 4.3.2.
- Lemma 4.18 is proved in Section 4.3.3.
- Lemma 4.26 and Lemma 4.28 are proved in Section 4.3.4.
- Lemma 4.27 is proved in Section 4.3.5.
- Lemma 4.23 and Lemma 4.29 are proved in Section 4.4.1.
- Lemma 4.21 is proved in Section 4.4.2.


### 4.3.2 2,3-maximal triples

Let $\mathbf{q}$ be a maximal triple and let $I=\left\{i \mid q_{i}>0\right\}$. From Lemma 4.21 (a) and (b), by taking partial derivatives of $\overline{\Phi^{S}}$ with respect to non-zero $R_{i}$ and $C_{i}$ 's and setting them to 0 , we get that the maximizer of $\overline{\Phi^{S}}$ satisfies

$$
\begin{array}{ll}
R_{0}^{1 / d} \propto C_{0} t^{2}+\left(q_{1} C_{1}+q_{2} C_{2}+q_{3} C_{3}\right) t, & R_{i}^{1 / d} \propto C_{0} t+q_{1} C_{1}+q_{2} C_{2}+q_{3} C_{3}-C_{i} ; \\
C_{0}^{1 / d} \propto R_{0} t^{2}+\left(q_{1} R_{1}+q_{2} R_{2}+q_{3} R_{3}\right) t, & C_{i}^{1 / d} \propto R_{0} t+q_{1} R_{1}+q_{2} R_{2}+q_{3} R_{3}-R_{i} \tag{4.13}
\end{array}
$$

for $i \in I$.
First assume $\mathbf{q}$ is 3-maximal, for any $i \neq j$ it holds that $R_{i} \neq R_{j}$ and $C_{i} \neq C_{j}$. From Lemma 4.21 (a) and (b), we may assume the following strict ordering

$$
\begin{equation*}
R_{1}>R_{2}>R_{3}>0 \text { and } 0<C_{1}<C_{2}<C_{3} . \tag{4.14}
\end{equation*}
$$

Lemma 4.30. Suppose $R_{i}$ 's and $C_{i}$ 's satisfy (4.12), (4.13) and (4.14). We have the following:
(a) If $R_{1} / R_{3} \neq C_{3} / C_{1}$, then $\partial \overline{\Phi^{S}} / \partial q_{1}-\partial \overline{\Phi^{S}} / \partial q_{3} \neq 0$.
(b) If $R_{1} / R_{3}=C_{3} / C_{1}$, then $\partial \overline{\Phi^{S}} / \partial q_{1}-\partial \overline{\Phi^{S}} / \partial q_{2} \neq 0$.

For the sake of convenience, we further set

$$
r_{0}^{d}:=R_{0} / R_{3}, r_{1}^{d}:=R_{1} / R_{3}, r_{2}^{d}:=R_{2} / R_{3}, \text { and } c_{0}^{d}:=C_{0} / C_{1}, c_{2}^{d}:=C_{2} / C_{1}, c_{3}^{d}:=C_{3} / C_{1} .
$$

which means

$$
\begin{equation*}
r_{1}>r_{2}>1 \text { and } c_{3}>c_{2}>1 \tag{4.15}
\end{equation*}
$$

We will need these notations in later sections too. With them, from (4.12) and (4.13), we obtain that

$$
\begin{array}{ll}
r_{0}=\frac{c_{0}^{d} t^{2}+\left(q_{1}+q_{2} c_{2}^{d}+q_{3} c_{3}^{d}\right) t}{c_{0}^{d} t+q_{1}+q_{2} c_{2}^{d}+\left(q_{3}-1\right) c_{3}^{d}}, & c_{0}=\frac{r_{0}^{d} t^{2}+\left(q_{1} r_{1}^{d}+q_{2} r_{2}^{d}+q_{3}\right) t}{r_{0}^{d} t+\left(q_{1}-1\right) r_{1}^{d}+q_{2} r_{2}^{d}+q_{3}} . \\
r_{1}=\frac{c_{0}^{d} t+q_{1}-1+q_{2} c_{2}^{d}+q_{3} c_{3}^{d}}{c_{0}^{d} t+q_{1}+q_{2} c_{2}^{d}+\left(q_{3}-1\right) c_{3}^{d}}, & c_{3}=\frac{r_{0}^{d} t+q_{1} r_{1}^{d}+q_{2} r_{2}^{d}+q_{3}-1}{r_{0}^{d} t+\left(q_{1}-1\right) r_{1}^{d}+q_{2} r_{2}^{d}+q_{3}}, \tag{4.17}
\end{array}
$$

Proof of Lemma 4.30. From (4.12) and (4.13), we get

$$
\begin{equation*}
\frac{r_{1}-1}{r_{2}-1}=\frac{c_{3}^{d}-1}{c_{3}^{d}-c_{2}^{d}}, \text { yielding that } r_{2}=\frac{r_{1} c_{3}^{d}-1-c_{2}^{d}\left(r_{1}-1\right)}{c_{3}^{d}-1} \tag{4.18}
\end{equation*}
$$

Similarly, we obtain that

$$
\begin{equation*}
\frac{c_{3}-1}{c_{2}-1}=\frac{r_{1}^{d}-1}{r_{1}^{d}-r_{2}^{d}}, \text { yielding that } r_{2}^{d}=\frac{r_{1}^{d} c_{3}-1-c_{2}\left(r_{1}^{d}-1\right)}{c_{3}-1}, \tag{4.19}
\end{equation*}
$$

From (4.12) and (4.13), we have that $r_{2}=\frac{c_{0}^{d} t+q_{1}+\left(q_{2}-1\right) c_{2}^{d}+q_{3} c_{3}^{d}}{c_{0}^{d}+q_{1}+q_{2} c_{2}^{d}+\left(q_{3}-1\right) c_{3}^{d}}$ which combined with (4.18) gives that

$$
\begin{equation*}
\frac{c_{0}^{d} t+q_{1}+\left(q_{2}-1\right) c_{2}^{d}+q_{3} c_{3}^{d}}{c_{0}^{d} t+q_{1}+q_{2} c_{2}^{d}+\left(q_{3}-1\right) c_{3}^{d}}=\frac{r_{1} c_{3}^{d}-1-c_{2}^{d}\left(r_{1}-1\right)}{c_{3}^{d}-1} \tag{4.20}
\end{equation*}
$$

Symmetrically we obtain that

$$
\begin{equation*}
\frac{r_{0}^{d} t+q_{1} r_{1}^{d}+\left(q_{2}-1\right) c_{2}^{d}+q_{3}}{r_{0}^{d} t+\left(q_{1}-1\right) r_{1}^{d}+q_{2} c_{2}^{d}+q_{3}}=\frac{r_{1}^{d} c_{3}-1-r_{2}^{d}\left(c_{3}-1\right)}{r_{1}^{d}-1} \tag{4.21}
\end{equation*}
$$

We can view (4.20) and (4.21) as a linear system in $q_{1}$ and $q_{3}$ after clearing the denominators, which yields that

$$
\begin{align*}
& q_{1} \cdot\left(r_{1}^{d} c_{3}^{d}-1\right)=c_{0}^{d} t+q_{2} c_{2}^{d}+\frac{1-r_{1} c_{3}^{d}}{r_{1}-1}-c_{3}^{d}\left(r_{0}^{d} t+q_{2} r_{2}^{d}+\frac{1-c_{3} r_{1}^{d}}{c_{3}-1}\right),  \tag{4.22}\\
& q_{3} \cdot\left(r_{1}^{d} c_{3}^{d}-1\right)=r_{0}^{d} t+q_{2} r_{2}^{d}+\frac{1-r_{1}^{d} c_{3}}{c_{3}-1}-r_{1}^{d}\left(c_{0}^{d} t+q_{2} c_{2}^{d}+\frac{1-r_{1} c_{3}^{d}}{r_{1}-1}\right), \tag{4.23}
\end{align*}
$$

From (4.12) and (4.13), we also obtain that

$$
\begin{equation*}
r_{0}^{d} t=\frac{r_{1}^{d}-1}{c_{3}-1}-\left(q_{1}-1\right) r_{1}^{d}-q_{2} r_{2}^{d}-q_{3}, \quad c_{0}^{d} t=\frac{c_{3}^{d}-1}{r_{1}-1}-q_{1}-q_{2} c_{2}^{d}-\left(q_{3}-1\right) c_{3}^{d} . \tag{4.24}
\end{equation*}
$$

We can now show the following:

$$
\begin{equation*}
\text { if } r_{1}=c_{3} \text {, then (i) } r_{2}=c_{2} \text {, (ii) } q_{1}=q_{3} \text {, and (iii) } r_{0}=c_{0} . \tag{4.25}
\end{equation*}
$$

The proof of (i) in (4.25) is identical to that in [GŠV15, Lemma 7.20] using the expressions for $r_{2}, r_{2}^{d}$ in (4.18) and (4.19), respectively. From (i) and the assumption that $r_{1}=c_{3}$, we obtain from (4.22) and (4.23) that

$$
\begin{equation*}
q_{3}-q_{1}=\left(r_{0}^{d}-c_{0}^{d}\right) \cdot \frac{t}{r_{1}^{d}-1} \tag{4.26}
\end{equation*}
$$

Furthermore, the equations in (4.16) can also be regarded as a linear system in $q_{1}$ and $q_{3}$. Using the assumption $r_{1}=c_{3}$ and $r_{2}=c_{2}$, we obtain that

$$
q_{3}-q_{1}=\frac{t\left[\left(r_{0}^{d}-c_{0}^{d}\right) t^{2}+\left(r_{0} c_{0}^{d}-c_{0} r_{0}^{d}+c_{0}^{1+d}-r_{0}^{1+d}\right) t+\left(r_{0}^{d}-c_{0}^{d}\right) r_{0} c_{0}-\left(r_{0}-c_{0}\right) r_{1}^{d}\right]}{\left(r_{0}-t\right)\left(c_{0}-t\right)\left(r_{1}^{d}-1\right)} .
$$

which, together with (4.26), implies $r_{0}=c_{0}$ and hence $q_{1}=q_{3}$. This finishes proving (4.25).

We are now ready to give the proof of Lemma 4.30. For part (a), Lemma 4.29 yields that

$$
\begin{aligned}
& \frac{\partial \overline{\Phi^{S}}}{\partial q_{1}}-\frac{\partial \overline{\Phi^{S}}}{\partial q_{3}}=\frac{1}{S} \times\left[\left(r_{1}^{d}-1\right)\left(c_{0}^{d} t+q_{1}+c_{2}^{d} q_{2}\right)-\left(c_{3}^{d}-1\right)\left(r_{0}^{d} t+q_{3}+r_{2}^{d} q_{2}\right)\right. \\
&\left.+(d-1)\left(r_{1}^{d}-c_{3}^{d}\right)+r_{1}^{d} c_{3}^{d}\left(q_{3}-q_{1}\right)+r_{1}^{d} q_{1}-c_{3}^{d} q_{3}\right]
\end{aligned}
$$

where $S>0$. Then plug in the expression of $q_{1}$ and $q_{3}$ in (4.22) and (4.23), we get

$$
\begin{equation*}
\frac{\partial \overline{\Phi^{S}}}{\partial q_{1}}-\frac{\partial \overline{\Phi^{S}}}{\partial q_{3}}=-\frac{g\left(r_{1}, c_{3}\right)}{S} \tag{4.27}
\end{equation*}
$$

where $g\left(r_{1}, c_{3}\right):=\left(r_{1}-c_{3}\right)\left(r_{1}^{d}-1\right)\left(c_{3}^{d}-1\right)-d\left(r_{1}-1\right)\left(c_{3}-1\right)\left(r_{1}^{d}-c_{3}^{d}\right)$. This quantity was shown to have the same sign as $r_{1}-c_{3}$ (see Equation (123) in the proof of Lemma 7.19 in [GŠV15]), and specifically, non-zero when $r_{1} \neq c_{3}$, concluding part (a).

Now we prove part (b) of Lemma 4.30. From (4.25), the assumption $r_{1}=c_{3}$ implies $r_{2}=c_{2}, q_{1}=q_{3}$ and $r_{0}=c_{0}$. Applying Lemma 4.29 based on these, we get

$$
\frac{\partial \overline{\Phi^{S}}}{\partial q_{1}}-\frac{\partial \overline{\Phi^{S}}}{\partial q_{2}}
$$

$=q_{1}\left(1+r_{1}^{d}\right)\left(1+r_{1}^{d}-2 r_{2}^{d}\right)+r_{2}^{d}\left(q_{2}-\left(2 q_{2}+d-1\right) r_{2}^{d}-2 r_{0}^{d} t\right)+r_{0}^{d} t+r_{1}^{d}\left(r_{0}^{d} t+q_{2} r_{2}^{d}+d-1\right)$
$=-\frac{(d-1)\left(r_{1}-1\right) r_{2}^{2 d}+2\left(r_{1}^{d+1}-1\right) r_{2}^{d}-\left(r_{1}^{2 d+1}+d r_{1}^{d+1}-d r_{1}^{d}-1\right)}{r_{1}-1}$,
where in the second line we use (4.24). This quantity was shown to be non-zero in the proof of Lemma 7.19 in [GŠV15] (from Equation (124) onwards) under (4.18), concluding part (b).

Now we assume $\mathbf{q}=\left(q_{1}, q_{2}, q_{3}\right)$ is a 2-maximal triple, and assume $q_{2}=0$ without loss of generality. The result here is analogous to Lemma 4.30 (a).

Lemma 4.31. Under the assumption that $q_{2}=0$, suppose $R_{i}$ 's and $C_{i}$ 's $(i \neq 2)$ satisfy (4.12), (4.13) and (4.14). For any $q_{1}, q_{3}>0$, it holds that $\partial \overline{\Phi^{S}} / \partial q_{1}-\partial \overline{\Phi^{S}} / \partial q_{3} \neq 0$, unless $q_{1}=q_{3}$ and $R_{1} / R_{3}=C_{3} / C_{1}$.

Proof. First, note that the values of $R_{2}$ and $C_{2}$ do not affect the value of derivatives $\partial \overline{\Phi^{S}} / \partial q_{1}$ and $\partial \overline{\Phi^{S}} / \partial q_{3}$ when $q_{2}=0$. In addition, the expressions of $q_{1}$ and $q_{3}$ in (4.22) and (4.23) still hold for $q_{2}=0$. Therefore, one can carry out the proof of Lemma 4.30 (a) once again for this case, showing $\partial \overline{\Phi^{S}} / \partial q_{1}-\partial \overline{\Phi^{S}} / \partial q_{3}=0$ only when $R_{1} / R_{3}=C_{3} / C_{1}$. Assuming this, one can show $q_{1}=q_{3}$ by going through the proof of (4.25).

We conclude this subsection with Lemma 4.24.
Proof of Lemma 4.24. This comes after Lemma 4.30, Lemma 4.31 and the second part of Lemma 4.29.

### 4.3.3 Stability of maximal $(q / 2, q / 2,0)$ fixpoints

In the next two subsections, we focus on the (in)stability of candidate fixpoints that may maximize $\Psi_{1}$. The condition of Jacobian stability is given by the following Lemma.

Lemma 4.32 (cf. [GŠV15, Lemma 4.16]). Suppose ( $R_{0}, R_{1}, \cdots, R_{q}, C_{0}, C_{1}, \cdots, C_{q}$ ) is a fixpoint of the tree recursion (4.8). Let $\alpha_{i}:=\sum_{j=0}^{q} B_{i j} R_{i} C_{j}$ and $\beta_{j}:=\sum_{i=0}^{q} B_{i j} R_{i} C_{j}$. Define the matrix $\boldsymbol{A}:=\left(a_{i j}\right)_{0 \leq i, j \leq q}$ as $a_{i j}=B_{i j} R_{i} C_{j} / \sqrt{\alpha_{i} \beta_{j}}$, and the matrix $\boldsymbol{L}:=\left[\begin{array}{c}0 \\ \boldsymbol{A}^{\top} \\ 0\end{array}\right]$. Then $\boldsymbol{L}$ has symmetric real spectrum (symmetry means if $a$ is an eigenvalue then so is $-a$ ), and $\pm 1$ is a pair of its eigenvalues. The condition for the fixpoint to be stable is that the second largest eigenvalue of $\boldsymbol{L}$ is less than $1 / d$.

We will also need the following lemma which is proved in Section 4.4.3.
Lemma 4.33. For any $q \geq 4, k^{\prime} \geq 2$ and $d \geq 3 q^{k^{\prime}}$, the function

$$
h(x):=\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d} t^{d+1}-\frac{x^{d}-1}{x-1}+q^{\prime}+\left(q^{\prime}-1\right) x^{d}
$$

has exactly one root in the region $(1, \infty)$.
We are now ready to prove Lemma 4.18.

Proof of Lemma 4.18. Define $q^{\prime}:=q / 2$. We first prove the uniqueness of 2-maximal ( $q^{\prime}, q^{\prime}, 0$ ) fixpoint (up to scaling). According to the proof of Lemma 4.31, fixpoints of type $\left(q^{\prime}, q^{\prime}, 0\right)$ maximize $\bar{\Phi}$ only when $r_{1}=c_{3}$. Now denote $x:=r_{1}=c_{3}$. To prove the first part of this lemma, we show there exists exactly one possible $x>1$ when $d \geq 3 q^{k^{\prime}}$. By (4.16) and (4.17), we get

$$
\frac{r_{0} / t-r_{1}}{r_{1}-1}=\frac{1}{c_{3}^{d}-1}
$$

Combining this with (4.24), $x>1$ satisfies $h(x)=0$, where

$$
h(x):=\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d} t^{d+1}-\frac{x^{d}-1}{x-1}+q^{\prime}+\left(q^{\prime}-1\right) x^{d} .
$$

By Lemma 4.33, $h(x)$ has exactly one root $x>1$.
Next, we construct the matrices $\boldsymbol{A}$ and $\boldsymbol{L}$. Note that both matrices are scale-free with respect to $R_{i}$ and $C_{i}$. Directly plug in the formula in Lemma 4.32 to get

$$
\boldsymbol{A}:=\left[\begin{array}{ccc}
c^{2} & b c \mathbf{1}^{\mathrm{T}} & a c \mathbf{1}^{\mathrm{T}} \\
a c \mathbf{1} & a b \boldsymbol{J} & a^{2} \boldsymbol{J}^{\prime} \\
b c \mathbf{1} & b^{2} \boldsymbol{J}^{\prime} & a b \boldsymbol{J}
\end{array}\right]
$$

where $a:=\sqrt{x^{d-1} \frac{x-1}{x^{d}-1}}, b:=\sqrt{\frac{x-1}{x^{d}-1}}$ and $c:=\sqrt{\frac{x^{d+1}-1-q^{\prime}(x-1)\left(x^{d}+1\right)}{x^{d+1}-1}}, \boldsymbol{J}$ is the $q^{\prime} \times q^{\prime}$ matrix with zeros on the diagonal and ones elsewhere, $\boldsymbol{J}^{\prime}$ is the $q^{\prime} \times q^{\prime}$ matrix with ones everywhere, and $\mathbf{1}$ is the $q^{\prime} \times 1$ matrix. The eigenvalues of $\boldsymbol{L}=\left[\begin{array}{cc}0 & \boldsymbol{A} \\ \boldsymbol{A}^{\top} & 0\end{array}\right]$ consist of $\pm a b$ (each by multiplicity $q-2$ ) and $\pm \lambda_{1}, \pm \lambda_{2}, \pm \lambda_{3}$, where $\lambda_{1}, \lambda_{2}, \lambda_{3}$ are the zeros of the following cubic function

$$
f(z)=z^{3}-\left(q^{\prime} a^{2}+q^{\prime} b^{2}+c^{2}\right) z^{2}+\left(2 q^{\prime}-1\right) a^{2} b^{2} z+a^{2} b^{2} c^{2} .
$$

We claim that $a b$ is the second largest eigenvalue. To prove this, recall that 1 is the eigenvalue of $\boldsymbol{L}$. We can assume $\lambda_{1}=1$ (because $a b<1$, which means 1 must be among $\lambda_{1,2,3}$ ) and hence $f(1)=0$. In addition, $f(z)$ is monic and $f(0)>0$. This means it suffices to show $f(-a b) \leq 0$ and $f(a b) \leq 0$, which are true since

$$
f(a b)=-a^{2} b^{2} q(a-b)^{2}<0, \quad f(-a b)=-a^{2} b^{2} q(a+b)^{2}<0
$$

It remains to prove $a b=x^{(d-1) / 2} \frac{x-1}{x^{d}-1}<1 / d$ which follows from $\frac{x^{d}-1}{x-1}=x^{d-1}+\ldots+$ $1>d x^{(d-1) / 2}$, where the last inequality is an application of the AM-GM inequality when $x>1$.

### 4.3.4 (In)stability of $(q, 0,0)$ fixpoints

Set $x:=R_{0} / R_{1}$ and $y:=C_{0} / C_{1}$. Then by rewriting the tree recursion, one can see $x, y$ satisfies the system

$$
\begin{equation*}
x=t^{d}\left(\frac{t y+q}{t y+q-1}\right)^{d}, \quad y=t^{d}\left(\frac{t x+q}{t x+q-1}\right)^{d} . \tag{4.28}
\end{equation*}
$$

Before analysing the stability of the original $(q+1)$-spin system, we first need to study this 2 -spin system. By replacing $\beta:=t / q, \gamma=(q-1) / t$ and $\lambda=q^{d}$, the system above is actually the tree recursion of a general anti-ferromagnetic Ising model with parameter ( $\beta, \gamma, \lambda$ ). It follows that such system has either one solution ( $Q^{*}, Q^{*}$ ) (uniqueness) or three solutions $\left(Q^{+}, Q^{-}\right),\left(Q^{*}, Q^{*}\right),\left(Q^{-}, Q^{+}\right)$(non-uniqueness) where $Q^{+}>Q^{*}>Q^{-}$ (see [MSW07, Section 6.2] or [GŠV16, Theorem 7]). First and foremost, if $d \geq 5 q^{k^{\prime}}$, the system (4.28) is actually the latter case.

Lemma 4.34. When $q \geq 4, k \geq 2$ and $d \geq 5 q^{k^{\prime}}$, the system (4.28) lies in non-uniqueness region.

One way to prove Lemma 4.34 is to verify the non-uniqueness condition in [LLY13]. However, in our case, that would cause pages of tedious calculation, and we could not get crucial quantitative information about solutions, which is the key to the stability of the original $(q+1)$-spin system. Hence, we show the non-uniqueness by locating the solutions directly, as in the next two lemmata. Also note that, when $x=R_{0} / R_{1}=C_{0} / C_{1}$, the two-step recursion (4.28) can be simplified into the following one-step recursion

$$
\begin{equation*}
x=\left(\frac{t^{2} x+q t}{t x+q-1}\right)^{d} . \tag{4.29}
\end{equation*}
$$

Lemma 4.35. Let ( $x, x$ ) be the solution of (4.28) i.e., $x$ be the solution of (4.29). When $q \geq 4, k^{\prime} \geq 2$ and $d \geq 5 q^{k}$, it holds that $t x+q-1<d$.

Lemma 4.36. When $q \geq 4, k^{\prime} \geq 2$ and $d \geq 5 q^{k^{\prime}}$, there exists a solution ( $x, y$ ) to (4.28) satisfying (a) $x>y$, and (b) $x>\frac{d}{q^{k^{\prime}}-q} \cdot d$.

We give the proof of Lemma 4.35 and Lemma 4.36 in Section 4.4.4.
Proof of Lemma 4.34. This directly follows from Lemma 4.35 and Lemma 4.36.
Now we are ready to analyse the stability of ( $q, 0,0$ )-type fixpoints. In the following it will be convenient to let $\boldsymbol{J}$ be the $q \times q$ matrix with 0 s on the diagonal and 1 s elsewhere, and $\mathbf{1}$ to be the $q \times 1$ vector with all ones.

Proof of Lemma 4.26. Let $x=R_{0} / R_{1}$ and $y=C_{0} / C_{1}$ be the solution of (4.28) with $x>y$. Set $a:=\sqrt{\frac{1}{t x+q-1}}, b:=\sqrt{\frac{1}{t y+q-1}}, r:=\sqrt{\frac{t y}{t x+q}}$ and $s:=\sqrt{\frac{t x}{t y+q}}$. By applying the formula in Lemma 4.32, the $(q+1) \times(q+1)$ matrix $\boldsymbol{A}$ can be written in the block form

$$
\boldsymbol{A}=\left[\begin{array}{cc}
r s & a s \mathbf{1}^{\mathrm{T}} \\
b r \mathbf{1} & a b \boldsymbol{J}
\end{array}\right]
$$

The eigenvalues of $\boldsymbol{L}=\left[\begin{array}{cc}0 & \boldsymbol{A} \\ \boldsymbol{A}^{\top} & 0\end{array}\right]$ consist of $\pm a b$ (with multiplicity $q-1$ respectively) and $\pm \lambda_{1}, \pm \lambda_{2}$, where $\pm \lambda_{1}, \pm \lambda_{2}$ are the zeros of the following biquadratic function

$$
f(z)=z^{4}-\left((q-1)^{2} a^{2} b^{2}+q b^{2} r^{2}+q a^{2} s^{2}+r^{2} s^{2}\right) z^{2}+a^{2} b^{2} r^{2} s^{2} .
$$

Again, we assume $\lambda_{1}=1$ (note that $a b \neq 1$ ). By Vieta's formula, $\lambda_{2}=a b r s$. Since $r s<$ 1, this means $a b$ is the second largest eigenvalue. Now it suffices to prove $a b<1 / d$, which is equivalent to showing $(t x+q-1)(t y+q-1)>d^{2}$. Note that $t y>t^{d+1}=q^{k^{\prime}}-$ $q$, and Lemma 4.36 gives $x>d \frac{d}{q^{k^{\prime}}-q}$. Therefore $(t x+q-1)(t y+q-1)>t x y>d^{2}$.

Remark 4.37. It is worth noting that the Jacobian stable fixpoints of the system (4.28) do not necessarily induce ( $q, 0,0$ ) -type Jacobian stable fixpoints of the original $(q+1)$-spin system. This is because the eigenvalue $a b$ from the $(q+1)$-spin system is missing in the 2 -spin system. Interestingly, by directly applying results over 2 -spin system (e.g., [GŠV16, Lemma 8]), what we get is $a b r s<1 / d$ instead of $a b<1 / d$. There is an interval of $d$ such that the former holds but the latter does not. Thus here we cannot only analyze the simplified 2 -spin system.

Proof of Lemma 4.28. According to the formula in Lemma 4.32, we construct the following $(q+1) \times(q+1)$ matrix $\boldsymbol{A}$ with block form

$$
\boldsymbol{A}=\left[\begin{array}{cc}
b & \sqrt{a b} \mathbf{1}^{\mathrm{T}} \\
\sqrt{a b} \mathbf{1} & a \boldsymbol{J}
\end{array}\right]
$$

where $a:=\frac{1}{q-1+t x}, b:=\frac{t x}{t x+q}$, and $x$ is the solution of equation (4.29). Because $\boldsymbol{A}$ is symmetric, the spectral radius of $\boldsymbol{L}=\left[\begin{array}{cc}0 & \boldsymbol{A} \\ \boldsymbol{A}^{\top} & 0\end{array}\right]$ is the same as that of $\boldsymbol{A}$. It is not hard to see that $-a$ is an eigenvalue of $\boldsymbol{A}$ by multiplicity $q-1$. From Lemma 4.35, we have that $1 / d<a$, and $a<1$ from $q \geq 2$ and $x>0$. Therefore, the fixpoint is unstable.

### 4.3.5 ( $q, 0,0$ ) fixpoint is not maximal

Let $q_{1}=q, q_{2}=q_{3}=0$ and $R_{0} / R_{1} \neq C_{0} / C_{1}$. Due to stability, it is difficult to analyse this kind of fixpoint's global optimality (recall that it corresponds to a local maxima of $\Psi_{1}$ ). However, observe that changing the value of $R_{3}$ and $C_{3}$ will not affect the value of $\overline{\Phi^{S}}$. Therefore, we can force $R_{3}$ and $C_{3}$ to be subject to (4.12) and (4.13). As we will show later, doing so allows us to reuse some lemmata we have utilized in our argument regarding 2-maximal fixpoints, among which the most important one is the perturbation argument. We define $r_{0}, r_{1}, c_{0}, c_{3}$ analogously, and without loss of generality, suppose $r_{1}, c_{3}>1$.

The next proposition shows how we choose $r_{1}$ and $c_{3}$.
Lemma 4.38. Let $x=r_{1}$ and $y=c_{3}$ be a pair of solutions to the following system

$$
\begin{array}{r}
f_{1}(x, y):=(x-1)\left(\left(1+\frac{x^{d}(y-1)}{x^{d}-1}\right)^{d} t^{d+1}+q-y^{d}\right)-y^{d}+1=0 ;  \tag{4.30}\\
f_{2}(x, y):=(y-1)\left(\left(1+\frac{y^{d}(x-1)}{y^{d}-1}\right)^{d} t^{d+1}+q x^{d}-x^{d}\right)-x^{d}+1=0,
\end{array}
$$

with $x, y>1$. Then there exists $r_{0}$ and $c_{0}$ such that (4.16) and (4.17) are satisfied for $q_{1}=q, q_{2}=q_{3}=0$.

Proof. The $r_{0}$ and $c_{0}$ we choose are defined by

$$
\begin{equation*}
r_{0} / t:=\frac{r_{1}-1}{c_{3}^{d}-1}+r_{1}, \quad c_{0} / t:=\frac{c_{3}-1}{r_{1}^{d}-1}+c_{3} . \tag{4.31}
\end{equation*}
$$

Combining (4.31) with the expression of $f_{2}\left(r_{1}, c_{3}\right)=0$, it holds that

$$
c_{0}^{d} t+q-c_{3}^{d}-\frac{c_{3}^{d}-1}{r_{1}-1}=0
$$

which is exactly (4.24), and is equivalent to the expression for $r_{1}$ in (4.17). The same argument holds for the $c_{3}$ expression in (4.17). In addition, plugging (4.31) back into (4.17) yields the expressions for $r_{0}, c_{0}$ in (4.16).

Be cautious that we do not assume $R_{0} / R_{1}=C_{0} / C_{1}$ in Lemma 4.38. Even if we managed to find a pair of solutions $r_{1}>c_{3}>1$ to (4.30), it does not imply that we can find $R_{3}$ and $C_{3}$ for the case $R_{0} / R_{1} \neq C_{0} / C_{1}$, because it is possible for such a pair to correspond to the other case $R_{0} / R_{1}=C_{0} / C_{1}$. We will handle this in Lemma 4.47 after finding a special solution to (4.30).

To study the solution of the system (4.30), we need to look into the properties of both functions. To clarify the intuition of our approach, we plot both functions for the case $q=6, k^{\prime}=3, d=5 q^{k^{\prime}}$ (see Figure 4.3a). In this setting, the two functions have three intersections in the region $(1,+\infty)^{2}$ : one above $y=x$, one near $y=x$ (but still below $y=x$; see Figure 4.3b) and one far below $y=x$. Experimentally, only the first two intersections correspond to the case $R_{0} / R_{1} \neq C_{0} / C_{1}$. Hence we would only be interested in them. Moreover, as we will see at the end of this subsection, a solution such that $x>y$ is required. For this purpose, the rest of the subsection endeavours to prove the existence of the intersection near $y=x$ before finishing the proof of Lemma 4.27. Doing so also avoids the need of fully characterising the shape of both curves $f_{i}(x, y)=0$.


Figure 4.3: (a): Shape of the curve $f_{1}(x, y)=0, f_{2}(x, y)=0$, and $y=x$. (b): Zoom in on the intersection near $y=x$.

Now we formalize our argument. Note that, by mimicking the proof of Lemma 4.33, one can show $f_{2}(x, x)=0$ has exactly one solution $x^{* *}>1$. Moreover, for any $x \in\left(1, x^{* *}\right), f_{2}(x, x)<0$, and for any $x>x^{* *}, f_{2}(x, x)>0$. A detailed proof is given in Section 4.4.3.

Lemma 4.39. For any $q \geq 4, k^{\prime} \geq 2$ and $d \geq 3 q^{k^{\prime}}$, the function

$$
h_{2}(x):=\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d} t^{d+1}-\frac{x^{d}-1}{x-1}+(q-1) x^{d}
$$

has exactly one root $x^{* *}$ in the region $x>1$.
For $f_{1}$, we do not need the uniqueness of its intersection with the line $y=x$.

Lemma 4.40. For any $q \geq 4, k^{\prime} \geq 2$ and $d \geq 3 q^{k^{\prime}}$, the function

$$
h_{1}(x):=\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d} t^{d+1}-\frac{x^{d}-1}{x-1}+q-x^{d}
$$

has at least one root in the region $x>1$. Let $x^{*}$ be its smallest root. Then $h_{1}(x)<0$ for $x \in\left(1, x^{*}\right)$. Moreover, $x^{*}>x^{* *}$, and consequently $h_{2}\left(x^{*}\right)>0$.

Proof. The first part of the lemma is similar to the proof of Lemma 4.39 and Lemma 4.33, by computing $\lim _{x \rightarrow 1} h_{1}(x)<0$ and $\lim _{x \rightarrow+\infty} h_{1}(x)=+\infty$. To prove the second part, note that $h_{2}(x)>h_{1}(x)$ for all $x>1$.

The next property will be useful later.
Proposition 4.41. If $f_{1}(x, y)=0$, then $x<1+\frac{1}{t^{d+1}-1}$. If $f_{2}(x, y)=0$, then $y<1+\frac{1}{t^{d+1}-1}$. Proof. Suppose $x \geq 1+\frac{1}{t^{d+1}-1}$. Then

$$
\begin{aligned}
f_{1}(x, y) & \geq \frac{1}{t^{d+1}-1}\left(\left(1+\frac{x^{d}(y-1)}{x^{d}-1}\right)^{d} t^{d+1}+q-y^{d}\right)-y^{d}+1 \\
& >\frac{1}{t^{d+1}-1}\left(y^{d} t^{d+1}+q-y^{d}\right)-y^{d}+1=\frac{q}{t^{d+1}-1}+1>0
\end{aligned}
$$

A similar argument holds for $f_{2}$.

Then we study the shape of $f_{1}$ below the line $y=x$.
Lemma 4.42. $\operatorname{Let} g(x):=\frac{\left(x^{d}-1\right)^{d}}{\left(x^{d+1}-1\right)^{d-1}(x-1)}$ and assume that $d \geq 3 q^{k^{\prime}}$. Then
(a) there is a unique $x_{0} \in(1, \infty)$ such that $g\left(x_{0}\right)=t^{d+1}$;
(b) for any $1<x<x_{0}, \frac{\partial f_{1}}{\partial y}<0$ for $y \in(1, x]$; and
(c) $x_{0}>x^{*}$, where $x^{*}>1$ is the smallest solution to $f_{1}(x, x)=0$ (see Lemma 4.40).

Moreover, for any $1<x<x_{0}, f_{1}(x, y)$ is decreasing for $y \in(1, x]$.
Proof. We first show that $g(x)$ is decreasing for $x>1$. By direct calculation,

$$
g^{\prime}(x)=\frac{\left(x^{d}-1\right)^{d-1}}{x(x-1)^{2}\left(x^{d+1}-1\right)^{d}}\left(x^{d} d^{2}(x-1)^{2}-x\left(x^{d}-1\right)^{2}\right)<0
$$

where the last inequality has already been shown in the proof of Lemma 4.18 for $x>1$. Notice that $\lim _{x \rightarrow 1} g(x)=\frac{d^{d}}{(d+1)^{d-1}}$ and $\lim _{x \rightarrow \infty} g(x)=1$. As $\frac{d^{d}}{(d+1)^{d-1}}>\frac{d}{e}>q^{k^{\prime}}>t^{d+1}=$ $q^{k^{\prime}}-q>2$, there is a unique $x_{0}$ such that $g\left(x_{0}\right)=t^{d+1}$ and for $x \in\left(1, x_{0}\right), g(x)>t^{d+1}$. This shows part (a).

For part (b), we have $\frac{\partial f_{1}}{\partial y}=-x d y^{d-1}+\frac{(x-1) d x^{d} t^{d+1}\left(x^{d} y-1\right)^{d-1}}{\left(x^{d}-1\right)^{d}}$ and thus, $\frac{\partial f_{1}}{\partial y}<0$ is equivalent to

$$
\left(x^{d+1}-\frac{\left(x^{d}-1\right)^{d /(d-1)}}{(x-1)^{1 /(d-1)} t^{\frac{d+1}{d-1}}}\right) y<x
$$

As the range of $y$ we consider is $(1, x]$, we only need to show that

$$
x^{d+1}-\frac{\left(x^{d}-1\right)^{d /(d-1)}}{(x-1)^{1 /(d-1)} t^{d+1}}<1,
$$

which, after rearranging, is equivalent to $g(x)>t^{d+1}$. This is guaranteed by part (a) of the lemma.

To prove the third part, by Lemma 4.40, it suffices to show $h_{1}\left(x_{0}\right)=f_{1}\left(x_{0}, x_{0}\right)>0$. Note that $x_{0}$ satisfies

$$
x_{0}^{d+1}-\frac{\left(x_{0}^{d}-1\right)^{d /(d-1)}}{\left(x_{0}-1\right)^{1 /(d-1)} t^{\frac{d+1}{d-1}}}=1, \text { or equivalently, }\left(\frac{x_{0}^{d+1}-1}{x_{0}^{d}-1}\right)^{d-1}=\frac{x_{0}^{d}-1}{t^{d+1}\left(x_{0}-1\right)}
$$

By multiplying this with $\frac{x_{0}^{d+1}-1}{x_{0}^{d}-1}$, we have $\left(\frac{x_{0}^{d+1}-1}{x_{0}^{d}-1}\right)^{d} t^{d+1}=\frac{x_{0}^{d+1}-1}{x_{0}-1}$ and plugging into the expression for $f_{1}(x, x)$ we get $f_{1}\left(x_{0}, x_{0}\right)=q\left(x_{0}-1\right)>0$, yielding part (c).

By Lemma 4.42 (b) and (c), the partial derivative $\partial f_{1} / \partial y \neq 0$ at all points $(x, y)$ such that $f_{1}(x, y)=0$ and $1<y \leq x \leq x^{*}$. Applying the implicit function theorem, $f_{1}$ yields a continuous function between $x$ and $y$ in the region $1<y \leq x \leq x^{*}$.

Corollary 4.43. The set $\mathcal{P}_{1}^{+}:=(1,1)+\left\{(x, y): f_{1}(x, y)=0, x \geq y>1, x \leq x^{*}\right\}$ forms a continuous curve from $(1,1)$ to $x^{*}, x^{*}$, where $x^{*}>1$ is the smallest solution to $f_{1}(x, x)=0$.

Regarding the shape of $f_{2}$, we have the next lemma.
Lemma 4.44. For any $1<y<1+\frac{1}{q-1}$, there are at most two $x>1$ such that $f_{2}(x, y)=0$. Moreover, if $1<y<x^{* *}$, where $x^{* *}>1$ is the unique value such that $f_{2}\left(x^{* *}, x^{* *}\right)=0$ (see Lemma 4.39), then there is exactly one $x>y$ such that $f_{2}(x, y)=0$.

Proof. The crucial idea of this proof is to study the sign of $f_{2}(x, y)$ at its critical points w.r.t. $x$ (i.e., $x^{\prime}$ such that $\partial f_{2}(x, y) / \partial x=0$ at $x=x^{\prime}$ ).

Fix $y$ in the range and define $g(x):=f_{2}(x, y)$. By direct calculation, if $g^{\prime}(x)=0, x^{\prime}$ satisfies

$$
t^{d+1}\left(1+\frac{y^{d}\left(x^{\prime}-1\right)}{y^{d}-1}\right)^{d}=\frac{x^{\prime d-1}\left(y-q\left(y^{\prime}-1\right)\right)\left(x^{\prime} y^{d}-1\right)}{(y-1) y^{d}}
$$

Plugging it back to $g$, we get

$$
g\left(x^{\prime}\right)=1-\frac{x^{\prime d-1}(y-q(y-1))}{y^{d}} .
$$

Because $y-q(y-1)>0$, for any critical point $x^{\prime}$ of $g$,
(a) if $x^{\prime}<\chi$, then $g\left(x^{\prime}\right)>0$;
(b) if $x^{\prime}=\chi$, then $g\left(x^{\prime}\right)=0$;
(c) if $x^{\prime}>\chi$, then $g\left(x^{\prime}\right)<0$,
where $\chi$ is defined by $\chi:=\left(\frac{y^{d}}{y-q(y-1)}\right)^{1 /(d-1)}$.
As $g(x)=f_{2}(x, y)$ for the fixed $y, g(x)$ is a polynomial in $x$. Moreover, $g(1)=$ $(y-1)\left(t^{d+1}+q-1\right)>0$ and $\lim _{x \rightarrow+\infty} g(x)=+\infty$. It implies that $g(x)$ must have an even number of roots. If $g(x)$ does not have any root greater than 1 then we are done. Otherwise, let $x_{1}>1$ be the smallest root and $x_{2}$ be the largest root.

- If $x_{1}<\chi$, this means the next critical point $x^{\prime} \geq x_{1}$ cannot be $x_{1}$, or otherwise, $g\left(x^{\prime}\right)=0$, contradicting with item (a) above. Therefore, $g\left(x^{\prime}\right)<0$, which means $x^{\prime}>\chi$. If there exists another zero $x^{\prime}<x_{3}<x_{2}$, then either $g^{\prime}\left(x_{3}\right)<0$ or $g^{\prime}\left(x_{3}\right)>0$ (otherwise, it contradicts with item (b)). In the former case, there must exist another critical point $x^{\prime \prime}$ such that $\chi<x^{\prime}<x^{\prime \prime}<x_{3}$ and $g\left(x^{\prime \prime}\right)>0$, which contradicts to item (c). In the latter case, there must exist another critical point $x^{\prime \prime \prime}$ such that $x_{3}<x^{\prime \prime \prime}<x_{2}$ and $g\left(x^{\prime \prime \prime}\right)>0$, violating item (c) as well.
- If $x_{1}>\chi$, this means all the critical points $x^{\prime}$ in $\left[x_{1}, x_{2}\right]$ must have function value $g\left(x^{\prime}\right)<0$, which implies there is not any other root in $\left(x_{1}, x_{2}\right)$.
- If $x_{1}=\chi$ and $x_{1}$ is not a critical point, then $g^{\prime}\left(x_{1}\right)<0$ and the argument of the previous case still applies.
- If $x_{1}=\chi$ and $x_{1}$ is a critical point, then for any other critical point (if exists) $x^{\prime}>x_{1}$, it must holds that $g\left(x^{\prime}\right)<0$. Namely once $g(x)$ becomes positive as $x$ increases, the sign of $g^{\prime}(x)$ will not change. It implies that $x_{2}$ is the only root larger than $x_{1}$ in this case. If no critical point $x^{\prime}>x_{1}$ exists, then $x_{1}=x_{2}$ is the only root.

In all cases, $g(x)$ has at most two roots greater than 1 . This finishes the first part of the lemma.

For the second part, notice that if $y<x^{* *}$ then $g(y)<0$, and recall $\lim _{x \rightarrow \infty} g(x)=$ $\infty$. The number of zeros larger than $y$ must be odd, and by the first part, it must be unique. Proposition 4.41 guarantees that $x^{* *}<1+\frac{1}{t^{d+1}-1}<1+\frac{1}{q-1}$.

We then argue there is a point on $\mathcal{P}_{1}^{+}$(except $\left.(1,1)\right)$ such that $f_{2}$ takes zero. To establish this, we first find a point $E$ with $f_{2}\left(x_{E}, y_{E}\right)=0$ such that it lies to the right of $\mathcal{P}_{1}^{+}$(with some extra conditions, and later we will apply Lemma 4.44). To simplify the calculation, we only consider the case $d=5 q^{k^{\prime}}$. The proof of the next lemma consists some detailed calculations, which we postpone till Section 4.4.5.

Lemma 4.45. Suppose $d=5 q^{k^{\prime}}$. There exists a point $E$ with $f_{2}\left(x_{E}, y_{E}\right)=0$ such that it lies to the right of $\mathcal{P}_{1}^{+}$. More specifically, (a) $y_{E}=1+\frac{0.5}{t^{d+1}-1}$; (b) $y_{E}<x^{* *}$; and (c) $x_{E}>1+\frac{1}{t^{d+1}-1}$.

This yields the following lemma.
Lemma 4.46. Suppose $d=5 q^{k^{\prime}}$. The system (4.30) has a solution ( $x, y$ ) such that $x>$ $y>y_{E}$.

Proof. Consider the following point $M$ on $\mathcal{P}_{1}^{+}: y_{M}=y_{E}$, and $x_{M}$ is the largest one such that $\left(x_{M}, y_{M}\right) \in \mathcal{P}_{1}^{+} .{ }^{7}$ Lemma 4.45 (b) asserts that $y_{E}<x^{* *}$, which allows us to invoke Lemma 4.44: for any $y_{E}<x<x_{E}$, we have $f_{2}\left(x, y_{E}\right)<0$. More specifically, $f_{2}\left(x_{M}, y_{M}\right)<0$ because $x_{M}<x^{*}<1+\frac{1}{t^{d+1}-1}<x_{E}$, where the second and third inequalities come from Proposition 4.41 and Lemma 4.45 (c) respectively.

Now consider the path of $\mathcal{P}_{1}^{+}$between the point $M$ and $\left(x^{*}, x^{*}\right)$. It is continuous and bounded away from both $x=1$ and $y=1$, and the function $f_{2}(x, y)$ is continuous over $(1,+\infty) \times(1,+\infty)$. This means as one walks along the path, the value of $f_{2}$ changes continuously; otherwise, it violates the continuity of $f_{2}$ by a simple $\varepsilon-\delta$ argument. Moreover, by the second part of Lemma 4.40, $f_{2}\left(x^{* *}, x^{* *}\right)>0$. This means there must be a point $(x, y)$ on the path such that $f_{2}(x, y)=0$. Moreover, by the choice of $x_{M}$, it must hold that $y>y_{M}=y_{E}$.

Now we argue that the solution we find actually satisfies $R_{0} / R_{1} \neq C_{0} / C_{1}$.

Lemma 4.47. If $r_{1}>c_{3}>1$ yields $R_{0} / R_{1}=C_{0} / C_{1}$, then it must hold that $c_{3}<y_{E}$.

[^10]Proof. In this case, $C_{0} / C_{1}$ is the solution to the one-step recursion (4.29). Define $u:=$ $c_{0} / t=\left(C_{0} / C_{1}\right)^{1 / d} / t$. By rewriting (4.29), one can see that $u$ is the unique solution to the following equation

$$
h(u):=u-\left(1+\frac{1}{t^{d+1} u^{d}+q-1}\right)=0 .
$$

Note that $u>1$. Using this notation, (4.31) yields that $c_{3}=\frac{u\left(r_{1}^{d}-1\right)+1}{r_{1}^{d}}$, giving that $c_{3}<u$.
It remains to show $u<1+\frac{0.5}{t^{d+1}-1}$. Note that the system (4.29) has a unique fixpoint, namely that $h(u)$ has a unique solution over $u>1$. Because $h(1)<0$ and $\lim _{u \rightarrow \infty} h(u)=\infty$, it suffices to prove $h\left(1+\frac{0.5}{t^{d+1}-1}\right)>0$. After plugging in the expression and clearing the denominator, it turns out to be equivalent to

$$
1+3 q-2 q^{k^{\prime}}+\left(q^{k^{\prime}}-q\right)\left(1+\frac{1}{2 q^{k^{\prime}}-2(q+1)}\right)^{5 q^{k^{\prime}}}>0
$$

which is true for any $q \geq 4$ and $k^{\prime} \geq 2$.
We can finally conclude Lemma 4.27.

Proof of Lemma 4.27. Lemma 4.46 guarantees the existence of $r_{1}>c_{3}>y_{E}$ satisfying (4.30) with $x=r_{1}$ and $y=c_{3}$. By Lemma 4.38, given $r_{1}$ and $c_{3}$, we can choose $R_{0}, R_{1}, R_{3}, C_{0}, C_{1}, C_{3}$ to satisfy (4.12) and (4.13). Lemma 4.47 implies that for this choice, $R_{0} / R_{1} \neq C_{0} / C_{1}$. Moreover, the 2 -spin system regarding $R_{0} / R_{1}$ and $C_{0} / C_{1}$ lies in non-uniqueness region, and hence the values of $R_{0} / R_{1}$ and $C_{0} / C_{1}$ are unique up to the swap of $R$ and $C$ (see Section 4.3.4).

Because $R_{3}$ and $C_{3}$ are subject to (4.12) and (4.13), the first part of the proof in Lemma 4.29 still holds, even when $q_{3}=0$ (since we only require (4.32)). Therefore the expression of $\partial \overline{\Phi^{S}} / \partial q_{3}$ still applies. Based on this, by going through the proof of Lemma 4.30 (a), we can see (4.27) still holds, i.e.,

$$
\operatorname{sgn}\left(\frac{\partial \overline{\Phi^{S}}}{\partial q_{1}}-\frac{\partial \overline{\Phi^{S}}}{\partial q_{3}}\right)=-\operatorname{sgn}\left(r_{1}-c_{3}\right)
$$

Hence under this choice, $\partial \overline{\Phi^{S}} / \partial q_{1}-\partial \overline{\Phi^{S}} / \partial q_{3}<0$. Now consider a new $\mathbf{q}$ vector $(q-\varepsilon, 0, \varepsilon)$. When $\varepsilon$ is small enough, the value of $\overline{\Phi^{s}}$ increases, and feasibility in (4.11) still holds. Because the value of $\overline{\Phi^{S}}$ at $\mathbf{q}=(q, 0,0)$ is irrelavent to $R_{3}, C_{3}$, and the value of $\overline{\Phi^{S}}$ is the same for all fixpoints of type $(q, 0,0)$ and $R_{0} / R_{1} \neq C_{0} / C_{1}$, it means $\bar{\Phi}$ does not take the maximum at fixpoints of such type.

Remark 4.48. Our approach in fact jumps out of the local area around the fixpoint. Intuitively, the argument considers a new "imaginary" fixpoint where $\varepsilon$ portion of the $q$ entries $R_{1}$ (resp. $C_{1}$ ) is changed into $R_{3}$ (resp. $C_{3}$, recall that $R_{3}$ and $C_{3}$ are bounded away from $R_{1}$ and $C_{1}$ ), and compares its value of the original induced matrix norm with the one of ( $R_{0}, R_{1}, \cdots, R_{1}, C_{0}, C_{1}, \cdots, C_{1}$ ). This is another reason why optimizing $\overline{\Phi^{S}}$ over all nonnegative q's instead of integer q's helps a lot.

### 4.4 Remaining proofs of this chapter

### 4.4.1 Proof of Lemma 4.29 and Lemma 4.23

Proof of Lemma 4.29 and Lemma 4.23. We first prove Lemma 4.29. Let

$$
\begin{gathered}
S:=R_{0} C_{0} t^{2}+\left(\sum_{j=1}^{3} C_{j} q_{j}\right) R_{0} t+\left(\sum_{j=1}^{3} R_{j} q_{j}\right) C_{0} t+\left(\sum_{j=1}^{3} R_{j} q_{j}\right)\left(\sum_{j=1}^{3} C_{j} q_{j}\right)-\left(\sum_{j=1}^{3} R_{j} C_{j} q_{j}\right), \\
R:=R_{0}^{(d+1) / d}+\left(\sum_{j=1}^{3} R_{j}^{(d+1) / d} q_{j}\right), \quad C:=C_{0}^{(d+1) / d}+\sum_{j=1}^{3} C_{j}^{(d+1) / d} q_{j} .
\end{gathered}
$$

By direct calculation,

$$
\begin{aligned}
\frac{\partial \overline{\Phi^{S}}}{\partial q_{i}}= & \frac{(d+1)}{S}\left[R_{i} C_{0} t+R_{0} C_{i} t-R_{i} C_{i}+R_{i}\left(\sum_{j=1}^{3} C_{j} q_{j}\right)+C_{i}\left(\sum_{j=1}^{3} R_{j} q_{j}\right)\right] \\
& -d\left(\frac{R_{i}^{(d+1) / d}}{R}+\frac{C_{i}^{(d+1) / d}}{C}\right) .
\end{aligned}
$$

Note that if $q_{i}>0$ and $R_{i} \neq 0$, then it must holds that $\partial \overline{\Phi^{S}} / \partial R_{i}=0$, and hence (4.12) applies, which gives

$$
\begin{align*}
& R_{0}^{(d+1) / d} \propto R_{0}\left(C_{0} t^{2}+\left(q_{1} C_{1}+q_{2} C_{2}+q_{3} C_{3}\right) t\right) \\
& R_{i}^{(d+1) / d} \propto R_{i}\left(C_{0} t+q_{1} C_{1}+q_{2} C_{2}+q_{3} C_{3}-C_{i}\right) . \tag{4.32}
\end{align*}
$$

Therefore,

$$
\frac{R_{i}^{(d+1) / d}}{R}=\frac{R_{i} C_{0} t+R_{i}\left(\sum_{j=1}^{3} C_{j} q_{j}\right)-R_{i} C_{i}}{S}
$$

and similarly,

$$
\frac{C_{i}^{(d+1) / d}}{C}=\frac{C_{i} R_{0} t+C_{i}\left(\sum_{j=1}^{3} R_{j} q_{j}\right)-R_{i} C_{i}}{S}
$$

Note that these two equations also hold trivially when $R_{i}=0$ or $C_{i}=0$, respectively. Putting these together yields the desired expression for $\frac{\partial \overline{\Phi^{s}}}{\partial q_{i}}$ in Lemma 4.29.

For the second part of Lemma 4.29, without loss of generality, suppose $q_{1}, q_{2}>0$ and $\partial \overline{\Phi^{S}} / \partial q_{1}-\partial \overline{\Phi^{S}} / \partial q_{2}>0$. Take a positive $\varepsilon$ and consider $\left(q_{1}+\varepsilon, q_{2}-\varepsilon, q_{3}\right)$. When $\varepsilon$ is small enough, the entries $q_{1}+\varepsilon$ and $q_{2}-\varepsilon$ are positive, the value of $\overline{\Phi^{s}}$ increases, and feasibility in (4.11) still holds. Hence $\left(q_{1}, q_{2}, q_{3}\right)$ does not maximize $\bar{\Phi}$.

Finally we prove Lemma 4.23. Here we have an extra condition that $\mathbf{q}$ is $m$ maximal. This means there exists a maximizer $\mathbf{r}, \mathbf{c}$ such that for every $i \neq j$ such that $q_{i}, q_{j}>0$, it holds that $R_{i} \neq R_{j}$ and $C_{i} \neq C_{j}$. From (4.12) and (4.13), we obtain that $\mathbf{r}, \mathbf{c}$ specify an $m$-supported fixpoint of the tree recursion (4.8).

### 4.4.2 Proof of Lemma 4.21

Proof of Lemma 4.21. We first show that the maximum in (4.10) cannot be achieved at $R_{0}=0$ or $C_{0}=0$. Assume otherwise. If $R_{0}=0$, we have that

$$
\left.\frac{\partial \overline{\Phi^{S}}}{\partial R_{0}}\right|_{R_{0}=0}=\frac{(d+1) t}{S} \cdot\left(C_{0} t+q_{1} C_{1}+q_{2} C_{2}+q_{3} C_{3}\right)>0
$$

where $S>0$. Therefore, increasing $R_{0}$ by a sufficiently small amount increases also the value of $\overline{\Phi^{S}}$, contradiction. An analogous argument applies for $C_{0}$.

Next, we show that at least one of $R_{1}, R_{2}, R_{3}, C_{1}, C_{2}, C_{3}$ are non-zero. Assume otherwise, then

$$
\left.\frac{\partial \overline{\Phi^{S}}}{\partial R_{1}}\right|_{R_{1}=0}=\frac{d+1}{S} \cdot\left(C_{0} t+\left(q_{1}-1\right) C_{1}+q_{2} C_{2}+q_{3} C_{3}\right)=\frac{d+1}{S} C_{0} t>0,
$$

and therefore we obtain a contradiction as above.
Consider now a triple ( $q_{1}, q_{2}, q_{3}$ ) with positive entries, and assume w.l.o.g. that the maximum is taken when $R_{1}=0$. We claim that $C_{1}>0$. Otherwise, by the first part of Lemma 4.29, we have $\partial \overline{\Phi^{S}} / \partial q_{1}=0$, and $\partial \overline{\Phi^{S}} / \partial q_{i}>0$ for some $i \in\{2,3\}$ since we cannot have $R_{2}=R_{3}=C_{2}=C_{3}=0$. This yields a contradiction to the second part of Lemma 4.29, and therefore $C_{1}>0$. Observe also that

$$
\left.\frac{\partial \overline{\Phi^{S}}}{\partial R_{1}}\right|_{R_{1}=0}=\frac{d+1}{S} \cdot\left(C_{0} t+\left(q_{1}-1\right) C_{1}+q_{2} C_{2}+q_{3} C_{3}\right)
$$

so by the argument above we conclude that $C_{0} t+\left(q_{1}-1\right) C_{1}+q_{2} C_{2}+q_{3} C_{3} \leq 0$ and therefore $q_{1}<1$ (since $C_{0}, C_{1}>0$ ). This yields that

$$
C_{1} \geq \frac{1}{1-q_{1}}\left(C_{0} t+q_{2} C_{2}+q_{3} C_{3}\right)>C_{0}
$$

On the other hand, since both of $C_{0}, C_{1}$ are nonzero, to achieve the maximum, (4.13) must hold for $i=1$, which gives $C_{0}>C_{1}$, contradiction. Therefore we have $R_{1}>0$ for triples with positive entries.

Exactly the same argument works for triples of type $\left(q_{1}, 0, q_{3}\right)$ with $q_{1}, q_{3}>0$. For the case $(q, 0,0)$, note that $q \geq 4>1$, which means the partial derivatives with respect to both $R_{1}$ and $C_{1}$ are positive at $R_{1}=0$ and $C_{1}=0$ respectively, and hence the maximum cannot be taken at either $R_{1}=0$ or $C_{1}=0$.

To prove the final part of the lemma, suppose that $q_{i}, q_{j}>0$. Since $R_{i}, C_{i}, R_{j}, C_{j}>$ 0 , we have that (4.12) and (4.13) apply, which yields that $R_{i}=R_{j}$ iff $C_{i}=C_{j}$.

### 4.4.3 Proof of Lemma 4.33 and Lemma 4.39

Proof of Lemma 4.33. We put the expression of $h$ here for convenient reference.

$$
\begin{equation*}
h(x):=\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d} t^{d+1}-\frac{x^{d}-1}{x-1}+q^{\prime}+\left(q^{\prime}-1\right) x^{d} . \tag{4.33}
\end{equation*}
$$

We have that $h$ is continuous over $x \in(1,+\infty)$ and $\lim _{x \rightarrow+\infty} h(x)=+\infty$. Using that $t^{d+1}=t^{\Delta}=q^{k^{\prime}}-q$, we have that

$$
\lim _{x \downarrow 1} h(x)=\left(\frac{d+1}{d}\right)^{d} t^{d+1}-d+q-1<\mathrm{e} q^{k^{\prime}}-\mathrm{e} q-d+q-1<\mathrm{e} q^{k^{\prime}}-d<0
$$

This implies the existence of $x$ with $h(x)=0$. To prove the uniqueness of the root, we will show that for any root $x>1$ of $h^{\prime}(x)$, it holds that $h(x)<0$ (note if such $x$ does not exist then we are already done), using the fact that $h$ is differentiable and its derivative is continuous. To see the reason why it is sufficient, note that the number of roots of $h(x)$ over $x>1$ must be odd (because any critical point of $h$ has value less than zero). Assuming towards contradiction, let $x_{2}>x_{1}>1$ be the smallest two roots. Then $h^{\prime}\left(x_{1}\right)>0$ and $h^{\prime}\left(x_{2}\right)<0$, indicating there must be some $x^{*} \in\left(x_{1}, x_{2}\right)$ such that $h^{\prime}\left(x^{*}\right)=0$. However, in this case $h\left(x^{*}\right)>0$, which leads to contradiction.

Next we prove our claim. Take the derivative of $h$ and let it be zero:
$h^{\prime}(x)=d\left(q^{\prime}-1\right) x^{d-1}-\frac{d x^{d-1}}{x-1}+\frac{x^{d}-1}{(x-1)^{2}}+\frac{d t^{d+1} x^{d-1}\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d-1}\left(d-d x+x\left(x^{d}-1\right)\right)}{\left(x^{d}-1\right)^{2}}=0$,
or equivalently,

$$
\begin{equation*}
\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d} t^{d+1}=\frac{\left(x^{d}-1\right)\left(x^{d+1}-1\right)\left(x-x^{d}\left(d\left(q^{\prime}(x-1)-x\right)(x-1)+x\right)\right)}{d(x-1)^{2} x^{d}\left(d-d x+x\left(x^{d}-1\right)\right)} \tag{4.34}
\end{equation*}
$$

Combining (4.33) and (4.34), we obtain that for any $x$ such that $h^{\prime}(x)=0$, it holds that

$$
h(x)=\frac{g\left(x, d, q^{\prime}\right)}{d(x-1)^{2} x^{d-1}\left(d-d x+x\left(x^{d}-1\right)\right)}
$$

where

$$
\begin{align*}
& g\left(x, d, q^{\prime}\right):=d q^{\prime}(x-1)^{2}(x+1)\left(x^{d}-1\right) x^{d-1}-\left(x^{d}-1\right)^{2}\left(x^{d+1}-1\right) \\
&-d^{2} x^{d-1}(x-1)^{2}\left(1-x^{1+d}+q^{\prime}(x-1)\left(x^{d}+1\right)\right) \tag{4.35}
\end{align*}
$$

It is not hard to see that $d-d x+x\left(x^{d}-1\right)>0$ for any $x>1$, so, to show $h(x)<0$, it suffices to prove $g\left(x, d, q^{\prime}\right)<0$ for all $x>1$. This will follow by showing that

$$
\begin{equation*}
g(x, d, 0)<0 \text { and } g\left(x, d, q^{\prime}\right) \text { is decreasing in } q^{\prime} \text {, for any } x>1 \text { and } d \geq 3, \tag{4.36}
\end{equation*}
$$

We have $g(x, d, 0) /\left(x^{d+1}-1\right)=\left(d^{2}(x-1)^{2} x^{d-1}-\left(x^{d}-1\right)^{2}\right)$; the last quantity has been shown negative for all $x>1$ in the proof of Lemma 4.18. To prove the monotonicity w.r.t. $q^{\prime}$ note that

$$
\frac{\partial g}{\partial q^{\prime}}=-d(x-1)^{2} x^{d-1}\left(-(x+1) x^{d}+d(x-1)\left(x^{d}+1\right)+x+1\right)=: d x^{d-1}(x-1)^{2} g_{1}(x)
$$

where $g_{1}(x):=-\left(-(x+1) x^{d}+d(x-1)\left(x^{d}+1\right)+x+1\right)$. Note that

$$
g_{1}^{\prime}(x)=(d+1)\left(x^{d-1}(d+x-d x)-1\right)<0 \text { for } x>1
$$

Since $g_{1}(1)=0$, we obtain $g_{1}(x)<0$ for all $x>1$, proving (4.36) and concluding the proof of Lemma 4.33.

Proof of Lemma 4.39. Recall that $h_{2}(x):=\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d} t^{d+1}-\frac{x^{d}-1}{x-1}+(q-1) x^{d}$. We adopt the same idea as the proof of Lemma 4.33 by showing that $h_{2}$ takes negative values at critical points. The estimation of $\lim _{x \rightarrow 1} h_{2}(x)$ is the same as we did in Lemma 4.33.

Taking the derivative of $h_{2}$ and setting it to zero, we get

$$
d q x^{d-1}+\frac{d t^{d+1}\left(\frac{x-1}{x^{d}-1}+x\right)^{d-1}\left(x\left(x^{d}-1\right)+d(-x)+d\right) x^{d-1}}{\left(x^{d}-1\right)^{2}}-\frac{(d+1) x^{d}}{x-1}+\frac{x^{d+1}-1}{(x-1)^{2}}=0
$$

or equivalently,

$$
\left(\frac{x^{d+1}-1}{x^{d}-1}\right)^{d} t^{d+1}=\frac{x^{-d}\left(x^{d}-1\right)\left(x^{d+1}-1\right)\left(d q x^{d}-2 d q x^{d+1}+d q x^{d+2}+d x^{d+1}+x^{d+1}-d x^{d+2}-x\right)}{d(x-1)^{2}\left(-x^{d+1}+d x-d+x\right)} .
$$

By plugging this back into the expression for $h_{2}(x)$ and simplifying, we obtain that for any $x$ such that $h_{2}^{\prime}(x)=0$ it holds that $h_{2}(x)=\frac{g(x, d, q)}{d(x-1)^{2}\left(d-d x+x\left(x^{d}-1\right)\right)}$, where $g(x, d, q):=-d^{2}(x-1)^{2}\left(x^{d}(q(x-1)-x)+1\right)+d q(x-1)^{2}\left(x^{d}-1\right)-\left(x^{d}-1\right)^{2}\left(x^{d+1}-1\right) x^{1-d}$.

Since $d-d x+x\left(x^{d}-1\right)>0$ for any $x>1$, it remains to prove that $g(x, d, q)<0$. Note that $\frac{\partial g(x, d, q)}{\partial q}=d(x-1)^{2}\left(x^{d}(d(-x)+d+1)-1\right)<0$ for $x>1$ and therefore $g(x, d, q)<$ $g(x, d, 0)$. We also have that $g(x, d, 0) /\left(x^{d+1}-1\right)=\left(d^{2}(x-1)^{2}-x^{1-d}\left(x^{d}-1\right)^{2}\right)<0$, where the inequality follows from the argument below (4.36). Therefore $g(x, d, 0)<0$ for all $x>1$, as desired, finishing the proof.

### 4.4.4 Proof of Lemma 4.35 and Lemma 4.36

We will use the following inequality.

$$
\begin{equation*}
\exp \{a\}>\left(1+\frac{a}{b}\right)^{b}>\exp \left\{\frac{a b}{a+b}\right\} \quad \text { for all } a, b>0 \tag{4.37}
\end{equation*}
$$

Proof of Lemma 4.35. Let $p:=t x+q-1$ and assume for the sake of contradiction that $p \geq d$. Let $w:=p / q^{k^{\prime}}$ and $c:=d / q^{k^{\prime}}$, so tha the assumptions of the lemma imply that $w \geq c \geq 5$. (4.29) gives

$$
p=q-1+t^{d+1}\left(1+\frac{1}{p}\right)^{d} q-1+t^{d+1} \exp \left\{\frac{d}{p}\right\}<q-1+q^{k^{\prime}} \exp \left\{\frac{c}{w}\right\} .
$$

Therefore, $w<\frac{q-1}{q^{k^{\prime}}}+\exp \left\{\frac{c}{w}\right\}<\frac{1}{q^{k^{\prime}-1}}+\mathrm{e}<3$, contradicting $w \geq 5$.
Proof of Lemma 4.36. For any solution ( $x, y$ ) of (4.28), $x$ satisfies the two-step recursion $f(x)=0$, where

$$
f(z):=t^{d}\left(1+\frac{1}{t \cdot t^{d}\left(1+\frac{1}{t z+q-1}\right)^{d}+q-1}\right)^{d}-z .
$$

Take $x$ as the largest root of $f$. Define $c:=d / q^{k^{\prime}}$. Because $\lim _{x \rightarrow \infty} f(x)=-\infty$, to show (b), it suffices to prove $f\left(c^{2} q^{k^{\prime}} \frac{q^{k^{\prime}}}{q^{k^{\prime}}-q}\right)>0$, or equivalently,

$$
\begin{equation*}
\left(1+\frac{1}{\left(q^{k^{\prime}}-q\right) D+q-1}\right)^{d}>t c^{2}\left(\frac{q^{k^{\prime}}}{q^{k^{\prime}}-q}\right)^{2} \quad \text { where } \quad D:=\left(1+\frac{1}{t\left(c^{2} q^{k^{\prime}} \frac{q^{k^{\prime}}}{q^{k^{\prime}}-q}\right)+q-1}\right)^{d} . \tag{4.38}
\end{equation*}
$$

Because $D<\exp \left\{\frac{d}{c^{2} q^{k^{\prime}}}\right\}<\exp \left\{\frac{1}{c}\right\}<1.2215$,

$$
\text { LHS of }(4.38)>\left(1+\frac{1}{1.2215\left(q^{k^{\prime}}-q\right)+q-1}\right)^{c q^{k^{\prime}}}>2.2674^{c}
$$

where the last inequality follows from (4.37). Moreover, for any $q \geq 4, k^{\prime} \geq 2, d \geq$ $5 q^{k^{\prime}}$, we have $\left(q^{k^{\prime}} /\left(q^{k^{\prime}}-q\right)\right)^{2}<1.7778$ and $t<1.0312$. Therefore, RHS of (4.38) $<1.8332 c^{2}$, which is smaller than $2.2674^{c}$ whenever $c \geq 5$. This concludes (b). Part (a) follows from (b) and Lemma 4.35.

### 4.4.5 Proof of Lemma 4.45

Proof of Lemma 4.45. Define $s:=\frac{d}{t^{d+1}-1}$. By Proposition 4.41, any point on $x=1+\frac{s}{d}$ must be on the right of $\mathcal{P}_{1}^{+}$. Therefore we are interested in the point $(x, y)$ where $x=1+\frac{s}{d}$ and $y=1+\frac{s}{2 d}$. Specifically, we will show $f_{2}(x, y)<0$, which, together with the fact that $\lim _{x \rightarrow+\infty} f_{2}(x, y)=+\infty$ for any fixed $y>1$, implies the existence of $x_{E}>x$ such that $f_{2}\left(x_{E}, y\right)=0$. However, in order to apply Lemma 4.44, we further need to show $y<x^{* *}$. The latter can be done by proving $f_{2}(y, y)<0$ due to Lemma 4.39.

We deal with the latter one first. Assume $q \geq 4, k^{\prime} \geq 3$, or $q \geq 12, k^{\prime} \geq 2$. Then $5<s<5.4962, \frac{k^{k^{\prime}}-q}{2\left(q^{k^{\prime}}-q-1\right)}<0.5085$ and $\left(1-\frac{q-1}{2\left(q^{k^{\prime}}-q-1\right)}\right)>0.9580$. Set $D:=\left(1+\frac{s}{2 d}\right)^{d}$. By using (4.37), one can show $D>\exp \{5 / 2\}>12.1824$. Therefore,

$$
\begin{aligned}
f_{2}\left(1+\frac{s}{2 d}, 1+\frac{s}{2 d}\right) & =1+\left(1+\frac{s\left(1+\frac{1}{-1+D}\right)}{2 d}\right)^{d} \frac{q^{k^{\prime}}-q}{2\left(q^{k^{\prime}}-q-1\right)}-D\left(1-\frac{q-1}{2\left(q^{k^{\prime}}-q-1\right)}\right) \\
& <1+\exp \left\{\frac{s}{2}\left(1+\frac{1}{-1+D}\right)\right\} \frac{q^{k^{\prime}}-q}{2\left(q^{k^{\prime}}-q-1\right)}-D\left(1-\frac{q-1}{2\left(q^{k^{\prime}}-q-1\right)}\right) \\
& <1+0.5085 \exp \left\{\frac{5.4962}{2}\left(1+\frac{1}{-1+D}\right)\right\}-0.9580 D<0
\end{aligned}
$$

where in the last inequality we use the fact that the function is decreasing in $D$. The cases $\left(q, k^{\prime}\right)=(4,2),(6,2),(8,2),(10,2)$ also holds by directly computing $f_{2}$.

The first one can be handled similarly. Denote $E:=\left(1+\frac{s}{d}\right)^{d}$. Then $D>E^{1 / 2}$. By using (4.37) again, $E>\exp \{5\}$. Consider the case $q \geq 8, k^{\prime} \geq 3$, or $q \geq 28, k^{\prime} \geq 2$. Then $5<s<5.1921, \frac{q^{k^{\prime}-q}}{2\left(q^{k^{\prime}}-q-1\right)}<0.5010$ and $\left(1-\frac{q-1}{2\left(q^{k^{\prime}}-q-1\right)}\right)>0.9821$. Therefore,

$$
\begin{aligned}
f_{2}\left(1+\frac{s}{d}, 1+\frac{s}{2 d}\right) & =1+\left(1+\frac{s\left(1+\frac{1}{-1+D}\right)}{d}\right)^{d} \frac{q^{k^{\prime}}-q}{2\left(q^{k^{\prime}}-q-1\right)}-E\left(1-\frac{q-1}{2\left(q^{k^{\prime}}-q-1\right)}\right) \\
& <1+\exp \left\{s\left(1+\frac{1}{-1+D}\right)\right\} \frac{q^{k^{\prime}}-q}{2\left(q^{k^{\prime}}-q-1\right)}-E\left(1-\frac{q-1}{2\left(q^{k^{\prime}}-q-1\right)}\right) \\
& <1+\exp \left\{s\left(1+\frac{1}{-1+E^{1 / 2}}\right)\right\} \frac{q^{k^{\prime}}-q}{2\left(q^{k^{\prime}}-q-1\right)}-E\left(1-\frac{q-1}{2\left(q^{\left.k^{\prime}-q-1\right)}\right.}\right) \\
& <1+0.5010 \exp \left\{5.1921\left(1+\frac{1}{-1+E^{1 / 2}}\right)\right\}-0.9821 E<0,
\end{aligned}
$$

where in the last inequality we use the fact that the function is decreasing in $E$. The remaining cases $\left(q, k^{\prime}\right)=(4,3),(6,3),(4,2),(6,2), \cdots,(26,2)$ also holds by directly computing $f_{2}$.

## Chapter 5

## FPRAS for linear hypergraphs

We move on to the algorithmic result of this thesis. Recall the definition of a linear hypergraph; that is, each two hyperedges intersect at at most one vertex. The main result of this chapter is stated as follows.

Theorem 5.1. For any $\delta>0$, there is a sampling algorithm such that given any $\epsilon \in$ $(0,1)$, a $k$-uniform linear hypergraph $H=(V, E)$ with maximum degree $\Delta$, where $k \geq$ $\frac{20(1+\delta)}{\delta}$, and an integer $q \geq 100 \Delta^{\frac{2+\delta}{k-4 / \delta-4}}$, it returns a random $q$-colouring that is $\epsilon$-close to $\mu$ in total variation distance in time $\tilde{O}\left(k^{5} \Delta^{2} n\left(\frac{n \Delta}{\epsilon}\right)^{0.01}\right)$, where $n=|V|$ and $\tilde{O}$ hides a polylog $(n, \Delta, q, 1 / \epsilon)$ factor.

A few quick remarks are in order. First of all, the exponent of $n$ in the running time can be made even closer to 1 if more colours are given. See Theorem 5.6 for the full technical statement. Secondly, our algorithm can be modified into a perfect sampler by applying the bounding chain method [Hub98] based on coupling from the past (CFTP) [PW96a], following the same lines of [HSW21]. Moreover, using known reductions from approximate counting to sampling [JVV86, ŠVV09, Hub15, Kol18] (see [FGYZ21b] for simpler arguments specialized to local lemma settings), one can efficiently and approximately count the number of proper colourings in linear hypergraphs under the same conditions in Theorem 5.1.

The exponent (roughly $2 / k$ ) of $\Delta$ in Theorem 5.1 is unlikely to be tight, although it appears to be the limit of current techniques. In fact, we conjecture that the computational transition for sampling $q$-colourings in linear hypergraphs happens around the same threshold of the local lemma (namely, the exponent should be roughly $1 / k$ ). This conjecture is supported by the hardness result in the previous chapter, and by the algorithm of Frieze and Anastos [FA17] for $q=\Omega(\log n)$. Note that for a linear
$k$-uniform hypergraph with maximum degree $\Delta$, Frieze and Mubayi [FM13] showed that the chromatic number $\chi(H) \leq C_{k}\left(\frac{\Delta}{\log \Delta}\right)^{\frac{1}{k-1}}$ where $C_{k}$ depends only on $k$. Their bound is asymptotically better than the bound given by the local lemma. Thus there may still be a gap between the searching threshold and the sampling threshold.

Technique overview. Our algorithm follows the recent projected Markov chain approach [FGYZ21b] with state compression [FHY21]. Roughly speaking, instead of assigning colours to vertices, we split $[q]$ into $\sqrt{q}$ buckets of size $\sqrt{q}$ each and assign buckets to vertices. We run a (systematic scan) Markov chain on these bucket assignments to generate a sample, and then conditional on this sample to draw a nearly uniform $q$-colouring. The benefit of this bucketing is that, under the conditions of Theorem 5.1, conditional on the assignments of all but one vertices, the assignment of the remaining vertex is close to uniformly at random. This implies that any atomic event ${ }^{1}$ is exponentially unlikely in the number of distinct vertices it depends on. In order to show that this approach works, we need to show two things: 1) the projected Markov chain is rapidly mixing; 2) each step of the Markov chain can be efficiently implemented. For general hypergraphs, the previous $q \gtrsim \Delta^{3 /(k-4)}$ bound comes from balancing the conditions so that the two claims are true simultaneously. However, there is no room left for relaxation on either claim. This means that, for our improvements in linear hypergraphs, new ingredients are required for both claims.

For rapid mixing, we take the information percolation approach [HSZ19, JPV21a, HSW21], where the main effort is to trace discrepancies through a one-step greedy coupling, and to show that they are unlikely after a sufficient amount of time. In linear hypergraphs, an individual discrepancy path through time has more distinct updates of vertices than in the general case, and are thus more unlikely. This allows us to relax the condition. Our mixing time analysis is largely inspired by the work of Hermon, Sly, and Zhang [HSZ19], although we do need to handle some new complicacies, such as hyperedges whose vertices are consecutively updated in the discrepancy path.

For efficient implementation, we use rejection sampling. Here we want to sample the colour/bucket of a vertex conditional on the buckets of all other vertices. We can safely prune hyperedges containing vertices of different buckets. The remaining connected component containing the update vertex needs to have logarithmic size to guarantee efficiency of our rejection sampling. The standard approach to bound

[^11]its size is to do a union bound over certain combinatorial structures with sufficiently many distinct vertices. Most previous analysis is based on enumerating so-called " 2 trees", a notion first introduced by Alon [Alo91]. Unfortunately, under the conditions of Theorem 5.1, there are too many "2-trees" to our need. Instead, we introduce a new structure called "2-block-trees" (see Definition 5.12). Here each "block" is a collection of $\theta$ connected hyperedges, and these blocks satisfy connectivity properties similar to a 2-tree. Since the hypergraph is linear, a block has at least $\theta k-\binom{\theta}{2}$ distinct vertices. As long as $\theta \ll k$, we have a good lower bound on the number of distinct vertices, which in turn implies a good upper bound on the probability of these structures showing up. To finish off with the union bound, we give a new counting argument for the number of 2-block-trees, which is based on finding a good encoding of these structures.

Outline of this chapter. The algorithm and its overall analysis are presented in Section 5.2 and Section 5.3, followed by more detailed analysis, where the implementation of the projected chain is studied in Section 5.4, and its mixing time is studied in Section 5.5.

### 5.1 Preliminaries of this chapter

### 5.1.1 List hypergraph colouring and local uniformity

We are actually considering a more general version of the colouring problem where each vertex may have different colours to choose from than other vertices, namely the list hypergraph colouring problem. Let $H=(V, \mathcal{E})$ be a $k$-uniform hypergraph with maximum degree $\Delta$. Let $\left(Q_{v}\right)_{v \in V}$ be a set of colour lists. We say $\mathbf{X} \in \otimes_{\nu \in V} Q_{v}$ is a proper list colouring if no hyperedge in $H$ is monochromatic with respect to $\mathbf{X}$. Let $\mu$ denote the uniform distribution of all proper list hypergraph colourings. The following local uniformity property holds for the distribution $\mu$. Its proof follows from the argument in [GLLZ19]. We include it here for completeness.

Lemma 5.2 (local uniformity [GLLZ19]). Let $q_{0}=\min _{v \in V}\left|Q_{v}\right|$ and $q_{1}=\max _{v \in V}\left|Q_{v}\right|$. For any $r \geq k \geq 2$, if $q_{0}^{k} \geq \mathrm{e} q_{1} r \Delta$, then for any $v \in V$ and $c \in Q_{v}$,

$$
\frac{1}{\left|Q_{v}\right|} \exp \left(-\frac{2}{r}\right) \leq \mu_{v}(c) \leq \frac{1}{\left|Q_{v}\right|} \exp \left(\frac{2}{r}\right),
$$

where $\mu_{v}$ is the marginal distribution on $v$ induced by $\mu$.

Proof. Let $\mathcal{D}$ denote the product distribution where each $v \in V$ samples a colour in $Q_{v}$ uniformly at random. For each $e \in \mathcal{E}$, let $B_{e}$ be the bad event that $e$ is monochromatic. Let $x(e)=\frac{1}{r \Delta}$ for all $e \in \mathcal{E}$. Note that $r \geq k$. We have

$$
\operatorname{Pr}_{\mathcal{D}}\left[B_{e}\right] \leq \frac{q_{1}}{q_{0}^{k}} \leq \frac{1}{\mathrm{e} r \Delta} \leq \frac{1}{r \Delta}\left(1-\frac{1}{r \Delta}\right)^{k(\Delta-1)} \leq x\left(B_{e}\right) \prod_{B \in \Gamma\left(B_{e}\right)}(1-x(B)) .
$$

By Theorem 2.4, it holds that

$$
\mu_{v}(c) \leq \frac{1}{\left|Q_{v}\right|}\left(1-\frac{1}{r \Delta}\right)^{-\Delta} \leq \frac{1}{\left|Q_{v}\right|} \exp \left(\frac{2}{r}\right)
$$

For the lower bound, consider each hyperedge $e$ such that $v \in e$. Let Block $_{e}$ be the event that all vertices in $e$ except $v$ have the colour $c$. If none of Block $_{e}$ occurs, then $v$ has colour $c$ with probability at least $1 /\left|Q_{v}\right|$. By Theorem 2.4 , we have

$$
\mu_{v}(c) \geq \frac{1}{\left|Q_{v}\right|} \mathbf{P r}_{\mu}\left[\bigwedge_{e \ni v} \overline{\text { Block }_{e}}\right] \geq \frac{1}{\left|Q_{v}\right|}\left(1-\sum_{e \ni v} \mathbf{P r}_{\mu}\left[\text { Block }_{e}\right]\right) .
$$

Note that $\operatorname{Pr}_{\mathcal{D}}\left[\right.$ Block $\left._{e}\right] \leq q_{0}^{-k+1}$ and $\mid \Gamma\left(\right.$ Block $\left._{e}\right) \mid \leq k(\Delta-1)+1$. We have

$$
\operatorname{Pr}_{\mu}\left[\text { Block }_{e}\right] \leq q_{0}^{-k+1}\left(1-\frac{1}{r \Delta}\right)^{-k(\Delta-1)-1} \leq q_{0}^{-k+1} \mathrm{e} \leq \frac{1}{r \Delta}
$$

where the last inequality holds because $q_{0}^{-k+1} \mathrm{e} \leq q_{0}^{-k} q_{1} \mathrm{e} \leq \frac{1}{r \Delta}$, which implies

$$
\mu_{v}(c) \geq \frac{1}{\left|Q_{v}\right|}\left(1-\sum_{e \ni v} \mathbf{P r}_{\mu}\left[\text { Block }_{e}\right]\right) \geq \frac{1}{\left|Q_{v}\right|}\left(1-\frac{1}{r}\right) \geq \frac{1}{\left|Q_{v}\right|} \exp \left(-\frac{2}{r}\right) .
$$

### 5.1.2 Projection scheme, projected distribution and conditional distribution

Our sampling algorithm is based on the following projection scheme introduced in [FHY21]. Suppose $H=(V, \mathcal{E})$ is a $k$-uniform hypergraph and $[q]$ a set of colours, and let $\mu$ denote the uniform distribution of its proper hypergraph colourings.

Definition 5.3 (projection scheme [FHY21]). Let $1 \leq s \leq q$ be an integer. A (balanced) projection scheme with image size $s$ is a function $h:[q] \rightarrow[s]$ such that for any $j \in[s],\left|h^{-1}(j)\right|=\left\lfloor\frac{q}{s}\right\rfloor$ or $\left|h^{-1}(j)\right|=\left\lceil\frac{q}{s}\right\rceil$.

For any $\mathbf{X} \in[q]^{V}$, define the projection image $\mathbf{Y} \in[s]^{V}$ of $\mathbf{X}$ by

$$
\forall v \in V, \quad Y_{v}=h\left(X_{v}\right) .
$$

For simplicity, we often denote $\mathbf{Y}=h(\mathbf{X})$, and for any subset $\Lambda \subseteq V$, we denote $\mathbf{Y}_{\Lambda}=$ $h\left(\mathbf{X}_{\Lambda}\right)$.

Given a projection scheme, the following projected distribution can be naturally defined.

Definition 5.4 (projected distribution). Given a projection scheme $h$, the projected distribution $v$ is the distribution of $\mathbf{Y}=h(\mathbf{X})$, where $\mathbf{X} \sim \mu$.

Given an image of the projection, we can define the following conditional distribution over [ $q]^{V}$.

Definition 5.5 (conditional distribution). Let $\Lambda \subseteq V$ be a subset of vertices. Given a (partial) image $\sigma_{\Lambda} \in[s]^{\Lambda}$, the conditional distribution $\mu^{\sigma_{\Lambda}}$ is the distribution of $\mathbf{X} \sim \mu$ conditional on $h\left(\mathbf{X}_{\Lambda}\right)=\sigma_{\Lambda}$.

By definition, $\mu^{\sigma_{\Lambda}}$ is a distribution over $[q]^{V}$. We use $\mu_{S}^{\sigma_{\Lambda}}$ to denote the marginal distribution on $S \subseteq V$ projected from $\mu^{\sigma_{\Lambda}}$, and we simply denote $\mu_{\{v\}}^{\sigma_{\Lambda}}$ by $\mu_{v}^{\sigma_{\Lambda}}$.

### 5.2 Algorithm

In this section and what follows, we always assume that all vertices in $V$ are labeled by $\{0,1, \ldots, n-1\}$. We also fix the parameter $s=\lceil\sqrt{q}\rceil$.

### 5.2.1 The sampling algorithm

Given a projection scheme $h$ with image size $s$, our sampling algorithm first samples $\mathbf{Y} \in[s]^{V}$ from the projected distribution $v$, and then uses it to sample a random hypergraph colouring from the conditional distribution $\mu^{\mathbf{Y}}$. The pseudocode is given in Algorithm 1.

The main ingredient of Algorithm 1 is the part that samples $\mathbf{Y}$ (Line 1 to Line 5). It is basically a systematic scan version of the Glauber dynamics for $v$. In order to update the state of a particular vertex, we invoke a subroutine Sample, given in Algorithm 2, to sample $X_{v}^{\prime}$ first from the distribution conditional on $\mathbf{Y}_{V \backslash\{v\}}$. Also, Sample is used to generate the random colouring conditional on $\mathbf{Y}$ in Line 6. The subroutine Sample in fact returns an approximate sample with high probability. Here we have to settle with some small error because exactly calculating the conditional distribution is intractable. To implement Sample, we use standard rejection sampling, which

```
Algorithm 1: Sampling algorithm for hypergraph colouring
    Input: A hypergraph \(H=(V, \mathcal{E})\), a set of colours [ \(q\) ], an error bound
            \(0<\epsilon<1\), and a balanced projection scheme \(h:[q] \rightarrow[s]\), where
            \(s=\lceil\sqrt{q}\rceil\)
    Output: A random colouring \(\mathbf{X} \in[q]^{V}\)
    sample \(\mathbf{Y} \in[s]^{V}\) uniformly at random;
    for \(t\) from 1 to \(T=\left\lceil 50 n \log \frac{2 n \Delta}{\epsilon}\right\rceil\) do
        let \(v\) be the vertex with label \((t \bmod n)\);
        \(X_{v}^{\prime} \leftarrow\) Sample \(\left(H, h,\{v\}, \mathbf{Y}_{V \backslash\{v\}}, \frac{\epsilon}{4 T}\right)\);
        /* The Sample subroutine is given in Algorithm 2. */
        \(Y_{v} \leftarrow h\left(X_{v}^{\prime}\right) ;\)
    return \(\mathbf{X} \leftarrow\) Sample \(\left(H, h, V, \mathbf{Y}, \frac{\epsilon}{4 T}\right)\);
```

is described in Algorithm 3. Showing the correctness and efficiency of Algorithm 2 and Algorithm 3 is one of our main contributions.

In the following we flesh out the outline above. Let $\Lambda \subseteq V$ and $\mathbf{Y}_{\Lambda} \in[s]^{\Lambda}$. Note that during the execution of Algorithm 1, $\mathbf{Y}_{\Lambda}$ is a random input to Sample. Let $S \subseteq V$ and $\zeta \in(0,1)$. The subroutine Sample $\left(H, h, S, \mathbf{Y}_{\Lambda}, \zeta\right)$ in Algorithm 1 returns a random sample $\mathbf{X}_{S} \in[q]^{S}$ such that with probability at least $1-\zeta$, the total variation distance between $\mathbf{X}_{S}$ and $\mu_{S}^{\mathbf{Y}_{\Lambda}}$ is at most $\zeta$, where the probability is taken over the randomness of the input $\mathbf{Y}_{\Lambda}$.

In the $t$-th step of the systematic scan in Algorithm 1, we pick the vertex $v$ with label $(t \bmod n)$, and use Line 4 and Line 5 to update the value of $Y_{v}$. Ideally, we want to resample the value of $Y_{v}$ according to the conditional distribution $\nu_{v}^{\mathbf{Y}_{V \backslash\{v\}}}$, where $v$ is the distribution projected from $\mu$. However, exactly computing the conditional distribution is not tractable, and we approximate it by projecting from the random sample $X_{v}^{\prime} \in[q]$ given by Sample in Line 4. It is straightforward to verify that $Y_{v}$ approximately follows the law of ${v_{v}}_{\mathbf{Y}_{V \backslash\{v\}}}$ as long as $X_{v}^{\prime}$ approximately follows the law of $\mu_{v}^{\mathbf{Y}_{V \backslash\{v\}}}$. In the last step, we use Sample to draw approximate samples from the conditional distribution $\mu^{\mathbf{Y}}$.

We explain the details of Sample $\left(H, h, S, \mathbf{Y}_{\Lambda}, \zeta\right)$ next. First we need some notations. Given a partial image $\mathbf{Y}_{\Lambda}$, we say an hyperedge $e \in \mathcal{E}$ is satisfied by $\mathbf{Y}_{\Lambda}$ if there exists $u, v \in e \cap \Lambda$ such that $Y_{u} \neq Y_{v}$. In other words, for all $\mathbf{X} \in[q]^{V}$ such that $\mathbf{Y}_{\Lambda}=h\left(\mathbf{X}_{\Lambda}\right)$, the hyperedge $e$ is not monochromatic with respect to $\mathbf{X}$, and thus $e$ is
always "satisfied" given $\mathbf{Y}_{\Lambda}$. Let $H^{\mathbf{Y}_{\Lambda}}=\left(V, \mathcal{E}^{\mathbf{Y}_{\Lambda}}\right)$ be the hypergraph obtained from $H$ by removing all hyperedges satisfied by $\mathbf{Y}_{\Lambda}$. Let $H_{1}^{\mathbf{Y}_{\Lambda}}, H_{2}^{\mathbf{Y}_{\Lambda}}, \ldots, H_{m}^{\mathbf{Y}_{\Lambda}}$ denote the connected components of $H^{\mathbf{Y}_{\Lambda}}$, where $H_{i}^{\mathbf{Y}_{\Lambda}}=\left(V_{i}, \mathcal{E}_{i}^{\mathbf{Y}_{\Lambda}}\right)$. The following fact is straightforward to verify

$$
\mu^{\mathbf{Y}_{\Lambda}}=\mu_{1}^{\mathbf{Y}_{\Lambda \cap V_{1}}} \times \mu_{2}^{\mathbf{Y}_{\Lambda \cap V_{2}}} \times \ldots \times \mu_{m}^{\mathbf{Y}_{\Lambda \cap V_{m}}}
$$

where $\mu_{i}$ is the uniform distribution over proper $q$-colourings of the sub-hypergraph $H_{i}^{\mathbf{Y}_{\Lambda}}$ (namely, $\mu_{i}^{\mathbf{Y}_{\Lambda \cap V_{i}}}$ is the uniform distribution over list colourings of $H_{i}^{\mathbf{Y}_{\Lambda}}$ conditional on $\mathbf{Y}_{\Lambda \cap V_{i}}$ ). Without loss of generality, we assume $S \cap V_{j} \neq \varnothing$ for $1 \leq j \leq \ell$. To draw a random sample from $\mu_{S}^{\mathbf{Y}_{\Lambda}}$, it suffices to draw a random sample from the product distribution $\mu_{1}^{\mathbf{Y}_{\Lambda \cap V_{1}}} \times \mu_{2} \mathbf{Y}_{\Lambda \cap V_{2}} \times \ldots \times \mu_{\ell}^{\mathbf{Y}_{\Lambda \cap V_{\ell}}}$, which we will do by drawing from each $\mu_{i} \mathbf{Y}_{\wedge \cap V_{i}}$ individually using standard rejection sampling (given in Algorithm 3).

One final detail about Algorithm 2 and Algorithm 3 is about their efficiency. Basically we set some thresholds to guard against two unlikely bad events. We break out from the normal execution immediately and return an arbitrary random sample if one of the following two bad events occur:

- for some $1 \leq i \leq \ell,\left|\mathcal{E}_{i}^{\mathbf{Y}_{\Lambda}}\right|>4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right) ;$
- for some $1 \leq i \leq \ell$, the rejection sampling for $\mu_{i}^{\mathbf{Y}_{\Lambda \cap V_{i}}}$ fails after $R$ trials, where

$$
\begin{equation*}
R:=\left\lceil 10\left(\frac{n \Delta}{\zeta}\right)^{\frac{1}{1000 \eta}} \log \frac{n}{\zeta}\right\rceil \quad \text { and } \quad \eta:=\frac{1}{\Delta}\left(\frac{q}{100}\right)^{\frac{k-3}{2}} \tag{5.1}
\end{equation*}
$$

In other words, the first bad event corresponds to the case that a connected component is so large that the rejection sampling is unlikely to succeed. In the analysis (see Lemma 5.9), we will show that both of the two bad events above occur with low probability, and thus with high probability the Sample subroutine returns an approximate sample with desired accuracy.

### 5.2.2 Proof of the main theorem

Let $H=(V, \mathcal{E})$ be a linear $k$-uniform hypergraph with maximum degree $\Delta$. Let $[q]$ be a set of $q$ colours. Recall $s=\lceil\sqrt{q}\rceil$, where $s$ is the parameter of projection scheme $h$ (Definition 5.3). To construct $h$, we partition $[q]$ into $s$ intervals, where the first $(q \bmod s)$ of them contains $\lceil q / s\rceil$ elements each while the rest contains $\lfloor q / s\rfloor$ elements each. For each $i \in[q]$, set

$$
\begin{equation*}
h(i)=j \quad \text { where } i \text { belongs to the } j \text {-th interval. } \tag{5.2}
\end{equation*}
$$

```
Algorithm 2: Sample \(\left(H, h, S, \mathbf{Y}_{\Lambda}, \zeta\right)\)
    Input: A hypergraph \(H=(V, \mathcal{E})\), a projection scheme \(h:[q] \rightarrow[s]\), a subset
        \(S \subseteq V\), a (partial) image \(\mathbf{Y}_{\Lambda} \in[s]^{\Lambda}\) where \(\Lambda \subseteq V\), and an error bound
        \(\zeta \in(0,1)\)
    Output: A random (partial) colouring \(\mathbf{X}_{S} \in[q]^{S}\)
    remove all hyperedges in \(H\) that are satisfied by \(\mathbf{Y}_{\Lambda}\) to obtain
    \(H^{\mathbf{Y}_{\Lambda}}=\left(V, \mathcal{E}^{\mathbf{Y}_{\Lambda}}\right) ;\)
    let \(H_{i}=\left(V_{i}, \mathcal{E}_{i}^{\mathbf{Y}_{\Lambda}}\right)\) for \(1 \leq i \leq \ell\) be the connected components such that
        \(V_{i} \cap S \neq \varnothing ;\)
    if \(\exists 1 \leq i \leq \ell\) such that \(\left|\mathcal{E}_{i}^{\mathbf{Y}_{\Lambda}}\right|>4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)\) then
        return arbitrarily, for example \(\mathbf{X}_{S} \in[q]^{S}\) uniformly at random;
    for \(i\) from 1 to \(\ell\) do
        \(\mathbf{X}_{i} \leftarrow\) RejectionSampling \(\left(H_{i}, h, \mathbf{Y}_{\Lambda \cap V_{i}}, R\right)\), where \(R=\left\lceil 10\left(\frac{n \Delta}{\zeta}\right)^{\frac{1}{1000 \eta}} \log \frac{n}{\zeta}\right\rceil ;\)
        /* The RejectionSampling subroutine is given in Algorithm 3.
            */
        if \(\mathbf{X}_{i}=\perp\) then
            return arbitrarily, for example \(\mathbf{X}_{S} \in[q]^{S}\) uniformly at random;
    return \(\mathbf{X}_{S}\) where \(\mathbf{X}=\biguplus_{i=1}^{\ell} \mathbf{X}_{i}\);
```

Note that this $h$ satisfies Definition 5.3. In our algorithm, $h$ is implemented as an oracle, supporting the following two types of queries.

- Evaluation: given $i$, the oracle returns $h(i)$.
- Inversion: given $j$, the oracle returns a uniform element in $h^{-1}(j)$.

Obviously, each query can be answered in time $O(\log q)$ because of the construction of $h$.

The next theorem is a stronger form of Theorem 5.1. It shows that our algorithm can run in time arbitrarily close to linear in $n$, the number of vertices, as long as sufficiently many colours are available.

Theorem 5.6. The following result holds for any $\delta>0$ and $0<\alpha \leq 1$. Given any $\epsilon \in(0,1)$, any (list-)q-colouring instance on $k$-uniform linear hypergraph $H=(V, E)$ with maximum degree $\Delta$, and a balanced projection scheme, if $k \geq \frac{20(1+\delta)}{\delta}$ and $q \geq$

```
Algorithm 3: RejectionSampling \(\left(H, h, \mathbf{Y}_{\Lambda}, R\right)\)
    Input: A hypergraph \(H=(V, \mathcal{E})\), a projection scheme \(h:[q] \rightarrow[s]\), a
            (partial) image \(\mathbf{Y}_{\Lambda} \in[s]^{\Lambda}\) where \(\Lambda \subseteq V\) and an integer \(R\)
    Output: A random colouring \(\mathbf{X} \in[q]^{V}\) or a special symbol \(\perp\)
    for each \(v \in V\), let \(Q_{v} \leftarrow h^{-1}\left(Y_{v}\right)\) if \(v \in \Lambda\), and \(Q_{v} \leftarrow[q]\) if \(v \notin \Lambda\);
    for \(i\) from 1 to \(R\) do
        sample \(X_{v} \in Q_{v}\) uniformly at random for all \(v \in V\) and let \(\mathbf{X}=\left(X_{v}\right)_{v \in V}\);
        if \(\mathbf{X}\) is a proper hypergraph colouring of \(H\) then
            return \(\mathbf{X}\);
    return \(\perp\);
```

$100\left(\frac{\Delta}{\alpha}\right)^{\frac{2+\delta}{k-4 / \delta-4}}$, Algorithm 1 returns a random colouring that is $\epsilon$-close to $\mu$ in total variation distance in time $O\left(\Delta^{2} k^{5} n\left(\frac{n \Delta}{\epsilon}\right)^{\alpha / 100} \log ^{4}\left(\frac{n \Delta q}{\epsilon}\right)\right)$.

Remark 5.7. The parameter $\alpha$ captures the relation between the local lemma condition and the running time of the algorithm. If $\alpha$ becomes smaller, the condition is more confined, and the running time is closer to linear. In particular, Theorem 5.1 is implied by setting $\alpha=1$.

We need two lemmas to prove Theorem 5.6. The first lemma analyses the mixing time of the idealised systematic scan. Let $v$ be the projected distribution. The idealised systematic scan for $v$ is defined as follows. Initially, let $\mathbf{X}_{0} \in[s]^{V}$ be an arbitrary initial configuration. In the $t$-th step, the systematic scan does the following update steps.

- Pick the vertex $v \in V$ with label $(t \bmod n)$ and let $X_{t}(V \backslash\{v\}) \leftarrow X_{t-1}(V \backslash\{v\})$.
- Sample $X_{t}(v) \sim v_{v}^{\mathbf{X}_{t-1}(V \backslash\{v\})}$.

Lemma 5.8. If $q \geq 40 \Delta \frac{2}{k-4}$ and $k \geq 20$, the systematic scan chain $\mathbf{P}_{\text {scan }}$ for $v$ is irreducible, aperiodic and reversible with respect to $v$. Furthermore, the mixing time satisfies

$$
\forall 0<\epsilon<1, \quad T_{\text {mix }}\left(\mathbf{P}_{\text {scan }}, \epsilon\right) \leq\left\lceil 50 n \log \frac{n \Delta}{\epsilon}\right\rceil .
$$

Lemma 5.8 is shown in Section 5.5.
Our next lemma analyzes the Sample subroutine. Let $\left(\mathbf{Y}_{t}\right)_{t=0}^{T}$ denote the sequence of random configurations in $[s]^{V}$ generated by Algorithm 1, where $\mathbf{Y}_{0} \in[s]^{V}$ is the initial configuration and $\mathbf{Y}_{t}$ is the configuration after the $t$-th iteration of the for-loop.

For any $1 \leq t \leq T+1$, consider the $t$-th invocation of Sample and define the following two bad events:

- $\mathcal{B}_{\text {com }}(t)$ : in the $t$-th invocation, $\mathbf{X}_{S}$ is returned by Line 4 in Algorithm 2;
- $\mathcal{B}_{\text {rej }}(t)$ : in the $t$-th invocation, $\mathbf{X}_{S}$ is returned by Line 8 in Algorithm 2.

Note that the $(T+1)$-th invocation of the subroutine Sample is in Line 6 in Algorithm 1. Let $H=(V, \mathcal{E})$ denote the input hypergraph of Algorithm 1.

Lemma 5.9. For any $1 \leq t \leq T+1$, the $t$-th invocation of the subroutine $\operatorname{Sample}\left(H, h, S, \mathbf{Y}_{\Lambda}, \zeta\right)$, where $h$ is given by (5.2), satisfies

1. the running time of the subroutine is bounded by $O\left(|S| \Delta^{2} k^{5}\left(\frac{n \Delta}{\zeta}\right)^{\frac{1}{1000 \eta}} \log ^{3}\left(\frac{n \Delta q}{\zeta}\right)\right)$;
2. conditional on neither $\mathcal{B}_{\mathrm{com}}(t)$ nor $\mathcal{B}_{\mathrm{rej}}(t)$ occurs, the subroutine returns a perfect sample from $\mu_{S}^{\mathbf{Y}_{\Lambda}}$;
3. if $q \geq 100 \Delta^{\frac{2}{k-3}}$ and $k \geq 20$, then $\operatorname{Pr}\left[\mathcal{B}_{\mathrm{rej}}(t)\right] \leq \zeta$;
4. for any $\delta>0$, ifk $\geq \frac{20(\delta+1)}{\delta}, q \geq 100 \Delta^{\frac{2+\delta}{k-4 / \delta-3}}$, and $H$ is linear, then $\operatorname{Pr}\left[\mathcal{B}_{\mathrm{com}}(t)\right] \leq$ $\zeta$.

Lemma 5.9 is proved in Section 5.3 and 5.4.
Now we are ready to prove our main result, Theorem 5.6.
Proof of Theorem 5.6. First note that the condition in Theorem 5.6 implies all the conditions in Lemma 5.8 and Lemma 5.9. Denote the output of Algorithm 1 by $\mathbf{X}_{\text {alg }}$. To prove the correctness of our algorithm, the goal is to show

$$
d_{\mathrm{TV}}\left(\mathbf{X}_{\mathrm{alg}}, \mu\right) \leq \epsilon
$$

We first consider an idealized algorithm which, instead of simulating the transitions by the Sample subroutine, is able to run the ideal Glauber dynamics to obtain $\mathbf{Y}_{\text {ideal }}$ before sampling $\mathbf{X}_{\text {ideal }}$ from the distribution $\mu^{\mathbf{Y}_{\text {ideal }}}$. By Lemma 5.8, running this systematic scan for $T=\left\lceil 50 n \log \frac{2 n \Delta}{\epsilon}\right\rceil$ steps ensures $d_{\mathrm{TV}}\left(\mathbf{Y}_{\text {ideal }}, v\right) \leq \frac{\varepsilon}{2}$. On the other hand, a perfect sample $\mathbf{X} \sim \mu$ can be drawn by sampling $\mathbf{Y} \sim v$ first, followed by sampling $\mathbf{X} \sim \mu^{\mathbf{Y}}$ based on that. The upper bound on total variation distance allows us to couple the perfect $\mathbf{Y}$ and $\mathbf{Y}_{\text {ideal }}$ such that $\mathbf{Y} \neq \mathbf{Y}_{\text {ideal }}$ with probability no more than $\frac{\epsilon}{2}$. Conditional on $\mathbf{Y}=\mathbf{Y}_{\text {ideal }}$, the samples $\mathbf{X}$ and $\mathbf{X}_{\text {ideal }}$ on original distribution can be perfectly coupled. Together with the coupling lemma (Lemma 2.3), we have

$$
d_{\mathrm{TV}}\left(\mathbf{X}_{\text {ideal }}, \mu\right) \leq \frac{\epsilon}{2}
$$

Hereinafter, we couple the idealized algorithm with Algorithm 1. The nature of systematic scan warrants that both algorithms pick the same vertex in the same step on Line 3. We then try to couple the vertex update as much as possible. That is, at Step $t$, if none of $\mathcal{B}_{\text {com }}(t)$ or $\mathcal{B}_{\text {rej }}(t)$ happens, then the output of Sample subroutine at Line 4 in Algorithm 1 is perfect, and hence we can couple it with the idealized systematic scan perfectly. The remaining coupling error emerges from the occurrence of $\mathcal{B}_{\text {com }}(t)$ or $\mathcal{B}_{\mathrm{rej}}(t)$. By the coupling lemma (Lemma 2.3) and Lemma 5.9 , we have

$$
d_{\mathrm{TV}}\left(\mathbf{X}_{\text {alg }}, \mathbf{X}_{\text {ideal }}\right) \leq \mathbf{P r}\left[\bigvee_{i=1}^{T}\left(\mathcal{B}_{\mathrm{com}}(t) \vee \mathcal{B}_{\mathrm{rej}}(t)\right)\right]=2 T \zeta=\frac{\epsilon}{2}
$$

where the last equality is due to the selection of $\zeta$ in Algorithm 1. Finally, a straightforward application of triangle inequality yields

$$
d_{\mathrm{TV}}\left(\mathbf{X}_{\mathrm{alg}}, \mu\right) \leq d_{\mathrm{TV}}\left(\mathbf{X}_{\text {alg }}, \mathbf{X}_{\text {ideal }}\right)+d_{\mathrm{TV}}\left(\mathbf{X}_{\text {ideal }}, \mu\right)=\epsilon
$$

as desired.
There are $T+1$ invocations to the Sample subroutine in total, with the first $T$ calls each costing

$$
T_{\text {step }}:=O\left(\Delta^{2} k^{5}\left(\frac{n \Delta}{\epsilon / 4 T}\right)^{\frac{1}{1000 \eta}} \log ^{3}\left(\frac{n \Delta q}{\epsilon / 4 T}\right)\right)
$$

and the final call on Line 6 costing

$$
T_{\text {final }}:=O\left(n \Delta^{2} k^{5}\left(\frac{n \Delta}{\epsilon / 4 T}\right)^{\frac{1}{1000 \eta}} \log ^{3}\left(\frac{n \Delta q}{\epsilon / 4 T}\right)\right) .
$$

Summing up, the total running time is

$$
\begin{equation*}
T_{\text {total }}=T \cdot T_{\text {step }}+T_{\text {tinal }}=O\left((T+n) \Delta^{2} k^{5}\left(\frac{n \Delta}{\epsilon / 4 T}\right)^{\frac{1}{1000 \eta}} \log ^{3}\left(\frac{n \Delta q}{\epsilon / 4 T}\right)\right) \tag{5.3}
\end{equation*}
$$

where

$$
\begin{equation*}
T=50 n \log \frac{2 n \Delta}{\epsilon} \quad \text { and } \quad \eta=\frac{1}{\Delta}\left(\frac{q}{100}\right)^{\frac{k-3}{2}} . \tag{5.4}
\end{equation*}
$$

Note that the condition $q \geq 100\left(\frac{\Delta}{\alpha}\right)^{\frac{2+\delta}{k-4 / \delta-4}}$ implies

$$
\eta=\frac{1}{\Delta}\left(\frac{q}{100}\right)^{\frac{k-3}{2}} \geq \frac{1}{\Delta}\left(\left(\frac{\Delta}{\alpha}\right)^{\frac{2+\delta}{k-4 / \delta-4}}\right)^{\frac{k-3}{2}} \geq \frac{1}{\alpha} \Delta^{\frac{(k-3)(1+\delta / 2)}{k-4 / \delta-4}-1} \geq \frac{1}{\alpha}
$$

and hence

$$
\begin{equation*}
\left(\frac{n \Delta}{\epsilon / 4 T}\right)^{\frac{1}{1000 \eta}} \leq\left(\frac{200 n^{2} \Delta \log \frac{2 n \Delta}{\epsilon}}{\epsilon}\right)^{\alpha / 1000}=O\left(\left(\frac{n \Delta}{\epsilon}\right)^{\alpha / 100}\right) \tag{5.5}
\end{equation*}
$$

Plugging (5.4) and (5.5) back into (5.3), we get

$$
T_{\text {total }}=O\left(\Delta^{2} k^{5} n\left(\frac{n \Delta}{\epsilon}\right)^{\alpha / 100} \log ^{4}\left(\frac{n \Delta q}{\epsilon}\right)\right)
$$

as desired.

### 5.3 Analysis of the Sample subroutine

In this section, we analyse the subroutine Sample and prove Lemma 5.9. Properties 1,2 , and 3 in Lemma 5.9 can be proved using techniques developed in [FGYZ21b, FHY21]. The proofs are given in Section 5.3.1 and Section 5.3.2. We remark that proofs of the first three properties in Lemma 5.9 hold for general hypergraphs, not necessarily linear hypergraphs. It is property 4 that requires a linear hypergraph as the input. The proof of property 4 is quite involved and is left to Section 5.4.

### 5.3.1 Proof of running time and correctness

Proof of Property 1 and 2, Lemma 5.9. Property 2 is straightforwardly implied by the nature of rejection sampling. We now deal with Property 1.

Assume all hypergraphs are stored as incidence lists. We first calculate the time cost of Line 2. Starting from each $v \in S$, we perform depth-first search (DFS) on $H$, and for each edge we encounter, we can check whether it is in $H^{Y_{\Lambda}}$ in time $O(k)$. This procedure can work simultaneously with Line 3, that once the current component reaches $\operatorname{size} 4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)$, the subroutine exits in Line 4 . The number of visits by DFS itself will be upper-bounded by the number of edges times maximum edge degree which is no larger than $\Delta k$. In all, the time complexity of DFS has a crude upper bound

$$
T_{\mathrm{DFS}}=O\left(|S| \cdot k \cdot 4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right) \cdot \Delta k\right)=O\left(|S| \Delta^{2} k^{5} \log \left(\frac{n \Delta}{\zeta}\right)\right) .
$$

For the time cost of Line 6 , be aware $\ell$ is at most $|S|$. Suppose the cost of sampling a uniformly random colour from a colour list $Q \subseteq[q]$ is $O(\log q)$. Each invocation of RejectionSampling contains $R$ rounds, each of which colours the subgraph $H_{i}$ and check if it is a proper colouring. The cost depends to the number of vertices in $H_{i}$, which is upper-bounded by $k \cdot 4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)$. The total cost is then

$$
T_{\mathrm{Rej}}=O\left(|S| \cdot R \cdot \Delta k^{4} \log \left(\frac{n \Delta}{\zeta}\right) \log q\right) \leq O\left(|S| \Delta k^{4}\left(\frac{n \Delta}{\zeta}\right)^{\frac{1}{1000 \eta}} \log ^{3}\left(\frac{n \Delta q}{\zeta}\right)\right)
$$

The total running time of Sample is hence given by

$$
T_{\mathrm{Sample}}=T_{\mathrm{DFS}}+T_{\mathrm{Rej}}=O\left(|S| \Delta^{2} k^{5}\left(\frac{n \Delta}{\zeta}\right)^{\frac{1}{1000 \eta}} \log ^{3}\left(\frac{n \Delta q}{\zeta}\right)\right)
$$

### 5.3.2 Bound the probability of $\mathcal{B}_{\mathrm{rej}}(t)$

Proof of Property 3, Lemma 5.9. By the definition of $\eta$ in (5.1) and the condition in Lemma 5.9, it holds that

$$
q=100(\eta \Delta)^{\frac{2}{k-3}}, \quad \eta \geq 1, \quad \text { and } \quad q \geq 100
$$

Consider Line 6 in Algorithm 2. In the rejection sampling, the input is a hyperedge $H=(V, \mathcal{E})$ with at most $4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)$ hyperedges. The size of the color list for each vertex $v \in V$ satisfies

$$
\left|Q_{v}\right| \geq\left\lfloor\frac{q}{s}\right\rfloor=\left\lfloor\frac{q}{\lceil q\rceil}\right\rfloor \stackrel{(*)}{\geq} \frac{4}{5} \sqrt{q},
$$

where inequality ( $*$ ) holds because $q \geq 100$.
Let $\mathcal{D}$ denote the product distribution that each $v \in V$ samples a colour from $Q_{v}$ uniformly at random. For each hyperedge $e \in \mathcal{E}$, let $\mathcal{B}_{e}$ denote the bad event that $e$ is monochromatic. Note that $\left|Q_{v}\right| \leq q$ for all $v \in V$. We have for any $e \in \mathcal{E}$,

$$
\operatorname{Pr}_{\mathcal{D}}\left[\mathcal{B}_{e}\right] \leq \frac{q}{\left(\frac{4}{5} \sqrt{q}\right)^{k-1}}=\left(\frac{5}{4}\right)^{k-1} q^{\frac{3-k}{2}}=\left(\frac{5}{4}\right)^{k-1} 100^{\frac{3-k}{2}} \frac{1}{\eta \Delta} \leq \frac{1}{10000 \mathrm{e} k^{3} \eta \Delta}
$$

where the last inequality holds because $k \geq 20$. For each $e \in \mathcal{E}$, define $x(e)=\frac{1}{10000 \eta \Delta k^{3}}$. Note that $\eta \geq 1$. It is straightforward to verify that

$$
\operatorname{Pr}_{\mathcal{D}}\left[\mathcal{B}_{e}\right] \leq x(e) \prod_{e^{\prime}: \mathcal{B}_{e^{\prime}} \in \Gamma\left(B_{e}\right)}\left(1-x\left(e^{\prime}\right)\right)
$$

By Lovász local lemma in Theorem 2.4, it holds that

$$
\operatorname{Pr}_{\mathcal{D}}\left[\bigwedge_{e \in \mathcal{E}} \overline{\mathcal{B}(e)}\right] \geq\left(1-\frac{1}{10000 \eta \Delta k^{3}}\right)^{\Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)} \geq \exp \left(-\frac{\log \left(\frac{n \Delta}{\zeta}\right)}{5000 \eta}\right) \geq\left(\frac{\zeta}{n \Delta}\right)^{\frac{1}{1000 \eta}}
$$

The rejection sampling repeats for $R=\left\lceil 10\left(\frac{n \Delta}{\zeta}\right)^{\frac{1}{1000 \eta}} \log \frac{n}{\zeta}\right\rceil$ times. Hence, the probability that the rejection sampling fails on one connected component is at most

$$
\left(1-\left(\frac{\zeta}{n \Delta}\right)^{\frac{1}{1000 \eta}}\right)^{R} \leq \exp \left(-10 \log \frac{n}{\zeta}\right) \leq\left(\frac{\zeta}{n}\right)^{2}
$$

Since there are at most $n$ connected components, by a union bound, we have

$$
\operatorname{Pr}\left[\mathcal{B}_{\mathrm{rej}}(t)\right] \leq \zeta
$$

### 5.4 Analysis of connected components

In this section, we prove Property 4 in Lemma 5.9. We assume that the input hypergraph $H$ is linear in this section. Fix $1 \leq t \leq T+1$. Consider the $t$-th invocation of the subroutine Sample. If $1 \leq t \leq T$, we use $v_{t}$ to denote the vertex picked by the $t$-th step of the systematic scan, i.e. $v_{t}$ is the vertex with label $(t \bmod n)$. Recall that $\mathbf{Y}_{t} \in[s]^{V}$ is the random configuration generated by Algorithm 1 after the $t$-th iteration of the for-loop. Denote

$$
\Lambda=\left\{\begin{array}{ll}
V \backslash\left\{v_{t}\right\} & \text { if } 1 \leq t \leq T  \tag{5.6}\\
V & \text { if } t=T+1
\end{array} \quad \text { and } \quad \mathbf{Y}=\mathbf{Y}_{t-1}(\Lambda)\right.
$$

so that the input partial configuration to Sample is $\mathbf{Y}$ (see Algorithm 1). Hence, we consider the subroutine Sample ( $H, h, S, \mathbf{Y}, \zeta$ ), where $\mathbf{Y} \in[s]^{\Lambda}$ is a random configuration.

Let $H=(V, \mathcal{E})$ denote the input linear hypergraph. Since $\mathbf{Y} \in[s]^{\Lambda}$ is a random configuration, $H^{\mathbf{Y}}$ is a random hypergraph, where $H^{\mathbf{Y}}$ is obtained by removing all the hyperedges in $H$ satisfied by $\mathbf{Y}$. Fix an arbitrary vertex $v \in V$. We use $H_{v}^{\mathbf{Y}}=\left(V_{v}^{\mathbf{Y}}, \mathcal{E}_{v}^{\mathbf{Y}}\right)$ to denote the connected component in $H^{\mathbf{Y}}$ that contains the vertex $v$. Note that $\mathcal{E}_{v}^{\mathbf{Y}}$ can be an empty set. A hyperedge $e \in \mathcal{E}$ is incident to $v$ in the hypergraph $H$ if $v \in e$. We prove the following lemma, which implies property 4.

Lemma 5.10. For any $\delta>0$, if $k \geq \frac{20(1+\delta)}{\delta}, q \geq 100 \Delta^{\frac{2+\delta}{k-4+\delta-3}}$, and $H$ is linear, then for any $v \in V$, anye incident to $v$ in $H$, it holds that

$$
\operatorname{Pr}_{\mathbf{Y}}\left[e \in \mathcal{E}_{v}^{\mathbf{Y}} \wedge\left|\mathcal{E}_{v}^{\mathbf{Y}}\right| \geq 4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)\right] \leq \frac{\zeta}{n \Delta}
$$

We now show that property 4 is a corollary of Lemma 5.10. Since there are at most $\Delta$ hyperedges incident to $v$, by a union bound, we have for all $v \in V$,

$$
\operatorname{Pr}_{\mathbf{Y}}\left[\left|\mathcal{E}_{v}^{\mathbf{Y}}\right| \geq 4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)\right] \leq \sum_{e \ni v} \operatorname{Pr}_{\mathbf{Y}}\left[e \in \mathcal{E}_{v}^{\mathbf{Y}} \wedge\left|\mathcal{E}_{v}^{\mathbf{Y}}\right| \geq 4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)\right] \leq \frac{\zeta}{n}
$$

By a union bound over all vertices $v \in V$, we have

$$
\operatorname{Pr}_{\mathbf{Y}}\left[\exists v \in V \text { s.t. }\left|\mathcal{E}_{v}^{\mathbf{Y}}\right| \geq 4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)\right] \leq \zeta .
$$

This implies the property 4 in Lemma 5.9. The rest of this section is dedicated to the proof of Lemma 5.10.

### 5.4.1 Proof of Lemma 5.10

Denote by $L_{H}=\left(V_{L}, E_{L}\right)=\operatorname{Lin}(H)$ the line graph of $H$ (recall Definition 2.2). Let $e$ be the hyperedge in Lemma 5.10 and let $u=u_{e}$ be the vertex in $L_{H}$ corresponding to $e$. Let $L_{H}^{\mathbf{Y}}=\left(V_{L}^{\mathbf{Y}}, E_{L}^{\mathbf{Y}}\right)$ denote the line graph of $H^{\mathbf{Y}}$. Note that $L_{H}^{\mathbf{Y}}$ is random, and the randomness of $L_{H}^{\mathbf{Y}}$ is determined by the randomness of $\mathbf{Y}$. Equivalently, the graph $L_{H}^{\mathbf{Y}}$ can be generated as follows:

- remove all vertices $w \in V_{L}$ such that the corresponding hyperedges in $H$ are satisfied by $\mathbf{Y}$; let $V_{L}^{\mathbf{Y}} \subseteq V_{L}$ denote the set of remaining vertices;
- let $L_{H}^{\mathbf{Y}}=L_{H}\left[V_{L}^{\mathbf{Y}}\right]$ be the subgraph of $L_{H}$ induced by $V_{L}^{\mathbf{Y}}$.

Let $C \subseteq V_{L}$ denote the random set of all vertices in the connected component of $L_{H}^{\mathbf{Y}}$ that contains the vertex $u$. If $u \notin V_{L}^{\mathbf{Y}}$, let $C=\varnothing$. Define an integer parameter $\theta:=\left\lceil\frac{4}{\delta}\right\rceil$. To prove Lemma 5.10, it suffices to show that

$$
\begin{equation*}
\forall M>\theta, \quad \operatorname{Pr}_{\mathbf{Y}}[|C| \geq M] \leq\left(\frac{1}{2}\right)^{\frac{M}{2 \theta k^{2} \Delta}-1} \tag{5.7}
\end{equation*}
$$

This is because $k \geq \frac{20(\delta+1)}{\delta}>\left\lceil\frac{4}{\delta}\right\rceil+1=\theta+1$, and setting $M=4 \Delta k^{3} \log \left(\frac{n \Delta}{\zeta}\right)$ proves Lemma 5.10.

Define the following collection of subsets

$$
\operatorname{Con}_{u}(M):=\left\{C \subseteq V_{L}|u \in C \wedge| C \mid=M \wedge L_{H}[C] \text { is connected }\right\} .
$$

It is straightforward to verify that

$$
\operatorname{Pr}_{\mathbf{Y}}[|C| \geq M] \leq \operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right]
$$

In our proof, we partition the set $\operatorname{Con}_{u}(M)$ into two disjoint subsets

$$
\operatorname{Con}_{u}(M)=\operatorname{Con}_{u}^{(1)}(M) \uplus \operatorname{Con}_{u}^{(2)}(M),
$$

and we bound the probability separately

$$
\begin{equation*}
\operatorname{Pr}_{\mathbf{Y}}[|C| \geq M] \leq \operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(1)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right]+\operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(2)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right] . \tag{5.8}
\end{equation*}
$$

We use Algorithm 4 to partition the set $\operatorname{Con}_{u}(M)$. Taking as an input any $C \in$ $\operatorname{Con}_{u}(M)$, Algorithm 4 outputs an integer $\ell=\ell(C)$ and disjoint sets $C_{1}, C_{2}, \ldots, C_{\ell} \subseteq C$. Let

$$
\forall C \in \operatorname{Con}_{u}(M), \quad C \in \begin{cases}\operatorname{Con}_{u}^{(1)}(M) & \text { if } \ell(C) \geq \frac{M}{2 \theta k^{2} \Delta}  \tag{5.9}\\ \operatorname{Con}_{u}^{(2)}(M) & \text { if } \ell(C)<\frac{M}{2 \theta k^{2} \Delta} .\end{cases}
$$

We remark that Algorithm 4 is only used for analysis, and we do not need to implement this algorithm.

```
Algorithm 4: 2-block-tree generator
    Input: the parameter \(\delta \in(0,1)\) in Lemma 5.10, the line graph \(L_{H}\), an integer
        \(M>\theta\), a vertex \(u\) in \(L_{H}\), and a subset \(C \in \operatorname{Con}_{u}(M)\)
    Output: an integer \(\ell\) and connected subgraphs \(C_{1}, \cdots, C_{\ell} \subseteq C\)
    let \(G=L_{H}[C]=\left(C, E_{C}\right)\) be the subgraph of \(L_{H}\) induced by \(C\);
    \(\theta \leftarrow\left\lceil\frac{4}{\delta}\right\rceil, \ell \leftarrow 0, V \leftarrow C ;\)
    while \(|V| \geq \theta\) do
        \(\ell \leftarrow \ell+1 ;\)
        if \(\ell=1\) then \(u_{\ell} \leftarrow u\);
        if \(\ell>1\) then let \(u_{\ell}\) be an arbitrary vertex in \(\Gamma_{G}(C \backslash V)\);
        let \(C_{\ell} \subseteq V\) be an arbitrary connected subgraph in \(G\) such that \(\left|C_{\ell}\right|=\theta\)
            and \(u_{\ell} \in C_{\ell}\);
        \(V \leftarrow V \backslash\left(C_{\ell} \cup \Gamma_{G}\left(C_{\ell}\right)\right) ;\)
        for each connected component \(G^{\prime}=\left(V^{\prime}, E^{\prime}\right)\) in \(G[V]\) such that \(\left|V^{\prime}\right|<\theta\) do
            \(V \leftarrow V \backslash V^{\prime} ;\)
    return \(\ell,\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\}\);
```

In Line 6 and Line 7 of Algorithm 4, we may use a specific rule to choose the vertex $u_{\ell}$ and the connected subgraph $C_{\ell}$ (e.g. pick the element with the smallest index according to an arbitrary but predetermined ordering). To explain this algorithm concretely, consider the first round of the while-loop running on the graph in Figure 5.1, with the parameter $\theta$ set to 3 .


Figure 5.1: The example graph where Algorithm 4 runs on.

In Line 7, the algorithm picks the connected subgraph $C_{1}$ containing $u$, represented by black circles. Then in Line 8 , the algorithm removes $C_{1}$, together with its neighbours, depicted by circles in dark grey, from the vertex set $V$. Afterwards, the


Figure 5.2: An example 2-block-tree with block size $\theta=3$ and tree size $\ell=3$. Each $C_{i}$ is indicated by a group of vertices of the same colour.
algorithm checks all remaining connected components, and removes those with size less than $\theta=3$ from $V$ in Line 10. In this example, the algorithm captures and deletes the component in the dotted box. Be aware that their neighbours (dark grey circles) have already been removed from $V$. As the algorithm goes into the second round of the while-loop, the next candidate starting point $u_{2}$ is selected, as of in Line 6, among the vertices depicted by white circles.

To formalize the properties of Algorithm 4, we begin with the following proposition, which asserts that Algorithm 4 is well defined. The proof is given in Section 5.4.2.

Proposition 5.11. Given the input $\delta, L_{H}, M, u$, and $C \in \operatorname{Con}_{u}(M)$, Algorithm 4 terminates and generates a unique output. Moreover, when Algorithm 4 terminates, $V=\varnothing$.

The next proposition, yet of more importance, establishes a few properties of the output of Algorithm 4. They will eventually be used to bound the probabilities on the right hand side (RHS) of (5.8). Before characterising these properties, we introduce a notion called "2-block-tree". (See Figure 5.2)

Definition 5.12 (2-block-tree). Let $\theta \geq 1$ be an integer. Let $G=(V, E)$ be a graph. A set $\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\}$ is a 2 -block-tree with block size $\theta$ and tree $\operatorname{size} \ell$ in $G$ if
(B1) for any $1 \leq i \leq \ell, C_{i} \subseteq V,\left|C_{i}\right|=\theta$ and the induced subgraph $G\left[C_{i}\right]$ is connected;
(B2) for any distinct $1 \leq i, j \leq \ell, \operatorname{dist}_{G}\left(C_{i}, C_{j}\right) \geq 2$;
(B3) $\left\{C_{1}, \cdots, C_{\ell}\right\}$ is connected on $G^{2}$. (Recall Definition 2.1 of graph powers.)
One can easily observe that the notion of 2-block-trees is a generalisation of 2trees in [Alo91] by setting $\theta=1$. The output of Algorithm 4 is a 2-block-tree in $L_{H}$. This explains the name " 2 -block-tree generator".

Proposition 5.13. The output $\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\}$ of Algorithm 4 satisfies that

1. $\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\}$ is a 2-block-tree in $L_{H}$ with block size $\theta$ satisfying $u \in C_{1}$ and $\cup_{i=1}^{\ell} C_{i} \subseteq C ;$
2. if all vertices in $\Gamma_{G}\left(C_{i}\right)$ are removed from $G$, where $G=L_{H}[C]$, then the resulting graph $G\left[C^{\prime}\right]$ is a collection of connected components whose sizes are at most $\theta$, where $C^{\prime}=C \backslash\left(\cup_{i=1}^{\ell} \Gamma_{G}\left(C_{i}\right)\right)$.

In Proposition 5.13, Item 1 is stated with respect to the line graph $L_{H}$, but Item 2 is stated with respect to the induced subgraph $L_{H}[C]$. The proof of Proposition 5.13 is also given in Section 5.4.2.

Finally, to bound the probabilities on the RHS of (5.8), we need the following lemma about the random configuration $\mathbf{Y} \in[s]^{\Lambda}$. The proof of Lemma 5.14 is given in Section 5.4.3.

Lemma 5.14. If $\lfloor q / s\rfloor^{k} \geq 2 \mathrm{e} q k \Delta$, then for any $R \subseteq \Lambda$, any $\sigma \in[s]^{R}$, it holds that

$$
\operatorname{Pr}\left[\mathbf{Y}_{R}=\sigma\right] \leq\left(\frac{1}{s}+\frac{1}{q}\right)^{|R|} \exp \left(\frac{|R|}{k}\right)
$$

The following result is a straightforward corollary of Lemma 5.14.
Corollary 5.15. Let $\delta>0$ and $R_{1}, R_{2}, \ldots, R_{\ell} \subseteq \Lambda$ be disjoint subsets. For each $1 \leq i \leq \ell$, let $\mathcal{S}_{i} \subseteq[s]^{R_{i}}$ be a subset of configurations (namely an event). If $k \geq \frac{20(\delta+1)}{\delta}$ and $q \geq$ $100 \Delta^{\frac{2+\delta}{k-4 \delta-3}}$, then it holds that

$$
\operatorname{Pr}\left[\bigwedge_{i=1}^{\ell}\left(\mathbf{Y}_{R_{i}} \in \mathcal{S}_{i}\right)\right] \leq \prod_{i=1}^{\ell}\left|\mathcal{S}_{i}\right|\left(\frac{1}{s}+\frac{1}{q}\right)^{\left|R_{i}\right|} \exp \left(\frac{\left|R_{i}\right|}{k}\right)
$$

Proof. Let $R=R_{1} \uplus R_{2} \uplus \ldots \uplus R_{\ell}$. Note that $\bigwedge_{i=1}^{\ell}\left(\mathbf{Y}_{R_{i}} \in \mathcal{S}_{i}\right)$ if and only if $\mathbf{Y}_{R} \in \mathcal{S}_{1} \otimes$ $\mathcal{S}_{2} \otimes \ldots \otimes \mathcal{S}_{\ell}$, where

$$
\mathcal{S}_{1} \otimes \mathcal{S}_{2} \otimes \ldots \otimes \mathcal{S}_{\ell}:=\left\{\sigma \in[s]^{R} \mid \forall 1 \leq i \leq \ell, \sigma_{R_{i}} \in \mathcal{S}_{i}\right\} .
$$

We now verify the condition in Lemma 5.14 that $\lfloor q / s\rfloor^{k} \geq 2 \mathrm{e} q k \Delta$. Since $s=\lceil\sqrt{q}\rceil$ and $q \geq 100,\lfloor q / s\rfloor \geq \sqrt{q} / 4$. Thus it suffices to verify $(\sqrt{q} / 4)^{k} \geq 2 \mathrm{e} q k \Delta$. The condition in Corollary 5.15 implies that $q \geq 100 \Delta \frac{2}{k-2}$ and $k \geq 20$, which implies $(\sqrt{q} / 4)^{k} \geq 2 \mathrm{e} q k \Delta$. Hence, the condition in Lemma 5.14 holds. We have

$$
\operatorname{Pr}\left[\bigwedge_{i=1}^{\ell}\left(\mathbf{Y}_{R_{i}} \in \mathcal{S}_{i}\right)\right]=\sum_{\sigma \in \mathcal{S}_{1} \uplus \mathcal{S}_{2} \uplus \ldots \uplus \mathcal{S}_{\ell}} \operatorname{Pr}\left[\mathbf{Y}_{R}=\sigma\right] \leq \prod_{i=1}^{\ell}\left|\mathcal{S}_{i}\right|\left(\frac{1}{s}+\frac{1}{q}\right)^{\left|R_{i}\right|} \exp \left(\frac{\left|R_{i}\right|}{k}\right) . \square
$$

Now, we are ready to bound the probabilities on the RHS of (5.8). We handle the two terms separately:

$$
\begin{align*}
& \operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(1)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right]<\left(\frac{1}{2}\right)^{\frac{M}{2 \theta k^{2} \Delta}} ;  \tag{5.10}\\
& \operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(2)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right]<\left(\frac{1}{2}\right)^{M}, \tag{5.11}
\end{align*}
$$

Combining (5.8) with (5.10) and (5.11), we have

$$
\begin{aligned}
\operatorname{Pr}_{\mathbf{Y}}[|C| \geq M] & \leq \operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(1)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right]+\operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(2)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right] \\
& \leq\left(\frac{1}{2}\right)^{\frac{M}{2 \theta k^{2} \Delta}}+\left(\frac{1}{2}\right)^{M} \leq\left(\frac{1}{2}\right)^{\frac{M}{2 \theta k^{2} \Delta}-1}
\end{aligned}
$$

This proves the desired inequality (5.7).
In the next two subsections, we give proofs of (5.10) and (5.11).

### 5.4.1.1 Proof of inequality (5.10)

We first prove (5.10). We need to use the following two properties of 2-block-trees, the proofs of which are deferred till Section 5.4.4.

Lemma 5.16. Let $\theta \geq 1$ be an integer. Let $G=(V, E)$ be a graph. For any integer $\ell \geq 2$, any vertex $v \in V$, if $G$ has a 2 -block-tree $\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\}$ with block size $\theta$ and tree size $\ell$ such that $v \in \cup_{i=1}^{\ell} C_{i}$, then there exists an index $1 \leq i \leq \ell$ such that $\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\} \backslash\left\{C_{i}\right\}$ is a 2-block-tree in $G$ with block size $\theta$ and tree size $\ell-1$ and $v \in \cup_{1 \leq j \leq \ell: j \neq i} C_{j}$.

Lemma 5.17. Let $\theta \geq 1$ be an integer. Let $G=(V, E)$ be a graph with maximum degree $d$. For any integer $\ell \geq 1$, any vertex $v \in V$, the number of 2-block-trees $\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\}$ with block size $\theta$ and tree size $\ell$ such that $v \in \cup_{i=1}^{\ell} C_{i}$ is at most $\left(\theta \mathrm{e}^{\theta} d^{\theta+1}\right)^{\ell}$.

In the rest of this subsection we fix $\ell=\left\lceil\frac{M}{2 \theta k^{2} \Delta}\right\rceil$. By (5.9), Proposition 5.13, and Lemma 5.16, for any $C \in \operatorname{Con}_{u}^{(1)}(M)$, there is a 2-block-tree tree $\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\}$ in the line graph $L_{H}$ with block size $\theta$ and tree size $\ell$ satisfying:
(P1) $u \in C_{1} \cup C_{2} \cup \ldots \cup C_{\ell}$;
(P2) $C_{1} \cup C_{2} \cup \ldots \cup C_{\ell} \subseteq C$.

We denote a 2 -block-tree tree with block size $\theta$ and tree size $\ell$ by $(\theta, \ell)$-2-block-tree. This implies that

$$
\begin{align*}
& \operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(1)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right] \\
\leq & \operatorname{Pr}_{\mathbf{Y}}\left[\exists(\theta, \ell) \text {-2-block-tree }\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\} \text { in } L_{H} \text { satisfying (P1) s.t. } \forall 1 \leq i \leq \ell, C_{i} \subseteq V_{L}^{\mathbf{Y}}\right] . \tag{5.12}
\end{align*}
$$

Note that we only need to consider $(\theta, \ell)$-2-block trees satisfying (P1), because (P2) implies the event that $\forall 1 \leq i \leq \ell, C_{i} \subseteq V_{L}^{\mathbf{Y}}$.

To bound the probability, we fix a $(\theta, \ell)$-2-block tree $\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\}$ in $L_{H}$ satisfying (P1). Fix an index $1 \leq j \leq \ell$. By Definition $5.12,\left|C_{j}\right|=\theta$. Note that each vertex in $C_{j}$ represents a hyperedge in the input hypergraph $H=(V, \mathcal{E})$. Let the hyperedges in $C_{j}$ be $e_{1}^{j}, e_{2}^{j}, \ldots, e_{\theta}^{j}$. For each $1 \leq t \leq \theta$, we define a subset of vertices $R_{t}^{j} \subseteq \Lambda($ in $H)$ by

$$
S_{t}^{j}:=e_{t}^{j} \backslash\left(\bigcup_{i \in[\theta]: i \neq t} e_{i}^{j}\right) \quad \text { and } \quad R_{t}^{j}:=S_{t}^{j} \cap \Lambda
$$

where $\Lambda$ is defined in (5.6). By definition, $R_{t}^{j} \subseteq e_{t}^{j}$ is a subset of vertices of the input hypergraph $H=(V, \mathcal{E})$, and $R_{t}^{j} \cap e_{i}^{j}=\varnothing$ for any $i \neq t$. This implies that $R_{1}^{j}, R_{2}^{j}, \ldots, R_{\theta}^{j}$ are mutually disjoint. Furthermore, since $H$ is linear and $|\Lambda| \geq|V|-1$, we have

$$
\begin{equation*}
\forall 1 \leq t \leq \theta: \quad\left|R_{t}^{j}\right| \geq k-(\theta-1)-1=k-\theta . \tag{5.13}
\end{equation*}
$$

The above inequality holds because (1) $\left|e_{t}^{j}\right|=k$; (2) for each $e_{i}^{j}$ with $i \neq t$, the intersection between $e_{t}^{j}$ and $e_{i}^{j}$ is at most one vertex; and (3) $|\Lambda| \geq|V|-1$. By Definition 5.12 of 2-block-trees, for $i \neq j$, $\operatorname{dist}_{L_{H}}\left(C_{i}, C_{j}\right) \geq 2$. Let $e \in \mathcal{E}$ be a hyperedge in $C_{i}$ and $e^{\prime} \in \mathcal{E}$ be a hyperedge in $C_{j}$, this implies that $e$ and $e^{\prime}$ are not adjacent in the line graph $L_{H}$, and thus $e \cap e^{\prime}=\varnothing$. Hence,

$$
\begin{equation*}
\left(R_{t}^{j}\right)_{1 \leq j \leq \ell, 1 \leq t \leq \theta} \text { are mutually disjoint. } \tag{5.14}
\end{equation*}
$$

We now bound the probability of $C_{j} \subseteq V_{L}^{\mathbf{Y}}$ for all $1 \leq j \leq \ell$. For all $1 \leq j \leq \ell$ and $1 \leq$ $t \leq \theta$, since $C_{j} \subseteq V_{L}^{\mathbf{Y}}$, the hyperedge $e_{t}^{j}$ is not satisfied by $\mathbf{Y}$, thus $e_{t}^{j}$ is monochromatic with respect to $\mathbf{Y}$, i.e. for all $v, v^{\prime} \in e_{t}^{j}$, it holds that $Y_{v}=Y_{v^{\prime}}$. Note that $R_{t}^{j} \subseteq e_{t}^{j}$. We have the following bound
$\operatorname{Pr}_{\mathbf{Y}}\left[\forall 1 \leq j \leq \ell, C_{j} \subseteq V_{L}^{\mathbf{Y}}\right] \leq \operatorname{Pr}_{\mathbf{Y}}\left[\forall 1 \leq j \leq \ell, 1 \leq t \leq \theta, R_{t}^{j}\right.$ is monochromatic w.r.t. $\left.\mathbf{Y}\right]$.

Let $\mathcal{S}_{t}^{j}$ be the set of all $s$ monochromatic configurations of $R_{t}^{j}$ (i.e. all vertices in $R_{t}^{j}$ take the same value $c$, where $c \in[s]$ ), or more formally,

$$
\mathcal{S}_{t}^{j}=\left\{\sigma \in\{c\}^{R_{t}^{j}} \mid c \in[s]\right\} .
$$

In particular, $\left|\mathcal{S}_{t}^{j}\right|=s$. By Corollary 5.15, (5.13), (5.14), and (5.15), it holds that

$$
\begin{aligned}
\operatorname{Pr}_{\mathbf{Y}}\left[\forall 1 \leq i \leq \ell, C_{i} \subseteq V_{L}^{\mathbf{Y}}\right] & \leq \operatorname{Pr}_{\mathbf{Y}}\left[\bigwedge_{j=1}^{\ell} \bigwedge_{t=1}^{\theta}\left(Y_{R_{t}^{j}} \in \mathcal{S}_{t}^{j}\right)\right] \leq \prod_{i=1}^{\ell} \prod_{t=1}^{\theta} s\left(\frac{1}{s}+\frac{1}{q}\right)^{\left|R_{t}^{j}\right|} \exp \left(\frac{\left|R_{t}^{j}\right|}{k}\right) \\
& \leq s^{\ell \theta} \prod_{i=1}^{\ell} \prod_{t=1}^{\theta}\left(\frac{1}{s}+\frac{1}{q}\right)^{\left|R_{t}^{j}\right|} \exp \left(\frac{\left|R_{t}^{j}\right|}{k}\right)^{\prime} \\
\left(\text { as } k-\theta \leq\left|R_{t}^{j}\right| \leq k\right) & \leq(\mathrm{e} s)^{\ell \theta}\left(\frac{1}{s}+\frac{1}{q}\right)^{\ell \theta(k-\theta)}=\left((\mathrm{es})^{\theta}\left(\frac{1}{s}+\frac{1}{q}\right)^{\theta(k-\theta)}\right)^{\ell} .
\end{aligned}
$$

Note that the maximum degree of $L_{H}$ is no more than $k \Delta$. By Lemma 5.17 and a union bound over all possible 2-block-trees, we have

$$
\begin{align*}
& \operatorname{Pr}_{\mathbf{Y}}\left[\exists(\theta, \ell)-2 \text {-block-tree }\left\{C_{1}, C_{2}, \ldots, C_{\ell}\right\} \text { in } L_{H} \text { satisfying (P1) s.t. } \forall 1 \leq i \leq \ell, C_{i} \subseteq V_{L}^{\mathbf{Y}}\right] \\
\leq & \left(\theta \mathrm{e}^{2 \theta}(k \Delta)^{\theta+1} s^{\theta}\left(\frac{1}{s}+\frac{1}{q}\right)^{\theta(k-\theta)}\right)^{\ell} \leq\left(\theta \mathrm{e}^{2 \theta} 2^{\theta(k-\theta)}(k \Delta)^{\theta+1} s^{\theta-\theta(k-\theta)}\right)^{\ell} \tag{5.16}
\end{align*}
$$

where the last inequality uses the fact that $\frac{1}{s}+\frac{1}{q} \leq \frac{2}{s}$. We will show that

$$
\begin{equation*}
\theta \mathrm{e}^{2 \theta} 2^{\theta(k-\theta)}(k \Delta)^{\theta+1} s^{\theta-\theta(k-\theta)} \leq \frac{1}{2} \tag{5.17}
\end{equation*}
$$

Recall that $k>\theta+1$, and consequently $\theta(k-\theta)-\theta>0$. It implies that $\theta \mathrm{e}^{2 \theta} 2^{\theta(k-\theta)}(k \Delta)^{\theta+1} s^{\theta-\theta(k-\theta)} \leq \frac{1}{2} \Longleftrightarrow s \geq \theta^{\frac{1}{\theta(k-\theta)-\theta}} \mathrm{e}^{\frac{2 \theta}{\theta(k-\theta)-\theta}} 2^{\frac{\theta(k-\theta)+1}{\theta(k-\theta)-\theta}}(k \Delta)^{\frac{\theta+1}{\theta(k-\theta)-\theta}}$. Recall that $s=\lceil\sqrt{q}\rceil \geq q^{1 / 2}$. It suffices to show that

$$
q \geq \theta^{\frac{2}{\theta(k-\theta)-\theta}} \mathrm{e}^{\frac{4 \theta}{\theta(k-\theta)-\theta}} 2^{\frac{2 \theta(k-\theta)+2}{\theta(k-\theta)-\theta}}(k \Delta)^{\frac{2 \theta+2}{\theta(k-\theta)-\theta}}=\theta^{\frac{2}{\overline{\theta(k-\theta)-\theta}} \mathrm{e}^{\frac{4}{k-\theta-1}} 2^{\frac{2(k-\theta)+2 / \theta}{k-\theta-1}}(k \Delta)^{\frac{2+2 / \theta}{k-\theta-1}} . . ~ . ~}
$$

Recall that $\theta=\left\lceil\frac{4}{\delta}\right\rceil$. If $\delta \geq 4$, then $\theta=1$. In this case, we only need to show that

$$
q \geq \mathrm{e}^{\frac{4}{k-2} 2^{\frac{2 k}{k-2}} \frac{4}{k-2}_{k-2+\delta / 2}^{k-2} .}
$$

Otherwise $0<\delta<4$, in which case we only need to show that

$$
q>2 \mathrm{e}^{\frac{4}{k-4 / \delta-2}} 2^{\frac{2 k-8 / \delta+\delta / 2}{k-4 / \delta-2}}(k \Delta)^{\frac{2+\delta / 2}{k-4 / \delta-2}},
$$

as $\theta^{\frac{2}{\theta(k-\theta)-\theta}}<2$ and $4 / \delta \leq \theta<4 / \delta+1$. The conditions $k \geq \frac{20(\delta+1)}{\delta}$ and $q \geq 100 \Delta^{\frac{2+\delta}{k-4 / \delta-3}}$ imply both conditions above. This finishes the proof of (5.17). Finally, (5.10) follows from combining (5.12), (5.16), and (5.17).

### 5.4.1.2 Proof of inequality (5.11)

We continue to show (5.11). Fix a connected component $C \in \operatorname{Con}_{u}^{(2)}(M)$. We analyse the probability of $C \subseteq V_{L}^{\mathbf{Y}}$. We run Algorithm 4 with the input $C$. The algorithm outputs an integer $\ell<\frac{M}{2 \theta k^{2} \Delta}$ and a set of connected components $C_{1}, C_{2}, \ldots, C_{\ell}$. Let $G=L_{H}[C]$ be the subgraph of $L_{H}$ induced by $C$. By Proposition 5.13, after removing all vertices of $\Gamma_{G}\left(C_{i}\right)$ for all $1 \leq i \leq \ell$, the graph $G$ is decomposed into connected components with vertex sets $D_{1}, D_{2}, \ldots, D_{m} \subseteq C$ such that $\left|D_{i}\right| \leq \theta$ for all $1 \leq j \leq m$. Note that given $C \in \operatorname{Con}_{u}^{(2)}(M)$, all the sets $D_{1}, D_{2}, \ldots, D_{m} \subseteq C$ are uniquely determined by Algorithm 4. We have

$$
\operatorname{Pr}_{\mathbf{Y}}\left[C \subseteq V_{L}^{\mathbf{Y}}\right] \leq \operatorname{Pr}_{\mathbf{Y}}\left[\bigwedge_{j=1}^{m}\left(D_{j} \subseteq V_{L}^{\mathbf{Y}}\right)\right]
$$

We then use an analysis similar to the last subsection but focused on the $D_{j}$ 's. For each $1 \leq j \leq m$, each vertex in $D_{j}$ represents a hyperedge in the input hypergraph $H=(V, \mathcal{E})$. Let $d(j)=\left|D_{j}\right|$. Let $e_{1}^{j}, e_{2}^{j}, \ldots, e_{d(j)}^{j}$ denote the hyperedges in $D_{j}$. For each $1 \leq t \leq d(j)$, we define

$$
S_{t}^{j}:=e_{t}^{j} \backslash\left(\bigcup_{i \in[d(j)]: i \neq t} e_{i}^{j}\right) \quad \text { and } \quad R_{t}^{j}:=S_{t}^{j} \cap \Lambda .
$$

Since $H$ is linear, $\left|D_{j}\right| \leq \theta$, and $|\Lambda| \geq|V|-1$, it holds that

$$
\begin{equation*}
\forall 1 \leq t \leq d(j): \quad\left|R_{t}^{j}\right| \geq k-(\theta-1)-1=k-\theta \tag{5.18}
\end{equation*}
$$

Next, note that $D_{1}, D_{2}, \ldots, D_{m} \subseteq C$ is a set of disjoint connected components in the induced subgraph $G[D]$, where $D=C \backslash\left(\cup_{i=1}^{\ell} \Gamma_{G}\left(C_{i}\right)\right)=\cup_{i=1}^{m} D_{i}$. For any two distinct $1 \leq i, j \leq m, \operatorname{dist}_{G}\left(D_{i}, D_{j}\right) \geq 2$, as otherwise $D_{i}$ and $D_{j}$ must have been merged into one component. As $G=L_{H}[C]$ is a subgraph of $L_{H}$ induced by $C$, for any two distinct $1 \leq i, j \leq m, \operatorname{dist}_{L_{H}}\left(D_{i}, D_{j}\right) \geq 2$. Hence, for any hyperedge $e \in \mathcal{E}$ in $D_{i}$, any hyperedge $e^{\prime} \in \mathcal{E}$ in $D_{j}$, it holds that $e \cap e^{\prime}=\varnothing$. It implies that

$$
\begin{equation*}
\left(R_{t}^{j}\right)_{1 \leq j \leq m, 1 \leq t \leq d(j)} \text { are mutually disjoint. } \tag{5.19}
\end{equation*}
$$

Again, let $\mathcal{S}_{t}^{j}$ denote the set of all $s$ monochromatic configurations of $R_{t}^{j}$ (i.e. all vertices in $R_{t}^{j}$ taking the same value $c$, where $c \in[s]$ ). By Corollary 5.15 and (5.19),
it holds that

$$
\begin{aligned}
\operatorname{Pr}_{\mathbf{Y}}\left[C \subseteq V_{L}^{\mathbf{Y}}\right] & \leq \operatorname{Pr}_{\mathbf{Y}}\left[\bigwedge_{j=1}^{m}\left(D_{j} \subseteq V_{L}^{\mathbf{Y}}\right)\right] \leq \operatorname{Pr}_{\mathbf{Y}}\left[\bigwedge_{j=1}^{m} \bigwedge_{t=1}^{d(j)}\left(R_{t}^{j} \subseteq V_{L}^{\mathbf{Y}}\right)\right]=\operatorname{Pr}_{\mathbf{Y}}\left[\bigwedge_{j=1}^{m} \bigwedge_{t=1}^{d(j)}\left(Y_{R_{t}^{j}} \in \mathcal{S}_{t}^{j}\right)\right] \\
& \leq \prod_{j=1}^{m} \prod_{t=1}^{d(j)}\left(s\left(\frac{1}{s}+\frac{1}{q}\right)^{\left|R_{t}^{j}\right|} \exp \left(\frac{\left|R_{t}^{j}\right|}{k}\right)\right) \leq \prod_{j=1}^{m} \prod_{t=1}^{d(j)}\left(\mathrm{e} s\left(\frac{1}{s}+\frac{1}{q}\right)^{\left|R_{t}^{j}\right|}\right),
\end{aligned}
$$

where the last equation holds because $\left|R_{t}^{j}\right| \leq k$. Define

$$
R:=\bigcup_{j=1}^{m} \bigcup_{t=1}^{d(j)} R_{t}^{j}
$$

as the (disjoint) union of all $R_{t}^{j}$. By the lower bound in (5.18), we have

$$
|R| \geq \sum_{j=1}^{m} \sum_{t=1}^{d(j)}(k-\theta)=(k-\theta) \sum_{j=1}^{m} d(j)=(k-\theta)\left(M-\left|\bigcup_{i=1}^{\ell} \Gamma_{G}\left(C_{i}\right)\right|\right)
$$

where the last equation holds because $\left\{D_{i}\right\}_{1 \leq i \leq m}$ is a partition of $C \backslash\left(\cup_{i=1}^{\ell} \Gamma_{G}\left(C_{i}\right)\right)$ and $|C|=M$. Note that for any $1 \leq i \leq \ell,\left|C_{i}\right|=\theta$ and the maximum degree of the line graph $L_{H}$ is at most $k \Delta$. We have

$$
|R| \geq(k-\theta)(M-\ell \theta k \Delta) .
$$

This implies
$\operatorname{Pr}_{\mathbf{Y}}\left[C \subseteq V_{L}^{\mathbf{Y}}\right] \leq \prod_{j=1}^{m} \prod_{t=1}^{d(j)}\left(\mathrm{e} s\left(\frac{1}{s}+\frac{1}{q}\right)^{\left|R_{t}^{j}\right|}\right)=(\mathrm{e} s)^{\sum_{i=1}^{m} d(j)}\left(\frac{1}{s}+\frac{1}{q}\right)^{|R|} \leq(\mathrm{e} s)^{M}\left(\frac{1}{s}+\frac{1}{q}\right)^{(k-\theta)(M-\ell \theta k \Delta)}$,
where we use the fact $\sum_{i=1}^{m} d(j) \leq M$ in the last inequality. Since $C \in \operatorname{Con}_{u}^{(2)}(M)$, it holds that $\ell<\frac{M}{2 \theta k^{2} \Delta}$. Combining with the fact that $\frac{1}{s}+\frac{1}{q} \leq \frac{2}{s}$, we have

$$
\operatorname{Pr}_{\mathbf{Y}}\left[C \subseteq V_{L}^{\mathbf{Y}}\right] \leq(\mathrm{e} s)^{M}\left(\frac{2}{s}\right)^{(k-\theta)\left(M-\frac{M}{2 k}\right)} \leq(\mathrm{e} s)^{M}\left(\frac{2}{s}\right)^{(k-\theta) M}\left(\frac{s}{2}\right)^{\frac{M}{2}}
$$

In order to give a rough bound on the number of connected subgraphs containing $u$, we will use the following well-known result by Borgs, Chayes, Kahn, and Lovász [BCKL13].

Lemma 5.18 ([BCKL13, Lemma 2.1]). Let $G=(V, E)$ be a graph with maximum degree $d$ and $v \in V$ be a vertex. Then the number of connected induced subgraphs of size $\ell$ containing $v$ is at most $(e d)^{\ell-1} / 2$.

The maximum degree of $L_{H}$ is at most $k \Delta$. By Lemma 5.18, the number of connected subgraphs of size $M$ containing $u$ in $L_{H}$ is at most $(\mathrm{e} \Delta k)^{M-1} / 2$. Hence $\left|\operatorname{Con}_{u}^{(2)}(M)\right|<$ $(\mathrm{e} \Delta k)^{M}$. By a union bound over all $C \in \operatorname{Con}_{u}^{(2)}(M)$, we have

$$
\begin{aligned}
& \operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(2)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right] \\
& \quad \leq(\mathrm{e} \Delta k)^{M}(\mathrm{e} s)^{M}\left(\frac{2}{s}\right)^{(k-\theta) M}\left(\frac{s}{2}\right)^{\frac{M}{2}}=\left(\mathrm{e}^{2} s \Delta k\left(\frac{2}{s}\right)^{(k-\theta)}\right)^{M}\left(\frac{s}{2}\right)^{\frac{M}{2}} .
\end{aligned}
$$

We claim that

$$
\mathrm{e}^{2} s \Delta k\left(\frac{2}{s}\right)^{(k-\theta)} \leq \frac{1}{s}
$$

Since $s=\lceil\sqrt{q}\rceil$, it suffices to show that

$$
q \geq \mathrm{e}^{\frac{4}{k-\theta-2}} 2^{\frac{2(k-\theta)}{k-\theta-2}} k^{\frac{2}{k-\theta-2}} \Delta^{\frac{2}{k-\theta-2}},
$$

which is, in turn, implied by $\theta=\left\lceil\frac{4}{\delta}\right\rceil, k \geq \frac{20(\delta+1)}{\delta}$ and $q \geq 100 \Delta^{\frac{2+\delta}{k-4+\delta-3}}$. Hence, we have

$$
\operatorname{Pr}_{\mathbf{Y}}\left[\exists C \in \operatorname{Con}_{u}^{(2)}(M) \text { s.t. } C \subseteq V_{L}^{\mathbf{Y}}\right] \leq\left(\frac{1}{s}\right)^{M}\left(\frac{s}{2}\right)^{\frac{M}{2}} \leq\left(\frac{1}{2}\right)^{M}
$$

where the last inequality holds because $s \geq \sqrt{q} \geq 10$.

### 5.4.2 Properties of the 2-block-tree generator

We begin with validating Algorithm 4, namely proving Proposition 5.11.
Proof of Proposition 5.11. We claim that the algorithm always succeeds in Line 6 and Line 7, which implies that the size of $V$ strictly decreases in every step and the algorithm halts eventually. Moreover, if $|V|<\theta$, then all vertices in $V$ will be removed in Line 9 and Line 10 . Also, so long as $u_{\ell}$ and $C_{\ell}$ are selected according to some (arbitrary but) deterministic rule, the output is deterministic.

For the claim, first notice that $V \subseteq C$ throughout the algorithm. For Line 6 , since $G=L_{H}[C]$ is connected and $V \neq \varnothing, \Gamma_{G}(C \backslash V) \neq \varnothing$ and thus $u_{\ell}$ exists. For Line 7, $C_{\ell}$ exists as long as the connected component containing $u_{\ell}$ in $G[V]$ has size at least $\theta$. In the first iteration of the while-loop, this holds true as $|V|=|C|=M>\theta$ and $G[V]=G$ is connected. In all iterations thereafter, the size of the component cannot be smaller than $\theta$, as otherwise it would have been removed in the previous iteration at Line 9 and Line 10 .

We then prove Proposition 5.13. The following observation will be useful.
Proposition 5.19. Let $\ell>1$ and $u_{\ell}$ be the vertex selected in Line 6. Then there exists some $1 \leq j<\ell$ such that $\operatorname{dist}_{G}\left(C_{j}, u_{\ell}\right)=2$.

Proof. Assume for contradiction that $\operatorname{dist}_{G}\left(C_{j}, u_{\ell}\right)>2$ for all $1 \leq j<\ell$. Consider the set $V$ when $u_{\ell}$ is selected. Because of Line 6 , we can find one of $u_{\ell}^{\prime} s$ neighbours that is in $C \backslash V$, say $v$. Consider the reason why $v$ was removed from $V$. If this happened on Line 8, then there must have been some $i$ such that $v \in C_{i}$ or $v \in \Gamma_{G}\left(C_{i}\right)$. The former case implies that $u_{\ell}$ must have been removed from $V$, which is impossible. The latter case indicates $\operatorname{dist}_{G}\left(C_{i}, u_{\ell}\right)=2$, a contradiction. Therefore, $v$ was removed in Line 10. However, this implies that $u_{\ell}$ would have been removed from $V$ too, because $u_{\ell}$ and $v$ must have been in the same component $V^{\prime}$, which is also a contradiction.

Proof of Proposition 5.13. The first part of this proposition requires us to verify that $\left\{C_{1}, \cdots, C_{\ell}\right\}$ is a 2-block-tree in $L_{H}$. To do so, we verify Items (B1), (B2), and (B3) of Definition 5.12 next. Notice that what we need to prove here is with respect to $L_{H}$, instead of $G=L_{H}[C]$.

- Item (B1) holds due to how $C_{i}$ is constructed in Line 7.
- For Item (B2), we first show $\operatorname{dist}_{G}\left(C_{i}, C_{j}\right) \geq 2$. For any $C_{i}$ generated by Algorithm 4, it is ensured that $\Gamma_{G}\left(C_{i}\right)$ gets removed from $V$, and therefore, no vertex in $\Gamma_{G}\left(C_{i}\right)$ will be in $C_{j}$ for any other $j$. To show $\operatorname{dist}_{L_{H}}\left(C_{i}, C_{j}\right) \geq 2$, note that $G$ is an induced subgraph of $L_{H}$. Any two vertices of distance more than 1 in $G$ cannot be neighbours in $L_{H}$, and this implies dist $L_{H}\left(C_{i}, C_{j}\right) \geq 2$.
- To verify (B3), it suffices to show that $\left\{C_{1}, \cdots, C_{\ell}\right\}$ is connected in $G^{2}$, because $G$ is a subgraph of $L_{H}$. This follows from a simple induction. Suppose $\left\{C_{1}, \cdots, C_{i}\right\}$, in the order of being generated by the algorithm, is connected in $G^{2}$. The base case of $i=1$ holds since $C_{1}$ is connected. Now consider $C_{i+1}$. By Proposition 5.19, there exists some $j$ such that $\operatorname{dist}_{G}\left(C_{i+1}, C_{j}\right)=2$, which implies that $\left\{C_{1}, \cdots, C_{i+1}\right\}$ is connected in $G^{2}$ as well.

For the second part, suppose towards contradiction that there is some connected component $C^{*}$ in $G\left[C^{\prime}\right]$ of size greater than $\theta$. All vertices in $C$ must have been removed from $V$ when the algorithm halts, according to Proposition 5.11. However, $C^{*}$ cannot be $C_{i}$ for any $i$, because $\left|C_{i}\right|=\theta$. It cannot contain any vertex in $\Gamma_{G}\left(C_{i}\right)$ either by the definition of $C^{\prime}$. Thus, no vertex in $C^{*}$ can be removed in Line 8, and
all vertices in $C^{*}$ must have been removed from $V$ in Line 10 . Because $C^{*}$ does not contain any vertex from either $C_{i}$ or $\Gamma_{G}\left(C_{i}\right)$, it does not split into smaller components whilst the algorithm is executed. Thus, the whole $C^{*}$ must have been removed from $V$ in a single step, which means $\left|C^{*}\right|<\theta$, a contradiction.

### 5.4.3 Property of random configurations

Proof of Lemma 5.14. Recall that $\mathbf{Y} \in[s]^{\Lambda}$, defined in (5.6), is the configuration at time $t-1$ on $\Lambda$. For each vertex $w \in V$, let $t(w)$ denote $\max _{1 \leq t^{\prime}<t}$ such that vertex $w$ is updated by the systematic scan in the $t^{\prime}$-th step (i.e. the label of $w$ is $t^{\prime} \bmod n$ ), and let $t(w)=0$ when such $t^{\prime}$ does not exist. With this notation $Y_{w}=Y_{t(w)}(w)$ for all $w \in \Lambda$. We assume $R=\left\{w_{1}, w_{2}, \ldots, w_{|R|}\right\}$ such that $t\left(w_{1}\right) \leq t\left(w_{2}\right) \leq \ldots \leq t\left(w_{|R|}\right)$. By the chain rule, we have $\operatorname{Pr}\left[\mathbf{Y}_{R}=\sigma\right]=\prod_{i=1}^{|R|} p_{i}$, where $p_{i}=\operatorname{Pr}\left[Y_{w_{i}}=\sigma_{w_{i}} \mid \bigwedge_{j=1}^{i-1} Y_{w_{j}}=\sigma_{w_{j}}\right]$. We now bound the value of each $p_{i}$ as follows. If $t\left(w_{i}\right)=0$, then it holds that $p_{i} \leq \frac{\lceil q / s\rceil}{q}$. If $t\left(w_{i}\right)>0$, then in the $t\left(w_{i}\right)$-th iteration, the algorithm first samples $X_{w_{i}}^{\prime}$ using Sample, and then sets $Y_{w_{i}}=h\left(X_{w_{i}}^{\prime}\right)$. Denote $\mathbf{Y}^{\prime}=\mathbf{Y}_{t\left(w_{i}\right)-1}\left(V \backslash\left\{w_{i}\right\}\right)$. There are two sub-cases:

- if $X_{w_{i}}^{\prime}$ is returned by Line 4 or Line 8 in Sample, then $X_{w_{i}}^{\prime}$ is sampled uniformly at random from $[q]$, which implies that $p_{i} \leq \frac{\lceil q / s]}{q}$;
- if $X_{w_{i}}^{\prime}$ is returned by Line 9 in Sample, by property 2 of Lemma 5.9, $X_{w_{i}}^{\prime}$ is sampled from the correct conditional distribution $\mu_{w_{i}}^{\mathbf{Y}^{\prime}}$. Note that for any $\tau \in$ $[s]^{V \backslash\left\{w_{i}\right\}}, \mu_{w_{i}}^{\tau}$ is the marginal distribution induced by a list hypergraph colouring instance where the colour list of any $w \neq w_{i}$ is $h^{-1}(\tau(w))$, where $h$ is the projection scheme, and $w_{i}$ 's colour list is [q]. By Definition 5.3 of projection schemes, for any $w \neq w_{i},\left|h^{-1}(\tau(w))\right| \geq\lfloor q / s\rfloor$. In other words, the upper bound on the size of the lists is $q$ and the lower bound is $\lfloor q / s\rfloor$. Since $\lfloor q / s\rfloor^{k} \geq 2 \mathrm{e} q k \Delta$, by Lemma 5.2, it holds that for all $\tau \in[s]^{V \backslash\left\{w_{i}\right\}}, c \in[q]$,

$$
\operatorname{Pr}\left[X_{w}^{\prime}=c \mid \mathbf{Y}^{\prime}=\tau \wedge X_{w_{i}}^{\prime} \text { is returned by Line } 9\right] \leq \frac{1}{q} \exp \left(\frac{1}{k}\right)
$$

which implies $p_{i} \leq \frac{\lceil q / s\rceil}{q} \exp \left(\frac{1}{k}\right)$.
Combining all the cases together, we have
$\operatorname{Pr}\left[\mathbf{Y}_{R}=\sigma\right] \leq\left(\frac{\lceil q / s\rceil}{q}\right)^{|R|} \exp \left(\frac{|R|}{k}\right) \leq\left(\frac{q / s+1}{q}\right)^{|R|} \exp \left(\frac{|R|}{k}\right)=\left(\frac{1}{s}+\frac{1}{q}\right)^{|R|} \exp \left(\frac{|R|}{k}\right)$.

### 5.4.4 Properties of 2-block-trees

In this subsection, we show Lemma 5.16 and Lemma 5.17. We begin with the first one, which is a simple observation.

Proof of Lemma 5.16. Given a 2-block-tree $\left\{C_{1}, \cdots, C_{\ell}\right\}$ of $G$ and the vertex $v$, construct the following graph $G_{C}$. Each vertex $u_{j}$ of $G_{C}$ corresponds to a block $C_{j}$, and two vertices $u_{j}, u_{j^{\prime}}$ are adjacent if and only if $\operatorname{dist}_{G}\left(C_{j}, C_{j^{\prime}}\right)=2$. By the definition of 2-block-tree, the graph $G_{C}$ is connected. Therefore, we can take an arbitrary spanning tree of it. To select the $C_{i}$ to drop, note that any tree containing at least 2 vertices has at least 2 vertices of degree 1 . Therefore, we just choose $u_{i}$ to be one such vertex where $v \notin C_{i}$. The rest of the tree is still connected, and so is $G_{C}-u_{i}$, which indicates that $\left\{C_{1}, \cdots, C_{\ell}\right\}-C_{i}$ still forms a 2-block-tree that contains $v$.

We proceed to show Lemma 5.17. We may apply Lemma 5.18 on $G^{2}$ due to property (B3). Unfortunately, this yields roughly $\left(e d^{2}\right)^{\theta \ell}$ and does not suffice for our purpose. Here, we give a refined estimation inspired by the original embedding argument of [Sta99, BCKL13].

Let $d^{\prime}:=(e d)^{\theta-1} / 2$, which, by Lemma 5.18 , upper bounds the number of size- $\theta$ connected induced subgraphs containing a given vertex in a graph with maximum degree $d$. Therefore, given $v$, we can encode each connected induced subgraph containing $v$ with a positive integer $\Xi \in\left[d^{\prime}\right]$. In other words, there exists an injective mapping $\Upsilon_{v}$ from all connected induced subgraphs of $G$ containing $v$ to $\{v\} \times\left[d^{\prime}\right]$.

Our counting argument will be based on encoding the whole 2-block-tree. Intuitively, the encoding contains $\ell+1$ components. The first one encodes how $C_{i}$ 's are connected in $G^{2}$, and the rest encodes each individual $C_{i}$ by an integer in [ $\left.d^{\prime}\right]$.

Let $\mathbb{T}_{\theta d^{2}}$ to be the infinite $\theta d^{2}$-ary tree. In the first step, the relation between blocks is encoded by a subtree of $\mathbb{T}_{\theta d^{2}}$ containing its root, which is basically a DFS tree. However, the order of visiting will affect the DFS tree we construct. For this reason, we need to specify this ordering. First, we order the vertices by their indices. That is, $v_{i}<v_{j}$ if $i<j$. Given a subset $C$ of vertices, consider the set $\Gamma^{2}(C)$ containing vertices of distance 2 from $C$. We can sort this set according to the ordering of vertices, and hence any vertex $u \in \Gamma^{2}(C)$ has a rank among $\Gamma^{2}(C)$, denoted by $\operatorname{Rank}_{C}(u)$. Suppose at some stage of our DFS algorithm, we have just finished handling some block $C$. Then we find the next unvisited vertex in $\Gamma^{2}(C)$, say $v^{\prime}$, which is in some block $C^{\prime}$ that needs to be encoded. Then $C^{\prime}$ will be encoded as the $\operatorname{Rank}_{C}\left(v^{\prime}\right)$-th child of current vertex in the DFS tree, together with the integer $\Upsilon_{v^{\prime}}\left(C^{\prime}\right) \in\left[d^{\prime}\right]$. The key of our proof
is to show that this encoding is injective, i.e., no two distinct 2-block-trees share the same encoding.

With all the preparation, we give the encoding algorithm as Algorithm 5. Once again, Algorithm 5 is for analysis only and does not need to be implemented.

```
Algorithm 5: Encoding
    Input: A graph \(G\), a vertex \(v \in G\), a 2-block-tree \(\left\{C_{1}, \cdots, C_{\ell}\right\}\) of block size \(\theta\)
        and tree size \(\ell\)
    Output: An encoding \(\left(T, \Xi_{1}, \cdots, \Xi_{\ell}\right)\), where \(T\) is a subtree of \(\mathbb{T}_{\theta d^{2}}\) of size \(\ell\)
    initialize visited [1. . \(\ell\) ] to be all False;
    let \(C_{j}\) be the component containing \(v\);
    let \(r\) be the root of \(\mathbb{T}_{\theta d^{2}}\);
    let \(T\) be an empty subtree;
    \(t \leftarrow 0\);
    DFS-Encode ( \(j, v, r\) );
    return \(\left(T, \Xi_{1}, \cdots, \Xi_{\ell}\right)\);
    Procedure DFS-Encode(i,u,w):
        visited[i] \(\leftarrow\) True;
        \(t \leftarrow t+1\);
        \(\Xi_{t} \leftarrow \Upsilon_{u}\left(C_{i}\right) ;\)
        add \(w\) into \(T\);
        for \(u^{\prime} \in \Gamma^{2}\left(C_{i}\right)\) do // enumerate \(u^{\prime} \in \Gamma^{2}\left(C_{i}\right)\) in order
            if there does not exist any \(i^{\prime}\) such that \(C_{i^{\prime}} \ni u^{\prime}\) then
                    continue;
            let \(i^{\prime}\) be the index such that \(C_{i^{\prime}} \ni u^{\prime}\);
            if visited[i']=False then
                    let \(w^{\prime}\) be the \(\operatorname{Rank}_{C_{i}}\left(u^{\prime}\right)\)-th child of \(w\) in \(\mathbb{T}_{\theta d^{2}}\);
                    DFS-Encode ( \(\left.i^{\prime}, u^{\prime}, w^{\prime}\right)\);
```

Lemma 5.20. Fix a graph $G$ and a vertex $v$. Any 2 -block-tree $\left\{C_{1}, \cdots, C_{\ell}\right\}$ of block size $\theta$ and tree size $\ell$ containing $v$ can be encoded by a tuple $\left(T, \Xi_{1}, \cdots, \Xi_{\ell}\right)$, where $T$ is a subtree of $\mathbb{T}_{\theta d^{2}}$ of size $\ell$ containing its root, and $\Xi_{i} \in\left[d^{\prime}\right]$. Moreover, no two distinct 2 -block-trees share the same encoding.

Proof. The first part of this lemma follows by going through Algorithm 5. There are two things to verify:

- The algorithm will always halt, outputting $\ell \Xi_{i}$ 's. To show this, one only needs to check that every $C_{i}$ will be visited exactly once, which is true due to property (B3) of Definition 5.12 and Line 17 of Algorithm 5.
- The algorithm can find such $w^{\prime}$ on Line 18 , or equivalently, $\operatorname{Rank}_{C_{i}}\left(u^{\prime}\right) \in\left[\theta d^{2}\right]$. This follows after a trivial upper bound on the number of distance-2 neighbours that $\left|\Gamma^{2}\left(C_{i}\right)\right| \leq \theta d^{2}$.

To prove the second part, suppose there are two 2-block-trees $\left\{C_{1}, \cdots, C_{\ell}\right\}$ and $\left\{C_{1}^{\prime}, \cdots, C_{\ell}^{\prime}\right\}$ with the same encoding $\left(T, \Xi_{1}, \cdots, \Xi_{\ell}\right)$. Without loss of generality, we can assume $C_{1}, \cdots, C_{\ell}$ (resp. $C_{1}^{\prime}, \cdots, C_{\ell}^{\prime}$ ) are sorted in the order of being visited by Algorithm 5. The goal is then to prove $C_{i}=C_{i}^{\prime}$ for all $i \in[\ell]$. To show this, we do a simple induction argument. More precisely, denote by $T_{t}$ and $T_{t}^{\prime}$ the subtrees constructed by the first $t$ calls to DFS-Encode respectively. We induce on $t$ to show that

$$
\begin{equation*}
C_{i}=C_{i}^{\prime} \text { for all } i \in[t] \text {, and } T_{t}=T_{t}^{\prime} . \tag{IH}
\end{equation*}
$$

Base case $t=1$. Note that $C_{1}=C_{1}^{\prime}$ follows from the injectivity of $\Upsilon_{v}$, and $T_{1}=T_{1}^{\prime}$ as they both contain only the root.

Induction step. Suppose (IH) holds for $t-1$. At this stage, we compare the progress of two copies of Encoding running on $C$ and $C^{\prime}$ respectively. Right before the for-loop in the ( $t-1$ )-th call to DFS-Encode, both copies get the same $w$ by (IH). Again by (IH), both copies get the same $C_{t-1}$ in the condition of the for-loop. In the enumeration of for-loop, both copies skip or keep the $u^{\prime}$ in Line 14 simultaneously, because $C_{i}=C_{i}^{\prime}$ for all $i \in[t-1]$. Note that each vertex of $\mathbb{T}_{\theta d^{2}}$ can be visited at most once. This means that if the two copies get different $u^{\prime}$ in Line 18 , then the final subtree will be different. Therefore, they must get the same $u^{\prime}$ and $i^{\prime}$, and hence the same $w^{\prime}$ because they have the same $C_{t-1}$, implying $T_{t}=T_{t}^{\prime}$. Moreover, the next calls to DFS-Encode have an identical input in both copies. Thus, $\Xi_{t}=\Upsilon_{u}\left(C_{t}\right)$ and $\Xi_{t}^{\prime}=\Upsilon_{u}\left(C_{t}^{\prime}\right)$. By assumption $\Xi_{t}=\Xi_{t}^{\prime}$. Injectivity of $\Upsilon_{u}$ implies that $C_{t}=C_{t}^{\prime}$, finishing the proof.

We conclude this subsection by proving Lemma 5.17.

Proof of Lemma 5.17. By Lemma 5.20, the number of 2-block-trees can be upperbounded by the number of possible encodings. To count the number of possible subtrees $T$, we simply apply Lemma 5.18 , which gives $\left(e \theta d^{2}\right)^{\ell-1} / 2$. The number of possible $\Xi_{i}$ sequences is $d^{\ell \ell}=(e d)^{\ell(\theta-1)} / 2^{\ell}$. Combining both parts yields the upper bound $\theta^{\ell-1} e^{\theta \ell-1} d^{(\theta+1) \ell-2} / 2^{\ell+1}$.

### 5.5 Mixing of systematic scan

In this section, we prove the mixing lemma for the projected systematic scan Markov chain of hypergraph colourings (Lemma 5.8). First, we verify that the systematic scan is irreducible, aperiodic with respect to $v$. This implies that the systematic scan has a unique stationary distribution, and it is straightforward to verify that $v$ is indeed the stationary distribution. Aperiodicity is also straightforward to verify. For irreducibility, it suffices to show that for any $\tau \in[s]^{V}, v(\tau)>0$, as our chain is a Glauber dynamics for $v$. Fix an arbitrary configuration $\tau \in[s]^{V}$. We show that there exists a proper colouring $\sigma \in[q]^{V}$ such that $h(\sigma)=\tau$, where $h$ is the projection scheme. This implies $v(\tau)>0$. To prove the existence of such a proper colouring, consider the list hypergraph colouring instance $\left(H,\left(Q_{v}\right)_{v \in V}\right)$, where $Q_{v}=h^{-1}\left(\tau_{v}\right)$ for all $v \in V$. We only need to show that this list colouring instance has a feasible solution. Note that $\left|Q_{v}\right| \geq\lfloor q /\lceil\sqrt{q}\rceil\rfloor \geq \sqrt{q} / 2$ for $q \geq 20$. By the Lovász local lemma, Theorem 2.4, we only need to verify that

$$
\mathrm{e} q\left(\frac{2}{\sqrt{q}}\right)^{k} \Delta k \leq 1
$$

which follows from $q \geq 40 \Delta^{\frac{2}{k-4}}$ and $k \geq 20$.
Next, we prove the mixing time result in Lemma 5.8. The analysis is based on an information percolation argument. We first define a coupling $C$ of the systematic scan $\left(\mathbf{X}_{t}, \mathbf{Y}_{t}\right)_{t \geq 0}$. Let $\mathbf{X}_{0}, \mathbf{Y}_{0} \in[s]^{V}$ be two arbitrary initial configurations. In the $t$-th transition step,

- let $v \in V$ be the vertex with label $(t \bmod n)$ and $\operatorname{set}\left(X_{t}(u), Y_{t}(u)\right) \leftarrow\left(X_{t-1}(u), Y_{t-1}(u)\right)$ for all other vertices $u \in V \backslash\{v\}$;
- sample $\left(X_{t}(v), Y_{t}(v)\right)$ from the optimal coupling between $v_{v}^{X_{t-1}(V \backslash\{v\})}$ and $v_{v}^{Y_{t-1}(V \backslash\{v\})}$.

We prove the following lemma in this section.

Lemma 5.21. Suppose $k \geq 20$ and $q \geq 40 \Delta^{\frac{2}{k-4}}$. For any initial configurations $\mathbf{X}_{0}, \mathbf{Y}_{0} \in$ $[s\rceil^{V}$, any $\epsilon \in(0,1)$, let $T=\left\lceil 50 n \log \frac{n \Delta}{\epsilon}\right\rceil$, it holds that

$$
\forall v \in V, \quad \operatorname{Pr}_{C}\left[X_{T}(v) \neq Y_{T}(v)\right] \leq \frac{\epsilon}{n} .
$$

By Lemma 5.21, a union bound over all vertices and the coupling lemma (Lemma 2.3), it holds that

$$
\max _{\mathbf{X}_{0}, \mathbf{Y}_{0} \in[s]^{V}} d_{\mathrm{TV}}\left(\mathbf{X}_{T}, \mathbf{Y}_{T}\right) \leq \operatorname{Pr}_{C}\left[\mathbf{X}_{T} \neq \mathbf{Y}_{T}\right] \leq \epsilon,
$$

which proves the mixing time part of Lemma 5.8 via (2.5). In the rest of this section, we use the information percolation technique to analyse the coupling $C$ and prove Lemma 5.21.

### 5.5.1 Information percolation analysis

Consider the coupling procedure $\left(\mathbf{X}_{t}, \mathbf{Y}_{t}\right)_{t \geq 0}$. For each $t \geq 1$, let $v_{t}$ denote the vertex picked in the $t$-th step of systematic scan, namely, $v_{t}$ is the vertex with label ( $t$ $\bmod n)$. Consider the $t$-th transition step, where $t>0$. Define the set of agreement vertices when updating $v_{t}$ at time $t$ by

$$
A_{t}:=\left\{v \in V \backslash\left\{v_{t}\right\} \mid X_{t-1}(v)=Y_{t-1}(v)\right\} .
$$

We say a hyperedge $e \in \mathcal{E}$ is satisfied by $A_{t}$ if there exist two distinct vertices $u, v \in$ $e \cap A_{t}$ such that $X_{t-1}(u) \neq X_{t-1}(v)$ (and hence $\left.Y_{t-1}(u) \neq Y_{t-1}(v)\right)$. We remove all the hyperedges $e \in \mathcal{E}$ satisfied by $A_{t}$ to obtain a sub-hypergraph $H_{t}$. Let $H_{t}^{v}$ denote the connected component in $H_{t}$ containing $v$.

Lemma 5.22. If $X_{t}\left(v_{t}\right) \neq Y_{t}\left(v_{t}\right)$ for some $t \geq 1$, then there exists $u \neq v_{t}$ in $H_{t}^{v_{t}}$ such that $X_{t-1}(u) \neq Y_{t-1}(u)$.

Proof. Note that $X_{t}\left(v_{t}\right)$ and $Y_{t}\left(v_{t}\right)$ are sampled from $v_{v_{t}}^{X_{t-1}\left(V \backslash\left\{v_{t}\right\}\right)}$ and $v_{v_{t}}^{Y_{t-1}\left(V \backslash\left\{v_{t}\right\}\right)}$ respectively. Let $\mu^{\prime}$ denote the uniform distribution of proper colourings of $H_{t}^{v}$. Let $\pi$ denote the projected distribution induced by $\mu^{\prime}$ and the projection scheme $h$. Let $V_{t}^{v_{t}}$ denote the vertex set of $H_{t}^{v_{t}}$ and let $S=V_{t}^{v_{t}} \backslash\left\{v_{t}\right\}$. We claim that (1) $v_{v_{t}}^{X_{t-1}\left(V \backslash\left\{v_{t}\right\}\right)}$ and $\pi_{v_{t}}^{X_{t-1}(S)}$ are identical distributions; (2) $v_{v_{t}}^{Y_{t-1}\left(V \backslash\left\{v_{t}\right\}\right)}$ and $\pi_{v_{t}}^{Y_{t-1}(S)}$ are identical distributions. Hence, if $X_{t-1}(u)=Y_{t-1}(u)$ for all $u \neq v_{t}$ in $H_{t}^{v_{t}}$, then $X_{t}\left(v_{t}\right)$ and $Y_{t}\left(v_{t}\right)$ must be perfectly coupled.

We verify that $v_{v_{t}}^{X_{t-1}\left(V \backslash\left\{v_{t}\right\}\right)}$ and $\pi_{v_{t}}^{X_{t-1}(S)}$ are identical distributions. The claim for $v_{v_{t}}^{Y_{t-1}\left(V \backslash\left\{v_{t}\right\}\right)}$ and $\pi_{v_{t}}^{Y_{t-1}(S)}$ can be verified by a similar proof. Consider the list colouring instance $\left(H,\left(Q_{v}\right)_{v \in V}\right)$, where $Q_{v}=[q]$ if $v=v_{t}$ and $Q_{v}=h^{-1}\left(X_{t-1}(v)\right)$ if $v \neq v_{t}$. Let $\mu_{\text {list }}$ denote the uniform distribution of all proper list colourings. If $X \sim \mu_{\text {list }}$, then $h\left(X_{v_{t}}\right) \sim v_{v_{t}}^{X_{t-1}\left(V \backslash\left\{v_{t}\right\}\right)}$. For any hyperedge $e$ satisfied by $A_{t}$, it holds that for any colouring $X \in \otimes_{v \in V} Q_{V}, e$ is not monochromatic. Let $H_{t}$ denote the hypergraph obtained from $H$ by removing all hyperedges satisfied by $A_{t}$. Hence, $\left(H,\left(Q_{v}\right)_{v \in V}\right)$ and $\left(H_{t},\left(Q_{v}\right)_{v \in V}\right)$ have the same set of proper list colourings. Recall that $H_{t}^{v_{t}}$ is the connected component in $H_{t}$ containing vertex $v_{t}$. Let $\mu_{\text {list }}^{\text {com }}$ denote the uniform distribution over all proper list colourings of $\left(H_{t}^{v_{t}},\left(Q_{v}\right)_{v \in V_{t}^{v_{t}}}\right)$. Hence, $\mu_{\text {list }}$ projected on $v_{t}$ is the same distribution as $\mu_{\text {list }}^{\text {com }}$ projected on $v_{t}$. If $X \sim \mu_{\text {list }}^{\text {com }}$, then $h\left(X_{v_{t}}\right) \sim \pi_{v_{t}}^{X_{t-1}(S)}$. This implies that $v_{v_{t}}^{X_{t-1}\left(V \backslash\left\{v_{t}\right\}\right)}$ and $\pi_{v_{t}}^{X_{t-1}(S)}$ are identical distributions.

We say that a hyperedge sequence $e_{1}, e_{2}, \ldots, e_{\ell}$ is a path in a hypergraph if for each $1<i \leq \ell, e_{i-1} \cap e_{i} \neq \varnothing$ and $e_{i-1} \neq e_{i}$. The following result is a straightforward corollary of Lemma 5.22.

Corollary 5.23. Let $t \geq 1$. If $X_{t}\left(v_{t}\right) \neq Y_{t}\left(v_{t}\right)$, then there exists a vertex $u \neq v_{t}$ satisfying $X_{t-1}(u) \neq Y_{t-1}(u)$ and $a$ path $e_{1}, e_{2}, \ldots, e_{\ell}$ in hypergraph $H$ such that

- $v \in e_{1}$ and $u \in e_{\ell} ;$
- for any hyperedge $e_{i}$ in the path, there exists $c \in[s]$ such that for all vertex $w \in e_{i}$ and $w \neq v_{t}$, either $X_{t-1}(w)=Y_{t-1}(w)=c$ or $X_{t-1}(w) \neq Y_{t-1}(w)$.

Proof. By Lemma 5.22, there is a vertex $u \neq v_{t}$ such that $X_{t-1}(u) \neq Y_{t-1}(u)$ and $u \in H_{t}^{v_{t}}$. As $u$ and $v_{t}$ are in the same connected component, there exist a path from $v_{t}$ to $u$. Moreover, for each hyperedge $e_{i}$ on this path, since $e_{i}$ is in $H_{t}^{v_{t}}$, it is not satisfied by $A_{t}$. This implies that for all $w \neq v_{t} \in e_{i}$ such that $X_{t-1}(w)=Y_{t-1}(w)$, their values in both chains must be the same $c \in[s]$. Lastly, note that any path in $H_{t}^{v_{t}}$ is also a path in $H$. This proves the corollary.

Corollary 5.23 is a key result for the information percolation analysis. For any time $0 \leq t \leq T$, any vertex $v \in V$, define the set of previous update times by

$$
S(v, t):=\left\{1 \leq i \leq t \mid v_{i}=v\right\}
$$

where $v_{i}$ is the vertex picked in the $i$-th transition step. Define the last update time for $v$ up to $t$ by

$$
\operatorname{time}_{u d}(v, t):= \begin{cases}\max _{i \in S(v, t)} i & \text { if } S(v, t) \neq \varnothing \\ 0 & \text { otherwise }\end{cases}
$$

By Corollary 5.23, if the coupling on vertex $v$ failed at time $t$, then there must exist a vertex $u$ such that the coupling on $u$ failed at time $t^{\prime}=$ time $_{\text {ud }}(u, t)$. We apply Corollary 5.23 recursively until we find a vertex $w$ such that $X_{0}(w) \neq Y_{0}(w)$. This gives us an update time sequence $t=t_{1}>t_{2}>\ldots>t_{\ell}=0$ such that the coupling of each $t_{i}$-th transition fails, together with a set of paths satisfying the properties in Corollary 5.23. We will show that such a update time sequence and the set of paths occur with small probability, which bounds the probability of $X_{t}\left(v_{t}\right) \neq Y_{t}\left(v_{t}\right)$. For this analysis, we will use the notions of extended hyperedges and extended hypergraphs introduced by He, Sun, and Wu [HSW21].

### 5.5.2 Extended hyperedges and the extended hypergraph

Fix an integer $T \geq 1$ to be the total number of transitions of the systematic scan. Define the set of extended vertex $V^{\text {ext }}$ by

$$
V^{\mathrm{ext}}=\left\{\left(t, v_{t}\right) \mid 1 \leq t \leq T\right\} \cup\{(0, v) \mid v \in V\},
$$

where $v_{t}$ is the vertex with label $(t \bmod n)$. Each vertex $(t, u) \in V^{\text {ext }}$ represents an update, i.e. $u$ is updated at the $t$-th transition step. We regard all vertices "updated" at the initial step $(t=0)$. Consider the systematic scan process $\left(\mathbf{X}_{t}\right)_{t \geq 0}$. For any hyperedge $e \in \mathcal{E}$, the configuration $X_{t}(e)$ of $e$ at time $t$ satisfies

$$
\forall u \in e, \quad X_{t}(u)=X_{t^{\prime}}(u), \quad \text { where } t^{\prime}=\operatorname{time}_{\mathrm{ud}}(u, t),
$$

namely, the value of $u$ at time $t$ is the same as the value of $u$ at time $t^{\prime}=\operatorname{time}_{u d}(u, t)$. Besides, the configuration of hyperedge $e$ remains unchanged until some vertex in $e$ is updated. This motivates the following definition of extended hyperedges and the extended hypergraph, introduced by He , Sun, and Wu [HSW21].

Definition 5.24. The set $\mathcal{E}^{\text {ext }}$ of extended hyperedges is defined by $\mathcal{E}^{\text {ext }}:=\cup_{t=0}^{T} \mathcal{E}_{t}^{\text {ext }}$, where

$$
\begin{gathered}
\mathcal{E}_{0}^{\mathrm{ext}}:=\bigcup_{e \in \mathcal{E}}\{(0, v) \mid v \in e\}, \\
\forall 1 \leq t \leq T, \quad \mathcal{E}_{t}^{\mathrm{ext}}:=\bigcup_{e: v_{t} \in e}\left\{\left(t^{\prime}, v\right) \mid v \in e \wedge t^{\prime}=\operatorname{time}_{\mathrm{ud}}(v, t)\right\} .
\end{gathered}
$$

The extended hypergraph is $H^{\text {ext }}=\left(V^{\text {ext }}, \mathcal{E}^{\text {ext }}\right)$.
At the beginning, each hyperedge $e \in \mathcal{E}$ takes its initial value, and thus we add all the extended hyperedges with $t=0$ to $\mathcal{E}_{0}^{\text {ext }}$. For each update at time $1 \leq t \leq T$, only the value of $v_{t}$ is updated. Thus the configurations of only the hyperedges containing $v_{t}$ are updated, and we add only those to $\mathcal{E}_{t}^{\text {ext }}$.

Corollary 5.23 shows that for any $t \geq 1$, if the coupling in the $t$-th transition step fails (i.e. $X_{t}\left(v_{t}\right) \neq Y_{t}\left(v_{t}\right)$ ), then we can find a specific path in the hypergraph $H$. Our next lemma lifts such a path to $H^{\text {ext }}$.

Lemma 5.25. Let $1 \leq t \leq T$ be an integer. Suppose $X_{t}\left(v_{t}\right) \neq Y_{t}\left(v_{t}\right)$. There exist a vertex $\left(t^{\prime}, u\right) \in V^{\text {ext }}$ satisfying $t^{\prime}<t$ and $X_{t^{\prime}}(u) \neq Y_{t^{\prime}}(u)$, together with a path $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$ in $H^{\text {ext }}$ such that

- $\left(t, v_{t}\right) \in e_{1}^{\text {ext }}$ and $\left(t^{\prime}, u\right) \in e_{\ell}^{\text {ext }} ;$
- for any hyperedge $e_{i}^{\text {ext }}$ in the path, there exists $c \in[s]$ such that for all $(j, w) \in e_{i}^{\text {ext }}$, either $X_{j}(w)=Y_{j}(w)=c$ or $X_{j}(w) \neq Y_{j}(w)$.

Proof. Let $u$ and $e_{1}, e_{2}, \ldots, e_{\ell}$ denote the vertex and the path in Corollary 5.23 respectively. For each vertex $w \in V$, let $t_{w}=\operatorname{time}_{\mathrm{ud}}(w, t)$. For each $1 \leq i \leq \ell$, define

$$
e_{i}^{\operatorname{ext}}=\left\{\left(t_{w}, w\right) \mid w \in e_{i}\right\} .
$$

To show that $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$ is a path in $H^{\text {ext }}$, we need to verify that each $e_{i}^{\text {ext }}$ defined above belongs to $\mathcal{E}^{\text {ext }}$ in Definition 5.24. Fix an $e_{i}^{\text {ext }}$. Let $t_{\max }=\max \left\{t \mid(t, w) \in e_{i}^{\text {ext }}\right\}$. It is straightforward to verify that $e_{i}^{\text {ext }} \in \mathcal{E}_{t_{\text {max }}}^{\text {ext }}$.

Next, we show that $t^{\prime}<t$ and $X_{t^{\prime}}(u) \neq Y_{t^{\prime}}(u)$. By definition, we have $t^{\prime}=t_{u}=$ time $_{\text {ud }}(u, t)<t$. As the value of any vertex does not change until the next update, we have that

$$
\begin{equation*}
\forall w \in V \backslash\left\{v_{t}\right\}, \quad X_{t-1}(w)=X_{t_{w}}(w) \text { and } Y_{t-1}(w)=Y_{t_{w}}(w) . \tag{5.20}
\end{equation*}
$$

By Corollary 5.23, it holds that $X_{t-1}(u) \neq Y_{t-1}(u)$. By (5.20), it holds that $X_{t^{\prime}}(u) \neq$ $Y_{t^{\prime}}(u)$.

Finally, we verify the two properties of the path. The first property $\left(t, v_{t}\right) \in e_{1}^{\text {ext }}$ and $\left(t^{\prime}, u\right) \in e_{\ell}^{\text {ext }}$ follows from the way $e_{i}^{\text {ext }}$ is constructed. By Corollary 5.23, for any $e_{i}$ in the path, there exists $c \in[s]$ such that for all vertices $w \in e_{i} \backslash\left\{v_{t}\right\}$, either $X_{t-1}(w)=$ $Y_{t-1}(w)=c$ or $X_{t-1}(w) \neq Y_{t-1}(w)$. By (5.20), for all extended vertices $(i, w) \in e_{i}^{\text {ext }}$ with $w \neq v_{t}$, either $X_{i}(w)=Y_{i}(w)=c$ or $X_{i}(w) \neq Y_{i}(w)$. Finally, consider the extended vertex $\left(t, v_{t}\right)$. By our assumption in the lemma, we have that $X_{t}\left(v_{t}\right) \neq Y_{t}\left(v_{t}\right)$.

We may repeatedly apply Lemma 5.25 to trace a discrepancy from some time $t$ to time 0 .

Lemma 5.26. Let $1 \leq t \leq T$ be an integer. Suppose $X_{t}\left(v_{t}\right) \neq Y_{t}\left(v_{t}\right)$. There exists a path $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$ in the extended hypergraph $H^{\text {ext }}$ such that

- $\left(t, v_{t}\right) \in e_{1}^{\text {ext }}, \min \left\{j \mid(j, w) \in e_{i}^{\text {ext }}\right\}>0$ for all $i<\ell$ and $\min \left\{j \mid(j, w) \in e_{\ell}^{\text {ext }}\right\}=0 ;$
- for any $1 \leq i, i^{\prime} \leq \ell$ satisfying $\left|i-i^{\prime}\right| \geq 2, e_{i}^{\text {ext }} \cap e_{i^{\prime}}^{\text {ext }}=\varnothing$;
- for any hyperedge $e_{i}^{\text {ext }}$ in the path, there exists $c \in[s]$ such that for all $(j, w) \in e_{i}^{\text {ext }}$, either $X_{j}(w)=Y_{j}(w)=c$ or $X_{j}(w) \neq Y_{j}(w)$.

Proof. We use Lemma 5.25 recursively. Namely, we use Lemma 5.25 for $\left(t, v_{t}\right)$ to find $\left(t^{\prime}, u\right)$. If $t^{\prime} \neq 0$, we apply Lemma 5.25 on $\left(t^{\prime}, u\right)$ again to find the previous discrepancy. Repeat this process until we find $\left(t^{\prime \prime}, w\right)$ such that $t^{\prime \prime}=0$. This gives a path $f_{1}^{\text {ext }}, f_{2}^{\text {ext }}, \ldots, f_{m}^{\text {ext }}$ in the extended hypergraph $H^{\text {ext }}$ such that $\left(t, v_{t}\right) \in f_{1}^{\text {ext }}$ and $\min \left\{j \mid(j, w) \in f_{m}^{\text {ext }}\right\}=0$. By Lemma 5.25 , this path $f_{1}^{\text {ext }}, f_{2}^{\text {ext }}, \ldots, f_{m}^{\text {ext }}$ satisfies the last property in Lemma 5.26.

We then construct the path $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$. First let $e_{1}^{\text {ext }}=f_{1}^{\text {ext }}, \ell=1$, and $p=1$. While $\min \left\{i \mid(i, w) \in e_{\ell}^{\text {ext }}\right\}>0$, we repeat the following process:

- let $p+1 \leq j \leq m$ be the largest index satisfying $f_{j}^{\text {ext }} \cap e_{\ell}^{\text {ext }} \neq \varnothing$;
- let $\ell \leftarrow \ell+1, e_{\ell}^{\text {ext }} \leftarrow f_{j}^{\text {ext }}$ and $p \leftarrow j$.

When the above process ends, we get the path $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$.
We first show that the process above is well-defined. Consider the beginning of each iteration of the while-loop. It holds that $e_{\ell}^{\text {ext }}=f_{p}^{\text {ext }}$. Since $\min \{i \mid(i, w) \in$ $\left.e_{\ell}^{\text {ext }}\right\}>0$, we know that $p<m$. The index $p+1 \leq j \leq m$ such that $f_{j}^{\text {ext }} \cap e_{\ell}^{\text {ext }} \neq \varnothing$ must exist because $f_{p+1}^{\text {ext }} \cap e_{\ell}^{\text {ext }}=f_{p+1}^{\text {ext }} \cap f_{p}^{\text {ext }} \neq \varnothing$. The while-loop must terminate eventually because $p$ always increase and $\min \left\{i \mid(i, w) \in f_{m}^{\text {ext }}\right\}=0$.

We claim that $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$ is indeed a path. We only need to show that for all $2 \leq i \leq \ell$, it holds that $e_{i-1}^{\text {ext }} \cap e_{i}^{\text {ext }} \neq \varnothing$ and $e_{i-1}^{\text {ext }} \neq e_{i}^{\text {ext }}$. The construction process guarantees that $e_{i-1}^{\text {ext }} \cap e_{i}^{\text {ext }} \neq \varnothing$. Suppose there is an index $2 \leq i \leq \ell$ such that $e_{i-1}^{\text {ext }}=e_{i}^{\text {ext }}=f_{i^{\prime}}^{\text {ext }}$ for some $i^{\prime} \leq m$. Since the construction process finds $e_{i}^{\text {ext }}$, we know that $\min \left\{t \mid(t, w) \in e_{i-1}^{\text {ext }}\right\}>0$. Thus $i^{\prime}<m$ and $f_{i^{\prime}+1}^{\text {ext }}$ exists. Since $f_{1}^{\text {ext }}, f_{2}^{\text {ext }}, \ldots, f_{m}^{\text {ext }}$ is a path, we know that $f_{i^{\prime}}^{\text {ext }} \cap f_{i^{\prime}+1}^{\text {ext }} \neq \varnothing$, which implies that $e_{i-1}^{\text {ext }} \cap f_{i^{\prime}+1}^{\text {ext }} \neq \varnothing$. When constructing $e_{i}^{\text {ext }}$, we look for the largest $j$ such that $e_{i-1}^{\text {ext }} \cap f_{j}^{\text {ext }} \neq \varnothing$. Hence, $e_{i}^{\text {ext }} \neq f_{i^{\prime}}^{\text {ext }}$, a contradiction.

Lastly, we verify the properties of the path.

- Since $e_{1}^{\text {ext }}=f_{1}^{\text {ext }}$ and $\left(t, v_{t}\right) \in f_{1}^{\text {ext }},\left(t, v_{t}\right) \in e_{1}^{\text {ext. }}$. The while-loop terminates once $\min \left\{j \mid(j, w) \in e_{\ell}^{\text {ext }}\right\}>0$. Hence, $\min \left\{j \mid(j, w) \in e_{i}^{\text {ext }}\right\}>0$ for all $i<\ell$ and $\min \left\{j \mid(j, w) \in e_{\ell}^{\text {ext }}\right\}=0$.
- For any $1 \leq i, i^{\prime} \leq \ell$ with $\left|i-i^{\prime}\right| \geq 2$, consider how $e_{i+1}^{\text {ext }}$ is constructed. We choose the largest index $j \leq m$ such that $f_{j}^{\text {ext }} \cap e_{\ell}^{\text {ext }} \neq \varnothing$ and $e_{i+1}^{\text {ext }} \leftarrow f_{j}^{\text {ext }}$. In other words, for any $j^{\prime}>j, f_{j^{\prime}}^{\text {ext }} \cap e_{i}^{\text {ext }}=\varnothing$. Since there is $j^{\prime}$ such that $e_{i^{\prime}}^{\text {ext }}=f_{j^{\prime}}^{\text {ext }}, e_{i}^{\text {ext }} \cap e_{i^{\prime}}^{\text {ext }}=\varnothing$.
- Since $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$ is a subsequence of $f_{1}^{\text {ext }}, f_{2}^{\text {ext }}, \ldots, f_{m}^{\text {ext }}$, the last property is satisfied as well.


### 5.5.3 Proof of Lemma 5.21

Recall that $T=\left\lceil 50 n \log \frac{n}{\epsilon}\right\rceil$ in Lemma 5.21. To prove Lemma 5.21, we need to show that

$$
\forall v \in V, \quad \operatorname{Pr}_{C}\left[X_{T}(v) \neq Y_{T}(v)\right] \leq \frac{\epsilon}{n}
$$

Fix a vertex $v$. By the same reason as (5.20), we only need to prove $\operatorname{Pr}_{C}\left[X_{T}(v) \neq Y_{T}(v)\right] \leq$ $\frac{\epsilon}{n}$ for a new $T$, where

$$
\begin{equation*}
T=\operatorname{time}_{\text {ud }}\left(v,\left\lceil 50 n \log \frac{n}{\epsilon}\right\rceil\right) \geq\left\lceil 40 n \log \frac{n}{\epsilon}\right\rceil . \tag{5.21}
\end{equation*}
$$

Note that $v$ is updated at time $T$, i.e. $v=v_{T}$.
Fix $T$ defined in (5.21). Define the following information percolation path (IPP).
Definition 5.27. We say a path $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$ of length $\ell$ in the extended hypergraph $H^{\text {ext }}$ is an information percolation path (IPP) if the following two properties are satisfied:

- $\left(T, v_{T}\right) \in e_{1}^{\text {ext }}, \min \left\{j \mid(j, w) \in e_{i}^{\text {ext }}\right\}>0$ for all $i<\ell$ and $\min \left\{j \mid(j, w) \in e_{\ell}^{\text {ext }}\right\}=0$;
- for any $1 \leq i, j \leq \ell$ such that $|i-j| \geq 2, e_{i}^{\text {ext }} \cap e_{j}^{\text {ext }}=\varnothing$.

Suppose $X_{T}(v) \neq Y_{T}(v)$. By Lemma 5.26 , we can find an IPP $e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$ in extended hypergraph $H^{\text {ext }}$. The following lemma lower bounds the length of the IPP.

Lemma 5.28. For any IPP of length $\ell, \ell \geq\lceil T / n\rceil$.

Proof. For any extended hyperedge $e_{i}^{\text {ext }}$, define the maximum and minimum update times in $e_{i}^{\text {ext }}$ by $t_{\text {max }}^{(i)}=\max \left\{t \mid(t, w) \in e_{i}^{\text {ext }}\right\}$ and $t_{\text {min }}^{(i)}=\min \left\{t \mid(t, w) \in e_{i}^{\text {ext }}\right\}$. In the systematic scan, we update vertices in order of their labels. By Definition 5.24, it holds that for any $i$,

$$
t_{\max }^{(i)}-t_{\min }^{(i)} \leq n-1 \leq n .
$$

Note that $e_{i}^{\text {ext }} \cap e_{i+1}^{\text {ext }} \neq \varnothing$, which implies

$$
t_{\min }^{(i)} \leq t_{\max }^{(i+1)} \leq t_{\min }^{(i+1)}+n .
$$

Note that $t_{\text {min }}^{(1)} \geq t_{\text {max }}^{(1)}-n=T-n$. We have

$$
T-n \leq t_{\min }^{(1)} \leq t_{\min }^{(\ell)}+(\ell-1) n=(\ell-1) n,
$$

where the last equation holds because $t_{\min }^{(\ell)}=0$. Since $\ell$ is an integer, we have $\ell \geq$ $\lceil T / n\rceil$.

Now fix an integer $\ell \geq T / n$ and an $\operatorname{IPP} \mathcal{P}=e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$ of length $\ell$. We define the bad event $\mathcal{B}(\mathcal{P})$ as: for any hyperedge $e_{i}^{\text {ext }}$ in the path, there exists $c \in[s]$ such that for all $(j, w) \in e_{i}^{\text {ext }}$, either $X_{j}(w)=Y_{j}(w)=c$ or $X_{j}(w) \neq Y_{j}(w)$. Namely, $\mathcal{B}(\mathcal{P})$ that implies $\mathcal{P}$ satisfies the third property in Lemma 5.26. By Lemma 5.26, Lemma 5.28 and a union bound over all IPPs of length at least $\ell$, the probability of $X_{T}(v) \neq Y_{T}(v)$ can be bounded as follows

$$
\begin{equation*}
\operatorname{Pr}_{C}\left[X_{T}(v) \neq Y_{T}(v)\right] \leq \sum_{\ell \geq\lceil T / n\rceil \mathcal{P}: ~ I P P ~ o f ~ l e n g t h ~} \sum_{\operatorname{Pr}} \operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})] . \tag{5.22}
\end{equation*}
$$

We bound $\operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})]$ in the RHS of (5.22) next. We need to use more delicate structures of the extended hypergraph $H^{\text {ext }}=\left(V^{\text {ext }}, \mathcal{E}^{\text {ext }}\right)$. By Definition 5.24, each extended hyperedge $e^{\text {ext }} \in \mathcal{E}^{\text {ext }}$ corresponds to a unique hyperedge edge $\left(e^{\text {ext }}\right) \in \mathcal{E}$ in the input hypergraph, or more formally,

$$
\text { edge }\left(e^{\mathrm{ext}}\right):=\left\{v \mid(t, v) \in e^{\mathrm{ext}}\right\} .
$$

We remark that different extended hyperedges may correspond to the same hyperedge. For each extended hyperedge $e^{\text {ext }} \in \mathcal{E}^{\text {ext }}$, we use $N\left(e^{\text {ext }}\right)$ to denote the neighbour extended hyperedges:

$$
N\left(e^{\mathrm{ext}}\right):=\left\{f^{\mathrm{ext}} \in \mathcal{E}^{\mathrm{ext}} \mid f^{\mathrm{ext}} \cap e^{\mathrm{ext}} \neq \varnothing \text { and } f^{\mathrm{ext}} \neq e^{\mathrm{ext}}\right\} .
$$

The following observation is straightforward to verify.

Observation 5.29. For any $e^{\text {ext }} \in \mathcal{E}^{\text {ext }}$ and $f^{\text {ext }} \in N\left(e^{\text {ext }}\right)$, edge $\left(e^{\text {ext }}\right) \cap \operatorname{edge}\left(f^{\text {ext }}\right) \neq \varnothing$. We further partition $N\left(e^{\text {ext }}\right)$ into self-neighbours and outside-neighbours as follows,

$$
\begin{aligned}
& N_{\text {self }}\left(e^{\mathrm{ext}}\right):=\left\{f^{\mathrm{ext}} \in N\left(e^{\mathrm{ext}}\right) \mid \operatorname{edge}\left(e^{\mathrm{ext}}\right)=\operatorname{edge}\left(f^{\mathrm{ext}}\right)\right\} ; \\
& N_{\text {out }}\left(e^{\mathrm{ext}}\right):=\left\{f^{\mathrm{ext}} \in N\left(e^{\mathrm{ext}}\right) \mid \operatorname{edge}\left(e^{\mathrm{ext}}\right) \neq \operatorname{edge}\left(f^{\mathrm{ext}}\right)\right\} .
\end{aligned}
$$

Observation 5.30. For any $e^{\text {ext }} \in \mathcal{E}^{\mathrm{ext}}$ and $f^{\mathrm{ext}} \in N_{\text {out }}\left(e^{\mathrm{ext}}\right),\left|e^{\mathrm{ext}} \cap f^{\text {ext }}\right|=1$.
Proof. Let $e=\operatorname{edge}\left(e^{\text {ext }}\right)$ and $f=\operatorname{edge}\left(f^{\text {ext }}\right)$. Since $f^{\text {ext }} \in N_{\text {out }}\left(e^{\text {ext }}\right)$, by Observation 5.29 and the fact that the input hypergraph is linear, $|e \cap f|=1$, which implies $\left|e^{\text {ext }} \cap f^{\text {ext }}\right|=1$.

The following lemma bounds the degree of the extended hypergraph.
Lemma 5.31. Let $\Delta$ be the maximum degree of the input hypergraph $H=(V, \mathcal{E})$. Then,

1. given $(v, t) \in V^{\text {ext }}$ and $e \in \mathcal{E}$ such that $v \in e$, the number of $e^{\text {ext }}$ such that $(v, t) \in$ $e^{\text {ext }}$ and edge $\left(e^{\text {ext }}\right)=e$ is at most $k$;
2. for any extended vertex $(v, t) \in V^{\mathrm{ext}}$, the number of extended hyperedges incident to $(v, t)$ is at most $d_{\mathrm{vtx}}:=\Delta k$;
3. for any extended hyperedge $e^{\text {ext }} \in \mathcal{E}^{\text {ext }}, N_{\text {self }}\left(e^{\text {ext }}\right) \leq d_{\text {self }}:=2 k, N_{\text {out }}\left(e^{\text {ext }}\right) \leq d_{\text {out }}:=$ $\Delta k^{2}$.

Proof. For Item 1, suppose such $e^{\text {ext }}$ is $\left\{\left(u_{j}, t_{j}\right) \mid 1 \leq j \leq k\right\}$ and $t_{1} \leq t_{2} \leq \ldots \leq t_{k}$. Moreover, for all $j$ such that $t_{j}=0$, we order $u_{j}$ according to their original label in $H$. As $(v, t) \in e^{\text {ext }}, t$ equals one of $t_{j}$. Then observe that $e^{\text {ext }}$ is uniquely determined if we know $t=t_{j}$ for some $1 \leq j \leq k$, and there are at most $k$ choices of $j$ (the number of choices can be less than $k$ if $t=0$ ). This shows the claim.

For Item 2, if $e^{\text {ext }}$ is incident to $(v, t)$, then edge $\left(e^{\text {ext }}\right)=e$ for some $e \ni v$. There are at most $\Delta$ choices of such hyperedge $e$ in $H$. Then the bound follows from Item 1.

For Item 3, let $e=$ edge $\left(e^{\text {ext }}\right)$, and again assume $e^{\text {ext }}$ is $\left\{\left(u_{j}, t_{j}\right) \mid 1 \leq j \leq k\right\}$ and $t_{1} \leq t_{2} \leq \ldots \leq t_{k}$ as in the proof of Item 1 .

To bound the number of self-neighbours, suppose $f^{\text {ext }} \in N_{\text {self }}\left(e^{\text {ext }}\right)$ such that edge $\left(f^{\text {ext }}\right)=$ $e$. Let $t_{\max }=\max \left\{t \mid(t, w) \in f^{\mathrm{ext}}\right\}$ and $t_{\text {min }}=\min \left\{t \mid(t, w) \in f^{\text {ext }}\right\}$. Note that if $t_{\max } \leq t_{k}$, then there are at most $k-1$ choices of $t_{\max }$, namely $t_{1}, t_{2}, \ldots, t_{k-1}$. Otherwise $t_{\max }>t_{k}$. Note that if $t_{\max } \geq t_{k}+n$, then $t_{\min } \geq t_{\max }-(n-1)>t_{k}$, which contradicts to $e^{\text {ext }} \cap f^{\text {ext }} \neq \varnothing$. It must hold that $t_{k}+1 \leq t_{\max } \leq t_{k}+n-1$. In the interval
[ $\left.t_{k}+1, t_{k}+n-1\right]$, there are at most $k-1$ times so that one of the vertices in $e$ is updated (this vertex cannot be $t_{k}$ as its update times are $t_{k}$ and $t_{k}+n$ ). Thus, there are $k-1$ choices of $t_{\max }$ again. Once $t_{\max }$ is fixed, since edge $\left(f^{\text {ext }}\right)=e, f^{\text {ext }}$ is also fixed. Overall, the number of $f^{\text {ext }} \in N_{\text {self }}\left(e^{\text {ext }}\right)$ is at most $2(k-1) \leq 2 k$.

To bound the number of outside-neighbours. We first choose one of the $k$ extended vertices in $e^{\text {ext }}$, say $\left(t_{i}, u_{i}\right)$. Then consider $f^{\text {ext }} \in N_{\text {out }}\left(e^{\text {ext }}\right)$ such that $\left(t_{i}, u_{i}\right) \in$ $f^{\text {ext }}$. By Item 2 , the number of such $f^{\text {ext }}$ is at most $\Delta k$, implying the overall bound of $\Delta k^{2}$.

Consider the $\operatorname{IPP} \mathcal{P}=e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$. Define the parameters $R_{\text {self }}$ and $R_{\text {out }}$ by

$$
\begin{aligned}
R_{\text {self }} & :=\left|\left\{2 \leq i \leq \ell \mid e_{i}^{\text {ext }} \in N_{\text {self }}\left(e_{i-1}^{\text {ext }}\right)\right\}\right| ; \\
R_{\text {out }} & :=\left|\left\{2 \leq i \leq \ell \mid e_{i}^{\text {ext }} \in N_{\text {out }}\left(e_{i-1}^{\text {ext }}\right)\right\}\right| .
\end{aligned}
$$

By definition, $R_{\text {self }}$ counts the number of consecutive self neighbours in $\mathcal{P}$ and $R_{\text {out }}$ counts the number of consecutive outside neighbours in $\mathcal{P}$. It holds that $R_{\text {self }}+R_{\text {out }}=$ $\ell-1$. We have the following lemma.
Lemma 5.32. Suppose $k \geq 20$ and $q \geq 40 \Delta^{\frac{2}{k-4}}$. For any $\operatorname{IPP} \mathcal{P}=e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext }}$, it holds that

$$
\operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})] \leq 10^{3} \Delta k^{6}\left(\frac{1}{10^{3} \Delta k^{6}}\right)^{R_{\text {out }}+\frac{1}{3}\left(R_{\text {self }}-b\right)}
$$

where $b$ is an integer satisfying $0 \leq b \leq \min \left\{R_{\text {self }}, 2 R_{\text {out }}\right\}$.
The proof of Lemma 5.32 is given in Section 5.5.4, where we will specify the value of the integer $b$. Now, we use Lemma 5.32 to prove Lemma 5.21. We remark that in the proof of Lemma 5.21, we do not use the specific value of $b$, we only use the fact that $0 \leq b \leq \min \left\{R_{\text {self }}, 2 R_{\text {out }}\right\}$.

Proof of Lemma 5.21. First fix an integer $\ell \geq\lfloor T / n\rfloor$ and an integer $0 \leq r \leq \ell-1$. Consider the $\operatorname{IPP} \mathcal{P}$ of length $\ell$ such that $R_{\text {out }}=r$ and $R_{\text {self }}=\ell-1-r$. By the definition of IPP (Definition 5.27) together with Lemma 5.31, the number of such path $\mathcal{P}$ is at most

$$
\binom{\ell-1}{r} d_{\mathrm{vtx}} d_{\mathrm{out}}^{r} d_{\mathrm{self}}^{\ell-1-r} \leq \Delta k\binom{\ell-1}{r}\left(\Delta k^{2}\right)^{r}(2 k)^{\ell-1-r}
$$

By Lemma 5.32 and the union bound in (5.22), we have

$$
\begin{aligned}
& \operatorname{Pr}_{C}\left[X_{T}(v) \neq Y_{T}(v)\right] \leq \sum_{\ell \geq\lceil T / n\rceil} \sum_{\mathcal{P}: \text { IPP of length } \ell} \operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})] \\
\leq & \sum_{\ell \geq\lceil T / n\rceil} \sum_{r=0}^{\ell-1} \Delta k\binom{\ell-1}{r}\left(\Delta k^{2}\right)^{r}(2 k)^{\ell-1-r} \cdot 10^{3} \Delta k^{6}\left(\frac{1}{10^{3} \Delta k^{6}}\right)^{r+\frac{1}{3}(\ell-1-r-b(\ell, r))},
\end{aligned}
$$

where $b(\ell, r)$ is an integer satisfying $0 \leq b(\ell, r) \leq \min \{\ell-1-r, 2 r\}$. Since $b(\ell, r) \leq$ $\ell-1-r$, it holds that $\left(\frac{1}{10^{3} \Delta k^{6}}\right)^{\frac{\ell-1-r-b(\ell, r)}{3}} \leq\left(\frac{1}{10^{3} k^{6}}\right)^{\frac{\ell-1-r-b(\ell, r)}{3}}$, which implies

$$
\begin{aligned}
\operatorname{Pr}_{C} & {\left[X_{T}(v) \neq Y_{T}(v)\right] } \\
& \leq \sum_{\ell \geq\lceil T / n\rceil} \sum_{r=0}^{\ell-1} \Delta k\binom{\ell-1}{r}\left(\Delta k^{2}\right)^{r}(2 k)^{\ell-1-r} \cdot 10^{3} \Delta k^{6}\left(\frac{1}{10^{3} \Delta k^{6}}\right)^{r}\left(\frac{1}{10^{3} k^{6}}\right)^{\frac{\ell-1-r-b(\ell, r)}{3}} \\
& =10^{3} \Delta^{2} k^{7} \sum_{\ell \geq\lceil T / n\rceil} \sum_{r=0}^{\ell-1}\binom{\ell-1}{r}\left(\frac{1}{5 k}\right)^{\ell-1-r}\left(\frac{1}{10^{3} k^{4}}\right)^{r}\left(\frac{1}{10 k^{2}}\right)^{-b(\ell, r)} .
\end{aligned}
$$

Note that $k \geq 20$. Since $0 \leq b(\ell, r) \leq 2 r$, we have $\left(\frac{1}{10 k^{2}}\right)^{-b(\ell, r)} \leq\left(\frac{1}{10 k^{2}}\right)^{-2 r}=\left(100 k^{4}\right)^{r}$, which imples

$$
\begin{aligned}
\operatorname{Pr}_{C}\left[X_{T}(v) \neq Y_{T}(v)\right] & \leq 10^{3} \Delta^{2} k^{7} \sum_{\ell \geq\lceil T / n\rceil}\left(\frac{1}{10}\right)^{\ell-1} \sum_{r=0}^{\ell-1}\binom{\ell-1}{r}=10^{3} \Delta^{2} k^{7} \sum_{\ell \geq\lceil T / n\rceil}\left(\frac{1}{5}\right)^{\ell-1} \\
& \leq 10^{3} \Delta^{2} k^{7}\left(\frac{1}{2}\right)^{T / n}
\end{aligned}
$$

Note that $T \geq 40 n \log \frac{n \Delta}{\epsilon}$ and $k \leq n$. We have

$$
\operatorname{Pr}_{C}\left[X_{T}(v) \neq Y_{T}(v)\right] \leq \frac{\epsilon}{n}
$$

### 5.5.4 Proof of Lemma 5.32

Fix an $\operatorname{IPP} \mathcal{P}=e_{1}^{\text {ext }}, e_{2}^{\text {ext }}, \ldots, e_{\ell}^{\text {ext. }}$. We define a total ordering among all extended hyperedges in $\mathcal{P}$. For any two extended hyperedges $e_{i}^{\text {ext }}$ and $e_{j}^{\text {ext }}$ in $\mathcal{P}$, we say $e_{i}^{\text {ext }}<e_{j}^{\text {ext }}$ if and only if $i<j$.

Lemma 5.33. There exists a subsequence $f_{1}^{\text {ext }}<f_{2}^{\mathrm{ext}}<\ldots<f_{m}^{\text {ext }}$ in IPP $\mathcal{P}$ such that

- for any $1 \leq i, j \leq m$ satisfying $|i-j| \geq 2, f_{i}^{\mathrm{ext}} \cap f_{j}^{\mathrm{ext}}=\varnothing$;
- for any $2 \leq i \leq m,\left|f_{i}^{\text {ext }} \cap f_{i-1}^{\text {ext }}\right| \leq 1$;
- $m \geq R_{\text {out }}+\frac{1}{3}\left(R_{\text {self }}-b\right)$ for some integer $0 \leq b \leq \min \left\{R_{\text {self }}, 2 R_{\text {out }}\right\}$.

Note that $\left\{f_{i}^{\text {ext }}\right\}$ given in Lemma 5.33 is not necessarily a path. What we do in Lemma 5.33 is to prune certain self-neighbours from $\mathcal{P}$ so that the second property holds. To be more precise, for a maximal sequence of consecutive self-neighbouring hyperedges, we prune all hyperedges that are in even positions of this sequence. We give a formal proof below.

Proof of Lemma 5.33. There are $\ell-1$ pairs of adjacent extended hyperedges, i.e. $e_{i-1}^{\text {ext }}$ and $e_{i}^{\text {ext }}$ are adjacent for $2 \leq i \leq \ell$. Define

$$
S_{\text {out }}:=\left\{\text { integer } i \in[2, \ell] \mid e_{i}^{\text {ext }} \in N_{\text {out }}\left(e_{i-1}^{\text {ext }}\right)\right\} .
$$

Note that $\left|S_{\text {out }}\right|=R_{\text {out }}$. Denote $R=R_{\text {out }}$. Suppose the elements in $S_{\text {out }}$ are $2 \leq i_{1}<$ $i_{2}<\ldots<i_{R} \leq \ell$. In addition, we define $i_{0}=1$ and $i_{R+1}=\ell+1$, although $i_{0} \notin S_{\text {out }}$ and $i_{R+1} \notin S_{\text {out }}$. Removing all the elements in $S_{\text {out }}$, the integers in the interval [2, $\left.\ell\right]$ splits into a set $I_{\text {self }}$ of sub-intervals:

$$
I_{\text {self }}:=\left\{[l, r] \mid \exists j \text { s.t. } 0 \leq j \leq R, l=i_{j}+1, r=i_{j+1}-1, \text { and } l \leq r\right\} .
$$

Equivalently, $I_{\text {self }}$ can be constructed by going through all $j$ from 0 to $R$, and adding the interval $\left[i_{j}+1, i_{j+1}-1\right]$ to the set $I_{\text {self }}$ if $i_{j}+1 \leq i_{j+1}-1$. For each interval $[l, r] \in$ $I_{\text {self }}$, the following properties hold

1. for each integer $i \in[l, r], e_{i}^{\text {ext }} \in N_{\text {self }}\left(e_{i-1}^{\text {ext }}\right)$;
2. either $l=2$ or $e_{l-1}^{\text {ext }} \in N_{\text {out }}\left(e_{l-2}^{\text {ext }}\right)$;
3. either $r=\ell$ or $e_{r+1}^{\text {ext }} \in N_{\text {out }}\left(e_{r}^{\text {ext }}\right)$.

In other words, each interval $[l, r] \in I_{\text {self }}$ represents a sequence of consecutive extended hyperedges in the $\operatorname{IPP} \mathcal{P}$ of length $r-l+1$ such that each extended hyperedge is a self-neighbour of its predecessor in $\mathcal{P}$, and this sequence is maximal.

Suppose the intervals in $I_{\text {self }}$ are $\left[l_{1}, r_{1}\right],\left[l_{2}, r_{2}\right], \ldots,\left[l_{a}, r_{a}\right]$ such that $l_{1} \leq r_{1}<l_{2} \leq$ $r_{2}<\ldots<l_{a} \leq r_{a}$, where $a=\left|I_{\text {self }}\right|$. It is straightforward to verify that

$$
\begin{equation*}
\sum_{i=1}^{a}\left(r_{i}-l_{i}+1\right)=R_{\text {self }} \tag{5.23}
\end{equation*}
$$

Define a subset $I_{\text {self }}^{(1)} \subseteq I_{\text {self }}$ by

$$
I_{\text {self }}^{(1)}:=\left\{[l, r] \in I_{\text {self }} \mid l=r\right\} .
$$

The quantity $b$ is the size of $I_{\text {self }}^{(1)}$, i.e. $b:=\left|I_{\text {self }}^{(1)}\right|$. Since $I_{\text {self }}^{(1)}$ is a subset of $I_{\text {self }}$, by (5.23), we have

$$
\begin{equation*}
b \leq R_{\text {self }} . \tag{5.24}
\end{equation*}
$$

Note that $\ell \geq T / n \geq 40 \log n \geq 20$. If $R_{\text {out }}=0$, then $I_{\text {self }}$ contains only a single interval [2, $\ell$. Thus $b=0$ and we have $b \leq 2 R_{\text {out }}$. Otherwise $R_{\text {out }} \geq 1$. By property 3 above,
for each $j \in[a]$, it holds that either $r_{j}=\ell$ or $e_{r_{j}+1}^{\text {ext }} \in N_{\text {out }}\left(e_{r_{j}}^{\text {ext }}\right)$ (namely $r_{j}+1 \in S_{\text {out }}$ ). This implies $b \leq R+1=R_{\text {out }}+1 \leq 2 R_{\text {out }}$, because there are at most one $\left(l_{j}, r_{j}\right) \in I_{\text {self }}^{(1)}$ satisfying $l_{j}=r_{j}=\ell$. Hence, in both cases, we have

$$
\begin{equation*}
b \leq 2 R_{\text {out }} . \tag{5.25}
\end{equation*}
$$

Combining (5.24) and (5.25) proves that $b \leq \min \left\{R_{\text {self }}, 2 R_{\text {out }}\right\}$.
Finally, we construct the the subsequence $f_{1}^{\text {ext }}<f_{2}^{\text {ext }}<\ldots<f_{m}^{\text {ext }}$ from IPP $\mathcal{P}$. We construct a subset $\mathcal{F}$ by the following procedure.

- For each $i \in S_{\text {out }}$, we add $e_{i}^{\text {ext }}$ into $\mathcal{F}$.
- For each interval $[l, r] \in I_{\text {self }}$, for all integers $j \in[l, r]$ such that $(j-l)$ is an odd number, we add $e_{j}^{\text {ext }}$ into $\mathcal{F}$. Note that by property 2 , if $l>2, e_{l-1}^{\text {ext }}$ is always in $\mathcal{F}$ because of the previous rule.
- To finish, we sort all extended hyperedges in $\mathcal{F}$ to obtain $f_{1}^{\text {ext }}<f_{2}^{\text {ext }}<\ldots<f_{m}^{\text {ext }}$. We now verify the three properties in Lemma 5.33.
- By the definition of IPP, for any $1 \leq i, j \leq \ell$ satisfying $|i-j| \geq 2, e_{i}^{\text {ext }} \cap e_{j}^{\text {ext }}=\varnothing$. Since $f_{1}^{\text {ext }}<f_{2}^{\text {ext }}<\ldots<f_{m}^{\text {ext }}$ is a subsequence of $\mathcal{P}$, the first property holds.
- Fix an index $2 \leq j \leq m$. Suppose $f_{j-1}^{\text {ext }}=e_{j_{1}}^{\text {ext }}$ and $f_{j}^{\mathrm{ext}}=e_{j_{2}}^{\text {ext. }}$. If $\left|j_{1}-j_{2}\right| \geq 2$, then $\left|f_{i}^{\text {ext }} \cap f_{i-1}^{\text {ext }}\right|=0$. Assume $j_{1}+1=j_{2}$, which means that $e_{j_{1}}^{\text {ext }}$ and $e_{j_{2}}^{\text {ext }}$ are neighbours in extended hypergraph. If $e_{j_{2}}^{\text {ext }} \in N_{\text {out }}\left(e_{j_{1}}^{\text {ext }}\right)$, by Observation 5.30, it holds that $\left|f_{i}^{\text {ext }} \cap f_{i-1}^{\text {ext }}\right|=1$. Otherwise, $e_{j_{2}}^{\text {ext }} \in N_{\text {self }}\left(e_{j_{1}}^{\text {ext }}\right)$. There must exist an interval $[l, r] \in I_{\text {self }}$ such that either $j_{1}, j_{2} \in[l, r]$ or $j_{1} \notin[l, r]$ but $j_{2} \in[l, r]$. The first case is impossible because we do not add two consecutive indices in any interval of $I_{\text {self }}$. The second case is also impossible because it implies $j_{1}=l-1$ and $j_{2}=l$, but $l$ cannot be added.
- All extendeds hyperedge in $S_{\text {out }}$ are added into $\mathcal{F}$. For each interval $[l, r] \in I_{\text {self }}$, $\left\lfloor\frac{r-l+1}{2}\right\rfloor$ extended hyperedges in $[l, r]$ are added into $\mathcal{F}$. Hence, if $l \neq r$, the number of vertices in $[l, r]$ added to $\mathcal{F}$ is at least $(r-l+1) / 3$ (with $r=l+2$ being the worst case). By (5.23), we have $m \geq R_{\text {out }}+\frac{1}{3}\left(R_{\text {self }}-b\right)$.

Hence, the subsequence $f_{1}^{\text {ext }}<f_{2}^{\text {ext }}<\ldots<f_{m}^{\text {ext }}$ satisfies all the properties in Lemma 5.33.

Now we are ready to prove Lemma 5.32.

Proof of Lemma 5.32. Let $f_{1}^{\text {ext }}<f_{2}^{\text {ext }}<\ldots<f_{m}^{\text {ext }}$ be the subsequence given in Lemma 5.33. For each $f_{i}^{\text {ext }}$ and $c \in[s]$, define a bad event $\mathcal{B}_{i}(c)$ that for all $(j, w) \in f_{i}^{\text {ext }}$, either $X_{j}(w) \neq Y_{j}(w)$ or $X_{j}(w)=Y_{j}(w)=c$. Note that $f_{1}^{\text {ext }}<f_{2}^{\text {ext }}<\ldots<f_{m}^{\text {ext }}$ is a subsequence in $\operatorname{IPP} \mathcal{P}$, the probability of $\mathcal{B}(\mathcal{P})$ can be bounded as follows

$$
\operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})] \leq \operatorname{Pr}_{C}\left[\forall i \in[m], \exists c_{i} \in[s] \text { s.t. } \mathcal{B}_{i}\left(c_{i}\right)\right] .
$$

By (5.21), it holds that $\ell \geq T / n \geq 40 \log n \geq 20$. By the last property in Lemma 5.33, $m \geq \frac{1}{3}\left(R_{\text {out }}+R_{\text {self }}\right)=\frac{\ell-1}{3}>6$. We further truncate the last element $f_{m}^{\text {ext }}$ and obtain the following inequality

$$
\begin{equation*}
\operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})] \leq \operatorname{Pr}_{C}\left[\forall i \in[m-1], \exists c_{i} \in[s] \text { s.t. } \mathcal{B}_{i}\left(c_{i}\right)\right] \leq \sum_{\mathbf{c} \in[s]^{m-1}} \operatorname{Pr}_{C}\left[\bigwedge_{i=1}^{m-1} \mathcal{B}_{i}\left(c_{i}\right)\right] \tag{5.26}
\end{equation*}
$$

where the second inequality follows from the union bound, and $\mathbf{c}=\left(c_{1}, \ldots, c_{m-1}\right) \in$ $[s]^{m-1}$. The truncation ensures that all elements $(j, w) \in \cup_{i=1}^{m-1} f_{i}^{\text {ext }}$ satisfy $j>0$. (See Definition 5.27 of IPPs.)

Fix $\mathbf{c} \in[s]^{m-1}$, we bound the probability of the event $\bigwedge_{i=1}^{m-1} \mathcal{B}_{i}\left(c_{i}\right)$. For each $1 \leq$ $i<m$, we define

$$
S_{i}^{\text {ext }}:= \begin{cases}f_{i}^{\text {ext }} & \text { if } i=1 \\ f_{i}^{\text {ext }} \backslash f_{i-1}^{\text {ext }} & \text { if } i>1\end{cases}
$$

Since $S_{i}^{\text {ext }} \subseteq f_{i}^{\text {ext }}$, we have the following bound
$\operatorname{Pr}_{C}\left[\bigwedge_{i=1}^{m-1} \mathcal{B}_{i}\left(c_{i}\right)\right] \leq \operatorname{Pr}_{C}\left[\bigwedge_{i=1}^{m-1}\left(\forall(j, w) \in S_{i}^{\mathrm{ext}},\left(X_{j}(w) \neq Y_{j}(w)\right) \vee\left(X_{j}(w)=Y_{j}(w)=c_{i}\right)\right)\right]$.
By the first property in Lemma 5.33, all $S_{i}^{\text {ext }}$ are mutually disjoint. Now we list all the extended vertices $\cup_{i=1}^{m-1} S_{i}^{\text {ext }}$ as $\left(j_{1}, w_{1}\right),\left(j_{2}, w_{2}\right), \ldots,\left(j_{M}, w_{M}\right)$, where $0<j_{1}<j_{2}<$ $\ldots<j_{M}$. For each $1 \leq p \leq M$, there is a unique $i$ such that $\left(j_{p}, w_{p}\right) \in S_{i}^{\text {ext }}$ and we denote $\operatorname{idx}\left(j_{p}\right):=i$. We define a bad event $\mathcal{A}(p)$ that either $X_{j_{p}}\left(w_{p}\right) \neq Y_{j_{P}}\left(w_{p}\right)$ or $X_{j_{p}}\left(w_{p}\right)=Y_{j_{P}}\left(w_{p}\right)=c_{\mathrm{idx}\left(j_{p}\right)}$. Using the chain rule for the RHS of the inequality above, it holds that

$$
\operatorname{Pr}_{C}\left[\bigwedge_{i=1}^{m-1} \mathcal{B}_{i}\left(c_{i}\right)\right] \leq \prod_{p=1}^{M} \operatorname{Pr}_{C}\left[\mathcal{A}(p) \mid \bigwedge_{p^{\prime}<p} \mathcal{A}\left(p^{\prime}\right)\right]
$$

Consider the probability of $\mathcal{A}(p)$ conditional on all $\mathcal{A}\left(p^{\prime}\right)$ for $p^{\prime}<p$. To simplify the notation, let $j=j_{p}>0$ and $w=w_{p}$. In the $j$-th update, $X_{j}(w)$ is sampled from the
distribution $v_{w}^{X_{j-1}(V \backslash\{w\})}$ and $Y_{j}(w)$ is sampled from the distribution $v_{w}^{Y_{j-1}(V \backslash\{w\})}$. For any $\tau \in[s]^{V \backslash\{w\}}$, it holds that

$$
\forall x \in[s], \quad v_{w}^{\tau}(x)=\sum_{y \in h^{-1}(x)} \mu_{w}^{\tau}(y) .
$$

Note that $\mu^{\tau}$ is actually the uniform distribution over a list colouring instance on $H$ where for each $u \neq w$, the colour list is $h^{-1}\left(\tau_{u}\right)$, and the colour list for $w$ is [ $\left.q\right]$. Hence, for each $u \neq w$, the size of colour list of $u$ is at least $\lfloor q / s\rfloor$, and the size of colour list of $w$ is $q$, where $s=\lceil\sqrt{q}\rceil$. Note that $q \geq 40 \Delta^{\frac{2}{k-4}}$ and $k \geq 20$ implies $\lfloor q / s\rfloor^{k} \geq 2 q^{2} k \Delta$. By Lemma 5.2, for all $\tau \in[s]^{V \backslash\{w\}}$, it holds that

$$
\forall y \in[q], \quad \frac{1}{q}\left(1-\frac{4}{k q}\right) \leq \frac{1}{q} \exp \left(-\frac{2}{k q}\right) \leq \mu_{w}^{\tau}(y) \leq \frac{1}{q} \exp \left(\frac{2}{k q}\right) \leq \frac{1}{q}\left(1+\frac{4}{k q}\right) .
$$

Hence, for any $\tau \in[s]^{V \backslash\{w\}}$, it holds that for any $x \in[s]$,

$$
\frac{\left|h^{-1}(x)\right|}{q}\left(1-\frac{4}{k q}\right) \leq v_{w}^{\tau}(x) \leq \frac{\left|h^{-1}(x)\right|}{q}\left(1+\frac{4}{k q}\right) .
$$

Note that all the events $\mathcal{A}\left(p^{\prime}\right)$ for $p^{\prime}<p$ are determined by the updates from time 1 to time $j-1$. The above bounds for $v_{w}^{\tau}(x)$ holds for any configuration $\tau \in[s]^{V \backslash\{w\}}$. In the $j$-th update step, since $X_{j}(w)$ and $Y_{j}(w)$ are coupled by the optimal coupling and $\left|h^{-1}(x)\right| \leq\lceil q / s\rceil$, we have the probability of $X_{j}(w) \neq Y_{j}(w)$ is at most $\frac{1}{2} \sum_{x \in[s]} \frac{\left|h^{-1}(x)\right|}{q}$. $\frac{8}{k q}=\frac{4}{k q}$, and the probability of $X_{j}(w)=Y_{j}(w)=c_{i}$ is at most $\frac{\lceil q / s\rceil}{q}\left(1+\frac{4}{k q}\right)$. Hence,

$$
\begin{aligned}
\operatorname{Pr}_{C}\left[\mathcal{A}(p) \mid \bigwedge_{p^{\prime}<p} \mathcal{A}\left(p^{\prime}\right)\right] & \leq \frac{4}{k q}+\frac{\lceil q / s\rceil}{q}\left(1+\frac{4}{k q}\right) \stackrel{(\star)}{\leq} \frac{\lceil q / s\rceil}{q}\left(1+\frac{5}{k}\right) \\
& \leq \frac{1.16}{\sqrt{q}}\left(1+\frac{5}{k}\right) .
\end{aligned}
$$

where ( $\star$ ) holds because $\frac{\lceil q / s\rceil}{k q} \geq \frac{4}{k q}$ if $q \geq 40$ and the last inequality is due to $\lceil q / s\rceil \leq$ $1.16 \sqrt{q}$. This implies

$$
\operatorname{Pr}_{C}\left[\bigwedge_{i=1}^{m-1} \mathcal{B}_{i}\left(c_{i}\right)\right] \leq \prod_{p=1}^{M}\left(\frac{1.16}{\sqrt{q}}\left(1+\frac{5}{k}\right)\right)=\prod_{i=1}^{m-1}\left(\frac{1.16}{\sqrt{q}}\left(1+\frac{5}{k}\right)\right)^{\left|S_{i}^{\text {ext }}\right|} .
$$

By the second property in Lemma 5.33 and the definition $S_{i}^{\text {ext }}$, it holds that

$$
\forall 1 \leq i \leq m, \quad\left|S_{i}^{\mathrm{ext}}\right| \geq k-1 .
$$

Combining with (5.26), we have

$$
\begin{aligned}
\operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})] & \leq \sum_{\mathbf{c} \in\left[s s^{m-1}\right.} \operatorname{Pr}_{C}\left[\bigwedge_{i=1}^{m-1} \mathcal{B}_{i}\left(c_{i}\right)\right] \leq \sum_{\mathbf{c} \in[s]^{m-1}}\left(\frac{1.16}{\sqrt{q}}\left(1+\frac{5}{k}\right)\right)^{(m-1)(k-1)} \\
& \leq\left(s\left(\frac{1.16}{\sqrt{q}}\left(1+\frac{5}{k}\right)\right)^{k-1}\right)^{m-1} .
\end{aligned}
$$

Now we claim that

$$
s\left(\frac{1.16}{\sqrt{q}}\left(1+\frac{5}{k}\right)\right)^{k-1} \leq \frac{1}{10^{3} \Delta k^{6}} .
$$

Using $s=\lceil\sqrt{q}\rceil \leq 1.16 \sqrt{q}$, it suffices to show that

$$
1.16 \times 10^{3}(1.16)^{k-1}\left(1+\frac{5}{k}\right)^{k-1} \Delta k^{6} \leq q^{(k-2) / 2}
$$

Using $\left(1+\frac{5}{k}\right)^{\frac{2(k-1)}{k-2}} \leq 1.7$ and $k^{12 /(k-2)} \leq 7.4$ for $k \geq 20$, we further simplifies the condition into

$$
q \geq 7.4 \times 1.7 \times\left(1.16 \times 10^{3}\right)^{2 /(k-2)}(1.16)^{2(k-1) /(k-2)} \Delta^{2 /(k-2)}
$$

which is implied by $q \geq 40 \Delta^{\frac{2}{k-4}}$ and $k \geq 20$.
The claim implies that

$$
\operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})] \leq\left(\frac{1}{10^{3} \Delta k^{6}}\right)^{m-1}=10^{3} \Delta k^{6}\left(\frac{1}{10^{3} \Delta k^{6}}\right)^{m}
$$

Finally, by the third property in Lemma 5.33, we have

$$
\operatorname{Pr}_{C}[\mathcal{B}(\mathcal{P})] \leq 10^{3} \Delta k^{6}\left(\frac{1}{10^{3} \Delta k^{6}}\right)^{R_{\text {out }}+\frac{1}{3}\left(R_{\text {self }}-b\right)}
$$

## Chapter 6

## Ferromagnetic Ising model

The Ising model is a classical statistical physics model for ferromagnetism that had far-reaching impact in many areas. In computer science / combinatorics terms, the model defines a weighted distribution over cuts of a graph. To be more precise, let $G=$ $(V, E)$ be a simple undirected graph. For each edge $e \in E$, we have the local interaction strength $\beta_{e} \in \mathbb{R}_{>0}$, and for each vertex $v \in V$, we have the external magnetic field (namely vertex weight) $\lambda_{v} \in \mathbb{R}_{>0}$. An Ising model is specified by the tuple $(G ; \beta, \lambda)$, where $\beta=\left(\beta_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$. We assign spins $\{0,1\}$ to the vertices $V$. For each spin configuration $\sigma \in\{0,1\}^{V}$, the weight of $\sigma$ is defined by

$$
\mathrm{wt}_{\mathrm{Ising}}(\sigma):=\prod_{e=(u, v) \in E} \beta_{e}^{\mathbb{[}[\sigma(u)=\sigma(v)]} \prod_{u \in V} \lambda_{u}^{\sigma(u)},
$$

where $\mathbb{I}[\sigma(u)=\sigma(v)]$ is the indicator variable of the event $\sigma(u)=\sigma(v)$. The Gibbs distribution $\pi_{\text {Ising }}$ is defined by

$$
\begin{equation*}
\forall \sigma \in\{0,1\}^{V}, \quad \pi_{\text {Ising }}(\sigma)=\frac{\mathrm{wt}_{\text {Ising }}(\sigma)}{Z_{\text {Ising }}}, \tag{6.1}
\end{equation*}
$$

where

$$
Z_{\text {Ising }}=Z_{\text {Ising }}(G ; \beta, \lambda):=\sum_{\tau \in\{0,1\}^{V}} \mathrm{wt}_{\text {Ising }}(\tau)
$$

is the partition function. In this part we focus on the ferromagnetic case, where $\beta_{e}>1$ for all $e \in E$, with consistent fields, where $\lambda_{v} \in(0,1]$ for all $v \in V$. Note that by flipping the spins, the last assumption is equivalent to assuming $\lambda_{v} \in[1, \infty)$ for all $v \in V$.

There is extensive computational interest in simulating the Ising model and in evaluating various quantities related to it. A major contribution in the rigorous algorithmic study of the model is the Jerrum-Sinclair algorithm [JS93], which is the
first fully polynomial-time randomised approximation scheme (FPRAS) for the partition function $Z_{\text {Ising }}$ of the ferromagnetic Ising model with consistent fields on any graph. The main ingredient of their algorithm is to show that a natural Markov chain mixes in polynomial-time to sample from the so-called "subgraph-world" model, which has the same partition function up to some easy to compute factors.

Usually, using self-reducibility, approximately evaluating the partition function is computationally inter-reducible to approximate sampling [JVV86]. However, in the case of the Ising model, the original algorithm by Jerrum and Sinclair does not directly yield a sampling algorithm for spin configurations. This is because inconsistent fields may be created during the self-reduction, making the algorithm no longer applicable. To circumvent this issue, Randall and Wilson [RW99] showed that when there is no external field, an efficient approximate sampler for spin configurations exists by doing self-reductions in the so-called random cluster model. This is a model introduced by Fortuin and Kasteleyn [FK72] and also has the same partition function as the previous two models up to some easy to compute factors. ${ }^{1}$

On the other hand, a different Markov chain introduced by Swendsen and Wang [SW87] has shown great performance on sampling Ising configurations in practice. This dynamics is best understood via the Edwards-Sokal distribution [ES88], which is a joint distribution on both edges and vertices. The marginal distribution on vertices is the Ising model, and the marginal distribution on edges is the random cluster model. Sokal and later Peres ${ }^{2}$ conjectured that the Swendsen-Wang (SW) dynamics mixes in polynomial-time for ferromagnetic Ising models, and this was resolved in affirmative by Guo and Jerrum [GJ18]. They showed that the edge-flipping dynamics for the random cluster model mixes in polynomial-time, and this dynamics is known to be no faster than the SW dynamics [Ull14]. Another consequence of [GJ18] is that there is a perfect sampler for the ferromagnetic Ising model and the corresponding random cluster model, improving upon the approximate sampler of [RW99]. This is done via monotone coupling from the past (CFTP) [PW96b] as the random cluster model is monotone.

One restriction of [GJ18] is that their result only applies to the ferromagnetic Ising model without external fields. The original random cluster formulation of [FK72] does not incorporate external fields, although it is not hard to do so by generalising to a weighted random cluster formulation. Indeed, Park, Jang, Galanis, Shin, Štefankovič,

[^12]and Vigoda $\left[\mathrm{PJG}^{+} 17\right]$ generalised the SW dynamics $P_{\mathrm{SW}}^{\text {Ising }}$ (see Section 6.1.2.2 for detailed description) in the presence of external fields. They also showed efficiency of this algorithm in certain parameter regimes and on random graphs. This left open the question if the generalised SW dynamics is efficient in general.

First, we show that the edge-flipping dynamics for the weighted random cluster model mixes in polynomial-time. By adapting [Ull14] to the case with fields, this implies that the generalised SW dynamics has a polynomial running time for any ferromagnetic Ising model with consistent fields on any graph, answering the question above.

Theorem 6.1. Let $1<\beta_{\min } \leq \beta_{\max }$ be constants. For any ferromagnetic Ising model on graph $G=(V, E)$ with parameters $\left(\beta_{e}\right)_{e \in E}$ and $\left(\lambda_{v}\right)_{v \in V}$, where $\beta_{\min } \leq \beta_{e} \leq \beta_{\max }$ and $0<\lambda_{\nu} \leq 1$, the mixing time of Swendsen-Wang dynamics is $O\left(N^{4} m^{2}\left(m+\log \frac{1}{\epsilon}\right)\right)$, where $N=\min \left\{n, \frac{1}{1-\lambda_{\text {max }}}\right\}, \lambda_{\max }=\max _{v \in V} \lambda_{v}, n=|V|$ and $m=|E|$.

Note that if $\beta_{e}=1$ for some $e \in E$, it is equivalent to remove such an edge. Also if $\lambda_{v}=0$ for some $v \in V$, it is equivalent to pin $v$ to 0 and then absorb $v$ into its neighbours external fields. Thus, any ferromagnetic Ising model with consistent external fields can be transformed into one satisfying the condition of Theorem 6.1. The big- $O$ notation hides a constant factor depending only on $\beta_{\min }$ and $\beta_{\max }$. See (6.30) for the details of the hidden constant.

The main technical innovation of ours is to introduce a grand model, which incorporates both the so-called subgraph world [JS93] and the random cluster model. The subgraph world assigns weights to subsets of edges, where each vertex of an odd degree in the induced graph suffers a penalty corresponding to its external field (or the lack thereof). Detailed definitions of the basic models are given in Section 6.1.1.

The main inspiration of our grand model is the coupling given by Grimmett and Janson [GJ07b] between the two models above without external fields. Our model assigns 3 states to each edge: $0,1,2$. A sample of our model can be generated as follows: first, we sample a subset of edges from the subgraph world model, and assign 1 to them; then, we assign 0 or 2 to each remaining independently with a carefully chosen probability. Detailed definitions are in Section 6.2.1. The marginal distribution of edges assigned 1 clearly follow the subgraph world distribution, and we show that the non-zero edges follow the weighted random cluster model (Lemma 6.12). This last step is done using Valiant's holographic transformations [Val08]. It is also a generalisation of [GJ07b] in the presence of external fields.

We give a polynomial upper bound of the mixing time of the Glauber dynamics for the grand model in Section 6.3 via the method of canonical paths [JS89]. Our construction of the canonical path is a variation of the original paths by Jerrum and Sinclair [JS93]. The projection of this dynamics to the non-zero edges is exactly the Glauber dynamics for the weighted random cluster model. We show that this project does not slow down the dynamics (Section 6.5), and therefore mixing time bounds for the weighted random cluster model is a direct consequence. This implies Theorem 6.1. When there is no field, our argument recovers the result of Guo and Jerrum [GJ18]. However, our argument is both simpler and more general.

Another important feature of the grand model is that it gives a Gibbs distribution, in the sense that variables are independent if we condition on a subset of edges which disconnect the graph. This is a feature absent in the random cluster models. Recently, there is a lot of progress in analysing the mixing time of dynamics for Gibbs distributions, especially using the notion of spectral independence [ALO20]. Since the domain in our case is not Boolean, we use a generalisation of [FGYZ21b] (see also [CGŠV21] for a different generalisation). An important development along this line is that in bounded degree graphs, spectral independence implies near-linear mixing time of dynamics for the Gibbs distribution [CLV21a]. To be more precise, they showed a constant decay rate for the relative entropy in this setting.

Back to the Ising model, when the maximum degree is bounded and all external fields are bounded away from 1, Chen, Liu, and Vigoda [CLV21b] established spectral independence for the subgraph world model. Using our grand model, this implies spectral independence for the random cluster model as well. However, since the random cluster model does not have conditional independence, the method of [CLV21a] does not apply. Instead, we show spectral independence for the grand model in this setting. Thus, via the method of [CLV21a] and exploiting the fact that the grand model is indeed a Gibbs distribution, we obtain a constant decay rate for the relative entropy for the (edge-flipping) Glauber dynamics for the weighted random cluster model. (We apply the result of projecting chains in Section 6.5 here again.)

However, this is still not quite enough to obtain desired mixing time bounds for the Swendsen-Wang dynamics. The reason is that the aforementioned comparison techniques of [Ull14] is an analysis of the eigenvalues of transition matrices, and thus it works only for spectral gaps but not for relative entropies. For this last step, we introduce a new comparison argument for the decay rate of relative entropies between the (edge-flipping) Glauber dynamics and the Swendsen-Wang dynamics in

Section 6.6.
To be more precise, we perform a careful analysis between the Glauber dynamics and the so-called "single-bond" dynamics introduced in [Ull14]. Our analysis utilises ideas from high-dimensional random walks [ALOV19, CGM21]. For both the Glauber dynamics and the single-bond dynamics, we decompose them into two sub-steps: the down-walk and the up-walk. Using our grand model, we bound the decay rate of relative entropy for the down-walk of Glauber dynamics. By a new comparison argument, we show that the relative entropy also decays for the down-walk of "singlebond" dynamics with a slightly weaker rate. Finally, we compare the down-walk of "single-bond" dynamics to the Swendsen-Wang dynamics via a simple application of the data processing inequality. Our analysis not only works for the decay of relative entropy, but also gives a very simple proof (see Remark 6.33) to the main result in [Ull14].

Theorem 6.2. Let $1<\beta_{\min } \leq \beta_{\max }, \Delta \geq 3$ and $0<\delta<1$ be constants. For any ferromagnetic Ising model on graph $G=(V, E)$ with parameters $\left(\beta_{e}\right)_{e \in E}$ and $\left(\lambda_{v}\right)_{v \in V}$, where $\beta_{\min } \leq \beta_{e} \leq \beta_{\text {max }}, 0<\lambda_{v} \leq 1-\delta$ and the maximum degree of $G$ is at most $\Delta$, the mixing time of Swendsen-Wang dynamics is $O\left(n \log \frac{n}{\epsilon}\right)$, where $n=|V|$.

By the same reasoning below Theorem 6.1, we do not lose generality by assum$\operatorname{ing} \beta_{\min }>1$ and $\lambda_{v}>0$ in Theorem 6.2. The big- $O$ notation hides a constant factor depending only on $\beta_{\min }, \beta_{\max }, \delta$ and $\Delta$. See (6.31) for the details of the hidden constant.

Comparing to Theorem 6.1, Theorem 6.2 has a faster mixing time bound but comes with two further assumptions: constant degree bound and no trivial field. It would be very interesting to relax either restriction. Essentially, the bottleneck in Theorem 6.1 comes from the overhead in the canonical path [JS93] or multicommodity flow method [Sin92] arguments. Unfortunately, there does not seem to be any progress in improving the mixing time bound of these methods in the last three decades. Instead, Theorem 6.2 relies on recent progress of analysing spin systems via high-dimensional random walks [CLV21a, CLV21b]. This method has very recently been generalised to bypass the bounded degree restriction [AJK ${ }^{+} 21$, CFYZ22, CE22] in various models. It is an interesting question whether this is also possible in the setting of Theorem 6.2. To bypass the no trivial field restriction, we would need a new spectral independence bound, for which there seems to be less progress. In particular, it seems hard to explain the $\Theta\left(n^{1 / 4}\right)$ mixing time on the complete graph without fields [LNNP14] with spectral independence.

Previously, most studies on Swendsen-Wang dynamics focus on the case without fields (with the exception of $\left[\mathrm{PJG}^{+} 17\right]$ discussed earlier), and are usually for the more general Potts model instead of just the Ising model. Very sharp mixing time bounds have been obtained recently, either for special cases of graphs such as $\mathbb{Z}^{d}\left[\mathrm{BCP}^{+} 21\right]$, or in the tree uniqueness region for general graphs [ $\mathrm{BCC}^{+} 22$ ]. Our Theorem 6.2 does not have these restrictions, but it only works with the presence of non-trivial external fields for the Ising model. In the settings of Theorem 6.2, we conjecture that the sharp mixing time bound is $O(\log n)$. The current argument reduces the analysis of SW dynamics to that of the single-bond dynamics, as the latter is "no-faster" in a technical sense. However, the single-bond dynamics has a $\Omega(n \log n)$ lower bound [HS07], making this line of argument difficult to approach the conjectured sharp bound for SW dynamics.

Lastly, by applying the monotone CFTP [PW96b], we obtain perfect sampling versions of the (edge-flipping) Glauber dynamics in Section 6.7 for the weighted random cluster models. Using that, we achieve perfectly sampling for the ferromagnetic Ising model with consistent external fields.

Theorem 6.3. Let $1<\beta_{\min } \leq \beta_{\max }$ be two constants. There is a perfect sampling algorithm such that given any ferromagnetic Ising model on graph $G=(V, E)$ with parameters $\left(\beta_{e}\right)_{e \in E}$ and $\left(\lambda_{v}\right)_{v \in V}$, where $\beta_{\min } \leq \beta_{e} \leq \beta_{\max }$ and $0<\lambda_{v}<1$, the algorithm returns a perfect sample in expected time $O\left(N^{4} m^{4} \log n\right)$, where $N=\min \left\{n, \frac{1}{1-\lambda_{\text {max }}}\right\}$ and $\lambda_{\text {max }}=\max _{v \in V} \lambda_{v}$.

Furthermore, if $G$ has bounded maximum degree $\Delta=O(1)$ and there exists a constant $\delta>0$ such that $\lambda_{v} \leq 1-\delta$ for all $v \in V$, the algorithm runs in expected time $O\left(n^{2} \log ^{2} n\right)$.

We remark that the overhead in monotone CFTP is $O(\log |V|)$ and there is an extra factor $m=|E|$ to implement each step of CFTP. The hidden constants can be found in (6.41).

A natural question is if we can relax the assumptions on the parameters in Theorem 6.1, 6.2, and 6.3. For anti-ferromagnetic Ising models, the sampling problem (either approximate or perfect) has no polynomial-time algorithm unless $\mathbf{N P}=\mathbf{R P}$ [JS93]. Even restricted to the ferromagnetic case, Goldberg and Jerrum [GJ07a] showed that the problem becomes \#BIS-equivalent when inconsistent fields are allowed, where \#BIS stands for counting bipartite independent sets. Its approximation complexity is a major open problem and is usually conjectured to have no polynomial-time algo-
rithm. Thus, it is unlikely to extend the range of parameters in in Theorem 6.1, 6.2, and 6.3.

Subsequent work. After our paper was posted on arXiv, Chen and Zhang [CZ23] gave a sampling algorithm of the ferromagnetic Ising model on any graph with running time $\tilde{O}(m)$, where $m$ is the number of edges, providing all the external fields are bounded away from 1 and all the edge interactions are consistent and bounded away from 1. This is a setting similar to our Theorem 6.2 without the bounded degree assumption. Their work relies heavily on our coupling result, Lemma 6.12. Furthermore, their algorithm is based on the field dynamics introduced in [CFYZ21], and does not imply mixing time bounds for either the Glauber dynamics or the Swendsen-Wang dynamics considered here.

### 6.1 Preliminaries of this chapter

This section involves numerous distributions (models) and Markov chains. The reader is welcomed to Figure 6.1 at the end of this section where a concrete example is provided.

### 6.1.1 The models and their equivalences

Here we define the weighted random cluster model, and the subgraph-world model. Then we give some equivalence results between them and the ferromagnetic Ising model.

### 6.1.1.1 Weighted random cluster model

The standard random cluster model (at $q=2$ ) is equivalent to the ferromagnetic Ising model without external fields. To handle Ising models with fields, we need to introduce weights to the random cluster model. Given a graph $G=(V, E)$, the parameters of this model are $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$, where $0<p_{e}<1$ and $\lambda_{v}>0$. The weight of any subset of edges $S \subseteq E$ is given by

$$
\begin{equation*}
\mathrm{wt}_{\mathrm{wrc}}(S):=\prod_{e \in S} p_{e} \prod_{f \in E \backslash S}\left(1-p_{f}\right) \prod_{C \in \kappa(V, S)}\left(1+\prod_{u \in C} \lambda_{u}\right), \tag{6.2}
\end{equation*}
$$

where $\kappa(V, S)$ is the set of all connected components of the graph $(V, S)$, where each $C \in \kappa(V, S)$ is a subset of vertices that forms a connected subgraph. The probability
that $S$ is drawn is

$$
\begin{equation*}
\pi_{\mathrm{wrc}}(S)=\frac{\mathrm{wt}_{\mathrm{wrc}}(S)}{Z_{\mathrm{wrc}}} \tag{6.3}
\end{equation*}
$$

where

$$
Z_{\mathrm{wrc}}=Z_{\mathrm{wrc}}(G ; \mathbf{p}, \lambda):=\sum_{S \subseteq E} \mathrm{wt}_{\mathrm{wrc}}(S)
$$

is the partition function of the weighted random cluster model. The (general) standard random cluster model allows a uniform weight $q$ for each connected component, and in the special case of $\lambda_{v}=1$ for all $v \in V$, the weighted random cluster model degenerates to the standard random cluster model at $q=2$. On the other hand, in our model the weight of each cluster depends on the vertices inside it, which makes it different from the standard random cluster models.

### 6.1.1.2 Subgraph-world model

Fix a graph $G=(V, E)$. For any subset of edges $S \subseteq E$, denote by odd $(S)$ the set of vertices with odd degree in $S$. The subgraph-world model [JS93] with parameters $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{\nu}\right)_{v \in V}$ is defined by following: each subset of edges $S$ has weight

$$
\begin{equation*}
\mathrm{wt}_{\mathrm{sg}}(S):=\prod_{e \in S} p_{e} \prod_{f \in E \backslash S}\left(1-p_{f}\right) \prod_{v \in \operatorname{odd}(S)} \eta_{v} \tag{6.4}
\end{equation*}
$$

The probability that $S$ is drawn is

$$
\begin{equation*}
\pi_{\mathrm{sg}}(S)=\frac{\mathrm{wt}_{\mathrm{sg}}(S)}{Z_{\mathrm{sg}}} \tag{6.5}
\end{equation*}
$$

where

$$
Z_{\mathrm{sg}}=Z_{\mathrm{sg}}(G ; \mathbf{p}, \eta):=\sum_{S \subseteq E} \mathrm{wt}_{\mathrm{sg}}(S)
$$

is the partition function of the subgraph-world model. In the special case where $p_{e}=$ $p \in(0,1)$ for all $e \in E$ and $\eta_{v}=0$ for all $v \in V$, the weight of any subgraph $S$ does not vanish if and only if $S$ is an even subgraph, i.e., $\operatorname{odd}(S)=\varnothing$. This yields the even subgraph model, or the so-called "high-temperature expansion" in the context of statistical mechanics.

### 6.1.1.3 Equivalences of the three models

We have the following equivalence result among the ferromagnetic Ising model with external fields, the weighted random cluster model, and the subgraph-world model. The proof of the equivalence result is given in Section 6.8.1 for completeness.

Proposition 6.4. Given any graph $G=(V, E)$, any $\beta=\left(\beta_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$ satisfying $\beta_{e}>1$ for all $e \in E$ and $0<\lambda_{v} \leq 1$ for all $v \in V$, it holds that

$$
\begin{equation*}
\left(\prod_{e \in E} \beta_{e}\right) \cdot Z_{w r c}(G ; 2 \mathbf{p}, \lambda)=Z_{\text {Ising }}(G ; \beta, \lambda)=\left(\prod_{v \in V}\left(1+\lambda_{v}\right)\right)\left(\prod_{e \in E} \beta_{e}\right) Z_{s g}(G ; \mathbf{p}, \eta) \tag{6.6}
\end{equation*}
$$

where $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ satisfying $p_{e}=\frac{1}{2}\left(1-\frac{1}{\beta_{e}}\right)$ and $\eta=\left(\eta_{v}\right)_{v \in V}$ satisfying $\eta_{v}=\frac{1-\lambda_{v}}{1+\lambda_{v}}$.
In addition, there are also probabilistic equivalence relations among the models, which will be the topic in Section 6.2.

Remark 6.5. For the ferromagnetic Ising model $(G ; \beta, \lambda)=(G ; \beta, 1)$, where $\beta_{e}=\beta>$ 1 for all $e \in E$ and $\lambda_{v}=1$ for all $v \in V$, its relationship with the even subgraph model and the random cluster model is well known (see e.g. [vdW41, FK72, Gri06]). Formally,

$$
\beta^{|E|} Z_{\mathrm{wrc}}(G ; 2 p, 1)=Z_{\mathrm{Ising}}(G ; \beta, 1)=2^{|V|} \beta^{|E|} Z_{\mathrm{sg}}(G ; p, 0) \text { where } p=\frac{1}{2}\left(1-\frac{1}{\beta}\right),
$$

which is a special case of Proposition 6.4.

### 6.1.2 Markov chains and down-up walks

In this part, we consider two Markov chains: Glauber dynamics and Swendsen-Wang dynamics. It will be convenient for us to view Glauber dynamics as a so-called "downup" walk, which we will define next.

Let $\Omega_{0}$ and $\Omega_{1}$ denote two finite state spaces. Let $\mu_{0}$ and $\mu_{1}$ denote two distributions over $\Omega_{0}$ and $\Omega_{1}$ respectively. For $f, g: \Omega_{i} \rightarrow \mathbb{R}$, define $\langle f, g\rangle_{\mu_{i}}=\sum_{x \in \Omega_{i}} \mu_{i}(x) f(x) g(x)$. Let $P^{\uparrow}: \Omega_{0} \times \Omega_{1} \rightarrow \mathbb{R}_{\geq 0}$ and $P^{\downarrow}: \Omega_{1} \times \Omega_{0} \rightarrow \mathbb{R}_{\geq 0}$ denote two transition matrices. We say $P^{\uparrow}$ and $P^{\downarrow}$ are a pair of adjoint operator if

$$
\forall f: \Omega_{0} \rightarrow \mathbb{R}, g: \Omega_{1} \rightarrow \mathbb{R}, \quad\left\langle f, P^{\uparrow} g\right\rangle_{\mu_{0}}=\left\langle P^{\downarrow} f, g\right\rangle_{\mu_{1}}
$$

The following equation holds for adjoint $P^{\uparrow}$ and $P^{\downarrow}$ :

$$
\forall x_{0} \in \Omega_{0}, x_{1} \in \Omega_{1}, \quad \mu_{0}\left(x_{0}\right) P^{\uparrow}\left(x_{0}, x_{1}\right)=\mu_{1}\left(x_{1}\right) P^{\downarrow}\left(x_{1}, x_{0}\right) .
$$

Moreover, for any distribution $v$ over $\Omega_{1}$ and $f=\frac{v}{\mu_{1}}$, it holds that

$$
D_{\mathrm{KL}}\left(v P^{\downarrow} \| \mu_{0}\right)=\operatorname{Ent}_{\mu_{1}}\left(P^{\uparrow} f\right) \text { and } D_{\chi^{2}}\left(v P^{\downarrow} \| \mu_{0}\right)=\operatorname{Var}_{\mu_{1}}\left(P^{\uparrow} f\right)
$$

It is straightforward to verify $P^{\vee}=P^{\downarrow} P^{\uparrow}$ and $P^{\wedge}=P^{\uparrow} P^{\downarrow}$ are self-adjoint, i.e. $\left\langle f, P^{\vee} g\right\rangle_{\mu_{1}}=$ $\left\langle P^{\vee} f, g\right\rangle_{\mu_{1}}$ and $\left\langle f, P^{\wedge} g\right\rangle_{\mu_{0}}=\left\langle P^{\wedge} f, g\right\rangle_{\mu_{0}}$. Hence, $P^{\vee}$ and $P^{\wedge}$ are reversible with respect to $\mu_{1}$ and $\mu_{0}$ respectively.

### 6.1.2.1 Glauber dynamics.

Given a distribution $\mu$ with support $Q^{V}$, let $\Omega_{1}=Q^{V}$ and $\Omega_{0}=\left\{\sigma \in Q^{V \backslash\{v\}} \mid v \in V\right\}$. and the current state $X \in \Omega$, the transition $X \rightarrow X^{\prime}$ of Glauber dynamics can be interpreted as the following two steps

- down walk $P_{\text {Glauber }}^{\downarrow}$ : pick $v \in V$ uniformly at random and $\operatorname{transform} X \in \Omega_{1}$ to $X_{V \backslash v} \in \Omega_{0} ;$
- up walk $P_{\text {Glauber }}^{\uparrow}$ : sample $c \sim \mu_{v}^{X_{V \backslash\{v\}}}$ and transform $X_{V \backslash v} \in \Omega_{0}$ to $X^{\prime} \in \Omega_{1}$ such that $X_{v}^{\prime}=c$ and $X_{V \backslash\{v\}}^{\prime}=X_{V \backslash\{v\}}$.

In other words, we clear the state of a vertex picked uniformly at random, and resample its new state proportional to the conditional probability.

Let $\mu_{0}:=\mu P_{\text {Glauber }}^{\downarrow}$ be a distribution over $\Omega_{0}$. Then $P_{\text {Glauber }}^{\downarrow}$ and $P_{\text {Glauber }}^{\uparrow}$ is a pair of adjoint operators with respect to distributions $\mu_{1}=\mu$ and $\mu_{0}$. Thus, Glauber dynamics is a down-up walk and is reversible with respect to $\mu$.

### 6.1.2 2 Swendsen-Wang dynamics.

Let $G=(V, E)$ be a graph. Consider the ferromagnetic Ising model on $G$ with parameters $\beta=\left(\beta_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$, where $\beta_{e}>1$ for all $e \in E$, and the weighted random cluster model on $G$ with parameters $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$, where $p_{e}=1-\frac{1}{\beta_{e}}$ for all $e \in E$. Recall $\pi_{\text {Ising }}$ from (6.1) and $\pi_{\text {wrc }}$ from (6.3).

Define the following two transformations between Ising and weighted random cluster models.

- $P_{I \rightarrow \mathcal{R}}:\{0,1\}^{V} \rightarrow 2^{E}$ : Given any Ising configuration $\sigma \in\{0,1\}^{V}, P_{I \rightarrow \mathcal{R}}$ transforms $\sigma$ into a weighted random cluster model configuration $S \subseteq E$. For each edge $e=\{u, v\} \in E$ with $\sigma(u)=\sigma(v)$, add $e$ independently into $S$ with probability $p_{e}=1-\frac{1}{\beta_{e}}$. Formally,
$\forall \sigma \in\{0,1\}^{V}, S \subseteq E, \quad P_{I \rightarrow \mathcal{R}}(\sigma, S)=\mathbb{I}[S \subseteq M(\sigma)] \cdot \prod_{e \in S}\left(1-\frac{1}{\beta_{e}}\right) \cdot \prod_{f \in M(\sigma) \backslash S} \frac{1}{\beta_{f}}$,
where $M(\sigma)=\left\{e=\{u, v\} \in E \mid \sigma_{u}=\sigma_{v}\right\}$ is the set of monochromatic edges with respect to $\sigma$.
- $P_{\mathcal{R} \rightarrow I}: 2^{E} \rightarrow\{0,1\}^{V}:$ Given any weighted random cluster model configuration $S \subseteq E, P_{\mathcal{R} \rightarrow I}$ transforms $S$ to an Ising configuration $\sigma \in\{0,1\}^{V}$. For each connected component $C \subseteq V$ in graph $G^{\prime}=(V, S)$, sample $x_{C} \in\{0,1\}$ independently according to the following distribution

$$
x_{C}= \begin{cases}1 & \text { with probability } \frac{\prod_{v \in C} \lambda_{v}}{1+\prod_{v \in C} \lambda_{v}} ; \\ 0 & \text { with probability } \frac{1}{1+\prod_{v \in C} \lambda_{v}},\end{cases}
$$

and then let $\sigma(v)=x_{C}$ for all vertices $v \in C$. Formally,

$$
\begin{equation*}
\forall \sigma \in\{0,1\}^{V}, S \subseteq E, \quad P_{\mathcal{R} \rightarrow I}(S, \sigma)=\mathbb{I}[S \subseteq M(\sigma)] \cdot \prod_{C \in \kappa(V, S)} \frac{\prod_{v \in C} \lambda_{v}^{\sigma(v)}}{1+\prod_{v \in C} \lambda_{v}}, \tag{6.8}
\end{equation*}
$$

where $\kappa(V, S)$ is the set of connected components in graph $G^{\prime}=(V, S)$.
The Swendsen-Wang dynamics for Ising models is defined by

$$
\begin{equation*}
P_{\mathrm{SW}}^{\mathrm{Ising}}:=P_{I \rightarrow \mathcal{R}} P_{\mathcal{R} \rightarrow I}, \tag{6.9}
\end{equation*}
$$

and the Swendsen-Wang dynamics for weighted random cluster models is defined by

$$
\begin{equation*}
P_{\mathrm{SW}}^{\mathrm{wrc}}:=P_{\mathcal{R} \rightarrow I} P_{I \rightarrow \mathcal{R}} . \tag{6.10}
\end{equation*}
$$

The following adjoint result about Swendsen-Wang dynamics is well-known. However, here we consider more general Ising models with external fields and weighted random cluster models. For completeness, we provide a proof of the following proposition in Section 6.8.2.

Proposition 6.6. For any functions $f:\{0,1\}^{V} \rightarrow \mathbb{R}$ and $g: 2^{E} \rightarrow \mathbb{R}$, it holds that

$$
\begin{equation*}
\left\langle f, P_{I \rightarrow \mathcal{R}} g\right\rangle_{\pi_{\text {Ising }}}=\left\langle P_{\mathcal{R} \rightarrow I} f, g\right\rangle_{\pi_{\mathrm{wrc}}} . \tag{6.11}
\end{equation*}
$$

By Proposition 6.6, it holds that $\pi_{\text {Ising }} P_{I \rightarrow \mathcal{R}}=\pi_{\text {wrc }}$ and $\pi_{\text {wrc }} P_{\mathcal{R} \rightarrow I}=\pi_{\text {Ising }}$. Both $P_{\mathrm{SW}}^{\mathrm{Is} i n g}$ and $P_{\mathrm{SW}}^{\mathrm{wrc}}$ are down-up walks, and their stationary distributions are $\pi_{\text {Ising }}$ and $\pi_{\text {wrc }}$ respectively.

Finally, the mixing times of $P_{\mathrm{SW}}^{\text {Ising }}$ and $P_{\mathrm{SW}}^{\mathrm{wrc}}$ have the following relationships:

$$
\begin{equation*}
T_{\mathrm{mix}}\left(P_{\mathrm{SW}}^{\mathrm{Ising}}, \epsilon\right) \leq T_{\mathrm{mix}}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}, \epsilon\right)+1 \quad \text { and } \quad T_{\mathrm{mix}}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}, \epsilon\right) \leq T_{\mathrm{mix}}\left(P_{\mathrm{SW}}^{\mathrm{Ising}}, \epsilon\right)+1 \tag{6.12}
\end{equation*}
$$

We prove the first one, the second one holds similarly. Let $T=T_{\text {mix }}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}, \epsilon\right)$. For any distribution $v$ over $\{0,1\}^{V}$, we have

$$
\begin{aligned}
d_{\mathrm{TV}}\left(v\left(P_{\mathrm{SW}}^{\mathrm{Ising}}\right)^{T+1}, \pi_{\mathrm{Ising}}\right) & =d_{\mathrm{TV}}\left(\left(v P_{I \rightarrow \mathcal{R}}\right)\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right)^{T} P_{\mathcal{R} \rightarrow I}, \pi_{\mathrm{wrc}} P_{\mathcal{R} \rightarrow I}\right) \\
\text { (by data processing inequality) } & \leq d_{\mathrm{TV}}\left(\left(v P_{I \rightarrow \mathcal{R}}\right)\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right)^{T}, \pi_{\mathrm{wrc}}\right) \leq \epsilon .
\end{aligned}
$$

### 6.1.3 Canonical paths and variance decay

Let $P$ denote a random walk over $\Omega$ that is reversible with respect to $\mu$. It is wellknown that $P$ has real eigenvalues $1=\lambda_{1} \geq \lambda_{2} \geq \ldots \geq \lambda_{|\Omega|}$. The spectral gap is defined by

$$
\mathfrak{b a p}(P)=1-\lambda_{2} .
$$

Define the Dirichlet form of $P$ by for any functions $f, g: \Omega \rightarrow \mathbb{R}$,

$$
\mathcal{E}_{P}(f, g)=\langle f,(I-P) g\rangle_{\mu}=\frac{1}{2} \sum_{x, y \in \Omega} \mu(x) P(x, y)(f(x)-f(y))(g(x)-g(y)) .
$$

We can also characterise the spectral gap $\mathfrak{G a p}(P)$ in a variational form:

$$
\begin{equation*}
\mathfrak{G a p}(P)=\inf \left\{\left.\frac{\mathcal{E}_{P}(f, f)}{\operatorname{Var}_{\mu}(f)} \right\rvert\, f: \Omega \rightarrow \mathbb{R} \wedge \operatorname{Var}_{\mu}(f) \neq 0\right\} \tag{6.13}
\end{equation*}
$$

A useful tool to analyse the spectral gap of a reversible Markov chains is the canonical path introduced by Jerrum and Sinclair [JS89]. Let $P$ be a reversible Markov chain over the state space $\Omega$ with stationary distribution $\pi$. Let $\gamma_{X Y}=\left(Z_{0}=X, Z_{1}, Z_{2}, \ldots, Z_{\ell}=\right.$ $Y$ ) be a path of length $\ell$ moving in the state space using transitions of $P$, i.e. for any $i \in[\ell], P\left(Z_{i-1}, Z_{i}\right)>0$. For each pair of $X, Y \in \Omega$, its path $\gamma_{X Y}$ is assigned a weight $w\left(\gamma_{X Y}\right)=\mu(X) \mu(Y)$. Let $\Gamma$ be the collection of all canonical paths. The congestion of $\Gamma$ is defined by

$$
\begin{equation*}
\varrho(\Gamma):=\max _{\left(Z, Z^{\prime}\right) \in \Omega^{2}, P\left(Z, Z^{\prime}\right)>0} \frac{L}{\mu(Z) P\left(Z, Z^{\prime}\right)} \sum_{\gamma \in \Gamma:\left(Z, Z^{\prime}\right) \in \gamma} w(\gamma) \tag{6.14}
\end{equation*}
$$

where $L$ is the maximum length of path in $\Gamma$. Sinclair [Sin92] showed that the congestion of any collection of paths $\Gamma$ for a Markov chain $P$ is an upper bound of the inverse of its spectral gap, namely,

$$
\frac{1}{\mathfrak{G a p}(P)} \leq \varrho(\Gamma)
$$

Consider the down-up walk $P^{\vee}=P^{\downarrow} P^{\uparrow}$ over $\Omega_{1}$, where $P^{\downarrow}: \Omega_{1} \times \Omega_{0} \rightarrow \mathbb{R}_{\geq 0}$ and $P^{\uparrow}: \Omega_{0} \times \Omega_{1} \rightarrow \mathbb{R}_{\geq 0}$ are a pair of adjoint operators with respect to distribution $\mu_{0}$ over $\Omega_{0}$ and $\mu_{1}$ over $\Omega_{1}$. For simplicity, we denote $\Omega_{1}$ by $\Omega$, and we denote $\mu_{1}$ by $\mu$. The following result holds for $P^{\vee}$.
Proposition 6.7. Let $P^{\vee}=P^{\downarrow} P^{\uparrow}$ be a down-up walk over $\Omega$ that is reversible with respect to $\mu$. For any $0<\delta<1$, the spectral gap $\mathfrak{G a p}\left(P^{\vee}\right) \geq \delta$ if and only if for any distribution $\nu$ over $\Omega$,

$$
\begin{equation*}
D_{\chi^{2}}\left(v P^{\downarrow} \| \mu P^{\downarrow}\right) \leq(1-\delta) D_{\chi^{2}}(v \| \mu) \tag{6.15}
\end{equation*}
$$

Proof. Let $f=\frac{v}{\mu}$. It holds that

$$
\mathcal{E}_{P^{\vee}}(f, f)=\langle f, f\rangle_{\mu}-\left\langle f, P^{\vee} f\right\rangle_{\mu}=\langle f, f\rangle_{\mu}-\left\langle P^{\uparrow} f, P^{\uparrow} f\right\rangle_{\mu_{0}}=\operatorname{Var}_{\mu}(f)-\operatorname{Var}_{\mu_{0}}\left(P^{\uparrow} f\right) .
$$

Then the lemma follows from $D_{\chi^{2}}\left(v P^{\downarrow} \| \mu P^{\downarrow}\right)=\operatorname{Var}_{\mu_{0}}\left(P^{\uparrow} f\right), D_{\chi^{2}}(v \| \mu)=\operatorname{Var}_{\mu}(f)$, and (6.13).

### 6.1.4 Spectral independence and entropy decay

Let $Q$ be a finite set. Let $\mu$ be a distribution with support $Q^{V}$. Fix a partial pinning $\tau \in Q^{\Lambda}$ for some $\Lambda \subseteq V$. Define the absolute influence matrix $\Psi_{\mu}^{\tau}$ by

$$
\begin{gathered}
\forall u, v \in V \backslash \Lambda \text { with } u \neq v, \quad \Psi_{\mu}^{\tau}(u, v):=\max _{i, j \in Q} d_{\mathrm{TV}}\left(\mu_{v}^{\tau \wedge(u \leftarrow i)}, \mu_{v}^{\tau \wedge(u \leftarrow j)}\right) \\
\forall v \in V \backslash \Lambda, \quad \Psi_{\mu}^{\tau}(v, v):=0 .
\end{gathered}
$$

where $d_{\mathrm{TV}}(\cdot, \cdot)$ denotes the total variation distance and $\mu_{v}^{\tau \wedge(u \leftarrow i)}$ denotes the marginal distribution on $v$ conditional on that variables in $\Lambda$ take the value $\tau$ and $u$ takes the value $i$. We say that the distribution $\mu$ is $\ell_{\infty}$-spectrally independent with parameter $\zeta$ if

$$
\forall \Lambda \subset V, \sigma \in Q^{\Lambda}, \quad\left\|\Psi_{\mu}^{\sigma}\right\|_{\infty}=\max _{u \notin \Lambda} \sum_{v \notin \Lambda} \Psi_{\mu}^{\sigma}(u, v) \leq \zeta .
$$

Call $\mu$ b-marginally bounded if

$$
\min _{\Lambda \subseteq V, v \notin \Lambda} \min _{\sigma \in Q^{\wedge}, c \in Q} \mu_{v}^{\sigma}(c) \geq b .
$$

In this part, we are particularly interested in Gibbs distributions. We will consider a slightly more general than usual version defined over hypergraphs. Let $H=(V, \mathcal{E})$ be a hypergraph. Given weight functions $\left(\phi_{v}\right)_{v \in V}$ and $\left(\phi_{e}\right)_{e \in \mathcal{E}}$, where $\phi_{v}: Q \rightarrow \mathbb{R}_{>0}$ and $\phi_{e}: Q^{e} \rightarrow \mathbb{R}_{>0}$, define the Gibbs distribution $\mu$ over $Q^{V}$ by

$$
\forall \sigma \in Q^{V}, \quad \mu(\sigma) \propto \prod_{v \in V} \phi_{v}\left(\sigma_{v}\right) \prod_{e \in \mathcal{E}} \phi_{e}\left(\sigma_{e}\right) .
$$

Let $G_{\mu}=(V, E)$ be a graph such that $\{u, v\} \in E$ if $u \in e^{\prime}$ and $v \in e^{\prime}$ for some $e^{\prime} \in \mathcal{E}$. For any disjoint $A, B, C \subseteq V$, if the removal of $C$ disconnects $A$ and $B$ in $G_{\mu}$, it holds that variables in $A$ and $B$ are independent in $\mu$ conditional on any assignment on $C$. Define maximum degree $D_{\mu}$ of the Gibbs distribution $\mu$ as the maximum degree of the graph $G_{\mu}$.

The spectral independence is related to the mixing time of Glauber dynamics. The following result is proved in [CLV21a, $\mathrm{BCC}^{+} 22$ ] (see also [CLV21b, Theorem 13])

Theorem 6.8 ([CLV21a, $\left.\mathrm{BCC}^{+} 22\right]$ ). Let $\zeta, b, D>0$. For any Gibbs distribution $\mu$ over $Q^{V}$, where $|V|=n$, if $\mu$ is $\ell_{\infty}$-spectrally independent with parameter $\zeta$, b-marginally bounded and has the maximum degree at most $D$, then the down walk of the Glauber dynamics satisfies that
$\forall$ distribution $v$ over $Q^{V}, \quad D_{\mathrm{KL}}\left(v P_{\text {Glauber }}^{\downarrow} \| \mu P_{\text {Glauber }}^{\downarrow}\right) \leq\left(1-\frac{1}{C n}\right) D_{\mathrm{KL}}(v \| \mu)$,
where $C=\left(\frac{D}{b}\right)^{1+2\left[\frac{\zeta}{b}\right]}>1$ is a constant depending only on $\zeta, b$ and $D$.
In [CLV21a, $\mathrm{BCC}^{+} 22$ ], they mainly establish the so-called "approximate tensorization of entropy" property for $\mu$. However this is equivalent to the contraction of relative entropy by $P_{\text {Glauber }}^{\downarrow}$ [CLV21a].

### 6.1.5 Holographic transformation

We will need holographic transformations [Val08] to show couplings between the subgraph-world model and the weighted random cluster model. Let $f:\{0,1\}^{d} \rightarrow$ $\mathbb{C}$ be a function. We may represent it by a vector (either row or column vector) $\left(f_{0}, \cdots, f_{x}, \cdots, f_{2^{d}-1}\right)$ where $f_{x}$ is the value of $f$ on $x \in\{0,1\}^{d}$ by regarding $x$ as a binary representation. In the symmetric case where $f$ is invariant under permutations of indices, we use a succinct "signature" $\left[f_{0}, \cdots, f_{w}, \cdots, f_{d}\right]$ to express $f$, where $f_{w}$ is the value of $f$ on inputs of Hamming weight $w$, i.e. all $x \in\{0,1\}^{d}$ satisfying $|x|=w$.

Given a bipartite graph $H=(V, E)$ with partition $V=V_{1} \uplus V_{2}$. Let $\mathcal{F}=\left(f_{v}\right)_{v \in V_{1}}$ and $\mathcal{G}=\left(g_{v}\right)_{v \in V_{2}}$ be two sets of functions such that the arity of the function is the degree of the corresponding vertex. The (bipartite) Holant (an edge weighted partition function) is defined by

$$
\operatorname{Holant}(H ; \mathcal{F} \mid \mathcal{G}):=\sum_{\sigma: E \rightarrow\{0,1\}} \prod_{v \in V_{1}} f_{v}\left(\left.\sigma\right|_{E(v)}\right) \prod_{u \in V_{2}} g_{u}\left(\left.\sigma\right|_{E(u)}\right),
$$

where $\left.\sigma\right|_{E(v)}$ stands for the restriction of the assignment $\sigma$ to the incident edges of $v .{ }^{3}$

Let $\boldsymbol{M}$ be a $2 \times 2$ matrix and $f$ be a function of arity $d$. If $f$ is represented by a column (resp. row) vector, we write $\boldsymbol{M} f=\boldsymbol{M}^{\otimes d} f$ (resp. $f \boldsymbol{M}=f \boldsymbol{M}^{\otimes d}$ ) as the transformed signature. Given $\operatorname{Holant}(H ; \mathcal{F} \mid \mathcal{G})$ and an invertible matrix $\boldsymbol{T} \in \mathbb{C}^{2 \times 2}$, we

[^13]view signatures in $\mathcal{F}$ as row vectors and define $\mathcal{F} \boldsymbol{T}=\left\{f_{v}^{\prime} \mid v \in V_{1} \wedge f_{v}^{\prime}=f_{v} \mathbf{T}\right\}$; and view signatures in $\mathcal{G}$ as column vectors and define $\boldsymbol{T}^{-1} \mathcal{G}=\left\{g_{v}^{\prime} \mid v \in V_{2} \wedge g_{v}^{\prime}=\mathbf{T}^{-1} g_{v}\right\}$. Valiant's celebrated Holant Theorem [Val08] states

Theorem 6.9. $\operatorname{Holant}(H ; \mathcal{F} \mid \mathcal{G})=\operatorname{Holant}\left(H ; \mathcal{F} \boldsymbol{T} \mid \boldsymbol{T}^{-1} \mathcal{G}\right)$ for any invertible $\boldsymbol{T} \in \mathbb{C}^{2 \times 2}$.

### 6.2 The grand model and a generalised GrimmettJanson coupling

We introduce a grand model, inspired by [GJ07b], that unifies the subgraph and random cluster models introduced in Section 6.1.1. We also generalise the coupling of Grimmett and Janson [GJ07b] for ferromagnetic Ising models with external fields. It is possible to also include vertex configurations in this grand model à la Edwards and Sokal [ES88], so that the Ising model is also unified under this framework. However it does not appear to have much benefit and we choose not to do so.

### 6.2.1 The grand model

Let $G=(V, E)$ be a simple undirected graph. The grand model, specified by parameters $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{v}\right)_{v \in V}$ where $0 \leq p_{e} \leq 1 / 2$ and $0 \leq \eta_{v} \leq 1$, defines a distribution $\pi_{\mathrm{gm}}$ over all configurations on the edges of three states $X: E \rightarrow\{0,1,2\}$. Given an assignment $X$ in the grand model, denote by $X^{-1}(q)$ the set of edges that are assigned $q$ under $X$ where $q=0,1,2$. The weight of each configuration is given by

$$
\begin{equation*}
\mathrm{wt}_{\mathrm{gm}}(X)=\prod_{e \in X^{-1}(\{1,2\})} p_{e} \prod_{f \in X^{-1}(0)}\left(1-2 p_{f}\right) \prod_{v \in O(X)} \eta_{v} \tag{6.16}
\end{equation*}
$$

where $O(X)$ is the set of vertices of odd degree in the subgraph $\left(V, X^{-1}(1)\right)$. The probability of each configuration $X$ is

$$
\begin{equation*}
\pi_{\mathrm{gm}}(X)=\frac{\mathrm{wt}_{\mathrm{gm}}(X)}{Z_{\mathrm{gm}}} \tag{6.17}
\end{equation*}
$$

where

$$
Z_{\mathrm{gm}}=Z_{\mathrm{gm}}(G ; \mathbf{p}, \eta):=\sum_{X \in \Omega_{\mathrm{gm}}(G)} \mathrm{wt} \mathrm{gm}_{\mathrm{gm}}(X)
$$

is the partition function of the grand model.
Equivalently, a random sample from the grand model can be generated by the following procedure.

- Step-I: Sample $S \sim \pi_{\mathrm{sg}}$, where $\pi_{\mathrm{sg}}$ is the distribution specified by the subgraphworld model with parameters ( $\mathbf{p}, \eta$ ); for each $e \in E$, let $X(e)=1$ if $e \in S$ and let $X(e)=*$ if $e \notin S$.
- Step-II: Independently for each $e \in E$ with $X_{e}=*$, set $X(e)=2$ with probability $\frac{p_{e}}{1-p_{e}}$, and $X(e)=0$ otherwise.

It is straightforward to verify that the outcome distribution is exactly the grand model distribution.

Recall the definition of a Gibbs distribution and its maximum degree in Section 6.1.4. The grand model is indeed a Gibbs distribution in the sense of Theorem 6.8. Each edge of $G$ corresponds to a variable, and each vertex $v \in V$ corresponds to a weight function. In other words, this is a Holant-type problem [CLX11]. Theorem 6.8 applies to Holant-type problems, as explained in [CLV21b, Section 2.2]. The underlying graph of $\pi_{\mathrm{gm}}$ (as defined in Section 6.1.4) is the line graph of $G$, whose maximum degree is at most $2 \Delta-1$. Thus we have the following observation.

Observation 6.10. The distribution $\pi_{\mathrm{gm}}$ is a Gibbs distribution with maximum degree $D \leq 2 \Delta-1$, where $\Delta$ is the maximum degree of the graph $G=(V, E)$.

The next lemma gives the relation among the grand model, the subgraph-world model and the random cluster model.

Lemma 6.11. Let $X \sim \pi_{\mathrm{gm}}$ be a random sample from the grand model with parameter $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{v}\right)_{v \in V}$, where $0 \leq p_{e} \leq 1 / 2$ and $0 \leq \eta_{v} \leq 1$. It holds that

- $\mathcal{S}=\{e \in E \mid X(e)=1\}$ follows the distribution specified by the subgraph-world model with parameters $(\mathbf{p}, \eta)$;
- $\mathcal{R}=\{e \in E \mid X(e)=1 \vee X(e)=2\}$ follows the distribution specified by the random cluster model with parameters $(2 \mathbf{p}, \lambda)$, where $\lambda_{v}=\frac{1-\eta_{v}}{1+\eta_{v}}$ for all $v \in V$.

Namely, $X(e)=1$ means $e$ is present in the subgraph-world model (Step-I), and $X(e)=2$ means $e$ is absent in the subgraph-world model, but gets added into the random cluster model in Step-II. $X(e)=0$ means $e$ is absent in both models.

The first part of Lemma 6.11 holds trivially. The second part is proved by a generalised Grimmett-Janson coupling [GJ07b]. The proof of the second part is given in Section 6.2.2.

### 6.2.2 Coupling via holographic transformation

Under the unweighted setting, Grimmett and Janson [GJ07b, Theorem 3.5] discovered a coupling between random even subgraphs and random cluster configurations. The following lemma is a generalisation to the weighted case via holographic transformations.

Lemma 6.12. Let $G=(V, E)$ be a graph, $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{v}\right)_{v \in V}$, where $0 \leq$ $p_{e} \leq 1 / 2$ for all $e \in E$ and $\eta_{v} \geq 0$ for all $v \in V$. Let $\mathcal{S} \subseteq E$ be a random sample from the subgraph-world model $(G ; \mathbf{p}, \eta)$. Let $\mathcal{R}$ be $\mathcal{S}$ with each remaining edge $e \in E \backslash \mathcal{S}$ added into $\mathcal{R}$ independently with probability $p_{e} /\left(1-p_{e}\right)$. Then the random subgraph $\mathcal{R}$ satisfies the distribution of the random cluster model with parameter $(2 \mathbf{p}, \lambda)$ where $\eta_{v}=\frac{1-\lambda_{v}}{1+\lambda_{v}}$ for all $v \in V$.

We remark that the second part of Lemma 6.11 is a straightforward consequence of Lemma 6.12. We need the following lemma to prove Lemma 6.12.

Lemma 6.13. Let $G=(V, E)$ be a graph. Let $\lambda=\left(\lambda_{v}\right)_{v \in V}$ where $0 \leq \lambda_{v}<1$ for all $v \in V$. For each $v \in V$, let $\eta_{v}=\frac{1-\lambda_{v}}{1+\lambda_{v}}$. It holds that

$$
\begin{equation*}
\prod_{C \in \kappa(V, E)}\left(1+\prod_{u \in C} \lambda_{u}\right)=\left(\prod_{v \in V}\left(1+\lambda_{v}\right)\right)\left(\frac{1}{2}\right)^{|E|} \sum_{E^{\prime} \subset E} \prod_{u \in \operatorname{odd}\left(E^{\prime}\right)} \eta_{u}, \tag{6.18}
\end{equation*}
$$

where $\kappa(V, E)$ is the set of connected components in graph $G=(V, E)$.
Proof. Define a bipartite graph $H$ with left part $V_{1}=V$ corresponding to vertices in $G$ and right part $V_{2}=E$ corresponding to edges in $G$. Two vertices $v \in V_{1}$ and $e \in V_{2}$ are adjacent in $H$ if $v$ is incident to $e$ in $G$. Let $d_{v}$ denote the degree of $v$ in $G$. Consider the following set of signatures

$$
\begin{aligned}
\mathcal{F}^{(1)} & =\left\{f_{v}^{(1)}=[1,0]^{\otimes d_{v}}+\lambda_{v}[0,1]^{\otimes d_{v}} \mid v \in V\right\}, \\
\mathcal{F}^{(2)} & =\left\{\left.f_{v}^{(2)}=\frac{1}{1+\lambda_{v}}\left([1,1]^{\otimes d_{v}}+\lambda_{v}[1,-1]^{\otimes d_{v}}\right) \right\rvert\, v \in V\right\}, \\
\mathcal{G} & =\left\{g_{e}=[1,0,1] \mid e \in E\right\} .
\end{aligned}
$$

We remark that $f_{v}^{(2)}=\left[1, \eta_{v}, 1, \eta_{v}, \ldots\right]$. Let $\mathbf{T}=\left(\begin{array}{cc}1 & 1 \\ 1 & -1\end{array}\right)$. Observe that $f_{v}^{(1)} \mathbf{T}=(1+$ $\left.\lambda_{v}\right) f_{v}^{(2)}$ and $\mathbf{T}^{-1} g_{e}=\frac{1}{2} g_{e}$. By Theorem 6.9, it holds that

$$
\begin{equation*}
\text { Holant }\left(H ; \mathcal{F}^{(1)} \mid \mathcal{G}\right)=\left(\prod_{v \in V}\left(1+\lambda_{v}\right)\right)\left(\frac{1}{2}\right)^{|E|} \operatorname{Holant}\left(H ; \mathcal{F}^{(2)} \mid \mathcal{G}\right) \tag{6.19}
\end{equation*}
$$

This equation is indeed (6.18) in disguise. The equivalence between the left-hand sides of (6.19) and (6.18) is a simple observation that the signature [1,0,1] on the edge forces the spins of vertices in each connected component $C$ to be the same. Each component contributes a weight $1+\prod_{u \in C} \lambda_{u}$. The equivalence between the right-hand sides of (6.19) and (6.18) follows from how $\mathcal{F}^{(2)}$ and $\mathcal{G}$ are defined. This proves the lemma.

Proof of Lemma 6.12. For each subgraph $R \subseteq E$ of $G=(V, E)$,

$$
\begin{align*}
\operatorname{Pr}[\mathcal{R}=R] & =\frac{1}{Z_{\mathrm{sg}}(G ; \mathbf{p}, \eta)} \sum_{S \subseteq R} \prod_{u \in \operatorname{odd}(S)} \eta_{u} \prod_{e \in S} p_{e} \prod_{f \in E \backslash S}\left(1-p_{f}\right) \prod_{g \in R \backslash S} \frac{p_{g}}{1-p_{g}} \prod_{h \in E \backslash R} \frac{1-2 p_{h}}{1-p_{h}} \\
& =\frac{1}{Z_{\mathrm{sg}}(G ; \mathbf{p}, \eta)} \sum_{S \subseteq R} \prod_{u \in \operatorname{odd}(S)} \eta_{u} \prod_{e \in R} p_{e} \prod_{f \in E \backslash R}\left(1-2 p_{f}\right) \\
& =\frac{1}{Z_{\mathrm{sg}}(G ; \mathbf{p}, \eta)} 2^{-|R|} \prod_{e \in R}\left(2 p_{e}\right) \prod_{f \in E \backslash R}\left(1-2 p_{f}\right) \sum_{S \subseteq R} \prod_{u \in \operatorname{odd}(S)} \eta_{u} \\
& =\frac{1}{Z_{\mathrm{sg}}(G ; \mathbf{p}, \eta)} \prod_{e \in R}\left(2 p_{e}\right) \prod_{f \in E \backslash R}\left(1-2 p_{f}\right) \prod_{v \in V} \frac{1}{1+\lambda_{v}} \prod_{C \in \kappa(V, R)}\left(1+\prod_{u \in C} \lambda_{u}\right) \\
& =\frac{1}{Z_{\mathrm{wrc}}(G ; 2 \mathbf{p}, \lambda)} \prod_{e \in R}\left(2 p_{e}\right) \prod_{f \in E \backslash R}\left(1-2 p_{f}\right) \prod_{C \in \kappa(V, R)}\left(1+\prod_{u \in C} \lambda_{u}\right) .
\end{align*}
$$

$$
=\pi_{\mathrm{wrc}}(R) .
$$

### 6.3 Variance decay of Glauber dynamics on the grand model

Let $G=(V, E)$ be a graph. Let $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{v}\right)_{v \in V}$, where $0<p_{e}<1 / 2$ and $0<\eta_{v}<1$. Let $\pi_{\mathrm{gm}}$ denote the distribution specified by the grand model with parameters $\mathbf{p}$ and $\eta$. Let $\Omega\left(\pi_{\mathrm{gm}}\right)$ denote the support of $\pi_{\mathrm{gm}}$. We use $P_{\text {GlauberGM }}$ to denote Glauber dynamics on $\pi_{\mathrm{gm}}$ as defined in Section 6.1.2.1.

Lemma 6.14. The Glauber dynamics $P_{\text {GlauberGM }}$ satisfies that for any distribution $v$ with support $\Omega(v) \subseteq \Omega\left(\pi_{\mathrm{gm}}\right)$,
$D_{\chi^{2}}\left(v P_{\text {GlauberGM }}^{\downarrow} \| \pi_{\mathrm{gm}} P_{\text {GlauberGM }}^{\downarrow}\right) \leq\left(1-\frac{\eta_{\min }^{4} \min \left\{p_{\min }, 1-2 p_{\max }\right\}}{m^{2}}\right) D_{\chi^{2}}\left(v \| \pi_{\mathrm{gm}}\right)$,
where $\eta_{\min }=\min _{v \in V} \eta_{v}$ and $m=|E|$.

By Proposition 6.7, we only need to bound the spectral gap of the Glauber dynamics. The rest of this section endeavours to show

$$
\begin{equation*}
\mathfrak{b a p}\left(P_{\mathrm{GlauberGM}}\right) \geq \frac{\eta_{\min }^{4}}{m^{2}} \min \left\{p_{\min }, 1-2 p_{\max }\right\} \tag{6.20}
\end{equation*}
$$

This will be proved using the canonical path method adapted from [JS93].

### 6.3.1 Construction of the canonical path

Below is the main lemma of this subsection.

Lemma 6.15. For any grand model on a graph $G=(V, E)$ with parameters $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{v}\right)_{v \in V}$, if $0<\eta_{v}<1$ for all $v \in V$, then there exists a set of canonical paths $\Gamma=\left\{\gamma_{X Y}: X, Y \in \Omega\right\}$ for the Glauber dynamics $P_{\mathrm{gm}}$ such that

1. $w_{\mathrm{gm}}(X, Y)=\pi_{\mathrm{gm}}(X) \pi_{\mathrm{gm}}(Y)$;
2. $\left|\gamma_{X Y}\right| \leq m$;
3. for any transition $\left(Z, Z^{\prime}\right)$ with $\left|\left\{e: Z(e) \neq Z^{\prime}(e)\right\}\right|=1$, where the only edge $e$ of discrepancy is assigned 1 in either $Z$ or $Z^{\prime}$, it holds that

$$
\begin{equation*}
\sum_{\gamma \in \Gamma:\left(Z, Z^{\prime}\right) \in \gamma} w_{\mathrm{gm}}(\gamma) \leq \eta_{\min }^{-4} \min \left\{\pi_{\mathrm{gm}}(Z), \pi_{\mathrm{gm}}\left(Z^{\prime}\right)\right\} \tag{6.21}
\end{equation*}
$$

where $\eta_{\text {min }}:=\min _{v} \eta_{v}$;
4. for any transition $\left(Z, Z^{\prime}\right)$ with $\left|\left\{e: Z(e) \neq Z^{\prime}(e)\right\}\right|=1$, where the only edge $e$ of discrepancy is assigned 1 in neither $Z$ nor $Z^{\prime}$, it holds that

$$
\begin{equation*}
\sum_{\gamma \in \Gamma:\left(Z, Z^{\prime}\right) \in \gamma} w_{\mathrm{gm}}(\gamma) \leq \min \left\{\pi_{\mathrm{gm}}(Z), \pi_{\mathrm{gm}}\left(Z^{\prime}\right)\right\} \tag{6.22}
\end{equation*}
$$

Proof. We begin the proof with the construction of the paths. Suppose all vertices and edges are indexed by distinct integers, and there is a fixed ordering $<$ for all paths and cycles of the graph $G$. For any pair of assignments $X, Y$ in the grand model, the canonical path $\gamma_{X Y}$ contains two stages, moving from $X$ to $W$ and $W$ to $Y$ respectively.

Stage 1. (1-edge mending.) Midst this stage we mend the edges assigned 1 in either $X$ or $Y$ but not the other. Denote the set of such edges $D:=X^{-1}(1) \oplus Y^{-1}(1)$. The resulting configuration $W$ has the property that (1) for any edge $e \in D$, it holds that $W(e)=Y(e)$, and (2) for any other edge $e \notin D$, it holds that $W(e)=X(e)$.

Let $2 k$ be the number of the odd-degree vertices in $D$. Then, $D$ can be decomposed into an edge-disjoint union of exactly $k$ paths $P_{1}, \cdots, P_{k}$ and cycles $C_{1}, \cdots, C_{k^{\prime}}$. We pick the unique one such that $P_{1}, \cdots, P_{k}, C_{1}, \cdots, C_{k^{\prime}}$ is the first one in the lexicographic order induced by $<$.

To move from $X$ to $W$, we process each of the paths and cycles one by one. For each of them, we first choose the vertex and edge to start with. When winding (handling) a path, the starting vertex is one of the two open vertices of the path that has a smaller index; when winding a cycle, the starting vertex is the one with the smallest index, and the next vertex (which together with the starting one defines a starting edge) is one of the two neighbours of the starting vertex of the cycle that has a smaller index than the other one. After deciding the starting vertex and edge, we just move along the path/cycle. For each of the edge, we set the assignment to it as that in $Y$. Obviously this gives $W$ satisfying the properties aforementioned because every edge in $D$ is mended while the rest are left untouched.

Stage 2. (0, 2-edge mending.) None of the conflicting edges between $W$ and $Y$ can be assigned 1 in either of them. In this stage, we simply change all remaining disagreeing edges from the value in $W$ to the value in $Y$ one by one according to the order of their indices.

We then show that the set of canonical paths $\Gamma$ constructed above fulfills Lemma 6.15. Assign weight $w_{\mathrm{gm}}(\gamma)=\pi_{\mathrm{gm}}(X) \pi_{\mathrm{gm}}(Y)$ to the path $\gamma_{X Y}$. The length (number of transitions) of each path $\gamma_{X Y}$ is at most $m$, because each edge is mended at most once.

We first prove (6.21). Let $\left(Z, Z^{\prime}\right)$ be a transition with $\left|\left\{e: Z(e) \neq Z^{\prime}(e)\right\}\right|=1$, where the only edge $e$ of discrepancy is assigned 1 in either $Z$ or $Z^{\prime}$. Note that $\left(Z, Z^{\prime}\right)$ will only be used by any path in its first stage described above. Define a mapping $\varphi_{Z, Z^{\prime}}: \Omega \times \Omega \rightarrow \Omega$ over any pair of configurations $X, Y$ whose corresponding path $\gamma_{X Y}$ uses the transition $\left(Z, Z^{\prime}\right)$ by

$$
\begin{equation*}
\varphi_{Z, Z^{\prime}}(X, Y)=U \quad \text { where } \quad U(e)=X(e)+Y(e)-Z(e), \forall e \in E(G) . \tag{6.23}
\end{equation*}
$$

We claim that $\varphi_{Z, Z^{\prime}}$ is an injection. Given $U$ and $Z$, we can recover $X(e)+Y(e)$ for any edge $e$. First we can find $D$, the set of conflicting 1-edge in Stage 1, as it is simply $\{e: X(e)+Y(e)=1$ or 3$\}$. This gives rise to the unique edge-disjoint decomposition $P_{1}, \cdots, P_{k}, C_{1}, \cdots, C_{k^{\prime}}$. By looking at $Z$ and $Z^{\prime}$, we know the edge that is currently being wound, and, together with the edge-disjoint decomposition, the stage of the whole winding process. Therefore, we can continue the winding from $Z^{\prime}$ with these information, and when finished, $W$ (defined in the process Stage 1) is obtained. To
further recover $Y$, note that $e$ gets mended in Stage 2 if any only if $U(e)+Z(e)=2$ and $Z(e) \neq 1$. This follows from the fact that $Z(e)$ (in the first stage) is in line with $X(e)$ so long as $Z(e) \neq 1$. Therefore, we can decide all such edges and mend the assignment to obtain $Y$. To get $X$, we just reverse the operations backwards from $Z$.

Given this injection, we compute $\sum_{\gamma \in \Gamma:\left(Z, Z^{\prime}\right) \in \gamma} w_{\mathrm{gm}}(\gamma)$. The goal here is to bound the following ratio

$$
\begin{equation*}
\frac{\pi_{\mathrm{gm}}(X) \pi_{\mathrm{gm}}(Y)}{\pi_{\mathrm{gm}}(U) \pi_{\mathrm{gm}}(Z)}, \quad \text { or equivalently, } \quad \frac{\mathrm{wt}_{\mathrm{gm}}(X) \mathrm{wt}_{\mathrm{gm}}(Y)}{\mathrm{wt}_{\mathrm{gm}}(U) \mathrm{wt}_{\mathrm{gm}}(Z)} . \tag{6.24}
\end{equation*}
$$

Recall that this ratio may contain two kinds of factors, emerging from both the vertices and edges. For the factor from edges, the construction of $U$ ensures that (1) if $X(e)+Y(e)=U(e)+Z(e) \in\{0,1,3,4\}$, or $X(e)+Y(e)=2$ and $X(e) \neq 1$, then it must holds that either $X(e)=U(e)$ and $Y(e)=Z(e)$, or $X(e)=Z(e)$ and $Y(e)=U(e) ;(2)$ if $X(e)=Y(e)=1$, then $e$ never gets mended throughout the canonical path, and hence $Z(e)=U(e)=1$. In either case, all the terms rising from the edges in the numerator and denominator cancel. The terms rising from the vertices come from those in $O(X), O(Y), O(U), O(Z)$. It is not hard to see that the ones that do not get cancelled only arise from the current cycle or path that is being processed, and more specifically, the vertex incident to the two edges wound before and after $Z$, which contributes twice, and the starting vertex of the current cycle, which contributes twice as well. Therefore,

$$
\begin{equation*}
\frac{\pi_{\mathrm{gm}}(X) \pi_{\mathrm{gm}}(Y)}{\pi_{\mathrm{gm}}(U) \pi_{\mathrm{gm}}(Z)} \leq \eta_{\mathrm{min}}^{-4}, \tag{6.25}
\end{equation*}
$$

as $0<\eta_{v}<1$ for all $v$.
Then, (6.21) follows from (6.25) that

$$
\begin{array}{rlr}
\sum_{\gamma \in \Gamma:\left(Z, Z^{\prime}\right) \in \gamma} w_{\mathrm{gm}}(\gamma) & =\sum_{X, Y:\left(Z, Z^{\prime}\right) \in \gamma_{X Y}} \pi_{\mathrm{gm}}(X) \pi_{\mathrm{gm}}(Y) \\
& \leq \eta_{\min }^{-4} \sum_{X, Y:\left(Z, Z^{\prime}\right) \in \gamma_{X Y}} \pi_{\mathrm{gm}}(Z) \pi_{\mathrm{gm}}\left(\varphi_{Z, Z^{\prime}}(X, Y)\right) \quad \text { (By definition) } \quad \text { (By (6.25)) } \\
& \leq \eta_{\min }^{-4} \pi_{\mathrm{gm}}(Z) .
\end{array}
$$

We construct the other mapping $\varphi_{Z, Z^{\prime}}^{\prime}(X, Y)$ by taking $\varphi_{Z, Z^{\prime}}(X, Y)(e)=X(e)+$ $Y(e)-Z^{\prime}(e)$. The same argument shows that $\sum_{\gamma \in \Gamma:\left(Z, Z^{\prime}\right) \in \gamma} w_{\mathrm{gm}}(\gamma) \leq \eta_{\min }^{-4} \pi_{\mathrm{gm}}\left(Z^{\prime}\right)$.

To prove (6.22), we look at the transition step $\left(Z, Z^{\prime}\right)$ with $\left|\left\{e: Z(e) \neq Z^{\prime}(e)\right\}\right|=1$ where the only edge $e$ of discrepancy is assigned 1 in neither $Z$ nor $Z^{\prime}$. We use the same mapping $\varphi_{Z, Z^{\prime}}(X, Y)$ as above, and claim it is still injective in this case. Recall that $e$ gets mended in Stage 2 if and only if $U(e)+Z(e)=2$ and $Z(e) \neq 1$, and
we can again determine the edges to be mended in Stage 2. Moreover, by looking at the difference of $Z$ and $Z^{\prime}$, we know the index of the edge being mended, and therefore we can continue this process manually according to the instruction of Stage 2, knowing which edges to mend, to obtain $Y$. To get $X$, we first go backwards from $Z$ to the beginning of Stage 2 to obtain $W$, and revert the whole Stage 1 using the same argument aforementioned.

To show (6.22), note that the edge factors in the ratio of (6.24) again cancel, and because no edge with assignment 1 is involved, the vertex factors cancel as well. Hence the ratio is exactly 1 , and (6.22) follows according to the same calculation.

### 6.3.2 Total congestion and rapid mixing

We next bound the total congestion for $\Gamma_{\mathrm{gm}}$. For each transition $\left(Z, Z^{\prime}\right)$ such that $\left|\left\{e: Z(e) \neq Z^{\prime}(e)\right\}\right|=1$, where the only edge of discrepancy is assigned 1 in either $Z$ or $Z^{\prime}$, we have

$$
\frac{L}{\pi_{\mathrm{gm}}(Z) P_{\mathrm{gm}}\left(Z, Z^{\prime}\right)} \sum_{\substack{\gamma \in \Gamma: \\\left(Z, Z^{\prime}\right) \in \gamma}} w_{\mathrm{gm}}(\gamma) \leq \frac{m \eta_{\min }^{-4} \min \left\{\pi_{\mathrm{gm}}(Z), \pi_{\mathrm{gm}}\left(Z^{\prime}\right)\right\}}{\pi_{\mathrm{gm}}(Z) P_{\mathrm{gm}}\left(Z, Z^{\prime}\right)}=:(\propto)
$$

by Lemma 6.15. To continue the calculation, there are several cases $\left(Z(e), Z^{\prime}(e)\right)=$ $(0,1),(2,1),(1,0),(1,2)$. Below we only prove the case $\left(Z(e), Z^{\prime}(e)\right)=(0,1)$. The rest cases can be argued the same way and yield the same bound. Let $e=(u, v)$. There are some more subcases, depending on if $u$ or $v$ is in $O(Z)$.

- $u, v \notin O(Z)$. In this case, setting the edge to 1 leads to extra factors from both vertices in $Z^{\prime}$. Cancelling all the edges and vertices not involved, we obtain
$(\bullet)=\frac{m^{2} \eta_{\min }^{-4} \min \left\{1-2 p_{e}, p_{e} \eta_{u} \eta_{v}\right\}}{\left(1-2 p_{e}\right) \frac{p_{e}{ }^{2}\left(\eta_{v}\right.}{\left(1-2 p_{e}\right)+\left(p_{e} \eta_{u} \eta_{v}\right)+p_{e}}} \leq \frac{m^{2} \eta_{\min }^{-4} \min \left\{1-2 p_{e}, p_{e} \eta_{u} \eta_{v}\right\}}{\left(1-2 p_{e}\right)\left(p_{e} \eta_{u} \eta_{v}\right)} \leq \frac{m^{2} \eta_{\min }^{-4}}{1-2 p_{e}}$ where we use the fact that $\eta_{u}, \eta_{v} \leq 1$.
- $u, v \in O(Z)$. In this case, setting the edge to 1 removes the factors from both vertices in $Z^{\prime}$. Cancelling all the edges and vertices not involved, we obtain

$$
\text { (৫) }=\frac{m^{2} \eta_{\min }^{-4} \min \left\{\left(1-2 p_{e}\right) \eta_{u} \eta_{v}, p_{e}\right\}}{\left(1-2 p_{e}\right) \eta_{u} \eta_{v} \frac{p_{e}}{\left(1-2 p_{e}\right) \eta_{u} \eta_{v}+p_{e}+p_{e} \eta_{u} \eta_{v}}} \leq \frac{m^{2} \eta_{\min }^{-4} \min \left\{\left(1-2 p_{e}\right) \eta_{u} \eta_{v}, p_{e}\right\}}{\left(1-2 p_{e}\right) \eta_{u} \eta_{v} p_{e}} \leq \frac{m^{2} \eta_{\min }^{-4}}{p_{e}}
$$

where we use the fact that $\eta_{u}, \eta_{v} \leq 1$ again.

- WLOG suppose $u \in O(Z), v \notin O(Z)$. In this case, setting the edge to 1 causes the vertex factor to switch. Cancelling all the edges and vertices not involved, we obtain

$$
(\star)=\frac{m^{2} \eta_{\min }^{-4} \min \left\{\left(1-2 p_{e}\right) \eta_{u}, p_{e} \eta_{v}\right\}}{\left(1-2 p_{e}\right) \eta_{u} \frac{p_{e} \eta_{v}}{\left(1-2 p_{e}\right) \eta_{u}+\left(p_{e} \eta_{v}\right)+\left(p_{e} \eta_{u}\right)}} .
$$

If $\eta_{u}<\eta_{v}$, then above becomes

$$
\frac{m^{2} \eta_{\min }^{-4} \min \left\{\left(1-2 p_{e}\right) \frac{\eta_{u}}{\eta_{v}}, p_{e}\right\}}{\left(1-2 p_{e}\right) \frac{\eta_{u}}{\eta_{v}} \frac{p_{e}}{\left(1-2 p_{e} \frac{\eta_{u}}{\eta_{v}}+p_{e}+p_{e} \frac{\eta_{u}}{\eta_{v}}\right.}} \leq \frac{m^{2} \eta_{\min }^{-4}}{p_{e}} .
$$

Otherwise, it can be written as

$$
\frac{m^{2} \eta_{\min }^{-4} \min \left\{\left(1-2 p_{e}\right), p_{e} \frac{\eta_{v}}{\eta_{u}}\right\}}{\left(1-2 p_{e}\right) \frac{p_{e} \frac{\eta_{v}}{\eta_{u}}}{\left(1-2 p_{e}\right)+p_{e} \frac{\eta_{v}+p_{e}}{\eta_{u}}}} \leq \frac{m^{2} \eta_{\min }^{-4}}{1-2 p_{e}} .
$$

For each transition $\left(Z, Z^{\prime}\right)$ such that $\left|\left\{e: Z(e) \neq Z^{\prime}(e)\right\}\right|=1$, where the only edge of discrepancy is assigned 1 in none of $Z$ or $Z^{\prime}$, the calculation is similar as above but simpler. WLOG assume $Z(e)=0$ and $Z^{\prime}(e)=2$.

$$
\begin{align*}
& \frac{L}{\pi_{\mathrm{gm}}(Z) P_{\mathrm{gm}}\left(Z, Z^{\prime}\right)} \sum_{\substack{\gamma \in \Gamma: \\
\left(Z, Z^{\prime}\right) \in \gamma}} w_{\mathrm{gm}}(\gamma) \leq \frac{m \min \left\{\pi_{\mathrm{gm}}(Z), \pi_{\mathrm{gm}}\left(Z^{\prime}\right)\right\}}{\pi_{\mathrm{gm}}(Z) P_{\mathrm{gm}}\left(Z, Z^{\prime}\right)} \quad \text { (Lemma 6.15) }  \tag{Lemma6.15}\\
\leq & \frac{m^{2} \min \left\{1-2 p_{e}, p_{e}\right\}}{\left(1-2 p_{e}\right) \frac{p_{e}}{1-2 p_{e}+p_{e}+p_{e} \frac{1}{\eta_{u} \eta_{\nu}}}} \leq \min \left\{\frac{1}{p_{e}}, \frac{1}{1-2 p_{e}}\right\} m^{2} \eta_{\min }^{-2} . \quad \text { (Worst case of } \eta \text { terms) }
\end{align*}
$$

There is no canonical path using the self loop $(Z, Z)$, so the congestion is zero. In all, the congestion is bounded by $m^{2} \eta_{\text {min }}^{-4} \max \left\{\frac{1}{p_{\text {min }}}, \frac{1}{1-2 p_{\text {max }}}\right\}$, from which (6.20) follows.

### 6.4 Entropy decay of Glauber dynamics on the grand model

In Section 6.3, we analysed the variance decay of Glauber dynamics on the grand model. We now continue to analyse its relative entropy decay. Let $G=(V, E)$ be a graph, and $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{v}\right)_{v \in V}$ be the parameters, where $0<p_{e}<1 / 2$ for any $e \in E$ and $\eta_{v}>0$ for any $v \in V$. Let $\pi_{\mathrm{gm}}$ denote the distribution specified by the grand model with parameters $\mathbf{p}$ and $\eta$. Let $\Omega\left(\pi_{\mathrm{gm}}\right)$ denote the support of $\pi_{\mathrm{gm}}$. We use $P_{\text {GlauberGM }}$ to denote Glauber dynamics on $\pi_{\mathrm{gm}}$.

Lemma 6.16. If $0<\eta_{v}<1$ for all $v \in V$, then for any distribution $v$ with support $\Omega(v) \subseteq \Omega\left(\pi_{\mathrm{gm}}\right)$, Glauber dynamics $P_{\text {GlauberGM }}$ satisfies

$$
D_{\mathrm{KL}}\left(v P_{\mathrm{GlauberGM}}^{\downarrow} \| \pi_{\mathrm{gm}} P_{\mathrm{GlauberGM}}^{\downarrow}\right) \leq\left(1-\frac{1}{C n}\right) D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{gm}}\right),
$$

where $C=C\left(\Delta, \eta_{\min }, p_{\min }, p_{\max }\right), \eta_{\min }=\min _{v \in V} \eta_{v}, p_{\min }=\min _{e \in E} p_{e}, p_{\max }=\max _{e \in E} p_{e}$, $\Delta$ is the maximum degree of $G$ and $n=|V|$.

Remark 6.17. For interested readers, the constant $C$ in the lemma above can be taken as

$$
C=\Delta\left(\frac{2 \Delta}{\eta_{\min }^{2} \min \left\{1-2 p_{\max }, p_{\min }\right\}}\right)^{2+\frac{16 \Delta^{2}}{\eta_{\min }^{4} \min \left\{1-2 p_{\max }, p_{\min }\right\}}}
$$

Lemma 6.16 is proved by Theorem 6.8. To apply Theorem 6.8, we need to verify (1) $\pi_{\mathrm{gm}}$ is a Gibbs distribution with maximum degree $D=2 \Delta-1$; (2) $\pi_{\mathrm{gm}}$ is $\ell_{\infty}$-spectrally independent; (3) $\pi_{\mathrm{gm}}$ is marginally bounded. The rest of this section is dedicated to the proof of Lemma 6.16.

Lemma 6.18. $\pi_{\mathrm{gm}}$ is $\ell_{\infty}$-spectrally independent with parameter $\zeta=O\left(\Delta^{2} / \eta_{\text {min }}^{2}\right)$.
We need the following result in [CLV21b] to prove Lemma 6.18. We view the subgraph world as a distribution over $\{0,1\}^{E}$, where each $Y \in\{0,1\}^{E}$ corresponds to $S=\left\{e \in E \mid Y_{e}=1\right\}$.

Lemma 6.19 ([CLV21b]). Let $G=(V, E)$ be a graph with the maximum degree $\Delta \geq 3$. Let $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{v}\right)_{v \in V}$, where $0 \leq p_{e}<1 / 2$ and $0<\eta_{v} \leq 1$. The distribution $\pi_{\text {sg }}$ specified by the subgraph-world model with parameters $(\mathbf{p}, \eta)$ is $\ell_{\infty}$-spectrally independent with parameter $\zeta=O\left(\Delta^{2} / \eta_{\text {min }}^{2}\right)$.
Remark 6.20. In [CLV21b], the authors only formalise the proof for the uniform case (i.e., all $\eta_{v}$ 's take the same value) while stating that the argument works for nonuniform case without a proof. This in fact holds true by going through the proof and taking the worst region of stability. The final spectral independence parameter is

$$
\zeta=8\left(\frac{\left(\frac{1+\eta_{\min }}{1-\eta_{\min }}\right)^{1 / \Delta}+1}{\left(\frac{1+\eta_{\min }}{1-\eta_{\min }}\right)^{1 / \Delta}-1}\right)^{2} \sim 8 \Delta^{2} / \eta_{\min }^{2}
$$

Note that the $\lambda$ in their paper is actually $p /(1-p)$ in our formulation of the subgraphworld model (under the uniform edge parameter setting). Also note that we are only considering the region $0<p<1 / 2$, so the $\lambda$ in their paper is bounded from above by 1.

Proof of Lemma 6.18. Fix a pinning $\sigma \in\{0,1,2\}^{\Lambda}$ for some $\Lambda \subseteq E$. According to the definition of the grand model, to draw $X \sim \pi_{\mathrm{gm}}$, we first sample $Y \sim \pi_{\mathrm{sg}}$ (where $Y \in$ $\{0,1\}^{E}$ as we view $\pi_{\text {sg }}$ as a distribution over $\{0,1\}^{E}$ ), then flip independent coins for each $e \in E$ with $Y_{e}=0$. Define the pinning $\tau \in\{0,1\}^{\Lambda}$ by $\tau_{e}=1$ if $\sigma_{e}=1$ and $\tau_{e}=0$ if $\sigma_{e}=0$ or $\sigma_{e}=2$. Consider the influence
$\Psi_{\pi_{\mathrm{gm}}}^{\sigma}(e, f)=\max \left\{d_{\mathrm{TV}}\left(\pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 0}, \pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 1}\right), d_{\mathrm{TV}}\left(\pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 0}, \pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 2}\right), d_{\mathrm{TV}}\left(\pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 1}, \pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 2}\right)\right\}$,
where $e, f \in E \backslash \Lambda$ and $e \neq f$. Since each coin flipping is independent with the random sample from $\pi_{\mathrm{gm}}$, we can couple two distributions $\pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 0}$ and $\pi_{\mathrm{gm}, f}^{\sigma \wedge \leftarrow 1}$ as follows:

- sample $Y_{f}, Y_{f}^{\prime}$ from the optimal coupling between $\pi_{\mathrm{sg}, f}^{\tau \lambda e \leftarrow 0}$ and $\pi_{\mathrm{sg}, f}^{\tau \lambda e \leftarrow 1}$;
- flip a coin $C$ independently with probability of HEADS being $\frac{p_{f}}{1-p_{f}}$;
- if $Y_{f}=1$, let $X_{f}=1$; otherwise, if the outcome of $C$ is HEADS, let $X_{f}=2$, if the outcome of $C$ is not HEADS, let $X_{f}=0$;
- if $Y_{f}^{\prime}=1$, let $X_{f}^{\prime}=1$; otherwise, if the outcome of $C$ is HEADS, let $X_{f}^{\prime}=2$, if the outcome of $C$ is not HEADS, let $X_{f}^{\prime}=0$;

It is straightforward to verify that ( $X_{f}, X_{f}^{\prime}$ ) is sampled from a coupling between $\pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 0}$ and $\pi_{\mathrm{gm}, f}^{\sigma \wedge \leftarrow 1}$. By the coupling lemma Lemma 2.3 and as $Y_{f}$ and $Y_{f}^{\prime}$ are optimally coupled, we have

$$
d_{\mathrm{TV}}\left(\pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 0}, \pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 1}\right) \leq \operatorname{Pr}\left[X_{f} \neq X_{f}^{\prime}\right]=\operatorname{Pr}\left[Y_{f} \neq Y_{f}^{\prime}\right]=d_{\mathrm{TV}}\left(\pi_{\mathrm{sg}, f}^{\tau \wedge e \leftarrow 0}, \pi_{\mathrm{sg}, f}^{\tau \wedge e \leftarrow 1}\right) .
$$

Similarly, we have

$$
d_{\mathrm{TV}}\left(\pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 0}, \pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 2}\right)=0 \quad \text { and } \quad d_{\mathrm{TV}}\left(\pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 1}, \pi_{\mathrm{gm}, f}^{\sigma \wedge e \leftarrow 2}\right) \leq d_{\mathrm{TV}}\left(\pi_{\mathrm{sg}, f}^{\tau \wedge e \leftarrow 0}, \pi_{\mathrm{sg}, f}^{\tau \wedge e \leftarrow 1}\right)
$$

Hence, by Lemma 6.19,

$$
\left\|\Psi_{\mathrm{gm}}^{\sigma}\right\|_{\infty} \leq\left\|\Psi_{\mathrm{sg}}^{\tau}\right\|_{\infty} \leq \zeta
$$

Lemma 6.21. $\pi_{\mathrm{gm}}$ is $b$-marginally bounded, where $b=\eta_{\min }^{2} \min \left\{1-2 p_{\max }, p_{\min }\right\}$.
Proof. Consider the marginal distribution of an edge $e=(u, v)$. Let $e_{1}, \ldots, e_{k}$ be the edges adjacent to either $u$ or $v$ (but not both). Suppose we have an arbitrary pinning $X$ on $\Lambda \subset E$ and $e \notin \Lambda$. Let $Y$ be an arbitrary pinning on $\Lambda \cup\left\{e_{1}, \ldots, e_{k}\right\}$ that is consistent with $X$. The true marginal of $e$ under $X$ is a linear combination of marginals
conditioned on all possibilities of $Y$ (namely, we first sample $Y$ and then sample $e$ conditioned on $Y$ ). Thus, to establish a lower bound, it suffices to establish a lower bound under any $Y$. Given $Y$, the marginal of $e$ depends only on $p_{e}$ and whether $u$ or $v$ is in $O(Y)$. These cases are verified as follows.

- $u, v \notin O\left(Y^{e \rightarrow 0}\right)$, where $Y^{e \rightarrow 0}$ is the configuration of $Y$ with $e$ further pinned to 0 . In this case the marginal is at least

$$
\frac{\min \left\{1-2 p_{e}, p_{e} \eta_{u} \eta_{v}, p_{e}\right\}}{1-2 p_{e}+p_{e} \eta_{u} \eta_{v}+p_{e}} \geq \min \left\{1-2 p_{e}, p_{e} \eta_{u} \eta_{v}\right\}
$$

Note that the denominator is no greater than 1 because $\eta_{u}, \eta_{v} \leq 1$.

- $u, v \in O\left(Y^{e \rightarrow 0}\right)$. Then the marginal is at least

$$
\frac{\min \left\{1-2 p_{e} \eta_{u} \eta_{v}, p_{e}, p_{e} \eta_{u} \eta_{v}\right\}}{\left(1-2 p_{e}\right) \eta_{u} \eta_{v}+p_{e}+p_{e} \eta_{u} \eta_{v}} \geq \min \left\{\left(1-2 p_{e}\right) \eta_{u} \eta_{v}, p_{e}\right\}
$$

- In the remaining cases, assume w.l.o.g. $u \in O\left(Y^{e \rightarrow 0}\right)$ while $v \notin O\left(Y^{e \rightarrow 0}\right)$. Then the marginal is at least

$$
\frac{\min \left\{\left(1-2 p_{e}\right) \eta_{u}, p_{e} \eta_{v}, p_{e} \eta_{u}\right\}}{\left(1-2 p_{e}\right) \eta_{u}+p_{e} \eta_{v}+p_{e} \eta_{u}}=\left\{\begin{array} { l l } 
{ \frac { \operatorname { m i n } \{ ( 1 - 2 p _ { e } e } { } \frac { \eta _ { u } } { \eta _ { v } } , p _ { e } \frac { \eta _ { u } } { \eta _ { u } \} } } \\
{ ( 1 - 2 p _ { e } ) \frac { \eta _ { u } } { \eta _ { v } + p _ { e } + p _ { e } } \overline { \eta } _ { v } }
\end{array} \operatorname { m i n } \left\{\left(1-2 p_{e}\right) \frac{\eta_{u}}{\eta_{v}}, p_{e} \frac{\eta_{u}}{\left.\frac{\eta_{v}}{\eta_{v}}\right\},} \quad \text { if } \eta_{u}<\eta_{v} ;\right.\right.
$$

In all cases, the value

$$
b=\eta_{\min }^{2} \min \left\{1-2 p_{\max }, p_{\min }\right\}
$$

suffices as a marginal lower bound.

Proof of Lemma 6.16. Combine Theorem 6.8, Observation 6.10, Lemma 6.18, Lemma 6.21 and $m \leq n \Delta$.

### 6.5 Rapid mixing of Glauber dynamics on the random cluster model

Let $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\eta=\left(\eta_{v}\right)_{v \in V}$, where $0<p_{e}<1 / 2$ and $0<\eta_{v}<1$. Let $\pi_{\text {wrc }}$ denote the distribution specified by the random cluster model with parameters $2 \mathbf{p}$ and $\lambda$, where $\lambda_{v}=\frac{1-\eta_{v}}{1+\eta_{v}}$. Let $\Omega\left(\pi_{\text {wrc }}\right)$ denote the support of $\pi_{\text {wrc }}$. We use $P_{\text {GlauberRC }}$ to denote Glauber dynamics on $\pi_{\text {wrc }}$.

Lemma 6.22. Let $\pi_{\text {wrc }}$ be the distribution specified by weighted random cluster model with parameters $(2 \mathbf{p}, \lambda)$. The Glauber dynamics $P_{\text {GlauberRC }}$ satisfies that for any distribution $v$ with support $\Omega(v) \subseteq \Omega\left(\pi_{\text {wrc }}\right)$,

- $D_{\chi^{2}}\left(v P_{\text {GlauberRC }}^{\downarrow} \| \pi_{\text {wrc }} P_{\text {GlauberRC }}^{\downarrow}\right) \leq\left(1-\frac{\alpha}{m^{2}}\right) D_{\chi^{2}}\left(v \| \pi_{\text {wrc }}\right)$,
- $D_{\mathrm{KL}}\left(v P_{\text {GlauberRC }}^{\downarrow} \| \pi_{\mathrm{wrC}} P_{\text {GlauberRC }}^{\downarrow}\right) \leq\left(1-\frac{1}{C n}\right) D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right)$,
where

$$
\begin{aligned}
& \alpha=\left(\frac{1-\lambda_{\max }}{1+\lambda_{\max }}\right)^{4} \min \left\{p_{\min }, 1-2 p_{\max }\right\}, \\
& C=\Delta\left(\frac{8 \Delta}{\left(1-\lambda_{\max }\right)^{2} \min \left\{1-2 p_{\max }, p_{\min }\right\}}\right)^{2+\frac{256 \Delta^{2}}{\left(1-\lambda_{\max }\right)^{4} \min \left\{1-2 p_{\max }, p_{\min }\right\}}},
\end{aligned}
$$

$\lambda_{\text {max }}=\max _{v \in V} \lambda_{v}, \lambda_{\text {min }}=\min _{v \in V} \lambda_{v}, p_{\max }=\max _{e \in E} p_{e}, p_{\text {min }}=\min _{e \in E} p_{e}, \Delta$ is the maximum degree of $G, n=|V|$ and $m=|E|$.

Lemma 6.22 projects the decay results (Lemma 6.14 and Lemma 6.16) from the grand model to the random cluster model. Lemma 6.22 is proved by a comparison lemma in Section 6.5.1 that works for general projections and $f$-divergences.

Lemma 6.22 provides an entropy decay rate and a $\chi^{2}$-divergence decay rate. When $\lambda_{\text {max }}$ is bounded away from 1 , the entropy decay rate is better. On the other hand, the $\chi^{2}$-divergence decay rate has a better dependency on $1-\lambda_{\text {max }}$. In particular, when $\lambda_{\text {max }}=1$, namely when some vertices do not have external fields, neither statement provides any decay. In such cases, we can perturb $\lambda$ by a factor of $1 / n$. This incurs a cost of a polynomial factor in $n$ for $\alpha$ and an exponentially large factor for $C$. Thus, we need to apply the $\chi^{2}$-divergence decay rate in Lemma 6.22 after perturbation in the $\lambda_{\max }=1$ case. Specifically, in Section 6.5.2 we showed the following.

Lemma 6.23. Let $\pi_{\mathrm{wrc}}$ be the distribution specified by the weighted random cluster model with parameters $(2 \mathbf{p}, \lambda)$. The Glauber dynamics $P_{\text {GlauberRC }}$ satisfies that for any distribution $v$ with support $\Omega(v) \subseteq \Omega\left(\pi_{\text {wrc }}\right)$,

$$
D_{\chi^{2}}\left(v P_{\text {GlauberRC }}^{\downarrow} \| \pi_{\text {wrc }} P_{\text {GlauberRC }}^{\downarrow}\right) \leq\left(1-\frac{\min \left\{p_{\min }, 1-2 p_{\max }\right\}}{10^{4} n^{4} m^{2}}\right) D_{\chi^{2}}\left(v \| \pi_{\mathrm{wrc}}\right) .
$$

We remark that both Lemma 6.22 and Lemma 6.23 consider the random cluster model specified by parameters ( $2 \mathbf{p}, \lambda$ ). Combining Lemma 6.22 and Lemma 6.23, we have the following mixing result for the Glauber dynamics on random cluster model.

Theorem 6.24. Let $G=(V, E)$ be a $n$-vertex and $m$-edge graph with maximum degree $\Delta$. Let $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$, where $0<p_{e}<1$ and $0<\lambda_{v} \leq 1$. Let $\pi_{\mathrm{wrc}}$ be the distribution specified by the random cluster model with parameters $(\mathbf{p}, \lambda)$. The mixing of Glauber dynamics $P_{\text {GlauberRC }}$ on $\pi_{\text {wrc }}$ satisfies

$$
T_{\operatorname{mix}}\left(P_{\mathrm{GlauberRC}}, \epsilon\right) \leq C_{1}\left(p_{\min }, p_{\max }\right) \cdot \min \left\{n^{4},\left(\frac{1}{1-\lambda_{\max }}\right)^{4}\right\} \cdot m^{2} \cdot\left(\log \frac{1}{\epsilon}+m\right)
$$

where $C_{1}\left(p_{\min }, p_{\max }\right)=O\left(\frac{1}{\min \left\{p_{\min }, 1-p_{\max }\right\}} \log \frac{1}{\min \left\{p_{\min }, 1-p_{\max }\right\}}\right)$.
Furthermore, if there exists $\delta>0$ such that $\lambda_{v} \leq 1-\delta$ for all $v \in V$, then the mixing time satisfies

$$
T_{\operatorname{mix}}\left(P_{\mathrm{GlauberRC}}, \epsilon\right) \leq C_{2}\left(\Delta, \delta, p_{\min }, p_{\max }\right) \cdot n\left(\log n+\log \frac{1}{\epsilon}\right)
$$

where $C_{2}\left(\Delta, \delta, p_{\min }, p_{\max }\right)=\left(\frac{\Delta}{\delta^{2} \min \left\{p_{\min }, 1-p_{\max }\right\}}\right)^{O\left(\frac{\Delta^{2}}{\delta^{4} \min \left\{p_{\min }, 1-p_{\max }\right\}}\right)}$.
Proof. Let $\pi_{\text {wre, min }}=\min _{S \subseteq E} \pi_{\text {wrc }}(S)$ denote the minimum probability in $\pi_{\text {wrc }}$. It is straightforward to verify that $\pi_{\text {wrc, } \min } \geq \min \left\{p_{\min }, 1-p_{\max }\right\}^{m} / 2^{m+n}$. By the data processing inequality,
$D_{f}\left(v P_{\text {GlauberRC }} \| \pi_{\text {wrc }}\right)=D_{f}\left(v P_{\text {GlauberRC }} \| \pi_{\text {wrc }} P_{\text {GlauberRC }}\right) \leq D_{f}\left(v P_{\text {GlauberRC }}^{\downarrow} \| \pi_{\text {wrc }} P_{\text {GlauberRC }}^{\downarrow}\right)$.
By Lemma 6.22 and Lemma 6.23, we know that after each transition step of Glauber dynamics, the $\chi^{2}$-divergence and KL-divergence between the current distribution of the stationary distribution decays by factors specified earlier. The $\chi^{2}$-divergence between the initial distribution and the stationary distribution is at most $\frac{1}{\pi_{\text {wr, min }}}$, and the KL-divergence is at most $\log \frac{1}{\pi_{\text {wr, min }}}$. By Lemma 6.22, Lemma 6.23, and (2.3),

$$
\begin{aligned}
T_{\text {mix }}\left(P_{\text {GlauberRC }}, \epsilon\right) & \leq \frac{10^{4}}{\min \left\{p_{\min } / 2,1-p_{\max }\right\}} \cdot \min \left\{n^{4},\left(\frac{1+\lambda_{\max }}{1-\lambda_{\max }}\right)^{4}\right\} \cdot m^{2}\left(\log \frac{1}{\epsilon^{2} \pi_{\mathrm{wrc}, \min }}\right) \\
& \leq C_{1}\left(p_{\min }, p_{\max }\right) \cdot \min \left\{n^{4},\left(\frac{1}{1-\lambda_{\max }}\right)^{4}\right\} \cdot m^{2} \cdot\left(\log \frac{1}{\epsilon}+m\right) .
\end{aligned}
$$

Note that $1<1+\lambda_{\text {max }} \leq 2$.
By Lemma 6.22, (2.4) and $m \leq \Delta n$, if for all $\lambda_{v} \leq 1-\delta$, then we have $1-\lambda_{\max } \geq \delta$ and

$$
\begin{aligned}
& T_{\text {mix }}\left(P_{\text {GlauberRC }}, \epsilon\right) \\
& \quad \leq \Delta\left(\frac{8 \Delta}{\delta^{2} \min \left\{1-p_{\max }, p_{\min } / 2\right\}}\right)^{2+\frac{2562^{2}}{\delta^{4} \min \left\{1-p_{\max }, p_{\min } / 2\right\}}} \cdot n\left(\log \log \frac{1}{\pi_{\mathrm{wrc}, \min }}+\log \frac{1}{2 \epsilon^{2}}\right) \\
& \quad \leq C_{2}\left(\Delta, \delta, p_{\min }, p_{\max }\right) \cdot n\left(\log n+\log \frac{1}{\epsilon}\right) .
\end{aligned}
$$

### 6.5.1 Comparing the decay rates of down walks

Here we consider a general projection from a larger state space to a smaller one. Let $Q$ and $R$ be two finite sets, and let $\Omega \subseteq Q^{V}$ be the state space. Consider a mapping $g: Q \rightarrow R$. (Note that here we can restrict $R$ to the range of $g$ without changing the rest of the argument. In other words, after the mapping the effective domain is never larger than $Q$, although we do not need to require $|Q| \geq|R|$ a priori.) Given any $\sigma \in \Omega$, we map $\sigma$ to $\tau=\left(\tau_{v}\right)_{v \in V}$, where $\tau_{v}=g\left(\sigma_{v}\right)$. We abuse the notation and denote $\tau=g(\sigma)$. Let $\Omega^{\prime}=\{g(\sigma) \mid \sigma \in \Omega\} \subseteq R^{V}$. Define the projection matrix $P: \Omega \times \Omega^{\prime} \rightarrow$ $\{0,1\}$ :

$$
\forall \sigma \in \Omega, \tau \in \Omega^{\prime}, \quad P(\sigma, \tau)=\mathbb{I}[\tau=g(\sigma)] .
$$

We remark that $P$ is a stochastic matrix.
Let $\pi$ be a distribution with support $\Omega$. Define the distribution $\mu=\pi P$ with support $\Omega^{\prime}$. Let $P_{\text {Glauber }, \pi}^{\downarrow}: \Omega \times \Omega_{\text {down }} \rightarrow \mathbb{R}_{\geq 0}$ denote the down walk of Glauber dynamics on $\pi$, where $\Omega_{\text {down }}=\left\{\sigma_{V \backslash\{v\}} \mid v \in V \wedge \sigma \in \Omega\right\}$. Given any configuration $\sigma \in \Omega$, $P_{\text {Glauber, } \pi}^{\downarrow}$ picks a variable $v \in V$ uniformly at random, and then transforms $\sigma$ to $\sigma_{V \backslash\{v\}}$ by dropping the value of $v$. Similarly, let $P_{\text {Glauber }, \mu}^{\downarrow}$ denote the down walk of Glauber dynamics on the distribution $\mu=\pi P$.

Lemma 6.25. Let $0<\delta<1$. Let $f: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ be a convex function with $f(1)=0$. If $P_{\text {Glauber, } \pi}^{\downarrow}$ satisfies that for any distribution $v$ with support $\Omega$,

$$
D_{f}\left(v P_{\text {Glauber }, \pi}^{\downarrow} \| \pi P_{\text {Glauber }, \pi}^{\downarrow}\right) \leq(1-\delta) D_{f}(v \| \pi),
$$

then $P_{\text {Glauber }, \mu}^{\downarrow}$ satisfies that for any distribution $\varphi$ with support $\Omega^{\prime}$,

$$
D_{f}\left(\varphi P_{\text {Glauber }, \mu}^{\downarrow} \| \mu P_{\text {Glauber }, \mu}^{\downarrow}\right) \leq(1-\delta) D_{f}(\varphi \| \mu) .
$$

Proof. Given any $\rho \in \Omega_{\text {down }}$, we can map $\rho$ to $\eta=g(\rho)$, where $\eta_{u}=g\left(\rho_{u}\right)$ for any variable $u$. Let $\Omega_{\text {down }}^{\prime}=\left\{g(\rho) \mid \rho \in \Omega_{\text {down }}\right\}$. Define the projection matrix $P^{\prime}: \Omega_{\text {down }} \times$ $\Omega_{\text {down }}^{\prime} \rightarrow\{0,1\}:$

$$
\forall \rho \in \Omega_{\mathrm{down}}, \eta \in \Omega_{\mathrm{down}}^{\prime}, \quad P^{\prime}(\rho, \eta)=\mathbb{I}[\eta=g(\rho)] .
$$

We remark that $P^{\prime}$ is a stochastic matrix. Since both $P$ and $P^{\prime}$ project the value of each variable independently, the following equation is straightforward to verify

$$
\begin{equation*}
P_{\text {Glauber }, \pi}^{\downarrow} \cdot P^{\prime}=P \cdot P_{\text {Glauber }, \mu}^{\downarrow} \tag{6.26}
\end{equation*}
$$

For any configuration $\tau \in \Omega^{\prime}$, define the distribution $\pi^{\tau}$ over $\Omega$ by

$$
\forall \sigma \in \Omega, \quad \pi^{\tau}(\sigma)=\frac{\mathbb{I}[g(\sigma)=\tau] \pi(\sigma)}{\mu(\tau)}
$$

For any $\sigma \in \Omega$, let $\tau=g(\sigma)$, it holds that $\pi(\sigma)=\mu(\tau) \pi^{\tau}(\sigma)$. Fix a distribution $\varphi$ with support $\Omega^{\prime}$. Define the distribution $v$ by

$$
\begin{equation*}
\forall \sigma \in \Omega, \quad v(\sigma)=\varphi(\tau) \pi^{\tau}(\sigma), \quad \text { where } \tau=g(\sigma) \tag{6.27}
\end{equation*}
$$

We have
$D_{f}(v \| \pi)=\mathbf{E}_{\sigma \sim \pi}\left[f\left(\frac{v(\sigma)}{\pi(\sigma)}\right)\right]=\mathbf{E}_{\tau \sim \mu} \mathbf{E}_{\sigma \sim \pi^{\tau}}\left[f\left(\frac{\varphi(\tau) \pi^{\tau}(\sigma)}{\mu(\tau) \pi^{\tau}(\sigma)}\right)\right]=\mathbf{E}_{\tau \sim \mu}\left[f\left(\frac{\varphi(\tau)}{\mu(\tau)}\right)\right]=D_{f}(\varphi \| \mu)$.

By the definition in (6.27), we have for all $\tau \in \Omega^{\prime}$,

$$
(v P)(\tau)=\sum_{\sigma: g(\sigma)=\tau} v(\sigma)=\varphi(\tau) \sum_{\sigma: g(\sigma)=\tau} \pi^{\tau}(\sigma)=\varphi(\tau)
$$

which implies $\varphi=v P$. Recall that $\mu=\pi P$. We have

$$
\begin{aligned}
D_{f}\left(\varphi P_{\text {Glauber }, \mu}^{\downarrow} \| \mu P_{\text {Glauber }, \mu}^{\downarrow}\right) & =D_{f}\left(v P P_{\text {Glauber }, \mu}^{\downarrow} \| \pi P P_{\text {Glauber }, \mu}^{\downarrow}\right) \\
(\text { by (6.26)) } & =D_{f}\left(v P_{\text {Glauber }, \pi}^{\downarrow} P^{\prime} \| \pi P_{\text {Glauber }, \pi}^{\downarrow} P^{\prime}\right) \\
\text { (by data processing inequality) } & \leq D_{f}\left(v P_{\text {Glauber }, \pi}^{\downarrow} \| \pi P_{\text {Glauber }, \pi}^{\downarrow}\right) \\
\text { (by assumption) } & \leq(1-\delta) D_{f}(v \| \pi) \\
\text { (by (6.28)) } & =(1-\delta) D_{f}(\varphi \| \mu) .
\end{aligned}
$$

We are now ready to prove Lemma 6.22.
Proof of Lemma 6.22. Let $\Omega=\{0,1,2\}^{E}$ denote the support of $\pi_{\mathrm{gm}}$. Define the map $g$ by $g(0)=0, g(1)=1$ and $g(2)=1$. By Lemma 6.11, it holds that $\pi_{\mathrm{wrc}}=\pi_{\mathrm{gm}} P$. Lemma 6.22 follows from Lemma 6.14, Lemma 6.16 and Lemma 6.25.

### 6.5.2 Faster mixing via perturbed chains

Given a subgraph-world model $(G ; \mathbf{p}, \eta)$, we define the "perturbed" model $(G ; \mathbf{p}, \widehat{\eta})$ by

$$
\widehat{\eta}_{v}= \begin{cases}\frac{1}{n}, & \text { if } 0 \leq \eta_{v} \leq \frac{1}{n}  \tag{6.29}\\ \eta_{v}, & \text { otherwise }\end{cases}
$$

Call the induced distribution $\widehat{\pi_{\mathrm{sg}}}$. Take a random subgraph $\mathcal{S}$ according to $\widehat{\pi_{\mathrm{sg}}}$, and add each remaining edge $e \in E \backslash \mathcal{S}$ with probability $p_{e} /\left(1-p_{e}\right)$ to obtain $\mathcal{R}$. By Lemma 6.12, the resulting distribution is $\pi_{\mathrm{wrc}}(G ; 2 \mathbf{p}, \widehat{\lambda})=: \widehat{\pi_{\mathrm{wrc}}}$, where $\widehat{\lambda}_{v}=\frac{1-\widehat{\eta}_{v}}{1+\widehat{\eta}_{v}}$. Let $\widehat{P_{\text {wrc }}}$ denote the Glauber dynamics on $\widehat{\pi_{\text {wrc }}}$. Let $\widehat{P_{\text {wrc }}} \downarrow$ denote the down-walk of $\widehat{P_{\text {wrc }}}$. Applying the first item of Lemma 6.22 to the perturbed random-cluster model $(G ; 2 \mathbf{p}, \widehat{\lambda})$ yields that for any distribution $v$,

$$
D_{\chi^{2}}\left(v{\widehat{P_{\mathrm{wrc}}}}^{\downarrow} \| \widehat{\pi_{\mathrm{wrc}}} \widehat{P_{\mathrm{wrc}}} \downarrow\right) \leq\left(1-\frac{\min \left\{p_{\min }, 1-2 p_{\max }\right\}}{m^{2} n^{4}}\right) D_{\chi^{2}}\left(v \| \widehat{\pi_{\mathrm{wrc}}}\right)
$$

By Proposition 6.7, we know that

$$
\mathfrak{G a p}\left(\widehat{P_{\mathrm{wrc}}}\right) \geq \frac{\min \left\{p_{\min }, 1-2 p_{\max }\right\}}{m^{2} n^{4}}
$$

Based on this, the main effort of this subsection is to bound the spectral gap of the original model ( $G ; 2 \mathbf{p}, \lambda$ ) via the bounds for $(G ; 2 \mathbf{p}, \widehat{\lambda})$.

We start with comparing the two distributions.
Lemma 6.26. For any $R \subseteq E$,

$$
\frac{1}{9} \leq \frac{\widehat{\pi_{\mathrm{wrc}}}(R)}{\pi_{\mathrm{wrc}}(R)}<\mathrm{e}
$$

Proof. Let $n=|V|$. If $n=1$, the only possible $R$ is $\varnothing$ and the lemma holds. We assume $n \geq 2$ in the rest. To prove the first inequality,

$$
\frac{\widehat{\pi_{\mathrm{wrc}}}(R)}{\pi_{\mathrm{wrc}}(R)}=\frac{Z_{\mathrm{wrc}}}{\widehat{Z_{\mathrm{wrc}}}} \cdot \frac{\widehat{\mathrm{wt}} \mathrm{wrrc}(R)}{\frac{Z_{\mathrm{wrc}}}{\mathrm{wt}_{\mathrm{wrc}}(R)}}=\frac{\widehat{Z_{\mathrm{wrc}}}}{} \cdot \prod_{C \in \kappa(V, S)} \frac{1+\prod_{u \in C} \widehat{\lambda}_{u}}{1+\prod_{u \in C} \lambda_{u}}
$$

Note that $\frac{Z_{\text {wre }}}{Z_{\text {wrc }}} \geq 1$ because $\widehat{\lambda}_{u} \leq \lambda_{u}$, which implies that the weight of each configuration of the random cluster model decreases after replacing $\lambda$ with $\widehat{\lambda}$. The second term can be handled by

$$
\prod_{C \in \kappa(V, S)} \frac{1+\prod_{u \in C} \hat{\lambda}_{u}}{1+\prod_{u \in C} \lambda_{u}} \geq \prod_{C \in \kappa(V, S)} \frac{\prod_{u \in C} \hat{\lambda}_{u}}{\prod_{u \in C} \lambda_{u}} \geq\left(\frac{n-1}{n+1}\right)^{n} \geq \frac{1}{9}
$$

as $n \geq 2$.
For the second inequality, the definition of $\pi_{\text {wrc }}$, together with the relation between $Z_{\text {wrc }}$ and $Z_{\text {sg }}$ in Equation (6.6), gives

$$
\frac{\widehat{\pi_{\mathrm{wrc}}}(R)}{\pi_{\mathrm{wrc}}(R)}=\frac{Z_{\mathrm{sg}}(G ; \mathbf{p}, \eta)}{Z_{\mathrm{sg}}(G ; \mathbf{p}, \widehat{\eta})} \cdot \frac{\prod_{v \in V} \frac{1}{1+\widehat{\lambda}_{v}}}{\prod_{v \in V} \frac{1}{1+\lambda_{v}}} \cdot \frac{\prod_{C \in \kappa(V, R)}\left(1+\prod_{u \in C} \widehat{\lambda}_{u}\right)}{\prod_{C \in \kappa(V, R)}\left(1+\prod_{u \in C} \lambda_{u}\right)}
$$

There are three terms. For the first one, note that $\widehat{\eta}_{v}>\eta_{v}$ for all $v$, indicating that the weight of each configuration of the subgraph-world model is increased after replacing $\eta$ with $\widehat{\eta}$. As such, it is less or equal than 1 . The third term is also less or equal than 1 due to $\widehat{\lambda}_{v}<\lambda_{v}$. The second term can be bounded by

$$
\frac{\prod_{v \in V} \frac{1}{1+\widehat{\lambda}_{v}}}{\prod_{v \in V} \frac{1}{1+\lambda_{v}}}=\frac{\prod_{v \in V}\left(1+\widehat{\eta}_{v}\right)}{\prod_{v \in V}\left(1+\eta_{v}\right)} \leq\left(1+\frac{1}{n}\right)^{n}<\mathrm{e}
$$

which concludes this lemma.
We also have a bound on the ratio of the transition probability between the original and perturbed model in the Glauber dynamics.

Lemma 6.27. Let $P_{\text {wrc }}$ and $\widehat{P_{\text {wrc }}}$ be the transition matrices of Glauber dynamics on the random cluster models $(G ; 2 \mathbf{p}, \lambda)$ and $(G ; 2 \mathbf{p}, \widehat{\lambda})$ respectively. Then it holds that

$$
\frac{1}{9 \mathrm{e}} \leq \frac{\widehat{P_{\mathrm{wrc}}}\left(Z, Z^{\prime}\right)}{P_{\mathrm{wrc}}\left(Z, Z^{\prime}\right)} \leq 9 \mathrm{e} \quad \text { for all }\left|Z \oplus Z^{\prime}\right|=1
$$

Proof. Assume $Z^{\prime}=Z+e$ where $e \notin Z$. The case $Z^{\prime}=Z-e$ where $e \in Z$ follows by a similar argument. We then have

$$
\frac{1}{9 \mathrm{e}} \leq \frac{\widehat{P_{\mathrm{wrc}}}\left(Z, Z^{\prime}\right)}{P_{\mathrm{wrc}}\left(Z, Z^{\prime}\right)}=\frac{\widehat{\pi_{\mathrm{wrc}}}\left(Z^{\prime}\right)\left(\pi_{\mathrm{wrc}}(Z)+\pi_{\mathrm{wrc}}\left(Z^{\prime}\right)\right)}{\left(\widehat{\pi_{\mathrm{wrc}}}(Z)+\widehat{\pi_{\mathrm{wrc}}}\left(Z^{\prime}\right)\right) \pi_{\mathrm{wrc}}\left(Z^{\prime}\right)} \leq 9 \mathrm{e}
$$

Now we are ready to prove Lemma 6.23.
Proof of Lemma 6.23. Fix a test function $f$. Denote by $\mathcal{E}(f, f), \widehat{\mathcal{E}}(f, f)$ the Dirichlet form of $P_{\text {wrc }}$ and $\widehat{P_{\text {wrc }}}$ respectively. Denote by $\operatorname{Var}[f]$ and $\widehat{\operatorname{Var}}[f]$ the variance of $f$ with respect to $\pi_{\text {wrc }}$ and $\widehat{\pi_{\text {wrc }}}$ respectively. Then by Lemma 6.26 and Lemma 6.27,

$$
\begin{aligned}
\frac{\mathcal{E}(f, f)}{\operatorname{Var}[f]}= & \frac{\sum_{\substack{X, Y \subseteq E \\
|X Y|=1}} \pi_{\mathrm{wrc}}(X) P_{\mathrm{wrc}}(X, Y)(f(X)-f(Y))^{2}}{\sum_{\substack{X, Y \subseteq E \\
|X \oplus Y|=1}} \pi_{\mathrm{wrc}}(X) \pi_{\mathrm{wrc}}(Y)(f(X)-f(Y))^{2}} \\
& \geq \frac{\frac{1}{9 \mathrm{e}^{2}} \sum_{\substack{X, Y \subseteq E \\
|X \oplus Y|=1}} \widehat{\pi_{\mathrm{wrc}}}(X) \widehat{P_{\mathrm{wrc}}}(X, Y)(f(X)-f(Y))^{2}}{81 \sum_{\substack{X, Y \subseteq E \\
|X \oplus Y|=1}} \widehat{\pi_{\mathrm{wrc}}}(X) \widehat{\pi_{\mathrm{wrc}}}(Y)(f(X)-f(Y))^{2}}>\frac{1}{10^{4}} \frac{\widehat{\mathcal{E}}(f, f)}{\widehat{\operatorname{Var}}[f]} .
\end{aligned}
$$

Therefore, $\mathfrak{G a p}\left(P_{\text {wrc }}\right) \geq \frac{1}{10^{4}} \mathfrak{G a p}\left(\widehat{P_{\text {wrc }}}\right) \geq \frac{\min \left\{p_{\min ,} 1-2 p_{\max }\right\}}{10^{4} h^{4} m^{2}}$. Lemma 6.23 follows from Proposition 6.7.

### 6.6 Rapid mixing of Swendsen-Wang dynamics

Having analysed the edge-flipping dynamics, now we turn to relating it with the Swendsen-Wang dynamics. From this point on, we no longer need the grand model. We first reiterate the settings for clarity. Let $G=(V, E)$ be a graph. We consider the Ising model on $G$ with parameters $\lambda=\left(\lambda_{v}\right)_{v \in V}$ and $\beta=\left(\beta_{e}\right)_{e \in E}$, where $0<\lambda_{v} \leq 1$ for all $v \in V$ and $\beta_{e}>1$ for all $e \in E$, as well as the weighted random cluster model on $G$ with parameters $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$, where $p_{e}=1-\frac{1}{\beta_{e}}$ for all $e \in E$. Let $\pi_{\text {Ising }}$ over $\Omega_{I}=\{0,1\}^{V}$ denote the Gibbs distribution of the Ising model, and $\pi_{\text {wro }}$ over $\Omega_{\mathcal{R}}=\{0,1\}^{E}$ denote the distribution of the weighted random cluster model. We remark that we view $\pi_{\text {wrc }}$ as a distribution over $\{0,1\}^{E}$ instead of $2^{E}$.

Let $P_{\mathrm{SW}}^{\mathrm{wrc}}=P_{\mathcal{R} \rightarrow I} P_{I \rightarrow \mathcal{R}}$ denote the transition matrix of the Swendsen-Wang dynamics for weighted random cluster models as defined in Section 6.1.2.2, and $P_{\text {GlauberRC }}$ denote the transition matrix of the Glauber dynamics for weighted random cluster models. In this section, we compare the Swendsen-Wang dynamics with the Glauber dynamics. Ullrich [Ull14] showed the following result about the variance decay (spectral gap) of the Swendsen-Wang dynamics.

Lemma 6.28 ([Ull14, Remark 2 and Theorem 5]). Suppose $0<\lambda_{v} \leq 1$ for all $v \in V$. It holds that

$$
\mathfrak{G a p}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \geq \frac{\mathfrak{G a p}\left(P_{\mathrm{GlauberRC}}\right)}{2}
$$

The above result is proved in [Ull14] in the case where $p_{e}=p \in(0,1)$ for all $e \in E$ and $\lambda_{v}=1$ for all $v \in V .^{4}$ The model we consider allows that each $e$ has different $p_{e} \in(0,1)$ and each $v$ has different $\lambda_{v} \in(0,1]$. However, there is no substantial change required to generalise it to our setting. Alternatively, we provide a somewhat simpler proof of Lemma 6.28 in Remark 6.33.

Lemma 6.28 only compares the decay rate of the variance. The main technical result in this section is the following comparison lemma on the decay rate of the relative entropy.

Lemma 6.29. Suppose $0<\lambda_{v} \leq 1$ for all $v \in V$. Let $0<\delta<1$. For any distribution $v$ over $\Omega_{\mathcal{R}}$, if

$$
D_{\mathrm{KL}}\left(v P_{\mathrm{GlauberRC}}^{\downarrow} \| \pi_{\mathrm{wrc}} P_{\mathrm{GlauberRC}}^{\downarrow}\right) \leq(1-\delta) D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right),
$$

[^14]then it holds that
$$
D_{\mathrm{KL}}\left(v P_{\mathrm{SW}}^{\mathrm{wrc}} \| \pi_{\mathrm{wrc}} P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \leq\left(1-\frac{\delta}{4}\right) D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right)
$$

We are now ready to prove the main results in Theorem 6.1 and Theorem 6.2.
Proofs of Theorem 6.1 and Theorem 6.2. Let $\pi_{\mathrm{wrc}, \min }=\min _{S \subseteq E} \pi_{\mathrm{wrc}}(S)$ denote the minimum probability in $\pi_{\text {wrc }}$. It is straightforward to verify that $\pi_{\text {wrc, } \min } \geq \min \left\{p_{\min }, 1-\right.$ $\left.p_{\max }\right\}^{m} / 2^{m+n}$. By the data processing inequality, Proposition 6.6 and Proposition 6.7, we have

$$
\begin{aligned}
D_{\chi^{2}}\left(v P_{\mathrm{SW}}^{\mathrm{wrc}} \| \pi_{\mathrm{wrc}}\right) & =D_{\chi^{2}}\left(v P_{\mathrm{SW}}^{\mathrm{wrc}} \| \pi_{\mathrm{wrc}} P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \leq D_{\chi^{2}}\left(v P_{\mathcal{R} \rightarrow I} \| \pi_{\mathrm{wrc}} P_{\mathcal{R} \rightarrow I}\right) \\
& \leq\left(1-\left(\mathfrak{G a p}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right)\right) D_{\chi^{2}}\left(v \| \pi_{\mathrm{wrc}}\right) .\right.
\end{aligned}
$$

A lower bound of $\left(\mathfrak{G a p}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right)\right.$ can be obtained by Proposition 6.7, Lemma 6.23 and Lemma 6.28. Let $C_{1}$ be the constant in Theorem 6.24. By a similar calculation as that in the proof of Theorem 6.24, we have

$$
T_{\operatorname{mix}}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}, \epsilon\right) \leq 2 C_{1}\left(p_{\min }, p_{\max }\right) \cdot \min \left\{n^{4},\left(\frac{1}{1-\lambda_{\max }}\right)^{4}\right\} \cdot m^{2} \cdot\left(\log \frac{1}{\epsilon}+m\right)
$$

By (6.12), the mixing time of Swendsen-Wang dynamics on Ising model satisfies

$$
T_{\operatorname{mix}}\left(P_{\mathrm{SW}}^{\mathrm{Ising}}, \epsilon\right) \leq C_{1}^{\prime}\left(\beta_{\min }, \beta_{\max }\right) \cdot \min \left\{n^{4},\left(\frac{1}{1-\lambda_{\max }}\right)^{4}\right\} \cdot m^{2} \cdot\left(\log \frac{1}{\epsilon}+m\right)
$$

where $p_{\text {min }}=1-\frac{1}{\beta_{\text {min }}}, p_{\text {max }}=1-\frac{1}{\beta_{\text {max }}}$, and thus

$$
\begin{align*}
C_{1}^{\prime}\left(\beta_{\min }, \beta_{\max }\right) & =O\left(\frac{1}{\min \left\{p_{\min }, 1-p_{\max }\right\}} \log \frac{1}{\min \left\{p_{\min }, 1-p_{\max }\right\}}\right) \\
& =O\left(\left(\frac{\beta_{\min }}{1-\beta_{\min }}+\beta_{\max }\right) \log \left(\frac{\beta_{\min }}{1-\beta_{\min }}+\beta_{\max }\right)\right) . \tag{6.30}
\end{align*}
$$

This proves Theorem 6.1.
For the decay of the relative entropy, the initial KL-divergence is at most $\log \frac{1}{\pi_{\text {wre, min }}}$. Let $C_{2}$ be the constant in Theorem 6.24. By Lemma 6.29, Lemma 6.22, and (2.4), we can use a similar calculation as that in the proof of Theorem 6.24 to obtain

$$
T_{\mathrm{mix}}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}, \epsilon\right) \leq 4 C_{2} \cdot\left(\Delta, \delta, p_{\min }, p_{\max }\right) \cdot n\left(\log n+\log \frac{1}{\epsilon}\right)
$$

By (6.12), the mixing time of Swendsen-Wang dynamics on Ising model satisfies

$$
T_{\mathrm{mix}}\left(P_{\mathrm{SW}}^{\mathrm{Ising}}, \epsilon\right) \leq C_{2}^{\prime}\left(\Delta, \delta, \beta_{\min }, \beta_{\max }\right) \cdot n\left(\log n+\log \frac{1}{\epsilon}\right)
$$

where

$$
\begin{align*}
C_{2}^{\prime}\left(\Delta, \delta, \beta_{\min }, \beta_{\max }\right) & =\left(\frac{\Delta}{\delta^{2} \min \left\{p_{\min }, 1-p_{\max }\right\}}\right)^{o\left(\frac{\Delta^{2}}{\delta^{4} \min \left\{p_{\min } 1-p_{\max }\right\}}\right)} \\
& =\left(\frac{\Delta}{\delta^{2}}\left(\frac{\beta_{\min }}{1-\beta_{\min }}+\beta_{\max }\right)\right)^{o\left(\frac{\Delta^{2}}{\left.\delta^{4}\left(\frac{\beta_{\min }}{1-\beta_{\min }}+\beta_{\max }\right)\right)}\right.} . \tag{6.31}
\end{align*}
$$

This proves Theorem 6.2.

The rest of this section is dedicated to the proof of Lemma 6.29.

### 6.6.1 FKES distribution and single-bond dynamics

To compare the Swendsen-Wang dynamics to the Glauber dynamics, we first introduce the FKES (Fortuin-Kasteleyn-Edwards-Sokal) distribution [FK72, ES88] $\pi_{\text {FKES }}$ over $\Omega_{I} \times \Omega_{\mathcal{R}}$, which couples the Ising distribution $\pi_{\text {Ising }}$ and the random cluster distribution $\pi_{\text {wrc }}$ :

$$
\begin{equation*}
\forall \sigma \in \Omega_{I}, \tau \in \Omega_{\mathcal{R}}, \quad \pi_{\mathrm{FKES}}(\sigma \tau):=\pi_{\mathrm{Ising}}(\sigma) P_{I \rightarrow \mathcal{R}}(\sigma, \tau) \stackrel{(\star)}{=} \pi_{\mathrm{wrc}}(\tau) P_{\mathcal{R} \rightarrow I}(\tau, \sigma) \tag{6.32}
\end{equation*}
$$

where $\Omega_{I}=\{0,1\}^{V}, \Omega_{\mathcal{R}}=\{0,1\}^{E}, P_{I \rightarrow \mathcal{R}}$ and $P_{\mathcal{R} \rightarrow I}$ are defined in (6.7) and (6.8) respectively. The equation ( $\star$ ) holds due to Proposition 6.6. We use $\Omega_{\mathrm{FKES}} \subseteq \Omega_{I} \times \Omega_{\mathcal{R}}$ to denote the support of the distribution $\pi_{\text {FKES }}$. The above equation shows that

- the marginal distribution projected from $\pi_{\mathrm{FKES}}$ to $\Omega_{\bar{I}}$ is $\pi_{\mathrm{Is} \text { ing }}$;
- the marginal distribution projected from $\pi_{\mathrm{FKES}}$ to $\Omega_{\mathcal{R}}$ is $\pi_{\mathrm{wrc}}$;
- conditional on $\sigma \in \Omega_{I}$, the marginal distribution projected from $\pi_{\text {FKES }}$ to $\Omega_{\mathcal{R}}$ is $\mathbf{P}_{I \rightarrow \mathcal{R}}(\sigma, \cdot)$;
- conditional on $\tau \in \Omega_{\mathcal{R}}$, the marginal distribution projected from $\pi_{\text {FKES }}$ to $\Omega_{\mathcal{I}}$ is $\mathbf{P}_{\mathcal{R} \rightarrow I}(\tau, \cdot)$.

Define the following stochastic matrix from the weighted random cluster model to the FKES model

$$
\forall \tau_{1} \in \Omega_{\mathcal{R}}, \sigma \tau_{2} \in \Omega_{\mathrm{FKES}}, \quad P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\tau_{1}, \sigma \tau_{2}\right)=P_{\mathcal{R} \rightarrow I}\left(\tau_{1}, \sigma\right) \cdot \mathbb{I}\left[\tau_{1}=\tau_{2}\right],
$$

The operator $P_{\mathcal{R} \rightarrow \mathrm{FKES}}$ maps from $L_{2}\left(\pi_{\mathrm{FKES}}\right)$ to $L_{2}\left(\pi_{\mathrm{wrc}}\right)$, where $L_{2}(\pi)$ is the vector space with the inner product $\langle\cdot, \cdot\rangle_{\pi}$. The adjoint operator $P_{\mathrm{FKES} \rightarrow \mathcal{R}}$ is defined by

$$
\forall \sigma \tau_{1} \in \Omega_{\mathrm{FKES}}, \tau_{2} \in \Omega_{\mathcal{R}}, \quad P_{\mathrm{FKES} \rightarrow \mathcal{R}}\left(\sigma \tau_{1}, \tau_{2}\right)=\mathbb{I}\left[\tau_{1}=\tau_{2}\right] .
$$

For any $f \in L_{2}\left(\pi_{\mathrm{FKES}}\right)$ and $g \in L_{2}\left(\pi_{\mathrm{wrc}}\right)$, it holds that $\left\langle P_{\mathcal{R} \rightarrow \mathrm{FKES}} f, g\right\rangle_{\pi_{\mathrm{wrc}}}=\left\langle f, P_{\mathrm{FKES} \rightarrow \mathcal{R}} g\right\rangle_{\pi_{\mathrm{FKES}}}$.
Next, we define the edge down-walk on the joint distribution. Fix an edge $e \in E$. Given $\sigma \tau \in \Omega_{\mathrm{FKES}}$, let $P_{e}^{\downarrow}$ denote the edge down-walk that drops the value on edge $e$. Formally, $P_{e}^{\downarrow}$ is defined on any $\sigma \tau \in \Omega_{\mathrm{FKES}}$ and any $\sigma^{\prime} \tau^{\prime} \in \Omega_{\mathrm{FKES}}^{e}$,

$$
P_{e}^{\downarrow}\left(\sigma \tau, \sigma^{\prime} \tau^{\prime}\right)=\mathbb{I}\left[\sigma=\sigma^{\prime} \wedge \tau^{\prime}=\tau_{E-e}\right]
$$

where we use $E-e$ to denote $E \backslash\{e\}$. Let $\pi_{\mathrm{FKES}}^{e}=\pi_{\mathrm{FKES}} P_{e}^{\downarrow}$. Let $\Omega_{\mathrm{FKES}}^{e}$ denote the support of $\pi_{\mathrm{FKES}} P_{e}^{\downarrow}$. Suppose $e=\{u, v\}$. We then define the edge up-walk $P_{e}^{\uparrow}$, for all $\sigma^{\prime} \tau^{\prime} \in \Omega_{\mathrm{FKES}}^{e}$ and $\sigma \tau \in \Omega_{\mathrm{FKES}}$,

$$
P_{e}^{\uparrow}\left(\sigma^{\prime} \tau^{\prime}, \sigma \tau\right)=\mathbb{I}\left[\sigma=\sigma^{\prime} \wedge \tau_{E-e}=\tau^{\prime}\right] \times \begin{cases}p_{e} & \text { if } \tau_{e}=1 \text { and } \sigma(u)=\sigma(v) \\ 1-p_{e} & \text { if } \tau_{e}=0 \text { and } \sigma(u)=\sigma(v) \\ 0 & \text { if } \tau_{e}=1 \text { and } \sigma(u) \neq \sigma(v) \\ 1 & \text { if } \tau_{e}=0 \text { and } \sigma(u) \neq \sigma(v)\end{cases}
$$

For any $f \in L_{2}\left(\pi_{\mathrm{FKES}}\right)$ and $g \in L_{2}\left(\pi_{\mathrm{FKES}}^{e}\right)$, it holds that $\left\langle P_{e}^{\uparrow} f, g\right\rangle_{\pi_{\mathrm{FKES}}^{e}}=\left\langle f, P_{e}^{\downarrow} g\right\rangle_{\pi_{\mathrm{FKES}}}$.
Since in each transition step of $P_{e}^{\uparrow}, \sigma^{\prime} \tau_{E-e}^{\prime}=\sigma \tau_{E-e}$ and the distribution of $\tau_{e}^{\prime}$ depends only on $\sigma_{u}$ and $\sigma_{v}$, the following observation is straightforward to verify.

Observation 6.30. For anye, $f \in E$, it holds that

- $\left(P_{e}^{\downarrow} P_{e}^{\uparrow}\right)\left(P_{f}^{\downarrow} P_{f}^{\uparrow}\right)=\left(P_{f}^{\downarrow} P_{f}^{\uparrow}\right)\left(P_{e}^{\downarrow} P_{e}^{\uparrow}\right)$.
- $P_{e}^{\downarrow} P_{e}^{\uparrow}=\left(P_{e}^{\downarrow} P_{e}^{\uparrow}\right)^{2}$.

The single bond dynamics $P_{\mathrm{SB}}: \Omega_{\mathcal{R}} \times \Omega_{\mathcal{R}} \rightarrow \mathbb{R}_{\geq 0}$ is defined as follows

$$
P_{\mathrm{SB}}=P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow} P_{e}^{\uparrow}\right) P_{\mathrm{FKES} \rightarrow \mathcal{R}} .
$$

Intuitively, given any $\tau \in \Omega_{\mathcal{R}}, P_{\mathrm{SB}}$ first transforms $\tau$ into a joint configuration $\sigma \tau \in$ $\Omega_{\mathrm{FKES}} ;$ samples an edge $e \in E$ uniformly at random; updates $\tau_{e}$ conditional on $\sigma$; drops $\sigma$ and keeps the random cluster configuration $\tau$. Similarly, we can decompose the single bond dynamics as $P_{\mathrm{SB}}=P_{\mathrm{SB}}^{\downarrow} P_{\mathrm{SB}}^{\uparrow}$ :

$$
\begin{equation*}
P_{\mathrm{SB}}^{\downarrow}=P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right) \text { and } P_{\mathrm{SB}}^{\uparrow}=P_{E}^{\uparrow} P_{\mathrm{FKES} \rightarrow \mathcal{R}}, \tag{6.33}
\end{equation*}
$$

where for convenience, we treat $\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right)$ as a matrix defined on $\Omega_{\mathrm{FKES}} \times\left(\cup_{e \in E} \Omega_{\mathrm{FKES}}^{e}\right)$ and $P_{E}^{\uparrow}:\left(\cup_{e \in E} \Omega_{\mathrm{FKES}}^{e}\right) \times \Omega_{\mathrm{FKES}} \rightarrow \mathbb{R}_{\geq 0}$ is defined by $P_{E}^{\uparrow}(x, y)=P_{e}^{\uparrow}(x, y)$ where $x \in$ $\Omega_{\mathrm{FKES}}^{e}$ for some $e \in E$ and $y \in \Omega_{\mathrm{FKES}}$. Note that once $x$ is given, $e$ is uniquely determined, and $P_{E}^{\uparrow}$ agrees with $P_{e}^{\uparrow}$. It is straightforward to check ( $\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}$ ) and $P_{E}^{\uparrow}$ is a pair of adjoint operators.

Lemma 6.31. Suppose $0<\lambda_{v} \leq 1$ for all $v \in V$. Let $0<\delta<1$. For any distribution $v$ over $\Omega_{\mathcal{R}}$, if

$$
D_{\mathrm{KL}}\left(v P_{\text {GlauberRC }}^{\downarrow} \| \pi_{\mathrm{wrc}} P_{\text {GlauberRC }}^{\downarrow}\right) \leq(1-\delta) D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right),
$$

then it holds that

$$
\begin{equation*}
D_{\mathrm{KL}}\left(v P_{\mathrm{SB}}^{\downarrow} \| \pi_{\mathrm{wrc}} P_{\mathrm{SB}}^{\downarrow}\right) \leq\left(1-\frac{\delta}{4}\right) D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right) \tag{6.34}
\end{equation*}
$$

The proof of Lemma 6.31 is deferred to Section 6.6.2. We prove Lemma 6.29 first.
Proof of Lemma 6.29. By Observation 6.30, the Swendsen-Wang dynamics $P_{S W}^{\mathrm{wrc}}$ can be written as

$$
\begin{aligned}
P_{\mathrm{SW}}^{\mathrm{wrc}} & =P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\prod_{e \in E} P_{e}^{\downarrow} P_{e}^{\uparrow}\right) P_{\mathrm{FKES} \rightarrow \mathcal{R}}=P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow} P_{e}^{\uparrow}\right)\left(\prod_{e \in E} P_{e}^{\downarrow} P_{e}^{\uparrow}\right) P_{\mathrm{FKES} \rightarrow \mathcal{R}} \\
& =P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right) P_{E}^{\uparrow}\left(\prod_{e \in E} P_{e}^{\downarrow} P_{e}^{\uparrow}\right) P_{\mathrm{FKES} \rightarrow \mathcal{R}}=P_{\mathrm{SB}}^{\downarrow} P_{E}^{\uparrow}\left(\prod_{e \in E} P_{e}^{\downarrow} P_{e}^{\uparrow}\right) P_{\mathrm{FKES} \rightarrow \mathcal{R}} .
\end{aligned}
$$

Hence, by the data processing inequality, we have

$$
D_{\mathrm{KL}}\left(v P_{\mathrm{SW}}^{\mathrm{wrc}} \| \pi_{\mathrm{wrc}} P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \leq D_{\mathrm{KL}}\left(v P_{\mathrm{SB}}^{\downarrow} \| \pi_{\mathrm{wrc}} P_{\mathrm{SB}}^{\downarrow}\right) \leq\left(1-\frac{\delta}{4}\right) D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right),
$$

where the last inequality holds due to Lemma 6.31.
Remark 6.32. The above proof can be illustrated by the diagrams below. Each step of transition of the single-bond dynamics can be decomposed as


While for the Swendsen-Wang dynamics, it is

$$
\tau \xrightarrow{P_{\mathcal{R} \rightarrow \mathrm{FKES}}} \sigma \tau \xrightarrow{\prod_{e}\left(P_{e}^{\downarrow} P_{e}^{\uparrow}\right)} \sigma \tau^{\prime \prime} \xrightarrow{P_{\mathrm{FKES} \rightarrow \mathcal{R}}} \tau^{\prime \prime}
$$

Note that the down-up operator $P_{e}^{\vee}:=P_{e}^{\downarrow} P_{e}^{\uparrow}$ is commutative and idempotent, allowing us to decompose Swendsen-Wang dynamics by performing the down walk of one edge first:

$$
\tau \underbrace{P_{\mathcal{R} \rightarrow \mathrm{FKES}}}_{P_{\mathrm{SB}}^{\downarrow}} \sigma \tau \xrightarrow{\frac{1}{m} \sum_{e} P_{e}^{\downarrow}} \sigma \tau \backslash e \xrightarrow{P_{E}^{\uparrow}} \sigma \tau^{\prime} \xrightarrow{\prod_{e}\left(P_{e}^{\downarrow} P_{e}^{\uparrow}\right)} \sigma \tau^{\prime \prime} \xrightarrow{P_{\mathrm{FKES} \rightarrow \mathcal{R}}} \tau^{\prime \prime} .
$$

Therefore, the variance (see the next remark) or entropy decay (see the next section) of each Swendsen-Wang transition is faster than that of the single-bond downwalk $P_{\text {SB }}^{\downarrow}$.

Remark 6.33 (a simple proof of the main result in [Ull14] and Lemma 6.28). If we replace KL-divergence in the above proof with $\chi^{2}$-divergence, the same proof shows that for any distribution $v$,

$$
D_{\chi^{2}}\left(v P_{\mathrm{SW}}^{\mathrm{wrc}} \| \pi_{\mathrm{wrc}} P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \leq D_{\chi^{2}}\left(v P_{\mathrm{SB}}^{\downarrow} \| \pi_{\mathrm{wrc}} P_{\mathrm{SB}}^{\downarrow}\right) .
$$

By Proposition 6.7, we have the following result

$$
\mathfrak{F a p}\left(\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right)^{2}\right) \geq \mathfrak{G a p}\left(P_{\mathrm{SB}}\right) \quad \Longrightarrow \quad \mathfrak{F a p}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \geq \frac{\mathfrak{F} \mathfrak{a} \mathfrak{p}\left(P_{\mathrm{SB}}\right)}{2}
$$

which recovers the main result in [Ull14] up to the factor 2.
The above analysis loses a factor of 2 because we compare $P_{\mathrm{SW}}^{\mathrm{wrc}}$ with $P_{\mathrm{SB}}^{\downarrow}$. Note that $P_{\mathrm{SW}}^{\mathrm{wrc}}$ can be decomposed as $P_{\mathcal{R} \rightarrow I} \cdot P_{I \rightarrow \mathcal{R}}$. This factor 2 can be saved by comparing the intermediate step $P_{\mathcal{R} \rightarrow I}$ with $P_{\mathrm{SB}}^{\downarrow}$. Define the intermediate state space $\Omega_{\mathcal{R}}^{*}=\{0,1, *\}^{E}$, where for any $\tau \in \Omega_{\mathcal{R}}^{*}$ and $e \in E, \tau_{e}=*$ means that $e$ is not assigned with any value, in other words, the value on $e$ is dropped. We can view $P_{e}^{\downarrow}$ as a random walk on $\Omega_{I} \times \Omega_{\mathcal{R}}^{*}$ such that given any $\sigma \tau \in \Omega_{I} \times \Omega_{\mathcal{R}}^{*}, P_{e}^{\downarrow}$ drops the value $\tau_{e}$ (i.e. sets $\tau_{e}=*$ ) and keeps $\sigma \tau_{E \backslash\{e\}}$ unchanged. It is straightforward to verify that $P_{\mathcal{R} \rightarrow I}$ is equivalent to $P_{\mathcal{R} \rightarrow \mathrm{FKES}} \prod_{e \in E} P_{e}^{\downarrow}$. Note that

$$
P_{\mathcal{R} \rightarrow \mathrm{FKES}} \prod_{e \in E} P_{e}^{\downarrow}=P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right) \prod_{e \in E} P_{e}^{\downarrow}
$$

as updating an edge twice is the same as updating it once. Recall (6.33) that $P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right)$ is equivalent to $P_{\text {SB }}^{\downarrow}$. By the data processing inequality, we have the following stronger result

$$
D_{\chi^{2}}\left(v P_{\mathcal{R} \rightarrow I} \| \pi_{\mathrm{wrc}} P_{\mathcal{R} \rightarrow I}\right) \leq D_{\chi^{2}}\left(v P_{\mathrm{SB}}^{\downarrow} \| \pi_{\mathrm{wrc}} P_{\mathrm{SB}}^{\downarrow}\right),
$$

which gives a better bound $\mathfrak{G a p}\left(P_{\mathrm{SW}}^{\mathrm{wrc}}\right) \geq \mathfrak{G a p}\left(P_{\mathrm{SB}}\right)$, matching [Ull14].
For Lemma 6.28, we still need to compare $\mathfrak{G a p}\left(P_{\mathrm{SB}}\right)$ with $\mathfrak{G a p}\left(P_{\text {GlauberRC }}\right)$. We claim that $\mathfrak{G a p}\left(P_{\mathrm{SB}}\right) \geq \mathfrak{G a p}\left(P_{\text {GlauberRC }}\right) / 2$. By a simple comparison argument through the Dirichlet form (see for example [LP17, Section 13.3]), it suffices to show

$$
\frac{P_{\mathrm{GlauberRC}}(A, B)}{P_{\mathrm{SB}}(A, B)} \leq 2
$$

for all $A, B \subseteq E$ such that $|A \oplus B|=1$. Let $e$ be the edge where $A$ and $B$ differ. By writing down the transition probability explicitly, the above ratio is 1 if $A$ and $B$ give the same connected components, and

$$
\frac{1}{1-\left(1-\frac{1+X Y}{(1+X)(1+Y)}\right) p_{e}}
$$

otherwise, where $X=\prod_{v \in C_{1}} \lambda_{v}, Y=\prod_{w \in C_{2}} \lambda_{w}$, and $C_{1}, C_{2}$ are the two components created by disconnecting $e$. Using the inequality that $1 / 2 \leq(1+X Y) /((1+X)(1+$ $Y)) \leq 1$ for all $0 \leq X, Y \leq 1$, the above ratio is bounded by 2 .

### 6.6.2 Comparing Glauber dynamics to single-bond dynamics

We first introduce some notations. Let $\mu$ be a distribution with support $\Omega \subseteq Q^{V}$.
For any $S \subseteq V$, we use $\mu_{S}$ to denote the marginal distribution on $S$ induced by $\mu$. Let $\Omega\left(\mu_{S}\right)$ denote the support of $\mu_{S}$. Given any $x_{S} \in \Omega\left(\mu_{S}\right)$, we use $\mu^{x_{S}}$ to denote the distribution over $\Omega$ obtained from $\mu$ conditional on $x_{S}$. Formally, for any $y \in \Omega$, $\mu^{x_{S}}(y)=\mathbb{I}\left[y_{S}=x_{S}\right] \mu(y) / \mu_{S}\left(x_{S}\right)$, where $y_{S}$ is the restriction of $y$ on $S$. For any $\Lambda \subseteq V$, we use $\mu_{\Lambda}^{x_{S}}$ to denote the marginal distribution on $\Lambda$ induced by $\mu^{x_{S}}$. We need the following chain rule of the KL-divergence. Such a result is very well-known. See for example [CP21, Lemma 3.1].

Lemma 6.34. For any distribution $v$ be a distribution over $\Omega$, any $S \subseteq V$, it holds that $D_{\mathrm{KL}}(v \| \mu)=D_{\mathrm{KL}}\left(v_{S} \| \mu_{S}\right)+\mathbf{E}_{x_{S} \sim v_{S}} D_{\mathrm{KL}}\left(v^{x_{S}} \| \mu^{x_{S}}\right)=D_{\mathrm{KL}}\left(v_{S} \| \mu_{S}\right)+\mu\left[\operatorname{Ent}_{V-S}(f)\right]$, where $V-S=V \backslash S$ and $f: \Omega \rightarrow \mathbb{R}_{\geq 0}$ is defined by $f(x)=v(x) / \mu(x)$ and

$$
\mu\left[\operatorname{Ent}_{V-S}(f)\right]=\sum_{x_{S} \in \Omega\left(\mu_{S}\right)} \mu_{S}\left(x_{S}\right) \operatorname{Ent}_{\mu^{x_{S}}}(f)
$$

Proof. The first equation $D_{\mathrm{KL}}(v \| \mu)=D_{\mathrm{KL}}\left(v_{S} \| \mu_{S}\right)+\mathbf{E}_{x_{S} \sim v_{S}} D_{\mathrm{KL}}\left(v^{x_{S}} \| \mu^{x_{S}}\right)$ follows directly from the standard chain rule of KL -divergence. To prove the second equation,
for any $x_{S} \in \Omega\left(v_{S}\right)$, define

$$
\forall y \in \Omega, \quad g^{x_{S}}(y):= \begin{cases}\frac{v^{x}(y)}{\mu^{x}(y)}=\frac{v(y) \mu_{S}\left(x_{S}\right)}{\mu(y) v_{S}\left(x_{S}\right)}=\frac{\mu_{S}\left(x_{S}\right)}{v_{S}\left(x_{S}\right)} f(y) & \text { if } y_{S}=x_{S} \\ 0 & \text { otherwise }\end{cases}
$$

Since $\Omega\left(v_{S}\right) \subseteq \Omega\left(\mu_{S}\right)$, we have

$$
\begin{aligned}
\mathbf{E}_{x_{S} \sim v_{S}} D_{\mathrm{KL}}\left(v^{x_{S}} \| \mu^{x_{S}}\right) & =\sum_{x_{S} \in \Omega\left(v_{S}\right)} v\left(x_{S}\right) \operatorname{Ent}_{\mu^{x_{S}}}\left(g^{x_{S}}\right)=\sum_{x_{S} \in \Omega\left(v_{S}\right)} v\left(x_{S}\right) \operatorname{Ent}_{\mu^{x_{S}}}\left(\frac{\mu_{S}\left(x_{S}\right)}{v_{S}\left(x_{S}\right)} f\right) \\
& =\sum_{x_{S} \in \Omega\left(v_{S}\right)} \mu\left(x_{S}\right) \operatorname{Ent}_{\mu^{x} S}(f) . \quad\left(\operatorname{as~}_{\operatorname{Ent}_{\mu} x_{S}}(c f)=\operatorname{cnt}_{\mu^{x_{S}}}(f)\right)
\end{aligned}
$$

Note that for all $\sigma \in \Omega$ such that $\sigma_{S} \in \Omega\left(\mu_{S}\right) \backslash \Omega\left(v_{S}\right)$, it holds that $f(\sigma)=\frac{v(\sigma)}{\mu(\sigma)}=0$, implying that $\operatorname{Ent}_{\mu} \sigma_{S}(f)=0$. We have

$$
\begin{aligned}
\mathbf{E}_{x_{S} \sim v_{S}} D_{\mathrm{KL}}\left(v^{x_{S}} \| \mu^{x_{S}}\right) & =\sum_{x_{S} \in \Omega\left(v_{S}\right)} \mu\left(x_{S}\right) \operatorname{Ent}_{\mu^{x_{S}}}(f)+\sum_{x_{S} \in \Omega\left(\mu_{S}\right) \backslash \Omega\left(v_{S}\right)} \mu\left(x_{S}\right) \operatorname{Ent}_{\mu^{x_{S}}}(f) \\
& =\sum_{x_{S} \in \Omega\left(\mu_{S}\right)} \mu\left(x_{S}\right) \operatorname{Ent}_{\mu^{x_{S}}}(f)=\mu\left[\operatorname{Ent}_{V-S}(f)\right] .
\end{aligned}
$$

Now we are ready to prove Lemma 6.31.
Proof of Lemma 6.31. For any $e \in E$, let $E-e=E \backslash\{e\}$, using Lemma 6.34, it holds that

$$
D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right)=D_{\mathrm{KL}}\left(v_{E-e} \| \pi_{\mathrm{wrc}, E-e}\right)+\pi_{\mathrm{wrc}}\left[\operatorname{Ent}_{e}(f)\right], \quad \text { where } f(\tau)=\frac{v(\tau)}{\pi_{\mathrm{wrc}}(\tau)}
$$

Averaging over all $e \in E$, we get

$$
\begin{aligned}
\frac{1}{m} \sum_{e \in E} \pi_{\mathrm{wrc}}\left[\operatorname{Ent}_{e}(f)\right] & =\frac{1}{m} \sum_{e \in E} D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right)-\frac{1}{m} \sum_{e \in E} D_{\mathrm{KL}}\left(v_{E-e} \| \pi_{\mathrm{wrc}, E-e}\right) \\
& =D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right)-D_{\mathrm{KL}}\left(v P_{\text {GlauberRC }}^{\downarrow} \| \pi_{\mathrm{wrc}} P_{\text {GlauberRC }}^{\downarrow}\right) .
\end{aligned}
$$

By the assumption of Lemma 6.31, we have

$$
\begin{equation*}
\frac{1}{m} \sum_{e \in E} \pi_{\mathrm{wrc}^{2}}\left[\operatorname{Ent}_{e}(f)\right] \geq \delta D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right) \tag{6.35}
\end{equation*}
$$

Next, by (6.33), we have

$$
\begin{aligned}
D_{\mathrm{KL}}\left(v P_{\mathrm{SB}}^{\downarrow} \| \pi_{\mathrm{wrc}} P_{\mathrm{SB}}^{\downarrow}\right) & =D_{\mathrm{KL}}\left(v P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right) \| \pi_{\mathrm{wrc}} P_{\mathcal{R} \rightarrow \mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right)\right) \\
& =D_{\mathrm{KL}}\left(v_{\text {joint }}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right) \| \pi_{\mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right)\right),
\end{aligned}
$$

where $v_{\mathrm{joint}}=v P_{\mathcal{R} \rightarrow \mathrm{FKES}}$ so that for any $\sigma \tau \in \Omega_{\mathrm{FKES}}, v_{\mathrm{joint}}(\sigma \tau)=v(\tau) \pi_{\mathrm{FKES}, V}^{\tau}(\sigma)$. Hence, we have

$$
D_{\mathrm{KL}}\left(v_{\mathrm{joint}} \| \pi_{\mathrm{FKES}}\right)=\sum_{\sigma \tau \in \Omega_{\mathrm{FKES}}} v_{\mathrm{joint}}(\sigma \tau) \log \frac{v(\tau) \pi_{\mathrm{FKES}, V}^{\tau}(\sigma)}{\pi_{\mathrm{wrc}}(\tau) \pi_{\mathrm{FKES}, V}^{\tau}(\sigma)}=D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right)
$$

With these two equations, our goal, (6.34), is equivalent to

$$
\begin{equation*}
D_{\mathrm{KL}}\left(v_{\mathrm{joint}} \| \pi_{\mathrm{FKES}}\right)-D_{\mathrm{KL}}\left(v_{\mathrm{joint}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right) \| \pi_{\mathrm{FKES}}\left(\frac{1}{m} \sum_{e \in E} P_{e}^{\downarrow}\right)\right) \geq \frac{\delta}{4} D_{\mathrm{KL}}\left(v_{\mathrm{joint}} \| \pi_{\mathrm{FKES}}\right) . \tag{6.36}
\end{equation*}
$$

Using Lemma 6.34, for any $e \in E$, let $V+E-e$ be $V \cup E \backslash\{e\}$, it holds that

$$
D_{\mathrm{KL}}\left(v_{\text {joint }} \| \pi_{\mathrm{FKES}}\right)=D_{\mathrm{KL}}\left(v_{\mathrm{joint}, V+E-e} \| \pi_{\mathrm{FKES}, V+E-e}\right)+\pi_{\mathrm{FKES}}\left[\operatorname{Ent}_{e}(\bar{f})\right],
$$

where

$$
\bar{f}(\sigma \tau)=\frac{v_{\mathrm{joint}}(\sigma \tau)}{\pi_{\mathrm{FKES}}(\sigma \tau)}=\frac{v(\tau) \pi_{\mathrm{FKES}, V}^{\tau}(\sigma)}{\pi_{\mathrm{wrc}}(\tau) \pi_{\mathrm{FKES}, V}^{\tau}(\sigma)}=\frac{v(\tau)}{\pi_{\mathrm{wrc}}(\tau)}=f(\tau) .
$$

Hence, (6.36) is equivalent to

$$
\frac{1}{m} \sum_{e \in E} \pi_{\mathrm{FKES}}\left[\operatorname{Ent}_{e}(\bar{f})\right] \geq \frac{\delta}{4} D_{\mathrm{KL}}\left(v_{\mathrm{joint}} \| \pi_{\mathrm{FKES}}\right)=\frac{\delta}{4} D_{\mathrm{KL}}\left(v \| \pi_{\mathrm{wrc}}\right)
$$

Given (6.35), to prove the above inequality, it suffices to show that for any $e \in E$,

$$
\begin{equation*}
4 \cdot \pi_{\mathrm{FKES}}\left[\operatorname{Ent}_{e}(\bar{f})\right] \geq \pi_{\mathrm{wrc}}\left[\operatorname{Ent}_{e}(f)\right] \tag{6.37}
\end{equation*}
$$

We now prove (6.37). We use $\sigma$ to denote the vertex configuration in $\{0,1\}^{V}$ and $\tau$ to denote the edge configuration $\tau \in\{0,1\}^{E}$. Suppose $e=\{u, \nu\}$. We use $\tau_{-e}$ to denote a configuration in $\{0,1\}^{E-e}$. To ease the notation, we use $\pi_{\text {FKES }}\left(\sigma \tau_{-e}\right)$ to denote $\pi_{\mathrm{FKES}, E-e}\left(\sigma \tau_{-e}\right)$. For any $\tau_{e} \in\{0,1\}$, we use $\tau_{-e} \tau_{e}$ to denote a full configuration $\tau$ in $\{0,1\}^{E}$. We have

$$
\left.\begin{array}{rl} 
& \pi_{\mathrm{FKES}}\left[\operatorname{Ent}_{e}(\bar{f})\right]=\sum_{\sigma \tau_{-e}} \pi_{\mathrm{FKES}}\left(\sigma \tau_{-e}\right) \operatorname{Ent}_{\pi_{\mathrm{FKES}}}^{\sigma \tau_{-e}}(\bar{f}) \\
= & \sum_{\sigma \tau_{-e}} \pi_{\mathrm{FKES}}\left(\sigma \tau_{-e}\right) \sum_{\tau_{e} \in\{0,1\}} \pi_{\mathrm{FKES}, e} \sigma \tau_{-e}
\end{array} \tau_{e}\right) \bar{f}\left(\sigma \tau_{-e} \tau_{e}\right) \log \frac{\bar{f}\left(\sigma \tau_{-e} \tau_{e}\right)}{\sum_{\tau_{e} \in\{0,1\}} \pi_{\mathrm{FKES}, e}^{\sigma \tau_{-e}}\left(\tau_{e}\right) \bar{f}\left(\sigma \tau_{-e} \tau_{e}\right)} .
$$

If $\sigma_{u} \neq \sigma_{v}$, then $\pi_{\mathrm{FKES}, e}^{\sigma \tau_{-e}}(0)=1$, and in this case

$$
\sum_{\tau_{e} \in\{0,1\}} \pi_{\mathrm{FKES}, e}^{\sigma \tau_{-e}}\left(\tau_{e}\right) \bar{f}\left(\sigma \tau_{-e} \tau_{e}\right) \log \frac{\bar{f}\left(\sigma \tau_{-e} \tau_{e}\right)}{\sum_{\tau_{e} \in\{0,1\}} \pi_{\mathrm{FKES}, e} \sigma \tau_{-e}}\left(\tau_{e}\right) \bar{f}\left(\sigma \tau_{-e} \tau_{e}\right) \quad=0 .
$$

Thus we only need to consider the case where the two endpoints of $e$ get the same spin. Note that this always happens if $\tau_{-e} \in C_{e}$, where $C_{e} \subseteq\{0,1\}^{E-e}$ is the set of $\tau_{-e}$ such that $u$ and $v$ are connected by edges assigned 1 in $\tau_{-e}$. Again, to ease the notation, let $\pi_{\mathrm{wrc}}\left(\tau_{-e}\right)$ be $\pi_{\mathrm{wrc}, E-e}\left(\tau_{-e}\right)$. Hence, we have

$$
\begin{align*}
\pi_{\mathrm{FKES}}\left[\operatorname{Ent}_{e}(\bar{f})\right] & =\sum_{\tau_{-e} \in C_{e}} \pi_{\mathrm{wrc}}\left(\tau_{-e}\right) h\left(p_{e}, \tau_{-e}\right)+\sum_{\tau_{-e} \notin C_{e}} \pi_{\mathrm{wrc}}\left(\tau_{-e}\right) \mathbf{P r}_{\sigma \sim \pi_{\mathrm{FKES}, V}^{\tau_{-e}}}\left[\sigma_{u}=\sigma_{v}\right] h\left(p_{e}, \tau_{-e}\right) \\
& \geq \sum_{\tau_{-e} \in C_{e}} \pi_{\mathrm{wrc}}\left(\tau_{-e}\right) h\left(p_{e}, \tau_{-e}\right)+\frac{1}{2} \sum_{\tau_{-e} \notin C_{e}} \pi_{\mathrm{wrc}}\left(\tau_{-e}\right) h\left(p_{e}, \tau_{-e}\right), \tag{6.38}
\end{align*}
$$

where

$$
\begin{aligned}
h\left(p_{e}, \tau_{-e}\right): & p_{e} f\left(\tau_{-e} 1\right) \log f\left(\tau_{-e} 1\right)+\left(1-p_{e}\right) f\left(\tau_{-e} 0\right) \log f\left(\tau_{-e} 0\right) \\
& -\left(p_{e} f\left(\tau_{-e} 1\right)+\left(1-p_{e}\right) f\left(\tau_{-e} 0\right)\right) \log \left(p_{e} f\left(\tau_{-e} 1\right)+\left(1-p_{e}\right) f\left(\tau_{-e} 0\right)\right)
\end{aligned}
$$

(Recall that $\tau_{-e} \tau_{e}$ is a full configuration on $E$, where $\tau_{e}=0$ or 1.) To see (6.38), since all external fields are consistent, $\operatorname{Pr}_{\sigma \sim \tau_{\text {FKES, } V}^{\tau-e}}\left[\sigma_{u}=\sigma_{v}\right] \geq 1 / 2$. This is because we can further condition on $\tau_{e}$ : if $\tau_{e}=1$, then $\sigma_{u}=\sigma_{v}$ with probability 1 , and if $\tau_{e}=0$, then $\sigma_{u}$ and $\sigma_{v}$ are independent and biased towards the same direction, in which case they are equal with probability at least $1 / 2$. The final probability is a linear combination of the two cases.

Similarly, we can expand the right hand side of (6.37),

$$
\begin{align*}
\pi_{\mathrm{wrc}}\left[\operatorname{Ent}_{e}(f)\right] & =\sum_{\tau_{-e}} \pi_{\mathrm{wrc}\left(\tau_{-e}\right) \operatorname{Ent}_{\pi_{\mathrm{wrc}}^{\tau_{-e}}}(f)} \\
& =\sum_{\tau_{-e}} \pi_{\mathrm{wrc}}\left(\tau_{-e}\right) \sum_{\tau_{e} \in\{0,1\}} \pi_{\mathrm{wrc}, e}^{\tau_{-e}}\left(\tau_{e}\right) f\left(\tau_{-e} \tau_{e}\right) \log \frac{f\left(\tau_{-e} \tau_{e}\right)}{\sum_{\tau_{e} \in\{0,1\}} \pi_{\mathrm{wrc}, e}^{\tau_{-}}\left(\tau_{e}\right) f\left(\tau_{-e} \tau_{e}\right)} \\
& =\sum_{\tau_{-e} \in C_{e}} \pi_{\mathrm{wrc}}\left(\tau_{-e}\right) h\left(p_{e}, \tau_{-e}\right)+\sum_{\tau_{-e} \notin C_{e}} \pi_{\mathrm{wrc}}\left(\tau_{-e}\right) h\left(\frac{p_{e}}{1-\alpha\left(\tau_{-e}\right)\left(p_{e}-1\right)}, \tau_{-e}\right) . \tag{6.39}
\end{align*}
$$

In the last step above we use $\frac{p_{e}}{1-\alpha\left(\tau_{-e}\right)\left(p_{e}-1\right)}=\pi_{\mathrm{wrc}, e}^{\tau_{-e}}(1)$ where $\alpha\left(\tau_{-e}\right)$ is a factor depending on $\tau_{-e}$, derived as follows. Suppose $e=\{u, v\}$. Consider the random cluster configuration with $e$ set not to be taken, and adding $e$ causes the two connected components $C_{1}$ and $C_{2}$ to be merged as one, where $u$ is in $C_{1}$ and $v$ is in $C_{2}$. Let $X=X\left(\tau_{-e}\right)=\prod_{w \in C_{1}} \lambda_{w}$ and $Y=Y\left(\tau_{-e}\right)=\prod_{w \in C_{2}} \lambda_{w}$. We have

$$
\pi_{\mathrm{wrc}, e}^{\tau_{-e}}(1)=\frac{p_{e}(1+X Y)}{p_{e}(1+X Y)+\left(1-p_{e}\right)(1+X)(1+Y)}=\frac{p_{e}}{1-\frac{X+Y}{1+X Y}\left(p_{e}-1\right)},
$$

which means we can take $\alpha\left(\tau_{-e}\right)=\frac{X+Y}{1+X Y}$. Moreover, we have $0 \leq \alpha\left(\tau_{-e}\right) \leq 1$ since $0<X \leq 1$ and $0<Y \leq 1$.

To finish the proof, define the following functions
$t(x, p, \alpha):=\frac{g(x, p)}{g(x, p /(1-\alpha(p-1)))} \quad$ where $g(x, p):=p x \log x-(p x+1-p) \log (p x+1-p)$
for $0 \leq p \leq 1$ and $0 \leq \alpha \leq 1$. Define $t(0, p, \alpha):=\lim _{x \downarrow 0} t(x, p, \alpha)$ and $t(1, p, \alpha):=$ $\lim _{x \rightarrow 1} t(x, p, \alpha)$. It is not hard to verify that $t(x, p, \alpha)$ is continuous with respect to $x$ over $[0, \infty)$ for any fixed $p$ and $\alpha$, and $t\left(\frac{f\left(\tau_{-e}, 1\right)}{f\left(\tau_{-e}, 0\right)}, p_{e}, \alpha\left(\tau_{-e}\right)\right)=\frac{h\left(p_{e}, \tau_{-e}\right)}{h\left(\frac{p_{e}}{1-\alpha\left(\tau_{-e}\right)\left(p_{e}-1\right)}, \tau_{-e}\right)}$. This function admits the following monotonicity property, whose proof is postponed till Section 6.8.3.

Lemma 6.35. For any $0 \leq p \leq 1$ and $0 \leq \alpha \leq 1, t(x, p, \alpha)$ is monotone decreasing in $x$ over $x \geq 0$.

Given this, $t(x, p, \alpha)$ has a lower bound

$$
t(x, p, \alpha) \geq \lim _{x \rightarrow \infty} t(x, p, \alpha)=\frac{(1-\alpha(p-1)) \log p}{\log p-\log (1-\alpha(p-1))}=: C(p, \alpha) .
$$

We remark that the constant $C=C(p, \alpha)$ satisfies

$$
\begin{equation*}
0.5 \leq C(p, \alpha) \leq 2 \tag{6.40}
\end{equation*}
$$

The proof is given in Section 6.8.3, too. Using this fact, we conclude (6.37) by comparing (6.38) with (6.39).

This finishes the proof of Lemma 6.31.

### 6.7 Perfect sampling via coupling from the past

In this section, we give a perfect sampler for the ferromagnetic Ising model with consistent fields. We first give a perfect sampler for the weighted random cluster model, then turn it into a perfect sampler for the Ising model.

Theorem 6.36. There exists a perfect sampling algorithm such that given any weighted random cluster model on graph $G=(V, E)$ with parameters $\mathbf{p}=\left(p_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$, if $0<p_{e}<1$ for all $e \in E$ and $0<\lambda_{v} \leq 1$ for all $v \in V$, the algorithm returns a perfect sample from weighted random cluster models in expected time $C_{1}\left(p_{\min }, p_{\max }\right) N^{4} m^{4} \log n$, where $N=\min \left\{n, \frac{1}{1-\lambda_{\max }}\right\}, C_{1}\left(p_{\min }, p_{\max }\right)=O\left(\frac{1}{\min \left\{p_{\min }, 1-p_{\max }\right\}} \log \frac{1}{\min \left\{p_{\min }, 1-p_{\max }\right\}}\right), \lambda_{\max }=$ $\max _{v \in V} \lambda_{v}, p_{\max }=\max _{e \in E} p_{e}$ and $p_{\min }=\min _{e \in E} p_{e}$.

Furthermore, if there exists $\delta>0$ such that $\lambda_{v} \leq 1-\delta$ for all $v \in V$, then the algorithm runs in time $C_{2}\left(\Delta, \delta, p_{\min }, p_{\max }\right) n^{2} \log ^{2} n$, where the constant $C_{2}\left(\Delta, \delta, p_{\min }, p_{\max }\right)$ is given by $\left(\frac{\Delta}{\delta^{2} \min \left\{p_{\min }, 1-p_{\max }\right\}}\right)^{o\left(\frac{\Delta^{2}}{\delta^{4} \min \left\{p_{\min }, 1-p_{\max }\right\}}\right)}$.

Note that if $p_{e}=0$, we can simply remove $e$, and if $p_{e}=1$, we can contract $e$. Similarly if $\lambda_{v}=0$, we may pin $v$ to 0 and absorb it into its neighbours external fields. Thus for any weighted random cluster model, we can modify it so that it satisfies the condition of Theorem 6.36.

### 6.7.1 Perfect ferromagnetic Ising sampler

We now prove Theorem 6.3. We give the perfect ferromagnetic Ising sampler assuming the algorithm in Theorem 6.36. Let $G=(V, E)$ be a graph. Let $\beta=\left(\beta_{e}\right)_{e \in E}$ and $\lambda=\left(\lambda_{v}\right)_{v \in V}$ be parameters for the Ising model, where $\beta_{e}>1$ for all $e \in E$ and $0<\lambda_{v}<1$ for all $v \in V$. Let $p_{e}=1-\frac{1}{\beta_{e}}$ for all $e \in E$. We first use algorithm in Theorem 6.36 to draw a perfect random sample $\mathcal{S} \subseteq E$ from the weighted random cluster model with parameters $\mathbf{p}$ and $\lambda$. Then we using the Markov chain $\mathcal{P}_{\mathcal{R} \rightarrow I}$ in (6.8) to transform $\mathcal{S}$ into a random Ising configuration $\sigma \in\{0,1\}^{V}$. By Proposition 6.6, since $\mathcal{S} \sim \pi_{\text {wrc }}, \sigma$ is a perfect sample from the Ising model. The running time of the transformation step is $O(n+m)$. Note that

$$
p_{\min }=1-\frac{1}{\beta_{\min }}, \quad 1-p_{\max }=\frac{1}{\beta_{\max }} .
$$

By Theorem 6.36, the total running time is $C_{1} N^{4} m^{4} \log n$ and $C_{2} n^{2} \log ^{2} n$ for all $\lambda_{v} \leq$ $1-\delta$, where

$$
\begin{align*}
& C_{1}=C_{1}\left(\beta_{\min }, \beta_{\max }\right)=O\left(\left(\beta_{\max }+\frac{\beta_{\min }}{\beta_{\min }-1}\right) \log \left(\beta_{\max }+\frac{\beta_{\min }}{\beta_{\min }-1}\right)\right), \\
& C_{2}=C_{2}\left(\Delta, \delta, \beta_{\min }, \beta_{\max }\right)=\left(\frac{\Delta}{\delta^{2}}\left(\beta_{\max }+\frac{\beta_{\min }}{\beta_{\min }-1}\right)\right)^{o\left(\frac{\Delta}{\delta^{4}}\left(\beta_{\max }+\frac{\beta_{\min }-1}{\beta_{\min }-1}\right)\right)} . \tag{6.41}
\end{align*}
$$

### 6.7.2 CFTP for weighted random cluster models

We give a perfect sampler for weighted random cluster models based on coupling form the past (CFTP) applied to the Glauber dynamics. Here is an equivalent definition of the Glauber dynamics. There is a one-to-one correspondence between vectors in $\{0,1\}^{E}$ and subsets in $2^{E}$ (i.e. for any $X \in\{0,1\}^{E}$, let $S_{X}=\left\{e \in E \mid X_{e}=1\right\}$ ). We assume that the Markov chain is defined over the state space $\{0,1\}^{E}$. The Glauber
dynamics starts from an arbitrary subset of edges $X_{0} \in\{0,1\}^{E}$. For the $t$-th transition step, the chain does the following:

- pick an edge $e_{t} \in E$ uniformly at random;
- sample a real number $r_{t} \in[0,1]$ uniformly at random; if $r_{t}<a_{t}$, let $X_{t}=X_{t-1}^{e \leftarrow 1}$; if $r_{t} \geq a_{t}$, let $X_{t}=X_{t-1}^{e \leftarrow 0}$, where $X_{t-1}^{e \leftarrow c}$ satisfies $X_{t-1}^{e \leftarrow c}(E \backslash\{e\})=X_{t-1}(E \backslash\{e\})$ and $X_{t-1}^{e \leftarrow c}(e)=c$, and

$$
\begin{equation*}
a_{t}=a\left(X_{t-1}, e\right):=\frac{\pi_{\mathrm{wrc}}\left(X_{t-1}^{e \leftarrow 1}\right)}{\pi_{\mathrm{wrc}}\left(X_{t-1}^{e \leftarrow 0}\right)+\pi_{\mathrm{wrc}}\left(X_{t-1}^{e \leftarrow 1}\right)} . \tag{6.42}
\end{equation*}
$$

The Glauber dynamics for weighted random cluster models admits a grand monotone coupling. Let $\Omega=\{0,1\}^{E}$. Let $P: \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$ denote the transition matrix of the Glauber dynamics. We use the function $\varphi(\cdot, \cdot)$ to represent each transition step of edge flipping dynamics. For any $t$, given the current configuration $X_{t-1} \in \Omega$, the next configuration can be generated by $X_{t}=\varphi\left(X_{t-1}, U_{t}\right)$, where $U_{t}$ is the randomness used in the $t$-th transition step. Specifically,

$$
U_{t} \sim \mathcal{D} \text { and } U_{t}=\left(e_{t}, r_{t}\right) \in \Omega_{R}=E \times[0,1]
$$

where $\mathcal{D}$ is a distribution such that $e_{t}$ is a uniform random edge in $E, r_{t}$ is a uniform random real number in $[0,1]$, and they are independent. The function $\varphi$ uses the transition rule defined above to map $X_{t-1}$ to a random state $X_{t}=\varphi\left(X_{t-1}, U_{t}\right)$, where the randomness of $X_{t}$ is determined by the randomness of $U_{t} \sim \mathcal{D}$. The function $\varphi(\cdot, \cdot)$ is called a grand coupling of flipping dynamics because

$$
\forall \sigma, \tau \in \Omega, \quad \operatorname{Pr}_{U \sim D}[\varphi(\sigma, U)=\tau]=P(\sigma, \tau) .
$$

Define a partial ordering $\leq$ among all vectors in $\{0,1\}^{E}$ : for any $X, Y \in\{0,1\}^{E}$,

$$
X \leq Y \quad \text { if } X(e) \leq Y(e) \text { for all } e \in E .
$$

Let $X^{\text {min }}=\mathbf{0}$ be the constant 0 vector and $X^{\text {max }}=\mathbf{1}$ be the constant 1 vector, so that $X^{\text {min }} \leq X \leq X^{\text {max }}$ for all $X \in\{0,1\}^{E}$. The next lemma shows that the grand coupling $\varphi$ is monotone with respect to the partial ordering $\leq$.

Lemma 6.37. Suppose $0 \leq p_{e}<1$ for all $e \in E$ and $0<\lambda_{v} \leq 1$ for all $v \in V$. The grand coupling $\varphi$ of the Glauber dynamics for weighted random cluster models is monotone, i.e. for any $\sigma, \tau \in \Omega$ with $\sigma \leq \tau$, any $U \in \Omega_{R}$, it holds that $\varphi(\sigma, U) \leq \varphi(\tau, U)$.

```
Algorithm 6: CFTP of the Glauber dynamics for weighted random cluster
models
    Input: a weighted random cluster model on graph \(G=(V, E)\) with
                parameters \(\lambda=\left(\lambda_{v}\right)_{v \in V}\) and \(\mathbf{p}=\left(p_{e}\right)_{e \in E}\), where \(0<p_{e}<1\) for all
        \(e \in E\) and \(0<\lambda_{v} \leq 1\) for all \(v \in V\).
    Output: a perfect sample \(X \sim \pi_{\text {wrc }}\), where \(\pi_{\text {wrc }}\) is the distribution over
            \(\{0,1\}^{E}\) defined by the input weighted random cluster model.
    generate \(U_{t} \sim \mathcal{D}\) independently for all integers \(t \in(-\infty,-1]\);
    \(T=1 ;\)
    repeat
        \(X^{\text {min }}=\mathbf{0}\) and \(X^{\text {max }}=\mathbf{1} ;\)
        for \(t=-T\) to -1 do
                \(X^{\text {min }} \leftarrow \varphi\left(X^{\text {min }}, U_{t}\right) ;\)
                \(X^{\max } \leftarrow \varphi\left(X^{\max }, U_{t}\right) ;\)
                // \(\varphi\) is the monotone grand coupling in Lemma 6.37
        \(T \leftarrow 2 T\)
    until \(X^{\text {min }}=X^{\text {max }}\);
    return \(X^{\text {min }}\);
```

The proof of Lemma 6.37 is deferred to Section 6.7.3. With the monotone grand coupling $\varphi$, we apply CFTP to the Glauber dynamics for weighted random cluster models in Algorithm 6.

Remark 6.38. In Algorithm 6, infinitely many $U_{t}$ are generated in Line 1. To implement the algorithm, we can first generate $U_{-1}$, and then generate $\left(U_{t}\right)_{-2 T \leq t<-T}$ when updating $T \leftarrow 2 T$.

Let $T_{\mathcal{D}}$ be the time cost for generating a random sample from $\mathcal{D}$. Let $T_{\varphi}$ be the time cost for computing the value of the function $\varphi$. Let $T_{\text {mix }}(\cdot)$ denote the mixing time of the edge flipping dynamics for weighted random cluster models. By the standard result of the CFTP for monotone systems [PW96b] (also see [LP17, Chapter 25]), we have the following proposition about Algorithm 6.

Proposition 6.39 ([PW96b]). Suppose the input weighted random cluster model satisfies $0 \leq p_{e}<1$ for all $e \in E$ and $0<\lambda_{v} \leq 1$ for all $v \in V$. Algorithm 6 returns a perfect sample for the stationary distribution of edge flipping dynamics for weighted random
cluster models, i.e. the distribution $\pi_{\text {wrc }}$. The expected running time of Algorithm 6 is $O\left(\left(T_{\mathcal{D}}+T_{\varphi}\right) T_{\text {mix }}\left(\frac{1}{4 e}\right) \log n\right)$.

Now, we are ready to prove Theorem 6.36.
Proof of Theorem 6.36. By definitions of $\mathcal{D}$ and $\varphi$, it is straightforward to verify that $T_{\mathcal{D}}=O(1)$ and $T_{\varphi}=O(n+m)$. The mixing time can be obtained from Theorem 6.24.

### 6.7.3 Proof of monotonicity

Here we prove Lemma 6.37. Fix $\sigma, \tau \in\{0,1\}^{E}$ such that $\sigma \leq \tau$. Fix $U=(e, r) \in \Omega_{R}$. Let $e=\{u, v\}$. Let $\sigma_{-e}$ and $\tau_{-e}$ denote $\sigma(E \backslash\{e\})$ and $\tau(E \backslash\{e\})$ respectively, and $G_{\sigma}$ and $G_{\tau}$ be the graphs with vertices $V$ and edges in $\sigma_{-e}$ and $\tau_{-e}$ respectively. Note that $G_{\sigma}$ is a subgraph of $G_{\tau}$. We prove the lemma by considering three cases (1) $u, v$ are connected in both $G_{\sigma}$ and $G_{\tau}(2) u, v$ are neither connected in neither $G_{\sigma}$ nor $G_{\tau}$ (3) $u, v$ are connected in $G_{\tau}$ but not in $G_{\sigma}$.

First suppose $u, v$ are connected in both $G_{\sigma}$ and $G_{\tau}$. In this case $a(\sigma, e)=a(\tau, e)=$ $p_{e}$, where $a(\cdot, \cdot)$ is defined in (6.42). The lemma holds trivially.

Next assume $u, v$ are neither connected in neither $G_{\sigma}$ nor $G_{\tau}$. Suppose $u, v$ belong to connected components $C_{1}, C_{2}$ (or $C_{1}^{\prime}, C_{2}^{\prime}$ ) in $G_{\sigma}$ (or $G_{\tau}$ ) respectively. Define

$$
x_{1}^{\sigma}:=\prod_{w \in C_{1}} \lambda_{w}, \quad x_{2}^{\sigma}:=\prod_{w \in C_{2}} \lambda_{w}, \quad x_{1}^{\tau}:=\prod_{w \in C_{1}^{\prime}} \lambda_{w}, \quad x_{2}^{\tau}:=\prod_{w \in C_{2}^{\prime}} \lambda_{w} .
$$

We have

$$
\begin{aligned}
a(\sigma, e) & =\frac{p_{e}\left(1+x_{1}^{\sigma} x_{2}^{\sigma}\right)}{p_{e}\left(1+x_{1}^{\sigma} x_{2}^{\sigma}\right)+\left(1-p_{e}\right)\left(1+x_{1}^{\sigma}\right)\left(1+x_{2}^{\sigma}\right)} \\
a(\tau, e) & =\frac{p_{e}\left(1+x_{1}^{\tau} x_{2}^{\tau}\right)}{p_{e}\left(1+x_{1}^{\tau} x_{2}^{\tau}\right)+\left(1-p_{e}\right)\left(1+x_{1}^{\tau}\right)\left(1+x_{2}^{\tau}\right)} .
\end{aligned}
$$

Since $\lambda_{w} \leq 1$ for all $w \in V, x_{1}^{\sigma} \geq x_{1}^{\tau}$ and $x_{2}^{\sigma} \geq x_{2}^{\tau}$, which implies

$$
\frac{\left(1+x_{1}^{\sigma}\right)\left(1+x_{2}^{\sigma}\right)}{\left(1+x_{1}^{\sigma} x_{2}^{\sigma}\right)} \geq \frac{\left(1+x_{1}^{\tau}\right)\left(1+x_{2}^{\tau}\right)}{\left(1+x_{1}^{\tau} x_{2}^{\tau}\right)}
$$

Hence $a(\sigma, e) \leq a(\tau, e)$, which implies the lemma.
Lastly suppose $u, v$ are connected in $G_{\tau}$ but not in $G_{\sigma}$. Suppose $u, v$ belong to connected components $C_{1}, C_{2}$ in $G_{\sigma}$. Define $x_{1}^{\sigma}$ and $x_{2}^{\sigma}$ in the same way.

$$
a(\sigma, e)=\frac{p_{e}\left(1+x_{1}^{\sigma} x_{2}^{\sigma}\right)}{p_{e}\left(1+x_{1}^{\sigma} x_{2}^{\sigma}\right)+\left(1-p_{e}\right)\left(1+x_{1}^{\sigma}\right)\left(1+x_{2}^{\sigma}\right)}, \quad a(\tau, e)=p_{e} .
$$

Since $\left(1+x_{1}^{\sigma}\right)\left(1+x_{2}^{\sigma}\right) \geq 1+x_{1}^{\sigma} x_{2}^{\sigma}, a(\sigma, e) \leq a(\tau, e)$, which implies the lemma.

### 6.8 Remaining proofs of this chapter

### 6.8.1 Proof of the equivalence result

### 6.8.1.1 Equivalence between Ising and weighted random cluster models

Fix a graph $G=(V, E)$. We first show the first equation in (6.6). Observe that we can decompose the Ising model interaction matrix as

$$
f_{e}^{\text {Ising }}=\left(\begin{array}{cc}
\beta_{e} & 1 \\
1 & \beta_{e}
\end{array}\right)=\left(\begin{array}{cc}
1 & 1 \\
1 & 1
\end{array}\right)+\left(\begin{array}{cc}
\beta_{e}-1 & 0 \\
0 & \beta_{e}-1
\end{array}\right)=: f_{e}^{(0)}+f_{e}^{(1)} .
$$

By definition, $f_{e}^{(1)}$ forces the two endpoints to take the same spin, while $f_{e}^{(0)}$ poses no requirements. In this way, we can perform an extra enumeration over all the assignments over the edges $\tau: E \rightarrow\{0,1\}$, the decompose the effect of $f_{e}^{\text {Ising }}$ into $f_{e}^{(0)}$ and $f_{e}^{(1)}$. The partition function of Ising model then becomes

$$
\begin{align*}
\sum_{\sigma \in\{0,1\}^{V}} \mathrm{wt}_{\mathrm{Ising}}(\sigma) & =\sum_{\sigma \in\{0,1\}^{V}} \prod_{e=(u, v) \in E} f_{e}^{\mathrm{Ising}}(\sigma(u), \sigma(v)) \prod_{u \in V} \lambda_{u}^{\sigma(u)} \\
& =\sum_{\sigma \in\{0,1\}^{V}} \prod_{e=(u, v) \in E}\left(\sum_{\tau(e) \in\{0,1\}} f_{e}^{(\tau(e))}(\sigma(u), \sigma(v))\right) \prod_{u \in V} \lambda_{u}^{\sigma(u)} \\
& =\sum_{\tau \in\{0,1\}^{E}} \sum_{\sigma \in\{0,1\}^{V}} \prod_{e=(u, v) \in E} f_{e}^{(\tau(e))}(\sigma(u), \sigma(v)) \prod_{u \in V} \lambda_{u}^{\sigma(u)} . \tag{*}
\end{align*}
$$

Fix $\tau$. Consider the subgraph $G^{\prime}=(V, S)$ where $S$ is the set of edges assigned to 1 under $\tau$. Each connected component $C \subseteq V$ of $G^{\prime}$ must take the same spin in $\sigma$, otherwise the contribution to the sum is 0 . Let $E_{C} \subseteq S$ denote all the edges in component $C$. The total weight of the component $C$ is $\prod_{e \in E_{C}}\left(\beta_{e}-1\right)\left(1+\prod_{u \in C} \lambda_{u}\right)$. Combining all components yields

$$
\sum_{\sigma \in\{0,1\}^{V}} \prod_{e=(u, v) \in E} f_{e}^{(\tau(e))}(\sigma(u), \sigma(v)) \prod_{u \in V} \lambda_{u}^{\sigma(u)}=\prod_{e \in S}\left(\beta_{e}-1\right) \prod_{C \in \kappa(V, S)}\left(1+\prod_{u \in C} \lambda_{u}\right) .
$$

And hence

$$
\begin{aligned}
(*) & =\sum_{S \subseteq E} \prod_{e \in S}\left(\beta_{e}-1\right) \prod_{C \in \kappa(V, S)}\left(1+\prod_{u \in C} \lambda_{u}\right) \\
& =\left(\prod_{e \in E} \beta_{e}\right) \cdot \sum_{S \subseteq E} \prod_{e \in S}\left(1-\frac{1}{\beta_{e}}\right) \prod_{f \in E \backslash S} \frac{1}{\beta_{f}} \prod_{C \in \kappa(V, S)}\left(1+\prod_{u \in C} \lambda_{u}\right)=Z_{\mathrm{wrc}}(G ; 2 \mathbf{p}, \lambda)
\end{aligned}
$$

by taking $2 p_{e}=1-1 / \beta_{e}$.

### 6.8.1.2 Equivalence between Ising and subgraph-world

To apply Theorem 6.9, we express the Ising model $(G=(V, E) ; \beta, \lambda)$ as a Holant problem. Given an Ising model on graph $G=(V, E)$. We define a bipartite graph $H$ with left part $V_{1}=V$ corresponding to vertices in $G$ and right part $V_{2}=E$ corresponding to edges in $G$. Two vertices $v \in V_{1}$ and $e \in V_{2}$ are adjacent in graph $H$ if $v$ is incident to $e$ in graph $G$. By definition, each edge $e=(u, v)$ in $G$ is decomposed into two half-edges $(v, e)$ and ( $u, e$ ) in graph $H$.

For any vertex $v \in V_{1}$, we force the assignment to its incident half-edges to be equal, and further more, if they are all ones, then we multiply the weight by $\lambda_{v}$. This yields the signature $\left[1,0, \cdots, 0, \lambda_{v}\right]=[1,0]^{\otimes d_{v}}+\lambda_{v}[0,1]^{\otimes d_{v}}$ on each vertex $v$, where $d_{v}$ is the degree of $v$ in $G$. For any edge $e$ in $G$, its signature is $\left[\beta_{e}, 1, \beta_{e}\right.$ ] to model the ferromagnetic Ising interaction. Define

$$
\mathcal{F}_{\text {Ising }}=\left\{[1,0]^{\otimes d_{v}}+\lambda_{v}[0,1]^{\otimes d_{v}} \mid v \in V\right\} \text { and } \mathcal{G}_{\text {Ising }}=\left\{\left[\beta_{e}, 1, \beta_{e}\right] \mid e \in E\right\} .
$$

It is straightforward to verify

$$
\begin{equation*}
\operatorname{Holant}\left(H ; \mathcal{F}_{\text {Ising }} \mid \mathcal{G}_{\text {Ising }}\right)=Z_{\text {Ising }}(G ; \beta, \lambda) \tag{6.43}
\end{equation*}
$$

For subgraph-world models, we define a Holant problem on the same bipartite graph $H$. The signature on each vertex $v$ is defined by $\left[1, \eta_{v}, 1, \eta_{v}, \cdots\right]$, and on each edge $e \in E$, it is defined by [ $1-p_{e}, 0, p_{e}$ ]. Define

$$
\mathcal{F}_{\mathrm{sg}}=\left\{\left[1, \eta_{v}, 1, \eta_{v}, \cdots\right] \mid v \in V\right\} \text { and } \mathcal{G}_{\mathrm{sg}}=\left\{\left[1-p_{e}, 0, p_{e}\right] \mid e \in E\right\} .
$$

It is straightforward to verify

$$
\begin{equation*}
\text { Holant }\left(H ; \mathcal{F}_{\mathrm{sg}} \mid \mathcal{G}_{\mathrm{sg}}\right)=Z_{\mathrm{sg}}(G ; \mathbf{p}, \eta) \tag{6.44}
\end{equation*}
$$

Take $T=\left(\begin{array}{cc}1 & 1 \\ 1 & -1\end{array}\right)$. Let $p_{e}=\frac{1}{2}\left(1-\frac{1}{\beta_{e}}\right)$. It holds that

$$
\left(\boldsymbol{T}^{-1}\right)^{\otimes 2}\left(\beta_{e}, 1,1, \beta_{e}\right)^{\top}=\left(\frac{\beta_{e}+1}{2}, 0,0, \frac{\beta_{e}-1}{2}\right)^{\top}=\beta_{e}\left[\frac{\beta_{e}+1}{2 \beta_{e}}, 0, \frac{\beta_{e}-1}{2 \beta_{e}}\right]=\beta_{e}\left[1-p_{e}, 0, p_{e}\right] .
$$

Let $\eta_{v}=\frac{1-\lambda_{v}}{1+\lambda_{v}}$. We have

$$
\left((1,0)^{\otimes d_{v}}+\lambda_{v}(0,1)^{\otimes d_{v}}\right) \mathbf{T}^{\otimes d_{v}}=(1,1)^{\otimes d_{v}}+\lambda_{v}(1,-1)^{\otimes d_{v}}=\left(1+\lambda_{v}\right)\left[1, \eta_{v}, 1, \eta_{v}, \cdots\right] .
$$

Combining Theorem 6.9, (6.43) and (6.44) with the above, it holds that

$$
Z_{\mathrm{Ising}}(G ; \beta, \lambda)=\left(\prod_{v \in V}\left(1+\lambda_{v}\right)\right)\left(\prod_{e \in E} \beta_{e}\right) Z_{\mathrm{sg}}(G ; \mathbf{p}, \eta)
$$

### 6.8.2 Proof of the adjointness

Proof of Proposition 6.6. Let $D_{\text {Ising }}=\operatorname{diag}\left(\pi_{\text {Ising }}\right)$ and $D_{\text {wrc }}=\operatorname{diag}\left(\pi_{\text {wrc }}\right)$ denote the diagonal matrices induced from vectors $\pi_{\text {Ising }}$ and $\pi_{\text {wrc }}$ respectively. We have

$$
\left\langle f, P_{I \rightarrow \mathcal{R}} g\right\rangle_{\pi_{\text {Ising }}}=f^{T} D_{\text {Ising }} P_{I \rightarrow \mathcal{R}} g \quad \text { and } \quad\left\langle P_{\mathcal{R} \rightarrow I} f, g\right\rangle_{\pi_{\mathrm{wrc}}}=f^{T} P_{\mathcal{R} \rightarrow I}^{T} D_{\mathrm{wrc}} g .
$$

For any $\sigma \in\{0,1\}^{V}$ and $S \subseteq E$, we show that

$$
\left(D_{\text {Ising }} P_{I \rightarrow \mathcal{R}}\right)(\sigma, S)=\left(P_{\mathcal{R} \rightarrow I}^{T} D_{\mathrm{wrc}}\right)(\sigma, S)
$$

Recall $M(\sigma)=\left\{\{u, v\} \in E \mid \sigma_{u}=\sigma_{v}\right\}$. It holds that

$$
\begin{align*}
\left(D_{\text {Ising }} P_{I \rightarrow \mathcal{R}}\right)(\sigma, S) & =\mathbb{I}[S \subseteq M(\sigma)] \cdot \pi_{\text {Ising }}(\sigma) \cdot \prod_{e \in S}\left(1-\frac{1}{\beta_{e}}\right) \prod_{f \in M(\sigma) \backslash S} \frac{1}{\beta_{f}} \\
& =\mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text {Ising }}} \cdot \prod_{v \in V} \lambda_{v}^{\sigma(v)} \prod_{h \in M(\sigma)} \beta_{h} \prod_{e \in S}\left(1-\frac{1}{\beta_{e}}\right) \prod_{f \in M(\sigma) \backslash S} \frac{1}{\beta_{f}} \\
& =\mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\text {Ising }}} \cdot \prod_{v \in V} \lambda_{v}^{\sigma(v)} \prod_{e \in S}\left(\beta_{e}-1\right) \tag{6.45}
\end{align*}
$$

Recall $\kappa(V, S)$ is the set of all connected components of graph $(V, S)$. It holds that

$$
\begin{align*}
& \left(P_{\mathcal{R} \rightarrow I}^{T} D_{\mathrm{wrc}}\right)(\sigma, S)=\mathbb{I}[S \subseteq M(\sigma)] \cdot \pi_{\mathrm{wrc}}(S) \cdot \prod_{C \in \kappa(V, S)} \frac{\prod_{v \in C} \lambda_{v}^{\sigma(v)}}{1+\prod_{v \in C} \lambda_{v}} \\
= & \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\mathrm{wrc}}} \cdot \prod_{e \in S}\left(1-\frac{1}{\beta_{e}}\right) \prod_{f \in E \backslash S} \frac{1}{\beta_{f}} \prod_{C \in \kappa(V, S)}\left(1+\prod_{u \in C} \lambda_{u}\right) \cdot \prod_{C \in \kappa(V, S)} \frac{\prod_{v \in C} \lambda_{v}^{\sigma(v)}}{1+\prod_{v \in C} \lambda_{v}} \\
= & \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\mathrm{wrc}}} \cdot \prod_{e \in S}\left(1-\frac{1}{\beta_{e}}\right) \prod_{f \in E \backslash S} \frac{1}{\beta_{f}} \prod_{v \in V} \lambda_{v}^{\sigma(v)} \\
= & \mathbb{I}[S \subseteq M(\sigma)] \cdot \frac{1}{Z_{\mathrm{wrc}}} \cdot \prod_{h \in E} \frac{1}{\beta_{h}} \prod_{v \in V} \lambda_{v}^{\sigma(v)} \prod_{e \in S}\left(\beta_{e}-1\right) \tag{6.46}
\end{align*}
$$

By Proposition 6.4, we know that

$$
\left(\prod_{e \in E} \beta_{e}\right) Z_{\mathrm{wrc}}=Z_{\mathrm{Ising}}
$$

Combining above equation with (6.45) and (6.46) gives

$$
\left(D_{\text {Ising }} P_{I \rightarrow \mathcal{R}}\right)(\sigma, S)=\left(P_{\mathcal{R} \rightarrow I}^{T} D_{\mathrm{wrc}}\right)(\sigma, S)
$$

### 6.8.3 Proof of analytic lemmata

This section of appendix proves Lemma 6.35 and (6.40).
Proof of Lemma 6.35. The goal is to show $\partial t(x, p, \alpha) / \partial x<0$ for all $x \in(0,1) \cup(1,+\infty)$. The lemma then follows by combining this with continuity.

A straightforward calculation shows that

$$
\frac{\partial t(x, p, \alpha)}{\partial x}=\frac{-(1-\alpha(1-p))(1-p) p}{\left(x p \log x-((1+\alpha)(1-p)+p x) \log \left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)\right)^{2}} s(x, p, \alpha)
$$

where

$$
\begin{aligned}
& s(x, p, \alpha):= \\
& \quad(1+\alpha)(\log x) \log \left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)-\left(\log x+\alpha \log \left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)\right) \log (1+p(x-1)) .
\end{aligned}
$$

This means $\operatorname{sgn}(\partial t(x, p, \alpha) / \partial x)=-\operatorname{sgn}(s(x, p, \alpha))$, and hence we only need to show $s(x, p, \alpha)>0$ whenever $x \in(0,1) \cup(1,+\infty)$.

From now on in this section, we use the notation $A \gtrless_{x} B$ to represent that $A>B$ when $x>1$, and $A<B$ when $0<x<1$. In other words, when $x>1, \gtrless_{x}$ should be read as $>$, and vice versa.

We first claim the following inequalities:

$$
\begin{gather*}
(1+\alpha) \log x-\alpha \log (1+p(x-1)) \gtrless_{x} 0 ;  \tag{6.47}\\
\log \left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right) \gtrless_{x} 0 ;  \tag{6.48}\\
\log (1+(x-1) p) \gtrless_{x} 0 . \tag{6.49}
\end{gather*}
$$

We focus on $s(x, p, \alpha)$ and postpone the proof of these simple inequalities till the end. By collecting terms of $\log \left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)$, one can find that $s(x, p, \alpha)>0$ if and only if

$$
((1+\alpha) \log x-\alpha \log (1+p(x-1))) \log \left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)>(\log x) \log (1+p(x-1)) .
$$

By using (6.47), it is equivalent to show that

$$
\log \left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right) \gtrless_{x} \frac{(\log x) \log (1+p(x-1))}{(1+\alpha) \log x-\alpha \log (1+p(x-1))},
$$

or equivalently, using (6.47)(6.48)(6.49), to show that

$$
\begin{equation*}
\frac{1}{\log (1+(x-1) p)} \gtrless_{x} \frac{\alpha}{1+\alpha} \cdot \frac{1}{\log x}+\frac{1}{1+\alpha} \cdot \frac{1}{\log \left(1+\frac{p(x-1)}{1+\alpha(1-p)}\right)} . \tag{6.50}
\end{equation*}
$$

Note that the following function

$$
u_{x, p}(y):=\frac{1}{\log \left(1+\frac{p(x-1)}{y}\right)}
$$

reveals the essence of (6.50) in the way that (6.50) is equivalent to

$$
\begin{equation*}
u_{x, p}(1) \gtrless_{x} \frac{\alpha}{1+\alpha} \cdot u_{x, p}(p)+\frac{1}{1+\alpha} \cdot u_{x, p}(1-\alpha(p-1)), \tag{6.51}
\end{equation*}
$$

and note that

$$
1=\frac{\alpha}{1+\alpha} \cdot p+\frac{1}{1+\alpha} \cdot(1-\alpha(p-1))
$$

This means (6.51) follows if for fixed $x>1$ (resp., $0<x<1$ ) and $p, u_{x, p}(y)$ is a concave (resp., convex) function over $y \in(p, 2) \supseteq(p, 1-\alpha(p-1))$, which would conclude the proof. We verify this as follows.

A straightforward calculation shows that

$$
\begin{aligned}
& \frac{\mathrm{d}^{2}}{\mathrm{~d} y^{2}} u_{x, p}(y)= \\
& \quad \frac{(x-1) p}{y(y+(x-1) p)^{2} \log ^{3}\left(1+\frac{p(x-1)}{y}\right)}\left(2 \cdot \frac{p(x-1)}{y}-\left(2+\frac{p(x-1)}{y}\right) \log \left(1+\frac{p(x-1)}{y}\right)\right) .
\end{aligned}
$$

It is not hard to verify that

$$
\begin{equation*}
\log \left(1+\frac{p(x-1)}{y}\right) \gtrless_{x} 0 \tag{6.52}
\end{equation*}
$$

which we prove later. With a bit more endeavour, we can also show that

$$
\begin{equation*}
-\left(2 \cdot \frac{p(x-1)}{y}-\left(2+\frac{p(x-1)}{y}\right) \log \left(1+\frac{p(x-1)}{y}\right)\right) \gtrless_{x} 0 \tag{6.53}
\end{equation*}
$$

whose proof is postponed as well. Concavity/Convexity is then established by combining the expression for the second-order derivative, (6.52) and (6.53).

Proof of (6.47), (6.48), (6.49), (6.52), and (6.53). For (6.47), because $\log x \gtrless_{x} 0$, we only need to show

$$
\frac{x}{1+(x-1) p} \gtrless_{x} 1 .
$$

Note that $1+(x-1) p$ is positive. The above is hence equivalent to

$$
(x-1)(1-p) \gtrless_{x} 0,
$$

which is obvious.

All of (6.48), (6.49) and (6.52), after simple calculation, are equivalent to $(x-1) p \gtrless_{x}$ 0 , which is obvious, too.

Finally, we show (6.53). Let $z:=p(x-1) / y$. LHS is then $r(z):=(2+z) \log (1+z)-$ $2 z$. It is not hard to show that $r(z)$ is monotone in $z$ over $z \in(-1,+\infty)$, by observing that $r^{\prime}(z)=\frac{1}{1+z}-1-\log \frac{1}{1+z}$, which is non-negative as $\log x \leq x-1$ for $x>0$. Moreover, $r(0)=0$. Therefore, when $x>1$, we have $z>0$, and (6.53) holds. When $0<x<1$, we have $-1<(x-1) \leq z<0$, and (6.53) holds too.

Proof of (6.40). For convenient reference, the expression of interest is

$$
C(p, \alpha):=\frac{(1-\alpha(p-1)) \log p}{\log p-\log (1-\alpha(p-1))}
$$

Taking derivative with respect to $\alpha$, we get

$$
\frac{\partial}{\partial \alpha} C(p, \alpha)=\frac{(1-p) \log (p)\left(1+\log \left(\frac{p}{1+\alpha(1-p)}\right)\right)}{\left(\log \left(\frac{p}{1+\alpha(1-p)}\right)\right)^{2}}
$$

A simple calculation shows that

- if $p \leq 1 / e$, then $C(p, \alpha)$ is increasing with $\alpha$, and hence lies between $C(p, 0)=1$ and $C(p, 1)=\frac{(2-p) \log p}{\log p-\log (2-p)}$;
- if $1 / e<p<2 /(1+e)$, then $C(p, \alpha)$ is decreasing within $\alpha \in(0,(e p-1) /(1-p))$ and increasing within $\alpha \in((e p-1) /(1-p), 1)$, and hence it lies between the lower bound $C(p,(e p-1) /(1-p))=-e p \log p \geq 2 e \log ((1+e) / 2) /(1+e)>$ 0.90 and the upper bound $\max \{C(p, 0), C(p, 1)\}$; and
- if $p \geq 2 /(1+e)$, then $C(p, \alpha)$ is decreasing, and hence lies between $C(p, 1)=$ $\frac{(2-p) \log p}{\log p-\log (2-p)}$ and $C(p, 0)=1$.

From the case-by-case analysis, it suffices to show that $0.5 \leq \frac{(2-p) \log p}{\log p-\log (2-p)} \leq 2$, which is a simple analytic exercise.



Random cluster $(p, \lambda)$
$P_{\mathcal{R} \rightarrow \text { FKES }}$ w.p. $\frac{1}{1+\lambda^{2}} \cdot \frac{\lambda^{6}}{1+\lambda^{6}}$ 0


FKES
I
$P_{e}^{\downarrow}$ with random $e$




Subgraph world $\left(p^{\prime}, \eta\right)$


Grand model: (6.16)
Random cluster: (6.2)
Subgraph world: (6.4)
Ising model: (6.1)
FKES: (6.32)
$P_{\text {SW }}$ : Section 6.1.2.2
$P_{\text {GlauberRC: }}$ Section 6.1.2.1
$P_{\mathrm{SB}}:$ Section 6.6.2
$p=1-1 / \beta, p^{\prime}=p / 2, \eta=(1-\lambda) /(1+\lambda)$

Figure 6.1: An example involving all the distributions and Markov chains of Chapter 6.

## Chapter 7

## Conclusion and open problems

We conclude this thesis by providing a list of open problems and potential future directions.

### 7.1 Towards sampling local lemma

The first major open problem is to derive a sampling version of the (variable) local lemma, namely an efficient algorithm that generate nearly uniform assignments to variables such that all bad events are avoided, providing the upper bound on the probability for each bad event being $p$ and the degree of dependency being $\Delta$ with

$$
p \Delta^{c} \lesssim 1
$$

for some constant $c$. The algorithmic Lovász local lemma that finds one such assignment in expected polynomial-time reaches $c=1$. However, the hardness result of this thesis concerning hypergraph colourings, together with a previous paper on hypergraph independent sets [ $\left.\mathrm{BGG}^{+} 19\right]$, defies any $c<2$.

Many of the algorithmic techniques used to derive our results can in fact be applied to more general settings, aside from the aforementioned hypergraph colouring problem and independent set problem (monotone $k$-SAT). One ceiling general form that includes these problems is the atomic CSP, in which each constraint is violated by one or very few, say the number of colours, forbidden assignments. The state of the art is an FPTAS that works in the regime with $c=5$ [HWY23a], which includes the general counting $k$-SAT problem. Going beyond into the non-atomic CSP setting, an efficient sampler also exists with $c=7$ [HWY22]. These two algorithm are based on the lazy marginal sampler introduced by Anand and Jerrum [AJ22].

This exponent advances further if we specify to some other problems. For counting hypergraph independent sets, it has already been confirmed that $c=2$ is the point of computational phase transition. However, the state space of the natural Glauber dynamics over all independent sets is connected, which somehow makes this problem significantly easier than many other prototypical problems in the local lemma regime. Perhaps counting hypergraph colourings is the most promising one to establish a sharp computational phase transition at the absence of this property. Though the state-of-art algorithms work for $c=3$, it is conjectured that this problem is tractable up to $c=2$. Formally, we expect the answer to the following question being true:

Question 7.1. Does there exist an FPRAS for counting $q$-colourings in $k$-uniform hypergraphs with maximum degree $\Delta$, when $\Delta \lesssim q^{k /(2+o(1))}$ ?

It is worth mentioning that the local uniformity property is indispensable to all previous works on hypergraph colourings. Specified to the marking/projecting framework, this property is used for establishing both rapid mixing of the projected chain and efficient implementation of it. In comparison, the tight-up-to-constant algorithm for hypergraph independent sets [HSZ19] does not rely on this. It would be an intriguing direction to obtain other algorithmic methods that bypass the connectivity issue.

Our algorithmic results make advancements when the hypergraph is linear. A remark is that our method would still work as long as the overlap of hyperedges is much smaller than $k$. The condition on the parameters will deteriorate slightly but would still be better than those for general hypergraphs. On the other end of the spectrum, if any two intersecting hyperedges intersect at at least $k / 2$ vertices, the algorithm by Guo, Jerrum, and Liu [GJL19] almost matches the hardness result. It is an interesting question how the size of overlaps affects the complexity of these sampling problems, or whether it is possible to improve sampling algorithms via a better use of the overlap information.

As another defect of the marking/projecting approach, it seems hard to obtain a rather tight bound on the maximum degree providing the number of colours $q$ is a fixed small constant. All algorithmic works require some constant lower bound, like $q \geq 5$ as in [HSW21] or $q \geq 650$ as in [FHY21]. THe special case $q=2$ - or in other words, the not-all-equal $k$-SATs ( $k$-NAESATs) - has never been covered. And note that the projection scheme is completely inapplicable. It is another intriguing question to study the threshold on maximum degree as $k$ approaches infinity for its
sampling and approximate counting problem.

Question 7.2. Find $c$ such that there exists an FPRAS for counting the solutions of any $\Delta$-degree $k$-NAESAT with $\Delta \lesssim 2^{k /(c+o(1))}$.

The proof of the hardness result of hypergraph colourings in this thesis still applies to $q=2$ without much change, yielding $c \geq 2$. We conjecture this to be the ground truth of the computation phase transition too.

Resembling the local lemma where instances are subject to the degree restriction, there are also a lot of works studying the behaviour of random instances, where a similar restriction is put upon the density, namely the number of constraints over the number of variables. For example, a random $k$-CNF formula consisting of $n$ variables and $m$ clauses is said to have density $\alpha:=m / n$. The density captures the average degree of each vertex rather than giving a definitive limit. Usually $\alpha$ is fixed to be a constant when studying the random instance, just like what we do for the maximum degree in the local lemma regime. However, if we uniformly draw a random $k$-CNF formula of density $\alpha$ with $n$ vertices, its maximum degree depends on $n$ and hence unbounded by any function of $\alpha$. This forbids the trivial application of Lovász local lemma, posing a well-known challenge to determine the threshold density $\alpha^{*}$ above or below which a random formula is satisfiable with probability 0 or 1 , known as the satisfiability conjecture. This challenge is resolved in a celebrated work by Ding, Sly and Sun [DSS22], giving $\alpha^{*}=2^{k} \ln 2-(1+\ln 2) / 2+o_{k}(1)$ as $k \rightarrow+\infty$.

Similar to the local lemma setting, we can also ask for algorithms searching for a solution and, more related to this thesis, estimating the size of the solution space, providing that the density is below some threshold. The first result that breaks into a "local-lemma-type" density regime is given by Galanis, Goldberg, Guo and Yang with $\alpha \lesssim 2^{k / 300}$. This gets further refined independently by several teams, with density upper bound being $2^{k / 74}$ [CMM23], $2^{k / 52}$ [GGGH22] and $2^{k / 3}$ [HWY23b] respectively.

As a notable property of random $k$-SAT instances, with high probability the sampled formula has overlap at most 2. In light of this, it is possible that our techniques might help with the improvement as the algorithm still works when the overlap is bounded by constants.

Question 7.3. Find $c$ such that there exists an algorithm that samples a uniform solution from a random $k$-CNF formula with density $\alpha \leqq 2^{k / c}$.

### 7.2 Parity issue

Recall from Chapter 4 that our hardness result only applies to even $q$ 's owing to the same technical issue as in the graph case [GŠV15]. Both proofs start with treating the integer optimisation problem over the number of colours of each type $\left(q_{1}, q_{2}, q_{3}\right)$ as a real optimisation problem in which way we can take derivatives with respect to $q_{i}$. To establish the optimality of the triplet $(q / 2, q / 2,0)$, we (1) argue that such maximiser always exists, and (2) show that any other point could never be a maximiser based on the derivatives. This inevitably requires $q$ to be even. On the other hand, for odd $q$ 's, even the real optimisation problem reaches its maximum at $(q / 2, q / 2,0)$, it barely says anything about the original integer version. Take $q=11$ as an instance to illustrate this. The perturbation argument regarding the derivatives hardly tells if $(5,5,1)$ is better than $(4,4,3)$ (subtracting both $q_{1}$ and $q_{2}$ simultaneously), as it only concerns the non-zeroness of the derivatives, disregarding if they are positive or negative. Although we manage to do this in the hypergraph colouring setting at the boundary, namely the stable $(11,0,0)$-fixpoints, it is already very complicated, and a similar analysis of other fixpoints seems out of reach. Let alone we still need to compare $(5,5,1)$ with $(6,5,0)$, where we do not even have a good intuition which one is better.

A potential approach is to look into the main result of [GŠV15] (Proposition 4.12 in this thesis). Due to a technical issue with the so-called small subgraph conditioning method, their theorem requires the dominant phases to be permutation symmetric and unbalanced. It would be helpful to our setting if, for example, one can establish the correctness of this theorem when the dominant phases lie in a constant number of orbits, each being permutation symmetric and unbalanced, rather than merely one.

Formally,
Question 7.4. Is it true that for all integer $q$ and $\Delta$ such that $4 \leq q<\Delta$, there is no FPRAS that approximates the number of proper $q$-colourings for graphs of maximum degree at most $\Delta$, assuming $\mathbf{N P} \neq \mathbf{R P}$ ?

### 7.3 Random cluster model

In Chapter 6, we establish the rapid mixing of both the edge-flipping dynamics and Swendsen-Wang dynamics for the ferromagnetic Ising model with consistent fields on arbitrary graphs. In the special case where the maximum degree of the graph is
bounded, and all fields are bounded away from 1, we further establish the optimal mixing of the edge-flipping dynamics, yielding a mixing time of $O(n \log n)$. This results in an $O(n \log n)$ mixing time of the Swendsen-Wang dynamics under the same setting.

As is discussed in the introduction of Chapter 6, we believe that none of the mixing time bounds here, except that of the edge-flipping dynamics under degree and field restrictions, is tight. Among these gaps, the one for Swendsen-Wang dynamics on general graphs is glaring: in the no-field case, the mixing time upper bound is $O\left(n^{4} m^{3}\right)$ [GJ18], while the best known lower bound is $O\left(n^{0.25}\right)$. Recall where the upper bound comes from:

- A "no-slower" comparison between Swendsen-Wang dynamics and edge-flipping dynamics carries the mixing time of the latter to the former.
- The winding step in the canonical path argument creates 2 open ends, leading to an $\eta^{4}$ factor that becomes $n^{4}$ after the perturbation.
- To bound the spectral gap as in (6.14), we lose a factor of $m$ due to minimum transition probability, and another $m$ due to maximum path length.
- To turn spectral gap into mixing time, we lose a factor $m$ which is the logarithm of minimum probability of all configurations.

Improving any of the items above seems out of reach. However, this does not hurdle it from being an exciting problem for the following reasons.

For the first item, there are several very recent works that give an $\Theta(\log n)$ mixing time of the Swendsen-Wang dynamics (for the more general Potts model), but for special cases of graphs like $\mathbb{Z}^{2}\left[B C P^{+} 21\right]$ or in the tree uniqueness region $\left[\mathrm{BCC}^{+} 22\right]$. It would be exciting to see to which extent we can employ the known techniques, and from where the $O\left(n^{0.25}\right)$ lower bound pops out.

For the next two items, the argument is quite similar as that for approximately counting matchings or the partition function of the subgraph-world model in general graphs. Its sharp mixing time bound is a notorious problem that has not been improved ever since the first time when a polynomial mixing is established [JS89, JS93].

And finally, the improvement of the last point seems possible if one thinks about the relation between the mixing time and a so-called modified log-Sobolev (MLS) constant [BT06]. There, the extra factor from the minimum probability is double logarithm, saving a factor of $m / \log m$. However, that would require us to bound the

MLS constant instead of the spectral gap, which in our work comes from the canonical path method. The implication from the congestion of canonical paths to the MLS constant is unfortunately not known yet. Any progress on this could be considered an exciting breakthrough.

In any case, it boils down to answering the following questions.
Question 7.5. What is the true mixing time of the Swendsen-Wang dynamics on ferromagnetic Ising models on general graphs without the presence of external fields?

Question 7.6. Does the Swendsen-Wang dynamics mix in $O(\log n)$ time on ferromagnetic Ising models on bounded-degree graphs with consistent non-trivial external fields?

The (weighted) random cluster model considered in this thesis corresponds to the special case $q=2$ of the original (yet unweighted) definition. The parameter $q$ there is the factor contributed by each connected components. The equivalence results of partition functions in the random cluster models and the Ising/Potts models carry over to general $q$ 's. A deeper connection to the important Tutte polynomial is also well-known.

The main open problem is to analyse the case where $1<q<2$, and answer the following question.

Question 7.7. Does the random cluster model for $1<q<2$ admit an FPRAS?

So far, rapid/torpid mixing of the edge-flipping Glauber dynamics, together with the computational complexity of approximating the partition function, are known for $q \in[0,1] \cup\{2\} \cup(2,+\infty)$. Listed below are the mixing times of the Glauber dynamics for the random cluster model on a graph of $n$ vertices and $m$ edges providing different $q$ 's:

| $q$ | $t_{\text {mix }}\left(P_{\mathrm{G}}\right)$ |
| :---: | :---: |
| $0 \leq q \leq 1$ | $O\left(m^{2} \log n\right)$ [AOV18, ALOV19] |
| $1<q<2$ | open problem (easy?) |
| $q=2$ | $O\left(n^{4} m^{3}\right)$ [GJ18] |
| $q>2$ | torpid mixing [ $\left.\mathrm{BCF}^{+} 99, \mathrm{GJ} 99\right]$ |



Figure 7.1: The complexity landscape of approximating the Tutte polynomial. The parameter $q$ in the random cluster model is given by $(x-1)(y-1)$ as in the Tutte polynomial. Legends: Green: exact counting in $\mathbf{P}$, including $q=1$, and $(1,1),(-1,-1),(-1,0),(0,-1)$. Blue: FPRASes, including $q=2, x>1$, and $0 \leq q<$ $1, x, y \geq 1$. Yellow: \#BIS-hard, including $q>2$ [GJ12]. Cyan: equivalent to approximately counting perfect matchings; including $q=2,0<x<1$. Red: NP-hard to approximate [GJ08]. White: open.

Figure 7.1 illustrates the known results about the (in)approximability of the more general Tutte polynomial over graphs. It would be interesting to determine the colour of any point that is currently white, including, for example, the notorious point $x=$ $2, y=0$ corresponding to the number of acyclic orientations.

### 7.4 Fine-grained complexity

Till now we assume the computational equivalence between the approximate counting and sampling problems, given that an interreduction does exist. However, when it comes to the study of their fine-grained complexity, such kinds of interreductions are no more satisfying. More specifically, the standard self-reduction [JVV86] exhibits
a quadratic blow-up of the running time in the number of sites. Consider the hardcore model (weighted independent sets) and colourings on bounded-degree graphs containing $n$ vertices. Even if we have an optimal $O(n \log n)$ mixing of single-site dynamics, the running time of the approximate counting algorithm would be $\tilde{O}\left(n^{3}\right)$. This is improved later by Štefankovič, Vempala and Vigoda [ŠVV09] via an adaptive simulated annealing, yielding $\tilde{O}\left(n^{2}\right)$ approximate counting algorithms. Yet on the other hand, no non-trivial fine-grained lower bound on approximate counting problems is known, up to the date when this thesis is written. This leaves a large field of unknowns for us to raid into.

Concretely, the hard-core model with parameter $\lambda$ is a distribution over all independent sets $I$ of a given $\Delta$-degree graph, with each of them taking weight $\lambda^{|I|}$. The partition function is defined accordingly. Estimating the partition function undergoes a very sharp computational phase transition that coincides with the physical phase transition. That is, there is some $\lambda_{c}(\Delta) \sim \mathrm{e} / \Delta$ such that the problem is tractable for all $\lambda<\lambda_{c}$ but NP-hard for all $\lambda>\lambda_{c}$. Up to now, all approximate counting algorithms meet a quadratic barrier even when $\lambda$ is sufficient small, say $\lambda=\Theta\left(1 / \Delta^{C}\right)$ for any constant $C \geq 1$. This leads people to ask the following natural question.

Question 7.8. Does the hard-core model with $\lambda=\Theta\left(1 / \Delta^{C}\right)$ admit a sub-quadratic FPRAS? ${ }^{1}$

In a very recent paper [ $\left.\mathrm{AFF}^{+} 23\right]$, the above question is answered positively but under some restrictions. For example, a new sub-quadratic FPRAS with running time $\tilde{O}\left(n^{1+1 /(2 C-2)}\right)$ is given there for any $1.5<C \leq 2$. However, we still do not know how such ideas can be carried over to other counting problems, like (hyper)graph colourings or hypergraph independent sets in the local lemma regime, even when the maximum degree is further restricted. Therefore, we ask

Question 7.9. Does there exist a sub-quadratic FPRAS that approximate the number of independent sets in a $k$-uniform $\Delta$-degree hypergraph providing $\Delta \lesssim 2^{k / 100}$ ? Does there exist a sub-quadratic FPRAS that approximate the number of $q$-colourings in a $k$-uniform $\Delta$-degree hypergraph providing $\Delta \lesssim q^{k / 100}$ ?

[^15]
## Appendix A

## A proof of \#P-hardness

For completeness, a proof of the \#P-hardness of computing the total variation distance between two product distributions is given in this appendix. Defined below are the computational problems we need.

Name ProdDTV
Instance $2 n$ rationals $p_{1}, p_{2}, \cdots, p_{n}$ and $q_{1}, q_{2}, \cdots q_{n}$.
Output $d_{\mathrm{TV}}(P, Q)$, where $P$ and $Q$ are defined as in Chapter 3, with each distribution there being $P_{i}(0)=1-p_{i}, P_{i}(1)=p_{i}$, and analogously for $Q_{i}$ 's.

## Name \#PMFEQUALS

Instance $n$ rationals $p_{1}, p_{2}, \cdots, p_{n}$, and a rational $v$.
Output The number of $x \in\{0,1\}^{n}$ such that $P(x)=v$, where $P$ is defined as above.

Name \#SubsetProd
Instance $n$ positive rationals $a_{1}, \cdots, a_{n}$, and a rational $T$.
Output The number of subsets $S \subseteq[n]$ such that $\prod_{i \in S} a_{i}=T$.

Name \#PerfectMatching

Instance A graph G.
Output The number of perfect matchings in $G$.

The last problem is the "canonical" \#P-complete problem:

Theorem A. 1 ([Val79]). \#PerfectMatching is \#P-complete.

We then begin our reduction. The notion $\leq_{T}^{p}$ indicates a polynomial-time Turing reduction. All rationals are represented by two strings, one for the numerator and one for the denominator. Note that for any $n$ defined as above, if each numerator and denominator are below $4^{n}$ for example, then the actual size of input is still polynomial in $n$, i.e., $O\left(n^{2}\right)$.

## Lemma A.2. \#PerfectMatching $\leq_{T}^{p} \#$ SubsetProd

Proof. Given an $n$-vertex graph $G=(V, E)$ of $\#$ PerfectMatching where $V=v_{1}, v_{2}, \cdots, v_{n}$, construct another instance of \#SubsetProd as follows.

- Let Prime $(i)$ be the $i$-th prime.
- For any edge $e_{k}=\left(v_{i}, v_{j}\right) \in E$, let $a_{k}:=\operatorname{Prime}(i) \operatorname{Prime}(j)$.
- Let $T:=\prod_{i=1}^{n} \operatorname{Prime}(i)$.

The correctness of this reduction immediately follows from the unique factorisation theorem. The reduction is in polynomial time due to the prime number theorem.

Lemma A.3. \#SUBSETPROD $\leq_{T}^{p} \# P M F E Q U A L S$
Proof. Let $a_{1}, \cdots, a_{n}$ and $T$ be the instance of \#SubsetProd. Construct the instance of \#PMFEQuals as follows.

- For any $i \in[n]$, take $p_{i}:=\frac{a_{i}}{a_{i}+1}$.
- Take $v:=T \cdot \prod_{i=1}^{n}\left(1-p_{i}\right)$.

To validate the reduction, compute that for any $S \subset[n]$ :

$$
\prod_{i \in S} a_{i}=T \Longleftrightarrow \prod_{i \in S} \frac{p_{i}}{1-p_{i}}=T \Longleftrightarrow \frac{\prod_{i \in S} p_{i} \prod_{i \notin S}\left(1-p_{i}\right)}{\prod_{i=1}^{n}\left(1-p_{i}\right)}=T \Longleftrightarrow P(x)=v
$$

where $x_{i}=\mathbb{I}[i \in S]$. The number of bits to represent each number $p_{i}$ blows up by a factor at most 3 , and the number of bits for $T$ is at most the sum of those for $p_{i}$ 's. All the arithmetic operations are hence in polynomial time of the size of input.

Lemma A. 4 ([ $\mathrm{BGM}^{+} 23$, Section 3.1]). \#PMFEQUALS $\leq_{T}^{p} P_{R O D D T V}$
Proof. Suppose the input instance of \#PMFEQUALS is $p_{1}, \cdots, p_{n}, v$ yielding the distribution $P$. The reduction contains two cases depending on $v$ :

Consider the case $v<1 / 2^{n}$ first. Construct a pair of auxiliary distributions $\hat{P}:=$ $\hat{P}_{1} \otimes \cdots \otimes \hat{P}_{n} \otimes \hat{P}_{n+1}$ and $\hat{Q}:=\hat{Q}_{1} \otimes \cdots \otimes \hat{Q}_{n} \otimes \hat{Q}_{n+1}$ over $\{0,1\}^{n+1}$ where

- For $i \in[n], \hat{P}_{i}(1)=p_{i}$ and $\hat{Q}_{i}(1)=1 / 2$.
- $\hat{P}_{n+1}(1)=1$ and $\hat{Q}_{n+1}(1)=v 2^{n}$.

Then it can be verified that

$$
d_{\mathrm{TV}}(\hat{P}, \hat{Q})=\sum_{x: P(x)>v}(P(x)-v)
$$

We then find $\beta>0$ such that

$$
\begin{aligned}
& \beta<\frac{1}{2} \cdot \frac{1-P(x) / v}{1+P(x) / v} \Longleftrightarrow P(x)\left(\frac{1}{2}+\beta\right)<v\left(\frac{1}{2}-\beta\right) \text { for any } x \text { such that } P(x)<v ; \\
& \beta<\frac{1}{2} \cdot \frac{1-v / P(x)}{1+v / P(x)} \Longleftrightarrow P(x)\left(\frac{1}{2}-\beta\right)>v\left(\frac{1}{2}+\beta\right) \text { for any } x \text { such that } P(x)>v .
\end{aligned}
$$

Note that this $\beta$ can be found without much computation. The number of bits to represent any $P(x) / v$ or $v / P(x)$ is bounded by some polynomial $O\left(n^{C}\right)$. Then choose $1 / \beta$ to be an integer that cannot be represented by $O\left(n^{C}\right)$ bits as a binary string. This can always been done in polynomial time.

Then construct another pair of auxiliary distributions $P^{\prime}:=P_{1}^{\prime} \otimes \cdots \otimes P_{n}^{\prime} \otimes P_{n+1}^{\prime} \otimes$ $P_{n+2}^{\prime}$ and $Q^{\prime}:=Q_{1}^{\prime} \otimes \cdots \otimes Q_{n}^{\prime} \otimes Q_{n+1}^{\prime} \otimes Q_{n+2}^{\prime}$ over $\{0,1\}^{n+2}$ where

- For $i \in[n], P_{i}^{\prime}(1)=\hat{P}_{i}(1)=p_{i}$.
- $P_{n+1}^{\prime}(1)=\hat{P}_{n+1}(1)=1$.
- $P_{n+2}^{\prime}(1)=1 / 2+\beta$.
- For $i \in[n], Q_{i}^{\prime}(1)=\hat{Q}_{i}(1)=1 / 2$.
- $Q_{n+1}^{\prime}(1)=\hat{Q}_{n+1}(1)=v 2^{n}$.
- $Q_{n+2}^{\prime}(1)=1 / 2-\beta$.

One can verify the following which concludes the reduction for the case when $v<1 / 2^{n}$ :

$$
|\{x: P(x)=v\}|=\frac{1}{2 \beta v}\left(d_{\mathrm{TV}}\left(P^{\prime}, Q^{\prime}\right)-d_{\mathrm{TV}}(\hat{P}, \hat{Q})\right)
$$

This is because ( $x$ 's in the summations below are always over $\{0,1\}^{n}$ )

$$
\begin{aligned}
d_{\mathrm{TV}}\left(P^{\prime}, Q^{\prime}\right)= & \sum_{x} \max \left\{0, P(x)\left(\frac{1}{2}-\beta\right)-v\left(\frac{1}{2}+\beta\right)\right\} \\
& +\sum_{x} \max \left\{0, P(x)\left(\frac{1}{2}+\beta\right)-v\left(\frac{1}{2}-\beta\right)\right\} \quad \text { (By definition) } \\
= & \sum_{x: P(x)>v}\left(P(x)\left(\frac{1}{2}-\beta\right)-v\left(\frac{1}{2}+\beta\right)\right) \\
& +\sum_{x: P(x) \geq v}\left(P(x)\left(\frac{1}{2}+\beta\right)-v\left(\frac{1}{2}-\beta\right)\right) \\
= & \sum_{x: P(x)>v}(P(x)-v) \\
& +2 \beta v|\{x: P(x)=v\}|
\end{aligned}
$$

where ( $\boldsymbol{\infty}$ ) is due to the previous choice of $\beta$.
The other case $v \geq 2^{-n}$ can be proved in the same way, but with a slight different choice of the distributions. More concretely, the construction is identical to the one above, but with $P_{n+1}^{\prime}(1)=\hat{P}_{n+1}(1)=1 /\left(v 2^{n}\right)$ and $Q_{n+1}^{\prime}(1)=\hat{Q}_{n+1}(1)=1$.

This chain of reductions implies
Theorem A.5. PRODDTV is \#P-hard.

## Bibliography

[AFF ${ }^{+}$23] Konrad Anand, Weiming Feng, Graham Freifeld, Heng Guo, and Jiaheng Wang. Approximate counting for spin systems in sub-quadratic time. CoRR, abs/2306.14867, 2023.
[AJ22] Konrad Anand and Mark Jerrum. Perfect sampling in infinite spin systems via strong spatial mixing. SIAM fournal on Computing, 51(4):12801295, 2022.
[AJK ${ }^{+}$21] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence II: Optimal sampling and concentration via restricted modified log-Sobolev inequalities. arXiv, abs/2111.03247, 2021.
[AK00] Paola Alimonti and Viggo Kann. Some APX-completeness results for cubic graphs. Theor. Comput. Sci., 237(1-2):123-134, 2000.
[Alo91] Noga Alon. A parallel algorithmic version of the local lemma. Random Struct. Algorithms, 2(4):367-378, 1991.
[ALO20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In FOCS, pages 1319-1330. IEEE, 2020.
[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In STOC, pages 1-12. ACM, 2019.
[AOV18] Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids. In FOCS, pages 35-46. IEEE Computer Society, 2018.
[ $\left.\mathrm{BCC}^{+} 22\right]$ Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: Coupling, spectral independence, and entropy factorization. In SODA, pages 36703692. SIAM, 2022.
[ $\mathrm{BCF}^{+} 99$ ] Christian Borgs, Jennifer T. Chayes, Alan M. Frieze, Jeong Han Kim, Prasad Tetali, Eric Vigoda, and Van H. Vu. Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics. In FOCS, pages 218-229. IEEE Computer Society, 1999.
[BCKL13] Christian Borgs, Jennifer Chayes, Jeff Kahn, and László Lovász. Left and right convergence of graphs with bounded degree. Random Struct. Algorithms, 42(1):1-28, 2013.
[ $\mathrm{BCP}^{+}$21] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen-Wang dynamics on $\mathbb{Z}^{d}$. In STOC, pages 1551-1564. ACM, 2021.
[BDK08] Magnus Bordewich, Martin E. Dyer, and Marek Karpinski. Path coupling using stopping times and counting independent sets and colorings in hypergraphs. Random Struct. Algorithms, 32(3):375-399, 2008.
[Bec91] József Beck. An algorithmic approach to the Lovász local lemma. I. Random Struct. Algorithms, 2(4):343-366, 1991.
[ $\left.\mathrm{BGG}^{+} 19\right]$ Ivona Bezáková, Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Daniel Štefankovič. Approximation via correlation decay when strong spatial mixing fails. SIAM 7. Comput., 48(2):279-349, 2019.
[ $\mathrm{BGM}^{+}$23] Arnab Bhattacharyya, Sutanu Gayen, Kuldeep S. Meel, Dimitrios Myrisiotis, Aduri Pavan, and N. V. Vinodchandran. On approximating total variation distance. In IFCAI, 2023. To appear.
[BT06] Sergey G. Bobkov and Prasad Tetali. Modified logarithmic Sobolev inequalities in discrete settings. f. Theoret. Probab., 19(2):289-336, 2006.
[CCGL12] Jin-Yi Cai, Xi Chen, Heng Guo, and Pinyan Lu. Inapproximability after uniqueness phase transition in two-spin systems. In COCOA, volume 7402 of Lecture Notes in Computer Science, pages 336-347. Springer, 2012.
[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for Markov chains. pages 110-122, 2022.
[CFMR96] Colin Cooper, Alan Frieze, Michael Molloy, and Bruce Reed. Perfect matchings in random $r$-regular, $s$-uniform hypergraphs. Combinatorics, Probability and Computing, 5(1):1-14, 1996.
[CFYZ21] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Rapid mixing of glauber dynamics via spectral independence for all degrees. In 62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022, pages 137-148. IEEE, 2021.
[CFYZ22] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In FOCS, pages 588-599. IEEE, 2022.
[CGM21] Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-Sobolev inequalities for strongly log-concave distributions. Ann. Probab., 49(1):506525, 2021.
[CGŠV21] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In SODA, pages 1548-1557. SIAM, 2021.
[CGSV22] Zongchen Chen, Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Sampling colorings and independent sets of random regular bipartite graphs in the non-uniqueness region. In SODA, pages 2198-2207. SIAM, 2022.
[CLV21a] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: entropy factorization via high-dimensional expansion. In STOC, pages 1537-1550. ACM, 2021.
[CLV21b] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Spectral independence via stability and applications to Holant-type problems. In FOCS, pages 149-160. IEEE, 2021.
[CLX11] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Computational complexity of Holant problems. SIAM 7. Comput., 40(4):1101-1132, 2011.
[CMM23] Zongchen Chen, Nitya Mani, and Ankur Moitra. From algorithms to connectivity and back: Finding a giant component in random $k$-SAT. In SODA, pages 3437-3470. SIAM, 2023.
[CP21] Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. Comm. Math. Phys., 388(2):793-818, 2021.
[CS00] Artur Czumaj and Christian Scheideler. Coloring nonuniform hypergraphs: A new algorithmic approach to the general Lovász local lemma. Random Struct. Algorithms, 17(3-4):213-237, 2000.
[CST01] Pierluigi Crescenzi, Riccardo Silvestri, and Luca Trevisan. On weighted vs unweighted versions of combinatorial optimization problems. Inf. Comput., 167(1):10-26, 2001.
[CZ23] Xiaoyu Chen and Xinyuan Zhang. A near-linear time sampler for the Ising model with external field. In SODA, pages 4478-4503. SIAM, 2023.
[DD20] Andreas Darmann and Janosch Döcker. On a simple hard variant of Not-All-Equal 3-Sat. Theor. Comput. Sci., 815:147-152, 2020.
[DFJ02] Martin E. Dyer, Alan M. Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. SIAM 7. Comput., 31(5):1527-1541, 2002.
[DSS22] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large $k$. Annals of Mathematics, 196(1):1-388, 2022.
[Dye03] Martin E. Dyer. Approximate counting by dynamic programming. In STOC, pages 693-699. ACM, 2003.
[EHK98] Thomas Emden-Weinert, Stefan Hougardy, and Bernd Kreuter. Uniquely colourable graphs and the hardness of colouring graphs of large girth. Combin. Probab. Comput., 7(4):375-386, 1998.
[EL75] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, pages 609-627. Colloq. Math. Soc. János Bolyai, Vol. 10. North-Holland, Amsterdam, 1975.
[ES88] Robert G. Edwards and Alan D. Sokal. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. Phys. Rev. D (3), 38(6):2009-2012, 1988.
[FA17] Alan M. Frieze and Michael Anastos. Randomly coloring simple hypergraphs with fewer colors. Inf. Process. Lett., 126:39-42, 2017.
[FGJW23] Weiming Feng, Heng Guo, Mark Jerrum, and Jiaheng Wang. A simple polynomial-time approximation algorithm for the total variation distance between two product distributions. TheoretiCS, 2, 2023.
[FGW22] Weiming Feng, Heng Guo, and Jiaheng Wang. Improved bounds for randomly colouring simple hypergraphs. In RANDOM, volume 245 of LIPIcs, pages 25:1-25:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. (full version in arXiv:2202.05554).
[ $\mathrm{FGW}^{+}$23a] Weiming Feng, Heng Guo, Chunyang Wang, Jiaheng Wang, and Yitong Yin. Towards derandomising Markov chain Monte Carlo. In FOCS. IEEE, 2023. To appear.
[FGW23b] Weiming Feng, Heng Guo, and Jiaheng Wang. Swendsen-wang dynamics for the ferromagnetic ising model with external fields. Information and Computation, 294:105066, 2023.
[FGYZ21a] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Fast sampling and counting $k$-SAT solutions in the local lemma regime. $\mathcal{F}$. $A C M$, 68(6):Art. 40, 42, 2021.
[FGYZ21b] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the Boolean domain. In SODA, pages 1558-1577. SIAM, 2021.
[FHY21] Weiming Feng, Kun He, and Yitong Yin. Sampling constraint satisfaction solutions in the local lemma regime. In STOC, pages 1565-1578. ACM, 2021.
[FK72] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. I. Introduction and relation to other models. Physica, 57:536-564, 1972.
[FM11] Alan M. Frieze and Páll Melsted. Randomly coloring simple hypergraphs. Inf. Process. Lett., 111(17):848-853, 2011.
[FM13] Alan Frieze and Dhruv Mubayi. Coloring simple hypergraphs. 7. Combin. Theory Ser. B, 103(6):767-794, 2013.
[GG16] Andreas Galanis and Leslie Ann Goldberg. The complexity of approximately counting in 2 -spin systems on $k$-uniform bounded-degree hypergraphs. Information and Computation, 251:36-66, 2016.
[GGGH22] Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Andrés HerreraPoyatos. Fast sampling of satisfying assignments from random k-SAT. CoRR, abs/2206.15308, 2022.
[GGGY21] Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Kuan Yang. Counting solutions to random CNF formulas. SIAM 7. Comput., 50(6):1701-1738, 2021.
[GGW22] Andreas Galanis, Heng Guo, and Jiaheng Wang. Inapproximability of counting hypergraph colourings. ACM Trans. Comput. Theory, 14(3-4):133, 2022.
[GJ99] Vivek Gore and Mark Jerrum. The Swendsen-Wang process does not always mix rapidly. Journal of Statistical Physics, 97:67-86, 1999.
[GJ07a] Leslie Ann Goldberg and Mark Jerrum. The complexity of ferromagnetic Ising with local fields. Comb. Probab. Comput., 16(1):43-61, 2007.
[GJ07b] Geoffrey R. Grimmett and Svante Janson. Random even graphs. Electronic fournal of Combinatorics, 16, 102007.
[GJ08] Leslie Ann Goldberg and Mark Jerrum. Inapproximability of the Tutte polynomial. Inf. Comput., 206(7):908-929, 2008.
[GJ12] Leslie Ann Goldberg and Mark Jerrum. Approximating the partition function of the ferromagnetic potts model. F. ACM, 59(5):25:1-25:31, 2012.
[GJ18] Heng Guo and Mark Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. Ann. Appl. Probab., 28(2):1292-1313, 2018.
[GJL19] Heng Guo, Mark Jerrum, and Jingcheng Liu. Uniform sampling through the Lovász local lemma. ․ ACM, 66(3):18:1-18:31, 2019.
[GLLZ19] Heng Guo, Chao Liao, Pinyan Lu, and Chihao Zhang. Counting hypergraph colorings in the local lemma regime. SIAM 7. Comput., 48(4):13971424, 2019.
[Gri06] Geoffrey R. Grimmett. The random-cluster model, volume 333. Springer Science \& Business Media, 2006.
[GST16] Heidi Gebauer, Tibor Szabó, and Gábor Tardos. The local lemma is asymptotically tight for SAT. 7. ACM, 63(5):43:1-43:32, 2016.
[GŠV15] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. $\mathcal{F}$. ACM, 62(6):50, 2015.
[GŠV16] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models. Combin. Probab. Comput., 25(4):500-559, 2016.
[HS07] Thomas P. Hayes and Alistair Sinclair. A general lower bound for mixing of single-site dynamics on graphs. Ann. Appl. Probab., 17(3):931-952, 2007.
[HSS11] Bernhard Haeupler, Barna Saha, and Aravind Srinivasan. New constructive aspects of the Lovász local lemma. $\mathcal{F}$. $A C M, 58(6): 28,2011$.
[HSW21] Kun He, Xiaoming Sun, and Kewen Wu. Perfect sampling for (atomic) Lovász local lemma. arXiv, abs/2107.03932, 2021.
[HSZ19] Jonathan Hermon, Allan Sly, and Yumeng Zhang. Rapid mixing of hypergraph independent sets. Random Struct. Algorithms, 54(4):730-767, 2019.
[Hub98] Mark Huber. Exact sampling and approximate counting techniques. In STOC, pages 31-40. ACM, 1998.
[Hub15] Mark Huber. Approximation algorithms for the normalizing constant of Gibbs distributions. Ann. Appl. Probab., 25(2):974-985, 2015.
[HWY22] Kun He, Chunyang Wang, and Yitong Yin. Sampling Lovász local lemma for general constraint satisfaction solutions in near-linear time. In FOCS, pages 147-158. IEEE, 2022.
[HWY23a] Kun He, Chunyang Wang, and Yitong Yin. Deterministic counting Lovász local lemma beyond linear programming. In SODA, pages 33883425. SIAM, 2023.
[HWY23b] Kun He, Kewen Wu, and Kuan Yang. Improved bounds for sampling solutions of random CNF formulas. In SODA, pages 3330-3361. SIAM, 2023.
[JPV21a] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. On the sampling Lovász local lemma for atomic constraint satisfaction problems. arXiv, abs/2102.08342, 2021.
[JPV21b] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. Towards the sampling Lovász local lemma. In FOCS, pages 173-183. IEEE, 2021.
[JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. SIAM f. Comput., 18(6):1149-1178, 1989.
[JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. SIAM 7. Comput., 22(5):1087-1116, 1993.
[JVV86] Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. Theor. Comput. Sci., 43:169-188, 1986.
[KKLP97] Viggo Kann, Sanjeev Khanna, Jens Lagergren, and Alessandro Panconesi. On the hardness of approximating max $k$-cut and its dual. Chic. 7. Theor. Comput. Sci., 1997, 1997.
[Kol18] Vladimir Kolmogorov. A faster approximation algorithm for the Gibbs partition function. In COLT, pages 228-249. PMLR, 2018.
[KST93] Jan Kratochvíl, Petr Savický, and Zsolt Tuza. One more occurrence of variables makes satisfiability jump from trivial to NP-complete. SIAM 7. Comput., 22(1):203-210, 1993.
[LLY13] Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In SODA, pages 67-84. SIAM, 2013. Full version from arXiv at abs/1111.7064.
[LNNP14] Yun Long, Asaf Nachmias, Weiyang Ning, and Yuval Peres. A power law of order $1 / 4$ for critical mean field Swendsen-Wang dynamics. Mem. Amer. Math. Soc., 232(1092):vi+84, 2014.
[LP17] David A. Levin and Yuval Peres. Markov chains and mixing times. American Mathematical Society, Providence, RI, 2017. Second edition of [ MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.
[Moi19] Ankur Moitra. Approximate counting, the Lovász local lemma, and inference in graphical models. F. ACM, 66(2):10:1-10:25, 2019.
[Mos09] Robin A. Moser. A constructive proof of the Lovász local lemma. In STOC, pages 343-350. ACM, 2009.
[MR98] Michael Molloy and Bruce A. Reed. Further algorithmic aspects of the local lemma. In STOC, pages 524-529. ACM, 1998.
[MSW07] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Fast mixing for independent sets, colorings, and other models on trees. Random Structures \& Algorithms, 31(2):134-172, 2007.
[MT10] Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. F. ACM, 57(2):11, 2010.
[MWW09] Elchanan Mossel, Dror Weitz, and Nicholas Wormald. On the hardness of sampling independent sets beyond the tree threshold. Probability Theory and Related Fields, 143(3-4):401-439, 2009.
[ $\mathrm{PJG}^{+}$17] Sejun Park, Yunhun Jang, Andreas Galanis, Jinwoo Shin, Daniel Štefankovič, and Eric Vigoda. Rapid mixing Swendsen-Wang sampler for stochastic partitioned attractive models. In AISTATS, volume 54 of Proceedings of Machine Learning Research, pages 440-449. PMLR, 2017.
[PP19] Konstantinos Panagiotou and Matija Pasch. Satisfiability thresholds for regular occupation problems. In ICALP, volume 132 of LIPIcs, pages 90:190:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
[PW96a] James G. Propp and David B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. Algorithms, 9(1-2):223-252, 1996.
[PW96b] James G. Propp and David B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. Algorithms, 9(1-2):223-252, 1996.
[QW23] Guoliang Qiu and Jiaheng Wang. Inapproximability of counting independent sets in linear hypergraphs. Information Processing Letters, page 106448, 2023.
[QWZ22] Guoliang Qiu, Yanheng Wang, and Chihao Zhang. A perfect sampler for hypergraph independent sets. In ICALP, volume 229 of LIPIcs, pages 103:1-103:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.
[RW99] Dana Randall and David B. Wilson. Sampling spin configurations of an Ising system. In SODA, pages 959-960. ACM/SIAM, 1999.
[Sin92] Alistair Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. Comb. Probab. Comput., 1:351-370, 1992.
[Sly10] Allan Sly. Computational transition at the uniqueness threshold. In FOCS, pages 287-296. IEEE Computer Society, 2010.
[Sri08] Aravind Srinivasan. Improved algorithmic versions of the Lovász local lemma. In SODA, pages 611-620. SIAM, 2008.
[SS14] Allan Sly and Nike Sun. Counting in two-spin models on $d$-regular graphs. Ann. Probab., 42(6):2383-2416, 2014.
[Sta99] Richard P. Stanley. Enumerative combinatorics. Vol. 2, volume 62 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 1999.
[ŠVV09] Daniel Štefankovič, Santosh Vempala, and Eric Vigoda. Adaptive simulated annealing: A near-optimal connection between sampling and counting. F. ACM, 56(3):18, 2009.
[SW87] Robert Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. Phys. Rev. Lett., 58:86-88, 1987.
[Ull14] Mario Ullrich. Swendsen-Wang is faster than single-bond dynamics. SIAM 7. Discret. Math., 28(1):37-48, 2014.
[Val79] Leslie G. Valiant. The complexity of enumeration and reliability problems. SIAM 7. Comput., 8(3):410-421, 1979.
[Val08] Leslie G. Valiant. Holographic algorithms. SIAM 7. Comput., 37(5):15651594, 2008.
[vdW41] Bartel L van der Waerden. Die lange reichweite der regelmässigen atomanordnung in mischkristallen. Zeitschrift für Physik, 118(7):473-488, 1941.


[^0]:    ${ }^{1}$ Here we treat $k$ and $\Delta$ as fixed constants, and the problem is parameterised by them. In other words, each pair of ( $k, \Delta$ ) defines a computation problem.
    ${ }^{2}$ In the spirit of the above footnote, this should be interpret as: for any constant $k$ and $\Delta$ such that the condition holds, the $(k, \Delta)$-SAT problem is hard. The input is still bounded-degree, despite the condition on $\Delta$ being a lower bound.

[^1]:    ${ }^{3}$ We remark here that the $\Delta \lesssim q^{k / 2}$ hardness bound heavily relies on the overlap being half of each hyperedge, which is very far from being linear.

[^2]:    ${ }^{1}$ For completeness, we provide a proof in Appendix A.

[^3]:    ${ }^{2}$ The running time remains polynomial under bit complexity.
    ${ }^{3}$ For example, consider $q=2, n=2$, and the two distributions are given by $P_{1}(1)=0.5, P_{2}(1)=0.5$; $Q_{1}(1)=0.4, Q_{2}(1)=0.3$. Then the discrepancy of the optimal coupling is 0.20 , while that of the greedy coupling is 0.28 .

[^4]:    ${ }^{1}$ A linear hypergraph in this paper actually corresponds to a configuration without 4-cycles in the context of [PP19, Theorem 2.4] (where one should plug in $\ell=2$ ), or a hypergraph without 2-cycles in [CFMR96].

[^5]:    ${ }^{2}$ Antiferromagnetism amounts to checking that $\boldsymbol{B}$ has all but one of its eigenvalues negative; it is not hard to see that $\boldsymbol{B}$ has -1 as an eigenvalue by multiplicity $q-1$, and therefore using trace/determinant we see that the other two eigenvalues have sum equal to $q-1+t^{2}$ and product $-t^{2}$. Ergodicity amounts to the fact that $\boldsymbol{B}$ is irreducible and aperiodic.

[^6]:    ${ }^{3}$ Technically, [GŠV15, Theorem 1.5] demands the assumption $\mathbf{N P} \neq \mathbf{R P}$ but that is merely to exclude randomised algorithms, the reduction itself is deterministic.

[^7]:    ${ }^{4}$ Here, and elsewhere, we use the notation $x_{i} \propto y_{i}$ for $i \in[\bar{q}]$ to denote that $x_{i}=A y_{i}$ for $i \in[\bar{q}]$, for some arbitrary $A$.

[^8]:    ${ }^{5}$ Any permutation over $q_{1}, q_{2}, q_{3}$ is considered equivalent. E.g., $(q / 2, q / 2,0)$ and $(q / 2,0, q / 2)$ are regarded as the same type.

[^9]:    ${ }^{6}$ In this example, $R_{3}$ and $C_{3}$ can be arbitrarily assigned because they do not affect the function $\overline{\Phi^{S}}$. Yet later on in Section 4.3.5, we might need to specify the assignments for the sake of "analytic-continuation"-style properties.

[^10]:    ${ }^{7} x_{M}$ is well-defined. This follows from the fact that $\left(x^{d}-1\right)^{d} f_{1}\left(x, y_{M}\right)$ is a non-zero polynomial with respect to $x$, thus having a finite number of zeros.

[^11]:    ${ }^{1}$ An event is atomic if each variable it depends on must take one particular value. In discrete spaces, any event can be decomposed into atomic ones.

[^12]:    ${ }^{1}$ The random cluster model has a parameter $q>0$. The Ising model corresponds to the case of $q=2$.
    ${ }^{2}$ Peres further conjectured that the sharp mixing time bound is $O\left(|V|^{1 / 4}\right)$.

[^13]:    ${ }^{3}$ Holant problems can also be defined for not necessarily bipartite graphs, but we do not need those here.

[^14]:    ${ }^{4}$ In [Ull14], Ullrich proved this for general random cluster models with an arbitrary $q \geq 1$, but when $q \neq 2$ that model cannot be easily translated to the notation we use.

[^15]:    ${ }^{1}$ FPRAS already requires a polynomial dependency on $1 / \varepsilon$ where $\varepsilon$ is the multiplicative error. We omit this from the running time.

