# Hamilton Cycles in Large Graphs and Hypergraphs: Existence and Counting

by

Stephen Gould

A thesis submitted to the University of Birmingham for the degree of Doctor of Philosophy in Pure Mathematics

Supervisor: Professor Deryk Osthus

Co-supervisor: Professor Daniela Kühn

School of Mathematics

College of Engineering and Physical Sciences

University of Birmingham

December 2022

## Abstract

The study of Hamilton cycles forms a central part of classical graph theory. In this thesis we present our contribution to the modern research on this topic. In Chapter 2, we prove that $k$-uniform hypergraphs satisfying a 'Dirac-like' condition on the minimum $(k-1)$-degree contain many of a natural hypergraph analogue of a Hamilton cycle. In Chapter 3, we show that almost all optimal edge-colourings of $K_n$ admit a Hamilton path whose edges all have distinct colours; that is, a *rainbow* Hamilton path. If $n$ is odd, we show that one is further able to find a rainbow Hamilton cycle. Chapter 4 is given to the proof that almost all optimal colourings of a directed analogue of $K_n$ we call $\overleftrightarrow{K_n}$ admit many rainbow directed Hamilton cycles; equivalently, almost all $n \times n$ Latin squares contain many structures we call 'Hamilton transversals'. Finally, in Chapter 5 we introduce an upcoming result which combines the Rödl Nibble with the Polynomial Method to substantially improve upon known results on the size of matchings in almost-regular hypergaphs. We also state an application of this result to the problem of finding rainbow almost-Hamilton directed cycles in any optimal colouring of $\overleftrightarrow{K_n}$.

# ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, Professors Deryk Osthus and Daniela Kühn. You put your faith in me, gave me good ideas to explore, and put me in touch with fantastic academics who would become my other co-authors. I would also like to acknowledge those co-authors, namely Stefan Glock, Felix Joos, and Tom Kelly. To all of the people mentioned above, your ideas, feedback, and different individual styles each helped to shape the papers presented in this thesis, but also the mathematician I have become.

Next, I want to mention my appreciation of my housemate and best friend Sam. You have been unfailingly supportive at every stage. Finally, my parents, neither of whom lived to see me embark on this PhD. Your sacrifices, your time, and your love gave me my love of life and my love of mathematics. I wish you could have seen this, and I hope that you would be proud.

# CONTENTS

# CHAPTER 1

# INTRODUCTION

## 1.1 Brief basics and history

A *cycle* $C$ is a graph whose vertices can be labelled, say $V(C) = \{v_1, v_2, \dots, v_k\}$, with $k \geq 3$, such that $E(C) = \{v_i v_{i+1} \colon i \in [k-1]\} \cup \{v_k v_1\}$. Here, $[n] \coloneqq \{1, 2, \dots, n\}$ for $n \in \mathbb{N}$. A *Hamilton* cycle (equivalently *spanning* cycle) in a graph $G$ is a cycle $C \subseteq G$ such that $V(C) = V(G)$. If $G$ contains a Hamilton cycle then $G$ is said to be *Hamiltonian*. Hamilton cycles were first studied by Kirkman [84] in 1856, but became named after Sir William Rowan Hamilton, who in 1857 described a game inspired by a group theory problem, with the aim of the game being for one player to find a Hamilton cycle in the dodecahedron graph, subject to containing a specified path on 5 of the vertices, this path being given by the other player.

Ideally, given any graph $G$ on any number of vertices, one would like to be able to efficiently determine if $G$ does or does not contain a Hamilton cycle. In computability terms, this decision problem is usually called 'HAM'. It is well-known that HAM is an NP-complete problem [34]. Loosely speaking, this means that firstly, if we are given a permutation of $V(G)$ for some $G$, then we can efficiently determine whether this permutation corresponds to a Hamilton cycle of $G$ or not. If yes, then output $G \in$ HAM; otherwise, guess another permutation. If there are no remaining unchecked permutations, then output $G \notin$ HAM. Secondly, HAM is at least as 'hard' as any other problem which can be resolved with this strategy. From a computational point of view, trying out all permutations of $V(G)$ is not a viable approach. However, if one can find a classification of the graphs that contain a Hamilton cycle, in terms of some set of conditions which are easier to check, then not only does one prove that one can efficiently determine whether or not a given $G \in$ HAM, but by cleverly transforming other NP problems into HAM, one shows that all such problems are computationally viable. Much of the research on Hamilton cycles was at least partly initially motivated by this fact. For a more thorough exposition to computational complexity, see for instance [28, 29].

A general classification of Hamiltonian graphs continues to prove elusive, and as such, many of the classical results on Hamilton cycles focus on natural sufficient or necessary conditions for the existence of a Hamilton cycle within a given graph $G$. The most famous result in this field is the following theorem of Dirac [36], which gives a minimum degree condition which is sufficient to deduce that a given graph is Hamiltonian.

**Theorem 1.1.1** (Dirac's Theorem [36], 1952)**.** *Let $G$ be a graph on $n \geq 3$ vertices and suppose that $\delta(G) \geq n/2$. Then $G$ contains a Hamilton cycle.*

We note that Theorem 1.1.1 is readily seen to be best-possible in the sense that the '$n/2$' cannot be reduced; for example if $n$ is odd then the bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ has minimum degree $\lfloor n/2 \rfloor$ but clearly cannot be Hamiltonian.

Ore [103] and Chvátal [24] generalised Theorem 1.1.1 by considering the sum of degrees of non-adjacent pairs of vertices, and the 'degree sequence' of $G$, respectively. Chvátal and Erdős [25] gave a different sufficient condition concerning the relationship between the independence number of $G$ and the connectivity of $G$. Most of these results appear in any undergraduate course on graph theory. To see elegant proofs of these results and other classical theorems, see for example the books of West [124], and Bondy and Murty [16].

Many of the above classical results have also given rise to the study of related problems in the setting of hypergraphs, which generalise graphs. In particular, we highlight that Rödl, Ruciński, and Szemerédi [111] proved that, for $k \geq 3$, $\gamma > 0$ and large $n$, if every $(k-1)$-set of vertices in a $k$-uniform hypergraph $H$ is contained in at least $(1/2 + \gamma)n$ edges, then $H$ contains a hypergraph analogue of a Hamilton cycle which we call a 'tight' Hamilton cycle. Here, a hypergraph is *$k$-uniform* if every edge contains precisely $k$ vertices, and a tight cycle is such that every set of $k$ consecutive vertices in the underlying cyclical ordering forms an edge. Further, the same authors [112] later 'removed the $\gamma$' in the case $k = 3$, obtaining an exact analogue of Dirac's Theorem in this case, provided $n$ is large. We give a more in-depth discussion on this topic in Chapter 2.

Throughout this thesis, we focus on the study of Hamilton cycles in 'large' graphs and hypergraphs, in the sense that most results hold only provided the number of vertices is at least some (usually unspecified) number. This is a natural restriction to consider, in the following sense. Firstly, 'small' graphs $G$ and hypergraphs $H$, having at most some fixed $n^*$ vertices say, can often be thought of as pathological examples within which common (for instance probabilistic) methodologies may not work correctly, and secondly, such $G$ and $H$ are only a vanishing proportion of all graphs and hypergraphs, regardless of the value of $n^*$. Intuitively, 'large' $G$ and $H$ capture the key *behaviour* of graphs and hypergraphs with respect to most properties of interest, for example Hamiltonicity. Further, if one is careful enough to identify a suitable $n^*$ for which all $G$ or $H$ on at least $n^*$ vertices have a given property, then it may sometimes be feasible to manually check all smaller $G$ or $H$ via computer, say. For the above reasons, much of the related literature from this century, and the end of the previous century, is primarily interested in the behaviour of large graphs or hypergraphs, and we do not deviate from that here.

## 1.2 Description of the chapters

In this section, we describe each of the remaining chapters of the thesis, including the names of the articles contained therein, together with the publishing journal where appropriate, and the list of co-authors. Though each chapter links to the overall theme of Hamilton cycles in large graphs and hypergraphs, they are each self-contained, and contain their own in-depth introductions. For this reason, we keep their introductions here brief, stating only the key motivations and our main contributions.

Chapters 2–4 each comprise a published (or accepted to be published) article for which the author of this thesis was a primary author. Chapter 2 is an article titled 'Counting Hamilton cycles in Dirac hypergraphs' [50], published by the 'Combinatorics, Probability and Computing' journal in 2021. This article was joint work with Stefan Glock, Felix Joos, Daniela Kühn, and Deryk Osthus. A natural followup question to Dirac's Theorem (Theorem 1.1.1) is the following: Is the Hamilton cycle obtained in Theorem 1.1.1 ever unique, or is the hypothesis

$\delta(G) \geq n/2$ already strong enough to ensure more than one Hamilton cycle? If the latter is true, how many? Sarközy, Selkow, and Szemerédi [115] showed in 2003 that graphs satisfying the hypotheses of Theorem 1.1.1 actually contain $\exp(n \ln n - \Theta(n))$ distinct (not necessarily disjoint) Hamilton cycles, and Cuckler and Kahn [31, 32] in 2009 improved the error term to a subexponential one, obtaining that such graphs have $\exp(n \ln n - n(1 + \ln 2) - o(n))$ Hamilton cycles. Here, we use standard asymptotic notation. One may similarly ask the analogous counting question for the result of Rödl, Ruciński, and Szemerédi [115] discussed in Section 1.1, that $k$-uniform hypergraphs satisfying a 'Dirac-like' minimum 'degree' condition contain a tight Hamilton cycle. Our main contribution (appearing as Theorem 2.1.1 in Chapter 2) is that such hypergraphs have $\exp(n \ln n - \Theta(n))$ tight Hamilton cycles. Here, for a $k$-uniform hypergraph $H$ we define $\delta(H)$ to be the minimum over all $(k-1)$-sets $S \subseteq V(H)$, of the number of edges of $H$ containing $S$.

**Theorem 1.2.1** (Glock, Gould, Joos, Kühn, and Osthus, 2021)**.** *For a fixed integer $k \geq 2$ and a fixed constant $\gamma > 0$, the number of tight Hamilton cycles of a $k$-uniform hypergraph $H$ on $n$ vertices with $\delta(H) \geq (1/2 + \gamma)n$ is $\exp(n \ln n - \Theta(n))$.*

Theorem 1.2.1 also immediately extends the same counting bound to some other hypergraph notions of Hamilton cycles (see Corollary 2.1.2). We remark that the main idea of the proof of Theorem 1.2.1 is to analyse a short random walk on the set of ordered $(k-1)$-tuples of vertices; each outcome of the random walk traces a portion of a tight Hamilton cycle. We show that the vertices of the walk are very likely to look like a uniformly random set of vertices of the appropriate size, which enables one to iterate this process and append these short tight paths to one another.

Chapter 3 comprises an article titled 'Almost all optimally coloured complete graphs contain a rainbow Hamilton path' [56], published by the 'Journal of Combinatorial Theory, Series B' journal in 2022. This article was joint work with Tom Kelly, Daniela Kühn, and Deryk Osthus. A *proper edge-colouring* of a graph $G$ is a colouring of the edges of $G$ such that no vertex is incident to more than one edge of the same colour, and such a colouring is *optimal* if it uses the minimum possible number $\chi'(G)$ of colours. It transpires that $\chi'(K_n) = n-1$ if $n$ is even, and $\chi'(K_n) = n$ otherwise. If $G$ is equipped with an edge-colouring and $H \subseteq G$, we say that $H$ is *rainbow* if the edges of $H$ all have distinct colours. It is obvious that $K_n$ contains a Hamilton cycle for $n \geq 3$, but it is less clear if any optimal edge-colouring of $K_n$ should necessarily contain a Hamilton cycle which is rainbow. Indeed, if $n$ is even then there are not enough colours for a rainbow Hamilton cycle, but in this case we may still ask for a rainbow Hamilton path, i.e. a rainbow Hamilton cycle with one edge removed. Maamoun and Meyniel [96] showed in 1984 that there is an optimal edge-colouring of $K_n$ for some even $n$ which does not contain such a path, but Andersen [9] in 1989 conjectured that in fact any proper edge-colouring of $K_n$ admits a rainbow path of length $n-2$, i.e. omitting one vertex from $K_n$. Our main contribution (appearing as Theorem 3.1.3 in Chapter 3) is the following.

**Theorem 1.2.2** (Gould, Kelly, Kühn, and Osthus, 2022)**.** *Let $\phi$ be a uniformly random optimal edge-colouring of $K_n$. Then with high probability,*

*(i) $\phi$ admits a rainbow Hamilton path, and*

*(ii) $\phi$ admits a rainbow cycle $F$ containing all of the colours.*

*In particular, if $n$ is odd, then $F$ is a rainbow Hamilton cycle.*

Here, 'with high probability' means 'with a probability which tends to 1 as $n$ tends to infinity'. Theorem 1.2.2 confirms the optimal colourings case of Andersen's Conjecture (Conjecture 3.1.2) in a strong sense, for all but a vanishing proportion of such colourings. To prove Theorem 1.2.2, we use a lemma of Glock, Kühn, Montgomery, and Osthus (see Lemma 3.4.10) to obtain that any optimal edge-colouring of $K_n$ admits a rainbow path of length $n - o(n)$, and we use delicate 'switchings' arguments to find subgraphs we call 'gadgets', together with the method of 'distributive absorption' introduced by Montgomery [101] to show that we can complete this long rainbow path into a rainbow Hamilton path (or rainbow Hamilton cycle if $n$ is odd) with high probability in a uniformly random optimal edge-colouring.

Chapter 4 comprises an article titled 'Hamilton transversals in random Latin squares' [55], accepted for publication by the 'Random Structures & Algorithms' journal. This article was joint work with Tom Kelly. Let $\overleftrightarrow{K_n}$ denote the digraph obtained from $K_n$ by replacing each edge with two arcs (one in each direction), and adding a directed loop at each vertex. Further, let $\Phi(\overleftrightarrow{K_n})$ denote the set of proper $n$-arc-colourings of $\overleftrightarrow{K_n}$ on a fixed colour set, say $[n]$. Here a proper $n$-arc-colouring of $\overleftrightarrow{K_n}$ is a colouring of the arcs of $\overleftrightarrow{K_n}$ with $n$ colours such that no vertex is the tail of more than one arc with the same colour, or the head of more than one arc with the same colour. Note that $\Phi(\overleftrightarrow{K_n})$ is in bijection with the set $\mathcal{L}_n$ of $n \times n$ Latin squares with symbols $[n]$; construct a coloured digraph $G \in \Phi(\overleftrightarrow{K_n})$ from a Latin square $L \in \mathcal{L}_n$ by colouring arc $(i, j)$ of $\overleftrightarrow{K_n}$ with the symbol in position $(i, j)$ of $L$, and note that $L$ is uniquely recovered from $G$ in the obvious way. Thus, questions and properties for $\mathcal{L}_n$ may be studied in $\Phi(\overleftrightarrow{K_n})$ and vice versa. In particular, we highlight that the famous Ryser-Brualdi-Stein Conjecture [114, 20, 119] (Conjecture 3.1.1 in the present thesis) corresponds to the statement that all $G \in \Phi(\overleftrightarrow{K_n})$ admit a rainbow directed subgraph $H \subseteq G$ with maximum in-degree and out-degree 1, with $n-1$ arcs. We remark that such an $H$ is thus a disjoint collection of directed paths and cycles, and is called a 'partial transversal' (of size $n-1$) in the Latin square setting. Kwan [90] showed that Conjecture 3.1.1 holds for almost all $G \in \Phi(\overleftrightarrow{K_n})$ in the very strong sense that such $G$ contain $((1-o(1))n/e^2)^n$ of the above subgraphs $H$ with $n$ arcs, rather than $n-1$ (simply 'transversals' in the Latin square setting), whence such $H$ are each a rainbow spanning collection of disjoint directed cycles. Our main contribution (appearing as Theorem 4.1.6 in Chapter 4) is the following stronger statement.

**Theorem 1.2.3** (Gould and Kelly, 2022+). *Almost all proper $n$-arc-colourings of $\overleftrightarrow{K_n}$ contain at least*

$$\left((1-o(1))\frac{n}{e^2}\right)^n$$

*rainbow directed Hamilton cycles.*

We remark that we use the term 'Hamilton transversals' for the special transversals in $L \in \mathcal{L}_n$ corresponding to rainbow directed Hamilton cycles in $G \in \Phi(\overleftrightarrow{K_n})$, so that Theorem 1.2.3 is equivalent to the statement that almost all $n \times n$ Latin squares have at least $((1-o(1))n/e^2)^n$ Hamilton transversals. Further, up to the error term, this attains the upper bound given by Taranenko [120] on the number of (general) transversals in a Latin square. Theorem 1.2.3 confirms a conjecture of Gyárfás and Sárközy [60] (Conjecture 4.1.3 in the present thesis) in a strong sense for almost all $G \in \Phi(\overleftrightarrow{K_n})$, and provides a simpler proof of Kwan's [90] result, which relied on Keevash's [75, 77] breakthrough results on combinatorial designs. To prove Theorem 1.2.3, we use a result of Kwan and Sudakov [93] (Theorem 4.4.7

in the present thesis, the proof of which is short and primarily involves standard concentration inequalities and comparisons between probability spaces) which essentially states that almost all $G \in \Phi(\overleftrightarrow{K_n})$ are in some sense 'quasirandom', to show that there are $((1 - o(1))n/e^2)^n$ spanning rainbow directed path forests with few components, in almost all $G \in \Phi(\overleftrightarrow{K_n})$. To complete such path forests into rainbow Hamilton cycles, we use delicate switchings arguments inspired by the methods presented in Chapter 3, to find well-distributed 'absorbing' subgraphs in random $G \in \Phi(\overleftrightarrow{K_n})$. The translation to the directed setting entails considerable added challenges, which required new arguments and 'gadgets'.

Finally, Chapter 5 is an extended introduction to an upcoming untitled paper concerning a substantial improvement to the known results on matchings (in terms of the size of the matching obtained) in large, almost-regular hypergraphs, using the 'semi-random method'. The work discussed therein is joint work with Tom Kelly. In 1985, Rödl [108], building on ideas of Ajtai, Komlós, and Szemerédi [3], introduced the 'Rödl Nibble' (usage of which is often called the semi-random method), which centres around the iteration of small random processes to yield almost-spanning substructures in combinatorial host structures. A number of authors [46, 57, 6, 85, 122] used the semi-random method to show that large almost-regular hypergraphs, for which the number of edges containing any pair of vertices is much smaller than the maximum vertex degree, contain almost-perfect matchings. Such results are frequently useful in the search for Hamilton cycles and similar structures; indeed, a hypergraph matching result of Alon and Yuster [8] (Theorem 3.5.3 in the present thesis, itself based on a hypergraph matching result of Pippenger and Spencer [104]) is essential in Chapter 3 both in the construction of an almost-spanning rainbow path, and in the construction of the 'absorption' structure. The state-of-the-art hypergraph matchings result based on the classical semi-random method (in the sense that the obtained matching is largest) was given by Vu [122] in 2000. Our main result (Theorem 5.2.1, whose statement we defer until Chapter 5) provides a matching with substantially smaller leftover than that given by Vu [122] in most settings (and only falling short by the error term, in limited extreme circumstances which we discuss). The proof of Theorem 5.2.1 centres on an involved application of the 'Polynomial Method' (discussed briefly in Section 5.5.1) during the application of a single 'Nibble'; we sketch the proof further in Section 5.5. We remark that the original motivation for Theorem 5.2.1 was to use such a result as part of the search towards improving upon the known results towards Conjecture 4.1.3 in *general* $G \in \Phi(\overleftrightarrow{K_n})$ (as opposed to almost all $G \in \Phi(\overleftrightarrow{K_n})$, as in Chapter 4). The best result towards this conjecture is that of Benzing, Pokrovskiy, and Sudakov [11], who showed that every $G \in \Phi(\overleftrightarrow{K_n})$ contains a rainbow directed path forest on all but $O(n^{2/3})$ vertices, and moreover, a rainbow directed cycle on all but $O(n^{4/5})$ vertices. In our upcoming paper, we improve both of these results by using Theorem 5.2.1 to show that all $G \in \Phi(\overleftrightarrow{K_n})$ contain a rainbow directed cycle on all but $O(n^{1/2+\delta})$ vertices, for any constant $\delta > 0$. This result is given as Theorem 5.4.1 in the present thesis, and is discussed in Section 5.4.

# CHAPTER 2

# COUNTING HAMILTON CYCLES IN DIRAC HYPERGRAPHS

## Abstract

A tight Hamilton cycle in a $k$-uniform hypergraph ($k$-graph) $G$ is a cyclic ordering of the vertices of $G$ such that every set of $k$ consecutive vertices in the ordering forms an edge. Rödl, Ruciński, and Szemerédi proved that for $k \geq 3$, every $k$-graph on $n$ vertices with minimum codegree at least $n/2 + o(n)$ contains a tight Hamilton cycle. We show that the number of tight Hamilton cycles in such $k$-graphs is $\exp(n \ln n - \Theta(n))$. As a corollary, we obtain a similar estimate on the number of Hamilton $\ell$-cycles in such $k$-graphs for all $\ell \in \{0, \ldots, k-1\}$, which makes progress on a question of Ferber, Krivelevich and Sudakov.

## 2.1 Introduction

### 2.1.1 Counting Hamilton cycles in graphs

The problem of determining sufficient conditions for the existence of Hamilton cycles in graphs is one of the central topics in graph theory, and has given rise to extensive research. A classical result of Dirac [36] states that graphs on $n \geq 3$ vertices with minimum degree at least $n/2$ (*Dirac graphs*) contain a Hamilton cycle, and there are natural families of graphs which show that $n/2$ is best possible.

Bollobás [14] and Bondy [15] asked for an asymptotic estimate for the number of distinct Hamilton cycles in Dirac graphs. In 2003, Sárközy, Selkow, and Szemerédi [115] made substantial progress on this question by showing that $n$-vertex Dirac graphs contain $\exp(n \ln n - \Theta(n))$ Hamilton cycles. They also posed the question of whether this is the right order of magnitude for graphs satisfying other conditions known to ensure Hamiltonicity, like those of Ore, Pósa, and Chvátal (see [15]). Further, they conjectured that the minimum number of Hamilton cycles in $n$-vertex Dirac graphs is $\exp(n \ln n - n(1 + \ln 2) - o(n))$.

Cuckler and Kahn [32] analysed a self-avoiding random walk on the vertices of Dirac graphs to verify this conjecture as a consequence of a more precise result. Moreover, in a separate paper [31], they used entropy considerations to provide an upper bound for the number of Hamilton cycles in Dirac graphs. More precisely, writing $\Psi(G)$ to denote the number of distinct Hamilton cycles of a graph $G$, the main results of [31] and [32] together state that for any $n$-vertex Dirac graph $G$, we have $\log_2 \Psi(G) = 2H(G) - n \log_2 e - o(n)$, where $H(G)$ is the *entropy* of $G$. Combined with the result [32, Theorem 1.3] that Dirac graphs $G$

on $n$ vertices satisfy $H(G) \geq \frac{n}{2} \log_2 \delta(G)$, this confirms the conjecture of [115]. Moreover, the parameter $H(G)$ is the maximum of a concave function subject to linear constraints, and can thus be efficiently estimated. This yields an efficient algorithm for estimating $\Psi(G)$ for Dirac graphs $G$, to within subexponential factors.

### 2.1.2 Hamilton cycles in hypergraphs

The study of Hamilton cycles in hypergraphs was initiated in a 1976 paper of Bermond, Germa, Heydemann, and Sotteau [12]. For $k$-uniform hypergraphs ($k$-*graphs*), we may sensibly define a cycle in a number of ways (see for example [87, 109, 127]). Let $k \geq 2$ be an integer and let $\ell \in \{0, \ldots, k-1\}$. We say that a $k$-uniform hypergraph $C$ is an $\ell$-cycle if there exists a cyclic ordering of the vertices of $C$ such that every edge of $C$ consists of $k$ consecutive vertices and such that every pair of consecutive edges (in the natural ordering of the edges) intersects in precisely $\ell$ vertices. A *Hamilton $\ell$-cycle* of a $k$-graph $G$ is a subgraph $C \subseteq G$, where $C$ is a $k$-uniform $\ell$-cycle with $V(C) = V(G)$. Thus, if $G$ contains a Hamilton $\ell$-cycle, then $k - \ell$ divides $|V(G)|$. Moreover, if $\ell = 0$ then a Hamilton $\ell$-cycle is just a perfect matching of $G$. We usually call a $(k-1)$-cycle a *tight cycle*, and we say that a Hamilton $(k-1)$-cycle of a $k$-graph $G$ is a *tight Hamilton cycle* of $G$.

We wish to generalise the study of Hamilton cycles in Dirac graphs to the setting of hypergraphs, and so we now need a natural hypergraph generalisation of the notion of degree. Given a $k$-graph $G$ and a set $S \subseteq V(G)$ of $k-1$ vertices, we say that the *codegree* of $S$ in $G$, denoted $d_G(S)$ (or simply $d(S)$ when $G$ is clear from the context), is the number of edges of $G$ containing $S$. For a $k$-graph $G$, we write $\delta(G)$ for the minimum codegree over all $(k-1)$-sets $S \subseteq V(G)$, and refer to this quantity as the *minimum codegree* of $G$.

Katona and Kierstead [74] gave a sufficient condition on the minimum codegree for $k$-graphs to have a tight Hamilton cycle. Further, they conjectured that for all integers $k \geq 2$, a minimum codegree of at least $n/2$ suffices for $n$-vertex $k$-graphs. Rödl, Ruciński, and Szemerédi proved an asymptotic version [110] of the $k = 3$ case of this conjecture, and then an exact version for large $n$ [112]. The work of [110] was shortly afterwards generalised to all integers $k \geq 3$ by the same authors [111]. Further results on tight Hamilton cycles can be found e.g. in [1, 107]. For $(k-\ell) \nmid k$, Kühn, Mycroft, and Osthus [86] asymptotically determined the threshold for the existence of a Hamilton $\ell$-cycle (this generalised previous results in [88, 78, 65]). Subsequently several exact results were proved in [33, 64]. It turns out that the threshold is significantly below $n/2$ if $(k - \ell) \nmid k$. For all other cases it follows from the result of [111] that the threshold is asymptotically $n/2$.

### 2.1.3 Our main result

Ferber, Krivelevich, and Sudakov [45] were the first to generalise the study of counting Hamilton cycles to the hypergraph setting (and also considered perfect matchings). They proved for $1 \leq \ell < k/2$ that if a $k$-graph $G$ on $n$ vertices with $(k - \ell) \mid n$ satisfies $\delta(G) \geq \alpha n$ for some $\alpha > 1/2$, then $G$ contains $(1 - o(1))^n \cdot n! \cdot \left(\frac{\alpha}{\ell!(k-2\ell)!}\right)^{\frac{n}{k-\ell}}$ Hamilton $\ell$-cycles. As a natural question, they asked whether this can be generalized to all $\ell$.

We adapt some ideas from the random walk analysis of [32] to show that any large $k$-graph whose minimum codegree is slightly above $n/2$ contains a large number of tight Hamilton cycles.

**Theorem 2.1.1.** *For a fixed integer $k \geq 2$ and a fixed constant $\gamma > 0$, the number of tight Hamilton cycles of a $k$-graph $G$ on $n$ vertices with $\delta(G) \geq (1/2+\gamma)n$ is* $\exp(n \ln n - \Theta(n))$.

Notice that we claim this number of tight Hamilton cycles holds with equality, up to the exponential error bound $\exp(-\Theta(n))$. We discuss this error bound further in the concluding remarks. It will suffice to show that the lower bound holds, since any $k$-graph on $n$ vertices trivially has at most $(n-1)!/2$ distinct tight Hamilton cycles. Theorem 2.1.1 easily yields the following corollary about the number of Hamilton $\ell$-cycles in a $k$-graph whose codegrees are slightly above $n/2$, for each $\ell \in \{0, \ldots, k-1\}$.

**Corollary 2.1.2.** *For a fixed integer $k \geq 2$ and a fixed constant $\gamma > 0$, the number of Hamilton $\ell$-cycles of a $k$-graph $G$ on $n$ vertices with $(k-\ell) \mid n$ and $\delta(G) \geq (1/2+\gamma)n$ is*

(i) $\exp\left(\left(1-\frac{1}{k}\right) n \ln n - \Theta(n)\right)$, *if* $\ell = 0$;

(ii) $\exp(n \ln n - \Theta(n))$, *if* $\ell \in [k-1]$.

This addresses the above mentioned question of Ferber, Krivelevich and Sudakov (though our result is less precise than theirs for $\ell < k/2$). We remark that the $-\frac{1}{k}$-term for the case of perfect matchings is missing in [45, Theorem 1.1], but follows from their proof. Finally, recall that the minimum codegree threshold for the existence of Hamilton $\ell$-cycles can be below $n/2$ when $\ell < k-1$. It would thus be a natural question to extend the counting results to this larger range. For the rest of the paper, we focus on counting tight Hamilton cycles.

## 2.2 Sketch of the proof of Theorem 2.1.1

In this section we provide a rough sketch of the proof of our main result.

### 2.2.1 Basic notation

We first need to introduce some notation that we use throughout the paper. For a set $V$ and a natural number $\ell$, we write $\binom{V}{\ell}$ to denote the set of all unordered $\ell$-subsets of distinct elements of $V$. We write $(V)_\ell$ to denote the set of all ordered $\ell$-subsets of distinct elements of $V$, so that $|(V)_\ell| = \ell!\left|\binom{V}{\ell}\right|$. We usually use boldface capital letters to denote unordered subsets $\mathbf{S} \in \binom{V}{\ell}$ of the fixed size $\ell$, and we exclusively use boldface capital letters with arrows above to denote ordered subsets $\overrightarrow{\mathbf{S}} \in (V)_\ell$. When an ordered tuple $\overrightarrow{\mathbf{S}} \in (V)_\ell$ is first given, the arrow will exclusively point to the right. We may subsequently drop the arrow to denote the unordered version of this $\ell$-set, so that if $\overrightarrow{\mathbf{S}}$ is the ordered sequence of $\ell$ distinct elements $(x_1, \dots, x_\ell)$, then $\mathbf{S}$ subsequently used without the arrow denotes the unordered set $\{x_1, \dots, x_\ell\}$. Moreover, we write $\overleftarrow{\mathbf{S}}$ to denote the ordered $\ell$-tuple obtained by reversing the ordering of $\overrightarrow{\mathbf{S}}$, so that $\overleftarrow{\mathbf{S}} = (x_\ell, x_{\ell-1}, \dots, x_1)$. Let $G = (V, E)$ be a hypergraph and let $U \subseteq V(G)$. Then the sub(hyper)graph of $G$ *induced* by $U$, denoted $G[U]$, is the hypergraph $H = (V(H), E(H))$, where $V(H) = U$, and $E(H)$ is precisely the set of all edges of $G$ containing only vertices in $U$. We write $G - U$ to denote the hypergraph $G' \subseteq G$ obtained from $G$ by deleting the vertices in $U$ and all edges of $G$ containing any vertex in $U$. We say that a $k$-graph $P$ is a *$k$-uniform tight path* (or simply *tight path* if $k$ is clear from the context) if $P$ admits an ordering of its vertices $V(P) = \{v_1, \dots, v_m\}$ such that $E(P) = \{\{v_i, \dots, v_{i+k-1}\}\colon 1 \le i \le m - (k-1)\}$. The *ends* of $P$ are the ordered

$(k-1)$-tuples $(v_1, \dots, v_{k-1})$ and $(v_m, \dots, v_{m-k+2})$. We also say that $P$ *connects* the ends of $P$. We say that a tight path $P$ with $m$ edges (and thus with $m+(k-1)$ vertices) is an *$m$-path*, and has *length $m$*. For a $k$-graph $G$ and an integer $t \geq k$ we say that a sequence $(v_1, v_2, \dots, v_t)$ of (not necessarily distinct) vertices is a *walk* in $G$ if every set of $k$ consecutive vertices in the sequence forms an edge. Let $\gamma > 0$ be a constant. A $k$-graph $G$ on $n$ vertices is called *$\gamma$-Dirac* if $\delta(G) \geq (1/2+\gamma)n$. Finally, given a hypergraph $G$, we say a weighting of the edges $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ is a *fractional matching* if we have $\sum_{e \ni v} \mathbf{x}\,(e) \leq 1$ for every $v \in V(G)$, and we say that $\mathbf{x}$ is *perfect* if $\sum_{e \ni v} \mathbf{x}\,(e) = 1$ for every $v \in V(G)$.

### 2.2.2 Outline of the argument

Let $\gamma > 0$, and let $G$ be an $n$-vertex $k$-graph satisfying $\delta(G) \geq (1/2+\gamma)n$, where $k \geq 2$ and $n$ is sufficiently large. The main step of our proof is to count tight paths of length $n - o(n)$ in $G$. Using the framework of Rödl, Ruciński, and Szemerédi [111], which is based on the absorption technique, we can complete each such long path into a tight Hamilton cycle of $G$. The key lemma (Lemma 2.5.1) in the proof of Theorem 2.1.1 states that we can find many paths of length $\sqrt{n}$ in $G$, all starting at the same ordered $(k-1)$-tuple $\overrightarrow{\mathbf{S}} \in (V(G))_{k-1}$, such that for each such path the remainder of $G$ still has minimum codegree at least $\left(\frac{1}{2} + \gamma - n^{-2/3}\right)(n - \sqrt{n})$. The proof of this 'iteration lemma' is the sole focus of Section 2.5, and involves the analysis of a self-avoiding random walk $\mathcal{X}$ on the vertices of $G$. In order to prove the iteration lemma, we first need to show that $G$ admits a perfect fractional matching which is 'normal', which means that each edge of $G$ has weight $\Theta(n^{-k+1})$. We construct such a normal perfect fractional matching $\mathbf{x}$ in Section 2.4 via a probabilistic argument based on switchings (it is not clear how to generalise the entropy-based approach of [32] to the hypergraph setting).

In Section 2.5, we use $\mathbf{x}$ to define the transition probabilities of the random walk $\mathcal{X}$. We construct $\mathcal{X}$ such that an outcome of $\mathcal{X}$ corresponds to a tight path in $G$ of length $\sqrt{n}$ which starts at some given $\overrightarrow{\mathbf{S}} \in (V(G))_{k-1}$. We wish to count the number of outcomes of $\mathcal{X}$ which essentially leave the $\gamma$-Dirac property of the remaining graph intact. Such outcomes of $\mathcal{X}$ are called *good* walks. It will suffice to show that $\mathcal{X}$ is good with probability at least $1/2$. To do this, we will show that it is likely that the vertices that $\mathcal{X}$ visits look roughly like a uniformly random subset of the vertices of $G$, of appropriate size. We will show that the behaviour of $\mathcal{X}$ over a small number of steps can be assumed to be very close to the behaviour of a modified version of $\mathcal{X}$, in which the walk is allowed to revisit vertices. We use the normality property of $\mathbf{x}$ to show that the modified walk mixes rapidly, and we use the fact that $\mathbf{x}$ is a perfect fractional matching to show that, under the stationary distribution, each vertex is essentially visited with the same probability. Thus, roughly speaking, the distribution of the vertices for $\mathcal{X}$ to visit at any step is close to uniform on $V(G)$. We give a more thorough sketch of the proof of the iteration lemma in Section 2.5.2.

In Section 2.6, we focus on repeatedly applying the iteration lemma to obtain many long paths in $G$. Let $P$ be a $\sqrt{n}$-path in $G$ obtained from the first iteration of the iteration lemma, let $\mathbf{T}$ be the unordered set consisting of the final $k-1$ vertices of $P$, and let $G_P \coloneqq G - (V(P) \setminus \mathbf{T})$. The idea is that, since the $\gamma$-Dirac property is essentially intact in $G_P$, we may find a new normal perfect fractional matching $\mathbf{x}_P \colon E(G_P) \to \mathbb{R}^+$ and can thus apply the iteration lemma to $G_P$. In this second iteration, we insist that all the walks $\mathcal{X}$ start at the final ordered $(k-1)$-tuple of $P$. Then we may attach any of the paths $P'$ from the second iteration onto $P$ to obtain a longer tight path in $G$ which still leaves the $\gamma$-Dirac property of the remaining graph essentially intact. We show that we may iterate this process until fewer than $n^{7/8}$ vertices of $G$ remain, and we multiplicatively use the count of paths given by the iteration lemma to deduce that the number of resulting long paths of $G$ is essentially the number given in the statement of Theorem 2.1.1. (Observe that each combination of paths yields a different concatenated path). Finally then, we complete the proof of Theorem 2.1.1 by absorbing the vertices left over by each such long path into a tight Hamilton cycle of $G$.

## 2.3 Preliminaries

In the following section, we collect further notation, as well as some results that we will use throughout the paper.

### 2.3.1 Notation

Let $\gamma > 0$ be a constant. We say that a $k$-graph $G$ is an $(n, k, \gamma)$-*graph* if $G$ has $n$ vertices and $G$ is $\gamma$-Dirac. When $G$ is clear from the context, we often write $V$ instead of $V(G)$. For a $k$-graph $G$ and $\mathbf{S} \in \binom{V}{k-1}$, we write $N_G(\mathbf{S}) \coloneqq \{v \in V : \mathbf{S} \cup \{v\} \in E(G)\}$. We say that $\mathbf{S}$ is *isolated* if $N_G(\mathbf{S}) = \emptyset$, and that $\mathbf{S}$ is *non-isolated* if $\mathbf{S}$ is not isolated. For a positive integer $\ell$, we say that a walk $(v_1, \dots, v_{\ell+k-1})$ on the vertices of $G$ is an $\ell$-*walk*. Let $\overrightarrow{\mathbf{S}}, \overrightarrow{\mathbf{T}} \in (V)_{k-1}$. We say an $\ell$-walk $(v_1, \dots, v_{\ell+k-1})$ in $G$ is an $\ell$-*walk from* $\overrightarrow{\mathbf{S}}$ *to* $\overrightarrow{\mathbf{T}}$ if $\overrightarrow{\mathbf{S}} = (v_1, \dots, v_{k-1})$ and $\overrightarrow{\mathbf{T}} = (v_{\ell+1}, \dots, v_{\ell+k-1})$. A *matching* $M$ of a $k$-graph $G$ is a collection of vertex-disjoint edges of $G$, and we say that $M$ is *perfect* if every vertex $v \in V(G)$ is included in some edge of $M$. Let $M$ be a matching in a $k$-graph $G$. Where it has no effect on the argument, we sometimes abuse notation and identify $M$ with the subgraph $M' \subseteq G$ satisfying $E(M') = M$ and $V(M') = \bigcup_{e \in M} e$. For finite sets $U \subseteq V$ and a function $f : V \to \mathbb{R}$, we define $f(U) \coloneqq \sum_{u \in U} f(u)$, and $||f||_\infty \coloneqq \max_{v \in V} f(v)$. We write $\mathbb{1}_U : V \to \{0, 1\}$ to be the *indicator function* for $U$, defined by $\mathbb{1}_U(x) = 1$ if $x \in U$, and $\mathbb{1}_U(x) = 0$ otherwise. For an event $\mathcal{E}$ in a probability space, we write $\mathcal{E}^c$ to denote the complement of $\mathcal{E}$. We write $\log x$ to mean $\log_2 x$, and we write $\ln x$ to mean $\log_e x$. We also write $a = (1 \pm b)c$ to mean $(1-b)c < a < (1+b)c$. For a natural number $n$ we write $[n] \coloneqq \{1, \ldots, n\}$. We write $x \ll y$ to mean that for any $y \in (0, 1]$ there exists an $x_0 \in (0, 1)$ such that for all $0 < x \leq x_0$ the subsequent statement holds. Hierarchies with more constants are defined similarly and should be read from the right to the left. Constants in hierarchies will always be real numbers in $(0, 1]$. Moreover, if $1/x$ appears in a hierarchy, this implicitly means that $x$ is a natural number. More precisely, $1/x \ll y$ means that for any $y \in (0, 1]$, there exists an $x_0 \in \mathbb{N}$ such that for all $x \in \mathbb{N}$ with $x \geq x_0$ the subsequent statement holds. We assume large numbers to be integers if this does not affect the argument.

### 2.3.2 Probabilistic tools

In this subsection we collect some probabilistic definitions and results that we will need throughout the paper.

The *total variation distance* between two probability measures $\mu$ and $\nu$ on a finite set $S$ is $d_{TV}(\mu, \nu) := \sup\{|\mu(T) - \nu(T)| \colon T \subseteq S\}$. It is well-known that the total variation distance satisfies

$$d_{TV}(\mu,\nu) = \frac{1}{2}\sum_{s\in S} |\mu(s) - \nu(s)| = \inf\{\mathbb{P}[X \neq Y]\}, \tag{2.3.1}$$

where the infimum is taken over coupled random variables $X$ and $Y$ having laws $\mu$ and $\nu$ respectively (see [37, p.119] for more details). We write $d_{TV}(X,Y)$ for the total variation distance between the laws of the random variables $X$ and $Y$.

Next, we need an inequality of [32], which follows easily from Azuma's inequality.

**Lemma 2.3.1** ([32, Lemma 5.3]). *Let $X_0, X_1, ...$ be random variables taking values in a set $V$, and let $g\colon V \to \mathbb{R}$. Then for any $t > 0$ and any $p, q \in \mathbb{N}$, we have*

$$\mathbb{P}\left[\left|\sum_{k=0}^{p} \left(g\left(X_{k+q}\right) - \mathbb{E}\left[g\left(X_{k+q}\right)|X_0, ..., X_k\right]\right)\right| > t||g||_\infty\sqrt{pq}\right] < 2qe^{-t^2/2}.$$

We will need the following Chernoff-type bound (see [23] and [67] for example).

**Lemma 2.3.2.** *Let $X$ be a random variable with a binomial or hypergeometric distribution. Suppose $\mathbb{E}[X] > 0$ and let $t > 0$. Then $\mathbb{P}[X \leq \mathbb{E}[X] - t] \leq e^{-t^2/(2\mathbb{E}[X])}$.*

We conclude this section with a result which shows that most small sets in an $(n,k,\gamma)$-graph inherit the Dirac condition.

**Proposition 2.3.3.** *Let $1/n \ll 1/m \ll \gamma, 1/k, 1/t, 1/\ell$, where $\ell \mid n$, and let $G$ be an $(n,k,\gamma)$-graph. Let $\mathcal{P}$ be a partition of $V$ into $\ell$-sets and let $\mathcal{P}_0 \subseteq \mathcal{P}$ be of size $|\mathcal{P}_0| = t$. Pick $\mathcal{P}' \subseteq \mathcal{P} \setminus \mathcal{P}_0$ of size $m$ uniformly at random. Then $\mathbb{P}[G[\bigcup(\mathcal{P}_0 \cup \mathcal{P}')]$ is $\gamma/2$-Dirac$] \geq 1 - e^{-\sqrt{m}}$.*

To prove Proposition 2.3.3, we first need the following Chernoff-type bound (see [23] and [67] for example), and a well-known result on the probability that a binomially distributed random variable assumes its mean value.

**Lemma 2.3.4.** *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with $\mathbb{P}[X_i = 1] = p_i$ for each $i \in [n]$. Let $a_1, \ldots, a_n \geq 0$ with $\sum_{i=1}^n a_i > 0$, set $X = \sum_{i=1}^n a_i X_i$, and define $\nu := \sum_{i=1}^n a_i^2 p_i$. Then $\mathbb{P}[X \leq \mathbb{E}[X] - t] \leq e^{-t^2/(2\nu)}$.*

**Lemma 2.3.5.** *Let $1/n \ll 1/m \leq 1$, with $m, n \in \mathbb{N}$, and let $X$ be a binomial random variable with parameters $n$ and $p := m/n$. Then $\mathbb{P}[X = m] \geq 1/(4\sqrt{m})$.*

We now prove Proposition 2.3.3.

**Proof of Proposition 2.3.3.** For any subset $\mathcal{Q} \subseteq \mathcal{P}$, let $V(\mathcal{Q})$ denote the set of all vertices in any $\ell$-set $A \in \mathcal{Q}$. Define $p := m/|\mathcal{P} \setminus \mathcal{P}_0|$. We construct a random set $X \subseteq \mathcal{P} \setminus \mathcal{P}_0$ by including each $\ell$-set $A \in \mathcal{P} \setminus \mathcal{P}_0$ independently with probability $p$. Let $Y := X \cup \mathcal{P}_0$, and define the events $\mathcal{E}_1 := \{|X| = m\}$ and $\mathcal{E}_2 := \bigcap_{\mathbf{S} \in \binom{V(Y)}{k-1}} \{d_{G[V(Y)]}(\mathbf{S}) \geq (1/2 + \gamma/2)\ell(m + t)\}$. Let $\mathbb{P}_b$ be the probability measure for the space corresponding to constructing $X$. Notice then that $\mathbb{P}[\,\cdot\,] = \mathbb{P}_b[\,\cdot \mid \mathcal{E}_1]$. It remains to prove that $\mathbb{P}_b[(\mathcal{E}_2)^c \mid \mathcal{E}_1] \leq e^{-\sqrt{m}}$. Write $M := \binom{V(G)}{k-1}$, and for each $\mathbf{S} \in M$ write $d_Y(\mathbf{S}) := |N_G(\mathbf{S}) \cap V(Y)|$, write $\mathcal{P}(\mathbf{S}) := \{A \in \mathcal{P} \colon \mathbf{S} \cap A \neq \emptyset\}$, write $\mathcal{P}_{\mathbf{S}} := \mathcal{P} \setminus (\mathcal{P}_0 \cup \mathcal{P}(\mathbf{S}))$, and write $J_{\mathbf{S}} := N_G(\mathbf{S}) \cap V(\mathcal{P}_{\mathbf{S}})$. Notice that $d_Y(\mathbf{S}) \geq d'_Y(\mathbf{S}) := |N_G(\mathbf{S}) \cap V(Y) \cap V(\mathcal{P}_{\mathbf{S}})|$. Fix $\mathbf{S} \in M$. Notice that $|J_{\mathbf{S}}| \geq (1/2 + 3\gamma/4)n$, since $|V(\mathcal{P}_0 \cup \mathcal{P}(\mathbf{S}))| \leq \ell(t + k)$. Observe that

$$\mathbb{E}_b[d'_Y(\mathbf{S})] = |J_{\mathbf{S}}|p \geq (1/2 + 3\gamma/4)np \geq (1/2 + 2\gamma/3)\ell(m + t).$$

For each $\ell$-set $A \in \mathcal{P}_{\mathbf{S}}$, let $Y_A$ be the indicator random variable for the event $\{A \in X\}$, and let $c_A := |A \cap N_G(\mathbf{S})|$. Then we have $d'_Y(\mathbf{S}) = \sum_{A \in \mathcal{P}_{\mathbf{S}}} c_A Y_A$, and by applying Lemma 2.3.4 to $d'_Y(\mathbf{S})$, we obtain $\mathbb{P}_b[d'_Y(\mathbf{S}) < (1/2 + \gamma/2)\ell(m + t)] \leq e^{-3\sqrt{m}}$. Note that for any $\mathbf{S} \in M$ and any $v \in J_{\mathbf{S}}$, the events $\{\mathbf{S} \subseteq V(Y)\}$

and $\{v \in V(Y)\}$ are independent by construction. Let $\tilde{\mathcal{P}}(\mathbf{S}) := \mathcal{P}(\mathbf{S}) \setminus \mathcal{P}_0$, and for each $0 \leq j \leq k-1$, let $M_j := \{\mathbf{S} \in M : |\tilde{\mathcal{P}}(\mathbf{S})| = j\}$. Note that for each $\mathbf{S} \in M_j$, we have $\mathbb{P}[\mathbf{S} \subseteq V(Y)] = p^j$, and note further that $|M_j| \leq (\ell t)^k n^j$, for each $j$. Then by a union bound over all $\mathbf{S} \in M$ we obtain

$$
\begin{aligned}
\mathbb{P}_b[(\mathcal{E}_2)^c] &\leq \sum_{j=0}^{k-1} \sum_{\mathbf{S} \in M_j} \mathbb{P}_b[\mathbf{S} \subseteq V(Y), d'_Y(\mathbf{S}) < (1/2 + \gamma/2)\ell(m+t)] \\
&\leq e^{-3\sqrt{m}} \sum_{j=0}^{k-1} (\ell t)^k (np)^j \leq e^{-3\sqrt{m}} \sum_{j=0}^{k-1} (\ell t)^k (2m\ell)^j \leq e^{-2\sqrt{m}}.
\end{aligned}
$$

Finally, by Lemma 2.3.5 we have $\mathbb{P}_b[\mathcal{E}_1] \geq 1/(4\sqrt{m})$, whence it follows that $\mathbb{P}_b[(\mathcal{E}_2)^c \mid \mathcal{E}_1] \leq \mathbb{P}_b[(\mathcal{E}_2)^c] / \mathbb{P}_b[\mathcal{E}_1] \leq e^{-\sqrt{m}}$. □

### 2.3.3 Tight Hamilton-connectedness

Let $G$ be an $(n, k, \gamma)$-graph and let $P$ be a tight path in $G$. We say that $P$ is a *tight Hamilton path* of $G$ if $V(P) = V$. We say that $G$ is *tight Hamilton-connected* if for any disjoint $\overrightarrow{\mathbf{S}}, \overrightarrow{\mathbf{T}} \in (V)_{k-1}$, there is a tight Hamilton path of $G$ which connects $\overrightarrow{\mathbf{S}}$ and $\overrightarrow{\mathbf{T}}$. We will deduce from the results in [111] that large $(n, k, \gamma)$-graphs are tight Hamilton-connected for $k \geq 3$. This will be important in the absorption step of our main argument, and also in the mixing part of our random walk analysis. We begin by stating the main theorem of [111].

**Theorem 2.3.6** ([111, Theorem 1.1])**.** *Let $1/n \ll \gamma, 1/k$, where $k \geq 3$, and let $G$ be an $(n, k, \gamma)$-graph. Then $G$ contains a tight Hamilton cycle.*

The next lemma ensures the existence of an 'absorbing path' $A$, which can absorb small sets of vertices into its interior.

**Lemma 2.3.7** ([111, Lemma 2.1])**.** *Let $1/n \ll \gamma, 1/k$, where $k \geq 3$, suppose that $\gamma \leq 1/(32k)$, set $\beta := 2^{k-4}\gamma^{2k}n$, and let $G$ be an $(n,k,\gamma)$-graph. Then there exists a tight path $A$ in $G$ with $|V(A)| \leq 16k\gamma^{k-1}n$ such that for every subset $U \subseteq V \setminus V(A)$ of size $|U| \leq \beta$, there is a tight path $A_U$ in $G$ with $V(A_U) = V(A) \cup U$ and such that $A_U$ has the same ends as $A$.*

The next lemma will enable us to find constant-length tight paths between any disjoint pair of ordered $(k-1)$-sets of vertices.

**Lemma 2.3.8** ([111, Lemma 2.4])**.** *Let $1/n \ll \gamma, 1/k$, where $k \geq 3$, and let $G$ be an $(n,k,\gamma)$-graph. Then for every $\overrightarrow{\mathbf{S}}, \overrightarrow{\mathbf{T}} \in (V)_{k-1}$ with $\mathbf{S} \cap \mathbf{T} = \emptyset$, there is an $\ell$-path $P$ in $G$ with $\ell \leq 2k/\gamma^2$ that connects $\overrightarrow{\mathbf{S}}$ and $\overrightarrow{\mathbf{T}}$.*

We are now ready to prove that large $(n,k,\gamma)$-graphs are tight Hamilton-connected.

**Lemma 2.3.9.** *Let $1/n \ll \gamma, 1/k$, where $k \geq 2$, and let $G$ be an $(n,k,\gamma)$-graph. Then $G$ is tight Hamilton-connected.*

**Proof.** Firstly, note that this result follows easily from Dirac's Theorem for the case $k = 2$. Now, suppose $k \geq 3$ and suppose without loss of generality that $\gamma > 0$ is sufficiently small in comparison to $k$. Let $\overrightarrow{\mathbf{S}}, \overrightarrow{\mathbf{T}} \in (V)_{k-1}$ be disjoint, and write $\overrightarrow{\mathbf{S}} = (s_1, \ldots, s_{k-1})$. Set $\gamma' := 3\gamma/4$, so that $G' := G - (\mathbf{S} \cup \mathbf{T})$ is $\gamma'$-Dirac, and set $n' := n - 2(k-1)$. We apply Lemma 2.3.7 to $G'$ to obtain a tight path $A$ in $G'$ with $|V(A)| \leq 16k(\gamma')^{k-1}n'$, with the properties as stated in Lemma 2.3.7. Choose a set $W \subseteq V \setminus (\mathbf{S} \cup \mathbf{T} \cup V(A))$ of size $(\gamma')^{3k}n'$ uniformly at random among all sets of that size. Then a simple application of Lemma 2.3.2 shows that with high probability, for all $\mathbf{M} \in \binom{V}{k-1}$ we have $|N_G(\mathbf{M}) \cap W| \geq (1/2 + \gamma/2)|W|$, and thus in particular, $G[W]$ is $\gamma/2$-Dirac. We fix such a choice of $W$. Set $G'' := G - (\mathbf{S} \cup \mathbf{T} \cup V(A) \cup W)$, and notice that $G''$ is $\gamma/2$-Dirac.

We apply Theorem 2.3.6 to $G''$ to obtain a Hamilton cycle $C$ of $G''$. Delete $k-1$ consecutive edges of $C$ to obtain a Hamilton path $P$ of $G''$ with ends $\overrightarrow{\mathbf{X}}$ and $\overrightarrow{\mathbf{Y}}$. We use the property that all $(k-1)$-tuples in $G$ have high codegree in $W$ to find, for each $i \in [k-1]$ in turn, a vertex $v_i \in W \setminus \{v_1, \ldots, v_{i-1}\}$ such that $\{s_i, \ldots, s_{k-1}, v_1, \ldots, v_i\}$ is an edge. Let $\overrightarrow{\mathbf{S'}} := (v_{k-1}, \ldots, v_1)$, so that we have found a $(k-1)$-path $P_S$ with ends $\overrightarrow{\mathbf{S}}$ and $\overrightarrow{\mathbf{S'}}$. Let $\overrightarrow{\mathbf{A}}_1$ and $\overrightarrow{\mathbf{A}}_2$ be the ends of $A$. We similarly find mutually disjoint $\overrightarrow{\mathbf{X'}}, \overrightarrow{\mathbf{Y'}}, \overrightarrow{\mathbf{A'_1}}, \overrightarrow{\mathbf{A'_2}}, \overrightarrow{\mathbf{T'}} \in (W)_{k-1}$ and $(k-1)$-paths $P_X, P_Y, P_{A_1}, P_{A_2}, P_T$ with the corresponding pairs of ends. Since $G[W]$ is $\gamma/2$-Dirac, we can apply Lemma 2.3.8 to obtain a path $P_{SX}$ of length at most $8k/\gamma^2$ in $G[W]$ which connects $\overleftarrow{\mathbf{S'}}$ and $\overleftarrow{\mathbf{X'}}$. Since $P_{SX}$ contains so few vertices, we can repeat the process to find disjoint paths $P_{YA_1}$ and $P_{A_2T}$ in $G[W]$ with ends $\overleftarrow{\mathbf{Y'}}$ and $\overleftarrow{\mathbf{A'_1}}$, and $\overleftarrow{\mathbf{A'_2}}$ and $\overleftarrow{\mathbf{T'}}$, respectively. Let $W' := W \setminus (V(P_{SX}) \cup V(P_{YA_1}) \cup V(P_{A_2T}))$. We apply the absorbing property of $A$ to obtain a tight path $A_{W'}$ in $G$ with $V(A_{W'}) = V(A) \cup W'$, such that $A_{W'}$ has ends $\overrightarrow{\mathbf{A_1}}$ and $\overrightarrow{\mathbf{A_2}}$.

Then $P_S \cup P_{SX} \cup P_X \cup P \cup P_Y \cup P_{YA_1} \cup P_{A_1} \cup A_{W'} \cup P_{A_2} \cup P_{A_2T} \cup P_T$ is a tight Hamilton path in $G$ which connects $\overrightarrow{\mathbf{S}}$ and $\overrightarrow{\mathbf{T}}$. $\square$

## 2.4 Normal perfect fractional matchings

Let $k \geq 2$ and let $G$ be a $k$-graph on $n$ vertices. We say that an edge weighting $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ is $C$-*normal* if

$$\frac{1}{Cn^{k-1}} \leq \mathbf{x}\,(e) \leq \frac{C}{n^{k-1}}, \qquad \text{for each } e \in E(G). \tag{2.4.1}$$

In this section we adapt some ideas of [30] to show that an $(n, k, \gamma)$-graph $G$ admits a normal perfect fractional matching (see Lemma 2.4.2). This will be an essential tool in our random walk analysis for showing that the random walk is roughly equally likely to visit any vertex. The idea is to construct a perfect fractional matching of $G$ in which the weight of any edge $e$ is set to be the probability that $e$ is included in a uniformly random perfect matching of $G$. (A sufficiently large $(n, k, \gamma)$-graph with $k \mid n$ has at least one perfect matching [89, 113]). A crucial feature of this approach is that any edge $e$ is roughly equally likely to be included in a uniformly random perfect matching of $G$. We show this using the so-called 'switching method' in a similar way as in [30]. Let $k \geq 2$, let $G$ be a $k$-graph, let $e \in E(G)$, and let $M_\ell$ be a perfect matching of $G$ containing precisely $\ell$ edges intersecting $e$. Supposing $0 \leq \ell \leq k-1$, we define an $(e, M_\ell)$-*upswitching* to be a matching $Y$ of $G$ satisfying

(i) $e \subseteq V(Y)$;

(ii) $Y$ contains precisely $\ell + 1$ edges intersecting $e$;

(iii) for all $e' \in M_\ell$, we have either $e' \subseteq V(Y)$ or $e' \cap V(Y) = \emptyset$.

Supposing instead that $\ell \in [k]$, we define an $(e, M_\ell)$-*downswitching* to be a matching $Y$ of $G$ satisfying

(i) $e \subseteq V(Y)$;

(ii) $Y$ contains precisely $\ell - 1$ edges intersecting $e$;

(iii) for all $e' \in M_\ell$, we have either $e' \subseteq V(Y)$ or $e' \cap V(Y) = \emptyset$.

Note that if $Y$ is an $(e, M_\ell)$-upswitching, then we can obtain a new perfect matching $M'$ from $M_\ell$ by replacing $M_\ell[V(Y)]$ with $Y$. Then $M'$ contains exactly $\ell + 1$ edges intersecting $e$. Similarly, if $Y$ is an $(e, M_\ell)$-downswitching, then $M'$ has exactly $\ell - 1$ edges intersecting $e$.

**Lemma 2.4.1.** *Let $1/n \ll 1/C \ll \gamma, 1/k$, where $k \geq 2$ and $k \mid n$. Let $G$ be an $(n, k, \gamma)$-graph, and let $M$ be a uniformly random perfect matching of $G$. Then for each $e \in E(G)$, we have*

$$\frac{1}{Cn^{k-1}} \leq \mathbb{P}\left[e \in M\right] \leq \frac{C}{n^{k-1}}. \tag{2.4.2}$$

**Proof.** Choose new integers $m$ and $B$ satisfying $1/n \ll 1/C \ll 1/B \ll 1/m \ll \gamma, 1/k$, and fix $e \in E(G)$. For each integer $\ell \in [k]$, let $\mathcal{M}_\ell$ be the set of perfect matchings of $G$ containing precisely $\ell$ edges intersecting $e$. Note that $\mathbb{P}\left[e \in M\right] = |\mathcal{M}_1|/(|\mathcal{M}_1| + \cdots + |\mathcal{M}_k|)$ and recall that there is at least one perfect matching of $G$ since $G$ is $\gamma$-Dirac (so the denominator here is nonzero). We first bound $|\mathcal{M}_\ell|/|\mathcal{M}_{\ell+1}|$ from above and below for each $\ell \in [k-1]$, and (2.4.2) will follow quickly. Let $\ell \in [k-1]$. We define an auxiliary bipartite multigraph $G^{\uparrow}_{e,\ell}$ with vertex bipartition $(\mathcal{M}_\ell, \mathcal{M}_{\ell+1})$. For each $M_\ell \in \mathcal{M}_\ell$ and each $(e, M_\ell)$-upswitching $Y$ of *size* $m$ (containing precisely $m$ edges), we add an edge in $G^{\uparrow}_{e,\ell}$ from $M_\ell$ to the matching $M_{\ell+1} \in \mathcal{M}_{\ell+1}$ obtained by replacing $M_\ell[V(Y)]$ with $Y$. Write $\delta^{e,\uparrow}_{\mathcal{M}_\ell}$ to denote the minimum degree in $G^{\uparrow}_{e,\ell}$ over all $M_\ell \in \mathcal{M}_\ell$, and write $\Delta^{e,\uparrow}_{\mathcal{M}_{\ell+1}}$ to denote the maximum degree in $G^{\uparrow}_{e,\ell}$ over all $M_{\ell+1} \in \mathcal{M}_{\ell+1}$. By double-counting $|E(G^{\uparrow}_{e,\ell})|$, we obtain $|\mathcal{M}_\ell|/|\mathcal{M}_{\ell+1}| \le \Delta^{e,\uparrow}_{\mathcal{M}_{\ell+1}}/\delta^{e,\uparrow}_{\mathcal{M}_\ell}$. To bound $\Delta^{e,\uparrow}_{\mathcal{M}_{\ell+1}}$, we fix $M_{\ell+1} \in \mathcal{M}_{\ell+1}$ and bound the number of pairs $(M_\ell, Y)$, where $M_\ell \in \mathcal{M}_\ell$ and $Y$ is an $(e, M_\ell)$-upswitching of size $m$ that produces $M_{\ell+1}$. Note that any such $Y$ must contain all vertices in the $\ell+1$ edges of $M_{\ell+1}$ intersecting $e$, and there are at most $n^{m-\ell-1}$ choices for the other $m-\ell-1$ edges of $M_{\ell+1}$ whose vertices to include in $V(Y)$. Once $V(Y)$ is fixed, there are at most $(mk)!$ choices for $M_\ell[V(Y)]$ (and hence for $M_\ell$). Thus, we have $\Delta^{e,\uparrow}_{\mathcal{M}_{\ell+1}} \le (mk)!n^{m-\ell-1}$.

To bound $\delta^{e,\uparrow}_{\mathcal{M}_\ell}$, we fix $M_\ell \in \mathcal{M}_\ell$ and bound the number of $(e, M_\ell)$-upswitchings of size $m$ from below. Let $U(M_\ell) \coloneqq \{e' \in M_\ell \colon e \cap e' \neq \emptyset\}$. Note that any $(e, M_\ell)$-upswitching $Y$ of size $m$ must include all the vertices in $U(M_\ell)$, and there are $\binom{n/k-\ell}{m-\ell}$ choices for the remaining $m-\ell$ edges of $M_\ell$ whose vertices to include in $V(Y)$. We apply Proposition 2.3.3 (with $\mathcal{P} = M_\ell$, $\mathcal{P}_0 = U(M_\ell)$, and with $m-\ell$, $k$, $\ell$ playing the roles of $m$, $\ell$, $t$, respectively) to deduce that there are at least $(1-e^{-\sqrt{m-\ell}})\binom{n/k-\ell}{m-\ell} \ge (mk)^{-m}n^{m-\ell}$ choices of $X \subseteq M_\ell \setminus U(M_\ell)$ of size $m-\ell$ such that $G[V(X \cup U(M_\ell))]$ is $\gamma/2$-Dirac. Note that for each such $X$, we may first choose a matching $U'$ of size $\ell+1$ in $G[V(X \cup U(M_\ell))]$ such that $e \subseteq V(U')$ and $e$ intersects every edge in $U'$, and then choose a perfect matching $Y'$ of $G[V(X \cup U(M_\ell)) \setminus V(U')]$. Then $Y \coloneqq Y' \cup U'$ is an $(e, M_\ell)$-upswitching of size $m$, unique to this choice of $X$. We deduce that $\delta^{e,\uparrow}_{\mathcal{M}_\ell} \ge (mk)^{-m}n^{m-\ell}$, and conclude that $|\mathcal{M}_\ell|/|\mathcal{M}_{\ell+1}| \le (mk)!(mk)^m/n \le B/n$.

We now bound the terms $|\mathcal{M}_\ell|/|\mathcal{M}_{\ell+1}|$ from below analogously. Let $\ell \in [k-1]$. We define an auxiliary bipartite multigraph $G^{\downarrow}_{e,\ell+1}$ with vertex bipartition $(\mathcal{M}_\ell, \mathcal{M}_{\ell+1})$. For each $M_{\ell+1} \in \mathcal{M}_{\ell+1}$ and each $(e, M_{\ell+1})$-downswitching $Y$ of size $m$, we add an edge in $G^{\downarrow}_{e,\ell+1}$ from $M_{\ell+1}$ to the matching $M_\ell \in \mathcal{M}_\ell$ obtained by replacing $M_{\ell+1}[V(Y)]$ with $Y$. Let $\delta^{e,\downarrow}_{\mathcal{M}_{\ell+1}}$ denote the minimum degree in $G^{\downarrow}_{e,\ell+1}$ among all $M_{\ell+1} \in \mathcal{M}_{\ell+1}$, and let $\Delta^{e,\downarrow}_{\mathcal{M}_\ell}$ denote the maximum degree in $G^{\downarrow}_{e,\ell+1}$ among all $M_\ell \in \mathcal{M}_\ell$. It is easy to see that $\Delta^{e,\downarrow}_{\mathcal{M}_\ell} \le (mk)!n^{m-\ell}$. Now fix $M_{\ell+1} \in \mathcal{M}_{\ell+1}$ and let $U(M_{\ell+1}) \coloneqq \{e' \in M_{\ell+1} \colon e \cap e' \neq \emptyset\}$. We apply Lemma 2.3.3 again (with $\mathcal{P} = M_{\ell+1}$, $\mathcal{P}_0 = U(M_{\ell+1})$, and with $m-\ell-1$, $k$, $\ell+1$ playing the roles of $m$, $\ell$, $t$, respectively) to deduce that $\delta^{e,\downarrow}_{\mathcal{M}_{\ell+1}} \ge (mk)^{-m}n^{m-\ell-1}$, and thus $|\mathcal{M}_\ell|/|\mathcal{M}_{\ell+1}| \ge 1/((mk)!(mk)^m n) \ge 1/(Bn)$.

Finally, note that

$$\begin{aligned}\mathbb{P}\left[e \in M\right] &= \frac{|\mathcal{M}_1|}{|\mathcal{M}_1| + \cdots + |\mathcal{M}_k|} \le \frac{|\mathcal{M}_1|}{|\mathcal{M}_k|} = \frac{|\mathcal{M}_1|}{|\mathcal{M}_2|} \cdot \frac{|\mathcal{M}_2|}{|\mathcal{M}_3|} \cdots \frac{|\mathcal{M}_{k-1}|}{|\mathcal{M}_k|} \le \frac{B^{k-1}}{n^{k-1}} \\ &\le \frac{C}{n^{k-1}},\end{aligned}$$

and similarly $\mathbb{P}\left[e \in M\right] \ge |\mathcal{M}_1|/(k|\mathcal{M}_k|) \ge 1/(kB^{k-1}n^{k-1}) \ge 1/(Cn^{k-1})$. □

Finally, we use Lemma 2.4.1 to show that an $(n,k,\gamma)$-graph admits a normal perfect fractional matching.

**Lemma 2.4.2.** *Let $1/n \ll 1/C \ll \gamma, 1/k$, where $k \ge 2$, and let $G$ be an $(n,k,\gamma)$-graph. Then there exists a $C$-normal perfect fractional matching of $G$.*

**Proof.** Let $i$ be the unique integer in $\{0,1,\ldots,k-1\}$ satisfying $n \equiv i \bmod k$. For each $\mathbf{S} \in \binom{V}{i}$, let $G_\mathbf{S} \coloneqq G - \mathbf{S}$. We define an edge weighting $\mathbf{x}_\mathbf{S} \colon E(G_\mathbf{S}) \to \mathbb{R}^+$ by setting $\mathbf{x}_\mathbf{S}(e) \coloneqq \mathbb{P}\left[e \in M_\mathbf{S}\right]$ for each $e \in E(G_\mathbf{S})$, where $M_\mathbf{S}$ is a uniformly random perfect matching in $G_\mathbf{S}$. We define an edge weighting $\mathbf{x} \colon E(G) \to \mathbb{R}^+$ by setting

$$\mathbf{x}\left(e\right) \coloneqq \binom{n-1}{i}^{-1} \sum_{\mathbf{S} \in \binom{V}{i}} \mathbf{x}_\mathbf{S}(e),$$

for each $e \in E(G)$, where we set $\mathbf{x}_\mathbf{S}(e)$ to be 0 for each $\mathbf{S}$ such that $e \notin E(G_\mathbf{S})$. Then, by Lemma 2.4.1, $\mathbf{x}$ is the desired $C$-normal perfect fractional matching of $G$. $\square$

## 2.5 Counting short paths

The aim of this section is to prove the following lemma, which guarantees many short tight paths in a $\gamma$-Dirac $k$-graph $G$, such that the $\gamma$-Dirac property of the graph $G'$ obtained from deleting any such path is still essentially intact.

**Lemma 2.5.1** (Iteration Lemma)**.** *Let $1/n \ll c \ll \gamma, 1/k$ where $k \geq 2$, let $G$ be an $(n, k, \gamma)$-graph, and let $\overrightarrow{\mathbf{S}} \in (V)_{k-1}$. There exists a set $\mathcal{P}$ of $\sqrt{n}$-paths in $G$ such that:*

(i) $|\mathcal{P}| \geq (cn)^{\sqrt{n}}$;

(ii) $\overrightarrow{\mathbf{S}}$ *is an end of each* $P \in \mathcal{P}$;

(iii) *if $P \in \mathcal{P}$ and $\overrightarrow{\mathbf{T}}$ is the non-$\overrightarrow{\mathbf{S}}$ end of $P$, then $G' \coloneqq G - (V(P) \setminus \mathbf{T})$ satisfies* $\delta(G') \geq (1/2 + \gamma - n^{-2/3})(n - \sqrt{n})$.

We now provide some important definitions and sketch the proof of Lemma 2.5.1. We then collect together a number of technical lemmas, and finally use these results to prove Lemma 2.5.1.

### 2.5.1 Random walk notation

We first define some random walks which will be of central importance to the proof of Lemma 2.5.1. Let $k \geq 2$ be an integer, let $G$ be a $k$-graph, and let $\mathbf{x}\colon E(G) \to \mathbb{R}^{+}$ be a positive edge weighting function. Each of our random walks $\mathcal{Z} = (Z_{-(k-2)}, Z_{-(k-3)}, \ldots)$ on $V$ will begin with an ordered $(k-1)$-tuple $(Z_{-(k-2)}, \ldots, Z_0)$

chosen according to some probability distribution $\mu\colon (V)_{k-1} \to \mathbb{R}^+$. We say that $\mu$ is the *initial distribution* of $\mathcal{Z}$. Random vertices will then be added one-by-one to each $\mathcal{Z}$ according to the transition probabilities of $\mathcal{Z}$. Suppose we are given a random walk $\mathcal{Z} = (Z_{-(k-2)}, \dots, Z_{j-1})$ on $V$, up to time $j-1$. Then we say that *semiviable* vertices for step $j$ are those vertices $v \in V$ satisfying $\{Z_{j-(k-1)}, \dots, Z_{j-1}\} \cup \{v\} \in E(G)$. We say that *viable* vertices for step $j$ are those vertices $v \in V$ which are semiviable and satisfy $v \notin \{Z_{-(k-2)}, \dots, Z_{j-1}\}$. Let $Q_j$ and $R_j$ denote the sets of semiviable and viable vertices for step $j$, given the random walk up to time $j-1$, respectively.

We say that a random walk $\mathcal{X} = (X_{-(k-2)}, X_{-(k-3)}, \dots)$ on the vertices of $G$, with any initial distribution $\mu$, is a *self-avoiding* $\mathbf{x}$*-walk* to mean that the transition probabilities of $\mathcal{X}$ for $j \geq 1$ are defined for all $v \in V$ by

$$\mathbb{P}\left[X_j = v \mid X_{-(k-2)}, \dots, X_{j-1}\right] \coloneqq \frac{\mathbf{x}\left(\{X_{j-(k-1)}, \dots, X_{j-1}\} \cup \{v\}\right) \mathbb{1}_{R_j}(v)}{\sum_{w \in R_j} \mathbf{x}\left(\{X_{j-(k-1)}, \dots, X_{j-1}\} \cup \{w\}\right)},$$

whenever $R_j$ is non-empty, otherwise we terminate the walk. Here, and throughout, we define $\mathbf{x}(S) \coloneqq 0$ for any $S \notin E(G)$. Note that $\mathcal{X} = (X_{-(k-2)}, X_{-(k-3)}, \dots)$ is equivalent to the random walk $\mathcal{X}^{(k-1)} = (\overrightarrow{\mathbf{X}}_0, \overrightarrow{\mathbf{X}}_1, \dots)$, where each $\overrightarrow{\mathbf{X}}_i$ is the ordered $(k-1)$-tuple $\overrightarrow{\mathbf{X}}_i = (X_{i-(k-2)}, \dots, X_i)$. We thus refer to both $\mathcal{X}$ and $\mathcal{X}^{(k-1)}$ as the self-avoiding $\mathbf{x}$-walk on $G$ with initial distribution $\mu$, since they are reformulations of each other.

Suppose now that $G$ has no isolated $(k-1)$-tuples (this will always be true for us). We say that a random walk $\mathcal{Y} = (Y_{-(k-2)}, Y_{-(k-3)}, \dots)$ on the vertices of $G$ (or $\mathcal{Y}^{(k-1)} = (\overrightarrow{\mathbf{Y}}_0, \overrightarrow{\mathbf{Y}}_1, \dots)$ on $(V)_{k-1}$, where $\overrightarrow{\mathbf{Y}}_i \coloneqq (Y_{i-(k-2)}, \dots, Y_i)$), with any initial distribution $\mu$, is a *simple* $\mathbf{x}$*-walk* to mean that the transition probabilities of $\mathcal{Y}$

for $j \geq 1$ are defined by

$$\mathbb{P}\left[Y_j = v \mid Y_{-(k-2)}, \dots, Y_{j-1}\right] := \frac{\mathbf{x}\left(\{Y_{j-(k-1)}, \dots, Y_{j-1}\} \cup \{v\}\right) \mathbb{1}_{Q_j}(v)}{\sum_{w \in Q_j} \mathbf{x}\left(\{Y_{j-(k-1)}, \dots, Y_{j-1}\} \cup \{w\}\right)}, \quad (2.5.1)$$

for all $v \in V$. Note that $\mathcal{Y}^{(k-1)}$ is a Markov chain on $(V)_{k-1}$ because the transition probabilities at any time depend only on the current state. When the stationary distribution of $\mathcal{Y}^{(k-1)}$ exists and is unique, we denote it by $\pi$, and we say that a simple $\mathbf{x}$-walk $\mathcal{W}^{(k-1)} = (\overrightarrow{\mathbf{W}}_0, \overrightarrow{\mathbf{W}}_1, \dots)$ on the ordered $(k-1)$-tuples of $V$ is the *stationary* $\mathbf{x}$*-walk* on $G$ if the initial distribution is $\pi$. Again, we have that $\mathcal{W}^{(k-1)} = (\overrightarrow{\mathbf{W}}_0, \overrightarrow{\mathbf{W}}_1, \dots)$ is equivalent to the walk $\mathcal{W} = (W_{-(k-2)}, W_{-(k-3)}, \dots)$ on the vertices of $G$, where each $\overrightarrow{\mathbf{W}}_i$ is the ordered $(k-1)$-tuple $\overrightarrow{\mathbf{W}}_i = (W_{i-(k-2)}, \dots, W_i)$. We also call $\mathcal{W}$ the stationary $\mathbf{x}$-walk, and use the vertex (or tuple) version whenever it is more convenient. Finally, whenever the initial distribution $\mu$ of $\mathcal{X}$ (or $\mathcal{Y}$) satisfies $\mu(\overrightarrow{\mathbf{S}}) = 1$ for some $\overrightarrow{\mathbf{S}} \in (V)_{k-1}$, we say that $\mathcal{X}$ (or $\mathcal{Y}$) has *starting tuple* $\overrightarrow{\mathbf{S}}$.

### 2.5.2 Further notation and sketch of the proof of Lemma 2.5.1

We now describe our approach to proving Lemma 2.5.1. Introduce a new constant $C$ satisfying $1/n \ll c \ll 1/C \ll \gamma, 1/k$, and let $G$ be an $(n, k, \gamma)$-graph. By Lemma 2.4.2, there exists a $C$-normal perfect fractional matching $\mathbf{x}$ of $G$. We fix such a $C$-normal $\mathbf{x}$ throughout this proof sketch. We will analyse a self-avoiding $\mathbf{x}$-walk $\mathcal{X}$ on $G$ with starting tuple $\overrightarrow{\mathbf{S}} \in (V)_{k-1}$. We stop the walk after time $\kappa := \sqrt{n}$, so that we may write $\mathcal{X} = (X_{-(k-2)}, \dots, X_\kappa)$. Note that each outcome of $\mathcal{X}$ will correspond to a tight $\kappa$-path in $G$, with $\overrightarrow{\mathbf{S}}$ as one end.

We define $V_j := V \setminus \{X_{-(k-2)}, \dots, X_{j-(k-1)}\}$ to be the set of all vertices of $G$

except for all vertices $\mathcal{X}^{(k-1)}$ has visited strictly before $\overrightarrow{\mathbf{X}}_j$. We say that $V_j$ is the *residual vertex set* of $G$ at time $j$. We also define $G_j \coloneqq G[V_j]$ and say that $G_j$ is the *residual graph* at time $j$. We also write $\mathcal{X}(j)$ to denote the walk $\mathcal{X}$ up to time $j$, specifically $\mathcal{X}(j) \coloneqq (X_{-(k-2)}, \ldots, X_j)$.

We will show that it is likely that the $\gamma$-Dirac property of the residual graph $G_\kappa$ is still essentially intact, by showing that it is likely that the vertices that $\mathcal{X}$ visits look roughly like a uniformly random subset of $V$ (see Lemma 2.5.7). For this, we will use the following 'tracking functions' to monitor the progress of $\mathcal{X}$, with respect to how the codegree of each $(k-1)$-tuple in the residual graph deteriorates over time.

For each unordered $(k-1)$-tuple $\mathbf{S} \in \binom{V}{k-1}$, we define a function $g_{\mathbf{S}} \colon V \to \mathbb{R}^+$ by setting $g_{\mathbf{S}}(v) \coloneqq \mathbb{1}_{N_G(\mathbf{S})}(v)$ for each $v \in V$, so that, in particular, if $\mathbf{S} \subseteq V_j$ then $g_{\mathbf{S}}(V_j) = d_{G_j}(\mathbf{S})$. We call the set $\mathcal{F} \coloneqq \{g_{\mathbf{S}} \colon \mathbf{S} \in \binom{V}{k-1}\}$ the set of *tracking functions* of $G$. We say that $\mathcal{X} = (X_{-(k-2)}, \ldots, X_\kappa)$ is *good* if

$$\sum_{i=-(k-2)}^{\kappa} g_{\mathbf{S}}(X_i) = \frac{\kappa}{n} g_{\mathbf{S}}(V) \pm n^{3/10} \qquad \text{for all } g_{\mathbf{S}} \in \mathcal{F}. \tag{2.5.2}$$

Thus, to say that $\mathcal{X}$ is good is to say that the set of $\kappa + k - 1$ vertices that $\mathcal{X}$ visits look roughly like a uniformly random subset of $V$, with respect to the codegrees of all $(k-1)$-tuples. In particular, for all $(k-1)$-tuples $\mathbf{S} \in \binom{V}{k-1}$, the proportion of vertices of $N_G(\mathbf{S})$ visited by $\mathcal{X}$ is approximately $\kappa/n$, and this is the property that will allow us to deduce condition (iii) of Lemma 2.5.1.

Let $(\overrightarrow{\mathbf{X}}_a, \ldots, \overrightarrow{\mathbf{X}}_b)$ be a suitable interval of $\mathcal{X}^{(k-1)}$, and let $\mathcal{Y}^{(k-1)} = (\overrightarrow{\mathbf{Y}}_0 = \overrightarrow{\mathbf{X}}_a, \overrightarrow{\mathbf{Y}}_1, \ldots)$ be the simple $\mathbf{x}$-walk on $G_a$ with starting tuple $\overrightarrow{\mathbf{X}}_a$. To show that the walk $\mathcal{X}$ is likely to be good, the following will be the main steps:

(i) Firstly we show that the behaviour of $(\overrightarrow{\mathbf{X}}_a, \dots)$ follows closely that of $\mathcal{Y}^{(k-1)}$ by exhibiting a coupling of the two walks such that the probability of $\mathcal{X}^{(k-1)}$ and $\mathcal{Y}^{(k-1)}$ being different is acceptably small, provided $b-a$ is small.

(ii) Next we see that $\mathcal{Y}^{(k-1)}$ *mixes* (converges to its stationary distribution $\pi$) rapidly.

(iii) We also show that the stationary $\mathbf{x}$-walk $\mathcal{W} = (W_0, W_1, \dots)$ is such that $\mathbb{P}[W_i = v] \approx 1/|V_a|$ for each $v \in V_a$.

Putting the above together, we see that even for small $q$, the distribution of each $X_i$, given the walk to time $i-q$, is typically close to the uniform distribution on $V_{i-q}$, and thus for each tracking function $g \in \mathcal{F}$, we have $\mathbb{E}[g(X_i)] \approx \frac{1}{n-(i-q)} g(V_{i-q}) \approx \frac{1}{n} g(V)$. Lastly, then:

(iv) We show that the actual values of the quantities $\sum_{a \le i \le b} g(X_i)$ are very likely to be close to their expectations, by using Lemma 2.3.1.

This completes the sketch of the proof that $\mathcal{X}$ is likely to be good. It only remains to count the number of good walks (outcomes of) $\mathcal{X}$. The count will be obtained by simply dividing a lower bound for the probability that $\mathcal{X}$ is good, by an upper bound for the probability of obtaining any specific outcome of $\mathcal{X}$. This completes the sketch of the proof of Lemma 2.5.1.

### 2.5.3 Random walk analysis

In this subsection we collect some of the tools that we will use to prove Lemma 2.5.1. We firstly define some convenient terminology for edge weightings. Let $\mathbf{x} \colon E(G) \to \mathbb{R}^+$ be a positive edge weighting of a $k$-graph $G$. We say that $\mathbf{x}$ is $a_1$-*lower-balanced* if for all non-isolated $\mathbf{S} \in \binom{V}{k-1}$ and all $v \in N_G(\mathbf{S})$ we have that

$\mathbf{x}\left(\mathbf{S}\cup\{v\}\right)/\left(\sum_{v'\in V}\mathbf{x}\left(\mathbf{S}\cup\{v'\}\right)\right) \geq a_1$. That is, all possible $\mathbf{x}$-walk transition probabilities are bounded below by $a_1$. Similarly, we say that $\mathbf{x}$ is $a_2$-*upper-balanced* if we have $\mathbf{x}\left(\mathbf{S}\cup\{v\}\right)/\left(\sum_{v'\in V}\mathbf{x}\left(\mathbf{S}\cup\{v'\}\right)\right) \leq a_2$ for all non-isolated $\mathbf{S} \in \binom{V}{k-1}$ and all $v \in N_G(\mathbf{S})$. We say that $\mathbf{x}$ is $(a_1, a_2)$-*balanced* if $\mathbf{x}$ is $a_1$-lower-balanced and $a_2$-upper-balanced. We now give a simple result which shows that we may couple the self-avoiding $\mathbf{x}$-walk and the simple $\mathbf{x}$-walk to behave very similarly over small distances.

**Lemma 2.5.2.** *Let $k \geq 2$, let $G$ be a $k$-graph, let $\overrightarrow{\mathbf{M}} \in (V)_{k-1}$, and let $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ be a positive $r$-upper-balanced edge weighting. Let $\mathcal{X} = (X_{-(k-2)}, \dots)$ and $\mathcal{Y} = (Y_{-(k-2)}, \dots)$ be, respectively, the self-avoiding $\mathbf{x}$-walk and the simple $\mathbf{x}$-walk on $G$, each with starting tuple $\overrightarrow{\mathbf{M}}$. Then for any positive integer $q \leq \delta(G)$, we have*

$$d_{TV}(X_q, Y_q) \leq q^2 r.$$

**Proof.** If $1 \leq i \leq \delta(G)$, and $\mathcal{X}$ and $\mathcal{Y}$ agree up to time $i-1$, say $X_j = Y_j = v_j \in V$ for each $j \in \{-k+2, \dots, i-1\}$, then we couple at the next step so that $X_i$ coincides with $Y_i$ whenever the choice of $Y_i$ is a viable choice for $X_i$, which is to say that $Y_i$ is not a vertex already seen. So with this coupling, for any positive integer $i \leq \delta(G)$ we have

$$\mathbb{P}\left[X_i \neq Y_i \mid X_j = Y_j \text{ for all } j \in \{-k+2, \dots, i-1\}\right] = \mathbb{P}\left[Y_i \in \{v_{-(k-2)}, \dots, v_{i-k}\}\right],$$

which is at most $ir$. Thus for any positive integer $q \leq \delta(G)$ we obtain

$$\mathbb{P}\left[X_q \neq Y_q\right] \leq \sum_{i=1}^{q} \mathbb{P}\left[X_i \neq Y_i \mid X_j = Y_j \text{ for all } j \in \{-k+2, \dots, i-1\}\right] \leq q^2 r.$$

The desired result now follows from (2.3.1). $\square$

We now aim to show that $\mathcal{Y}^{(k-1)}$ mixes rapidly. The key part of the proof will be the following argument that $\mathcal{Y}^{(k-1)}$ has many different choices for how to arrive at a specified target ordered tuple $\overrightarrow{\mathbf{T}} \in (V)_{k-1}$ in a fixed number of steps.

**Lemma 2.5.3.** *Let $1/n \ll \zeta \ll 1/\ell \ll \gamma, 1/k$, where $k \geq 2$, and let $G$ be an $(n,k,\gamma)$-graph. For any $\overrightarrow{\mathbf{S}}, \overrightarrow{\mathbf{T}} \in (V)_{k-1}$, there are at least $\zeta n^{\ell-(k-1)}$ $\ell$-walks from $\overrightarrow{\mathbf{S}}$ to $\overrightarrow{\mathbf{T}}$ in $G$. Further, if $\mathbf{S} \cap \mathbf{T} = \emptyset$, then there are at least $\zeta n^{\ell-(k-1)}$ $\ell$-paths in $G$ which connect $\overrightarrow{\mathbf{S}}$ and $\overleftarrow{\mathbf{T}}$.*

**Proof.** Let $\overrightarrow{\mathbf{S}}, \overrightarrow{\mathbf{T}} \in (V)_{k-1}$, set $a := \ell - (k-1)$, let $\mathcal{Q} := \binom{V \setminus (\mathbf{S} \cup \mathbf{T})}{a}$, let $G_Q := G[\mathbf{S} \cup Q \cup \mathbf{T}]$ for each $Q \in \mathcal{Q}$, and define $\mathcal{Q}' := \{Q \in \mathcal{Q} \colon G_Q \text{ is } (\gamma/2)\text{-Dirac}\}$. We apply Proposition 2.3.3 (with $\mathcal{P} = \{\{v\} \colon v \in V\}$, $\mathcal{P}_0 = \{\{v\} \colon v \in \mathbf{S} \cup \mathbf{T}\}$, and with $a$, 1, $|\mathbf{S} \cup \mathbf{T}|$ playing the roles of $m$, $\ell$, $t$, respectively) to deduce that $|\mathcal{Q}'| \geq |\mathcal{Q}|/2$. Fix $Q \in \mathcal{Q}'$ and write $\overrightarrow{\mathbf{T}} = (t_1, \ldots, t_{k-1})$. Since $G[Q \cup \mathbf{T}]$ is $\gamma/4$-Dirac, for each $i \in [k-1]$ in turn we may find a vertex $v_i \in Q \setminus \{v_1, \ldots, v_{i-1}\}$ such that $\{v_1, \ldots, v_i\} \cup \{t_1, \ldots, t_{k-i}\}$ is an edge. Write $\overrightarrow{\mathbf{T}'} := (v_{k-1}, \ldots, v_1)$. Thus, we obtain a $(k-1)$-path $P_Q^2$ in $G_Q$ which connects $\overrightarrow{\mathbf{T}'}$ and $\overleftarrow{\mathbf{T}}$. Since $G[\mathbf{S} \cup Q]$ is $\gamma/4$-Dirac, we may apply Lemma 2.3.9 (with $\gamma/4$ playing the role of $\gamma$) to find an $a$-path $P_Q^1$ in $G[\mathbf{S} \cup Q]$ which connects $\overrightarrow{\mathbf{S}}$ and $\overleftarrow{\mathbf{T}'}$, and the obvious concatenation of $P_Q^1$ and $P_Q^2$ is an $\ell$-walk $W_Q$ from $\overrightarrow{\mathbf{S}}$ to $\overrightarrow{\mathbf{T}}$ in $G_Q$. It is clear that these $\ell$-walks $W_Q$ are distinct for different choices of $Q \in \mathcal{Q}'$. It thus suffices to observe that $|\mathcal{Q}'| \geq \frac{1}{2}|\mathcal{Q}| \geq \zeta n^{\ell-(k-1)}$. If $\mathbf{S} \cap \mathbf{T} = \emptyset$, then the walks $W_Q$ do not revisit vertices and thus correspond to $\ell$-paths in $G$ which connect $\overrightarrow{\mathbf{S}}$ and $\overleftarrow{\mathbf{T}}$. $\square$

Lemma 2.5.3 shows that in an $(n,k,\gamma)$-graph $G$, the Markov chain $\mathcal{Y}^{(k-1)}$ is irreducible, and thus there is a unique stationary distribution $\pi$ of $\mathcal{Y}^{(k-1)}$. We will use this fact without stating it from now on. We note that it also follows from Lemma 2.5.3 that $\mathcal{Y}^{(k-1)}$ is aperiodic, which implies that the distribution of $\overrightarrow{\mathbf{Y}}_t$ converges to $\pi$ as $t \to \infty$. However, we need something stronger, namely that this convergence occurs quickly. This is achieved by the following lemma, which shows that $\mathcal{Y}^{(k-1)}$ mixes rapidly.

**Lemma 2.5.4** (Mixing Lemma)**.** *Let $1/n \ll 1/\lambda \ll \gamma, \tau, 1/k$, where $k \geq 2$, let $G$ be an $(n, k, \gamma)$-graph, let $\mu\colon (V)_{k-1} \to \mathbb{R}$ be a probability distribution, and let $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ be a positive $(\tau/n)$-lower-balanced edge weighting. Let $\mathcal{Y} = (Y_{-(k-2)}, \dots)$ be the simple $\mathbf{x}$-walk on $G$ with initial distribution $\mu$, and let $\mathcal{W} = (W_{-(k-2)}, \dots)$ be the stationary $\mathbf{x}$-walk on $G$. For any $\eta > 0$, if $q \geq \lambda \ln(1/\eta)$, then $d_{TV}(Y_q, W_q) < \eta$.*

**Proof.** Choose new constants $\zeta$ and $\ell$ satisfying $1/n \ll 1/\lambda \ll \zeta \ll 1/\ell \ll \gamma, \tau, 1/k$. We proceed by using Lemma 2.5.3 to show that for two simple $\mathbf{x}$-walks $\mathcal{Z}^{(k-1)}$ and $\mathcal{Z}'^{(k-1)}$ on $G$ given any initial distributions, we can find a coupling such that $\mathcal{Z}^{(k-1)}$ and $\mathcal{Z}'^{(k-1)}$ are relatively likely to meet after $\ell$ steps. Using this, we then exhibit a coupling of $\mathcal{Y}^{(k-1)}$ and $\mathcal{W}^{(k-1)}$ that will allow us to use (2.3.1) to upper bound $d_{TV}(Y_q, W_q)$.

Let $\mathcal{Z}^{(k-1)}$ be a simple $\mathbf{x}$-walk on $G$ with any initial distribution $\phi$, and fix any $\overrightarrow{\mathbf{T}} \in (V)_{k-1}$. By Lemma 2.5.3, $G$ has at least $\zeta n^{\ell-(k-1)}$ $\ell$-walks from $\overrightarrow{\mathbf{S}}$ to $\overrightarrow{\mathbf{T}}$, for each $\overrightarrow{\mathbf{S}} \in (V)_{k-1}$, and thus:

$$\begin{aligned}\mathbb{P}\left[\overrightarrow{\mathbf{Z}}_\ell = \overrightarrow{\mathbf{T}}\right] &= \sum_{\overrightarrow{\mathbf{S}} \in (V)_{k-1}} \phi(\overrightarrow{\mathbf{S}})\mathbb{P}\left[\overrightarrow{\mathbf{Z}}_\ell = \overrightarrow{\mathbf{T}} \mid \overrightarrow{\mathbf{Z}}_0 = \overrightarrow{\mathbf{S}}\right] \\ &\geq \sum_{\overrightarrow{\mathbf{S}} \in (V)_{k-1}} \phi(\overrightarrow{\mathbf{S}})\zeta n^{\ell-(k-1)}(\tau/n)^\ell = \zeta\tau^\ell/n^{k-1}.\end{aligned}$$

Thus we may construct a coupling of any pair of simple $\mathbf{x}$-walks $\mathcal{Z}^{(k-1)}$ and $\mathcal{Z}'^{(k-1)}$ such that $\mathbb{P}\left[\overrightarrow{\mathbf{Z}}_\ell = \overrightarrow{\mathbf{Z}}'_\ell = \overrightarrow{\mathbf{T}}\right] \geq \zeta\tau^\ell/n^{k-1}$ for all $\overrightarrow{\mathbf{T}} \in (V)_{k-1}$. Under this coupling, we have

$$\mathbb{P}\left[\overrightarrow{\mathbf{Z}}_\ell = \overrightarrow{\mathbf{Z}'}_\ell\right] \geq \frac{\zeta\tau^\ell}{n^{k-1}}\left|(V)_{k-1}\right| \geq \zeta\tau^\ell\left(\frac{n-(k-2)}{n}\right)^{k-1} \geq \zeta\tau^\ell/2. \tag{2.5.3}$$

We now construct a coupling of $\mathcal{Y}^{(k-1)}$ and $\mathcal{W}^{(k-1)}$ as follows. We partition the time steps into consecutive intervals of length $\ell$. In the first interval, we couple $\mathcal{Y}^{(k-1)}$ and $\mathcal{W}^{(k-1)}$ as in (2.5.3), so that $\mathbb{P}\left[\overrightarrow{\mathbf{Y}}_\ell = \overrightarrow{\mathbf{W}}_\ell\right] \geq \zeta\tau^\ell/2$. If $\overrightarrow{\mathbf{Y}}_\ell = \overrightarrow{\mathbf{W}}_\ell$, then we couple $\mathcal{Y}^{(k-1)}$ and $\mathcal{W}^{(k-1)}$ such that $\overrightarrow{\mathbf{Y}}_t = \overrightarrow{\mathbf{W}}_t$ for all $t \geq \ell$. Otherwise we again couple $\mathcal{Y}^{(k-1)}$ and $\mathcal{W}^{(k-1)}$ in the second time interval, as in (2.5.3), so that $\mathbb{P}\left[\overrightarrow{\mathbf{Y}}_{2\ell} = \overrightarrow{\mathbf{W}}_{2\ell} \mid \overrightarrow{\mathbf{Y}}_\ell \neq \overrightarrow{\mathbf{W}}_\ell\right] \geq \zeta\tau^\ell/2$. One can easily check that repeating this process yields a valid coupling of $\mathcal{Y}^{(k-1)}$ and $\mathcal{W}^{(k-1)}$. Note that, with this coupling and for any $q$ which is sufficiently large compared to $\ell$, we have

$$\mathbb{P}\left[\overrightarrow{\mathbf{Y}}_q \neq \overrightarrow{\mathbf{W}}_q\right] \leq \prod_{k=1}^{q/\ell}\mathbb{P}\left[\overrightarrow{\mathbf{Y}}_{k\ell} \neq \overrightarrow{\mathbf{W}}_{k\ell}\,\middle|\, \bigcap_{j\leq k-1}\{\overrightarrow{\mathbf{Y}}_{j\ell} \neq \overrightarrow{\mathbf{W}}_{j\ell}\}\right] \leq (1-\zeta\tau^\ell/2)^{q/\ell}.$$

We use this coupling and apply (2.3.1) to obtain

$$d_{TV}(Y_q, W_q) \leq \mathbb{P}\left[Y_q \neq W_q\right] \leq \mathbb{P}\left[\overrightarrow{\mathbf{Y}}_q \neq \overrightarrow{\mathbf{W}}_q\right] \leq (1-\zeta\tau^\ell/2)^{q/\ell} \leq \exp(-\zeta\tau^\ell q/2\ell) < \eta,$$

provided $q \geq \lambda\ln(1/\eta)$. □

It will be useful to have an explicit formula for the stationary distribution $\pi$ of $\mathcal{Y}^{(k-1)}$, and so we obtain this now. (Observe that the simple $\mathbf{x}$-walk $\mathcal{Y}^{(k-1)}$ on $(V)_{k-1}$ is in general not a symmetric Markov chain.)

**Proposition 2.5.5.** *Let $1/n \ll \gamma \ll 1/k$ where $k \geq 2$, and let $G$ be an $(n,k,\gamma)$-graph. Let $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ be a positive edge weighting of $G$, and for each $\overrightarrow{\mathbf{M}} \in (V)_{k-1}$, define*

$$\pi(\overrightarrow{\mathbf{M}}) := \frac{\sum_{v\in V} \mathbf{x}\left(\mathbf{M} \cup \{v\}\right)}{\sum_{\overrightarrow{\mathbf{B}}\in (V)_{k-1}} \sum_{v \in V} \mathbf{x}\left(\mathbf{B} \cup \{v\}\right)}. \tag{2.5.4}$$

*Then $\pi$ is the unique stationary distribution of the simple $\mathbf{x}$-walk $\mathcal{Y}^{(k-1)}$ on $(V)_{k-1}$.*

**Proof.** By standard results on the stationary distribution of a Markov chain (see [94, Proposition 1.20] for example), it suffices to prove that $\pi\colon (V)_{k-1} \to \mathbb{R}^+$ as defined in (2.5.4) is a probability distribution on $(V)_{k-1}$, and that

$$\sum_{\overrightarrow{\mathbf{S}}\in (V)_{k-1}} \pi(\overrightarrow{\mathbf{S}})P(\overrightarrow{\mathbf{S}}, \overrightarrow{\mathbf{T}}) = \pi(\overrightarrow{\mathbf{T}}) \qquad \text{for all } \overrightarrow{\mathbf{T}} \in (V)_{k-1}, \tag{2.5.5}$$

where $P(\overrightarrow{\mathbf{S}}, \overrightarrow{\mathbf{T}})$ denotes the (one-step) transition probability of $\mathcal{Y}^{(k-1)}$ from $\overrightarrow{\mathbf{S}}$ to $\overrightarrow{\mathbf{T}}$. It follows quickly from (2.5.4) that $\pi$ is a probability distribution on $(V)_{k-1}$, and (2.5.5) follows from applying (2.5.1) and (2.5.4). □

Next we show that, provided the edge weighting $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ is an 'almost-perfect' fractional matching, the stationary $\mathbf{x}$-walk on $G$ is such that each vertex of $G$ is roughly equally likely to be the current vertex at any time. Let $\varepsilon \in (0,1)$ and $k \geq 2$. For a $k$-graph $G$, we say that a fractional matching $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ is *$\varepsilon$-almost-perfect* if $\sum_{e\ni v} \mathbf{x}\left(e\right) \geq 1 - \varepsilon$ for all $v \in V$.

**Lemma 2.5.6.** *Let $1/n \ll \gamma, 1/k$, where $k \geq 2$, and let $G$ be an $(n,k,\gamma)$-graph. Suppose that $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ is an $n^{-2/5}$-almost-perfect fractional matching of $G$ and let $\mathcal{W} = (W_{-(k-2)}, W_{-(k-3)}, \dots)$ be the stationary $\mathbf{x}$-walk on $G$. Then for* *each $i \in \mathbb{N}$ and each $v \in V$ we have*

$$\mathbb{P}\left[W_i = v\right] = (1 \pm n^{-1/3}) \cdot \frac{1}{n}. \tag{2.5.6}$$

**Proof.** We reformulate $\mathcal{W}$ as $\mathcal{W} = (\overrightarrow{\mathbf{W}}_0, \overrightarrow{\mathbf{W}}_1, \dots)$, where $\overrightarrow{\mathbf{W}}_i \coloneqq (W_{i-(k-2)}, \dots, W_i)$. Let $v \in V$ and $i \in \mathbb{N}$. By the law of total probability, (2.5.1), and Proposition 2.5.5, we obtain:

$$\begin{aligned}
\mathbb{P}\left[W_i = v\right] &= \sum_{\overrightarrow{\mathbf{S}} \in (V)_{k-1}} \mathbb{P}\left[\overrightarrow{\mathbf{W}}_{i-1} = \overrightarrow{\mathbf{S}}\right] \mathbb{P}\left[W_i = v \mid \overrightarrow{\mathbf{W}}_{i-1} = \overrightarrow{\mathbf{S}}\right] \\
&= \sum_{\overrightarrow{\mathbf{S}} \in (V)_{k-1}} \frac{\sum_{v' \in V} \mathbf{x}\left(\mathbf{S} \cup \{v'\}\right)}{\sum_{\overrightarrow{\mathbf{B}} \in (V)_{k-1}} \sum_{v' \in V} \mathbf{x}\left(\mathbf{B} \cup \{v'\}\right)} \cdot \frac{\mathbf{x}\left(\mathbf{S} \cup \{v\}\right)}{\sum_{v' \in V} \mathbf{x}\left(\mathbf{S} \cup \{v'\}\right)} \\
&= \frac{\sum_{\mathbf{S} \in \binom{V}{k-1}} \mathbf{x}\left(\mathbf{S} \cup \{v\}\right)}{\sum_{\mathbf{B} \in \binom{V}{k-1}} \sum_{v' \in V} \mathbf{x}\left(\mathbf{B} \cup \{v'\}\right)} = \frac{\sum_{\mathbf{S} \in \binom{V}{k-1}} \mathbf{x}\left(\mathbf{S} \cup \{v\}\right)}{\sum_{v' \in V} \sum_{\mathbf{B} \in \binom{V}{k-1}} \mathbf{x}\left(\mathbf{B} \cup \{v'\}\right)} \\
&= \frac{\sum_{e \ni v} \mathbf{x}\left(e\right)}{\sum_{v' \in V} \sum_{e \ni v'} \mathbf{x}\left(e\right)}.
\end{aligned}$$

Applying $1 - n^{-2/5} \leq \sum_{e \ni v} \mathbf{x}\left(e\right) \leq 1$ for all $v \in V$, we obtain (2.5.6). □

We now show the crucial fact that, for any large set $U \subseteq V$, the probability that the self-avoiding $\mathbf{x}$-walk $\mathcal{X}$ is in $U$ after a small number of steps is roughly $|U|/n$. We will apply this fact to neighbourhoods of $(k-1)$-tuples in the proof that $\mathcal{X}$ is likely to be good. This in turn will be used to show that $\mathcal{X}$ will, in expectation, behave roughly uniformly with respect to the codegrees of all $(k-1)$-tuples.

**Lemma 2.5.7** (Uniformity Lemma)**.** *Let $1/n \ll 1/C \ll \gamma, 1/k$, where $k \geq 2$, and let $G$ be an $(n, k, \gamma)$-graph. Let $\mathbf{x} : E(G) \to \mathbb{R}^+$ be a $C$-normal, $n^{-2/5}$-almost-perfect fractional matching, let $\overrightarrow{\mathbf{S}} \in (V)_{k-1}$, and let $\mathcal{X} = (X_{-(k-2)}, X_{-(k-3)}, \dots)$ be the self-avoiding $\mathbf{x}$-walk on $G$ with starting tuple $\overrightarrow{\mathbf{S}}$. Then for any $q \in [(\ln n)^2, n^{1/5}]$*

*and any $U \subseteq V$ of size $|U| \geq n^{3/4}$, we have*

$$\mathbb{P}\left[X_q \in U\right] = (1 \pm n^{-3/10})\frac{|U|}{n}. \tag{2.5.7}$$

**Proof.** We first argue that $\mathbf{x}$ is $(1/C^2 n, 2C^2/n)$-balanced. Indeed, since $\mathbf{x}$ is $C$-normal and $G$ is $\gamma$-Dirac, we have

$$\frac{\max_{e \in E(G)} \mathbf{x}\left(e\right)}{\min_{\mathbf{S} \in \binom{V}{k-1}} \sum_{v \in V} \mathbf{x}\left(\mathbf{S} \cup \{v\}\right)} \leq \frac{C}{n^{k-1}} \cdot \frac{C n^{k-1}}{(1/2+\gamma)n} \leq \frac{2C^2}{n}.$$

The lower bound follows similarly. Now let $U \subseteq V$ be of size $|U| \geq n^{3/4}$, and fix $q \in [(\ln n)^2, n^{1/5}]$. Let $\mathcal{Y} = (Y_{-(k-2)}, Y_{-(k-3)}, \dots)$ be the simple $\mathbf{x}$-walk on $G$ with starting tuple $\overrightarrow{\mathbf{S}}$ and let $\mathcal{W} = (W_{-(k-2)}, W_{-(k-3)}, \dots)$ be the stationary $\mathbf{x}$-walk on $G$. We will show that $X_q$ is distributed similarly to $Y_q$ since $q$ is not too large, and that $Y_q$ is distributed similarly to $W_q$ since $q$ is large enough, and finally we will use Lemma 2.5.6 to show that $\mathbb{P}\left[W_q \in U\right]$ is roughly $|U|/n$.

Setting $c_1 := |\mathbb{P}\left[X_q \in U\right] - \mathbb{P}\left[Y_q \in U\right]|$ and applying Lemma 2.5.2 (with $2C^2/n$ playing the role of $r$), we obtain

$$\begin{aligned} c_1 &= \left|\sum_{v \in U}(\mathbb{P}\left[X_q = v\right] - \mathbb{P}\left[Y_q = v\right])\right| \leq \sum_{v \in V} |\mathbb{P}\left[X_q = v\right] - \mathbb{P}\left[Y_q = v\right]| \\ &\overset{(2.3.1)}{=} 2d_{TV}(X_q, Y_q) \leq 4C^2 q^2/n < \frac{1}{3} n^{-13/10}|U|. \end{aligned}$$

Consider the term $c_2 := |\mathbb{P}\left[Y_q \in U\right] - \mathbb{P}\left[W_q \in U\right]|$. We apply Lemma 2.5.4 (with $1/C^2$, $1/n^2$ playing the roles of $\tau$, $\eta$ respectively, and using $q \geq (\ln n)^2 = ((\ln n)/2) \cdot \ln(1/\eta)$) to obtain

$$c_2 \leq \sum_{v \in V} |\mathbb{P}\left[Y_q = v\right] - \mathbb{P}\left[W_q = v\right]| \overset{(2.3.1)}{=} 2d_{TV}(Y_q, W_q) < 2n^{-2} < \frac{1}{3} n^{-13/10}|U|.$$

We now apply Lemma 2.5.6 to the term $c_3 \coloneqq |\mathbb{P}[W_q \in U] - |U|/n|$ to obtain

$$c_3 = \left|\sum_{v\in U}\left(\mathbb{P}[W_q = v] - \frac{1}{n}\right)\right| \leq \sum_{v\in U}\left|\mathbb{P}[W_q = v] - \frac{1}{n}\right| \leq n^{-4/3}|U|.$$

Finally, by the triangle inequality we have $|\mathbb{P}[X_q \in U] - |U|/n| \leq c_1 + c_2 + c_3 < n^{-13/10}|U|$, which is equivalent to (2.5.7). $\square$

### 2.5.4 The walk is likely to be good

The aim of this section is to prove the following lemma, which states that under our assumptions, a self-avoiding $\mathbf{x}$-walk of length $\sqrt{n}$ is good with high probability.

**Lemma 2.5.8.** *Let $1/n \ll 1/C \ll \gamma, 1/k$, where $k \geq 2$, and let $G$ be an $(n,k,\gamma)$-graph. Let $\mathbf{x}\colon E(G) \to \mathbb{R}^+$ be a $C$-normal perfect fractional matching, let $\overrightarrow{\mathbf{S}} \in (V)_{k-1}$, let $\kappa \coloneqq \sqrt{n}$, and let $\mathcal{X} = (X_{-(k-2)}, \dots, X_\kappa)$ be a self-avoiding $\mathbf{x}$-walk on $G$ with starting tuple $\overrightarrow{\mathbf{S}}$. Then $\mathbb{P}[\mathcal{X}$ is good$] \geq 1 - 1/n$.*

To prove Lemma 2.5.8, we will need some results on the behaviour of the residual graphs $G_j$ as the walk $\mathcal{X}$ progresses, so that we can apply Lemma 2.5.7 to each $G_j$. We define $\mathbf{x}|_{G_j}$ to be the restriction of $\mathbf{x}$ to $E(G_j)$, so that $\mathbf{x}|_{G_j}\colon E(G_j) \to \mathbb{R}^+$ is a (not necessarily perfect) fractional matching of $G_j$.

**Proposition 2.5.9.** *Suppose the assumptions of Lemma 2.5.8 hold. Let $\mathcal{F}$ be the set of tracking functions of $G$. Then for any $g \in \mathcal{F}$ and any $j \in \{0, \dots, \kappa\}$, the following conditions hold deterministically:*

(i) $g(V_j) = \left(1 \pm n^{-1/4}\right)\frac{n-j}{n}g(V)$;

(ii) $G_j$ *is $\gamma/2$-Dirac;*

(iii) $\mathbf{x}|_{G_j}$ *is* $2C$*-normal;*

(iv) $\mathbf{x}|_{G_j}$ *is* $n^{-2/5}$*-almost-perfect.*

**Proof.** It suffices to prove that conditions (i)–(iv) hold for any $j \in \{0, \ldots, \kappa\}$ and any outcome $x(j) = (x_{-(k-2)}, \ldots, x_j)$ of $\mathcal{X}(j)$. Throughout the proof, we let $j \in \{0, \ldots, \kappa\}$ be fixed, and we let $x(j)$ be a fixed outcome of $\mathcal{X}(j)$, thus determining $V_j$ and $G_j$.

(i): Fix $g \in \mathcal{F}$. It is clear that $g(V) - \kappa \leq g(V_j) \leq g(V)$. Relaxing the upper bound and recalling that $g(V) \geq n/2$, we obtain $g(V_j) = (1 \pm 2\kappa/n)g(V)$. Note that $2\kappa/n = 2n^{-1/2} < n^{-1/4}(1 - \kappa/n) - \kappa/n \leq n^{-1/4}(1 - j/n) - j/n$, which implies (i).

(ii): Let $\mathbf{M} \in \binom{V_j}{k-1}$. By (i), we have

$$\begin{aligned} d_{G_j}(\mathbf{M}) &= g_{\mathbf{M}}(V_j) \geq \left(1 - n^{-1/4}\right) \frac{n-j}{n} g_{\mathbf{M}}(V) \geq \left(1 - n^{-1/4}\right) \left(\frac{1}{2} + \gamma\right)(n-j) \\ &\geq \left(\frac{1}{2} + \frac{\gamma}{2}\right)|V_j|. \end{aligned} \tag{2.5.8}$$

Since (2.5.8) holds for all $\mathbf{M} \in \binom{V_j}{k-1}$, we conclude that $G_j$ is $\gamma/2$-Dirac.

The calculations for (iii) and (iv) are straightforward. $\square$

For any $j \in \{0, \ldots, \kappa\}$ and any fixed outcome $x(j) = (x_{-(k-2)}, \ldots, x_j)$ of $\mathcal{X}(j)$, we write $\mathbb{P}_{x(j)}$ for the probability measure in the conditional probability space where we have fixed $\mathcal{X}(j) = x(j)$, so that $\mathbb{P}_{x(j)}[\,\cdot\,] = \mathbb{P}[\,\cdot \mid \mathcal{X}(j) = x(j)]$. We are now ready to prove Lemma 2.5.8.

**Proof of Lemma 2.5.8.** We need to prove that, with probability at least $1 - 1/n$,

$$\mathrm{Error}_g(\mathcal{X}) := \left| \left( \sum_{j=-(k-2)}^{\kappa} g(X_j) \right) - \frac{\kappa}{n} g(V) \right| < n^{3/10} \tag{2.5.9}$$

holds simultaneously for every $g \in \mathcal{F}$. It will suffice to prove that (2.5.9) holds for any fixed $g \in \mathcal{F}$ with probability at least $1 - 1/n^k$, say, since $|\mathcal{F}| \leq n^{k-1}$. Fix $g \in \mathcal{F}$ and set $q \coloneqq (\ln n)^2$. By breaking up $\mathrm{Error}_g(\mathcal{X})$ and repeatedly applying the triangle inequality, we obtain:

$$\begin{aligned}
\mathrm{Error}_g(\mathcal{X}) &= \left| \sum_{j=-(k-2)}^{0} g(X_j) + \sum_{j=1}^{q-1} \left( g(X_j) - \frac{g(V)}{n} \right) + \sum_{j=q}^{\kappa} \left( g(X_j) - \frac{g(V)}{n} \right) \right| \\
&\leq 2q + \left| \sum_{j=0}^{\kappa-q} \left( g(X_{j+q}) - \frac{g(V)}{n} \right) \right| \\
&\leq 2q + \left| \sum_{j=0}^{\kappa-q} (g(X_{j+q}) - \mathbb{E}\left[g(X_{j+q}) \mid \mathcal{X}(j)\right]) \right| \\
&+ \sum_{j=0}^{\kappa-q} \left| \mathbb{E}\left[g(X_{j+q}) \mid \mathcal{X}(j)\right] - \frac{g(V_j)}{n-j} \right| + \sum_{j=0}^{\kappa-q} \left| \frac{g(V_j)}{n-j} - \frac{g(V)}{n} \right|.
\end{aligned}$$

We now prove an upper bound for each of the three sums in the final expression above. To this end, fix $j \in \{0, \ldots, \kappa - q\}$, fix an outcome $x(j) = (x_{-(k-2)}, \ldots, x_j)$ of $\mathcal{X}(j)$, and let $\mathbf{T} \in \binom{V}{k-1}$ be such that $g = g_{\mathbf{T}}$. We apply Proposition 2.5.9 to deduce that $G_j$ is $\gamma/2$-Dirac, and that $\mathbf{x}|_{G_j}$ is $2C$-normal and $n^{-2/5}$-almost-perfect. We can now apply Lemma 2.5.7 to $G_j$ to deduce that

$$\begin{aligned}
\left| \mathbb{E}\left[g(X_{j+q}) \mid \mathcal{X}(j) = x(j)\right] - \frac{g(V_j)}{n-j} \right| &= \left| \mathbb{P}_{x(j)}\left[X_{j+q} \in N_{G_j}(\mathbf{T})\right] - \frac{|N_{G_j}(\mathbf{T})|}{n-j} \right| \\
&< (n-j)^{-13/10} |N_{G_j}(\mathbf{T})| < n^{-5/4} g(V).
\end{aligned}$$

We deduce that $|\mathbb{E}\left[g(X_{j+q}) \mid \mathcal{X}(j)\right] - g(V_j)/(n-j)| < n^{-5/4} g(V)$ for each $j \in \{0, \ldots, \kappa - q\}$. Next, we apply Proposition 2.5.9(i) to obtain that, for each $j \in \{0, \ldots, \kappa - q\}$, we have $|g(V_j)/(n-j) - g(V)/n| \leq n^{-5/4} g(V)$. Finally, applying Lemma 2.3.1 with $\log n$ playing the role of $t$, and using $||g||_\infty = 1$ (there are no isolated $(k-1)$-tuples), we see that with probability at least $1-2q\exp(-(\log n)^2/2)$, we have

$$\left|\sum_{j=0}^{\kappa-q}(g(X_{j+q})-\mathbb{E}\left[g(X_{j+q})\mid \mathcal{X}(j)\right])\right| \leq \log n\sqrt{q(\kappa-q)},$$

so that altogether, with probability at least $1-1/n^k$, we have

$$\mathrm{Error}_g(\mathcal{X}) \leq 2q+\log n\sqrt{q(\kappa-q)}+2(\kappa-q+1)n^{-5/4}g(V) < n^{3/10},$$

completing the proof of the lemma. $\square$

We now have all the tools we need to prove Lemma 2.5.1.

**Proof of Lemma 2.5.1.** Choose a new constant $C$ satisfying $1/n \ll c \ll 1/C \ll \gamma, 1/k$, and let $\mathbf{x}\colon E(G)\to\mathbb{R}^+$ be a $C$-normal perfect fractional matching (such an $\mathbf{x}$ exists by Lemma 2.4.2). Write $\kappa := \sqrt{n}$, and let $\mathcal{X} = (X_{-(k-2)},\dots,X_\kappa)$ be the self-avoiding $\mathbf{x}$-walk on $G$ with starting tuple $\overrightarrow{\mathbf{S}}$. It is clear from the definition of a self-avoiding $\mathbf{x}$-walk that any outcome of $\mathcal{X}$ corresponds to a $\kappa$-path in $G$ with $\overrightarrow{\mathbf{S}}$ as one end (note that the walk does not stop before time $\kappa$, since all codegrees are large enough). We argue now that good outcomes of $\mathcal{X}$ also satisfy condition (iii), where we say an outcome $X = (X_{-(k-2)},\dots,X_\kappa)$ of $\mathcal{X}$ is a *good outcome* if $X$ satisfies (2.5.2). Let $P$ be a tight path in $G$ corresponding to a good outcome $X$ of $\mathcal{X}$, let $\overrightarrow{\mathbf{T}}$ be the non-$\overrightarrow{\mathbf{S}}$ end of $P$, and let $G_\kappa$ denote the residual graph of $G$ at time $\kappa$ of $X$. Thus $G_\kappa = G-(V(P)\setminus\mathbf{T})$ and $|V_\kappa| = n-\kappa$. Let $\mathbf{M}\in\binom{V_\kappa}{k-1}$ and let $g_{\mathbf{M}}\in\mathcal{F}$ be the tracking function of $G$ corresponding to $\mathbf{M}$. Since $X$ is good, we obtain:

$$
\begin{aligned}
d_{G_\kappa}(\mathbf{M}) \quad &= \quad g_{\mathbf{M}}(V_\kappa) = g_{\mathbf{M}}(V) - \sum_{j=-(k-2)}^{\kappa} g_{\mathbf{M}}(X_j) + \sum_{j=\kappa-(k-2)}^{\kappa} g_{\mathbf{M}}(X_j) \\
&\overset{(2.5.2)}{\geq} \quad \frac{n-\kappa}{n} g_{\mathbf{M}}(V) - n^{3/10} \geq \left(\frac{1}{2} + \gamma - n^{-2/3}\right) |V_\kappa|. \qquad (2.5.10)
\end{aligned}
$$

Since (2.5.10) holds for all $\mathbf{M} \in \binom{V_\kappa}{k-1}$, we conclude that $\delta(G_\kappa) \geq (1/2 + \gamma - n^{-2/3})|V_\kappa|$.

Lastly then, it suffices to count the number of good outcomes of $\mathcal{X}$. We begin by finding an upper bound for the probability that $\mathcal{X}$ yields any particular fixed tight path. For any $j \in \{0, \ldots, \kappa\}$, we have by Proposition 2.5.9(ii)–(iii) that $G_j$ is $\gamma/2$-Dirac and $\mathbf{x}|_{G_j}$ is $2C$-normal. It follows that $\mathbf{x}|_{G_j}$ is $8C^2/(n-j)$-upper-balanced. In particular, setting $p \coloneqq 16C^2/n$, we have that all transition probabilities of $\mathcal{X}$ are bounded from above by $p$. Let $Q = (q_{-(k-2)}, \ldots, q_\kappa)$ be a fixed $\kappa$-path in $G$ with $\overrightarrow{\mathbf{S}} = (q_{-(k-2)}, \ldots, q_0)$ as one end. Then

$$
\mathbb{P}\left[\mathcal{X} = Q\right] = \prod_{j=1}^{\kappa} \mathbb{P}\left[X_j = q_j \middle| \bigcap_{i=-(k-2)}^{j-1} \{X_i = q_i\}\right] \leq p^\kappa.
$$

By Lemma 2.5.8, we have that $\mathbb{P}\left[\mathcal{X} \text{ is good}\right] \geq 1/2$, so we conclude that the number of good outcomes of $\mathcal{X}$ (and thus the number of tight $\kappa$-paths in $G$ satisfying (ii) and (iii)) is at least $(1/2)/p^\kappa \geq (cn)^\kappa$, which completes the proof of the lemma.

□

## 2.6 Counting and absorbing long paths

In this section we show how to iterate Lemma 2.5.1 to construct tight paths in $G$ which use almost all of the vertices of $G$, and we count the number of choices that can be made in this process to obtain a lower bound for the number of these long paths. Finally, we prove Theorem 2.1.1 by showing how these paths can be completed into tight Hamilton cycles of $G$.

**Lemma 2.6.1.** *Let $\gamma > 0$, let $k \geq 2$, and let $G$ be an $(n,k,\gamma)$-graph. There are at least* $\exp(n \ln n - \Theta(n))$ *tight paths in $G$ of length at least $n - n^{7/8}$.*

**Proof.** We describe an algorithm on $G$. Let $\overrightarrow{\mathbf{S}}_0 \in (V)_{k-1}$ be arbitrary, set $G_0 \coloneqq G$, set $n_0 \coloneqq n$, and set $\gamma_0 \coloneqq \gamma$. For each $i \geq 0$, set $n_{i+1} \coloneqq n_i - \sqrt{n_i}$, and set $\gamma_{i+1} \coloneqq \gamma_i - (n_i)^{-2/3}$. Set $L$ to be the smallest index such that $n_L < n^{7/8}$. Suppose we have already performed $i$ steps of the algorithm, and obtained a $k$-graph $G_i$ on $n_i$ vertices satisfying $\delta(G_i) \geq (1/2 + \gamma_i)n_i$, and we have obtained $\overrightarrow{\mathbf{S}}_i \in (V(G_i))_{k-1}$. If $\gamma_i < \gamma/2$ or $i = L$, then we terminate the algorithm. Otherwise, we apply Lemma 2.5.1 to $G_i$ to obtain a set $\mathcal{P}_{i+1}$ of $\sqrt{n_i}$-paths, each with chosen starting tuple $\overrightarrow{\mathbf{S}}_i$. Choose $P_{i+1} \in \mathcal{P}_{i+1}$ arbitrarily, let $\overrightarrow{\mathbf{T}}_{i+1}$ be the non-$\overrightarrow{\mathbf{S}}_i$ end of $P_{i+1}$, and put $\overrightarrow{\mathbf{S}}_{i+1} \coloneqq \overleftarrow{\mathbf{T}}_{i+1}$. Set $G_{i+1} \coloneqq G_i - (V(P_{i+1}) \setminus \mathbf{S}_{i+1})$. Observe that by Lemma 2.5.1(iii), we have $\delta(G_{i+1}) \geq (1/2 + \gamma_{i+1})n_{i+1}$.

Let $r_i \coloneqq \sum_{j=0}^{i-1} (n_j)^{-2/3}$. Note that, provided $\gamma_{i-1} \geq \gamma/2$, we have $\gamma_i = \gamma - r_i$. We claim that the algorithm does not terminate in the first $L$ steps. To see this, note that it suffices to show that $r_L = o(1)$. Write $\kappa_i \coloneqq \sqrt{n_i}$ and observe that $n_{L-1} = n - \sum_{j=0}^{L-2} \kappa_j \leq n - (L-1)\kappa_{L-1}$. Re-arranging, we obtain that $L \leq 2n/\kappa_{L-1}$.

Using $n_{L-1} \geq n^{7/8}$, we obtain that

$$r_L \leq \frac{L}{(n_{L-1})^{2/3}} \leq \frac{2n}{(n_{L-1})^{7/6}} \leq 2n^{-1/48} = o(1),$$

so the algorithm does not terminate in the first $L$ steps, as claimed. When the algorithm terminates, we have obtained tight paths $P_1, \ldots, P_L$. By construction, we may concatenate these paths, in order, to obtain a path $Q \coloneqq \bigcup_{i \leq L} P_i$ of length $n - n_L \geq n - n^{7/8}$. Let $N$ be the number of tight paths of length $n - n_L$ in $G$. By Lemma 2.5.1(i), there is a positive constant $c < 1$ such that the number of choices for $P_{i+1}$ is at least $(cn_i)^{\kappa_i}$, for each $i \in \{0, \ldots, L-1\}$. Thus, we obtain

$$N \geq \prod_{i=0}^{L-1} (cn_i)^{\kappa_i} \geq c^n \prod_{i=0}^{L-1} \frac{n_i!}{n_{i+1}!} = c^n \frac{n!}{n_L!} \geq c^n \frac{n!}{(n^{7/8})!} = \exp(n \ln n - \Theta(n)).$$

$\square$

We are now ready to prove Theorem 2.1.1.

**Proof of Theorem 2.1.1.** The upper bound holds trivially. To prove the lower bound, we choose a set $W \subseteq V$ of size $n^{9/10}$ uniformly at random. A simple application of Lemma 2.3.2 shows that there is a choice of $W$ such that $|N(\mathbf{S}) \cap W| \geq (1/2 + 3\gamma/4)|W|$ for all $\mathbf{S} \in \binom{V}{k-1}$. Fix such a choice of $W$, set $G' \coloneqq G - W$, and put $n' \coloneqq n - n^{9/10}$. Then $G'$ is $\gamma/2$-Dirac, and we apply Lemma 2.6.1 to $G'$ (with $\gamma/2$ playing the role of $\gamma$) to find a set $\mathcal{P}$ of tight paths of length at least $n' - (n')^{7/8}$ in $G'$, such that $|\mathcal{P}| \geq \exp(n' \ln n' - \Theta(n')) = \exp(n \ln n - \Theta(n))$. Fix $P \in \mathcal{P}$, let $\overrightarrow{\mathbf{S}}_P$ and $\overrightarrow{\mathbf{T}}_P$ be the ends of $P$, and let $U_P \coloneqq V(G') \setminus V(P)$, so that $|U_P| \leq (n')^{7/8} \leq n^{7/8}$. Notice that $G[W \cup U_P \cup \mathbf{S}_P \cup \mathbf{T}_P]$ is $\gamma/2$-Dirac. Thus, by Lemma 2.3.9, there is a tight Hamilton path $Q_P$ of $G[W \cup U_P \cup \mathbf{S}_P \cup \mathbf{T}_P]$ with ends $\overleftarrow{\mathbf{S}}_P$ and $\overleftarrow{\mathbf{T}}_P$. Then $C_P \coloneqq P \cup Q_P$ is a tight Hamilton cycle of $G$.

Define $\mathcal{C} \coloneqq \{C_P \colon P \in \mathcal{P}\}$ and note that for each $C \in \mathcal{C}$, the number of $P \in \mathcal{P}$ with $C = C_P$ is at most $n^2$, since $P$ must be a subpath of $C$. We conclude that $\mathcal{C}$ is a set of at least $n^{-2} \exp(n \ln n - \Theta(n)) = \exp(n \ln n - \Theta(n))$ tight Hamilton cycles in $G$. □

Finally, we prove Corollary 2.1.2.

**Proof of Corollary 2.1.2.** Let $\gamma > 0$ be fixed, let $k \geq 2$, $\ell \in \{0, \dots, k-1\}$, $(k-\ell) \mid n$, and let $G$ be a $k$-graph on $n$ vertices satisfying $\delta(G) \geq (1/2 + \gamma)n$. Firstly, suppose $\ell = 0$, and recall that the number of Hamilton 0-cycles of $G$ is precisely the number of perfect matchings of $G$. By considering the number of perfect matchings in the complete $k$-graph on $n$ vertices, it is easy to see that the upper bound of (i) holds. We now use Theorem 2.1.1 to show that the lower bound holds. Let $\mathcal{M}$ be the set of perfect matchings of $G$, and let $\mathcal{C}$ be the set of tight Hamilton cycles of $G$. Notice that, for any $M \in \mathcal{M}$, there are at most $(n/k)!(k!)^{n/k}$ choices of $C \in \mathcal{C}$ such that $M \subseteq E(C)$, because we may construct all vertex orderings corresponding to possible such $C$ by reordering the edges of $M$ and the vertices within them. By applying Theorem 2.1.1, we conclude that

$$|\mathcal{M}| \geq \frac{|\mathcal{C}|}{\left(\frac{n}{k}\right)!(k!)^{n/k}} = \exp\left(\left(1 - \frac{1}{k}\right) n \ln n - \Theta(n)\right).$$

For the case $\ell \in [k-1]$, firstly notice that $\frac{k-\ell}{2}(n-1)! = \exp(n \ln n - \Theta(n))$ is a trivial upper bound for the number of Hamilton $\ell$-cycles of $G$. Finally, it suffices to apply Theorem 2.1.1 to $G$ and observe that every tight Hamilton cycle of $G$ contains $k - \ell$ Hamilton $\ell$-cycles (since $(k-\ell) \mid n$), and each Hamilton $\ell$-cycle of $G$

is contained in at most $(k!)^{n/(k-\ell)}$ tight Hamilton cycles. $\square$

## 2.7 Concluding remarks

Though Theorem 2.1.1 holds in $\gamma$-Dirac $k$-graphs with equality, we believe that the error bound can be made more precise. More specifically, we believe the following hypergraph version of [32, Theorem 1.1] holds, giving a more accurate lower bound for the number of tight Hamilton cycles in such hypergraphs.

**Conjecture 2.7.1.** *For a fixed integer $k \geq 2$ and a fixed constant $\gamma > 0$, the number of tight Hamilton cycles of a $k$-graph $G$ on $n$ vertices with $\delta(G) \geq (1/2+\gamma)n$ is at least $(1/2 - o(1))^n n!$.*

It would of course be desirable to obtain a formula for the number of tight Hamilton cycles in $\gamma$-Dirac $k$-graphs $G$ which takes properties of $G$ like the degrees and codegrees into account. We recall that such a formula has already been obtained [32, Theorem 1.3, Theorem 1.5] in terms of the 'entropy of $G$' in the $k = 2$ case, and it would be interesting to see if this (or a similar) notion can be extended to $k \geq 3$.

Finally, we note that the results of [32] show that graphs with minimum degree precisely at the threshold for Hamiltonicity in fact have many Hamilton cycles. The exact minimum codegree threshold for existence of a tight Hamilton cycle in $k$-graphs on $n$ vertices is not yet known for $k \geq 4$, but is known to be $\lfloor n/2 \rfloor$ in the case $k = 3$ [112, Theorem 1.2], and it is of course a natural question to ask if the conclusions of Theorem 2.1.1 or indeed Conjecture 2.7.1 hold in this exact setting.

# CHAPTER 3

# ALMOST ALL OPTIMALLY COLOURED COMPLETE GRAPHS CONTAIN A RAINBOW HAMILTON PATH

**Abstract**

A subgraph $H$ of an edge-coloured graph is called rainbow if all of the edges of $H$ have different colours. In 1989, Andersen conjectured that every proper edge-colouring of $K_n$ admits a rainbow path of length $n-2$. We show that almost all optimal edge-colourings of $K_n$ admit both (i) a rainbow Hamilton path and (ii) a rainbow cycle using all of the colours. This result demonstrates that Andersen's Conjecture holds for almost all optimal edge-colourings of $K_n$ and answers a recent question of Ferber, Jain, and Sudakov. Our result also has applications to the existence of transversals in random symmetric Latin squares.

## 3.1 Introduction

### 3.1.1 Extremal results on rainbow colourings

We say that a subgraph $H$ of an edge-coloured graph is *rainbow* if all of the edges of $H$ have different colours. An *optimal edge-colouring* of a graph is a proper edge-colouring using the minimum possible number of colours. In this paper we study the problem of finding a rainbow Hamilton path in large optimally edge-coloured complete graphs.

The study of finding rainbow structures within edge-coloured graphs has a rich history. For example, the problem posed by Euler on finding orthogonal $n \times n$ Latin squares can easily be seen to be equivalent to that of finding an optimal edge-colouring of the complete bipartite graph $K_{n,n}$ which decomposes into edge-disjoint rainbow perfect matchings. It transpires that there are optimal colourings of $K_{n,n}$ without even a single rainbow perfect matching, if $n$ is even. However, an important conjecture, often referred to as the Ryser-Brualdi-Stein Conjecture, posits that one can always find an almost-perfect rainbow matching, as follows.

**Conjecture 3.1.1** (Ryser [114], Brualdi-Stein [20, 119]). *Every optimal edge-colouring of $K_{n,n}$ admits a rainbow matching of size $n-1$ and, if $n$ is odd, a rainbow perfect matching.*

Currently, the strongest result towards this conjecture for arbitrary optimal edge-colourings is due to Keevash, Pokrovskiy, Sudakov, and Yepremyan [79], who showed that there is always a rainbow matching of size $n - O(\log n/\log\log n)$. This result improved earlier bounds of Woolbright [125], Brouwer, de Vries, and Wieringa [18], and Hatami and Shor [66].

It is natural to search for spanning rainbow structures in the non-partite setting as well; that is, what spanning rainbow substructures can be found in properly edge-coloured complete graphs $K_n$? It is clear that one can always find a rainbow spanning tree – indeed, simply take the star rooted at any vertex. Kaneko, Kano, and Suzuki [71] conjectured that for $n > 4$, in any proper edge-colouring of $K_n$, one can find $\lfloor n/2 \rfloor$ edge-disjoint rainbow spanning trees, thus decomposing $K_n$ if $n$ is even, and almost decomposing $K_n$ if $n$ is odd. This conjecture was recently proved approximately by Montgomery, Pokrovskiy, and Sudakov [102], who showed that in any properly edge-coloured $K_n$, one can find $(1 - o(1))n/2$ edge-disjoint rainbow spanning trees.

For *optimal* edge-colourings, even more is known. Note firstly that if $n$ is even and $K_n$ is optimally edge-coloured, then the colour classes form a 1-*factorization* of $K_n$; that is, a decomposition of $K_n$ into perfect matchings. Throughout the paper, we will use the term 1-factorization synonymously with an edge-colouring whose colour classes form a 1-factorization. It is clear that if a 1-factorization of $K_n$ exists, then $n$ is even. Very recently, Glock, Kühn, Montgomery, and Osthus [54] showed that for sufficiently large even $n$, there exists a tree $T$ on $n$ vertices such that any 1-factorization of $K_n$ decomposes into edge-disjoint rainbow spanning trees isomorphic to $T$, thus resolving conjectures of Brualdi and Hollingsworth [19], and Constantine [27, 26]. See e.g. [106, 102, 80] for previous work on these conjectures.

The tree $T$ used in [54] is a path of length $n - o(n)$, together with $o(n)$ short paths attached to it. Thus it might seem natural to ask if one can find a rainbow Hamilton path in any 1-factorization of $K_n$. Note that such a path would contain all of the colours used in the 1-factorization, so it is not possible to find a rainbow Hamilton cycle in a 1-factorization of $K_n$. However, in 1984 Maamoun and Meyniel [96] proved the existence of a 1-factorization of $K_n$ (for $n \geq 4$ being any power of 2) without a rainbow Hamilton path. Sharing parallels with Conjecture 3.1.1 for the non-partite setting, Andersen [9] conjectured in 1989 that all proper edge-colourings of $K_n$ admit a rainbow path which omits only one vertex.

**Conjecture 3.1.2** (Andersen [9])**.** *All proper edge-colourings of $K_n$ admit a rainbow path of length $n-2$.*

Several variations of Andersen's Conjecture have been proposed. In 2007, Akbari, Etesami, Mahini, and Mahmoody [4] conjectured that all 1-factorizations of $K_n$ admit a Hamilton cycle whose edges collectively have at least $n-2$ colours. They also conjectured that all 1-factorizations of $K_n$ admit a rainbow cycle omitting only two vertices.

Although now known to be false, the following stronger form of Conjecture 3.1.2 involving the 'sub-Ramsey number' of the Hamilton path was proposed by Hahn [61]. Every (not necessarily proper) edge-colouring of $K_n$ with at most $n/2$ edges of each colour admits a rainbow Hamilton path. In light of the aforementioned construction of Maamoun and Meyniel [96], in 1986 Hahn and Thomassen [62] suggested the following slightly weaker form of Hahn's Conjecture, that all edge-colourings of $K_n$ with strictly fewer than $n/2$ edges of each colour admit a rainbow Hamilton path. However, even this weakening of Hahn's Conjecture is false – Pokrovskiy and Sudakov [105] proved the existence of such edge-colourings of $K_n$ in which the longest rainbow Hamilton path has length at most $n - \ln n/42$.

Andersen's Conjecture has led to a number of results, generally focussing on increasing the length of the rainbow path or cycle that one can find in an arbitrary 1-factorization or proper edge-colouring of $K_n$ (see e.g. [22, 58, 48, 59]). Alon, Pokrovskiy, and Sudakov [7] proved that all proper edge-colourings of $K_n$ admit a rainbow path with length $n - O(n^{3/4})$, and the error bound has since been improved to $O(\sqrt{n} \cdot \log n)$ by Balogh and Molla [10]. Further support for Conjecture 3.1.2 and its variants was provided by Montgomery, Pokrovskiy, and Sudakov [102] as well as Kim, Kühn, Kupavskii, and Osthus [80], who showed that if we consider proper edge-colourings where no colour class is larger than $n/2 - o(n)$, then we can even find $n/2 - o(n)$ edge-disjoint rainbow Hamilton cycles.

### 3.1.2 Random colourings

It is natural to consider these problems in a probabilistic setting, that is to consider random edge-colourings as well as random Latin squares. However, the 'rigidity' of the underlying structure makes these probability spaces very challenging to analyse. Recently significant progress was made by Kwan [90], who showed that almost all Latin squares contain a transversal, or equivalently, that almost all optimal edge-colourings of $K_{n,n}$ admit a rainbow perfect matching. His analysis was carried out in a hypergraph setting, which also yields the result that almost all Steiner triple systems contain a perfect matching. Recently, this latter result was strengthened by Ferber and Kwan [44], who showed that almost all Steiner triple systems have an approximate decomposition into edge-disjoint perfect matchings. Here we show that Hahn's original conjecture (and thus Andersen's Conjecture as well) holds for almost all 1-factorizations, answering a recent question of Ferber, Jain, and Sudakov [43]. In what follows, we say a property holds 'with high probability' if it holds with a probability that tends to 1 as the number of vertices $n$ tends to infinity.

**Theorem 3.1.3.** *Let $\phi$ be a uniformly random optimal edge-colouring of $K_n$. Then with high probability,*

*(i) $\phi$ admits a rainbow Hamilton path, and*

*(ii) $\phi$ admits a rainbow cycle $F$ containing all of the colours.*

*In particular, if $n$ is odd, then $F$ is a rainbow Hamilton cycle.*

As discussed in Section 3.8, there is a well-known correspondence between rainbow 2-factors in $n$-edge-colourings of $K_n$ and transversals in symmetric Latin squares, as a transversal in a Latin square corresponds to a permutation $\sigma$ of $[n]$ such that the entries in positions $(i, \sigma(i))$ are distinct for all $i \in [n]$. Based on this, we use Theorem 3.1.3(ii) to show that random symmetric Latin squares of odd order contain a *Hamilton transversal* with high probability. Here we say a transversal is *Hamilton* if the underlying permutation $\sigma$ is an $n$-cycle.

**Corollary 3.1.4.** *Let $n$ be an odd integer and $\mathbf{L}$ a uniformly random symmetric $n \times n$ Latin square. Then with high probability $\mathbf{L}$ contains a Hamilton transversal.*

Further results on random Latin squares were recently obtained by Kwan and Sudakov [93], who gave estimates on the number of intercalates in a random Latin square as well as their likely discrepancy. After the completion of the initial version of this paper, additional results on intercalates in random Latin squares were obtained by Kwan, Sah, and Sawhney [91], which, together with the results of [93], resolve an old conjecture of McKay and Wanless [98]. In addition, Gould and Kelly [55] showed that an analogue of Corollary 3.1.4 also holds when $\mathbf{L}$ is a uniformly random (not necessarily symmetric) $n \times n$ Latin square, strengthening the aforementioned result of Kwan [90].

## 3.2 Notation

In this section, we collect some definitions and notation that we will use throughout the paper.

For a graph $G$ and (not necessarily distinct) vertex sets $A, B \subseteq V(G)$, we define $E_G(A, B) \coloneqq \{e = ab \in E(G) \colon a \in A, b \in B\}$. We often simply write $E(A, B)$ when $G$ is clear from the context. We define $e(A, B) \coloneqq |E(A, B)|$. For a vertex $v \in V(G)$, we define $\partial_G(v)$ to be the set of edges of $G$ which are incident to $v$. For a proper colouring $\phi \colon E(G) \to \mathbb{N}$ and a colour $c \in \mathbb{N}$, we define $E_c(G) \coloneqq \{e \in E(G) \colon \phi(e) = c\}$ and say that an edge $e \in E_c(G)$ is a *$c$-edge* of $G$. For a vertex $v \in V(G)$, if $e$ is a $c$-edge in $G$ incident to $v$, then we say that the non-$v$ endpoint of $e$ is the *$c$-neighbour* of $v$. For a vertex $v \in V(G)$ and three colours $c_1, c_2, c_3 \in \mathbb{N}$, we say that the $c_3$-neighbour of the $c_2$-neighbour of the $c_1$-neighbour of $v$ is the *end* of the *$c_1c_2c_3$-walk starting at* $v$, if all such edges exist. For a set of colours $D \subseteq \mathbb{N}$, we define $N_D(v) \coloneqq \{w \in N_G(v) \colon \phi(vw) \in D\}$. For sets $A, B \subseteq V(G)$ and a colour $c \in \mathbb{N}$, we define $E_c(A, B) \coloneqq \{e \in E(A, B) \colon \phi(e) = c\}$. If $G$ is not clear from the context, we sometimes also write $E^c_G(A, B)$. For any subgraph $H \subseteq G$, we define $\phi(H) \coloneqq \{\phi(e) \colon e \in E(H)\}$. For a set of colours $D \subseteq [n-1]$, let $\mathcal{G}^{\mathrm{col}}_D$ be the set of pairs $(G, \phi_G)$, where $G$ is a $|D|$-regular graph on a vertex set $V$ of size $n$, and $\phi_G$ is a 1-factorization of $G$ with colour set $D$. Often, we abuse notation and write $G \in \mathcal{G}^{\mathrm{col}}_D$, and in this case we let $\phi_G$ denote the implicit 1-factorization of $G$, sometimes simply writing $\phi$ when $G$ is clear from the context. For $G \in \mathcal{G}^{\mathrm{col}}_{[n-1]}$ and a set of colours $D \subseteq [n-1]$, we define the *restriction of $G$ to $D$*, denoted $G|_D$, to be the spanning subgraph of $G$ containing precisely those edges of $G$ which have colour in $D$. Observe that $G|_D \in \mathcal{G}^{\mathrm{col}}_D$. A subgraph $H \subseteq G \in \mathcal{G}^{\mathrm{col}}_D$ inherits the colours of its edges from $G$. Observe that uniformly randomly choosing a 1-factorization $\phi$ of $K_n$ on vertex set $V$ and colour set $[n-1]$ is equivalent to uniformly randomly choosing $G \in \mathcal{G}^{\mathrm{col}}_{[n-1]}$. For any $D \subseteq [n-1]$, $G \in \mathcal{G}^{\mathrm{col}}_{D}$, and sets $V' \subseteq V$, $D' \subseteq D$, we define $E_{V',D'}(G) \coloneqq \{e = xy \in E(G) \colon \phi_G(e) \in D', x, y \in V'\}$, and we define $e_{V',D'}(G) \coloneqq |E_{V',D'}(G)|$. For a hypergraph $\mathcal{H}$, we write $\Delta^c(\mathcal{H})$ to denote the maximum codegree of $\mathcal{H}$; that is, the maximum number of edges containing any two fixed vertices of $\mathcal{H}$.

For a set $D$ of size $n$ and a partition $\mathcal{P}$ of $D$ into $m$ parts, we say that $\mathcal{P}$ is *equitable* to mean that all parts $P$ of $\mathcal{P}$ satisfy $|P| \in \{\lfloor n/m \rfloor, \lceil n/m \rceil\}$, and when it does not affect the argument, we will assume all parts of an equitable partition have size precisely $n/m$. For a set $S$ and a real number $p \in [0,1]$, a *$p$-random* subset $T \subseteq S$ is a random subset in which each element of $S$ is included in $T$ independently with probability $p$. A *$\beta$-random* subgraph of a graph $G$ is a spanning subgraph of $G$ where the edge-set is a $\beta$-random subset of $E(G)$. For an event $\mathcal{E}$ in any probability space, we write $\overline{\mathcal{E}}$ to denote the complement of $\mathcal{E}$. For real numbers $a, b, c$ such that $b > 0$, we write $a = (1 \pm b)c$ to mean that the inequality $(1-b)c \leq a \leq (1+b)c$ holds. For a natural number $n \in \mathbb{N}$, we define $[n] \coloneqq \{1, 2, \ldots, n\}$, and $[n]_0 \coloneqq [n] \cup \{0\}$. We write $x \ll y$ to mean that for any $y \in (0,1]$ there exists an $x_0 \in (0,1)$ such that for all $0 < x \leq x_0$ the subsequent statement holds. Hierarchies with more constants are defined similarly and should be read from the right to the left. Constants in hierarchies will always be real numbers in $(0,1]$. We assume large numbers to be integers if this does not affect the argument.

## 3.3 Overview of the proof

In this section, we provide an overview of the proof of Theorem 3.1.3. In Section 3.4 we prove Theorem 3.1.3 in the case when $n$ is even assuming two key lemmas which we prove in later sections. In particular, we assume that $n$ is even in Sections 3.3–3.7, so that the optimal edge-colouring $\phi$ we work with is a 1-factorization of $K_n$. In Section 3.8 we derive Theorem 3.1.3 in the case when $n$ is odd from the case when $n$ is even. We will also deduce Corollary 3.1.4 from Theorem 3.1.3(ii) in Section 3.8. Throughout the proof we work with constants $\varepsilon, \gamma, \eta$, and $\mu$ satisfying the following hierarchy:

$$1/n \ll \varepsilon \ll \gamma \ll \eta \ll \mu \ll 1. \tag{3.3.1}$$

Our proof uses the absorption method as well as switching techniques. Note that the latter is a significant difference to [90, 44], which rely on the analysis of the random greedy triangle removal process, as well as modifications of arguments in [95, 75] which bound the number of Steiner triple systems. Our main objective is to show that with high probability, in a random 1-factorization, we can find an absorbing structure inside a random subset of $\Theta(\mu n)$ reserved vertices, using a random subset of $\Theta(\mu n)$ reserved colours. A recent result [54, Lemma 16], based on hypergraph matchings, enables us to find a long rainbow path avoiding these reserved vertices and colours, and using our absorbing structure, we extend this path to a rainbow Hamilton path. More precisely, we randomly 'reserve' $\Theta(\mu n)$ vertices and colours and show that with high probability we can find an absorbing structure. This absorbing structure consists of a subgraph $G_{\mathrm{abs}}$ containing only reserved vertices and colours and all but at most $\gamma n$ of them. Moreover $G_{\text{abs}}$ contains 'flexible' sets of vertices and colours $V_{\text{flex}}$ and $C_{\text{flex}}$ each of size $\eta n$, with the following crucial property:

(†) for any pair of equal-sized subsets $X \subseteq V_{\text{flex}}$ and $Y \subseteq C_{\text{flex}}$ of size at most $\eta n/2$, the graph $G_{\text{abs}} - X$ contains a spanning rainbow path whose colours avoid $Y$.

In fact, this spanning rainbow path has the same end vertices, regardless of the choice of $X$ and $Y$. Given this absorbing structure, we find a rainbow Hamilton path in the following three steps:

1. **Long path step:** Apply [54, Lemma 16] to obtain a long rainbow path $P_1$ containing only non-reserved vertices and colours. Moreover, $P_1$ contains all but at most $\gamma n$ of them.
2. **Covering step:** 'Cover' the vertices and colours not in $G_{\text{abs}}$ or $P_1$ using the flexible sets, by greedily constructing a path $P_2$ containing them as well as sets $X \subseteq V_{\text{flex}}$ and $Y \subseteq C_{\text{flex}}$ of size at most $\eta n/2$.
3. **Absorbing step:** 'Absorb' the remaining vertices and colours, by letting $P_3$ be the rainbow path guaranteed by (†).

In the covering step, we can ensure that $P_2$ shares one end with $P_1$ and one end with $P_3$ so that $P_1 \cup P_2 \cup P_3$ is a rainbow Hamilton path, as desired. These steps are fairly straightforward, so the majority of the paper is devoted to building the absorbing structure, that is, the subgraph $G_{\text{abs}}$ which satisfies (†) with respect to 'flexible' sets $V_{\text{flex}}$ and $C_{\text{flex}}$. This argument is split into two parts. Lemma 3.4.9, proved in Section 3.6, asserts that, subject to some quasirandomness conditions, we can build our absorbing structure using our randomly reserved vertices and colours; Lemma 3.4.8, proved in Section 3.7, asserts that a typical 1-factorization of $K_n$ has

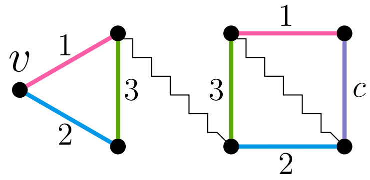



Figure 3.1: A $(v, c)$-absorber, where $\phi(e_i) = i$. The paths $P_1$ and $P_2$ are drawn as zigzags.

these quasirandom properties.

### 3.3.1 Absorption

To design our absorbing structure, we employ a strategy sometimes called 'distributed absorption', first introduced by Montgomery [101]. The details of this are presented in Section 3.4, but we provide an overview now. Our absorbing structure consists of many 'gadgets' pieced together in a particular way. In particular, for a vertex $v$ and colour $c$, a $(v, c)$-*absorber* (see Definition 3.4.1 and Figure 3.1) is a small subgraph containing both $v$ and an edge coloured $c$, with the following property: It contains a rainbow path which is spanning and which uses one of each colour assigned to its edges, and it also contains a rainbow path which includes all of its vertices except $v$ and an edge of every one if its colours except $c$; moreover, these paths have the same end vertices. We refer to the former path as the $(v, c)$-*absorbing path* and the latter as the $(v, c)$-*avoiding path* (again, see Definition 3.4.1).

We build our absorbing structure out of $(v, c)$-absorbers, along with short rainbow paths linking them together, using an auxiliary bipartite graph $H$ as a template (see Definition 3.4.2 and Figure 3.2), where one part of $H$ is a set of vertices (including $V_{\text{flex}}$) and the other part is a set of colours (including $C_{\text{flex}}$). For

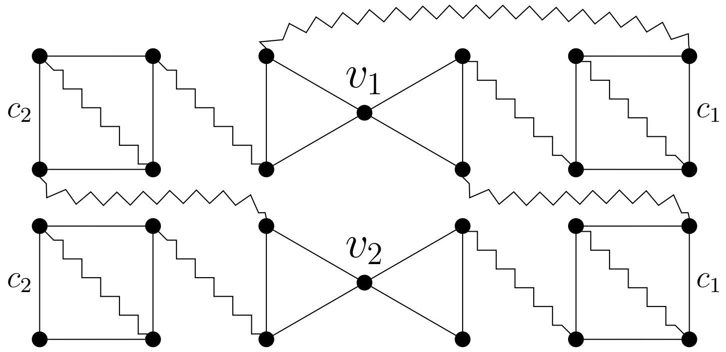



Figure 3.2: An $H$-absorber where $H \cong K_{2,2}$ with bipartition $(\{v_1, v_2\}, \{c_1, c_2\})$.

every edge $vc \in E(H)$, we will have a $(v, c)$-absorber in the absorbing structure. When proving (†), if $v$ or $c$ is in $X$ or $Y$, then the spanning rainbow path in $G_{\text{abs}} - X$ contains the $(v, c)$-avoiding path, and otherwise it may contain the $(v, c)$-absorbing path. More precisely, we find a perfect matching of $H - (X \cup Y)$, and we use the $(v, c)$-absorbing path for every matched pair of vertex $v$ and colour $c$.

A naive approach would be to use the complete bipartite graph with parts $V_{\text{flex}}$ and $C_{\text{flex}}$ as our template $H$; however, this would require too many absorbing gadgets. Instead, we choose a much sparser template graph $H$ that is *robustly matchable* with respect to $V_{\text{flex}}$ and $C_{\text{flex}}$ (see Definition 3.4.3); we use a result of Montgomery [101, Lemma 10.7] to construct a robustly matchable bipartite graph with maximum degree $O(1)$. Thus, we only need $\Theta(\eta n)$ absorbing gadets to build an absorbing structure satisfying (†). Using that $\eta \ll \mu$, we can build such an absorbing structure inside the random subset of $\Theta(\mu n)$ vertices and $\Theta(\mu n)$ colours (see Lemma 3.6.4). However, our absorbing structure needs to contain all but at most $\gamma n$ of the reserved vertices and colours. To that end, we attach a long rainbow path using almost all of the remaining reserved vertices and colours that we call a *tail* (see Definition 3.4.2); this is accomplished using the semi-random method, implemented via hypergraph matchings results (see Lemma 3.6.5). We use a similar approach in the long path step.

### 3.3.2 Analysing a random 1-factorization of $K_n$

To build the absorbing structure described in Section 3.3.1, we need to show that a typical 1-factorization of $K_n$ satisfies some quasirandom properties. We call these properties *local edge-resilience* and *robust gadget-resilience* (see Definitions 3.4.6 and 3.4.7), and we prove they hold for typical 1-factorizations in Lemma 3.4.8. Standard arguments can be used to show that these properties hold with high probability for a (not necessarily proper) edge-colouring of $K_n$ where each edge is assigned one of $n$ colours independently and uniformly at random; however, it is much more challenging to prove this for a random 1-factorization. We prove Lemma 3.4.8 using a 'coloured version' of switching arguments that are commonly used to study random regular graphs. Unfortunately, 1-factorizations of the complete graph $K_n$ are 'rigid' structures, in the sense that it is difficult to make local changes without global ramifications on such a 1-factorization. Thus, instead of analysing switchings between graphs in $\mathcal{G}^{\mathrm{col}}_{[n-1]}$, we will analyse switchings between graphs in $\mathcal{G}^{\mathrm{col}}_{D}$ for appropriately chosen $D \subsetneq [n-1]$. In the setting of random Latin squares, this approach was used by McKay and Wanless [98] and further developed by Kwan and Sudakov [93], and we build on their ideas.

We use results on the number of 1-factorizations of dense regular graphs due to Kahn and Lovász (see Theorem 3.7.10) and Ferber, Jain, and Sudakov (see Theorem 3.7.11) to study the number of completions of a graph $H \in \mathcal{G}^{\mathrm{col}}_{D}$ to a graph $G \in \mathcal{G}^{\mathrm{col}}_{[n-1]}$, and we use this information to compare the probability space corresponding to a uniform random choice of $\mathbf{H} \in \mathcal{G}_D^{\mathrm{col}}$, with the probability space corresponding to a uniform random choice of $\mathbf{G} \in \mathcal{G}_{[n-1]}^{\mathrm{col}}$. In particular, if a uniformly random $\mathbf{H} \in \mathcal{G}_D^{\mathrm{col}}$ is extended uniformly at random to obtain a colouring $\mathbf{H}' \in \mathcal{G}_{[n-1]}^{\mathrm{col}}$, then $\mathbf{H}'$ is not chosen uniformly at random from $\mathcal{G}_{[n-1]}^{\mathrm{col}}$, since different choices of $H \in \mathcal{G}_D^{\mathrm{col}}$ have different numbers of extensions; however, $\mathbf{H}'$ can be compared to a uniformly random $\mathbf{G} \in \mathcal{G}_{[n-1]}^{\mathrm{col}}$ as follows (see also Corollary 3.7.12). For an absolute constant $C$, and for each $K \in \mathcal{G}_{[n-1]}^{\mathrm{col}}$,

$$\mathbb{P}\left[\mathbf{G} = K\right] = \mathbb{P}\left[\mathbf{H}' = K\right] \cdot \exp(\pm n^{2-1/C}).$$

Therefore, any property that holds for $\mathbf{H}$ with probability at least $1 - \exp(-\Omega(n^2))$ also holds with high probability for $\mathbf{G}|_D$. Our switching arguments yield local edge-resilience and robust gadget-resilience for $\mathbf{H}$ with high enough probability (see Lemmas 3.7.1 and 3.7.8) to apply Corollary 3.7.12.

## 3.4 Proving Theorem 3.1.3

In this section, let $\phi$ be a 1-factorization of $K_n$ with vertex set $V$ and colour set $C = [n-1]$. We first present the details of our absorbing structure, and in Section 3.4.1, we prove Theorem 3.1.3 (in the case when $n$ is even) subject to its existence. We begin by introducing our absorbing gadgets in the following definition (see also Figure 3.1).

**Definition 3.4.1.** For every $v \in V$ and $c \in C$, a $(v, c)$-*absorbing gadget* is a subgraph of $K_n$ of the form $A = T \cup Q$ such that the following holds:

- $T \cong K_3$ and $Q \cong C_4$,

- $T$ and $Q$ are vertex-disjoint,
- $v \in V(T)$ and there is a unique edge $e \in E(Q)$ such that $\phi(e) = c$,
- if $e_1, e_2 \in E(T)$ are the edges incident to $v$, then there is matching $\{e'_1, e'_2\}$ in $Q$ not containing $e$ such that $\phi(e_i) = \phi(e'_i)$ for $i \in \{1, 2\}$,
- if $e_3 \in E(T)$ is the edge not incident to $v$, then there is an edge $e'_3 \in E(Q)$ such that $\{e'_3, e\}$ is a matching in $Q$ and $\phi(e_3) = \phi(e'_3) \neq c$.

In this case, a pair $P_1$, $P_2$ of paths *completes* the $(v, c)$-absorbing gadget $A = T \cup Q$ if

- the ends of $P_1$ are non-adjacent vertices in $Q$,
- one end of $P_2$ is in $Q$ but not incident to $e$ and the other end of $P_2$ is in $V(T) \setminus \{v\}$,
- $P_1$ and $P_2$ are vertex-disjoint and both $P_1$ and $P_2$ are internally vertex-disjoint from $A$,
- $P_1 \cup P_2$ is rainbow, and
- $\phi(P_1 \cup P_2) \cap \phi(A) = \varnothing$,

and we say $A' \coloneqq A \cup P_1 \cup P_2$ is a $(v, c)$-*absorber*. We also define the following.

- The path $P$ with edge-set $E(P_1) \cup E(P_2) \cup \{e'_1, e'_2, e_3\}$ is the $(v, c)$-*avoiding* path in $A'$, and the path $P'$ with edge-set $E(P_1) \cup E(P_2) \cup \{e_1, e_2, e'_3, e\}$ is the $(v, c)$-*absorbing* path in $A'$.
- A vertex in $V(A) \setminus \{v\}$, a colour in $\phi(A) \setminus \{c\}$, or an edge in $E(A)$ is *used* by the $(v, c)$-absorbing gadget $A$.

It is convenient for us to distinguish between a $(v, c)$-absorbing gadget and a $(v, c)$-absorber, because when we build our absorbing structure, we first find a $(v, c)$-absorbing gadget for every $vc \in E(H)$ and then find the paths completing each absorbing gadget. We also find an additional set of paths that 'links' the gadgets together, as in the following definition.

**Definition 3.4.2.** Let $H$ be a bipartite graph with bipartition $(V', C')$ where $V' \subseteq V$ and $C' \subseteq C$, and suppose $\mathcal{A} = \{A_{v,c} : vc \in E(H)\}$ where $A_{v,c}$ is a $(v,c)$-absorbing gadget.

- We say $\mathcal{A}$ *satisfies* $H$ if whenever $A_{v,c}, A_{v',c'} \in \mathcal{A}$ for some $(v,c) \neq (v',c')$, no vertex in $V(A_{v,c})$ or colour in $\phi(A_{v,c})$ is used by $A_{v',c'}$.
- If $\mathcal{P}$ is a collection of vertex-disjoint paths of length 4, then we say $\mathcal{P}$ *completes* $\mathcal{A}$ if the following holds:
  - $\bigcup_{P \in \mathcal{P}} P$ is rainbow,
  - no colour that is either in $C'$ or is used by a $(v,c)$-absorbing gadget $A_{v,c} \in \mathcal{A}$ appears in a path $P \in \mathcal{P}$,
  - no vertex that is either in $V'$ or is used by a $(v,c)$-absorbing gadget $A_{v,c} \in \mathcal{A}$ is an internal vertex of a path $P \in \mathcal{P}$,
  - for every $(v,c)$-absorbing gadget $A_{v,c} \in \mathcal{A}$ there is a pair of paths $P_1, P_2 \in \mathcal{P}$ such that $P_1$ and $P_2$ complete $A_{v,c}$ to a $(v,c)$-absorber $A'_{v,c}$, and
  - the graph $(\bigcup_{A \in \mathcal{A}} A \cup \bigcup_{P \in \mathcal{P}} P) \setminus V'$ is connected and has maximum degree three, and $\mathcal{P}$ is minimal subject to this property.
- We say $(\mathcal{A}, \mathcal{P})$ is an *$H$-absorber* if $\mathcal{A}$ satisfies $H$ and is completed by $\mathcal{P}$. See Figure 3.2.
- We say a rainbow path $T$ is a *tail* of an $H$-absorber $(\mathcal{A}, \mathcal{P})$ if
  - one of the ends of $T$, say $x$, is in a $(v,c)$-absorbing gadget $A_{v,c} \in \mathcal{A}$ such that $x \neq v$,

– $V(T) \cap V(A) \subseteq \{x\}$ for all $A \in \mathcal{A}$ and $V(T) \cap V(P) = \varnothing$ for all $P \in \mathcal{P}$, and

– $\phi(T) \cap \phi(A) = \varnothing$ for all $A \in \mathcal{A}$ and $\phi(T) \cap \phi(P) = \varnothing$ for all $P \in \mathcal{P}$.

- For every matching $M$ in $H$, we define the *path absorbing* $M$ in $(\mathcal{A}, \mathcal{P}, T)$ to be the rainbow path $P$ such that

  – $P$ contains $\bigcup_{P' \in \mathcal{P}} P' \cup T$ and

  – for every $vc \in E(H)$, if $vc \in E(M)$, then $P$ contains the $(v, c)$-absorbing path in the $(v, c)$-absorber $A'_{v,c}$ and $P$ contains the $(v, c)$-avoiding path otherwise (that is, $V(P) \cap V' = V(M) \cap V'$ and $\phi(P) \cap C' = V(M) \cap C'$).

Note that if $\mathcal{P}$ completes $\mathcal{A}$, then some of the paths in $\mathcal{P}$ will complete absorbing gadgets in $\mathcal{A}$ to absorbers, while the remaining set of paths $\mathcal{P}' \subseteq \mathcal{P}$ will be used to connect all the absorbing gadgets in $\mathcal{A}$. More precisely, there is an enumeration $A_1, \ldots, A_{|\mathcal{A}|}$ of $\mathcal{A}$ and an enumeration $P_1, \ldots, P_{|\mathcal{A}|-1}$ of $\mathcal{P}'$ such that each $P_i$ joins $A_i$ to $A_{i+1}$. In particular, for each $i \in [|\mathcal{A}|] \setminus \{1, |\mathcal{A}|\}$, each vertex in $A_i \setminus V'$ is the endpoint of precisely one path in $\mathcal{P}$ (and thus has degree three in $\bigcup_{A \in \mathcal{A}} A \cup \bigcup_{P \in \mathcal{P}} P$), while both $A_1 \setminus V'$ and $A_{|\mathcal{A}|} \setminus V'$ contain precisely one vertex which is not the endpoint of some path in $\mathcal{P}$ (and thus these two vertices have degree two in $\bigcup_{A \in \mathcal{A}} A \cup \bigcup_{P \in \mathcal{P}} P$). Any tail $T$ of an $H$-absorber $(\mathcal{A}, \mathcal{P})$ has to start at one of these two vertices. Altogether this means that, given any matching $M$ of $H$, the path absorbing $M$ in $(\mathcal{A}, \mathcal{P}, T)$ in the definition above actually exists.

Our absorbing structure is essentially an $H$-absorber $(\mathcal{A}, \mathcal{P})$ with a tail $T$ and flexible sets $V_{\text{flex}}, C_{\text{flex}} \subseteq V(H)$ for an appropriately chosen template $H$. If $H - (X \cup Y)$ has a perfect matching $M$, then the path absorbing $M$ in $(\mathcal{A}, \mathcal{P}, T)$ satisfies (†). This fact motivates the property of $H$ that we need in the next definition.

**Definition 3.4.3.** Let $H$ be a bipartite graph with bipartition $(A, B)$ such that $|A| = |B|$, and let $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| = |B'|$.

- We say $H$ is *robustly matchable* with respect to $A'$ and $B'$ if for every pair of sets $X$ and $Y$ where $X \subseteq A'$, $Y \subseteq B'$, and $|X| = |Y| \leq |A'|/2$, there is a perfect matching in $H - (X \cup Y)$.
- In this case, we say $A'$ and $B'$ are *flexible* and $A \setminus A'$ and $B \setminus B'$ are *buffer sets*.

This concept was first introduced by Montgomery [101]. If $H$ is robustly matchable with respect to $V_{\text{flex}}$ and $C_{\text{flex}}$, then an $H$-absorber $(\mathcal{A}, \mathcal{P})$ with tail $T$ satisfies (†). The last property of our absorbing structure that we need is that the flexible sets allow us to execute the covering step, which we capture in the following definition.

**Definition 3.4.4.** Let $V_{\text{flex}} \subseteq V$, let $C_{\text{flex}} \subseteq C$, and let $G_{\text{flex}}$ be a spanning subgraph of $K_n$.

- If $u, v \in V$ and $c \in C$, and $P \subseteq G_{\text{flex}}$ is a rainbow path of length four such that
  - $u$ and $v$ are the ends of $P$,
  - $u', w, v' \in V_{\text{flex}}$, where $uu', u'w, wv', vv' \in E(P)$,
  - $\phi(uu'), \phi(wv'), \phi(vv') \in C_{\text{flex}}$, and
  - $\phi(u'w) = c$,

  then $P$ is a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-*cover* of $u, v$, and $c$.
- If $P$ is a rainbow path such that $P = \bigcup_{i=1}^{k} P_i$ where $P_i$ is a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-cover of $v_i, v_{i+1}$, and $c_i$, then $P$ is a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-*cover* of $\{v_1, \ldots, v_{k+1}\}$

and $\{c_1, \ldots, c_k\}$.

- If $H$ is a regular bipartite graph with bipartition $(V', C')$ where $V' \subseteq V$ and $C' \subseteq C$ such that
  - $H$ is robustly matchable with respect to $V_{\text{flex}}$ and $C_{\text{flex}}$ where $|V_{\text{flex}}| \geq \delta n$, $|C_{\text{flex}}| \geq \delta n$, and
  - for every $u, v \in V$ and $c \in C$, there are at least $\delta n^2$ $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-covers of $u, v$, and $c$,

  then $H$ is a *$\delta$-absorbing template with flexible sets* $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$.
- If $(\mathcal{A}, \mathcal{P})$ is an $H$-absorber where $H$ is a $\delta$-absorbing template and $T$ is a tail for $(\mathcal{A}, \mathcal{P})$, then $(\mathcal{A}, \mathcal{P}, T, H)$ is a *$\delta$-absorber.*

A $36\gamma$-absorber has the properties we need to execute both the covering step and the absorbing step, which we make formal with the next proposition.

First, we introduce the following convenient convention. Given $V'' \subseteq V' \subseteq V$ and $C'' \subseteq C' \subseteq C$, we say that $(V'', C'')$ is *contained* in $(V', C')$ with *$\delta$-bounded remainder* if $V'' \subseteq V'$, $C'' \subseteq C'$, and $|V' \setminus V''|, |C' \setminus C''| \leq \delta n$. If $G$ is a spanning subgraph of $K_n$, $V' \subseteq V$, and $C' \subseteq C$, then we say a graph $G'$ is *contained* in $(V', C', G)$ with *$\delta$-bounded remainder* if $(V(G'), \phi(G'))$ is contained in $(V', C')$ with $\delta$-bounded remainder and $G' \subseteq G$.

**Proposition 3.4.5.** *If $(\mathcal{A}, \mathcal{P}, T, H)$ is a $\delta$-absorber and $P'$ is a rainbow path contained in $(V \setminus V', C \setminus C', G')$ with $\delta/18$-bounded remainder where*

- $V' = \bigcup_{A \in \mathcal{A}} V(A) \cup \bigcup_{P \in \mathcal{P}} V(P) \cup V(T)$
- $C' = \bigcup_{A \in \mathcal{A}} \phi(A) \cup \bigcup_{P \in \mathcal{P}} \phi(P) \cup \phi(T)$*, and*
- *$G'$ is the complement of* $\bigcup_{A \in \mathcal{A}} A \cup \bigcup_{P \in \mathcal{P}} P \cup T$,

*then there is both a rainbow Hamilton path containing $P'$ and a rainbow cycle* *containing $P'$ and all of the colours in $C$.*

**Proof.** Order the colours in $C \setminus (\phi(P') \cup C')$ as $c_1, \ldots, c_k$, and note that $k \leq \delta n/18$. Order the vertices in $V \setminus (V(P') \cup V')$ as $v_1, \ldots, v_\ell$. Using that $H$ is regular, it is easy to see that $|V'| = |C'| + 1$, and thus $\ell = k - 1$. Let $v_0$ and $v_0'$ be the ends of $P'$, let $u$ be the end of $T$ not in $V(A)$ for any $A \in \mathcal{A}$, and let $u'$ be the unique vertex in $\bigcup_{A \in \mathcal{A}} V(A) \setminus V(T)$ of degree two in $\bigcup_{A \in \mathcal{A}} A \cup \bigcup_{P \in \mathcal{P}} P$. Let $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$ be the flexible sets of $H$.

First we show that there is a rainbow Hamilton path containing $P'$. We claim there is a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-cover $P''$ of $\{v_0, \ldots, v_k\}$ and $\{c_1, \ldots, c_k\}$, where $v_k \coloneqq u$. Suppose for $j \in [k-1]$ and $i < j$ that $P_i$ is a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-cover of $v_i, v_{i+1}$, and $c_{i+1}$ such that $\bigcup_{i<j} P_i$ is a rainbow path. We show that there exists a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-cover $P_j$ of $v_j, v_{j+1}$, and $c_{j+1}$ that is internally-vertex- and colour-disjoint from $\bigcup_{i<j} P_i$, which implies that $\bigcup_{i \leq j} P_i$ is a rainbow path, and thus we can choose the path $P''$ greedily, proving the claim. Since each vertex in $V_{\text{flex}}$ and each colour in $C_{\text{flex}}$ is contained in at most $3n$ $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-covers of $v_j, v_{j+1}$, and $c_{j+1}$, and since $H$ is a $\delta$-absorbing template, there are at least $\delta n^2 - 18n \cdot j$ $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-covers of $v_j, v_{j+1}$, and $c_{j+1}$ not containing a vertex or colour from $\bigcup_{i<j} P_i$. Thus, since $j < k \leq \delta n/18$, there exists a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-cover $P_j$ of $v_j, v_{j+1}$, and $c_{j+1}$ such that $\bigcup_{i \leq j} P_i$ is a rainbow path, as desired, and consequently we can choose the path $P''$ greedily, as claimed.

Now let $X \coloneqq V(P'') \cap V_{\text{flex}}$, and let $Y \coloneqq \phi(P'') \cap C_{\text{flex}}$. Since $|X| = |Y| = 3k \leq |V_{\text{flex}}|/2$ and $H$ is robustly matchable with respect to $V_{\text{flex}}$ and $C_{\text{flex}}$, there is a perfect matching $M$ in $H - (X \cup Y)$. Let $P'''$ be the path absorbing $M$ in $(\mathcal{A}, \mathcal{P}, T)$. Then $P' \cup P'' \cup P'''$ is a rainbow Hamilton path, as desired.

Now we show that there is a rainbow cycle containing $P'$ and all of the colours in $C$. By the same argument as before, there is a $(V_{\rm flex}, C_{\rm flex}, G_{\rm flex})$-cover $P_1''$ of $\{v_0, \ldots, v_{\ell-1}, u\}$ and $\{c_1, \ldots, c_{k-1}\}$ as well as a $(V_{\rm flex}, C_{\rm flex}, G_{\rm flex})$-cover $P_2''$ of $v_0'$, $u'$, and $c_k$ such that $P_1''$ and $P_2''$ are vertex-and colour-disjoint. Letting $X :=V(P_1'' \cup P_2'') \cap V_{\rm flex}$ and $Y := \phi(P_1'' \cup P_2'') \cap C_{\rm flex}$, letting $M$ be a perfect matching in $H-(X \cup Y)$ and $P'''$ be the path absorbing $M$ in $(\mathcal{A}, \mathcal{P}, T)$ as before, $P' \cup P_1'' \cup P_2'' \cup P'''$ is a rainbow cycle using all the colours in $C$, as desired. $\square$

### 3.4.1 The proof of Theorem 3.1.3 when $n$ is even

In this subsection, we prove the $n$ even case of Theorem 3.1.3 subject to two lemmas, Lemmas 3.4.8 and 3.4.9, which we prove in Sections 3.7 and 3.6, respectively. The first of these lemmas, Lemma 3.4.8, states that almost all 1-factorizations have two key properties, introduced in the next two definitions. Lemma 3.4.9 states that if a 1-factorization has both of these properties, then we can build an absorber using the reserved vertices and colours with high probability.

Recall the hierarchy of constants $\varepsilon, \gamma, \eta, \mu$ from (3.3.1). Firstly, we will need to show that if $G \in \mathcal{G}^{\rm col}_{[n-1]}$ is chosen uniformly at random, then with high probability, for any $V' \subseteq V$, $C' \subseteq C$ that are not too small, $G$ admits many edges with colour in $C'$ and both endpoints in $V'$. This property will be used in the construction of the tail of our absorber.

**Definition 3.4.6.** For $D \subseteq C = [n-1]$, we say that $G \in \mathcal{G}^{\rm col}_D$ is $\varepsilon$*-locally edge-resilient* if for all sets of colours $D' \subseteq D$ and all sets of vertices $V' \subseteq V$ of sizes $|V'|, |D'| \geq \varepsilon n$, we have that $e_{V',D'}(G) \geq \varepsilon^3 n^2/100$.

Secondly, we will need that almost all $G \in \mathcal{G}^{\mathrm{col}}_{[n-1]}$ contain many $(v, c)$-absorbing gadgets for all $v \in V$, $c \in C$.

**Definition 3.4.7.** Let $D \subseteq C = [n-1]$.

- For $G \in \mathcal{G}^{\mathrm{col}}_{D}$, $x \in V$, $c \in D$, and $t \in \mathbb{N}_0$, we say that a collection $\mathcal{A}_{(x,c)}$ of $(x, c)$-absorbing gadgets in $G$ is *$t$-well-spread* if
  - for all $v \in V$, there are at most $t$ $(x, c)$-absorbing gadgets in $\mathcal{A}_{(x,c)}$ using $v$;
  - for all $e \in E(G)$, there are at most $t$ $(x, c)$-absorbing gadgets in $\mathcal{A}_{(x,c)}$ using $e$;
  - for all $d \in D$, there are at most $t$ $(x, c)$-absorbing gadgets in $\mathcal{A}_{(x,c)}$ using $d$.

  (Note that by definition of 'using' (see Definition 3.4.1), there are no $(x, c)$-absorbing gadgets using $x$ or $c$.)
- We say that $G \in \mathcal{G}^{\mathrm{col}}_{[n-1]}$ is *$\mu$-robustly gadget-resilient* if for all $x \in V$ and all $c \in C$, there is a $5\mu n/4$-well-spread collection of at least $\mu^4 n^2/2^{23}$ $(x, c)$-absorbing gadgets in $G$.

**Lemma 3.4.8.** *Suppose $1/n \ll \varepsilon, \mu \ll 1$. If $\phi$ is a 1-factorization of $K_n$ chosen uniformly at random, then $\phi$ is $\varepsilon$-locally edge-resilient and $\mu$-robustly gadget-resilient with high probability.*

As discussed, we prove Lemma 3.4.8 in Section 3.7 using switching arguments. The next lemma, which we prove in Section 3.6, is used to construct an absorber using the reserved vertices and colours.

**Lemma 3.4.9.** *Suppose $1/n \ll \varepsilon \ll \gamma \ll \eta \ll \mu \ll 1$, and let $p = q = \beta = 5\mu + 26887\eta/2 + \gamma/3 - 26880\varepsilon$. If $\phi$ is an $\varepsilon$-locally edge-resilient and $\mu$-robustly* *gadget-resilient 1-factorization of $K_n$ with vertex set $V$ and colour set $C$ and*

*(R1) $V'$ is a $p$-random subset of $V$,*

*(R2) $C'$ is a $q$-random subset of $C$, and*

*(R3) $G'$ is a $\beta$-random subgraph of $K_n$,*

*then with high probability there is a $36\gamma$-absorber $(\mathcal{A}, \mathcal{P}, T, H)$ such that $\bigcup_{A\in\mathcal{A}} A \cup \bigcup_{P\in\mathcal{P}} P \cup T$ is contained in $(V', C', G')$ with $\gamma$-bounded remainder.*

The final ingredient in the proof of Theorem 3.1.3 is the following lemma which follows from [54, Lemma 16], that enables us to find the long rainbow path whose leftover we absorb using the absorber from Lemma 3.4.9.

**Lemma 3.4.10.** *Suppose $1/n \ll \gamma \ll p$, and let $q = \beta = p$. For every 1-factorization $\phi$ of $K_n$ with vertex set $V$ and colour set $C$, if*

- *$V'$ is a $p$-random subset of $V$,*
- *$C'$ is a $q$-random subset of $C$, and*
- *$G$ is a $\beta$-random subgraph of $K_n$,*

*then with high probability there is a rainbow path contained in $(V', C', G)$ with $\gamma$-bounded remainder.*

We conclude this section with a proof of Theorem 3.1.3 in the case that $n$ is even, assuming Lemmas 3.4.8 and 3.4.9.

**Proof of Theorem 3.1.3, $n$ even case.** By Lemma 3.4.8, it suffices to prove that if $\phi$ is an $\varepsilon$-locally edge-resilient and $\mu$-robustly gadget-resilient 1-factorization, then there is a rainbow Hamilton path and a rainbow cycle containing all of the colours.

Let $p = q = \beta$ as in Lemma 3.4.9, let $V_1, V_2$ be a random partition of $V$ where $V_1$ is $p$-random and $V_2$ is $(1-p)$-random, let $C_1, C_2$ be a random partition of $C$

where $C_1$ is $q$-random and $C_2$ is $(1-q)$-random, and let $G_1$ and $G_2$ be $\beta$-random and $(1-\beta)$-random subgraphs of $K_n$ such that $E(G_1)$ and $E(G_2)$ partition the edges of $K_n$. By Lemma 3.4.9 applied with $V' = V_1$, $C' = C_1$, and $G' = G_1$, and by Lemma 3.4.10 applied with $V' = V_2$, $C' = C_2$, and $G = G_2$, the following holds with high probability. There exists

(i) a $36\gamma$-absorber $(\mathcal{A}, \mathcal{P}, T, H)$ such that $\bigcup_{A\in\mathcal{A}} A \cup \bigcup_{P\in\mathcal{P}} P \cup T$ is contained in $(V_1, C_1, G_1)$ with $\gamma$-bounded remainder, and

(ii) a rainbow path $P'$ contained in $(V_2, C_2, G_2)$ with $\gamma$-bounded remainder.

Now we fix an outcome of the random partitions $(V_1, V_2)$, $(C_1, C_2)$, and $(G_1, G_2)$ so that (i) and (ii) hold. By Proposition 3.4.5, there is both a rainbow Hamilton path containing $P'$ and a rainbow cycle containing $P'$ and all of the colours in $C$, as desired. $\square$

## 3.5 Tools

In this section, we collect some results that we will use throughout the paper.

### 3.5.1 Probabilistic tools

We will use the following standard probabilistic estimates.

**Lemma 3.5.1** (Chernoff Bound). *Let $X$ have binomial distribution with parameters $n, p$. Then for any $0 < t \leq np$,*

$$\mathbb{P}\left[|X - np| > t\right] \leq 2\exp\left(\frac{-t^2}{3np}\right).$$

Let $X_1, \ldots, X_m$ be independent random variables taking values in $\mathcal{X}$, and let $f : \mathcal{X}^m \to \mathbb{R}$. If for all $i \in [m]$ and $x_i', x_1, \ldots, x_m \in \mathcal{X}$, we have

$$|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_m)| \le c_i,$$

then we say $X_i$ *affects* $f$ by at most $c_i$.

**Theorem 3.5.2** (McDiarmid's Inequality)**.** *If $X_1, \ldots, X_m$ are independent random variables taking values in $\mathcal{X}$ and $f : \mathcal{X}^m \to \mathbb{R}$ is such that $X_i$ affects $f$ by at most $c_i$ for all $i \in [m]$, then for all $t > 0$,*

$$\mathbb{P}\left[|f(X_1, \ldots, X_m) - \mathbb{E}\left[f(X_1, \ldots, X_m)\right]| \ge t\right] \le \exp\left(-\frac{t^2}{\sum_{i=1}^m c_i^2}\right).$$

### 3.5.2 Hypergraph matchings

When we build our absorber in the proof of Lemma 3.4.9, we seek to efficiently use the vertices, colours, and edges of our random subsets $V' \subseteq V$, $C' \subseteq C$, $E' \subseteq E$, and to do this we make use of the existence of large matchings in almost-regular hypergraphs with small codegree. In fact, we will need the stronger property that there exists a large matching in such a hypergraph which is well-distributed with respect to a specified collection of vertex subsets. We make this precise in the following definition. Given a hypergraph $\mathcal{H}$ and a collection of subsets $\mathcal{F}$ of $V(\mathcal{H})$, we say a matching $\mathcal{M}$ in $\mathcal{H}$ is $(\gamma, \mathcal{F})$*-perfect* if for each $F \in \mathcal{F}$, at most $\gamma \cdot \max\{|F|, |V(\mathcal{H})|^{2/5}\}$ vertices of $\mathcal{F}$ are left uncovered by $\mathcal{M}$. The following theorem is a consequence of Theorem 1.2 in [8], and is based on a result of Pippenger and Spencer [104].

**Theorem 3.5.3.** *Suppose $1/n \ll \varepsilon \ll \gamma \ll 1/r$. Let $\mathcal{H}$ be an $r$-uniform hypergraph on $n$ vertices such that for some $D \in \mathbb{N}$, we have $d_{\mathcal{H}}(x) = (1 \pm \varepsilon)D$ for all $x \in V(\mathcal{H})$ and $\Delta^c(\mathcal{H}) \leq D/\log^{9r} n$. If $\mathcal{F}$ is a collection of subsets of $V(\mathcal{H})$ such that $|\mathcal{F}| \leq n^{\log n}$, then there exists a $(\gamma, \mathcal{F})$-perfect matching.*

We will use Theorem 3.5.3 in the final step of constructing an absorber (see Lemma 3.6.5). We construct an auxiliary hypergraph $\mathcal{H}$ whose edges represent structures we wish to find, and a large well-distributed matching in $\mathcal{H}$ corresponds to an efficient allocation of vertices, colours, and edges of the 1-factorization to construct almost all of these desired structures. We remark that this is also a key strategy in the proof of Lemma 3.4.10, and was first used in [80].

### 3.5.3 Robustly matchable bipartite graphs of constant degree

In this subsection, we prove that there exist large bipartite graphs which are robustly matchable as in Definition 3.4.3, and have constant maximum degree.

**Definition 3.5.4.** Let $m \in \mathbb{N}$.

- An $RMBG(3m, 2m, 2m)$ is a bipartite graph $H$ with bipartition $(A, B_1 \cup B_2)$ where $|A| = 3m$ and $|B_1| = |B_2| = 2m$ such that for any $B' \subsetneq B_1$ of size $m$, there is a perfect matching in $H - B'$. In this case, we say $H$ is *robustly matchable* with respect to $B_1$, and that $B_1$ is the identified *flexible set*.
- A $2RMBG(7m, 2m)$ is a bipartite graph $H$ with bipartition $(A, B)$ where $|A| = |B| = 7m$ such that $H$ is robustly matchable with respect to sets $A' \subsetneq A$ and $B' \subsetneq B$ where $|A'| = |B'| = 2m$.

By [101, Lemma 10.7], there exists an $RMBG(3m, 2m, 2m)$ with maximum degree at most 100 for all sufficiently large $m$. We use a one-sided (there is one flexible set) $RMBG(3m, 2m, 2m)$ exhibited in [54, Corollary 10] in which each of the vertex classes are regular, to construct a 256-regular two-sided (in that we identify a flexible set on each side of the vertex bipartition) $2RMBG(7m, 2m)$.

**Lemma 3.5.5.** *For all sufficiently large $m$, there is a $2RMBG(7m, 2m)$ that is 256-regular.*

**Proof.** Suppose that $m \in \mathbb{N}$ is sufficiently large. By [54, Corollary 10], there exists an $RMBG(3m, 2m, 2m)$ that is (256, 192)-regular (i.e. all vertices in the first vertex class have degree 256 and all vertices in the second vertex class of have degree 192). Let $H$ and $H'$ be two vertex-disjoint isomorphic copies of a $(256, 192)$-regular $RMBG(3m, 2m, 2m)$, and let $(A, B_1 \cup B_2)$ and $(A', B_1' \cup B_2')$ be the bipartitions of $H$ and $H'$ respectively such that $H$ is robustly matchable with respect to $B_1$ and $H'$ is robustly matchable with respect to $B_1'$.

Let $H''$ be a 64-regular bipartite graph with bipartition $(B_1 \cup B_2, B_1' \cup B_2')$ such that $H''[B_1 \cup B_1']$ contains a perfect matching $M$. We claim that $H \cup H' \cup H''$ is robustly matchable with respect to $B_1$ and $B_1'$. To that end, let $X \subseteq B_1$ and $Y \subseteq B_1'$ such that $|X| = |Y| \leq m$. It suffices to show that $H \cup H' \cup H'' - (X \cup Y)$ has a perfect matching. Since $H''[B_1 \cup B_1']$ contains a perfect matching, $H''[B_1 \cup B_1'] - (X \cup Y)$ contains a matching of size at least $2m - |X| - |Y| = 2(m - |X|)$. Thus, there exists a matching $M'$ in $H''[B_1 \cup B_1'] - (X \cup Y)$ of size $m - |X|$. Let $X' := X \cup (B_1 \cap V(M'))$ and $Y' := Y \cup (B_1' \cap V(M'))$, and note that $|X'| = |Y'| = m$. Since $H$ is an $RMBG(3m, 2m, 2m)$, $H - X'$ has a perfect matching $M_1$, and similarly $H' - Y'$ has a perfect matching $M_2$. Now $M' \cup M_1 \cup M_2$ is a perfect matching in $H \cup H' \cup H'' - (X \cup Y)$, as required. Since $H \cup H' \cup H''$ is 256-regular, the result follows. $\square$

## 3.6 Constructing the absorber

Throughout this section, let $\phi$ be an $\varepsilon$-locally edge-resilient and $\mu$-robustly gadget resilient 1-factorization of $K_n$ with vertex set $V$ and colour set $C$, let $E := E(K_n)$, and recall

$$1/n \ll \varepsilon \ll \gamma \ll \eta \ll \mu \ll 1.$$

Let $\tilde{H}$ be a 256-regular $2RMBG(7m, 2m)$ where $2m = (\eta - 2\varepsilon)n$, which exists by Lemma 3.5.5. We define the following probabilities:

$$\begin{aligned} p_{\text{flex}} &:= \eta, \\ p_{\text{buff}} &:= 5\eta/2, \\ p_{\text{abs}} &:= 6|E(\tilde{H})|/n + 2\mu, \quad (3.6.1) \\ p_{\text{link}} &:= 9|E(\tilde{H})|/n + 3\mu, \\ p'_{\text{link}} &:= \gamma/3, \end{aligned} \qquad \begin{aligned} q_{\text{flex}} &:= \eta, \\ q_{\text{buff}} &:= 5\eta/2, \\ q_{\text{abs}} &:= 3|E(\tilde{H})|/n + \mu, \\ q_{\text{link}} &:= 12|E(\tilde{H})|/n + 4\mu, \\ q'_{\text{link}} &:= \gamma/3, \end{aligned}$$

and we let $p_{\text{main}} := 1 - p_{\text{flex}} - p_{\text{buff}} - p_{\text{abs}} - p_{\text{link}} - p'_{\text{link}}$ and $q_{\text{main}} := 1 - q_{\text{flex}} - q_{\text{buff}} - q_{\text{abs}} - q_{\text{link}} - q'_{\text{link}}$. Note that $p_{\text{main}} = q_{\text{main}}$, and let $\beta := 1 - p_{\text{main}}$.

**Definition 3.6.1.** An *absorber partition of $V$, $C$, and $K_n$* is defined as follows:

$$\begin{aligned} V &= V_{\text{main}} \dot\cup V_{\text{flex}} \dot\cup V_{\text{buff}} \dot\cup V_{\text{abs}} \dot\cup V_{\text{link}} \dot\cup V'_{\text{link}}, \text{ and} \\ C &= C_{\text{main}} \dot\cup C_{\text{flex}} \dot\cup C_{\text{buff}} \dot\cup C_{\text{abs}} \dot\cup C_{\text{link}} \dot\cup C'_{\text{link}}, \end{aligned} \tag{3.6.2}$$

where $V_{\text{main}}$ is $p_{\text{main}}$-random, $V_{\text{flex}}$ is $p_{\text{flex}}$-random etc, and the sets of colours are defined analogously. Let $V' := V \setminus V_{\text{main}}$, $C' := C \setminus C_{\text{main}}$, and let $G'$ be a $\beta$-random subgraph of $K_n$.

Note that $V'$, $C'$, and $G'$ satisfy (R1)–(R3) in the statement of Lemma 3.4.9.

### 3.6.1 Overview of the proof

We now overview our strategy for proving Lemma 3.4.9. First we need the following definitions. A *link* is a rainbow path of length 4 with internal vertices in $V_{\text{link}} \cup V'_{\text{link}}$, ends in $V_{\text{abs}}$, and colours and edges in $C_{\text{link}} \cup C'_{\text{link}}$ and $G'$, respectively. A link with internal vertices in $V_{\text{link}}$ and colours in $C_{\text{link}}$ is a *main link*, and a link with internal vertices in $V'_{\text{link}}$ and colours in $C'_{\text{link}}$ is a *reserve link*. If $M$ is a matching and $\mathcal{P} = \{P_e\}_{e \in E(M)}$ is a collection of vertex-disjoint links such that $\bigcup_{P \in \mathcal{P}} P$ is rainbow and $P_{uv}$ has ends $u$ and $v$ for every $uv \in E(M)$, then $\mathcal{P}$ *links* $M$.

We aim to build a $36\gamma$-absorber $(\mathcal{A}, \mathcal{P}, T, H)$ such that $\bigcup_{A \in \mathcal{A}} A \cup \bigcup_{P \in \mathcal{P}} P \cup T$ is contained in $(V', C', G')$ with $\gamma$-bounded remainder and $H \cong \tilde{H}$. First, we show (see Lemma 3.6.3) that with high probability there is a $36\gamma$-absorbing template $H \cong \tilde{H}$, where

- $H$ has flexible sets $(V'_{\text{flex}}, C'_{\text{flex}}, G')$ and $(V'_{\text{flex}}, C'_{\text{flex}})$ is contained in $(V_{\text{flex}}, C_{\text{flex}})$ with $3\varepsilon$-bounded remainder, and
- $H$ has buffer sets $V'_{\text{buff}}$ and $C'_{\text{buff}}$ where $(V'_{\text{buff}}, C'_{\text{buff}})$ is contained in $(V_{\text{buff}}, C_{\text{buff}})$ with $6\varepsilon$-bounded remainder.

Then, we show that with high probability, there exists an $H$-absorber $(\mathcal{A}, \mathcal{P})$ where

- for every $vc \in E(H)$, the $(v, c)$-absorbing gadget $A_{v,c} \in \mathcal{A}$ uses vertices, colours, and edges in $V_{\text{abs}}$, $C_{\text{abs}}$, and $G'$, respectively, and

- every $P \in \mathcal{P}$ is a link.

In particular, if $\mathcal{A} = \{A_1, \ldots, A_k\}$, where $A_i$ is a $(v_i, c_i)$-absorbing gadget, then $\mathcal{P}$ links the matching $M_1 \cup M_2 \cup M_3$, where $V(M_1)$, $V(M_2)$, and $V(M_3)$ are pairwise vertex-disjoint, and

(M1) $M_1 = \{r_1s_1, \ldots, r_ks_k\}$, where $r_i$ and $s_i$ are non-adjacent vertices of the 4-cycle in $A_i$, for each $i \in [k]$,

(M2) $M_2 = \{w_1x_1, \ldots, w_kx_k\}$, where $w_i$ is a non-$v_i$ vertex of the triangle in $A_i$ and $x_i$ is a vertex of the 4-cycle in $A_i$, for each $i \in [k]$, and

(M3) $M_3 = \{y_1z_2, \ldots, y_{k-1}z_k\}$, where $y_i$ is a non-$v_i$ vertex of the triangle in $A_i$ for each $i \in [k-1]$, and $z_i$ is a vertex of the 4-cycle in $A_i$ for each $i \in [k] \setminus \{1\}$.

Finally, letting $V'_{\text{abs}}$ and $C'_{\text{abs}}$ be the vertices and colours in $V_{\text{abs}}$ and $C_{\text{abs}}$ not used by any $(v, c)$-absorbing gadget in $\mathcal{A}$, we show that with high probability there is a tail $T$ for $(\mathcal{A}, \mathcal{P})$ where $T$ is the union of

- a rainbow matching $M$ contained in $(V'_{\text{abs}}, C'_{\text{abs}}, G')$ with $6\varepsilon$-bounded remainder and
- a collection $\mathcal{T}$ of vertex-disjoint links where all but one vertex in $V(M)$ is the end of precisely one link.

In particular, if $E(M) = \{a_1b_1, \ldots, a_\ell b_\ell\}$, then $\mathcal{T}$ links $M_4$, where

(M4) $M_4$ is a matching of size $\ell$ with edges $b_ia_{i+1}$ for every $i \in [\ell - 1]$ and an edge $va_1$ where $v$ is one of the two vertices used by a gadget in $\mathcal{A}$ that is not in a link in $\mathcal{P}$.

See Figure 3.3.

**Fact 3.6.2.** *Suppose that $\mathcal{A}$ satisfies $H$. If $\mathcal{P} \cup \mathcal{T}$ links $M_1 \cup \cdots \cup M_4$, where $\mathcal{P}$*

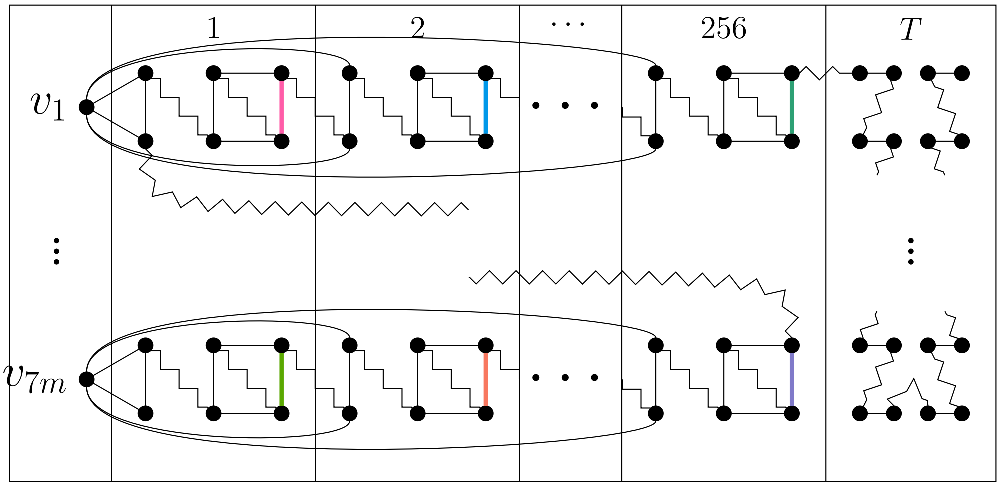



Figure 3.3: An absorber $(\mathcal{A}, \mathcal{P}, T, H)$, where $\mathcal{P}$ links $\bigcup_{i=1}^{3} M_i$ and $T = M \cup \bigcup_{P \in \mathcal{T}} P$, where $\mathcal{T}$ links $M_4$. Links are drawn as zigzags.

*links $M_1 \cup M_2 \cup M_3$ and $\mathcal{T}$ links $M_4$, then $\mathcal{P}$ completes $\mathcal{A}$ and thus $(\mathcal{A}, \mathcal{P})$ is an $H$-absorber. Moreover, $T := M \cup \bigcup_{P \in \mathcal{T}} P$ is a tail of $(\mathcal{A}, \mathcal{P})$. Thus $(\mathcal{A}, \mathcal{P}, T, H)$ is a $36\gamma$-absorber.*

We find these structures in the following steps. For Steps 1 and 2, see Lemma 3.6.4, and for Steps 3 and 4, see Lemma 3.6.5.

1) First, we find the collection $\mathcal{A}$ of absorbing gadgets greedily, using the robust gadget-resilience property of $\phi$,
2) then we greedily construct the matching $M$, using the local edge-resilience property of $\phi$.
3) Next, we construct an auxiliary hypergraph in which each hyperedge corresponds to a main link and apply Theorem 3.5.3 to choose most of the links in $\mathcal{P}$, and
4) finally we greedily choose the remainder of the links in $\mathcal{P}$ from the reserve links.

### 3.6.2 The absorbing template

**Lemma 3.6.3.** *Consider an absorber partition of $V$, $C$, and $K_n$. With high probability, there exists a $36\gamma$-absorbing template $H \cong \tilde{H}$, where*

*(3.6.3.1) $H$ has flexible sets $(V'_{\text{flex}}, C'_{\text{flex}}, G')$ where $(V'_{flex}, C'_{flex})$ is contained in $(V_{\text{flex}}, C_{\text{flex}})$ with $3\varepsilon$-bounded remainder, and*

*(3.6.3.2) $H$ has buffer sets $V'_{\text{buff}}$ and $C'_{\text{buff}}$ where $(V'_{\text{buff}}, C'_{\text{buff}})$ is contained in $(V_{\text{buff}}, C_{\text{buff}})$ with $6\varepsilon$-bounded remainder.*

**Proof.** For convenience, let $p \coloneqq p_{\text{flex}}$ and $q \coloneqq q_{\text{flex}}$. We claim that the following holds with high probability:

(a) $|V_{\text{flex}}|, |C_{\text{flex}}| = (\eta \pm \varepsilon)n$,

(b) $|V_{\text{buff}}|, |C_{\text{buff}}| = (5\eta/2 \pm \varepsilon)n$, and

(c) for every distinct $u, v \in V$ and $c \in C$, there are at least $p^3q^3\beta^4n^2/4$ $(V_{\text{flex}}, C_{\text{flex}}, G')$-covers of $u$, $v$, and $c$.

Indeed, (a) and (b) follow from the Chernoff Bound (Lemma 3.5.1). To prove (c), for each $u, v$, and $c$, we apply McDiarmid's Inequality (Theorem 3.5.2). Consider the random variable $f$ counting the number of $(V_{\text{flex}}, C_{\text{flex}}, G')$-covers of $u$, $v$, and $c$. Note that $f$ is determined by the following independent binomial random variables: $\{X_z\}_{z\in V}$, where $X_z$ indicates if $z \in V_{\text{flex}}$, $\{X_{c'}\}_{c'\in C}$, where $X_{c'}$ indicates if $c' \in C_{\text{flex}}$, and for each edge $e$, the random variable $X_e$ which indicates if $e \in E(G')$. We claim there are at least $2(n/2-2)(n-7)$ $(V, C, K_n)$-covers of $u$, $v$, and $c$. To that end, let $u'w$ be a $c$-edge with $u', w \in V \setminus \{u, v\}$. There are at least $n-7$ vertices $v' \in V \setminus \{u, v, u', w\}$ such that $\phi(vv'), \phi(wv') \notin \{\phi(uu'), c\}$, and for each such vertex $v'$ the path $uu'wv'v$ is a $(V, C, K_n)$-cover of $u$, $v$, and $c$. Similarly, there are at least $n-7$ $(V, C, K_n)$-covers of the form $uwu'v'v$. Altogether this gives at least

$2(n/2-2)(n-7) \geq n^2/2$ $(V, C, K_n)$-covers of $u$, $v$, and $c$, as claimed. Therefore $\mathbb{E}[f] \geq p^3q^3\beta^4n^2/2$. For each $z \in V$, $X_z$ affects $f$ by at most $3n$, and $X_{uz}$, and $X_{vz}$ each affect $f$ by at most $n$, and for each $c' \in C$, $X_{c'}$ affects $f$ by at most $3n$. For each edge $e$ not incident to $u$ or $v$, if $e$ is a $c$-edge, then $X_e$ affects $f$ by at most $2n$, and otherwise $e$ affects $f$ by at most two. Thus, by McDiarmid's Inequality applied with $t = \mathbb{E}[f]/2$, there are at least $p^3q^3\beta^4n^2/4$ $(V_{\text{flex}}, C_{\text{flex}}, G')$-covers of $u$, $v$, and $c$ with probability at least $1 - \exp(-p^6q^6\beta^8n^4/O(n^3))$. Thus by a union bound, (c) also holds with high probability.

Now we assume (a)–(c) holds, and we show there exists a $36\gamma$-absorbing template $H \cong \tilde{H}$ satisfying (3.6.3.1) and (3.6.3.2).

Since $m = (\eta/2 - \varepsilon)n$, by (a) and (b), there exists $V'_{\text{flex}} \subseteq V_{\text{flex}}$, $C'_{\text{flex}} \subseteq C_{\text{flex}}$, $V'_{\text{buff}} \subseteq V_{\text{buff}}$, and $C'_{\text{buff}} \subseteq C_{\text{buff}}$, such that $|V'_{\text{flex}}|, |C'_{\text{flex}}| = 2m$ and $|V'_{\text{buff}}|, |C'_{\text{buff}}| = 5m$, which we choose arbitrarily, and moreover, $|V_{\text{flex}} \setminus V'_{\text{flex}}|, |C_{\text{flex}} \setminus C'_{\text{flex}}| \leq 3\varepsilon n$ and $|V_{\text{buff}} \setminus V'_{\text{buff}}|, |C_{\text{buff}} \setminus C'_{\text{buff}}| \leq 6\varepsilon n$, as required. Choose bijections from $V'_{\text{flex}}, C'_{\text{flex}}$, $V'_{\text{buff}}$, and $C'_{\text{buff}}$ to the flexible sets and the buffer sets of $\tilde{H}$ arbitrarily, and let $H \cong \tilde{H}$ be the corresponding graph. Now $H$ satisfies (3.6.3.1) and (3.6.3.2), as required, so it remains to show that $H$ is a $36\gamma$-absorbing template. Since each vertex or colour in $V_{\text{flex}}$ or $C_{\text{flex}}$ is in at most $3n$ $(V_{\text{flex}}, C_{\text{flex}}, G')$-covers of $u$, $v$, and $c$, (a) and (c) imply that there are at least $p^3q^3\beta^4n^2/4 - 18\varepsilon n^2 \geq 36\gamma n^2$ $(V'_{\text{flex}}, C'_{\text{flex}}, G')$-covers of $u$, $v$, and $c$, so $H$ is a $36\gamma$-absorbing template, as desired. $\square$

### 3.6.3 Greedily building an $H$-absorber

**Lemma 3.6.4.** *Consider an absorber partition of $V$, $C$, and $K_n$. The following holds with high probability. Suppose $V_{res} \subseteq V_{flex} \cup V_{buff}$ and $C_{res} \subseteq C_{flex} \cup C_{buff}$. For* *every graph $H \cong \tilde{H}$ with bipartition $(V_{\mathrm{res}}, C_{\mathrm{res}})$, there exists*

(3.6.4.1) *a collection $\mathcal{A} = \{A_{vc} \colon vc \in E(H)\}$ such that $\mathcal{A}$ satisfies $H$ and such that for all $A_{vc} \in \mathcal{A}$ we have that $A_{vc}$ uses vertices, colours, and edges in $V_{\mathrm{abs}}$, $C_{\mathrm{abs}}$, and $G'$ respectively, and*

(3.6.4.2) *a rainbow matching $M$ contained in $(V'_{\mathrm{abs}}, C'_{\mathrm{abs}}, G')$ with $5\varepsilon$-bounded remainder, where $V'_{abs}$ and $C'_{abs}$ are the sets of vertices and colours in $V_{abs}$ and $C_{abs}$ not used by any absorbing gadget in $\mathcal{A}$.*

**Proof.** For convenience, let $p \coloneqq p_{\mathrm{abs}}$ and $q \coloneqq q_{\mathrm{abs}}$ in this proof.

Since $\phi$ is $\mu$-robustly gadget-resilient, for every $v \in V$, $c \in C$, there is a collection $\mathcal{A}_{v,c}$ of precisely $2^{-23}\mu^4 n^2$ $(v,c)$-absorbing gadgets such that every vertex, every colour, and every edge is used by at most $5\mu n/4$ of the $A \in \mathcal{A}_{v,c}$. (Recall from Definition 3.4.1 that a $(v,c)$-absorbing gadget does not use $v$ and $c$.) Fix $v \in V$, $c \in C$. The expected number of the $(v,c)$-absorbing gadgets in $\mathcal{A}_{v,c}$ using only vertices in $V_{\mathrm{abs}}$, colours in $C_{\mathrm{abs}}$, and edges in $G'$ is $p^6 q^3 \beta^7 |\mathcal{A}_{v,c}|$. Let $\mathcal{E}_{v,c}$ be the event that fewer than $p^6 q^3 \beta^7 |\mathcal{A}_{v,c}|/2$ of the $(v,c)$-absorbing gadgets in $\mathcal{A}_{v,c}$ use only vertices in $V_{\mathrm{abs}}$, colours in $C_{\mathrm{abs}}$ and edges in $G'$. We claim that $\mathbb{P}\left[\mathcal{E}_{v,c}\right] \leq \exp(-2^{-51} p^{12} q^6 \beta^{14} \mu^6 n)$.

To see this, for each $u \in V$, $d \in C$, $e \in E$, let $m_u$, $m_d$, and $m_e$ denote the number of $(v,c)$-absorbing gadgets in $\mathcal{A}_{v,c}$ using $u$, $d$, and $e$, respectively. We will apply McDiarmid's Inequality (Theorem 3.5.2) to the function $f_{v,c}$ which counts the number of $A \in \mathcal{A}_{v,c}$ using only vertices in $V_{\mathrm{abs}}$, colours in $C_{\mathrm{abs}}$, and edges in $G'$. We use independent indicator random variables $\{X_u\}_{u \in V} \cup \{X_d\}_{d \in C} \cup \{X_e\}_{e \in E}$ which indicate whether or not a vertex $u$ is in $V_{\mathrm{abs}}$, a colour $d$ is in $C_{\mathrm{abs}}$, and an edge $e$ is in $G'$. Each random variable $X_u$, $X_d$, $X_e$ affects $f_{v,c}$ by at most $m_u$, $m_d$, $m_e$, respectively. Since $m_u \leq 5\mu n/4$ for all $u \in V$ and $m_d \leq 5\mu n/4$ for all $d \in C$, we have $\sum_{u \in V} m_u^2, \sum_{d \in C} m_d^2 \leq 25\mu^2 n^3/16$. Since $\sum_{e \in E} m_e = 7|\mathcal{A}_{v,c}|$ and $m_e \leq 5\mu n/4$ for all $e \in E$, it follows that $\sum_{e \in E} m_e^2 \leq 35\mu n|\mathcal{A}_{v,c}|/4$. Therefore, by McDiarmid's Inequality, we have

$$\mathbb{P}\left[\mathcal{E}_{v,c}\right] \leq \exp\left(-\frac{p^{12}q^6\beta^{14}|\mathcal{A}_{v,c}|^2/4}{25\mu^2n^3/8 + 35\mu n|\mathcal{A}_{v,c}|/4}\right) \leq \exp(-2^{-51}p^{12}q^6\beta^{14}\mu^6 n),$$

as claimed. Thus, by a union bound, the probability that there exist $v \in V$, $c \in C$ such that $\mathcal{E}_{v,c}$ holds is at most $\exp(-2^{-52}p^{12}q^6\beta^{14}\mu^6 n)$.

We claim the following holds with high probability:

(a) $|V_{\text{abs}}| = (p \pm \varepsilon)n$ and $|C_{\text{abs}}| = (q \pm \varepsilon)n$;

(b) for every $v \in V$, $c \in C$, the event $\mathcal{E}_{v,c}$ does not hold;

(c) for every $V^\circ \subseteq V_{\text{abs}}$ and $C^\circ \subseteq C_{\text{abs}}$ such that $|V^\circ|, |C^\circ| \geq \varepsilon n$, there are at least $\beta\varepsilon^3 n^2/200$ edges in $G'$ with both ends in $V^\circ$ and a colour in $C^\circ$.

Indeed, (a) holds by the Chernoff Bound (Lemma 3.5.1), we have already shown (b), and since $\phi$ is $\varepsilon$-locally edge-resilient, (c) holds by applying the Chernoff Bound for each $V^\circ$ and $C^\circ$ and using a union bound.

Now we assume that (a)–(c) hold, we suppose $H \cong \tilde{H}$ has bipartition $(V_{\text{res}}, C_{\text{res}})$ contained in $(V_{\text{flex}} \cup V_{\text{buff}}, C_{\text{flex}} \cup C_{\text{buff}})$, and we show that (3.6.4.1) and (3.6.4.2) hold. Arbitrarily order the edges of $H$ as $e_1, \ldots, e_{|E(H)|}$. Let $i \in [|E(H)|]$ and suppose that for each $j < i$ we have found a $(v_j, c_j)$-absorbing gadget $A_j$, where $e_j = v_j c_j$, and further, the collection $\{A_1, \ldots, A_{i-1}\}$ satisfies the spanning subgraph of $H$ containing precisely the edges $e_1, \ldots, e_{i-1}$. Writing $e_i = v_i c_i$, by (b) there is a collection $\mathcal{A}^{\text{abs}}_{v_i,c_i}$ of at least $2^{-24}p^6q^3\beta^7\mu^4n^2$ $(v_i, c_i)$-absorbing gadgets each using only $V_{\text{abs}}$-vertices, $C_{\text{abs}}$-colours, and $G'$-edges, and moreover, each vertex in $V_{\text{abs}}$, colour in $C_{\text{abs}}$, and edge in $G'$ is used by at most $5\mu n/4$ of the $A \in \mathcal{A}^{\text{abs}}_{v_i,c_i}$. Thus, at most $20\mu n \cdot i \leq 20\mu n|E(H)| \leq 17920\eta\mu n^2$ of the $(v_i, c_i)$-absorbing gadgets in $\mathcal{A}^{\text{abs}}_{v_i,c_i}$ use a vertex, colour, or edge used by any of the $A_j$ for $j < i$. Since $|\mathcal{A}^{\text{abs}}_{v_i,c_i}| \geq 2^{-24}p^6q^3\beta^7\mu^4n^2$, we conclude that there is at least one $(v_i, c_i)$-absorbing gadget $A \in \mathcal{A}^{\text{abs}}_{v_i,c_i}$ using vertices, colours, and edges which are disjoint from the vertices, colours, and edges used by $A_j$, for all $j < i$. We arbitrarily choose such an $A$ to be $A_i$. Continuing in this way, it is clear that $\mathcal{A} \coloneqq \{A_i\}_{i=1}^{|E(H)|}$ satisfies $H$, so (3.6.4.1) holds.

Now we prove (3.6.4.2). Let $V'_{\text{abs}}$ and $C'_{\text{abs}}$ be the vertices, colours, and edges in $V_{\text{abs}}$ and $C_{\text{abs}}$ not used by any $(v, c)$-absorbing gadget in $\mathcal{A}$. By (a) and (3.6.1), we have $|V'_{\text{abs}}| = (2\mu \pm \varepsilon)n$ and $|C'_{\text{abs}}| = (\mu \pm \varepsilon)n$. Thus, by (c), we can greedily choose a rainbow matching $M$ in $(V'_{\text{abs}}, C'_{\text{abs}}, G')$ of size at least $(\mu - 2\varepsilon)n$, and $M$ satisfies (3.6.4.2). $\square$

### 3.6.4 Linking

Lastly, we need the following lemma, inspired by [54, Lemma 20], which we use to both complete the set of absorbing gadgets obtained by Lemma 3.6.4 to an $H$-absorber and also construct its tail. Recall that links were defined at the beginning of Section 3.6.1.

**Lemma 3.6.5.** *Consider an absorber partition of $V$, $C$, and $K_n$. The following holds with high probability. For every matching $M$ such that $V(M) \subseteq V_{\text{abs}}$ and $|V_{\text{abs}} \setminus V(M)| \leq \varepsilon n$, there exists a collection $\mathcal{P}$ of links in $G'$ such that*

*(3.6.5.1) $\mathcal{P}$ links $M$ and*

*(3.6.5.2)* $\bigcup_{P\in\mathcal{P}} P\setminus V(M)$ *is contained in* $(V_{\mathrm{link}}\cup V'_{\mathrm{link}}, C_{\mathrm{link}}\cup C'_{\mathrm{link}}, G')$ *with* $\gamma/2$*-bounded remainder.*

**Proof.** We choose a new constant $\delta$ such that $\varepsilon \ll \delta \ll \gamma$. For convenience, let $p \coloneqq p_{\mathrm{link}}$ and $q \coloneqq q_{\mathrm{link}}$, let $G_1$ be the spanning subgraph of $G'$ consisting of edges with a colour in $C_{\mathrm{link}}$, and let $G_2$ be the spanning subgraph of $G'$ consisting of edges with a colour in $C'_{\mathrm{link}}$. First we claim that with high probability the following holds:

(a) $|V_{\mathrm{link}}| = (p \pm \varepsilon)n$, $|C_{\mathrm{link}}| = (q \pm \varepsilon)n$, $|V'_{\mathrm{link}}| = (\gamma/3 \pm \varepsilon)n$, and $|C'_{\mathrm{link}}| = (\gamma/3 \pm \varepsilon)n$,

(b) $|V_{\mathrm{abs}}| = (1 \pm \varepsilon)p_{\mathrm{abs}}n = (1 \pm \varepsilon)2pn/3$,

(c) for all $v \in V$, we have

  (i) $|N_{G_1}(v) \cap V_{\mathrm{abs}}| = (1 \pm \varepsilon)p_{\mathrm{abs}}\beta qn = (1 \pm \varepsilon)2p\beta qn/3$ and

  (ii) $|N_{G_1}(v) \cap V_{\mathrm{link}}| = (1 \pm \varepsilon)p\beta qn$,

(d) for all $c \in C$, we have

  (i) $|E^c_{G'}(V_{\mathrm{abs}}, V_{\mathrm{link}})| = (1 \pm \varepsilon)p_{\mathrm{abs}}p\beta n = (1 \pm \varepsilon)2p^2\beta n/3$ and

  (ii) $|E^c(G'[V_{\mathrm{link}}])| = (1 \pm \varepsilon)p^2\beta n/2$,

(e) for all distinct $u, v \in V$, we have $|N_{G_1}(u) \cap N_{G_1}(v) \cap V_{\mathrm{link}}| = (1 \pm \varepsilon)p\beta^2q^2n$, and

(f) for all $u, v \in V$ we have $|N_{G_2}(u) \cap N_{G_2}(v) \cap V'_{\mathrm{link}}| \geq \gamma^6 n$.

Indeed (a)–(d) follow from (3.6.1) and the Chernoff Bound. We prove (e) and (f) using McDiarmid's Inequality. To prove (e), for each $u, v \in V$, we apply McDiarmid's Inequality to the random variable $f$ counting $|N_{G_1}(u) \cap N_{G_1}(v) \cap V_{\mathrm{link}}|$ with respect to independent binomial random variables $\{X_w, X_{uw}, X_{vw}\}_{w\in V}$ and $\{X_c\}_{c\in C}$, where $X_w$ indicates if $w \in V_{\mathrm{link}}$, $X_{uw}$ and $X_{vw}$ indicate if the edges $uw$ and

$vw$ respectively are in $G'$, and $X_c$ indicates if $c \in C_{\text{link}}$. For each $w \in V$, $X_w$, $X_{uw}$, and $X_{vw}$ affect $f$ by at most one, and for each $c \in C$, $X_c$ affects $f$ by at most two. Thus, by McDiarmid's Inequality with $t = \varepsilon\mathbb{E}\left[f\right]/2$, we have $|N_{G_1}(u) \cap N_{G_1}(v) \cap V_{\text{link}}| = (1 \pm \varepsilon)p\beta^2q^2n$ with probability at least $1 - \exp\left(-(\varepsilon p\beta^2q^2n/2)^2/7n\right)$. By a union bound, (e) also holds with high probability. The proof of (f) is similar, so we omit it.

Now we assume (a)–(f) hold, we suppose $M$ is a matching such that $V(M) \subseteq V_{\text{abs}}$ and $|V_{\text{abs}} \setminus V(M)| \leq \varepsilon n$, and we show that (3.6.5.1) and (3.6.5.2) hold with respect to $M$. Since $|V_{\text{abs}} \setminus V(M)| \leq \varepsilon n$, (b) implies that

$$|V(M)| = (1 \pm \sqrt{\varepsilon})2pn/3. \tag{3.6.3}$$

We apply Theorem 3.5.3 to the following 8-uniform hypergraph $\mathcal{H}$: the vertex-set is $E(M) \cup V_{\text{link}} \cup C_{\text{link}}$, and for every $xy \in E(M)$, $v_1, v_2, v_3 \in V_{\text{link}}$, and $c_1, c_2, c_3, c_4 \in C_{\text{link}}$, $\mathcal{H}$ contains the hyperedge $\{xy, v_1, v_2, v_3, c_1, c_2, c_3, c_4\}$ if there is a main link $P$ such that

- $P$ has ends $x$ and $y$,
- $v_1$, $v_2$, and $v_3$ are the internal vertices in $P$, and
- $\phi(P) = \{c_1, c_2, c_3, c_4\}$.

*Claim 1: $d_{\mathcal{H}}(v) = (1 \pm 2\sqrt{\varepsilon})p^3\beta^4q^4n^3$ for all $v \in V(\mathcal{H})$.*

*Proof of claim:* Let $xy \in E(M)$. By (a), there are $(1 \pm \varepsilon)pn$ vertices $v_1 \in V_{\text{link}}$ that can be in a link $P$ with ends $x$ and $y$ corresponding to a hyperedge in $\mathcal{H}$, where $v_1$ is not adjacent to $x$ or $y$. For each such $v_1 \in V_{\text{link}}$, by (e), there are $(1 \pm \varepsilon)p\beta^2q^2n$ choices for the vertex in $V_{\text{link}}$ adjacent to both $x$ and $v_1$ in $P$, and for each such $v_2 \in V_{\text{link}}$, again by (e), there are $(1 \pm 2\varepsilon)p\beta^2q^2n$ choices for the vertex in $V_{\rm link}$ adjacent to both $v_1$ and $y$ in $P$ such that $P$ is a main link. Thus, $d_{\mathcal{H}}(xy) = (1 \pm 5\varepsilon)p^3\beta^4q^4n^3$, as required.

Now let $v_1 \in V_{\rm link}$. First, we count the number of hyperedges in $\mathcal{H}$ containing $v_1$ corresponding to a link $P$ where $v_1$ is adjacent to one of the ends. By (c), and since $|V_{\rm abs} \setminus V(M)| \leq \varepsilon n$, there are $(1 \pm \sqrt{\varepsilon})2p\beta qn/3$ choices of the vertex $x \in V(M)$ adjacent to $v_1$ in $P$. For each such $x$, again by (c), there are $(1 \pm 2\varepsilon)p\beta qn$ choices of the vertex $v_2 \in V_{\rm link}$ adjacent to $y$ in $P$ where $xy \in E(M)$. For each such $v_2 \in V_{\rm link}$, by (e), there are $(1 \pm 2\varepsilon)p\beta^2q^2n$ choices of the vertex $v_3 \in V_{\rm link}$ adjacent to both $v_1$ and $v_2$ in $P$. Thus, the number of hyperedges in $\mathcal{H}$ containing $v_1$ corresponding to a link where $v_1$ is adjacent to one of the ends is $(1 \pm 2\sqrt{\varepsilon})2p^3\beta^4q^4n^3/3$.

Next, we count the number of hyperedges in $\mathcal{H}$ containing $v_1$ corresponding to a link $P$ where $v_1$ is not adjacent to one of the ends. By (3.6.3), there are $(1 \pm \sqrt{\varepsilon})pn/3$ choices for the edge $xy \in E(M)$ where $x$ and $y$ are the ends of $P$. For each such $xy \in E(M)$, by (e), there are $(1 \pm \varepsilon)p\beta^2q^2n$ choices of the vertex $v_2 \in V_{\rm link}$ such that $v_2$ is adjacent to $x$ and $v_1$ in $P$, and again by (e), for each such $v_2 \in V_{\rm link}$, there are $(1 \pm 2\varepsilon)p\beta^2q^2n$ choices of the vertex $v_3 \in V_{\rm link}$ adjacent to both $y$ and $v_1$ in $P$. Thus, the number of hyperedges in $\mathcal{H}$ containing $v_1$ corresponding to a link where $v_1$ is not adjacent to one of the ends is $(1 \pm 2\sqrt{\varepsilon})p^3\beta^4q^4n^3/3$, so

$$d_{\mathcal{H}}(v_1) = (1 \pm 2\sqrt{\varepsilon})(2p^3\beta^4q^4n^3/3) + (1 \pm 2\sqrt{\varepsilon})p^3\beta^4q^4n^3/3 = (1 \pm 2\sqrt{\varepsilon})p^3\beta^4q^4n^3,$$

as required.

Now let $c_1 \in C_{\rm link}$. First we count the number of hyperedges in $\mathcal{H}$ containing $c_1$ corresponding to a link $P$ where $c_1$ is the colour of one of the edges incident to an end of $P$. By (d), and since $|V_{\rm abs} \setminus V(M)| \leq \varepsilon n$, there are $(1 \pm \sqrt{\varepsilon})2p^2\beta n/3$

choices of the edge $xv_1$ in $P$ where $x \in V(M)$ is an end of $P$ and $\phi(xv_1) = c_1$. For each such edge $xv_1$, by (c), there are $(1 \pm 2\varepsilon)p\beta qn$ choices of the vertex $v_2 \in V_{\text{link}}$ adjacent to $y$ in $P$ where $xy \in E(M)$. For each such vertex $v_2$, by (e), there are $(1 \pm 2\varepsilon)p\beta^2 q^2 n$ choices of the vertex $v_3$ adjacent to both $v_1$ and $v_2$ in $P$. Thus, the number of hyperedges in $\mathcal{H}$ containing $c_1$ corresponding to a link where $c_1$ is the colour of one of the edges incident to an end of $P$ is $(1 \pm 2\sqrt{\varepsilon})2p^4\beta^4 q^3 n^3/3$.

Next, we count the number of hyperedges in $\mathcal{H}$ containing $c_1$ corresponding to a link $P$ where $c_1$ is the colour of one of the edges with both ends in $V_{\text{link}}$. By (d), there are $(1 \pm \varepsilon)p^2\beta n/2$ choices for the edge $v_1v_2$ in $P$ such that $\phi(v_1v_2) = c_1$, and thus $(1 \pm \varepsilon)p^2\beta n$ choices for the edge if we assume $v_1$ is adjacent to an end in $P$. For each such edge $v_1v_2$, by (c), and since $|V_{\text{abs}} \setminus V(M)| \leq \varepsilon n$, there are $(1 \pm \sqrt{\varepsilon})2p\beta qn/3$ choices for the vertex $x \in V(M)$ adjacent to $v_1$ in $P$. For each such vertex $x$, by (e), there are $(1 \pm 2\varepsilon)p\beta^2 q^2 n$ choices for the vertex $v_3$ adjacent to both $y$ and $v_2$ in $P$, where $xy \in E(M)$. Thus, the number of hyperedges in $\mathcal{H}$ containing $c_1$ corresponding to a link where $c_1$ is the colour of one of the edges with both ends in $V_{\text{link}}$ is $(1 \pm 2\sqrt{\varepsilon})2p^4\beta^4 q^3 n^3/3$, so by (3.6.1)

$$d_{\mathcal{H}}(c_1) = (1 \pm 2\sqrt{\varepsilon})4p^4\beta^4 q^3 n^3/3 = (1 \pm 2\sqrt{\varepsilon})p^3\beta^4 q^4 n^3,$$

as required to prove Claim 1. –

*Claim 2:* $\Delta^c(\mathcal{H}) \leq 100n^2$.

This can be proved similarly as above (with room to spare). Let $\mathcal{F} := \{E(M), V_{\text{link}}, C_{\text{link}}\}$. By Theorem 3.5.3, $\mathcal{H}$ has a $(\delta, \mathcal{F})$-perfect matching $\mathcal{M}$. Let $\mathcal{P}_1$ be the collection of links corresponding to $\mathcal{M}$, and let $M'$ be the matching consisting of all those $xy \in E(M)$ that are not covered by $\mathcal{M}$. To complete the proof, we greedily find a collection $\mathcal{P}_2$ of reserve links that links $M'$.

Write $E(M') = \{x_1y_1, \dots, x_ky_k\}$, and suppose $P_i$ is a reserve link with ends $x_i$ and $y_i$ for $i < j$, where $j \in [k]$. We show that that there is a reserve link $P_j$ that is vertex- and colour-disjoint from $\bigcup_{i<j} P_i$, which implies that $\bigcup_{i=1}^{j} P_i$ links $\{x_1y_1, \dots, x_jy_j\}$, and thus we can choose $\mathcal{P}_2$ greedily. Since $k \leq \delta n$ and each link has at most three vertices in $V'_{\mathrm{link}}$, by (a), there is a vertex $v \in V'_{\mathrm{link}} \setminus \bigcup_{i<j} V(P_i)$. By (f), there are at least $\gamma^6 n - 11j$ vertices $v_1 \in (N_{G_2}(x_j) \cap N_{G_2}(v) \cap V'_{\mathrm{link}}) \setminus \bigcup_{i<j} V(P_i)$ such that $\phi(x_jv_1), \phi(v_1v) \notin \bigcup_{i<j} \phi(P_i)$, and since $j/n \leq \delta \ll \gamma$, we may let $v_1$ be such a vertex. Similarly, by (f), there is a vertex $v_2 \in (N_{G_2}(y_j) \cap N_{G_2}(v) \cap V'_{\mathrm{link}}) \setminus \bigcup_{i<j} V(P_i)$ such that $\phi(y_jv_2), \phi(v_2v) \notin \bigcup_{i<j} \phi(P_i) \cup \{\phi(x_jv_1), \phi(v_1v)\}$. Now there is a reserve link $P_j$ with ends $x_j$ and $y_j$ and internal vertices $v$, $v_1$, and $v_2$ that is vertex- and colour-disjoint from $\bigcup_{i<j} P_i$, as claimed, and therefore there exists a collection $\mathcal{P}_2$ of reserve links that links $M'$. Now $\mathcal{P}_1 \cup \mathcal{P}_2$ links $M$, so (3.6.5.1) holds. By (a), and since $\mathcal{M}$ is $(\delta, \mathcal{F})$-perfect, (3.6.5.2) holds, as required. $\square$

### 3.6.5 Proof

We now have all the tools we need to prove Lemma 3.4.9.

**Proof of Lemma 3.4.9.** Consider an absorber partition of $V$, $C$, and $K_n$. By Lemmas 3.6.3, 3.6.4, and 3.6.5, there exists an outcome of the absorber partition satisfying the conclusions of these lemmas simultaneously. In particular, by Lemmas 3.6.3 and 3.6.4 there exists $H, \mathcal{A}$, and $M$ such that, writing $(V_{\mathrm{res}}, C_{\mathrm{res}})$ for the bipartition of $H$,

- $H \cong \tilde{H}$ is a $36\gamma$-absorbing template satisfying (3.6.3.1) and (3.6.3.2),

- $\mathcal{A}$ and $H$ satisfy (3.6.4.1), and
- $M$ satisfies (3.6.4.2).

Write $\mathcal{A} = \{A_1, \dots, A_k\}$ and $E(M) = \{a_1b_1, \dots, a_\ell b_\ell\}$. Consider $M_1 \dot\cup M_2 \dot\cup M_3 \dot\cup M_4$, where $M_i$ is a matching satisfying (Mi) for $i \in [4]$ (see Section 3.6.1). By (3.6.4.2) we have $|V_{\text{abs}} \setminus V(M_1 \cup \dots \cup M_4)| \leq 5\varepsilon n + 1 \leq 6\varepsilon n$. Thus by Lemma 3.6.5 there exist collections of links $\mathcal{P}$ and $\mathcal{T}$ in $G'$ such that

- $\mathcal{P} \cup \mathcal{T}$ is a collection of links satisfying (3.6.5.1) with respect to $\bigcup_{i=1}^4 M_i$ and
- $\mathcal{P} \cup \mathcal{T}$ satisfies (3.6.5.2).

In particular $\mathcal{P}$ links $\bigcup_{i=1}^3 M_i$ and $\mathcal{T}$ links $M_4$. Let $T \coloneqq M \cup \bigcup_{P \in \mathcal{T}} P$. By Fact 3.6.2, $(\mathcal{A}, \mathcal{P}, T, H)$ is a $36\gamma$-absorber, as desired. Moreover, since $H$ satisfies (3.6.3.1) and (3.6.3.2), $M$ satisfies (3.6.4.2), and $\mathcal{P} \cup \mathcal{T}$ satisfies (3.6.5.2), we have $\bigcup_{A \in \mathcal{A}} A \cup \bigcup_{P \in \mathcal{P}} P \cup T$ is contained in $(V', C', G')$ with $\gamma$-bounded remainder, as required.

$\square$

## 3.7 Finding many well-spread absorbing gadgets

The aim of this section is to prove Lemma 3.4.8, which states that, for appropriate $\mu, \varepsilon$, almost all 1-factorizations of $K_n$ are $\varepsilon$-locally edge-resilient and $\mu$-robustly gadget-resilient. We will use switchings in $\mathcal{G}_D^{\text{col}}$ for appropriate $D \subsetneq [n-1]$ to analyse the probability that a uniformly random $G \in \mathcal{G}_D^{\text{col}}$ satisfies the necessary properties, and then use a 'weighting factor' (see Corollary 3.7.12) to make comparisons to the probability space corresponding to a uniform random choice of $G \in \mathcal{G}_{[n-1]}^{\text{col}}$.

### 3.7.1 Switchings

We begin by analysing the property of $\varepsilon$-local edge-resilience.

**Lemma 3.7.1.** *Suppose $1/n \ll \varepsilon \ll 1$, and let $D \subseteq [n-1]$ have size $|D| = \varepsilon n$. Suppose $\mathbf{G} \in \mathcal{G}_D^{col}$ is chosen uniformly at random. Then $\mathbb{P}[\mathbf{G}$ is $\varepsilon$-locally edge-resilient$] \geq 1 - \exp(-\varepsilon^3 n^2/1000)$.*

**Proof.** Note that if $G \in \mathcal{G}_D^{\mathrm{col}}$ has at least $\varepsilon^3 n^2/100$ edges with endpoints in $V'$ for all choices of $V' \subseteq V$ of size precisely $\varepsilon n$, then $G$ is $\varepsilon$-locally edge-resilient. Fix $V' \subseteq V$ of size precisely $\varepsilon n$. For any $G \in \mathcal{G}_D^{\mathrm{col}}$, we say that a subgraph $H \subseteq G$ together with a labelling of its vertices $V(H) = \{u, v, w, x, y, z\}$ is a *spin system* of $G$ if $E(H) = \{vw, xy, zu\}$, where $u, v \in V'$, $w, x, y, z \in V \setminus V'$, $uv, wx, yz \notin E(G)$, and $\phi_G(vw) = \phi_G(xy) = \phi_G(zu)$. (Note that different labellings of a subgraph $H \subseteq G$ that both satisfy these conditions will be considered to correspond to different spin systems of $G$.) We now define the *spin* switching operation. Suppose $G \in \mathcal{G}_D^{\mathrm{col}}$ and $H \subseteq G$ is a spin system. Then we define $\mathrm{spin}_H(G)$ to be the coloured graph obtained from $G$ by deleting the edges $vw, xy, zu$, and adding the edges $uv, wx, yz$, each with colour $\phi_G(vw)$. Writing $G' := \mathrm{spin}_H(G)$, we have $G' \in \mathcal{G}_D^{\mathrm{col}}$ and $e_{V',D}(G') = e_{V',D}(G) + 1$.

We define a partition $\{M_s\}_{s=0}^{\binom{\varepsilon n}{2}}$ of $\mathcal{G}_D^{\mathrm{col}}$ by setting $M_s := \{G \in \mathcal{G}_D^{\mathrm{col}} \colon e_{V',D}(G) = s\}$, for each $s \in [\binom{\varepsilon n}{2}]_0$. For each $s \in [\binom{\varepsilon n}{2} - 1]_0$ we define an auxiliary bipartite multigraph $B_s$ with vertex bipartition $(M_s, M_{s+1})$, where for each $G \in M_s$ and each spin system $H \subseteq G$ we put an edge in $B_s$ with endpoints $G \in M_s$ and $\mathrm{spin}_H(G) \in M_{s+1}$. Define $\delta_s := \min_{G \in M_s} d_{B_s}(G)$ and $\Delta_{s+1} := \max_{G \in M_{s+1}} d_{B_s}(G)$. Observe, by double counting $e(B_s)$, that $|M_s|/|M_{s+1}| \leq \Delta_{s+1}/\delta_s$. To bound $\Delta_{s+1}$ from above, we fix $G' \in M_{s+1}$ and bound the number of pairs $(G, H)$, where

$G \in M_s$ and $H$ is a spin system of $G$ such that $\mathrm{spin}_H(G) = G'$. There are $s+1$ choices for the edge $e \in E_{V',D}(G')$ created by a spin operation, and 2 choices for which endpoint of $e$ played the role of $u$ in a spin, and which played the role of $v$. Now there are at most $(n/2)^2$ choices for two edges with colour $\phi_{G'}(e)$ in $G'$ with both endpoints outside of $V'$, and at most 8 choices for which endpoints of these edges played the roles of $w, x, y, z$ in a spin operation yielding $G'$. We deduce that $\Delta_{s+1} \leq 4(s+1)n^2$.

Suppose that $s \leq \varepsilon^3 n^2/80$. To bound $\delta_s$ from below, we fix $G \in M_s$ and find a lower bound for the number of spin systems $H \subseteq G$. For a vertex $v \in V'$, let $D^*_G(v) \subseteq D$ denote the set of colours $c \in D$ such that the $c$-neighbour of $v$ is not in $V'$, in $G$. Let $V^*_G := \{v \in V' \colon |D^*_G(v)| \geq 9\varepsilon n/10\}$, and suppose for a contradiction that $|V^*_G| < 9\varepsilon n/10$. Then there are at least $\varepsilon n/10$ vertices $v \in V'$ for which there are at least $\varepsilon n/10$ colours $c \in D$ such that the $c$-neighbour of $v$ is in $V'$, in $G$, whence $s = e_{V',D}(G) \geq \varepsilon^2 n^2/200 > \varepsilon^3 n^2/80 \geq s$, a contradiction. Note further that, since $s \leq \varepsilon^3 n^2/80$, there are at least $\binom{9\varepsilon n/10}{2} - \varepsilon^3 n^2/80 \geq \varepsilon^2 n^2/4$ pairs $\{a, b\} \in \binom{V^*_G}{2}$ such that $ab \notin E(G)$. For each such choice of $\{a, b\}$, there are two choices of which vertex will play the role of $u$ and which will play the role of $v$ in a spin system. Since $u, v \in V^*_G$, there are at least $4\varepsilon n/5$ colours $c \in D$ such that the $c$-neighbour $z$ of $u$, and the $c$-neighbour $w$ of $v$, are such that $w, z \in V \setminus V'$, in $G$. Finally, there are at least $n/2 - 3\varepsilon n \geq n/4$ edges coloured $c$ in $G$ with neither endpoint in $V' \cup N_G(w) \cup N_G(z)$, and two choices of which endpoint of such an edge will play the role of $x$, and which will play the role of $y$. We deduce that $\delta_s \geq \varepsilon^3 n^4/5$. Altogether, we conclude that if $s \leq \varepsilon^3 n^2/80$ and $M_s$ is non-empty, then $M_{s+1}$ is non-empty and $|M_s|/|M_{s+1}| \leq 20(s+1)n^2/\varepsilon^3 n^4 \leq 1/2$.

Now, fix $s \leq \varepsilon^3 n^2/100$. If $M_s$ is empty, then $\mathbb{P}\left[e_{V',D}(\mathbf{G}) = s\right] = 0$. If $M_s$ is non-empty, then

$$\mathbb{P}\left[e_{V',D}(\mathbf{G})=s\right]=\frac{|M_s|}{|\mathcal{G}_D^{\mathrm{col}}|}\leq\frac{|M_s|}{|M_{\varepsilon^3n^2/80}|}=\prod_{j=s}^{\varepsilon^3n^2/80-1}\frac{|M_j|}{|M_{j+1}|}\leq\left(\frac{1}{2}\right)^{\varepsilon^3n^2/80-s},$$

and thus

$$\mathbb{P}\left[e_{V',D}(\mathbf{G})\leq\varepsilon^3n^2/100\right]\leq\sum_{s=0}^{\varepsilon^3n^2/100}\exp(-(\varepsilon^3n^2/80-s)\ln 2)\leq\exp\left(-\frac{\varepsilon^3n^2}{800}\right).$$

A union bound over all choices of $V'\subseteq V$ of size $\varepsilon n$ now completes the proof. $\square$

We now turn to showing that for suitable $D\subseteq[n-1]$, almost all $G\in\mathcal{G}_D^{\mathrm{col}}$ are robustly gadget-resilient, which turns out to be a much harder property to analyse than local edge-resilience, and we devote the rest of this section to it. We first need to show that almost all $G\in\mathcal{G}_D^{\mathrm{col}}$ are 'quasirandom', in the sense that small sets of vertices do not have too many crossing edges.

**Definition 3.7.2.** Let $D\subseteq[n-1]$. We say that $G\in\mathcal{G}_D^{\mathrm{col}}$ is *quasirandom* if for all sets $A,B\subseteq V$, not necessarily distinct, such that $|A|=|B|=|D|$, we have that $e_G(A,B)<8(|D|-1)^3/n$. We define $\mathcal{Q}_D^{\mathrm{col}}:=\{G\in\mathcal{G}_D^{\mathrm{col}}\colon G\,\mathrm{is\,quasirandom}\}$.

When we are analysing switchings to study the property of robust gadget-resilience (see Lemma 3.7.8), it will be important to condition on this quasirandomness. One can use another switching argument to show that almost all $G\in\mathcal{G}_D^{\mathrm{col}}$ are quasirandom.

**Lemma 3.7.3.** *Suppose that $1/n\ll\mu\ll1$, let $D\subseteq[n-1]$ have size $|D|=\mu n+1$. Suppose that $\mathbf{G}\in\mathcal{G}_D^{col}$ is chosen uniformly at random. Then $\mathbb{P}\left[\mathbf{G}\in\mathcal{Q}_D^{col}\right]\geq 1-\exp(-\mu^3n^2)$.*

**Proof.** Fix $A, B \subseteq V$ satisfying $|A| = |B| = \mu n + 1$. For any $G \in \mathcal{G}_D^{\mathrm{col}}$, we say that a subgraph $H \subseteq G$ together with a labelling of its vertices $V(H) = \{a, b, v, w\}$ is a *rotation system* of $G$ if $E(H) = \{ab, vw\}$, where $a \in A$, $b \in B$, $v, w \notin A \cup B$, $aw, bv \notin E(G)$, and $\phi_G(ab) = \phi_G(vw)$. We now define the *rotate* switching operation. Suppose $G \in \mathcal{G}_D^{\mathrm{col}}$ and $H \subseteq G$ is a rotation system. Then we define $\mathrm{rot}_H(G)$ to be the coloured graph obtained from $G$ by deleting the edges $ab, vw$, and adding the edges $aw, bv$, each with colour $\phi_G(ab)$. Writing $G' \coloneqq \mathrm{rot}_H(G)$, notice that $G' \in \mathcal{G}_D^{\mathrm{col}}$ and $e_{G'}(A, B) = e_G(A, B) - 1$.

Lemma 3.7.3 follows by analysing the degrees of auxiliary bipartite multigraphs $B_s$ in a similar way as in the proof of Lemma 3.7.1. We omit the details.

□

Next we will use a switching argument to find a large set of well-spread absorbing gadgets (cf. Definition 3.4.7). For this, we consider slightly more restrictive substructures than the absorbing gadgets defined in Definition 3.4.1. These additional restrictions (an extra edge $f$ as well as an underlying partition $\mathcal{P}$ of the colours) give us better control over the switching process: they allow us to argue that we do not create more than one additional gadget per switch. Let $D \subseteq [n-1]$, $c \in [n-1] \setminus D$, write $D^* \coloneqq D \cup \{c\}$, and let $G \in \mathcal{G}_{D^*}^{\mathrm{col}}$. Suppose that $\mathcal{P} = \{D_i\}_{i=1}^4$ is an (ordered) partition of $D$ into four subsets, and let $x \in V$.

**Definition 3.7.4.** An $(x, c, \mathcal{P})$*-gadget* in $G$ is a subgraph $J = A \cup \{f\}$ of $G$ the following form (see Figure 3.4):

(i) $A$ is an $(x, c)$-absorbing gadget in $G$;

(ii) there is an edge $e_1 \in \partial_A(x)$ such that $\phi(e_1) \in D_1$, and the remaining edge $e_2 \in \partial_A(x)$ satisfies $\phi(e_2) \in D_2$;

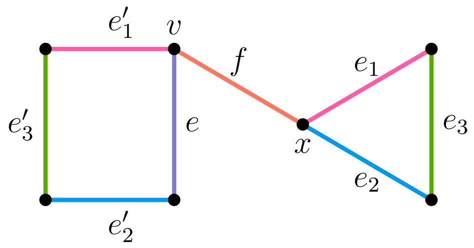



Figure 3.4: An $(x, c, \mathcal{P})$-gadget. Here, $\phi(f) \in D_4$, $\phi(e) = c$, and $\phi(e_i) = \phi(e_i') \in D_i$ for each $i \in [3]$.

(iii) the edge $e_3$ of $A$ which is not incident to $x$ but shares an endvertex with $e_1$ and an endvertex with $e_2$ satisfies $\phi(e_3) \in D_3$;

(iv) $f = xv$ is an edge of $G$, where $v$ is the unique vertex of $A$ such that $\phi(\partial_A(v)) = \{c, \phi(e_1)\}$;

(v) $\phi(f) \in D_4$.

We now define some terminology that will be useful for analysing how many $(x, c, \mathcal{P})$-gadgets there are in a graph $G \in \mathcal{G}_{D^*}^{\text{col}}$, and how well-spread these gadgets are. Each of the terms we define here will have a dependence on the choice of the triple $(x, c, \mathcal{P})$, but since this triple will always be clear from context, for presentation we omit the $(x, c, \mathcal{P})$-notation.

**Definition 3.7.5.** We say that an $(x, c, \mathcal{P})$-gadget $J$ in $G$ is *distinguishable* in $G$ if the edges $e_3, e_3'$ of $J$ such that $\phi(e_3) = \phi(e_3') \in D_3$ are such that there is no other $(x, c, \mathcal{P})$-gadget $J' \neq J$ in $G$ such that $e_3 \in E(J')$ or $e_3' \in E(J')$.

We will aim only to count distinguishable $(x, c, \mathcal{P})$-gadgets, which will ensure the collection of gadgets we find is well-spread across the set of edges in $G \in \mathcal{G}_{D^*}^{\text{col}}$ that can play the roles of $e_3$, $e_3'$. We also need to ensure that the collection of gadgets we find is well-spread across the $c$-edges of $G$.

**Definition 3.7.6.**

- For each $c$-edge $e$ of $G \in \mathcal{G}^{\mathrm{col}}_{D^*}$, we define the *saturation* of $e$ in $G$, denoted $\mathrm{sat}_G(e)$, or simply $\mathrm{sat}(e)$ when $G$ is clear from context, to be the number of distinguishable $(x, c, \mathcal{P})$-gadgets of $G$ which contain $e$. We say that $e$ is *unsaturated* in $G$ if $\mathrm{sat}(e) \leq |D| - 1$, *saturated* if $\mathrm{sat}(e) \geq |D|$, and *supersaturated* if $\mathrm{sat}(e) \geq |D| + 6$. We define $\mathrm{Sat}(G)$ to be the set of saturated $c$-edges of $G$, and $\mathrm{Unsat}(G) \coloneqq E_c(G) \setminus \mathrm{Sat}(G)$.
- We define the function $r\colon \mathcal{G}^{\mathrm{col}}_{D^*} \to [n|D|/2]_0$ by

$$r(G) \coloneqq |D||\mathrm{Sat}(G)| + \sum_{e \in \mathrm{Unsat}(G)} \mathrm{sat}(e).$$

In Lemma 3.7.8, we will use switchings to show that $r(G)$ is large (for some well-chosen $\mathcal{P}$) in almost all quasirandom $G \in \mathcal{G}^{\mathrm{col}}_{D^*}$. In Lemma 3.7.9, we use distinguishability, saturation, and the fact that any non-$x$ vertex in an $(x, c, \mathcal{P})$-gadget must be incident to an edge playing the role of either $e_3$, $e_3'$, or the $c$-edge, to show that $r(G)$ being large means that there are many well-distributed $(x, c, \mathcal{P})$-gadgets in $G$, and thus many well-spread $(x, c)$-absorbing gadgets. We now define a relaxation of $\mathcal{Q}^{\mathrm{col}}_{D^*}$, which will be a convenient formulation for ensuring that quasirandomness is maintained when we use switchings to find $(x, c, \mathcal{P})$-gadgets. For each $s \in [n|D|/2]_0$, we write $A_s^{D^*} \coloneqq \{G \in \mathcal{G}^{\mathrm{col}}_{D^*} \colon r(G) = s\}$, and we write $Q_s^{D^*}$ for the set of $G \in \mathcal{G}^{\mathrm{col}}_{D^*}$ such that $e_G(A, B) < 8|D|^3/n + 6s$ for all $A, B \subseteq V$ such that $|A| = |B| = |D|$. We also define $T_s^{D^*} \coloneqq A_s^{D^*} \cap Q_s^{D^*}$ and $\widetilde{\mathcal{Q}}^{\mathrm{col}}_{D^*} \coloneqq \bigcup_{s=0}^{n|D|/2} T_s^{D^*}$. Notice that

$$\mathcal{Q}^{\mathrm{col}}_{D^*} \subseteq \widetilde{\mathcal{Q}}^{\mathrm{col}}_{D^*}. \tag{3.7.1}$$

Finally, we discuss the switching operation that we will use in Lemma 3.7.8.

**Definition 3.7.7.** For any $G \in \mathcal{G}^{\mathrm{col}}_{D^*}$, we say that a subgraph $H \subseteq G$ together with a labelling of its vertices $V(H) = \{x, u_1, u_2, \ldots, u_{14}\}$ is a *twist system* of $G$ if (see Figure 3.5):

(i) $E(H) = \{u_1u_2, u_3u_5, u_4u_6, u_5u_7, u_6u_8, u_7u_8, u_7x, xu_9, xu_{10}, u_9u_{11}, u_{10}u_{12}, u_{13}u_{14}\}$;

(ii) $\phi(u_5u_7) = \phi(xu_9) \in D_1$;

(iii) $\phi(u_6u_8) = \phi(xu_{10}) \in D_2$;

(iv) $\phi(u_1u_2) = \phi(u_3u_5) = \phi(u_4u_6) = \phi(u_9u_{11}) = \phi(u_{10}u_{12}) = \phi(u_{13}u_{14}) \in D_3$;

(v) $\phi(u_7x) \in D_4$;

(vi) $\phi(u_7u_8) = c$;

(vii) $u_1u_3, u_2u_4, u_5u_6, u_9u_{10}, u_{11}u_{13}, u_{12}u_{14} \notin E(G)$.

For a twist system $H$ of $G$, we define $\mathrm{twist}_H(G)$ to be the coloured graph obtained from $G$ by deleting the edges $u_1u_2$, $u_3u_5$, $u_4u_6$, $u_9u_{11}$, $u_{10}u_{12}$, $u_{13}u_{14}$, and adding the edges $u_1u_3$, $u_2u_4$, $u_5u_6$, $u_9u_{10}$, $u_{11}u_{13}$, $u_{12}u_{14}$, each with colour $\phi_G(u_1u_2)$. The $(x, c, \mathcal{P})$-gadget in $\mathrm{twist}_H(G)$ with edges $u_5u_6$, $u_5u_7$, $u_6u_8$, $u_7u_8$, $u_7x$, $xu_9$, $xu_{10}$, $u_9u_{10}$ is called the *canonical* $(x, c, \mathcal{P})$*-gadget* of the twist.

We simultaneously switch two edges into the positions $u_5u_6$ and $u_9u_{10}$ because it is much easier to find structures as in Figure 3.5 than it is to find such a structure with one of these edges already in place. Moreover, the two 'switching cycles' we use have three edges and three non-edges (rather than two of each, as in the rotation switching) essentially because of the extra freedom this gives us when choosing the edges $u_1u_2$ and $u_{13}u_{14}$. This extra freedom allows us to ensure that in almost all twist systems, one avoids undesirable issues like inadvertently creating

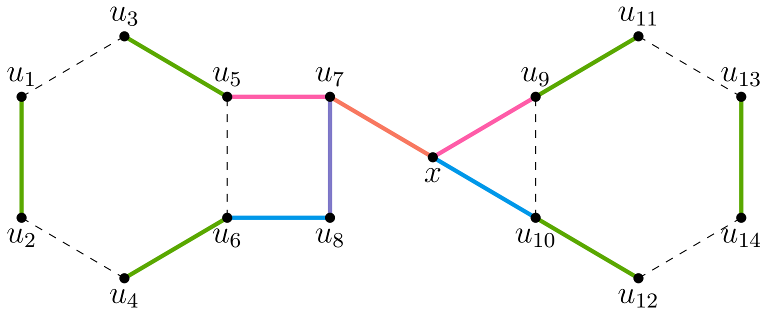



Figure 3.5: A twist system of $G$. Here, dashed edges represent non-edges of $G$, and the colours of the edges satisfy (ii)–(vi) in the definition of twist system.

more than one new gadget when one performs the twist.

The proof of Lemma 3.7.8 proceeds with a similar strategy to those of Lemmas 3.7.1 and 3.7.3, but it is much more challenging this time to show that graphs with low $r(G)$-value admit many ways to switch to yield a graph $G' \in \mathcal{G}^{\mathrm{col}}_{D^*}$ satisfying $r(G') = r(G) + 1$.

**Lemma 3.7.8.** *Suppose that $1/n \ll \mu \ll 1$, and let $D \subseteq [n-1]$ have size $|D| = \mu n$. Let $x \in V$, let $c \in [n-1] \setminus D$, and let $\mathcal{P} = \{D_i\}_{i=1}^4$ be an equitable partition of $D$. Suppose that $\mathbf{G} \in \mathcal{G}^{col}_{D \cup \{c\}}$ is chosen uniformly at random. Then*

$$\mathbb{P}\left[r(\mathbf{G}) \leq \frac{\mu^4 n^2}{2^{23}} \,\middle|\, \mathbf{G} \in \widetilde{\mathcal{Q}}^{col}_{D \cup \{c\}}\right] \leq \exp\left(-\frac{\mu^4 n^2}{2^{24}}\right).$$

**Proof.** Write $D^* := D \cup \{c\}$. Consider the partition $\{T_s^{D^*}\}_{s=0}^{nk/2}$ of $\widetilde{\mathcal{Q}}^{\mathrm{col}}_{D^*}$, where $k := |D|$. For each $s \in [nk/2 - 1]_0$, we define an auxiliary bipartite multigraph $B_s$ with vertex bipartition $(T_s^{D^*}, T_{s+1}^{D^*})$ and an edge between $G$ and $\mathrm{twist}_H(G)$ whenever:

(a) $G \in T_s^{D^*}$;

(b) $H$ is a twist system in $G$ for which $G' := \mathrm{twist}_H(G) \in T_{s+1}^{D^*}$ and $G'$ satisfies $\mathrm{sat}_{G'}(e) = \mathrm{sat}_G(e) + 1 \leq k$ for the $c$-edge $e = u_7u_8$ of $H$, with the canonical $(x, c, \mathcal{P})$-gadget of the twist $G'$ being the only additional distinguishable

$(x, c, \mathcal{P})$-gadget using this $c$-edge.

Define $\delta_s \coloneqq \min_{G \in T_s^{D^*}} d_{B_s}(G)$ and $\Delta_{s+1} \coloneqq \max_{G \in T_{s+1}^{D^*}} d_{B_s}(G)$. Thus $|T_s^{D^*}|/|T_{s+1}^{D^*}| \leq \Delta_{s+1}/\delta_s$. To bound $\Delta_{s+1}$ from above, we fix $G' \in T_{s+1}^{D^*}$ and bound the number of pairs $(G, H)$, where $G \in T_s^{D^*}$ and $H$ is a twist system of $G$ such that $\mathrm{twist}_H(G) = G'$ and (b) holds. Firstly, note that

$$\sum_{\substack{e \in E_c(G') \\ \mathrm{sat}_{G'}(e) \leq k}} \mathrm{sat}_{G'}(e) \leq r(G') = s + 1.$$

Thus, it follows from condition (b) that there are at most $s + 1$ choices for the canonical $(x, c, \mathcal{P})$-gadget of a twist yielding $G'$ for which we record an edge in $B_s$. Fixing this $(x, c, \mathcal{P})$-gadget fixes the vertices of $V$ which played the roles of $x$, $u_5$, $u_6$, $\dots$, $u_{10}$ in a twist yielding $G'$. To determine all possible sets of vertices playing the roles of $u_1$, $u_2$, $u_3$, $u_4$, $u_{11}$, $u_{12}$, $u_{13}$, $u_{14}$ (thus determining $H$ and $G$ such that $\mathrm{twist}_H(G) = G'$), it suffices to find all choices of four edges of $G'$ with colour $\phi_{G'}(u_5 u_6)$ satisfying the necessary non-adjacency conditions. There are at most $(n/2)^4$ choices for these four edges, and at most $4! \cdot 2^4$ choices for which endpoints of these edges play which role. We deduce that $\Delta_{s+1} \leq 24n^4(s + 1)$.

Suppose that $s \leq k^4/2^{22}n^2$. To bound $\delta_s$ from below, we fix $G \in T_s^{D^*}$ and find a lower bound for the number of twist systems $H \subseteq G$ for which we record an edge between $G$ and $\mathrm{twist}_H(G)$ in $B_s$. To do this, we will show that there are many choices for a set of four colours and two edges, such that each of these sets uniquely identifies a twist system in $G$ for which we record an edge in $B_s$. Note that since $s \leq k^4/2^{22}n^2$ and $G \in Q_s^{D^*}$, we have

$$e_G(A, B) \leq 10k^3/n \quad \text{for all sets } A, B \subseteq V \text{ of sizes } |A| = |B| = k. \tag{3.7.2}$$

We begin by finding subsets of $D_3$ and $D_4$ with some useful properties in $G$.

*Claim 1: There is a set $D_3^{good} \subseteq D_3$ of size $|D_3^{good}| \geq k/8$ such that for all $d \in D_3^{good}$ we have*

(i) *$|E_d(N_{D_1}(x), N_{D_2}(x))| \leq 200k^2/n$;*

(ii) *there are at most $64k^3/n^2$ $d$-edges $e$ in $G$ with the property that $e$ lies in some distinguishable $(x, c, \mathcal{P})$-gadget in $G$ whose $c$-edge is not supersaturated.*

*Proof of claim:* Observe that $|N_{D_1}(x)| = |N_{D_2}(x)| = k/4$. Then, by (arbitrarily extending $N_{D_1}(x)$, $N_{D_2}(x)$ and) applying (3.7.2), we see that $e(N_{D_1}(x), N_{D_2}(x)) \leq 10k^3/n$. Thus there is a set $\hat{D}_3 \subseteq D_3$ of size $|\hat{D}_3| \geq 3k/16$ such that each $d \in \hat{D}_3$ satisfies (i). Next, notice that, since $r(G) = s$, there are at most $s/k \leq k^3/2^{22}n^2$ saturated $c$-edges in $G$. Suppose for a contradiction that at least $k/16$ colours $d \in D_3$ are such that there are at least $64k^3/n^2$ $d$-edges $e$ in $G$ with the property that $e$ lies in some distinguishable $(x, c, \mathcal{P})$-gadget in $G$ whose $c$-edge is not supersaturated. Then, by considering the contribution of these distinguishable $(x, c, \mathcal{P})$-gadgets to $r(G)$, and accounting for saturated $c$-edges, we obtain that $r(G) \geq (k/16) \cdot 32k^3/n^2 - 5k^3/2^{22}n^2 > s$, a contradiction. Thus there is a set $\tilde{D}_3 \subseteq D_3$ of size $|\tilde{D}_3| \geq 3k/16$ such that each $d \in \tilde{D}_3$ satisfies (ii). We define $D_3^{\text{good}} := \hat{D}_3 \cap \tilde{D}_3$, and note that $|D_3^{\text{good}}| \geq k/8$. $-$

We also define $D_4^{\text{good}} \subseteq D_4$ to be the set of colours $d_4 \in D_4$ such that the $c$-edge $e$ incident to the $d_4$-neighbour of $x$ in $G$ satisfies $\text{sat}(e) \leq k - 1$. Observe that $|D_4^{\text{good}}| \geq k/8$, since otherwise there are at least $k/16$ saturated $c$-edges in $G$, whence $r(G) \geq k^2/16 > s$, a contradiction.

We now show that there are many choices of a vector $(d_1, d_2, d_3, d_4, \overrightarrow{f_1}, \overrightarrow{f_2})$ where each $d_i \in D_i$ and each $\overrightarrow{f_j}$ is an edge $f_j \in E_{d_3}(G)$ together with an identification of which endpoints will play which role, such that each vector uniquely gives rise to a candidate of a twist system $H \subseteq G$. We can begin to construct such a candidate by choosing $d_4 \in D_4^{\text{good}}$ and letting $u_7$ denote the $d_4$-neighbour of $x$ in $G$, and letting $u_8$ denote the $c$-neighbour of $u_7$. Secondly, we choose $d_1 \in D_1$, avoiding the colour of the edge $xu_8$ (if it is present), and let $u_5$ denote the $d_1$-neighbour of $u_7$, and let $u_9$ denote the $d_1$-neighbour of $x$. Next, we choose $d_2 \in D_2$, avoiding the colours of the edges $u_5u_8$, $u_5x$, $u_8x$, $u_8u_9$ in $G$ (if they are present), and let $u_6$ denote the $d_2$-neighbour of $u_8$, and let $u_{10}$ denote the $d_2$-neighbour of $x$. Then, we choose $d_3 \in D_3^{\text{good}}$, avoiding the colours of all edges in $E_G(\{x, u_5, u_6, \ldots, u_{10}\})$. We let $u_3, u_4, u_{11}, u_{12}$ denote the $d_3$-neighbours of $u_5, u_6, u_9, u_{10}$, respectively. Finally, we choose two distinct edges $f_1, f_2 \in E_{d_3}(G)$ which are not incident to any vertex in $\{x, u_3, u_4, \ldots, u_{12}\}$, and we choose which endpoint of $f_1$ will play the role of $u_1$ and which will play the role of $u_2$, and choose which endpoint of $f_2$ will play the role of $u_{13}$ and which will play the role of $u_{14}$. Let $\Lambda$ denote the set of all possible vectors $(d_1, d_2, d_3, d_4, \overrightarrow{f_1}, \overrightarrow{f_2})$ that can be chosen in this way, so that $|\Lambda| \geq \frac{k}{8} \cdot \frac{3k}{16} \cdot \frac{k}{8} \cdot \frac{k}{16} \cdot \frac{n}{4} \cdot 2 \cdot \frac{n}{4} \cdot 2 = 3k^4n^2/2^{16}$. Further, let $H(\lambda) \subseteq G$ denote the labelled subgraph of $G$ corresponding to $\lambda \in \Lambda$ in the above way. If $H(\lambda)$ is a twist system, then we sometimes say that we 'twist on $\lambda$' to mean that we perform the twist operation to obtain $\text{twist}_{H(\lambda)}(G)$ from $G$.

It is clear that $H(\lambda)$ is unique for all vectors $\lambda \in \Lambda$, and that $H(\lambda)$ satisfies conditions (i)–(vi) of the definition of a twist system. However, some $H(\lambda)$ may fail to satisfy (vii), and some may fail to satisfy condition (b) in the definition of adjacency in $B_s$. We now show that only for a small proportion of $\lambda \in \Lambda$ do either of these problems occur. We begin by ensuring that most $\lambda \in \Lambda$ give rise to twist systems.

*Claim 2: There is a subset $\Lambda_1 \subseteq \Lambda$ such that $|\Lambda_1| \geq 9|\Lambda|/10$ and $H(\lambda)$ is a twist system for all $\lambda \in \Lambda_1$.*

*Proof of claim:* Fix any choice of $d_3 \in D_3^{\text{good}}$, $d_4 \in D_4^{\text{good}}$ and $\overrightarrow{f_1}, \overrightarrow{f_2}$ appearing concurrently in some $\lambda \in \Lambda$, and note that there are at most $(k/4)^2 \cdot n^2$ such choices. Here and throughout the remainder of the proof of Lemma 3.7.8, we write $u_7$ for the $d_4$-neighbour of $x$, we write $u_8$ for the $c$-neighbour of $u_7$, and so on, where the choice of $d_1$, $d_2$, $d_3$, $d_4$, $\overrightarrow{f_1}$, $\overrightarrow{f_2}$ will always be clear from context. Note that fixing $d_3$, $d_4$ only fixes the vertices $x$, $u_7$, $u_8$. There are at most $10k^3/n$ pairs $(d_1, d_2)$ with each $d_i \in D_i$ such that there is an edge $u_5u_6 \in E(G)$, since otherwise $e(N_{D_1}(u_7), N_{D_2}(u_8)) > 10k^3/n$, contradicting (3.7.2). Similarly, there are at most $10k^3/n$ pairs $(d_1, d_2)$ with each $d_i \in D_i$ such that $u_9u_{10}$ is an edge of $G$. We deduce that there are at most $(20k^3/n) \cdot (k/4)^2 \cdot n^2 = 5k^5n/4$ vectors $\lambda \in \Lambda$ for which $H(\lambda)$ is such that either $u_5u_6$ or $u_9u_{10}$ is an edge of $G$. Now fix instead $d_1$, $d_2$, $d_3$, $d_4$, $\overrightarrow{f_2}$. Note that $|N_G(u_3) \cup N_G(u_4)| \leq 2k + 2$ so that there are at most $4k + 4$ choices of $\overrightarrow{f_1}$ such that either $u_1u_3$ or $u_2u_4$ is an edge of $G$. Analysing the pairs $u_{11}u_{13}$ and $u_{12}u_{14}$ similarly, we deduce that altogether, there are at most $5k^5n/4 + 2((k/4)^4 \cdot n \cdot (4k + 4)) \leq 2k^5n \leq |\Lambda|/10$ vectors $\lambda \in \Lambda$ for which $H(\lambda)$ fails to be a twist system. $-$

We now show that only for a small proportion of $\lambda \in \Lambda_1$ does $H(\lambda)$ fail to give rise to an edge in $B_s$, by showing that most $H(\lambda)$ satisfy the following properties:

(P1) $\text{twist}_{H(\lambda)}(G) \in Q_{s+1}^{D^*}$;

(P2) Deletion of the six $d_3$-edges in $H(\lambda)$ does not decrease $r(G)$;

(P3) The canonical $(x, c, \mathcal{P})$-gadget of the twist $\text{twist}_{H(\lambda)}(G)$ is distinguishable, and it is the only $(x, c, \mathcal{P})$-gadget which is in $\text{twist}_{H(\lambda)}(G)$ but not in $G$.

Firstly, since $G \in Q_s^{D^*}$ and we only create six new edges in any twist, it is clear that $H(\lambda)$ satisfies (P1) for all $\lambda \in \Lambda_1$.

*Claim* 3: *There is a subset* $\Lambda_2 \subseteq \Lambda_1$ *such that* $|\Lambda_2| \geq 9|\Lambda_1|/10$ *and* $H(\lambda)$ *satisfies property (P2) for all* $\lambda \in \Lambda_2$.

*Proof of claim:* Fix $d_1 \in D_1$, $d_3 \in D_3^{\mathrm{good}}$, $d_4 \in D_4^{\mathrm{good}}$, $\overrightarrow{f_1}$, $\overrightarrow{f_2}$ appearing concurrently in some $\lambda \in \Lambda_1$. Let $F_{d_3}(G) \subseteq E_{d_3}(G)$ be the set of $d_3$-edges $e$ in $G$ with the property that $e$ is in some distinguishable $(x, c, \mathcal{P})$-gadget in $G$ whose $c$-edge is not supersaturated. Recall that $|F_{d_3}(G)| \leq 64k^3/n^2$ since $d_3 \in D_3^{\mathrm{good}}$. Observe then that there are at most $128k^3/n^2$ colours $d_2 \in D_2$ such that $u_{10}$ is the endpoint of an edge in $F_{d_3}(G)$. Thus for all but at most $(k/4)^3 \cdot n^2 \cdot 128k^3/n^2 = 2k^6$ choices of $\lambda = (d_1, d_2, d_3, d_4, \overrightarrow{f_1}, \overrightarrow{f_2}) \in \Lambda_1$, the edge $u_{10}u_{12}$ is not in $F_{d_3}(G)$. Now fix instead $d_1$, $d_2$, $d_3$, $d_4$, $\overrightarrow{f_2}$. Then since $d_3 \in D_3^{\mathrm{good}}$, there are at most $128k^3/n^2$ choices of $\overrightarrow{f_1}$ such that $f_1 \in F_{d_3}(G)$, so that for all but at most $(k/4)^4 \cdot n \cdot 128k^3/n^2 = k^7/2n$ vectors $\lambda \in \Lambda_1$, $H(\lambda)$ is such that $f_1 \notin F_{d_3}(G)$. Similar analyses show that there are at most $8k^6 + k^7/n \leq 9k^6 \leq |\Lambda_1|/10$ choices of $\lambda \in \Lambda_1$ such that $\{u_1u_2, u_3u_5, u_4u_6, u_9u_{11}, u_{10}u_{12}, u_{13}u_{14}\} \cap F_{d_3}(G) \neq \emptyset$. By definition of $F_{d_3}(G)$ and supersaturation of a $c$-edge, we deduce that for all remaining $\lambda \in \Lambda_1$, $H(\lambda)$ is such that deleting the edges $u_1u_2$, $u_3u_5$, $u_4u_6$, $u_9u_{11}$, $u_{10}u_{12}$, $u_{13}u_{14}$ does not decrease $r(G)$. —

When we perform a twist operation on a twist system $H$ in $G$, since the only new edges we add have some colour in $D_3$, we have that for any new distinguishable $(x, c, \mathcal{P})$-gadget $J$ we create in the twist, one of the new edges $u_1u_3$, $u_2u_4$, $u_5u_6$, $u_9u_{10}$, $u_{11}u_{13}$, $u_{12}u_{14}$ of the twist is playing the role of either $v_5v_6$ or $v_9v_{10}$ in $J$. (Here and throughout the rest of the proof, we imagine completed $(x, c, \mathcal{P})$-gadgets $J$

as having vertices labelled $x, v_5, \ldots, v_{10}$, where the role of $v_i$ corresponds to the role of $u_i$ in Figure 3.5.) We now show that for most $\lambda \in \Lambda_2$, $H(\lambda)$ satisfies property (P3). This is the most delicate part of the argument, and we break it into three more claims.

*Claim 4: There is a subset $\Lambda_3 \subseteq \Lambda_2$ such that $|\Lambda_3| \geq 9|\Lambda_2|/10$ and all $\lambda \in \Lambda_3$ are such that if $J$ is an $(x, c, \mathcal{P})$-gadget that is in $\mathrm{twist}_{H(\lambda)}(G)$ but not in $G$, then the pair $u_9u_{10}$ of $H(\lambda)$ plays the role of $v_9v_{10}$.*

*Proof of claim:* Since the only edges added by any twist operation all have colour in $D_3$, it suffices to show that at most $|\Lambda_2|/10$ vectors $\lambda \in \Lambda_2$ are such that twisting on $\lambda$ creates an $(x, c, \mathcal{P})$-gadget $J$ for which either

(i) one of the pairs $u_1u_3$, $u_2u_4$, $u_5u_6$, $u_{11}u_{13}$, $u_{12}u_{14}$ of $H(\lambda)$ plays the role of $v_9v_{10}$, or

(ii) the edge $v_9v_{10}$ of $J$ is present in $G$.

To address (i), we show that $u_1, u_2, u_5, u_{11}, u_{12} \notin N_G(x)$ for all but at most $|\Lambda_2|/20$ vectors $\lambda \in \Lambda_2$. Note firstly that at most $10k^3/n$ pairs $(d_1, d_4)$ where $d_1 \in D_1$, $d_4 \in D_4^{\mathrm{good}}$ are such that $u_5 \in N_G(x)$, since otherwise $e(N_{D_4}(x), N_G(x)) > 10k^3/n$, contradicting (3.7.2). Thus, at most $(k/4)^2 \cdot n^2 \cdot 10k^3/n = 5k^5n/8$ choices of $\lambda \in \Lambda_2$ are such that $u_5 \in N_G(x)$. Now fix $d_1$, $d_2$, $d_3$, $d_4$, $\overrightarrow{f_2}$ appearing concurrently in some $\lambda \in \Lambda_2$. Notice that there are at most $2k+2$ choices of $\overrightarrow{f_1}$ such that $f_1$ has at least one endpoint in $N_G(x)$. Analysing $\overrightarrow{f_2}$ similarly, we deduce that there are at most $5k^5n/8 + 2(k/4)^4(2k+2)n \leq |\Lambda_2|/20$ choices of $\lambda \in \Lambda_2$ such that at least one of $u_1$, $u_2$, $u_5$, $u_{13}$, $u_{14}$ lies in $N_G(x)$.

Turning now to (ii), we show that at most $|\Lambda_2|/20$ vectors $\lambda \in \Lambda_2$ are such that twisting on $\lambda$ creates any $(x, c, \mathcal{P})$-gadgets $J$ for which the edge $v_9v_{10}$ of $J$ is present in $G$ (and thus one of the pairs $u_1u_3, u_2u_4, u_5u_6, u_9u_{10}, u_{11}u_{13}, u_{12}u_{14}$ of $H(\lambda)$ plays the role of $v_5v_6$). To do this, we use some of the properties of $D_3^{\text{good}}$. Fix $d_2 \in D_2$, $d_3 \in D_3^{\text{good}}$, $d_4 \in D_4^{\text{good}}$, $\overrightarrow{f_1}$, $\overrightarrow{f_2}$ appearing concurrently in some $\lambda \in \Lambda_3$. Note that since $d_3 \in D_3^{\text{good}}$, there are at most $200k^2/n$ pairs $(d_1', d_2')$ where $d_1' \in D_1$, $d_2' \in D_2$, such that there is a $K_3$ in $G$ with vertices $x$, $w_1$, $w_2$, where $w_i$ is the $d_i'$-neighbour of $x$ for $i \in \{1, 2\}$, and the edge $w_1w_2$ is coloured $d_3$. Let the set of these pairs $(d_1', d_2')$ be denoted $L(d_3)$. For each pair $\ell = (d_1', d_2') \in L(d_3)$, let $z_\ell^1$ be the end of the $d_1'cd_2'$-walk starting at $u_{10}$. Similarly, let $z_\ell^2$ denote the end of the $d_2'cd_1'$-walk starting at $u_{10}$. Define $M \coloneqq \bigcup_{\ell \in L(d_3)}\{z_\ell^1, z_\ell^2\}$, so that $|M| \leq 400k^2/n$. Since there are at most $400k^2/n$ choices of $d_1 \in D_1$ for which we obtain $u_9 \in M$, we deduce that for all but at most $(k/4)^3 \cdot n^2 \cdot 400k^2/n = 25k^5n/4$ vectors $\lambda \in \Lambda_2$, $H(\lambda)$ is such that adding the edge $u_9u_{10}$ in colour $d_3$ does not create a new $(x, c, \mathcal{P})$-gadget $J$ where $u_9u_{10}$ plays the role of $v_5v_6$ in $J$ and the edge playing the role of $v_9v_{10}$ in $J$ is already present in $G$ before the twist. One can observe similarly that for all but at most $25k^5n/4$ vectors $\lambda \in \Lambda$, $H(\lambda)$ is such that adding the edge $u_5u_6$ in colour $d_3$ does not create a new $(x, c, \mathcal{P})$-gadget $J$ where $u_5u_6$ plays the role of $v_5v_6$ in $J$ and the edge playing the role of $v_9v_{10}$ in $J$ is already present in $G$ before the twist.

Now fix instead $d_1$, $d_2$, $d_3$, $d_4$, $\overrightarrow{f_2}$ appearing concurrently in some $\lambda \in \Lambda_2$. For each $\ell = (d_1', d_2') \in L(d_3)$, let $y_\ell^1$ be the end of the $d_1'cd_2'$-walk starting at $u_4$, let $y_\ell^2$ be the end of the $d_2'cd_1'$-walk starting at $u_4$, let $z_\ell^1$ be the end of the $d_1'cd_2'$-walk starting at $u_3$, and let $z_\ell^2$ be the end of the $d_2'cd_1'$-walk starting at $u_3$. Define $M \coloneqq \bigcup_{\ell \in L(d_3)}\{y_\ell^1, y_\ell^2, z_\ell^1, z_\ell^2\}$, and notice that $|M| \leq 800k^2/n$. We deduce that there are at most $1600k^2/n$ choices of $\overrightarrow{f_1}$ such that $f_1$ has an endpoint in $M$, and that for all remaining choices of $\overrightarrow{f_1}$, twisting on $\lambda = (d_1, d_2, d_3, d_4, \overrightarrow{f_1}, \overrightarrow{f_2})$ cannot create a new $(x, c, \mathcal{P})$-gadget $J$ where the new $d_3$-edges $u_1u_3$ or $u_2u_4$ play the role of $v_5v_6$ in $J$ and the edge $v_9v_{10}$ of $J$ is present in $G$. Analysing $\overrightarrow{f_2}$ similarly, we conclude that for all but at most $13k^5n \leq |\Lambda_2|/20$ choices of $\lambda \in \Lambda_2$, twisting on $\lambda$ cannot create a new $(x, c, \mathcal{P})$-gadget $J$ for which the edge $v_9v_{10}$ of $J$ is present in $G$. $-$

*Claim 5: There is a subset $\Lambda_4 \subseteq \Lambda_3$ such that $|\Lambda_4| \geq 9|\Lambda_3|/10$ and all $\lambda \in \Lambda_4$ are such that if $J$ is an $(x, c, \mathcal{P})$-gadget that is in $twist_{H(\lambda)}(G)$ but not in $G$, then the pair $u_5u_6$ of $H(\lambda)$ plays the role of $v_5v_6$.*

*Proof of claim:* By Claim 4, it will suffice to show that at most $|\Lambda_3|/10$ vectors $\lambda \in \Lambda_3$ are such that twisting on $\lambda$ creates an $(x, c, \mathcal{P})$-gadget $J$ for which either

(i) one of the pairs $u_1u_3, u_2u_4, u_{11}u_{13}, u_{12}u_{14}$ of $H(\lambda)$ plays the role of $v_5v_6$, and $u_9u_{10}$ plays the role of $v_9v_{10}$, or

(ii) the edge $v_5v_6$ of $J$ is present in $G$ and the pair $u_9u_{10}$ of $H(\lambda)$ plays the role of $v_9v_{10}$.

To address (i), fix $d_1$, $d_2$, $d_3$, $d_4$, $\overrightarrow{f_2}$ appearing concurrently in some $\lambda \in \Lambda_3$. Let $a_1$ be the end of the $d_1cd_2$-walk starting at $u_4$, let $a_2$ be the end of the $d_2cd_1$-walk starting at $u_4$, let $b_1$ be the end of the $d_1cd_2$-walk starting at $u_3$, and let $b_2$ be the end of the $d_2cd_1$-walk starting at $u_3$. Since there are at most 8 choices of $\overrightarrow{f_1}$ with an endpoint in $\{a_1, a_2, b_1, b_2\}$, we deduce that for all remaining choices of $\overrightarrow{f_1}$, twisting on $\lambda = (d_1, d_2, d_3, d_4, \overrightarrow{f_1}, \overrightarrow{f_2})$ cannot create an $(x, c, \mathcal{P})$-gadget $J$ for which the new $d_3$-edges $u_1u_3$ or $u_2u_4$ play the role of $v_5v_6$ in $J$ and $u_9u_{10}$ plays the role of $v_9v_{10}$. Analysing $\overrightarrow{f_2}$ similarly, we conclude that we must discard at most $k^4n/16 \leq |\Lambda_3|/20$ vectors $\lambda \in \Lambda_3$ to account for (i).

Turning now to (ii), write $D_4 = \{d_4^1, d_4^2, \ldots, d_4^{k/4}\}$. For each $d_4^i \in D_4$, let $y_i$ be the $d_4^i$-neighbour of $x$, let $z_i$ be the $c$-neighbour of $y_i$, define $R_i := N_{D_1}(y_i)$ and

$S_i := N_{D_2}(z_i)$. Notice that $\sum_{i=1}^{k/4} e(R_i, S_i) \leq 5k^4/2n$, since otherwise we obtain a contradiction to (3.7.2) for some pair $(R_i, S_i)$. We deduce that there are at most $5k^4/2n$ triples $(d_1, d_2, d_3)$ with each $d_i \in D_i$ for which adding the edge $u_9u_{10}$ in colour $d_3$ creates an $(x, c, \mathcal{P})$-gadget $J$ for which $u_9u_{10}$ plays the role of $v_9v_{10}$ in $J$ and the edge playing the role of $v_5v_6$ is already present in $G$, whence at most $(5k^4/2n) \cdot (k/4) \cdot n^2 = 5k^5n/8 \leq |\Lambda_3|/20$ choices of $\lambda \in \Lambda_3$ are such that twisting on $\lambda$ creates an $(x, c, \mathcal{P})$-gadget of this type. —

*Claim 6: There is a subset $\Lambda_5 \subseteq \Lambda_4$ such that $|\Lambda_5| \geq 9|\Lambda_4|/10$ and all $\lambda \in \Lambda_5$ are such that if $J$ is an $(x, c, \mathcal{P})$-gadget that is in twist$_{H(\lambda)}(G)$ but not in $G$ and the pairs $u_5u_6, u_9u_{10}$ of $H(\lambda)$ play the roles of the edges $v_5v_6, v_9v_{10}$ of $J$ respectively, then $J$ is the canonical $(x, c, \mathcal{P})$-gadget of the twist.*

*Proof of claim:* Fix $d_3$, $d_4$, $\overrightarrow{f_1}$, $\overrightarrow{f_2}$ appearing concurrently in some $\lambda \in \Lambda_4$. By (3.7.2), we have that $e(N_{D_2}(u_8), N_{D_4}(x)) \leq 10k^3/n$. We deduce that there are at most $10k^3/n$ choices of the pair $(d_1, d_2)$ such that the $d_1$-neighbour of $u_6$ lies in $N_{D_4}(x)$, whence for all but at most $5k^5n/8 \leq |\Lambda_4|/10$ choices of $\lambda \in \Lambda_4$, the canonical $(x, c, \mathcal{P})$-gadget of the twist is the only new $(x, c, \mathcal{P})$-gadget for which $u_5u_6, u_9u_{10}$ play the roles of $v_5v_6, v_9v_{10}$ respectively. —

Note that, by Claims 4–6, the canonical $(x, c, \mathcal{P})$-gadget of a twist on $\lambda \in \Lambda_5$ is clearly distinguishable in twist$_{H(\lambda)}(G)$ since its edges $v_5v_6$ and $v_9v_{10}$ with colours in $D_3$ were added by the twist and performing this twist creates no other $(x, c, \mathcal{P})$-gadgets. Thus Claims 4–6 imply that $H(\lambda)$ satisfies (P3) for all $\lambda \in \Lambda_5$. Recalling that $\mathrm{sat}_G(e) \leq k - 1$ for the $c$-edge $e$ of $H(\lambda)$ for all $\lambda \in \Lambda$ and also using Claim 3, we now deduce that $r(\mathrm{twist}_{H(\lambda)}(G)) = r(G) + 1$, and thus $\mathrm{twist}_{H(\lambda)}(G) \in A_{s+1}^{D^*}$, for all $\lambda \in \Lambda_5$. Since $H(\lambda)$ satisfies (P1) for all $\lambda \in \Lambda_1$, we deduce that

$\mathrm{twist}_{H(\lambda)}(G) \in T_{s+1}^{D^*}$ for all $\lambda \in \Lambda_5$, and that $\delta_s \geq |\Lambda_5| \geq |\Lambda|/2 \geq 3k^4n^2/2^{17}$. We conclude that if $s \leq k^4/2^{22}n^2$ and $T_s^{D^*}$ is non-empty, then $T_{s+1}^{D^*}$ is non-empty and $|T_s^{D^*}|/|T_{s+1}^{D^*}| \leq 2^{17} \cdot 24n^4(s+1)/3k^4n^2 \leq 1/2$. Now, fix $s \leq \mu^4n^2/2^{23}$. If $T_s^{D^*}$ is empty, then $\mathbb{P}\left[r(\mathbf{G}) = s \mid \mathbf{G} \in \widetilde{\mathcal{Q}}_{D^*}^{\mathrm{col}}\right] = 0$. If $T_s^{D^*}$ is non-empty, then

$$\mathbb{P}\left[r(\mathbf{G}) = s \mid \mathbf{G} \in \widetilde{\mathcal{Q}}_{D^*}^{\mathrm{col}}\right] = \frac{|T_s^{D^*}|}{|\widetilde{\mathcal{Q}}_{D^*}^{\mathrm{col}}|} \leq \frac{|T_s^{D^*}|}{|T_{k^4/2^{22}n^2}^{D^*}|} = \prod_{j=s}^{k^4/2^{22}n^2-1} \frac{|T_j^{D^*}|}{|T_{j+1}^{D^*}|} \leq (1/2)^{k^4/2^{22}n^2-s},$$

and thus,

$$\begin{aligned}\mathbb{P}\left[r(\mathbf{G}) \leq \mu^4n^2/2^{23} \mid \mathbf{G} \in \widetilde{\mathcal{Q}}_{D^*}^{\mathrm{col}}\right] &\leq \sum_{s=0}^{\mu^4n^2/2^{23}} \exp(-(k^4/2^{22}n^2 - s)\ln 2) \\ &\leq \exp\left(-\frac{\mu^4n^2}{2^{24}}\right),\end{aligned}$$

which completes the proof of the lemma. □

Next, we show that in order to find many well-spread $(x,c)$-absorbing gadgets in $G \in \mathcal{G}_{D\cup\{c\}}^{\mathrm{col}}$, it suffices to show that $r(G)$ is large for some equitable partition $\mathcal{P}$ of $D$ into four parts. (Recall that 'well-spread' was defined in Definition 3.4.7.)

**Lemma 3.7.9.** *Suppose that $1/n \ll \mu$, and let $D \subseteq [n-1]$ be such that $|D| \leq \mu n$. Let $x \in V$, let $c \in [n-1] \setminus D$, and let $\mathcal{P} = \{D_i\}_{i=1}^4$ be an equitable partition of $D$. Then for any integer $t \geq 0$ and any $G \in \mathcal{G}_{D\cup\{c\}}^{col}$, if $r(G) \geq t$, then $G$ contains a $5\mu n/4$-well-spread collection of $t$ distinct $(x,c)$-absorbing gadgets.*

**Proof.** Let $G \in \mathcal{G}_{D\cup\{c\}}^{\mathrm{col}}$, let $t \geq 0$ be an integer, and suppose that $r(G) \geq t$. Then, since $|D| \leq \mu n$ and by definition of $r$, we deduce that there is a collection $\mathcal{A}_{(x,c,\mathcal{P})}$ of $t$ distinct $(x,c,\mathcal{P})$-gadgets satisfying the following conditions:

(i) Each edge of $G$ with colour in $D_3$ is contained in at most one $(x,c,\mathcal{P})$-gadget

$J \in \mathcal{A}_{(x,c,\mathcal{P})}$;

(ii) Each $c$-edge of $G$ is contained in at most $\mu n$ $(x, c, \mathcal{P})$-gadgets $J \in \mathcal{A}_{(x,c,\mathcal{P})}$.

Fix $v \in V \setminus \{x\}$. Let $e$ be the $c$-edge of $G$ incident to $v$ and for each $d \in D_3$ let $f_d$ be the $d$-edge of $G$ incident to $v$. Then by conditions (i) and (ii) there are at most $5\mu n/4$ $(x, c, \mathcal{P})$-gadgets $J \in \mathcal{A}_{(x,c,\mathcal{P})}$ containing any of the edges in $\{e\} \cup \bigcup_{d \in D_3} \{f_d\}$. Note that if $v$ is contained in some $J \in \mathcal{A}_{(x,c,\mathcal{P})}$, then $v$ is incident to either the $c$-edge in $J$, or to one of the edges in $J$ with colour in $D_3$. We thus conclude that $v$ is contained in at most $5\mu n/4$ $(x, c, \mathcal{P})$-gadgets $J \in \mathcal{A}_{(x,c,\mathcal{P})}$. It immediately follows that no edge of $G$ is contained in more than $5\mu n/4$ $(x, c, \mathcal{P})$-gadgets $J \in \mathcal{A}_{(x,c,\mathcal{P})}$.

For each $d \in D_1 \cup D_2 \cup D_4$, there are at most $5\mu n/4$ $J \in \mathcal{A}_{(x,c,\mathcal{P})}$ with $d \in \phi(J)$ since each such $J$ must contain the $d$-neighbour of $x$ in $G$. For each $d \in D_3$, there are at most $\mu n/2$ $d$-edges $f$ in $G$ such that both endpoints of $f$ are neighbours of $x$. Any $J \in \mathcal{A}_{(x,c,\mathcal{P})}$ for which $d \in \phi(J)$ must contain one of these edges $f$. Thus by (i), there are at most $\mu n/2$ $J \in \mathcal{A}_{(x,c,\mathcal{P})}$ such that $d \in \phi(J)$.

Finally, define a function $g$ on $\mathcal{A}_{(x,c,\mathcal{P})}$ by setting $g(J) := J - f$, where $f$ is the unique edge of $J$ with colour in $D_4$, for each $J \in \mathcal{A}_{(x,c,\mathcal{P})}$. Then it is clear that $g$ is injective and that $g(J)$ is an $(x, c)$-absorbing gadget, for each $J \in \mathcal{A}_{(x,c,\mathcal{P})}$. Thus, $g(\mathcal{A}_{(x,c,\mathcal{P})})$ is a $5\mu n/4$-well-spread collection of $t$ distinct $(x, c)$-absorbing gadgets in $G$, as required. $\square$

### 3.7.2 Weighting factor

We now state two results on the number of 1-factorizations in dense $d$-regular graphs $G$, where a 1-factorization of $G$ consists of an ordered set of $d$ perfect matchings in $G$. We will use these results to find a 'weighting factor' (see Corollary 3.7.12), which we will use to compare the probabilities of particular events occurring in different probability spaces. For any graph $G$, let $M(G)$ denote the number of distinct 1-factorizations of $G$, and for any $n, d \in \mathbb{N}$, let $\mathcal{G}_d^n$ denote the set of $d$-regular graphs on $n$ vertices. Firstly, the Kahn-Lovász Theorem (see e.g. [5]) states that a graph with degree sequence $r_1, \ldots, r_n$ has at most $\prod_{i=1}^n (r_i!)^{1/2r_i}$ perfect matchings. In particular, an $n$-vertex $d$-regular graph has at most $(d!)^{n/2d}$ perfect matchings. To determine an upper bound for the number of 1-factorizations of a $d$-regular graph $G$, one can simply apply the Kahn-Lovász Theorem repeatedly to obtain $M(G) \leq \prod_{r=1}^d (r!)^{n/2r}$. Using Stirling's approximation, we obtain the following result.

**Theorem 3.7.10.** *Suppose $n \in \mathbb{N}$ is even with $1/n \ll 1$, and $d \geq n/2$. Then every $G \in \mathcal{G}_d^n$ satisfies*

$$M(G) \leq \left( \left(1 + n^{-1/2}\right) \frac{d}{e^2} \right)^{dn/2}.$$

On the other hand, Ferber, Jain, and Sudakov [43] proved the following lower bound for the number of distinct 1-factorizations in dense regular graphs.

**Theorem 3.7.11** ([43, Theorem 1.2])**.** *Suppose $C > 0$ and $n \in \mathbb{N}$ is even with $1/n \ll 1/C \ll 1$, and $d \geq (1/2 + n^{-1/C})n$. Then every $G \in \mathcal{G}_d^n$ satisfies*

$$M(G) \geq \left( \left(1 - n^{-1/C}\right) \frac{d}{e^2} \right)^{dn/2}.$$

Theorems 3.7.10 and 3.7.11 immediately yield the following corollary:

**Corollary 3.7.12.** *Suppose $C > 0$ and $n \in \mathbb{N}$ is even with $1/n \ll 1/C \ll 1$, and*

$d \geq (1/2 + n^{-1/C})n$. *Then*

$$\frac{M(G)}{M(H)} \leq \exp\left(2n^{1-1/C}d\right),$$

*for all* $G, H \in \mathcal{G}_d^n$.

Recall that for $G \in \mathcal{G}_{[n-1]}^{\mathrm{col}}$ and a set of colours $D \subseteq [n-1]$, $G|_D$ is be the spanning subgraph of $G$ containing precisely those edges of $G$ which have colour in $D$. We now have all the tools we need to prove Lemma 3.4.8.

**Proof of Lemma 3.4.8.** Let $C > 0$ be the constant given by Corollary 3.7.12 and suppose that $1/n \ll 1/C, \mu, \varepsilon$. Let $\mathbb{P}$ denote the probability measure for the space corresponding to choosing $\mathbf{G} \in \mathcal{G}_{[n-1]}^{\mathrm{col}}$ uniformly at random. Fix $D \subseteq [n-1]$ such that $|D| = \varepsilon n$, and let $\mathbb{P}_D$ denote the probability measure for the space corresponding to choosing $\mathbf{H} \in \mathcal{G}_D^{\mathrm{col}}$ uniformly at random. Let $\mathcal{G}_D^{\mathrm{bad}}$ denote the set of $H \in \mathcal{G}_D^{\mathrm{col}}$ such that $H$ is not $\varepsilon$-locally edge-resilient. For $H \in \mathcal{G}_D^{\mathrm{col}}$, write $N_H$ for the number of distinct completions of $H$ to an element $G \in \mathcal{G}_{[n-1]}^{\mathrm{col}}$; that is, $N_H$ is the number of 1-factorizations of the complement of $H$. Then

$$\begin{aligned}
\mathbb{P}\left[\mathbf{G}|_D \,\text{is not}\, \varepsilon\text{-locally edge-resilient}\right] &= \frac{\sum_{H \in \mathcal{G}_D^{\mathrm{bad}}} N_H}{\sum_{H' \in \mathcal{G}_D^{\mathrm{col}}} N_{H'}} \\
&\leq \mathbb{P}_D\left[\mathbf{H} \in \mathcal{G}_D^{\mathrm{bad}}\right] \cdot \exp\left(2n^{2-1/C}\right) \\
&\leq \exp\left(-\varepsilon^3 n^2/2000\right),
\end{aligned}$$

where we have used Lemma 3.7.1 and Corollary 3.7.12. Then, union bounding over choices of $D$, we deduce that

$$\mathbb{P}\left[\mathbf{G}\,\text{is not}\,\varepsilon\text{-locally edge-resilient}\right] \le \binom{n-1}{\varepsilon n}\exp\left(-\frac{\varepsilon^3 n^2}{2000}\right) \le \exp\left(-\frac{\varepsilon^3 n^2}{4000}\right). \tag{3.7.3}$$

Now, fix $x \in V$, and fix $c \in [n-1]$. Choose $F \subseteq [n-1] \setminus \{c\}$ of size $|F| = \mu n$ arbitrarily. Write $F^* := F \cup \{c\}$, and let $\mathbb{P}_{F^*}$ denote the probability measure for the space $\mathcal{S}$ corresponding to choosing $\mathbf{H} \in \mathcal{G}^{\text{col}}_{F^*}$ uniformly at random. Let $\mathcal{P}$ be an equitable (ordered) partition of $F$ into four subsets. Let $A^{(x,c)}_{F^*} \subseteq \mathcal{G}^{\text{col}}_{F^*}$ be the set of $H \in \mathcal{G}^{\text{col}}_{F^*}$ such that $H$ has a $5\mu n/4$-well-spread collection of at least $\mu^4 n^2/2^{23}$ $(x,c)$-absorbing gadgets. Then, considering $A^{(x,c)}_{F^*}$, $\mathcal{Q}^{\text{col}}_{F^*}$, $\widetilde{\mathcal{Q}}^{\text{col}}_{F^*}$ as events in $\mathcal{S}$, observe that

$$\begin{aligned}
\mathbb{P}_{F^*}\left[\overline{A^{(x,c)}_{F^*}}\right] &\le \mathbb{P}_{F^*}\left[\widetilde{\mathcal{Q}}^{\text{col}}_{F^*}\right]\mathbb{P}_{F^*}\left[\overline{A^{(x,c)}_{F^*}}\;\middle|\;\widetilde{\mathcal{Q}}^{\text{col}}_{F^*}\right] + \mathbb{P}_{F^*}\left[\overline{\widetilde{\mathcal{Q}}^{\text{col}}_{F^*}}\right] \\
&\overset{(3.7.1)}{\le} \mathbb{P}_{F^*}\left[\overline{A^{(x,c)}_{F^*}}\;\middle|\;\widetilde{\mathcal{Q}}^{\text{col}}_{F^*}\right] + \mathbb{P}_{F^*}\left[\overline{\mathcal{Q}^{\text{col}}_{F^*}}\right].
\end{aligned}$$

Thus, applying Lemma 3.7.9, Lemma 3.7.3, and Lemma 3.7.8, we obtain

$$\begin{aligned}
\mathbb{P}_{F^*}\left[\overline{A^{(x,c)}_{F^*}}\right] &\le \mathbb{P}_{F^*}\left[r(\mathbf{H}) \le \mu^4 n^2/2^{23}\;\middle|\;\mathbf{H} \in \widetilde{\mathcal{Q}}^{\text{col}}_{F^*}\right] + \mathbb{P}_{F^*}\left[\mathbf{H} \notin \mathcal{Q}^{\text{col}}_{F^*}\right] \\
&\le \exp\left(-\frac{\mu^4 n^2}{2^{24}}\right) + \exp\left(-\mu^3 n^2\right) \le \exp\left(-\frac{\mu^4 n^2}{2^{25}}\right).
\end{aligned}$$

Then by Corollary 3.7.12,

$$\begin{aligned}
\mathbb{P}\left[\mathbf{G}|_{F^*} \notin A^{(x,c)}_{F^*}\right] &= \frac{\sum_{H \in \overline{A^{(x,c)}_{F^*}}} N_H}{\sum_{H' \in \mathcal{G}^{\text{col}}_{F^*}} N_{H'}} \le \mathbb{P}_{F^*}\left[\mathbf{H} \notin A^{(x,c)}_{F^*}\right] \cdot \exp\left(2n^{2-1/C}\right) \\
&\le \exp\left(-\frac{\mu^4 n^2}{2^{26}}\right).
\end{aligned}$$

In particular, with probability at least $1 - \exp(-\mu^4 n^2/2^{26})$, $\mathbf{G}$ has a $5\mu n/4$-well-spread collection of at least $\mu^4 n^2/2^{23}$ $(x, c)$-absorbing gadgets. Now, union bounding over all vertices $x \in V$ and all colours $c \in [n-1]$, we deduce that

$$\mathbb{P}\left[\mathbf{G} \text{ is not } \mu\text{-robustly gadget-resilient}\right] \leq n^2 \cdot \exp\left(-\frac{\mu^4 n^2}{2^{26}}\right) \leq \exp\left(-\frac{\mu^4 n^2}{2^{27}}\right). \tag{3.7.4}$$

The result now follows by combining (3.7.3) and (3.7.4). □

## 3.8 Modifications and Corollaries

In this section we show how to derive the $n$ odd case of Theorem 3.1.3 from the case when $n$ is even. We also show how Theorem 3.1.3(ii) implies Corollary 3.1.4.

### 3.8.1 A rainbow Hamilton cycle for $n$ odd

We actually derive the $n$ odd case of Theorem 3.1.3 from the following slightly stronger version of Theorem 3.1.3(ii) in the case when $n$ is even.

**Theorem 3.8.1.** *If $n$ is even and $\phi$ is a uniformly random 1-factorization of $K_n$, then for every vertex $v$, with high probability, $\phi$ admits a rainbow cycle containing all of the colours and all of the vertices except $v$.*

We now argue that our proof of Theorem 3.1.3 for $n$ even is sufficiently robust to also obtain this strengthening. In particular, we can strengthen Lemma 3.4.9 so that the absorber does not contain $v$, since (a)–(c) in Lemma 3.6.3, (a)–(c) in Lemma 3.6.4, and (a)–(f) in Lemma 3.6.5 all hold after deleting $v$ from any part in the absorber partition. The proof of Lemma 3.4.10 is also sufficiently robust to guarantee that the rainbow path from the lemma does not contain $v$, but we do not need this strengthening, since we can instead strengthen Proposition 3.4.5 to obtain a rainbow cycle containing $P' - v$ and all of the colours, as follows. If $v \in V(P')$, then we replace $v$ in $P'$ with a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-cover by deleting $v$ and adding a $(V_{\text{flex}}, C_{\text{flex}}, G_{\text{flex}})$-cover of $w$, $w'$, and $\phi(vw)$, where $w$ and $w'$ are the vertices adjacent to $v$ in $P'$. The remainder of the proof proceeds normally, letting $v_\ell := v$ to ensure $v \notin V(P_1'')$. In this procedure, we need to assume that $P'$ is contained in $(V \setminus V', C \setminus C', G')$ with $\delta/19$-bounded remainder (rather than $\delta/18$), but in Lemma 3.4.9 we can find a $38\gamma$-absorber, which completes the proof.

Now we show how Theorem 3.8.1 implies the odd $n$ case of Theorem 3.1.3.

**Proof of Theorem 3.1.3, $n$ odd case.** When $n$ is odd, any optimal edge-colouring of $K_n$ has $n$ colour classes, each containing precisely $(n-1)/2$ edges. For every colour $c$, there is a unique vertex which has no incident edges of colour $c$, and for every vertex $v$, there is a unique colour such that $v$ has no incident edges of this colour. Thus, we can obtain a 1-factorization $\phi'$ of $K_{n+1}$ from an optimal edge-colouring $\phi$ of $K_n$ in the following way. We add a vertex $z$, and for every other vertex $v$, we add an edge $zv$, where $\phi'(zv)$ is the unique colour $c$ such that $v$ is not incident to a $c$-edge in $K_n$. Note that this operation produces a bijection from the set of $n$-edge-colourings of $K_n$ to the set of 1-factorizations of $K_{n+1}$. Thus, if $n$ is odd and $\phi$ is a uniformly random optimal edge-colouring of $K_n$, then $\phi'$ is a uniformly random optimal edge-colouring of $K_{n+1}$. By Theorem 3.8.1, with high probability there is a rainbow cycle $F$ in $K_{n+1}$ containing all of the colours and all of the vertices except $z$, so $F$ is a rainbow Hamilton cycle in $K_n$, satisfying

Theorem 3.1.3(ii). Deleting any edge from $F$ gives a rainbow Hamilton path, as required in Theorem 3.1.3(i). $\square$

### 3.8.2 Symmetric Latin squares

Now we use Theorem 3.1.3 to prove Corollary 3.1.4.

**Proof of Corollary 3.1.4.** Suppose that $n \in \mathbb{N}$ is odd. Firstly, note that there is a one-to-one correspondence between the set $\mathcal{L}_n^{\mathrm{sym}}$ of symmetric $n \times n$ Latin squares with symbols in $[n]$ (say) and the set $\Phi_n$ of optimal edge-colourings of $K_n$ on vertices $[n]$ and with colours in $[n]$. Indeed, let $\phi \in \Phi_n$. Then we can construct a unique symmetric Latin square $L_\phi \in \mathcal{L}_n^{\mathrm{sym}}$ by putting the symbol $\phi(ij)$ in position $(i,j)$ for all edges $ij \in E(K_n)$, and for each position $(i,i)$ on the leading diagonal we now enter the unique symbol still missing from row $i$. Conversely, let $L \in \mathcal{L}_n^{\mathrm{sym}}$. We can obtain a unique element $\phi_L \in \Phi_n$ from $L$ in the following way. Colour each edge $ij$ of the complete graph $K_n$ on vertex set $[n]$ with the symbol in position $(i,j)$ of $L$. It is clear that $\phi_L$ is proper, and thus $\phi_L$ is optimal. Moreover, it is clear that we can uniquely recover $L$ from $\phi_L$.

Now, let $K_n^\circ$ be the graph obtained from $K_n$ by adding a loop $ii$ at every vertex $i \in [n]$, and for every $\phi \in \Phi_n$, let $\phi^\circ$ be the unique proper $n$-edge-colouring of $K_n^\circ$ such that the restriction of $\phi^\circ$ to the underlying simple graph is $\phi$. The rainbow 2-factors in $K_n^\circ$ admitted by $\phi^\circ$ correspond to transversals in $L_\phi$ in the following way. If $L \in \mathcal{L}_n^{\mathrm{sym}}$ and $T$ is a transversal of $L$, then the subgraph of $K_n^\circ$ induced by the edges $ij$ where $(i,j) \in T$ is a rainbow 2-factor. If $\sigma$ is the underlying permutation of $T$, then the cycles of this rainbow 2-factor are precisely the cycles in the cycle decomposition of $\sigma$, up to orientation. Therefore a rainbow Hamilton cycle in $K_n^\circ$ corresponds to two disjoint Hamilton transversals in $L_\phi$.

By these correspondences, for $n$ odd, if $\mathbf{L} \in \mathcal{L}_n^{\mathrm{sym}}$ is a uniformly random symmetric $n \times n$ Latin square, then $\phi_{\mathbf{L}}$ is a uniformly random optimal edge-colouring of $K_n$. By Theorem 3.1.3(ii), $\phi_{\mathbf{L}}$ admits a rainbow Hamilton cycle $F$ with high probability. Since $F$ is also a rainbow Hamilton cycle in $K_n^\circ$, the corresponding transversals in $\mathbf{L}$ are Hamilton, as desired. $\square$

Note that, if $n$ is odd, the leading diagonal of any $L \in \mathcal{L}_n^{\mathrm{sym}}$ is also a transversal, disjoint from any Hamilton transversal. Indeed, by symmetry all symbols appear an even number of times off of the leading diagonal, and therefore an odd number of times (and thus exactly once) on the leading diagonal.

# CHAPTER 4

# HAMILTON TRANSVERSALS IN RANDOM LATIN SQUARES

**Abstract**

Gyárfás and Sárközy conjectured that every $n \times n$ Latin square has a 'cycle-free' partial transversal of size $n-2$. We confirm this conjecture in a strong sense for almost all Latin squares, by showing that as $n \to \infty$, all but a vanishing proportion of $n \times n$ Latin squares have a *Hamilton* transversal, i.e. a full transversal for which any proper subset is cycle-free. In fact, we prove a counting result that in almost all Latin squares, the number of Hamilton transversals is essentially that of Taranenko's upper bound on the number of full transversals. This result strengthens a result of Kwan (which in turn implies that almost all Latin squares also satisfy the famous Ryser-Brualdi-Stein conjecture).

## 4.1 Introduction

### 4.1.1 Transversals in Latin squares

An $n \times n$ *Latin square* is an arrangement of $n$ symbols into $n$ rows and $n$ columns, such that each row and each column contains precisely one instance of each symbol. A *(full) transversal* in an $n \times n$ Latin square is a collection of $n$ positions of the Latin square that use each row, column, and symbol exactly once, and a *partial transversal* is a collection of at most $n$ positions not using any row, column, or symbol more than once. The most famous open problem on the topic of transversals in Latin squares is the following.

**Conjecture 4.1.1** (Ryser, Brualdi, and Stein [20, 114, 119])**.** *All $n \times n$ Latin squares have a partial transversal of size $n-1$.*

Conjecture 4.1.1 would be best-possible, because for even $n$ the addition table of the integers modulo $n$ is a Latin square which has no transversal. If $n$ is odd, it is actually conjectured that all $n \times n$ Latin squares have a full transversal. For nearly forty years the best result towards Conjecture 4.1.1 was the theorem of Hatami and Shor [66, 118] (improving [125, 18]) that all $n \times n$ Latin squares have a partial transversal of size $n - O(\log^2 n)$. Recently however, Keevash, Pokrovskiy, Sudakov, and Yepremyan [79] improved the error term to $O(\log n / \log\log n)$.

Conjecture 4.1.1 is related to the following conjecture of Andersen [9]. An edge-coloured graph is *rainbow* if all of its edges have different colours, and an edge-colouring is *proper* if no two edges of the same colour share a vertex.

**Conjecture 4.1.2** (Andersen [9])**.** *All proper edge-colourings of $K_n$, the complete graph on $n$ vertices, admit a rainbow path of length $n-2$.*

In light of the result of Maamoun and Meyniel [96] that for infinitely many $n$ there are proper edge-colourings of $K_n$ without a rainbow Hamilton path, Conjecture 4.1.2 would be best-possible. Similarly to Conjecture 4.1.1, progress towards Conjecture 4.1.2 has largely focussed on increasing the length of the longest rainbow path known to exist for any proper edge-colouring of $K_n$ (see for example [22, 48, 58, 59]). Alon, Pokrovskiy, and Sudakov [7] were the first to asymptotically prove Conjecture 4.1.2 by exhibiting the existence of a rainbow path of length $n - O(n^{3/4})$, with the best known error bound now being $O(n^{1/2}\log n)$, provided by Balogh and Molla [10].

Let $\overleftrightarrow{K_n}$ be the digraph obtained from the complete $n$-vertex graph $K_n$ by replacing each edge with two arcs (one in each direction) and adding a directed loop at each vertex. For every $n \times n$ Latin square we can uniquely associate an arc-colouring of $\overleftrightarrow{K_n}$ as follows: for every position $(i, j)$ of the Latin square, assign the symbol of $(i, j)$ as a colour to the arc in $\overleftrightarrow{K_n}$ with tail $i$ and head $j$. Importantly, a partial transversal corresponds to a rainbow subgraph of $\overleftrightarrow{K_n}$ with maximum in-degree and out-degree one. A set of positions is a *cycle* if the corresponding subgraph of $\overleftrightarrow{K_n}$ is a directed cycle, and a partial transversal is *cycle-free* if it contains no cycle. Thus, cycle-free partial transversals correspond to linear directed forests in $\overleftrightarrow{K_n}$. Gyárfás and Sárközy [60] proposed the following conjecture, which combines aspects of Conjectures 4.1.1 and 4.1.2.

**Conjecture 4.1.3** (Gyárfás and Sárközy [60])**.** *All $n \times n$ Latin squares have a cycle-free partial transversal of size $n - 2$.*

A *proper $k$-arc-colouring* of a digraph is a colouring of its arcs with $k$ colours such that no two arcs of the same colour have a common head, or a common tail.

The set of $n \times n$ Latin squares is in fact in bijection with the set of proper $n$-arc-colourings of $\overleftrightarrow{K_n}$, with the correspondence described above. Thus, Conjecture 4.1.3 is equivalent to the following: all proper $n$-arc-colourings of $\overleftrightarrow{K_n}$ contain a rainbow directed linear forest with at least $n-2$ arcs. No undirected analogue of this conjecture is known – Balogh and Molla [10] proved that for every proper edge-colouring of $K_n$, there is a rainbow linear forest with at least $n - O(\log^2 n)$ edges, and this bound is the best known.

Less is known in the directed setting. Gyárfás and Sárközy [60] proved that every $n \times n$ Latin square has a cycle-free partial transversal of size $n - O(n \log\log n/ \log n)$, and Benzing, Pokrovskiy, and Sudakov [11] improved the error bound to $O(n^{2/3})$. Benzing, Pokrovskiy, and Sudakov [11] also proved that every proper arc-colouring of $\overleftrightarrow{K_n}$ contains a rainbow directed cycle of size $n - O(n^{4/5})$ and asked by how much this bound can be improved. We believe it is also interesting to consider by how much this bound can be improved if one considers both rainbow directed cycles and paths, i.e. rainbow *connected* subgraphs of maximum in-degree and out-degree at most one. We conjecture the following.

**Conjecture 4.1.4.** *All proper arc-colourings of $\overleftrightarrow{K_n}$ admit a rainbow directed cycle or path of length at least $n-1$.*

We define a set of positions in a Latin square to be *connected* if the corresponding subgraph of $\overleftrightarrow{K_n}$ is (weakly) connected, and we say a transversal is *Hamilton* if it is both full and connected. For the case of $n$-arc-colourings, Conjecture 4.1.4 is equivalent to the following: all $n \times n$ Latin squares have a connected transversal of size $n-1$. If true, Conjecture 4.1.3 implies that every $n \times n$ Latin square has a partial transversal of size one less than what is predicted by Conjecture 4.1.1 and also that every proper $n$-edge-colouring of $K_n$ contains either a rainbow path of length $n-2$, as predicted by Conjecture 4.1.2, or a spanning rainbow forest with two components. Conjecture 4.1.4, if true, implies all of Conjectures 4.1.1–4.1.3.

### 4.1.2 Random Latin squares

In this paper, we study the above conjectures in the probabilistic setting. Recently, Kwan [90] proved that at most a vanishing proportion of Latin squares fail to satisfy the statement of Conjecture 4.1.1, even finding (many) full transversals in most Latin squares, as follows.

**Theorem 4.1.5** (Kwan [90]). *Almost all $n \times n$ Latin squares have at least*

$$\left((1-o(1))\frac{n}{e^2}\right)^n$$

*transversals.*

Equivalently, a uniformly random $n\times n$ Latin square has at least $((1-o(1))n/e^2)^n$ transversals with high probability. We note that it was proven by Taranenko [120] (with a simpler proof later found by Glebov and Luria [49]) that $n\times n$ Latin squares can have at most $((1+o(1))n/e^2)^n$ transversals, so that the counting term given in Theorem 4.1.5 is best possible, up to the exponential error term. Analogously, the authors, together with Kühn and Osthus [56], proved that almost all optimal edge-colourings (proper edge-colourings using the minimum possible number of colours) of $K_n$ admit a rainbow Hamilton path, which proves a stronger statement than Conjecture 4.1.2 for all but a vanishing proportion of such colourings.

The main result of this paper is the following strengthening of Theorem 4.1.5.

**Theorem 4.1.6.** *Almost all proper $n$-arc-colourings of $\overleftrightarrow{K_n}$ contain at least*

$$\left((1 - o(1))\frac{n}{e^2}\right)^n$$

*rainbow directed Hamilton cycles. Equivalently, almost all $n \times n$ Latin squares have at least $((1 - o(1))n/e^2)^n$ Hamilton transversals.*

Theorem 4.1.6 implies that a uniformly random proper $n$-arc-colouring satisfies Conjecture 4.1.4 with high probability, which in turn implies that a uniformly random $n \times n$ Latin square satisfies Conjectures 4.1.1 and 4.1.3 with high probability as well. We note that the number of optimal edge-colourings of $K_n$ is a vanishing fraction of the number of $n \times n$ Latin squares, so Theorem 4.1.6 does not imply the result of [56].

Random Latin squares can be difficult to analyze, in part due to their 'rigidity' and lack of independence. To prove Theorem 4.1.5, Kwan [90] – using Keevash's [77, 75] breakthrough results on the existence of combinatorial designs – developed a method for approximating a uniformly random Latin square by an outcome of the 'triangle-removal process', which is in comparison much easier to analyze. Prior to Kwan's [90] work, a limited number of results (e.g. [21, 93, 98, 123]) were proved using so-called 'switching' methods. Our proof, notably, does not rely on Keevash's [77, 75] results and instead introduces new techniques for analyzing 'switchings' to study Latin squares, thus providing a more elementary proof of Theorem 4.1.5.

### 4.1.3 Organization of the paper

In Section 4.2 we clarify some notation and definitions that we will use throughout the paper. We overview the proof of Theorem 4.1.6 in Section 4.3, and give some preliminary probabilistic results and useful theorems of other authors in Section 4.4. Sections 4.5–4.7 are devoted to the proof of Theorem 4.1.6.

## 4.2 Notation

For a natural number $n$ we define $[n] \coloneqq \{1, 2, \ldots, n\}$ and $[n]_0 \coloneqq [n] \cup \{0\}$. We say that a partition $\mathcal{P} = \{D_i\}_{i=1}^m$ of a finite set $D$ into $m$ parts is *equitable* if $|D_i| \in \{\lfloor |D|/m \rfloor, \lceil |D|/m \rceil\}$ for all $i \in [m]$, and when $|D|$ is large we assume that each part $D_i$ has the same size $|D|/m$, where this does not affect the argument.

For a digraph $G$, we write the arc set of $G$ as $E(G)$, and we denote an arc from a vertex $u$ to a vertex $v$ as $uv$, and we say that $u$ is the *tail* of the arc $e = uv$, denoted $u = \mathrm{tail}\,(e)$, and that $v$ is the *head* of $e$, denoted $v = \mathrm{head}\,(e)$. We say that any vertex $v$ such that $uv \in E(G)$ is an *out-neighbour* of $u$ in $G$, and that any $v$ such that $vu \in E(G)$ is an *in-neighbour* of $u$ in $G$. We define $N_G^+(v)$ to be the set of out-neighbours of $v$ in $G$, sometimes dropping the subscript $G$ when $G$ is clear from context, and we call $N_G^+(v)$ the *out-neighbourhood* of $v$ in $G$. We define the *in-neighbourhood* of $v$ in $G$, denoted $N_G^-(v)$, analogously, and we define the *neighbourhood* of $v$ in $G$ to be $N_G(v) \coloneqq N_G^+(v) \cup N_G^-(v)$. We define $d_G^+(v) \coloneqq |N_G^+(v)|$ and $d_G^-(v) \coloneqq |N_G^-(v)|$. For (not necessarily distinct) vertex sets $A, B \subseteq V(G)$ we define $E_G(A, B) \coloneqq \{ab \in E(G) \colon a \in A, b \in B\}$, and $e_G(A, B) \coloneqq |E_G(A, B)|$. Suppose now that $G$ is equipped with an arc-colouring $\phi_G$ in colour set $C$. Then for a colour $c \in C$ and an arc $e \in E(G)$ we write $\phi_G(e) = c$ to mean that $e$ has colour $c$ in the colouring $\phi_G$ of $G$. We frequently drop the notation $G$ when $G$ is clear from context. Further, if $\phi(e) = c$ then we say that $e$ is a $c$-arc, and in the case that $e$ is a loop we say that $e$ is a $c$-loop. We write $E_c(G)$ for the set of $c$-arcs in $G$ (including $c$-loops), and we refer to $E_c(G)$ as the *colour class* of $c$. Fix $u \in V(G)$. If $d \in C$ is such that there is a $d$-arc $uv$ in $G$, then the (unique) vertex $v$ is called the *$d$-out-neighbour* of $u$, which we denote by $N_d^+(u)$. We define the *$d$-in-neighbour* $N_d^-(u)$ of $u$ analogously. For $D \subseteq C$ we define $N_D^+(u) \coloneqq \{N_d^+(u) : d \in D\}$ and $N_D^-(u) \coloneqq \{N_d^-(u) : d \in D\}$, and for $A, B \subseteq V$ we define $E_{G,D}(A,B) \coloneqq \{e \in E_G(A,B) \colon \phi(e) \in D\}$ and $e_{G,D}(A,B) \coloneqq |E_{G,D}(A,B)|$. For a subdigraph $H \subseteq G$ we define $\phi_G(H) \coloneqq \{\phi_G(e) \colon e \in E(H)\}$.

With a slight abuse of notation, we often refer to a pair $(H, \phi)$ where $H$ is a digraph and $\phi$ is a proper arc-colouring of $H$ as a 'coloured digraph' $H$ implicitly equipped with a proper arc-colouring $\phi_H$ (or simply $\phi$ if it is clear from the context). Using this convention, we let $\Phi(\overleftrightarrow{K_n})$ denote the set of all properly $n$-arc-coloured digraphs $G \cong \overleftrightarrow{K_n}$ with vertex set and colour set $[n]$. (That is, the set of pairs $(G, \phi)$ where $G \cong \overleftrightarrow{K_n}$ and $\phi$ is a proper $n$-arc-colouring of $G$). For a coloured digraph $G \in \Phi(\overleftrightarrow{K_n})$ and a set of colours $D \subseteq [n]$ we define $G|_D$ to be the coloured digraph obtained by deleting all arcs of $G$ having colours not in $D$, and we set $\mathcal{G}_D^n \coloneqq \{G|_D \colon G \in \Phi(\overleftrightarrow{K_n})\}$, though we always drop the $n$ in the superscript as $n$ will be clear from context. By symmetry of the roles of rows, columns, and symbols in Latin squares, the correspondence between Latin squares and elements $G \in \Phi(\overleftrightarrow{K_n})$, and the well-known result that any Latin rectangle has a completion to a Latin square, it is clear that $\mathcal{G}_D$ could be equivalently defined as the set of all pairs $(H, \phi_H)$, where $H$ is a $|D|$-regular digraph on vertices $[n]$, and $\phi_H$ is a proper arc-colouring of $H$ in colours $D$. Throughout the paper we will use the letter $G$ for an element of $\Phi(\overleftrightarrow{K_n})$, and the letter $H$ for an element of $\mathcal{G}_D$ (for any $D$). We often write random variables and objects in bold notation. For an event $\mathcal{E}$ in any probability space we use the notation $\overline{\mathcal{E}}$ to denote the complement of $\mathcal{E}$.

## 4.3 Overview of the proof

The proof of Theorem 4.1.6 proceeds in two key steps. We first analyze uniformly random $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$ and show that with high probability, $\mathbf{G}$ satisfies three key properties. It then suffices to suppose that a fixed $G \in \Phi(\overleftrightarrow{K_n})$ satisfies these three properties, and use that hypothesis to build many rainbow directed Hamilton cycles in $G$. Before describing these properties, we discuss our strategy for building rainbow directed Hamilton cycles. To that end, we introduce the following definition.

**Definition 4.3.1.** A subgraph $H \subseteq G \in \Phi(\overleftrightarrow{K_n})$ is *robustly rainbow-Hamiltonian* (with respect to *flexible sets* $V_{\text{flex}} \subseteq V(H)$ and $C_{\text{flex}} \subseteq \phi(H)$ of vertices and colours, and *initial* vertex $u \in V(H)$ and *terminal* vertex $v \in V(H)$), if for any pair of equal-sized subsets $X \subseteq V_{\text{flex}}$ and $Y \subseteq C_{\text{flex}}$ of size at most $\min\{|V_{\text{flex}}|/2, |C_{\text{flex}}|/2\}$, the graph $H - X$ contains a rainbow directed Hamilton path with tail $u$ and head $v$, not containing a colour in $Y$.

We show that for almost all $G \in \Phi(\overleftrightarrow{K_n})$ and arbitrary sets $V_{\text{flex}}, C_{\text{flex}} \subseteq [n]$ of sizes $|V_{\text{flex}}| = |C_{\text{flex}}| = \Omega(n/\log^3 n)$, $G$ contains a robustly rainbow-Hamiltonian subgraph $H$ with flexible sets $V_{\text{flex}}$ and $C_{\text{flex}}$, such that $H$ has $O(n/\log^3 n)$ vertices and arcs in total. We construct rainbow directed Hamilton cycles by using the popular 'absorption' method, and $H$ will form the key absorbing structure. More

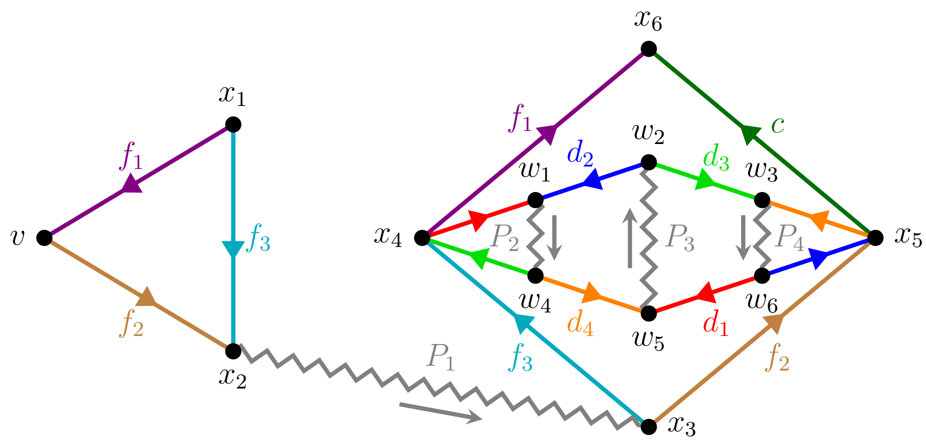



Figure 4.1: A $(v, c)$-absorber. Here $\phi(x_4w_1) = d_1$, $\phi(w_6x_5) = d_2$, $\phi(w_4x_4) = d_3$, and $\phi(x_5w_3) = d_4$. $P_1, \dots, P_4$ are rainbow directed paths with directions as indicated, sharing no colours with each other or with the rest of the $(v, c)$-absorber.

precisely, we find a rainbow directed path $P$ having the terminal vertex $v$ of $H$ as its tail, the initial vertex $u$ of $H$ as its head, such that $V(G) \setminus V(H) \subseteq V(P)$, $V(P) \cap V(H)$ is a subset of $V_{\text{flex}}$ of size at most $|V_{\text{flex}}/2|$, and likewise for the colours. Letting $X := V(P) \cap V(H)$ and $Y := \phi(P) \cap \phi(H)$, the robust rainbow-Hamiltonicity of $H$ guarantees there is a rainbow directed Hamilton path $P'$ in $H - X$ with tail $u$ and head $v$, not containing a colour in $Y$, and $P \cup P'$ is a rainbow directed Hamilton cycle.

We find $H$ by piecing together smaller building blocks we call 'absorbers' in a delicate way, where each absorber has the ability to 'absorb' a vertex $v$ and a colour $c$ not used by $P$. We delay a definition of a $(v, c)$-absorber to Definition 4.6.1, but we give a figure now (see Figure 4.1). Notice that a $(v, c)$-absorber has a rainbow directed Hamilton path with tail $x_1$ and head $x_6$, and a rainbow directed path with the same head and tail using all vertices except $v$ and all colours except $c$. This is the key property of a $(v, c)$-absorber, and by piecing these together in a precise way we ensure that the resulting union of absorbers (with the sets of specified vertices '$v$' and colours '$c$' forming $V_{\text{flex}}$ and $C_{\text{flex}}$ respectively) has the desired robustly rainbow-Hamiltonian property. For technical reasons, we find $(v, c)$-absorbers by piecing together two smaller structures we call $(v, c)$-*absorbing gadgets* and $(y, z)$-*bridging gadgets* (see Definitions 4.5.1 and 4.5.2, respectively), together with the short rainbow directed paths $P_1$, $P_2$, $P_3$, $P_4$ as in Figure 4.1.

Thus, the first key property that we need almost all $G \in \Phi(\overleftrightarrow{K_n})$ to satisfy, is that $G$ contains many absorbing gadgets and bridging gadgets, in a 'well-spread' way that enables us to construct an appropriate robustly rainbow-Hamiltonian subgraph. We prove this in Section 4.5 using 'switchings' in Latin rectangles, then using permanent estimates (see [17, 38, 42], encapsulated by Proposition 4.4.4 in the current paper) to compare a uniformly random $k \times n$ Latin rectangle to the first $k$ rows of a uniformly random $n \times n$ Latin square. Lemma 4.5.6 ensures the existence of the absorbing gadgets we need, and Lemma 4.5.10 accomplishes the same for the bridging gadgets. This approach of using permanent estimates to translate statements between these probability spaces was pioneered by McKay and Wanless [98], who investigated the typical prevalence of $2 \times 2$ Latin subsquares (also called 'intercalates') in a uniformly random Latin square. For further insight into the usage of this method to study intercalates in random Latin squares, see for example [93, 91, 92]. As the substructures we seek are more complex than intercalates, and we moreover require that they are 'well-spread', our proof introduces new techniques for switching arguments in Latin rectangles. We note that in [56], the authors, with Kühn and Osthus, used switching arguments to analyze a uniformly random 1-factorization of $K_n$ and show that with high probability there is a large collection of subgraphs of a form analogous to that of our $(v, c)$-absorbing gadgets in the undirected setting. Fortunately, this argument also works in the directed setting with only minor changes, so we defer the proof of Lemma 4.5.6 to the appendix (appearing in the present thesis as Section 4.8). Thus, Section 4.5 is primarily devoted to the proof of Lemma 4.5.9.

The second property of almost all $G \in \Phi(\overleftrightarrow{K_n})$ that we will need concerns the colours of the loops. Clearly, if we seek to find any rainbow directed Hamilton cycle of $G \in \Phi(\overleftrightarrow{K_n})$, we need to know that there is no colour appearing only on loops in $G$, and this is given for almost all $G \in \Phi(\overleftrightarrow{K_n})$ (in the context of Latin squares and in considerably stronger form) by Lemma 4.4.6.

The third and final property of almost all $G \in \Phi(\overleftrightarrow{K_n})$ that we will need is an appropriate notion of 'lower-quasirandomness', which roughly states that for any two subsets $U_1, U_2$ of vertices of $G$ and any set $D$ of colours, the number of arcs in $G$ with tail in $U_1$, head in $U_2$, and colour in $D$, is close to what we would expect if the colours of the arcs of $G$ were assigned independently and uniformly at random. We delay the precise definition of lower-quasirandomness of $G \in \Phi(\overleftrightarrow{K_n})$ to Definition 4.7.1. The desired property that almost all $G \in \Phi(\overleftrightarrow{K_n})$ are lower-quasirandom will follow immediately from [93, Theorem 2] (see Theorem 4.4.7 of the current paper), originally stated in the context of 'discrepancy' of random Latin squares.

Armed with the three properties of typical $G \in \Phi(\overleftrightarrow{K_n})$ described above, it then suffices to fix such a $G$ and build many rainbow directed Hamilton cycles. In Section 4.6, we show that the existence of many well-spread absorbing and bridging gadgets enables us to greedily build a small robustly rainbow-Hamiltonian subgraph $H \subseteq G$ with arbitrary flexible sets $V_{\text{flex}}$ and $C_{\text{flex}}$ of size $\Theta(n/\log^3 n)$, and in Section 4.7, we use this to prove Theorem 4.1.6. The rough idea is to first choose the flexible sets $V_{\text{flex}}$ and $C_{\text{flex}}$ randomly. Next, we use the lower-quasirandomness property of $G$ to build a rainbow directed spanning path forest $Q$ of $G - H$, one arc at a time, until $Q$ has very few components. Then, we use the random choice of $V_{\text{flex}}$ and $C_{\text{flex}}$, together with Lemma 4.4.6, to find short rainbow directed paths linking the components of $Q$ and the designated start and end of $H$, which use all remaining colours of $G - H$, and at most half of $V_{\text{flex}}$ and $C_{\text{flex}}$. Finally, we use the key robustly rainbow-Hamiltonian property of $H$ to absorb the remaining vertices and colours in $V_{\text{flex}}$ and $C_{\text{flex}}$ as described above, completing the rainbow directed Hamilton cycle of $G$. To obtain the counting result on the number of rainbow directed Hamilton cycles in $G$, it suffices to count the number of choices we can make whilst building the rainbow directed spanning path forest $Q$ of $G - H$.

We remark that this particular absorption strategy, wherein we create an absorbing structure with 'flexible' sets, is an instance of the 'distributive absorption' method, which was introduced by Montgomery [101] in 2018 and has been found to have several applications since. In particular, this method is also used in [90] and [56] to find transversals in random Latin squares and rainbow Hamilton paths in random 1-factorizations, respectively. Our approach differs from that of [90] and [56] in a few key ways. First, the 'asymmetry' of proper $n$-arc-colourings of $\overleftrightarrow{K_n}$ (in comparison to proper edge-colourings of $K_n$ with at most $n$ colours, which correspond to proper $n$-arc-colourings of $\overleftrightarrow{K_n}$ with monochromatic digons) and 'connectedness' of rainbow Hamilton cycles/ Hamilton transversals (in comparison to general transversals in Latin squares) necessitate a more complex absorbing structure than the one of either [90] or [56], which is more challenging to create and construct. Nevertheless, as mentioned, we show that switching arguments are sufficient for finding our absorbing structure, yielding a more elementary proof than that of [90], and moreover, by choosing our flexible sets randomly, we avoid complications involving vertices with few out- or in-neighbours in $V_{\text{flex}}$ on arcs with colour in $C_{\text{flex}}$, providing a further simplification of the approach in [90]. In [56], results [104, 8] on nearly perfect matchings in nearly regular hypergraphs are applied to auxiliary hypergraphs to construct both the absorbing structure and a nearly spanning rainbow path in a random 1-factorization of $K_n$, but since the absorbers we use here (minus the internal vertices of the linking paths $P_1, \ldots, P_4$) are not regular, the analogous approach fails in the directed setting (as the corresponding auxiliary hypergraphs are not regular). However, as we show, the 'lower-quasirandomness' of typical $G \in \Phi(\overleftrightarrow{K_n})$ is enough for us to find $Q$, the nearly spanning rainbow path forest, without these hypergraph matching results, and our absorbing structure is robust enough to augment it to a rainbow directed path.

## 4.4 Preliminaries

In this brief section we state some results that we will use in the proof of Theorem 4.1.6. We begin with a well-known concentration inequality for independent random variables.

Let $X_1, \ldots, X_m$ be independent random variables taking values in $\mathcal{X}$, and let $f : \mathcal{X}^m \to \mathbb{R}$. If for all $i \in [m]$ and $x_i', x_1, \ldots, x_m \in \mathcal{X}$, we have

$$|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_m)| \leq c_i,$$

then we say $X_i$ *affects* $f$ by at most $c_i$.

**Theorem 4.4.1** (McDiarmid's Inequality [97]). *If $X_1, \ldots, X_m$ are independent* *random variables taking values in $\mathcal{X}$ and $f : \mathcal{X}^m \to \mathbb{R}$ is such that $X_i$ affects $f$ by at most $c_i$ for all $i \in [m]$, then for all $t > 0$,*

$$\mathbb{P}\left[|f(X_1, \ldots, X_m) - \mathbb{E}\left[f(X_1, \ldots, X_m)\right]| \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^m c_i^2}\right).$$

Next, we need the notion of 'robustly matchable' bipartite graphs, which will form a key part of our absorption argument.

**Definition 4.4.2.** Let $T$ be a bipartite graph with bipartition $(A, B)$ such that $|A| = |B|$.

- We say $T$ is *robustly matchable* with respect to *flexible sets* $A' \subseteq A$ and $B' \subseteq B$, if for every pair of equal-sized subsets $X \subseteq A'$ and $Y \subseteq B'$ of size at most $\min\{|A'|/2, |B'|/2\}$, there is a perfect matching in $T - (X \cup Y)$.
- For $m \in \mathbb{N}$, we say $T$ is a $2RMBG(7m, 2m)$ if $|A| = |B| = 7m$ and $T$ is robustly matchable with respect to flexible sets $A' \subseteq A$ and $B' \subseteq B$ where $|A'| = |B'| = 2m$.

The concept of using robustly matchable bipartite graphs in absorption arguments was first introduced by Montgomery [101]. We need the following observation of the authors, Kühn, and Osthus [56, Lemma 4.5], which is based on the work of Montgomery.

**Lemma 4.4.3** (Gould, Kelly, Kühn, and Osthus [56])**.** *For all sufficiently large $m$, there is a $2RMBG(7m, 2m)$ that is $256$-regular.*

For a coloured digraph $H \in \mathcal{G}_D$, we define $\mathrm{comp}(H)$ to be the number of distinct ways to complete $H$ to an element $G \in \Phi(\overleftrightarrow{K_n})$, or more precisely the number of $H' \in \mathcal{G}_{[n]\setminus D}$ having $E(H) \cap E(H') = \emptyset$ (and therefore $E(H) \cup E(H') = E(\overleftrightarrow{K_n})$).

We will use the following proposition to compare the probabilities of events in the probability spaces corresponding to uniformly random $\mathbf{H} \in \mathcal{G}_D$ (for some small $D \subseteq [n]$) and uniformly random $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$ (see for example the proof of Lemma 4.5.10).

**Proposition 4.4.4.** *For any $D \subseteq [n]$ and $H, H' \in \mathcal{G}_D$ we have*

$$\frac{\mathrm{comp}(H)}{\mathrm{comp}(H')} \leq \exp(O(n \log^2 n)).$$

Proposition 4.4.4 follows immediately from (for example) [93, Proposition 5] as $\mathcal{G}_D$ can easily be seen to be equivalent to the set of $|D| \times n$ Latin rectangles.

Next, we show (in the context of Latin squares) that a uniformly random $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$ does not have too many loops of a fixed colour. We first need the following well-known result on the number of fixed points of a random permutation.

**Lemma 4.4.5.** *Let $\boldsymbol{\sigma}$ be a uniformly random permutation of $[n]$, and let $\mathbf{X}$ denote the number of fixed points of $\boldsymbol{\sigma}$. Then, for $k \in [n]_0$, we have $\mathbb{P}[\mathbf{X} = k] = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}$.*

**Lemma 4.4.6.** *Let $\mathbf{L}$ be a uniformly random $n \times n$ Latin square with entries in $[n]$, and suppose $t \geq 3 \log n / \log\log n$. Let $\mathbf{X}$ be the random variable which returns the maximum (over the symbol set $[n]$) number of times that any symbol appears on the leading diagonal, in $\mathbf{L}$. Then $\mathbb{P}[\mathbf{X} \geq t] \leq \exp(-\Omega(t \log t))$.*

**Proof.** Let $\mathcal{L}_n$ be the set of $n \times n$ Latin squares with symbols $[n]$, and for $L, L' \in \mathcal{L}_n$, write $L \sim L'$ if $L'$ can be obtained from $L$ via a permutation of the rows. Clearly, $\sim$ is an equivalence relation on $\mathcal{L}_n$. Note that $\mathbf{L}$ can be obtained by first choosing an equivalence class $\mathbf{S} \in \mathcal{L}_n/{\sim}$ uniformly at random and then choosing $\mathbf{L} \in \mathbf{S}$ uniformly at random. We actually prove the stronger statement that for every equivalence class $S \in \mathcal{L}_n/\sim$, if $\mathbf{L} \in S$ is chosen uniformly at random, then $\mathbb{P}[\mathbf{X} \geq t] \leq \exp(-\Omega(t \log t))$.

Each equivalence class $S \in \mathcal{L}_n/\sim$ has size $n!$ and contains a unique representative $L_{S,i}$ with every symbol on the leading diagonal being $i$, for each $i \in [n]$. Applying a uniformly random row permutation $\boldsymbol{\sigma}$ to $L_{S,i}$ yields a uniformly random element $\mathbf{L}$ of $S$, and the number of appearances $\mathbf{X}_i$ of $i$ on the leading diagonal of $\mathbf{L}$ is equal to the number of fixed points of $\boldsymbol{\sigma}$. Then, if $t \geq 3 \log n / \log \log n$ and $n$ is sufficiently large, we have by Lemma 4.4.5 and Stirling's formula that

$$\mathbb{P}[\mathbf{X}_i \geq t] = \sum_{k=t}^{n} \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \leq \sum_{k=t}^{n} \frac{1}{k!} \leq \frac{n}{t!} \leq \exp\left(-\frac{1}{2} t \log t\right),$$

where we have used the simple observation that $\sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \leq 1$ for all $k \in [n]_0$. A union bound over symbols $i \in [n]$ now completes the proof. □

Finally, we need the following theorem of Kwan and Sudakov [93, Theorem 2], originally stated in the context of 'discrepancy' of random Latin squares. Theorem 4.4.7 ensures in particular that almost all $G \in \Phi(\overleftrightarrow{K_n})$ are 'lower-quasirandom' (see Definition 4.7.1), which we will use when building and counting the almost-spanning rainbow directed path forests (see Lemma 4.7.2) that we later absorb into rainbow directed Hamilton cycles.

**Theorem 4.4.7** (Kwan and Sudakov [93]). *Let $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$ be chosen uniformly at random. Then with high probability, for all (not necessarily distinct) sets*

$U_1, U_2, D \subseteq [n]$, *we have that*

$$\left| e_{G,D}(U_1, U_2) - \frac{|U_1||U_2||D|}{n} \right| = O\left( \sqrt{|U_1||U_2||D|} \log n + n \log^2 n \right).$$

## 4.5 Absorbers via switchings

The aim of this section is to prove that almost all $G \in \Phi(\overleftrightarrow{K_n})$ have many well-distributed absorbing gadgets and bridging gadgets, which we define now (see also Figure 4.2).

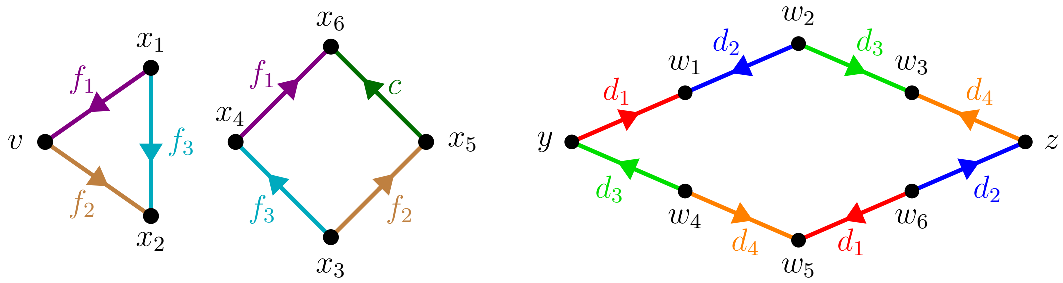
**Definition 4.5.1.** For a vertex $v$ and a colour $c$, a $(v, c)$-*absorbing gadget* is a digraph $A$ having vertex set $V(A) = \{v, x_1, x_2, \ldots, x_6\}$ and arcs $E(A) = \{x_1v, vx_2, x_1x_2, x_3x_4, x_3x_5, x_4x_6, x_5x_6\}$, equipped with a proper arc-colouring $\phi_A$, such that the following holds:

- $\phi_A(x_1v) = \phi_A(x_4x_6) =: f_1$;
- $\phi_A(vx_2) = \phi_A(x_3x_5) =: f_2$;
- $\phi_A(x_1x_2) = \phi_A(x_3x_4) =: f_3$;
- $\phi_A(x_5x_6) = c$;
- the colours $f_1, f_2, f_3, c$ are distinct.

In this case, we say $(x_4, x_5)$ is the pair of *abutment vertices* of $A$.

**Definition 4.5.2.** For distinct vertices $y$ and $z$, a $(y, z)$-*bridging gadget* is a digraph $B$ with $V(B) = \{y, z, w_1, w_2, \ldots, w_6\}$ and arcs $yw_1$, $w_2w_1$, $w_2w_3$, $zw_3$, $w_4y$, $w_4w_5$, $w_6w_5$, $w_6z$, equipped with a proper arc-colouring $\phi_B$, such that the following holds:

- $\phi_B(yw_1) = \phi_B(w_6w_5) =: d_1$;




(a) A $(v, c)$-absorbing gadget.

(b) A $(y, z)$-bridging gadget.

Figure 4.2: The key building blocks for the absorbing structure we build in Section 4.6.

- $\phi_B(w_2w_1) = \phi_B(w_6z) =: d_2$;
- $\phi_B(w_4y) = \phi_B(w_2w_3) =: d_3$;
- $\phi_B(w_4w_5) = \phi_B(zw_3) =: d_4$;
- the colours $d_1, \ldots, d_4$ are distinct.

As discussed in Section 4.3, the union of a $(v, c)$-absorbing gadget and an $(x_4, x_5)$-bridging gadget (together with some short rainbow directed paths) forms a structure we will call a $(v, c)$-*absorber* (see Figure 4.1 and Definition 4.6.1), which is the key building block of our absorption structure. To show that almost all $G \in \Phi(\overleftrightarrow{K_n})$ contain the gadgets we need, we analyze switchings in the probability space corresponding to uniformly random $\mathbf{H} \in \mathcal{G}_D$ (recall that $\mathcal{G}_D$ is the set of digraphs obtained from the digraphs in $\Phi(\overleftrightarrow{K_n})$ by deleting all arcs with colour not in $D$) for small $D \subseteq [n]$, before applying Proposition 4.4.4 to compare this probability space with that of uniformly random $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$ (see the proof of Lemma 4.5.10).

First, we need the following lemma, which asserts that for small $D \subseteq [n]$, a uniformly random $\mathbf{H} \in \mathcal{G}_D$ does not have too many more arcs than we would expect between any pair of vertex sets, each of size $|D|$.

**Definition 4.5.3.** For $D \subseteq [n]$, we say that $H \in \mathcal{G}_D$ is *$\ell$-upper-quasirandom* if $e_H(A, B) \leq (1+\ell)|D|^3/n$ for all (not necessarily distinct) vertex sets $A, B \subseteq V(H)$ of sizes $|A| = |B| = |D|$. We define $\mathcal{Q}_D^\ell \coloneqq \{H \in \mathcal{G}_D \colon H \text{ is } \ell\text{-upper-quasirandom}\}$.

For a colour $c \in D$ and uniformly random $\mathbf{H} \in \mathcal{G}_D$, we write $\mathbf{F}_c = F_c(\mathbf{H})$ for the random colour class of $c$ in $\mathbf{H}$ ($F$ here standing for 'factor'), so that $\mathbf{H}$ is determined by the random variables $\{\mathbf{F}_c\}_{c\in D}$.

**Lemma 4.5.4.** *Suppose $D \subseteq [n]$ has size $|D| = n/10^6$. Fix $c \in D$, let $\mathbf{H} \in \mathcal{G}_D$ be chosen uniformly at random, and let $\mathbf{F}_c = F_c(\mathbf{H})$. Then for any outcome $F$ of $\mathbf{F}_c$ we have*

$$\mathbb{P}\left[\mathbf{H} \in \mathcal{Q}_D^1 \mid \mathbf{F}_c = F\right] \geq 1 - \exp(-\Omega(n^2)).$$

The authors of [56] proved a lemma ([56, Lemma 6.3]) analogous to Lemma 4.5.4 in the undirected setting. The proof of Lemma 4.5.4 is similar so we omit it here. In the appendix (appearing as Section 4.8 in the present thesis), we describe how the proof of [56, Lemma 6.3] can be modified to obtain a proof of Lemma 4.5.4.

We condition on versions of upper-quasirandomness when we are using switching arguments to show that almost all $H \in \mathcal{G}_D$ admit many absorbing gadgets and bridging gadgets. Further, we will need that $H$ does not have many $c$-loops in order to find many $(v, c)$-absorbing gadgets, for any $v \in V(H)$. Lemma 4.5.4 enables us to 'uncondition' from these two events, so as to study simply the probability that a uniformly random $\mathbf{H}$ has many absorbing gadgets.

Since, as discussed in Section 4.3, we eventually piece together gadgets in a greedy fashion to build an absorbing structure in a typical $G \in \Phi(\overleftrightarrow{K_n})$, it will be important to know that we can find collections $\mathcal{A}$ of gadgets which are 'well-spread', in that no vertex or colour of $G$ is contained in too many $A \in \mathcal{A}$. We formalise this notion in the following definition.

**Definition 4.5.5.** Suppose that $G$ is an $n$-vertex directed, arc-coloured digraph with vertices $V$ and colours $C$. Fix $v \in V$, $c \in C$, and fix $y, z \in V$ distinct. We say that a collection $\mathcal{A}$ of $(v,c)$-absorbing gadgets in $G$ is *well-spread* if for all $u \in V \setminus \{v\}$ and $d \in C \setminus \{c\}$, there are at most $n$ distinct $A \in \mathcal{A}$ which contain $u$, and at most $n$ distinct $A \in \mathcal{A}$ which contain $d$. We say that a collection $\mathcal{B}$ of $(y,z)$-bridging gadgets in $G$ is *well-spread* if for all $u \in V \setminus \{y, z\}$ and $d \in C$, there are at most $n$ distinct $B \in \mathcal{B}$ which contain $u$, and at most $n$ distinct $B \in \mathcal{B}$ which contain $d$.

The next lemma ensures that almost all $G \in \Phi(\overleftrightarrow{K_n})$ contain the collections of well-spread absorbing gadgets that we need.

**Lemma 4.5.6.** *Let* $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$ *be chosen uniformly at random, and let* $\mathcal{E}$ *be the event that for all* $v, c \in [n]$, $\mathbf{G}$ *contains a well-spread collection of at least* $n^2/2^{100}$ $(v,c)$*-absorbing gadgets. Let* $\mathcal{C}$ *be the event that no colour class of* $\mathbf{G}$ *has more than* $n/10^9$ *loops. Then* $\mathbb{P}[\mathcal{E} \mid \mathcal{C}] \geq 1 - \exp(-\Omega(n^2))$*, and in particular,* $\mathbb{P}[\mathcal{E}] \geq 1 - \exp(-\Omega(n \log n))$ *by Lemma 4.4.6.*

As with Lemma 4.5.4, the authors of [56] proved an analogous lemma ([56, Lemma 3.8]) in the undirected setting with a similar proof, so we omit it here but provide details in the appendix (appearing as Section 4.8 in the present thesis) of how the proof of [56, Lemma 3.8] may be modified to prove Lemma 4.5.6.

The rest of this section is dedicated to showing that almost all $G \in \Phi(\overleftrightarrow{K_n})$ have large well-spread collections of bridging gadgets (recall Figure 4.2b). For technical reasons that make the switching argument a little easier to analyze, we instead actually look for a slightly more special structure. In particular, we add some extra

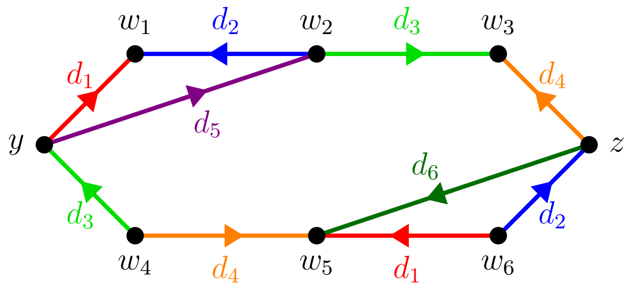



Figure 4.3: A $(y, z, \mathcal{P})$-bridge, with $\mathcal{P} = (D_i)_{i=1}^6$ and $d_i \in D_i$ for each $i \in [6]$.

arcs so that all vertices we find are in the neighbourhood of $y$ or of $z$, we partition the colours to limit the number of 'roles' certain arcs can play when we apply the switching operation, and we introduce the notion of *distinguishability*, which will be useful when arguing that the gadgets we find are well-spread.

**Definition 4.5.7.** Let $D \subseteq [n]$, let $H \in \mathcal{G}_D$, and let $\mathcal{P} = (D_i)_{i=1}^6$ be an equitable (ordered) partition of $D$ into six parts. Let $y, z \in [n]$ be distinct vertices.

- We say that a subgraph $B \subseteq H$ is a $(y, z, \mathcal{P})$-*bridge* (see Figure 4.3) if $B$ is the union of a $(y, z)$-bridging gadget $B'$ (with vertex- and colour-labelling as in Definition 4.5.2) and the extra arcs $yw_2$, $zw_5$, such that $d_i \in D_i$ for all $i \in [4]$, $\phi_H(yw_2) \in D_5$, and $\phi_H(zw_5) \in D_6$;
- we say that a $(y, z, \mathcal{P})$-bridge $B$ is *distinguishable* in $H$ if $B$ is the only $(y, z, \mathcal{P})$-bridge in $H$ containing any of the arcs $w_2w_1, w_2w_3, w_4w_5, w_6w_5$;
- we write $r_{(y,z,\mathcal{P})}(H)$ for the number of distinguishable $(y, z, \mathcal{P})$-bridges in $H$;
- for $s \in [n|D|]_0$, we write $M_s^{(y,z,\mathcal{P})}$ for the set of $H \in \mathcal{G}_D$ such that $r_{(y,z,\mathcal{P})}(H) = s$ and $e_H(A, B) \leq 2|D|^3/n + 12s$ for all $A, B \subseteq [n]$ of size $|A| = |B| = |D|$, and we define $\widehat{\mathcal{Q}}_D^{(y,z,\mathcal{P})} := \bigcup_{s=0}^{n|D|} M_s^{(y,z,\mathcal{P})}$.

We frequently drop the $(y, z, \mathcal{P})$-notation in the terminology introduced above when the tuple $(y, z, \mathcal{P})$ is clear from context. For every distinct $y, z \in [n]$ and equitable partition $\mathcal{P} = (D_i)_{i=1}^6$,

$$\mathcal{Q}_D^1 \subseteq \widehat{\mathcal{Q}}_D, \text{ and if } s \leq |D|^4/(10^{24}n^2), \text{ then } M_s \subseteq \mathcal{Q}_D^2. \tag{4.5.1}$$

In Lemma 4.5.9 we use switchings on some $H \in \mathcal{G}_D$ to produce some $H' \in \mathcal{G}_D$ having $r(H') = r(H) + 1$. As mentioned earlier, we condition on upper-quasirandomness in this lemma; more specifically, we will condition that $\mathbf{H} \in \widehat{\mathcal{Q}}_D$. The notion of distinguishability of $(y, z, \mathcal{P})$-bridges is useful because, as we show in Lemma 4.5.10, Claim 1, a collection of distinguishable $(y, z, \mathcal{P})$-bridges in $G|_D$ is necessarily well-spread (recall Definition 4.5.5).

We now discuss the switching operation that forms the backbone of the proof of Lemma 4.5.9.

**Definition 4.5.8.** Let $D \subseteq [n]$, let $H \in \mathcal{G}_D$, let $\mathcal{P} = (D_i)_{i=1}^6$ be a partition of $D$ and suppose $y, z \in [n]$ are distinct. Let $u_1, \ldots, u_6, u_1', \ldots, u_8', u_1'', \ldots, u_8'' \in [n] \setminus \{y, z\}$, where $u_1, \ldots, u_6, u_1'', \ldots, u_8''$ are distinct and $\{u_1', \ldots, u_8'\} \cap \{u_1, \ldots, u_6, u_1'', \ldots, u_8''\} = \emptyset$. Let $U^{\text{int}} := \{u_1, u_2, \ldots, u_6\}, U^{\text{mid}} := \{u_1', u_2', \ldots, u_8'\}, U^{\text{ext}} := \{u_1'', u_2'', \ldots, u_8''\}$. Then we say that a subgraph $T \subseteq H[\{y, z\} \cup U^{\text{int}} \cup U^{\text{mid}} \cup U^{\text{ext}}]$ is a *twist system* (see Figure 4.4) of $H$ if:

(i) $E(T) = \{yu_1, yu_2, u_4y, zu_3, zu_5, u_6z, u_1'u_1, u_2u_2', u_2u_3', u_4'u_3, u_4u_5', u_6'u_5, u_7'u_5, u_6u_8', u_2''u_1'', u_3''u_4'', u_5''u_6'', u_8''u_7''\}$;

(ii) $\phi_H(yu_1) = \phi_H(u_7'u_5) = \phi_H(u_8''u_7'') = \phi_H(u_6u_8') \in D_1$;

(iii) $\phi_H(u_6z) = \phi_H(u_2u_2') = \phi_H(u_2''u_1'') = \phi_H(u_1'u_1) \in D_2$;

(iv) $\phi_H(u_4y) = \phi_H(u_2u_3') = \phi_H(u_3''u_4'') = \phi_H(u_4'u_3) \in D_3$;

(v) $\phi_H(zu_3) = \phi_H(u_4u_5') = \phi_H(u_5''u_6'') = \phi_H(u_6'u_5) \in D_4$;

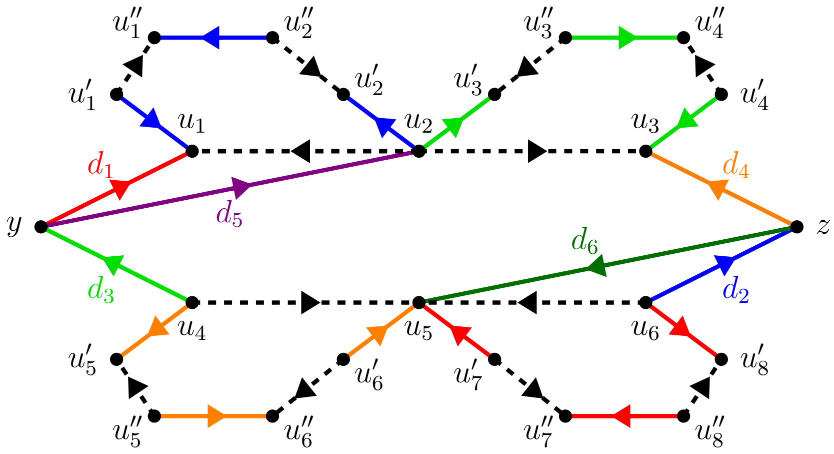



Figure 4.4: A twist system. Here, $d_i \in D_i$ for each $i \in [6]$, and dashed arcs indicate an arc which is absent in $H$.

(vi) $\phi_H(yu_2) \in D_5$ and $\phi_H(zu_5) \in D_6$;

(vii) $u_2u_1, u_1'u_1'', u_2''u_2', u_2u_3, u_3''u_3', u_4'u_4'', u_4u_5, u_5''u_5', u_6'u_6'', u_6u_5, u_7'u_7'', u_8''u_8' \notin E(H)$.

For a twist system $T \subseteq H$, we define $\text{twist}_T(H)$ to be the coloured digraph obtained from $H$ by deleting the arcs $u_1'u_1$, $u_2''u_1''$, $u_2u_2'$, $u_2u_3'$, $u_3''u_4''$, $u_4'u_3$, $u_4u_5'$, $u_5''u_6''$, $u_6'u_5$, $u_6u_8'$, $u_8''u_7''$, $u_7'u_5$, and adding the arcs $u_6u_5$, $u_7'u_7''$, $u_8''u_8'$ each in colour $\phi_H(yu_1)$, the arcs $u_4u_5$, $u_6'u_6''$, $u_5''u_5'$ each in colour $\phi_H(zu_3)$, the arcs $u_2u_3$, $u_4'u_4''$, $u_3''u_3'$ each in colour $\phi_H(u_4y)$, and the arcs $u_2u_1$, $u_1'u_1''$, $u_2''u_2'$ each in colour $\phi_H(u_6z)$. The $(y, z, \mathcal{P})$-bridge in $\text{twist}_T(H)$ with arc set $\{yu_1, yu_2, u_2u_1, u_2u_3, zu_3, u_6z, zu_5, u_6u_5, u_4u_5, u_4y\}$ is called the *canonical* $(y, z, \mathcal{P})$*-bridge* of the twist.

Notice that if $H \in \mathcal{G}_D$ and $T$ is a twist system of $H$, then $\text{twist}_T(H) \in \mathcal{G}_D$, even if $u_1', \ldots, u_8'$ are not distinct. We now use the twist switching operation to argue that almost all $H \in \widehat{\mathcal{Q}}_D$ have many distinguishable $(y, z, \mathcal{P})$-bridges, for fixed $y, z, \mathcal{P}$, and appropriately sized $D$.

**Lemma 4.5.9.** *Suppose $D \subseteq [n]$ has size $|D| = n/10^6$. Let $y, z \in [n]$ be distinct,* *and let $\mathcal{P} = (D_i)_{i=1}^6$ be an equitable partition of $D$. Let $\mathbf{H} \in \mathcal{G}_D$ be chosen uniformly at random. Then*

$$\mathbb{P}\left[r(\mathbf{H}) \leq \frac{n^2}{10^{50}} \,\middle|\, \mathbf{H} \in \widehat{\mathcal{Q}}_D\right] \leq \exp\left(-\Omega\left(n^2\right)\right).$$

**Proof.** Let $k := |D|$. Recall that $M_1, \ldots, M_{nk}$ is a partition of $\widehat{\mathcal{Q}}_D$ (see Definition 4.5.7). For each $s \in [nk-1]_0$ we define an auxiliary bipartite digraph $B_s$ with vertex bipartition $(M_s, M_{s+1})$ by putting an arc $HH'$ whenever $H \in M_s$ contains a twist system $T$ for which the canonical $(y, z, \mathcal{P})$-bridge of the twist is distinguishable in $\text{twist}_T(H) =: H'$ and $H' \in M_{s+1}$. Define $\delta_s^+ := \min_{H \in M_s} d_{B_s}^+(H)$ and $\Delta_{s+1}^- := \max_{H' \in M_{s+1}} d_{B_s}^-(H')$, and note that $|M_s|/|M_{s+1}| \leq \Delta_{s+1}^-/\delta_s^+$. We will show that $|M_s|/|M_{s+1}| \leq 1/10$, if $M_s$ is non-empty. To that end, we first obtain an upper bound for $\Delta_{s+1}^-$. Fix $H' \in M_{s+1}$. There are $s+1$ choices of a distinguishable $(y, z, \mathcal{P})$-bridge $B$ in $H'$ which could have been the canonical $(y, z, \mathcal{P})$-bridge of a twist of a graph $H \in M_s$ producing $H'$. There are then at most $n^8$ choices for the eight additional arcs added by a twist whose canonical bridge is $B$ since the colours of these arcs are determined by $B$ and there are $n$ arcs of each colour in $H'$. For any such sequence of choices, there is a unique $H \in M_s$ and twist system $T \subseteq H$ such that $twist_T(H) = H'$, so we determine that $\Delta_{s+1}^- \leq (s+1)n^8$, for all $s \in [nk-1]_0$.

We now find a lower bound for $\delta_s^+$, in the case where $s \leq k^4/(10^{24}n^2)$. Fix $H \in M_s$. We proceed by finding a large collection $\mathcal{T}$ of distinct twist systems in $H$, such that for each $T \in \mathcal{T}$, the canonical $(y, z, \mathcal{P})$-bridge of the twist is distinguishable in $\text{twist}_T(H) =: H'$ and $H' \in M_{s+1}$. We do this by ensuring that for any $T \in \mathcal{T}$, the arc deletions involved in twisting on $T$ do not decrease $r(H)$, and that the only $(y, z, \mathcal{P})$-bridge created by the arc additions involved in twisting on $T$ is the canonical $(y, z, \mathcal{P})$-bridge $B$ of the twist (whence $B$ is evidently distinguishable in $H'$). We first use the assumption on $s$ to argue that $H$ is not far from being upper-quasirandom (as per Definition 4.5.3). Indeed, since $H \in M_s$ and $s \le k^4/(10^{24}n^2)$ we have by Definition 4.5.7 that for any sets $W_1, W_2 \subseteq [n]$ of sizes $|W_1| = |W_2| = k$,

$$e_H(W_1, W_2) \le \frac{2k^3}{n} + 12s \le \frac{3k^3}{n}. \tag{4.5.2}$$

We now use (4.5.2) to find a large set $\Lambda$ of choices for a sequence of colours and arcs $\lambda = (d_1, d_2, \dots, d_6, e_1, \dots, e_4)$ with $d_i \in D_i$, and $e_i \in E_{d_i}(H)$, such that $\lambda$ has a number of desirable properties. We simultaneously use such a sequence $\lambda$ to choose vertices $u_1, \dots, u_6, u'_1, \dots, u'_8,\ u''_1, \dots, u''_8$ as in Definition 4.5.8 and a subgraph $T_\lambda \subseteq H[\{y,z\} \cup U_{\text{int}} \cup U^{\text{mid}} \cup U^{\text{ext}}]$, thus constructing a set $\mathcal{T}$ of such $T_\lambda$ by ranging over all $\lambda \in \Lambda$. We will then use the known properties of the sequences $\lambda \in \Lambda$ to verify that each $T_\lambda \in \mathcal{T}$ is a twist system for which the arc $H\text{twist}_{T_\lambda}(H)$ is in $B_s$.

*Claim 1: There is a set $D_{1,2}^{good}$ of pairs $(d_1, d_2) \in D_1 \times D_2$ such that $|D_{1,2}^{good}| \ge k^2/100$ and each $(d_1, d_2) \in D_{1,2}^{good}$ satisfies the following, where $u_1 \coloneqq N^+_{d_1}(y)$, $u'_1 \coloneqq N^-_{d_2}(u_1)$, $u_6 \coloneqq N^-_{d_2}(z)$, $u'_8 \coloneqq N^+_{d_1}(u_6)$.*

($D_1$1) *There are at most $10^8$ loops with colour $d_1$ in $H$;*

($D_1$2) *$u_1$ has at most $300k^2/n$ in-neighbours in the set $N^+_{D_5}(y)$;*

($D_1$3) *there are at most $k/100$ arcs $e$ coloured $d_1$ in $H$ such that $e$ is contained in a distinguishable $(y, z, \mathcal{P})$-bridge in $H$;*

($D_2$1) *there are at most $10^8$ loops with colour $d_2$ in $H$;*

($D_2$2) *$u_6$ has at most $300k^2/n$ out-neighbours in the set $N^+_{D_6}(z)$;*

($D_2$3) *there are at most $k/100$ arcs $e$ coloured $d_2$ in $H$ such that $e$ is contained in a distinguishable $(y, z, \mathcal{P})$-bridge in $H$;*

($V_{1,2}$) *the vertices* $y, z, u_1, u_6$ *are distinct, and* $u_1', u_8' \notin \{y, z, u_1, u_6\}$*;*

($R_{1,2}$) *there is no distinguishable* $(y, z, \mathcal{P})$*-bridge in* $H$ *containing the arc* $u_1'u_1$ *or the arc* $u_6u_8'$*.*

*Proof of claim:* For $i \in [3]$, let $D_{1,i}$ be the set of colours $d_1 \in D_1$ that fail to satisfy ($D_1i$). Since $H$ contains at most $n$ loops, $|D_{1,1}| \leq n/10^8 = k/100$. Since $e_H(N^+_{D_5}(y), N^+_{D_1}(y)) \leq 3k^3/n$ by (4.5.2), $|D_{1,2}| \leq k/100$. Since any $d_1$-arc of $H$ whose tail is not $y$ is contained in at most one distinguishable $(y, z, \mathcal{P})$-bridge $B$ and each $B$ contains two such arcs, $r(H) \geq |D_{1,3}|(k/100 - 1)/2$. Thus, $|D_{1,3}| \leq k/1000$. Let $D_1' \coloneqq D_1 \setminus (D_{1,1} \cup D_{1,2} \cup D_{1,3})$, and notice that $|D_1'| \geq k/6 - k/50 - k/1000 \geq 7k/50$. Similarly there is a set $D_2' \subseteq D_2$ of size at least $7k/50$ such that each $d_2 \in D_2'$ satisfies ($D_21$)–($D_23$). At most two colours $d_1 \in D_1'$ yield $u_1 \in \{y, z\}$, and for any $d_1 \in D_1'$ there are at most three choices of $d_2 \in D_2'$ such that $u_1' \in \{y, z, u_1\}$, and at most three choices of $d_2 \in D_2'$ such that $u_6 \in \{y, z, u_1\}$. For fixed $d_1 \in D_1'$, by ($D_11$) there are at most $10^8$ choices of $d_2 \in D_2'$ such that $u_8' = u_6$, and at most three choices of $d_2$ such that $u_8' \in \{y, z, u_1\}$. Finally, if $u_1$ and $z$ are distinct, then we have $u_1' \neq u_6$, since otherwise $u_1 = z$ has two distinct $d_2$-out-neighbours. Since $|D_1'|, |D_2'| \leq k/6$, we deduce that we can remove at most $2|D_2'| + (10^8 + 9)|D_1'| \leq 10^8 k/3$ pairs from $D_1' \times D_2'$ to ensure that all remaining pairs satisfy ($V_{1,2}$). To address ($R_{1,2}$), notice that for fixed $d_1 \in D_1'$, by ($D_13$) there are at most $k/100$ choices of $d_2 \in D_2'$ such that $u_6u_8'$ is contained in a distinguishable $(y, z, \mathcal{P})$-bridge $B$ in $H$. Handling $u_1'u_1$ analogously we deduce that we may remove at most $2 \cdot \frac{k}{100} \cdot \frac{k}{6}$ pairs from $D_1' \times D_2'$ to ensure all remaining pairs satisfy ($R_{1,2}$). In total the number of pairs in $D_1' \times D_2'$ satisfying ($V_{1,2}$) and ($R_{1,2}$) is at least $(7k/50)^2 - 10^8 k/3 - k^2/300 \geq k^2/100$ as claimed. –

*Claim 2: For any $(d_1, d_2) \in D_{1,2}^{good}$ there is a set $D_{3,4}^{good} = D_{3,4}^{good}(d_1, d_2)$ of pairs $(d_3, d_4) \in D_3 \times D_4$ such that $|D_{3,4}^{good}| \geq k^2/100$ and each $(d_3, d_4) \in D_{3,4}^{good}$ satisfies the following, where $u_4 := N_{d_3}^-(y)$, $u_5' := N_{d_4}^+(u_4)$, $u_3 := N_{d_4}^+(z)$, $u_4' := N_{d_3}^-(u_3)$, and $u_1, u_1', u_6$, and $u_8'$ are defined as in Claim 1.*

($D_3$1) *There are at most $10^8$ loops with colour $d_3$ in $H$;*

($D_3$2) *$u_4$ has at most $300k^2/n$ out-neighbours in the set $N_{D_6}^+(z)$;*

($D_3$3) *there are at most $k/100$ arcs $e$ coloured $d_3$ in $H$ such that $e$ is contained in a distinguishable $(y, z, \mathcal{P})$-bridge in $H$;*

($D_4$1) *there are at most $10^8$ loops with colour $d_4$ in $H$;*

($D_4$2) *$u_3$ has at most $300k^2/n$ in-neighbours in the set $N_{D_5}^+(y)$;*

($D_4$3) *there are at most $k/100$ arcs $e$ coloured $d_4$ in $H$ such that $e$ is contained in a distinguishable $(y, z, \mathcal{P})$-bridge in $H$;*

($V_{3,4}$) *$y, z, u_1, u_3, u_4, u_6$ are distinct vertices, and $u_1', u_4', u_5', u_8' \notin \{y, z, u_1, u_3, u_4, u_6\}$;*

($R_{3,4}$) *there is no distinguishable $(y, z, \mathcal{P})$-bridge in $H$ containing the arc $u_4'u_3$ or the arc $u_4u_5'$.*

The proof is similar to that of Claim 1, so we omit it.

*Claim 3: For any $(d_1, d_2) \in D_{1,2}^{good}$ and $(d_3, d_4) \in D_{3,4}^{good}(d_1, d_2)$, there is a set $D_{5,6}^{good}$ (depending on $(d_1, \ldots, d_4)$) of pairs $(d_5, d_6) \in D_5 \times D_6$ such that $|D_{5,6}^{good}| \geq k^2/100$ and each $(d_5, d_6) \in D_{5,6}^{good}$ satisfies the following, where $u_2 := N_{d_5}^+(y)$, $u_2' := N_{d_2}^+(u_2)$, $u_3' := N_{d_3}^+(u_2)$, $u_5 := N_{d_6}^+(z)$, $u_6' := N_{d_4}^-(u_5)$, $u_7' := N_{d_1}^-(u_5)$, and $u_1, u_3, u_4, u_6, u_1', u_4', u_5', u_8'$ are defined as in Claims 1 and 2.*

($D_5$1) *$u_2u_1, u_2u_3 \notin E(H)$;*

($D_6$1) *$u_4u_5, u_6u_5 \notin E(H)$;*

($R_5$) *there is no distinguishable $(y, z, \mathcal{P})$-bridge in $H$ containing the arc $u_2u_2'$ or* *the arc $u_2u_3'$;*

($R_6$) *there is no distinguishable $(y,z,\mathcal{P})$-bridge in $H$ containing the arc $u_6'u_5$ or the arc $u_7'u_5$;*

($V_{5,6}$) *$y,z,u_1,u_2,\dots,u_6$ are distinct, and $u_1',u_2',\dots,u_8' \notin \{y,z,u_1,u_2,\dots,u_6\}$;*

($A_{5,6}1$) *for each $(d_1',d_2') \in F_{1,2}(d_5)$, where $F_{1,2}(d_5)$ is the set of pairs $(d_1',d_2') \in D_1 \times D_2$ such that $N^+_{d_1'}(y) = N^+_{d_2'}(u_2)$, we have $N^-_{d_1'}(u_5) \neq N^-_{d_2'}(z)$;*

($A_{5,6}2$) *for each $(d_3',d_4') \in F_{3,4}(d_5)$, where $F_{3,4}(d_5)$ is the set of pairs $(d_3',d_4') \in D_3 \times D_4$ such that $N^+_{d_4'}(z) = N^+_{d_3'}(u_2)$, we have $N^-_{d_4'}(u_5) \neq N^-_{d_3'}(y)$.*

*Proof of claim:* Let $\widetilde{D}_5$ be the set of colours $d_5 \in D_5$ which fail to satisfy ($D_51$) and ($R_5$), let $\widetilde{D}_6$ be the set of colours $d_6 \in D_6$ which fail to satisfy ($D_61$) and ($R_6$), and define $D_5' \coloneqq D_5 \setminus \widetilde{D}_5$ and $D_6' \coloneqq D_6 \setminus \widetilde{D}_6$. By ($D_12$) and ($D_42$), there are at most $600k^2/n$ colours $d_5 \in D_5$ which fail to satisfy ($D_51$). By ($D_23$) and ($D_33$), at most $k/50$ choices of $d_5 \in D_5$ give $u_2$ to be the tail of a $d_2$-arc or $d_3$-arc contained in a distinguishable $(y,z,\mathcal{P})$-bridge in $H$, and all other choices of $d_5 \in D_5$ satisfy ($R_5$). Using ($D_22$), ($D_32$), ($D_13$), and ($D_43$) similarly, we deduce that $|D_5'|, |D_6'| \geq 7k/50$. By ($D_21$) and ($D_31$), for any $d_6 \in D_6'$ there are at most $2 \cdot 10^8 + 10$ choices of $d_5 \in D_5'$ such that $u_5 \in \{y,z,u_1,u_3,u_4,u_6,u_1',u_4',u_5',u_8'\}$ or $u_5$ is incident to a loop of colour $d_2$ or $d_3$. Further, for any $d_6 \in D_6'$, there are at most 12 choices of $d_5$ such that $u_2 \in \{N^-_d(w) : d \in \{d_2,d_3\}, w \in \{y,z,u_1,u_3,u_4,u_6\}\}$. Thus there are at most $(2 \cdot 10^8 + 22)|D_6|$ choices of pair $(d_5,d_6) \in D_5' \times D_6'$ such that the choice of $d_5$ causes ($V_{5,6}$) to fail. Using ($D_11$) and ($D_41$) to address the choice of $d_6$ similarly, we conclude that we can remove a set of at most $(5 \cdot 10^8)k/6$ pairs $(d_5,d_6) \in D_5' \times D_6'$ such that ($V_{5,6}$) holds for all remaining pairs. For ($A_{5,6}1$), notice that (4.5.2) implies $e_H(N^+_{D_5}(y), N^+_{D_1}(y)) \leq 3k^3/n$, so that there are at most $k/100$ colours $d_5 \in D'_5$ for which $u_2$ has at least $300k^2/n$ out-neighbours in the set $N^+_{D_1}(y)$. Deleting all pairs $(d_5, d_6)$ using such a $d_5$, we have in particular that all remaining pairs satisfy $|F_{1,2}(d_5)| \leq 300k^2/n$. For a remaining pair $(d_5, d_6)$ and each $(d'_1, d'_2) \in F_{1,2}(d_5)$ we define the vertex $v_{d'_1,d'_2} := N^+_{d'_1}\left(N^-_{d'_2}(z)\right)$. Now further deleting all (at most $\frac{k}{6} \cdot 300\frac{k^2}{n}$) pairs $(d_5, d_6)$ such that $u_5 \in \{v_{d'_1,d'_2} : (d'_1, d'_2) \in F_{1,2}(d_5)\}$, all remaining pairs satisfy $(A_{5,6}1)$. We address $(A_{5,6}2)$ similarly. In total, the number of pairs in $D'_5 \times D'_6$ satisfying $(V_{5,6})$, $(A_{5,6}1)$, and $(A_{5,6}2)$ is at least $(7k/50)^2 - 5 \cdot 10^8 k/6 - 2(k^2/600 + 50k^3/n) \geq k^2/100$, finishing the proof of the claim. –

*Claim 4: For any $(d_1, d_2) \in D^{good}_{1,2}$, $(d_3, d_4) \in D^{good}_{3,4}$, $(d_5, d_6) \in D^{good}_{5,6}$, there is a set $E^{good} = E^{good}(d_1, d_2, \ldots, d_6) \subseteq \prod_{i\in[4]} E_{d_i}(H)$ such that $|E^{good}| \geq 9n^4/10$ and each $(e_1, \ldots, e_4) \in E^{good}$ (with $e_i \in E_{d_i}(H)$ for each $i \in [4]$) satisfies the following, where $u''_1 := head(e_2)$, $u''_2 := tail(e_2)$, $u''_3 := tail(e_3)$, $u''_4 := head(e_3)$, $u''_5 := tail(e_4)$, $u''_6 := head(e_4)$, $u''_7 := head(e_1)$, $u''_8 := tail(e_1)$, and $u_1, \ldots, u_6, u'_1, \ldots, u'_8$ are defined as in Claims 1–3.*

($V_E$) *$u''_1, u''_2, \ldots, u''_8$ are distinct vertices and $u''_1, \ldots, u''_8 \notin \{y, z, u_1, \ldots, u_6, u'_1, \ldots, u'_8\}$;*

($E1$) *$u'_1u''_1, u''_2u'_2, u''_3u'_3, u'_4u''_4, u''_5u'_5, u'_6u''_6, u'_7u''_7, u''_8u'_8 \notin E(H)$;*

($E2$) *$\{u''_1, u''_2, \ldots, u''_8\} \cap (N(y) \cup N(z)) = \emptyset$;*

($R_E$) *there is no distinguishable $(y, z, \mathcal{P})$-bridge in $H$ containing any of the arcs $e_1, \ldots, e_4$.*

*Proof of claim:* At most 32 $d_2$-arcs of $H$ have head or tail amongst $y$, $z$, $u_1, \ldots, u_6$, $u'_1, \ldots, u'_8$, and by $(D_21)$, at most $10^8$ arcs in $E_{d_2}(H)$ are loops. Choosing any other $d_2$-arc to be $e_2$, and proceeding similarly for $e_3, e_4, e_1$ (also avoiding the vertices of previously chosen such arcs), we deduce that we can delete at most $4 \cdot$

$10^9 n^3$ tuples from $\prod_{i=1}^4 E_{d_i}(H)$ so that the remaining tuples satisfy $(V_E)$. Notice that at most $2k$ choices of $e_2 \in E_{d_2}(H)$ have head in $N^+(u'_1)$ or tail in $N^-(u'_2)$, and any other $e_2$ satisfies $u'_1u''_1, u''_2u'_2 \notin E(H)$. Dealing with $e_3, e_4, e_1$ similarly addresses $(E1)$. Similarly, for $(E2)$ it suffices to notice that at most $8k$ choices of $e_2$ (for example) have either head or tail in $N(y) \cup N(z)$. Finally, by $(D_i3)$ for each $i \in [4]$, at most $k/100$ choices of each $e_i$ fail to satisfy $(R_E)$. In total, at least $n^4 - 4n^3(10^9 - 2k - 8k - k/100) \geq 9n^4/10$ tuples in $\prod_{i=1}^4 E_{d_i}(H)$ satisfy $(V_E)$, $(E1)$, $(E2)$, and $(R_E)$, as desired. $-$

Let $\Lambda$ be the set of tuples $\lambda = (d_1, \ldots, d_6, e_1, \ldots, e_4)$ satisfying all properties in Claims 1–4. For each $\lambda \in \Lambda$ we define a subgraph $T_\lambda \subseteq H$ with the vertices $V(T_\lambda) = \{y, z, u_1, \ldots, u_6\} \cup \{u'_1, \ldots, u'_8, u''_1, \ldots, u''_8\}$ as defined in Claims 1–4 by the choice of $\lambda$, and whose arcs are as in Definition 4.5.8(i). (These arcs each exist in $H$ by the way the vertices of $T_\lambda$ were defined.) Then since $d_i \in D_i$ for each $i \in [6]$ and since $(D_51)$, $(D_61)$, and $(E1)$ hold for each $\lambda \in \Lambda$, we have that each condition of Definition 4.5.8 is satisfied, so that $T_\lambda$ is a twist system of $H$ for each $\lambda \in \Lambda$. Further, clearly the $T_\lambda$ are distinct. Define $\mathcal{T} := \{T_\lambda \colon \lambda \in \Lambda\}$, and notice that $|\mathcal{T}| \geq (k^2/100)^3 \cdot 9n^4/10 = 9k^6n^4/10^7$.

*Claim 5: For any $T_\lambda \in \mathcal{T}$, the only $(y, z, \mathcal{P})$-bridge in twist$_{T_\lambda}(H)$ that is not in $H$ is the canonical $(y, z, \mathcal{P})$-bridge of the twist.*

*Proof of claim:* Fix $T_\lambda \in \mathcal{T}$ (fixing the notation of all the vertices and arcs as above), and let $B$ be the canonical $(y, z, \mathcal{P})$-bridge of the twist (which has vertices $V(B) = \{y, z, u_1, \ldots, u_6\}$ and colours $d_1, \ldots, d_6$). Suppose that $B'$ is a $(y, z, \mathcal{P})$-bridge in twist$_{T_\lambda}(H)$ that is not in $H$, and label the vertices of $B'$ as $V(B') = \{y, z, v_1, \ldots, v_6\}$ (where the role of $v_i$ in $B'$ corresponds to that of $u_i$ in $B$), and label the colours of $B'$ as $d'_i \in D_i$, for $i \in [6]$. By $(V_{1,2})$, $(V_{3,4})$, and $(V_{5,6})$, all arcs we add when producing $\mathrm{twist}_{T_\lambda}(H)$ from $H$ do not have $y$ or $z$ as an endpoint, and thus one (or more) of the arcs $v_2v_1$, $v_2v_3$, $v_4v_5$, $v_6v_5$ is added by the twist operation. Further, due to the colour partition $\mathcal{P}$, $v_2v_1$ must either be in $H$, or be one of the added arcs $u_2u_1$, $u'_1u''_1$, $u''_2u'_2$ with colour $d'_2 \in D_2$. But since we do not add any arcs incident to $y$, by $(E2)$, $u''_1$ and $u''_2$ are not in the neighbourhood of $y$ in $\mathrm{twist}_{T_\lambda}(H)$, whence $u''_1$ cannot be $v_1$, and $u''_2$ cannot be $v_2$. Thus $v_2v_1$ must either be in $H$, or be $u_2u_1$. Similarly $v_2v_3$, $v_4v_5$, $v_6v_5$ must be $u_2u_3$, $u_4u_5$, $u_6u_5$ respectively, or be in $H$, in some combination. We now split the analysis into cases, depending on how many arcs in $F := \{v_2v_1, v_2v_3, v_4v_5, v_6v_5\}$ are in $H$. In each case we show either that that case does not occur or that $B' = B$, which will complete the proof of the claim. Since $B' \nsubseteq H$, at most three arcs in $F$ are in $H$.

Case 1: Precisely three arcs in $F$ are in $H$.

Let $e$ be the arc in $F$ that is not in $H$. Suppose $e = v_2v_1$, which implies that $e = u_2u_1$, and $d'_2 = d_2$. Since $N^+_{d'_1}(y) = v_1 = u_1 = N^+_{d_1}(y)$, we have $d'_1 = d_1$. Then $v_6 = N^-_{d'_2}(z) = N^-_{d_2}(z) = u_6$, and $v_5$ is the $d'_1$-out-neighbour of $v_6$ in $\mathrm{twist}_{T_\lambda}(H)$, which is the $d_1$-out-neighbour of $u_6$ in $\mathrm{twist}_{T_\lambda}(H)$, namely $u_5$. This is a contradiction, since $v_6v_5$ is in $H$, but $u_6u_5$ is not. One similarly obtains a contradiction if $e$ is $v_2v_3$, $v_4v_5$, or $v_6v_5$, so we deduce that this case does not occur.

Case 2: Precisely two arcs in $F$ are in $H$.

Suppose that $v_2v_1 = u_2u_1$ and $v_6v_5 = u_6u_5$. Then $d'_1 = d_1$, $d'_2 = d_2$, and $v_2v_3$, $v_4v_5$ are in $H$. Hence $v_3 \neq u_3$, and the arc $u_2v_3$ is in $H$ with colour $d'_3$. Moreover the arc $zv_3$ is in $H$ with colour $d'_4$. In particular, $(d'_3, d'_4) \in F_{3,4}(d_5)$. Similarly $v_4 \neq u_4$, we have $v_4y \in E(H)$ has colour $d'_3$, and $v_4u_5 \in E(H)$ has colour $d'_4$. That is, $N^-_{d'_4}(u_5) = v_4 = N^-_{d'_3}(y)$, which contradicts $(A_{5,6}2)$ of Claim 3. Similarly one can use ($A_{5,6}1$) to show that assuming $v_2v_3 = u_2u_3$ and $v_4v_5 = u_4u_5$ yields a contradiction. Suppose instead that $v_2v_1 = u_2u_1$ and $v_2v_3 = u_2u_3$. Then $d'_2 = d_2$, and $v_1 = u_1$ so that $d'_1 = d_1$. Further, $d'_3 = d_3$, and $v_3 = u_3$ so that $d'_4 = d_4$. But this now also determines that $v_4v_5 = u_4u_5$ and $v_6v_5 = u_6u_5$, a contradiction since then no arcs in $F$ are in $H$. All remaining possibilities yield a contradiction similarly, whence this case does not occur.

Case 3: Precisely one arc in $F$ is in $H$.

In particular either we have $v_2v_1 = u_2u_1$ and $v_2v_3 = u_2u_3$ or we have $v_4v_5 = u_4u_5$ and $v_6v_5 = u_6u_5$. But as in Case 2, either way yields that no arcs in $F$ are in $H$. We deduce that this case does not occur.

Case 4: No arcs in $F$ are in $H$.

It is easy to see in this case that $v_i = u_i$ for all $i \in [6]$ whence $B' = B$. $\dashv$

For any $T_\lambda \in \mathcal{T}$, we have by Claim 5 that the canonical $(y, z, \mathcal{P})$-bridge of the twist is distinguishable in $\text{twist}_{T_\lambda}(H)$, since the four arcs $u_2u_1$, $u_2u_3$, $u_4u_5$, $u_6u_5$ are each added by the twist, and do not create any other $(y, z, \mathcal{P})$-bridge. Further, since the twist operation only adds 12 arcs, we obtain from (4.5.2) that $e_{\text{twist}_{T_\lambda}(H)}(W_1, W_2) \le 2k^3/n + 12(s+1)$ for all $W_1, W_2 \subseteq [n]$ of sizes $|W_1| = |W_2| = k$. Thus by ($R_{1,2}$), ($R_{3,4}$), ($R_5$), ($R_6$), ($R_E$), and Claim 5, we have that $\text{twist}_{T_\lambda}(H) \in M_{s+1}$. We now give a final claim which ensures that $|\{\text{twist}_{T_\lambda}(H) : T_\lambda \in \mathcal{T}\}| = |\mathcal{T}|$.

*Claim* 6: *Fix $H' \in M_{s+1}$, let $\lambda, \lambda' \in \Lambda$, and suppose that $\text{twist}_{T_\lambda}(H) = \text{twist}_{T_{\lambda'}}(H) = H'$. Then $\lambda = \lambda'$.*

*Proof of claim:* Let $\lambda = (d_1, \dots, d_6, e_1, \dots, e_4)$, $\lambda' = (d'_1, \dots, d'_6, e'_1, \dots, e'_4)$, with corresponding vertices $V(T_\lambda) = \{y, z, u_1, \dots, u_6\} \cup \{u'_1, \dots, u'_8, u''_1, \dots, u''_8\}$, $V(T_{\lambda'}) = \{y, z, v_1, \dots, v_6\} \cup \{v'_1, \dots, v'_8, v''_1, \dots, v''_8\}$. Let $B$ and $B'$ be the canonical

$(y, z, \mathcal{P})$-bridges of the twists corresponding to $\lambda$ and $\lambda'$, respectively. By Claim 5, $B$ and $B'$ are each the unique $(y, z, \mathcal{P})$-bridge that is in $H'$ but not in $H$, and thus $B = B'$. In particular, $d_i = d_i'$ and $u_i = v_i$ for all $i \in [6]$. By considering the partition $\mathcal{P}$ of $D$, the four arcs in $E(H) \setminus E(H')$ with no endvertex in $\{u_1, \ldots, u_6\}$ must be $e_1 = e_1'$, $e_2 = e_2'$, $e_3 = e_3'$, and $e_4 = e_4'$. We conclude that $\lambda = \lambda'$, as required. $-$

We determine that if $s \leq k^4/(10^{24}n^2)$ and $M_s$ is non-empty, then $M_{s+1}$ is non-empty, and $\delta_s^+ \geq |\mathcal{T}| \geq 9k^6n^4/10^7$, whence $|M_s|/|M_{s+1}| \leq \Delta_{s+1}^-/\delta_s^+ \leq 1/10$. Recalling that $\mathbf{H} \in \mathcal{G}_D$ is uniformly random, it follows that if $s \leq k^4/(10^{25}n^2)$ then

$$\begin{aligned}\mathbb{P}\left[r(\mathbf{H}) = s \,\middle|\, \mathbf{H} \in \widehat{\mathcal{Q}}_D\right] &\leq \frac{|M_s|}{|M_{k^4/(10^{24}n^2)}|} = \prod_{t=s}^{k^4/(10^{24}n^2)-1} \frac{|M_t|}{|M_{t+1}|} \leq \left(\frac{1}{10}\right)^{k^4/(10^{24}n^2)-s} \\ &\leq \exp\left(-\frac{9k^4 \log 10}{10^{25}n^2}\right).\end{aligned}$$

Moreover, we note that if $M_s$ is empty, then clearly $\mathbb{P}\left[r(\mathbf{H}) = s \mid \mathbf{H} \in \widehat{\mathcal{Q}}_D\right] = 0$. Since $k = n/10^6$, we obtain that $\mathbb{P}\left[r(\mathbf{H}) \leq k^4/(10^{25}n^2) \mid \mathbf{H} \in \widehat{\mathcal{Q}}_D\right] \leq (k^4/(10^{25}n^2)+1)\exp(-\Omega(n^2)) = \exp(-\Omega(n^2))$. Since $n^2/10^{50} \leq k^4/(10^{25}n^2)$, the result follows.

$\square$

We are now ready to argue that almost all $G \in \Phi(\overleftrightarrow{K_n})$ contain large well-spread collections of bridging gadgets, which will complete our study of the properties we need to be satisfied by uniformly random $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$.

**Lemma 4.5.10.** *Let $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$ be chosen uniformly at random, and let $\mathcal{E}$ be the event that for all distinct $y, z \in [n]$, $\mathbf{G}$ contains a well-spread collection of at least $n^2/10^{50}$ distinct $(y, z)$-bridging gadgets. Then $\mathbb{P}[\mathcal{E}] \geq 1 - \exp(-\Omega(n^2))$.*

**Proof.** For $D \subseteq [n]$, let $\mathcal{E}|_D$ denote the event (in $\Phi(\overleftrightarrow{K_n})$) that $\mathbf{G}|_D$ contains a well-spread collection of $n^2/10^{50}$ distinct $(y,z)$-bridging gadgets, for each distinct $y,z,\in[n]$. Let $\mathbf{H}\in\mathcal{G}_D$ be chosen uniformly at random, let $\mathbb{P}_D$ denote the measure for this probability space, and let $\mathcal{E}_D^{(y,z)}$ denote the event (in $\mathcal{G}_D$) that $\mathbf{H}$ contains a well-spread collection of $n^2/10^{50}$ distinct $(y,z)$-bridging gadgets, and define $\mathcal{E}_D := \bigcap_{y,z\in[n]\,\text{distinct}} \mathcal{E}_D^{(y,z)}$.

*Claim* 1*: Suppose $D\subseteq[n]$ has size $|D|\le 3n/4$, let $\mathcal{P}=(D_i)_{i\in[6]}$ be an equitable partition of $D$ into six parts, and fix $y,z,\in[n]$ distinct. Then we have that $\mathbb{P}_D\left[r_{(y,z,\mathcal{P})}(\mathbf{H})\ge n^2/10^{50}\right]\le \mathbb{P}_D\left[\mathcal{E}_D^{(y,z)}\right]$.*

*Proof of claim:* Suppose that $H\in\mathcal{G}_D$ and that $r(H)\ge n^2/10^{50}$. By definition of $r(H)$ (see Definition 4.5.7), there is a collection $\mathcal{B}$ of $n^2/10^{50}$ distinct $(y,z,\mathcal{P})$-bridges such that for each $i\in[4]$, any $d_i\in D_i$, and any arc $e\in E_{d_i}(H)$ for which $e$ does not have $y$ nor $z$ as an endvertex, we have that $e$ is contained in at most one $B\in\mathcal{B}$. Note that for each $u\in[n]\setminus\{y,z\}$, we have for any $B\in\mathcal{B}$ which contains $u$, that $B$ must contain an arc $e$ incident to $u$ with colour in $D_i$ for some $i\in[4]$ such that $e$ is not incident to $y$ nor $z$. Therefore $u$ is contained in at most $4\cdot 2\cdot|D|/6\le n$ distinct $B\in\mathcal{B}$. For any colour $d\in D_1$, any $B\in\mathcal{B}$ which uses the colour $d$ must be such that $N_d^+(y)\notin\{y,z\}$ and must contain the vertex $N_d^+(y)$, and thus the colour $d$ is used by at most $n$ distinct $B\in\mathcal{B}$ (and similarly for $d\in D_i$ for all $i\in[6]$). Now, forming a collection $\mathcal{B}'$ of $(y,z)$-bridging gadgets in $H$ by deleting the arcs with colours in $D_5\cup D_6$ for each $B\in\mathcal{B}$, it is clear that $\mathcal{B}'$ witnesses that $H\in\mathcal{E}_D^{(y,z)}$. The claim follows. $-$

Arbitrarily fix $c\in[n]$ and $D\subseteq[n]$ of size $|D|=n/10^6$, and let $\mathcal{F}$ be the set of all possible colour classes for a proper $n$-arc colouring of $\overleftrightarrow{K_n}$ (more precisely, $\mathcal{F}$ is the collection of all sets $F$ of $n$ arcs of $\overleftrightarrow{K_n}$ such that every vertex of $\overleftrightarrow{K_n}$ is the head of precisely one arc in $F$ and the tail of precisely one arc in $F$). Observe that for a fixed equitable partition $\mathcal{P} = (D_i)_{i=1}^6$ of $D$ into six parts, and for fixed distinct $y, z \in [n]$, by (4.5.1), the law of total probability, and Lemma 4.5.4,

$$\mathbb{P}_D\left[\overline{\widehat{\mathcal{Q}}_D}\right] \leq \mathbb{P}_D\left[\overline{\mathcal{Q}_D^1}\right] = \sum_{F\in\mathcal{F}} \mathbb{P}_D\left[\mathbf{F}_c = F\right]\mathbb{P}_D\left[\overline{\mathcal{Q}_D^1} \mid \mathbf{F}_c = F\right] \leq \exp(-\Omega(n^2)). \tag{4.5.3}$$

Then the law of total probability, (4.5.3), and Lemma 4.5.9 give

$$\mathbb{P}_D\left[r_{(y,z,\mathcal{P})}(\mathbf{H}) \leq \frac{n^2}{10^{50}}\right] \leq \mathbb{P}_D\left[r_{(y,z,\mathcal{P})}(\mathbf{H}) \leq \frac{n^2}{10^{50}} \,\middle|\, \widehat{\mathcal{Q}}_D\right] + \mathbb{P}_D\left[\overline{\widehat{\mathcal{Q}}_D}\right] \leq \exp(-\Omega(n^2)). \tag{4.5.4}$$

By (4.5.4) and Claim 1, $\mathbb{P}_D\left[\overline{\mathcal{E}_D^{(y,z)}}\right] \leq \exp(-\Omega(n^2))$, so by a union bound we have $\mathbb{P}_D\left[\overline{\mathcal{E}_D}\right] \leq \exp(-\Omega(n^2))$. Then by Proposition 4.4.4, we have

$$\mathbb{P}\left[\overline{\mathcal{E}}\right] \leq \mathbb{P}\left[\overline{\mathcal{E}|_D}\right] = \frac{\sum_{H\in\overline{\mathcal{E}_D}} \mathrm{comp}(H)}{\sum_{H\in\mathcal{G}_D} \mathrm{comp}(H)} \leq \mathbb{P}_D\left[\overline{\mathcal{E}_D}\right] \cdot \exp(O(n\log^2 n)) \leq \exp(-\Omega(n^2)),$$

which completes the proof of the lemma. □

## 4.6 Absorption

The aim of this section is to show that if $G \in \Phi(\overleftrightarrow{K_n})$ satisfies the conclusions of Lemmas 4.5.6 and 3.4.8, then $G$ admits a small robustly rainbow-Hamiltonian subdigraph $H$ (recall Definition 4.3.1), with arbitrarily chosen flexible sets of appropriate size. In Section 4.7, $H$ will form the key 'absorbing structure'.

**Definition 4.6.1.** Let $A$ be a $(v, c)$-absorbing gadget with abutment vertices $(x_4, x_5)$, and let $B$ be a $(y, z)$-bridging gadget (with all vertices retaining their notation as defined in Definitions 4.5.1 and 4.5.2). If $(y, z) = (x_4, x_5)$, $V(A) \cap V(B) = \{y, z\}$, and $\phi_A(A) \cap \phi_B(B) = \emptyset$, then we say $B$ *bridges* $A$. In this case, we also say a collection $(P_1, P_2, P_3, P_4)$ of properly arc-coloured directed paths *completes* the pair $(A, B)$ if:

- $P_1$ has tail $x_2$ and head $x_3$;
- $P_2$ has tail $w_1$ and head $w_4$;
- $P_3$ has tail $w_5$ and head $w_2$;
- $P_4$ has tail $w_3$ and head $w_6$;
- $P_1, \dots, P_4$ are mutually vertex-disjoint and the internal vertices of $P_1, \dots, P_4$ are disjoint from $V(A) \cup V(B)$;
- $\bigcup_{i=1}^4 P_i$ is rainbow and shares no colour with $A \cup B$.

In this case, we say that $A^* := A \cup B \cup \bigcup_{i=1}^4 P_i$ is a $(v, c)$-*absorber* (see Figure 4.1), and we also define the following.

- The *initial vertex* of $A^*$ is $x_1$, and the *terminal vertex* of $A^*$ is $x_6$.
- The $(v, c)$-*absorbing path* in $A^*$ is the directed path with arc set $\{x_1v, vx_2, x_3x_4, x_4w_1, w_4w_5, w_2w_3, w_6x_5, x_5x_6\} \cup \bigcup_{i=1}^4 E(P_i)$.
- The $(v, c)$-*avoiding path* in $A^*$ is the directed path with arc set $\{x_1x_2, x_3x_5, x_5w_3, w_6w_5, w_2w_1, w_4x_4, x_4x_6\} \cup \bigcup_{i=1}^4 E(P_i)$.

Observe that the $(v, c)$-absorbing path and $(v, c)$-avoiding path of a $(v, c)$-absorber satisfy the following key properties.

(4.6.1) The initial (resp. terminal) vertex of a $(v, c)$-absorber is the tail (resp. head) of both the $(v, c)$-absorbing path and the $(v, c)$-avoiding path.

(4.6.2) The $(v, c)$-absorbing path in a $(v, c)$-absorber contains all of the vertices.

(4.6.3) The $(v, c)$-absorbing path in a $(v, c)$-absorber is rainbow and contains all of the colours.

(4.6.4) The $(v,c)$-avoiding path in a $(v,c)$-absorber contains all of the vertices except $v$.

(4.6.5) The $(v,c)$-avoiding path in a $(v,c)$-absorber is rainbow and contains all of the colours except $c$.

We now use $(v,c)$-absorbers to define a '$T$-absorber' for a bipartite graph $T$, which will essentially form our absorbing structure in almost all $G \in \Phi(\overleftrightarrow{K_n})$, for suitably chosen $T$. The role of $T$ is to provide the 'template' for which pairs $(v,c)$ must provide a $(v,c)$-absorbing path to the rainbow directed Hamilton cycle we are building, and which pairs must provide a $(v,c)$-avoiding path.

**Definition 4.6.2.** Let $T$ be a bipartite graph with bipartition $(A,B)$. A digraph $H$ equipped with a proper arc-colouring $\phi$ is a $T$*-absorber* if the following holds.

(i) There exist injections $f_V : A \to V(H)$ and $f_C : B \to \phi(H)$ such that for every $ab \in E(T)$, there is a unique $(v,c)$-absorber $A_{ab} \subseteq H$, where $v = f_V(a)$ and $c = f_C(b)$, satisfying the following.

  (a) For every $ab \in E(T)$, if $V(A_{ab}) \cap V(A_{a'b'}) \neq \emptyset$ for some $a'b' \in E(T)$ where $a'b' \neq ab$, then $a = a'$ and $V(A_{a,b}) \cap V(A_{a'b'}) = \{f_V(a)\}$.

  (b) For every $ab \in E(T)$, if $\phi(A_{ab}) \cap \phi(A_{a'b'}) \neq \emptyset$ for some $a'b' \in E(T)$ where $a'b' \neq ab$, then $b = b'$ and $\phi(A_{ab}) \cap \phi(A_{a'b'}) = \{f_C(b)\}$.

(ii) There exist pairwise vertex-disjoint length-three paths $P_1, \dots, P_{|E(T)|-1}$, each contained in $H$, satisfying the following.

  (a) $\bigcup_{i=1}^{|E(T)|-1} P_i$ is rainbow and $\phi\left(\bigcup_{i=1}^{|E(T)|-1} P_i\right) \cap \phi\left(\bigcup_{e \in E(T)} A_e\right) = \emptyset$.

  (b) For some enumeration $(e_1, \dots, e_{|E(T)|})$ of $E(T)$, for each $i \in [|E(T)|-1]$, the tail of $P_i$ is the terminal vertex of $A_{e_i}$ and the head of $P_i$ is the initial vertex of $A_{e_{i+1}}$.

(c) For each $i \in [|E(T)|-1]$, $P_i$ is internally vertex-disjoint from $\bigcup_{e \in E(T)} A_e$.

(iii) Subject to (i) and (ii), $H$ is minimal.

In this case, we say a vertex $f_V(a)$ for some $a \in A$ is a *root vertex* of $H$ and a colour $f_C(b)$ for some $b \in B$ is a *root colour* of $H$. Moreover, we say the *initial vertex* of $H$ is the initial vertex of $A_{e_1}$ and the *terminal vertex* of $H$ is the terminal vertex of $A_{e_{|E(T)|}}$.

Suppose that a bipartite graph $T$ is robustly matchable (recall Definition 4.4.2) with respect to flexible sets $A'$ and $B'$. The following lemma shows that a $T$-absorber is robustly rainbow-Hamiltonian (recall Definition 4.3.1) with respect to the root vertices and colours corresponding to $A'$ and $B'$. This (together with Lemmas 4.5.6 and 3.4.8) reduces the task of finding such a subdigraph in almost all $G \in \Phi(\overleftrightarrow{K_n})$ to the task of using large well-spread collections of $(v, c)$-absorbing gadgets and $(y, z)$-bridging gadgets to embed a $T$-absorber, for an appropriate robustly matchable $T$.

**Lemma 4.6.3.** *Let $G \in \Phi(\overleftrightarrow{K_n})$ with proper $n$-arc-colouring $\phi$. Let $T$ be a bipartite graph with bipartition $(A, B)$, let $H \subseteq G$ be a $T$-absorber, and let $u$ and $v$ be the initial and terminal vertices of $H$, respectively. Let $A' \subseteq A$ and $B' \subseteq B$, and let $V'$ and $C'$ be the set of root vertices and colours of $H$ corresponding to $A'$ and $B'$, respectively. If $T$ is robustly matchable with respect to flexible sets $A'$ and $B'$, then $H$ is robustly rainbow-Hamiltonian with respect to flexible sets $V'$ and $C'$ and initial and terminal vertices $u$ and $v$.*

**Proof.** Let $X \subseteq V'$ and $Y \subseteq C'$ such that $|X| = |Y| \leq \min\{|V'|/2, |C'|/2\}$. It suffices to show that $H - X$ contains a rainbow directed Hamilton path which starts at $u$ and ends at $v$, not containing a colour in $Y$. Since $H$ is a $T$-absorber, by Definition 4.6.2(i), there exist injections $f_V : A \to V(H)$ and $f_C : B \to \phi(H)$ such that for every $ab \in E(T)$, there is a unique $(v, c)$-absorber $A_{ab} \subseteq H$, where $v \coloneqq f_V(a)$ and $c \coloneqq f_C(b)$, satisfying (i)(a) and (i)(b). By Definition 4.6.2(ii), there also exist pairwise vertex-disjoint length-three paths $P_1, \dots, P_{|E(T)|-1}$, each contained in $H$, satisfying (ii)(a), (ii)(b), and (ii)(c). Let $(e_1, \dots, e_{|E(T)|})$ be the enumeration of $E(T)$ guaranteed by (ii)(b).

Since $V'$ and $C'$ are the sets of root vertices and colours of $H$ corresponding to $A'$ and $B'$, respectively, $f_V^{-1}(X) \subseteq A'$ and $f_C^{-1}(Y) \subseteq B'$. Thus, since $T$ is robustly matchable with respect to $A'$ and $B'$, there exists a perfect matching $M$ in $T - (f_V^{-1}(X) \cup f_C^{-1}(Y))$. For each $ab \in E(T)$, define a directed path $P_{ab}$ as follows. If $ab \in M$, then let $P_{ab}$ be the $(f_V(a), f_C(b))$-absorbing path in $A_{ab}$, and otherwise let $P_{ab}$ be the $(f_V(a), f_C(b))$-avoiding path in $A_{ab}$.

Now let $P \coloneqq \bigcup_{e \in E(T)} P_e \cup \bigcup_{i=1}^{|E(T)|-1} P_i$. We claim that $P$ is a rainbow directed Hamilton path in $H - X$ which starts at $u$, ends at $v$, and does not contain a colour in $Y$. To that end, we first show the following:

(a) $u$ has out-degree one and in-degree zero in $P$;

(b) $v$ has in-degree one and out-degree zero in $P$;

(c) every $w \in V(P) \setminus \{u, v\}$ has in-degree and out-degree one in $P$;

(d) $V(P) \cap X = \emptyset$;

(e) $\phi(P) \cap Y = \emptyset$;

(f) $V(H) \setminus X \subseteq V(P)$.

Indeed, (a) and (b) follow from (4.6.1), 4.6.2(i)(a), 4.6.2(ii)(b), and 4.6.2(ii)(c), and (c) follows from (4.6.1), (4.6.4), 4.6.2(i)(a), 4.6.2(ii)(b), and 4.6.2(ii)(c).

To prove (d), note that if $w \in X$, then $a \coloneqq f_V^{-1}(w) \notin V(M)$. Thus $w$ is not in the paths $P_{ab}$ for $b \in N_T(a)$ by (4.6.4) as they are all $(w, f_C(b))$-avoiding. Therefore (d) follows again by 4.6.2(i)(a), 4.6.2(ii)(b), and 4.6.2(ii)(c). The proof of (e) is the same, with (4.6.5) instead of (4.6.4), 4.6.2(i)(b) instead of 4.6.2(i)(a), and 4.6.2(ii)(a) instead of 4.6.2(ii)(b) and 4.6.2(ii)(c). To prove (f), first note that if $w \in V' \setminus X$, then there exists $ab \in M$, where $a \coloneqq f_V^{-1}(w)$. Thus, $w \in V(P_{ab})$ by (4.6.2) since $P_{ab}$ is $(w, f_C(b))$-absorbing. In particular, $V' \setminus X \subseteq V(P)$. By (4.6.2) and (4.6.4), 4.6.2(iii) implies that $V(H) \setminus V' \subseteq V(P)$. Thus, $V(H) \setminus X \subseteq V(P)$, as desired.

By (4.6.1), 4.6.2(ii)(b), 4.6.2(ii)(c), and (a)-(c), $P$ contains no cycle, so (a)-(e) imply that $P$ is indeed a directed path in $H - X$, not containing a colour in $Y$, which starts at $u$ and ends at $v$, and (f) implies that $P$ is Hamilton in $H - X$, as required. It remains to show that $P$ is rainbow. By (4.6.3), (4.6.5), and 4.6.2(i)(b), $\bigcup_{e \in E(T)} P_e$ is rainbow, so 4.6.2(ii)(a) implies that $P$ is rainbow, as required. $\square$

The following proposition implies that there are many short rainbow paths that are 'well-spread' in *all* $G \in \Phi(\overleftrightarrow{K_n})$. This enables us to embed these paths in any such $G$ in a vertex- and colour-disjoint way whilst constructing a $T$-absorber for suitably chosen $T$, and whilst absorbing the colours unused by the large rainbow directed path forests we find in Section 4.7.

**Proposition 4.6.4.** *Let $G \in \Phi(\overleftrightarrow{K_n})$ with proper $n$-arc-colouring $\phi$, let $V \coloneqq V(G)$, and let $C \coloneqq \phi(G)$. Let $u, v \in V$ such that $u \neq v$, and let $c \in C$. The following holds for $n$ sufficiently large.*

*(i) There are at least $n^2/3$ length-three directed rainbow paths in $G$ with head $v$ and tail $u$.*

*(ii) If at most $n/2$ loops in $G$ are coloured $c$, then there are at least $n^2/5$ length-*

*four directed rainbow paths in $G$ with head $v$ and tail $u$ such that the second arc is coloured $c$.*

(iii) *For every $w \in V$, there are at most $2n$ length-three directed rainbow paths in $G$ with head $v$ and tail $u$ that contain $w$ as an internal vertex.*

(iv) *For every $w \in V$, there are at most $3n$ length-four directed rainbow paths in $G$ with head $v$ and tail $u$ that contain $w$ as an internal vertex such that the second arc is coloured $c$.*

(v) *For every $d \in C$, there are at most $3n$ length-three directed rainbow paths in $G$ with head $v$ and tail $u$ that contain an arc coloured $d$.*

(vi) *For every $d \in C \setminus \{c\}$, there are at most $3n$ length-four directed rainbow paths in $G$ with head $v$ and tail $u$ that contain an arc coloured $d$ such that the second arc is coloured $c$.*

**Proof.** First, for each vertex $w \in V$, we let $B_w \coloneqq \{x \in V \setminus \{w, v\} : \phi(wx) = \phi(xv)\}$, we say $w$ is *bad* if $|B_w| > n/2$, and we let $B \subseteq V$ be the set of bad vertices. We claim that there is at most one bad vertex; that is, $|B| \leq 1$. To that end, suppose for a contradiction that distinct vertices $w$ and $w'$ are bad. Since $|B_w| + |B_{w'}| > n$, we have $B_w \cap B_{w'} \neq \emptyset$. Thus, there exists some $x \in B_w \cap B_{w'}$, so $\phi(wx) = \phi(xv) = \phi(w'x)$, contradicting that $\phi$ is a proper arc-colouring.

Now we prove (i). Since $|B| \leq 1$, there are at least $(n-3)(n/2-5)$ choices of an ordered pair $(w_1, w_2)$ where $w_1 \in V \setminus (B \cup \{u, v\})$ and $w_2 \in V \setminus (B_{w_1} \cup \{u, v, w_1\})$ such that $\phi(uw_1) \notin \{\phi(w_1w_2), \phi(w_2v)\}$. For each such pair, there is a distinct directed path $P_{w_1,w_2} \coloneqq uw_1w_2v$ in $G$. Since $\phi(uw_1) \notin \{\phi(w_1w_2), \phi(w_2v)\}$ and $w_2 \notin B_{w_1}$, $P_{w_1,w_2}$ is rainbow. Therefore there are at least $n^2/3$ directed length-three rainbow paths in $G$ with head $v$ and tail $u$, as desired.

Now we prove (ii). Let $X \coloneqq \{w \in V : \phi(ww) = c\}$; by assumption, $|X| \leq n/2$. Thus, since $|B| \leq 1$, there are at least $n/2 - 6$ choices of an ordered pair $(w_1, w_2)$ where $w_2 \in V \setminus (X \cup B \cup \{u, v\})$, $w_1 \in V \setminus \{u, v, w_2\}$, $\phi(w_1w_2) = c$, and $\phi(uw_1) \neq c$. For each such pair, there are at least $n/2 - 8$ choices of a vertex $w_3 \in V \setminus (B_{w_2} \cup \{u, v, w_1, w_2\})$ such that $\{\phi(w_2w_3), \phi(w_3v)\} \cap \{\phi(uw_1), c\} = \emptyset$. For each such choice of $(w_1, w_2, w_3)$, there is a distinct directed length-four rainbow path $P_{w_1,w_2,w_3} \coloneqq uw_1w_2w_3v$ with head $v$ and tail $u$ such that the second arc is coloured $c$. Therefore there are at least $(n/2 - 7)(n/2 - 8) \geq n^2/5$ such paths, as desired.

The proofs of (iii)-(vi) are similar, so we only provide a complete proof of (vi). Fix $d \in C \setminus \{c\}$, and let $\mathcal{P}_{(vi)}$ be the set of length-four rainbow directed paths in $G$ with head $v$ and tail $u$ that contain an arc coloured $d$ such that the second arc is coloured $c$. Partition $\mathcal{P}_{(vi)}$ into sets $\mathcal{P}_{(vi),1}, \ldots, \mathcal{P}_{(vi),4}$ such that for $i \in [4]$, a path $P \in \mathcal{P}_{(vi),i}$ if the arc coloured $d$ is the $i$th arc of $P$. Since $d \neq c$, $\mathcal{P}_{(vi),2} = \emptyset$. Every path in $\mathcal{P}_{(vi),1}$ is uniquely determined by the vertex in the path adjacent to $v$, every path in $\mathcal{P}_{(vi),3}$ is uniquely determined by an ordered pair $(w_2, w_3)$ such that $\phi(w_2w_3) = d$, and every path in $\mathcal{P}_{(vi),4}$ is uniquely determined by an ordered pair $(w_1, w_2)$ such that $\phi(w_1w_2) = c$. Therefore $|\mathcal{P}_{(vi)}| \leq 3n$, as desired.

We conclude by outlining the necessary changes to the proof of (vi) to obtain proofs of (iii)-(v). Define $\mathcal{P}_{(iii)}$, $\mathcal{P}_{(iv)}$, and $\mathcal{P}_{(v)}$ in an analogous way. Partition $\mathcal{P}_{(iii)}$ and $\mathcal{P}_{(iv)}$ based on the position of $w$ in each path, and partition $\mathcal{P}_{(v)}$ based on the position of the arc coloured $d$. Finally, show that each part contains at most $n$ paths. $\square$

For any $m = \omega(1)$ (in Section 4.7 we will set $m \coloneqq \lfloor n/\log^3 n \rfloor$, and we assume $n$

to be sufficiently large) we have by Lemma 4.4.3 that there exists a 256-regular $2RMBG(7m, 2m)$, say $T$ (recall Definition 4.4.2). In the following lemma, we show that if $m < n/\log n$, then we may greedily embed a $T$-absorber in any $G \in \Phi(\overleftrightarrow{K_n})$ that satisfies the conclusions of Lemma 4.5.6 and 3.4.8, by choosing each absorber successively in three steps: for each edge $vc$ of $T$, we first embed a $(v, c)$-absorbing gadget $A$, then choose a bridging gadget $B$ that bridges $A$ (recall Definition 4.6.1), and finally use Proposition 4.6.4 to embed the extra short rainbow paths required to complete the $(v, c)$-absorber and connect it to the previously embedded absorber.

**Lemma 4.6.5.** *Let $G \in \Phi(\overleftrightarrow{K_n})$ with proper $n$-arc-colouring $\phi$, let $V := V(G)$, and let $C := \phi(G)$. Let $m < n/\log n$, let $T$ be a 256-regular $2RMBG(7m, 2m)$, let $U \subseteq V$ and $D \subseteq C$ such that $|U| = |D| = 7m$, and let $V' \subseteq U$ and $C' \subseteq D$ such that $|V'| = |C'| = 2m$. For $n$ sufficiently large, if*

- *for all $v \in V$ and $c \in C$, $G$ contains a well-spread collection $\mathcal{A}_{v,c}$ of at least $n^2/2^{100}$ $(v, c)$-absorbing gadgets and*
- *for all distinct $y, z \in V$, $G$ contains a well-spread collection $\mathcal{B}_{y,z}$ of at least $n^2/10^{50}$ $(y, z)$-bridging gadgets,*

*then $G$ contains a $T$-absorber $H$ rooted on vertices $U$ and colours $D$ such that $V'$ and $C'$ are the sets of root vertices and colours of $H$ corresponding to the flexible sets of $T$.*

**Proof.** Denote the bipartition of $T$ by $(A, B)$, let $A' \subseteq A$ and $B' \subseteq B$ be the flexible sets of $T$, and enumerate the edges of $T$ as $e_1, \ldots, e_{|E(T)|}$. Let $f_V : A \to U$ and $f_C : B \to D$ be bijections, chosen such that $f_V(A') = V'$ and $f_C(B') = C'$. For each $j \in [|E(T)|]$, we inductively choose

(i) a $(v, c)$-absorbing gadget $A_j$ in $G$, where $ab := e_j$, $v := f_V(a)$, and $c := f_V(b)$,

(ii) a $(y_j, z_j)$-bridging gadget $B_j$ in $G$, where $(y_j, z_j)$ is the pair of abutment vertices of $A_j$, which bridges $A_j$, and

(iii) length-three rainbow directed paths $P_{j,1}, \ldots, P_{j,5}$ in $G$, such that $P_{j,1}, \ldots, P_{j,4}$ complete the pair $(A_j, B_j)$ to a $(v, c)$-absorber $A_j^*$,

such that the following holds:

(a)$_j$ $V(A_j) \cap U = \{f_V(a)\}$, $V(A_j) \cap V(P_{\ell,5}) = \emptyset$ for all $\ell < j$, and if $V(A_j) \cap V(A_\ell^*) \neq \emptyset$ for some $\ell < j$, then $V(A_j) \cap V(A_\ell^*) = \{f_V(a)\}$, where $a \in e_j \cap e_\ell \subseteq A$;

(b)$_j$ $\phi(A_j) \cap D = \{f_C(b)\}$, $\phi(A_j) \cap \phi(P_{\ell,5}) = \emptyset$ for all $\ell < j$, and if $\phi(A_j) \cap \phi(A_\ell^*) \neq \emptyset$ for some $\ell < j$, then $\phi(A_j) \cap \phi(A_\ell^*) = \{f_C(b)\}$, where $b \in e_j \cap e_\ell \subseteq B$;

(c)$_j$ $V(B_j) \cap V(A_\ell^* \cup P_{\ell,5}) = \emptyset$ for all $\ell < j$, and $V(B_j) \cap U = \emptyset$;

(d)$_j$ $\phi(B_j) \cap \phi(A_\ell^* \cup P_{\ell,5}) = \emptyset$ for all $\ell < j$, and $\phi(B_j) \cap D = \emptyset$;

(e)$_j$ $V(P_{j,k}) \cap V(A_\ell^* \cup P_{\ell,5}) = \emptyset$ for all $k \in [4]$ and $\ell < j$, and $V(P_{j,k}) \cap U = \emptyset$ for all $k \in [5]$;

(f)$_j$ $\phi(P_{j,k}) \cap \phi(A_\ell^* \cup P_{\ell,5}) = \emptyset$ for all $k \in [4]$ and $\ell < j$, and $\phi(P_{j,k}) \cap D = \emptyset$ for all $k \in [5]$;

(g)$_j$ $\phi(P_{j,5}) \cap \phi(A_\ell^*) = \emptyset$ for all $\ell \leq j$, and $\phi(P_{j,5}) \cap \phi(P_{\ell,5}) = \emptyset$ for all $\ell < j$;

(h)$_j$ if $j > 1$, then the tail of $P_{j,5}$ is the terminal vertex of $A_{j-1}^*$, the head of $P_{j,5}$ is the initial vertex of $A_j^*$, $P_{j,5}$ is internally vertex-disjoint from $A_\ell^*$ for all $\ell \leq j$, $V(P_{j,5}) \cap V(P_{\ell,5}) = \emptyset$ for all $\ell < j$, and if $j = 1$, then $P_{j,5} = \emptyset$.

To that end, we let $i \in [|E(T)|]$ and assume $A_j$, $B_j$, and $P_{j,1}, \ldots, P_{j,5}$ satisfying (a)$_j$-(h)$_j$ have been chosen for $j < i$, and we show that we can indeed choose $A_j$, $B_j$, and $P_{j,1}, \ldots, P_{j,5}$ according to (i)-(iii) satisfying (a)$_j$-(h)$_j$ for $j = i$. Let $U' := U \cup \bigcup_{\ell<j} V(A_\ell^* \cup P_{\ell,5})$, and let $D' := D \cup \bigcup_{\ell<j} \phi(A_\ell^* \cup P_{\ell,5})$. For every $\ell < j$, we have by (i)–(iii) that $|V(A_\ell^* \cup P_{\ell,5})|$, $|\phi(A_\ell^* \cup P_{\ell,5})| \leq 24$. Thus, since $T$ is

256-regular and $m < n/\log n$,

$$|U'|, |D'| \leq 7m + 24j \leq 43008m < 43008n/\log n. \tag{4.6.6}$$

First we show that we can choose a $(v, c)$-absorbing gadget $A_j$ according to (i) satisfying (a)$_j$ and (b)$_j$. By assumption, $G$ contains a well-spread collection $\mathcal{A}_{v,c}$ of $n^2/2^{100}$ $(v, c)$-absorbing gadgets. For each $u \in U'$, let $\mathcal{A}_u := \{A^{\mathrm{gdgt}} \in \mathcal{A}_{v,c} : u \in V(A^{\mathrm{gdgt}})\}$, and for each $d \in D'$, let $\mathcal{A}_d := \{A^{\mathrm{gdgt}} \in \mathcal{A}_{v,c} : d \in \phi(A^{\mathrm{gdgt}})\}$. Let $\mathcal{A}'_{v,c} := \mathcal{A}_{v,c} \setminus (\bigcup_{u \in U' \setminus \{v\}} \mathcal{A}_u \cup \bigcup_{d \in D' \setminus \{c\}} \mathcal{A}_d)$. Since $\mathcal{A}_{v,c}$ is well-spread, $|\mathcal{A}_u|, |\mathcal{A}_d| \leq n$ for every $u \in U' \setminus \{v\}$ and $d \in D' \setminus \{d\}$, so by (4.6.6), $|\mathcal{A}'_{v,c}| \geq |\mathcal{A}_{v,c}| - 86016n^2/\log n > 0$. In particular, there exists $A_j \in \mathcal{A}'_{v,c}$. By construction of $\mathcal{A}'_{v,c}$, $A_j$ satisfies (a)$_j$ and (b)$_j$, as desired.

Now let $(y_j, z_j)$ be the abutment vertices of $A_j$. By a similar argument, we can choose a $(y_j, z_j)$-bridging gadget $B_j$ according to (ii) satisfying (c)$_j$ and (d)$_j$.

Finally, we show that we can choose $P_{j,1}, \ldots, P_{j,5}$ according to (iii) satisfying (e)$_j$-(h)$_j$. The argument is again similar, using (i), (iii), and (v) of Proposition 4.6.4 instead of the existence of a well-spread collection of gadgets, so we omit the proof.

To complete the proof, we show that $H := \bigcup_{i=1}^{|E(T)|} A_i^* \cup \bigcup_{i=1}^{|E(T)|-1} P_{i+1,5}$ is a $T$-absorber rooted on vertices $U$ and colours $D$. Since (a)$_j$-(f)$_j$ hold for every $j \in [|E(T)|]$, $H$ satisfies 4.6.2(i), and since (g)$_j$ and (h)$_j$ hold for every $j \in [|E(T)|]$, $H$ satisfies 4.6.2(ii). Clearly $H$ is minimal with respect to these properties, so $H$ also satisfies 4.6.2(iii). Thus, $H$ is a $T$-absorber rooted on $U$ and $D$, as desired. Moreover, by the choice of $f_V$ and $f_C$, $V'$ and $C'$ are the sets of root vertices and colours of $H$ corresponding to the flexible sets of $T$, as desired. $\square$

## 4.7 Proof of Theorem 4.1.6

In this section we use the results we have obtained thus far to prove Theorem 4.1.6. We begin by arguing that a 'lower-quasirandomness' condition in $G \in \Phi(\overleftrightarrow{K_n})$ (which holds in almost all $G \in \Phi(\overleftrightarrow{K_n})$ by Theorem 4.4.7) is enough to ensure the existence of many rainbow directed path forests spanning all but a small arbitrary set of vertices, avoiding a small arbitrary forbidden set of colours, and having few components. We remark that the method we use to count the rainbow directed path forests is inspired by the method used by Kwan (see the proof of [90, Lemma 5.5]) to count large matchings in random Steiner triple systems.

**Definition 4.7.1.** Let $G \in \Phi(\overleftrightarrow{K_n})$ with proper $n$-arc-colouring $\phi$, let $V := V(G)$, and let $C := \phi(G)$. We say that $G$ is *lower-quasirandom* if for all (not necessarily distinct) sets $U_1, U_2 \subseteq V$ and $D \subseteq C$, we have that $e_{G,D}(U_1, U_2) \geq |U_1||U_2||D|/n - n^{5/3}$.

**Lemma 4.7.2.** *Let $G \in \Phi(\overleftrightarrow{K_n})$ with proper $n$-arc-colouring $\phi$, let $V := V(G)$, and let $C := \phi(G)$. Let $U \subseteq V$ and let $D \subseteq C$ be equal-sized sets of size at most $n/\log^2 n$. If $G$ is lower-quasirandom, then there are at least $((1 - o(1))n/e^2)^n$ spanning rainbow directed path forests $Q$ of $G - U$ such that $\phi(Q) \cap D = \emptyset$ and $Q$ has at most $n^{9/10}$ components.*

**Proof.** Throughout the proof we implicitly assume $n$ is sufficiently large for certain inequaities to hold. We say a rainbow directed path forest in $G$ is *valid* if it has no vertices in $U$ and no colours in $D$. Let $n' := n - |U| - \lfloor n^{9/10} \rfloor$, and let $n'' := n - |U|$. If $Q$ is a spanning directed path forest of $G - U$, then the number of components of $Q$ is equal to $|V \setminus U| - |E(Q)|$, so $Q$ has at most $n^{9/10}$ components if and only if it has at least $n'$ arcs. Thus, it suffices to count the number of valid rainbow directed path forests in $G$ that have $n'$ arcs. To that end, we first count the number of *ordered* sequences of arcs $(e_1, \ldots, e_{n'})$ such that $\bigcup_{i=1}^{n'} e_i$ is a valid rainbow directed path forest in $G$. We claim that for every $j \in [n']$, if $e_1, \ldots, e_{j-1}$ are arcs in $G$ such that $\bigcup_{i=1}^{j-1} e_i$ is a valid rainbow directed path forest, then there are at least $(1 - o(1)) \left(n'' - (j-1)\right)^3 / n$ choices of an arc $e_j \in E(G)$ such that $\bigcup_{i=1}^{j} e_j$ is also a valid rainbow directed path forest. Let $V_H$ be the set of vertices $u \in V \setminus U$ such that $u$ has no out-neighbor in $\bigcup_{i=1}^{j-1} e_i$, let $V_T$ be the set of vertices $u \in V \setminus U$ such that $u$ has no in-neighbor in $\bigcup_{i=1}^{j-1} e_i$, and let $D' \coloneqq (C \setminus D) \setminus \bigcup_{i=1}^{j-1} \phi(e_i)$. Since $G$ is lower-quasirandom, since $|V_H| = |V_T| = |D'| = n - |U| - (j-1)$, and since $j \leq n' = n'' - \lfloor n^{9/10} \rfloor$, we have that

$$e_{G,D'}(V_H, V_T) \geq \frac{(n - |U| - (j-1))^3}{n} - n^{5/3} \geq (1 - o(1)) \frac{(n'' - (j-1))^3}{n}. \quad (4.7.1)$$

The spanning path forest in $G - U$ with edge set $\{e_1, \ldots, e_{j-1}\}$ has $k \coloneqq n'' - (j-1)$ components, which we denote $P_1, \ldots, P_k$. For every $i \in [k]$, there is a unique arc $f_i \in E(G)$ (whose head is the tail of $P_i$ and whose tail is the head of $P_i$) such that $P_i \cup f_i$ is a directed cycle. Let $F \coloneqq \{f_1, \ldots, f_k\}$, and let $e_j \in E_{G,D'}(V_H, V_T) \setminus F$. By the choice of $V_H$, $V_T$, $D'$, and $F$, we have that $\bigcup_{i=1}^{j} e_i$ is rainbow, has maximum in-degree and out-degree one, and contains no cycle. Hence, it is a valid rainbow directed path forest, as required. By (4.7.1), $|E_{G,D'}(V_H, V_T) \setminus F| \geq (1 - o(1)) \left(n'' - (j-1)\right)^3 / n$, so the claim follows.

Therefore, the number of ordered sequences of arcs $e_1, \ldots, e_{n'} \in E(G)$ such that

$\bigcup_{i=1}^{n'} e_i$ is a valid rainbow directed path forest is at least

$$\prod_{j=1}^{n'} (1-o(1)) \frac{(n''-(j-1))^3}{n} = \left((1-o(1)) \frac{(n'')^3}{n}\right)^{n'} \exp\left(3\sum_{j=1}^{n'} \log\left(1-\frac{j-1}{n''}\right)\right). \tag{4.7.2}$$

Since $|U| \le n/\log^2 n$, we have $n' \ge n - 2n/\log^2 n$ and $(n'')^3/n \ge (1-3/\log^2 n)n^2$. Hence,

$$\left(\frac{(n'')^3}{n}\right)^{n'} \ge \left(\left(1-\frac{3}{\log^2 n}\right)n^2\right)^n n^{-4n/\log^2 n} \ge \left((1-o(1))\, n^2\right)^n. \tag{4.7.3}$$

Since the function $x \mapsto \log(1-x)$ is negative and monotonically decreasing for $x \in [0,1)$,

$$\sum_{j=1}^{n'} \frac{1}{n''} \log\left(1-\frac{j-1}{n''}\right) \ge \int_0^{n'/n''} \log(1-x)\, dx \ge \int_0^1 \log x\, dx = -1. \tag{4.7.4}$$

By substituting (4.7.3) and (4.7.4) into the right side of (4.7.2), the expression in (4.7.2) is at least

$$\left((1-o(1))\, n^2\right)^n e^{-3n''} \ge \left((1-o(1))\, \frac{n^2}{e^3}\right)^n. \tag{4.7.5}$$

By Stirling's approximation, $n! = (1+O(1/n))\sqrt{2\pi n}(n/e)^n \le ((1+o(1))n/e)^n$. Hence, since (4.7.5) provides a lower bound on the number of ordered sequences of edges $e_1, \ldots, e_{n'} \in E(G)$ such that $\bigcup_{i=1}^{n'} e_i$ is a valid rainbow directed path forest, the total number of valid rainbow directed path forests in $G$ with at most $n^{9/10}$ components is at least

$$\left((1-o(1))\, \frac{n^2}{e^3}\right)^n \Big/ n! \ge \left((1-o(1))\, \frac{n}{e^2}\right)^n,$$

as desired. $\square$

We now have all the tools we need to prove Theorem 4.1.6.

**Proof of Theorem 4.1.6.** Let $G \in \Phi(\overleftrightarrow{K_n})$ with proper $n$-arc-colouring $\phi$, let $V := V(G)$, and let $C := \phi(G)$. By Lemma 4.4.6, Theorem 4.4.7, and Lemmas 4.5.6 and 3.4.8, it suffices to show that if

(4.7.6) for every $c \in C$, at most $n/2$ loops in $G$ are coloured $c$,

(4.7.7) $G$ is lower-quasirandom,

(4.7.8) $G$ contains a well-spread collection of at least $n^2/2^{100}$ $(v, c)$-absorbing gadgets for every $v \in V$ and $c \in C$, and

(4.7.9) $G$ contains a well-spread collection of at least $n^2/10^{50}$ $(y, z)$-bridging gadgets for every $y, z \in V$,

then $G$ contains at least $((1 - o(1))\, n/e^2)^n$ rainbow directed Hamilton cycles.

Let $m = \lfloor n/\log^3 n \rfloor$, and let $T$ be a 256-regular $2RMBG(7m, 2m)$ (which exists by Lemma 4.4.3). We build a $T$-absorber $H$ in $G$ using Lemma 4.6.5, but first we need the following claim to choose the roots of $H$ that will correspond to the flexible sets of $T$.

*Claim 1: There exist $V' \subseteq V$ and $C' \subseteq C$ such that $|V'| = |C'| = 2m$ and for every $u, v \in V$ such that $u \neq v$ and for every $c \in C$, there are at least $n^{99/50}$ directed paths $P$ in $G$ such that*

*(i) $P$ is rainbow and has length four,*

*(ii) $P$ has head $v$ and tail $u$,*

*(iii) the second arc of $P$ is coloured $c$,*

*(iv) $V(P) \setminus \{u, v\} \subseteq V'$, and*

*(v) $\phi(P) \setminus \{c\} \subseteq C'$.*

*Proof of claim:* Let $p \coloneqq (2m + n^{9/10})/n$, and let $U' \subseteq V$ and $D' \subseteq C$ be chosen randomly by including every $v \in V$ in $U'$ and every $c \in C$ in $D'$ independently with probability $p$. We claim that the following holds with high probability:

(a) $|U'|, |D'| = pn \pm n^{4/5}$, and

(b) for every $u, v \in V$ such that $u \neq v$ and for every $c \in C$, there are at least $p^6 n^2/10$ directed paths $P$ in $G$ satisfying (i)-(iii), such that $V(P) \setminus \{u, v\} \subseteq U'$ and $\phi(P) \setminus \{c\} \subseteq D'$.

Indeed, (a) follows from a standard application of the Chernoff Bound. To prove (b), we use McDiarmid's Inequality. To that end, fix $u, v \in V$ distinct and $c \in C$. We let $f$ denote the random variable counting the number of paths satisfying (b). Note that $f$ is determined by the independent binomial random variables $\{X_w : w \in V\} \cup \{X_d : d \in C\}$, where $X_w$ indicates if $w \in U'$ and $X_d$ indicates if $d \in D'$. By (4.7.6) and Proposition 4.6.4(ii), there are at least $n^2/5$ paths satisfying (i), (ii), and (iii), and each such path satisfies $V(P) \setminus \{u, v\} \subseteq U'$ with probability $p^3$ and $\phi(P) \setminus \{c\} \subseteq D'$ with probability $p^3$, independently. Hence, $\mathbb{E}[f] \geq p^6 n^2/5$. For each $w \in V$, by Proposition 4.6.4(iv), $X_w$ affects $f$ by at most $3n$, and for each $d \in C$, by Proposition 4.6.4(vi), $X_d$ affects $f$ by at most $3n$. Therefore by McDiarmid's Inequality (Theorem 4.4.1) applied with $t \coloneqq \mathbb{E}[f]/2$, there are at least $p^6 n^2/10$ paths satisfying (b) with probability at least $1 - \exp(-\Omega(p^{12} n))$. Hence, by a union bound, (b) holds for every $u, v \in V$ with $u \neq v$ and every $c \in C$ with high probability, as claimed.

Now we fix a choice of $U'$ and $D'$ satisfying both (a) and (b) simultaneously. By (a), $|U'|, |D'| \geq 2m$, so there exists $V' \subseteq U'$ and $C' \subseteq D'$ such that $|V'| = |C'| = 2m$, as required. Moreover, by (a), $|U' \setminus V'|, |D' \setminus C'| \leq 2n^{9/10}$. Thus, by Proposition 4.6.4(iv) and 4.6.4(vi), (b) implies that for every $u, v \in V$ distinct and $c \in C$, there are at least $p^6n^2/10 - 2(2n^{9/10})(3n) \geq n^{99/50}$ directed paths satisfying (i)-(v), as desired. $-$

Now let $U \subseteq V$ and $D \subseteq C$ such that $|U| = |D| = 7m$, $V' \subseteq U$, and $C' \subseteq D$. By Lemma 4.6.5, (4.7.8), and (4.7.9), there is a $T$-absorber $H$ in $G$ rooted on $U$ and $D$ such that $V'$ and $C'$ are the sets of root vertices and colours of $H$ corresponding to the flexible sets of $T$. By Lemma 4.6.3, $H$ is robustly rainbow-Hamiltonian with respect to flexible sets $V'$ and $C'$ and initial and terminal vertices $t$ and $h$, where $t$ is the initial vertex of $H$ and $h$ is the terminal vertex of $H$. Note that

$$|V(H)| = 22|E(T)| - 2 + |U| \text{ and } |\phi(H)| = 22|E(T)| - 3 + |D|. \tag{4.7.10}$$

*Claim 2: If $Q$ is a spanning rainbow path forest in $G - V(H)$, sharing no colour with $H$, with at most $n^{9/10}$ components, then $G$ has a rainbow directed Hamilton cycle $F$ such that $F - V(H) = Q$.*

*Proof of claim:* Let $P_1, \dots, P_k$ be the components of $Q$, and for each $i \in [k]$, let $h_i$ be the head of $P_i$, and let $t_i$ be the tail of $P_i$. Let $t_{k+1}$ denote the initial vertex of $H$, and let $h_0$ denote the terminal vertex of $H$. That is, $t_{k+1} \coloneqq t$ and $h_0 \coloneqq h$. Let $c_1, \dots, c_{k'}$ be an enumeration of the colours in $C \setminus (\phi(H) \cup \phi(Q))$. Since $Q$ is rainbow, $|\phi(Q)| = |E(Q)| = |V \setminus V(H)| - k$, so by (4.7.10), $k' = k + 1$.

By Claim 1, for each $j \in [k+1]$, there is a collection $\mathcal{P}_j$ of at least $n^{99/50}$ directed paths satisfying (i)-(v) where $u = h_{j-1}$, $v = t_j$, and $c = c_j$. For each $j \in [k+1]$, we inductively choose a path $P'_j \in \mathcal{P}_j$ such that

(a)$_j$ $\phi(P'_j) \cap \phi(P'_\ell) = \emptyset$ for every $\ell < j$ and

(b)$_j$ $V(P'_j) \cap V(P'_\ell) \cap V' = \emptyset$ for every $\ell < j$.

To that end, we let $i \in [k+1]$ and assume $P'_j \in \mathcal{P}_j$ has been chosen to satisfy (a)$_j$

and (b)$_j$ for each $j < i$, and we show that we can indeed choose such a $P'_j$ for $j = i$. Let $U_j := V' \cap \bigcup_{\ell=1}^{j-1} V(P'_\ell)$, and let $D_j := D' \cap \bigcup_{\ell=1}^{j-1} \phi(P'_\ell)$. For each $u \in U_j$, let $\mathcal{P}_u := \{P \in \mathcal{P}_j : u \in V(P)\}$, and for each $d \in D_j$, let $\mathcal{P}_d := \{P \in \mathcal{P}_j : d \in \phi(P)\}$. Let $\mathcal{P}'_j := \mathcal{P}_j \setminus \left(\bigcup_{u \in U_j} \mathcal{P}_u \cup \bigcup_{d \in D_j} \mathcal{P}_d\right)$. By Proposition 4.6.4(iv), $|\mathcal{P}_u| \leq 3n$ for each $u \in U_j$, and by Proposition 4.6.4(vi), $|\mathcal{P}_d| \leq 3n$ for each $d \in D_j$. Since $P'_\ell$ has length four for each $\ell < j$, we have $|U_j|, |D_j| \leq 3k$. Hence, $|\mathcal{P}'_j| \geq n^{99/50} - (6k)(3n) > 0$. In particular, there exists $P'_j \in \mathcal{P}'_j$. By construction of $\mathcal{P}'_j$, $P'_j$ satisfies (a)$_j$ and (b)$_j$, as desired.

Since $Q$ shares no vertices or colours with $H$, for each $j \in [k+1]$, since the internal vertices of $P'_j$ are in $V' \subseteq V(H)$ and since $\phi(P'_j) \setminus D' = \{c_j\}$, it follows that $P'_j \cup P_j$ is a rainbow directed path with tail $h_{j-1}$ and head $h_j$. Moreover, by induction, using (a)$_j$ and (b)$_j$, if $j \leq k$, then $\bigcup_{\ell=1}^{j}(P'_\ell \cup P_\ell)$ is a rainbow directed path with tail $h_0$ and head $h_j$, and in particular, $P^*_1 := P'_{k+1} \cup \bigcup_{\ell=1}^{k}(P'_\ell \cup P_\ell)$ is a rainbow directed path with tail $h_0$ and head $t_{k+1}$.

Let $X := V' \cap \bigcup_{j=1}^{k+1} V(P'_j)$, and let $Y := D' \cap \bigcup_{j=1}^{k+1} \phi(P'_j)$. Note that $|X|, |Y| \leq 3(k+1) \leq m$. Therefore, since $H$ is robustly rainbow-Hamiltonian, there exists a rainbow directed Hamilton path $P^*_2$ in $H - X$ with tail $t_{k+1}$ and head $h_0$, not containing a colour in $Y$. Now $F := P^*_1 \cup P^*_2$ is a rainbow directed Hamilton cycle in $G$, and $F - V(H) = Q$, as desired. $-$

By Lemma 4.7.2 and (4.7.10), there is a collection $\mathcal{Q}$ of at least $((1 - o(1))n/e^2)^n$ spanning rainbow directed path forests in $G - V(H)$ that share no colours with $\phi(H)$ and have at most $n^{9/10}$ components. By Claim 2, for every $Q \in \mathcal{Q}$, there is a rainbow directed Hamilton cycle $F_Q$ such that $F - V(H) = Q$. Therefore $\{F_Q : Q \in \mathcal{Q}\}$ is a collection of at least $((1 - o(1))n/e^2)^n$ distinct rainbow directed

Hamilton cycles in $G$, as desired. $\square$


## Acknowledgements

We are grateful to both anonymous referees for their helpful comments. In particular, we thank one of the referees for suggesting a simplification to the proof of Lemma 4.4.6.


## 4.8 Appendix: Proofs of Lemmas 4.5.4 and 4.5.6

In this section we make clear the changes one needs to make to the arguments of [56] to obtain Lemmas 4.5.4 and 4.5.6. Lemma 4.5.6 is analogous to [56, Lemma 3.8] (minus the notion of 'edge-resilience'), and Lemma 4.5.4 is a direct analogue of [56, Lemma 6.3]. The proof of Lemma 4.5.4 can be obtained via a straightforward modification of the proof of [56, Lemma 6.3], and we describe this first. Then, we show how to adapt the proof of Lemma 4.5.6 from the proof of [56, Lemma 3.8], which can be be summarized as follows:

- the main ingredients in the proof of [56, Lemma 3.8] are [56, Lemmas 6.3, 6.8, and 6.9], and we will need an analogue of each of these;
- as mentioned, Lemma 4.5.4 is the analogue of [56, Lemma 6.3], and the proof adapts easily to this setting;
- we will need the obvious directed analogues of [56, Definitions 6.4–6.7] used in [56, Lemmas 6.8 and 6.9];
- we will need an analogue of [56, Lemma 6.8] (namely Lemma 4.8.1), which

we break down into three claims, of which only the second does not quickly follow as an analogue of claims in the proof of [56, Lemma 6.8];

- the proof of [56, Lemma 6.9] adapts easily to this setting;
- the proof of [56, Lemma 3.8] from [56, Lemmas 6.3, 6.8, and 6.9] is essentially the same (see Lemma 4.8.2), with an extra step to deal with the conditioning on the number of loops in each colour class.

To prove Lemma 4.5.4, the main idea is that we can perform a switching operation very similar to that in the proof of [56, Lemma 6.3], whilst avoiding any arcs coloured $c$, thus leaving the colour class of $c$ unchanged.

**Proof of Lemma 4.5.4.** Modify the proof of [56, Lemma 6.3] by fixing an outcome $F$ of $\mathbf{F}_c$ and analyzing the following 'rotate' switching operation on the set of $H \in \mathcal{G}_D$ whose $c$-colour class is $F$. For fixed $A, B \subseteq [n]$ satisfying $|A| = |B| = |D| = n/10^6$, we instead say that a 'rotation system' of $H$ is a subdigraph $R \subseteq G$ together with a labelling of its vertices $V(R) = \{a, b, v, w\}$ such that $E(R) = \{ab, vw\}$ where $a \in A$, $b \in B$, $v \notin A$, $w \notin B$, $aw, vb \notin E(H)$, and $\phi_H(ab) = \phi_H(vw) \neq c$. Then the 'rotate' switching operation replaces the arcs $ab$ and $vw$ with the arcs $aw$ and $vb$, each in colour $\phi_H(ab) \neq c$, thus leaving the $c$-colour class, $F$, unchanged. Analyzing auxiliary bipartite graphs $B_s$ as in the proof of [56, Lemma 6.1] and defining $\delta_s$ and $\Delta_{s-1}$ to be the analogous quantities, where an edge captures a rotation operation destroying an arc from $A$ to $B$ in some $H \in M_s$, it is simple to see that $\Delta_{s-1} \leq |D|^3$, and $\delta_s \geq (n - |A| - |B| - 2|D|)(s - |A|) = \frac{999996}{10^6} n(s - |D|)$ (here the '$-|D|$' term occurs due to avoiding any arc of colour $c$ to be the one that the rotation switching operation destroys between $A$ and $B$). Then for $s \geq \frac{3|D|^3}{2n}$ and $n$ sufficiently large we obtain that $|M_s|/|M_{s-1}| \leq 9/10$.

Proceeding as in the end of the proof of [56, Lemma 6.1], the result follows. $\square$

The rest of the appendix is dedicated to the proof of Lemma 4.5.6, which we split into two lemmas (Lemmas 4.8.1 and 4.8.2), which roughly speaking correspond to [56, Lemmas 6.8 and 6.9] and the proof of [56, Lemma 3.8], respectively. We begin by defining six events (in addition to $\mathcal{E}$ and $\mathcal{C}$ defined in the lemma statement and $\mathcal{Q}_D^1$ defined in Definition 4.5.3) that we will use, as well as giving some extra notation. For fixed $c \in [n]$, we define $\mathcal{E}^c$ to be the event in the probability space $\mathcal{S}$ corresponding to uniformly random choice of $\mathbf{G} \in \Phi(\overleftrightarrow{K_n})$, that $\mathbf{G}$ contains a well-spread collection of $n^2/2^{100}$ $(v,c)$-absorbing gadgets, for all $v \in [n]$. We define $\mathcal{C}^c$ to be the event in $\mathcal{S}$ that there are at most $n/10^9$ $c$-loops in $\mathbf{G}$, and for fixed $D \subseteq [n]$ we define $(\mathcal{E}^c|_D)$ to be the event in $\mathcal{S}$ that $\mathbf{G}|_D$ contains a well-spread collection of $n^2/2^{100}$ $(v,c)$-absorbing gadgets, for all $v \in [n]$.

For fixed $D \subseteq [n]$ and $c \in D$ we define $\mathcal{E}_D^c$ to be the event in the probability space $\mathcal{S}_D$ corresponding to uniformly random choice of $\mathbf{H} \in \mathcal{G}_D$, that $\mathbf{H}$ contains a well-spread collection of $n^2/2^{100}$ $(v,c)$-absorbing gadgets, for all $v \in [n]$, and we define $\mathcal{C}_D^c$ to be the event in $\mathcal{S}_D$ that there are at most $n/10^9$ $c$-loops in $\mathbf{H}$. Finally, we define the event $\widetilde{\mathcal{Q}}_D$ in $\mathcal{S}_D$ in an analogous way to the definition of $\widetilde{\mathcal{Q}}_{D^*}^{\mathrm{col}}$ in [56] (see the text preceding Definition 6.7 in the cited paper). Indeed, for an equitable partition $(D_i)_{i=1}^4$ of $D$ we define '$(v,c,\mathcal{P})$-gadgets', 'distinguishability' of $(v,c,\mathcal{P})$-gadgets, 'saturation' of $c$-arcs, and the function $r_{(v,c,\mathcal{P})}\colon \mathcal{G}_D \to [n|D|]_0$ in ways corresponding to [56, Definitions 6.4–6.6]. (Just add the necessary directions as per Definition 4.5.1 of the current paper, orienting the $d_4$-arc (with $d_4 \in D_4$) of a $(v,c,\mathcal{P})$-gadget away from $v$ (say).) Then $\widetilde{\mathcal{Q}}_D$ is the set of $H \in \mathcal{G}_D$ such that if $r_{(v,c,\mathcal{P})}(H) = s$, then $e_H(A,B) \leq 2|D|^3/n + 6s$ for all

$A, B \subseteq [n]$ such that $|A| = |B| = |D|$. $\widetilde{\mathcal{Q}}_D$ is just a reformulation of upper-quasirandomness which is closed under the switching operation we use to find $(v, c, \mathcal{P})$-gadgets. The following lemma plays a role analogous to [56, Lemmas 6.8 and 6.9]. First, we discuss the switching operation that we will use in the proof. Suppose $D \subseteq [n]$, $v, c \in [n]$, let $H \in \mathcal{G}_D$, and fix a partition $\mathcal{P} = (D_i)_{i=1}^4$ of $D$. Let $u_1, u_2, \ldots, u_{14} \in [n] \setminus \{v\}$, where $u_1, u_2, u_5, u_6, u_7, u_8, u_9, u_{10}, u_{13}, u_{14}$ are distinct and $\{u_1, u_2, u_5, u_6, u_7, u_8, u_9, u_{10}, u_{13}, u_{14}\} \cap \{u_3, u_4, u_{11}, u_{12}\} = \emptyset$. Then we say that a subgraph $T \subseteq H[\{v, u_1, u_2, \ldots, u_{14}\}]$ is a *twist system* of $H$ if $T$ satisfies [56, Definition 6.7(i)–(vii)] (with directions added as discussed above), and we define the switching operation 'twist$_T$' and the 'canonical $(v, c, \mathcal{P})$-gadget of the twist' analogously to [56, Definition 6.7]. We use the notation $\mathbb{P}_D$ for the measure of the probability space $\mathcal{S}_D$.

**Lemma 4.8.1.** *Suppose $D \subseteq [n]$ has size $|D| = n/10^6$, and fix $c \in D$. Then*

$$\mathbb{P}_D\left[\mathcal{E}_D^c \,\middle|\, \widetilde{\mathcal{Q}}_D \cap \mathcal{C}_D^c\right] \geq 1 - \exp\left(-\Omega(n^2)\right).$$

**Proof.** It is simple to repurpose the arguments of [56, Lemma 6.9] to show that if $r_{(v,c,\mathcal{P})}(\mathbf{H}) \geq n^2/2^{100}$ for all $v$ then $\mathcal{E}_D^c$ occurs, whence it suffices to show that

$$\mathbb{P}_D\left[\left(\exists v \in [n] \colon r_{(v,c,\mathcal{P})}(\mathbf{H}) \leq \frac{n^2}{2^{100}}\right) \,\middle|\, \widetilde{\mathcal{Q}}_D \cap \mathcal{C}_D^c\right] \leq \exp\left(-\Omega\left(n^2\right)\right), \tag{4.8.1}$$

for an arbitrary fixed equitable partition $\mathcal{P} = (D_i)_{i=1}^4$. Fix such a $\mathcal{P}$. We show that (4.8.1) holds by analysing the twist switching operation. We will ensure that we only perform twists which increase $r(H)$ (by precisely one). Note that the set $\widetilde{\mathcal{Q}}_D \cap \mathcal{C}_D^c$ is closed under such twists, since we only add six arcs, and we do not add nor delete any arc coloured $c$. Let $k := |D| = n/10^6$, and fix $v \in [n]$. Analyzing auxiliary bipartite graphs $B_s$ analogous to those in the proof of [56, Lemma 6.8], it is simple to see that $\Delta_{s+1} \leq (s+1)n^4$ for all $s \in [nk-1]_0$. We now seek a lower bound for $\delta_s$ in the case that $s \leq k^4/2^{16}n^2$. To that end, we fix such an $s$, fix $H$ in (the analogue of) $T_s^D$, and bound from below the degree of $H$ in $B_s$ by finding many twist systems in $H$ with desirable properties. Note that [56, Lemma 6.8, Equation (6.2)] holds as stated, with $e_G(A,B)$ replaced by $e_H(A,B)$. The remainder of the proof now largely splits into three claims.

*Claim 1: There is a set $D_3^{good} \subseteq D_3$ of size $|D_3^{good}| \geq k/10$ such that for all $d \in D_3^{good}$ we have*

(i) $|E_d(N^-_{D_1}(x), N^+_{D_2}(x))| \leq 200k^2/n$ *(in $H$);*

(ii) *there are at most $64k^3/n^2$ $d$-arcs $e$ in $H$ with the property that $e$ lies in some distinguishable $(v,c,\mathcal{P})$-gadget in $H$ whose $c$-arc is not supersaturated;*

(iii) *there are at most $n/100$ $d$-loops in $H$.*

*Proof of claim:* We have that (i) and (ii) hold for at least $k/8$ colours $d \in D_3$ analogously to [56, Lemma 6.8, Claim 1], so it suffices to note that at most 100 colours fail condition (iii). –

*Claim 2: There is a set $\Lambda \subseteq \{(d_1,d_2,d_3,d_4,f_1,f_2)\colon d_1 \in D_1, d_2 \in D_2, d_3 \in D_3^{good}, d_4 \in D_4, f_1,f_2 \in E_{d_3}(H)\}$ of size $|\Lambda| \geq k^4n^2/10000$ such that for each $(d_1,d_2,d_3,d_4,f_1,f_2) \in \Lambda$ the following holds:*

(i) $v \notin \{u_1,u_2,\ldots,u_{14}\}$*;*

(ii) $u_1,u_2,u_5,u_6,u_7,u_8,u_9,u_{10},u_{13},u_{14}$ *are distinct;*

(iii) $\{u_1,u_2,u_5,u_6,u_7,u_8,u_9,u_{10},u_{13},u_{14}\} \cap \{u_3,u_4,u_{11},u_{12}\} = \emptyset$,

*where for each $(d_1,d_2,d_3,d_4,f_1,f_2) \in \Lambda$, we set* $u_7 := N^+_{d_4}(v)$, $u_5 := N^-_{d_1}(u_7)$, $u_3 := N^-_{d_3}(u_5)$, $u_8 := N^-_c(u_7)$, $u_6 := N^-_{d_2}(u_8)$, $u_4 := N^+_{d_3}(u_4)$, $u_2 :=$ *tail*$(f_1)$, $u_1 :=$

$head(f_1), u_9 \coloneqq N^-_{d_1}(v), u_{11} \coloneqq N^+_{d_3}(u_9), u_{10} \coloneqq N^+_{d_2}(v), u_{12} \coloneqq N^-_{d_3}(u_{10}), u_{13} \coloneqq tail(f_2), u_{14} \coloneqq head(f_2)$.

*Proof of claim:* We define $R \coloneqq \left\{N^-_c\left(N^+_{d_4}(v)\right) : d_4 \in D_4\right\}$. Since $|R| \leq k$ we have that $e_H(N^+_{D_2}(v), R) \leq 10k^3/n$, whence for all $d_4 \in D_4$ but a set $D_4^{\mathrm{bad}}$ of size at most $k/10$, there are at most $100k^2/n$ colours $d_2 \in D_2$ for which there is an arc from $N^+_{d_2}(v)$ to $N^-_c\left(N^+_{d_4}(v)\right)$. We now begin construction of the tuples $\lambda \in \Lambda$ by selecting $d_4 \in D_4$ avoiding any colours for which we have a $d_4$-loop at $v$, or $d_4 \in D_4^{\mathrm{bad}}$, or the $c$-arc with head $N^+_{d_4}(v)$ is saturated or a loop. Due to our conditioning on $\mathcal{C}^c_D$, we have at least $k/4 - 1 - k/1000 - k/16 - k/10 \geq k/20$ acceptable choices for $d_4$. Next we choose $d_1 \in D_1$ avoiding the colours of the arcs $vv$, $u_7u_7$, $vu_8$ (if they are present in $H$) so that we have at least $k/5$ acceptable choices. Now we choose $d_2 \in D_2$ avoiding the colours of 8 arcs which if chosen would cause us to give the label $u_6$ or $u_{10}$ to a vertex already labelled, and also avoiding any of the at most $100k^2/n = n/10^{10}$ choices of $d_2$ for which there is an arc from $u_{10}$ to $u_8$. (This ensures that $u_6$ and $u_{10}$ are distinct vertices.) There are at least $k/5$ acceptable choices for $d_2$. There are at most 21 colours $d_3 \in D_3^{\mathrm{good}}$ we must avoid for relabelling reasons. Our choice of $d_3$ may cause $u_3 = u_4$ and/or $u_{11} = u_{12}$, or instead may cause $u_3 = u_{11}$ and/or $u_4 = u_{12}$, but this seems difficult to avoid, and does not cause any problems. Thus there are at least $k/12$ acceptable choices of $d_3 \in D_3^{\mathrm{good}}$. We now choose $f_1$ to be a non-loop $d_3$-arc (recall condition (iii) of Claim 1) with neither endvertex labelled so far. Choosing $f_2$ similarly, we conclude that $|\Lambda| \geq \frac{k}{20} \cdot \frac{k}{5} \cdot \frac{k}{5} \cdot \frac{k}{12} \cdot \frac{19n}{20} \cdot \frac{19n}{20} \geq \frac{k^4n^2}{10000}$. $-$

For each tuple $\lambda = (d_1, d_2, d_3, d_4, f_1, f_2) \in \Lambda$, define $T_\lambda$ to be the subgraph of $H$ with vertex set $V(T_\lambda) = \{v, u_1, u_2, \dots, u_{14}\}$ and arcs $u_2u_1$, $u_3u_5$, $u_6u_4$, $u_5u_7$,

$u_6u_8$, $u_8u_7$, $vu_7$, $u_9v$, $vu_{10}$, $u_9u_{11}$, $u_{12}u_{10}$, $u_{13}u_{14}$ (with the vertices $u_i$ defined as in Claim 2).

*Claim 3: There is a set $\Lambda^* \subseteq \Lambda$ of size at least $|\Lambda|/2$ such that each $\lambda \in \Lambda^*$ satisfies the following properties:*

($Q1$) *$T_\lambda$ is a twist system of $H$;*

($Q2$) *deleting the six $d_3$-arcs in $T_\lambda$ does not decrease $r(H)$;*

($Q3$) *the canonical $(v, c, \mathcal{P})$-gadget of the twist $\mathrm{twist}_{T_\lambda}(H)$ is distinguishable, and it is the only $(v, c, \mathcal{P})$-gadget that is in $\mathrm{twist}_{T_\lambda}(H)$ but not in $H$.*

*Proof of claim:* For all $\lambda \in \Lambda$, by the definition of $u_1, u_2, \ldots, u_{14}$ and the result of Claim 2, we have that $T_\lambda$ satisfies the analogues of [56, Definition 6.7(i)–(vi)]. Thus, to check that there is a set $\Lambda_1 \subseteq \Lambda$ of size $|\Lambda_1| \geq 9|\Lambda|/10$ such that each $\lambda \in \Lambda_1$ satisfies ($Q1$), it suffices to check that at most $|\Lambda|/10$ tuples $\lambda \in \Lambda$ fail to satisfy the analogue of [56, Definition 6.7(vii)]. To do this, we use the analogue of [56, Lemma 6.8, Equation (6.2)] in the same way as in the proof of [56, Lemma 6.8, Claim 2], so we omit the details here.

An analogue of [56, Lemma 6.8, Claim 3] (with the same proof) now proves that at most $|\Lambda|/10$ tuples $\lambda \in \Lambda_1$ fail to satisfy ($Q2$), and analogues of [56, Lemma 6.8, Claims 4 and 5] (with the same proof) address ($Q3$). We remark that we do not require an analogue of [56, Lemma 6.8, Claim 6] in this setting. $-$

Observe that each twist adds only six arcs, and that each $\lambda \in \Lambda^*$ produces a unique twist system $T_\lambda \subseteq H$. Thus, from Claims 2 and 3 we deduce that if $s \leq k^4/2^{16}n^2$ then $\delta_s \geq |\Lambda^*| \geq k^4n^2/20000$, whence either $T_s = \emptyset$ or $|T_s|/|T_{s+1}| \leq \Delta_{s+1}/\delta_s \leq 2/3$, where $T_s$ is the set of $H \in \widetilde{\mathcal{Q}}_D \cap \mathcal{C}_D^c$ for which $r(H) = s$. We now obtain $\mathbb{P}_D\left[r_{(v,c,\mathcal{P})}(\mathbf{H}) \leq n^2/2^{100} \,\middle|\, \widetilde{\mathcal{Q}}_D \cap \mathcal{C}_D^c\right] \leq \exp(-\Omega(n^2))$ by repurposing the calculations at the end of the proof of [56, Lemma 6.8]. By a union bound over $v \in [n]$, (4.8.1) holds, which completes the proof of the lemma. $\square$

The following lemma does most of the analogous work to the proof of [56, Lemma 3.8].

**Lemma 4.8.2.** *Fix $c \in [n]$. Then we have $\mathbb{P}[\mathcal{E}^c \mid \mathcal{C}^c] \geq 1 - \exp(-\Omega(n^2))$.*

**Proof.** Arbitrarily select $D \subseteq [n]$ such that $c \in D$ and $|D| = n/10^6$. Define $\mathcal{F}$ to be the collection of all sets $F$ of $n$ arcs of $\overleftrightarrow{K_n}$ such that every vertex of $\overleftrightarrow{K_n}$ is the head of precisely one arc in $F$ and the tail of precisely one arc in $F$, and define $\mathcal{F}^{\text{good}} \subseteq \mathcal{F}$ to be the set of $F \in \mathcal{F}$ such that $F$ contains at most $n/10^9$ loops. Note that

$$\begin{aligned}
\mathbb{P}_D\left[\overline{\widetilde{\mathcal{Q}}_D} \mid \mathcal{C}_D^c\right] &\leq \mathbb{P}_D\left[\overline{\mathcal{Q}_D^1} \mid \mathcal{C}_D^c\right] = \sum_{F \in \mathcal{F}^{\text{good}}} \mathbb{P}_D\left[\mathbf{F}_c = F \mid \mathcal{C}_D^c\right] \mathbb{P}_D\left[\overline{\mathcal{Q}_D^1} \mid \mathbf{F}_c = F\right] \\
&\leq \exp(-\Omega(n^2)),
\end{aligned} \tag{4.8.2}$$

where we have used Lemma 4.5.4 and the fact that $\mathcal{Q}_D^1 \subseteq \widetilde{\mathcal{Q}}_D$. Now using Lemma 4.8.1, (4.8.2), and the law of total probability, we obtain that

$$\mathbb{P}_D\left[\overline{\mathcal{E}_D^c} \mid \mathcal{C}_D^c\right] \leq \mathbb{P}_D\left[\overline{\mathcal{E}_D^c} \mid \widetilde{\mathcal{Q}}_D \cap \mathcal{C}_D^c\right] + \mathbb{P}_D\left[\overline{\widetilde{\mathcal{Q}}_D} \mid \mathcal{C}_D^c\right] \leq \exp(-\Omega(n^2)). \tag{4.8.3}$$

By Proposition 4.4.4 and (4.8.3), we obtain

$$\begin{aligned}
\mathbb{P}\left[\overline{\mathcal{E}^c} \mid \mathcal{C}^c\right] &\leq \mathbb{P}\left[\overline{(\mathcal{E}^c|_D)} \mid \mathcal{C}^c\right] = \frac{\sum_{H \in \overline{\mathcal{E}_D^c} \cap \mathcal{C}_D^c} \operatorname{comp}(H)}{\sum_{H' \in \mathcal{C}_D^c} \operatorname{comp}(H')} \\
&\leq \mathbb{P}_D\left[\overline{\mathcal{E}_D^c} \mid \mathcal{C}_D^c\right] \cdot \exp\left(O\left(n \log^2 n\right)\right) \leq \exp\left(-\Omega\left(n^2\right)\right),
\end{aligned}$$

completing the proof of the lemma. $\square$

Lemma 4.5.6 now follows very quickly from Lemmas 4.8.1 and 4.8.2.

**Proof of Lemma 4.5.6.** For fixed $c \in [n]$, we can use Lemma 4.8.2 to see that

$$\mathbb{P}\left[\overline{\mathcal{E}^c} \mid \mathcal{C}\right] = \frac{|\overline{\mathcal{E}^c} \cap \mathcal{C}|}{|\mathcal{C}|} \leq \frac{|\overline{\mathcal{E}^c} \cap \mathcal{C}^c|}{|\mathcal{C}^c|} \cdot \frac{|\mathcal{C}^c|}{|\mathcal{C}|} \leq 2\mathbb{P}\left[\overline{\mathcal{E}^c} \mid \mathcal{C}^c\right] \leq \exp\left(-\Omega\left(n^2\right)\right),$$

where we have used $|\mathcal{C}^c| \leq |\Phi(\overleftrightarrow{K_n})|$ and $|\mathcal{C}| \geq |\Phi(\overleftrightarrow{K_n})|/2$, the latter following from Lemma 4.4.6. A union bound over $c \in [n]$ now gives that $\mathbb{P}\left[\overline{\mathcal{E}} \mid \mathcal{C}\right] \leq \exp\left(-\Omega\left(n^2\right)\right)$, as desired. The final claim in the statement now follows by applying Lemma 4.4.6 with $t = n/10^9$, and using the law of total probability. $\square$

CHAPTER 5

# ADVANCING THE SEMI-RANDOM METHOD: AN UPCOMING APPROXIMATELY OPTIMAL NIBBLE-TYPE RESULT

In this chapter we describe an upcoming result based on joint work with Tom Kelly, which represents a substantial strengthening and generalisation of many known results in the field of matchings in approximately regular hypergraphs. The paper itself is not quite ready to be included in the present thesis, as we are still in the process of investigating how much further we can optimise the main result by making small tweaks to the proof, and we are still exploring applications of our result. Further, the word limit of this thesis prevents the inclusion of the proof of our main result (Theorem 5.2.1). We anticipate that the paper will be submitted for publication within the first half of 2023.

In Section 5.1, we give a short introduction to the general ideas and applicability of the 'semi-random method' (used essentially interchangeably with the 'Rödl Nibble'), and cover the progression of the key results involving using the semi-random method to find large matchings in approximately regular hypergraphs, with emphasis towards more efficiently using the known information on the codegrees to produce matchings with smaller leftover. In Section 5.2, we state our result (Theorem 5.2.1), and discuss the extent to which it strengthens and generalises the results presented in Section 5.1. We discuss the ways in which our result could yet be strengthened even further in Section 5.3. In Section 5.4, we discuss an application of Theorem 5.2.1 to the study of 'almost-Hamilton cyclical partial transversals' in general Latin squares, which ties in naturally with [55] (Chapter 4 in the present thesis), and indeed was the original motivation behind Theorem 5.2.1. Finally, in Section 5.5, we sketch briefly the key elements of the proof of Theorem 5.2.1.

## 5.1 Introduction

### 5.1.1 Ideas and applicability

The 'Rödl Nibble' (usage of which is generally referred to as the 'semi-random method') is a method introduced by Rödl [108] in 1985, motivated by the study of the existence of designs (see [41, 76] for more details). Roughly speaking, the semi-random method is useful for finding very large, typically almost-spanning substructures within combinatorial host structures. If one seeks, for instance, an almost-perfect matching in a dense $k$-uniform hypergraph, and one simply chooses $n/k$ edges randomly at once (independently with equal probability, say) then one likely will need to delete many of the obtained edges to obtain a subset which is a matching, whence the obtained matching is not close to spanning. Rödl's key observation in [108] (building on some ideas of Ajtai, Komlós, and Szemerédi [3]) is that if instead one only chooses a *small* set of edges at random ('takes a *nibble*'), then not only does one only have to 'clean up' a much smaller proportion of the resulting set of edges to obtain a matching, but the remaining hypergraph still looks very close to the original hypergraph. One may therefore iterate this process, piecing together all the obtained small matchings to yield the desired almost-perfect matching, with the process only needing to stop once the accumulated errors are too large to continue iterating within the ever-smaller leftover hypergraph. This methodology applies equally to, for example, edge-colourings [104, 70, 73], list-edge-colourings [69, 100], independent sets [3, 2], and thus vertex colourings [82, 68, 99]. For a more thorough exposition of these topics, see for example the survey of Kang, Kelly, Kühn, Methuku, and Osthus [72].

Throughout this chapter, we focus only on use of the semi-random method to obtain large matchings in hypergraphs, though we suspect the methodology we use in the proof of Theorem 5.2.1 transfers naturally to the use of the semi-random method for the study of other structures. Perhaps the true beauty of hypergraph matching results is their applicability. Indeed, the proof of the main result of Chapter 3 (namely Theorem 3.1.3) uses a hypergraph matching result (namely Theorem 3.5.3) in two places: in the construction of an appropriate 'absorber', and also as an important part of the proof of Lemma 3.4.10 (see [54]), which provides the bulk of the rainbow Hamilton path (or cycle) given by Theorem 3.1.3. In each of these applications, the key is that if one seeks an almost-spanning structure in a 'smooth' graph or hypergraph $G$, one can often construct an almost-regular auxiliary hypergraph $\mathcal{H}$, where an almost-perfect matching of $\mathcal{H}$ *corresponds* to the desired almost-spanning structure in $G$ (recall for instance the proof of Lemma 3.6.5). This is a general phenomenon, indeed prompting Füredi to remark that "almost all combinatorial questions can be reformulated as either a matching or covering problem of a hypergraph" in his survey [47] into results on such problems.

### 5.1.2 Improving the asymptotics of the leftover

For the remainder of Section 5.1, we give a brief history of the results which use the semi-random method to produce large matchings (thus having small 'leftover') in regular or almost-regular hypergraphs. We focus primarily on the quest to increase the size of obtained matching.

In 1985, Frankl and Rödl [46] proved (under more general conditions than in [108]) that a $D$-regular, $n$-vertex hypergraph whose codegrees are much smaller than $D$ (say at most $D/\log^4 n$), admits a matching covering all but at most $o(n)$ vertices. Here, we say the *codegree* of a distinct pair of vertices is the number of edges containing that pair. We often denote by $C$ the maximum codegree among all pairs of vertices, so that Frankl and Rödl's [46] hypothesis is $C \leq D/\log^4 n$. This is the first result which clearly exhibits the following phenomenon: One can use 'space' between the degree $D$ and the maximum codegree $C$ (in the sense that $D/C$ is large) to find a matching with small leftover. This is not a surprise, since larger values of $D/C$ mean that the vertex degrees are more independent of one another during the analysis of a random nibble, and this usually means that we obtain smaller error and tighter concentration in a single nibble. In turn, this allows one to iterate the process for longer, providing a larger matching, and smaller final leftover. Pippenger (unpublished, see [47] for a proof) weakened the hypothesis of Frankl and Rödl [46] on the space between the codegrees whilst providing the same leftover.

**Theorem 5.1.1** (Pippenger, unpublished)**.** *Suppose that $1/D \ll \delta \ll \varepsilon, 1/k \leq 1$. Let $H$ be a $(k+1)$-uniform, $D$-regular hypergraph with $n$ vertices. If $C \leq \delta D$, then there is a matching $M$ of $H$ covering all but at most $\varepsilon n$ vertices.*

In many applications it is useful to obtain a stronger bound on the number of uncovered vertices; that is, to improve the asymptotics $o(n)$ given by Theorem 5.1.1 for the size of the leftover hypergraph. Grable [57] achieved this, as follows.

**Theorem 5.1.2** (Grable [57], 1999)**.** *Let $H$ be a $(k+1)$-uniform, $D$-regular hypergraph with $n$ vertices, and suppose that $C = o(D/\log n)$ (as $n \to \infty$). Then there is a matching $M$ of $H$ covering all but at most $O(n(C \log n/D)^{1/(2k+1+o(1))})$ vertices.*

In 1997 (appearing before [57] reached publication), Kostochka and Rödl greatly improved upon Theorem 5.1.2 (relying on and generalising work of Alon, Kim, and Spencer [6] on hypergraphs with maximum codegree 1) by obtaining the following result.

**Theorem 5.1.3** (Kostochka and Rödl [85], 1997)**.** *Suppose that $1/D \ll 1/k \leq 1$, and let $0 < \delta, \gamma < 1$ be fixed real numbers. There is $c = c(k, \delta, \gamma)$ such that the following holds. Let $k \geq 2$ and let $H$ be a $(k+1)$-uniform, $D$-regular hypergraph on $n$ vertices with maximum codegree $C \leq D^{1-\gamma}$. Then there is a matching $M$ of $H$ which covers all but at most $cn(C/D)^{1/k-\delta}$ vertices.*

Note that the hypothesis on the size of $D/C$ is a little stronger in Theorem 5.1.3 than in 5.1.2, assuming $D$ is at least polynomial in $n$. Ignoring for now the $c(D/C)^{\delta}$ error, we remark that producing leftover $n(C/D)^{1/k}$ in some sense constitutes the 'best usage that one can expect' of the space between $D$ and $C$ to produce matchings with small leftover, without any extra assumptions on $H$. Indeed, the 'leftover set' of vertices uncovered by the matching produced by the semi-random method, if it has size approximately $pn$, say, can heuristically be thought of as being similar to a set chosen 'binomially', each vertex independently being uncovered with probability $p$. But in such a scheme, the expected degree of a leftover vertex is roughly $Dp^k$, whence $p = (C/D)^{1/k}$ corresponds to leftover vertices having degree roughly $C$. Once the leftover vertices have degree as low as this, then it is impossible to successfully apply concentration arguments during any further iterations of the Nibble, since as discussed above, the space between $D$ and $C$ in some sense measures the independence between each vertex's degree. Thus, classical Rödl Nibble arguments cannot push past this point.

The next key result in this direction, and still the state-of-the-art, was provided by Vu [122] in 2000. We describe this result now. For a set $W \subseteq V(H)$ of at least two vertices in a $(k+1)$-uniform hypergraph $H$, we say that the *codegree* of $W$ (in $H$) is the number of edges of $H$ containing $W$. For $2 \leq j \leq k+1$, the *maximum $j$-codegree* of $H$, denoted $C_j(H)$, is defined to be the maximum codegree amongst all vertex subsets of size $j$. Vu [122] used the 'Polynomial Method' (which we cover briefly in Section 5.5.1) during the concentration arguments within a random nibble to make the following key observation: If one knows useful values of $C_j(H)$ for values $j > 2$, so for instance one has information on the 'codegree sequence' $D, C_2(H), C_3(H), \ldots, C_s(H)$ for some $s \leq k+1$, then actually the space between $C_{s-1}(H)$ and $C_s(H)$ can be used more efficiently to create matchings with smaller leftover, than the space between $D$ and $C_2(H)$. Indeed, the larger one can take $s$ (ie. the more information one has about the codegree sequence), the larger the matching one finds using the semi-random method.

**Theorem 5.1.4** (Vu [122], 2000)**.** *Suppose that $1/D \ll 1/k \leq 1$. There is $c = c(k)$ such that the following holds. Let $H$ be a $(k+1)$-uniform, $D = D_1$-regular hypergraph on $n$ vertices, and suppose that there is $2 \leq s \leq k+1$ and quantities $D_j$ for $2 \leq j \leq s$ such that $C_j(H) \leq D_j$ for all $2 \leq j \leq s$, and $D_s \geq 1$. Assume further that there is $x > 0$ such that the following conditions hold:*

(i) $x^3 \leq \frac{D_j}{D_{j+1}}$ *for all $1 \leq j \leq s-1$;*

(ii) $x^{k-s+2} \leq \frac{D_{s-1}}{D_s}$.

*Then there is a matching $M$ of $H$ which covers all but at most $n(\log^c n)/x$ vertices.*

Notice that if $s = 2$, then $x^{k-s+2} = x^k$, so that in this case condition (ii) of Theorem 5.1.4 is $x \leq (D/C)^{1/k}$, whereas for larger $s$, $x^{k-s+2} < x^k$, whence condition (ii) gives a less strict inequality on $x$ (assuming $D/D_2$ and $D_{s-1}/D_s$ are comparable in size, as they often are in applications) and a larger $x$ yields a smaller leftover. Notice also that the $s = 2$ case of Theorem 5.1.4 is a strict strengthening of Theorem 5.1.3 for large hypergraphs, because the error has been reduced from polynomial to logarithmic, and Theorem 5.1.4 does not make the extra hypothesis $C \leq D^{1-\gamma}$ on the space between $D$ and $C$. The intuition behind Theorem 5.1.4 is that, if one has information about codegrees beyond $C_2(H)$, then the degradation of $C_2(H)$ can also be concentrated (using $C_3(H) < C_2(H)$) during a random nibble, and carefully tracked throughout the iteration of the nibble. Since the gradual degradation of $D, C_2(H), \ldots, C_{s-1}(H)$ can all be monitored in this way, one is only forced to abandon the process when $C_{s-1}(H)$ becomes degraded to the point where it is not usefully larger than $C_s(H)$ (we sometimes call this 'collapsing $C_{s-1}(H)$ to $C_s(H)$' during this chapter), and we have not been able to decrease $C_s(H)$ at any point in the process. By this time, one has been able to push the process further, decreasing $D$ far past the point at which the stationary $C_2(H)$ stopped being useful in the proof of Theorem 5.1.3, using more of the hypergraph, leaving fewer vertices uncovered by the matching. We remark that Vu [122] comments that it is possible to modify his proof such that the $x^3$ expression in condition (i) can be replaced by the improved $x^2$, but he did not provide a proof of this improved theorem.

## 5.2 Main result

We are now ready to state our main result. We show that, provided the codegree sequence isn't too unusual and one doesn't run into other hard bottlenecks, one can use *all* of the space in the *entire known codegree sequence* to reduce the set of vertices uncovered by the matching. In most circumstances, this represents a substantial strengthening of Theorem 5.1.4. Since it will help us to compare the strength of Theorem 5.2.1 to Theorem 5.1.4, we remark briefly that the rough idea behind the proof of Theorem 5.2.1 is that one can 'collapse $D_{s-1}$ to $D_s$', as in Theorem 5.1.4, and track the effect this has on earlier codegrees, but then one can 'forget' $D_s$, effectively setting the new $s$ to be $s-1$, then collapse $D_{s-2}$ to $D_{s-1}$, and repeat until one runs out of codegrees to forget. For a more thorough sketch of the proof, see Section 5.5.

### 5.2.1 Statement and optimality

We say a hypergraph $H$ is *$(n, D, \varepsilon)$-regular* if $H$ has $n$ vertices, and $(1-\varepsilon)D \leq \deg(v) \leq (1+\varepsilon)D$ holds for all $v \in V(H)$. We have a proof of the following.

**Theorem 5.2.1** (Main Result, Gould and Kelly, 2023+). *Suppose that $1/D \ll \gamma \ll 1/k \leq 1$. There is $A = A(\gamma)$ such that the following holds. Let $H$ be a $(k+1)$-uniform, $(n, D, \varepsilon)$-regular hypergraph, and suppose that there is $2 \leq s \leq k+1$ and quantities $D_j$ for $2 \leq j \leq s$ such that $C_j(H) \leq D_j$ for all $2 \leq j \leq s$, and $D_s \geq 1$. Set*

$$x := \min\left\{\left(\frac{D}{D_s}\right)^{1/k}, \left(\frac{D}{D_2}\right)^{1/2}, \frac{1}{\varepsilon}\right\},$$

*and suppose that the following conditions hold:*

(i) $\frac{D_j}{D_{j+1}} \geq \left(\frac{D_j}{D_s}\right)^{1/(k-j+1)}$ *for all* $2 \leq j \leq s-2$*;*

(ii) $x \geq \log^A D$.

*Then there is a matching $M$ of $H$ which covers all but at most $n/x^{1-\gamma}$ vertices.*

We remark that if $x$ takes the value $(D/D_s)^{1/k}$ in Theorem 5.2.1, then the size of the leftover is approximately $n(D_s/D)^{1/k}$, which, similarly to the discussion after Theorem 5.1.3, cannot be improved upon via classical Rödl Nibble methods. Indeed, if the leftover graph is this small, then the remaining vertices have expected degree approximately $D_s$, at which point future concentration arguments fail. Thus, the expression $(D/D_s)^{1/k}$ being the bottleneck for $x$ corresponds to 'using up all the space in the codegree sequence', to an extent which cannot be improved upon via classical Nibble, and it is in this sense that we claim that Theorem 5.2.1 is approximately 'optimal' (invoking also that one is free to feed into Theorem 5.2.1 *all* of the known codegrees). If $x$ takes one of the other two values in Theorem 5.2.1, this corresponds to the iteration process hitting some other natural bottleneck before managing to use all the space between $D$ and $D_s$; for example $x = 1/\varepsilon$ corresponds to the initial control of the relative vertex degree errors being too poor to sustain iteration for this long, given that one must anticipate the possible error growing a little with each Nibble. Though it is more difficult to intuitively pin down why or if $(D/D_2)^{1/2}$ is a naturally impassable bottleneck, we remark that it is a substantial improvement upon the corresponding bottleneck $(D/D_2)^{1/k}$ of Theorem 5.1.3 due to the independence from $k$ (and note that Theorem 5.1.3 does not apply for $k = 1$). Further, the bottleneck $x = (D/D_2)^{1/2}$ can be thought of as the $j = 1$ case of (the unproven improvement of) condition (i) of Theorem 5.1.4.

### 5.2.2 Comparison with Theorem 5.1.4

To give a concrete example which illustrates how strong Theorem 5.2.1 is compared to the other results already presented in this chapter (and how weak a hypothesis condition (i) often is, which we return to in Section 5.2.3), consider the following easy corollary of Theorem 5.2.1.

**Corollary 5.2.2.** *Suppose $1/D \ll \gamma \ll 1/k \leq 1$, and let $H$ be a $(k+1)$-uniform, $D$-regular hypergraph on $n$ vertices. Suppose further that $D/C_2(H) = \Theta(n)$, and $C_j(H)/C_{j+1}(H) = \Theta(n)$ for all $2 \leq j \leq \lceil k/2 \rceil$, and $C_{\lceil k/2 \rceil+1}(H) \geq 1$. Then $H$ has a matching which covers all but at most $n^{1/2+\gamma}$ vertices.*

**Proof.** Set $s := \lceil k/2 \rceil + 1$ (which satisfies $2 \leq s \leq k+1$ for all $k \in \mathbb{N}$), and put $D_j := C_j(H)$ for all $j \in [\lceil k/2 \rceil + 1] \setminus \{1\}$. We aim to apply Theorem 5.2.1. To check condition (i), we require $D_j/D_{j+1} \geq \Theta(n^{(s-j)/(k+1-j)})$ for all $2 \leq j \leq \lceil k/2 \rceil - 1$ (if such $j$ exist), but this holds very comfortably since the left hand-side is $\Theta(n)$, and the right hand-side is at most $O(\sqrt{n})$. We can set (say) $\varepsilon := 1/n$, and notice that $(D/D_2)^{1/2} = \Theta(\sqrt{n})$, and $(D/D_s)^{1/k} = \Theta(n^{\lceil k/2 \rceil \cdot 1/k}) = \Omega(\sqrt{n})$, so that $x = \Theta(\sqrt{n})$. Since $D \leq n^k$, condition (ii) is clear, and thus Theorem 5.2.1 gives a matching of $H$ covering all but at most $n/x^{1-\gamma} = \Theta(n^{1-1/2+\gamma/2}) \leq n^{1/2+\gamma}$ vertices. □

Theorem 5.1.4 applied to the $H$ in Corollary 5.2.2 can (due to condition (i) of that theorem) never give leftover smaller than approximately $\Theta(n^{2/3})$. It is easy to check that for $k \geq 4$, the improved version (with $x^3$ replaced by $x^2$ in condition (i)) of Theorem 5.1.4 would give a matching with leftover of size approximately $\Theta(n^{1-1/(\lfloor k/2 \rfloor+1)})$, which, in particular, is $\Omega(n^{2/3})$ for $k = 4$, and is larger for larger $k$. Indeed, for large $k$, the leftover given by the improved version of Theorem 5.1.4 for this $H$ has size approximately $n^{1-2/k}$, which is considerably larger than the $n^{1/2+\gamma}$ given by Theorem 5.2.1 in Corollary 5.2.2. Note that such a hypergraph $H$ arises naturally during the proof of Theorem 5.4.1, as we will discuss in Section 5.4. Indeed, we believe that many hypergraphs arising naturally in applications have the property that more codegrees than just $C_2(H)$ can be bounded, and often these bounds are such that the ratios $C_j(H)/C_{j+1}(H)$ are comparable (say, on the same asymptotic order of $n$). In such cases, Theorem 5.2.1 should apply and give a matching with very small leftover.

We remark that Theorem 5.1.4 (or at least, the claimed improvement) is never stronger than Theorem 5.2.1, except for the logarithmic error term (as opposed to our polynomial one). Further, generally speaking, there are only two situations in which Theorem 5.2.1 is not strictly stronger. The first of these is the case $s = 2$; in particular if $k \geq 2$ then Theorems 5.1.3, 5.1.4, and 5.2.1 align to give leftover approximately $n(C/D)^{1/k}$, which occurs since the Theorem 5.1.3 heuristic of 'using up all the space between $D$ and $C$', the Theorem 5.1.4 heuristic of 'using up all the space between $D_{s-1}$ and $D_s$', and the Theorem 5.2.1 heuristic of 'using up all the space between $D$ and $D_s$' align when $s = 2$. The second situation in which Theorem 5.2.1 is not strictly stronger than the improved version of Theorem 5.1.4 is when $s = k + 1$, or is very close to this maximum possible value, and $D_{s-1}/D_s$ is large with respect to $D/D_2$ (though it cannot be too large otherwise condition (i) becomes the bottleneck of Theorem 5.1.4). Indeed, for very large $s$, recall that the space between $D_{s-1}$ and $D_s$ can be used more effectively for reducing the leftover, and if this ratio also happens to be large, then during the process of collapsing $D_{s-1}$ to $D_s$ in the proof of Theorem 5.2.1, one may reach other bottlenecks before using up all of the space between $D_{s-1}$ and $D_s$. Thus, we are never able to begin collapsing other 'spaces' in the codegree sequence, which is what distinguishes Theorem 5.2.1 from Theorem 5.1.4. Outside of these extreme cases however, Theorem 5.2.1 will give strictly smaller leftover than Theorem 5.1.4, and the difference will often be substantial, as observed in the discussion following Corollary 5.2.2, above.

Finally, we note that hypothesis (i) of Theorem 5.2.1 is often a much weaker hypothesis on 'middle ratios' $D_j/D_{j+1}$ than the $j \neq 1$ case of hypothesis (i) of Theorem 5.1.4, in the following sense. Small ratios $D_j/D_{j+1}$ can present the bottleneck for $x$ in Theorem 5.1.4, without being too small for Theorem 5.2.1 to apply. Indeed, in the proof of Corollary 5.2.2, recall that some of the ratios $D_j/D_{j+1}$ would have been permitted to be $O(\sqrt{n})$. In fact, if $k$ is large, ratios $D_j/D_{j+1}$ for $j$ close to $k/2$ would have been permitted to be extremely small, say even $\Theta(n^{10/k})$, whilst not having a big effect on the leftover (i.e. $x$ still attains approximately $\sqrt{n}$) if the other ratios remain $\Theta(n)$. By contrast, the $x$ in (the improved) Theorem 5.1.4 would be forced to be at most the square root of such a ratio.

### 5.2.3 Remarks on hypothesis (i)

Note that condition (i) of Theorem 5.2.1 is effectively the statement that the sequence of codegree bounds $D_j$ is 'not too convex'. Early ratios are allowed to be larger than later ratios (to any extent), and depending on the relationship between $s$ and $k$, early ratios are also allowed to be smaller than later ratios, often by a considerable factor, but they cannot be *too* small with respect to later ratios. We note that if $D_j/D_{j+1} \geq D_{j+1}/D_{j+2}$ holds for all $2 \leq j \leq s-2$ then condition (i) is satisfied, for any $s$, but this is much stronger than is required if $s \leq k$. In general, one is not allowed as much convexity by (i) if $s$ is larger. This fact is, to an extent, an artifact of the proof, in that the proof of Theorem 5.2.1 is geared to being more effective for less extreme values of $s$. Though (i) is still not a particularly strict condition in the case $s = k+1$ (especially, as we will come back to shortly, as one is free to choose the values $D_j$ subject only to $C_j(H) \leq D_j$), one could tweak the proof of Theorem 5.2.1 to obtain a theorem which makes a weaker assumption on the codegree ratios if $s = k + 1$, and we may include this in the paper, but we do not discuss it further here.

We note that it is natural that a condition like (i) should be required in Theorem 5.2.1, in that the process of 'collapsing' $D_{s-1}$ to $D_s$ also entails a small amount of degradation to all ratios $D_j/D_{j+1}$ for $j < s - 1$, so we require that these earlier ratios are at least large enough to tolerate this damage. Further, though Theorem 5.1.4 does not explicitly state such a hypothesis on the 'shape' of the codegree sequence, its effect is hidden in the statement, because if one of the ratios of consecutive codegrees is so small that it violates the desired 'shape' of the sequence, then this ratio becomes the bottleneck for $x$ in Theorem 5.1.4, limiting the size of the matching one obtains.

Even in the event that the codegree sequence of $H$ is so 'bumpy' (in that some early ratios of consecutive codegrees are too small with respect to the later ratios) that condition (i) fails, there are several options for still applying the result. Firstly one has the flexibility to increase some of the values $D_j$ (at the cost of making $D_{j-1}/D_j$ smaller) since $C_j(H) \leq D_j$ will still be satisfied. Determining for which values $j$ condition (i) originally fails should assist in determining which values $D_j$ to increase. If one can successfully perform this redistribution in a way that does not increase $D_2$ or $D_s$, then the leftover given by Theorem 5.2.1 is still the same as it would have been if one could have applied the theorem immediately. However, if one is forced to increase $D_2$ or $D_s$ in such a way that the value bottlenecking $x$ becomes worse, then one obtains some 'damage' to the size of the leftover as a result of smoothing out the codegree sequence. Alternatively (or in addition), one can choose to 'forget' some codegrees, effectively choosing instead to consider a shorter codegree sequence, and a smaller $s$. Since (i) allows more convexity of the codegree sequence for smaller $s$, this effectively loosens this hypothesis. Again, if this process of forgetting codegrees results in no change to the value of the bottleneck for $x$, then one has still found an 'optimal' application of Theorem 5.2.1. Finally, we remark that the full strength of Theorem 5.2.1 is actually contained within Lemma 5.5.3, so that depending on the circumstances of $H$, one may choose instead to iterate Lemma 5.5.3 in a tailored way for that $H$. We made natural hypotheses on $H$ in order to give Theorem 5.2.1 in as general form as possible, but in some circumstances it may be prudent to 'manually' iterate Lemma 5.5.3 for a given $H$.

## 5.3 Possible additions to Theorem 5.2.1

In this section, we discuss three ways in which Theorem 5.2.1 may yet be added to, or improved.

### 5.3.1 Augmenting

Recently, several authors have made progress in answering the following question. If one can impose stronger hypotheses on the input hypergraph $H$, can one continue to push for smaller leftover even after the space between $D$ and $C$ is used up? In 2020, Kang, Kühn, Methuku, and Osthus [73] answered this question in the affirmative by adding the assumptions that $H$ is dense and that there is more space between $D$ and $C$ than was shown by Theorem 5.1.4 to be necessary otherwise.

**Theorem 5.3.1** (Kang, Kühn, Methuku, and Osthus [73], 2020)**.** *Let $k \geq 3$, let*

*$0<\gamma,\mu<1$, and let $0<\eta<\frac{k-2}{k^4+k^3-2k^2+2k}$. Then there exists $n_0=n_0(k,\gamma,\eta,\mu)$ such that the following holds for all $n\geq n_0$ and $D\geq\exp(\log^{\mu} n)$. Let $H$ be a $(k+1)$-uniform, $D$-regular hypergraph on $n$ vertices with $C_2(H)\leq D^{1-\gamma}$. Then $H$ contains a matching covering all but at most $n(C/D)^{1/k+\eta}$ vertices.*

The key to the proof of Theorem 5.3.1 is that one can concentrate and track, in addition to the degradation of the usual hypergraph parameters, the continued existence and distribution of a useful set of subgraphs called 'augmenting stars', whose purpose is to augment and improve upon the matching obtained at the end of the iteration of the classical Rödl Nibble process.

Also in 2020, Keevash, Pokrovskiy, Sudakov, and Yepremyan [79] used the semi-random method, in conjunction with a similar augmentation strategy, to prove that certain *linear* (i.e. $C=1$) hypergraphs with very strong quasirandomness and expansion properties admit a matching covering all but at most $c\log n/\log\log n$ vertices, for some constant $c$. More precisely, they first used the semi-random method to find a matching leaving at most $o(n)$ vertices uncovered, with the added result that the matching has very strong 'expansion' properties, then repeatedly used these expansion properties to shuffle and increase the matching, until the leftover is remarkably small. Indeed, this was the key approach in their paper which provides the still best-known result towards Conjecture 4.1.1 (Latin squares are readily seen to be equivalent to linear 3-partite hypergraphs with the required strong quasirandomness and expansion properties to apply their augmented semi-random approach).

It is an interesting question to ask if one could improve Theorem 5.2.1 in a similar way under some added assumptions; i.e. could one assume say, an extra hypothesis on the amount of space available in the codegree sequence, and use the tracking of augmenting subgraphs to push past the barriers provided by our main results? Could one even use stronger quasirandomness or expansion assumptions, in conjunction with the methodology within the proof of Theorem 5.2.1, say, to prove that less 'well-spread' structures than Latin squares still admit matchings covering all but a logarithmic number of vertices?

### 5.3.2 Quasirandom and conflict-free matchings

It is often desirable in applications to have that the matching found by the semi-random method is itself in some way quasirandom, or that the leftover is quasirandom. For example, note that such a theorem of Alon and Yuster [8] (Theorem 3.5.3) was essential in Chapter 3 not only because it gave a hypergraph matching with small leftover, but because it shows that the leftover is also well-distributed with respect to any (not too large) collection of vertex sets from the host hypergraph. In the proof of Lemma 3.6.5, we used this well-distributed property of the leftover to ensure that the vertices, colours, and desired 'link endpoints' remaining from the large hypergraph matching (which found most of the desired 'links' in the absorber) were well-distributed enough for us to greedily finish the construction of the links.

In 2020, Ehard, Glock, and Joos [39] strengthened Theorem 3.5.3 (in the sense that the obtained matching has stronger quasirandomness properties), given a stronger assumption on the space between the (maximum) degree $D$ and the codegree $C$. The following result follows quickly from [39, Theorem 1.2].

**Theorem 5.3.2** (Ehard, Glock, and Joos [39], 2020)**.** *For every $k \geq 1$ and $\delta \in (0, 1)$,* *there exists $D_0$ such that the following holds for all $D \geq D_0$. Set $\gamma := \delta/50k^2$, let $H$ be a $(k+1)$-uniform hypergraph, and let $\mathcal{F}$ be a collection of subsets of $V(H)$ such that $|\mathcal{F}| \leq \exp(D^{\gamma^2})$ and $\sum_{v\in S} d(v) \geq (k+1)D^{1+\delta}$ for every $S \in \mathcal{F}$. Suppose that the following conditions hold:*

(i) $\Delta(H) \leq D$;

(ii) $C_2(H) \leq D^{1-\delta}$;

(iii) $e(H) \leq \exp(D^{\gamma^2})$.

*Then there is a matching $M$ of $H$ such that every $S \in \mathcal{F}$ satisfies $|S \cap V(M)| = (1 \pm D^{-\gamma}) \sum_{v\in S} d(v)/D$. In particular, if $H$ is $D$-regular and $V(H) \in \mathcal{F}$, then $M$ covers all but at most $nD^{-\gamma}$ vertices.*

We remark further that [73, Theorem 7.1] is a 'quasirandom version' of Theorem 5.3.1, and also constitutes a strengthening of Theorem 3.5.3.

Instead of seeking a matching which is 'well-spread' or 'quasirandom' with respect to collections of vertex sets as above, several authors have recently been interested in the more complicated notion of finding matchings which avoid 'forbidden configurations', or are 'conflict-free'. For example, Glock, Kühn, Lo, and Osthus [52], and independently Bohman and Warnke [13], approximately confirmed a conjecture of Erdős [40] on high-girth, high-density partial Steiner triple systems, by randomly constructing a partial Steiner triple system, one edge at a time, whilst maintaining the 'high-girth' property; that is, avoiding short 'cycles'. As observed by Delcourt and Postle [35] and Glock, Joos, Kim, Kühn, and Lichev [51], the aforementioned work on this problem could be formulated in terms of finding large hypergraph matchings in which one wishes to avoid the configurations corresponding to short cycles in the partial Steiner triple system. The authors of [35] and [51] (independently proving similar results to each other) generalized the work of [52] and [13], by finding large matchings which are free from a general given set of conflicts (of which the above short cycles would be a specific example) in hypergraphs $H$ which are approximately $D$-regular, and have $C_2(H)$ bounded away from $D$. In particular, we note that the proof in [35] uses the Rödl Nibble (as opposed to the random greedy algorithm used in [51]), and allows for more sparse hypergraphs, though [51] entails counting results not present in [35].

One wonders if such ideas could be incorporated into the proof of Theorem 5.2.1. As discussed, quasirandom and/or conflict-free matchings are often useful in applications, and such a result would combine these notions with the key property of our main result that the leftover is just so small.

### 5.3.3 Logarithmic error

It would be an improvement to Theorem 5.2.1 if one could replace the polynomial error $x^\gamma$ with a polylogarthmic error, in a similar way to which the $s = 2$ case of Theorem 5.1.4 improves Theorem 5.1.3. We are not attempting to achieve this in our upcoming paper, but we claim that we know how one could do this, if one has a slightly better concentration inequality for the key step of our Nibble process, than the concentration inequalities we have been able to find. We discuss this more concretely in Section 5.5.3.

## 5.4 A thematic application of Theorem 5.2.1

In this section, we present an application of Theorem 5.2.1 to the setting of almost-spanning cyclical partial transversals in Latin squares, which, by the transformation described in Section 4.1, corresponds to rainbow almost-Hamilton directed cycles in $G \in \Phi(\overleftrightarrow{K_n})$. This application (Theorem 5.4.1) will appear in our upcoming paper on Theorem 5.2.1, and was our original motivation. Recall that the main result of Chapter 4 concerns finding (many) rainbow directed *Hamilton* cycles in *almost all* $G \in \Phi(\overleftrightarrow{K_n})$. Here, we pursue *almost-Hamilton* rainbow directed cycles in *general* $G \in \Phi(\overleftrightarrow{K_n})$.

Recall (from Section 4.1) that Conjecture 4.1.3 is equivalent to the statement that any $G \in \Phi(\overleftrightarrow{K_n})$ has a rainbow directed linear forest with at least $n-2$ arcs, and that Gyárfás and Sárközy [60] proved that this is at least approximately true, obtaining such a linear forest with at least $n - O(n \log\log n/ \log n)$ arcs. One is naturally interested in improving the asymptotic error bound, and indeed Benzing, Pokrovskiy, and Sudakov [11] used path exchange arguments to improve the error to $O(n^{2/3})$. Further, they again used path exchange arguments to show that if one is willing to sacrifice some of the tightening of the error, one can even 'close' the rainbow directed linear forest into a rainbow directed cycle, obtaining such a cycle on all but $O(n^{4/5})$ vertices.

In our upcoming paper, we will show that applying Theorem 5.2.1 quickly yields that all $G \in \Phi(\overleftrightarrow{K_n})$ have a rainbow directed linear forest on $n - O(n^{1/2+\delta})$ arcs, for any constant $\delta > 0$. Further, we use some brief Polynomial Method arguments (see Section 5.5.1) to close such a linear forest into a cycle, without changing the error.

**Theorem 5.4.1** (Gould and Kelly, 2023+)**.** *Let $\delta > 0$ and suppose that $1/n \ll \delta$.*

*All $G \in \Phi(\overleftrightarrow{K_n})$ have a rainbow directed cycle which contains all but at most $n^{1/2+\delta}$ vertices. Equivalently, all $n \times n$ Latin squares have a cyclical partial transversal with at least $n - n^{1/2+\delta}$ positions.*

To be more specific about our application of Theorem 5.2.1 to this setting, we note that the approach is to first randomly reserve a set of $n^{1/2+\delta}$ colours, leaving the host graph $G$ essentially entirely intact. We use the Polynomial Method to argue, quite briefly, that this random set of colours has the desired absorption properties we require, for any almost-spanning rainbow directed linear forest in the remainder of the host graph. One then uses random partitioning arguments in the main part of $G$ to split the graph into a very 'smooth', very regular multipartite structure, at which point one can construct an auxiliary hypergraph $H$ which captures the information of this structure, such that an almost-perfect matching in $H$ corresponds to the desired linear forest in $G$. It then suffices to use the accuracy of the random partitioning arguments to justify that $H$ is sufficiently smooth to apply Theorem 5.2.1. In particular, edges in $H$ actually correspond to rainbow directed $(\ell+1)$-paths in $G$, and contain an $H$-vertex for each of the $\ell$ internal vertices, for each of the $\ell+1$ colours, and a vertex to signify the ordered pair of endpoints of this path. Here, $\ell$ is some large constant. Thus, $H$ is $(2\ell+2)$-uniform. One uses the random partitioning arguments to check that $C_j(H)/C_{j+1}(H) = \Theta(n)$ for $2 \leq j \leq \ell$, and similarly $D/C_2(H) = \Theta(n)$. Then, applying Theorem 5.2.1 to $H$ (with $s \coloneqq \ell + 1$) yields a matching with leftover $O(n^{1/2+\gamma})$ for appropriate $\gamma$. One then uses the absorption properties of the randomly reserved colours to 'close' the obtained rainbow directed linear forest into a rainbow directed cycle. Note that this $H$ is similar to the hypergraph in Corollary 5.2.2 (though $H$ here is not exactly regular, one is able to use the random partitioning arguments to check that $\varepsilon$ is not large enough to be the bottleneck for $x$ when applying Theorem 5.2.1).

We remark that, since one can obtain an element of $\Phi(\overleftrightarrow{K_n})$ from an optimal edge colouring of $K_n$ by replacing each edge with a monochromatic digon, and adding a loop at each vertex with the only remaining colour available at that vertex, Theorem 5.4.1 implies the (optimal colourings case of the) theorem of Balogh and Molla [10] in the undirected setting, up to our polynomial error term. This provides an alternate proof (up to that error) with the added benefit of the inherent flexibility of the Nibble (to, for example, push for stronger results as discussed in Sections 5.3.1 and 5.3.2), as opposed to the path exchange arguments employed by Balogh and Molla [10].

### 5.4.1 Further applications

We believe that the approach outlined above for applying Theorem 5.2.1 can likely be applied quite generally to situations in which the host combinatorial structure is in some sense 'very smooth'. When equipped with such information, the key to situationally applying Theorem 5.2.1 will be in the defining of an auxiliary hypergraph $H$ about which reasonable information on (ideally many) codegrees can be gleaned. One topic that we would like to highlight is the following claimed result of Kim [81] on partial Steiner systems. A partial Steiner system $S_p(n, k, t)$ is a collection of $k$-sets from a global $n$-set such that any $t$-set is contained in at most one such $k$-set, so that for instance, an $S_p(n, 3, 2)$ is a partial triangle-decomposition of $K_n$.

**Theorem 5.4.2.** *Let $2 \leq t \leq k-2$. There is a partial Steiner system $S_p(t, k, n)$*

*of size at least*

$$\left(1 - O\left(\left(\frac{\log n}{n}\right)^{-(k-t)/\left(\binom{k}{t}-1\right)}\right)\right)\binom{n}{t}/\binom{k}{t}.$$

We remark that Theorem 5.4.2 appeared in a conference proceedings, and to our knowledge a proof was never provided by Kim. The breakthrough results of Keevash [76, 77] and Glock, Kühn, Lo, and Osthus [53] on the existence of designs are much stronger results (even if the divisibility conditions don't apply, one can use these results to obtain designs which are as close to full as permitted by the parameters), but we highlight Theorem 5.4.2 as Kim [81] claims this result is obtained purely via the Rödl Nibble, and further, that he "believes that the Nibble method gives no better bound up to a logarithmic factor". We believe that philosophically, Theorem 5.2.1 should be applicable to prove Theorem 5.4.2 (though not quite obtaining logarithmic error), and we believe that the idea of Kim's proof was likely some form of 'completely collapsing the entire codegree sequence'.

We also believe that, provided $p$ is large enough for Nibble ideas to be used at all, one should be able to apply Theorem 5.2.1 to the context of clique-packing problems in $G(n, p)$, or similar random hypergraph spaces, since the obtained random graph or hypergraph is with high probability very 'smooth'. This problem is mentioned by Yuster [126], and is likely very difficult, due to the connection to the existence of designs. Further, we remark that the methods used in [76, 77, 53] apply to this setting, but only with constant, or close to constant, values of $p$. An advantage of our results is that they would apply to settings in which $p$ is allowed to be much smaller than constant (which is an inherent advantage of the Polynomial Method, which is key to our proof of Theorem 5.2.1, and we discuss further in the next section).

Finally, we believe that our methodology for proving Theorem 5.4.1 may extend in quite a natural way to the study of approximate 'Hamilton 2-plexes' in Latin squares. Here, we remark that an $n \times n$ Latin square corresponds to a proper edge-colouring of $K_{n,n}$ with $n$ colours, and in this setting a Hamilton 2-plex is precisely a Hamilton cycle of $K_{n,n}$ using each colour precisely twice. Halasz [63] conjectured that for $n \geq 5$, every $n \times n$ Latin square has a Hamilton 2-plex, and proved this conjecture approximately, by showing that each properly $n$-edge-coloured $K_{n,n}$ for $n \geq 5$ admits a path forest on $2n - o(n)$ edges, containing each colour at most twice. We believe that one could use Theorem 5.2.1 in conjunction with some random partitioning arguments, as in the proof of Theorem 5.4.1, to improve the $o(n)$ error bound of Halasz [63] to $n^{1/2+\delta}$ for any constant $\delta > 0$, and further, we anticipate that one could use absorption arguments to 'close' such a path forest into a cycle, still using each colour at most twice.

## 5.5 Sketch of the proof of Theorem 5.2.1

In this section, we sketch the most important parts of the proof of Theorem 5.2.1. Put simply, the idea (at least originally) proceeds as follows: Use Theorem 5.1.4 to 'collapse the penultimate codegree $D_{s-1}$ to the final codegree $D_s$', by which we mean iterate the Rödl Nibble whilst tracking the degradation of $D_{s-1}$ until it is not usefully larger than $D_s$, to create a large matching. As one does this, one also tracks the degradation of $D, D_2, \ldots, D_{s-2}$, and observes that the ratios $D_j/D_{j+1}$ all reduce by a factor of $x$, for $j \neq s-1$ (with $x$ as in Theorem 5.1.4). Then, one simply 'forgets' about $D_s$, taking the degraded $D_{s-1}$ to be the new final entry in the codegree sequence, and applies Theorem 5.1.4 again to 'collapse $D_{s-2}$ to $D_{s-1}$', further reducing the size of the leftover. Keep repeating this process until one runs out of codegrees (we call this using all the space in the codegree sequence), or one hits some other bottleneck, for example that $\varepsilon$ was not tight enough to begin with to allow us to iterate this long.

Thus, one needs a form of Theorem 5.1.4 which better lends itself to iteration. The key obstacle here is that the proof of Theorem 5.1.4 only works for $\varepsilon \approx 1/x$; one seemingly cannot use a smaller $\varepsilon$ and expect this $\varepsilon$ to have not degraded too much by the end of the process. Since we intend to iterate a rough equivalent of Theorem 5.1.4 many times, and since $\varepsilon$ grows by a factor of $x$ with each iteration, we will usually want to allow $\varepsilon$ to be much, much smaller than $1/x$. To be a little more precise about where this causes a problem, we note that in the proof of Theorem 5.1.4, Vu [122] performs Nibbles which each shrink the leftover by a factor of $\theta \approx 1/x$. In order to control the error in vertex degrees, one throws out a waste set $W$ of vertices at each Nibble, where $|W|$ is on the order $\varepsilon\theta n$, and cannot be larger, otherwise the accumulated waste sets eventually grow to a point at which they are larger than the leftover hypergraph. Further, during the Nibble, one must ignore those chosen edges which intersect in order to obtain a matching, and one expects roughly $\theta^2 n$ vertices to be involved in such intersections. In the proof of Theorem 5.1.4, one has $\varepsilon\theta n \approx \theta^2 n$, and so all such intersection vertices can be thrown into the waste set. Since we need to permit $\varepsilon\theta n \ll \theta^2 n$ during the proof of Theorem 5.2.1, it is important that we put such vertices back into the leftover hypergraph in order to keep the waste sets small, so one key difference is that we need to consider a 'non-wasteful' Nibble, in this sense. We remark that Alon, Kim, and Spencer [6] use such a 'non-wasteful' Nibble procedure, but their concentration arguments (which revolve around martingale inequalities) do not seem to extend to the setting of hypergraphs which are not linear.

The most challenging obstacle is the need to concentrate the degrees of vertices after a Nibble, within very tight error, in this 'non-wasteful' regime, where the hypergraph need not be linear. We note that to ensure that $\varepsilon$ does not degrade too much during a Nibble, one must ensure that the vertex degrees are concentrated within an error of $\varepsilon\theta D$. This is the same error expression that Vu [122] permits in his concentration arguments, but since we must permit $\varepsilon$ to be considerably smaller than $\theta$ (as opposed to $\varepsilon \approx \theta$ in [122]), we are usually effectively asking for much smaller error in the concentration of the vertex degrees. To achieve this, we employ a very involved application of the Polynomial Method. Before proceeding further, we briefly discuss the Polynomial Method now.

### 5.5.1 The Polynomial Method

The central idea of the Polynomial Method is that one can often bound a complicated random variable (in our application, the vertex degrees after a Nibble) from above and below in terms of two polynomials of simpler, mutually independent random variables (in our application, the indicator random variables for the events of given edges being chosen for the matching, or given vertices being chosen for the waste set). One can then use concentration inequalities to tightly concentrate these bounding polynomials, effectively bounding the original random variable of interest. Concentration inequalities for polynomials of random variables were first produced by Kim and Vu [83]. They noticed that such inequalities can be powerful in the sense that they allow one to avoid using a 'worst-case Lipschitz constant' (for instance all the values $c_i$ in McDiarmid's Inequality (Theorem 3.5.2)), instead only taking into account the *average* adverse effects that the changes in individual random variables can have on the polynomial. This is especially useful in analysis of sparse graphs, or in 'small' processes like the Nibble, because in such cases, the Lipschitz constants usually are much larger than the expectations of most random variables of interest, prohibiting the usage of concentration inequalities like Theorem 3.5.2.

We now state a polynomial concentration inequality of Vu [121], which is the key inequality we use in our concentration arguments. We say that a polynomial $\mathbf{Y} = Y(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ of the mutually independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_m$ is a *positive polynomial* if all of the coefficients of $\mathbf{Y}$ are non-negative. Given an index set $A = \{i_1, \ldots, i_j\} \in [m]^j$, we define $\mathbf{Y}_A \coloneqq \partial^j \mathbf{Y} / \partial \mathbf{X}_{i_1} \ldots \partial \mathbf{X}_{i_j}$. If $\mathbf{Y}$ has degree $\ell$ then for $j \in [\ell]_0$ we define $\mathbb{E}_j(\mathbf{Y}) \coloneqq \max_{A \in \binom{[m]}{j}} \mathbb{E}[\mathbf{Y}_A]$. Notice that $\mathbb{E}_0(\mathbf{Y}) = \mathbb{E}[\mathbf{Y}]$.

**Theorem 5.5.1** (Vu [121]). *There is $f \colon \mathbb{N} \to \mathbb{N}$ such that the following holds. Suppose that $\mathbf{Y}$ is a degree-$\ell$ positive polynomial of the mutually independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_m$. Suppose further that there are $\lambda, \mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_\ell > 0$ satisfying $\mathbb{E}_j(\mathbf{Y}) \leq \mathcal{E}_j$ for $j \in [\ell]_0$ and $\mathcal{E}_j / \mathcal{E}_{j+1} > \lambda + (j+1) \log m$ for $j \in [\ell - 1]_0$. Then*

$$\mathbb{P}\left[ |\mathbf{Y} - \mathbb{E}[\mathbf{Y}]| > f(\ell) \sqrt{\lambda \mathcal{E}_0 \mathcal{E}_1} \right] \leq 3^\ell \exp(-\lambda).$$

We remark that Theorem 5.5.1 is also the main concentration inequality used by Vu in [122], though he only uses it for polynomials of degree at most 2, whereas we require the full generality.

### 5.5.2 Key lemmas

Returning to sketching the proof of Theorem 5.2.1, we note that the polynomials that we use to bound the vertex degrees in the course of a Nibble, are such that the sum of all degree-$i$ terms has expectation on the order $\theta^i D$. In particular, since $\varepsilon \approx \theta$ in the proof of Theorem 5.1.4 (so that $\theta^2 D \approx \varepsilon\theta D$), Vu [122] uses polynomials of degree 2, and applies Theorem 5.5.1, obtaining an acceptable amount of growth in the value of $\varepsilon$ during a Nibble. In our paper, we must use bounding polynomials of degree at least $R$, where $\theta^R \approx \varepsilon\theta$. In fact, we use polynomials of degree approximately $R^6$. Even arguing the correct form of, let alone performing the required calculus and expectation calculations on these polynomials is a considerable challenge. To give a glimpse into the obtaining of these polynomials, we note that our 'Nibble' is obtained from the outcomes of the mutually independent indicator random variables $\{\mathbf{a}_u\}_{u \in V(H)} \cup \{\mathbf{t}_f\}_{f \in E(H)}$, where $\mathbf{a}_u$ indicates if the vertex $u$ is chosen for the 'waste set' $\mathbf{W}$, and $\mathbf{t}_f$ indicates if the edge $f$ is chosen for a set called $\mathbf{X}$. We then let the random matching $\mathbf{M}$ be the set of those $\mathbf{X}$-edges which do not intersect any other $\mathbf{X}$-edges. In this regime, the degree of a vertex $v$ after the Nibble may be written as $\mathbf{deg}(v) = \sum_{e \ni v} \mathbf{y}_e$, where $\mathbf{y}_e$ indicates that $e$ does not lose any non-$v$ vertices to $\mathbf{W}$ or $\mathbf{M}$. More specifically, where $\Lambda_e := \{f \in E(H) \colon f \cap (e \setminus \{v\}) \neq \emptyset\}$,

$$\mathbf{y}_e = \left( \prod_{u \in e \setminus \{v\}} (1 - \mathbf{a}_u) \right) \cdot \left( \prod_{f \in \Lambda_e} (1 - \mathbf{b}_f) \right), \tag{5.5.1}$$

where $\mathbf{b}_f$ indicates if $f \in \mathbf{M}$; that is,

$$\mathbf{b}_f = \mathbf{t}_f \prod_{\substack{h \in E(H) \setminus \{f\}: \\ h \cap f \neq \emptyset}} (1 - \mathbf{t}_h).$$

We cannot apply Theorem 5.5.1 to this full polynomial representation of the random variable $\mathbf{deg}(v)$ directly (even after we decompose it into positive polynomials) because the degree (as a polynomial) of $\mathbf{deg}(v)$ is too large to obtain tight concentration. To deal with this issue, observe that one may expand the right hand-side of (5.5.1) as

$$1 - \sum_{u \in e \setminus \{v\}} \mathbf{a}_u - \sum_{f \in \Lambda_e} \mathbf{b}_f + \sum_{\substack{u \in e \setminus \{v\} \\ f \in \Lambda_e}} \mathbf{a}_u \mathbf{b}_f + \sum_{U \in \binom{e \setminus \{v\}}{2}} \prod_{u \in U} \mathbf{a}_u + \sum_{F \in \binom{\Lambda_e}{2}} \prod_{f \in F} \mathbf{b}_f - \dots,$$

continuing to 'longer' products of the indicator random variables, with the sign alternating depending on the number of random variables in the terms. One can check that if we truncate the above expansion after writing all terms of at most a given 'length' $i$ (i.e. $i$ is the number of random variables within the term), then the resulting polynomial is deterministically an upper bound for $\mathbf{y}_e$ if $i$ is even, and a lower bound if $i$ is odd. It would seem to suffice, then, to bound $\mathbf{y}_e$ from above and below in terms of these expansions, truncated to lengths $R$ and $R+1$ respectively (if $R$ was even, say) where $\theta^R = \varepsilon\theta$, since then in particular the expected difference between the two approximations for $\mathbf{deg}(v)$ is on the order $\varepsilon\theta D$. However, we still cannot apply Theorem 5.5.1 to these polynomials, since the random variables $\{\mathbf{b}_f\}_{f \in E(H)}$ are not mutually independent. We must therefore also perform a similar expansion to each of the random variables $\mathbf{b}_f$, and substitute appropriate truncations of each $\mathbf{b}_f$ into our existing truncated polynomials for $\mathbf{y}_e$, ensuring that we preserve the desired global upper or lower bound. This, together with other confounding factors, means that the final polynomials we consider involving only the mutually independent random variables $\{\mathbf{a}_u\}_{u\in V(H)} \cup \{\mathbf{t}_f\}_{f\in E(H)}$ have degree closer to $R^6$. We then apply Theorem 5.5.1 to each of (the positive polynomials within) the obtained bounding polynomials for $\mathbf{deg}(v)$. We omit any further detail here.

We now state the 'Nibble Lemma', which performs the main probabilistic analysis in our paper; in particular the Nibble Lemma incorporates the checking of the concentration of the vertex degrees via the aforementioned Polynomial Method arguments, as well as the concentration of the various codegrees (for which we also use the Polynomial Method, though this application is simpler), and checking that the leftover set and waste set have the right size.

**Lemma 5.5.2** ('Nibble Lemma', Gould and Kelly, 2023+). *Suppose $1/d \ll 1/R, 1/k \leq 1$, with $R \geq 3$ odd. Let $H$ be an $(m, d, \varepsilon)$-regular, $(k+1)$-uniform hypergraph, and suppose there is $2 \leq s \leq k+1$ and quantities $b_z$ for $2 \leq z \leq s$ such that $C_z(H) \leq b_z$ for $2 \leq z \leq s$ and $b_s \geq 1$. Suppose further that there is $0 < \theta < 1$ such that the following conditions are satisfied:*

*(N1)* $\theta^{R-1} \leq \varepsilon$;

*(N2)* $\varepsilon \leq \theta$;

*(N3)* $\varepsilon^2\theta\frac{d}{b_2} \geq \log^2 d$;

*(N4)* $\theta^2\frac{b_z}{b_{z+1}} \geq \log^2 d$ *for all* $2 \leq z \leq s-1$;

*(N5)* $\theta \leq 1/\log^{5R^6} d$.

*Then there is a matching $M$ of $H$ and a set $W \subseteq V(H)$ of size $|W| \leq 10\varepsilon\theta m$, together with numbers $m'$ and $d'$ satisfying $m(1-\theta-2\theta^{3/2}) \leq m' \leq m(1-\theta+2\theta^{3/2})$*

*and* $d(1-k\theta-\theta^{3/2}) \leq d' \leq d(1-k\theta+\theta^{3/2})$*, such that the hypergraph* $H'$ *induced by* $V(H) \setminus (V(M) \cup W)$ *is* $(m', d', \varepsilon')$*-regular, where* $\varepsilon' := \varepsilon(1+\theta)$*. Further,* $C_z(H') \leq b_z(1-(k-z+1)\theta+\theta^{3/2})$ *for all* $2 \leq z \leq s-1$*.*

In particular, hypothesis (N1) and the definition of $R$ effectively ensure that we can obtain acceptable growth in $\varepsilon$ whilst still using only 'constant' degree bounding polynomials, which we require in order to keep the concentration sufficiently tight when applying Theorem 5.5.1.

One then iterates Lemma 5.5.2 similarly to how Vu [122] iterates his version of the Nibble Lemma, in order to obtain the following result, which we call the 'Chomp Lemma'.

**Lemma 5.5.3** ('Chomp Lemma', Gould and Kelly, 2023+)**.** *Suppose* $1/D \ll 1/K, 1/k, \gamma \leq 1$*. Let* $H$ *be an* $(n, D, \varepsilon)$*-regular,* $(k+1)$*-uniform hypergraph, and suppose there is* $2 \leq s \leq k+1$ *and quantities* $D_j$ *for* $2 \leq j \leq s$ *such that* $C_j(H) \leq D_j$ *for* $2 \leq j \leq s$ *and* $D_s \geq 1$*. Suppose further that there is* $x > 0$ *such that following conditions hold:*

*(C1)* $\frac{1}{x^K} \leq \varepsilon$;

*(C2)* $\varepsilon \leq \frac{1}{ex^{1+\gamma}}$;

*(C3)* $\frac{\varepsilon^2 D}{x^{1+\gamma} D_2} \geq e^{k+12} \log^2 D$;

*(C4)* $\frac{D_j}{x^{2+2\gamma} D_{j+1}} \geq e^{k+7} \log^2 D$ *for all* $2 \leq j \leq s-1$;

*(C5)* $\frac{D_{s-1}}{x^{k-s+2+2\gamma} D_s} \geq e^{k+2} \log^2 D$;

*(C6)* $x \geq \log^{\frac{5}{\gamma}\left(\frac{K}{\gamma}+3\right)^6} D$.

*Then there is a matching* $M$ *of* $H$ *and a set* $W \subseteq V(H)$ *of size* $|W| \leq 10e^7 \varepsilon n \log x$*, together with numbers* $n'$ *and* $D'$ *satisfying* $n/e^6 x \leq n' \leq e^7 n/x$ *and* $D/e^6 x^k \leq D' \leq e^{k+6} D/x^k$*, such that the hypergraph* $H'$ *induced by* $V(H) \setminus (V(M) \cup W)$

*is* $(n', D', \varepsilon')$*-regular, where* $\varepsilon' := \varepsilon x$*. Further,* $C_j(H') \leq e^{k+6} D_j / x^{k-j+1}$ *for all* $2 \leq j \leq s-1$*.*

Lemma 5.5.3 is similar to Theorem 5.1.4 in the appearance of conditions (C4) and (C5), but conditions (C1) and (C3) are key differences in the sense that (C1) permits $\varepsilon$ to be much smaller than $1/x$, and (C3) effectively contains the information that allows us to conclude that $\varepsilon$ only grows by the amount we expect (a factor of $x$) after each 'Chomp'. Note that (if $D = \omega(1)$) a 'Nibble' reduces the leftover vertex set only by a factor of $o(1)$, whereas a Chomp reduces the leftover vertex set by a factor of $\omega(1)$. In order to minimise error, we then iterate the Chomp Lemma, each time choosing $x \approx (D_{s-1}/D)^{\gamma}$ (instead of the more immediately natural choice of setting $x$ to meet one of the bottlenecks provided by (C1)–(C6)), until after $1/\gamma$ Chomps, one has 'collapsed' $D_{s-1}$ to $D_s$. Then one forgets about $D_s$, and begins the process of gradually collapsing $D_{s-2}$ to $D_{s-1}$ via small Chomps, and so on, until one runs out of codegrees, or hits one of the other two bottlenecks for $x$ given in Theorem 5.2.1 (in particular hypothesis (i) of Theorem 5.2.1 ensures that the iteration of Lemma 5.5.3 does not stop due to a 'middle' ratio $D_j/D_{j+1}$ degrading too much).

### 5.5.3 Improving the error, revisited

We return to the discussion in Section 5.3.3 concerning the possibility of replacing the polynomial error in Theorem 5.2.1 with a polylogarithmic one. The main problem is the requirement that $R$ must be constant in (N1) in order to have sufficiently tight concentration in Theorem 5.5.1. If one could instead bound the random variable corresponding to a vertex's degree, by polynomials of logarithmic degree, and still obtain tight concentration, then one would not need to suppose (C1), instead allowing for smaller Chomps (say polylogarithmic values of $x$, even while $\varepsilon = n^{-\Omega(1)}$), which should ultimately incur only the desired logarithmic error by the end of the process.

One needs a replacement for Theorem 5.5.1, then, which allows random polynomials of degree $\omega(1)$ without losing tight concentration. Such concentration inequalities do exist, for instance those of Schudy and Sviridenko [116, 117], but each seem to cause other problems; for example the cited results do not also allow for the terms to have as many repeated instances of a single random variable as they unavoidably do in our application, without losing concentration. Given an ideal concentration inequality though, we believe one could run our proof of Theorem 5.2.1 and obtain only polylogarithmic error.

# BIBLIOGRAPHY


[1] E. Aigner-Horev and G. Levy. Tight Hamilton cycles in cherry-quasirandom 3-uniform hypergraphs. *Combin. Probab. Comput.* 30.3 (2021), pp. 412–443.

[2] M. Ajtai, J. Komlós, J. Pintz, J. Spencer and E. Szemerédi. Extremal uncrowded hypergraphs. *J. Combin. Theory Ser. A* 32 (1982), pp. 321–335.

[3] M. Ajtai, J. Komlós and E. Szemerédi. A dense infinite Sidon sequence. *European J. Combin.* 2 (1981), pp. 1–11.

[4] S. Akbari, O. Etesami, H. Mahini and M. Mahmoody. On rainbow cycles in edge coloured complete graphs. *Australas. J. Combin.* 37 (2007), pp. 33–42.

[5] N. Alon and S. Friedland. The maximum number of perfect matchings in graphs with a given degree sequence. *Electron. J. Combin.* 15.1 (2008), N13.

[6] N. Alon, J.-H. Kim and J. Spencer. Nearly perfect matchings in regular simple hypergraphs. *Israel J. Math.* 100.1 (1997), pp. 171–187.

[7] N. Alon, A. Pokrovskiy and B. Sudakov. Random subgraphs of properly edge-coloured complete graphs and long rainbow cycles. *Israel J. Math.* 222.1 (2017), pp. 317–331.

[8] N. Alon and R. Yuster. On a hypergraph matching problem. *Graphs Combin.* 21.4 (2005), pp. 377–384.

[9] L. D. Andersen. Hamilton circuits with many colours in properly edge-coloured complete graphs. *Math. Scand.* 64 (1989), pp. 5–14.

[10] J. Balogh and T. Molla. Long rainbow cycles and Hamiltonian cycles using many colors in properly edge-colored complete graphs. *European J. Combin.* 79 (2019), pp. 140–151.

[11] F. Benzing, A. Pokrovskiy and B. Sudakov. Long directed rainbow cycles and rainbow spanning trees. *European J. Combin.* 88 (2020), p. 103102.

[12] J.-C. Bermond, A. Germa, M.-C. Heydemann and D. Sotteau. Hypergraphes Hamiltoniens. *Prob. Comb. Theorie Graph Orsay* 260.9 (1976), pp. 39–43.

[13] T. Bohman and L. Warnke. Large girth approximate Steiner triple systems. *J. Lond. Math. Soc.* 100 (2019), pp. 895–913.

[14] B. Bollobás. 'Extremal graph theory'. *Handbook of Combinatorics*. Elsevier Science, 1995, pp. 1231–1292.

[15] J. A. Bondy. 'Basic graph theory: Paths and circuits'. *Handbook of Combinatorics*. Elsevier Science, 1995, pp. 3–110.

[16] J. A. Bondy and U. S. R. Murty. *Graph Theory*. Springer, 2008.

[17] L. M. Brègman. Certain properties of nonnegative matrices and their permanents. *Dokl. Akad. Nauk SSSR* 211 (1973), pp. 27–30.

[18] A. E. Brouwer, A. J. de Vries and R. M. A. Wieringa. A lower bound for the length of partial transversals in a Latin square. *Nieuw Arch. Wiskd.* 26.2 (1978), pp. 330–332.

[19] R. A. Brualdi and S. Hollingsworth. Multicolored trees in complete graphs. *J. Combin. Theory Ser. B* 68.2 (1996), pp. 310–313.

[20] R. A. Brualdi and H. J. Ryser. *Combinatorial matrix theory.* Cambridge University Press, 1991.

[21] N. J. Cavenagh, C. Greenhill and I. M. Wanless. The cycle structure of two rows in a random Latin square. *Random Structures Algorithms* 33.3 (2008), pp. 286–309.

[22] H. Chen and X. Li. 'Long rainbow path in properly edge-coloured complete graphs'. arXiv preprint arXiv:1503.04516. 2015.

[23] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics* 3.1 (2006), pp. 79–127.

[24] V. Chvátal. On Hamilton's ideals. *J. Combin. Theory Ser. B* 12.2 (1972), pp. 163–168.

[25] V. Chvátal and P. Erdős. A note on Hamiltonian circuits. *Discrete Math.* 2.2 (1972), pp. 111–113.

[26] G. M. Constantine. Edge-disjoint isomorphic multicolored trees and cycles in complete graphs. *SIAM J. Discrete Math.* 18.4 (2004), pp. 577–580.

[27] G. M. Constantine. Multicolored parallelisms of isomorphic spanning trees. *Discrete Math. Theor. Comput. Sci.* 5.1 (2002), pp. 121–125.

[28] S. A. Cook. 'The complexity of theorem-proving procedures'. *Proceedings of the third annual ACM symposium on Theory of computing.* 1971, pp. 151–158.

[29] S. B. Cooper. *Computability Theory.* Taylor & Francis Inc, 2003.

[30] M. Coulson, P. Keevash, G. Perarnau and L. Yepremyan. Rainbow factors in hypergraphs. *J. Combin. Theory Ser. A* 172 (2020), p. 105184.

[31] B. Cuckler and J. Kahn. Entropy bounds for perfect matchings and Hamiltonian cycles. *Combinatorica* 29.3 (2009), pp. 327–335.

[32] B. Cuckler and J. Kahn. Hamiltonian cycles in Dirac graphs. *Combinatorica* 29.3 (2009), pp. 299–326.

[33] A. Czygrinow and T. Molla. Tight codegree condition for the existence of loose Hamilton cycles in 3-graphs. *SIAM J. Discrete Math.* 28.1 (2014), pp. 67–76.

[34] E. Dahlhaus, P. Hajnal and M. Karpinski. On the parallel complexity of Hamiltonian cycle and matching problem on dense graphs. *Journal of Algorithms* 15.3 (1993), pp. 367–384.

[35] M. Delcourt and L. Postle. 'Finding an almost perfect matching in a hypergraph avoiding forbidden submatchings'. arXiv preprint arXiv:2204.08981. 2022.

[36] G. A. Dirac. Some theorems on abstract graphs. *Proc. Lond. Math. Soc.* 3.1 (1952), pp. 69–81.

[37] R. Durrett. *Probability: theory and examples.* Vol. 49. Camb. Ser. Stat. Probab. Math. Cambridge University Press, 2019.

[38] G. P. Egorychev. The solution of van der Waerden's problem for permanents. *Adv. Math.* 42.3 (1981), pp. 299–305.

[39] S. Ehard, S. Glock and F. Joos. Pseudorandom hypergraph matchings. *Combin. Probab. Comput.* 29.6 (2020), pp. 1–18.

[40] P. Erdős. 'Problems and results in combinatorial analysis'. *Colloq. Internat. Theor. Combin. Rome.* 1973, pp. 3–17.

[41] P. Erdős and H. Hanani. On a limit theorem in combinatorial analysis. *Publ. Math. Debrecen* 10 (1963), pp. 10–13.

[42] D. I. Falikman. Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix. *Mat. Zametki* 29.6 (1981), pp. 931–938, 957.

[43] A. Ferber, V. Jain and B. Sudakov. Number of 1-factorizations of regular high-degree graphs. *Combinatorica* 40 (2020), pp. 315–344.

[44] A. Ferber and M. Kwan. Almost all Steiner triple systems are almost resolvable. *Forum Math. Sigma* 8 (2020), E39.

[45] S. Ferber, M. Krivelevich and B. Sudakov. Counting and packing Hamilton $\ell$-cycles in dense hypergraphs. *Journal of Combinatorics* 7.1 (2016), pp. 135–157.

[46] P. Frankl and V. Rödl. Near perfect coverings in graphs and hypergraphs. *European J. Combin.* 6 (1985), pp. 317–326.

[47] Z. Füredi. Matchings and covers in hypergraphs. *Graphs Combin.* 4 (1988), pp. 115–206.

[48] H. Gebauer and F. Mousset. 'On rainbow cycles and paths'. arXiv preprint arXiv:1207.0840. 2012.

[49] R. Glebov and Z. Luria. On the maximum number of Latin transversals. *J. Combin. Theory Ser. A* 141 (2016), pp. 136–146.

[50] S. Glock, S. Gould, F. Joos, D. Kühn and D. Osthus. Counting Hamilton cycles in Dirac hypergraphs. *Combin. Probab. Comput.* 30.4 (2021), pp. 631–653.

[51] S. Glock, F. Joos, J. Kim, M. Kühn and L. Lichev. 'Conflict-free hypergraph matchings'. arXiv preprint arXiv:2205.05564. 2022.

[52] S. Glock, D. Kühn, A. Lo and D. Osthus. On a conjecture of Erdős on locally sparse Steiner triple systems. *Combinatorica* 40.3 (2020), pp. 363–403.

[53] S. Glock, D. Kühn, A. Lo and D. Osthus. The existence of designs via iterative absorption: hypergraph $F$-designs for arbitrary $F$. *Mem. Amer. Math. Soc.* (to appear).

[54] S. Glock, D. Kühn, R. Montgomery and D. Osthus. Decompositions into isomorphic rainbow spanning trees. *J. Combin. Theory Ser. B* 146 (2021), pp. 439–484.

[55] S. Gould and T. Kelly. Hamilton transversals in random Latin squares. *Random Structures Algorithms* (to appear).

[56] S. Gould, T. Kelly, D. Kühn and D. Osthus. Almost all optimally coloured complete graphs contain a rainbow Hamilton path. *J. Combin. Theory Ser. B* 156 (2022), pp. 57–100.

[57] D. A. Grable. More-than-nearly-perfect packings and partial designs. *Combinatorica* 19 (1999), pp. 221–239.

[58] A. Gyárfás and M. Mhalla. Rainbow and orthogonal paths in factorizations of $K_n$. *J. Combin. Des.* 18.3 (2010), pp. 167–176.

[59] A. Gyárfás, M. Ruszinkó, G. Sárközy and R. Schelp. Long rainbow cycles in proper edge-colourings of complete graphs. *Australas. J. Combin.* 50 (2011), pp. 45–53.

[60] A. Gyárfás and G. N. Sárközy. Rainbow matchings and cycle-free partial transversals of Latin squares. *Discrete Math.* 327 (2014), pp. 96–102.

[61] G. Hahn. Un jeu de colouration. *Actes du Colloque de Cerisy* 12 (1980), p. 18.

[62] G. Hahn and C. Thomassen. Path and cycle sub-Ramsey numbers and an edge-colouring conjecture. *Discrete Math.* 62.1 (1986), pp. 29–33.

[63] K. Halasz. 'Among graphs, groups, and latin squares'. PhD thesis, Simon Fraser University. 2021.

[64] J. Han and Y. Zhao. Minimum codegree threshold for Hamilton $\ell$-cycles in $k$-uniform hypergraphs. *J. Combin. Theory Ser. A* 132 (2015), pp. 194–223.

[65] H. Hàn and M. Schacht. Dirac-type results for loose Hamilton cycles in uniform hypergraphs. *J. Combin. Theory Ser. B* 100.3 (2010), pp. 332–346.

[66] P. Hatami and P. W. Shor. A lower bound for the length of a partial transversal in a Latin square. *J. Combin. Theory Ser. A* 115.7 (2008), pp. 1103–1113.

[67] S. Janson, T. Łuczak and A. Ruciński. *Random graphs.* Wiley-Intersci. Ser. Discrete Math. Optim. New York: Wiley-Interscience, 2000, pp. xii+333.

[68] A. Johansson. 'Asymptotic choice number for triangle free graphs'. Tech. report, DIMACS technical report. 1996.

[69] J. Kahn. Asymptotically good list-colorings. *J. Combin. Theory Ser. A* 73 (1996), pp. 1–59.

[70] J. Kahn. Coloring nearly-disjoint hypergraphs with $n + o(n)$ colors. *J. Combin. Theory Ser. A* 59 (1992), pp. 31–39.

[71] A. Kaneko, M. Kano and K. Suzuki. 'Three edge-disjoint Multicolored Spanning Trees in Complete Graphs'. preprint. 2002.

[72] D. Kang, T. Kelly, D. Kühn, A. Methuku and D. Osthus. Graph and hypergraph colouring via nibble methods: a survey. *Proceedings of the 8th European Congress of Mathematics* (to appear).

[73] D. Kang, D. Kühn, A. Methuku and D. Osthus. 'New bounds on the size of nearly perfect matchings in almost regular hypergraphs'. arXiv preprint arXiv:2010.04183. 2020.

[74] G. Y. Katona and H. A. Kierstead. Hamiltonian chains in hypergraphs. *J. Graph Theory* 30.3 (1999), pp. 205–212.

[75] P. Keevash. Counting designs. *J. Eur. Math. Soc. (JEMS)* 20.4 (2018), pp. 903–927.

[76] P. Keevash. 'The existence of designs'. arXiv preprint arXiv:1401.3665. 2014.

[77] P. Keevash. 'The existence of designs II'. arXiv preprint arXiv:1802.05900. 2018.

[78] P. Keevash, D. Kühn, R. Mycroft and D. Osthus. Loose Hamilton cycles in hypergraphs. *Discrete Math.* 311.7 (2011), pp. 544–559.

[79] P. Keevash, A. Pokrovskiy, B. Sudakov and L. Yepremyan. New bounds for Ryser's conjecture and related problems. *Trans. Amer. Math. Soc. Ser. B* 9.08 (2022), pp. 288–321.

[80] J. Kim, D. Kühn, A. Kupavskii and D. Osthus. Rainbow structures in locally bounded colourings of graphs. *Random Structures Algorithms* 56.4 (2020), pp. 1171–1204.

[81] J. H. Kim. Nearly optimal partial Steiner systems. *Electronic notes in Discrete Mathematics* 7 (2001), pp. 74–77.

[82] J. H. Kim. On Brooks' theorem for sparse graphs. *Combin. Probab. Comput.* 4 (1995), pp. 97–132.

[83] J. H. Kim and V. H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica* 20.3 (2000), pp. 417–434.

[84] T. P. Kirkman. On the representation of polyhedra. *Philos. Trans. Roy. Soc. London Ser. A* 146 (1856), pp. 413–418.

[85] A. Kostochka and V. Rödl. Partial Steiner systems and matchings in hypergraphs. *Random Structures Algorithms* 13.3–4 (1997), pp. 335–347.

[86] D. Kühn, R. Mycroft and D. Osthus. Hamilton $\ell$-cycles in uniform hypergraphs. *J. Combin. Theory Ser. A* 117.7 (2010), pp. 910–927.

[87] D. Kühn and D. Osthus. 'Hamilton cycles in graphs and hypergraphs: an extremal perspective'. *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV.* Kyung Moon Sa, Seoul, 2014, pp. 381–406.

[88] D. Kühn and D. Osthus. Loose Hamilton cycles in 3-uniform hypergraphs of high minimum degree. *J. Combin. Theory Ser. B* 96.6 (2006), pp. 767–821.

[89] D. Kühn and D. Osthus. Matchings in hypergraphs of large minimum degree. *J. Graph Theory* 51.4 (2006), pp. 269–280.

[90] M. Kwan. Almost all Steiner triple systems have perfect matchings. *Proc. Lond. Math. Soc.* 121.6 (2020), pp. 1468–1495.

[91] M. Kwan, A. Sah and M. Sawhney. Large deviations in random Latin squares. *Bull. Lond. Math. Soc.* 54.4 (2022), pp. 1420–1438.

[92] M. Kwan, A. Sah, M. Sawhney and M. Simkin. 'Substructures in Latin squares'. arXiv preprint arXiv:2202.05088. 2022.

[93] M. Kwan and B. Sudakov. Intercalates and discrepancy in random Latin squares. *Random Structures Algorithms* 52.2 (2018), pp. 181–196.

[94] D. A. Levin and Y. Peres. *Markov chains and mixing times.* American Mathematical Soc., 2017.

[95] N. Linial and Z. Luria. An upper bound on the number of Steiner triple systems. *Random Structures Algorithms* 43.4 (2013), pp. 399–406.

[96] A. Maamoun and H. Meyniel. On a problem of G. Hahn about coloured Hamiltonian paths in $K_{2^t}$. *Discrete Math.* 51.2 (1984), pp. 213–214.

[97] C. McDiarmid. 'On the method of bounded differences'. *Surveys in Combinatorics, 1989: Invited Papers at the Twelfth British Combinatorial Conference.* Vol. 141. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, 1989, pp. 148–188.

[98] B. D. McKay and I. M. Wanless. Most Latin squares have many subsquares. *J. Combin. Theory Ser. A* 86.2 (1999), pp. 323–347.

[99] M. Molloy. The list chromatic number of graphs with small clique number. *J. Combin. Theory Ser B.* 134 (2019), pp. 264–284.

[100] M. Molloy and B. Reed. Near-optimal list colorings. *Random Structures Algorithms* 17 (2000), pp. 376–402.

[101] R. Montgomery. Spanning trees in random graphs. *Adv. Math.* 356 (2019), 106793, pp.1–92.

[102] R. Montgomery, A. Pokrovskiy and B. Sudakov. Decompositions into spanning rainbow structures. *Proc. Lond. Math. Soc.* 119.4 (2019), pp. 899–959.

[103] Ø. Ore. Note on Hamilton circuits. *Am. Mat. Monthly* 67 (1960), p. 55.

[104] N. Pippenger and J. Spencer. Asymptotic behaviour of the chromatic index for hypergraphs. *J. Combin. Theory Ser. A* 51.1 (1989), pp. 24–42.

[105] A. Pokrovskiy and B. Sudakov. A counterexample to Stein's equi-$n$-square conjecture. *Proc. Amer. Math. Soc.* 147.6 (2019), pp. 2281–2287.

[106] A. Pokrovskiy and B. Sudakov. Linearly many rainbow trees in properly edge-coloured complete graphs. *J. Combin. Theory Ser. B* 132 (2018), pp. 134–156.

[107] C. Reiher, V. Rödl, A. Ruciński, M. Schacht and E. Szemerédi. Minimum vertex degree condition for tight Hamiltonian cycles in 3-uniform hypergraphs. *Proc. Lond. Math. Soc.* 119.2 (2019), pp. 409–439.

[108] V. Rödl. On a packing and covering problem. *European J. Combin.* 6.1 (1985), pp. 69–78.

[109] V. Rödl and A. Ruciński. 'Dirac-type questions for hypergraphs: a survey (or more problems for Endre to solve)'. *An irregular mind.* Springer, 2010, pp. 561–590.

[110] V. Rödl, A. Ruciński and E. Szemerédi. A Dirac-type theorem for 3-uniform hypergraphs. *Combin. Probab. Comput.* 15.1-2 (2006), pp. 229–251.

[111] V. Rödl, A. Ruciński and E. Szemerédi. An approximate Dirac-type theorem for $k$-uniform hypergraphs. *Combinatorica* 28.2 (2008), pp. 229–260.

[112] V. Rödl, A. Ruciński and E. Szemerédi. Dirac-type conditions for Hamiltonian paths and cycles in 3-uniform hypergraphs. *Adv. Math.* 227.3 (2011), pp. 1225–1299.

[113] V. Rödl, A. Ruciński and E. Szemerédi. Perfect matchings in large uniform hypergraphs with large minimum collective degree. *J. Combin. Theory Ser. A* 116.3 (2009), pp. 613–636.

[114] H. J. Ryser. 'Neuere Probleme der Kombinatorik'. *Vorträge über Kombinatorik.* 1967, pp. 69–91.

[115] G. N. Sárközy, S. M. Selkow and E. Szemerédi. On the number of Hamiltonian cycles in Dirac graphs. *Discrete Math.* 265.1-3 (2003), pp. 237–250.

[116] W. Schudy and M. Sviridenko. 'Bernstein-like concentration and moment inequalities for polynomials of independent random variables: multilinear case'. arXiv preprint arXiv:1109.5193. 2011.

[117] W. Schudy and M. Sviridenko. 'Concentration and moment inequalities for polynomials of independent random variables'. arXiv preprint arXiv:1104.4997. 2011.

[118] P. W. Shor. A lower bound for the length of a partial transversal in a Latin square. *J. Combin. Theory Ser. A* 33.1 (1982), pp. 1–8.

[119] S. K. Stein. Transversals of Latin squares and their generalizations. *Pacific J. Math.* 59.2 (1975), pp. 567–575.

[120] A. A. Taranenko. Multidimensional permanents and an upper bound on the number of transversals in Latin squares. *J. Combin. Des.* 23.7 (2015), pp. 305–320.

[121] V. H. Vu. Concentration of non-Lipschitz functions and applications. *Random Structures Algorithms* 20.3 (2002), pp. 262–316.

[122] V. H. Vu. New bounds on nearly perfect matchings in hypergraphs: Higher codegrees do help. *Random Structures Algorithms* 17.1 (2000), pp. 29–63.

[123] I. M. Wanless. Cycle switches in Latin squares. *Graphs Combin.* 20 (2004), pp. 545–570.

[124] D. B. West. *Introduction to Graph Theory, 2nd ed.* Pearson, 2000.

[125] D. Woolbright. An $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols. *J. Combin. Theory Ser. A* 24.2 (1978), pp. 235–237.

[126] R. Yuster. Combinatorial and computational aspects of graph packing and graph decomposition. *Computer Science Review* 1.1 (2007), pp. 12–26.

[127] Y. Zhao. ‘Recent advances on Dirac-type problems for hypergraphs’. *Recent trends in Combinatorics*. Springer, 2016, pp. 145–165.