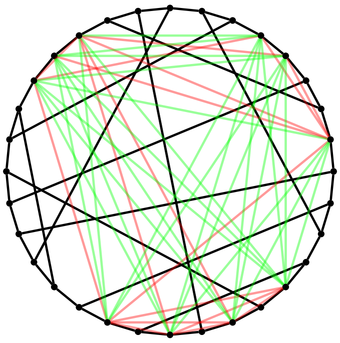
האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM
الجامعة العبرية في اورشليم القدس

Finding Structure with Randomness



A thesis submitted for the degree of
Doctor of Philosophy

By
Michael Simkin

Submitted to the senate of the Hebrew University
March 2020

This work was carried out under the supervision of

Nati Linial

Acknowledgements

“It takes a village,” we’re told, “to raise a child.” Besides parents, a child needs many support circles to create an environment just right for them to flourish. This is surely true also when raising a scholar.

Beginning with the innermost circle, I wish to thank Nati Linial, my academic parent. Besides the obvious - combinatorics - you’ve taught me to have high aspirations, and not to shy away from tough questions. More importantly, you’ve taught by example that a true person of learning - a *philosopher* - should not focus narrowly on their own field of expertise. Rather, they must seek out a broader view of science and society, and find ways to make their contribution.

Moving outwards, I wish to thank my collaborators, Zur Luria and Roman Glebov. Zur - thank you for teaching me the intricacies of Keevash’s methods, but also that when faced with a challenging problem, it might be most productive to visit downtown Zürich and enjoy the *Sechseläuten.* Roman - you taught me the value of an aggressive approach to mathematics (“This problem is not solved until we have a hitting time result!”). Thank you both for being such a pleasure to work with.

I was inspired to take an academic path by many people, but I owe a special debt to Reuven Gellman ז”ל. At a young age I was impressed by the depth and breadth of your knowledge, and obtaining my own understanding of the world around us became a goal. I would have loved to share with you the first fruits of my own research. Sadly, these arrived too late.

I was lucky to be raised in a home where scholarship was a central value. I have my parents, Shlomo and Elana, to thank for that. Abba - thank you for modeling the hard work necessary to produce good technical writing, and for giving me a copy of *Strunk & White.* Ima - I’ve learned much pedagogy from you over the years. In particular, that enthusiasm is a teacher’s best friend and never to underestimate a good prop.

The intellectual journey chronicled in this thesis took place against the backdrop of my life with Shanee. Five years ago we married and, with Noam Yonat’s arrival, became a family. Noam - it is a constant joy watching you learn and grow. I can’t wait to see what we'll discover together in the years to come. Shanee - these years have been the most fulfilling of my life. Together we've experienced the greatest joys and profound loss. Alongside these, it has been a privilege sharing my personal academic journey with you. Thanks for putting up with my habit of sitting in the living room, physically present, but oblivious to the goings-on around while following my thoughts down an ill-begotten mathematical path. Most of all, thank you for supporting me through thick and thin, whenever needed. I lovingly dedicate this thesis to you.

Abstract

This thesis explores the use of randomized algorithms to construct and study regular combinatorial objects, especially in high dimensions.

High-dimensional permutations are one of our focal points. Continuing a line of investigation initiated by Linial and Luria, we define a d-dimensional permutation of order n as an $n \times n \times \ldots \times n = [n]^{d+1}$ array taking values in $\{0, 1\}$, with every axis-parallel line containing a single 1. Thus, a one-dimensional permutation is just a permutation matrix, and a two-dimensional permutation is synonymous with a Latin square.

In Chapter 2 we generalize the Erdős-Szekeres theorem to high-dimensional permutations. We show that every order-n d-dimensional permutation has a monotone subsequence of length $\Omega_d(\sqrt{n})$, and this bound is tight. We also show that the length of the longest monotone subsequence in a typical (i.e., uniformly random) d-dimensional permutation of order n is with high probability $\Theta_d\left(n^{d/(d+1)}\right)$ [2].

Chapter 3 studies the *threshold problem for Latin squares.* For positive integers $m \leq n \leq k$ we say that an $m \times n \times k$ $0-1$ array is a *Latin box* if it contains exactly mn 1s, and every axis-parallel line contains at most one 1. Thus, a Latin box with $m = n = k$ coincides with an order-n Latin square. When m and k are close to n, this may be viewed as an approximate Latin square. Let $\mathcal{M}(m, n, k; p)$ be the distribution on $m \times n \times k$ $0-1$-arrays where each entry is 1 with probability p, independently. The threshold problem for Latin squares is to determine for which p it holds with high probability that $\mathcal{M}(n, n, n; p)$ supports a Latin square. We ask more generally when $\mathcal{M}(m, n, k; p)$ supports a Latin box with high probability. For every $\varepsilon > 0$, we give an asymptotically tight answer for the cases where either $m = n$ and $k \geq (1 + \varepsilon)n$ or $m \leq (1 - \varepsilon)n$ and $k = n$. In both cases, the threshold is $\Theta(\log(n)/n)$ [4].

Chapter 4 revisits the foundational result of Erdős and Rényi that the threshold at which a perfect matching appears in random bipartite graphs is the same as the one at which all isolated vertices disappear. We consider the random process where a k-regular bipartite graph G with $2n$ vertices is revealed edge by edge in

a uniformly random order. We show that if $k = \omega\left(n/\log^{1/3}(n)\right)$ then with high probability a perfect matching appears at the very moment that the last isolated vertex disappears. When $G = K_{n,n}$ this is the well-known hitting-time version, due to Bollobás and Thomason, of the classical Erdős-Rényi result. On the other hand, we show that for k as large as $\Omega\left(n/(\log(n)\log\log(n))\right)$ there exist graphs in which the last isolated vertex disappears well before any perfect matching appears [1].

Chapter 5 describes a new random greedy algorithm to construct regular graphs with large girth. Let $k \geq 3$ and $c < 1$ be fixed. Let n be even and set $g = c\log_{k-1}(n)$. Begin with a Hamilton cycle G on n vertices. As long as $\delta(G) < k$, choose, uniformly at random, two vertices u, v of degree $\delta(G)$ subject to the condition that their distance is at least $g - 1$. If there are no such pairs, abort. Otherwise, add the edge uv to $E(G)$. We show that with high probability this process terminates with a k-regular graph, which by definition of the algorithm has girth at least g. Our analysis also yields a lower bound on the number of high-girth regular graphs [3].

Bibliography

[1] Roman Glebov, Zur Luria, and Michael Simkin, *Perfect matchings in random subgraphs of regular bipartite graphs*, arXiv preprint arXiv:1805.06944 (2018).

[2] Nathan Linial and Michael Simkin, *Monotone subsequences in high-dimensional permutations*, Combinatorics, Probability and Computing **27** (2018), no. 1, 69–83.

[3] Nati Linial and Michael Simkin, *A randomized construction of high girth regular graphs*, arXiv preprint arXiv:1911.09640 (2019).

[4] Zur Luria and Michael Simkin, *On the threshold problem for Latin boxes*, Random Structures & Algorithms **55** (2019), no. 4, 926–949.

Letter of Contribution

This thesis consists of four manuscripts:

1. “Monotone subsequences in high-dimensional permutations”, that was published in *Combinatorics, Probability and Computing*, is a joint work with my advisor Nati Linial.
2. “On the threshold problem for Latin boxes”, that was published in *Random Structures & Algorithms*, is a joint work with Zur Luria.
3. “Perfect matchings in random subgraphs of regular bipartite graphs”, that is submitted for publication, is a joint work with Roman Glebov and Zur Luria.
4. “A randomized construction of high-girth regular graphs”, that is submitted for publication, is a joint work with my advisor Nati Linial.

I am a primary writer and contributor to all parts of this thesis.

Contents

Chapter 1

Introduction

INTRODUCTION

1. Finding structure with randomness

This thesis investigates the construction of regular combinatorial objects, mostly high dimensional ones, using *randomized algorithms.*

As a motivating example of high-dimensional regularity, we consider *high-dimensional permutations.* An order-n permutation can be viewed as an $n \times n$ permutation matrix. Correspondingly, we define an order-n **d-dimensional permutation** as a $(0, 1)$-array indexed by $n \times n \times \ldots \times n = [n]^{d+1}$, where each axis-parallel line contains exactly one 1. It is of great interest to study the *extremal* properties of high-dimensional permutations as well as their *typical* behavior.

We also pursue a line of work in which we observes interesting phenomena in the realm of (one-dimensional) permutations and ask whether they have interesting analogues in high dimensions.

Chapter 2 is a case in point. It is a joint work with Nati Linial [27], that generalizes the Erdős-Szekeres theorem to high-dimensional permutations, and answers extremal and probabilistic questions relating to monotone subsequences. While working on these results it became apparent that a fundamental difficulty in understanding high-dimensional regularity is the dearth of probabilistic techniques applicable to this setting. This has led us to a change of perspective. Rather than directly studying high-dimensional regularity, we began a search for (randomized) algorithms to find and construct regular structures in various constrained settings. Thus, in Chapter 3, which is a joint work with Zur Luria [31], we ask when an approximate Latin square can be found in a random hypergraph. In Chapter 4, coauthored with Roman Glebov and Zur Luria [13], we consider the random process wherein a graph G is reconstructed edge by edge in a random order. We ask at which moment of such a process the evolving graph contains a perfect matching. Following Krivilevich, Lee, and Sudakov [24], We view this as a measure of *robustness* for the property "G contains a perfect matching". In Chapter 5, written with Nati Linial [28], we introduce a random greedy algorithm for constructing regular graphs with high girth. Finally, in Chapter 6 we suggest several research directions related to the various results in this thesis.

Each one of Chapters 2-5 is a standalone paper, and their progression traces the author's intellectual journey over the last few years. Each paper has its own introduction which we hope gives sufficient motivation and background for the results therein. Naturally, in most cases there have been pertinent developments in the time since publication. In the remainder of this introduction we complement the papers by painting a broader picture of the landscape as we currently see it.

2. High-dimensional permutations

We have already defined high-dimensional permutations as $(0, 1)$-valued $[n]^{d+1}$-arrays in which each axis-parallel line contains a single 1. This is equivalent to the following:

- An order-n d-dimensional Latin hypercube, i.e., an $[n]^d$-array in which each axis-parallel line is a permutation of $[n]$. In two dimensions this is a **Latin square** - an $n \times n$ matrix in which each row and each column contains every integer between 1 and n.
- A (labeled) $(d+1)$**-clique-decomposition** of the complete balanced $(d+1)$-partite hypergraph with $(d+1)n$ vertices, i.e., a $(d+1)$-uniform $(d+1)$-partite hypergraph in which each partite d-set is contained in exactly one hyperedge.

Latin squares are well-known and classical objects, and their study extends back at least to the early 18th century [6, p. 11]. Nevertheless, even such basic questions as their enumeration were open until the late 20th century: using permanent inequalities, van-Lint and Wilson showed that there are $((1 \pm o(1))n/e^2)^{n^2}$ order-n Latin squares [38, Theorem 17.3]. More recently, this enumeration was extended to all dimensions: There are $\left((1 \pm o_d(1))\, n/e^d\right)^{n^d}$ order-n d-dimensional permutations. The upper bound follows from the *entropy method* of Linial and Luria [26] and the lower bound from Keevash's method of *randomized algebraic constructions* [23, Theorem 1.8].

In Chapter 2 (which is a joint work with Nati Linial [27]) we consider *monotone subsequences* in high-dimensional permutations. A basic result regarding (one-dimensional) permutations is the Erdős-Szekeres theorem: every order-n permutation contains a monotone subsequence of length $\sqrt{n}$, and this is tight [9]. With the advent of probabilistic combinatorics and computer experimentation began the study of monotone subsequences in random permutations [37]. A series of advances [17, 29, 39] culminated in the understanding that in a uniformly random permutation, w.h.p.[1] the longest monotone subsequence has length $(1 \pm o(1))2\sqrt{n}$.

We generalize these results to the setting of high-dimensional permutations. Deferring the precise definitions to Chapter 2, our main results are:

- As in the one-dimensional case, every order-n d-dimensional permutation contains a monotone subsequence of length $\Omega_d(\sqrt{n})$, and this is tight;
- On the other hand, for $d \geq 2$, the typical case differs significantly from the extremal case: w.h.p. the longest monotone subsequence in a d-dimensional permutation has length $\Theta_d(n^{d/(d+1)})$.

The second result adds to the short list of facts known about typical Latin squares (on which we focus for simplicity). While pleasing, its proof uses very limited randomness. Indeed, the proof proceeds by showing that the result holds for the distribution obtained by taking an *arbitrary* Latin square and then permuting its symbols uniformly at random. This raises more questions than it answers: clearly, this technique is inapplicable to isotopy-invariant[2] properties - and these are precisely the properties of interest from the perspective of hypergraph theory.

In the time passed since Chapter 2 was written, there has been substantial progress on this front. Most notably, Kwan [25] developed a method to transfer certain properties from the *triangle removal process* to random Latin squares (and related combinatorial designs). This was then applied to show that almost all Latin squares contain a transversal. Kwan's method is especially suited to proving lower bounds

[1] We say that a sequence of events occurs **with high probability** (**w.h.p.**) if the probabilities of their occurrence tend to 1.

[2] Two Latin squares are **isotopic** if one can be obtained by permuting the rows, columns, and symbols of the other.

on subgraph counts in random Latin squares (see [34] for an example). However, the technique applies only to a restricted class of properties: crucially, only those which hold with extremely high probability in the triangle removal process. As a result, there is still much to look forward to.

3. Regular structures in random hypergraphs

In what follows, we denote by $\mathcal{H}_k(n;p)$ the random binomial distribution on k-partite k-uniform hypergraphs with n vertices in each part, where each hyperedge is present independently with probability p. Thus, $\mathcal{H}_2(n;p)$ is the usual Erdős-Rényi model $\mathcal{G}(n,n;p)$ for random bipartite graphs.

Beginning with the seminal work of Erdős and Rényi [7, 8], a central concern in the study of random (hyper)graphs has been finding *thresholds* for interesting properties. A particular focus is the emergence of spanning regular structures. For example, $\log(n)/n$ is the threshold[3] for $\mathcal{G}(n,n;p)$ to contain a perfect matching [8, Theorem 1]. Despite its apparent simplicity, generalizing this result to higher dimensions has proved challenging. In the simplest instance, we may ask what the threshold is for $\mathcal{H}_k(n;p)$ to contain a perfect matching. This became known as *Shamir's problem* [33], and was open for nearly three decades before its resolution in breakthrough work of Johansson, Kahn, and Vu [18] and Kahn [20]. In a very recent breakthrough, Frankston, Kahn, Narayanan, and Park [12] found a simpler solution via the notion of *fractional expectation thresholds.* Of course, perfect matchings are but one example of high-dimensional regularity, and it is just as natural to study the emergence of high-dimensional permutations (and other combinatorial designs) in random hypergraphs. For simplicity, we focus on Latin squares (which we view hypergraphically as triangle-decompositions of $K_{n,n,n}$). Unfortunately, it seems neither of the techniques used to solve Shamir's problem are applicable to Latin squares (see [12, Section 8]). This sets the stage for the following question, which we call "the threshold problem for Latin squares".

Question 3.1. *What is the threshold $p_{LS}(n)$ for $H \sim \mathcal{H}_3(n;p)$ to contain an order-n Latin square?*

There is an obvious lower bound: If $H \sim \mathcal{H}_3(n;p)$ contains a Latin square, then every edge in $K_{n,n,n}$ is contained in at least one triangle of H. Put differently, if H contains a Latin square then it does not have any isolated edges. It is not difficult to show that the threshold for H having no isolated edges is $2\log(n)/n$, implying $p_{LS}(n) \geq 2\log(n)/n$. In Shamir's problem, the threshold for the appearance of perfect matchings turned out to be the same as the analogous lower bound, namely, the disappearance of isolated vertices. It is thus tantalizing to raise:

Conjecture 3.2. *The threshold for the appearance of spanning Latin squares in $H \sim \mathcal{H}_3(n;p)$ is $2\log(n)/n$.*

Unfortunately, this remains purely conjectural. Nevertheless, this thesis contains some results that support Conjecture 3.2. In Chapter 3, which is a joint work with Zur Luria [31], we determine the threshold for the appearance of an approximation of Latin squares - which we term **Latin boxes** - in random hypergraphs. The main

[3] By this we mean that if $p = p(n) = o(\log(n)/n)$ then w.h.p. $\mathcal{G}(n,n;p)$ does not contain a perfect matching, whereas if $p = \omega(\log(n)/n)$ then w.h.p. it does. Sharper results are known - see Section 4.

tool is the use of random greedy algorithms to find the desired objects in random settings. We find that the threshold for the appearance of Latin boxes in random hypergraphs is indeed the same as a lower bound analogous to the notion of isolated edges.

With respect to upper bounds on p_{LS}, the best bound currently in the literature is due to Keevash, and follows from his method of randomized algebraic constructions: there exists an (exceedingly small) $\varepsilon > 0$ such that $p_{LS}(n) \leq n^{-\varepsilon}$ [23, Theorem 1.7].

In fact, based on limited numerical evidence, we wonder if it is true not not only that $p_{LS}(n) = 2\log(n)/n$, but that this holds in the sharpest form possible: Consider the random process where, beginning with an empty tripartite balanced hypergraph on $3n$ vertices, triangles are added one by one in a uniformly random order. Let τ_I be the time at which there are no longer any isolated edges, and let τ_{LS} be the time at which the hypergraph contains an order-n Latin square. Clearly, $\tau_I \leq \tau_{LS}$.

Question 3.3. *In the process above, is it true that w.h.p. $\tau_{LS} = \tau_I$?*

4. Robustness of graph properties

A recurring theme in the study of random graphs - mentioned in passing in the previous section - is that the threshold for the appearance of spanning structures is often the same as the threshold for the disappearance of local obstructions. Thus, the threshold for the appearance of perfect matchings in $\mathcal{G}(n;p)$ (for n even) is the same as for the disappearance of isolated vertices [8]. Similarly, Hamilton cycles appear in $\mathcal{G}(n;p)$ when the minimum degree becomes two [32]. As the study of random graphs advanced, the following understanding emerged: if we consider the random process where, beginning with n isolated vertices, we construct a sequence of graphs by adding edges one by one in a uniformly random order, then w.h.p. a perfect matching appears as soon as there are no isolated vertices [5], and a Hamilton cycle appears as soon as the minimum degree is two [4, 1].

Krivelevich, Lee, and Sudakov [24] suggest interpreting these results as a measure of *robustness*: K_n is as robust as possible with respect to Hamiltonicity (resp., containing a perfect matching), since w.h.p. a random subgraph of K_n fails to be Hamiltonian (resp., contain a perfect matching) only for trivial reasons - having minimum degree less than 2 (resp., having an isolated vertex). With this in mind, we consider the following problem: Let G be a graph satisfying a nontrivial increasing property $\mathcal{P}$. For $p \in [0,1]$, let $G(p)$ be a random binomial subgraph of G, where each edge is retained with probability p. What is the threshold for $G(p)$ to have property $\mathcal{P}$? Similarly, consider the process where G is reconstructed edge by edge in a uniformly random order. What is the hitting time for property $\mathcal{P}$? Do these quantities have natural interpretations in terms of local obstructions?

In Chapter 4, a joint work with Roman Glebov and Zur Luria [13], we study the robustness of regular bipartite graphs with respect to containing a perfect matching. Our main result is that for $k = \omega\left(n/\log^{1/3}(n)\right)$, if G is a k-regular bipartite graph on $2n$ vertices, then with high probability the hitting time for the appearance of perfect matchings is the same as for the disappearance of isolated vertices. This implies that the (sharp) threshold for $G(p)$ to contain a perfect matching is $p = \log(n)/k$. Thus, such graphs are as robust as possible with respect to containing perfect matchings. For this range of degrees, this improves on a result of Goel, Kapralov, and Khanna

[16], who showed that $p = O\left(n\log(n)/k^2\right)$ is an upper bound on the threshold for $G(p)$ to contain a perfect matching.

This notion of robustness has algorithmic implications (in fact, these were the original motivation for [16]). For example, consider the problem of finding a perfect matching in a graph G to which our result applies. Rather than searching for a perfect matching directly in G (which has $\tilde{\Theta}(n^2)$ edges), one can first sample a subgraph $G(p)$, with $p = (1 + o(1))\log(n)/k$. By our result, w.h.p. $G(p)$ contains a perfect matching. Furthermore, it has only $(1 + o(1))n\log(n)$ edges. One can now search for a perfect matching in the much smaller graph $G(p)$. Perhaps surprisingly, the number of edges in $G(p)$ depends only on n, and not on k. This leads to an algorithm for finding perfect matchings with runtime independent of the number of edges in the input graph. We stress that this idea appeared first in [16], and in this context our only improvement is to shave a polylogarithmic factor from the runtime. This is because for our result to apply we require $k = \omega\left(n/\log^{1/3}(n)\right)$. On the other hand, [16] guarantees a perfect matching in a random subgraph of G with $O\left(kn \cdot n\log(n)/k^2\right) = o\left(n\log^{4/3}(n)\right)$ edges.

The main tool used to prove our result is a graph-decomposition similar in flavor to the regularity lemma. This allows us to classify those vertex sets in G that are most likely to violate Hall's condition in $G(p)$. Together with a union bound, this is enough to show that w.h.p. when no isolated vertices remain there are also no violations of Hall's condition, and so there is a perfect matching.

Notably, our proof is quite specific to perfect matchings. Similarly, Krivelevich, Lee, and Sudakov [24] and Johansson [19] used techniques specific to Hamilton cycles to study the robustness of Hamiltonicity in Dirac graphs. This is not surprising, since the original proofs pertaining to thresholds in $\mathcal{G}(n;p)$ are also specific to the properties of interest. However, a new development has led to a unified proof of many threshold results in $\mathcal{G}(n;p)$, and promises to shed light on the relationship between global properties and local obstructions more generally.

As already mentioned, Kahn, Narayanan, and Park [12] related the notion of fractional expectation thresholds with threshold functions for properties of random graphs. This is a major breakthrough, proving a conjecture of Talagrand [36], which in turn was inspired by Kahn and Kalai's conjectures [21] relating thresholds with easy-to-understand obstructions. In short, they reduce the problem of determining the threshold for $\mathcal{G}(n;p)$ to contain a perfect matching (resp., a Hamilton cycle) to the enumerative problem of showing that for every set of edges $S \subseteq E(K_n)$ there are at most $(O(n^{-1}))^{|S|}P_n$ (resp., $(O(n^{-1}))^{|S|}H_n$) perfect matchings (resp., Hamilton cycles) containing S. Here, P_n and H_n are, respectively, the number of perfect matchings and Hamilton cycles in K_n. Similar calculations would imply threshold results for graph properties in $G(p)$, for arbitrary G. However, it is not clear how hard these calculations are in practice.

5. The surprising success of random greedy algorithms

Chapter 5, the last paper included in this thesis, is a joint work with Nati Linial [28]. In it, we return to the primary motivation of the probabilistic method: constructions of interesting objects. Concretely, we describe a *random greedy algorithm* to construct k-regular graphs with n vertices and girth[4] $(1 - o_k(1)) \log_{k-1}(n)$. Such algorithms have their roots in the *Rödl nibble* [11], which was first used to show the existence of approximate Steiner systems[5]. Later, Spencer [35] showed that the following random greedy algorithm succeeds, w.h.p., in constructing an approximate ((n, q, r)-Steiner system (with q, r fixed)): Begin with the empty hypergraph on n vertices. Then, for as long as possible, choose a q-set e uniformly at random (and independently of previous choices) subject to the condition that it intersects each previously chosen set on at most $r - 1$ vertices. Add the edge e to the hypergraph.

Since Rödl and Spencer, random greedy algorithms have played an increasingly prominent role in probabilistic constructions. In particular, they were heavily used by Keevash [22] and Glock, Kühn, Lo, and Osthus [14] in their breakthrough constructions of (exact) Steiner systems. As another example, Glock, Kühn, Lo, and Osthus [15], and independently Bohman and Warnke [3], used a random greedy algorithm to construct approximate Steiner triple systems (i.e., $(n, 3, 2)$-Steiner systems) with large girth[6]. It is remarkable that such simple algorithms do so well in constructing constrained combinatorial objects. Furthermore, their analysis typically proceeds via tight bounds on the number of choices available to the algorithm at each step. Thus, they often give a lower bound on the count of the desired object. Finally, as in Chapter 5, they are often efficient, leading to concrete examples of the desired object which in turn allow its empirical study.

To put the results of Chapter 5 in context, we note that high-girth regular graphs are one of the few cases where algebraic constructions do significantly better than probabilistic ones. Specifically, probabilistic techniques have yielded essentially no improvement on the very first construction of Erdős and Sachs [10], who described k-regular graphs with girth $(1-o_k(1)) \log_{k-1}(n)$ (where n is the number of vertices). On the other hand, beginning with Biggs and Hoare [2] and Lubotzky, Phillips, and Sarnak [30], a series of advances has shown that Cayley graphs exist with girth $\geq \frac{4}{3} \log_{k-1}(n)$. It is therefore interesting to understand, precisely, the obstacles to improving the probabilistic constructions. To this end, it is helpful to have at our disposal many different probabilistic constructions of high-girth regular graphs.

References

[1] Miklós Ajtai, János Komlós, and Endre Szemerédi, *First occurrence of Hamilton cycles in random graphs*, North-Holland Mathematics Studies **115** (1985), no. C, 173–178.

[2] NL Biggs and MJ Hoare, *The sextet construction for cubic graphs*, Combinatorica **3** (1983), no. 2, 153–165.

[3] Tom Bohman and Lutz Warnke, *Large girth approximate Steiner triple systems*, Journal of the London Mathematical Society (2018).

[4] Béla Bollobás, *The evolution of sparse graphs*, Graph Theory and Combinatorics, Academic Press, London, 1984, pp. 35–57.

[5] Béla Bollobás and Andrew Thomason, *Random graphs of small order*, North-Holland Mathematics Studies, vol. 118, Elsevier, 1985, pp. 47–97.

[6] Charles J Colbourn and Jeffrey H Dinitz, *Handbook of combinatorial designs*, 2 ed., CRC press, 2006.

[4] The **girth** of a graph is the length of its shortest cycle.

[5] For $n, q, r \in \mathbb{N}$, an (n, q, r)**-Steiner system** is a q-uniform hypergraph on n vertices in which every r-set is contained in exactly one hyperedge.

[6] The **girth** of a Steiner triple system is the smallest number of vertices $k \geq 4$ that span at least $k - 2$ hyperedges. This naturally generalizes graphical girth, which is the smallest number k of vertices that span k edges.

[7] Paul Erdős and Alfréd Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci **5** (1960), 17–61.

[8] ______, *On random matrices*, Magyar Tud. Akad. Mat. Kutató Int. Közl **8** (1964), 455–461.

[9] Paul Erdős and George Szekeres, *A combinatorial problem in geometry*, Compositio mathematica **2** (1935), 463–470.

[10] Paul Erdős and Horst Sachs, *Reguläre graphen gegebener taillenweite mit minimaler knotenzahl*, Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe **12** (1963), no. 251-257, 22.

[11] Péter Frankl and Vojtech Rödl, *Near perfect coverings in graphs and hypergraphs*, European Journal of Combinatorics **6** (1985), no. 4, 317–326.

[12] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park, *Thresholds versus fractional expectation-thresholds*, arXiv preprint arXiv:1910.13433 (2019).

[13] Roman Glebov, Zur Luria, and Michael Simkin, *Perfect matchings in random subgraphs of regular bipartite graphs*, arXiv preprint arXiv:1805.06944 (2018).

[14] Stefan Glock, Daniela Kühn, Allan Lo, and Deryk Osthus, *The existence of designs via iterative absorption*, arXiv preprint arXiv:1611.06827 (2016).

[15] ______, *On a conjecture of Erdős on locally sparse Steiner triple systems*, arXiv preprint arXiv:1802.04227 (2018).

[16] Ashish Goel, Michael Kapralov, and Sanjeev Khanna, *Perfect matchings via uniform sampling in regular bipartite graphs*, ACM Transactions on Algorithms (TALG) **6** (2010), no. 2, 27.

[17] John M Hammersley et al., *A few seedlings of research*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics, The Regents of the University of California, 1972.

[18] Anders Johansson, Jeff Kahn, and Van Vu, *Factors in random graphs*, Random Structures & Algorithms **33** (2008), no. 1, 1–28.

[19] Tony Johansson, *On Hamilton cycles in Erdős-Rényi subgraphs of large graphs*, Random Structures & Algorithms (2020).

[20] Jeff Kahn, *Asymptotics for Shamir's problem*, arXiv preprint arXiv:1909.06834 (2019).

[21] Jeff Kahn and Gil Kalai, *Thresholds and expectation thresholds*, Combinatorics, Probability and Computing **16** (2007), no. 3, 495–502.

[22] Peter Keevash, *The existence of designs*, arXiv preprint arXiv:1401.3665 (2014).

[23] ______, *The existence of designs II*, arXiv preprint arXiv:1802.05900 (2018).

[24] Michael Krivelevich, Choongbum Lee, and Benny Sudakov, *Robust hamiltonicity of dirac graphs*, Transactions of the American Mathematical Society **366** (2014), no. 6, 3095–3130.

[25] Matthew Kwan, *Almost all Steiner triple systems have perfect matchings*, arXiv preprint arXiv:1611.02246 (2016).

[26] Nathan Linial and Zur Luria, *An upper bound on the number of high-dimensional permutations*, Combinatorica **34** (2014), no. 4, 471–486.

[27] Nathan Linial and Michael Simkin, *Monotone subsequences in high-dimensional permutations*, Combinatorics, Probability and Computing **27** (2018), no. 1, 69–83.

[28] Nati Linial and Michael Simkin, *A randomized construction of high girth regular graphs*, arXiv preprint arXiv:1911.09640 (2019).

[29] Benjamin F Logan and Larry A Shepp, *A variational problem for random Young tableaux*, Advances in mathematics **26** (1977), no. 2, 206–222.

[30] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak, *Ramanujan graphs*, Combinatorica **8** (1988), no. 3, 261–277.

[31] Zur Luria and Michael Simkin, *On the threshold problem for Latin boxes*, Random Structures & Algorithms **55** (2019), no. 4, 926–949.

[32] Lajos Pósa, *Hamiltonian circuits in random graphs*, Discrete Mathematics **14** (1976), no. 4, 359–364.

[33] Jeanette Schmidt and Eli Shamir, *A threshold for perfect matchings in random d-pure hypergraphs*, Discrete mathematics **45** (1983), no. 2-3, 287–295.

[34] Michael Simkin, *Lecture notes: methods for analyzing random designs - IIAS special day on combinatorial design theory*, `http://math.huji.ac.il/~michaels/files/papers/pasch_in_sts.pdf`, 2018.

[35] Joel Spencer, *Asymptotic packing via a branching process*, Random Structures & Algorithms **7** (1995), no. 2, 167–172.

[36] Michel Talagrand, *Are many small sets explicitly small?*, Proceedings of the forty-second ACM symposium on Theory of computing, 2010, pp. 13–36.

[37] Stanislaw M Ulam, *Monte Carlo calculations in problems of mathematical physics*, Modern Mathematics for the Engineers (1961), 261–281.

[38] JH Van Lint and Richard M Wilson, *A course in combinatorics. 1992.*

[39] Anatolii Moiseevich Vershik and Sergei Vasilevich Kerov, *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux*, Doklady Akademii Nauk, vol. 233, Russian Academy of Sciences, 1977, pp. 1024–1027.

Chapter 2

Monotone subsequences in high-dimensional permutations

Nati Linial and Michael Simkin. *Combinatorics, Probability and Computing* 27 (2018), no. 1, 69–83.

MONOTONE SUBSEQUENCES IN HIGH-DIMENSIONAL PERMUTATIONS

NATHAN LINIAL AND MICHAEL SIMKIN

Abstract. This paper is part of the ongoing effort to study high-dimensional permutations. We prove the analogue to the Erdős–Szekeres Theorem: For every $k \geq 1$, every order-n k-dimensional permutation contains a monotone subsequence of length $\Omega_k\left(\sqrt{n}\right)$, and this is tight. On the other hand, and unlike the classical case, the longest monotone subsequence in a random k-dimensional permutation of order n is asymptotically almost surely $\Theta_k\left(n^{\frac{k}{k+1}}\right)$.

1. Introduction

The study of monotone subsequences in permutations began with the famous Erdős–Szekeres Theorem [5]. Since then numerous proofs and generalizations have emerged (see Steele's survey [14]). We recall the theorem.

Theorem 1.1. *Every permutation in S_n contains a monotone subsequence of length at least $\lceil\sqrt{n}\rceil$, and this is tight: for every n there exists some permutation in S_n in which all monotone subsequences are of length at most $\lceil\sqrt{n}\rceil$.*

In order to derive a high-dimensional analogue of Theorem 1.1 we need to define high-dimensional permutations and their monotone subsequences. If we view a permutation as a sequence of distinct real numbers, it is suggestive to consider sequences of points in $\mathbb{R}^k$, with coordinatewise monotonicity. The following argument is attributed by Kruskal [9] to de Bruijn: Repeatedly apply Theorem 1.1 to conclude that every sequence $x_1, x_2, \ldots, x_n \in \mathbb{R}^k$ must have a coordinatewise monotone subsequence of length $n^{\frac{1}{2^k}}$, and this is tight up to an additive constant. In [9] one considers projections of the points to a line and defines the length of the longest monotone subsequence according to the line with the longest such subsequence. Szabó and Tardos [15] consider sequences in $\mathbb{R}^k$ that avoid at least one of the 2^k coordinatewise orderings.

Here we adopt the perspective of [11] of a high-dimensional analogue of permutation matrices, and monotone subsequences are defined by strict coordinatewise monotonicity. We show (Theorem 2.2) that every k-dimensional permutation of order n has a monotone subsequence of length $\Omega_k\left(\sqrt{n}\right)$, and this is tight up to the implicit multiplicative constant.

A related question, posed by Ulam [16] in 1961, concerns the distribution of H_n^1, the length of the longest increasing subsequence in a random member of S_n. In 1972 Hammersley [6] showed that there exists some $C > 0$ s.t. $H_n^1 n^{-\frac{1}{2}}$ converges to C in probability. In 1977 Logan and Shepp [12] showed that $C \geq 2$ and Vershik and Kerov [17] demonstrated that $C \leq 2$. This yields the next theorem.

2000 *Mathematics Subject Classification.* Primary 05B15.

Supported by ERC grant 339096 *High-Dimensional Combinatorics.*

Theorem 1.2. *Let H_n^1 be the length of the longest increasing subsequence in a uniformly random member of S_n. Then $\lim_{n\to\infty} H_n^1 n^{-\frac{1}{2}} = 2$ in probability.*

This result was famously refined in 1999 by Baik, Deift, and Johansson [1] who related the limiting distribution of H_n^1 to the Tracy–Widom distribution.

Using coordinatewise monotonicity Bollobás and Winkler [3] extended Theorem 1.2 to show that the longest increasing subsequence among n independently random points in $[0,1]^k$ is typically of length $c_k n^{\frac{1}{k}}$ for some $c_k \in (0,e)$. We show (Theorem 4.1) that the longest monotone subsequence of a typical k-dimensional permutation of order n has length $\Theta_k\left(n^{\frac{k}{k+1}}\right)$. A k-dimensional permutation can be viewed as a set of n^k points in $[0,1]^k$, and it is interesting to note this asymptotic match with Bollobás and Winkler's result.

2. Definitions and Main Results

Note: Throughout the paper all asymptotic expressions are in terms of $n \to \infty$ and k fixed.

As discussed in [11] and [10], we equate a permutation with the corresponding permutation matrix, i.e., an $n \times n$ $(0,1)$-matrix in which each row or column (henceforth, *line*) contains a single 1. We correspondingly define an *order-n k-dimensional permutation* as an $[n]^{k+1}$ $(0,1)$-array in which each line contains precisely one 1. A *line* in an $[n]^{k+1}$ array is comprised of all the positions obtained by fixing k coordinates and varying the remaining coordinate. We denote the set of order-n k-dimensional permutations by L_n^k.

For a given $A \in L_n^k$ and $\alpha \in [n]^k$, there is a unique $t \in [n]$ s.t. $A(\alpha, t) = 1$. Since t is uniquely defined by α, we can write $t = f_A(\alpha)$. The function f_A has the property that if we fix $k-1$ coordinates and vary the remaining coordinate, the result is a permutation of $[n]$. In fact, the mapping $A \mapsto f_A$ is a bijection between L_n^k and the family of $[n]^k$ arrays in which every line is a permutation of $[n]$. In dimension one this is exactly the identification between permutation matrices and permutations. This shows in particular that two-dimensional permutations, i.e., members of L_n^2, are order-n *Latin squares.*

We denote by G_A the *support* of $A \in L_n^k$, i.e., the set of $\alpha \in [n]^{k+1}$ s.t. $A(\alpha) = 1$.

The next definition generalizes monotonicity to higher dimensions.

Definition 2.1. A length-m monotone subsequence in $A \in L_n^k$ is a sequence $\alpha^1, \alpha^2, \dots, \alpha^m \in G_A$ s.t. for every $1 \le j \le k+1$ the sequence $\alpha_j^1, \alpha_j^2, \dots, \alpha_j^m$ is *strictly* monotone.

In dimension one this clearly coincides with the definition of a monotone subsequence in a permutation $\pi \in S_n$.

We are now ready to state a high-dimensional analogue of the Erdős–Szekeres Theorem.

Theorem 2.2. *Every member of L_n^k contains a monotone subsequence of length $\Omega_k(\sqrt{n})$. The bound is tight up to the implicit multiplicative constant: for every n and k there exists some $A \in L_n^k$ s.t. every monotone subsequence in A has length $O_k(\sqrt{n})$.*

The next theorem is a high dimensional analogue of Theorem 1.2.

Theorem 2.3. *Let H_n^k be the length of the longest monotone subsequence in a uniformly random element of L_n^k. Then $\mathbb{E}\left[H_n^k\right] = \Theta_k\left(n^{\frac{k}{k+1}}\right)$ and $H_n^k = \Theta_k\left(n^{\frac{k}{k+1}}\right)$ a.a.s.*

Remark 2.4. Aside from *strong monotonicity* as in definition 2.1 it is interesting to consider *weak monotonicity.* A sequence of pairwise distinct $\alpha^1, \alpha^2, \ldots, \alpha^m$ in $[n]^{k+1}$ is called weakly monotone if it is weakly monotone in every coordinate. In the spirit of the Hales–Jewett Theorem one may also consider the case where every coordinate is either strictly monotone or constant.

We strive throughout to deal with the harder of the two cases, namely prove large lower bounds for strongly monotone subsequences and small upper bounds for the weakly monotone case. The one exception is that the proof of the upper bound in Theorem 2.2, applies only to the strongly monotone case. It remains an interesting open problem to determine the correct upper bound for weakly monotone subsequences.

Remark 2.5. Note the following symmetries of high-dimensional permutations:

(1) S_{k+1} acts on L_n^k by permuting the coordinates.
(2) For each $1 \leq i \leq k+1$, the group S_n acts on L_n^k by permuting the values of the i-th coordinate of each $A \in L_n^k$. Actions on different coordinates commute, and so this defines an S_n^{k+1}-action on L_n^k.
(3) A special case of (2), is reversal, i.e. applying the map $a \mapsto n+1-a$ on the i-th coordinate.

Note that actions (1) and (3) preserve monotonicity.

3. A High-Dimensional Analogue of the Erdős–Szekeres Theorem

We begin by proving Theorem 2.2. Due to the Erdős–Szekeres Theorem it suffices to consider the case $k \geq 2$.

We define two partial orders on $[n]^{k+1}$: Let $\alpha, \beta \in [n]^{k+1}$. We write $\alpha <_1 \beta$ if for all $1 \leq i \leq k+1$, $\alpha_i < \beta_i$, and we write $\alpha <_2 \beta$ if for all $1 \leq i \leq k$, $\alpha_i < \beta_i$ and $\alpha_{k+1} > \beta_{k+1}$. For $\alpha, \beta \in [n]^k$ we write $\alpha < \beta$ if for all $1 \leq i \leq k$, $\alpha_i < \beta_i$.

Recall that the *height* $h(P)$ of a poset P is the size of the largest chain in P and its *width* $w(P)$ is the size of its largest anti-chain. The next lemma is an easy consequence of Dilworth's Theorem [4] or Mirsky's Theorem [13].

Lemma 3.1. *For every finite poset P there holds $h(P) \cdot w(P) \geq |P|$.*

We use Lemma 3.1 to show that if A has no long monotone subsequences, then there is a large $S \subseteq G_A$ that is an anti-chain in both $<_1$ and $<_2$. On the other hand, the next two lemmas give an upper bound on the size of anti-chains common to $<_1$ and $<_2$. This yields the theorem.

Lemma 3.2. *Let X be an $M \times N$ matrix in which every two entries in the same column are distinct. Let S be a set of positions in X such that $X_a = X_b$ for every $a, b \in S$ with a to the left and above b. Then $|S| \leq M + 2N$.*

Proof. If either $M = 1$ or $N = 1$, this is obvious. We prove the claim inductively by showing that either S has at most two positions in the rightmost column of X or at most one element in the topmost row of X. Indeed, if S has at least three entries in the rightmost column, then at least two of them, say a and b, are not in the top row. But there are no repetitions in the same column, so $X_a \neq X_b$. It follows that the only element S may have in the top row is at the top-right corner, for any other such element must equal both X_a and X_b, which is impossible. □

We are now ready to prove Theorem 2.2.

Proof. For the lower bound, let $A \in L_n^k$ and consider the $n \times n$ matrix X defined by $X_{a,b} = f_A(a, b, b, \ldots, b)$. We define two partial orders on $[n]^2$: Let $\alpha, \beta \in [n]^2$. We write $\alpha <_1 \beta$ if $\alpha_i < \beta_i$, $i = 1, 2$ and $X_\alpha < X_\beta$. We write $\alpha <_2 \beta$ if $\alpha_i < \beta_i$, $i = 1, 2$ and $X_\alpha > X_\beta$. Clearly, a sequence $\alpha^1 <_1 \alpha^2 <_1 \ldots <_1 \alpha^m$ corresponds to a monotone subsequence in A, and similarly for $<_2$.

Assume that $[n]^2$ contains no $<_1$-monotone subsequences of length $r = \left\lfloor \frac{\sqrt{n}}{3} \right\rfloor$. By Lemma 3.1 there is an $<_1$-anti-chain $S_1 \subseteq [n]^2$ of size at least $\frac{n^2}{r}$. Order S_1 by $<_2$ and let $S \subseteq S_1$ be an anti-chain. S is an anti-chain w.r.t. both $<_1$ and $<_2$, hence if $\alpha \in S$ is above and to the left of $\beta \in S$ we have $X_\alpha = X_\beta$. Every column in X is a permutation of $[n]$, so X and S satisfy the conditions of Lemma 3.2 and therefore $|S| \leq 3n$. This is true for every anti-chain in S_1 and so $w(S_1) \leq 3n$. Applying Lemma 3.1 again we conclude: $h(S_1) \geq \frac{|S_1|}{w(S_1)} \geq \frac{n^2}{3nr} \geq r = \left\lfloor \frac{\sqrt{n}}{3} \right\rfloor$. The height of S_1 is realized by a monotone subsequence of length $h(S_1)$ in A, yielding the lower bound.

For the second part of the theorem, for every n and k we construct $A \in L_n^k$ with all monotone subsequences having length $O(\sqrt{n})$. We first assume n is prime, and use a simple construction similar to one that shows the tightness of the Erdős–Szekeres Theorem. We later modify the construction to deal with composite n. Assuming n is prime, let $M = \left\lfloor \sqrt{\frac{n}{k+1}} \right\rfloor$, and define A as follows:

$$A(\alpha_1, \alpha_2, \ldots, \alpha_{k+1}) = 1 \iff M \sum_{i=1}^{k} \alpha_i + \alpha_{k+1} = 0 \,(\mathrm{mod} n)$$

Since n is prime it follows easily that A is a k-dimensional permutation.

We'll show that if $\alpha, \beta \in G_A$ differ in every coordinate then $\|\alpha - \beta\|_1 \geq M$. This is sufficient, since if $\alpha^1, \alpha^2, \ldots, \alpha^m \in G_A$ is a monotone subsequence, then for every $1 \leq j < m$, α^j, α^{j+1} differ on every coordinate and so $M(m-1) \leq \sum_{j=1}^{m-1} \|\alpha^{j+1} - \alpha^j\|_1$. On the other hand, by monotonicity we have $\sum_{j=1}^{m-1} \|\alpha^{j+1} - \alpha^j\|_1 = \|\alpha^m - \alpha^1\|_1 \leq (k+1)n$. It follows that $m \leq \sqrt{(k+1)n} + 1 = O(\sqrt{n})$.

Assume $\alpha, \beta \in G_A$ differ in every coordinate. We have:

$$M \sum_{i=1}^{k} (\alpha_i - \beta_i) + (\alpha_{k+1} - \beta_{k+1}) = 0 \,(\mathrm{mod} n)$$

Now $Mx + y = 0 \,(\mathrm{mod} n)$ implies either $|y| \geq M, |x| \geq \frac{n}{M} - 1 \geq M$ or $x = y = 0$. Setting $x = \sum_{i=1}^{k} (\alpha_i - \beta_i)$ and $y = (\alpha_{k+1} - \beta_{k+1})$, we have by assumption $y \neq 0$ and so $\|\alpha - \beta\|_1 \geq |x| + |y| \geq M$.

In this construction we need M and n to be relatively prime. For composite n this isn't necessarily the case, and we offer two remedies: The first is an appeal to number theory to produce $M \approx \sqrt{\frac{n}{k+1}}$ coprime to n. It is known [2] that for large x, there is always a prime in the interval $[x - x^{0.525}, x]$. Therefore, we can find three distinct primes in an interval $\left[\sqrt{\frac{n}{k+1}}, (1+o(1))\sqrt{\frac{n}{k+1}}\right]$. At least one of these must be coprime to n, since their product exceeds n for large n. This implies that all monotone subsequences have length $\leq (2+o(1))\sqrt{(k+1)n}$.

The second approach is easy to generalize, as done in the proof of Theorem 3.5. Take $M = \left\lfloor \sqrt{\frac{n}{k+1}} \right\rfloor$ as before. Let $g = \gcd(M,n)$ and define the permutation $\pi \in S_n$ as follows (all values are taken modulo n):

$$\pi = \left(M, 2M, \ldots, \frac{n}{g}M, 1+M, \ldots, 1+\frac{n}{g}M, \ldots, g-1+M, \ldots, g-1+\frac{n}{g}M\right)$$

Set $f_A(\alpha_1, \alpha_2, \ldots, \alpha_k) = -\pi\left(\sum_{i=1}^{k} \alpha_i\right)$. Note that if $\gcd(M,n) = 1$, this coincides with the construction above. As before, we show that if $\alpha, \beta \in G_A$ differ on all coordinates then $\|\alpha - \beta\|_1 \geq M$, which is enough.

Assume $\alpha, \beta \in G_A$ differ on all coordinates. We then have:

$$M\sum_{i=1}^{k}(\alpha_i - \beta_i) + (\alpha_{k+1} - \beta_{k+1}) = r \,(\mathrm{mod} n)$$

for some $|r| < g \leq M$. If $r = 0$ we have the same situation as before and we may conclude $\|\alpha - \beta\|_1 \geq M$. Otherwise, by definition of π, we must have either $\|\alpha - \beta\|_1 \geq \left|\sum_{i=1}^{k}(\alpha_i - \beta_i)\right| \geq \frac{n}{g} - 1 \geq \frac{n}{M} - 1 \geq M$ or else $|\alpha_{k+1} - \beta_{k+1}| \geq M$. □

Most proofs of Theorem 1.1 actually yield the following, more general, statement.

Theorem 3.3. *Let r, s and n be positive integers with $rs < n$. Then every permutation in S_n contains either an increasing subsequence of length $r+1$, or a decreasing subsequence of length $s+1$. The bound is tight: if $rs \geq n$ then there is a permutation in S_n with neither an increasing subsequence of length $r+1$ nor a decreasing subsequence of length $s+1$.*

It is possible to extend Theorem 2.2 in a similar fashion. To this end we refine our notion of monotonicity. In dimension one we distinguish between ascending and descending subsequences, and we need something similar in higher dimensions.

Definition 3.4. A vector $\vec{c} \in \{0,1\}^{k+1}$ induces a partial order $x <_{\vec{c}} y$ on $\mathbb{R}^{k+1}$ as follows: $x <_{\vec{c}} y$ if for every $1 \leq i \leq k+1$ s.t. $c_i = 1$, $x_i < y_i$, and $y_i < x_i$ otherwise.

Theorem 3.5. *Let $\vec{c}, \vec{d} \in \{0,1\}^{k+1}$ differ in exactly one coordinate. Let $rs < \frac{n}{3(k-1)}$. Then every $A \in L_n^k$, contains either a $<_{\vec{c}}$-monotone subsequence of length r or a $<_{\vec{d}}$-monotone subsequence of length s.*

The bound is tight up to the multiplicative constants: If $r, s \geq 9(k+10)$ and $rs > 5kn$, then there exists $A \in L_n^k$ with neither a $<_{\vec{c}}$-monotone subsequence of length r nor a $<_{\vec{d}}$-monotone subsequence of length s.

Proof. Using the symmetries from Remark 2.5 we may assume w.l.o.g. that $\vec{c} = (1,1,\ldots,1)$ and $\vec{d} = (1,1,\ldots,1,0)$.

The proof of the lower bound is similar to the proof of the lower bound in Theorem 2.2, and we provide only a sketch. As in the proof of Theorem 2.2, consider the matrix X and the partial orders $<_1, <_2$. Lemma 3.2 gives an upper bound of $3n$ on the size of any anti-chain under both $<_1$ and $<_2$. Two applications of Lemma 3.1 yield the lower bound.

For the upper bound, assume w.l.o.g. that $r \geq s$. We construct $\pi \in S_n$ and $A \in L_n^k$ as before, with $M = \left\lfloor \frac{s}{2k} \right\rfloor$. Let $\alpha^1, \alpha^2, \ldots, \alpha^m \in G_A$ be a $<_{\vec{c}}$-monotone subsequence. Then the sequence is increasing in every coordinate. For all j, if $\alpha_{k+1}^{j+1} - \alpha_k^j < M$ then $\sum_{i=1}^k \left(\alpha_i^{j+1} - \alpha_i^j\right) \geq \frac{n}{g} \geq \frac{n}{M}$. Thus

$$m \leq \frac{n}{M} + \frac{kn}{\frac{n}{M}} + 1 = \frac{n}{M} + kM + 1 \leq \frac{2kn}{s}\left(1 + \frac{2k}{s}\right) + \frac{s}{2} + 1$$

Using the assumptions that $\frac{r}{5k} > \frac{n}{s}$ and $r \geq s \geq 9\,(k+10)$, we have:

$$m \leq r\left(\frac{2}{5}\left(1 + \frac{2}{9}\right) + \frac{1}{2} + \frac{1}{r}\right) \leq r$$

Now, let $\alpha^1, \alpha^2, \ldots, \alpha^m \in G_A$ be a $<_{\vec{d}}$-monotone subsequence. For $1 \leq j \leq m$ define $s_j = M\sum_{i=1}^k \alpha_i^j$. This is an increasing sequence, and $s_{j+1} - s_j \geq M$ for all j. By definition of A, $\alpha_{k+1}^j = s_j\,(\mathrm{mod} n) + r_j$ for some $0 \leq r_j < M$. Because $\alpha_{k+1}^1, \alpha_{k+1}^2, \ldots, \alpha_{k+1}^m$ is decreasing, if for some j, s_j and s_{j+1} fall in the same interval of the form $[dn+1, (d+1)\,n]$ (for $d \in \mathbb{Z}$), then $s_j + r_j > s_{j+1} \implies s_{j+1} - s_j < r_j < M$, a contradiction. Therefore the s_j's fall into distinct intervals of the form $[dn+1, (d+1)\,n]$. But for every j, $0 < s_j \leq Mkn$. Since $[0, Mkn]$ contains only $\left\lceil \frac{Mkn}{n} \right\rceil \leq Mk + 1$ intervals of length n, we have $m \leq Mk + 1 \leq \frac{s}{2} + 1 < s$. □

4. Monotone Subsequences in Random High-Dimensional Permutations

As mentioned in the introduction, the longest monotone subsequence of a random permutation is typically of length $2\sqrt{n}$. In view of the Erdős–Szekres Theorem this means that the random case and the worst case are of the same order of magnitude and differ by only a constant factor. In higher dimensions this is no longer the case. The longest monotone subsequence of a typical element in L_n^k has length $\Theta_k\left(n^{\frac{k}{k+1}}\right)$.

We define the random variable H_n^k - the length of the longest monotone subsequence in a uniformly random element of L_n^k, and prove the next theorem.

Theorem 4.1. *For every $k \in \mathbb{N}$:*

(1) For every $\varepsilon > 0$, $H_n^k n^{-\frac{k}{k+1}} \in \left[\frac{1}{k+1}, e + \varepsilon\right]$ asymptotically almost surely.

(2) $1 - \frac{\ln k+1}{k+1} - o_k(1) \leq \mathbb{E}\left[H_n^k n^{-\frac{k}{k+1}}\right] \leq e + o_k(1)$.

There are 2^{k+1} distinct order types of monotone subsequences, indexed by binary vectors $\vec{c} \in \{0,1\}^{k+1}$. By reversing some of the coordinates (operation 3 in Remark 2.5) we see that the distribution of the longest $<_{\vec{c}}$-monotone subsequence in a random element of L_n^k is independent of $\vec{c}$. Thus it suffices to prove Theorem 4.1 for $<_{(1,1,\ldots,1)}$-monotone subsequences. For brevity of notation we write $<$ in place of $<_{(1,1,\ldots,1)}$.

The following lemmas are useful in dealing with uniformly random elements of L_n^k.

Lemma 4.2. *Given $A \in L_n^k$ and $\pi = (\pi_1, \pi_2, \ldots, \pi_{k+1}) \in S_n^{k+1}$, let $\pi(A) \in L_n^k$ be the k-dimensional permutation given by*

$$\pi(A)(x_1, x_2, \ldots, x_{k+1}) = A(\pi_1(x_1), \pi_2(x_2), \ldots, \pi_{k+1}(x_{k+1}))$$

(equivalently, $\pi(A)$ is obtained by permuting the ith coordinate of G_A according to π_i^{-1}). If A is chosen uniformly at random from L_n^k and π is independently chosen from any *distribution on S_n^{k+1}, then $\pi(A)$ is uniformly distributed in L_n^k.*

Proof. This follows immediately from the fact that S_n^{k+1} acts on L_n^k in the way described. □

Lemma 4.3. *Let $\alpha^1, \alpha^2, \ldots, \alpha^m \in [n]^{k+1}$ be a weakly monotone sequence of positions. For a uniformly drawn $A \in L_n^k$,*

$$\mathbb{P}\left[A\left(\alpha^1\right) = A\left(\alpha^2\right) = \ldots = A\left(\alpha^m\right) = 1\right] \leq \frac{(n-m)!}{n!}$$

Proof. Assume w.l.o.g. that the sequence is weakly monotone according to $<$.

We define a distribution $\mathcal{D}$ on S_n^{k+1} s.t. if $\pi \sim \mathcal{D}$ and A is drawn independently and uniformly from L_n^k, then $\mathbb{P}\left[\pi(A)\left(\alpha^1\right) = \pi(A)\left(\alpha^2\right) = \ldots = \pi(A)\left(\alpha^m\right) = 1\right] \leq \frac{(n-m)!}{n!}$. The conclusion follows from Lemma 4.2.

In order to define $\mathcal{D}$ we construct distributions $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_m$ on S_n^{k+1}, and we let $\pi = \pi_m \pi_{m-1} \cdot \ldots \cdot \pi_1$ where for each i, π_i is drawn independently from $\mathcal{D}_i$. We then define $\pi(A)$ via

$$A \to A_1 = \pi_1(A) \to A_2 = \pi_2(A_1) \to \ldots \to A_m = \pi_m(A_{m-1}) = \pi(A)$$

We'll define the distributions $\mathcal{D}_i$ s.t. the following properties hold:

- For all $1 \leq i < j \leq m$, $A_j\left(\alpha^i\right) = A_i\left(\alpha^i\right)$, so the value at position α^i remains fixed from stage i onward.
- For $1 \leq i \leq m$, $\mathbb{P}\left[A_i\left(\alpha^1\right) = A_i\left(\alpha^2\right) = \ldots = A_i\left(\alpha^i\right) = 1\right] \leq \frac{(n-i)!}{n!}$.

Let $\mathcal{D}_1$ be uniformly distributed on $S_n \times \{I\}^k$, where $I \in S_n$ is the identity element. There is a unique x s.t. $A\left(x, \alpha_2^1, \ldots, \alpha_{k+1}^1\right) = 1$, and therefore $\mathbb{P}\left[A_1\left(\alpha^1\right) = 1\right] = \mathbb{P}\left[A\left(\pi_1\left(\alpha_1^1\right), \alpha_2^1, \ldots, \alpha_{k+1}^1\right) = 1\right] = \mathbb{P}\left[\pi_1\left(\alpha_1^1\right) = x\right] = \frac{1}{n}$.

Now suppose that $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_i$ are already defined and have the properties above. The sequence $\alpha^1, \alpha^2, \ldots, \alpha^m$ is weakly increasing so there exists some coordinate $1 \leq j \leq k+1$ s.t. $\alpha_j^i < \alpha_j^{i+1}$. Let $T \subseteq S_n$ be the set of permutations that fix $\left\{\alpha_j^1, \alpha_j^2, \ldots, \alpha_j^i\right\}$, and let $\mathcal{D}_{i+1}$ be the uniform distribution on $\{I\}^{j-1} \times T \times \{I\}^{k+1-j}$. We write $\pi_{i+1} = (I, \ldots, I, \tau, I, \ldots, I)$ and verify the properties above:

- For $1 \leq \ell \leq i$, by definition $A_{i+1}\left(\alpha^\ell\right) = A_i\left(\alpha_1^\ell, \ldots, \alpha_{j-1}^\ell, \tau\left(\alpha_j^\ell\right), \alpha_{j+1}^\ell, \ldots, \alpha_{k+1}^\ell\right)$. But τ fixes α_j^ℓ, so $A_{i+1}\left(\alpha^\ell\right) = A_i\left(\alpha^\ell\right) = A_\ell\left(\alpha^\ell\right)$ where the last equality follows by induction.
- We have:

$$\mathbb{P}\left[A_i\left(\alpha^1\right) = A_i\left(\alpha^2\right) = \ldots = A_{i+1}\left(\alpha^{i+1}\right) = 1\right]$$

$$= \mathbb{P}\left[A_{i+1}\left(\alpha^{i+1}\right) = 1 | A_{i+1}\left(\alpha^1\right) = A_{i+1}\left(\alpha^2\right) = \ldots = A_{i+1}\left(\alpha^i\right) = 1\right] \times$$

$$\times \mathbb{P}\left[A_{i+1}\left(\alpha^1\right) = A_{i+1}\left(\alpha^2\right) = \ldots = A_{i+1}\left(\alpha^i\right) = 1\right]$$

By the inductive assumption:

$$\mathbb{P}\left[A_{i+1}\left(\alpha^1\right) = A_{i+1}\left(\alpha^2\right) = \ldots = A_{i+1}\left(\alpha^i\right) = 1\right]$$

$$= \mathbb{P}\left[A_i\left(\alpha^1\right) = A_i\left(\alpha^2\right) = \ldots = A_i\left(\alpha^i\right) = 1\right] \leq \frac{(n-i)!}{n!}$$

Now, $\alpha_{i+1}^j \notin \left\{\alpha_1^j, \alpha_2^j, \ldots, \alpha_i^j\right\}$, so that $\tau\left(\alpha_{i+1}^j\right)$ is distributed uniformly on a set of cardinality $\geq n-i$, and is independent of $A_{i+1}\left(\alpha^1\right), A_{i+1}\left(\alpha^2\right), \ldots, A_{i+1}\left(\alpha^i\right)$. Thus:

$$\mathbb{P}\left[A_{i+1}\left(\alpha^{i+1}\right) = 1 | A_{i+1}\left(\alpha^1\right) = A_{i+1}\left(\alpha^2\right) = \ldots = A_{i+1}\left(\alpha^i\right) = 1\right] \leq \frac{1}{n-i}$$

We conclude:

$$\mathbb{P}\left[A_i\left(\alpha^1\right) = A_i\left(\alpha^2\right) = \ldots = A_{i+1}\left(\alpha^{i+1}\right) = 1\right] \leq \frac{1}{n-i}\frac{(n-i)!}{n!} = \frac{(n-(i+1))!}{n!}$$

as desired.

□

We first prove the upper bounds in Theorem 4.1.

Proposition 4.4.

(1) For every $\varepsilon > 0$ there holds $\mathbb{P}\left[H_n^k n^{-\frac{k}{k+1}} > e + \varepsilon\right] = o(1)$.

(2) $\mathbb{E}\left[H_n^k\right] n^{-\frac{k}{k+1}} \leq e + o(1)$.

Proof. We bound the expected number of length-m (weakly) monotone subsequences in a random k-dimensional permutation. For every increasing sequence of positions $\alpha = \alpha^1, \alpha^2, \ldots, \alpha^m \in [n]^{k+1}$ and $A \in L_n^k$ we define

$$X_\alpha(A) = \begin{cases} 1 & A\left(\alpha^1\right) = A\left(\alpha^2\right) = \ldots = A\left(\alpha^m\right) \\ 0 & otherwise \end{cases}$$

By Lemma 4.3 $\mathbb{E}\left[X_\alpha(A)\right] = \mathbb{P}\left[X_\alpha(A) = 1\right] \leq \frac{(n-m)!}{n!}$ for a uniform $A \in L_n^k$. Let S be the set of all length-m increasing sequences of positions in $[n]^k$. Clearly, $|S| \leq \left(\begin{array}{c} n+m-1 \\ m \end{array}\right)^{k+1}$ so by linearity of expectation:

$$\mathbb{P}\left[H_n^k \geq m\right] = \mathbb{P}\left[\sum_{\alpha \in S} X_\alpha(A) > 0\right] \leq \mathbb{E}\left[\sum_{\alpha \in S} X_\alpha(A)\right]$$

$$\leq \left(\begin{array}{c} n+m-1 \\ m \end{array}\right)^{k+1} \frac{(n-m)!}{n!} \leq \left(\frac{e(n+m)}{m}\right)^{(k+1)m} \frac{1}{(n-m)^m}$$

Let $c = e + \varepsilon$ for some $\varepsilon > 0$, and let $m = \left\lceil cn^{\frac{k}{k+1}}\right\rceil$. Then:

$$\mathbb{P}\left[H_n^k n^{-\frac{k}{k+1}} > c\right] = \mathbb{P}\left[H_n^k \geq m\right] \leq \left((1+o(1))\, e^{k+1}\frac{n^k}{m^{k+1}}\right)^m$$

$$\leq \left((1+o(1))\frac{e}{c}\right)^{(k+1)cn^{\frac{k}{k+1}}} = o(1)$$

proving the first claim in the proposition. Further:

$$\mathbb{E}\left[H_n^k\right] n^{-\frac{k}{k+1}} \leq \left(m\mathbb{P}\left[H_n^k < m\right] + n\mathbb{P}\left[H_n^k \geq m\right]\right) n^{-\frac{k}{k+1}}$$

$$\leq c + n^{\frac{1}{k+1}}\left(\frac{e}{c}\right)^{(k+1)cn^{\frac{k}{k+1}}} + o(1) = c + o(1)$$

which proves the second claim. □

The proof of the lower bounds is more intricate. Fix some $C > 0$ and let $m = \left\lceil Cn^{\frac{1}{k+1}} \right\rceil$. For $1 \leq i \leq \left\lfloor \frac{n}{m} \right\rfloor$, let $D_i = \left[(i-1)m+1, im\right]^{k+1}$ be the diagonal subcubes of $[n]^{k+1}$. For a uniformly random $A \in L_n^k$ let Z_i be the indicator random variable of the event that A is not all zero on D_i. Clearly, $H_n^k \geq \sum_{1\leq i\leq \frac{n}{m}} Z_i$, since $\alpha < \beta$ if $\alpha \in D_i, \beta \in D_j$, and $i < j$. Indeed we prove lower bounds on H_n^k by bounding $\sum_{1\leq i\leq \frac{n}{m}} Z_i$. It is convenient to express everything in terms of the random variable $Y_n = n^{-\frac{k}{k+1}} \sum_{1\leq i\leq \frac{n}{m}} Z_i$. We show that for an appropriate choice of C (see below) Y_n converges in probability to a constant in $(0, 1)$. These are our main steps:

(1) Note that $Y_n \leq \frac{1}{C} + o(1)$ (trivially).
(2) Prove that $\mathbb{E}[Y_n] \geq \frac{C^k}{C^{k+1}+1} - o(1)$ (Proposition 4.6).
(3) Show that if $C < 1$, then $\mathbb{P}\left[Y_n > C^{k+1} + \varepsilon\right] = o(1)$ for every $\varepsilon > 0$ (Corollary 4.9).
(4) By letting $1 > C > 0$ be the unique solution to $\frac{C^k}{1+C^{k+1}} = C^{k+1}$, conclude that $\mathbb{P}\left[Y_n < C^{k+1} - \varepsilon\right] = o(1)$ for every $\varepsilon > 0$ (Proposition 4.10). Hence $\lim_{n\to\infty} Y_n = C^{k+1}$ in probability.

In step 1 we assume only that $C > 0$. The claim in step 2 applies to all $C > 0$, and we optimize the bound on $\mathbb{E}[Y_n]$ by a particular choice of C. Step 3 applies to all $1 > C > 0$. Finally in step 4 we assign a value to C to derive the conclusion that Y_n converges in probability to C^{k+1}.

We start with step 2, a lower bound on $\mathbb{E}[Y_n]$:

Lemma 4.5. *For* $1 \leq i \leq \frac{n}{m}$, $\mathbb{P}[Z_i = 1] \geq \frac{C^{k+1}}{C^{k+1}+1} - o(1)$.

Proof. Let $X_i = \sum_{\alpha\in D_i} A(\alpha)$ be the number of non-zero entries in D_i. Note that $X_i > 0 \iff Z_i = 1$. We prove a lower bound on the probability of this event by a second moment argument.

Clearly, $\mathbb{E}[X_i] = \frac{|D_i|}{n} = C^{k+1} + o(1)$, since $\mathbb{P}[A(\alpha) = 1] = \frac{1}{n}$ for every $\alpha \in [n]^{k+1}$. We next seek an upper bound on $\mathbb{E}[X_i^2]$.

$$\mathbb{E}\left[X_i^2\right] = \sum_{\alpha,\beta\in D_i} \mathbb{E}[A(\alpha)A(\beta)] = \sum_{\alpha,\beta\in D_i} \mathbb{P}[A(\alpha)A(\beta) = 1]$$

There are m^{k+1} terms with $\alpha = \beta$, each being $\frac{1}{n}$. For $\alpha \neq \beta$, Lemma 4.3 gives $\mathbb{P}[A(\alpha)A(\beta) = 1] \leq \frac{1}{n(n-1)}$. There are fewer than $m^{2(k+1)}$ such pairs $\alpha, \beta \in D_i$, so

$$\mathbb{E}\left[X_i^2\right] = \sum_{\alpha,\beta\in D_i} \mathbb{P}[A(\alpha)A(\beta) = 1] \leq m^{k+1}\left(\frac{1}{n} + \frac{m^{k+1}}{n(n-1)}\right) = \frac{m^{k+1}}{n}\left(1 + \frac{m^{k+1}}{n-1}\right)$$

Noting that $\mathbb{E}[X_i] = \frac{m^{k+1}}{n} = C^{k+1} + o(1)$, we have:

$$\mathbb{E}\left[X_i^2\right] \leq \mathbb{E}[X_i]\left(1 + \frac{n}{n-1}\mathbb{E}[X_i]\right) = C^{k+1}\left(1 + \frac{n}{n-1}C^{k+1}\right) + o(1)$$

The second moment method yields:

$$\mathbb{P}[Z_i = 1] = \mathbb{P}[X_i > 0] \geq \frac{\mathbb{E}[X_i]^2}{\mathbb{E}\left[X_i^2\right]} = \frac{C^{k+1}}{C^{k+1} + 1 + o(1)} \geq \frac{C^{k+1}}{C^{k+1}+1} - o(1)$$

□

Proposition 4.6.
$\mathbb{E}\left[Y_n\right] \geq \frac{C^k}{C^{k+1}+1} - o\left(1\right)$, *consequently* $\mathbb{E}\left[n^{-\frac{k}{k+1}} H_n^k\right] \geq 1 - \frac{\ln k+1}{k+1} - o(1)$.

Proof. As observed earlier:

$$\mathbb{E}\left[Y_n\right] = \mathbb{E}\left[n^{-\frac{k}{k+1}} \sum_{1 \leq i \leq \frac{n}{m}} Z_i\right] = n^{-\frac{k}{k+1}} \left\lfloor \frac{n}{m} \right\rfloor \mathbb{P}\left[Z_i = 1\right]$$

So, by Lemma 4.5:

$$\mathbb{E}\left[Y_n\right] \geq \frac{C^k}{C^{k+1}+1} - o\left(1\right)$$

For all C, $\mathbb{E}\left[n^{-\frac{k}{k+1}} H_n^k\right] \geq \mathbb{E}\left[Y_n\right]$. The optimal bound is attained when $C = k^{\frac{1}{k+1}}$, yielding:

$$\mathbb{E}\left[n^{-\frac{k}{k+1}} H_n^k\right] \geq \frac{k^{\frac{k}{k+1}}}{k+1} - o\left(1\right) \geq 1 - \frac{\ln k + 1}{k+1} - o\left(1\right)$$

□

To prove the lower bound in Theorem 4.1 part (1), we apply a Chernoff bound to the events $\{Z_i = 1\}_{1 \leq i \leq \frac{n}{m}}$. To overcome the dependencies among these events we utilize the following version of the Chernoff inequality from [7] (Theorem 1.1).

Theorem 4.7. *Let* $0 \leq \alpha \leq \beta \leq 1$ *and let* $\{X_i\}_{i \in [N]}$ *be Boolean random variables such that for all* $S \subseteq [N]$, $\mathbb{P}\left[\prod_{i \in X} X_i = 1\right] \leq \alpha^{|S|}$. *Then* $\mathbb{P}\left[\sum_{i \in [N]} X_i \geq \beta N\right] \leq e^{-ND(\beta \| \alpha)}$, *where* $D\left(\beta \parallel \alpha\right) = \beta \ln\left(\frac{\beta}{\alpha}\right) + \left(1-\beta\right) \ln\left(\frac{1-\beta}{1-\alpha}\right)$ *is the relative entropy function.*

Lemma 4.8. *Assume* $C < 1$. *Let* $S \subseteq \left\{1, 2, \ldots, \left\lfloor \frac{n}{m} \right\rfloor\right\}$. *Then* $\mathbb{P}\left[\prod_{i \in S} Z_i = 1\right] \leq \alpha^{|S|}$ *for all* $C^{k+1} < \alpha < 1$ *and large enough* n.

Proof. Note that $Z_i = 1$ for all $i \in S$ iff there exist positions $\{\beta^i\}_{i \in S}$ s.t. $\beta^i \in D_i$ for all $i \in S$ and $A_{\beta^i} = 1$ for all i. We bound the probability of this occurrence using a union bound.

Let $\{\beta^i\}_{i \in S}$ be positions s.t. $\beta^i \in D_i$ for all $i \in S$. If the indices in S are taken in order this is a monotone subsequence, and so by Lemma 4.3 $\mathbb{P}\left[\wedge_{i \in S} A\left(\beta^i\right) = 1\right] \leq \frac{(n-|S|)!}{n!}$. There are $m^{(k+1)|S|}$ such coordinate sequences, and so, by a union bound:

$$\mathbb{P}\left[\prod_{i \in S} Z_i = 1\right] \leq m^{(k+1)|S|} \frac{(n-|S|)!}{n!} \leq \left(\frac{m^{k+1}}{n-|S|}\right)^{|S|}$$

We have: $|S| \leq \frac{n}{m} = \frac{1}{C} n^{\frac{k}{k+1}} + o\left(1\right)$. Thus:

$$\mathbb{P}\left[\prod_{i \in S} Z_i = 1\right] \leq \left(\left(1 + o\left(1\right)\right) \frac{C^{k+1} n}{n - \frac{1}{C} n^{\frac{k}{k+1}}}\right)^{|S|} = \left(\left(1 + o\left(1\right)\right) C^{k+1}\right)^{|S|}$$

and the result follows. □

Lemma 4.8 allows us to apply Theorem 4.7 to the variables $\{Z_i\}_{1 \leq i \leq \frac{n}{m}}$ to obtain the next corollary.

Corollary 4.9. *For all $\beta > C^{k+1}$, for large enough n it holds:*

$$\mathbb{P}\left[Y_n > \beta\right] \leq \exp\left(-n^{\frac{k}{k+1}}\gamma\right)$$

for some $\gamma > 0$.

We are now ready to complete the proof of Theorem 4.1.

Proposition 4.10. *Let $1 > C > 0$ be the unique solution to the equation $C\left(1 + C^{k+1}\right) = 1$. Then $\mathbb{P}\left[Y_n < \frac{1}{k+2}\right] = o\left(1\right)$.*

Proof. By Proposition 4.6

$$\mathbb{E}\left[Y_n\right] \geq \frac{C^k}{C^{k+1}+1} - o\left(1\right) = C^{k+1} - o\left(1\right) \tag{1}$$

For an integer n and $0 < x < C^{k+1}$, let $p_n = \mathbb{P}\left[Y_n \leq x\right]$. Since $Y_n \leq \frac{1}{C} + o(1)$ for every $\varepsilon > 0$:

$$\mathbb{E}\left[Y_n\right] \leq p_n x + \left(1 - p_n\right)\left(C^{k+1} + \varepsilon\right) + \left(\frac{1}{C} + o\left(1\right)\right)\mathbb{P}\left[Y_n \geq C^{k+1} + \varepsilon\right] \tag{2}$$

Corollary 4.9 yields:

$$\mathbb{P}\left[Y_n \geq C^{k+1} + \varepsilon\right] = o\left(1\right)$$

Combining inequalities 1 and 2 and rearranging:

$$p_n\left(C^{k+1} - x\right) \leq \varepsilon\left(1 - p_n\right) + o\left(1\right)$$

But this holds for all $\varepsilon > 0$, so that $\lim_{n\to\infty} p_n = 0$.

The result follows by taking $x = \frac{1}{k+1} < C^{k+1}$. □

5. Concluding Remarks and Open Problems

- As mentioned in section 2, we do not know what the analogous statement of Theorem 2.2 is for weakly monotone subsequences.
- What are the best constant factors in Theorems 2.2 and 3.5? For the sake of clarity we have neglected to optimize the constants, and our bounds can certainly be somewhat improved with some additional effort. However, we suspect that getting the correct bounds would require some new ideas. While we find the correct exponent of n in the problems addressed here, we are still unable to determine the dependency of the relevant coefficients on the dimension k. Perhaps the most pressing question of this sort is to derive a sharp result on the existence of long monotone subsequences in Latin squares.
- For $A \in L_n^k$ and $\vec{c} \in \{0,1\}^{k+1}$, let $\ell_{\vec{c}}\left(A\right)$ be the length of the longest $<_{\vec{c}}$-monotone subsequence in A. Let $\ell\left(A\right) = \left(\ell_{\vec{c}}\left(A\right)\right)_{\vec{c}\in\{0,1\}^{k+1}}$. We seek a better description of the set $\ell_n^k = \left\{\ell\left(A\right) : A \in L_n^k\right\}$. By Theorem 2.2 we know that $\min_{x\in\ell_n^k}\left\|x\right\|_\infty = \Theta\left(\sqrt{n}\right)$. Theorem 3.5 gives fairly tight sufficient conditions under which we can conclude that $x_{\vec{c}} \geq r \vee x_{\vec{d}} \geq s$ for $\vec{c}, \vec{d} \in \{0,1\}^{k+1}$ that differ in precisely one coordinate.

- The proof of Theorem 4.1 uses only a very limited amount of randomness. Recall that L_n^k splits into *isotopy classes* where permutations are reachable from each other by applications of symmetries (2) in Remark 2.5. That theorem applies even when the high-dimensional permutation is drawn uniformly from a particular isotopy class, rather than from all of L_n^k. Beyond the randomness inherent in these symmetries, we have little insight concerning the structure of random high-dimensional permutations. In our view, it's a major challenge in this field to understand (fully) random high-dimensional permutations. In particular, we do not know how to uniformly sample elements of L_n^k. Even for Latin squares, the best known method is Jacobson and Matthews' Markov chain [8], which is not known to be rapidly mixing.

- We believe Theorem 4.1 can be strengthened, and there exist constants c_k s.t. $H_n^k n^{-\frac{k}{k+1}} \to c_k$ in probability. This is borne out by numerical experiments, which indicate that $H_n^2 n^{-\frac{2}{3}}$ is concentrated in a small interval. We do not know how to prove this, but perhaps an approach based on super-additive ergodic theorems à la Hammersley [6] may apply. If these constants c_k do, in fact, exist, their dependence on k is of interest. We note that analogous results for random points in $[0,1]^k$ are known [3].

Acknowledgments. The authors wish to thank the anonymous referee for her thorough review and insightful comments.

References

[1] Jinho Baik, Percy Deift, and Kurt Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations*, Journal of the American Mathematical Society **12** (1999), no. 4, 1119–1178.

[2] Roger C Baker, Glyn Harman, and János Pintz, *The difference between consecutive primes, II*, Proceedings of the London Mathematical Society **83** (2001), no. 03, 532–562.

[3] Béla Bollobás and Peter Winkler, *The longest chain among random points in Euclidean space*, Proceedings of the American Mathematical Society **103** (1988), no. 2, 347–353.

[4] Robert P Dilworth, *A decomposition theorem for partially ordered sets*, Annals of Mathematics (1950), 161–166.

[5] Paul Erdős and George Szekeres, *A combinatorial problem in geometry*, Compositio Mathematica **2** (1935), 463–470.

[6] JM Hammersley, *A few seedlings of research*, Proc. of the Sixth Berkeley Symp. Math. Statist. and Probability, vol. 1, University of California Press, 1972, pp. 345–394.

[7] Russell Impagliazzo and Valentine Kabanets, *Constructive proofs of concentration bounds*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, Springer, 2010, pp. 617–631.

[8] Mark T Jacobson and Peter Matthews, *Generating uniformly distributed random Latin squares*, Journal of Combinatorial Designs **4** (1996), no. 6, 405–437.

[9] Joseph B Kruskal, *Monotonic subsequences*, Proceedings of the American Mathematical Society **4** (1953), no. 2, 264–274.

[10] Nathan Linial and Zur Luria, *On the vertices of the d-dimensional Birkhoff polytope*, Discrete & Computational Geometry **51** (2014), no. 1, 161–170.

[11] ———, *An upper bound on the number of high-dimensional permutations*, Combinatorica **34** (2014), no. 4, 471–486.

[12] Benjamin F Logan and Larry A Shepp, *A variational problem for random Young tableaux*, Advances in mathematics **26** (1977), no. 2, 206–222.

[13] Leon Mirsky, *A dual of Dilworth's decomposition theorem*, American Mathematical Monthly (1971), 876–877.

[14] J Michael Steele, *Variations on the monotone subsequence theme of Erdős and Szekeres*, Discrete probability and algorithms, Springer, 1995, pp. 111–131.

[15] Tibor Szabó and Gábor Tardos, *A multidimensional generalization of the Erdős–Szekeres lemma on monotone subsequences*, Combinatorics, Probability and Computing **10** (2001), no. 06, 557–565.

[16] Stanislaw M Ulam, *Monte Carlo calculations in problems of mathematical physics*, Modern Mathematics for the Engineers (1961), 261–281.

[17] Anatoly M Vershik and Sergei V Kerov, *Asymptotics of Plancherel measure of symmetrical group and limit form of Young tables*, Doklady Akademii Nauk SSSR **233** (1977), no. 6, 1024–1027.

School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
E-mail address: nati@cs.huji.ac.il

Institute of Mathematics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
E-mail address: menahem.simkin@mail.huji.ac.il

Chapter 3

On the threshold problem for Latin boxes

Zur Luria and Michael Simkin. *Random Structures & Algorithms* 55 (2019), no. 4, 926–949.

ON THE THRESHOLD PROBLEM FOR LATIN BOXES

ZUR LURIA AND MICHAEL SIMKIN

Abstract. Let $m \leq n \leq k$. An $m \times n \times k$ 0-1 array is a *Latin box* if it contains exactly mn ones, and has at most one 1 in each line. As a special case, Latin boxes in which $m = n = k$ are equivalent to Latin squares.

Let $\mathcal{M}(m, n, k; p)$ be the distribution on $m \times n \times k$ 0-1 arrays where each entry is 1 with probability p, independently of the other entries. The threshold question for Latin squares asks when $\mathcal{M}(n, n, n; p)$ contains a Latin square with high probability. More generally, when does $\mathcal{M}(m, n, k; p)$ support a Latin box with high probability?

Let $\varepsilon > 0$. We give an asymptotically tight answer to this question in the special cases where $n = k$ and $m \leq (1 - \varepsilon) n$, and where $n = m$ and $k \geq (1 + \varepsilon) n$. In both cases, the threshold probability is $\Theta(\log(n)/n)$. This implies threshold results for Latin rectangles and proper edge-colorings of $K_{n,n}$.

1. Introduction

An order-n *Latin square* is equivalent to an $n \times n \times n$ 0-1 array with a single 1 in each line, where a line is the set of elements obtained by fixing the values of two indices and letting the third vary over $[n] := \{1, ..., n\}$. With this in mind, the following definition is natural.

Definition 1.1. Let $m \leq n \leq k$. An $m \times n \times k$ 0-1 array is a *Latin box* if it contains exactly mn ones, and at most one 1 in each line.

An $m \times n \times k$ Latin box is equivalent to a 3-uniform tripartite hypergraph on $m + n + k$ vertices such that each pair of vertices is contained in at most one edge, and the number of edges is maximal subject to this constraint. Thus, Latin boxes can be viewed as a 3-uniform version of matchings of size m in unbalanced bipartite graphs on $m + n$ vertices.

As additional motivation, consider the two following special cases. An $n \times n \times k$ Latin box A is equivalent to a proper edge-coloring of the complete bipartite graph $K_{n,n}$ using k colors. One obtains such a coloring from A by coloring the edge $\{i, j\}$ with the unique index c such that $A(i, j, c) = 1$. The Latin box property ensures that this is a proper coloring. In addition, an $m \times n \times n$ Latin box A is equivalent to an $m \times n$ Latin rectangle R over the symbol set $[n]$, by setting $R(i, j)$ to be the index of the unique 1 in $A(i, j, \cdot)$.

In this paper, we ask when a random three-dimensional 0-1 array contains a Latin box with high probability. Formally, let $\mathcal{M}(m, n, k; p)$ be the distribution over $m \times n \times k$ 0-1 arrays where each element is 1 with probability p. A property of such an array is *monotone* if changing zeros to ones cannot violate the property.

Part of this research was conducted at the Institute of Theoretical Studies, ETH, 8092 Zurich, Switzerland. Partially supported by Dr. Max Rössler, the Walter Haefner Foundation and the ETH Foundation.

Definition 1.2. Let $m = m(n), k = k(n)$ satisfy $m \le n \le k$ for all $n \in \mathbb{N}$. $p_0 = p(n)$ is a *threshold* for a monotone property $\mathcal{P}$ if

$$\lim_{n\to\infty} \Pr[M \sim \mathcal{M}(m, n, k; p) \text{ satisfies } \mathcal{P}] = \begin{cases} 0 & \text{if } p/p_0 \to 0 \\ 1 & \text{if } p/p_0 \to \infty \end{cases}.$$

p_0 is a *sharp threshold* for $\mathcal{P}$ if for every $\eta > 0$,

$$\lim_{n\to\infty} \Pr[M \sim \mathcal{M}(m, n, k; p) \text{ satisfies } \mathcal{P}] = \begin{cases} 0 & \text{if } p < (1-\eta)p_0 \\ 1 & \text{if } p > (1+\eta)p_0 \end{cases}.$$

Our first result addresses the motivating case of $n = m = k$, namely, Latin squares. Here, and throughout the paper, we abuse notation and refer to $n \times n \times n$ Latin boxes as Latin squares.

Theorem 1.3. *There is an infinite family $F \subseteq \mathbb{N}$ and $p < 1$ such that*

$$\lim_{n\in F, n\to\infty} \Pr[M \sim \mathcal{M}(n, n, n; p) \text{ contains a Latin square}] = 1.$$

This theorem is proved in Section 2. It is actually an easy consequence of a stronger result of Andrén, Casselgren, and Öhman [2], who showed that an analogous minimum-degree result holds. We include it here because the proof is short and elegant. We also note that Keevash's method of randomized algebraic constructions [10, 11] can likely be used to show the existence of some $\varepsilon > 0$ for which $\mathcal{M}(n, n, n; n^{-\varepsilon})$ contains a Latin square with high probability. Showing this, however, is beyond the scope of this paper.

A recurring theme in the study of threshold properties is that an obvious obstruction for a property is essentially the *only* obstruction for that property. For example, in the $G(n; p)$ model, a random graph contains a perfect matching w.h.p. whenever it contains no isolated vertices. In the case of Latin squares, the obvious obstruction is a line with no 1s, corresponding to a threshold of $p = \log(n)/n$. This leads us to the following conjecture.

Conjecture 1.4. *The threshold for $M \sim \mathcal{M}(n, n, n; p)$ to contain a Latin square is $p = \log(n)/n$.*

A similar conjecture was proposed by Casselgren and Häggkvist [6, Conjecture 1.4], although the underlying probability models are different.

The next theorem deals with the case $m < n = k$. It can be interpreted as a result on Latin rectangles. Following a common abuse of notation, here and in the rest of the paper we round large reals to the nearest integer. By an argument of van-Lint and Wilson [15, Theorem 17.3], the number of Latin squares is $((1 + o(1))n/e^2)^{n^2}$. Essentially the same argument implies that for fixed $\varepsilon \in (0, 1)$, the number of $(1-\varepsilon)n \times n$ Latin rectangles is asymptotically $\left((1 + o(1))\left(\frac{1}{\varepsilon}\right)^{\varepsilon/(1-\varepsilon)} \frac{n}{e^2}\right)^{(1-\varepsilon)n^2}$. For the sake of completeness, we prove this assertion in Appendix B.

Theorem 1.5. *Let $\varepsilon > 0$. The threshold for $M \sim \mathcal{M}((1-\varepsilon)n, n, n; p)$ to contain a Latin box is* $\log(n)/n$. *Furthermore, if $p = \omega(\log(n)/n)$, then with high probability M supports $\left((1 \pm o(1))\left(\frac{1}{\varepsilon}\right)^{\varepsilon/(1-\varepsilon)} \frac{n}{e^2}p\right)^{(1-\varepsilon)n^2}$ Latin boxes.*

We prove this theorem in Section 3. A recent work by Casselgren and Häggkvist [6] proved a similar result for $1 - o(n^{-1/2}) < \varepsilon < 1$. Our theorem can be viewed as a strengthening of their result to any constant $\varepsilon > 0$.

The next theorem can be interpreted as a result on edge-coloring $K_{n,n}$ with $(1+\varepsilon)\,n$ colors. It is proved in Section 4.

Theorem 1.6. *Let $\varepsilon > 0$. The threshold for $M \sim \mathcal{M}\left(n, n, (1+\varepsilon)n; p\right)$ to contain a Latin box is $p = \frac{2\log n}{(1+\varepsilon)n}$, and this threshold is sharp.*

In fact, we prove a stronger result (Theorem 4.3): In the random process where, starting with the all zeros array, at each step we flip a randomly chosen 0 to 1, then with high probability the first time at which the array contains a Latin box is equal to the time at which every line of the form $(r, c, \cdot)$ contains at least one 1.

1.1. **Notation.** We use asymptotic notation in the usual way. For example, if $f, g : \mathbb{N} \to (0, \infty)$, then $f(n) = O\left(g(n)\right)$ means that $\limsup_{n\to\infty} f(n)/g(n) < \infty$. We also make use of asymptotic notation in arithmetic expressions. For example, by $f(n) = n + e^{O(g(n))}$ we mean that there exists a function h satisfying $h(n) = O\left(g(n)\right)$ and $f(n) = n + e^{h(n)}$.

2. Proof of Theorem 1.3

Proof of Theorem 1.3. For $p \in (0,1)$, we define $F = \{2^k : k \in \mathbb{N}\}$, and give a recursive bound on

$$p_k = \Pr\left[M \sim \mathcal{M}\left(2^k, 2^k, 2^k; p\right) \text{ contains a Latin square}\right].$$

Consider first the case $k = 1$. The probability that $M \sim \mathcal{M}\left(2, 2, 2; p\right)$ contains a given order-2 Latin square is p^4. As there are exactly two such Latin squares, and they are disjoint, by the inclusion-exclusion principle the probability that M contains a Latin square is $q(p) := 2p^4 - p^8$.

For $k > 1$, we view $M \sim \mathcal{M}\left(2^k, 2^k, 2^k; p\right)$ as a $2 \times 2 \times 2$ block array, where each block is distributed according to $\mathcal{M}\left(2^{k-1}, 2^{k-1}, 2^{k-1}; p\right)$. If there is an order-2 Latin square L such that the blocks in M corresponding to the 1s of L all contain order-2^{k-1} Latin squares, then the union of these squares is a Latin square contained in M.

The probability that this happens is $q(p_{k-1})$, and so we have $p_k \geq q(p_{k-1})$ and $p_1 = q(p)$. We note that the equation $q(x) = x$ has a unique solution $p^* \in (0,1)$, and that for $x \in (p^*, 1)$, $q(x) > x$. Therefore, if $p > p^* \approx 0.9206$, the sequence $\{p_k\}_{k=1}^{\infty}$ is monotone increasing and bounded and hence convergent. Let $p' = \lim_{k\to\infty} p_k$. As q is continuous and increasing on $(p^*, 1]$ we have $p' \leq q\left(p'\right) = \lim_{k\to\infty} q(p_k) \leq \lim_{k\to\infty} p_{k+1} = p'$. The unique fixed point of q in the interval $(p^*, 1]$ is 1, and so $p' = 1$.

□

In order to obtain a better bound on p, one can in principle repeat the above argument for any fixed n_0. The probability that $M \sim \mathcal{M}\left(n_0, n_0, n_0; p\right)$ contains a Latin square is given by some polynomial $q_{n_0}(p)$. One can compute $q_{n_0}(p)$ by listing all order-n_0 Latin squares and applying the inclusion-exclusion principle to calculate the probability that M contains one of them.

It is possible to show that there exists some $p^*_{n_0} \in (0,1)$, such that for p between $p^*_{n_0}$ and 1, $q_{n_0}(p) > p$. Indeed, fix two disjoint order-n_0 Latin squares, and let

$\tilde{q}(p) = 2p^{n_0^2} - p^{2n_0^2}$ be the probability that $M \sim \mathcal{M}(n_0, n_0, n_0; p)$ contains at least one of them. Clearly, $q_{n_0}(p) \geq \tilde{q}(p)$, and when $1 - 1/(2n_0^4) < p < 1$ one can check that $\tilde{q}(p) > p$.

Set $F = \{n_0^k : k \in \mathbb{N}\}$. Consider an $n_0^k \times n_0^k \times n_0^k$ 0-1 array A as an $n_0 \times n_0 \times n_0$ block array consisting of n_0^3 blocks, each of which is an $n_0^{k-1} \times n_0^{k-1} \times n_0^{k-1}$ 0-1 array. We say that A is a *block* Latin square if there is some order-n_0 Latin square L such that each block of A is an order-n_0^{k-1} Latin square if the corresponding element of L is 1, or the all zero array otherwise.

Now, the probability p_k that $M \sim \mathcal{M}(n_0, n_0, n_0; p)$ contains a Latin square is bounded below by the probability that it contains a block Latin square, which is $q_{n_0}(p_{k-1})$. Therefore, if $p > p^*_{n_0}$, we have $\lim_{k\to\infty} p_k = 1$. For example, performing this calculation for $n_0 = 3$ gives $p^*_3 \approx 0.86$. As a practical matter, however, this procedure seems computationally infeasible for much larger values of n_0.

3. Proof of Theorem 1.5

It is easy to show that for small p, with high probability $M \sim \mathcal{M}((1-\varepsilon)n, n, n; p)$ has an empty line of the form $M(i, j, \cdot)$. Indeed, the number of such lines is distributed binomially with parameters $(1-\varepsilon)n^2, (1-p)^n$, and when $p < \frac{1}{2}\log(n)/n$, a second moment argument shows that with high probability there is such an empty line. In this case M does not contain a Latin box.

For the upper bound, we show that for every $\varepsilon > 0$, there is a constant $C > 0$ depending only on ε such that if $p \geq C\log(n)/n$, then w.h.p. $M \sim \mathcal{M}((1-\varepsilon)n, n, n; p)$ contains a Latin box. We present a randomized algorithm for finding a Latin box, and show that with high probability it succeeds.

Note that a Latin box in M is a sequence of $(1-\varepsilon)\,n$ disjoint permutation matrices P_i, one in each plane of the form $M_i := M(i, \cdot, \cdot)$. Therefore, a natural algorithm to consider is to deal with these plane one by one, at each step choosing a permutation matrix supported by M_i that does not conflict with previous choices.

To analyze this algorithm, consider the i-th step. At this stage, $(i-1)$ permutation matrices have already been chosen, ruling out exactly $(i-1)$ entries in each row and column of M_i. Our task is to find a permutation matrix supported by the remaining elements of M_i.

Any $n \times n$ 0-1 matrix is the biadjacency matrix of a bipartite graph on $n + n$ vertices. In this language, the elements that have not been ruled out correspond to a regular bipartite graph G_i, and we want to find a perfect matching of a random subgraph of G_i, in which we keep each edge with probability p.

It is well known that with high probability a random bipartite graph has a perfect matching when it has no isolated vertices, which happens around $p = \log(n)/n$. The same holds for a random subgraph of a dense regular bipartite graph: Goel, Kapralov, and Khanna [9, Theorem 2.1] proved that there exists a constant $C > 0$ s.t. if G is a k-regular bipartite graph on $2n$ vertices, then a random subgraph of G in which each edge is retained independently with probability $p = Cn\log(n)/k^2$ contains a perfect matching with high probability. A careful analysis of their proof shows that if C is large enough then the probability of failure is $o(1/n)$. In our context this implies that if $p \geq C\log(n)/(\varepsilon^2 n)$, then w.h.p. $M \sim \mathcal{M}((1-\varepsilon)n, n, n; p)$ contains a Latin box.

The arguments above determine the threshold for the appearance of Latin boxes in $\mathcal{M}((1-\varepsilon)n, n, n; p)$. In order to prove that w.h.p. M contains close to the expected number of Latin boxes, we modify the algorithm by requiring that each permutation matrix be chosen uniformly at random. As we will show, this ensures that with high probability the graphs G_i are all pseudorandom. We then prove that with high probability, a random subgraph of a sufficiently dense pseudorandom regular bipartite graph has many perfect matchings.

Suppose that $f(n) = \omega(1)$ and $p = f(n)\frac{\log(n)}{n}$. Set $\delta = \max(f(n)^{-1/3}, 1/n) = o(1)$. Wherever necessary we assume that n is sufficiently large for asymptotic inequalities to hold. Formally, at the i-th step we choose a permutation matrix uniformly at random from the set of permutation matrices supported by M_i that are disjoint from previous choices. Now, set $k = k(i) := n-i+1$, and set $L = L(i) := (1-\delta)kp$. If the number of choices at step i is less than $L^n \frac{n!}{n^n}$, the algorithm aborts. We will show that with high probability, the algorithm does not abort, and therefore it succeeds in finding a Latin box. This implies the enumeration result.

Indeed, let A be the number of Latin boxes supported by M. The probability of a specific Latin box being chosen by the algorithm is at most

$$Q = \prod_{i=1}^{(1-\varepsilon)n} \left(L(i)^n \frac{n!}{n^n} \right)^{-1}.$$

Therefore, the probability that the algorithm succeeds is at most AQ. On the other hand, the algorithm succeeds w.h.p. and therefore, applying Stirling's approximation,

$$A \geq (1-o(1))/Q = \left((1 \pm o(1)) \left(\frac{1}{\varepsilon}\right)^{\varepsilon/(1-\varepsilon)} \frac{n}{e^2} p \right)^{(1-\varepsilon)n^2}.$$

The upper bound on A follows from Markov's inequality, together with the observation that

$$\mathbb{E}[A] = \left((1+o(1)) \left(\frac{1}{\varepsilon}\right)^{\varepsilon/(1-\varepsilon)} \frac{n}{e^2} p \right)^{(1-\varepsilon)n^2}.$$

As described above, let G_i be the k-regular bipartite graph corresponding to the elements that were not ruled out by previous choices. Let H_i be the intersection of G_i with the graph corresponding to M_i. Thus, H_i is distributed as a random subgraph of G_i, where each edge is kept with probability p. It suffices to show that with high probability the graphs H_i all have sufficiently many perfect matchings.

We say that a k-regular bipartite graph $G = \langle U \cup V, E \rangle$ is *c-pseudorandom* if for every $X \subseteq U, Y \subseteq V$ such that $|X|, |Y| \geq \frac{\varepsilon}{10} n$, the number $E_G(X,Y)$ of edges between X and Y is at least $(1-c)|X||Y|\frac{k}{n}$. Our general strategy is to show that with high probability the graphs G_i are all sufficiently pseudorandom, and that this implies the desired property for the graphs H_i.

The following lemma will enable us to bound the number of perfect matchings in H_i provided that G_i is pseudorandom.

Lemma 3.1. *Let G be a k-regular $\delta^{1/3}$-pseudorandom graph, and let H be a random subgraph of G, in which each edge of G survives with probability p. With probability $1 - n^{-\omega(1)}$, the graph H contains an L-factor, i.e. an L-regular spanning subgraph, where $L = (1-\delta)kp$.*

The next lemma asserts that if the algorithm did not abort before the i-th step, then with high probability G_i is $\delta^{1/3}$-pseudorandom. Its proof is reminiscent of the proof of [13, Theorem 2].

Lemma 3.2. *Let $1 \leq i \leq (1-\varepsilon)n$. Conditioned on the number of perfect matchings in H_j being at least $L(j)^n \frac{n!}{n^n}$ for every $j < i$, the probability that G_i is not $\delta^{1/3}$-pseudorandom is at most $\exp(-\Omega(n))$.*

The Egorychev–Falikman theorem [7, 8] states that the permanent of an order-n doubly stochastic matrix is minimized by the matrix whose entries are all $1/n$, and is equal to $\frac{n!}{n^n}$. As the biadjacency matrix of an L-regular bipartite graph on $2n$ vertices is L times a doubly stochastic matrix, this theorem implies that such a graph has at least $L^n \frac{n!}{n^n}$ perfect matchings. In particular, if H contains an L-factor, then H has at least $L^n \frac{n!}{n^n}$ perfect matchings.

We now show how Lemmas 3.1 and 3.2 imply that w.h.p. the algorithm does not abort.

Let A_i be the event that G_i is not $\delta^{1/3}$-pseudorandom, and let B_i be the event that H_i has less than $L(i)^n \frac{n!}{n^n}$ perfect matchings. For convenience, for $1 \leq i \leq m+1$ we define $C_i = \cup_{j<i} B_j$. We want to show that $\Pr[C_{m+1}]$, which is the probability that the algorithm aborts, is $o(1)$.

We prove this by induction. We assume that $\Pr[C_i] = o(1)$, and prove that $\Pr[C_{i+1}] = o(1)$.

$$\Pr[C_{i+1}] = \Pr[\cup_{j:j\leq i} B_j] \leq \sum_{j\leq i} \Pr[B_j | \cap_{\ell<j} \overline{B_\ell}] = \sum_{j \leq i} \Pr[B_j|\overline{C_j}].$$

Now,

$$\Pr\left[B_j|\overline{C_j}\right] \leq \Pr\left[B_j|\overline{C_j}, \overline{A_j}\right] + \Pr\left[A_j|\overline{C_j}\right],$$

and

$$\Pr\left[B_j|\overline{C_j}, \overline{A_j}\right] \leq \frac{1}{\Pr\left[\overline{C_j}|\overline{A_j}\right]} \Pr\left[B_j|\overline{A_j}\right].$$

Applying Bayes' theorem, this is equal to

$$\frac{\Pr\left[\overline{A_j}\right]}{\Pr\left[\overline{C_j}\right]\Pr\left[\overline{A_j}|\overline{C_j}\right]} \Pr\left[B_j|\overline{A_j}\right] \leq (1+o(1))\Pr\left[B_j|\overline{A_j}\right].$$

The inequality follows from the induction hypothesis and Lemma 3.2. Thus,

$$\Pr[C_{i+1}] \leq (1+o(1)) \sum_{j\leq i} \Pr\left[B_j|\overline{A_j}\right] + \sum_{j\leq i} \Pr\left[A_j|\overline{C_j}\right].$$

Now, by Lemma 3.1, $\Pr\left[B_j|\overline{A_j}\right] = n^{-\omega(1)}$, and by Lemma 3.2 we have $\Pr\left[A_j|\overline{C_j}\right] = e^{-\Omega(n)}$. This implies that $\Pr[C_{i+1}] = o(1)$, completing the inductive proof. Thus, w.h.p. the algorithm does not abort.

We turn to prove the lemmas. In what follows, we will make repeated use of the following version of Chernoff's inequality.

Theorem 3.3 (Chernoff's inequality)**.** *Let $X_1, ..., X_N$ be i.i.d. Bernoulli random variables with $\Pr(X_i = 1) = p$ for all i, and let $Z = \sum_{i=1}^N X_i$. Then for all $\alpha > 0$ it* *holds that*

$$\Pr[Z < (1-\alpha)Np] \leq \exp\left(-\frac{\alpha^2 Np}{2}\right),$$
$$\Pr[Z > (1+\alpha)Np] \leq \exp\left(-\frac{\alpha^2 Np}{3}\right).$$

Proof of Lemma 3.1: We use the following generalization of Hall's theorem, which can be found, e.g., in [1, Theorem 3].

Theorem 3.4. *Let $G = \langle U \cup V, E\rangle$ be a balanced bipartite graph on $2n$ vertices. Then G has an L-factor if and only if for all $X \subseteq U, Y \subseteq V$ it holds that*

$$E_G(U \setminus X, V \setminus Y) \geq (n - |X| - |Y|)L.$$

Let $X \subseteq U, Y \subseteq V$, and assume without loss of generality that $|X| \geq |Y|$ and that $|X| + |Y| < n$. We say that the pair X, Y is a *bad pair* if the random variable $Z = Z_{X,Y} := E_H(U \setminus X, V \setminus Y)$ is smaller than $(n - |X| - |Y|)L$. Our goal is to show that the probability that there exists a bad pair in H is $n^{-\omega(1)}$.

We consider three cases.

Case 1: $n - |X| \geq \frac{\varepsilon}{10}n$ and $|Y| \geq \frac{\varepsilon}{10}n$. Now, $n - |Y| \geq \frac{\varepsilon}{10}n$, because $|Y| \leq |X|$, and so by the pseudorandomness of G, we have

$$E_G(U \setminus X, V \setminus Y) \geq (1 - \delta^{1/3})(n - |X|)(n - |Y|)\frac{k}{n}.$$

Hence,

$$\mathbb{E}[Z] \geq (1 - \delta^{1/3})(n - |X|)(n - |Y|)\frac{kp}{n} = \Omega\left(n \log n\right).$$

Now, we want to bound the probability that $Z < (1-\delta)kp(n - |X| - |Y|)$. Note that

$$\begin{aligned}
(1-\delta)kp(n - |X| - |Y|) &\leq \frac{(1-\delta)kp(n - |X| - |Y|)}{(1-\delta^{1/3})(n-|X|)(n-|Y|)\frac{kp}{n}}\mathbb{E}[Z] \\
&= \frac{1-\delta}{1-\delta^{1/3}} \frac{n(n-|X|-|Y|)}{(n-|X|)(n-|Y|)}\mathbb{E}[Z] \\
&= \frac{1-\delta}{1-\delta^{1/3}}\left(1 - \frac{|X||Y|}{(n-|X|)(n-|Y|)}\right)\mathbb{E}[Z] \\
&\leq \frac{1-\delta}{1-\delta^{1/3}}\left(1 - \left(\frac{\varepsilon}{10}\right)^2\right)\mathbb{E}[Z] \leq \left(1 - \frac{\varepsilon^2}{200}\right)\mathbb{E}[Z].
\end{aligned}$$

The last inequality holds for large enough n. Therefore,

$$\Pr\left[Z < (1-\delta)kp(n - |X| - |Y|)\right] \leq \Pr\left[Z < \left(1 - \frac{\varepsilon^2}{200}\right)\mathbb{E}[Z]\right].$$

By Chernoff's inequality, this is at most

$$\exp\left(-\frac{1}{2}\left(\frac{\varepsilon^2}{200}\right)^2 \mathbb{E}[Z]\right) = \exp\left(-\Omega\left(n \log n\right)\right).$$

As there are less than 4^n possible pairs X, Y, we can apply a union bound. We conclude that the probability that a pair X, Y considered in this case is bad is at most $\exp\left(-\Omega(n \log n)\right)$.

Case 2: Y is the empty set. In this case, $Z_{X,Y}$ is the total number of edges with an endpoint in $U \setminus X$, and so it suffices to show that with sufficiently high probability, all degrees are at least L. Indeed, the expected degree of any fixed vertex v is kp, so by Chernoff's inequality,

$$\Pr\left[\deg_H(v) < (1-\delta)kp\right] \leq \exp\left(-\delta^2 kp/2\right) \leq n^{-\omega(1)}.$$

Therefore, a union bound on all $2n$ vertices implies that the probability that there is a vertex whose degree is less than L is $n^{-\omega(1)}$.

Case 3: Assume now that $t := |Y| < \frac{\varepsilon}{10}n$ and $t > 0$. Since $|X| + |Y| < n$, we have $t < s := n - |X|$.

If $s \geq \varepsilon n$, then as the number of edges from $U \setminus X$ to Y is at most kt, we have

$$E_G(U \setminus X, V \setminus Y) \geq sk - kt = k(s-t).$$

Therefore, $\mathbb{E}[Z] \geq k(s-t)p \geq s\frac{kp}{2}$. We want

$$Z \geq (1-\delta)kp(n - |X| - |Y|) = (1-\delta)kp(s-t) \leq (1-\delta)\mathbb{E}[Z].$$

On the other hand, if $s < \varepsilon n$ then the fact that the number of edges from $U \setminus X$ to Y is at most st implies that

$$E_G(U \setminus X, V \setminus Y) \geq sk - st = s(k-t).$$

Therefore, $\mathbb{E}[Z] \geq s(k-t)p \geq s\frac{kp}{2}$. We want

$$Z \geq (1-\delta)kp(n-|X|-|Y|) = (1-\delta)kp(s-t) \leq$$
$$\left(\frac{(1-\delta)kp(s-t)}{s(k-t)p}\right)\mathbb{E}[Z] =$$
$$(1-\delta)\frac{1-(t/s)}{1-(t/k)}\mathbb{E}[Z] \leq (1-\delta)\mathbb{E}[Z].$$

In either case, by Chernoff's inequality, we have

$$\Pr\left[Z < (1-\delta)kp(n-|X|-|Y|)\right] \leq \exp\left(-\delta^2 s\frac{kp}{4}\right) \leq \left(n^{-\omega(1)}\right)^s.$$

We now apply a union bound over all such pairs X, Y. Note that $s > t \geq 1$, so the probability that one of the pairs considered in the last two cases is bad is at most

$$2\sum_{s=2}^{n}\binom{n}{s}\sum_{t=1}^{\min(s-1,(\varepsilon/10)n)}\binom{n}{t}\left(n^{-\omega(1)}\right)^s \leq n^{-\omega(1)}.$$

□

To prove Lemma 3.2 we will need the following upper bound on the number of regular bipartite graphs that are not pseudorandom.

Lemma 3.5. *Let $\varepsilon n \leq k \leq n$. The number of k-regular bipartite graphs on $2n$ vertices that are not $\delta^{1/3}$-pseudorandom is bounded from above by:*

$$\binom{n}{k}^n \exp\left(-\Omega\left(\delta^{2/3}n^2\right)\right).$$

Proof. Let $R \sim G(n, n; k/n)$ be a balanced, bipartite, binomial random graph with $2n$ vertices, vertex partition $U \cup V$, and edge probability k/n. Let $\mathcal{B}$ be the event that for some $X \subseteq U$, $Y \subseteq V$ satisfying $|X|, |Y| \geq \varepsilon n/10$, the number of edges between X and Y satisfies $E_R(X, Y) \leq \left(1 - \delta^{1/3}\right) |X||Y|k/n$. As $E_R(X, Y)$ is distributed binomially with parameters $|X||Y|, k/n$, by Chernoff's inequality and a union bound over all such pairs X, Y:

$$\Pr[\mathcal{B}] \leq 4^n \exp\left(-\frac{\delta^{2/3}\varepsilon^3}{200} n^2\right) = \exp\left(-\Omega\left(\delta^{2/3} n^2\right)\right).$$

On the other hand, let $\mathcal{C}$ be the event that R is k-regular. By various estimates (e.g. [14, Proposition 2.2]), the number of k-regular bipartite graphs on $2n$ vertices is at least $\binom{n}{k}^{2n} \left(\frac{k}{n}\right)^{kn} \left(1 - \frac{k}{n}\right)^{n(n-k)}$. As k-regular graphs have precisely kn edges:

$$\begin{aligned}\Pr[\mathcal{C}] &\geq \binom{n}{k}^{2n} \left(\frac{k}{n}\right)^{2kn} \left(1 - \frac{k}{n}\right)^{2n(n-k)} \\ &= \left(\frac{n!}{k!(n-k)!} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k}\right)^{2n} = \exp\left(-O\left(n \log n\right)\right),\end{aligned}$$

where the final equality follows from Stirling's approximation. Recall that $\delta \geq 1/n$, and so $\delta^{2/3} n^2 = \omega\left(n \log n\right)$. Therefore,

$$\Pr[\mathcal{B}|\mathcal{C}] \leq \frac{\Pr[\mathcal{B}]}{\Pr[\mathcal{C}]} \leq \exp\left(-\Omega\left(\delta^{2/3} n^2\right)\right).$$

Observe that conditioning on $\mathcal{C}$ gives the uniform distribution on k-regular bipartite graphs with $2n$ vertices. As the number of k-regular bipartite graphs with $2n$ vertices is bounded from above by $\binom{n}{k}^n$, the lemma follows. □

Proof of Lemma 3.2. The proof is by induction on i. Recall that for $1 \leq i \leq m$, A_i denotes the event that G_i is not $\delta^{1/3}$-pseudorandom, B_i is the event that H_i contains fewer than $L(i)^n \frac{n!}{n^n}$ perfect matchings, and $C_i = \cup_{j<i} B_j$. We want to show that for all $1 < i \leq m + 1$, $\Pr[A_i|\overline{C_i}] = \exp\left(-\Omega(n)\right)$.

Recall that $k(i) = n - i + 1$, and thus G_i is a $k(i)$-regular graph. For $1 \leq j < (1 - \varepsilon)n$ define the error function $\alpha(j) = \sum_{s=1}^{j-1} \frac{\log(s)+2}{4s}$. Note that for all j, $\alpha(j) \leq \log^2(n)$. We will show that the following conditions hold for every $1 \leq j < (1 - \varepsilon)n$:

(1) For every graph G it holds that

$$\Pr[G_j = G|\overline{C_j}] \leq \binom{n}{k(j)}^{-n} e^{2n(\delta j + \alpha(j))}.$$

(2) $\Pr[A_j|\overline{C_j}] = \exp\left(-\Omega\left(n\right)\right)$.

(3) $\Pr[C_{j+1}|\overline{C_j}] = n^{-\omega(1)}$.

Observe that $\Pr\left[\overline{C_1}\right] = 1$. When $j = 1$, the first condition holds trivially. The second condition follows from the fact that $G_1 = K_{n,n}$. The third condition follows from Lemma 3.1. Assume inductively that for $2 \leq i < (1 - \varepsilon)n$, the conditions hold for $j = i - 1$. We will show that they hold for $j = i$. This suffices to prove the lemma.

Let $k = k(i)$. For a bipartite graph G let $\overline{G}$ denote its complement, i.e., the bipartite graph on the same vertex set with all bipartite edges not in G. Let $M(\overline{G})$ be the set of perfect matchings in $\overline{G}$.

(1) Let G be a k-regular bipartite graph. Let μ be the perfect matching chosen by the algorithm at step $(i-1)$. We apply the law of total probability to the choice of μ. We have $G_i = G$ only if $\mu = \nu$ for some $\nu \in M(\overline{G})$ and G_{i-1} is equal to the union of G_i and ν. Thus:

$$\Pr[G_i = G|\overline{C_i}] = \sum_{\nu \in M(\overline{G})} \Pr[\mu = \nu | G_{i-1} = G \cup \nu, \overline{C_i}] \Pr[G_{i-1} = G \cup \nu | \overline{C_i}]. \tag{1}$$

We bound the probabilities in the sum separately. Once again applying the law of total probability, while observing that conditioning on $\nu \notin M(H_{i-1})$ implies that $\mu \neq \nu$:

$$\Pr[\mu = \nu | G_{i-1} = G \cup \nu, \overline{C_i}] =$$
$$\Pr\left[\mu = \nu | \nu \in M(H_{i-1}), G_{i-1} = G \cup \nu, \overline{C_i}\right] \Pr\left[\nu \in M(H_{i-1}) | G_{i-1} = G \cup \nu, \overline{C_i}\right].$$

The event $\overline{C_i}$ implies that H_{i-1} contains at least $L(i-1)^n \frac{n!}{n^n}$ perfect matchings. As μ is chosen uniformly at random from $M(H_{i-1})$:

$$\Pr\left[\mu = \nu | \nu \in M(H_{i-1}), G_{i-1} = G \cup \nu, \overline{C_i}\right] \leq \frac{n^n}{L(i-1)^n n!}.$$

In order to bound the probability that $\nu \in H(M_{i-1})$, we note that every perfect matching contains n edges, and H_{i-1} is a random subgraph of G_{i-1} in which each edge survives with probability p. This would suggest a probability of p^n. However, we must be careful not to condition on properties of H_{i-1} itself. With this in mind, we replace the conditioning on $\overline{C_i}$ with conditioning on $\overline{C_{i-1}}$, and use the induction hypothesis to obtain:

$$\begin{aligned}\Pr\left[\nu \in M(H_{i-1}) | G_{i-1} = G \cup \nu, \overline{C_i}\right] &\leq \frac{\Pr\left[\nu \in M(H_{i-1}) | G_{i-1} = G \cup \nu, \overline{C_{i-1}}\right]}{\Pr\left[\overline{C_i} | \overline{C_{i-1}}\right]} \\ &\leq (1 + o(1)) p^n.\end{aligned}$$

Therefore:

$$\Pr[\mu = \nu | G_{i-1} = G \cup \nu, \overline{C_i}] \leq (1 + o(1)) \frac{n^n}{L(i-1)^n n!} p^n. \tag{2}$$

Using the induction hypothesis, we bound the second probability in inequality (1) as follows:

(3)
$$\Pr\left[G_{i-1} = G \cup \nu | \overline{C_i}\right] \leq \frac{\Pr\left[G_{i-1} = G \cup \nu | \overline{C_{i-1}}\right]}{\Pr\left[\overline{C_i} | \overline{C_{i-1}}\right]} \leq (1 + o(1)) \binom{n}{k+1}^{-n} e^{2n(\delta(i-1) + \alpha(i-1))}.$$

Together, (1), (2), and (3) imply:

$$\Pr[G_i = G|\overline{C_i}] \leq (1 + o(1)) \left|M\left(\overline{G}\right)\right| \frac{n^n p^n}{L(i-1)^n n!} \binom{n}{k+1}^{-n} e^{2n(\delta(i-1) + \alpha(i-1))}.$$

Finally, we bound $\left|M\left(\overline{G}\right)\right|$ by using Brégman's permanent inequality [5]. It implies that the number of perfect matchings in a d-regular bipartite graph on $2n$ vertices is at most $(d!)^{n/d}$. Since $\overline{G}$ is an $(i-1)$-regular bipartite graph on $2n$ vertices, we have $\left|M\left(\overline{G}\right)\right| \leq ((i-1)!)^{n/(i-1)}$. Therefore, using the inequality $\sqrt{2\pi\ell}(\ell/e)^{\ell} \leq \ell! \leq e\sqrt{\ell}(\ell/e)^{\ell}$, which holds for all natural ℓ:

$$\begin{aligned}\Pr[G_i = G|\overline{C_i}] &\leq (1+o(1))\,((i-1)!)^{n/(i-1)}\,\frac{n^n p^n}{L(i-1)^n n!}\binom{n}{k+1}^{-n} e^{2n(\delta(i-1)+\alpha(i-1))}\\ &\leq (1+o(1))\left(e\sqrt{i-1}\right)^{n/(i-1)}\left(\frac{n}{e}\right)^n \frac{1}{n!(1-\delta)^n}\binom{n}{k}^{-n} e^{2n(\delta(i-1)+\alpha(i-1))}\\ &\leq \binom{n}{k}^{-n} e^{2n(\delta i+\alpha(i))}\end{aligned}$$

as desired.

(2) Let $\mathcal{G}$ be the set of k-regular graphs on $2n$ vertices that are not $\delta^{1/3}$-pseudorandom. Then:

$$\Pr\left[A_i|\overline{C_i}\right] = \sum_{G\in\mathcal{G}} \Pr\left[G_i = G|\overline{C_i}\right].$$

We have already shown that for any G it holds that $\Pr\left[G_i = G|\overline{C_i}\right] \leq \binom{n}{k}^{-n} e^{2n(\delta i+\alpha(i))}$. Furthermore, by Lemma 3.5: $|\mathcal{G}| \leq \binom{n}{k}^n \exp\left(-\Omega\left(\delta^{2/3}n^2\right)\right)$. Therefore:

$$\Pr\left[A_i|\overline{C_i}\right] \leq \exp\left(2\delta n^2 - \Omega\left(\delta^{2/3}n^2\right) + n\log^2(n)\right) = \exp\left(-\Omega(n)\right).$$

(3) We have:

$$\Pr\left[C_{i+1}|\overline{C_i}\right] \leq \Pr\left[C_{i+1}|\overline{C_i},\overline{A_i}\right]\Pr\left[\overline{A_i}|\overline{C_i}\right] + \Pr\left[C_{i+1}|\overline{C_i},A_i\right]\Pr\left[A_i|\overline{C_i}\right].$$

We have already shown that $\Pr\left[A_i|\overline{C_i}\right] = n^{-\omega(1)}$. Furthermore, Lemma 3.1 implies that $\Pr\left[C_{i+1}|\overline{C_i},\overline{A_i}\right] = n^{-\omega(1)}$. Therefore:

$$\Pr\left[C_{i+1}|\overline{C_i}\right] = n^{-\omega(1)}.$$

□

4. Proof of Theorem 1.6

Let $M_{n,m,k}$ be the set of all $n\times m\times k$ 0-1 arrays. For $M \in M_{n,m,k}$ we denote by $|M|$ the number of 1s in M.

Definition 4.1. For integers $n,m \in \mathbb{N}$, an (n,n,m)-*array process* is a sequence $\{M_i\}_{i=0}^{n^2m} \subseteq M_{n,n,m}$, where M_0 is the all 0s array, and M_{i+1} is obtained from M_i by changing a single 0 to 1.

We denote a generic (n,n,m)-array process by $\tilde{M} = \{M_i\}_{i=0}^{n^2m}$ and write $\tilde{\mathcal{M}}(n,n,m)$ for the uniform distribution on such processes.

Definition 4.2. Let Q be a non-trivial monotone increasing property of $M_{n,n,m}$, and let $\tilde{M}$ be an array process. The *hitting time* of Q w.r.t. $\tilde{M}$ is defined as:

$$\tau\left(\tilde{M};Q\right) = \min\{t : M_t \text{ has } Q\}.$$

We are interested in the hitting time for the property that $\tilde{M} \sim \tilde{\mathcal{M}}(n,n,m)$, where $m \geq n$, supports a Latin box.

Let $M \in M_{n,n,m}$. For $1 \leq r,c \leq n$ we refer to a line of the form $(r,c,\cdot)$ as a *shaft*. The shaft is *empty* if $M(r,c,1) = M(r,c,2) = \ldots = M(r,c,m) = 0$. Clearly, a necessary condition for M to support a Latin box is that it have no empty shafts. We show that for m slightly larger than n this is asymptotically almost surely a sufficient condition.

Theorem 4.3. *For every $\varepsilon > 0$, if $\tilde{M} \sim \tilde{\mathcal{M}}\left(n, n, (1+\varepsilon)\, n\right)$ then asymptotically almost surely:*

$$\tau\left(\tilde{M} \text{ has no empty shafts}\right) = \tau\left(\tilde{M} \text{ supports a Latin box}\right).$$

Theorem 1.6 follows from a standard coupling between random processes and their binomial counterparts.

Proof of Theorem 1.6. Let $\varepsilon > 0$ and let $M \sim \mathcal{M}\left(n, n, (1+\varepsilon)n; p\right)$. For a fixed pair r and c, the probability that $M(r, c, \cdot)$ is empty is $(1-p)^{(1+\varepsilon)n}$. The different shafts are independent, and so the probability that there are no empty shafts is

$$q(p) = \left(1 - (1-p)^{(1+\varepsilon)n}\right)^{n^2}.$$

If, for some $\delta > 0$, $p \leq (1-\delta)\frac{2}{1+\varepsilon}\frac{\log n}{n}$, then $q(p) \to 0$, and therefore w.h.p. M contains empty shafts. In this case M does not support a Latin box.

On the other hand, if $p \geq (1+\delta)\frac{2}{1+\varepsilon}\frac{\log n}{n}$, then $q(p) \to 1$, and so w.h.p. M contains no empty shafts. Consider the following random process. For each triple r, c, v of indices, choose a real number $\alpha_{r,c,v} \sim U[0,1]$ uniformly at random from the interval $[0,1]$, all choices independent. Now, M is identically distributed to the array M' in which all entries with $\alpha_{r,c,v} < p$ are set to 1, and all other entries are 0. Furthermore, let $\tilde{N}'$ be the array process obtained by flipping the entries of the all zeros array to 1 in ascending order of α. Note that $\tilde{N}'$ is a uniformly random array process.

Let $t = |M'|$. Observe that $M' = N'_t$, and therefore, w.h.p. N'_t contains no empty shafts. Thus, by Theorem 4.3, w.h.p. N'_t contains a Latin box, which implies that M' contains a Latin box. Since M and M' are identically distributed, w.h.p. M contains a Latin box. □

Henceforth, fix $\varepsilon > 0$ and $m = (1+\varepsilon)\, n$. Wherever necessary we assume that n is arbitrarily large and ε is arbitrarily small.

To prove Theorem 4.3, we introduce a new model for random arrays, denoted $\mathcal{M}\left(n, n, m; p; \geq 1\right)$, whose sample space consists of $n \times n \times m$ $(0,1)$-arrays where each 1 is colored either green or blue. The green values are an array $M_G \sim \mathcal{M}\left(n, n, m; p\right)$. Then, from each empty shaft in M_G, a position is chosen uniformly at random (all choices independent), changed to 1, and colored blue. Denote by M_B the array of blue values, and set $M = M_G + M_B$. The next proposition shows that it is enough to prove that w.h.p. $M \sim \mathcal{M}\left(n, n, m; p; \geq 1\right)$ supports a Latin box, for a suitable choice of p.

Proposition 4.4. *Let Q be a monotone property of $M_{n,n,m}$ implying that there are no empty shafts. Let $p = \frac{2}{1+\varepsilon}\frac{\log n - \log\log n}{n}$. If Q holds w.h.p. for $M \sim \mathcal{M}\left(n, n, m; p; \geq 1\right)$, then for almost every $\tilde{M} \sim \tilde{\mathcal{M}}\left(n, n, m\right)$:*

$$\tau\left(\tilde{M}; Q\right) = \tau\left(\tilde{M} \text{ has no empty shafts}\right).$$

Proposition 4.4 is similar to analogous claims used to prove hitting time results in random graph and hypergraph processes (for example [4, Lemma 7.9] and [12, Lemma 1]).

Proof. As in the proof of Theorem 1.6, for each triple r, c, v let $\alpha_{r,c,v} \in [0,1]$ be drawn uniformly at random and independently. Now M_G is identically distributed to the array M' in which all entries with $\alpha_{r,c,v} < p$ are set to 1, and all other entries are 0. Furthermore, M_B is identically distributed to the array M'' in which, for each empty shaft r, c in M', the element with minimal $\alpha_{r,c,v}$ is set to 1. As before, let $\tilde{N}$ be the (uniformly random) array process where elements are set to 1 in ascending order of α. Let N_t be the first array in which there are no empty shafts. Recall that w.h.p. M' has empty shafts. Therefore, w.h.p., $\mathit{supp}(M' + M'') \subseteq \mathit{supp}(N_t)$. Now, $M' + M'' \sim M$, w.h.p. $M \in Q$, and Q is a monotone property. Therefore, $N_t \in Q$ w.h.p. □

Henceforth, let $M = M_G + M_B \sim \mathcal{M}\left(n, n, m; p; \geq 1\right)$, with p as in the statement of Proposition 4.4. Unless stated otherwise all probabilities refer to this distribution. For $1 \leq r, c \leq n$ set:

$$d(r,c) = \sum_{i=1}^{m} M\left(r,c,i\right),$$

$$d_m(r,c) = \sum_{i=n+1}^{m} M\left(r,c,i\right).$$

In what follows, we think of an $n \times n \times m$ Latin box as a function $L : [n]^2 \to [m]$ such that $L\left(a,b\right) \neq L\left(c,d\right)$ whenever (a,b) and (c,d) have exactly one coordinate in common. A function $B : S \to [m]$ is a *partial Latin box* if $S \subseteq [n]^2$ and $B\left(a,b\right) \neq B\left(c,d\right)$ whenever (a,b) and (c,d) have exactly one coordinate in common. We call the positions in S *covered*, and those in $[n]^2 \setminus S$ *uncovered*. B is *supported* by M if for all $(r,c) \in S, M\left(r,c,B(r,c)\right) = 1$. We will occasionally use the adjective "proper" to distinguish a Latin box from a partial one.

Assuming Proposition 4.4, it suffices to prove that w.h.p. M supports a Latin box. We will show that w.h.p. we can construct partial Latin boxes B_1, B_2, B_3, B_4 supported by M, and then show that w.h.p. B_4 can be completed to a proper Latin box B, also supported by M. The stages of the construction are roughly as follows:

- Construct B_1 by covering all positions (r,c) s.t. $d(r,c) - d_m(r,c) \leq \log\log n$ and $d_m(r,c) \leq \frac{\varepsilon}{1+\varepsilon}\log n$.
- Extend B_1 to B_2 by covering all positions (r,c) for which $d_m(r,c) \leq \frac{\varepsilon}{1+\varepsilon}\log n$.
- Construct B_3 using only symbols from $[n]$, and covering all but $o\left(n\right)$ positions in each row and column.
- Combine B_2 and B_3 to construct B_4, in which all but $o\left(n\right)$ positions in each row and column are covered, and in addition each uncovered position (r,c) satisfies $d_m(r,c) \geq \frac{\varepsilon}{1+\varepsilon}\log n$.
- Extend B_4 to a proper Latin box B by covering the remaining positions with values from $\{n+1, \ldots, m\}$.

B_1, B_2, and B are found via a simple randomized algorithm. To construct B_3, we use a random greedy algorithm. B_4 is constructed by "overwriting" B_3 with B_2, and erasing any values from B_3 that collide with B_2. We now prove that these steps can, in fact, be successfully completed w.h.p.

The following lemma constructs B_2. The construction of B_1 is an ingredient in the proof.

Lemma 4.5. *W.h.p. M supports a partial Latin box B_2 covering only $o(n)$ positions in each row and column s.t. if $(r,c) \in [n]^2$ is not covered by B_2 then $d_m(r,c) \geq \frac{\varepsilon}{1+\varepsilon}\log n$.*

Proof. For a position (r,c) let $X_{r,c} = \sum_{i=n+1}^{m} M_G(r,c,i)$, and let $Y_{r,c}$ be the indicator of the event $X_{r,c} < \frac{\varepsilon}{1+\varepsilon}\log n$. Then $X_{r,c}$ are i.i.d. binomial random variables with distribution $Bin(\varepsilon n, p)$, and so by Chernoff's inequality (Theorem 3.3):

$$\Pr\left[X_{r,c} \leq \frac{\varepsilon}{1+\varepsilon}\log n\right] \leq n^{-\frac{\varepsilon}{4(1+\varepsilon)}+o(1)}.$$

Thus $Y_{r,c} \sim Ber(q)$ for some $q \leq n^{-\frac{\varepsilon}{4(1+\varepsilon)}+o(1)}$. Now, the expected number of positions in each row or column for which $Y_{r,c} = 1$ is $nq \leq n^{1-\frac{\varepsilon}{4(1+\varepsilon)}+o(1)}$, and again applying Chernoff's inequality we obtain that w.h.p. there are at most $n^{1-\delta}$ such positions in each row and column, for some $\delta > 0$.

Let $S = \left\{(r,c) \in [n]^2 : Y_{r,c} = 1\right\}$. By the above, w.h.p. S contains only $o(n)$ positions in each row and column. We show that w.h.p. we can find a partial Latin box B_2 supported by M whose domain is S.

We do this in two stages: We first cover all $(r,c) \in S$ s.t. $d(r,c) - d_m(r,c)$ is small with a partial Latin box B_1. We then show that w.h.p. B_1 can be extended to the desired B_2.

Let $T = \{(r,c) \in S : d(r,c) - d_m(r,c) \leq \log\log n\}$. Observe that

$$d(r,c) - d_m(r,c) = \sum_{i=1}^{n} M(r,c,i) \geq \sum_{i=1}^{n} M_G(r,c,i) \sim Bin(n,p).$$

We have:

$$\begin{aligned}\Pr\left[(r,c) \in T \middle| (r,c) \in S\right] &\leq \sum_{k=0}^{\log\log n} \binom{n}{k} p^k (1-p)^{n-k} = (1-p)^n \sum_{k=0}^{\log\log n} \binom{n}{k}\left(\frac{p}{1-p}\right)^k \\ &\leq \left(\frac{\log n}{n}\right)^{\frac{2}{1+\varepsilon}} \left(1 + \sum_{k=1}^{\log\log n}\left(\frac{2e\log n}{k}\right)^k\right) \\ &\leq \left(\frac{\log n}{n}\right)^{\frac{2}{1+\varepsilon}} \log\log n \left(6\log n\right)^{\log\log n} = \frac{e^{O\left((\log\log n)^2\right)}}{n^{\frac{2}{1+\varepsilon}}}.\end{aligned}$$

Applying Markov's inequality we conclude that w.h.p. $|T| \leq n^{3\varepsilon}$.

We construct B_1 by covering T. Note that for every $(r,c) \in T$, $d(r,c) \geq 1$. For each $(r,c) \in T$, choose $B_1(r,c)$ uniformly at random from $\{i : M(r,c,i) = 1\} \cap [n]$ if this set is non-empty; otherwise choose $B_1(r,c)$ uniformly at random from $\{i : M(r,c,i) = 1\} \subseteq [m] \setminus [n]$. Note that in the former case $B_1(r,c)$ is distributed uniformly amongst $[n]$ and in the latter case $B_1(r,c)$ is distributed uniformly amongst $[m] \setminus [n]$. Therefore, $\{B_1(r,c)\}_{(r,c)\in T}$ is a collection of $O(n^{3\varepsilon}) = o(\sqrt{n})$ values, each chosen uniformly at random and independently from a set of size $\Omega(n)$. Hence w.h.p. no value appears more than once. This implies that w.h.p. B_1 is indeed a partial Latin box covering T.

The remaining positions $(r,c) \in S \setminus T$ all satisfy $d(r,c) - d_m(r,c) \geq \log\log n$. For each $(r,c) \in S \setminus T$ let $V'(r,c) := \{i \in [n] : M(r,c,i) = 1\}$. Choose $V(r,c) \subseteq V'(r,c)$ of size $\log\log n$ uniformly at random and independently. Note that $\{V(r,c)\}_{(r,c)\in S\setminus T}$

is a collection of uniformly random and independent elements of $\binom{[n]}{\log\log n}$. We construct B_2 by extending B_1 greedily while avoiding collisions: For each $(r,c) \in S \setminus T$, we choose $B_2(r,c)$ uniformly at random from the values in $V(r,c)$ that have not yet been used in row r or column c. W.h.p. this procedure succeeds: Indeed, when choosing the value of B_2 for any heretofore uncovered $(r,c) \in S$, there are at most $2n^{1-\delta} + n^{3\varepsilon} \leq 3n^{1-\delta}$ previously covered positions in row r and column c. Thus there are at most $3n^{1-\delta}$ forbidden values. Therefore the probability that $V(r,c)$ contains only forbidden values is at most:

$$\frac{\binom{3n^{1-\delta}}{\log\log n}}{\binom{n}{\log\log n}} = n^{-\omega(1)}.$$

Applying a union bound to the $O\left(n^2\right)$ steps in the greedy algorithm, we see that w.h.p. the algorithm succeeds in constructing B_2.

□

To construct B_3 we need the following lemma.

Lemma 4.6. *Let $q = \omega\left(\frac{1}{n}\right)$ and let $M \sim \mathcal{M}\left(n,n,n;q\right)$. W.h.p. M supports a partial Latin box with at most $o\left(n\right)$ uncovered positions in each row and column.*

We prove Lemma 4.6 by showing that w.h.p. a random greedy algorithm succeeds in finding an appropriate partial Latin box. The proof is deferred to Appendix A.

Proof of Theorem 4.3. Recall that $M = M_G + M_B \sim \mathcal{M}\left(n,n,m;p;\geq 1\right)$. W.h.p. M supports a partial Latin box B_2 as per the conclusion of Lemma 4.5. By Lemma 4.6, w.h.p. $M_G \sim \mathcal{M}\left(n,n,m;p\right)$ supports a partial Latin box B_3 covering all but at most $o\left(n\right)$ positions in each row and column, and using only values from $[n]$.

For $i = 2,3$ let S_i be the set of positions covered by B_i. Define the partial Latin box B_4 as follows: For all $(r,c) \in S_2$ set $B_4(r,c) = B_2(r,c)$. For all $(r,c) \in S_3 \setminus S_2$ s.t. $B_3(r,c)$ isn't used by B_2 in row r or column c, set $B_4(r,c) = B_3(r,c)$. B_4 is thus a partial Latin box covering all but at most $o\left(n\right)$ positions in each row and column, and in which each row and column uses at most $o\left(n\right)$ values from $\{n+1, n+2, \ldots, m\}$. Additionally, if (r,c) isn't covered by B_4 then (since (r,c) is not covered by B_2) $d_m(r,c) \geq \frac{\varepsilon}{1+\varepsilon}\log n$. We now show that w.h.p. a random greedy algorithm succeeds in extending B_4 to a proper Latin box.

In a manner similar to the proof of Lemma 4.5, for each uncovered (r,c) let $W'(r,c) = \{v \in [m] \setminus [n] : M\left(r,c,v\right) = 1\}$, and let $W(r,c) \subseteq W'(r,c)$ be uniformly random subsets of size $\frac{\varepsilon}{1+\varepsilon}\log n$ chosen independently.

Iterate over the uncovered elements in an arbitrary order. For every uncovered (r,c) choose $B(r,c)$ uniformly at random from $W(r,c)$ that have not previously been used in the same row or column. At each step of the algorithm, there are at most $o\left(n\right)$ forbidden values, so the probability that all available values are forbidden is at most:

$$\frac{\binom{o(n)}{\frac{\varepsilon}{1+\varepsilon}\log n}}{\binom{\varepsilon n}{\frac{\varepsilon}{1+\varepsilon}\log n}} = n^{-\omega(1)}.$$

There are $O\left(n^2\right)$ steps in the greedy algorithm so, applying a union bound, the probability of failure is $o\left(1\right)$.

□

Acknowledgment. We wish to thank the anonymous referees for their helpful comments, which greatly improved the exposition, as well as for finding an error in a previous version of the paper.

Appendix A. Random Greedy Packing in Random Hypergraphs

In this section we prove Lemma 4.6. Although it is a statement about random arrays, it is convenient to reformulate it in terms of random hypergraphs. This is because the random greedy algorithm we are about to introduce is similar to the *triangle removal process* analyzed, among others, by Wormald [16, Section 7.2] and Bohman, Frieze, and Lubetzky [3].

A.1. Notation and Terminology.

We denote by $H_3(n)$ the set of tripartite, 3-uniform hypergraphs whose vertex set is $[n] \sqcup [n] \sqcup [n]$. A *triangle* is a partite vertex set of size 3 and an *edge* is a partite vertex set of size 2. To avoid unnecessary delimiters we sometimes write abc for the triangle $\{a, b, c\}$, and ab for the edge between a and b. We denote by $\mathcal{H}_3(n;p)$ the distribution on $H_3(n)$ where each triangle is included in the hypergraph with probability p, independently of the other triangles, and we denote by $\mathcal{H}_3(n;m)$ the distribution on $H_3(n)$ where the triangle set is a uniformly random element of $\binom{[n]^3}{m}$.

Let $H \in H_3(n)$, and let $T(H)$ denote the set of its triangles. A set $S \subseteq T(H)$ is a *set of edge-disjoint triangles* (*SET*) in H if for all $t_1, t_2 \in S$, $|t_1 \cap t_2| \geq 2 \implies t_1 = t_2$. If uv is an edge, we say it is *covered* by S if there exists some $t \in S$ s.t. $\{u, v\} \subseteq t$. In this case we write $uv \in G(S)$. We say a triangle t is *edge-disjoint from* S if none of its edges are covered by S. For $v \in V(H)$, let $d_S(v)$ be the number of triangles in S containing v.

For $a, b \in \mathbb{R}$, we write $a \pm b$ to indicate some quantity in the interval $[a - |b|, a + |b|]$. We say that an event occurs *with very high probability* (**w.v.h.p.**) if it occurs with probability $1 - n^{-\omega(1)}$.

A.2. From Arrays to Hypergraphs.

Let $M \in M_{n,n,n}$. We define the hypergraph $H_M \in H_3(n)$ by setting $T(H_M) = \{(i,j,k) \in [n]^3 : M(i,j,k) = 1\}$. This induces a natural correspondence between SETs in H_M and partial Latin boxes supported by M.

Lemma 4.6 now follows from:

Lemma A.1. *Let $p = \omega\left(\frac{1}{n}\right)$ and let $H \sim \mathcal{H}_3(n;p)$. W.h.p. $T(H)$ contains an SET S s.t. for every vertex v, $d_S(v) = (1 - o(1))\, n$.*

A.3. Proof of Lemma A.1.

In the *random hypergraph process*, the triangles of the complete tripartite 3-uniform hypergraph $K^{(3)}_{n,n,n}$ are considered one by one in a uniformly random order $t_1, t_2, \ldots, t_{n^3}$. This process generates a sequence of hypergraphs $H_0, H_1, \ldots, H_{n^3} \in H_3(n)$, where $T(H_0) = \emptyset$ and $T(H_{i+1}) = T(H_i) \cup \{t_{i+1}\}$. We couple this with the following process: $S_0 = \emptyset$, and $S_{i+1} = S_i \cup \{t_{i+1}\}$ if t_{i+1} is edge disjoint from S_i, and $S_{i+1} = S_i$ otherwise. Observe that for every i, S_i is an SET in H_i. The next proposition says that w.v.h.p. the vertex degrees in S_i are concentrated.

Proposition A.2. *There exists some $\delta > 0$ s.t. w.v.h.p. for every $v \in [n] \sqcup [n] \sqcup [n]$ and every $0 \leq m \leq n^{2+\delta}$:*

$$d_{S_m}(v) = \left(1 - (1 \pm o(1)) \frac{1}{\sqrt{1 + 2\frac{m}{n^2}}}\right) n.$$

Before proving Proposition A.2, we first describe how Lemma A.1 follows from Proposition A.2.

The *random greedy packing algorithm in $H \in H_3(n)$* is the following probabilistic procedure: Set $S = \emptyset$. As long as there are triangles in $T(H)$ that are edge disjoint from S, choose one uniformly at random and add it to S. If there are no such triangles, halt. Say that a hypergraph in $H_3(n)$ is *good* if it satisfies the conclusion of Lemma A.1. Let $H \sim \mathcal{H}_3(n; p)$. We claim that w.h.p. H is good, and this is witnessed by the result of the random greedy packing algorithm in H.

Clearly, the distribution of H conditioned on $|T(H)| = m$ is identical to H_m. Moreover, given H_m, S_m is distributed identically to the result of the random greedy packing algorithm in H_m. Note also that the probability that H_m is good is increasing in m. As $|T(H)| \sim Bin(n^3, p)$ and $p = \omega\left(\frac{1}{n}\right)$, there exists some $k = \omega(n^2)$, s.t. w.h.p. $|T(H)| \geq k$. Proposition A.2 implies that H_k is good w.v.h.p. Therefore,

$$\begin{aligned}
\Pr[H \text{ is good}] &\geq \sum_{m=k}^{n^3} \Pr[H \text{ is good} | |T(H)| = m] \Pr[|T(H)| = m] \\
&= \sum_{m=k}^{n^3} \Pr[H_m \text{ is good}] \Pr[|T(H)| = m] \\
&\geq \Pr[|T(H)| \geq k] \Pr[H_k \text{ is good}] = 1 - o(1).
\end{aligned}$$

We turn to prove Proposition A.2.

Proof. We prove the proposition for $\delta = \frac{1}{100}$.

In the spirit of the differential equation method of Wormald [16] we track a set of random variables throughout the hypergraph process by modeling their evolution on a system of differential equations.

We make use of the following version of the Azuma-Hoeffding inequality, which follows from [16, Lemma 4.2].

Lemma A.3. *Let $A_0 \subseteq A_1 \subseteq \ldots \subseteq A_N$ be a filtration of a finite probability space. Let $X_0, X_1, \ldots, X_N$ be a sequence of random variables s.t. for every i, X_i is A_i-measurable. Assume that for some $C > 0$, $|X_{i+1} - X_i| \leq C$ for all i. Assume further that for all i, $\mathbb{E}[X_{i+1} - X_i | A_i] \leq 0$, i.e. $X_0, X_1, \ldots, X_N$ is a supermartingale. Finally, assume $X_0 \leq 0$. Then, for all $\lambda > 0$:*

$$\Pr[X_N > \lambda] \leq \exp\left(-\frac{\lambda^2}{2NC^2}\right).$$

We define the following functions on $[0, \infty)$, whose relevance will become apparent presently:

$$\begin{aligned} y(x) &= \frac{1}{\sqrt{1+2x}} \\ z(x) &= \frac{1}{1+2x} \end{aligned}$$

These satisfy the differential equations:

$$\begin{aligned} y' &= -yz \\ z' &= -2z^2 \end{aligned}$$

We now define the variables we want to track. For every vertex v and $0 \leq i \leq n^{2+\delta}$ write:

$$c_v(i) = n - d_{S_i}(v).$$

Next, for $0 \leq i \leq n^{2+\delta}$ we define the set of *permissible* triangles:

$$A_i = \left\{ t_j : i < j \leq n^3, \forall t \in S_i, |t \cap t_j| \leq 1 \right\}.$$

In words, A_i is the set of triangles not in H_i that, if selected at time $i+1$, will be included in S_{i+1}.

For every uncovered edge uv, we track the number of permissible triangles containing it. For convenience, we associate a random variable to covered edges as well:

$$d_{uv}(i) = \begin{cases} |\{t \in A_i : \{u,v\} \subseteq t\}| & uv \notin G(S_i) \\ nz\left(\frac{i}{n^2}\right) & otherwise \end{cases}.$$

We will show that w.v.h.p. for every $0 \leq i \leq n^{2+\delta}$, every vertex v, and every edge uv:

$$\begin{aligned} c_v(i) &= (1 \pm o(1))\, ny\left(\frac{i}{n^2}\right) \\ d_{uv}(i) &= (1 \pm o(1))\, nz\left(\frac{i}{n^2}\right). \end{aligned} \tag{4}$$

In particular, this will prove the proposition.

We first consider the evolution of the random variables d_{uv}. Note that if uv is covered by S_i then (by definition) $d_{uv}(i) = nz\left(\frac{i}{n^2}\right)$ and there is nothing to prove. So assume that uv is not covered by S_i. How might d_{uv} change during step $i+1$? Well, if uv remains uncovered, then d_{uv} will change if and only if some permissible triangle $t \in A_i$ containing uv is no longer in A_{i+1}. Now, $uv \subseteq t \in A_i$ will not be in A_{i+1} if $|t_{i+1} \cap t| = 2$, and $t_{i+1} \in A_i$. In this case, d_{uv} decreases by 1. For every $uvw \in A_i$ there are $d_{uw}(i) + d_{vw}(i) - 2$ triangles in A_i that have this effect. Thus, observing that at step i there are $\left(1 - O\left(n^{\delta-1}\right)\right) n^3$ triangles remaining to be considered that do not contain uv:

$$\begin{aligned} &\Pr\left[d_{uv}(i+1) \neq d_{uv}(i) \,|H_i, S_i, uv \notin G(S_{i+1})\right] \\ &= \frac{1}{(1 - O(n^{\delta-1}))\, n^3} \sum_{uvw \in A_i} \left(d_{uw}(i) + d_{vw}(i) - 2\right) \leq \frac{4}{n}. \end{aligned} \tag{5}$$

Note that since the underlying graph is tripartite, so long as uv remains uncovered, d_{uv} can decrease by at most 1 in a single step. Therefore:

$$\begin{aligned}&\mathbb{E}\left[d_{uv}\left(i+1\right)-d_{uv}\left(i\right)|H_i,S_i,uv\notin G\left(S_{i+1}\right)\right]\\&\quad=-\Pr\left[d_{uv}\left(i+1\right)\neq d_{uv}\left(i\right)|H_i,S_i,uv\notin G\left(S_{i+1}\right)\right].\end{aligned}\tag{6}$$

Lemma A.3 (the Azuma–Hoeffding inequality) requires control over the maximal one-step change of the sequence of random variables. Although the maximum change in d_{uv} is 1, this is too large for Lemma A.3 to be useful. Therefore, we show that d_{uv} cannot change too much in any n consecutive steps, which, after rescaling, will enable an application of Lemma A.3. First, note that for any $1 \le j < n$, we have:

$$\Pr\left[uv\in G(S_{i+n})|H_{i+j-1},S_{i+j-1},uv\notin G(S_{i+j})\right]\le\frac{2}{n}.\tag{7}$$

Indeed, the triangles $t_{i+j+1},\dots,t_{i+n}$ are a uniformly random subset of size $n-j\le n$, that are chosen from a set of size at least $n^3/2$. Furthermore, the number of triangles containing uv is bounded from above by n. Thus, the probability that one of these was chosen is at most $2n^2/n^3 = 2/n$.

Now, let $1\le i\le n^{2+\delta}-n$ and $1\le j<n$. By the law of total probability and Inequalities (5) and (7), for any H_{i+j},S_{i+j} s.t. $uv\notin G(S_{i+j})$, it holds that

$$\begin{aligned}&\Pr\left[d_{uv}(i+j)\neq d_{uv}(i+j-1)|H_{i+j-1},S_{i+j-1},uv\notin G(S_{i+n})\right]\\&\le\frac{\Pr\left[d_{uv}(i+j)\neq d_{uv}(i+j-1)|H_{i+j-1},S_{i+j-1},uv\notin G(S_{i+j})\right]}{\Pr\left[uv\notin G(S_{i+n})|H_{i+j-1},S_{i+j-1}uv\notin G(S_{i+j})\right]}\le\frac{5}{n}.\end{aligned}\tag{8}$$

Now, for $T\in\binom{[n]}{\log n}$ let B_T denote the event that for every $j\in T$, $d_{uv}(i+j)\neq d_{uv}(i+j-1)$. Applying a chain of conditional probabilities together with Inequality (8), for any $T\in\binom{[n]}{\log n}$:

$$\Pr\left[B_T|S_i,H_i,uv\notin G(S_{i+n})\right]\le\left(\frac{5}{n}\right)^{\log n}.$$

We will now use a union bound over all events B_T to show that w.v.h.p. in any n consecutive steps of the hypergraph process and for any uncovered edge uv, conditioning on the event that uv remains uncovered during these steps, d_{uv} changes by at most $\log n$. Indeed:

$$\begin{aligned}\Pr\left[d_{uv}\left(i+n\right)\le d_{uv}\left(i\right)-\log n|H_i,S_i,uv\notin G\left(S_{i+n}\right)\right]&\le\binom{n}{\log n}\left(\frac{5}{n}\right)^{\log n}\\&\le\left(\frac{5en}{n\log n}\right)^{\log n}=n^{-\omega(1)}.\end{aligned}\tag{9}$$

We treat the evolution of the variables c_v in a similar fashion. At time $i+1$, c_v decreases iff $v\in t_{i+1}\in A_i$, in which case $c_v\left(i+1\right)=c_v\left(i\right)-1$. Thus:

$$\begin{aligned}\Pr\left[c_v\left(i+1\right)\neq c_v\left(i\right)|H_i,S_i\right]&=\frac{1}{\left(1-O\left(n^{\delta-1}\right)\right)n^3}\frac{1}{2}\sum_{vu\notin G(S_i)}d_{vu}\left(i\right)\le\frac{2}{n}\\\mathbb{E}\left[c_v\left(i+1\right)-c_v\left(i\right)|H_i,S_i\right]&=-\Pr\left[c_v\left(i+1\right)\neq c_v\left(i\right)|H_i,S_i\right].\end{aligned}\tag{10}$$

By reasoning similar to that above, for any vertex v and any $0\le i\le n^{2+\delta}-n$:

$$\Pr\left[c_v\left(i+n\right)\le c_v\left(i\right)-\log n|H_i,S_i\right]\le n^{-\omega(1)}.\tag{11}$$

At this point it is convenient to rescale our variables. For $0 \leq T \leq n^{1+\delta}$, a vertex v, and an edge uv we define:

$$\begin{aligned} C_v(T) &= c_v(nT) \\ D_{uv}(T) &= d_{uv}(nT). \end{aligned}$$

Let $\varepsilon = \frac{1}{2}$. We will prove that for all $T < n^{1+\delta}$ of the form $T = kn^{\varepsilon}$ (where $k \in \left\{0, 1, \ldots, n^{2+\delta-\varepsilon}\right\}$), every vertex v, and every edge uv:

$$\begin{aligned} C_v(T) &= ny\left(\frac{T}{n}\right) \pm \alpha(T) \\ D_{uv}(T) &= nz\left(\frac{T}{n}\right) \pm \alpha(T). \end{aligned} \tag{12}$$

Where:

$$\begin{aligned} \alpha(0) &= n^{1+\delta-\frac{\varepsilon}{3}} = n^{\frac{253}{300}} \\ \alpha(T+n^{\varepsilon}) &= \alpha(T)\left(1 + \frac{20n^{\varepsilon}}{n+2T}\right). \end{aligned}$$

It is straightforward to verify that for all T:

$$\begin{aligned} &\alpha(T) \leq \alpha\left(n^{1+\delta}\right) = O\left(n^{1+11\delta-\frac{\varepsilon}{3}}\right) = O\left(n^{\frac{91}{100}}\right) \\ &ny\left(\frac{T}{n}\right) \geq nz\left(\frac{T}{n}\right) \geq nz\left(\frac{n^{1+\delta}}{n}\right) = \Omega\left(n^{1-\delta}\right) = \Omega\left(n^{\frac{99}{100}}\right). \end{aligned}$$

And so:

$$\alpha(T) = o\left(nz\left(\frac{T}{n}\right)\right), o\left(ny\left(\frac{T}{n}\right)\right). \tag{13}$$

Together, Equalities (12) and (13) imply the proposition.

We will prove that if (12) holds for T, then w.v.h.p. (12) also holds for $T + n^{\varepsilon}$. Since $C_v(0) = D_{uv}(0) = n$, an inductive argument completes the proof.

Assume (12) holds for some T. Let uv be an edge. Our first order of business is to calculate the expected change in D_{uv} in a single time step. Let $T \leq i < T + n^{\varepsilon}$. If uv is covered at time $i+1$ then (12) holds by definition. Therefore we condition on $uv \notin G(S_{i+1})$. For compactness, we set $F_i = \left(H_{in}, S_{in}, uv \notin G\left(S_{(i+1)n}\right)\right)$. Now, by definition:

$$\mathbb{E}\left[D_{uv}(i+1) - D_{uv}(i) \,|F_i\right] = \sum_{j=1}^{n} \mathbb{E}\left[d_{uv}(in+j) - d_{uv}(in+j-1)\,|F_i\right].$$

Equality (6) holds for any choice of $t_{in+1}, \ldots, t_{in+j-1}$. Furthermore, $uv \notin G\left(S_{(i+1)n}\right)$ implies $uv \notin G(S_{in+j})$. Therefore:

$$\mathbb{E}\left[d_{uv}(in+j) - d_{uv}(in+j-1)\,|F_i\right] = -\Pr\left[d_{uv}(in+j) \neq d_{uv}(in+j-1)\,|F_i\right].$$

Now, for $j \in [n]$ let B_j be the event that for some edge $ab \notin G\left(S_{(i+1)n}\right)$:

$$\left|d_{ab}(in+j-1) - d_{ab}(Tn)\right| \geq (i+1-T)\log n.$$

By Inequality (9), $\Pr\left[B_j | F_i\right] = n^{-\omega(1)}$. Note that if $\overline{B_j}$ holds, then for all $ab \notin G\left(S_{(i+1)n}\right)$, it holds that $d_{ab}(in+j-1) = d_{ab}(Tn) \pm \alpha(T) = nz\left(\frac{T}{n}\right) \pm 2\alpha(T)$.

Therefore, applying the law of total probability:

$$\begin{aligned}
&\Pr\left[d_{uv}\left(in+j\right)\neq d_{uv}\left(in+j-1\right)|F_i\right]\\
&\quad=\Pr\left[d_{uv}\left(in+j\right)\neq d_{uv}\left(in+j-1\right)|F_i,\overline{B_j}\right]\pm n^{-\omega(1)}\\
&\quad=\frac{2\left(nz\left(\frac{T}{n}\right)\pm 4\alpha\left(T\right)\right)^2}{\left(1-O\left(n^{\delta-1}\right)\right)n^3}=\left(1\pm O\left(n^{\delta-1}\right)\right)\frac{2\left(z\left(\frac{T}{n}\right)\pm 4\alpha\left(T\right)\right)^2}{n}.
\end{aligned}$$

Therefore:

$$\begin{aligned}
\mathbb{E}\left[D_{uv}\left(i+1\right)-D_{uv}\left(i\right)|F_i\right]&=-\left(1\pm O\left(n^{\delta-1}\right)\right)2\left(z\left(\frac{T}{n}\right)\pm 4\alpha\left(T\right)\right)^2\\
=-2z^2\left(\frac{T}{n}\right)\pm\frac{18z\left(\frac{T}{n}\right)}{n}\alpha\left(T\right)&=z'\left(\frac{T}{n}\right)\pm\frac{18}{n+2T}\alpha\left(T\right).
\end{aligned}$$

We cannot apply the Azuma-Hoeffding inequality to D_{uv} directly, as the change in a single time step might be as large as $\Omega(n)$, resulting in a meaningless bound. However, as we have already shown, this is unlikely to happen. We will therefore apply the Azuma-Hoeffding inequality to the conditional probability space in which the random variables we are tracking do not change too much in a single time step. Let B be the event that for some i, $|D_{uv}\left(T\left(i+1\right)\right)-T\left(i\right)|\geq\log n$. By Inequality (9) $\Pr\left[B|F_i\right]=n^{-\omega(1)}$. Therefore, applying the law of total expectation:

(14)

$$\begin{aligned}
&\mathbb{E}\left[D_{uv}\left(i+1\right)-D_{uv}\left(i\right)|F_i,\overline{B}\right]=\\
&\quad\frac{1}{\Pr\left[\overline{B}|F_i\right]}\mathbb{E}\left[D_{uv}\left(i+1\right)-D_{uv}\left(i\right)|F_i\right]-\frac{\Pr[B|F_i]}{\Pr\left[\overline{B}|F_i\right]}\mathbb{E}\left[D_{uv}\left(i+1\right)-D_{uv}\left(i\right)|F_i,B\right]\\
&\quad=z'\left(\frac{T}{n}\right)\pm\left(\frac{18}{n+2T}\alpha\left(T\right)+n^{-\omega(1)}\right)=z'\left(\frac{T}{n}\right)\pm\frac{19}{n+2T}\alpha\left(T\right).
\end{aligned}$$

We will prove the upper bound in Equation (12). The proof of the lower bound is similar. To do so we transform D_{uv} into a supermartingale. Define, for $T\leq i\leq T+n^{\varepsilon}$:

$$D'_{uv}\left(i\right):=D_{uv}\left(i\right)-nz\left(\frac{i}{n}\right)-\left(1+\frac{19\left(i-T\right)}{n+2T}\right)\alpha\left(T\right).$$

Observe that, conditioning on $\overline{B}$:

$$\begin{aligned}
&|D'_{uv}\left(i+1\right)-D'_{uv}\left(i\right)|\\
&\leq|D_{uv}\left(i+1\right)-D_{uv}\left(i\right)|+n\left|z\left(\frac{i+1}{n}\right)-z\left(\frac{i}{n}\right)\right|+\alpha\left(T\right)\frac{19}{n+2T}=O\left(\log n\right).
\end{aligned}$$

We next show that $\mathbb{E}\left[D'_{uv}\left(i+1\right)-D'_{uv}\left(i\right)|H_{in},S_{in},\overline{B}\right]\leq 0$, i.e., D'_{uv} is a supermartingale. Taking Equation (14) into account:

$$\begin{aligned}
&\mathbb{E}\left[D'_{uv}\left(i+1\right)-D'_{uv}\left(i\right)|H_{in},S_{in},\overline{B}\right]\\
&\quad=\mathbb{E}\left[D_{uv}\left(i+1\right)-D_{uv}\left(i\right)|H_{in},S_{in},\overline{B}\right]-nz\left(\frac{i+1}{n}\right)+nz\left(\frac{i}{n}\right)-\frac{19}{n+2T}\alpha\left(T\right)\\
&\quad\leq z'\left(\frac{T}{n}\right)+\frac{19}{n+2T}\alpha\left(T\right)-n\left(z\left(\frac{i+1}{n}\right)-z\left(\frac{i}{n}\right)\right)-\frac{19}{n+2i}\alpha\left(i\right)\\
&\quad\leq z'\left(\frac{T}{n}\right)-n\left(z\left(\frac{i+1}{n}\right)-z\left(\frac{i}{n}\right)\right).
\end{aligned}$$

By the mean value theorem there exists some $s\in[i/n,(i+1)/n]$ s.t. $n\left(z\left(\frac{i+1}{n}\right)-z\left(\frac{i}{n}\right)\right)=z'(s)$. Since z' is increasing it holds that $z'(s)\geq z'\left(\left(\frac{T}{n}\right)\right)$. Therefore:

$$\mathbb{E}\left[D'_{uv}\left(i+1\right)-D'_{uv}\left(i\right)|H_{in},S_{in},\overline{B}\right]\leq z'\left(\frac{T}{n}\right)-z'(s)\leq 0.$$

Finally, we apply the Azuma-Hoeffding inequality (Lemma A.3) with respect to the filtration induced by the random variables $\{H_{in},S_{in}\}_{i=T}^{T+n^{\varepsilon}}$, conditioned on $\overline{B}$.

$$\Pr\left[D'_{uv}\left(T+n^{\varepsilon}\right)\left(i\right)>\frac{n^{\varepsilon}}{n+2T}\alpha\left(T\right)|\overline{B}\right]\leq\exp\left(-\Omega\left(\frac{\left(\frac{n^{\varepsilon}}{n+2T}\alpha\left(T\right)\right)^{2}}{n^{\varepsilon}\log^{2}n}\right)\right)=n^{-\omega(1)}.$$

Since $\overline{B}$ holds w.v.h.p. we have, for all edges uv, w.v.h.p.:

$$D_{uv}\left(T+n^{\varepsilon}\right)\leq nz\left(\frac{T+n^{\varepsilon}}{n}\right)+\alpha\left(T\right)+\frac{20n^{\varepsilon}}{n+2T}\alpha\left(T\right)=nz\left(\frac{T+n^{\varepsilon}}{n}\right)+\alpha\left(T+n^{\varepsilon}\right).$$

We analyze C_v analogously, while omitting calculations very similar to those above. We focus on the most important step: calculating the expected difference. Assume Equality (12) holds for T and let $T\leq i\leq T+n^{\varepsilon}$. Let B_i be the event where for some v, $|C_v\left(i+1\right)-C_v\left(i\right)|\geq\log n$ or for some $uv\notin G\left(S_{(i+1)n}\right)$, $|D_{uv}\left(i+1\right)-D_{uv}\left(i\right)|\geq\log n$. Let $\mathcal{B}_i=\cup_{T\leq j\leq i}B_j$. By Inequality (11) $\Pr\left[\mathcal{B}_i|H_{in},S_{in}\right]=n^{-\omega(1)}$. If $\overline{\mathcal{B}_i}$ holds, then $C_v\left(i\right)=C_v\left(T\right)\pm\left(i-T\right)\log n$ and $D_{uv}\left(i\right)=D_{uv}\left(T\right)\pm\left(i-T\right)\log n$. For j between in and $(i+1)n$, each t_j is chosen uniformly at random from $n^3\left(1-O\left(n^{\delta-1}\right)\right)$ triangles. Thus, by the inductive hypothesis and Equation (10):

$$\begin{aligned}
&\mathbb{E}\left[C_v\left(i+1\right)-C_v\left(i\right)|H_{in},S_{in}\right]=\sum_{j=1}^{n}\mathbb{E}\left[c_v\left(in+j\right)-c_v\left(in+j-1\right)|H_{in},S_{in}\right]\\
&\quad=-n\cdot\frac{\left(ny\left(\frac{T}{n}\right)\pm2\alpha\left(T\right)\right)\left(nz\left(\frac{T}{n}\right)\pm2\alpha\left(T\right)\right)}{n^3\left(1\pm O\left(n^{\delta-1}\right)\right)}=y'\left(\frac{T}{n}\right)\pm\frac{6\alpha\left(T\right)z\left(\frac{T}{n}\right)}{n}\\
&\quad=y'\left(\frac{T}{n}\right)\pm\frac{6}{n+2T}\alpha\left(T\right).
\end{aligned}$$

As above, we can apply the Azuma-Hoeffding inequality to an appropriate shifted variable to obtain the result.

□

Appendix B. Asymptotic Enumeration of Latin Rectangles

In this section we show that for any $\varepsilon > 0$ the number of $(1-\varepsilon)n \times n$ Latin rectangles is asymptotically

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

Note that a $(1-\varepsilon)n \times n$ Latin rectangle can be viewed as a sequence of $(1-\varepsilon)n$ disjoint $n \times n$ permutation matrices. We count the number of ways to construct such a sequence matrix by matrix. Suppose we have chosen disjoint permutation matrices $P_1, \ldots, P_{i-1}$. Let A_i be the $(0,1)$-matrix of available entries, i.e., $A_i(s,t) = 1$ iff for all $1 \leq j < i$, $P_j(s,t) = 0$. Then A_i has $k(i) = n - i + 1$ ones in each row and column, and P_i can be any permutation matrix supported by A_i. By the permanent bounds of Egorychev–Falikman [7, 8] and Brégman [5], the permanent of an $n \times n$ $(0,1)$-matrix M with k ones in each row and column satisfies:

$$\left(\frac{k}{e}\right)^n \leq Per(M) \leq (k!)^{n/k}.$$

Thus, the number of choices for the whole process is at least:

$$\prod_{k=\varepsilon n}^{n}\left(\frac{k}{e}\right)^n = \left(\frac{1}{e}\right)^{(1-\varepsilon)n^2}\left(\frac{n!}{(\varepsilon n)!}\right)^n \geq \left(\frac{1}{e}\right)^{(1-\varepsilon)n^2}\frac{(n/e)^{n^2}}{(\varepsilon n/e)^{\varepsilon n^2}}$$
$$= \left(\left(\frac{1}{\varepsilon}\right)^{\varepsilon/(1-\varepsilon)}\frac{n}{e^2}\right)^{(1-\varepsilon)n^2}.$$

On the other hand, the number of choices is bounded from above by

$$\prod_{k=\varepsilon n}^{n}(k!)^{n/k} = \prod_{k=\varepsilon n}^{n}\left((1+o(1))\frac{k}{e}\right)^n = \left((1+o(1))\left(\frac{1}{\varepsilon}\right)^{\varepsilon/(1-\varepsilon)}\frac{n}{e^2}\right)^{(1-\varepsilon)n^2},$$

as desired.

References

[1] Noga Alon, Vojtech Rödl, and Andrzej Rucinski, *Perfect matchings in ϵ-regular graphs*, the electronic journal of combinatorics **5** (1998), no. R13, 1–4.

[2] Lina J Andrén, Carl Johan Casselgren, and Lars-Daniel Öhman, *Avoiding arrays of odd order by latin squares*, Combinatorics, Probability and Computing **22** (2013), no. 2, 184–212.

[3] Tom Bohman, Alan Frieze, and Eyal Lubetzky, *Random triangle removal*, Advances in Mathematics **280** (2015), 379–438.

[4] Béla Bollobás, *Random graphs*, Modern Graph Theory, Springer, 1998, pp. 215–252.

[5] Lev M Bregman, *Some properties of nonnegative matrices and their permanents*, Soviet Math. Dokl, vol. 14, 1973, pp. 945–949.

[6] Carl Johan Casselgren and Roland Häggkvist, *Coloring complete and complete bipartite graphs from random lists*, Graphs and Combinatorics **32** (2016), no. 2, 533–542.

[7] Gregory P Egorychev, *The solution of van der waerden's problem for permanents*, Advances in Mathematics **42** (1981), no. 3, 299–305.

[8] Dmitry I Falikman, *Proof of the van der waerden conjecture regarding the permanent of a doubly stochastic matrix*, Mathematical notes of the Academy of Sciences of the USSR **29** (1981), no. 6, 475–479.

[9] Ashish Goel, Michael Kapralov, and Sanjeev Khanna, *Perfect matchings via uniform sampling in regular bipartite graphs*, ACM Transactions on Algorithms (TALG) **6** (2010), no. 2, 27.

[10] Peter Keevash, *The existence of designs*, arXiv preprint arXiv:1401.3665 (2014).

[11] ———, *Counting designs*, arXiv preprint arXiv:1504.02909 (2015).

[12] Michael Krivelevich, *Perfect fractional matchings in random hypergraphs*, Random Structures & Algorithms **9** (1996), no. 3, 317–334.

[13] Matthew Kwan and Benny Sudakov, *Intercalates and discrepancy in random latin squares*, Random Structures & Algorithms (2017).

[14] Erik Ordentlich and Ron M Roth, *Two-dimensional weight-constrained codes through enumeration bounds*, IEEE Transactions on Information Theory **46** (2000), no. 4, 1292–1301.

[15] Jacobus Hendricus Van Lint and Richard Michael Wilson, *A course in combinatorics*, Cambridge university press, 2001.

[16] Nicholas C Wormald, *The differential equation method for random graph processes and greedy algorithms*, Lectures on approximation and randomized algorithms (1999), 73–155.

Israel Institute for Advanced Studies.
Email address: zluria@gmail.com

Institute of Mathematics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Israel.
Email address: menahem.simkin@mail.huji.ac.il

Chapter 4

Perfect matchings in random subgraphs of regular bipartite graphs

Roman Glebov, Zur Luria and Michael Simkin. Submitted for publication.

PERFECT MATCHINGS IN RANDOM SUBGRAPHS OF REGULAR BIPARTITE GRAPHS

ROMAN GLEBOV, ZUR LURIA, AND MICHAEL SIMKIN

Abstract. Consider the random process in which the edges of a graph G are added one by one in a random order. A classical result states that if G is the complete graph K_{2n} or the complete bipartite graph $K_{n,n}$, then typically a perfect matching appears at the moment at which the last isolated vertex disappears. We extend this result to arbitrary k-regular bipartite graphs G on $2n$ vertices for all $k = \omega\left(\frac{n}{\log^{1/3} n}\right)$.

Surprisingly, this is not the case for smaller values of k. Using a construction due to Goel, Kapralov and Khanna, we show that there exist bipartite k-regular graphs in which the last isolated vertex disappears long before a perfect matching appears.

1. Introduction

The study of the random graph model $G(n;p)$ began with two influential papers by Erdős and Rényi [7, 8]. In [7] and [9], they considered the range $p = \Theta(\log n/n)$ and the appearance of spanning structures in that regime. Later, several papers [1, 3, 14, 15, 16, 21] led to the following understanding. Consider a random graph process on n vertices, in which edges are added one by one in a random order. Asymptotically almost surely[1], the first edge that makes the minimum degree one connects the graph, and creates a perfect matching. Likewise, when the minimum degree becomes two, the graph immediately contains a Hamilton cycle. Philosophically, spanning structures appear once local obstructions disappear.

For a graph $G = (V, E)$ and $p \in [0, 1]$, let $G(p)$ denote the distribution on subgraphs of G in which each edge is retained with probability p, independently of the other edges. Recently, a series of papers [17, 12, 18, 22] extended the above philosophy to $G(p)$ for various G. For example, in [17] it was shown that if G is a Dirac graph, then the threshold for Hamiltonicity of $G(p)$ remains $\Theta(\log n/n)$. See [23] for a survey of these and related results.

In this paper we consider the threshold p_0 for the appearance of a perfect matching in $G(p)$ where G is a k-regular bipartite graph on $2n$ vertices. The celebrated permanent inequalities of Bregman [5] and Egorychev–Falikman [6, 10] imply that the number of perfect matchings in G is $\left((1+o(1))\frac{k}{e}\right)^n$. In particular, this number depends little on the specific structure of G. It is therefore natural to conjecture that p_0 depends only on n and k. Furthermore, the logical candidate is the threshold for the disappearance of isolated vertices in $G(p)$, which is $p = \Theta(\log n/k)$.

[1] An event occurs "asymptotically almost surely" (a.a.s.) if the probability of its occurrence tends to 1 as $n \to \infty$. We say that a property holds for "almost every" element of a set if it holds a.a.s. for a uniformly random element of the set.

Indeed, Goel, Kapralov and Khanna [13, Theorem 2.1] showed that there exists a constant c such that for any $k \leq n$, if $p = cn \log n/k^2$, then with high probability $G(p)$ contains a perfect matching. In particular, if $k = \Omega(n)$, $p = O(\log n/k)$ suffices.

For $k = \omega\left(\frac{n}{\log^{1/3} n}\right)$ we considerably strengthen this result. Namely, we show that if one reconstructs G by adding its edges one by one in a random order, then typically a perfect matching appears at the same moment that the last isolated vertex vanishes. As a consequence, it follows that for any $C > 1$, if $p = C\log(n)/k$, then with high probability $G(p)$ contains a perfect matching.

Formally, a **graph process in** $G = (V, E)$ is a sequence of graphs

$$(V, \emptyset) = G_0, G_1, \ldots, G_{|E|} = G$$

on the vertex set V, where for each i, G_i is obtained from G_{i-1} by adding a single edge of G. The **hitting time** of a monotone graph property P with respect to a graph process is $\min\{t : G_t \in P\}$.

For a graph process $\tilde{G}$, let $\tau_M(\tilde{G})$ and $\tau_I(\tilde{G})$ denote the hitting times for containing a perfect matching and having no isolated vertices, respectively. Clearly, for every graph process $\tilde{G}$ we have $\tau_M(\tilde{G}) \geq \tau_I(\tilde{G})$. Our main result is that if G is sufficiently dense and $\tilde{G}$ is chosen uniformly at random, equality a.a.s. holds.

Theorem 1.1. *Let $k = \omega\left(\frac{n}{\log^{1/3} n}\right)$, let G be a k-regular bipartite graph on $2n$ vertices, and let $\tilde{G}$ be a uniformly random graph process in G. Then, a.a.s. $\tau_M(\tilde{G}) = \tau_I(\tilde{G})$.*

Corollary 1.2. *For G and k as above,*

- *If $p = \frac{\log n - \omega(1)}{k}$, then a.a.s. $G(p)$ does not contain a perfect matching.*
- *If $p = \frac{\log n + \omega(1)}{k}$, then a.a.s. $G(p)$ contains a perfect matching.*

Quite surprisingly, it turns out that these results fail when k is significantly smaller than $n/\log^{1/3} n$. We analyze a construction of Goel, Kapralov, and Khanna [13] in which the threshold for a perfect matching is much larger than the threshold for the disappearance of isolated vertices.

Proposition 1.3. *There exist infinitely many k-regular bipartite graphs G on n vertices, with $k = \Omega\left(\frac{n}{\log(n)\cdot\log(\log(n))}\right)$, such that a.a.s. the random subgraph $G(p)$ does not contain a perfect matching for any $p \leq 2\log n/k$. On the other hand, if $p = (\log n + \omega(1))/k$, then a.a.s. $G(p)$ contains no isolated vertices.*

We prove Proposition 1.3 in Appendix A.

Theorem 1.1 is almost a triviality if one assumes that G is pseudorandom (cf. [20, Lemma 3.1]). The main element needed in our proof is a way to control induced subgraphs of G with high discrepancy. To this end we prove a result on the structure of high discrepancy sets in sufficiently dense, regular, bipartite graphs (Lemma 2.4).

The remainder of this paper is organized as follows. Section 1.1 introduces our notation. In Section 2 we prove Lemma 2.4, and in Section 3 we establish some probabilistic tools. Finally, in Section 4, we prove Theorem 1.1.

1.1. Notation.

Throughout the paper, we disregard floor and ceiling signs to improve readability. Large real numbers should be rounded to the nearest integer. We denote by "log" the natural logarithm.

For an integer $m \in \mathbb{N}$, we define $[m] = \{1, 2, \ldots, m\}$. Let X be a set and let $f : [|X|] \to \mathbb{R}$. We sometimes abuse notation by writing

$$\sum_{|S|=1}^{m} \binom{|X|}{|S|} f(|S|) = \sum_{S \subseteq X : |S| \in [m]} f(|S|).$$

Let $f, g : \mathbb{N} \to \mathbb{R}$. We write $f = \tilde{O}(g)$ if, for some $c > 0$ and all large enough $n \in \mathbb{N}$, $f(n) \le g(n) \log^c (g(n))$.

Let $G = (V, E)$ be a graph. For $A, B \subseteq V$, denote by $E_G(A, B)$ the set of edges incident to both A and B, and let $e_G(A, B) = |E_G(A, B)|$. Let $N_G(A)$ denote the set of **neighbors** of A, i.e., the set $\{v \in V : \exists a \in A \text{ s.t. } av \in E\} \setminus A$. We define $G \setminus A$ to be the induced graph on the vertex set $V(G) \setminus A$.

Suppose G is a bipartite graph with vertex partition X, Y. A vertex set A is **partite** if $A \subseteq X$ or $A \subseteq Y$. We denote by A^c the complement of A w.r.t. its own part, i.e., $X \setminus A$ if $A \subseteq X$ and $Y \setminus A$ if $A \subseteq Y$. If A is empty, it will be clear from context whether $A^c = X$ or $A^c = Y$.

By a common abuse of notation, we speak of $G(p)$ as having a certain property, instead of saying that $G \sim G(p)$ has that property.

In certain places we will need to show that events not only occur a.a.s., but that the probability of their non-occurence decays at a polynomial rate. We will say that such events occur **with very high probability (w.v.h.p.)**. Formally, we say that a sequence of events $\{A_n\}_{n\in\mathbb{N}}$ occurs w.v.h.p. if $\log\left(\mathbb{P}[A_n^c]\right) = -\Omega\left(\log n\right)$.

2. A Structural Lemma

Throughout this section $G = (X \dot\cup Y, E)$ is a k-regular bipartite graph on $2n$ vertices. A **cut** in G is a pair (S, T) where $S \subseteq X$ and $T \subseteq Y$. We call (S, T) a **Hall cut** if $|S| > |T|$ and $N(S) \subseteq T$. Hall's marriage theorem states that a balanced bipartite graph contains a perfect matching if and only if it contains no Hall cuts. The main idea in the proof of Theorem 1.1 is to show that a.a.s. G_{τ_I} does not contain a Hall cut.

Let (S, T) be a cut in G. We call $E_G(S, T^c)$ the **outgoing edges** of (S, T). The **cross edges** of G with respect to (S, T) are those in $E(S \cup T, S^c \cup T^c)$. We call the remaining edges **parallel**. For a vertex $x \in V(G)$, we denote by $\deg^{\mathrm{Par}}_{G,S,T}(x)$ and $\deg^{\mathrm{Cr}}_{G,S,T}(x)$ the number of parallel and cross edges incident to x, respectively. Similarly, we denote by $N^{\mathrm{Par}}_{G,S,T}(x)$ the set of neighbors of x that are connected to x by a parallel edge. If the cut (S, T) is clear from the context, we sometimes write $\deg^{\mathrm{Par}}_G(x)$ and $\deg^{\mathrm{Cr}}_G(x)$.

We define the following distance function on the set of cuts in G:

$$d((S_1, T_1), (S_2, T_2)) = |S_1 \setminus S_2| + |S_2 \setminus S_1| + |T_1 \setminus T_2| + |T_2 \setminus T_1|.$$

For $C \in \mathbb{R}$, we say that two cuts are **C-close** if their distance is at most C.

Observation 2.1. Let (S, T) be a cut in G. Then $e(S, T^c) = k \cdot (|S| - |T|) + e(S^c, T)$.

Proof. Since G is k-regular we have:

$$e(S, T) + e(S, T^c) = k \cdot |S|$$
$$e(S, T) + e(S^c, T) = k \cdot |T|.$$

Subtracting the second equation from the first yields the result. □

Observation 2.2. Let (S,T) be a cut in G with $|S| > |T|$. Let $C = e_G(S \cup T, S^c \cup T^c)$ be the number of cross edges in G w.r.t. (S,T). Then, for any $p \in (0,1)$, it holds that:

$$\mathbb{P}\left[(S,T) \text{ is a Hall cut in } G(p)\right] \leq (1-p)^{C/2}.$$

Proof. We have $e_G(S \cup T, S^c \cup T^c) = e_G(S, T^c) + e_G(S^c, T)$. As $|S| > |T|$, Observation 2.1 implies that $e_G(S,T^c) > e_G(S^c,T)$ and therefore $e_G(S,T^c) > e_G(S \cup T, S^c \cup T^c)/2$. The probability that none of these cross edges are edges in $G(p)$ (and thus (S,T) is a Hall cut) is therefore bounded from above by $(1-p)^{C/2}$, as desired. □

The following structural lemma is the heart of our proof. Observation 2.2 implies that if G has almost no cuts with few cross edges, a union bound is enough to show that a.a.s. $G(p)$ contains no Hall cuts. This is the case, for example, in random regular graphs. However, in an arbitrary graph this need not hold. Therefore, we must understand the behavior of cuts with few cross edges, and hence a significant chance of being Hall cuts in $G(p)$. We show that in any sufficiently dense, regular, bipartite graph, all such cuts can be grouped into a small (specifically, subpolynomial) number of equivalence classes. This allows us to control the contribution of these cuts to the probability that $G(p)$ contains a Hall cut.

Definition 2.3. Let $c > 0$. A cut (S,T) is **c-internal** if it has at most $4cnk/\log n$ cross edges.

If a cut is 1-internal, we sometimes just say that it is **internal**. Note that (S,T) is c-internal if and only if its complement (S^c,T^c) is c-internal. Indeed, both cuts have the same cross edges.

Lemma 2.4. *Let $G = (X \dot\cup Y, E)$ be a k-regular bipartite graph on $2n$ vertices, with $k = \omega\left(\frac{n}{\log^{1/3} n}\right)$ and n sufficiently large. Set $\varepsilon = \frac{n}{k \log^{1/3}(n)} = o(1)$. There exist $m = 2^{\Theta(n/k)}$ and cuts $(S_1,T_1), \dots, (S_m,T_m)$ with the following properties.*

1. *For every $i \in [m]$ and $x \in V(G)$, we have $\deg^{Cr}_{G,S_i,T_i}(x) \leq (1+\varepsilon)\frac{k}{2}$.*
2. *Every internal cut (S,T) with $|S| > |T|$ is εk-close to (S_i,T_i) for some $i \in [m]$.*

Remark 2.5. For graphs satisfying $k = \Omega(n)$ it is relatively straightforward to derive Lemma 2.4 from Szemerédi's regularity lemma (albeit with a vastly larger bound on m). Alternatively, one could use the decomposition of dense regular graphs into "robust components" (induced subgraphs with good expansion properties) due to Kühn, Lo, Osthus, and Staden [19, Theorem 3.1].

We begin by defining a lattice-like structure on the internal cuts. Let $\delta = k/n$ be the density of G.

Claim 2.6. *Let (S_1,T_1) and (S_2,T_2) be two c-internal cuts. Then*

$$d((S_1,T_1),(S_2,T_2)) \leq \frac{40cn}{\log n} \text{ or } d((S_1,T_1),(S_2,T_2)) \geq k/10.$$

This implies that for sufficiently small c, the distance between two c-internal cuts is either very large or very small. Throughout this section, c will always be bounded by $O(1/\delta^2)$.

Proof. Let $d = d((S_1, T_1), (S_2, T_2))$. Without loss of generality assume that $|S_2 \setminus S_1| \geq d/4$. As (S_1, T_1) is c-internal, we have:

$$e_G(S_2 \setminus S_1, T_1) \leq e_G(S_1^c, T_1) \leq \frac{4cnk}{\log n}.$$

Similarly, since (S_2, T_2) is c-internal, we have:

$$e_G(S_2 \setminus S_1, T_2) = e_G(S_2 \setminus S_1, Y) - e_G(S_2 \setminus S_1, T_2^c) \geq e_G(S_2 \setminus S_1, Y) - \frac{4cnk}{\log n}.$$

Since G is k-regular, we have:

$$e_G(S_2 \setminus S_1, Y) = k\,|S_2 \setminus S_1| \geq \frac{kd}{4}.$$

Therefore:

$$e_G(S_2 \setminus S_1, T_2 \setminus T_1) \geq e_G(S_2 \setminus S_1, T_2) - e_G(S_2 \setminus S_1, T_1) \geq \frac{kd}{4} - \frac{8cnk}{\log n}. \tag{1}$$

On the other hand, it is certainly true that $e_G(S_2 \setminus S_1, T_2 \setminus T_1) \leq |S_2 \setminus S_1|\,|T_2 \setminus T_1|$. As $|S_2 \setminus S_1| + |T_2 \setminus T_1| \leq d$, we have:

$$e_G(S_2 \setminus S_1, T_2 \setminus T_1) \leq \frac{d^2}{4}. \tag{2}$$

Combining (1) and (2) and rearranging yields:

$$d\,(k - d) \leq \frac{32cnk}{\log n}.$$

Suppose $d < k/10$. Then $k - d > 9k/10$. We thus obtain the inequality:

$$d \leq \frac{320cn}{9 \log n} < \frac{40cn}{\log n},$$

as desired. □

We say that two c-internal cuts are **equivalent** if they are $(\varepsilon k/100)$-close. The triangle inequality, together with Claim 2.6, implies that this is an equivalence relation. Let $\mathcal{X}_c$ be the set of equivalence classes of c-internal cuts. We say that a cut is **trivial** if it is equivalent to $(\emptyset, \emptyset)$.

We now define an intersection operation on equivalence classes. Note that if the cuts (S_1, T_1) and (S_2, T_2) are c_1-internal and c_2-internal, respectively, then the cut $(S_1 \cap S_2, T_1 \cap T_2)$ is $(c_1 + c_2)$-internal. This follows from the fact that any cross edge of the intersection is a cross edge in at least one of the cuts.

Denote the equivalence class of a c-internal cut (S, T) by $[(S, T)]_c$. When the value of c is clear from the context, we omit the subscript.

Definition 2.7. Let c_1 and c_2 satisfy $c_1 + c_2 \leq 40/\delta$. The **intersection** between $[(S_1, T_1)] \in \mathcal{X}_{c_1}$ and $[(S_2, T_2)] \in \mathcal{X}_{c_2}$ is

$$[(S_1, T_1)] \cap [(S_2, T_2)] = [(S_1 \cap S_2, T_1 \cap T_2)] \in \mathcal{X}_{c_1 + c_2}.$$

The fact that this is well-defined, in the sense that it does not depend on the choice of representatives, follows from Claim 2.6. Indeed, different choices of representatives may only change the intersection by at most $80(c_1 + c_2)n/\log n$ vertices. This is smaller than $\varepsilon k/100$, and therefore the two intersections are equivalent.

Definition 2.8. Two classes $[(S_1, T_1)] \in \mathcal{X}_{c_1}$ and $[(S_2, T_2)] \in \mathcal{X}_{c_2}$ are **disjoint** if their intersection is trivial. We say that $[(S_1, T_1)]$ **contains** $[(S_2, T_2)]$ if $[(S_1^c, T_1^c)] \cap [(S_2, T_2)]$ is trivial.

Note that if $[(S, T)]$ is a non-trivial class then by Claim 2.6 $|S \cup T| \geq k/10$.

Observation 2.9. If $[(S_1, T_1)]$ is not contained in $[(S_2, T_2)]$, then for every $(S', T') \in [(S_1, T_1)] \cap [(S_2, T_2)]$, we have $|S_1 \cup T_1| - |S' \cup T'| \geq k/20$.

Proof. Since $[(S_1, T_1)]$ is not contained in $[(S_2, T_2)]$, it holds that $(S_1 \setminus S_2, T_1 \setminus T_2)$ is non-trivial, and therefore $|S_1 \setminus S_2| + |T_1 \setminus T_2| \geq k/10$. Note that S_1 is the disjoint union of $S_1 \cap S_2$ and $S_1 \setminus S_2$, and that a similar statement holds for T_1. This implies that $|S_1 \cup T_1| \geq |S_1 \cap S_2| + |T_1 \cap T_2| + k/10$. As (S', T') is equivalent to $(S_1 \cap S_2, T_1 \cap T_2)$, the observation follows. □

Consider the following process:

(1) Initialize $[(S_1, T_1)] \in \mathcal{X}_1$ to be an arbitrary non-trivial internal equivalence class.
(2) As long as there exists a class $[(S^*, T^*)] \in \mathcal{X}_1$ that is neither disjoint from nor containing $[(S_i, T_i)]$, set $[(S_{i+1}, T_{i+1})] = [(S_i, T_i)] \cap [(S^*, T^*)]$.

We call the equivalence classes that can be obtained at the end of this process **atoms**.

Observation 2.10.

(1) The process halts after at most $40/\delta$ steps.
(2) All of the atoms are $(40/\delta)$-internal.
(3) All of the atoms are non-trivial.
(4) The atoms are pairwise disjoint.
(5) There are at most $30/\delta$ atoms.

Proof. Item 1 follows from the fact that $|S_1 \cup T_1| \leq 2n$, and Observation 2.9 implies that for each i, $|S_i \cup T_i| \leq |S_{i-1} \cup T_{i-1}| - k/20$. Item 2 follows from Item 1 and the fact that the intersection of a c-internal cut with a 1-internal cut is $(c+1)$-internal. Item 3 holds because at each stage of the process, $[(S_i, T_i)]$ and $[(S^*, T^*)]$ are not disjoint.

For Item 4, assume that $[(S, T)]$ and $[(S', T')]$ are distinct atoms. Since they are not equivalent, $d((S, T), (S', T')) \geq \varepsilon k/100$. By Claim 2.6 this implies that in fact $d((S, T), (S', T')) \geq k/10$. Without loss of generality, we may assume that $|S \setminus S'| + |T \setminus T'| \geq k/20$. Now, $[(S', T')]$ is the intersection of at most $40/\delta$ internal classes. Therefore at least one of these classes, $[(S^*, T^*)]$, satisfies

$$|S \setminus S^*| + |T \setminus T^*| \geq \frac{k}{20} \cdot \frac{\delta}{40}.$$

Since $k = \omega\left(\frac{n}{\log^{1/3} n}\right)$, this is larger than $\frac{40}{\delta} \cdot \frac{40n}{\log n}$. Therefore, by Claim 2.6, this implies that $[(S, T)] \cap [(S^{*c}, T^{*c})]$ is not trivial, and therefore $[(S, T)]$ is not contained in $[(S^*, T^*)]$. Suppose, for a contradiction, that $[(S, T)]$ and $[(S', T')]$ are not disjoint. Then $[(S, T)]$ and $[(S^*, T^*)]$ are not disjoint. Therefore $[(S, T)]$, by definition, is not an atom, which is a contradiction.

Item 5 is true because there are $2n$ vertices, each atom contains at least $k/10$ vertices, and the atoms are pairwise disjoint. Therefore, by Claim 2.6, all intersections have at most $\frac{1600n^2}{k \log n}$ vertices. Letting A denote the number of atoms, the inclusion-exclusion formula implies that for all $a \leq A$:

$$a \cdot \binom{k}{10} - \binom{a}{2} \frac{1600n^2}{k \log n} \leq 2n.$$

The inequality does not hold for $a = 30/\delta$, and therefore $A \leq 30/\delta$. □

The proof of Lemma 2.4 will make repeated use of the following claim.

Claim 2.11. *Suppose $[(S_A, T_A)]$ is an atom, $[(S_B, T_B)]$ is a nontrivial $O(1/\delta^2)$-internal class, and their intersection is trivial. Then there exists a 1-internal cut (S_C, T_C) such that $[(S_C, T_C)]$ and $[(S_A, T_A)]$ are disjoint and $[(S_C, T_C)] \cap [(S_B, T_B)]$ is non-trivial.*

Proof. Since $[(S_A, T_A)]$ and $[(S_B, T_B)]$ have trivial intersection, $|S_A \cap S_B| + |T_A \cap T_B| < k/10$. Since their intersection is an $O(1/\delta^2)$-internal class, by Claim 2.6:

$$|S_A \cap S_B| + |T_A \cap T_B| = O\left(\frac{n}{\delta^2 \log n}\right).$$

Let $[(S_1, T_1)], \ldots, [(S_\ell, T_\ell)]$ be a sequence of 1-internal classes whose intersection is (S_A, T_A). Observe that for each $i \in [\ell]$, $[(S_i^c, T_i^c)]$ is 1-internal and disjoint from $[(S_A, T_A)]$. Furthermore, for each $i \in [\ell]$, $[(S_i^c, T_i^c)] \cap [(S_B, T_B)]$ is 2-internal (as the intersection of two 1-internal classes). We will show that (at least) one of these intersections is non-trivial. Suppose, for a contradiction, that for every $i \in [\ell]$, $[(S_i^c, T_i^c)] \cap [(S_B, T_B)]$ is trivial. Then, by Claim 2.6, for every $i \in [\ell]$, $|S_B \cap S_i^c| + |T_B \cap T_i^c| \leq \frac{80n}{\log n}$. However, by assumption, $[(S_B, T_B)]$ is non-trivial. We may therefore assume w.l.o.g. that $|S_B| \geq k/20$. Then:

$$O\left(\frac{n}{\delta^2 \log n}\right) \geq |S_B \cap S_A| \geq |S_B| - \sum_{i=1}^{\ell} |S_B \cap S_i^c| \geq \frac{k}{20} - \ell \frac{80n}{\log n}.$$

Together with the fact that $\ell = O(1/\delta)$ this implies that $k = O\left(\frac{n}{\log^{1/3} n}\right)$, a contradiction. □

Proof of Lemma 2.4. We first construct the cuts $(S_1, T_1), \ldots, (S_m, T_m)$. Let $\mathcal{A}$ be the set of atoms. For each $\alpha \in \mathcal{A}$, fix a representative (S_α, T_α). For $\mathcal{S} \subseteq \mathcal{A}$, let $(S'_\mathcal{S}, T'_\mathcal{S})$ be the cut $(\cup_{\alpha \in \mathcal{S}} S_\alpha, \cup_{\alpha \in \mathcal{S}} T_\alpha)$. Finally, define the sets:

$$S_\mathcal{S} := \left\{x \in X : \deg^{\text{Par}}_{G, S'_\mathcal{S}, T'_\mathcal{S}}(x) \geq \frac{k}{2}\right\}, T_\mathcal{S} := \left\{y \in Y : \deg^{\text{Par}}_{G, S'_\mathcal{S}, T'_\mathcal{S}}(y) \geq \frac{k}{2}\right\}.$$

Let $(S_1, T_1), \ldots, (S_m, T_m)$ be a list of the cuts $\{(S_\mathcal{S}, T_\mathcal{S})\}_{\mathcal{S} \subseteq \mathcal{A}}$. By Observation 2.10, $m \leq 2^{|\mathcal{A}|} = 2^{O(n/k)}$. It remains to prove that in each of these cuts, every vertex is incident to few (i.e., less than $(1+\varepsilon)k/2$) cross edges, and that every internal cut (S, T) with $|S| > |T|$ is εk-close to one of the (S_i, T_i)s.

Let $\mathcal{S} \subseteq \mathcal{A}$. Since, by Observation 2.10, every atom is $(40/\delta)$-internal, the number of cross edges with respect to $(S'_\mathcal{S}, T'_\mathcal{S})$ is at most $|\mathcal{A}| \frac{40}{\delta} \frac{4nk}{\log n} = O\left(\frac{n^3}{k \log n}\right)$. Therefore there are at most $\varepsilon k/2$ vertices $x \in V(G)$ s.t. $\deg^{\text{Cr}}_{G, S'_\mathcal{S}, T'_\mathcal{S}}(x) > k/2$. Thus $d\left((S'_\mathcal{S}, T'_\mathcal{S}), (S_\mathcal{S}, T_\mathcal{S})\right) \leq \varepsilon k/2$. Let $x \in V(G)$. By construction,

$$\deg^{\text{Cr}}_{G, S_\mathcal{S}, T_\mathcal{S}}(x) \leq \frac{k}{2} + d\left((S'_\mathcal{S}, T'_\mathcal{S}), (S_\mathcal{S}, T_\mathcal{S})\right) \leq (1+\varepsilon)\frac{k}{2},$$

as desired.

For the second property, let (S,T) be an internal cut. Let $\mathcal{S} = \{\alpha_1, \dots, \alpha_m\}$ be the set of atoms contained in $[(S,T)]$. By the triangle inequality it suffices to show that $(S'_{\mathcal{S}}, T'_{\mathcal{S}})$ is equivalent to (S,T).

Suppose, for a contradiction, that $(S'_{\mathcal{S}}, T'_{\mathcal{S}})$ is not equivalent to (S,T). Define

$$(S^1,T^1) = (S,T) \cap (S'_{\mathcal{S}}, T'_{\mathcal{S}})^c = (S,T) \cap \left(\bigcap_{\alpha \in \mathcal{S}} (S_\alpha, T_\alpha)^c \right).$$

Observe that by construction, (S^1,T^1) does not contain any atoms. Furthermore, the cut (S^1,T^1) is $\left(|\mathcal{A}|\frac{40}{\delta}+1\right) = O(1/\delta^2)$-internal and by assumption non-trivial. Therefore $|S^1|+|T^1| \geq k/10$ by Claim 2.6. We apply Claim 2.11 with $[(S_A,T_A)] = \alpha_1$ and $[(S_B,T_B)] = [(S^1,T^1)]$ to obtain a 1-internal cut (S^1_C, T^1_C) such that $[(S^1_C,T^1_C)]$ and α_1 are disjoint and $[(S^1,T^1)] \cap [(S^1_C,T^1_C)]$ is non-trivial. We set $[(S^2,T^2)] = [(S^1,T^1)] \cap [(S^1_C,T^1_C)]$. Since $[(S^2,T^2)]$ is non-trivial, we may proceed in this fashion; having obtained $[(S^i,T^i)]$ we apply the claim with $[(S_A,T_A)] = \alpha_i$ to obtain (S^i_C,T^i_C) and $[(S^{i+1},T^{i+1})] = [(S^i,T^i)] \cap [(S^i_C,T^i_C)]$. Finally, we obtain $[(S^{m+1},T^{m+1})]$ which is $O(1/\delta^2)$-internal, nontrivial, and contained in $[(S,T)] \cap (\cap_{i=1}^m [(S^i_C,T^i_C)])$. As the nontrivial intersection of 1-internal classes, this cut contains an atom α that is contained in $[(S,T)]$. Furthermore, for each i, $[(S^i_C,T^i_C)]$ and α_i are disjoint. Therefore $\alpha \notin \mathcal{S}$, a contradiction. □

3. Properties of Random Subgraphs

Let $k = \delta n$, with $\delta = \omega(\log^{-1/3} n)$, and fix a k-regular bipartite graph $G = (X \dot\cup Y, E)$ on $2n$ vertices. In this section we collect properties of random subgraphs of G that are essential for our proof.

Set

$$p_1 = \frac{\log n - \log\log\log\log n}{k},$$
$$p_2 = \frac{\log n + \log\log\log\log n}{k}.$$

We define the following random subgraphs of G:

$$G_2 \sim G(p_2),$$
$$G_1 \sim G_2\left(\frac{p_1}{p_2}\right).$$

Observe that $G_1 \sim G(p_1)$. Furthermore, the same distribution on G_1, G_2 can be obtained as follows. Let $G_1 \sim G(p_1)$, $G' \sim G\left(\frac{p_2-p_1}{1-p_1}\right)$, and set $G_2 = G_1 \cup G'$.

We will show presently that a.a.s. G_1 contains isolated vertices, while G_2 does not. Furthermore, the distance between any two vertices that are isolated in G_1 is at least 2 in G_2. This motivates the following construction: let $G_1 \subseteq G_H \subseteq G_2$ be the random graph obtained by adding, for each isolated vertex v in G_1, an edge drawn uniformly at random from $\{e \in E(G_2) : v \in e\}$. If any of these sets are empty, or if there are two isolated vertices in G_1 that are connected in G_2, set $G_H = G_1$. The next claim, a variation of Lemma 7.9 in [4], establishes that it is sufficient to prove that a.a.s. G_H contains a perfect matching.

Claim 3.1. *Let Q be a monotone increasing property of subgraphs of G. If Q holds a.a.s. for G_H then, in almost every graph process in G, Q holds for G_{τ_I}, the first graph in which there are no isolated vertices.*

We defer the proof until after establishing some properties of G_1 and G_2.

Claim 3.2. *A.a.s. G_1 contains isolated vertices and G_2 does not. Furthermore, a.a.s. there is no pair x, y of vertices that are isolated in G_1 and $xy \in E(G_2)$.*

Proof. The probability that a specific vertex is isolated in $G(p)$ is $(1-p)^k$. The expected number of isolated vertices in G_2 is therefore:

$$2n(1-p_2)^k \leq 2n \exp\left(-p_2 k\right) = 2n\frac{1}{\omega(n)} = o\left(1\right).$$

Applying Markov's inequality, a.a.s. G_2 contains no isolated vertices.

By a similar calculation, the probability that a specific vertex is isolated in G_1 is $\omega(1/n)$. Let the random variable I be the number of vertices in X that are isolated in G_1. Then $\mathbb{E}\left[I\right] = \omega\left(1\right)$. Furthermore, the events that two vertices $x, y \in X$ are each isolated in G_1 are independent. Thus $\operatorname{Var}\left[I\right] \leq \mathbb{E}\left[I\right]$, and by Chebychev's inequality, a.a.s. $I > 0$ and G_1 contains isolated vertices.

For the second part of the claim, observe that by the calculations above a.a.s. the number of isolated vertices in G_1 is $O(\log n)$. Therefore the expected number of edges between these vertices in G_2 is $o\left(1\right)$, and so by Markov's inequality a.a.s. there are none. □

Proof of Claim 3.1. We describe a coupling that relates G_1, G_2 and G_H to G_{τ_I}. Consider the following random process. For each edge e of G, choose a real number $\alpha_e \sim U[0,1]$ uniformly at random from the interval $[0,1]$, all choices independent. Let G_1' and G_2' be the subgraphs of G whose edges are $E(G_i') = \{e \in E(G) : \alpha_e \leq p_i\}$ for $i \in \{1,2\}$. Let G_H' be the random graph obtained by adding, for each isolated vertex in G_1', the edge incident to it in G_2' whose α value is minimal. If G_2' contains isolated vertices or an edge between two vertices that are isolated in G_1', set instead $G_H' = G_1'$.

Observe that the distributions of (G_1, G_2, G_H) and (G_1', G_2', G_H') are identical. Furthermore, the distribution of the random graph process is identical to that of the process in which edges are revealed in increasing order of α. With respect to this process, a.a.s. G_H' is a subgraph of G_{τ_I}. Therefore, if Q holds a.a.s. for G_H, then Q holds a.a.s. for G_H', and as Q is monotone increasing, a.a.s. $G_{\tau_I} \in Q$.

□

For a cut (S,T) we define the set

$$\Gamma\left(S,T\right) = \left\{x \in V : \deg^{\mathrm{Cr}}_{G,S,T}(x) \geq n^{-1/20}k\right\}$$

of vertices with high cross degree. We remind the reader that a sequence of events occurs with very high probability (w.v.h.p.) if the probabilities of their non-occurence decay at a polynomial rate.

Lemma 3.3. *Let (S,T) be a cut. Suppose $R \subseteq V \setminus \Gamma(S,T)$ is a set of $O(\log n)$ isolated vertices in G_1. Then w.v.h.p. for every $x \in R$, $\deg^{Cr}_{G_H}(x) = 0$.*

Proof. It suffices to show that a.a.s. for every $x \in R$, $\deg^{\mathrm{Cr}}_{G_2}(x) = 0$. Indeed, the expected number of cross edges incident to x in G_2 is bounded above by

$n^{-1/20}k\frac{p_2-p_1}{1-p_1} = \tilde{O}\left(n^{-1/20}\right)$. The conclusion follows by applying Markov's inequality and a union bound over the $O(\log n)$ vertices. □

Lemma 3.4. *Let (S,T) be a cut, and let $V' = V \setminus \Gamma(S,T)$. Then, w.v.h.p. for every $x \in V'$, $\deg_{G_1}^{Cr}(x) \le 30$.*

Proof. Observe that $\deg_{G_1}^{\mathrm{Cr}}(x) \sim Bin\left(\deg_G^{\mathrm{Cr}}(x), p_1\right)$. Therefore:

$$\mathbb{P}\left[\deg_{G_1}^{\mathrm{Cr}}(x) \ge 30\right] \le \binom{n^{-1/20}k}{30} p_1^{30} = \tilde{O}\left(\frac{1}{n^{3/2}}\right).$$

The lemma follows by applying a union bound over all $O(n)$ vertices in V'. □

Lemma 3.5. *Let (S,T) be a cut, and let V_{low} be the set of vertices $x \in V \setminus \Gamma(S,T)$ s.t. $\deg_{G_1}^{Par}(x) \le \frac{1}{1000}\log n$. W.v.h.p. the following hold:*

(1) $|V_{low}| \le n^{0.01}$.

(2) For each $x, y \in V_{low}$, the distance between x and y in G_H is at least 6.

Proof. We first show that a.a.s. $|V_{low}| \le n^{0.01}$. Indeed, suppose $x \in V$ satisfies $\deg_G^{\mathrm{Cr}}(x) < n^{-1/20}k$. The probability that $x \in V_{low}$ is at most

$$\sum_{i=0}^{\frac{1}{1000}\log n} \Pr\left[\deg_{G_1}^{\mathrm{Par}} = i\right] \le \log n \binom{k}{\frac{1}{1000}\log n} p_1^{\frac{1}{1000}\log n}(1-p_1)^{\left(1-O(n^{-1/20})\right)k}$$
$$\le \left(\frac{1.1e}{\frac{1}{1000}}\right)^{\frac{1}{1000}\log n} \cdot \tilde{O}\left(\frac{1}{n}\right) < n^{-0.991}.$$

Thus, $\mathbb{E}\left[|V_{low}|\right] < n^{0.009}$. Therefore, by Markov's inequality, $\mathbb{P}\left[|V_{low}| \ge n^{0.01}\right] \le n^{-0.001}$.

The proof of 2 is similar to the proof of [2, Claim 4.4] and property (P2) in [11, Lemma 5.1.1]. Fix two distinct vertices $u, w \in V \setminus \Gamma(S,T)$ and consider a path $(u = v_0, \ldots, v_r = w)$ in G, where $1 \le r \le 5$. Denote by $\mathcal{A}$ the event that for every $0 \le i \le r-1$, we have $\{v_i, v_{i+1}\} \in E(G_2)$, i.e., the path exists in G_2. Denote by $\mathcal{B}$ the event that $u, w \in V_{low}$. Clearly, $\mathbb{P}[\mathcal{A}] = p_2^r$, hence

$$\mathbb{P}[\mathcal{B} \wedge \mathcal{A}] = p_2^r \cdot \mathbb{P}[\mathcal{B}|\mathcal{A}].$$

Let X denote the random variable which counts the number of parallel edges in G_2 incident with u or w disregarding the pairs $\{u, v_1\}$, $\{v_{r-1}, w\}$, and $\{u, w\}$. Observing that $X \sim \mathrm{Bin}\left((1-o(1))2k, p_2\right)$ and using standard concentration inequalities, we have

$$\mathbb{P}[\mathcal{B}|\mathcal{A}] \le \mathbb{P}\left[X < 2\frac{1}{1000}\log n\right] < n^{-1.8}.$$

Fixing the two endpoints u, w, the number of such sequences is at most k^{r-1}. Applying a union bound over all pairs of vertices and possible paths between them, we conclude that the probability of a path in G_2 of length $r \le 5$, connecting two distinct vertices of V_{low} is at most

$$\sum_{r=1}^{5} n^2 \cdot k^{r-1} \cdot p_2^r \cdot n^{-1.8} = \tilde{O}\left(\frac{1}{n^{0.8}}\right).$$

This completes the proof of the lemma. □

Recall that a set $A \subset V$ is **partite** if $A \subset X$ or $A \subset Y$. A vertex is a **parallel neighbor** of A if it is connected to a vertex in A via a parallel edge.

Lemma 3.6. *Let (S,T) be a cut. W.v.h.p. the following holds. If $A \subseteq V$ is a partite set satisfying*

- $|A| \le n^{0.9}$, *and*
- *for every $x \in A$, $\deg_{G_1}^{Par}(x) \ge \frac{1}{1000}\log n$,*

then A has at least $|A|\frac{1}{2000}\log n$ parallel neighbors in G_1.

Proof. Let $A \subseteq V$ be a partite set, and let $t = t(A) = |A|\frac{1}{2000}\log n$. Let $\mathcal{P}(A)$ be the event that the minimum parallel degree of a vertex in A is at least $\frac{1}{1000}\log n$.

For any fixed set B,

$$\mathbb{P}\left[N_{G_1}(A) \subseteq B \wedge \mathcal{P}(A)\right] \le \mathbb{P}\left[e_{G_1}(A,B) \ge 2t\right] \le \binom{e_G(A,B)}{2t} p_1^{2t}$$
$$\le \binom{|A||B|}{2t} p_1^{2t} \le \left(\frac{e \cdot |A||B| p_1}{2t}\right)^{2t}.$$

Applying a union bound, we have:

$$\mathbb{P}\left[\exists A \text{ s.t. } |A| \le n^{0.9} \wedge \mathcal{P}(A) \wedge |N_{G_1}(A)| \le t(A)\right]$$
$$\le 2\sum_{|A|=1}^{n^{0.9}} \binom{n}{|A|}\binom{n}{t(A)}\left(\frac{e|A|t(A)p_1}{2t(A)}\right)^{2t(A)} \le 2\sum_{|A|=1}^{n^{0.9}}\left(\frac{ne}{|A|}\right)^{|A|}\left(\frac{ne^3|A|^2p_1^2}{4t(A)}\right)^{t(A)}$$
$$\le 2\sum_{|A|=1}^{n^{0.9}}\left(\frac{ne}{|A|}\right)^{|A|}\left(\frac{e^3|A|^2\log^2(n)}{4\delta^2 t(A) n}\right)^{t(A)} \le \sum_{|A|=1}^{n^{0.9}} n^{-t(A)/20} = O\left(\frac{1}{n^{1/20}}\right).$$

□

4. Proof of Theorem 1.1

4.1. **Outline.** As mentioned previously, we prove Theorem 1.1 by showing that a.a.s. G_H does not contain a Hall cut. This is similar to the approach used in [9] to show that $p = \log n/n$ is the threshold for $K_{n,n}(p)$ to contain a perfect matching. There, a union bound over all cuts (S,T) (satisfying certain conditions) was sufficient for the result. In this regard, the crucial property of $K_{n,n}$ is that every cut has many outgoing edges. Essentially the same approach was utilized in the proof of [20, Lemma 3.1] to show that if a k-regular, bipartite graph G satisfies a certain expansion property, then the threshold for $G(p)$ to contain a perfect matching is $p = \Theta(\log n/k)$. However, for arbitrary G, there may be many cuts (S,T) with few outgoing edges, potentially foiling the union bound. Indeed, this is the case in the counterexamples described in Appendix A.

To overcome this we take a more delicate approach, wherein we group the various cuts in G into families that we treat separately. Informally, the steps are as follows:

(1) The first family contains all cuts that are not internal (in the sense of Definition 2.3), and therefore have many outgoing edges in G. Here a simple union bound suffices to show that a.a.s. none of these cuts are Hall cuts in G_H (Claim 4.1).

(2) At this point we apply Lemma 2.4 to conclude that any cut not covered in the previous step is close to one of the $m = 2^{\Theta(n/k)}$ cuts from the lemma. We fix one of these cuts, (S', T'), and show that conditioned on G_H having no isolated vertices, w.v.h.p. none of the cuts that are εk-close to (S', T') become a Hall cut in G_H. As m is subpolynomial, a union bound implies that a.a.s. G_H does not contain a Hall cut.

For a cut (S, T), we define the set of **shifted** vertices:

$$\Delta = \Delta(S, T) = (S \setminus S') \cup (S' \setminus S) \cup (T \setminus T') \cup (T' \setminus T)\,.$$

We make use of the natural correspondence $\Delta \leftrightarrow (S, T)$, and interchange between them freely. We recall the definition of the set

$$\Gamma = \Gamma(S', T') = \left\{x \in V : \deg^{\mathrm{Cr}}_{G,S',T'}(x) \geq n^{-1/20}k\right\}$$

of vertices with many cross edges in G w.r.t. (S', T'). We emphasize that Γ is deterministic, i.e., depends only G and (S', T').

(3) The second family of cuts consists of those with $|\Delta| \geq n^{0.9}$. The insight here is that shifting a large number of vertices w.r.t. (S', T') creates many cross edges. Here too, a union bound suffices to show that w.v.h.p. none of these are Hall cuts in G_H (Claim 4.2).

(4) The third family consists of cuts satisfying $|\Delta| \leq \frac{n^{-1/20}}{10}|\Gamma|$. As the vertices in Γ have, by definition, many cross edges, if Δ is much smaller than Γ then most of these cross edges are unaffected. This allows us to employ a union bound here as well (Claim 4.3).

(5) We now argue that w.v.h.p. we may remove from G_H a small matching M covering all vertices that have low degree in G_1, leaving the residual graph $\widetilde{G}_H = G_H \setminus V(M)$. In Observation 4.8, we show that if G_H contains a Hall cut that is C-close to (S', T'), then $\widetilde{G}_H$ contains a Hall cut that is C-close to

$$(S' \setminus V(M), T' \setminus V(M)).$$

It therefore suffices to consider cuts in $\widetilde{G}_H$.

(6) It remains to consider Δ such that $\frac{n^{-1/20}}{10}|\Gamma| < |\Delta| < n^{0.9}$. We show that w.v.h.p. there is no such Hall cut if either $|\Delta| \geq \log n/\log\log n$ or $\Delta \cap \Gamma \neq \emptyset$ (Claim 4.11). Here we take advantage of the fact that once the low degree vertices have been removed from G_1, the remaining vertices satisfy an expansion property (Claim 4.9).

(7) Finally, we argue that w.v.h.p. there is no Hall cut satisfying $|\Delta| \leq \log n/\log\log n$ and $\Delta \cap \Gamma = \emptyset$ (Claim 4.12). Here we use the expansion property to show that for such a cut to exist, w.v.h.p. (S', T') contains many outgoing edges in G_1, and that it is impossible to make all these edges parallel by shifting only $\log n/\log\log n$ vertices.

4.2. **The proof.** We first show that if a cut has many outgoing edges, the probability that it is a Hall cut in G_H is very small.

Claim 4.1. *A.a.s. G_H contains no Hall cut (S, T) that is not internal.*

Proof. Since $G_1 \subseteq G_H$ it suffices to prove the statement with G_H replaced by G_1. Suppose (S, T) is not internal. Then it has at least $4nk/(\log n)$ cross edges. By

Observation 2.2 the probability that it is a Hall cut is less than

$$(1-p_1)^{e_G(S,T^c)} \leq \exp\left(-p_1\frac{2nk}{\log n}\right) = \exp(-(1-o(1))2n) = o\left(4^{-n}\right).$$

The claim follows by applying a union bound over all 4^n cuts in G. □

We now apply Lemma 2.4 to obtain the cuts $(S_1,T_1)\ldots,(S_m,T_m)$. By Claim 4.1, Lemma 2.4, and the fact that m is subpolynomial, it suffices to show that w.v.h.p. all cuts (S,T) that are εk-close to (S_i,T_i) for some i are not Hall cuts in G_H.

Fix an index $i\in[m]$, set $S'=S_i, T'=T_i$, and define Γ and Δ with respect to (S',T') as in the outline. Henceforth, cross edges, parallel edges, cross degrees and parallel degrees are with respect to (S',T').

Claim 4.2. *W.v.h.p. for every Δ such that $n^{0.9}\leq|\Delta|\leq\varepsilon k$, (S,T) is not a Hall cut in G_H.*

Proof. By Lemma 2.4, each $x\in\Delta$ satisfies $\deg_G^{\mathrm{Par}}(x)\geq(1-\varepsilon)\frac{k}{2}$. Most of these parallel edges - all those with an endpoint not in Δ - are cross edges w.r.t. (S,T). Thus the number of cross edges satisfies:

$$e_G(S,T^c)+e_G(S^c,T)\geq(1-\varepsilon)|\Delta|\frac{k}{2}-|\Delta|^2\geq\frac{|\Delta|k}{3}.$$

By Observation 2.2 the probability that (S,T) is a Hall cut in G_1 is at most:

$$(1-p_1)^{|\Delta|k/6}\leq\left(\frac{1}{n}\right)^{|\Delta|/7}.$$

Applying a union bound, the probability that there exists such a Hall cut is at most

$$\sum_{|\Delta|=n^{0.9}}^{\varepsilon n}\binom{2n}{|\Delta|}\left(\frac{1}{n}\right)^{|\Delta|/7}=O\left(\frac{1}{n}\right).$$

□

Claim 4.3. *W.v.h.p. for all Δ s.t. $|\Delta|\leq\frac{n^{-1/20}}{10}|\Gamma|$, the corresponding cut (S,T) is not a Hall cut in G_H.*

Proof. Suppose Δ satisfies the claim's hypothesis. By Lemma 2.4 and the definition of Γ, each $x\in\Gamma$ satisfies

$$\min\left\{\deg_G^{\mathrm{Cr}}(x),\deg_G^{\mathrm{Par}}(x)\right\}\geq n^{-1/20}k.$$

Ignoring, for the moment, the possibility that $N_G(x)\cap\Delta\neq\emptyset$, this means that every $x\in\Gamma$ is incident to at least $n^{-1/20}k$ cross edges w.r.t. (S,T), regardless of whether $x\in\Delta$. There are at most $|\Delta|\min\{|\Gamma|,k\}$ edges between Δ and Γ. Accounting for possible double counting of the edges incident to Γ, we obtain:

$$e_G(S,T^c)+e_G(S^c,T)\geq\frac{|\Gamma|n^{-1/20}k}{2}-|\Delta|\min\{|\Gamma|,k\}\geq4|\Delta|k.$$

Applying Observation 2.2, the probability that (S,T) is a Hall cut in G_1 is at most

$$(1-p_1)^{4|\Delta|k/2}\leq\left(\frac{1}{n}\right)^{1.9|\Delta|}.$$

We now observe that if $\Delta=\emptyset$ (i.e., $(S,T)=(S',T')$), then the probability that (S,T) is a Hall cut in G_1 is at most $(1-p_1)^k=\tilde{O}(1/n)$. Let X be the number of cuts satisfying the claim's hypothesis that are Hall cuts in G_1. Applying a union bound, we have:

$$\mathbb{P}\left[X > 0\right] \leq \tilde{O}\left(\frac{1}{n}\right) + \sum_{1 \leq |\Delta| \leq \min\left\{\frac{n^{-1/20}}{10}|\Gamma|, n^{0.9}\right\}} \binom{2n}{|\Delta|}\left(\frac{1}{n}\right)^{1.9|\Delta|} = O\left(\frac{1}{n^{0.9}}\right).$$

□

Remark 4.4. As a consequence of Claim 4.3, if $|\Gamma| \geq 10n^{19/20}$ then w.v.h.p. none of the cuts that are εk-close to (S', T') become Hall cuts in G_H. This is because Claim 4.2 covers all cases where $|\Delta| \geq n^{0.9}$, and the previous claim covers all cases where $|\Delta| \leq \frac{n^{-1/20}}{10}|\Gamma|$. Therefore, we proceed under the assumption that $|\Gamma| < 10n^{19/20}$.

Before continuing to steps 6 and 7, we modify G_H by removing a small matching covering the low degree vertices that are not in Γ. Moreover, this matching contains only parallel edges. The following claim, together with Lemma 3.5, implies that conditioned on G_H having no isolated vertices, w.v.h.p. such a matching exists.

Claim 4.5. *Conditioned on there being no isolated vertices in G_H, w.v.h.p. every vertex in $V \setminus \Gamma$ is incident to at least one parallel edge in G_H.*

Proof. Assuming there are no isolated vertices in G_H, if there exists some $v \in V \setminus \Gamma$ s.t. $\deg^{\text{Par}}_{G_H}(v) = 0$, then $\deg^{\text{Par}}_{G_1}(v) = 0$ and $\deg^{\text{Cr}}_{G_2}(v) > 0$. We use the first moment method to show that w.v.h.p. there are no vertices $v \notin \Gamma$ for which this holds. Indeed, if $v \notin \Gamma$ then the probability of this occurring is bounded from above by

$$\frac{k}{n^{1/20}} p_2 (1 - p_1)^{k\left(1 - n^{-1/20}\right)} = \tilde{O}\left(\frac{1}{n^{21/20}}\right).$$

Therefore the expected number of such vertices is $\tilde{O}\left(n^{-1/20}\right)$. By Markov's inequality, w.v.h.p. there are none. □

Recall that

$$V_{low} = \left\{x \in V \setminus \Gamma : \deg^{\text{Par}}_{G_1}(x) \leq \frac{1}{1000}\log n\right\}.$$

Conditioning on the conclusions of Lemma 3.5 and Claim 4.5 holding, there exists a matching $M \subseteq G_H$ of size $|V_{low}|$ consisting of parallel edges that contains V_{low}.

Claim 4.6. *W.v.h.p. $N_{G_H}(V_{low}) \cap \Gamma = \emptyset$.*

Proof. By Remark 4.4 we may assume $|\Gamma| < n^{0.96}$. Fix an arbitrary vertex $x \notin \Gamma$. Then

$$\begin{aligned}
\mathbb{P}\left[x \in V_{low} \wedge N_{G_H}(x) \cap \Gamma \neq \emptyset\right] &\leq \sum_{y \in \Gamma} \mathbb{P}\left[x \in V_{low} \wedge y \in N_{G_H}(x)\right] \\
&\leq \sum_{y \in \Gamma} \mathbb{P}\left[\left|N^{Par}_{G_1}(x) \setminus \{y\}\right| \leq \frac{1}{1000}\log n \wedge y \in N_{G_2}(x)\right] \\
&= \sum_{y \in \Gamma} \mathbb{P}\left[\left|N^{Par}_{G_1}(x) \setminus \{y\}\right| \leq \frac{1}{1000}\log n\right] \cdot \mathbb{P}\left[y \in N_{G_2}(x)\right] \\
&\leq |\Gamma|\left(\frac{1}{n}\right)^{0.99} \cdot p_2 = O\left(\frac{1}{n^{1.01}}\right),
\end{aligned}$$

where the probability of the first event is estimated as in the proof of Lemma 3.5. The equality between the second and third lines is due to the fact that the events $\left|N^{Par}_{G_1}(x) \setminus \{y\}\right| \leq \frac{1}{1000}\log n$ and $y \in N_{G_2}(x)$ are independent. The statement of the claim follows from a union bound over all $O(n)$ choices of x. □

Claim 4.7. *W.v.h.p. the number of cross edges incident to $V(M)$ in G is $o\left(n^{0.99}\right)$.*

Proof. By Lemma 3.5 and Claim 4.6, we may assume that $|M| = |V_{low}| < n^{0.01}$ and $V(M) \cap \Gamma = \emptyset$. Therefore, each vertex in $V(M)$ is incident to $O\left(n^{0.95}\right)$ cross edges, and the claim follows. □

Observe that the identity of M depends only on the parallel edges of G_H w.r.t. (S', T'). This allows us to think of G_1 as being exposed in two independent stages. In the first stage the parallel edges of G_1 are exposed, and in the second the cross edges are exposed.

Set

$$\begin{aligned} \tilde{G} &= G \setminus V(M),\\ \tilde{G}_1 &= G_1 \setminus V(M),\\ \tilde{G}_H &= G_H \setminus V(M),\\ \left(\tilde{S}, \tilde{T}\right) &= \left(S' \setminus V(M), T' \setminus V(M)\right), \end{aligned}$$

and

$$\Gamma_{bad} = \left\{x \in \Gamma : \deg^{\mathrm{Par}}_{G_1}(x) < \frac{1}{1000}\log n\right\}.$$

Observation 4.8. Suppose that in $\tilde{G}_H$, there is no Hall cut that is C-close to $\left(\tilde{S}, \tilde{T}\right)$ for some given C. Then there is no Hall cut in G_H that is C-close to (S', T').

Proof. We prove the contrapositive. Suppose there exists a Hall cut (S, T) in G_H that is C-close to (S', T'). Observe that there is no edge connecting S and T^c, and therefore $(S \setminus V(M), T \setminus V(M))$ is also a Hall cut in $\tilde{G}_H$. Furthermore, since we only removed vertices, this cut is C-close to $(\tilde{S}, \tilde{T})$. □

As before, we identify cuts that are close to $(\tilde{S}, \tilde{T})$ with their set of shifted vertices Δ. By Observation 4.8 it suffices to show that w.v.h.p. there is no Hall cut in $\tilde{G}_H$ with $\frac{n^{-1/20}}{10}|\Gamma| \leq |\Delta| \leq n^{0.9}$. We first explore pseudo-random properties of $\tilde{G}_H$.

Claim 4.9. *W.v.h.p every partite set $A \subseteq V \setminus (V(M) \cup \Gamma_{bad})$ of size at most $n^{0.9}$ satisfies*

$$\left|N^{Par}_{\tilde{G}_1}(A)\right| \geq |A|\frac{1}{3000}\log n.$$

Proof. Suppose the conclusion does not hold, i.e., there is a partite set $A \subseteq V(\tilde{G}_H) \setminus \Gamma_{bad}$ with $|A| \leq n^{0.9}$ s.t. $\left|N_{\tilde{G}_1}(A)\right| < |A|\frac{1}{3000}\log n$. Then, for every $x \in A$, $\deg^{\mathrm{Par}}_{G_H}(x) \geq \frac{1}{1000}\log n$. Furthermore,

$$N^{\mathrm{Par}}_{G_1}(A) \subseteq N^{\mathrm{Par}}_{\tilde{G}_1}(A) \cup \left(\left(V_{low} \cup N^{\mathrm{Par}}_{G_H}(V_{low})\right) \cap N^{\mathrm{Par}}_{G_1}(A)\right).$$

However, by Lemma 3.5, $\left|\left(V_{low} \cup N^{\mathrm{Par}}_{G_H}(V_{low})\right) \cap N^{\mathrm{Par}}_{G_1}(A)\right| \leq |A|$, because if a vertex in A has two neighbors in $V_{low} \cup N^{\mathrm{Par}}_{G_H}(V_{low})$, then there are two vertices in V_{low} whose distance in G_H is at most 4. Therefore:

$$\left|N^{\mathrm{Par}}_{G_1}(A)\right| \le \left|N^{\mathrm{Par}}_{\tilde{G}_1}(A)\right| + |A| \le |A|\frac{1}{3000}\log n + |A| < |A|\frac{1}{2000}\log n.$$

The set A does not satisfy the conclusion of Lemma 3.6, which holds w.v.h.p. Therefore the conclusion of the present claim holds w.v.h.p. as well. □

Claim 4.10. *W.v.h.p.* $|\Gamma_{bad}| \le |\Gamma|/n^{0.4}$.

Proof. By Lemma 2.4, every vertex has parallel degree in G at least $(1-\varepsilon)\frac{k}{2}$. Therefore the probability that a vertex's parallel degree in G_1 is less than $\frac{1}{1000}\log n$ is bounded above by

$$\frac{1}{1000}\log n\binom{(1-\varepsilon)\frac{k}{2}}{\frac{1}{1000}\log n}p_1^{\frac{1}{1000}\log n}(1-p_1)^{(1-\varepsilon)k/2-\frac{1}{1000}\log n} = O\left(n^{-0.49}\right).$$

The conclusion follows from an application of Markov's inequality. □

Henceforth, unless otherwise specified, parallel degrees, cross degrees, etc., are with respect to the vertex set $V\setminus V(M)$ and the cut $\left(\tilde{S},\tilde{T}\right)$.

Claim 4.11. *The following holds w.v.h.p. Suppose* Δ *of size* $\frac{n^{-1/20}}{10}|\Gamma| \le |\Delta| \le n^{0.9}$ *satisfies one of:*

(1) $|\Delta| \ge \log n/\log\log n$.

(2) $\Delta \cap \Gamma \ne \emptyset$.

Then (S,T) *is not a Hall cut in* $\tilde{G}_H$.

Proof. Set:

$$a = \tilde{S}\setminus S, b = \tilde{S}^c\setminus S^c, c = \tilde{T}^c\setminus T^c, d = \tilde{T}\setminus T.$$

We first observe that if either $N^{\mathrm{Par}}_{\tilde{G}_1}(b) \not\subseteq c$ or $N^{\mathrm{Par}}_{\tilde{G}_1}(d) \not\subseteq a$, then (S,T) is not a Hall cut. Thus we may assume that $N^{\mathrm{Par}}_{\tilde{G}_1}(b) \subseteq c$ and $N^{\mathrm{Par}}_{\tilde{G}_1}(d) \subseteq a$. The conclusion of Claim 4.9 then implies that

$$|a| \ge |d\setminus\Gamma_{bad}|\frac{1}{3000}\log n, |c| \ge |b\setminus\Gamma_{bad}|\frac{1}{3000}\log n.$$

As by the previous claim and the hypothesis $|\Gamma_{bad}| \le |\Gamma| = o\left(|\Delta|/\log n\right)$:

$$\begin{aligned}|b|+|d| &= |b\setminus\Gamma_{bad}| + |b\cap\Gamma_{bad}| + |d\setminus\Gamma_{bad}| + |d\cap\Gamma_{bad}|\\ &\le O\left(\frac{|a|+|c|}{\log n}\right) + |\Gamma_{bad}| = O\left(\frac{|\Delta|}{\log n}\right).\end{aligned}$$

We may assume that $|S| > |T|$, for otherwise (S,T) is not a Hall cut by definition. It also holds that:

$$\begin{aligned}|S'|-|T'| &= |\tilde{S}|-|\tilde{T}| = (|S|+|a|-|b|)-(|T|+|d|-|c|)\\ &> |a|+|c|-(|b|+|d|) = \left(1-O\left(\frac{1}{\log n}\right)\right)|\Delta|.\end{aligned}$$

Since G is k-regular,

$$e_G(\tilde{S},\tilde{T}^c) \ge \left(1-O\left(\frac{1}{\log n}\right)\right)|\Delta|k.$$

By Claim 4.7 the number of cross edges in G that are not cross edges in $\tilde{G}$ is at most $n^{0.99}$. Thus:

$$e_{\tilde{G}}(\tilde{S},\tilde{T}^c) \geq \left(1 - O\left(\frac{1}{\log n}\right)\right)|\Delta|k.$$

Set $\Delta_1 = \Delta \cap \Gamma, \Delta_2 = \Delta \setminus \Gamma$. We then have:

$$\begin{aligned}\left|E_{\tilde{G}}(S,T^c) \cap E_{\tilde{G}}(\tilde{S},\tilde{T}^c)\right| &\geq e_{\tilde{G}}(\tilde{S},\tilde{T}^c) - |\Delta_1|(1+\varepsilon)\frac{k}{2} - |\Delta_2|\frac{k}{n^{1/20}} \\ &\geq \left(\frac{1}{2} - \varepsilon\right)|\Delta_1|k + \left(1 - O\left(\frac{1}{\log n}\right)\right)|\Delta_2|k.\end{aligned}$$

Therefore, the probability that none of the cross edges is in $\tilde{G}_1$ is at most:

$$(1-p_1)^{\left(\frac{1}{2}-\varepsilon\right)|\Delta_1|k}(1-p_1)^{\left(1-O\left(\frac{1}{\log n}\right)\right)|\Delta_2|k}.$$

Suppose 1 holds. Let $m = \max\left\{\frac{n^{-1/20}}{10}|\Gamma|, \log n/\log\log n\right\}$. Then, applying a union bound over all choices of $\Delta_1 \subseteq \Gamma$ and Δ_2:

$$\begin{aligned}\alpha &:= \sum_{\substack{|\Delta|\in\{m,\ldots,n^{0.9}\} \\ |\Delta_1|+|\Delta_2|=|\Delta|}} \binom{|\Gamma|}{|\Delta_1|}\binom{2n}{|\Delta_2|}(1-p_1)^{\left(\frac{1}{2}-\varepsilon\right)|\Delta_1|k}(1-p_1)^{\left(1-O\left(\frac{1}{\log n}\right)\right)|\Delta_2|k} \\ &\leq \sum_{\substack{|\Delta|\in\{m,\ldots,n^{0.9}\} \\ |\Delta_1|+|\Delta_2|=|\Delta|}} |\Delta_1|^{-|\Delta_1|}|\Delta_2|^{-|\Delta_2|}\left(e|\Gamma|\left(\frac{1}{n}\right)^{1/2-2\varepsilon}\right)^{|\Delta_1|}(O(\log\log\log n))^{|\Delta_2|}.\end{aligned}$$

Now

$$|\Delta_1|^{-|\Delta_1|}|\Delta_2|^{-|\Delta_2|} \leq \left(\frac{2}{|\Delta|}\right)^{|\Delta|} = \left(\frac{2}{|\Delta|}\right)^{|\Delta_1|}\left(\frac{2}{|\Delta|}\right)^{|\Delta_2|}.$$

Thus:

$$\begin{aligned}\alpha &\leq \sum_{\substack{|\Delta|\in\{m,\ldots,n^{0.9}\} \\ |\Delta_1|+|\Delta_2|=|\Delta|}} \left(\frac{2e|\Gamma|}{|\Delta|}\left(\frac{1}{n}\right)^{1/2-2\varepsilon}\right)^{|\Delta_1|}(O(\log\log\log n/|\Delta|))^{|\Delta_2|} \\ &\leq \sum_{\substack{|\Delta|\in\{m,\ldots,n^{0.9}\} \\ |\Delta_1|+|\Delta_2|=|\Delta|}} \left(\frac{1}{n^{2/5}}\right)^{|\Delta_1|}\left(O\left(\frac{\log|\Delta|}{|\Delta|}\right)\right)^{|\Delta_2|} \leq n^{0.9}\left(\tilde{O}\left(\frac{1}{\log n}\right)\right)^{\log n/\log\log n} \\ &\leq n^{0.9}\frac{1}{n^{1-o(1)}} = O\left(\frac{1}{n^{0.05}}\right).\end{aligned}$$

Otherwise, 2 holds. Then $|\Delta_1| \geq 1$. By a similar application of a union bound, the probability that there exists any Hall cut (S,T) in $\tilde{G}_1$ satisfying the hypothesis is bounded above by:

$$\sum_{\substack{|\Delta|\in[\log n/\log\log n] \\ |\Delta_1|+|\Delta_2|=|\Delta|}} \left(\frac{1}{n^{2/5}}\right)^{|\Delta_1|}\left(O\left(\frac{\log|\Delta|}{|\Delta|}\right)\right)^{|\Delta_2|} = \tilde{O}\left(\frac{1}{n^{2/5}}\right).$$

□

It remains to show that w.v.h.p. there is no Hall cut with $|\Delta| \leq \log n/\log\log n$ and $\Delta \cap \Gamma = \emptyset$.

Claim 4.12. *W.v.h.p. there exists no Hall cut (S,T) in $\tilde{G}_H$ with $|\Delta| < \log n / \log\log n$ and $\Delta \cap \Gamma = \emptyset$.*

Proof. We first show that if $\left|\tilde{S}\right| \leq \left|\tilde{T}\right|$ then w.v.h.p. there is no Hall cut satisfying the claim's hypothesis. Suppose, for a contradiction, that $|\Delta| < \log n / \log\log n$ and $\Delta \cap \Gamma = \emptyset$ corresponds to a Hall cut. Since $\Delta \cap \Gamma = \emptyset$, we have $b, d \subseteq V \setminus (V(M) \cup \Gamma)$. Therefore for every $x \in b \cup d$, $N_{\tilde{G}_1}(x) \subseteq \Delta$. However, by Claim 4.9, w.v.h.p. for every such x, $\left|N_{\tilde{G}_1}(x)\right| = \Omega(\log n)$. Assuming this, and since $|\Delta| < \log n / \log\log n$, we conclude that $b = d = \emptyset$. Thus:

$$|S| - |T| = \left|\tilde{S}\right| - \left|\tilde{T}\right| - |a| + |b| - |c| + |d| = \left|\tilde{S}\right| - \left|\tilde{T}\right| - |a| - |c| \leq \left|\tilde{S}\right| - \left|\tilde{T}\right| \leq 0.$$

Therefore, (S,T) is not a Hall cut.

We now assume that $\left|\tilde{S}\right| - \left|\tilde{T}\right| > 0$. We will show presently that w.v.h.p. $e_{\tilde{G}_1}\left(\tilde{S}, \tilde{T}^c\right) = \Omega(\log n)$. Suppose (S,T) is a cut satisfying the claim's hypothesis. Then Δ must contain a vertex cover of $E_{\tilde{G}_1}\left(\tilde{S}, \tilde{T}^c\right)$. However, since $\Delta \cap \Gamma = \emptyset$, by Lemma 3.4 w.v.h.p. each vertex in Δ is incident to at most 30 cross edges in G_1. Since $|\Delta| < \log n / \log\log n$, Δ does not contain a vertex cover of $E_{\tilde{G}_1}\left(\tilde{S}, \tilde{T}^c\right)$, and so (S,T) is not a Hall cut.

It remains to show that w.v.h.p. $e_{\tilde{G}_1}\left(\tilde{S}, \tilde{T}^c\right) = \Omega(\log n)$. Since G is k-regular, we have $e_G\left(\tilde{S}, \tilde{T}\right) \geq k\left(\left|\tilde{S}\right| - \left|\tilde{T}\right|\right) \geq k$. By Claim 4.7, the number of cross edges of G incident to $V(M)$ is $o\left(n^{0.99}\right)$, so $\tilde{G}$ has at least $C = (1 - o(1))k$ cross edges. Now, $e_{\tilde{G}_1}\left(\tilde{S}, \tilde{T}\right) \sim Bin(C, p_1)$. By an application of Chernoff's inequality, w.v.h.p. $e_{\tilde{G}_1}\left(\tilde{S}, \tilde{T}\right) \geq \frac{1}{2}\mathbb{E}[C] = \Omega(\log n)$.

□

Appendix A. Proof of Proposition 1.3

It may be intuitive to think at first - as all three of us did - that the conclusion of Theorem 1.1 holds for all large regular bipartite graphs, i.e., the requirement $k = \omega\left(\frac{n}{\log^{1/3} n}\right)$ is not necessary. In this section, we analyze a construction of Goel, Kapralov, and Khanna [13] to show that this is not true, and indeed for small values of k, $G(p)$ might not contain a perfect matching even for relatively large p.

The intuition for all of our counterexamples comes from the following simple construction.

Definition A.1. A ***k*-resistor** between two vertices x and y is the following bipartite graph: The vertex set is $\{x,y\} \dot{\cup} X' \dot{\cup} Y'$, where X' and Y' have cardinality k. Let $x' \in X', y' \in Y'$ be "special" vertices. The edge set is:

$$\{xx', yy'\} \cup \left(\{ab : a \in X', b \in Y'\} \setminus \{x'y'\}\right).$$

In other words, starting from the complete bipartite graph on X' and Y', the edge $x'y'$ is removed, and the edges xx' and yy' are added.

Notice that of the $2k+2$ vertices of a k-resistor between x and y, all but x and y have degree k. Furthermore, if a spanning subgraph of the resistor contains a perfect matching, both edges xx' and yy' are present. This leads to the following construction.

Proposition A.2. *Construct a k-regular, $n = (2k^2 + 2)$-vertex bipartite graph G as follows. Let x and y be two initial vertices. Add k distinct k-resistors between x and y. Then, a.a.s. the random subgraph $G(p)$ does not contain a perfect matching for any $p = o\left(n^{-1/4}\right)$. On the other hand, a.a.s. $G(p)$ contains no isolated vertices for any $p = \omega\left(\log n/\sqrt{n}\right)$.*

Proof. Both conclusions follow from the first moment method.

Let $H \sim G(p)$. Note that H contains a perfect matching only if for one of the resistors, both edges xx' and yy' are present. This occurs with probability p^2. As there are k different resistors, and they are all edge-disjoint, the expected number of such pairs is kp^2. Since $k = \Theta(\sqrt{n})$, if $p = o\left(n^{-1/4}\right)$, a.a.s. there is no such pair in H.

The expected number of isolated vertices in H is $n(1-p)^k \leq \exp\left(\log n - pk\right)$. When $p = \omega\left(\log n/\sqrt{n}\right)$ this tends to zero, and a.a.s. there are no isolated vertices. □

In this example we had $k = \Theta(\sqrt{n})$, leaving a large gap between it and the range $k = \Theta(n)$ in Theorem 1.1. We reduce this gap as follows.

Definition A.3. A **(k, ℓ, r)-series of resistors** between two vertices x and y is constructed as follows. Let $K_1, K_2, \ldots, K_\ell$ be ℓ copies of the complete bipartite graph $K_{k,k}$, with respective vertex sets $X_1 \dot\cup Y_1, X_2 \dot\cup Y_2, \ldots, X_\ell \dot\cup Y_\ell$. For each $1 \leq i \leq \ell$, let $x_i^1, x_i^2, \ldots, x_i^r \in X_i, y_i^1, y_i^2, \ldots, y_i^r \in Y_i$ be distinct. Remove all edges of the form $x_i^j y_i^j$, and add all edges of the form $y_i^j x_{i+1}^j$, as well as $xx_1^j, y_\ell^j y$.

The following proposition uses a construction similar to the one in Proposition A.2.

Proposition A.4. *For $n = 2 + 20k \log k \log\log k$, construct a k-regular n-vertex bipartite graph G as follows. Starting with two vertices x and y, add $\log k$ distinct $(k, 10\log\log k, \frac{k}{\log k})$-series of resistors between x and y. A.a.s. the random subgraph $G(p)$ does not contain a perfect matching for any $p \leq 2\log n/k$. On the other hand, $p = \left(\log n + \omega\left(1\right)\right)/k$ suffices for $G(p)$ to contain no isolated vertices a.a.s.*

Proof. For consistency with Definition A.3, let $\ell = 10\log\log k$ and $r = k/\log k$. For a spanning subgraph $G' \subseteq G$ to contain a perfect matching, there must be at least one series of resistors containing at least one edge of the form xx_1^j, at least one edge of the form $y_\ell^j y$, and one edge of the form $y_i^j x_{i+1}^j$ for each i between 1 and $\ell - 1$. Therefore, applying the union bound over all k/r choices of the (k, ℓ, r)-series, we obtain

$$\mathbb{P}\left[G(p) \text{ contains a perfect matching}\right] \leq \frac{k}{r}\left[1 - (1-p)^r\right]^{\ell+1}.$$

Let $p = 2\log n/k$. Then $(1-p)^r \sim e^{-2}$, and therefore

$$\mathbb{P}\left[G(p) \text{ contains a perfect matching}\right] \leq \log k\left(1 - e^{-2}\right)^{10\log\log k} = o\left(1\right).$$

The statement about isolated vertices follows from an argument similar to the one in the proof of Proposition A.2. □

References

[1] Miklós Ajtai, János Komlós, and Endre Szemerédi, *The longest path in a random graph*, Combinatorica **1** (1981), no. 1, 1–12.

[2] Sonny Ben-Shimon, Michael Krivelevich, and Benny Sudakov, *On the resilience of Hamiltonicity and optimal packing of Hamilton cycles in random graphs*, SIAM J. Discrete Math. **25** (2011), no. 3, 1176–1193.

[3] Béla Bollobás, *The evolution of sparse graphs*, Graph Theory and Combinatorics (Cambridge 1983).

[4] ———, *Random graphs*, Academic Press, 1985.

[5] Lev Bregman, *Some properties of nonnegative matrices and their permanents*, Soviet Math. Dokl, vol. 14, 1973, pp. 945–949.

[6] Gregory Egorychev, *The solution of van der waerden's problem for permanents*, Advances in Mathematics **42** (1981), no. 3, 299–305.

[7] Paul Erdős and Alfréd Rényi, *On random graphs I*, Publ. Math. Debrecen **6** (1959), 290–297.

[8] ———, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci **5** (1960), 17–61.

[9] ———, *On random matrices*, Magyar Tud. Akad. Mat. Kutató Int. Közl **8** (1964), 455–461.

[10] Dmitry Falikman, *Proof of the van der waerden conjecture regarding the permanent of a doubly stochastic matrix*, Mathematical notes of the Academy of Sciences of the USSR **29** (1981), no. 6, 475–479.

[11] Roman Glebov, *On hamilton cycles and other spanning structures*, Ph.D. thesis, Free University of Berlin, 2013.

[12] Roman Glebov, Humberto Naves, and Benny Sudakov, *The threshold probability for long cycles*, Combinatorics, Probability and Computing **26** (2017), no. 2, 208–247.

[13] Ashish Goel, Michael Kapralov, and Sanjeev Khanna, *Perfect matchings via uniform sampling in regular bipartite graphs*, ACM Transactions on Algorithms (TALG) **6** (2010), no. 2, 27.

[14] János Komlós and Endre Szemerédi, *Hamilton cycles in random graphs*, Infinite and finite sets **2** (1973), 1003–1010.

[15] ———, *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. **43** (1983), no. 1, 55–63.

[16] Alexey Korshunov, *Solution of a problem of Erdős and Rényi on Hamilton cycles in nonoriented graphs*, Soviet Math. Dokl., vol. 17, 1976, pp. 760–764.

[17] Michael Krivelevich, Choongbum Lee, and Benny Sudakov, *Robust Hamiltonicity of Dirac graphs*, Trans. Amer. Math. Soc. **366** (2014), no. 6, 3095–3130.

[18] ———, *Long paths and cycles in random subgraphs of graphs with large minimum degree*, Random Structures & Algorithms **46** (2015), no. 2, 320–345.

[19] Daniela Kühn, Allan Lo, Deryk Osthus, and Katherine Staden, *The robust component structure of dense regular graphs and applications*, Proceedings of the London Mathematical Society **110** (2014), no. 1, 19–56.

[20] Zur Luria and Michael Simkin, *On the threshold problem for Latin boxes*, Random Structures & Algorithms (2019), 1–24, `https://doi.org/10.1002/rsa.20855`.

[21] Lajos Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), no. 4, 359–364.

[22] Oliver Riordan, *Long cycles in random subgraphs of graphs with large minimum degree*, Random Structures & Algorithms **45** (2014), no. 4, 764–767.

[23] Benny Sudakov, *Robustness of graph properties*, Surveys in Combinatorics 2017 **440** (2017), 372.

Hebrew University of Jerusalem, Jerusalem 91904, Israel
Email address: `roman.l.glebov@gmail.com`

Software Department, Jerusalem College of Engineering
Email address: `zluria@gmail.com`

Institute of Mathematics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
Email address: `menahem.simkin@mail.huji.ac.il`

Chapter 5

A randomized construction of high girth regular graphs

Nati Linial and Michael Simkin. Submitted for publication.

A RANDOMIZED CONSTRUCTION OF HIGH GIRTH REGULAR GRAPHS

NATI LINIAL AND MICHAEL SIMKIN

Abstract. We describe a new random greedy algorithm for generating regular graphs of high girth: Let $k \geq 3$ and $c \in (0,1)$ be fixed. Let $n \in \mathbb{N}$ be even and set $g = c \log_{k-1}(n)$. Begin with a Hamilton cycle G on n vertices. As long as the smallest degree $\delta(G) < k$, choose, uniformly at random, two vertices $u, v \in V(G)$ of degree $\delta(G)$ whose distance is at least $g - 1$. If there are no such vertex pairs, abort. Otherwise, add the edge uv to $E(G)$.

We show that with high probability this algorithm yields a k-regular graph with girth at least g. Our analysis also implies that there are $(\Omega(n))^{kn/2}$ labeled k-regular n-vertex graphs with girth at least g.

1. Introduction

The **girth** of a graph is the length of its shortest cycle. It is a classical challenge to determine $g(k,n)$, the largest possible girth of k-regular graphs with n vertices. Here we only concern ourselves with fixed $k \geq 3$ and large n. Moore's bound says that $g(k,n) \leq (1 \pm o(1)) \cdot 2\log_{k-1}(n)$. Although the argument is very simple, this remains our best asymptotic upper bound.

The study of high-girth graphs has a long history. Using a combinatorial argument, Erdős and Sachs [12] proved in 1963 that $g(k,n) \geq (1 \pm o(1)) \log_{k-1}(n)$. Twenty years later, Biggs and Hoare [3] gave an algebraic construction of a family of cubic graphs later shown [35] to have girth at least $(1 - o(1))\frac{4}{3}\log_2(n)$. Then, for k an odd prime plus one, Lubotzky, Phillips, and Sarnak [26] constructed their celebrated Ramanujan graphs, with girth $(1 \pm o(1))\frac{4}{3}\log_{k-1}(n)$. As observed in [20, Introduction], this implies that $g(k,n) \geq (1 - o(1))c(k)\log_{k-1}(n)$, where $c(k) > 1$ for *every* $k \geq 3$, and $\lim_{k\to\infty} c(k) = 4/3$. Cayley graphs attaining this bound were found by Dahan [11]. Along the way, advances by Chiu [10], Morgenstern [28], and Lazebnik, Ustimenko, and Woldar [25] broadened the range of degrees for which similar constructions are known. We further refer the reader to Biggs's survey [4] of the best known constructions for cubic graphs.

In contrast, and notwithstanding considerable research efforts, the Erdős-Sachs bound remains the best asymptotic lower bound on $g(k,n)$ that is derived by combinatorial and probabilistic techniques. This is one of very few examples where explicit algebraic constructions beat the probabilistic method. We believe that the road to constructing high-girth graphs using such methods goes via better understanding of the *large-scale geometry of graphs*. In our open problem section we mention several additional mysteries in this domain.

Here we describe a *random greedy algorithm* to construct regular high-girth graphs. In recent years, random greedy algorithms have become a powerful tool for constructing constrained combinatorial structures. Thus, Glock, Kühn, Lo, and Osthus [18],

Supported by Israel Science Foundation grant 659/18.

and independently Bohman and Warnke [6], used this method to prove the existence of approximate Steiner triple systems that are locally sparse. This methodology has also played prominent roles in the proofs by Keevash [21] and Glock, Kühn, Lo, and Osthus [17] of the existence of combinatorial designs.

Random greedy algorithms have also been studied in their own right. For example, in the "triangle-free" graph process (e.g., [13, 5]), edges are randomly added to a graph one by one and subject to the constraint that no triangle is created. Similarly, various authors studied "H-free" processes for other fixed graphs H, including stars [32] and cycles [29, 7, 30, 33, 31]. In another relevant paper Krivelevich, Kwan, Loh, and Sudakov [23] studied the process where edges are randomly added to a graph as long as the matching number remains below a fixed value which may depend on the number of vertices. This is indeed just a tiny sample of a rich and beautiful body of literature.

Here is a simple method to generate random k-regular graphs on n vertices: Start with a Hamilton cycle, and repeatedly add perfect matchings uniformly at random until the desired degree is attained. Since the present paragraph is intended only as background, we do not go into detail, and do not dwell on how to avoid double edges. We consider here a sequential variant of this algorithm, which produces graphs of girth at least g. Let $G = (V, E)$ be a graph on n vertices with all vertex degrees at most k (in our main application, G is a Hamilton cycle, and $k \geq 3$). Let $g \leq n$. Set $G_0 = G = (V, E_0)$. We obtain $G_{t+1} = (V, E_{t+1})$ from G_t as follows:

- If G_t is k-regular, set $G_{t+1} = G_t$.
- Otherwise:
  - Let $d < k$ be the smallest vertex degree in G_t, and let W_t be the set of **unsaturated** vertices in G_t, i.e., those with degree d.
  - We say that $u, v \in W_t$ is an **available** pair of vertices if their distance in G_t is at least $g - 1$. Let $\mathcal{A}_t$ be the set of available pairs, and let H_t be the graph $(W_t, \mathcal{A}_t)$.
  - If $\mathcal{A}_t = \emptyset$, set $G_{t+1} = G_t$.
  - Otherwise, choose $e_{t+1} \in \mathcal{A}_t$ uniformly at random, and set $E_{t+1} = E_t \cup \{e_{t+1}\}$.

We call this the (G, g, k)**-high-girth-process**. We say that the process **saturates** if for some t, G_t is k-regular. We note that in this case $\operatorname{girth}(G_t) \geq \min\{g, \operatorname{girth}(G)\}$.

Our main result is that with proper choice of parameters, this algorithm yields high-girth regular graphs.

Theorem 1.1. *Let $1 > c > 0$, $k \geq 3$ an integer, and n an even integer. Let $g = g(n) \leq c \log_{k-1}(n)$, and G be a Hamilton cycle on n vertices. Then, w.h.p.*[1]*, the (G, g, k)-high-girth-process saturates.*

A byproduct of the analysis of this algorithm is a lower bound on the number of high-girth regular graphs.

Theorem 1.2. *Let $1 > c > 0$, $k \geq 3$ an integer, and n an even integer. There are at least $(\Omega(n))^{kn/2}$ labeled k-regular graphs G on n vertices with* $\operatorname{girth}(G) \geq c \log_{k-1}(n)$.

Remark 1.3. We do not give Theorem 1.2 in the best form known to us, since we believe this is in any rate far from the truth.

[1] We say that a sequence of events occurs **with high probability** (**w.h.p.**) if the probabilities of their occurrence tend to 1.

We also mention that for $c < 1/2$, a remarkably accurate enumeration is given by McKay, Wormald, and Wysocka [27, Corollary 2] who studied the distribution of the number of cycles in random regular graphs. However, they do not give a construction, and their method applies only when $c < 1/2$.

Theorem 1.2 illustrates one advantage of probabilistic constructions over algebraic ones: While the latter achieve higher girth, they are sporadic and provide only a small supply of examples. In contrast, probabilistic techniques provide a viewpoint from which to study a very large family of high-girth graphs.

In comparison with other results in the literature, ours is the first probabilistic algorithm that constructs graphs with unbounded girth that are also regular. For constant g, Osthus and Taraz [29, Corollary 4] determined (up to polylog factors) the final number of edges in the $\mathcal{H}$-free process, where $\mathcal{H}$ is the collection of all cycles shorter than g. Bayati, Montanari, and Saberi [1] studied a similar sequential process which samples uniformly from the family of girth-g graphs with m edges, where g is a constant and $m = O\left(n^{1+\alpha(g)}\right)$, for some non-negative function α. Chandran [9] considered a (deterministic) greedy algorithm to construct graphs with girth $(1 + o\,(1)) \log_k(n)$ and average degree k. However, none of these constructions produce regular graphs. Closer to the algebraic end of the spectrum, Gamburd, Hoory, Shahshahani, Shalev, and Virág [16] showed that for various families of groups, random k-regular Cayley graphs have unbounded girth that in some cases is as high as $(1 - o(1)) \log_{k-1}(n)$.

The rest of this paper is organized as follows. The remainder of this section introduces some notations. In Section 2 we prove Theorem 1.1, modulo two technical claims which are proved in Sections 3 and 4. We prove Theorem 1.2 in Section 5. We close with some remarks and open problems in Section 6.

1.1. **Notation.**

The vertex and edge sets of a graph G are denoted by $V(G)$, resp. $E(G)$. We write $e(G) = |E(G)|$. The neighbor set of vertex $v \in V(G)$ is denoted $\Gamma_G(v)$. The distance between $u, v \in V(G)$ is denoted $\delta_G(u, v)$. The graph of G induced by $U \subseteq V(G)$ is denoted $G[U]$.

The set $\{1, 2, \ldots, a\}$ is denoted by $[a]$. Also, $[a]_0 := \{0, 1, 2, \ldots, a\}$, and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For $x, y \in \mathbb{R}$, we write $x \pm y$ to indicate an unspecified number in the interval $[x - |y|, x + |y|]$.

2. Constructing high-girth graphs: proof of Theorem 1.1

Let $G'_0, G'_1, \ldots$ be a (G', g, k)-high-girth-process, where G' is a Hamilton cycle, and c, k, n and $g \leq c \log_{k-1}(n)$ are as in the theorem. We argue by induction on k, starting with $k = 3$. Now, suppose Theorem 1.1 holds for $k - 1 \geq 3$. Then, since $g \leq c \log_{k-1}(n) < c \log_{k-2}(n)$, it follows by induction that w.h.p. $G'_{(k-2)n/2}$ is $(k - 1)$-regular. It is thus sufficient to prove the following proposition (which covers both the base case and the inductive step).

Proposition 2.1. *Let G be a $(k-1)$-regular graph on n vertices, with n even and $k \geq 3$. Let $c < 1$ and let $g \leq c \log_{k-1}(n)$. Then, w.h.p., the (G, g, k)-high-girth-process saturates.*

Let $G_0, G_1, \ldots$ be a (G, g, k)-high-girth-process, and let $e_1, e_2, \ldots$ be the edges added to the graph at each step. Clearly, a necessary and sufficient condition for the process to saturate is that $|E_t| = (k - 1)n/2 + t$ for every $0 \leq t \leq n/2$.

We say that the process **freezes at time** t if t is the smallest integer such that $E_t = E_{t+1}$. We denote this time by T_{freeze} (so that the process saturates if and only if $T_{freeze} = n/2$).

Our proof deals separately with two phases of the process. In Section 2.1 we show that in the first phase it holds with certainty that almost all vertices saturate, and H_t is almost complete.

Section 2.2 is devoted to the more involved, "nibbling"-based analysis of the second phase. We divide the remainder of the process into a bounded number of steps. We show that in each step, the number of unsaturated vertices is reduced by a polynomial factor, and that certain pseudorandomness conditions are preserved from step to step. We then argue that w.h.p. the graph obtained at the end of the penultimate step has a combinatorial property that implies the process saturates with certainty.

2.1. **The early evolution of the process.** Let $0 < \varepsilon < 1 - c$, and let

$$T = \frac{1}{2}\left(n - n^{c+\varepsilon}\right).$$

The following observation follows from the Moore bound.

Observation 2.2. Let H be a graph with maximal degree at most k, and let $\ell \in \mathbb{N}$. For every $v \in V(H)$ there are at most $k \cdot (k-1)^{\ell-1}$ vertices at distance ℓ from v, and at most $2k \cdot (k-1)^{\ell}$ vertices at distance at most ℓ from v.

We next use this observation to show that for n sufficiently large $T_{freeze} \geq T$ with certainty, and for every $t \leq T$, the graph H_t is nearly complete.

Lemma 2.3. *There exists an integer $n_0 = n_0(c, \varepsilon)$ such that for all $n \geq n_0$ and every $t \leq T$ it holds with certainty that:*

(1) All vertex degrees in $H_t = (W_t, \mathcal{A}_t)$ are at least $|W_t| - O(n^c)$.
(2) $|\mathcal{A}_t| = \frac{1}{2}|W_t|^2\left(1 - O(n^c/|W_t|)\right)$.
(3) $|W_t| = n - 2t$, and hence
(4) $T_{freeze} \geq T$.

Proof. The two vertices of every edge in $E_t \setminus E_0$ have degree k, and every vertex of degree k is in exactly one edge from $E_t \setminus E_0$. Therefore: $|W_t| = n - 2|E_t \setminus E_0| \geq n - 2t$, with equality if and only if $t \leq T_{freeze}$.

Let $v \in W_t$. By Observation 2.2, there are $O(n^c)$ vertices in G_t with distance at most $g - 2$ to v. In H_t, v is adjacent to all other vertices in W_t. Therefore $d_{H_t}(v) \geq |W_t| - O(n^c)$, as claimed. Hence,

$$|\mathcal{A}_t| = \frac{1}{2}\sum_{v \in W_t} d_{H_t}(v) = \left(1 - O(n^c/|W_t|)\right)\frac{1}{2}|W_t|^2,$$

as desired.

Finally, $T_{freeze} \geq T$ as long as $\mathcal{A}_T \neq \emptyset$. As observed:

$$|W_T| \geq n - 2T = n^{c+\varepsilon}.$$

Hence, by 2:

$$|\mathcal{A}_T| = \frac{1}{2}|W_T|^2\left(1 - O\left(\frac{n^c}{|W_T|}\right)\right) = \Omega\left(n^{2(c+\varepsilon)}\right).$$

Thus, if n is large enough, then $\mathcal{A}_T$ is nonempty with certainty, implying 3, 4. □

The set $\mathcal{B}_t$ of **forbidden edges** is comprised of those pairs $u, v \in W_t$ with $uv \notin \mathcal{A}_t$. We will show that for $t \geq T$ w.h.p. the number of forbidden edges in G_t does not exceed the bound given by the following heuristic argument. Let $u \in W_t$. By Observation 2.2, at most n^c vertices $v \in V$ satisfy $\delta_{G_t}(u, v) \leq g - 2$. So, if v is chosen randomly from V, then $\mathbb{P}[v \in W_t] = |W_t|/n$ and $\mathbb{P}[\delta_{G_t}(u, v) \leq g - 2] \leq n^c/n$. Had these events been independent, we expect there to be at most $|W_t|^2 n^c/n$ pairs $u, v \in W_t$ with $\delta_H(u, v) \leq g - 2$. Hence, when $|W_t| \lll \sqrt{n/n^c}$, we expect that $\mathcal{B}_t = \emptyset$. We now show that the latter condition implies that the process saturates with certainty.

Definition 2.4. Let $G = (V, E)$ be a graph with all degrees either $k - 1$ or k. We say that G is **safe** if every two vertices of degree $k - 1$ are at distance $\geq g - 1$.

Clearly G_t is safe if and only if $\mathcal{B}_t = \emptyset$.

Lemma 2.5. *If for some t, G_t is safe, then the process saturates with certainty.*

Proof. We first observe that if G_t is safe then H_t is the complete graph on W_t. Thus, it is enough to show that if $t < n/2$, then $\mathcal{A}_t \neq \emptyset$ and G_{t+1} is also safe. Suppose, for a contradiction, that $e_{t+1} = uv$ for some $u, v \in W_t$, and that G_{t+1} is not safe. Namely, there exist two vertices $a, b \in W_{t+1}$ such that $\delta_{G_{t+1}}(a, b) \leq g - 2$. Let P be a shortest ab-path in G_{t+1}. By assumption, its length is at most $g - 2$. But G_t is safe, whence $\delta_{G_t}(a, b) \geq g - 1$, so that necessarily $uv \in P$. It follows that in G_t there is a path of length $\leq g - 2$ from one of the vertices a, b to one of u, v contrary to the assumption that G_t is safe. □

Here is the main technical ingredient in the analysis of the first T steps in the process. An edge uv is a **chord** if it is not in the initial graph G. E.g., all edges chosen by the process are chords.

Claim 2.6. *Let $a, b \leq \log^2(n)$. Let $s_1, s_2, \ldots, s_a$ be distinct chords and let $U \subseteq V$ be a set of b vertices. Let $0 \leq t_1, t_2, \ldots, t_a < T$. Let A be the event that for every $1 \leq i \leq a$, the process chooses chord s_i at step t_i (i.e., $e_{t_i} = s_i$), and that $U \subseteq W_T$. Then*

$$\mathbb{P}[A] \leq (1 \pm o(1)) \left(\frac{2}{n^2}\right)^a \left(1 - \frac{2T}{n}\right)^b.$$

It is easy to see where this expression comes from. Since $|W_T| = n - 2T$, it is plausible that $\mathbb{P}[v \in W_T] \approx 1 - 2T/n$ for every $v \in V$. Also, if edges are chosen uniformly at random, ignoring the degree and girth constraints, then the probability of the event $e_t = s$ is $(1 \pm o(1))2n^{-2}$. The bound on $\mathbb{P}[A]$ says that the constraints can only reduce this probability. This heuristic will be justified by Lemma 2.3: Throughout the first T steps of the process, the graph of available edges H_t is nearly complete. Therefore, in each of the first T steps, both the number of available edges and the number of available edges incident to U are very close to what these values would be in the unconstrained graph process. As a consequence, the two processes exhibit similar behavior

Proof. Note that there is no loss in assuming that

- $s_1, \ldots, s_a$ form a matching,
- $t_1, \ldots, t_a$ are all distinct, and
- U is disjoint from the vertices in $s_1, \ldots, s_a$,

for otherwise $\mathbb{P}[A] = 0$ and the conclusion follows trivially.

The sequential nature of the process suggests that we express A as an intersection of events $B_0, B_1, \dots, B_T$, where B_t depends only on the chord selected at step t. For $0 \le t < T$, let $S_t = \{s_i : t_i > t\}$ be the set of chords that are to be chosen after step t. Let $U_t = U \cup \{u : \exists s \in S_t, u \in s\}$. The definition of B_t depends on whether or not $t = t_i$ for some i. If so, we let B_t be the event that chord s_i is selected at step t. Otherwise, it the event that we select at step t a chord disjoint from U_t. Clearly,

$$A = B_0 \cap B_1 \cap \ldots \cap B_T.$$

Therefore:

$$\mathbb{P}[A] = \mathbb{P}[B_0] \times \mathbb{P}[B_1|B_0] \times \mathbb{P}[B_2|B_1 \cap B_0] \times \ldots \times \mathbb{P}[B_{T-1}|B_0 \cap \ldots \cap B_{T-2}]. \tag{1}$$

By Lemma 2.3, for every $t < T$, we have

$$|\mathcal{A}_t| = \left(1 \pm O\left(\frac{n^c}{n-2t}\right)\right)\frac{(n-2t)^2}{2}.$$

Therefore, for every $i \in [a]$, it holds that

$$\mathbb{P}[B_{t_i}|B_{t_i-1} \cap \ldots \cap B_0] \le \left(1 + O\left(\frac{n^c}{n-2t_i}\right)\right)\frac{2}{(n-2t_i)^2}.$$

It will be useful to note also that

$$\begin{aligned}\prod_{i=1}^{a}\mathbb{P}[B_{t_i}|B_{t_i-1} \cap \ldots \cap B_0] &\le \prod_{i=1}^{a}\left(1 + O\left(\frac{n^c}{n-2t_i}\right)\right)\frac{2}{(n-2t_i)^2} \\ &\le (1 \pm o(1))\prod_{i=1}^{a}\left(1 + O\left(\frac{n^c}{n-2t_i}\right)\right)\frac{2}{(n-2t_i)^2}\left(1 - \frac{2|U_{t_i}|}{n-2t_i}\right).\end{aligned} \tag{2}$$

Consider next the case $t \notin \{t_1, \dots, t_a\}$. The event $B_0 \cap \ldots \cap B_{t-1}$ implies that $U_t \subseteq W_t$, so by Lemma 2.3, U_t intersects at least

$$\left(1 \pm O\left(\frac{n^c}{n-2t}\right)\right)|U_t|(n-2t) \pm \binom{|U_t|}{2} = \left(1 \pm O\left(\frac{n^c}{n-2t}\right)\right)|U_t|(n-2t)$$

chords in $\mathcal{A}_t$. Thus:

$$\begin{aligned}\mathbb{P}[B_t|B_{t-1} \cap \ldots \cap B_0] &\le 1 - \left(1 \pm O\left(\frac{n^c}{n-2t}\right)\right)\frac{|U_t|(n-2t)}{|\mathcal{A}_t|} \\ &\le 1 - \left(1 \pm O\left(\frac{n^c}{n-2t}\right)\right)\frac{2|U_t|}{n-2t}.\end{aligned} \tag{3}$$

Therefore, by (1), (2), and (3):

$$\begin{aligned}\mathbb{P}[A] &\le (1 \pm o(1))\left(\prod_{i=1}^{a}\frac{2}{(n-2t_i)^2}\right)\left(\prod_{t=0}^{T}\left(1 - \left(1 \pm O\left(\frac{n^c}{n-2t}\right)\right)\frac{2|U_t|}{n-2t}\right)\right) \\ &\le (1 \pm o(1))\left(\prod_{i=1}^{a}\frac{2}{(n-2t_i)^2}\right)\exp\left(-\sum_{t=0}^{T}\left(1 \pm O\left(\frac{n^c}{n-2t}\right)\right)\frac{2|U_t|}{n-2t}\right) \\ &\le (1 \pm o(1))\left(\prod_{i=1}^{a}\frac{2}{(n-2t_i)^2}\right)\exp\left(-\sum_{t=0}^{T}\frac{2|U_t|}{n-2t}\right).\end{aligned} \tag{4}$$

We now estimate the sum in the exponent. By definition of U_t, we have:

$$|U_t| = |U| + 2\,|S_t| = b + 2\,|S_t|\,.$$

It follows that:

$$\sum_{t=0}^{T} \frac{2|U_t|}{n-2t} = b\sum_{t=0}^{T} \frac{2}{n-2t} + \sum_{i=1}^{a}\sum_{t=0}^{t_i} \frac{4}{n-2t} = b\sum_{t=0}^{T} \frac{1}{n/2-t} + 2\sum_{i=1}^{a}\sum_{t=0}^{t_i} \frac{1}{n/2-t}.$$

We recall that $\sum_{k=\ell}^{L} 1/k \geq \log\left(L/\ell\right)$ holds whenever $\ell \leq L$. Therefore:

$$\sum_{t=0}^{T} \frac{2|U_t|}{n-2t} \geq b\log\left(\frac{n}{n-2T}\right) + 2\sum_{i=1}^{a} \log\left(\frac{n}{n-2t_i}\right).$$

Plugging this into (4), we obtain:

$$\begin{aligned}
\mathbb{P}[A] &\leq (1 \pm o\,(1)) \left(\prod_{i=1}^{a} \frac{2}{(n-2t_i)^2}\right) \exp\left(b\log\left(\frac{n-2T}{n}\right) + 2\sum_{i=1}^{a} \log\left(\frac{n-2t_i}{n}\right)\right)\\
&\leq (1 \pm o\,(1)) \left(\prod_{i=1}^{a} \left(\frac{2}{(n-2t_i)^2}\left(\frac{n-2t_i}{n}\right)^2\right)\right) \left(1-\frac{2T}{n}\right)^b\\
&\leq (1 \pm o\,(1)) \left(\frac{2}{n^2}\right)^a \left(1-\frac{2T}{n}\right)^b,
\end{aligned}$$

as claimed. □

Claim 2.6 helps us bound the probability that E_T contains a given set of edges:

Lemma 2.7. *Let S be a set of $a \leq \log^2 n$ chords, and let $u, v \in V$ be distinct vertices. The probability that $S \subseteq E_T$ and that $u, v \in W_T$ is $O(n^{-(a+2(1-c-\varepsilon))})$.*

Proof. Let $s_1, \ldots, s_a$ be an ordering of the chords in S. We employ a union bound over all times $t_1, \ldots, t_a$ such that for every i, the chord chosen at step t_i is s_i. By Claim 2.6 the probability that $S \subseteq E_T$ and both u and v have degree 2 in G_T is at most

$$\begin{aligned}
\sum_{0\leq t_1,\ldots,t_a\leq T} (1 \pm o\,(1)) \left(\frac{2}{n^2}\right)^a \left(1-\frac{2T}{n}\right)^2 &= O\left(T^a \left(\frac{2}{n^2}\right)^a \left(\frac{n^{c+\varepsilon}}{n}\right)^2\right)\\
&= O\left(\frac{1}{n^{a+2(1-c-\varepsilon)}}\right),
\end{aligned}$$

as desired. □

Lemma 2.7 allows us to bound the number of edges in $\mathcal{B}_T$. If $uv \in \mathcal{B}_T$, then G_T contains a path P from u to v of length $\leq g-2$ such that:

- No two consecutive edges in P are chords.
- The first and the last edge in P are not chords.

A length-ℓ path in K_n satisfying these conditions is said to be **ℓ-threatening**.

We also introduce several random variables that will allow us to bound the size of $\mathcal{B}_t$ throughout the process. For $\ell \leq g-2$, we denote by $P_\ell(G_t)$ the number of pairs $u, v \in W_t$ such that $\delta_{G_t}(u,v) = \ell$. Additionally, for $v \in V$, we denote by $P_\ell(G_t, v)$ the number of vertices $u \in W_t$ such that $\delta_{G_t}(u,v) = \ell$.

Lemma 2.8. *Let $\ell \in \mathbb{N}$ and let $a \in \mathbb{N}_0$ such that $2a+1 \leq \ell$. Then K_n contains fewer than $n^{a+1}(k-1)^\ell$ ℓ-threatening paths containing a chords.*

Proof. We prove the lemma by considering the number of ℓ-threatening path with a chords. There are n choices for the initial vertex. Since the first edge is not a chord, it must be one of the $k-1$ edges in G that are incident to the initial vertex. Then, for each subsequent edge, there are two possibilities:

- If the previous edge was a chord, the next edge must be one of the $k-1$ edges in G incident to the current vertex.
- Otherwise, the next edge is either one of the $k-2$ non-backtracking edges incident to the current vertex, or else it is a chord. In this case there are $n-k<n$ choices for the chord.

Put differently, at each step, there are $k-1$ basic choices: Either the $k-1$ edges incident to the current vertex, or else the $k-2$ non-backtracking edges incident to the current vertex together with the choice "chord". If the choice is "chord", there are (fewer than) n further choices of the specific chord. Since we are considering length-ℓ paths with a chords, the total number of choices is at most $n^{a+1}(k-1)^\ell$, as desired. □

We next give upper bounds on $P_\ell(G_T)$ and $P_\ell(G_T, v)$ for $\ell \leq g-2$ and $v \in W_T$.

Lemma 2.9. *The following hold w.h.p.:*

(1) For every $\ell \leq g-2$, $P_\ell(G_T) \leq |W_T|^2 \frac{(k-1)^\ell}{n} \log^3(n)$.
(2) For every $\ell \leq g-2$ and every $v \in W_T$, $P_\ell(G_T, v) \leq (k-1)^\ell$.

Proof. We calculate the expected number of length-ℓ paths between vertices in W_T. By Lemma 2.8, there are at most $n^{a+1}(k-1)^\ell$ ℓ-threatening paths in K_n with a chords. By Lemma 2.7, for each such path, the probability that it is contained in $E(G_T)$ and that its two endpoints are in W_T is at most $O\left(n^{-a-2(1-c-\varepsilon)}\right)$. Therefore:

$$\mathbb{E}\left[P_\ell(G_T)\right] = O\left(\sum_{a=0}^{(\ell-1)/2} \frac{n^{a+1}(k-1)^\ell}{n^{a+2(1-c-\varepsilon)}}\right) = O\left(\frac{n^{2(c+\varepsilon)}}{n^2} \sum_{a=0}^{(\ell-1)/2} \frac{n^{a+1}(k-1)^\ell}{n^a}\right)$$
$$= O\left(\frac{|W_T|^2}{n} \sum_{a=1}^{(\ell-1)/2} (k-1)^\ell\right) = O\left(|W_T|^2 \frac{(k-1)^\ell}{n} \log(n)\right).$$

Therefore, by Markov's inequality, for every ℓ, it holds that

$$\mathbb{P}\left[P_\ell(G_T) \geq |W_T|^2 \frac{(k-1)^\ell}{n} \log^3(n)\right] = O\left(\frac{1}{\log^2(n)}\right).$$

Applying a union bound to the $O(\log(n))$ random variables $P_1(G_T), \ldots, P_{g-2}(G_T)$, we conclude that w.h.p., for every $1 \leq \ell \leq g-2$, it holds that

$$P_\ell(G_T) \leq |W_T|^2 \frac{(k-1)^\ell}{n} \log^3(n),$$

as desired.

Part 2 follows from Moore's bound (Observation 2.2). For every $v \in W_T$ there are at most $(k-1)^\ell$ vertices $u \in V$ with $\delta_{G_T}(u,v) = \ell$. In particular, there are at most $(k-1)^\ell$ such vertices in W_T. □

2.2. **The latter evolution of the process.** Let

$$\varepsilon = \frac{c(1-c)}{3}, \quad T = \frac{1}{2}(n - n^{c+\varepsilon}), \quad T_{safe} = \frac{1}{2}(n - n^{\varepsilon}).$$

Our plan is to show that w.h.p. $G_{T_{safe}}$ is safe.

Our analysis of the first T steps of the process used rather crude tools: Moore's bound, and a first-moment calculation. Analyzing the remaining steps of the process is more involved. However, this more involved analysis is not necessary if already G_T is safe. This is indeed the case w.h.p. if $c < 1/3$, as we now show.

Proof of Theorem 1.1 when $c < 1/3$*.* Suppose $c < 1/3$. By Lemma 2.9, w.h.p., for every $1 \leq \ell \leq g-2$, we have:

$$\begin{aligned} P_\ell(G_T) &\leq |W_T|^2 \frac{(k-1)^\ell}{n} \log^3(n) \leq n^{2(c+\varepsilon)} \frac{(k-1)^\ell}{n} \log^3(n) \\ &\leq n^{2c+2\varepsilon} \frac{n^c}{n} \log^3(n) \leq n^{3c+2\varepsilon-1} \log^3(n) = o(1). \end{aligned}$$

Therefore, for every $1 \leq \ell \leq g-2$, it holds that $P_\ell(G_T) = 0$. In other words G_T is safe, as claimed. □

We return to our main narrative with $1 > c \geq 1/3$. We define:

$$\beta = \varepsilon/10, \quad \alpha = \beta/100.$$

We have chosen these particular constants for concreteness; all we need is that β is sufficiently smaller than ε and that α is sufficiently smaller than β. By Lemmas 2.3 and 2.9 the following hold w.h.p. (in fact, 1 - 3 hold with certainty):

1. $T_{freeze} \geq T$.
2. $|W_T| = n - 2T = n^{c+\varepsilon}$.
3. For every $v \in W_T$ and $\ell \leq g-2$, there holds $P_\ell(G_T, v) \leq (k-1)^\ell = \frac{(k-1)^\ell}{n^{c+\varepsilon}}|W_T|$.
4. For every $\ell \leq g-2$, there holds $P_\ell(G_T) \leq \frac{|W_T|^2(k-1)^\ell}{n} \log^3(n)$. In particular, $|\mathcal{B}_T| \leq \frac{|W_T|^2 n^c}{n} \log^4(n)$.

These are pseudorandom properties of G_T: The number of pairs of vertices in W_T at distance $\ell \leq g-2$ does not exceed its expectation by more than a polylog factor, and no vertex in W_T is close to too many other vertices in W_T. As we show, if these pseudorandomness conditions hold at step $T_{safe} \geq t \geq T$, then w.h.p. they persist until step $t' = (n - (n-2t)n^{-\alpha})/2$. In particular, between times t and t' the number of unsaturated vertices gets multiplied by $n^{-\alpha}$, while the number of forbidden pairs is multiplied by $n^{-2\alpha}$ (ignoring polylog factors). If we repeat this process $(c+\varepsilon)/\alpha = O(1)$ times, then w.h.p. no forbidden pairs remain, i.e., the graph is safe. By Lemma 2.5, this implies that the process saturates.

We make this precise in the next lemma, which is the heart of our proof. It is useful to introduce the following real function L:

$$L(\ell, t) = \max\left\{1, \frac{(k-1)^\ell (n-2t)}{n^{c+\varepsilon}}\right\}.$$

We now formally define the pseudorandomness properties.

Definition 2.10. For $C > 0$ we say that G_t is C**-path-bounded** if:

1. $P_\ell(G_t, v) \leq L(\ell, t) \log^C(n)$ for every $v \in W_t$ and every $\ell \leq g-2$.

(2) $P_\ell(G_t) \leq \frac{|W_t|^2(k-1)^\ell}{n}\log^C(n)$ for every $\ell \leq g-2$.

We remark that for every C there exists some $n_0 = n_0(C)$ such that if $n \geq n_0$ and G_t is C-path-bounded for some $T \geq t \geq T_{safe}$, then $|W_t| = n - 2t$. This is because 2 implies that $|\mathcal{B}_t| = O\left(|W_t|^2 n^c \log^C(n)/n\right) = o(|W_t|^2)$. Therefore, if n is large enough then $\mathcal{A}_t$ is not empty, meaning $T_{freeze} \geq t$, and $|W_t| = n-2t$.

Lemma 2.11. *There is a function $D = D(C)$ such that for every $T_{safe} \geq t \geq T$, if G_t is C-path-bounded then, for $t' = (n-(n-2t)n^{-\alpha})/2$, w.h.p. $G_{t'}$ is D-path-bounded.*

Lemma 2.11, yields Proposition 2.1, and hence Theorem 1.1.

Proof of Proposition 2.1. Define the sequence of integers $t_0 = T$, and for $i \geq 0$, $t_{i+1} := (n-(n-2t_i)n^{-\alpha})/2$. Let $m := (c+\varepsilon)/\alpha$. Clearly $m = O(1)$. We observe that for every $i \leq m$, there holds $(n-2t_i)^2 = (n-2t)^2 n^{-2\alpha i}$. In particular, $(n - 2t_m)^2 n^{c-1} = n^{-\Omega(1)}$.

As observed above, $G_{t_0} = G_T$ is $C_0 := 3$-path-bounded. Therefore, by Lemma 2.11, w.h.p. G_{t_1} is $C_1 := D(C_0)$-path-bounded. Proceeding by induction, we conclude that w.h.p. G_{t_m} is C_m-path-bounded, with $C_m = O(1)$. In particular, $|W_{t_m}| = n-2t_m$ and $|\mathcal{B}_{t_m}| \leq |W_{t_m}|^2 n^{c-1}\log^{C_m}(n) = (n-2t_m)^2 n^{c-1} = o\left(1\right)$. Namely, $\mathcal{B}_{t_m} = \emptyset$, i.e., G_{t_m} is safe, and By Lemma 2.5 the process saturates. □

We turn to prove Lemma 2.11. We wish to analyze the $t'-t = |W_t|(1-n^{-\alpha})/2$ steps of the process $G_t, G_{t+1}, \ldots, G_{t'}$. We do so by viewing the process as taking place in two stages: Recall that $H_t = (W_t, \mathcal{A}_t)$ is the graph of available edges. In the first stage, we take a random subgraph $H \subseteq H_t$, where $V(H) = W_t$ and every edge in $E(H_t) = \mathcal{A}_t$ is included in $E(H)$ with probability $p := n^\beta/|W_t|$ (with all choices independent). We also define the graph $G' = (V, E(G_t) \cup E(H))$. In the second stage, we run the high-girth process beginning from G_t, *but using only the edges in* $E(H)$. This is similar to the "honest nibble" used by Grable [19] to analyze random greedy triangle packing.

The advantage of this approach is that we can use standard tools for analyzing random binomial graphs to obtain properties of H and G'. Significantly, there are very few ways that adding a matching from H to G_t might create a cycle shorter than g. This implies that the high-girth process run "inside" G' behaves similarly to the random greedy matching algorithm in H. Finally, H is sufficiently regular that the random greedy matching algorithm in H succeeds, with high probability, in matching all but at most $|W_t|n^{-\alpha}$ vertices in W_t.

Formally, we define the process $G'_t, G'_{t+1}, \ldots$ as follows. To start, $G'_t = G_t$. Given G'_i, if there exist edges $uv \in E(H)$ such that $d_{G'_i}(u) = d_{G'_i}(v) = k-1$, and $\delta_{G'_i}(u,v) \geq g-1$, then choose such an edge e uniformly at random and set $G'_{i+1} = (V(G), E(G'_i) \cup \{e\})$. If no such edges exist, set $G'_{i+1} = G'_i$. Let T'_{freeze} be the smallest integer i such that $G'_i = G'_{i+1}$.

We couple $G_t, G_{t+1}, \ldots, G_{t'}$ and $G'_t, G'_{t+1}, \ldots, G'_{t'}$ by setting $G_i = G'_i$ for every $t < i \leq T'_{freeze}$. For $i > T'_{freeze}$, we obtain G_{i+1} from G_i independently of G'_{i+1}.

Clearly, for every $i \geq t$, it holds that $E(G'_i) \subseteq E(G')$. We will show that w.h.p. $T'_{freeze} \geq t'$, and hence $G_{t'} = G'_{t'}$. It will then follow from the analysis of the process $G'_0, \ldots, G'_{t'}$ that $G_{t'}$ is D-path-bounded for an appropriate D.

In order to track the process $G'_t, G'_{t+1}, \ldots$, we first identify the pairs of vertices $u, v \in W_t$ that *might* have distance $\leq g-2$ in $G'_{t'}$. We observe that since $G'_{t'} \subseteq G'$, if $\delta_{G'_{t'}}(u,v) = \ell \leq g-2$ then there exists a sequence of vertices $w_0, w_1, \ldots, w_{2m-1}$ in W_t such that:

- $w_0 = u$ and $w_{2m-1} = v$.
- For every $1 \leq i \leq m-1$, it holds that $w_{2i-1}w_{2i} \in E(H)$.
- $m - 1 + \delta_{G_t}(w_0, w_1) + \delta_{G_t}(w_2, w_3) + \ldots + \delta_{G_t}(w_{2m-2}, w_{2m-1}) = \ell$.

This is similar to the observation in Section 2.1 that $uv \in \mathcal{B}_T$ only if the chords from a threatening path in K_n were chosen in the first T steps of the process. We say that a pair of vertices $u, v \in W_t$ is **ℓ-threatened** if there exists a sequence of vertices satisfying these conditions. In this case, we say that the sequence $w_0, \ldots, w_{2m-1}$ **witnesses** this fact.

We remark that the notion of an ℓ-threatened pair of vertices is similar, but distinct from, the notion of an ℓ-threatening path. Indeed, the latter refers to the specific *path*, while the former to the endpoints. Furthermore, ℓ-threatening paths are allowed to use any chord from K_n, whereas if $w_0, \ldots, w_{2m-1}$ is a witness that W_0, w_{2m-1} are ℓ-threatened then the edges $w_1w_2, w_3w_4, \ldots, w_{2m-2}w_{2m-1}$ must be in the random graph H. path between ℓ-threatened vertices may only use edges from the random graph H.

For $1 \leq \ell \leq g-2$, let T_ℓ denote the number of ℓ-threatened pairs in W_t. For a vertex $v \in W_t$, let $T_\ell(v)$ denote the number of ℓ-threatened pairs that include v.

In the next claim we establish pseudorandom properties of H and G'. These follow from standard techniques in the analysis of Boolean functions of independent random variables. In order not to interrupt the narrative, we defer the proof to Section 3.

Claim 2.12. *There exists a function $Q = Q(C)$ such that for every $T_{safe} \geq t \geq T$ if G_t is C-path-bounded then, with H and G' defined as above, the following hold w.h.p.:*

(1) For every $v \in W_t$, $d_H(v) = \left(1 \pm n^{-0.4\beta}\right) n^\beta$.

(2) For every $\ell \leq g-2$ it holds that $T_\ell \leq |W_t|^2 \frac{(k-1)^\ell}{n} \log^Q(n)$.

(3) For every $v \in W_t$ and every $\ell \leq g-2$ it holds that $T_\ell(v) \leq L(\ell, t) \log^Q(n)$.

(4) For every $v \in W_t$ there are at most $\log(n)$ vertices $u \in W_t$ such that $uv \in E(H)$ and u, v are ℓ-threatened for some $\ell \leq g-2$.

(5) For every $S \subseteq W_t$ such that $|S| \leq |W_t|/n^{\varepsilon/2}$, it holds that $e\left(H[S]\right) \leq |S| n^{0.9\beta}$.

We turn to establish properties of the process $G'_t, G'_{t+1}, \ldots$. For $s \in \mathbb{N}_0$, let U_s denote the set of degree-$(k-1)$ vertices in G'_{t+s}. Our intuition is that for every $s \leq t' - t$, U_s resembles a random subset of W_t with density $1 - 2s/|W_t|$. This implies, first, that $T'_{freeze} \geq t'$, and therefore $G_{t'} = G'_{t'}$. Second, this means that for $\ell \leq g-2$, the number of ℓ-threatened pairs in G' in which both vertices remain unsaturated in $G_{t'}$ is approximately $\left(|U_{t'}|/|W_t|\right)^2 T_\ell = n^{-2\alpha} T_\ell$. A similar statement holds for ℓ-threatened pairs that contain a specific vertex. We conclude that $G_{t'}$ is, w.h.p., D-path-bounded for an appropriate D.

Claim 2.13. *There exists a function $D = D(Q)$ such that if H and G' satisfy conclusions 1-5 of Claim 2.12 then, for $t' = (n - (n-2t)n^{-\alpha})/2$, w.h.p. $G_{t'}$ is D-path-bounded.*

While Claim 2.13 follows from methods developed to study random greedy hypergraph matching (e.g., [2] and [24, Section 4]), it does not seem to follow directly from any explicit result in the literature. For completeness' sake we prove it in Section 4.

We are now ready to prove Lemma 2.11.

Proof of Lemma 2.11. Suppose that for $T_{safe} \geq t \geq T$ and $C \in \mathbb{R}$, G_t is C-path-bounded. Let H and G' be as above and let $t' = (n-(n-2t)n^{-\alpha})/2$ be as in the statement of Lemma 2.11. Then, w.h.p. H and G' satisfy the conclusions of Claim 2.12 for some $Q = Q(C)$. Consequently, by Claim 2.13, w.h.p. $G_{t'}$ is D-path-bounded for some $D = D(Q)$. □

3. Proof of Claim 2.12

Claim 2.12 follows from standard arguments in the analysis of functions of independent random variables. We recall the following version of Chernoff's inequality.

Theorem 3.1 (Chernoff's Inequality)**.** *Let $X_1, X_2, \ldots, X_N$ be independent Bernoulli random variables, let $X = \sum_{i=1}^{N} X_i$, and let $\delta \in (0,1)$. Then:*

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \delta \mathbb{E}[X]\right] \leq 2\exp\left(-\frac{1}{3}\delta^2 \mathbb{E}[X]\right).$$

We use a theorem of Kim and Vu to show concentration of multivariate polynomials. The setup is this: Let $Y = (V(Y), E(Y))$ be a hypergraph where the largest hyperedge size is $K = O(1)$ and $|V(Y)| = \omega(1)$. Let $\{X_v\}_{v \in V(Y)}$ be a collection of independent Bernoulli random variables. Consider the random variable:

$$X = \sum_{e \in E(Y)} \prod_{v \in e} X_v.$$

For every $S \subseteq V(Y)$, define the random variable:

$$X_S = \sum_{S \subseteq e \in E(Y)} \prod_{v \in e \setminus S} X_v.$$

For every $i \in [K]_0$ let

$$\mu_i = \max_{S \in \binom{V(Y)}{i}} \mathbb{E}[X_S].$$

Finally, let $\mu = \max_{0 \leq i \leq K} \mu_i$. Here is a consequence of the Main Theorem in [22].

Theorem 3.2. *In the setup above, there exists a constant $D > 0$ such that*

$$\mathbb{P}\left[X \geq \mu \log^D(|V(Y)|)\right] = \exp\left(-\Omega\left(\log^2(|V(Y)|)\right)\right).$$

Proof of Claim 2.12. Recall that we have defined the function

$$L(\ell, t) = \max\left\{1, \frac{(k-1)^\ell (n-2t)}{n^{c+\varepsilon}}\right\} = \max\left\{1, \frac{(k-1)^\ell |W_t|}{n^{c+\varepsilon}}\right\}.$$

For brevity, throughout this proof we write

$$L(\ell) := L(\ell, t).$$

By assumption, for every $v \in W_t$ and every $\ell \leq g-2$, it holds that $P_\ell(G_t, v) \leq L(\ell)\log^C(n)$. Therefore, the number of forbidden edges incident to v does not exceed

$$\sum_{\ell=1}^{g-2} P_\ell(G_t, v) \leq \sum_{\ell=1}^{g-2} L(\ell)\log^C(n) \leq \log^C(n) \sum_{\ell=1}^{g-2}\left(1 + \frac{(k-1)^\ell |W_t|}{n^{c+\varepsilon}}\right)$$
$$\leq \left(1 + \frac{|W_t|}{n^{\varepsilon}}\right)\log^{C+1}(n).$$

Recalling that $t \leq T_{safe} = (n - n^{\varepsilon})/2$, it follows that $|W_t| \geq n^{\varepsilon}$. Thus:

$$\sum_{\ell=1}^{g-2} P_{\ell}(G_t, v) \leq \frac{|W_t|}{n^{\varepsilon}} \log^{C+2}(n).$$

Therefore:

$$d_{H_t}(v) = |W_t| \pm \frac{|W_t|}{n^{\varepsilon}} \log^{C+2}(n).$$

By definition, $d_H(v)$ is distributed binomially with parameters $(d_{H_t}(v), p)$. In particular, $\mathbb{E}\left[d_H(v)\right] = (1 \pm n^{-\varepsilon} \log^{C+2}(n))n^{\beta}$. Applying Chernoff's inequality we obtain:

$$\mathbb{P}\left[|d_H(v) - pd_{H_t}(v)| > \frac{n^{0.6\beta}}{2|W_t|} d_{H_t}(v)\right] \leq \exp\left(-\Omega\left(n^{0.2\beta}\right)\right).$$

Next apply a union bound to the vertices in W_t, and conclude that w.h.p. for every $v \in W_t$:

$$d_H(v) = pd_{H_t}(v) \pm \frac{n^{0.6\beta}}{2|W_t|} d_{H_t}(v) = \left(1 \pm n^{-0.4\beta}\right) n^{\beta},$$

as desired.

We now prove 2. Suppose that $u, v \in W_t$ are ℓ-threatened for some $\ell \leq g - 2$. Then there is a sequence $u = w_0, \ldots, w_{2m-1} = v$ in W_t witnessing this fact. Now, the number of sequences $w_0, \ldots, w_{2m-1} \in W_t$ such that $\delta_{G_t}(w_0, w_1) + \delta_{G_t}(w_2, w_3) + \ldots + \delta_{G_t}(w_{2m-2}, w_{2m-1}) = \ell - m + 1$ is bounded from above by

$$\sum_{\ell_1 + \ldots + \ell_m = \ell - m + 1} \prod_{i=1}^{m} P_{\ell_i}(G_t)$$

For each such sequence, the probability that all of the edges $w_1 w_2, \ldots, w_{2m-3} w_{2m-2}$ are in $E(H)$ is p^{m-1}. This yields the following bound:

$$\begin{aligned}
\mathbb{E}\left[T_{\ell}\right] &\leq \sum_{m=1}^{(\ell+1)/2} \sum_{\ell_1 + \ldots + \ell_m = \ell - m + 1} p^{m-1} \prod_{i=1}^{m} P_{\ell_i}(G_t) \\
&\leq \sum_{m=1}^{(\ell+1)/2} \sum_{\ell_1 + \ldots + \ell_m = \ell - m + 1} p^{m-1} \prod_{i=1}^{m} \frac{(k-1)^{\ell_i}}{n} |W_t|^2 \log^C(n) \\
&\leq \frac{(k-1)^{\ell}}{p} \sum_{m=1}^{(\ell+1)/2} \left(\frac{p|W_t|^2 \log^C(n)}{n}\right)^m \binom{\ell - m}{m-1} \\
&\leq \frac{(k-1)^{\ell}}{p} \sum_{m=1}^{(\ell+1)/2} \left(\frac{n^{\beta}|W_t| \ell \log^C(n)}{n}\right)^m = O\left(\frac{(k-1)^{\ell}}{n} |W_t|^2 \log^{C+1}(n)\right).
\end{aligned}$$

Hence, by Markov's inequality:

$$\mathbb{P}\left[T_{\ell} \geq \frac{(k-1)^{\ell}}{n} |W_t|^2 \log^{C+3}(n)\right] = O\left(\frac{1}{\log^2(n)}\right).$$

Applying a union bound to the $O(\log(n))$ random variables $T_1, \ldots, T_{g-2}$ we conclude that w.h.p., for every $\ell \leq g - 2$, it holds that $T_{\ell} \leq \frac{(k-1)^{\ell}}{n} |W_t|^2 \log^{C+3}(n)$, as desired.

In order to prove 3, we must bound $|W_t|(g-2) = O(|W_t|\log(n))$ random variables. For a union bound, it suffices to show that there exists a constant $D = D(C)$ such that for every $v \in W_t$ and $\ell \leq g-2$,

$$\mathbb{P}\left[T_\ell(v) \geq L(\ell)\log^D(n)\right] = O\left(\frac{1}{|W_t|\log^2(n)}\right).$$

Let $v \in W_t$, $\ell \leq g-2$, and $m \in \mathbb{N}$. Let $T_\ell^m(v)$ be the number of vertices $u \in W_t$ such that u, v are ℓ-threatened and there exists a witness with $2m$ vertices. Then $T_\ell(v) \leq \sum_{m=1}^{(\ell+1)/2} T_\ell^m(v)$. We plan to use Theorem 3.2 to bound $T_\ell(v)$. However, Theorem 3.2 applies only to polynomials with constant degree, whereas $T_\ell(v)$ corresponds to a polynomial with unbounded degree. To get around this, we first show that for large values of m, $T_\ell^m(v) = 0$ with sufficiently high probability. Let $M = \lceil (1-c-\varepsilon-\beta)^{-1}\rceil = O(1)$. We will show that

$$\mathbb{P}\left[\sum_{m=M}^{(\ell+1)/2} T_\ell^m(v) \geq L(\ell)\log^{C+1}(n)\right] = O\left(\frac{1}{|W_t|\log^2(n)}\right). \tag{5}$$

We first observe that:

$$\begin{aligned}\left(\frac{|W_t|n^\beta\log^{C+1}(n)}{n}\right)^{M-1} &\leq \frac{1}{|W_t|}|W_t|^M n^{-(M-1)(1-\beta)}\log^{O(1)}(n)\\ \leq \frac{1}{|W_t|}n^{M(c+\varepsilon)-(M-1)(1-\beta)}\log^{O(1)}(n) &\leq \frac{1}{|W_t|}n^{1-\beta-M(1-c-\varepsilon-\beta)}\log^{O(1)}(n)\\ \leq \frac{n^{-\Omega(1)}}{|W_t|} \leq \frac{1}{|W_t|\log^2(n)}.\end{aligned} \tag{6}$$

Now, by considerations similar to those used to bound $\mathbb{E}[T_\ell]$:

$$\begin{aligned}\mathbb{E}\left[\sum_{m=M}^{(\ell+1)/2} T_\ell^m(v)\right] &\leq \sum_{m=M}^{(\ell+1)/2} p^{m-1}\sum_{\ell_1+\ldots+\ell_m=\ell-m+1} P_{\ell_1}(G_t,v)\prod_{i=2}^m P_{\ell_i}(G_t)\\ &\leq \sum_{m=M}^{(\ell+1)/2} L(\ell)\log^C(n)\left(\ell p\frac{|W_t|^2\log^C(n)}{n}\right)^{m-1}\\ &\leq L(\ell)\log^{C+1}(n)\left(\frac{|W_t|n^\beta\log^{C+1}(n)}{n}\right)^{M-1}\\ &\overset{(6)}{\leq} L(\ell)\log^{C+1}(n)\frac{1}{|W_t|\log^2(n)}.\end{aligned}$$

Inequality (5) follows from Markov's inequality.

We now use Theorem 3.2 to bound $T_\ell^m(v)$, for $m < M$. For every $e \in E(H_t)$, let X_e be the indicator of the event $e \in E(H)$. For $m < M$, let $\mathcal{T}_\ell^m(v)$ be the collection of *potential* witnesses of length $2m$ to the fact that for some $u \in W_t$, u, v are ℓ-threatened. In other words, $\mathcal{T}_\ell^m(v)$ is the collection of sequences $v = w_0, w_1, \ldots, w_{2m-1} \in W_t$ such that:

- For every $1 \leq i \leq m-1$, $w_{2i-1}w_{2i} \in E(H_t)$ and
- $\delta_{G_t}(w_0, w_1) + \delta_{G_t}(w_2, w_3) + \ldots + \delta_{G_t}(w_{2m-2}, w_{2m-1}) = \ell - m + 1$.

For conciseness, for $P \in \mathcal{T}_\ell^m(v)$ and $i \in [m-1]$, we write $e_i(P) = w_{2i-1}w_{2i}$. We also write $\ell_i(P) = \delta_{G_t}(w_{2(i-1)}, w_{2i-1})$. Now consider the polynomial

$$Y_\ell(v) = \sum_{m=1}^{M-1} T_\ell^m(v) = \sum_{m=1}^{M-1} \sum_{P \in \mathcal{T}_\ell^m(v)} \prod_{i=1}^{m-1} X_{e_i(P)}.$$

We will show that there exists a constant $D > 0$ (independent of ℓ and v) such that:

$$\mathbb{P}\left[Y_\ell(v) \geq L(\ell)\log^D(n)\right] = \exp\left(-\Omega\left(\log^2(n)\right)\right).$$

For $S \subseteq E(H_t)$ and $m \in [M-1]$, define the set:

$$\mathcal{T}_\ell^m(v, S) = \left\{P \in \mathcal{T}_\ell^m(v) : S \subseteq \{e_1(P), \dots, e_{m-1}(P)\}\right\}.$$

Since $\deg Y_\ell(v) < M = O(1)$, by Theorem 3.2 it suffices to show that there exists a constant B such that for every $S \subseteq E(H_t)$ of cardinality $M-1$ or less, it holds that

$$\mathbb{E}\left[Y_\ell(v)_S\right] = \sum_{m=1}^{M-1} p^{m-1-|S|} \left|\mathcal{T}_\ell^m(v, S)\right| = O\left(L(\ell)\log^B(n)\right).$$

For this it is enough to show that for every $S \subseteq E(H_t)$ and every $m < M$:

$$\left|\mathcal{T}_\ell^m(v, S)\right| = O\left(p^{|S|-m+1}L(\ell)\log^B(n)\right). \tag{7}$$

Let $S \subseteq E(H_t)$ satisfy $|S| \leq M-1$. We make the following observations: Suppose that for $m < M-1$, $P = w_0, \dots, w_{2m-1} \in \mathcal{T}_\ell^m$ satisfies $S \subseteq \{e_1(P), \dots, e_{m-1}(P)\}$. In particular, $m \geq |S| + 1$. Furthermore, there exists an index set $I \in \binom{[m-1]}{|S|}$ such that for every $i \in I$, $e_i(P) \in S$. This implies that for every $i \in I$, w_{2i} is contained in one of the edges in S. Therefore, since G_t is C-path bounded, w_{2i+1} is one of the $O(P_{\ell_i(P)}(G_t, w_{2i})) = O(L(\ell_i(P))\log^C(n))$ vertices at distance $\ell_i(P)$ from S. Using these insights, we may now bound $|\mathcal{T}_\ell^m(v, S)|$.

Start with the case where $|S| = m-1$. In this case, for every $P = w_0, \dots, w_{2m-1} \in \mathcal{T}_\ell^m(v)$, it holds that $S = \{e_1(P), \dots, e_{m-1}(P)\}$. Therefore, each of $w_1, \dots, w_{2m-2}$ is contained in an edge in S, and the only remaining freedom is in the choice of w_{2m-1}. Additionally, w_{2m-1} is at distance at most ℓ from S. As mentioned, there are $O(L(\ell)\log^C(n))$ such vertices. Therefore, in this case:

$$\left|\mathcal{T}_\ell^m(v, S)\right| = O\left(L(\ell)\log^C(n)\right),$$

confirming (7).

We now assume that $|S| < m-1$. In this case:

$$\begin{aligned}&\left|\mathcal{T}_\ell^m(v, S)\right| \\ &\leq \sum_{\ell_1+\ldots+\ell_m=\ell-m+1} P_{\ell_1}(G_t, v) \sum_{I \in \binom{[m-1]}{|S|}} \left(\prod_{i-1 \in I} O\left(L(\ell_i)\log^C(n)\right)\right)\left(\prod_{i-1 \in [m-1]\setminus I} P_{\ell_i}(G_t)\right).\end{aligned}$$

Since G_t is C-path bounded, for every $a \leq g-2$ it holds that

$$P_a(G_t) \leq \frac{(k-1)^a |W_t|^2 \log^C(n)}{n}.$$

Similarly, $P_a(G_t,v) \le L(a)\log^C(n) \le (k-1)^a \log^C(n)$. Therefore, for every $\ell_1+\ldots+\ell_m = \ell-m+1$ and every $I \in \binom{[m-1]}{|S|}$ it holds that:

$$\begin{aligned}
&P_{\ell_1}(G_t,v)\left(\prod_{i-1\in I} O\left(L(\ell_i)\log^C(n)\right)\right)\left(\prod_{i-1\in[m-1]\setminus I} P_{\ell_i}(G_t)\right)\\
&\quad= O\left(\log^{Cm}(n)(k-1)^{\ell_1+\ldots+\ell_m}\left(\frac{|W_t|^2}{n}\right)^{m-1-|S|}\right)\\
&\quad= O\left(\log^{Cm}(n)(k-1)^{\ell}\left(\frac{|W_t|^2}{n}\right)^{m-1-|S|}\right).
\end{aligned}$$

Furthermore, it holds that $\binom{m-1}{|S|} = O(1)$. Thus:

$$\begin{aligned}
&|\mathcal{T}_\ell^m(v,S)|\\
&= O\left(\log^{Cm}(n)\sum_{\ell_1+\ldots+\ell_m=\ell-m+1}(k-1)^\ell\left(\frac{|W_t|^2}{n}\right)^{m-1-|S|}\right)\\
&= O\left(\log^{Cm}(n)\left(\frac{n^\beta|W_t|}{pn}\right)^{m-1-|S|}\ell^m(k-1)^\ell\right)\\
&= O\left(p^{|S|-m+1}\log^{(C+1)m}(n)\frac{n^\beta|W_t|}{n}(k-1)^\ell\right)\\
&= O\left(p^{|S|-m+1}\frac{(k-1)^\ell|W_t|}{n^{c+\varepsilon}}\right) = O\left(p^{|S|-m+1}L(\ell)\right),
\end{aligned}$$

as desired.

We prove 4 by exposing $E(H)$ in two rounds: Let $v \in W_t$, and let $E(H_t,v)$ denote the set of edges in $E(H_t)$ that are incident to v. We first expose $E_1 := E(H)\setminus E(H_t,v)$, and then $E_2 := E(H)\cap E(H_t,v)$. We note that E_1 and E_2 are independent. Furthermore, the random variables $Y_1(v), Y_2(v), \ldots, Y_{g-2}(v)$, as well as the set $W \subseteq W_t$ of vertices $u \in W_t$ such that u,v are ℓ-threatened are determined by E_1. From the proof of 3 it follows that there exists a constant $D = D(C)$ such that $\mathbb{P}\left[|W| \ge \frac{|W_t|\log^D(n)}{n^\varepsilon}\right] = o\left(|W_t|^{-1}\right)$.

We want to bound $|W \cap \Gamma_H(v)|$. We observe that given E_1 (and hence W), $W\cap\Gamma_H(v)$ is a binomial random subset of W with density parameter p. Therefore, for any $s \le \frac{|W_t|\log^D(n)}{n^\varepsilon}$, conditioned on $|W| = s$, it holds that:

$$\begin{aligned}
\mathbb{P}\left[|W\cap\Gamma_H(v)| \ge \log(n)\right] &\le \binom{s}{\log(n)}p^{\log(n)} \le (sp)^{\log(n)}\\
&\le \left(\frac{|W_t|\log^D(n)}{n^\varepsilon}\cdot\frac{n^\beta}{|W_t|}\right)^{\log(n)} = \exp\left(-\Omega\left(\log^2(n)\right)\right).
\end{aligned}$$

Therefore, using the law of total probability:

$$\begin{aligned}
&\mathbb{P}\left[|W \cap \Gamma_H(v)| \geq \log(n)\right] \\
\leq &\mathbb{P}\left[|W \cap \Gamma_H(v)| \geq \log(n) \middle| |W| \leq \frac{|W_t| \log^D(n)}{n^\varepsilon}\right] \mathbb{P}\left[|W| \leq \frac{|W_t| \log^D(n)}{n^\varepsilon}\right] \\
&+ \mathbb{P}\left[|W| > \frac{|W_t| \log^D(n)}{n^\varepsilon}\right] = o\left(\frac{1}{|W_t|}\right).
\end{aligned}$$

By applying a union bound to all $|W_t|$ vertices, we conclude that w.h.p., for every $v \in W_t$, $|W \cap \Gamma_H(v) \leq \log(n)$, as desired.

Finally, we prove 5. For $\emptyset \neq S \subseteq W_t$ such that $|S| \leq |W_t|/n^{\varepsilon/2}$, let X_S be the indicator of the event that $e\left(H[S]\right) \geq |S|n^{0.9\beta}$. Now, $e\left(H[S]\right)$ is distributed binomially with parameters $e\left(H_t[S]\right) \leq |S|^2/2$ and p. Hence, by a union bound:

$$\mathbb{E}[X_S] = \mathbb{P}\left[e\left(H[S]\right) \geq |S|n^{0.9\beta}\right] \leq \binom{|S|^2/2}{|S|n^{0.9\beta}} p^{|S|n^{0.9\beta}}.$$

Applying the inequality $\binom{a}{b} \leq (ea/b)^b$, we obtain:

$$\mathbb{E}[X_S] \leq \left(\frac{e|S|^2 p}{|S|n^{0.9\beta}}\right)^{|S|n^{0.9\beta}} \leq \left(\frac{e|S|n^\beta}{n^{0.9\beta}|W_t|}\right)^{|S|n^{0.9\beta}} \leq \left(\frac{en^{0.1\beta}}{n^{\varepsilon/2}}\right)^{|S|n^{0.9\beta}}.$$

Applying a union bound over all such sets S, we have:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{k=1}^{|W_t|n^{-\varepsilon/2}} \sum_{S \in \binom{W_t}{k}} X_S\right] \leq \sum_{k=1}^{|W_t|n^{-\varepsilon/2}} \binom{|W_t|}{k} \left(\frac{en^{0.1\beta}}{n^{\varepsilon/2}}\right)^{kn^{0.9\beta}} \\
&\leq \sum_{k=1}^{|W_t|n^{-\varepsilon/2}} \left(\frac{e|W_t|}{k} \left(\frac{e}{n^{\varepsilon/2-0.1\beta}}\right)^{n^{0.9\beta}}\right)^k \leq \sum_{k=1}^{|W_t|n^{-\varepsilon/2}} \left(\left(\frac{2e}{n^{\varepsilon/2-0.1\beta}}\right)^{n^{0.9\beta}}\right)^k = o(1).
\end{aligned}$$

By Markov's inequality, w.h.p., for every such set S, it holds that $X_S = 0$, which is equivalent to 5. □

4. Proof of Claim 2.13

We prove Claim 2.13 by showing that w.h.p. various parameters associated with the process $G'_t, G'_{t+1}, \ldots$ remain close to their expected trajectories. This is motivated by the differential equation method of Wormald [36], and is similar to the approach taken by Bennett and Bohman [2] in their analysis of random greedy hypergraph matching. As similar results are abundant in the literature, we aim for the simplest exposition and not the sharpest analysis.

We use the following supermartingale inequality of Warnke [34, Lemma 2.2 and Remark 10]. This is a variation on a martingale inequality of Freedman [14, Theorem 1.6].

Theorem 4.1. *Let $X_0, X_1, \ldots$ be a supermartingale with respect to a filtration $\mathcal{F}_0, \mathcal{F}_1, \ldots$. Suppose that $|X_{i+1} - X_i| \leq K$ for all i, and let $V(j) = \sum_{i=0}^{j-1} \mathbb{E}\left[(X_{i+1} - X_i)^2 \,|\mathcal{F}_i\right]$. Then, for any $\lambda, v > 0$,*

$$\mathbb{P}\left[X_i > X_0 + \lambda \text{ and } V(i) \leq v \text{ for some } i\right] \leq \exp\left(-\frac{\lambda^2}{2(v + K\lambda/3)}\right).$$

In our application, we find some v such that $V(i) \leq v$ for all i. For this v, Theorem 4.1 tells us that for every $\lambda > 0$:

$$\mathbb{P}\left[X_i > X_0 + \lambda\right] \leq \exp\left(-\frac{\lambda^2}{2(v + K\lambda/3)}\right).$$

We now introduce the random variables we wish to track. Recall that U_s is the set of unsaturated (i.e., degree-$(k-1)$) vertices in G'_{t+s}, where s is a non-negative integer. In particular, $U_0 = W_t$ and for every $s \leq T'_{freeze} - t$, it holds that $U_s = W_{t+s}$ and $|U_s| = |W_t| - 2s$. For a vertex $v \in W_t$, let $N(v,s) = |\Gamma_H(v) \cap U_s|$ be the number of neighbors of v in H that are unsaturated at time $s+t$. We also define the functions

$$p(s) = 1 - \frac{2s}{|W_t|}, \quad n(s) = n^{\beta} p(s), \quad \varepsilon(s) = \frac{n^{0.6\beta}}{p(s)^8}.$$

We observe that for $0 \leq T'_{freeze} - t$ it holds that $p(s) = |U_s|/|W_t|$.

Recall that $t' = (n - |W_t| n^{-\alpha})/2$. We show that w.h.p., for every $v \in W_t$ and every $0 \leq s \leq t' - t$, it holds that

$$N(v,s) = n(s) \pm \varepsilon(s). \tag{8}$$

The guiding intuition is that $\Gamma_H(v) \cap U_s$ behaves like a random subset of $\Gamma_H(v)$ with density $p(s)$. Since by Claim 2.12 1 $\Gamma_H(v) \approx n^{\beta}$, it follows that $N(v,s)$ is approximated by $n(s)$.

We define the stopping time τ as the minimum between $t' - t$ and the first time that (8) fails for some $v \in W_t$.

A naive attempt to prove (8) might be to show that $N(v,s) - n(s)$ is a martingale. However, this is not quite true, as the expected one-step change might be non-zero. To remedy this, we consider two *shifted* random variables that are obtained from $N(v,s) - n(s)$ and $-(N(v,s) - n(s))$, respectively, by subtracting an error term. These turn out to be supermartingales, enabling us to apply Theorem 4.1. For every $v \in W_t$, we define:

$$N^+(v,s) = \begin{cases} N(v,s) - n(s) - \frac{1}{2}\varepsilon(s) & s \leq \tau \\ N^+(v,s-1) & s > \tau \end{cases},$$

$$N^-(v,s) = \begin{cases} -N(v,s) + n(s) - \frac{1}{2}\varepsilon(s) & s \leq \tau \\ N^-(v,s-1) & s > \tau \end{cases}.$$

The fact that these random variables "freeze" at time τ is crucial as it allows us to assume that (8) holds when calculating the maximal and expected one-step changes.

We say that $uv \in E(H)$ is **available at time** s if $u, v \in U_s$ and $\delta_{G'_{t+s}}(u,v) \geq g-1$. For $v \in U_s$, let $A(v,s)$ be the set of available edges at time s that are incident to v. We note that $A(v,s) \subseteq \{vu : u \in \Gamma_H(v) \cap U_s\}$ (the cardinality of this last set is $N(v,s)$). In general, this inclusion might be strict because there may be vertices $u \in \Gamma_H(v) \cap U_s$ with $\delta_{G'_{t+s}}(u,v) < g-1$ (however, it is true that $A(v,0) = \{vu : u \in \Gamma_H(v) \cap U_0\}$). Nevertheless, the difference between the sets is small.

Claim 4.2. *For every $0 \leq s \leq t' - t$ and every $v \in U_s$, it holds that $|A(v,s)| \geq N(v,s) - \log(n)$. Consequently, if $\tau > s$, then $|A(v,s)| = n(s) \pm 1.1\varepsilon(s)$.*

Proof. By assumption, G' satisfies the conclusions of Claim 2.12. In particular, 4 implies that there are at most $\log(n)$ vertices $u \in U_s \subseteq W_t$ such that $u \in \Gamma_H(v)$ and such that u, v are ℓ-threatened for some $\ell < g-1$. Since G'_{t+s} is a subgraph of G', this holds for G'_{t+s} as well, and the claim follows.

We now observe that for every $s \le t'-t$, $\varepsilon(s) \ge \varepsilon(0) = n^{0.6\beta} > 10\log(n)$. Therefore, if $\tau > s$:

$$|A(v,s)| = N(v,s) \pm \log(n) = n(s) \pm (\varepsilon(s) + \log(n)) = n(s) \pm 1.1\varepsilon(s).$$

□

Claim 4.3. *For every $v \in W_t$, the sequences $\{N^+(v,s)\}_{s=0}^{\infty}$ and $\{N^-(v,s)\}_{s=0}^{\infty}$ are supermartingales with respect to the filtration induced by $\{G'_{t+s}\}_{s=0}^{\infty}$. Furthermore, for every $0 \le s \le t'-t$, it holds that*

$$\mathbb{E}\left[\left|N^-(v,s+1) - N^-(v,s)\right|\right], \mathbb{E}\left[\left|N^+(v,s+1) - N^+(v,s)\right|\right] \le \frac{5n^\beta}{|W_t|}.$$

Proof. We show that $\{N^+(v,s)\}_{s=0}^{\infty}$ is a supermartingale for every $v \in W_t$. The proof for $\{N^-(v,s)\}_{s=0}^{\infty}$ is similar. We need to show that for every $s \ge 0$, it holds that

$$\mathbb{E}\left[N^+(v,s+1) - N^+(v,s) | G'_t, G'_{t+1}, \ldots, G'_{t+s}\right] \le 0. \tag{9}$$

We apply the law of total probability. We first observe that if $\tau \le s$, then, by definition, $N^+(v,s+1) = N^+(v,s)$, so (9) holds. It thus suffices to show that

$$\mathbb{E}\left[N^+(v,s+1) - N^+(v,s) | G'_{t+s} \wedge \tau \ge s+1\right] \le 0.$$

We therefore assume that $\tau \ge s+1$. By Claim 4.2 this implies that for every $u \in U_s$, it holds that $|A(u,s)| = n(s) \pm 1.1\varepsilon(s)$. In particular, this holds for every vertex in $\Gamma_H(v) \cap U_s$. Finally, the number of available edges is equal to $|U_s|\left(n(s) \pm 1.1\varepsilon(s)\right)/2$. Therefore:

$$\mathbb{E}\left[N(v,s+1) - N(v,s) | G'_{t+s} \wedge \tau \ge s+1\right] = -\frac{2}{|U_s|\left(n(s) \pm 1.1\varepsilon(s)\right)} \sum_{u \in \Gamma_H(v) \cap U_s} |A(u,s)|$$
$$= -\frac{2N(v,s)(n(s) \pm 1.1\varepsilon(s))}{|U_s|(n(s) \pm 1.1\varepsilon(s)} = -\frac{2(n(s) \pm 1.1\varepsilon(s))^2}{|U_s|(n(s) \pm 1.1\varepsilon(s))} = -\frac{2n(s)}{|U_s|}\left(1 \pm \frac{3.5\varepsilon(s)}{n(s)}\right)$$
$$= -\frac{2n^\beta}{|W_t|} \pm \frac{7\varepsilon(s)}{|U_s|}.$$

Next, we observe that:

$$n(s+1) - n(s) = -\frac{2n^\beta}{|W_t|}.$$

Finally, we note that:

$$\varepsilon(s+1) - \varepsilon(s) = n^{0.6\beta}\left(\frac{1}{p(s+1)^8} - \frac{1}{p(s)^8}\right) = \frac{n^{0.6\beta}}{p(s)^8}\left(\frac{p(s)^8}{p(s+1)^8} - 1\right)$$
$$= \varepsilon(s)\left(\left(\frac{|W_t| - 2s}{|W_t| - 2s - 2}\right)^8 - 1\right)$$
$$= \varepsilon(s)\left(\left(1 - \frac{2}{|U_s|}\right)^{-8} - 1\right).$$

Hence, by Taylor's Theorem (recalling that $|U_s| \geq n^{-\alpha}|W_t| = \omega(1)$):

$$\varepsilon(s+1) - \varepsilon(s) = \varepsilon(s)\left(\frac{16}{|U_s|} + O\left(\frac{1}{|U_s|^2}\right)\right) \in \left[\frac{16\varepsilon(s)}{|U_s|}, \frac{18\varepsilon(s)}{|U_s|}\right].$$

Therefore:

$$\begin{aligned}
&\mathbb{E}\left[N^+(v,s+1) - N^+(v,s)|G'_{t+s} \wedge \tau \geq s+1\right] \\
&= \mathbb{E}\left[N(v,s+1) - N(v,s)|G'_{t+s} \wedge \tau \geq s+1\right] - (n(s+1) - n(s)) - \frac{1}{2}(\varepsilon(s+1) - \varepsilon(s)) \\
&\leq \left(-\frac{2n^\beta}{|W_t|} \pm \frac{7\varepsilon(s)}{|U_s|}\right) + \frac{2n^\beta}{|W_t|} - \frac{8\varepsilon(s)}{|U_s|} \leq 0,
\end{aligned}$$

as desired.

We also observe that the estimates above imply:

$$\begin{aligned}
&\mathbb{E}\left[\left|N^+(v,s+1) - N^+(v,s)\right|\right] \\
&\leq \mathbb{E}\left[N(v,s) - N(v,s+1)\right] + n(s) - n(s+1) + \frac{1}{2}(\varepsilon(s+1) - \varepsilon(s)) \\
&\leq \frac{2n^\beta}{|W_t|} + \frac{7\varepsilon(s)}{|U_s|} + \frac{2n^\beta}{|W_t|} + \frac{9\varepsilon(s)}{|U_s|} \leq \frac{4n^\beta}{|W_t|} + \frac{16\varepsilon(t'-t)}{|U_{t'-t}|} = \frac{4n^\beta}{|W_t|} + \frac{16n^{0.6\beta}}{n^{-8\alpha}|W_t|}.
\end{aligned}$$

Recalling that $\alpha = \beta/100$, we obtain:

$$\mathbb{E}\left[\left|N^+(v,s+1) - N^+(v,s)\right|\right] \leq \frac{5n^\beta}{|W_t|},$$

as claimed. □

In order to apply Theorem 4.1, we first note that the maximal one-step change in $N(v,s)$ is 2. Furthermore, for every $s \leq t'-t$, $|n(s+1)-n(s)|, \varepsilon(s+1)-\varepsilon(s) = o(1)$. Therefore, the maximal one-step change in $N^+(v,s)$ and $N^-(v,s)$ is bounded from above by 3. Hence, for every $1 \leq s \leq t'-t$, it holds that:

$$\begin{aligned}
V(s) \leq V(t'-t) &\leq 3\sum_{i=0}^{t'-t-1} \mathbb{E}\left[\left|N^+(v,i+1) - N^+(v,i)\right| |G'_{t+i}\right] \\
&\leq 3(t'-t)\frac{5n^\beta}{|W_t|} = O\left(n^\beta\right).
\end{aligned}$$

By applying Theorem 4.1 with $K = 3$, $\lambda = \varepsilon(s)/2$ and $v = n^\beta \log(n)$, we conclude that for every $w \in W_t$ and every $0 \leq s \leq t'-t$:

$$\mathbb{P}\left[N^+(w,s) \geq \varepsilon(s)/2\right] \leq \exp\left(-\frac{\varepsilon(s)^2/4}{2(n^\beta\log(n) + 3\varepsilon(s)/2)}\right) = \exp\left(-\Omega\left(n^{\beta/100}\right)\right).$$

Applying a union bound to the $O\left(|W_t|^2\right)$ choices for v and s, we conclude that w.h.p., for every $v \in W_t$ and $0 \leq s \leq t'-t$, it holds that $N^+(v,s) \leq \frac{1}{2}\varepsilon(s)$. A similar calculation implies the analogous result for $N^-(v,s)$. This implies that $\tau \geq t'-t$. therefore, w.h.p., for every $v \in W_t$ and $0 \leq s \leq t'-t$, it holds that

$$N(v,s) = n(s) \pm \varepsilon(s).$$

In particular, this implies that $T'_{freeze} \geq t'$. Therefore, $G_{t'} = G'_{t'}$. Thus $U_{t'-t} = W_{t'}$.

In order to show that $G_{t'}$ is path-bounded we estimate the probability that a given set of vertices is in $U_{t'-t}$.

Claim 4.4. *Let $A \subseteq W_t$ satisfy $|A| \le |W_t|/n^{\varepsilon/2}$. Then:*

$$\mathbb{P}\left[A \subseteq U_{t'-t}\right] = (1 \pm o\left(1\right))n^{-\alpha|A|}.$$

Proof. Similar to the proof of Claim 2.6, we denote by B_s the event that $A \subseteq U_s$, and observe that:

$$\mathbb{P}\left[A \subseteq U_{t'-t}\right] = \mathbb{P}\left[B_1\right] \times \mathbb{P}\left[B_2|B_1\right] \times \ldots \times \mathbb{P}\left[B_{t'-t}|B_{t'-t-1}\right].$$

Using the law of total probability, for every $s \le t' - t$ it holds that

$$\mathbb{P}\left[B_s|B_{s-1}\right] = \mathbb{P}\left[B_s|B_{s-1} \wedge \tau \ge s\right]\mathbb{P}\left[\tau \ge s\right] + \mathbb{P}\left[B_s|B_{s-1} \wedge \tau < s\right]\mathbb{P}\left[\tau < s\right].$$

Now, if $\tau \ge s$, then every vertex in U_{s-1} is incident to $\left(1 \pm n^{-0.3\beta}\right)n(s)$ available edges, and there are $\left(1 \pm n^{-0.3\beta}\right)|U_s|n(s)/2$ available edges in total. Furthermore, by assumption, H satisfies Claim 2.12 5. Therefore $e\left(H[A]\right) \le |A|n^{0.9\beta}$. Hence A is incident to $\left(1 \pm n^{-0.3\beta}\right)|A|n(s) \pm |A|n^{0.9\beta} = \left(1 \pm n^{-0.05\beta}\right)|A|n(s)$ available edges. Thus:

$$\mathbb{P}\left[B_s|B_{s-1} \wedge \tau \ge s\right] = \left(1 - \frac{\left(1 \pm n^{-0.05\beta}\right)|A|n(s)}{\left(1 \pm n^{-0.3\beta}\right)|U_s|n(s)/2}\right) = \left(1 - \left(1 \pm n^{-0.04\beta}\right)\frac{2|A|}{|U_s|}\right).$$

Since $\mathbb{P}\left[\tau \le t'\right] = \exp\left(-\Omega(n^{\beta/100})\right)$, we conclude that:

$$\mathbb{P}\left[B_s|B_{s-1}\right] = \left(1 - \left(1 \pm n^{-0.04\beta}\right)\frac{2|A|}{|U_s|}\right)\left(1 \pm \exp\left(-\Omega\left(n^{\beta/100}\right)\right)\right).$$

Thus:

$$\mathbb{P}\left[A \subseteq U_{t'-t}\right] = \prod_{s=1}^{t'-t}\mathbb{P}\left[B_s|B_{s-1}\right] = (1 \pm o\left(1\right))n^{-\alpha|A|}.$$

□

We can now use Markov's inequality to show that $G_{t'}$ satisfies Definition 2.10 2. Recall that for $\ell \le g-2$, T_ℓ is the number of ℓ-threatened pairs in G'. By Claim 2.12 2, $T_\ell \le \frac{(k-1)^\ell}{n}|W_t|^2\log^Q(n)$. By Claim 4.4, the expected number of these pairs that are also contained in $U_{t'-t}$ is at most $(1 \pm o\left(1\right))n^{-2\alpha}T_\ell \le \frac{(k-1)^\ell}{n}|W_t|^2 n^{-2\alpha}\log^Q(n) = \frac{(k-1)^\ell}{n}|W_{t'}|^2\log^Q(n)$. Applying Markov's inequality and a union bound, we conclude that w.h.p., for every $\ell \le g-2$, $P_\ell(G_{t'}) \le \frac{(k-1)^\ell}{n}|W_{t'}|^2\log^{Q+2}(n)$.

Finally, we show that $G_{t'}$ satisfies Definition 2.10 1 . Let $v \in W_t$ and let $\ell \le g-2$. Let $A = A_\ell(v)$ be the set of vertices $u \in W_t$ such that u, v is ℓ-threatened in G'. Then $|A| = T_\ell(v)$, and by assumption $T_\ell(v) \le L(\ell, t)\log^Q(n)$. We will bound the probability that $|A \cap W_{t'}| \ge L(\ell, t')\log^{Q+1}(n)$. By Claim 4.4, for every $B \in \binom{A}{L(\ell,t')\log^{Q+1}(n)}$, it holds that:

$$\mathbb{P}\left[B \subseteq W_{t'}\right] = (1 \pm o\left(1\right))n^{-\alpha L(\ell,t')\log^{Q+1}(n)}.$$

Therefore, by a union bound:

$$\mathbb{P}\left[|A \cap W_{t'}| \ge L(\ell, t')\log^{Q+1}(n)\right] \le \binom{|A|}{L(\ell, t')\log^{Q+1}(n)}(1 \pm o\left(1\right))n^{-\alpha L(\ell,t')\log^{Q+1}(n)}.$$

Applying the inequality $\binom{a}{b} \le (ea/b)^b$, it follows that:

$$\mathbb{P}\left[|A \cap W_{t'}| \ge L(\ell, t')\log^{Q+1}(n)\right] \le (1 \pm o\left(1\right))\left(\frac{e|A|}{n^\alpha L(\ell, t')\log^{Q+1}(n)}\right)^{L(\ell,t')\log^{Q+1}(n)}.$$

Observing that $L(\ell,t) \le n^\alpha L(\ell,t')$, and that $|A| \le L(\ell,t)\log^Q(n)$, we have:

$$\begin{aligned}\mathbb{P}\left[|A \cap W_{t'}| \ge L(\ell,t')\log^{Q+1}(n)\right] &\le (1 \pm o(1))\left(\frac{eL(\ell,t)\log^Q(n)}{n^\alpha L(\ell,t')\log^{Q+1}(n)}\right)^{L(\ell,t')\log^{Q+1}(n)}\\ &\le (1 \pm o(1))\left(\frac{e}{\log(n)}\right)^{L(\ell,t')\log^{Q+1}(n)} = n^{-\omega(1)}.\end{aligned}$$

We apply a union bound over the $O(|W_t|\log(n))$ choices of v and ℓ to conclude that w.h.p., for every $v \in W_{t'}$ and every $\ell \le g-2$, it holds that $|A_\ell(v) \cap W_{t'}| \le L(\ell,t')\log^{Q+1}(n)$. Therefore, for $D = Q+2$, w.h.p. $G_{t'}$ is D-path bounded.

5. Counting high-girth graphs: proof of Theorem 1.2

Let k, c, n be as in the statement of Theorem 1.2. Let $g = c\log_{k-1}(n)$. We prove the theorem by considering the number of (labeled) graphs that can be produced by the (G,g,k)-high-girth process, with G a random Hamilton cycle on n vertices. First, there are $n!/(2n)$ choices for the Hamilton cycle G. In Lemma 2.3 we showed that if, for $d \ge 3$, G' is a $(d-1)$-regular graph on n vertices, then for every $0 \le t \le T := (n - n^{c+\varepsilon})/2$, the (G',g,d)-high-girth process has $(1-o(1))(n-2t)^2/2$ available edges. Thus, the total number of choices in this phase is equal to

$$\begin{aligned}N(d) := \prod_{t=0}^{T}\left((1-o(1))\frac{(n-2t)^2}{2}\right) &= \left((1-o(1))\frac{n^2}{2}\right)^{T+1}\prod_{t=0}^{T}\left(1-\frac{2t}{n}\right)^2\\ &= \left((1-o(1))\frac{n^2}{2e^2}\right)^{n/2}\end{aligned}$$

By Proposition 2.1, w.h.p. the (G',g,d)-high-girth-process succeeds in constructing a d-regular graph. Therefore the number of successful runs of the algorithm is at least $(1-o(1))N(d)$. Returning to the (G,g,k)-high-girth-process, we conclude that the number of successful runs for this algorithm is at least

$$(1-o(1))\frac{n!}{2n}N(3)\times N(4)\times\ldots\times N(k) = \left((1\pm o(1))\frac{n}{e}\right)^n\left((1-o(1))\frac{n^2}{2e^2}\right)^{(k-2)n/2}.$$

Let H be one of the k-regular graphs that the (G,g,k)-high-girth-process can produce. Then H is the disjoint union of the Hamilton cycle G and the $(k-2)$-regular graph H' of the chords chosen by the process. There are fewer than $e(H')!$ ways in which the process can construct H', according to the order in which the edges of H' are added. Furthermore, since H is k-regular, it contains fewer than k^n Hamilton cycles. This serves as an upper bound on the number of possible choices for G. Since $e(H') = (k-2)n/2$, the algorithm can produce at least

$$\frac{((1\pm o(1))n/e)^n\left((1-o(1))n^2/\left(2e^2\right)\right)^{(k-2)n/2}}{k^n\left((k-2)n/2\right)!} = (\Omega(n))^{kn/2}$$

different graphs, as claimed.

6. Concluding remarks and open problems

- A natural and interesting variation of our algorithm starts with n isolated vertices rather than a Hamilton cycle. At each step we add a uniformly chosen edge subject to the constraints that all vertex degrees remain $\le k$

and the girth remains $\geq g$. Ruciński and Wormald [32] studied this process without the girth constraint, and showed that w.h.p. the process yields a regular graph. We believe that ideas from the present work can be modified to show that even for $g = c\log_{k-1}(n)$ (with $c < 1$) the process is likely to produce a k-regular graph. However, new complications arise, which presumably require Wormald's differential equation method. We leave this to future work.

- To what extent do our graphs resemble random regular graphs? Numerical experiments that we have conducted suggest that they are Ramanujan, or at least nearly Ramanujan. For reference recall Friedman's famous result [15] that almost all regular graphs are nearly Ramanujan.
- The large-scale geometry of graphs holds many open questions. Thus, it is not hard to show that every n-vertex k-regular graph of girth g has at most $\frac{nk}{g}(k-1)^{g/2}$ cycles of length g. On the other hand in LPS graphs the number is $\tilde{\Omega}(n^{4/3})$ [8], and no graphs are known for which this number is larger. Our numerical calculations suggest that in our graphs this number is in fact $\Theta_k\left((k-1)^g/g\right)$. It would also be interesting to determine the smallest $\gamma = \gamma(n, k, g)$ such that every girth-g k-regular graph on n vertices has a set of γ edges that intersects every g-cycle.

  The possible relation between a graph's girth and its diameter is particularly intriguing. It follows from [12] and Moore's bound that

$$2 \geq \limsup \frac{\text{girth}(G)}{\text{diam}(G)} \geq 1,$$

  where the lim sup ranges over all graphs where all vertex degrees are ≥ 3. Nothing better seems to be known at the moment.

  Even more remarkably, we do not know whether

$$\sup(\text{girth}(G) - \text{diam}(G))$$

  is finite or not. The sup is over all G in which all vertex degrees are ≥ 3.

References

[1] Mohsen Bayati, Andrea Montanari, and Amin Saberi, *Generating random graphs with large girth*, Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2009, pp. 566–575.

[2] Patrick Bennett and Tom Bohman, *A natural barrier in random greedy hypergraph matching*, Combinatorics, Probability and Computing (2012), 1–10.

[3] NL Biggs and MJ Hoare, *The sextet construction for cubic graphs*, Combinatorica **3** (1983), no. 2, 153–165.

[4] Norman Biggs, *Constructions for cubic graphs with large girth*, the Electronic Journal of Combinatorics **5** (1998), no. 1, 1.

[5] Tom Bohman, *The triangle-free process*, Advances in Mathematics **221** (2009), no. 5, 1653–1677.

[6] Tom Bohman and Lutz Warnke, *Large girth approximate Steiner triple systems*, Journal of the London Mathematical Society (2018).

[7] Béla Bollobás and Oliver Riordan, *Constrained graph processes*, the electronic journal of combinatorics **7** (2000), no. 1, 18.

[8] Maxime Fortier Bourque and Bram Petri, *Kissing numbers of regular graphs*, arXiv preprint arXiv:1909.12817 (2019).

[9] L Sunil Chandran, *A high girth graph construction*, SIAM journal on Discrete Mathematics **16** (2003), no. 3, 366–370.

[10] Patrick Chiu, *Cubic Ramanujan graphs*, Combinatorica **12** (1992), no. 3, 275–285.

[11] Xavier Dahan, *Regular graphs of large girth and arbitrary degree*, Combinatorica **34** (2014), no. 4, 407–426.

[12] Paul Erdős and Horst Sachs, *Reguläre graphen gegebener taillenweite mit minimaler knotenzahl*, Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe **12** (1963), no. 251-257, 22.

[13] Paul Erdös, Stephen Suen, and Peter Winkler, *On the size of a random maximal graph*, Random Structures & Algorithms **6** (1995), no. 2-3, 309–318.

[14] David A Freedman, *On tail probabilities for martingales*, the Annals of Probability **3** (1975), no. 1, 100–118.

[15] Joel Friedman, *A proof of Alon's second eigenvalue conjecture and related problems*, American Mathematical Soc., 2008.

[16] Alexander Gamburd, Shlomo Hoory, Mehrdad Shahshahani, Aner Shalev, and Balint Virág, *On the girth of random Cayley graphs*, Random Structures & Algorithms **35** (2009), no. 1, 100–117.

[17] Stefan Glock, Daniela Kühn, Allan Lo, and Deryk Osthus, *The existence of designs via iterative absorption*, arXiv preprint arXiv:1611.06827 (2016).

[18] ______, *On a conjecture of Erdős on locally sparse Steiner triple systems*, arXiv preprint arXiv:1802.04227 (2018).

[19] David A Grable, *On random greedy triangle packing*, the Electronic Journal of Combinatorics **4** (1997), no. 1, 11.

[20] Shlomo Hoory, *On graphs of high girth*, Ph.D. thesis, Hebrew University of Jerusalem Israel, 2002.

[21] Peter Keevash, *The existence of designs*, arXiv preprint arXiv:1401.3665 (2014).

[22] Jeong Han Kim and Van H Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434.

[23] Michael Krivelevich, Matthew Kwan, Po-Shen Loh, and Benny Sudakov, *The random k-matching-free process*, Random Structures & Algorithms **53** (2018), no. 4, 692–716.

[24] Matthew Kwan, *Almost all Steiner triple systems have perfect matchings*, arXiv preprint arXiv:1611.02246 (2016).

[25] Felix Lazebnik, Vasiliy A Ustimenko, and Andrew J Woldar, *A new series of dense graphs of high girth*, Bulletin of the American mathematical society **32** (1995), no. 1, 73–79.

[26] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak, *Ramanujan graphs*, Combinatorica **8** (1988), no. 3, 261–277.

[27] Brendan D McKay, Nicholas C Wormald, and Beata Wysocka, *Short cycles in random regular graphs*, the electronic journal of combinatorics **11** (2004), no. 1, 66.

[28] Moshe Morgenstern, *Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q*, Journal of Combinatorial Theory, Series B **62** (1994), no. 1, 44–62.

[29] Deryk Osthus and Anusch Taraz, *Random maximal H-free graphs*, Random Structures & Algorithms **18** (2001), no. 1, 61–82.

[30] Michael E Picollelli, *The final size of the C_4-free process*, Combinatorics, Probability and Computing **20** (2011), no. 6, 939–955.

[31] ______, *The final size of the C_ℓ-free process*, SIAM Journal on Discrete Mathematics **28** (2014), no. 3, 1276–1305.

[32] Andrzej Ruciński and Nicholas C Wormald, *Random graph processes with degree restrictions*, Combinatorics, Probability and Computing **1** (1992), no. 2, 169–180.

[33] Lutz Warnke, *The C_ℓ-free process*, Random Structures & Algorithms **44** (2014), no. 4, 490–526.

[34] ______, *On the method of typical bounded differences*, Combinatorics, Probability and Computing **25** (2016), no. 2, 269–299.

[35] Alfred Weiss, *Girths of bipartite sextet graphs*, Combinatorica **4** (1984), no. 2-3, 241–245.

[36] Nicholas C Wormald, *The differential equation method for random graph processes and greedy algorithms*, Lectures on approximation and randomized algorithms **73** (1999), 73–155.

Department of Computer Science, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
Email address: nati@cs.huji.ac.il

Institute of Mathematics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
Email address: menahem.simkin@mail.huji.ac.il

Chapter 6

Discussion and open problems

DISCUSSION AND OPEN PROBLEMS

In this chapter we suggest several directions for future research, organized by topic.

High-dimensional permutations. As mentioned in Chapter 2, a major challenge in the study of high-dimensional permutations is to develop efficient algorithms to sample them uniformly at random. Even in the simplest case of Latin squares, the best algorithm currently known is the Markov chain of Jacobson and Matthews [4], which is not known to mix rapidly. In higher dimensions, the situation is even worse: at present, there is not even a candidate algorithm for generating (say) three-dimensional permutations uniformly at random. In fact, there are currently no practical algorithms at all for sampling three-dimensional permutations with a reasonable amount of randomness. Although Keevash's constructions [6] are, essentially, polynomial-time randomized algorithms to construct such objects, they do not seem to be practical, and the objects they construct are, most likely, quite atypical.

Sampling high-dimensional permutations is related to understanding their typical structure. As laid out in the introduction, the techniques currently available to study random high-dimensional permutations are quite limited. Thus the major question remains: what does a typical high-dimensional permutation look like?

Combinatorial designs in random hypergraphs. The motivating problem of Chapter 3 is to determine the threshold at which a Latin square (viewed as a two-dimensional permutation) appears in a random array. Our main conjecture in this regard is that this threshold is the same as the threshold at which no axis parallel line is all-zero.

This is a a special case of the general *threshold problem for combinatorial designs*. Recall that an (n, q, r)-Steiner system is a q-uniform hypergraph on n vertices in which every r-set is covered exactly once. Keevash [5] proved that for fixed q, r these exist for all but finitely many n satisfying necessary arithmetic conditions. We denote by $\mathcal{H}_q(n; p)$ the binomial distribution on q-uniform hypergraphs in which every q-set is present with probability p. It is tempting to make the following conjecture.

Conjecture. *Fix integers $q > r$. For n satisfying the necessary arithmetic conditions, the threshold for the appearance of an (n, q, r)-Steiner system in $\mathcal{H}_q(n; p)$ is the same as the one at which every r-set is contained in some chosen q-set, i.e., $p = \Theta\left(\log(n)/n^{q-r}\right)$.*

Robustness of graph properties. In Chapter 4 we studied the robustness of perfect matchings in regular bipartite graphs. It is interesting to study this notion with respect to other graph properties and other graph families. For example, a classical result of Corrádi and Hajnal [2] states that every graph G with n vertices and $\delta(G) \geq 2n/3$ has a triangle factor, provided $3|n$. Let G be such a graph. As in Chapter 4, denote by $G(p)$ the distribution on subgraphs of G where each edge is retained with probability p. What is the threshold for $G(p)$ to contain a triangle factor with high probability?

More broadly (and as mentioned in the introduction), we wonder if there exist methods to unify the proofs for the robustness of different properties. The recent breakthrough of Frankston, Kahn, Narayanan, and Park [3] relates thresholds to fractional expectation thresholds. This provides a single framework which at once determines the thresholds for perfect matchings and Hamilton cycles in $\mathcal{G}(n;p)$, as well as many other graph properties. Can this framework be applied to study robustness of graph properties?

Probabilistic constructions of high-girth regular graphs. In Chapter 5 we described a random greedy algorithm to construct k-regular graphs with n vertices and girth $\geq (1 - o_k(1)) \log_{k-1}(n)$. We originally studied this problem in the hope of finding a probabilistic construction of such graphs with girth $\geq (1+\varepsilon)\log_{k-1}(n)$, for some $\varepsilon > 0$. Chapter 5 closes with some questions that we hope will guide the way to such a construction. For example, consider the number of g-cycles in a graph with girth g (sometimes called the graph's **kissing number**, e.g., [1]). How large can this number be? How large is it typically? Also, what is the relationship between the girth and diameter of a graph? Concretely, do there exist cubic graphs G for which $\operatorname{diam}(G) - \operatorname{girth}(G)$ is arbitrarily large?

References

[1] Maxime Fortier Bourque and Bram Petri, *Kissing numbers of regular graphs*, arXiv preprint arXiv:1909.12817 (2019).

[2] Keresztély Corradi and András Hajnal, *On the maximal number of independent circuits in a graph*, Acta Mathematica Hungarica **14** (1963), no. 3-4, 423–439.

[3] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park, *Thresholds versus fractional expectation-thresholds*, arXiv preprint arXiv:1910.13433 (2019).

[4] Mark T Jacobson and Peter Matthews, *Generating uniformly distributed random Latin squares*, Journal of Combinatorial Designs **4** (1996), no. 6, 405–437.

[5] Peter Keevash, *The existence of designs*, arXiv preprint arXiv:1401.3665 (2014).

[6] ______, *The existence of designs II*, arXiv preprint arXiv:1802.05900 (2018).

התהליך שלגרף הנוצר יש מותן g לפחות. הניתוח של האלגוריתם נותן גם חסם תחתון על מספר הגרפים הרגולריים עם מותן גבוה.

תקציר

בעבודה זו אנו חוקרים שימוש באלגוריתמים הסתברותיים במטרה לבנות מבנים קומבינטוריים רגולריים, ובפרט בממדים גבוהים.

אנחנו שמים דגש מיוחד על **תמורות רב-ממדיות**. ברוח עבודתם של ליניאל ולוריא, אנו מגדירים תמורה מממד d ומסדר n כמערך $n \times n \times \ldots \times n = [n]^{d+1}$ המקבל ערכים ב-$\{0,1\}$, כך שבכל שורה המקבילה לצירים ישנו 1 יחיד. כך, תמורה מממד אחד זהה למטריצת תמורה, והמושג של תמורה מממד שניים זהה לריבוע לטיני.

בפרק 2 אנו מכלילים את משפט ארדש-סקרש לתמורות מממדים גבוהים. אנחנו מראים שבכל תמורה בממד d מסדר n ישנה תת-סדרה מונוטונית מאורך $\Omega_d\left(\sqrt{n}\right)$, והחסם הזה הדוק. אנו גם מראים כי בתמורה טיפוסית בממד d מסדר n מתקיים בהסתברות גבוהה שהאורך המרבי של תת-סדרה מונוטונית הוא $\Theta_d\left(n^{d/(d+1)}\right)$.

פרק 3 עוסק בבעיית הסף של ריבועים לטיניים. יהיו $m \le n \le k$ מספרים טבעיים. אומרים שמערך $m \times n \times k$ המקבל ערכים ב-$\{0,1\}$ הוא **תיבה לטינית** אם מספר האחדים בו הוא בדיוק mn, ובכל שורה המקבילה לצירים יש לכל היותר 1 יחיד. כך, תיבה לטינית עם הפרמטרים $\mathrm{m} = \mathrm{n} = k$ היא ריבוע לטיני. כאשר m ו-k קרובים ל-n, ניתן לראות זאת כריבוע לטיני מקורב. תהי $M(m,n,k;p)$ ההתפלגות על מערכים $m \times n \times k$ שבה כל ערך מקבל 1 בהסתברות p (בלתי תלויה בשאר הערכים), ו-0 אחרת. בעיית הסף של הריבועים הלטיניים היא לקבוע עבור איזה p מתקיים בהסתברות גבוהה ש-$M(n,n,n;p)$ מכיל ריבוע לטיני. באופן כללי יותר אנו שואלים מתי $M(m,n,k;p)$ מכיל בהסתברות גבוהה תיבה לטינית. לכל $\varepsilon > 0$ אנו נותנים תשובה הדוקה אסימפטוטית בתחומי הערכים $m = n$ ו-$k \ge (1+\varepsilon)n$ או $m \le (1-\varepsilon)n$ ו-$k = n$. בשני המקרים, הסף הוא $\Theta(\log(n)/n)$.

בפרק 4 אנו בוחנים מחדש את התוצאה היסודית של ארדש ורניי האומרת שהסף להופעת זיווג מושלם בגרף דו-צדדי מקרי והסף להיעלמותם של קודקודים מבודדים – חד הם. יהיה G גרף k-רגולרי דו-צדדי עם $2n$ קודקודים. מתבוננים בתהליך שבו אנו משחזרים את G צלע אחר צלע בסדר מקרי. אנו מראים כי אם $k = \omega\left(n/(\log n)^{1/3}\right)$ אזי בהסתברות גבוהה, זיווג מושלם מופיע בתהליך זה בדיוק באותו הרגע שבו נעלם הקדקוד המבודד האחרון. כאשר $G = K_{n,n}$ זוהי גרסת זמן עצירה ידועה, אשר הוכחה על ידי בולובאש ותומאסון, למשפט הנ"ל של ארדש ורניי. לעומת זאת, אנו מראים כי ישנם גרפים שבהם $\mathrm{k} = \Omega\left(n/(\log(n)\log\log(n))\right)$ ובהם הקודקוד המבודד האחרון נעלם זמן רב לפני שמופיע לראשונה זיווג מושלם.

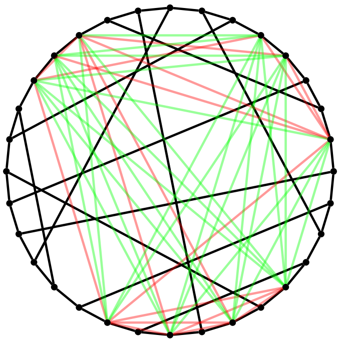
פרק 5 מתאר אלגוריתם חמדני מקרי לבנייה של גרפים רגולריים בעלי מותן גבוה. יהיו $k \ge 3$ טבעי ו-$c < 1$ קבועים. עבור מספר זוגי n נגדיר $g = c\log_{k-1} n$. מתחילים עם מעגל המילטוני G על n קודקודים. כל עוד $\delta(G) < k$, בוחרים בהתפלגות אחידה זוג קודקודים u,v מדרגה $\delta(G)$ בכפוף לאילוץ שמרחקם הוא לפחות $g - 1$. אם אין זוג כזה, עוצרים. אחרת מוסיפים את הצלע uv לגרף G. אנו מראים כי בהסתברות גבוהה התהליך מסתיים בגרף k-רגולרי. ברור מהגדרת

עבודה זו נעשתה בהדרכתו של

נתי ליניאל

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM
الجامعة العبرية في اورشليم القدس

שימוש באקראיות למציאת מבנה



חיבור לשם קבלת תואר
דוקטור לפילוסופיה

מאת
מיכאל סימקין

הוגש לסנט האוניברסיטה העברית בירושלים
אדר תש"ף