# A TREATISE ON THE BINOMIAL THEOREM

**BY PATRICK DEVLIN**

**A dissertation submitted to the**

**Graduate School—New Brunswick**

**Rutgers, The State University of New Jersey**

**in partial fulfillment of the requirements**

**for the degree of**

**Doctor of Philosophy**

**Graduate Program in Mathematics**

**Written under the direction of**

**Jeff Kahn**

**and approved by**

**New Brunswick, New Jersey**

**May, 2017**

## ABSTRACT OF THE DISSERTATION

# A treatise on the binomial theorem


**by PATRICK DEVLIN**

**Dissertation Director: Jeff Kahn**



This dissertation discusses four problems taken from various areas of combinatorics—stability results, extremal set systems, information theory, and hypergraph matchings. Though diverse in content, the unifying theme throughout is that each proof relies on the machinery of probabilistic combinatorics. The first chapter offers a summary.

In the second chapter, we prove a stability version of a general result that bounds the permanent of a matrix in terms of its operator norm. More specifically, suppose $A$ is an $n \times n$ matrix over $\mathbb{C}$ (resp. $\mathbb{R}$), and let $\mathcal{P}$ denote the set of $n \times n$ matrices over $\mathbb{C}$ (resp. $\mathbb{R}$) that can be written as a permutation matrix times a unitary diagonal matrix. Then it is known that the permanent of $A$ satisfies $|\mathrm{perm}(A)| \leq \|A\|_2^n$ with equality iff $A/\|A\|_2 \in \mathcal{P}$ (where $\|A\|_2$ is the operator 2-norm of $A$). We show a stability version of this result asserting that unless $A$ is very close (in a particular sense) to one of these extremal matrices, its permanent is exponentially smaller (as a function of $n$) than $\|A\|_2^n$. In particular, for any fixed $\alpha, \beta > 0$, we show that $|\mathrm{perm}(A)|$ is exponentially smaller than $\|A\|_2^n$ unless all but at most $\alpha n$ rows contain entries of modulus at least $\|A\|_2(1 - \beta)$.

In the third chapter, we prove a randomized result extending the classical Erdős–Ko–Rado theorem. Namely, let $K_p(n, k)$ denote the random subgraph of the usual

Kneser graph $K(n,k)$ in which edges appear independently, each with probability $p$. Answering a question of Bollobás, Narayanan, and Raigorodskii, we show that there is a fixed $p < 1$ such that *almost surely* (i.e., with probability tending to 1) the maximum independent sets of $K_p(2k+1,k)$ are precisely the sets $\{A \in V(K(2k+1,k)) : x \in A\}$ $(x \in [2k+1])$. We also complete the determination of the order of magnitude of the "threshold" for the above property for general $k$ and $n \geq 2k+2$. This is new for $k \sim n/2$, while for smaller $k$ it is a recent result of Das and Tran.

In the fourth chapter, we prove the following conjecture of Leighton and Moitra. If $\sigma$ is a random (not necessarily uniform) permutation of $[n]$ such that for all $i,j$ $|\mathbb{P}(\sigma(i) < \sigma(j)) - 1/2| > \varepsilon$, then the binary entropy of $\sigma$ is at most $(1-\vartheta_\varepsilon)\log_2 n!$ for some (fixed) positive $\vartheta_\varepsilon$. If we further assume $\mathbb{P}(\sigma(i) < \sigma(j)) > 1/2 + \varepsilon$ for all $i < j$, the theorem is due to Leighton and Moitra; for this case we give a short proof with a better $\vartheta_\varepsilon$.

Finally, in the fifth chapter, we extend the notion of (random) $k$-out graphs and consider when a $k$-out hypergraph is likely to have a perfect fractional matching. In particular, we show that for each $r$ there is a $k = k(r)$ such that the $k$-out $r$-uniform hypergraph on $n$ vertices almost surely has a perfect fractional matching and prove an analogous result for $r$-uniform $r$-partite hypergraphs. This is based on a new notion of hypergraph expansion and the observation that sufficiently expansive hypergraphs admit perfect fractional matchings. As a further application, we give a short proof of a stopping-time result originally due to Krivelevich.

## Acknowledgements

I first want to thank my amazing wife, Nora, who knows more graduate-level combinatorics than any other education theorist on the planet. You are sincerely the most brilliant person I have ever met, and I love watching you excel. In my life, each success is as much yours as mine, and without your love and support, even my loftiest ambitions would pale in comparison to who I already am with you today.

For over a decade you have been telling me that I want to be a professor. When I was overwhelmed by depression and anxiety, your prayers, love, and encouragement comforted me and reminded me of the person I want to become. As so many can attest to, the wisdom I gained from this journey was as much transmitted by you as by any other experience, mentor, or friend. My love for you is deeper than ever before, because of how much you have cared for me in ways that keep me true to my hopes and dreams for myself and for our life together. Thank you for believing in me, doing jigsaw puzzles with me, and sharing in all my joys and sorrows. I love you.

I want to thank my family, who get excited about my mathematical adventures in a way that only deep familial love could inspire. I want to thank my friends, who always make me smile, especially Mookie, whose passion for math has grown side-by-side with my own ever since we were boyhood friends. I want to thank the countless mathematicians throughout my entire academic career who have warmly welcomed me into their community despite my persistent ignorance.

Thank you to the members of my dissertation committee—Professors Beck, Kahn, Kopparty, and Vu—for all your help and continued support.

I would like to thank any readers not here acknowledged; however, I regret to inform you that—with overwhelming probability—you do not exist.

And finally…

# His Last Bow

To Jeff:

It is a truth universally acknowledged, that an available professor in possession of a good publication record must be in want of a graduate student. However little known the feelings or views of such a man may be on his first entering a classroom, this truth is so well fixed in the minds of the surrounding students, that he is considered as the rightful property of some one or other of them.[1]

Looking for an advisor, I thought of you and remarked: *His career has been an extraordinary one. He is a man of good birth and excellent education, endowed by nature with a phenomenal mathematical faculty. In all seriousness, if I could beat that man, I should feel that my own career had reached its summit.* Though your brilliance is humbling, there is a stubbornness about me that never can bear to be frightened at the will of others. My courage always rises with every attempt to intimidate me. And—as I have not been in the habit of brooking disappointment—I asked to be your student.

You were a fantastic advisor, and I have been most anxious to acknowledge to you how gratefully I feel it. Were it known to the rest of my family, I should not have merely my own gratitude to express.[2]

Your discussions, support, and feedback have been invaluable to me in my development as a mathematician, and I hope some day to mentor my own students as you have for me.

[1] That the wish of giving happiness to you might add force to the other inducements which lead me on, I shall not attempt to deny.

[2] Thanks also for introducing me to Austen, Conan Doyle, Groucho, and so many others.

# Dedication

李文博教授

To Wenbo Li,

who imparted so much

(including a love of mathematics)

I know you would be proud of me right now,

and you are missed.


九连环

Traditional nine rings puzzle

Image reproduced with permission from chinesepuzzles.org

# Table of Contents

# Chapter 1

# A study in scarlet regalia

*So much is known to the world, but what I am telling you now is what I have myself discovered.* — Sir Arthur Conan Doyle, *The Final Problem*

This dissertation primarily consists of four papers, each of which is given its own chapter. In this introductory chapter, we offer quick, high-level accounts of each of the four before they are discussed in full detail. The chapter concludes with a brief collection of common notation used throughout.

## A case of identity (and other permutation matrices)

[*A paper joint with Ross Berkowitz*] Leonid Gurvits [30] proved that the permanent of an $n \times n$ matrix $A$ over $\mathbb{C}$ is bounded by its operator norm via $|\mathrm{perm}(A)| \leq \|A\|^n$. Motivated by questions related to boson sampling and quantum computing, Scott Aaronson and Travis Hance [1] asked for a characterization of matrices for which this bound is nearly tight. Appealing to inverse Littlewood-Offord theory (developed by Terry Tao and Van Vu), several papers [2, 46] attempted to address this question with only limited successs. We settle this in Chapter 2 by proving (a quantified version of) the following, which was a conjecture of Aaronson.

**Theorem 1.1.** *If $|perm(A)| \geq \|A\|^n/n^{100}$, then $A$ must have a readily identifiable form where virtually every row and column is dominated by a single entry of very large modulus.*

Although the result deterministically holds for all matrices over $\mathbb{C}$, our proof is entirely probabilistic, using Talagrand's inequality, hypercontractivity, and Khintchine's inequality. It would be interesting to extend our results to obtain deterministic algorithms approximating $|\mathrm{perm}(A)|$ to within an additive error of $\pm\varepsilon\|A\|^n$, which was a driving motivation of [1].

## The adventure of the missing three-quarter

[*A paper joint with Jeff Kahn*] For $n \geq 2k+1$, the *Kneser graph*, $K(n,k)$, has as vertices the $k$-element subsets of $[n] := \{1, 2, \ldots, n\}$ with vertices $A$ and $B$ adjacent iff $A \cap B = \emptyset$. Recall that the *independence number*, $\alpha(G)$, of a graph $G$ is the maximum size of a set of vertices containing no edges. In the language of Kneser graphs, the classical Erdős–Ko–Rado theorem [19] says $\alpha(K(n,k)) = \binom{n-1}{k-1}$ (and that a largest independent set consists of all $k$-sets containing some fixed element of $[n]$).

Following a trend of considerable recent interest, Béla Bollobás and various co-authors [11] considered this classical result in a probabilistic setting and asked when the same behavior is likely to hold in the random subgraph $K_p(n,k) \subseteq K(n,k)$ gotten by retaining edges independently with probability $p$.

Combined with earlier work, our results completely determine the order of magnitude of the threshold for this property. A full discussion is given in in Chapter 3. Here, I just highlight one case addressing what both [11] and [8] identified as the most interesting aspect of the problem:

**Theorem 1.2.** *There is a fixed $\varepsilon > 0$ such that for $n = 2k+1$ and $p > 1 - \varepsilon$,*

$$\lim_{n \to \infty} \mathbb{P}[\alpha(K_p(n,k)) = \alpha(K(n,k))] = 1.$$

The key steps of the proof rely on spectral techniques and results from *Fourier analysis on the slice* to show that certain collections of vertices contain many edges in the Kneser graph. A natural conjecture is that the above result should hold for all $p > 3/4$, however we were ultimately unable to prove this.

## The adventure of the three students

[*A paper joint with Hüseyin Acan and Jeff Kahn*] In Chapter 4, we prove the following, which was conjectured by Tom Leighton and Ankur Moitra [44] in connection with the algorithmic problem of sorting under partial information.

**Theorem 1.3.** *If $\sigma$ is a random (not necessarily uniform) permutation of $\{1, 2, \ldots, n\}$ satisfying for fixed $\varepsilon > 0$*

$$|\mathbb{P}(\sigma(i) < \sigma(j)) - 1/2| > \varepsilon \qquad \forall i \neq j, \tag{1.1}$$

*then $\sigma$ has entropy at most $(1 - \delta)\log(n!)$, where $\delta > 0$ depends only on $\varepsilon$.*

That is, the assumption (1.1) implies a significant loss of information relative to the entropy of a uniform distribution (namely $\log(n!)$). Leighton and Moitra proved this in the special case where $\mathbb{P}(\sigma(i) < \sigma(j)) > 1/2 + \varepsilon$ for all $i < j$. Our proof uses a mix of probabilistic and graph theoretic techniques including a version of Szemerédi's regularity lemma, a coupling argument, and martingale concentration results. We also provide a short proof improving on the result in [44].

## The final problem

[*A paper joint with Jeff Kahn*] Hypergraphs are extremely useful generalizations of graphs but are notoriously difficult to work with. An *$r$-uniform hypergraph $\mathcal{H}$* on vertex set $V$ is a collection of $r$-subsets of $V$—thus 2-uniform hypergraphs coincide with graphs. A *perfect matching* of a hypergraph is a subset of the members of $\mathcal{H}$ ("edges") that partitions the vertex set. The linear programming relaxation of this is a *perfect fractional matching*, that is a nonnegative weighting of the edges of $\mathcal{H}$ for which the weights of the edges containing any $v \in V$ sum to 1.

For any $r > 2$, determining if an $r$-uniform hypergraph has a perfect matching is an NP-complete problem [38], making the question both important and computationally intractible (unless P = NP). Motivated by this and by a conjecture of Alan Frieze and Gregory Sorkin [26], we prove the following, which extends earlier results for graphs to hypergraphs. (The natural $k$-out model is defined in Chapter 5.)

**Theorem 1.4.** *For each $r$, there is a $k$ such that with high probability the $r$-uniform $k$-out hypergraph admits a perfect fractional matching and $w \equiv 1/r$ is the only fractional cover of weight $n/r$.*

We prove an analogous result for $r$-uniform $r$-partite hypergraphs as well. A key step in our proof is establishing that certain expansion-type properties deterministically imply the existence of perfect fractional matchings in $r$-uniform hypergraphs. As a further application, we also provide a short proof of a stopping-time result originally due to Krivelevich.

## The sign of the four

These four papers have the following common notation

- All asymptotics are taken as $n \to \infty$ (usually a parameter related to the number of vertices of a graph or hypergraph)
- $f(n) = \Omega(g(n))$ iff $g(n) = \mathcal{O}(f(n))$ iff $\limsup_n g(n)/f(n) < \infty$
- $f(n) = \omega(g(n))$ iff $g(n) = o(f(n))$ iff $\limsup_n g(n)/f(n) = 0$
- $[n] = \{1, 2, 3, \ldots, n\}$
- $\binom{X}{t}$ is the collection of all $t$-element subsets of $X$
- We say a statement holds *with high probability* (w.h.p.) or *almost surely* (a.s.) iff it holds with probability tending to 1
- $\mathfrak{S}_n$ is the set of permutations of $[n]$
- Following a common abuse, we pretend all large numbers are integers whenever convenient
- Throughout, we use $\log(\cdot)$ to mean $\ln(\cdot)$, and we use $\lg(\cdot)$ to mean $\log_2(\cdot)$

# Chapter 2

# A stability result using the matrix norm to bound the permanent

Ross Berkowitz[†] Pat Devlin[‡]

*Abstract:* We prove a stability version of a general result that bounds the permanent of a matrix in terms of its operator norm. More specifically, suppose $A$ is an $n \times n$ matrix over $\mathbb{C}$ (resp. $\mathbb{R}$), and let $\mathcal{P}$ denote the set of $n \times n$ matrices over $\mathbb{C}$ (resp. $\mathbb{R}$) that can be written as a permutation matrix times a unitary diagonal matrix. Then it is known that the permanent of $A$ satisfies $|\mathrm{perm}(A)| \le \|A\|_2^n$ with equality iff $A/\|A\|_2 \in \mathcal{P}$ (where $\|A\|_2$ is the operator 2-norm of $A$). We show a stability version of this result asserting that unless $A$ is very close (in a particular sense) to one of these extremal matrices, its permanent is exponentially smaller (as a function of $n$) than $\|A\|_2^n$. In particular, for any fixed $\alpha, \beta > 0$, we show that $|\mathrm{perm}(A)|$ is exponentially smaller than $\|A\|_2^n$ unless all but at most $\alpha n$ rows contain entries of modulus at least $\|A\|_2(1-\beta)$.

## Introduction

The *permanent* of an $n \times n$ matrix, $A$, has long been an important quantity in combinatorics and computer science, and more recently it has also had applications to physics and linear-optical quantum computing. It is defined as

$$\mathrm{perm}(A) := \sum_{\sigma \in \mathfrak{S}_n} \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

---

AMS 2010 subject classification: 05D40, 15A15, 68R05, 05B20

Key words and phrases: permanent, matrix norm, stability, probabilistic method

[†]Rutgers University. Email: `rkb73@math.rutgers.edu`

[‡]Rutgers University. Email: `prd41@math.rutgers.edu`. Work done while supported by NSF grant DMS1501962

For instance, if $A$ only has entries in $\{0,1\} \subseteq \mathbb{R}$, then the permanent counts the number of perfect matchings in the bipartite graph whose bipartite adjacency matrix is $A$.

The definition of the permanent is of course reminiscent of that for the determinant; however, whereas the determinant is rich in algebraic and geometric meaning, the more combinatorial permanent is notoriously difficult to understand. For example, computing $\mathrm{perm}(A)$ even for $\{0,1\}$-matrices is the prototypical #P-complete problem (Valiant [53]).

On the other hand, the *operator 2-norm* (also called the *operator norm*) of a matrix is a particularly nice parameter. For an $n \times n$ matrix $A$ with entries in $\mathbb{C}$, it is defined as

$$\|A\|_2 = \sup_{\|\vec{x}\|_2 \leq 1,\ \vec{x} \in \mathbb{C}^n} \|A\vec{x}\|_2,$$

where $\|\vec{v}\|_p$ is the usual $l_p$ norm (i.e., $\|\vec{v}\|_p^p = \sum_i |v_i|^p$ for $p \in (0,\infty)$, and $\|\vec{v}\|_\infty = \max |v_i|$). The operator norm of a matrix has the advantages of being both algebraically and analytically well-behaved as well as computationally easy to determine (as this amounts to finding the largest singular value of $A$).

Considering how differently behaved the permanent and operator norm are, it is perhaps strange to think that there would be much of a connection between them. Nonetheless, they are related by the following extremal result, which is due to Gurvits [30] (see also [1, 2]).

**Theorem 2.1.** *Suppose $A$ is an $n \times n$ matrix over $\mathbb{C}$ (resp. $\mathbb{R}$), and let $\mathcal{P}$ denote the set of $n \times n$ matrices over $\mathbb{C}$ (resp. $\mathbb{R}$) that can be written as a permutation matrix times a unitary diagonal matrix. Then $|perm(A)| \leq \|A\|_2^n$ with equality iff $A$ is a scalar multiple of a matrix in $\mathcal{P}$.*

Note that this extremal set $\mathcal{P}$ is simply the set of matrices with exactly $n$ non-zero entries, each having modulus 1, and no two of which are in the same row or column. Such a matrix $P \in \mathcal{P}$ has $\|P\|_2 = |\mathrm{perm}(P)| = 1$ and satisfies

$$\|AP\|_2 = \|PA\|_2 = \|A\|_2, \qquad \text{and} \qquad |\mathrm{perm}(AP)| = |\mathrm{perm}(PA)| = |\mathrm{perm}(A)|$$

for all matrices $A$ (which is equivalent to membership in $\mathcal{P}$). Moreover, $\mathcal{P}$ is a subgroup of the group of unitary matrices, and as a set, it has a very tractable topological structure.

Motivated by algorithmic questions related to approximating the permanent, Aaronson and Hance [1] asked whether one could prove a stability version of Theorem 2.1:

**Question 2.A:** If $|\mathrm{perm}(A)|$ is close to $\|A\|_2^n$, must $A/\|A\|_2$ be 'close' to a matrix in $\mathcal{P}$?

A somewhat more concrete version was suggested by Aaronson and Nguyen [2]:

**Question 2.B:** Characterize $n \times n$ matrices $A$ such that $\|A\|_2 \leq 1$ and there exists a constant $C > 0$ such that $|\mathrm{perm}(A)| \geq n^{-C}$.

Using techniques of inverse Littlewood-Offord theory, Aaronson and Nguyen gave a substantial answer to an analogous question under the (stronger) assumptions that $A$ is orthogonal and that the intersection of the hypercube $\{\pm 1\}^n$ with its image under $A$ is large. They also proved something like (actually slightly stronger than) our results below for stochastic matrices. Further results in the direction of Question 2.B were given by Nguyen [46].

The two main results of this chapter are Theorems 2.2 and 2.4 below. The first provides a positive answer to Question 2.A for matrices over $\mathbb{C}$ (or $\mathbb{R}$), and the second is a more refined result that (depending on your philosophical views) at least partially addresses Question 2.B for matrices over $\mathbb{R}$. More specifically, we bound $\mathrm{perm}(A)$ in terms of the following easily computed parameters.

**Definition:** Let $A$ be a matrix with rows $r_1, r_2, \ldots, r_n$, and $p \in \mathbb{R} \cup \{\infty\}$. Then the parameter $h_p(A)$ is defined as $h_p(A) = h_p = \frac{1}{n}\sum_i \|r_i\|_p$.

We will only consider $h_\infty$ and $h_2$. First note $0 \leq h_\infty(A) \leq h_2(A) \leq \|A\|_2$. Moreover, it is easy to show $h_2(A) = \|A\|_2$ iff $A/\|A\|_2$ is a unitary matrix, and $h_\infty(A) = \|A\|_2$ iff $A/\|A\|_2$ is in $\mathcal{P}$. Thus, in some sense, the quantity $1 - h_2(A)/\|A\|_2 \in [0,1]$ measures how close $A/\|A\|_2$ is to being unitary, and $1 - h_\infty(A)/\|A\|_2 \in [0,1]$ measures how close

$A/\|A\|_2$ is to being in $\mathcal{P}$. Broadly speaking, $h_\infty/\|A\|_2$ is close to 1 precisely when most of the rows of $A$ each have one entry of modulus close to $\|A\|_2$ and all the other entries in that row are close to 0.

Before stating the first of our main results, notice that in addressing either of the above questions, we lose no generality in assuming $\|A\|_2 \leq 1$, since Question 2.A is invariant under scaling. However, to facilitate any application of our results, we state them in the "more general" case that $\|A\|_2 \leq T$.

**Theorem 2.2.** *Let $A$ be an $n \times n$ matrix over $\mathbb{C}$ and $\|A\|_2 \leq T \neq 0$. Then*

*(i)* $|perm(A)| \leq 2T^n \exp\left[ -3n\Big(1 - \frac{\sqrt{\pi}}{2}h_2/T - \Big(1 - \frac{\sqrt{\pi}}{2}\Big)\, h_\infty/T\Big)^2/100\right]$,

*(ii)* $|perm(A)| \leq 2T^n \exp[-n(1-h_\infty/T)^2/10^5]$.

As discussed above, this provides a positive answer to Question 2.A by viewing $h_\infty$ (and to a lesser extent $h_2$) as a proxy for 'closeness' of a matrix $A$ to those in $\mathcal{P}$. As an easy corollary, if $\alpha, \beta \geq 0$ satisfy $|\text{perm}(A)| \geq 2T^n \exp[-n\alpha^2\beta^2/10^5]$, then all but at most $\alpha n$ of the rows of $A$ contain an entry whose modulus is at least $T(1-\beta)$. And since the $l_2$ norm of any row of $A$ is at most $\|A\|_2$, no entry of $A$ can have modulus larger than $T$. Thus, entries of modulus $T(1-\beta)$ are nearly as large as possible. Moreover, if a row (or column) has an entry with very large modulus, then the remaining entries must have very small moduli (again since its $l_2$ norm is at most $\|A\|_2$). Thus, this theorem also provides a *qualitative* stability result stating that matrices with large permanent must have many very large entries, and a row (or column) containing a large entry must have all its other entries small.

Note that Theorem 2.2 is only useful for values of $h_\infty/T$ that are not very close to 1—namely when $1 - h_\infty/T \gg n^{-1/2}$. Although this does well in many cases, we believe that for large values of $h_\infty/T$, it is not optimal. For comparison, if $A$ is $\delta$ times the identity matrix, and $\delta \approx 1$, then $|\text{perm}(A)| \approx e^{-n(1-\delta)} = e^{-n(1-h_\infty)}$, and we conjecture that this is essentially tight.

**Conjecture 2.3.** *There is some constant $C > 0$ and some polynomial $f(n)$ such that the following holds. If $A$ is an $n \times n$ matrix with complex entries and $\|A\|_2 \leq 1$, then $|perm(A)| \leq f(n)e^{-Cn(1-h_\infty)}$.*

As a step in this direction, we are able to prove the following, which better addresses Question 2.B for matrices over $\mathbb{R}$.

**Theorem 2.4.** *Let $A$ be an $n \times n$ matrix over $\mathbb{R}$ and $\|A\|_2 \leq T \neq 0$. Then*

$$|perm(A)| \leq T^n(n+6)\exp\left[\frac{-\sqrt{n(1-h_\infty/T)}}{400}\right].$$

As with Theorem 2.2, a result like Theorem 2.4 that involves $h_2$ is also possible, and it essentially falls out of our proof directly. Theorem 2.4 is an improvement over Theorem 2.2 when $n^{-1/3} \gg 1 - h_\infty/T$ and gives a meaningful bound provided $1 - h_\infty/T \gg \log(n)^2/n$. Although this yields a quantitatively better understanding for matrices over $\mathbb{R}$, we cannot shake the belief that neither of our main results (i.e., Theorems 2.2 and 2.4) is best possible, and we discuss this further in Section 2.5.

### Structure

The chapter is devoted to proving Theorems 2.2 and 2.4, which goes roughly as follows. First, we appeal to a result of Glynn [29] that allows us to convert the problem of estimating the permanent into a problem about estimating the expected value of a certain random variable (Section 2.2). We then use standard probabilistic tools to show certain concentration results for the random variable of interest, which in turn yield the estimates needed for our results. This is done for the complex-valued case in Section 2.3, which proves Theorem 2.2. In Section 2.4, we consider the real-valued case, where we analyze the corresponding random variable more carefully to obtain Theorem 2.4. We conclude in Section 2.5 with several open questions and conjectures, as well as a discussion of Question 2.B.

### Definitions and set-up with random variables

We first need to use an observation due to Glynn [29] whereby the permanent of a matrix is expressed as the expectation of a certain random variable. We will work over the field $\mathbb{K}$, which will either be $\mathbb{R}$ or $\mathbb{C}$.

Given an $n \times n$ matrix $A$ over $\mathbb{K}$ and $x \in \mathbb{K}^n$, set $y = Ax$, and define the *Glynn estimator* of $A$ at $x$ to be

$$Gly_x(A) = \prod_{i=1}^{n} \overline{x}_i \times \prod_{i=1}^{n} y_i,$$

where $\overline{z}$ denotes the complex conjugate of $z$. Let $X \in \mathbb{K}^n$ be the random variable whose coordinates are independently selected uniformly on $|z| = 1$, and let $Y = AX$ (note: if $\mathbb{K} = \mathbb{C}$, then each coordinate of $X$ is distributed continuously over the unit circle, whereas if $\mathbb{K} = \mathbb{R}$, then $X$ is chosen uniformly from the discrete set $\{-1, 1\}^n$). Then

$$\text{perm}(A) = \mathbb{E}[Gly_X(A)] = \mathbb{E}\left[\prod_{i=1}^{n} \overline{X}_i Y_i\right],$$

obtained simply by expanding out the product in the Glynn estimator and using the fact that the $X_i$ are independent with mean 0 and variance 1 (see the original proof due to Glynn [29] or also [30, 1, 2]). Therefore, by convexity (which we are about to use twice), we have

$$|\text{perm}(A)| \leq \mathbb{E}\left[\prod_{i=1}^{n} |\overline{X}_i Y_i|\right] = \mathbb{E}\left[\prod_{i=1}^{n} |Y_i|\right] \leq \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} |Y_i|\right)^n\right] = \mathbb{E}\left[\left(\frac{\|AX\|_1}{n}\right)^n\right].$$

Note that from here, we could say (by Cauchy-Schwartz)

$$\frac{\|AX\|_1}{n} \leq \frac{\|AX\|_2}{\sqrt{n}} = \frac{\|AX\|_2}{\|X\|_2} \leq \|A\|_2,$$

thus obtaining the inequality $|\text{perm}(A)| \leq \|A\|_2^n$ of Theorem 2.1 (the equality case follows by considering equality in the above estimates).

### Specializing to norm at most 1

Note that to prove our results, it suffices to prove them for the case $\|A\|_2 \leq 1$. This is because otherwise, we could simply scale the matrix by some $\alpha$ to have norm at most 1, and because $\text{perm}(A) = \alpha^n \text{perm}(A/\alpha)$, our results would follow. As such, we will henceforth assume $\|A\|_2 \leq 1$ (explicitly making note of when we do), but this choice is simply for notational ease. We remark that the set-up thus far has also been employed in several other papers [30, 1, 2]; however, the remainder of this chapter deviates from the previous literature.

### Proof of Theorem 2.2 ($\mathbb{K} = \mathbb{C}$)

In the setting where $\|A\|_2 \leq 1$, the permanent is always bounded above by 1 (as shown above), and we want to conclude that under certain conditions, it must be (exponentially) small. We know (since $0 \leq \|AX\|_1/n \leq \|A\|_2 \leq 1$) that for all $\varepsilon \geq 0$ and all $\tilde{\mu} \geq 0$,

$$|\mathrm{perm}(A)| \leq \mathbb{E}\left[\left(\frac{\|AX\|_1}{n}\right)^n\right] \leq (\tilde{\mu}/n + \varepsilon)^n + \mathbb{P}(\|AX\|_1 \geq \tilde{\mu} + \varepsilon n).$$

We will pick $\tilde{\mu}$ suitably small with $\tilde{\mu} \geq \mathbb{E}[\|AX\|_1]$ and then argue that $\|AX\|_1$ is tightly concentrated about its mean, which will complete the proof.

### The mean of $\|AX\|_1$

We appeal to a theorem of König, Schütt, and Tomczak-Jaegermann [40], which is a variant of Khintchine's inequality conveniently well-suited for our situation (in fact, $X$ was chosen in part so that we could apply this result directly).

**Theorem 2.5** (König et al. [40], 1999)**.** *Let $\mathbb{K}$ be $\mathbb{R}$ or $\mathbb{C}$. Suppose $\vec{a} = (a_1, \ldots, a_n) \in \mathbb{K}^n$ is fixed, and suppose each coordinate of $\xi \in \mathbb{K}^n$ is independently distributed uniformly on $|z| = 1$. Then*

$$\left|\mathbb{E}\left[\left|\sum_i a_i \xi_i\right|\right] - \Lambda_{\mathbb{K}}\|\vec{a}\|_2\right| \leq (1 - \Lambda_{\mathbb{K}})\,\|\vec{a}\|_\infty,$$

*where $\Lambda_{\mathbb{R}} = \sqrt{2/\pi}$ and $\Lambda_{\mathbb{C}} = \sqrt{\pi}/2$.*

Applying this to each row of $A$ (and using linearity of expectation) gives

**Proposition 2.6.** *With $A$ and $X \in \mathbb{C}^n$ as in Section 2.2, we have*

$$\mathbb{E}[\|AX\|_1/n] \leq \frac{1}{n}\sum_{i=1}^n \left[\sqrt{\pi}/2\|r_i\|_2 + \left(1 - \sqrt{\pi}/2\right)\|r_i\|_\infty\right] = \frac{\sqrt{\pi}}{2}h_2(A) + \left(1 - \frac{\sqrt{\pi}}{2}\right)h_\infty(A).$$

### Concentration about mean

To show concentration of $\|AX\|_1$ about its mean, we use a very general and useful result of Talagrand (a form of "Talagrand's inequality"), which can be found in chapter 1 of his book [42].

**Theorem 2.7** (Talagrand [42], 1991)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is such that $|f(x) - f(y)| \leq \sigma\|x - y\|_2$ for all $x, y \in \mathbb{R}^n$, and define the random variable $F = f(\xi_1, \xi_2, \ldots, \xi_n)$, where the $\xi_i$ are independent standard normal random variables. Then for all $t \geq 0$,*

$$\mathbb{P}(F > \mathbb{E}[F] + t) \leq e^{-2t^2/(\pi\sigma)^2}.$$

We apply this result to our setting by way of a now standard trick that expresses our random variable of interest as a function of standard Gaussians. In fact, this trick is even discussed in [42], so we could have saved a few lines of the following argument by simply citing a "more applicable" version of Theorem 2.7 (i.e., one for which this trick has already been incorporated); however, the trick so nicely captures the usefulness of Theorem 2.7, that we thought it worth recalling here.

**Proposition 2.8.** *Suppose $\|A\|_2 \leq 1$, and let $X \in \mathbb{C}^n$ be as in Section 2.2. Then for all $t \geq 0$,*

$$\mathbb{P}(\|AX\|_1 > \mathbb{E}[\|AX\|_1] + tn) \leq e^{-nt^2/\pi^3}.$$

*Proof.* To make use of Theorem 2.7, we need to define a suitable $f : \mathbb{R}^n \to \mathbb{R}$, which we do in pieces. First define $\Phi : \mathbb{R} \to \mathbb{R}$ via

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-x^2/2} \, \mathrm{d}x,$$

which is the probability that a standard Gaussian is at most $u$. Then define $g : \mathbb{R}^n \to \mathbb{C}^n$ as

$$g(x_1, \ldots, x_n) = \begin{pmatrix} e^{2\pi i\Phi(x_1)} \\ e^{2\pi i\Phi(x_2)} \\ \vdots \\ e^{2\pi i\Phi(x_n)} \end{pmatrix},$$

and, finally, set $f(x) = \|Ag(x)\|_1$.

Notice that if $\xi_1, \xi_2, \ldots, \xi_n$ are independently sampled from the standard normal distribution, then each $\Phi(\xi_i)$ is distributed uniformly on $[0,1]$. Therefore $g(\xi_1, \ldots, \xi_n)$ has the same distribution as $X$, and so $F := f(\xi_1, \ldots, \xi_n)$ has the same distribution as $\|AX\|_1$.

Now let $x, y \in \mathbb{R}^n$ be arbitrary. Then we have

$$\begin{aligned}|f(x) - f(y)| &= \Big| \|Ag(x)\|_1 - \|Ag(y)\|_1 \Big| \le \|Ag(x) - Ag(y)\|_1 \le \sqrt{n}\|A(g(x) - g(y))\|_2 \\ &\le \sqrt{n}\|A\|_2\|g(x) - g(y)\|_2 \le \sqrt{n}\|g(x) - g(y)\|_2.\end{aligned}$$

Using the fact that $|e^{i\alpha} - 1| \le |\alpha|$ for all $\alpha \in \mathbb{R}$, we further bound the above by

$$\begin{aligned}\|g(x) - g(y)\|_2^2 &= \sum_{j=1}^n |e^{2\pi i\Phi(x_j)} - e^{2\pi i\Phi(y_j)}|^2 = \sum_{j=1}^n |e^{2\pi i(\Phi(x_j) - \Phi(y_j))} - 1|^2 \\ &\le (2\pi)^2 \sum_{j=1}^n |\Phi(x_j) - \Phi(y_j)|^2 \le 2\pi \sum_{j=1}^n |x_j - y_j|^2 = 2\pi\|x - y\|_2^2.\end{aligned}$$

Thus, $|f(x) - f(y)| \le \sqrt{2\pi n}\|x - y\|_2$, and appealing to Theorem 2.7 with $\sigma = \sqrt{2\pi n}$ yields

$$\mathbb{P}(\|AX\|_1 > \mathbb{E}[\|AX\|_1] + tn) = \mathbb{P}(F > \mathbb{E}[F] + tn) \le e^{-2(nt)^2/(\pi\sqrt{2\pi n})^2} = e^{-nt^2/\pi^3}. \quad \square$$

## Finishing the proof for $\mathbb{K} = \mathbb{C}$

**Proposition 2.9.** *Let $\|A\|_2 \le 1$ and $X \in \mathbb{C}^n$ be as in Section 2.2. If $\mathbb{E}[\|AX\|_1/n] = \mu$, then*

$$\mathbb{E}[(\|AX\|_1/n)^n] \le 2\exp[-3n(1-\mu)^2/100].$$

*Proof.* Let $L = t\mu + (1-t)$ with $t \in [0,1]$ to be determined. Since $0 \le \|AX\|_1/n \le 1$, we have (appealing to Proposition 2.8 for the last inequality)

$$\begin{aligned}\mathbb{E}[(\|AX\|_1/n)^n] &\le L^n + \mathbb{P}(\|AX\|_1/n > L) \\ &\le \exp[-n(1-L)] + \mathbb{P}(\|AX\|_1/n - \mu > (1-t)(1-\mu)) \\ &\le \exp[-nt(1-\mu)] + \exp[-n(1-t)^2(1-\mu)^2/\pi^3],\end{aligned}$$

We now take $2t(1-\mu) = \pi^3 + 2 - 2\mu - \pi^{3/2}\sqrt{\pi^3 + 4 - 4\mu}$ (for which $t$ does lie in the interval $[0,1]$), so as to make the exponents equal. For this $t$, we obtain

$$\mathbb{E}[(\|AX\|_1/n)^n] \le 2\exp\Big[-n(2\mu + \pi^{3/2}\sqrt{\pi^3 + 4 - 4\mu} - \pi^3 - 2)/2\Big].$$

Then appealing to the Taylor series at $\mu = 1$, we see that for all $\mu \in [0,1]$,

$$\frac{2\mu + \pi^{3/2}\sqrt{\pi^3 + 4 - 4\mu} - \pi^3 - 2}{2} \geq \frac{(1-\mu)^2}{\pi^3} - \frac{2(1-\mu)^3}{\pi^6} \geq \frac{3(1-\mu)^2}{100}. \qquad \square$$

We then readily obtain Theorem 2.2 simply by combining Propositions 2.6 and 2.9 and using the fact that if $\|A\|_2 \leq 1$, then $0 \leq h_\infty(A) \leq h_2(A) \leq 1$.

### Proof of Theorem 2.4 (better results for $\mathbb{K} = \mathbb{R}$)

For matrices over $\mathbb{R}$, our general strategy is the same as before, but we first partition the rows of $A$ into those that contain 'big' entries and those that do not. We show that the contribution due to rows with large entries has small variance, and although the rows without large entries may each contribute something of high variance, we benefit from the fact that there simply aren't that many such rows. In this way, we are able to obtain better concentration of $\|AX\|_1$ about its mean, which in turn gives a better bound on $\mathrm{perm}(A)$.

We are not sure exactly how to adapt this argument when $\mathbb{K} = \mathbb{C}$, although we admittedly didn't try very hard to do so. We feel confident (especially in light of Theorem 2.4) that Theorem 2.2 can be improved, but we do not think that Theorem 2.4 is best possible either (which is why we haven't worried so much about extending it to $\mathbb{K} = \mathbb{C}$). See Section 2.5 for a discussion of several related conjectures (some perhaps more true than others) and open problems.

### Set-up for the real-valued case

As in Section 2.2, we let $A$ be an $n \times n$ matrix over $\mathbb{R}$ with $\|A\|_2 \leq 1$. Define $t = 1 - h_\infty(A)$. Then to prove Theorem 2.4, our goal is to show

$$|\mathrm{perm}(A)| \leq (n+6)\exp[-\sqrt{nt}/400].$$

Let $\varepsilon > 0$ and $1/10 > \lambda > 0$ be parameters to be determined (we will end up choosing $\varepsilon = t/10$ and $\lambda = 64/\sqrt{nt}$). We now partition the rows of $A$ into "big rows" (those containing an element of absolute value at least $1-\lambda$) and "small rows" (the rest). Suppose there are $b$ big rows and $l = n - b$ small rows. Recall that because

$\|A\|_2 \le 1$, each row and column of $A$ has $l_2$-norm at most 1. Thus, 'large' entries (those of absolute value at least $1-\lambda$) must appear in different rows and columns. By multiplying $A$ by appropriate permutation matrices and the appropriate $\pm 1$-diagonal matrix (which changes neither the norm, nor the absolute value of the permanent, nor the values of $t, b$, or $l$), we can assume $A$ is of the form:

$$A = \begin{pmatrix} B \\ L \end{pmatrix},$$

where $B$ is a $b \times n$ matrix, the $(i,i)$-entries of $B$ are all positive with size at least $1-\lambda$, and all the rest of the entries in $A$ have absolute value less than $1-\lambda$. For convenience, we will assume $b > 0$ and $l > 0$, for if not, our same argument would apply with only superficial alterations.

We recall our earlier set-up as in the complex-case (but with $X \in \mathbb{R}^n$ now uniformly distributed over $\{-1,1\}^n$). Then for all $\tilde{\mu}_B, \tilde{\mu}_L \ge 0$, we have

$$\begin{aligned} |\mathrm{perm}(A)| &\le \mathbb{E}_X\left[\left(\frac{\|AX\|_1}{n}\right)^n\right] = \mathbb{E}_X\left[\left(\frac{\|LX\|_1 + \|BX\|_1}{n}\right)^n\right] \\ &\le \left(\frac{\tilde{\mu}_L + \tilde{\mu}_B}{n} + 2\varepsilon\right)^n + \mathbb{P}\left(\|LX\|_1 \ge \tilde{\mu}_L + \varepsilon n\right) + \mathbb{P}\left(\|BX\|_1 \ge \tilde{\mu}_B + \varepsilon n\right), \end{aligned} \tag{2.1}$$

where (as before) the last inequality is justified by the fact that the random variable within the expected value is bounded above by 1.

We choose

$$\begin{aligned} \tilde{\mu}_B &= \sum_{i=1}^{b}\left[\sqrt{\frac{2}{\pi}} + \left(1 - \sqrt{\frac{2}{\pi}}\right)\|r_i\|_\infty\right] = \sum_{i=1}^{b}\left[1 - \left(1 - \sqrt{\frac{2}{\pi}}\right)(1 - \|r_i\|_\infty)\right], \quad \text{and} \\ \tilde{\mu}_L &= \sum_{i>b}^{n}\left[\sqrt{\frac{2}{\pi}} + \left(1 - \sqrt{\frac{2}{\pi}}\right)\|r_i\|_\infty\right] = \sum_{i>b}^{n}\left[1 - \left(1 - \sqrt{\frac{2}{\pi}}\right)(1 - \|r_i\|_\infty)\right], \end{aligned}$$

where (again) $r_i$ is the $i^{\text{th}}$ row of $A$ (note, $\|r_i\|_\infty = b_{i,i}$ for all $i \le b$). Then by Theorem 2.5 (this time with $\mathbb{K} = \mathbb{R}$), we have $\tilde{\mu}_L \ge \mathbb{E}[\|LX\|_1]$ and $\tilde{\mu}_B \ge \mathbb{E}[\|BX\|_1]$, and by the definitions

$$\frac{\tilde{\mu}_L + \tilde{\mu}_B}{n} = 1 - \left(1 - \sqrt{\frac{2}{\pi}}\right)\frac{1}{n}\sum_{i=1}^{n}\left(1 - \|r_i\|_\infty\right) = 1 - \left(1 - \sqrt{\frac{2}{\pi}}\right)t. \tag{2.2}$$

To take advantage of (2.1), we need only exhibit concentration bounds for $\|LX\|_1$ and $\|BX\|_1$.

## Concentration of $\|LX\|_1$

To show concentration of $\|LX\|_1$ about its mean, we will again apply a version of Talagrand's inequality (but this time suited for the discrete distribution over $\{-1,1\}^n$). Instead of showing the derivation of this from the corresponding general result in [42] (as we did before), we will simply cite [3], in which the following statement appears as Theorem 3.3.

**Theorem 2.10.** *Suppose $M$ is a $k \times n$ real-valued matrix such that $\|M\vec{x}\|_1 \leq \sigma\|\vec{x}\|_2$ for all $\vec{x} \in \mathbb{R}^n$. Let $\xi \in \mathbb{R}^n$ be chosen uniformly from $\{-1,1\}^n$, and let $m$ be a median of $\|M\xi\|_1$. Then for all $\gamma \geq 0$, we have $\mathbb{P}(|\|M\xi\|_1 - m| > \gamma) \leq 4e^{-\gamma^2/(8\sigma^2)}$.*

**Lemma 2.11.** *With notation as before, if $\varepsilon n \geq 16\sqrt{nt\log(n)/\lambda}$, then*

$$\mathbb{P}\left(\|LX\|_1 \geq \tilde{\mu}_L + \varepsilon n\right) \leq 4\exp\left[\frac{-\varepsilon^2 n\lambda}{32t}\right].$$

*Proof.* Note that for all $\vec{x} \in \mathbb{R}^n$, we have $\|L\vec{x}\|_1 \leq \sqrt{l}\|L\vec{x}\|_2 \leq \sqrt{l}\|A\vec{x}\|_2 \leq \sqrt{l}\|\vec{x}\|_2$. Thus, if $m$ is a median of $\|LX\|_1$, then by Theorem 2.10, we have

$$\mathbb{P}(|\|LX\|_1 - m| > \gamma) \leq 4e^{-\gamma^2/(8l)}. \tag{2.3}$$

From this, we see that $\|LX\|_1$ is tightly concentrated about its *median*. However, this also implies

$$m \leq \mathbb{E}[\|LX\|_1] + 8\sqrt{l\log n}, \tag{2.4}$$

since otherwise, we would have

$$\begin{aligned}
\mathbb{E}[\|LX\|_1] &\geq \left(\mathbb{E}[\|LX\|_1] + 4\sqrt{l\log n}\right)\cdot\mathbb{P}\left(|\|LX\|_1 - m| \leq 4\sqrt{l\log n}\right)\\
&\geq \left(\mathbb{E}[\|LX\|_1] + 4\sqrt{l\log n}\right)\cdot(1 - 4/n^2)\\
&= \mathbb{E}[\|LX\|_1] + 4\sqrt{l\log n} - \left(\mathbb{E}[\|LX\|_1] + 4\sqrt{l\log n}\right)\cdot 4/n^2.
\end{aligned}$$

And subtracting $\mathbb{E}[\|LX\|_1]$ from both sides and rearranging, we would obtain

$$n^2 \leq 4 + \frac{\mathbb{E}[\|LX\|_1]}{\sqrt{l\log n}} \leq 4 + \frac{n}{\sqrt{\log n}},$$

which is a contradiction if $n > 2$ (whereas for $n \leq 2$, the desired bound on $m$ is implied by $m \leq n$ [not that it matters]). Therefore, appealing to (2.4), we have

$$\mathbb{P}\left(\|LX\|_1 \geq \tilde{\mu}_L + \varepsilon n\right) \leq \mathbb{P}\left(\|LX\|_1 \geq \mathbb{E}[\|LX\|_1] + \varepsilon n\right) \leq \mathbb{P}\left(\|LX\|_1 \geq m + \varepsilon n - 8\sqrt{l\log n}\right).$$

Furthermore, if $\varepsilon n \geq 16\sqrt{l \log n}$, then we can combine this with (2.3) to obtain

$$\text{if } \varepsilon n \geq 16\sqrt{l \log n}, \text{ then} \qquad \mathbb{P}\left(\|LX\|_1 \geq \tilde{\mu}_L + \varepsilon n\right) \leq 4 \exp\left[\frac{-\varepsilon^2 n^2}{32l}\right]. \tag{2.5}$$

Finally, since $nt \geq \sum_{i=b+1}^{n}(1 - \|r_i\|_\infty) \geq l\lambda$, we know $l \leq nt/\lambda$, completing the proof by (2.5). □

## Concentration of $\|BX\|_1$

We now focus on getting an upper bound on $\mathbb{P}(\|BX\|_1 \geq \tilde{\mu}_B + \varepsilon n)$. We first recall the following classical concentration result.

**Proposition 2.12** (Hoeffding's inequality)**.** *Let $a_1, \dots, a_k$ be real numbers (not all of which are $0$), and let $\xi_1, \xi_2, \dots, \xi_k$ be independent each distributed uniformly on $\{-1, 1\}$. Then for all $\gamma \geq 0$,*

$$\mathbb{P}\left(\sum_{i=1}^{k} a_i \xi_i \geq \gamma\right) \leq \exp\left[\frac{-\gamma^2}{2\sum_{i=1}^{k} a_i^2}\right].$$

Let $\tilde{B} = \begin{pmatrix} B \\ 0 \end{pmatrix}$ be the $n \times n$ matrix whose first $b$ rows are given by $B$ and the rest are 0. Our key step here is replacing $\|BX\|_1$ with $\langle X, \tilde{B}X\rangle$, via the following lemma[1].

**Lemma 2.13.** *With notation as before, if $\lambda < 0.1$ then*

$$\mathbb{P}(\|BX\|_1 \geq \tilde{\mu}_B + \varepsilon n) \leq \mathbb{P}(\langle X, \tilde{B}X\rangle \geq \tilde{\mu}_B + \varepsilon n) + ne^{-1/(5\lambda)}.$$

*Proof.* It suffices to show $\mathbb{P}(\|BX\|_1 \neq \langle X, \tilde{B}X\rangle) \leq ne^{-1/(5\lambda)}$. The idea is that since each row of $B$ is dominated by a single large entry (namely $b_{i,i}$), each entry of $BX$ is a random sum dominated by a single large term (namely $X_i b_{i,i}$). Thus, it is very unlikely that any entry of $BX$ would have a different sign than $X_i b_{i,i}$. This is made rigorous as follows.

Recall that we ordered the columns of $B$ so that the $(i,i)$-entry is the largest in its row, and that $b_{i,i} \geq 1 - \lambda$. Letting $Y_i$ be the $i^{\text{th}}$ coordinate of $BX$, we have, by a simple union bound,

$$\mathbb{P}(\|BX\|_1 \neq \langle X, \tilde{B}X\rangle) \leq \sum_{i=1}^{b} \mathbb{P}(|Y_i| \neq X_i Y_i) = \sum_{i=1}^{b} \mathbb{P}(X_i Y_i < 0) = \sum_{i=1}^{b} \mathbb{P}\left(\sum_{j=1}^{n} X_i X_j b_{i,j} < 0\right).$$

[1] Extending this step is the main obstacle to applying the present argument when $\mathbb{K} = \mathbb{C}$.

Using the fact that for any given $i$, the random vector $(X_iX_j)_{j\neq i}$ has the same joint distribution as $(X_j)_{j\neq i}$ (and that $X_i^2 = 1$), we obtain by Proposition 2.12

$$\sum_{i=1}^{b} \mathbb{P}\left(\sum_{j=1}^{n} X_iX_jb_{i,j} < 0\right) = \sum_{i=1}^{b} \mathbb{P}\left(b_{i,i} < \sum_{j\neq i}^{n} X_jb_{i,j}\right) \leq \sum_{i=1}^{b} \exp\left[\frac{-b_{i,i}^2}{2\sum_{i\neq j} b_{i,j}^2}\right].$$

Since $b_{i,i} \geq 1-\lambda$ and $\sum_j b_{i,j}^2 \leq 1$, this in turn is bounded by

$$\sum_{i=1}^{b} \exp\left[\frac{-b_{i,i}^2}{2\sum_{i\neq j} b_{i,j}^2}\right] \leq n\exp\left[\frac{-(1-\lambda)^2}{2(1-(1-\lambda)^2)}\right] \leq ne^{-1/(5\lambda)},$$

where the last inequality is justified because $0 < \lambda < 0.1$. □

We can now exploit the fact that $\langle X, \tilde{B}X\rangle$ is a degree two polynomial over $\{-1,1\}^n$, allowing us to use any of a variety of concentration inequalities. We will use an inequality of Bonami [13], which was the first *hypercontractivity inequality* of its type. A detailed exposition of such results can be found in chapter 9 of O'Donnell's book [47], and a comparison of this to more recent polynomial concentration inequalities can be found in [51].

**Theorem 2.14** (Bonami [13], 1970)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a degree $k$ polynomial, and consider the random variable $Z = F(\xi_1, \xi_2, \ldots, \xi_n)$, where the $\xi_i$ are independent with each distributed uniformly over $\{-1,1\}$. Then for all $q \geq 2$, we have $\mathbb{E}[|Z|^q] \leq \left((q-1)^k\mathbb{E}[Z^2]\right)^{q/2}$.*

**Lemma 2.15.** *With notation as before, if $\varepsilon n \geq 4e\sqrt{nt}$, then*

$$\mathbb{P}(\langle X, \tilde{B}X\rangle \geq \tilde{\mu}_B + \varepsilon n) \leq \exp\left(\frac{-\varepsilon n}{2e\sqrt{nt}}\right).$$

*Proof.* For $\vec{x} \in \mathbb{R}^n$, define $F(x_1, x_2, \ldots, x_n) = \langle \vec{x}, \tilde{B}\vec{x}\rangle - \sum_{i=1}^{b} b_{i,i}$, and define the random variable $Z = F(X_1, \ldots, X_n)$. Then $\mathbb{P}(\langle X, \tilde{B}X\rangle \geq \tilde{\mu}_B + \varepsilon n) \leq \mathbb{P}(Z \geq \varepsilon n)$, since[2] $\tilde{\mu}_B \geq \sum_{i\leq b} b_{i,i}$. Now $F(x_1, x_2, \ldots, x_n)$ is a degree 2 polynomial, and moreover, by expanding out the sums and using the fact that terms such as $\mathbb{E}[X_iX_j]$ vanish when

[2] In fact, we could have simply taken $\tilde{\mu}_B = \sum_{i\leq b} b_{i,i}$, but we chose instead to define it similarly to $\tilde{\mu}_L$, a change which only affects the constants in our end result.

$i \neq j$, we obtain

$$\begin{aligned}
\mathbb{E}[Z^2] &= \mathbb{E}\left[\left(\sum_{i=1}^{b}\left[-b_{i,i}+\sum_{j=1}^{b}X_iX_jb_{i,j}\right]+\sum_{i=1}^{b}\sum_{j=b+1}^{n}X_iX_jb_{i,j}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{b}\left[-b_{i,i}+\sum_{j=1}^{b}X_iX_jb_{i,j}\right]\right)^2\right]+\mathbb{E}\left[\left(\sum_{i=1}^{b}\sum_{j=b+1}^{n}X_iX_jb_{i,j}\right)^2\right] \\
&= \sum_{i=1}^{b}\sum_{j<i}(b_{i,j}+b_{j,i})^2+\sum_{i=1}^{b}\sum_{j=b+1}^{n}b_{i,j}^2 \leq 2\sum_{i=1}^{b}\sum_{j<i}(b_{i,j}^2+b_{j,i}^2)+2\sum_{i=1}^{b}\sum_{j=b+1}^{n}b_{i,j}^2 \\
&= 2\sum_{i=1}^{b}\left(-b_{i,i}^2+\sum_{j=1}^{n}b_{i,j}^2\right) \leq 2\sum_{i=1}^{b}(1-b_{i,i}^2) \leq 4\sum_{i=1}^{b}(1-b_{i,i}) \leq 4nt.
\end{aligned}$$

Applying Theorem 2.14 with $q = \varepsilon n/(2e\sqrt{nt})$—which is valid since by hypothesis this ratio is at least 2—together with Markov's inequality, we obtain

$$\mathbb{P}(Z \geq \varepsilon n) \leq \mathbb{P}(|Z|^q \geq (\varepsilon n)^q) \leq \frac{\mathbb{E}[|Z|^q]}{(\varepsilon n)^q} \leq \left(\frac{(q-1)2\sqrt{nt}}{\varepsilon n}\right)^q \leq \exp\left(\frac{-\varepsilon n}{2e\sqrt{nt}}\right). \quad \square$$

### Finishing the proof for $\mathbb{K} = \mathbb{R}$

We now need to pick $\varepsilon$ and $\lambda$ to optimize the tradeoffs between our various upper bounds. We need the assumptions of Lemmas 2.11, 2.13, and 2.15—namely (i) $\varepsilon n \geq 16\sqrt{nt\log(n)/\lambda}$, (ii) $\lambda < 0.1$, and (iii) $\varepsilon n \geq 4e\sqrt{nt}$—in which case we can combine these lemmas with (2.1) and (2.2) to obtain

$$\begin{aligned}
|\mathrm{perm}(A)| &\leq \left(2\varepsilon+\frac{\tilde{\mu}_L+\tilde{\mu}_B}{n}\right)^n+\mathbb{P}\left(\|LX\|_1 \geq \tilde{\mu}_L+\varepsilon n\right)+\mathbb{P}\left(\|BX\|_1 \geq \tilde{\mu}_B+\varepsilon n\right) \\
&\leq \left(2\varepsilon+1-\left(1-\sqrt{\frac{2}{\pi}}\right)t\right)^n+4\exp\left[\frac{-\varepsilon^2 n\lambda}{32t}\right]+ne^{-1/(5\lambda)} \\
&\qquad +\exp\left(\frac{-\varepsilon n}{2e\sqrt{nt}}\right).
\end{aligned}$$

We will take $\varepsilon = t/10$ and $\lambda = 64/\sqrt{nt}$, for which we claim that conditions (i), (ii), and (iii) are satisfied. Note that since our goal is to show $|\mathrm{perm}(A)| \leq (n+6)\exp[-\sqrt{nt}/400]$, we may assume $\sqrt{nt}/\log(n+6) \geq 400$ (or the bound we are trying for is worse than the trivial bound of 1) (of course, in any case we are really more interested in large $n$). Notice that with $\varepsilon$ and $\lambda$ as above:

(i) $\varepsilon n \geq 16\sqrt{nt\log(n)/\lambda}$ is equivalent to $\sqrt{nt} \geq 400\log n$;

(ii) $\lambda < 0.1$ is equivalent to $\sqrt{nt} > 640$; and

(iii) $\varepsilon n \geq 4e\sqrt{nt}$ is equivalent to $\sqrt{nt} \geq 40e$.

Thus, these choices of $\lambda$ and $\varepsilon$ allow us to appeal to the aforementioned results, obtaining

$$\begin{aligned}
|\mathrm{perm}(A)| \quad &\leq \quad \left(2\varepsilon + 1 - \left(1 - \sqrt{\frac{2}{\pi}}\right) t\right)^n + 4\exp\left[\frac{-\varepsilon^2 n\lambda}{32t}\right] + ne^{-1/(5\lambda)} \\
&\qquad + \exp\left(\frac{-\varepsilon n}{2e\sqrt{nt}}\right) \\
&\leq \quad \exp\left[-nt\left(1 - \sqrt{2/\pi} - 0.2\right)\right] + 4\exp\left[\frac{-\sqrt{nt}}{50}\right] + n\exp\left[-\frac{\sqrt{nt}}{320}\right] \\
&\qquad + \exp\left[\frac{-\sqrt{nt}}{20e}\right] \\
&\leq \quad (n+6)\exp\left[\frac{-\sqrt{nt}}{400}\right],
\end{aligned}$$

which completes the proof of Theorem 2.4.

## Conclusion

Our biggest (and most natural) open question concerns the optimality of our main results. Namely, a proof of Conjecture 2.3 as stated in Section 2.1 would be very interesting. The main barrier preventing us from proving this conjecture is our reliance on Talagrand's inequality. For $\mathbb{K} = \mathbb{R}$, we partially mitigated the cost of using this inequality via Lemma 2.11, but the application of Theorem 2.10 was still a crucial (though not the only) bottleneck. Our argument could conceivably be pushed further either by a more careful analysis that better uses (2.5) or by a more nuanced argument that splits the matrix $A$ into more than two pieces.

One could also try to avoid using Talagrand's inequality altogether. It is possible that some stronger inequality could replace it (by taking advantage of some aspects particular to our situation), but a more likely "quick fix" of this sort would be a more direct estimate of $\mathbb{E}[(\|AX\|_1/n)^n]$ (in the real case, $AX$ is simply a vector-valued Rademacher sum, which is a well-studied random variable). On the other hand, it could be that the convexity bounds on the Glynn estimator already give away too much to recover anything stronger than what we have.

An entirely different approach would be to determine among matrices with given norm and $h_\infty$, which ones maximize $|\mathrm{perm}(A)|$ (it does not seem impossible that this maximum is always attained by a circulant matrix with all real entries). A characterization of these extremal matrices would certainly be very appealing, and one might hope that thinking along these lines would suggest a more combinatorial approach.

As far as Question 2.B is concerned, we feel that there is still more to be said beyond the present results. Namely, our results only provide a necessary condition for a matrix to have a large permanent (i.e., $h_\infty$ must be large). But there is no clean converse to this statement; consider for example a diagonal matrix with most of its diagonal entries equal to 1 except for one of them equal to 0 (this has large $h_\infty$ and permanent 0). To continue the spirit of the question, we state the following variation of Question 2.B (essentially echoing a question of [1]):

**Problem 2.B$'$:** Find a (deterministic) polynomial-time algorithm that takes an $n \times n$ matrix $A$ of norm 1 and decides whether $|\mathrm{perm}(A)| < n^{-100}$ or $|\mathrm{perm}(A)| > n^{-10}$ (with the understanding that the input matrix will satisfy one of these inequalities).

We attempted this along the following lines: "if the matrix has large permanent, it must have many rows each of which is dominated by a single large entry. If the matrix is of this form, then [heuristic] hopefully that means the permanent is dominated by terms that use at least most of these large entries. Since there are so many large entries, we can efficiently compute the exact contribution of these dominant terms." However, our current results do not allow us to conclude that there are enough rows with large entries (we would like all but about $\log n$ of the rows but are limited to all but about $\log^2 n$ when $\mathbb{K} = \mathbb{R}$ and $\sqrt{n \log n}$ when $\mathbb{K} = \mathbb{C}$). And in fact, even if we could improve our result to the conjectured (and best possible) bound mentioned above, we still do not quite see how to make this heuristic argument yield a polynomial-time algorithm. We should note that Gurvits [30] found a *randomized* algorithm accomplishing the goal of Problem 2.B$'$, and in the deterministic setting, progress towards Problem 2.B$'$ was made in [1] which gives an algorithm in the case that the entries of $A$ are non-negative.

### Further remarks

- We note that there is a lot of freedom in choosing the random variable $X \in \mathbb{K}^n$ for the Glynn estimator ($X$ just needs to have independent components each satisfying $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[|X_i|^2] = 1$). For example, when $\mathbb{K} = \mathbb{R}$, it is tempting to replace $X \in \mathbb{R}^n$ with an $n$-dimensional Gaussian and bound the Glynn estimator by something like

$$|\mathrm{perm}(A)| = \left|\mathbb{E}\left[\prod_i X_i Y_i\right]\right| \leq \mathbb{E}\left[\prod_i |X_i Y_i|\right] \leq \mathbb{E}\left[\left(\frac{1}{n}\sum_i |X_i Y_i|\right)^n\right].$$

  But even if $A$ is the identity matrix this is already (exponentially) larger than 1, which illustrates the difficulty with this approach.

- Via an entirely different method, we were also able to get an upper bound on the permanent for matrices having only non-negative real entries by appealing to the results of [31]. Unfortunately, the bound we obtained is strictly weaker than the results of the present chapter, so it is omitted.

**Acknowledgement:** We thank Hoi Nguyen for introducing us to this problem and sharing [46].

# Chapter 3

# Stability in the Erdős–Ko–Rado theorem

Pat Devlin† Jeff Kahn†

*Abstract:* Denote by $K_p(n,k)$ the random subgraph of the usual Kneser graph $K(n,k)$ in which edges appear independently, each with probability $p$. Answering a question of Bollobás, Narayanan, and Raigorodskii, we show that there is a fixed $p<1$ such that a.s. (i.e., with probability tending to 1 as $k\to\infty$) the maximum independent sets of $K_p(2k+1,k)$ are precisely the sets $\{A\in V(K(2k+1,k)) : x\in A\}$ $(x\in[2k+1])$.

We also complete the determination of the order of magnitude of the "threshold" for the above property for general $k$ and $n\geq 2k+2$. This is new for $k\sim n/2$, while for smaller $k$ it is a recent result of Das and Tran.

## Introduction

The broad context of this chapter is an effort, which has been one of the most interesting and successful combinatorial trends of the last couple decades, to understand how far some of the subject's classical results remain true in a random setting. Since several nice accounts of these developments are available, we will not attempt a review (see, for example, the survey [48] or [11, 8] for discussions closer to present concerns) and mainly confine ourselves to the problem at hand.

Recall that, for integers $0<k<n/2$, the *Kneser graph*, $K(n,k)$ has vertices the $k$-subsets of $[n]$, with two vertices adjacent if and only if they are disjoint sets. In

AMS 2010 subject classification: 05C35, 05D40, 05C80, 05C65, 05C69

Key words and phrases: Erdős-Ko-Rado Theorem, Kneser graph, random subgraph, threshold

†Rutgers University. Work done while supported by NSF grant DMS1201337.

what follows we set $\mathcal{K} = \binom{[n]}{k}$ (the vertex set of $K(n,k)$). A *star* is one of the sets $\mathcal{K}_x := \{A : x \in A\}$ ($x \in [n]$). We also set $M = \binom{n-1}{k-1}$ (the size of a star) and write $\mathcal{C}$ for the collection of $M$-subsets of $\mathcal{K}$ that are *not* stars.

In Kneser-graph terms the classical Erdős-Ko-Rado Theorem [19] says that for $k < n/2$, the independence number of $K(n,k)$ is $M$ and, moreover, the only independent sets of this size are the stars.

Say a spanning subgraph $H$ of $K(n,k)$ has the *EKR property* or *is EKR* if each of its largest independent sets is a star. We are interested in this property for $H = K_p(n,k)$, the random subgraph of $K(n,k)$ in which edges appear independently, each with probability $p$. In particular we are interested in a question suggested and first studied by Bollobás, Narayanan, and Raigorodskii [11], *viz.*

**Question 3.1.** *For what $p = p(n,k)$ is $K_p(n,k)$ likely to be EKR?*

Formally, we would like to estimate the "threshold," $p_c = p_c(n,k)$, which we define to be the unique $p$ satisfying

$$\Pr(K_p(n,k) \text{ is EKR}) = 1/2 \tag{3.1}$$

(which does turn out to be a threshold in the original Erdős-Rényi sense). Ideally (or nearly so) one hopes to identify some $p_0$, necessarily close to $p_c$, such that for fixed $\varepsilon > 0$, $K_p(n,k)$ is a.s. EKR if $p > (1+\varepsilon)p_0$ and a.s. *not* EKR if $p < (1-\varepsilon)p_0$.

Successively stronger results (some of this "ideal" type, some less precise) have been achieved by the aforementioned Bollobás *et al.* [11] and then by Balogh, Bollobás, and Narayanan [8] and Das and Tran [16]. Here we briefly discuss only [16], which subsumes the others.

A natural guess is that the value of $p_c$ is driven by the need to avoid independence of any $\mathcal{F} \in \mathcal{C}$ that, for some $x$, satisfies $|\mathcal{F} \setminus \mathcal{K}_x| = 1$. This turns out to suggest that $p_c = p_c(n,k)$ should be asymptotic to

$$p_0 = p_0(n,k) := \begin{cases} \binom{n-k-1}{k-1}^{-1} \log(n\binom{n-1}{k}) & \text{if } n \geq 2k+2 \\ 3/4 & \text{if } n = 2k+1. \end{cases}$$

Namely, [16] shows (strictly speaking only for $n \geq 2k+2$) that for $p < (1-\varepsilon)p_0$ (with $\varepsilon > 0$ fixed), $K_p(n,k)$ a.s. does contain independent $\mathcal{F}$'s as above (implying

$p_c > (1-o(1))p_0$), while it is easy to see that for $p > (1+\varepsilon)p_0$ it a.s. does not. (Note $n = 2k+1$ is not really special here: the form of $p_0$ changes because we lose the approximation of $1-p$ by $e^{-p}$.)

In fact, Das and Tran show that, for some specified constant $C$, $p_c$ is indeed asymptotic to $p_0$ if $k < n/(3C)$, and is less than $Cnp_0/(n-2k)$ whenever $n \geq 2k+2$, whence $p_c = O(p_0)$ if $k < (1/2 - \Omega(1))n$ (where the first implied constant depends on the second). Of course the estimate becomes less satisfactory as $k/n \to 1/2$, and in particular gives nothing for $n \in \{2k+2, 2k+1\}$. On the other hand, both [11] and [8] suggest that $n = 2k+1$ is the most interesting case of the problem and ask whether one can at least show that $K_p(2k+1,k)$ is a.s. EKR for some $p$ bounded away from 1. Here we prove such a result and also show that $p_c = O(p_0)$ remains true for general $n$ and $k$.

**Theorem 3.2.** *There is a fixed $p < 1$ such that (for every $k$) $K_p(2k+1,k)$ is a.s. EKR. There is a fixed $C$ such that for every $n$ and $k$, $K_p(n,k)$ is a.s. EKR for $p > Cp_0(n,k)$.*

Again, one expects that $p_c \sim p_0$ in all cases and in particular that, as suggested in [8], $p_c(2k+1,k) \to 3/4$; but we do not come close to these asymptotics and make no attempt to squeeze the best possible $\varepsilon$ and/or $C$ from our arguments.

It may be worth (briefly) comparing the present question with a similar one, introduced earlier by Balogh, Bohman and Mubayi [7], in which one considers a random *induced* subgraph of $K(n,k)$. Thus, one specifies only $\mathcal{H} = \mathcal{K}_p$ (the random set in which each $A \in \mathcal{K}$ is present with probability $p$, independent of other choices) and asks when the subgraph induced by $\mathcal{H}$ has the *EKR property*, now meaning that each largest independent set (that is, intersecting subfamily of $\mathcal{H}$) is a *star* $\mathcal{H}_x = \{A \in \mathcal{H} : x \in A\}$ for some $x$.

For $n = 2k+1$ the situation here is similar to the one above: EKR should hold (a.s.) for any fixed $p > 3/4$, but even proving this for $p > 1-\varepsilon$ with a fixed $\varepsilon > 0$—a problem suggested in [7]—does not seem easy; such a proof was given in [34] (using methods unrelated to those employed here).

But the resemblance may be superficial, and in fact the induced problem seems considerably subtler than the one considered here (as should probably be expected, e.g. since (i) the size of a largest star is itself a moving target and (ii) the most likely violators of EKR are not always families that are close to stars). See [33] for a guess as to what ought to be true here and [9, 33] for what's known at this time.

The rest of the chapter is devoted to the proof of Theorem 3.2. A single argument will suffice for both assertions, though, as noted below, not all of what we do is needed for $n = 2k + 1$.

### Proof

*Notation.* From now on we take $n = 2k+c$ and write $V$ for $[n]$ (so $\mathcal{K} = \binom{V}{k}$). For $\mathcal{H} \subseteq \mathcal{K}$, we let $\mathcal{H}_x = \{A \in \mathcal{H} : x \in A\}$, $\mathcal{H}_{\overline{x}} = \mathcal{H} \setminus \mathcal{H}_x$ ($x \in V$) and $\Delta_{\mathcal{H}} = \max\{|\mathcal{H}_x| : x \in V\}$. As usual $|\mathcal{H}_x|$ is the *degree* of $x$ in $\mathcal{H}$. We use $M$ and $\mathcal{C}$ as above and set $N = \binom{n-1}{k}$. For $\mathcal{F} \in \mathcal{C}$ we set $a_{\mathcal{F}} = M - \Delta_{\mathcal{F}}$ and $e(\mathcal{F}) = |\{\{A, B\} : A, B \in \mathcal{F}, A \cap B = \emptyset\}|$ (the number of Kneser edges in $\mathcal{F}$).

In view of [16] we may assume

$$k > 6c. \tag{3.2}$$

We also assume henceforth that

$$p > \begin{cases} 1 - \varepsilon & \text{if } c = 1, \\ Cp_0(n, k) & \text{if } c \geq 2 \end{cases} \tag{3.3}$$

for suitable fixed $C$ and $\varepsilon > 0$ (namely, ones that support our arguments) and want to show that then

$$\Pr(\text{some } \mathcal{F} \in \mathcal{C} \text{ is independent in } K_p(n, k)) = o(1).$$

Perhaps surprisingly, this is given by a straight union bound; that is, there are lower bounds on the sizes of the various $e(\mathcal{F})$'s that imply (with $\mathcal{F}$ running over $\mathcal{C}$)

$$\sum (1 - p)^{e(\mathcal{F})} = o(1). \tag{3.4}$$

This contrasts with (e.g.) [34], where a naive union bound gives nothing.

The rest of our discussion is devoted to the proof of (3.4), and we assume from now on that $\mathcal{F} \in \mathcal{C}$. Notice that we always have

$$a_{\mathcal{F}}/N \leq k/n \tag{3.5}$$

(since the trivial $\Delta_{\mathcal{F}} \geq k|\mathcal{F}|/n = kM/n$ gives $a_{\mathcal{F}} \leq (1-k/n)M = kN/n$).

The next assertion is the main point.

**Lemma 3.3.** *There is a fixed $\vartheta > 0$ such that for any $\mathcal{F} \in \mathcal{C}$,*

$$e(\mathcal{F}) > \vartheta k^{-1}\tbinom{n-k-1}{k-1} a_{\mathcal{F}} \log(N/a_{\mathcal{F}}). \tag{3.6}$$

We first observe that this easily gives (3.4). Noting that (for any $a$) the number of $\mathcal{F}$'s with $a_{\mathcal{F}} = a$ is at most $n\binom{M}{a}\binom{N}{a}$ (choose a maximum degree vertex $x$ of $\mathcal{F}$ and then the $a$-sets $\mathcal{K}_x \setminus \mathcal{F} \subseteq \mathcal{K}_x$ and $\mathcal{F}_{\overline{x}} \subseteq \mathcal{K}_{\overline{x}}$), we find that, with $\vartheta$ as in Lemma 3.3, the sum in (3.4) is then less than

$$n \sum \left\{ \tbinom{M}{a}\tbinom{N}{a} \exp[-\xi\vartheta k^{-1}\tbinom{n-k-1}{k-1} a \log(N/a)] : 0 < a \leq kN/n \right\}, \tag{3.7}$$

where $\xi = \log(1/\varepsilon)$ if $n = 2k+1$ and otherwise $\xi = p$. We may bound the summand using $\binom{M}{a}\binom{N}{a} < \exp[2a\log(eN/a)]$ and

$$\xi\vartheta k^{-1}\tbinom{n-k-1}{k-1} \geq \begin{cases} \vartheta \log(1/\varepsilon) & \text{if } n = 2k+1, \\ C\vartheta k^{-1} \log(n\binom{n-1}{k}) & \text{otherwise,} \end{cases}$$

and the expression in (3.7) is then easily seen to be small if (say) $\varepsilon < e^{-5/\vartheta}$ or $C > 4/\vartheta$ (for $n = 2k+1$ and $n \geq 2k+2$ respectively).

∎

The proof of Lemma 3.3 divides into three regimes, depending on $a_{\mathcal{F}}$. The first of these—$a_{\mathcal{F}}$ not too small—is handled as in [16], from which we recall only what we need (see their Theorem 1.2):

**Theorem 3.4.** *There is a fixed $K$ such that for any $2 \leq k < n/2$: if $\mathcal{F} \in \mathcal{C}$ satisfies $a_{\mathcal{F}} > K\zeta \frac{n}{c} M$ with $\zeta \leq \frac{c}{(10K)^2 n}$, then $e(\mathcal{F}) > \zeta M \binom{n-k-1}{k-1}$.*

It will be convenient to assume (as we may) that $K \geq 1$. Theorem 3.4 gives (3.6) for any $\mathcal{F}$ satisfying $a_{\mathcal{F}} > M/(100K)$ (with $\vartheta$ something like $.01K^{-2}$), so we assume from now on that this is not the case.

For smaller values we need to say a little about graphs belonging to the "Johnson scheme" (e.g. [45]). For positive integers $k \leq m$ we use $J_i(m,k)$ for the graph on $V_{m,k} := \binom{[m]}{k}$ with $A, B$ adjacent ($A \sim_i B$) iff $|A \Delta B| = 2i$. Here we take $m = n-1$ and will be interested in $i \in \{1, c\}$. Uniform measure on $V_{m,k}$ will be denoted $\mu_k$.

We use $\beta_i(\mathcal{A})$ for the size of the *edge boundary* of $\mathcal{A} \subseteq V_{m,k}$ in $J_i(m,k)$; that is,

$$\beta_i(\mathcal{A}) = |\{\{A, A'\} : A \in \mathcal{A},\ A' \in (V_{m,k} \setminus \mathcal{A}),\ A \sim_i A'\}|.$$

The following lower bounds on $\beta_c$ and $\beta_1$ will suffice for our purposes.

For $\beta_c$ we use a standard version of the eigenvalue-expansion connection due to Alon and Milman [5] (see e.g. [6, Theorem 9.2.1]), which (here) says that for any $\mathcal{A} \subseteq V_{m,k}$,

$$\beta_c(\mathcal{A}) \geq \lambda |\mathcal{A}|(1 - \mu_k(\mathcal{A})), \tag{3.8}$$

with $\lambda$ the smallest positive eigenvalue of the Laplacian of $J_c(m,k)$ (the matrix $DI_N - A$, where $D = \binom{k}{c}\binom{m-k}{c}$ and $A$ are the degree and adjacency matrix of $J_c(m,k)$). We assert that (assuming (3.2))

$$\lambda = \frac{m}{k}\binom{k}{c}\binom{m-k-1}{c-1}. \tag{3.9}$$

*Proof.* (This ought to be known, but we couldn't find a reference.) The eigenvalues of $A$ are (again, see e.g. [45])

$$\lambda_j := \textstyle\sum_{i=0}^{c}(-1)^i S_i^j, \qquad j = 0, \ldots, k,$$

where $S_i^j := \binom{j}{i}\binom{k-j}{c-i}\binom{m-k-j}{c-i}$. In particular, $\lambda_0 = \binom{k}{c}\binom{m-k}{c}$,

$$\lambda_1 = \tbinom{k-1}{c}\tbinom{m-k-1}{c} - \tbinom{k-1}{c-1}\tbinom{m-k-1}{c-1} = \tbinom{k-1}{c}\tbinom{m-k-1}{c}\tfrac{km-k^2-cm}{km-k^2-cm+c^2} \tag{3.10}$$

and $\lambda_0 - \lambda_1 = \lambda$ (the value in (3.9)), so we just need to show that $\lambda_j \leq \lambda_1$ for $j \geq 2$. In fact it is enough to show that

$$S_i^j \leq \lambda_1 \quad \text{whenever } j \geq 2, \tag{3.11}$$

since log-concavity of the sequences $\left(\binom{a}{\ell}\right)_\ell$ implies log-concavity of $(S_i^j)_i$ and thus $\lambda_j \leq \max_i S_i^j$.

Routine manipulations (using the expression for $\lambda_1$ in (3.10)) give

$$S_i^j/\lambda_1 = \frac{km-k^2}{km-k^2-cm}\frac{\binom{c}{i}\binom{k-c}{j-i}}{\binom{k}{j}}\frac{\binom{m-k-c}{j-i}\binom{c}{i}}{\binom{m-k}{j}}\frac{1}{\binom{j}{i}} \leq \frac{1}{\binom{j}{i}}\frac{km-k^2}{km-k^2-cm}.$$

For $0 < i < j$, the r.h.s. is less than 1 since $km - k^2 < 2(km - k^2 - cm)$, as follows easily from (3.2). On the other hand, it is easy to see (using (3.2)) that each of $S_0^2$ and $S_2^2$ is less than $\lambda_1$, which gives (3.11) for $i \in \{0, j\}$, since $S_0^j$ and $S_j^j$ are decreasing in $j$.

∎

For $\beta_1$ we use an instance of a result of Lee and Yau [43] (estimating the log-Sobolev constant for $J_1(m,k)$): there is a fixed $\gamma > 0$ such that, for any $k$ as in (3.2) and $\mathcal{A} \subseteq \binom{[m]}{k}$,

$$\beta_1(\mathcal{A}) > \gamma m|\mathcal{A}|\log(1/\mu_k(\mathcal{A})). \tag{3.12}$$

*Proof of Lemma 3.3.* As already noted, Theorem 3.4 gives Lemma 3.3 when $a_{\mathcal{F}} > M/(100D)$, so we assume this is not the case.

We assume (w.l.o.g.) that $x = n$ is a maximum degree vertex of $\mathcal{F}$ and set $\mathcal{A} = \mathcal{F}_{\overline{x}}$ and

$$\mathcal{B} = \{V \setminus T : T \in \mathcal{K}_x \setminus \mathcal{F}\}$$

(so $|\mathcal{A}| = |\mathcal{B}| = a_{\mathcal{F}}$).

As above we take $m = n - 1$. The rest of our discussion takes place in the universe $V \setminus \{x\} = [m]$. We use $\Gamma_l$ for $\binom{[m]}{l}$—thus $\mathcal{A} \subseteq \Gamma_k$ (our earlier $V_{m,k}$) and $\mathcal{B} \subseteq \Gamma_{k+c}$—and set $\bar{\mathcal{A}} = \Gamma_k \setminus \mathcal{A}$ and $\bar{\mathcal{B}} = \Gamma_{k+c} \setminus \mathcal{B}$.

For $\mathcal{S} \subseteq \Gamma_k$ and $\mathcal{T} \subseteq \Gamma_{k+c}$, set

$$\Lambda(\mathcal{S}, \mathcal{T}) = |\{(A, B) \in \mathcal{S} \times \mathcal{T} : A \subseteq B\}|.$$

Notice that

$$e(\mathcal{F}) = \Lambda(\mathcal{A}, \bar{\mathcal{B}}) + e(\mathcal{A}) \geq \Lambda(\mathcal{A}, \bar{\mathcal{B}}). \tag{3.13}$$

We next observe that lower bounds on the $\beta$'s imply lower bounds on the quantities $\Lambda(\mathcal{A}, \bar{\mathcal{B}})$:

**Proposition 3.5.** *For any* $\mathcal{F} \in \mathcal{C}$*,*

$$\Lambda(\mathcal{A}, \bar{\mathcal{B}}) \geq \max\left\{(2\tbinom{k}{c})^{-1}\beta_c(\mathcal{A}), \tfrac{1}{2ck}\tbinom{k+c-2}{c-1}\beta_1(\mathcal{A})\right\} - \tbinom{k+c-1}{c-1}|\mathcal{A}|/2. \tag{3.14}$$

Of course in view of (3.13) this gives the same lower bound on $e(\mathcal{F})$.

*Proof.* The combination of

$$\Lambda(\mathcal{A}, \mathcal{B}) + \Lambda(\mathcal{A}, \bar{\mathcal{B}}) = \Lambda(\mathcal{A}, \Gamma_{k+c}) = \tbinom{k+c-1}{c}|\mathcal{A}|$$

and

$$\Lambda(\mathcal{A}, \mathcal{B}) + \Lambda(\bar{\mathcal{A}}, \mathcal{B}) = \Lambda(\Gamma_k, \mathcal{B}) = \tbinom{k+c}{c}|\mathcal{B}| = \tbinom{k+c}{c}|\mathcal{A}|$$

gives

$$\Lambda(\bar{\mathcal{A}}, \mathcal{B}) = \Lambda(\mathcal{A}, \bar{\mathcal{B}}) + \tbinom{k+c-1}{c-1}|\mathcal{A}|. \tag{3.15}$$

For the second bound in (3.14) we work in the ("Johnson") graph $J_1(m,k)$. Write $\Phi$ for the number of triples $(A, B, A') \in \mathcal{A} \times \Gamma_{k+c} \times \bar{\mathcal{A}}$ with $A \sim_1 A'$ and $A \cup A' \subseteq B$. Since each relevant pair $(A, A')$ admits exactly $\binom{k+c-2}{c-1}$ choices of $B$, we have

$$\Phi = \beta_1(\mathcal{A})\tbinom{k+c-2}{c-1}. \tag{3.16}$$

On the other hand, for each of the above triples, either $(A, B)$ is one of the pairs counted by $\Lambda(\mathcal{A}, \bar{\mathcal{B}})$ or $(A', B)$ is one of the pairs counted by $\Lambda(\bar{\mathcal{A}}, \mathcal{B})$ (and not both). In the first case the number of choices of $A'$ is at most the number of neighbors of $A$ contained in $B$, namely $ck$, and similarly in the second case. This with (3.15) gives

$$\Phi \leq (\Lambda(A, \bar{\mathcal{B}}) + \Lambda(\bar{\mathcal{A}}, \mathcal{B}))ck = ck(2\Lambda(A, \bar{\mathcal{B}}) + \tbinom{k+c-1}{c-1}|\mathcal{A}|),$$

and then combining with (3.16) yields the stated bound.

The argument for the first bound is similar and we just indicate the changes. We work in $J_c(m,k)$ and consider triples as above but with $A \sim_c A'$ (so $B = A \cup A'$). The number of triples, which is now just $\beta_c(\mathcal{A})$, is bounded above by

$$(\Lambda(A, \bar{\mathcal{B}}) + \Lambda(\bar{\mathcal{A}}, \mathcal{B}))\tbinom{k}{c} = \tbinom{k}{c}(2\Lambda(A, \bar{\mathcal{B}}) + \tbinom{k+c-1}{c-1}|\mathcal{A}|)$$

($\binom{k}{c}$ being the number of neighbors—now in $J_c(m,k)$—of $A$ contained in $B$ when $A \in \Gamma_k$, $B \in \Gamma_{k+c}$ and $A \subseteq B$), and the desired bound follows.

∎

Finally, combining (3.14) with (3.13) and our earlier bounds on the $\beta$'s (see (3.8)-(3.12)) yields (with $\gamma$ as in (3.12))

$$e(\mathcal{F}) \geq \tfrac{|\mathcal{A}|}{2}\tbinom{k+c-2}{c-1}\max\{\left(1-\tfrac{m}{k}\mu_k(\mathcal{A})\right), \tfrac{\gamma m}{ck}\log\tfrac{1}{\mu_k(\mathcal{A})} - \tfrac{k+c-1}{k}\}. \tag{3.17}$$

(Replacing $\beta_c(\mathcal{A})$ in (3.14) by the lower bound provided by (3.8) and (3.9) gives

$$\begin{aligned} e(\mathcal{F}) &\geq (2\tbinom{k}{c})^{-1}\tfrac{m}{k}\tbinom{k}{c}\tbinom{m-k-1}{c-1}|\mathcal{A}|(1-\mu_k(\mathcal{A})) - \tbinom{k+c-1}{c-1}\tfrac{|\mathcal{A}|}{2} \\ &= \tfrac{|\mathcal{A}|}{2}\left[\tfrac{m}{k}\tbinom{m-k-1}{c-1}(1-\mu_k(\mathcal{A})) - \tbinom{k+c-1}{c-1}\right] \\ &= \tfrac{|\mathcal{A}|}{2}\tbinom{k+c-2}{c-1}\left[1-\tfrac{m}{k}\mu_k(\mathcal{A})\right], \end{aligned}$$

and replacing $\beta_1$ by the lower bound from (3.12) yields

$$e(\mathcal{F}) \geq \tfrac{1}{2ck}\tbinom{k+c-2}{c-1}\gamma m|\mathcal{A}|\log\tfrac{1}{\mu_k(\mathcal{A})} - \tbinom{k+c-1}{c-1}\tfrac{|\mathcal{A}|}{2},$$

which is easily seen to be equal to the second bound in (3.17).)

It only remains to observe that this does what we want, namely that (for suitable $\vartheta$) the expression in (3.17) is at least as large as the bound in (3.6), or, equivalently, that the max in (3.17) is at least

$$\tfrac{2(k+c-1)}{ck}\vartheta\log(N/a_{\mathcal{F}}) < (4/c)\vartheta\log(1/\mu_k(\mathcal{A})); \tag{3.18}$$

we assert that this is true provided $\vartheta < \gamma/5$.

If $\log(1/\mu_k(\mathcal{A})) \leq c/\gamma$, then the r.h.s. of (3.18) is less than the first term in the max (which is essentially 1 since $\mu_k(\mathcal{A}) = |\mathcal{A}|/N < |\mathcal{A}|/M$, which we are assuming is less than $.01D^{-1}$; see following Theorem 3.4).

If, on the other hand, $\log(1/\mu_k(\mathcal{A})) > c/\gamma$, then the second term in the max is at least

$$\tfrac{\gamma}{c}\tfrac{2k+c-1}{k}\log\tfrac{1}{\mu_k(\mathcal{A})} - \tfrac{\gamma}{c}\tfrac{k+c-1}{k}\log\tfrac{1}{\mu_k(\mathcal{A})} = \tfrac{\gamma}{c}\log\tfrac{1}{\mu_k(\mathcal{A})},$$

which is again greater than the r.h.s. of (3.18). ∎

For $n = 2k+1$ we could avoid the machinery used above for intermediate values of $|\mathcal{A}|$ (namely (3.8), (3.9) and the first bound in (3.14)) by choosing $\zeta$ in Theorem 3.4 to handle $\log(1/\mu_k(\mathcal{A})) \leq c/\gamma$ (and adjusting $\vartheta$ accordingly).

# Chapter 4

# Proof of an entropy conjecture of Leighton and Moitra

Hüseyin Acan[†] Pat Devlin[‡] Jeff Kahn[‡]

*Abstract:* We prove the following conjecture of Leighton and Moitra. Let $T$ be a tournament on $[n]$ and $\mathfrak{S}_n$ the set of permutations of $[n]$. For an arc $uv$ of $T$, let $A_{uv} = \{\sigma \in \mathfrak{S}_n : \sigma(u) < \sigma(v)\}$.

**Theorem.** For a fixed $\varepsilon > 0$, if $\mathbb{P}$ is a probability distribution on $\mathfrak{S}_n$ such that $\mathbb{P}(A_{uv}) > 1/2 + \varepsilon$ for every arc $uv$ of $T$, then the binary entropy of $\mathbb{P}$ is at most $(1 - \vartheta_\varepsilon) \log_2 n!$ for some (fixed) positive $\vartheta_\varepsilon$.

When $T$ is transitive the theorem is due to Leighton and Moitra; for this case we give a short proof with a better $\vartheta_\varepsilon$.

## Introduction

In what follows we use lg for $\log_2$ and $H(\cdot)$ for binary entropy. The purpose of this note is to prove the following natural statement, which was conjectured by Tom Leighton and Ankur Moitra [44] (and told to the third author by Moitra in 2008).

**Theorem 4.1.** *Let $T$ be a tournament on $[n]$ and $\sigma$ a random (not necessarily uniform) permutation of $[n]$ satisfying:*

$$\text{for each arc } uv \text{ of } T, \;\; \mathbb{P}(\sigma(u) < \sigma(v)) > 1/2 + \varepsilon. \tag{4.1}$$

---

AMS 2010 subject classification: 05C20, 05D40, 94A17, 06A07

Key words and phrases: entropy, permutations, tournaments, regularity

[†]Rutgers University. Work done while supported by National Science Foundation Fellowship (Award No. 1502650)

[‡]Rutgers University. Work done while supported by NSF grant DMS1501962

*Then*

$$H(\sigma) \leq (1 - \vartheta) \lg n!, \tag{4.2}$$

*where* $\vartheta > 0$ *depends only on* $\varepsilon$*.*

(We will usually think of permutations as bijections $\sigma : [n] \to [n]$). The original motivation for Leighton and Moitra came mostly from questions about sorting partially ordered sets; see [44] for more on this.

For the special case of *transitive* $T$, Theorem 4.1 was proved in [44] with $\vartheta_\varepsilon = C\varepsilon^4$. Note that for a *typical* (a.k.a. *random*) $T$, the conjecture's hypothesis is unachievable, since, as shown long ago by Erdős and Moon [20], no $\sigma$ agrees with $T$ on more than a $(1/2 + o(1))$-fraction of its arcs. In fact, it seems natural to expect that transitive tournaments are the *worst* instances, being the ones for which the hypothesized agreement is easiest to achieve. From this standpoint, what we do here may be considered somewhat unsatisfactory, as our $\vartheta$'s are quite a bit worse than those in [44]. For transitive $T$ it's easy to see [44, Claim 4.14] that one can't take $\vartheta$ greater than $2\varepsilon$, which seems likely to be close to the truth. We make some progress on this, giving a surprisingly simple proof of the following improvement of [44].

**Theorem 4.2.** *For* $T$*,* $\mathbb{P}$*,* $\sigma$ *as Theorem 4.1 with* $T$ *transitive,*

$$H(\sigma) \leq (1 - \varepsilon^2/8) n \lg n. \tag{4.3}$$

The proof of Theorem 4.1 is given in Section 4.3 following brief preliminaries in Section 4.2. The underlying idea is similar to that of [44], which in turn was based on the beautiful tournament ranking bound of W. Fernandez de la Vega [17]; see Section 4.3 (end of "Sketch") for an indication of the relation to [44]. Theorem 4.2 is proved in Section 4.4.

## Preliminaries

### *Usage*

In what follows we assume $n$ is large enough to support our arguments.

As usual $G[X]$ is the subgraph of $G$ induced by $X$; we use $G[X,Y]$ for the bipartite subgraph induced (in the obvious sense) by disjoint $X$ and $Y$. For a digraph $D$, $D[X]$ and $D[X,Y]$ are used analogously. For both graphs and digraphs, we use $|\cdot|$ for number of edges (or arcs).

Also as usual, the *density* of a pair $(X,Y)$ of disjoint subsets of $V(G)$ is $d(X,Y) = d_G(X,Y) = |G[X,Y]|/(|X||Y|)$, and we extend this to bipartite digraphs $D$ in which

$$\text{at most one of } D\cap(X\times Y),\ D\cap(Y\times X) \text{ is nonempty.} \tag{4.4}$$

For a digraph $D$, $D^r$ is the digraph gotten from $D$ by reversing its arcs.

For $\sigma\in\mathfrak{S}_n$, we use $T_\sigma$ for the corresponding (transitive) tournament on $[n]$ (that is, $uv\in T_\sigma$ iff $\sigma(u)<\sigma(v)$) and for a digraph $D$ (on $[n]$) define

$$\mathrm{fit}(\sigma,D) = |D\cap T_\sigma| - |D^r\cap T_\sigma|$$

(e.g. when $D$ is a tournament, this is a measure of the quality of $\sigma$ as a ranking of $D$).

*Regularity*

Here we need just Szemerédi's basic notion [52] of a regular pair and a very weak version (Lemma 4.3) of his Regularity Lemma. As usual a bipartite graph $H$ on disjoint $X\cup Y$ is $\delta$-*regular* if

$$|d_H(X',Y') - d_H(X,Y)| < \delta$$

whenever $X'\subseteq X$, $Y'\subseteq Y$, $|X'|>\delta|X|$ and $|Y'|>\delta|Y|$, and we extend this in the obvious way to the situation in (4.4). It is easy to see that if a bigraph $H$ is $\delta$-regular then its bipartite complement is as well; this implies that for a tournament $T$ on $[n]$ and $X$, $Y$ disjoint subsets of $[n]$,

$$T\cap(X\times Y) \text{ is } \delta\text{-regular if and only if } T\cap(Y\times X) \text{ is.} \tag{4.5}$$

The following statement should perhaps be considered folklore, though similar results were proved by János Komlós, circa 1991 (see [39, Sec. 7.3]).

**Lemma 4.3.** *For each $\delta>0$ there is a $\beta > 2^{-\delta^{-O(1)}}$ such that for any bigraph $H$ on $X\cup Y$ with $|X|,|Y|\geq n$, there is a $\delta$-regular pair $(X',Y')$ with $X'\subseteq X, Y'\subseteq Y$ and each of $|X'|,|Y'|$ at least $\beta n$.*

**Corollary 4.4.** *For each $\delta > 0$, $\beta$ as in Lemma 4.3 and digraph $G = (V, E)$, there is a partition $L \cup R \cup W$ of $V$ such that $E \cap (L \times R)$ is $\delta$-regular and* $\min\{|L|, |R|\} \geq \beta|V|/2$.

*Proof.* Let $X \cup Y$ be an (arbitrary) equipartition of $V$ and apply Lemma 4.3 to the undirected graph $H$ underlying the digraph $G \cap (X \times Y)$. □

## Proof of Theorem 4.1

We now assume that $\sigma$ drawn from the probability distribution $\mathbb{P}$ on $\mathfrak{S}_n$ satisfies (4.1) and try to show (4.2) (with $\vartheta$ TBA). We use $\mathbb{E}$ for expectation w.r.t. $\mathbb{P}$ and $\mu$ for uniform distribution on $\mathfrak{S}_n$.

*Sketch and connection with* [44]

We will produce $S_1, \ldots, S_m \subseteq T$ with $S_i \subseteq L_i \times R_i$ for some disjoint $L_i, R_i \subseteq [n]$, satisfying:

(i) with $\|S_i\| := \min\{|L_i|, |R_i|\}$, $\sum \|S_i\| = \Omega(n \lg n)$ (where the implied constant depends on $\varepsilon$);

(ii) each $S_i$ is $\delta$-regular (with $\delta = \delta_\varepsilon$ TBA);

(iii) for all $i < j$, either $(L_i \cup R_i) \cap (L_j \cup R_j) = \emptyset$ or $L_j \cup R_j$ is contained in one of $L_i, R_i$ (note this implies the $S_i$'s are disjoint).

Let $A_i = \{\mathrm{fit}(\sigma, S_i) > \varepsilon|S_i|\}$ and $Q = \{\sum\{\|S_i\| : A_i \text{ occurs}\} = \Omega(n \lg n)\}$. The main points are then:

(a) $\mathbb{P}(Q)$ is bounded below by a positive function of $\varepsilon$. (This is just (i) together with a couple applications of Markov's Inequality.)

(b) Regularity of $S_i$ implies $\mu(A_i) \leq \exp[-\Omega(\|S_i\|)]$.

(c) Under (iii), for any $I \subseteq [m]$,

$$\mu(\cap_{i \in I} A_i) < \exp[-\textstyle\sum_{i \in I} \Omega(\|S_i\|)]$$

(a weak version of independence of the $A_i$'s under $\mu$).

And these points easily combine to give (4.2) (see (4.7) and (4.9)).

For the transitive case in [44] most of this argument is unnecessary; in particular, regularity disappears and there is a natural *decomposition* of $T$ into $S_i$'s: Supposing $T = \{ab : a < b\}$ and (for simplicity) $n = 2^k$, we may take the $S_i$'s to be the sets $L_i \times R_i$ with $(L_i, R_i)$ running over pairs

$$([(2s-2)2^{-j}n + 1, (2s-1)2^{-j}n], [(2s-1)2^{-j}n + 1, 2s2^{-j}n]), \tag{4.6}$$

with $j \in [k]$ and $s \in [2^{j-1}]$. (As mentioned earlier, this decomposition of the (identity) permutation $(1, \ldots, n)$ also provides the framework for [17].) After some translation, our argument (really, a fairly small subset thereof) then specializes to essentially what's done in [44]. $\square$

Set $\delta = .03\varepsilon$ and let $\beta$ be half the $\beta$ of Lemma 4.3 and Corollary 4.4. We use the corollary to find a rooted tree $\mathcal{T}$ each of whose internal nodes has degree (number of children) 2 or 3, together with disjoint subsets $S_1, S_2, \ldots, S_m$ of (the arc set of) $T$, corresponding to the internal nodes of $\mathcal{T}$. The nodes of $\mathcal{T}$ will be subsets of $[n]$ (so the *size,* $|U|$, of a node $U$ is its size as a set).

To construct $\mathcal{T}$, start with root $V_1 = [n]$ and repeat the following for $k = 1, \ldots$ until each unprocessed node has size less than (say) $t := \sqrt{n}$. Let $V_k$ be an unprocessed node of size at least $t$ and apply Corollary 4.4 to $T[V_k]$ to produce a partition $V_k = L_k \cup R_k \cup W_k$, with $|L_k|, |R_k| > \beta|V_k|$ and $S_k := T \cap (L_k \times R_k)$ $\delta$-regular of density at least 1/2. (Note (4.5) says we can reverse the roles of $L_k$ and $R_k$ if the density of $T \cap (L_k \times R_k)$ is less than 1/2.) Add $L_k, R_k, W_k$ to $\mathcal{T}$ as the children of $V_k$ and mark $V_k$ "processed." (Note the $V_k$'s are the *internal* nodes of $\mathcal{T}$; nodes of size less then $t$ are not processed and are automatically leaves. Note also that there is no restriction on $|W_k|$ and that, for $k > 1$, $V_k$ is equal to one of $L_i$, $R_i$, $W_i$ for some $i < k$.)

Let $m$ be the number of internal nodes of $\mathcal{T}$ (the final tree). Note that the leaves of $\mathcal{T}$ have size at most $t$ and that the $S_i$'s satisfy (ii) and (iii) of the proof sketch; that they also satisfy (i) is shown by the next lemma.

Set

$$\Lambda = \textstyle\sum_{i=1}^m |V_i|;$$

this quantity will play a central role in what follows.

**Lemma 4.5.** $\Lambda \geq \frac{1}{2} n \log_3 n$

*Proof.* This will follow easily from the next general (presumably known) observation, for which we assume $\mathcal{T}$ is a tree satisfying:

- the nodes of $\mathcal{T}$ are subsets of $S$, an $s$-set which is also the root of $\mathcal{T}$;
- the children of each internal node $U$ of $\mathcal{T}$ form a partition of $U$ with at most $b$ blocks;
- the leaves of $\mathcal{T}$ are $U_1, \ldots, U_r$, with $|U_i| = u_i \leq t$ (any $t$) and depth $d_i$.

**Lemma 4.6.** *With the setup above,* $\sum u_i d_i \geq s \log_b(s/t)$.

(Of course this is exact if $\mathcal{T}$ is the complete $b$-ary tree of depth $d$ and all leaves have size $2^{-b}s$).

*Proof.* Recall that the *relative entropy* between probability distributions $p$ and $q$ on $[r]$ is

$$D(p\|q) = \sum p_i \lg(q_i/p_i) \leq 0$$

(the inequality given by the concavity of the logarithm). We apply this with $p_i = u_i/s$ and $q_i$ the probability that the ordinary random walk down the tree ends at $u_i$. In particular $q_i \geq b^{-d_i}$, which, with nonpositivity of $D(p\|q)$ and the assumption $u_i \leq t$, gives

$$\begin{aligned} \sum (u_i/s) d_i \log b &\geq \sum (u_i/s) \lg(1/q_i) \\ &\geq \sum (u_i/s) \lg(s/u_i) \geq \lg(s/t). \end{aligned}$$

The lemma follows. □

This gives Lemma 4.5 since $\sum |V_i| = \sum_U |U| d(U)$, with $U$ ranging over leaves of $\mathcal{T}$ (and $d(\cdot)$ again denoting depth). □

**Lemma 4.7.** *The number* $m$ *of internal nodes of* $\mathcal{T}$ *is less than* $n$.

*Proof.* Let $m(V_i)$ denote the number of internal nodes of the subtree starting at $V_i$. For any internal node $m(V_i) = 1 + \sum m(U)$, where the sum is taken over the children of $V_i$, and—since each internal node has degree at least 2—the claim follows readily by induction on $n$. $\square$

Recalling that $A_i = \{\sigma \in \mathfrak{S}_n : \mathrm{fit}(\sigma, S_i) \geq \varepsilon|S_i|\}$ and that $\mathbb{E}$ refers to $\mathbb{P}$, we have $\mathbb{E}[\mathrm{fit}(\sigma, S_i)] \geq 2\varepsilon|S_i|$, which with

$$\mathbb{E}[\mathrm{fit}(\sigma, S_i)] \leq \mathbb{P}(A_i)|S_i| + (1 - \mathbb{P}(A_i))\varepsilon|S_i| \leq (\mathbb{P}(A_i) + \varepsilon)|S_i|$$

gives $\mathbb{P}(A_i) \geq \varepsilon$ (essentially Markov's Inequality applied to $|S_i| - \mathrm{fit}(\sigma, S_i)$).

Set $\xi_i = |V_i|\mathbf{1}_{A_i}$ and $\xi = \sum_i \xi_i$, and let $Q$ be the event $\{\xi \geq \varepsilon\Lambda/2\}$. Then $\mathbb{E}[\xi_i] = |V_i|\mathbb{P}(A_i) \geq \varepsilon|V_i|$, implying $\mathbb{E}[\xi] = \sum \mathbb{E}[\xi_i] \geq \varepsilon\Lambda$, and (since $\xi_i \leq |V_i|$) $\xi \leq \Lambda$. So using Markov's Inequality as above gives $\mathbb{P}(Q) \geq \varepsilon/2$.

Thus, with $\sigma$ chosen from $\mathfrak{S}_n$ according to $\mathbb{P}$, we have

$$\begin{aligned} H(\sigma) &\leq 1 + (1 - \mathbb{P}(Q)) \lg n! + \mathbb{P}(Q) \lg |Q| \\ &= 1 + \lg n! + \mathbb{P}(Q) \lg \mu(Q) \leq 1 + \lg n! + (\varepsilon/2) \lg \mu(Q) \end{aligned} \tag{4.7}$$

(recall $\mu$ is the uniform measure on $\mathfrak{S}_n$).

Let

$$\mathcal{J} = \{I \subseteq [m] : \textstyle\sum_{i \in I} |V_i| \geq \varepsilon\Lambda/2\}$$

and, for $I \in \mathcal{J}$, let $A_I = \cap_{i \in I} A_i$. Set

$$b = \varepsilon^2\delta\beta^3/33 \tag{4.8}$$

(see (4.13) for the reason for the choice of $b$). We will show, for each $I \in \mathcal{J}$,

$$\mu(A_I) \leq e^{-b\varepsilon\Lambda/2}, \tag{4.9}$$

which implies

$$\lg \mu(Q) = \lg \mu(\cup_{I \in \mathcal{J}} A_I) \leq \lg |\mathcal{J}| - (b\varepsilon\Lambda \lg e)/2 \leq n - (b\varepsilon\Lambda \lg e)/2,$$

the second inequality following from $|\mathcal{J}| \le 2^m$ together with Lemma 4.7. With $c = \varepsilon^3\delta\beta^3/150 < (b\varepsilon \log_3 e)/4$, this bounds (for large $n$) the r.h.s. of (4.7) as

$$H(\sigma) \le (1 - \varepsilon c/2) \lg n!,$$

which proves Theorem 4.1 with $\vartheta = \varepsilon^4\delta\beta^3/300 = \exp[-\varepsilon^{-O(1)}]$. □

The rest of our discussion is devoted to the proof of (4.9). For a digraph $D \subseteq L \times R$ with $L, R$ disjoint subsets of $V$, say a pair $(X, Y)$ of disjoint subsets of $[n]$ with $|X| = |L|$, $|Y| = |R|$ is *safe* for $D$ if

$$\mathrm{fit}(\tau, D) < \varepsilon|L||R|/4 \tag{4.10}$$

for every bijection $\tau : L \cup R \to X \cup Y$ with $\tau(L) = X$ (where $\mathrm{fit}(\tau, D)$ has the obvious meaning). We also say $\sigma \in \mathfrak{S}_n$ is *safe* for $D$ if $(\sigma(L), \sigma(R))$ is. Note that since $S_i$ has density at least 1/2 in $L_i \times R_i$, the $\sigma$'s in $A_i$ are unsafe for $S_i$.

**Lemma 4.8.** *Assume the above setup with $|L| + |R| = l$ and $|L| = \gamma l$, and set $\lambda = 2\delta$ and $\zeta = \varepsilon\delta\gamma(1-\gamma)/4$. Let $I_1 \cup \cdots \cup I_r$ be the natural partition of $X \cup Y$ into intervals of size $\lambda l$. If $D$ is $\delta$-regular and*

$$|X \cap I_j| = (\gamma\lambda \pm \zeta)l \quad \forall j \in [r], \tag{4.11}$$

*then $(X, Y)$ is safe for $D$.*

(Of course an *interval* of $Z = \{i_1 < \cdots < i_u\}$ is one of the sets $\{i_s, \ldots, i_{s+t}\}$.)

*Proof.* For $\tau$ as in the line after (4.10), let $L_j = L \cap \tau^{-1}(I_j)$ and $R_j = R \cap \tau^{-1}(I_j)$ ($j \in [r]$). Then

$$|\mathrm{fit}(\tau, D)| \le \sum_{1 \le i < j \le r} ||D \cap (L_i \times R_j)| - |D \cap (L_j \times R_i)|| + \gamma(1-\gamma)\lambda l^2. \tag{4.12}$$

Here the last term is an upper bound on the contribution of pairs contained in the $I_j$'s: if $|L_j| = \gamma_j|I_j| = \gamma_j\lambda l$ (so $|R_j| = (1-\gamma_j)\lambda l$ and $\sum \gamma_j = \gamma/\lambda$), then

$$\sum \gamma_j(1-\gamma_j) \le \sum \gamma_j - (\sum \gamma_j)^2/r = (\gamma - \gamma^2)/\lambda$$

gives

$$\sum |L_j||R_j| = \sum \gamma_j(1-\gamma_j)\lambda^2 l^2 \le \gamma(1-\gamma)\lambda l^2.$$

On the other hand, regularity and (4.11) (which implies $|L_i| > \delta|L|$ $(= \delta\gamma l)$ since $\gamma\lambda - \zeta > \gamma\delta$, and similarly $|R_i| > \delta|R|$) give, for all $i \neq j$,

$$|D \cap (L_i \times R_j)| = (d \pm \delta)|L_i||R_j|,$$

where $d$ is the density of $D$. Combining this with (4.11) bounds each of the summands in (4.12) by

$$\begin{aligned}[]&[(d+\delta)(\gamma\lambda+\zeta)((1-\gamma)\lambda+\zeta) - (d-\delta)(\gamma\lambda-\zeta)((1-\gamma)\lambda-\zeta)]l^2 \\ &\qquad\qquad = 2[\lambda\zeta d + \delta(\gamma(1-\gamma)\lambda^2+\zeta^2)]l^2\end{aligned}$$

and the r.h.s. of (4.12) by

$$\left\{2\tbinom{r}{2}[\lambda\zeta d + \delta(\gamma(1-\gamma)\lambda^2+\zeta^2)] + \gamma(1-\gamma)\lambda\right\} l^2 < \varepsilon\gamma(1-\gamma)l^2/4.$$

(The main term on the l.h.s. is the one with $\lambda\zeta d$, which, since $r^{-1} = \lambda = 2\delta$, is less than half the r.h.s. The second and third terms are much smaller (the second since $\delta$ is much smaller than $\varepsilon$).) □

**Corollary 4.9.** *For $D$ and parameters as in Lemma 4.8, and $\sigma$ uniform from $\mathfrak{S}_n$,*

$$\Pr(\sigma \textit{ is unsafe for } D) < 2r\exp[-2\zeta^2 l/\lambda].$$

*Proof.* Let $(X,Y) = (\sigma(L), \sigma(R))$. Once we've chosen $X \cup Y$ (determining $I_1, \ldots, I_r$), $2\exp[-2\zeta^2 l/\lambda]$ is the usual Hoeffding bound [35, Eq. (2.3)] on the probability that $X$ violates (4.11) for a given $j$. (The bound may be more familiar when elements of $X \cup Y$ are in $X$ independently, but also applies to the *hypergeometric* r.v. $|X \cap I_j|$; see e.g. [36, Thm. 2.10 and (2.12)].) □

*Proof of* (4.9). Let

$$B_i = \{\sigma \in \mathfrak{S}_n : \sigma \text{ is unsafe for } S_i\}$$

and $B_I = \cap_{i \in I} B_i$. Then $A_i \subseteq B_i$ (as noted above) and (therefore) $A_I \subseteq B_I$. Moreover—perhaps the central point—the $B_i$'s are independent, since $B_i$ depends only on the relative positions of $\sigma(L_i)$ and $\sigma(R_i)$ within $\sigma(V_i)$.

On the other hand, Corollary 4.9, applied with $D = S_i$ (so $L = L_i$, $R = R_i$, $l = |L_i| + |R_i|$ and $\gamma = |L_i|/l \in (\beta, 1-\beta)$) gives

$$\begin{aligned} \Pr(B_i) &< 2r\exp[-2\zeta^2 l/\lambda] < 2r\exp[-\varepsilon^2\delta\beta^2 l/64] \\ &< 2r\exp[-\varepsilon^2\delta\beta^3|V_i|/32] < e^{-b|V_i|}. \end{aligned} \tag{4.13}$$

(Recall $b$ was defined in (4.8); since we assume $|V_i|$ is large ($|V_i| > t = \sqrt{n}$), the choice leaves a little room to absorb the $2r$.) And of course (4.13) and the independence of the $B_i$'s give (4.9). □

### Back to the transitive case

Theorem 4.2 is an easy consequence of the next observation.

**Lemma 4.10.** *Let $2r^2$ a random $m$-subset of $[2m]$ satisfying*

$$\mathbb{E}|\{(a,b) : a < b, a \in [2m] \setminus 2r^2, b \in 2r^2\}| > (\tfrac{1}{2} + \varepsilon)m^2. \tag{4.14}$$

*Then $H(2r^2) < (1 - \varepsilon^2/8)2m$.*

To get Theorem 4.2 from this, let $T = \{ab : a < b\}$ and, for simplicity, $n = 2^k$, and decompose $T = \bigcup(L_i \times R_i)$ as in (4.6). For each $i$, say with $|L_i|$ $(= |R_i|) = m_i$, let $2r_i^2 \subseteq [2m_i]$ consist of the indices of positions within $\sigma(L_i \cup R_i)$ occupied by $\sigma(R_i)$; that is, if $\sigma(L_i \cup R_i) = \{j_1 < \cdots < j_{2m_i}\}$, then $2r_i^2 = \{l : j_l \in \sigma(R_i)\}$. Then Lemma 4.10 (its hypothesis provided by (4.1)) gives

$$H(2r_i^2) \leq (1 - \varepsilon^2/8)2m_i;$$

so, since $\sigma$ is determined by the $2r_i^2$'s, we have

$$H(\sigma) \leq \sum H(2r_i^2) \leq (1 - \varepsilon^2/8)\sum(2m_i) = (1 - \varepsilon^2/8)n\lg n. \quad \square$$

*Remark.* Note that the $\Omega(\varepsilon^2)$ of Theorem 4.2 is the best one can do without more fully exploiting (4.1) (that is, beyond (4.14) for the $(L_i, R_i, Y_i)$'s, which is all we are using).

*Proof of Lemma 4.10.* For $a \in [2m]$, set $\mathbb{P}(a \in 2r^2) = 1/2 + \delta_a$. Then

$$H(2r^2) \leq \sum_a H(1/2 + \delta_a) \leq \sum_a (1 - 2\delta_a^2)$$

(where the 2 could actually be $2 \lg e$); so it is enough to show

$$\sum \delta_a^2 \geq \varepsilon^2 m/8.$$

For a given $m$-subset $Y$ of $[2m]$, we have

$$\begin{array}{rcl} f(Y) & := & |\{(a,b) : a < b, a \in [2m] \setminus Y, b \in Y\}| \\ & = & \sum_{b \in Y}(b-1) - \mathbb{C}m2 = \sum_{b \in Y} b - \mathbb{C}m + 12. \end{array}$$

(the first sum counts pairs $(a,b)$ with $a < b$ and $b \in Y$, and $\mathbb{C}m2$ is the number of such pairs with $a$ also in $Y$); so we have

$$(\tfrac{1}{2} + \varepsilon)m^2 < \mathbb{E}f(2r^2) = \sum(\tfrac{1}{2} + \delta_b)b - \mathbb{C}m + 12 = \sum \delta_b b + m^2/2,$$

implying $\sum \delta_b b > \varepsilon m^2$. Combining this with $2m \sum_{\delta_b > 0} \delta_b \geq \sum \delta_b b$, we have $\sum_{\delta_b > 0} \delta_b > \varepsilon m/2$ and then, using Cauchy-Schwarz,

$$\sum \delta_b^2 \geq \sum_{\delta_b > 0} \delta_b^2 \geq \tfrac{1}{2m}(\varepsilon m/2)^2 = \varepsilon^2 m/8. \qquad \square$$

# Chapter 5

# Fractional matchings in $k$-out hypergraphs

Pat Devlin† Jeff Kahn†

*Abstract:* Extending the notion of (random) $k$-out graphs, we consider when the $k$-out hypergraph is likely to have a perfect fractional matching. In particular, we show that for each $r$ there is a $k = k(r)$ such that the $k$-out $r$-uniform hypergraph on $n$ vertices has a perfect fractional matching with high probability (i.e., with probability tending to 1 as $n \to \infty$) and prove an analogous result for $r$-uniform $r$-partite hypergraphs. This is based on a new notion of hypergraph expansion and the observation that sufficiently expansive hypergraphs admit perfect fractional matchings. As a further application, we give a short proof of a stopping-time result originally due to Krivelevich.

## Introduction

Hypergraphs constitute a far-reaching generalization of graphs and a basic combinatorial construct but are notoriously difficult to work with. A *hypergraph* is a collection $\mathcal{H}$ of subsets (*"edges"*) of a set $V$ of *"vertices."* Such an $\mathcal{H}$ is *r-uniform* (or an *r-graph*) if each edge has cardinality $r$ (so 2-graphs are graphs). A *perfect matching* in a hypergraph is a collection of edges partitioning the vertex set. For any $r > 2$, deciding whether an $r$-graph has a perfect matching is an NP-complete problem [38]; so instances of the problem tend to be both interesting and difficult. Of particular interest here has been trying to understand conditions under which a *random* hypergraph is likely to

---

AMS 2010 subject classification: 05C65, 05C70, 05D40, 05C80, 90C05, 90C32

Key words and phrases: random hypergraphs, perfect fractional matchings, k-out model, hypergraph expansion

†Rutgers University. Work done while supported by NSF grant DMS1501962

have a perfect matching.

The most natural model of a random $r$-graph is the "Erdős-Rényi" model, in which each $r$-set is included in $\mathcal{H}$ with probability $p$, independent of other choices. One is then interested in the "threshold," roughly, the order of magnitude of $p = p_r(n)$ required to make a perfect matching likely. Here the graph case was settled by Erdős and Rényi [21, 22], but for $r > 2$ the problem—which became known as Shamir's Problem following [18]—remained open until [37]. In each case, the obvious obstruction to containing a perfect matching is existence of an isolated vertex (that is, a vertex contained in no edges), and a natural guess is that this is the *main* obstruction. A literal form of this assertion—the *stopping time* version—says that if we choose random edges *sequentially*, each uniform from those as yet unchosen, then we w.h.p. have a perfect matching as soon as all vertices are covered. This nice behavior does hold for graphs [12], but for hypergraphs remains conjectural (though at least the value it suggests for the threshold is correct).

An interesting point here is that taking $p$ large enough to avoid isolated vertices produces many more edges than other considerations—e.g., wanting a large *expected* number of perfect matchings—suggest. This has been one motivation for the substantial body of work on models of random graphs in which isolated vertices are automatically avoided, notably random *regular* graphs (e.g., [55]) and the $k$-out model. The generalization of the latter to hypergraphs, which we now introduce, will be our main focus here.

> **The $k$-out model.** For a ("host") hypergraph $\mathcal{H}$ on $V$, $\mathcal{H}(k\text{-out})$ is the random subhypergraph $\cup_{v \in V} E_v$, where $E_v$ is chosen uniformly from the $k$-subsets of $\mathcal{H}_v := \{A \in \mathcal{H} : v \in A\}$ (or—but we won't see this—$E_v = \mathcal{H}_v$ if $|\mathcal{H}_v| < k$), these choices made independently.

The $k$-out model for $\mathcal{H} = K_{n,n}$ (the complete bipartite graph) was introduced by Walkup [54], who showed that w.h.p. $K_{n,n}(2\text{-out})$ is Hamiltonian, so in particular contains a perfect matching, and Frieze [27] proved the nonbipartite counterpart of the matching result, showing that $K_{2n}$(2-out) has a perfect matching w.h.p. (Hamiltonicity in the latter case turned out to be more challenging; it was studied in [23, 28, 14] and finally resolved by Bohman and Frieze [10], who proved $K_n$(3-out) is Hamiltonian w.h.p.). The idea of a general host $G$ was introduced by Frieze and T. Johansson [25]; see also e.g., Ferber *et al.* [24] for (*inter alia*) a nice connection with $G_{n,p}$.

For *hypergraphs* the $k$-out model seems not to have been studied previously (random regular hypergraphs *have* been considered, e.g., in [15]). Here the two most important examples would seem to be $\mathcal{H} = K_n^{(r)}$ (the complete $r$-graph on $n$ vertices) and $\mathcal{H} = K_{[n]^r}$ (the complete $r$-partite $r$-graph with $n$ vertices in each part). It is natural to expect that for each of these there is some $k = k(r)$ for which $\mathcal{H}$($k$-out) has a perfect matching w.h.p.. Note that, while almost certainly correct, these are likely to be difficult, as either would imply the aforementioned resolution of Shamir's Problem; still, we would like to regard the following linear relaxations as a small step in this direction. (Relevant definitions are recalled in Section 5.2.)

**Theorem 5.1.** *For each $r$, there is a $k$ such that w.h.p. $K_n^{(r)}$($k$-out) admits a perfect fractional matching and $w \equiv 1/r$ is the only fractional cover of weight $n/r$.*

**Theorem 5.2.** *For each $r$, there is a $k$ such that w.h.p. $\mathcal{H} = K_{[n]^r}$($k$-out) admits a perfect fractional matching and each minimum weight fractional cover of $\mathcal{H}$ is constant on each block of the $r$-partition.*

Our upper bounds on the $k$'s are quite large (roughly $r^r$), but in fact we don't even know that they must be larger than 2 (though this sounds optimistic), and we make no attempt to optimize. In the more interesting case of (ordinary) perfect matchings, consideration of the expected number of perfect matchings shows that $k$ does need to be be at least exponential in $r$.

We will make substantial use of the next observation (or, in the $r$-partite case, of the analogous Proposition 5.8, whose statement we postpone), in which the notion of expansion may be of some interest. Recall that an *independent set* in a hypergraph is a set of vertices containing no edges.

**Proposition 5.3.** *Suppose $\mathcal{H}$ is an $r$-graph in which, for all disjoint $X, Y \subseteq V$ with $X$ independent and*

$$|Y| < (r-1)|X|, \tag{5.1}$$

*there is some edge meeting $X$ but not $Y$. Then $\mathcal{H}$ has a perfect fractional matching. If, moreover we replace "$<$" by "$\leq$" in* (5.1)*, then $w \equiv 1/r$ is the only fractional cover of weight $n/r$.*

It's not hard to see that for $r > 2$ the proof of this can be tweaked to give the stronger conclusion even under the weaker hypothesis. (For $r = 2$ this is clearly false, e.g., if $G$ is a matching.)

Related notions of expansion (respectively stronger than and incomparable to ours) appear in [41] and [32]. An additional application of Proposition 5.3, given in Section 5.4, is a short alternate proof of the following result of Krivelevich [41].

**Theorem 5.4.** *Let $\{\mathcal{H}_t\}_{t\geq 0}$ denote the random $r$-graph process on $V$ in which each step adds an edge chosen uniformly from the current non-edges, let $T$ denote the first $t$ for which $\mathcal{H}_t$ has no isolated vertices. Then $\mathcal{H}_T$ has a perfect fractional matching w.h.p..*

**Outline.** Section 5.2 includes definitions and brief linear programming background. Section 5.3 treats $K_n^{(r)}$, proving Proposition 5.3 and Theorem 5.1, and the corresponding results for $K_{[n]^r}$ are proved in Section 5.4. Finally, Section 5.5 returns to $K_n^{(r)}$, using Proposition 5.3 to give an alternate proof of Theorem 5.4.

## Preliminaries

Except where otherwise specified, $\mathcal{H}$ is an $r$-graph on $V = [n]$.

We need to recall a minimal amount of linear programming background (see e.g., [50] for a more serious discussion). For a hypergraph $\mathcal{H}$, a *fractional (vertex) cover* is a map $w : V \to [0,1]$ such that $\sum_{v\in e} w(v) \geq 1$ for all $e \in \mathcal{H}$; the *weight* of a cover $w$ is $|w| = \sum_v w(v)$; and the *fractional cover number*, $\tau^*(\mathcal{H})$, is the largest such weight. Similarly a *fractional matching* of $\mathcal{H}$ is a $\varphi : \mathcal{H} \to [0,1]$ such that $\sum_{e\ni v} \varphi(e) \leq 1$ for all $v \in V$; the weight of such a $\varphi$ is defined as for fractional covers; and the *fractional matching number*, $\nu^*(\mathcal{H})$, is the *maximum* weight of a fractional matching.

In this context, LP-duality says that $\nu^*(\mathcal{H}) = \tau^*(\mathcal{H})$ for any hypergraph. For $r$-graphs the common value is trivially at most $n/r$ (e.g., since $w \equiv 1/r$ is a fractional cover). A fractional matching in an $r$-graph is *perfect* if it achieves this bound; that is, if $\sum \varphi_e = n/r$ (equivalently $\sum_{e \ni v} \varphi_e = 1\ \forall v$, which would be the definition of perfection in a nonuniform $\mathcal{H}$).

Finally, given $\mathcal{H}$ we say a nonempty $X \subseteq V$ is $\lambda$-*expansive* if for all $Y \subseteq V \setminus X$ of size at most $\lambda|X|$, there is some edge meeting $X$ but not $Y$.

## Proofs of Proposition of 5.3 and Theorem 5.1

*Proof of Proposition 5.3.* It is enough to show that if $w$ is a fractional cover with $t_0 := 1/r - \min_v w(v) > 0$, then $|w| \geq n/r$, with the inequality strict if we assume the stronger version of (5.1). We give the argument under this stronger assumption; for the weaker, just replace the few strict inequalities below by nonstrict ones. Given $w$ as above, set, for each $t > 0$,

$$W_t = \{v \in [n] \ : \ w(v) \leq \tfrac{1}{r} - t\}, \quad W^t = \{v \in [n] \ : \ w(v) \geq \tfrac{1}{r} + t\}.$$

Since $w$ is a fractional cover, each edge meeting $W_t$ must also meet $W^{t/(r-1)}$ (or the weight on the edge would be less than 1); so, since $W_t$ is independent, the hypothesis of Proposition 5.3 gives $|W^{t/(r-1)}| > (r-1)|W_t|$ for $t \in (0, t_0]$ (the $t$'s for which $W_t \neq \emptyset$).

For $s \in \mathbb{R}$, define $f(s) = |\{v \in [n] \ : \ w(v) \geq s\}|$. Then

$$\begin{aligned} \int_0^1 f(s)\ \mathrm{d}s &= \int_0^1 \sum_{v \in [n]} \mathbb{1}_{\{w(v) \geq s\}}\ \mathrm{d}s \\ &= \sum_{v \in [n]} \int_0^1 \mathbb{1}_{\{w(v) \geq s\}}\ \mathrm{d}s = \sum_{v \in [n]} w(v) = \tau^*(\mathcal{H}). \end{aligned}$$

We also have $|W^t| = f(1/r + t)$ and $|W_t| \geq n - f(1/r - t)$, implying

$$f(1/r + t/(r-1)) \geq (r-1)(n - f(1/r - t)),$$

with the inequality strict if $t \in (0, t_0]$. Thus,

$$\begin{aligned}
\tau^*(\mathcal{H}) &= \int_0^1 f(s)\ \mathrm{d}s = \int_0^{1/r} f(s)\ \mathrm{d}s + \int_{1/r}^1 f(s)\ \mathrm{d}s \\
&= \int_0^{1/r} f(1/r - t)\ \mathrm{d}t + \int_0^{(r-1)^2/r} \frac{f(1/r + t/(r-1))}{r-1}\ \mathrm{d}t \\
&\geq \int_0^{1/r} \left[ f(1/r - t) + \frac{f(1/r + t/(r-1))}{r-1} \right]\ \mathrm{d}t \\
&> \int_0^{1/r} \left[ f(1/r - t) + (r-1)\frac{n - f(1/r - t)}{r-1} \right]\ \mathrm{d}t = \frac{n}{r}.
\end{aligned}$$

□

We should perhaps note that the converse of Proposition 5.3 is not true in general (failing, e.g., if $r > 2$ and $\mathcal{H}$ is itself a perfect matching). But in the graphic case ($r = 2$) the converse *is* true (and trivial), and the proposition provides an alternate proof of the following characterization, which is [49, Thm. 2.2.4] (and is also contained in [4, Thm. 2.1], e.g.).

**Corollary 5.5.** *A graph has a perfect fractional matching iff* $|N(I)| \geq |I|$ *for all independent* $I$.

(where $N(I)$ is the set of vertices with at least one neighbor in $I$).

*Proof of Theorem 5.1.* Given $r$, let (without trying to optimize) $k = (2r^2)^r$ and $c = k^{-1/r} = 1/(2r^2)$, and let $\mathcal{H} = K_n^{(r)}(k\text{-out})$. Theorem 5.1 (with this $k$) is an immediate consequence of Proposition 5.3 and the next two routine lemmas. (As usual $\alpha(\mathcal{H})$ is the size of a largest independent set in $\mathcal{H}$.)

**Lemma 5.6.** *W.h.p.* $\alpha(\mathcal{H}) < cn$.

**Lemma 5.7.** *W.h.p. every* $X \subseteq V(\mathcal{H})$ *with* $|X| \leq cn$ *is* $(r-1)$*-expansive.*

*Proof of Lemma 5.6.* The probability that $S \in \mathbb{C}[n]s$ is independent in $\mathcal{H}$ is

$$\left[1 - \tfrac{(s-1)_{r-1}}{(n-1)_{r-1}}\right]^{sk} < \exp\left[-sk\left(\tfrac{s-r}{n}\right)^{r-1}\right].$$

(where $(a)_b = a(a-1)\cdots(a-b+1)$), and summing this over $S$ of size $cn$ bounds $\mathbb{P}(\alpha \geq cn)$ by

$$2^n \exp\left[-cnk(c - r/n)^{r-1}\right] = \exp\left[n\left(\log 2 - (1 - o(1))kc^r\right)\right],$$

which tends to 0 as desired. □

*Proof of Lemma 5.7.* For $X$, $Y$ disjoint subsets of $[n]$, let $B(X,Y)$ be the event that $Y$ meets all edges meeting $X$. Then, with $x = |X|$ and $y = |Y|$,

$$\mathbb{P}(B(X,Y)) \leq \left[1 - \tfrac{(n-y-1)_{r-1}}{(n-1)_{r-1}}\right]^{kx} \leq \left[1 - \left(\tfrac{n-y-r}{n}\right)^{r-1}\right]^{kx} \leq \left[\tfrac{r(y+r)}{n}\right]^{kx},$$

the last inequality following from

$$1-(1-x)^m \leq mx \tag{5.2}$$

(valid for $x \in [0,1]$ and nonnegative integer $m$). The probability that the conclusion of the lemma fails is thus less than

$$\begin{aligned}\sum \tbinom{n}{rx}\tbinom{rx}{x}\left[\tfrac{r(y+r)}{n}\right]^{kx} &< \sum \left(\tfrac{ne}{rx}\right)^{rx} 2^{rx}\left[\tfrac{r(y+r)}{n}\right]^{kx}\\ &= \sum \left[(2e)^r \left(\tfrac{rx}{n}\right)^{k-r}((r-1)+r/x)^k\right]^x\\ &< \sum \left[(4er)^r (r(2r-1)x/n)^{k-r}\right]^x = o(1),\end{aligned}$$

where the sums are over $1 \leq x \leq cn$. □

□

## Proof of Theorem 5.2

As in the proof of Theorem 5.1 we first show that the conclusions of Theorem 5.2 are implied (deterministically) by sufficiently good expansion and then show that $K_{[n]^r}$($k$-out) w.h.p. expands as desired. We take $V = V_1 \cup \cdots \cup V_r$ to be our $r$-partition (so $|V_i| = n$ $\forall i$) and below always assume $\mathcal{H} \subseteq K_{[n]^r}$.

**Proposition 5.8.** *Suppose $\varepsilon \in (0, 1/2)$ and $\lambda > 2r^2$ are fixed and $\mathcal{H}$ satisfies: for any $i \in [r]$, $T \subseteq V_i$, $U_j \subseteq V_j$ for $j \neq i$ and $U = \cup_{j\neq i} U_j$, there is an edge meeting $T$ but not $U$ provided either*

*(i) $|T| \leq \varepsilon n$ and $|U_j| \leq \lambda|T|$ $\forall j \neq i$, or*

*(ii) $|T| \geq \varepsilon n$ and $|U_j| \leq (1-\varepsilon)n$ $\forall j \neq i$.*

*Then $\mathcal{H}$ admits a perfect fractional matching, and every minimum weight fractional cover of $\mathcal{H}$ is constant on each $V_i$.*

*Proof.* Define a *balanced assignment* to be a $w : V \to \mathbb{R}$ with $\sum_{v \in V_i} w(v) = 0$ and $w(e) \geq 0$ for all $e \in \mathcal{H}$.

We claim that (under our hypotheses) the only balanced assignment is the trivial $w \equiv 0$. To get Proposition 5.8 from this, let $f$ be a minimum weight fractional cover, and let $w_f(v) = f(v) - \sum_{u \in V_i} f(u)/n$, for each $i$ and $v \in V_i$. Then $w_f$ is a balanced assignment: $\sum_{v \in V_i} w_f(v) = 0$ is obvious and nonnegativity holds since $f(e) \geq 1$ and, by minimality, $\sum_{v \in V} f(v) \leq n$. Thus $w_f \equiv 0$, implying $f$ is as promised.

Suppose then that $w$ is a balanced assignment. For $X \subseteq V$ and $t \geq 0$, set $X^t = \{v \in X : w(v) \geq t\}$, $X_t = \{v \in X : w(v) < -t\}$, $X^+ = X^0$ and $X^- = X_0$, and define the *value* of $X$ to be $\psi(X) = \sum_{v \in X} |w(v)|$. Let $S = \{i \in [r] : |V_i^-| \leq \varepsilon n\}$ and $B = [r] \setminus S$.

**Lemma 5.9.** *If $X \subseteq V^-$ and $|X| \leq \varepsilon n$, then $\psi(X) \leq r\psi(V^+)/\lambda$.*

*Proof.* For any $t > 0$, note that every edge meeting $X_t$ meets $V^{t/(r-1)}$ since otherwise, we could find an edge of negative weight. So since $|X_t| \leq |X| \leq \varepsilon n$, condition (i) implies $|V^{t/(r-1)}| \geq \lambda |X_t|$. Thus,

$$\begin{aligned}\psi(V^+) &= \int_0^\infty |V^u| \, \mathrm{d}u = \frac{1}{r-1} \int_0^\infty |V^{t/(r-1)}| \, \mathrm{d}t \\ &\geq \frac{\lambda}{r-1} \int_0^\infty |X_t| \, \mathrm{d}t = \frac{\lambda}{r-1} \psi(X). \quad \square\end{aligned}$$

**Lemma 5.10.** *If $|(V_i)_t| \leq \varepsilon n$, then $\max_{j \in S} |V_j^{t/(r-1)}| \geq (1-\varepsilon)n$.*

*Proof.* Since any edge meeting $(V_i)_t$ meets $\cup_{j \neq i} V_j^{t/(r-1)}$ and $|V_j^+| \leq (1-\varepsilon)n$ for $j \in B$, there must (see (ii)) be some $j \in S$ with $|V_j^{t/(r-1)}| \geq (1-\varepsilon)n$. $\square$

We now claim $\psi(V_i) \leq 2r^2\psi(V)/\lambda$ for all $i$. For $i \in S$, we do a little better: Lemma 5.9 gives $\psi(V_i^-) \leq r\psi(V^+)/\lambda$, and balance (of $w$) then implies $\psi(V_i) = 2\psi(V_i^-) \leq r\psi(V)/\lambda$. For $i \in B$ write $W$ for $V_i$ (just to avoid some double subscripts) and set $T = \sup\{t \ : \ |W_t| \geq \varepsilon n\}$. Then

$$\psi(W^-) = \psi(W_T) + \psi(W^- \setminus W_T) \leq \psi(W_T) + T|W^- \setminus W_T|.$$

Since $|W_T| < \varepsilon n$, Lemma 5.9 gives $\psi(W_T) \le r\psi(V^+)/\lambda$. On the other hand, $|W_t| \ge \varepsilon n$ for $t \in [0,T)$, with Lemma 5.10, implies that there is a $j \in S$ with $|V_j^{t/(r-1)}| \ge (1-\varepsilon)n$ for all such $t$. Thus

$$\begin{aligned}(1-\varepsilon)T|W^- \setminus W_T| \ &\le\ (1-\varepsilon)nT \ \le\ \textstyle\int_0^T |V_j^{t/(r-1)}|\,\mathrm{d}t \le \int_0^\infty |V_j^{t/(r-1)}|\,\mathrm{d}t \\ &=\ (r-1)\psi(V_j^+) \ \le\ r^2\psi(V^+)/\lambda.\end{aligned}$$

So, combining, we have $\psi(W) = 2\psi(W^-) \le 2r^2\psi(V)/\lambda$ (establishing the claim) and

$$\psi(V) = \textstyle\sum_i \psi(V_i) \le 2r^3\psi(V)/\lambda.$$

But since $2r^3 < \lambda$, this forces $\psi(V) = 0$ and so $w \equiv 0$. □

*Proof of Theorem 5.2.* Set $\lambda = 4r^3$, $\varepsilon = (2r\lambda)^{-1}$ and $k = 2r\varepsilon^{-r}$ (so $k$ is a little more than $r^{4r}$). We show that w.h.p. $\mathcal{H} = K_{[n]^r}(k\text{-out})$ is as in Proposition 5.8. As earlier, let $B(X,Y)$ be the event that every edge meeting $X$ meets $Y$.

Suppose first that $T$ and $U$ are fixed with $|U_i| = \lambda|T| \le \lambda\varepsilon n$. Then

$$\mathbb{P}(B(T,U)) \le \left[1 - \left(1 - \tfrac{\lambda|T|}{n}\right)^{r-1}\right]^{k|T|} \le \left(\tfrac{r\lambda|T|}{n}\right)^{k|T|}.$$

Summing over choices of $T$ and $U$ bounds the probability that $\mathcal{H}$ violates the assumptions of the proposition for some $T$ and $U$ as in (i) by

$$\begin{aligned} r\textstyle\sum_{t=1}^{\varepsilon n} \binom{n}{t}\binom{n}{\lambda t}^{r-1}\left(\frac{r\lambda t}{n}\right)^{kt} \ &\le\ r\textstyle\sum_{t=1}^{\varepsilon n}\left(\frac{en}{t}\right)^t\left(\frac{en}{\lambda t}\right)^{\lambda t(r-1)}\left(\frac{r\lambda t}{n}\right)^{kt} \\ &\le\ \textstyle\sum_{t=1}^{\varepsilon n}\left[(r\lambda t/n)^{k-r\lambda}\lambda(er)^{r\lambda}\right]^t = o(1).\end{aligned}$$

Now say $T$ and $U$ are fixed with $|T| = \varepsilon n$ and $|U_i| = (1-\varepsilon)n$. Then

$$\mathbb{P}(B(T,U)) \le (1-\varepsilon^{r-1})^{k|T|} \le \exp\left[-k|T|\varepsilon^{r-1}\right] \le \exp\left[-kn\varepsilon^r\right].$$

So summing over possibilities for $(T,U)$ bounds the probability of a violation with $T$ and $U$ as in (ii) by

$$r2^{nr}\exp\left[-kn\varepsilon^r\right] \le \exp\left[n(r - k\varepsilon^r)\right] = o(1).$$ □

### Proof of Theorem 5.4

We now turn to our proof of Theorem 5.4, for which we work with the following standard device for handling the process $\{\mathcal{H}_t\}$.

Let $\xi_S$, $S \in \binom{[n]}{r}$, be independent random variables, each uniform from $[0,1]$, and for $\lambda \in [0,1]$, let $G(\lambda)$ be the $r$-graph on $[n]$ with edge set $\mathcal{E}(\lambda) = \{S \ : \ \xi_S \leq \lambda\}$. Members of $\mathcal{E}(\lambda)$ will be called $\lambda$-*edges*. Note that with probability one, $G(0)$ is empty, $G(1)$ is complete, and the $\xi_S$'s are distinct.

Provided the $\xi_S$'s are distinct, this defines the discrete process $\{\mathcal{H}_t\}$ in the natural way, namely by adding edges $S$ in the order in which their associated $\xi_S$'s appear in $[0,1]$. We will work with the following quantities, where $\gamma = \varepsilon \log n$ for some small fixed (positive) $\varepsilon$ and $g$ is a suitably slow $\omega(1)$.

- $\Lambda = \min\{\lambda : G(\lambda) \text{ has no isolated vertices}\}$;
- $W_\lambda = \{v \in [n] \ : \ d_{G(\lambda)}(v) \leq \gamma\}$;
- $\sigma = \frac{\log n - g(n)}{\binom{n-1}{r-1}}$ and $\beta = \frac{\log n + g(n)}{\binom{n-1}{r-1}}$;
- $N = \{v : \exists e \in \mathcal{E}(\beta),\ v \in e,\ e \cap W_\sigma \neq \emptyset\}$

(so $N$ is $W_\sigma$ together with its $\mathcal{E}(\beta)$-neighbors).

**Preview.** With the above framework, our assignment is to show that $G(\Lambda)$ has a perfect matching w.h.p.. Perhaps the nicest part of this—and the point of coupling the different $G(\lambda)$'s—is that, so long as $\Lambda \in [\sigma, \beta]$, which we will show holds w.h.p., the desired assertion on $G(\Lambda)$ follows *deterministically* from a few properties ((b)-(d)) of Lemma 5.11) involving $G(\sigma)$, $G(\beta)$ *or both*; so by showing that the latter properties hold w.h.p. we avoid the need for a union bound to cover possibilities for $\Lambda$. Production of the fractional matching is then similar to (though somewhat simpler than) what happens in [41]: the relatively few vertices of $W_\Lambda$ (and some others) are covered by an (ordinary) matching, and the hypergraph induced by what's left has the expansion needed for Proposition 5.3.

**Lemma 5.11.** *With the above setup (for fixed $r$) and $Z = n(\log n)^{-1/r}$, w.h.p.*

*(a)* $\Lambda \in [\sigma, \beta]$*;*

*(b)* $\alpha(G(\sigma)) < Z$*;*

*(c) no $\beta$-edge meets $W_\sigma$ more than once and no $u \notin W_\sigma$ lies in more than one $\beta$-edge meeting $N \setminus \{u\}$;*

*(d) each $X \subseteq V \setminus W_\sigma$ of size at most $Z$ is $r$-expansive in $G(\sigma)$.*

*Proof.* For (a), note that the expected number of isolated vertices in $G(\lambda)$ is $h(\lambda) := n(1-\lambda)^{\mathbb{C}n-1r-1}$. The upper bound (i.e. $\Lambda < \beta$ w.h.p.) then follows from $h(\beta) = o(1)$, and the lower bound is given by Chebyshev's Inequality (applied to the number of isolated vertices).

For (b), we have

$$\begin{aligned} \mathbb{P}(\alpha(G(\beta)) \geq Z) &< \tbinom{n}{Z}(1-\beta)^{\binom{Z}{r}} < (en/Z)^Z \exp\left[-\beta\tbinom{Z}{r}\right] \\ &= \exp\left[Z\log(en/Z) - (1-o(1))(n/r)\log n(Z/n)^r\right] \\ &= \exp\left[Z\log(en/Z) - \Omega(n)\right] = o(1). \end{aligned}$$

The proofs of (c) and (d) are similarly routine but take a little longer. Aiming for (c), set $p = \mathbb{P}(\zeta \leq \gamma)$, where $\zeta$ is binomial with parameters $\mathbb{C}n - 2r - 1$ and $\sigma$. Since $\mu := \mathbb{E}\zeta \sim \log n$, a standard large deviation estimate (e.g., [36, Thm. 2.1]) gives

$$p < \exp[-\mu\varphi(-(\mu-\gamma)/\mu)] < n^{-1+\delta},$$

where $\varphi(x) = (x+1)\log(x+1) - x$ for $x \geq -1$ and $\delta \approx \varepsilon\log(1/\varepsilon)$.

Failure of the first assertion in (c) implies existence of $S \in K_n^{(r)}$ and (distinct) $u, v \in S$ with $S \in G(\beta)$ and $u, v \in W_\sigma$. The probability that this occurs for a given $S, u, v$ is less than $\beta p^2$ (the $p^2$ bounding the probability that each of $u, v$ lies in at most $\gamma$ edges not containing the other), so the probability that the assertion fails is less than

$$\tbinom{n}{r}r^2\beta p^2 \sim nr(\log n)p^2 = o(1).$$

If the second part of (c) fails, then we must be able to find a $u \notin W_\sigma$ as well as one of the following configurations, in which $x, y \in W_\sigma$, $S_i \in G(\beta)$, and $a, b \in [n]$ (and vertices and edges within a configuration are distinct):

(i) $x, S_1, S_2$ with $x, u \in S_1 \cap S_2$;

(ii) $x, y, S_1, S_2$ with $x, u \in S_1$, $y, u \in S_2$;

(iii) $x, a, S_1, S_2, S_3$ with $x, u \in S_1$, $x, a \in S_2$, $u, a \in S_3$;

(iv) $x, y, a, S_1, S_2, S_3$ with $x, u \in S_1$, $y, a \in S_2$, $u, a \in S_3$;

(v) $x, a, S_1, S_2, S_3$ with $x, a \in S_1$, $u, a \in S_2 \cap S_3$;

(vi) $x, a, b, S_1, S_2, S_3, S_4$ with $x, a \in S_1$, $x, b \in S_2$, $u, a \in S_3$, $u, b \in S_4$;

(vii) $x, y, a, b, S_1, S_2, S_3, S_4$ with $x, a \in S_1$, $y, b \in S_2$, $u, a \in S_3$, $u, b \in S_4$;

(viii) $x, a, b, S_1, S_2, S_3$ with $x, a, b \in S_1, u, a \in S_2, u, b \in S_3$.

Thus, with $M = \mathbb{C}n - 2r - 2$, summing probabilities for these possibilities bounds the probability of violating the second part of (c) by

$$n^2pM^2\beta^2 + n^3p^2M^2\beta^2 + n^3pM^3\beta^3 + n^4p^2M^3\beta^3 + n^3pM^3\beta^3 \\ + n^4pM^4\beta^4 + n^5p^2M^4\beta^4 + n^4pM^2\tbinom{n-3}{r-3}\beta^3 = o(1).$$

For (d) it is enough to bound (by $o(1)$) the probability that for some (nonempty) $X \subseteq V$ of size $x \le Z$ and $Y \subseteq V \setminus X$ of size $rx$,

$$\textit{there are at least } \gamma x/r\ \sigma\textit{-edges meeting both } X \textit{ and } Y. \tag{5.3}$$

For given $X, Y$ the expected number of such edges is less than

$$x \cdot rx\tbinom{n-2}{r-2}\sigma < xr^2\tfrac{Z \log n}{n-1} =: bx.$$

(The first inequality is a significant giveaway for small $x$, but we have lots of room.) So, again using [36, Thm. 2.1], we find that the probability of (5.3) is less than

$$\exp[-(\gamma x/r)\log(\gamma/(erb)] < \exp[-\Omega(\gamma x \log\log n)],$$

while the number of possibilities for $(X, Y)$ is less than

$$\binom{n}{x}\binom{n}{rx} < \exp[(r+1)x\log(n/x)] = \exp[O(x\log n)],$$

and the desired $o(1)$ bound follows. □

*Proof of Theorem 5.4.* By Lemma 5.11 it is enough to show that if (a)-(d) of the lemma hold then $G(\Lambda)$ has a perfect fractional matching; so we assume we have these conditions and proceed (working in $G(\Lambda)$).

According to (c) (and the definition of $\Lambda$), $G(\Lambda)$ admits a *matching*, $M$, covering $W_\sigma$ (each edge of which contains exactly one vertex of $W_\sigma$). Let $W$ be the set of vertices covered by $M$ (so $W$ consists of $W_\sigma$ plus some subset of $N\setminus W_\sigma$), and $H = G(\Lambda) - W$ (as usual meaning that the edges of $H$ are the edges of $G(\Lambda)$ that miss $W$). It is enough to show that $H$ has a perfect fractional matching, which will follow from Proposition 5.3 if we show

$$\text{each independent set } X \text{ of } H \text{ is } (r-1)\text{-expansive.} \tag{5.4}$$

*Proof.* Since such an $X$ is also independent in $G(\sigma)$, (b) gives $|X| \le Z$, and (d) then says $X$ is $r$-expansive in $G(\sigma)$, *a fortiori* in $G(\Lambda)$. On the other hand, since $X \cap W_\sigma = \emptyset$, (c) guarantees that the $\beta$-edges (so also the $\Lambda$-edges) meeting $X$ and *not* contained in $V(H)$ can be covered by some $U \subseteq W$ of size at most $|X|$ (namely, (c) says each $x \in X$ lies in at most one such edge). It follows that the $\Lambda$-edges meeting $X$ that *do* belong to $H$ cannot be covered by $(r-1)|X|$ vertices of $V(H) \setminus X$. □

□

# Bibliography

[1] S. Aaronson and T. Hance, *Generalizing and derandomizing Gurvits's approximation algorithm for the permanent*, Quantum Inf. Comput., 14 (2014), pp. 541–559.

[2] S. Aaronson and H. Nguyen, *Near invariance of the hypercube*, Israel Journal of Mathematics, (2016).

[3] N. Ailon and E. Liberty, *Fast dimension reduction using Rademacher series on dual BCH codes*, Discrete Comput. Geom., 42 (2009), pp. 615–630.

[4] N. Alon and E. Lubetzky, *Independent sets in tensor graph powers*, Journal of Graph Theory, 54 (2007), pp. 73–87.

[5] N. Alon and V. D. Milman, *Eigenvalues, expanders and superconcentrators*, in FOCS, Citeseer, 1984, pp. 320–322.

[6] N. Alon and J. H. Spencer, *The probabilistic method*, John Wiley & Sons, 2004.

[7] J. Balogh, T. Bohman, and D. Mubayi, *Erdős–Ko–Rado in random hypergraphs*, Combinatorics, Probability and Computing, 18 (2009), pp. 629–646.

[8] J. Balogh, B. Bollobás, and B. P. Narayanan, *Transference for the Erdős–Ko–Rado theorem*, in Forum of Mathematics, Sigma, vol. 3, Cambridge Univ Press, 2015, p. e23.

[9] J. Balogh, S. Das, M. Delcourt, H. Liu, and M. Sharifzadeh, *Intersecting families of discrete structures are typically trivial*, Journal of Combinatorial Theory, Series A, 132 (2015), pp. 224–245.

[10] T. Bohman and A. Frieze, *Hamilton cycles in 3-out*, Random Structures & Algorithms, 35 (2009), pp. 393–417.

[11] B. Bollobás, B. P. Narayanan, and A. M. Raigorodskii, *On the stability of the Erdős–Ko–Rado theorem*, Journal of Combinatorial Theory, Series A, 137 (2016), pp. 64–78.

[12] B. Bollobás and A. Thomason, *Random graphs of small order*, North-Holland Mathematics Studies, 118 (1985), pp. 47–97.

[13] A. Bonami, *Étude des coefficients de Fourier des fonctions de* $L^p(G)$, Ann. Inst. Fourier (Grenoble), 20 (1970), pp. 335–402 (1971).

[14] C. Cooper and A. Frieze, *Hamilton cycles in a class of random directed graphs*, Journal of Combinatorial Theory, Series B, 62 (1994), pp. 151–163.

[15] C. Cooper, A. Frieze, M. Molloy, and B. Reed, *Perfect matchings in random r-regular, s-uniform hypergraphs*, Combinatorics, Probability and Computing, 5 (1996), pp. 1–14.

[16] S. Das and T. Tran, *Removal and stability for erdős–Ko–Rado*, SIAM Journal on Discrete Mathematics, 30 (2016), pp. 1102–1114.

[17] W. F. de la Vega, *On the maximum cardinality of a consistent set of arcs in a random tournament*, Journal of Combinatorial Theory, Series B, 35 (1983), pp. 328–332.

[18] P. Erdős, *On the combinatorial problems which I would most like to see solved*, Combinatorica, 1 (1981), pp. 25–42.

[19] P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, The Quarterly Journal of Mathematics, 12 (1961), pp. 313–320.

[20] P. Erdős and J. Moon, *On sets of consistent arcs in a tournament*, Canad. Math. Bull, 8 (1965), pp. 269–271.

[21] P. Erdos and A. Renyi, *On random matrices*, Magyar Tud. Akad. Mat. Kutató Int. Közl, 8 (1964), p. 1964.

[22] P. Erdős and A. Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta Mathematica Hungarica, 17 (1966), pp. 359–368.

[23] T. I. Fenner and A. M. Frieze, *Hamiltonian cycles in random regular graphs*, Journal of Combinatorial Theory, Series B, 37 (1984), pp. 103–112.

[24] A. Ferber, G. Kronenberg, F. Mousset, and C. Shikhelman, *Packing a randomly edge-colored random graph with rainbow $k$-outs*, arXiv preprint arXiv:1410.1803, (2014).

[25] A. Frieze and T. Johansson, *On random $k$-out sub-graphs of large graphs*, arXiv preprint arXiv:1405.2129, (2014).

[26] A. Frieze and G. B. Sorkin, *Efficient algorithms for three-dimensional axial and planar random assignment problems*, Random Structures & Algorithms, 46 (2015), pp. 160–196.

[27] A. M. Frieze, *Maximum matchings in a class of random graphs*, Journal of Combinatorial Theory, Series B, 40 (1986), pp. 196–212.

[28] A. M. Frieze and T. Luczak, *Hamiltonian cycles in a class of random graphs: one step further*, in Proceedings of Random Graphs, vol. 87, 1987, pp. 53–59.

[29] D. G. Glynn, *The permanent of a square matrix*, European J. Combin., 31 (2010), pp. 1887–1891.

[30] L. Gurvits, *On the complexity of mixed discriminants and related problems*, in Mathematical Foundations of Computer Science 2005, Springer, 2005, pp. 447–458.

[31] L. Gurvits and A. Samorodnitsky, *Bounds on the permanent and some applications*, in 2014 IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2014, pp. 90–99.

[32] S. Haber and M. Krivelevich, *On fractional k-factors of random graphs*, Random Structures & Algorithms, 30 (2007), pp. 441–463.

[33] A. Hamm and J. Kahn, *On Erdős–Ko–Rado for random hypergraphs I*, arXiv preprint arXiv:1412.5085, (2014).

[34] ——, *On Erdős–Ko–Rado for random hypergraphs II*, arXiv preprint arXiv:1406.5793, (2014).

[35] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American statistical association, 58 (1963), pp. 13–30.

[36] S. Janson, T. Luczak, and A. Rucinski, *Random graphs*, vol. 45, John Wiley & Sons, 2011.

[37] A. Johansson, J. Kahn, and V. Vu, *Factors in random graphs*, Random Structures & Algorithms, 33 (2008), pp. 1–28.

[38] R. M. Karp, *Reducibility among combinatorial problems*, in Complexity of computer computations, Springer, 1972, pp. 85–103.

[39] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, (1996).

[40] H. König, C. Schütt, and N. Tomczak-Jaegermann, *Projection constants of symmetric spaces and variants of Khintchine's inequality*, J. Reine Angew. Math., 511 (1999), pp. 1–42.

[41] M. Krivelevich, *Perfect fractional matchings in random hypergraphs*, Random Structures & Algorithms, 9 (1996), pp. 317–334.

[42] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Classics in Mathematics, Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[43] T.-Y. Lee, H.-T. Yau, et al., *Logarithmic sobolev inequality for some models of random walks*, The Annals of Probability, 26 (1998), pp. 1855–1873.

[44] T. Leighton and A. Moitra, *On entropy and extensions of posets*, 2011.

[45] F. J. MacWilliams and N. J. A. Sloane, *The theory of error correcting codes*, Elsevier, 1977.

[46] H. Nguyen, *On matrices of large permanent.* Private communication, 2016.

[47] R. O'Donnell, *Analysis of boolean functions*, Cambridge University Press, 2014.

[48] V. Rödl and M. Schacht, *Extremal results in random graphs*, in Erdős Centennial, Springer, 2013, pp. 535–583.

[49] E. R. Scheinerman and D. H. Ullman, *Fractional graph theory: a rational approach to the theory of graphs*, Courier Corporation, 2011.

[50] A. Schrijver, *Theory of linear and integer programming*, John Wiley & Sons, 1998.

[51] W. Schudy and M. Sviridenko, *Concentration and moment inequalities for polynomials of independent random variables*, in Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, 2012, pp. 437–446.

[52] E. Szemerédi, *Regular partitions of graphs*, tech. rep., DTIC Document, 1975.

[53] L. G. Valiant, *The complexity of computing the permanent*, Theoret. Comput. Sci., 8 (1979), pp. 189–201.

[54] D. W. Walkup, *Matchings in random regular bipartite digraphs*, Discrete Mathematics, 31 (1980), pp. 59–64.

[55] N. C. Wormald, *Models of random regular graphs*, London Mathematical Society Lecture Note Series, (1999), pp. 239–298.