A FEW RESULTS REGARDING THRESHOLDS

By

KEITH FRANKSTON

A dissertation submitted to the

School of Graduate Studies

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

October, 2020

ABSTRACT OF THE DISSERTATION

A Few Results Regarding Thresholds

By KEITH FRANKSTON

Dissertation Director: Jeff Kahn

This thesis consists of four parts, each regarding a topic from extremal combinatorics. While only Chapters 3 and 4 are directly related, each in some way concerns the concept of thresholds, whether providing a new sharp threshold result for regular properties (in the case of Chapter 2), proving specific graph theoretic thresholds (in the case of Chapter 5), or showing how different thresholds are related (as in Chapters 3 and 4).

In Chapter 2, we answer a question of Cameron, Frankl, and Kantor from 1989, extending a result of Ellis and Narayanan. They verified a conjecture of Frankl, that any 3-wise intersecting family of subsets of $\{1, 2, \ldots, n\}$ admitting a transitive automorphism group has cardinality $o(2^n)$. However, a construction of Frankl demonstrates that the same conclusion need not hold under the weaker constraint of being regular. We show that the restriction of admitting a transitive automorphism group may be relaxed significantly: we prove that any 3-wise intersecting family of subsets of $\{1, 2, \ldots, n\}$ that is regular *and increasing* has cardinality $o(2^n)$.

In Chapter 3, we prove a conjecture of Talagrand, itself a fractional version of the "expectation-threshold" conjecture of Kahn and Kalai. We show that for any increasing family $\mathcal{F}$ on a finite set X, we have $p_c(\mathcal{F}) = O(q_f(\mathcal{F}) \ln \ell(\mathcal{F}))$, where $p_c(\mathcal{F})$ and $q_f(\mathcal{F})$ are the threshold and "fractional expectation-threshold" of $\mathcal{F}$, and $\ell(\mathcal{F})$ is the maximum size of a minimal member of $\mathcal{F}$. This easily implies several heretofore difficult results and

conjectures in probabilistic combinatorics, including thresholds for perfect hypergraph matchings (Johansson–Kahn–Vu), bounded degree spanning trees (Montgomery), and bounded degree graphs (new). We also resolve (and vastly extend) the "axial" version of the random multi-dimensional assignment problem (earlier considered by Martin–Mézard–Rivoire and Frieze–Sorkin). Our approach builds on a breakthrough of Alweiss, Lovett, Wu and Zhang on the Erdős–Rado "Sunflower Conjecture."

In Chapter 4, we address a special case of a conjecture of Talagrand relating the "expectation" and "fractional-expectation" thresholds of an increasing family $\mathcal{F}$ of a finite set X. The full conjecture implies the equivalence of the so-called "fractional expectation-threshold" conjecture shown in Chapter 3 to the "expectation-threshold" conjecture of Kahn and Kalai. The conjecture discussed in this chapter states there is a fixed J such that if $p \in [0,1]$ admits $\lambda : X \to [0,1]$ with

$$\sum_{S \subseteq F} \lambda_S \geq 1 \quad \forall F \in \mathcal{F}$$

and

$$\sum_S \lambda_S p^{|S|} \leq 1/2$$

(a.k.a. $\mathcal{F}$ is *weakly p-small*), then p/J admits such a λ taking values in $\{0,1\}$ ($\mathcal{F}$ is *(p/J)-small*). Talagrand showed this when λ is supported on singletons and suggested, as a more challenging test case, proving it when λ is supported on pairs. This chapter presents such a proof.

Expanding on work on the rigidity of random graph structures going back to Erdős and Rényi, Chapter 5 introduces a new notion of "local" rigidity. Say H is *locally t-rigid* if all its induced subgraphs on t vertices are rigid. Then for what $t = t(n,p)$ is $G_{n,p}$ is locally t-rigid? To answer this question, we produce machinery which allows for more careful analysis of the probability of appearance of non-trivial automorphisms based on their "type." In particular, for any cycle type, λ, we give a threshold $t(\lambda)$ for the appearance of automorphisms of that type such that, if $m(\lambda)$ is the size of the largest induced subgraph of $G_{n,1/2}$ whose automorphism group has a permutation of type λ, then with high probability $m < t + \sqrt{5n \log n}$ for all λ (and with high probability $|t - m| < \sqrt{5n \log n}$ for any fixed choice of λ).

Preface

The result given in Chapter 2 is joint work with Jeff Kahn and Bhargav Narayanan and has previously been published as [20]. The results in Chapter 3 are joint work with Jeff Kahn, Bhargav Narayanan, and Jinyoung Park and are contained in [21] which is under review for publication. The result given in Chapter 4 is joint work with Jeff Kahn and Jinyoung Park.

Acknowledgements

No one can do it alone and, fortunately, I've had the help of more people than I can possibly list here.

Thank you to all my fellow graduate students. For all the support, whether I was asking for help or needed a place to vent. For the games, the random conversations, for providing a community for the last 6 years that I know I will miss.

I'd especially like thank the following people: Matthew Russell and Nathan Fox, for keeping me sane by driving me insane. Jinyoung Park, for being my co-author in this final year; I look forward to reading whatever you work on next. Yonah Biers-Ariel and Justin Semonsen, for providing an office of help and distraction (in possibly disproportionate measures) for the past 4 years, and for being wonderful friends, hopefully for many more years to come. Pat Devlin, for always being an infuriating source of amazing help, and without whom this document might still be in four distinct pieces.

Érik de Amorim and Abigail Raz, for being the best roommates I could have asked for, and becoming two of my closest friends. Érik, I will miss never knowing what topic a conversation will turn to next, from every branch of math under the sun, to philosophy, into dying laughing at the dumbest thing. Keep being insane, but try to be less Brazilian. Abigail, for being a source of support in every aspect of life and someone I knew I could turn to for anything. For being someone I could share with, whether it was a love of sports or fiction, thoughts on the world around us, or even just what happened that day. Thank you.

My mom, who made every step of this journey possible. For being there for me whenever I needed you and being patient with me when I was too stubborn to admit I did. My dad, for being a treasure trove of knowledge that I never got tired of exploring. For each drive to someplace random that turned into a deep discussion on something random. To both of you, for encouraging my curiosity at every turn, for changing "I don't know" into "let's figure it out."

All of the Frankstons and Warshaws, for your support along the way.

All the amazing teachers who helped me get to this point. While there have been too many to count, I'd particularly like to thank: Meredith Haas, for seeing a kid who couldn't pay attention in calculus and handing him an analysis book. Ira Gessel and Olivier Bernardi for introducing me to research, and all the frustrations and wonders therein. Ervin Győri for giving me my first taste of extremal combinatorics and getting me hooked. To Ann Olga Koloski-Ostrow, for teaching me to love the classical world and ensuring I got a BA in the liberal arts, not just math.

Thank you to Mike Saks, Bhargav Narayanan, and Gil Kalai for serving on my committee, and for all your patience in bearing with a longer process than you signed up for.

Finally, thank you to my advisor, Jeff Kahn, for everything. I really can't thank you enough for all the teaching, advice, and sharing of knowledge about math and being a mathematician, coupled with patience that went beyond the call of duty.

Dedication

To Michael Hanau,

for guiding me through math when it stopped making sense.

Table of Contents

Chapter 1

Introduction

This thesis consists of four papers, each of which is given its own chapter. Here we give a brief high-level overview of each topic, leaving details to their individual chapters. At the end of this chapter we collect some common notation and conventions that are used throughout.

Chapter 2 is primarily concerned with intersecting families: for an integer $r \geq 2$, a family of sets $\mathcal{A}$ is said to be *r-wise intersecting* if any r of the sets in $\mathcal{A}$ have nonempty intersection. There is by now a large body of work studying the extremal properties of families of sets under various intersection requirements; we refer the reader to the surveys [9, 45] for an overview. A common theme that arises when studying the extremal properties of intersecting families is that the extremal constructions are often highly asymmetric; indeed, this is the case with many of the classical results in the field, such as the Erdős–Ko–Rado theorem [13] and the Ahlswede–Khachatrian theorem [1] to name just two. It is therefore natural to ask what, if anything, changes when one considers intersecting families subject to requirements of "symmetry", and this is the line of questioning that we pursue here.

Specifically, we expand upon work of Ellis and Narayanan to answer a question of Cameron, Frankl, and Kantor from 1989, showing that any 3-wise intersecting family of subsets of $\{1, 2, \dots, n\}$ that is regular and increasing has cardinality $o(2^n)$. The main technical tool that we develop to prove this is a lemma demonstrating the existence of threshold-type behaviour for increasing regular families.

Chapters 3 and 4 both address the so-called "expectation-threshold" and fractional "expectation-threshold" of an increasing family. Thresholds have been a—maybe *the*—central concern of the study of random discrete structures (random graphs and hypergraphs, for example) since its initiation by Erdős and Rényi [15], with much work around identifying thresholds for specific properties (see [4, 30]), though it was not observed until [5] that *every* increasing $\mathcal{F}$ admits a threshold (in the Erdős–Rényi sense; see below). See also [24] for developments, since [23], on the very interesting question of *sharpness* of thresholds.

The focus of Chapter 3 is the proof of a conjecture of Talagrand [53] that is a fractional version of the "expectation-threshold" conjecture of Kalai and Kahn [33]. For an increasing family $\mathcal{F}$ on a finite set X, we write (with definitions below) $p_c(\mathcal{F})$, $q_f(\mathcal{F})$, and $\ell(\mathcal{F})$ for the threshold, fractional expectation-threshold, and size of a largest minimal element of $\mathcal{F}$. In this language, our main result is the following.

Theorem 1.0.1. *There is a universal K such that for every finite X and increasing $\mathcal{F} \subseteq 2^X$,*

$$p_c(\mathcal{F}) \leq K q_f(\mathcal{F}) \ln \ell(\mathcal{F}).$$

As we will see, $q_f(\mathcal{F})$ is a more or less trivial lower bound on $p_c(\mathcal{F})$, and Theorem 1.0.1 says this bound is never far from the truth. (Apart from the constant K, the upper bound is tight in many of the most interesting cases; see 3.7.)

The second main result in Chapter 3 is Theorem 3.1.6, which was motivated by work of Frieze and Sorkin [26] on the "random multi-dimensional assignment problem." The statement is postponed until the chapter itself.

In Chapter 4, we present initial work towards another conjecture of Talagrand which, together with Theorem 1.0.1, would give the Kahn-Kalai "expectation-threshold" conjecture. Specifically, Talagrand proposed that $q(\mathcal{F}) = O(q_f(\mathcal{F}))$ for any increasing family $\mathcal{F}$ of a finite set X. We approach this statement from a slightly different point of view and verify it for a special case.

We say a non-negative weight assignment $\lambda : 2^X \to \mathbb{R}^+$ certifies $\mathcal{F}$ is weakly p-small if for every $F \in \mathcal{F}$ we have $\sum_{S \subseteq F} \lambda_S \geq 1$ and $\sum_S \lambda_s p^{|S|} \leq 1/2$. If, additionally, λ is the indicator function for some collection $\mathcal{G}$, then we say the collection $\mathcal{G}$ certifies $\mathcal{F}$ is

p-small. $q(\mathcal{F})$ (or $q_f(\mathcal{F})$) is the largest p such that there exists a collection $\mathcal{G}$ (resp. λ) certifying $\mathcal{F}$ is p-small (resp. weakly p-small).

Thus, the statement $q(\mathcal{F}) = O(q_f(\mathcal{F}))$ is equivalent to showing that there is some universal $J > 0$ such that given any λ and p, if λ certifies $\mathcal{F}$ is weakly p-small, then there is a $\mathcal{G}$ which certifies $\mathcal{F}$ is p/J-small. Talagrand showed the statement holds when λ is supported on singletons and suggested pairs as a further test case. We verify that the statement holds when restricted to $\mathrm{supp}(\lambda) \subseteq \binom{X}{2}$.

Finally, in Chapter 5 we present a method for analyzing the appearance of non-trivial automorphisms in random graphs. We say a graph H is *rigid* (sometimes referred to as *asymmetric*) if its automorphism group, denoted $\mathrm{Aut}(H)$, is trivial (i.e. consists only of the identity map). The *Erdős-Rényi random graph*, denoted $G_{n,p}$, is the random graph on n vertices where edges are present independently, each with probability p. Early in their work on these random structures, Erdős and Rényi [16] showed the following result.

Theorem 1.0.2 (Erdős-Rényi '63)**.** *Asymptotically almost all labelled graphs on n vertices are rigid, i.e.*

$$\Pr\left[\left|\mathrm{Aut}(G_{n,1/2})\right| = 1\right] = 1 - o(1).$$

This result has since been extended to $G_{n,p}$ as follows[1].

Theorem 1.0.3. *For $1/2 \geq p = \frac{\ln n + \omega(1)}{n}$,*

$$\Pr\left[|\mathrm{Aut}(G_{n,p})| = 1\right] = 1 - o(1).$$

(Note that the lower bound on p is best possible as isolated vertices begin to appear in $G_{n,p}$ when p is below this threshold.)

An analogous result for the random graph with m edges, $G_{n,m}$, was originally announced by Erdős and Rényi in [16], though we are unaware of its subsequent publication. Wright's work counting unlabelled m-edge graphs [57, 58] produced the following stronger result.

[1]As the automorphism group of G is identical to that of its complement, $\overline{G}$, we will usually assume $p \leq 1/2$.

Theorem 1.0.4 (Wright '74)**.** *For* $\binom{n}{2}/2 \geq m = \frac{n}{2}(\ln n + \omega(1))$, *asymptotically almost all* unlabelled *m-edge graphs on n vertices are rigid.*

More recently, the random d-regular graph, $G_{n,d}$, was shown to be rigid by Kim, Vu, and Sudakov [36].

Theorem 1.0.5. *For* $3 \leq d \leq n-4$,

$$\Pr[|\mathrm{Aut}(G_{n,d})| = 1] = 1 - o(1).$$

See also [39, 6] for other recent work.

The work in this chapter was motivated by a desire to extend these lines of inquiry to a notion of "local" rigidity. Say H is *locally t-rigid* if all its induced subgraphs on t vertices are rigid. The question is then, for what $t = t(n,p)$ is $G_{n,p}$ is locally t-rigid w.h.p.[2]?

We expand upon techniques in [36, 46] to show the following essentially optimal result in the case $p = 1/2$.

Lemma 1.0.6. *$G_{n,1/2}$ is locally $(n/2 + \sqrt{5n\log n})$-rigid w.h.p.*

This follows from a much stronger general result on what "types" of automorphisms exist among all induced subgraphs of a given size t (stated formally as Theorem 5.3.1). We also give a similar (but weaker) result for general p (which we expect can be strengthened to produce a similar result to our main theorem for some restricted ranges of p).

Notation and Conventions. Given a finite set X, write 2^X for the power set of X and, for $p \in [0,1]$, μ_p for the product measure on 2^X given by $\mu_p(S) = p^{|S|}(1-p)^{|X \setminus S|}$. An $\mathcal{F} \subseteq 2^X$ is *increasing* if $B \supseteq A \in \mathcal{F} \Rightarrow B \in \mathcal{F}$. For $\mathcal{G} \subseteq 2^X$ we use $\langle \mathcal{G} \rangle$ for the increasing family generated by $\mathcal{G}$, namely $\{B \subseteq X : \exists A \in \mathcal{G}, B \supseteq A\}$.

If $\mathcal{F}$ is increasing (and $\mathcal{F} \neq 2^X, \emptyset$), then $\mu_p(\mathcal{F})(:= \sum\{\mu_p(S) : S \in \mathcal{F}\})$ is strictly increasing in p, and we define the *threshold*, $p_c(\mathcal{F})$, to be the unique p for which $\mu_p(\mathcal{F}) = 1/2$. (This is finer than the original Erdős–Rényi notion, according to which $p^* = p^*(n)$

[2] *With high probability*, i.e. with probability tending to 1 as $n \to \infty$.

is $\boldsymbol{a}$ threshold for $\mathcal{F} = \mathcal{F}_n$ if $\mu_p(\mathcal{F}) \to 0$ when $p \ll p^*$ and $\mu_p(\mathcal{F}) \to 1$ when $p \gg p^*$. That $p_c(\mathcal{F})$ *is* always an Erdős–Rényi threshold follows from [5].)

Throughout we use ln for the natural logarithm and reserve log for $\log_2$.

Chapter 2

On regular 3-wise intersecting families

Joint work with Jeff Kahn and Bhargav Narayanan.

2.1 Introduction

For a positive integer $n \in \mathbb{N}$, let us write $[n]$ for the set $\{1, 2, \dots, n\}$, and $2^{[n]}$ for the power-set of $[n]$. We say that a family $\mathcal{A} \subseteq 2^{[n]}$ is *symmetric* if the automorphism group of $\mathcal{A}$ is transitive on $[n]$, *regular* if every element of $[n]$ belongs to the same number of sets in $\mathcal{A}$, and *increasing* if $\mathcal{A}$ is closed under taking supersets. We stress that the families we shall study in this chapter will be non-uniform, i.e., their members need not all be of the same size; for related work on uniform intersecting families, see the paper of Ellis, Kalai and Narayanan [11] addressing the symmetric case, and the results of Ihringer and Kupavskii [29] addressing the regular case.

The family $\{x \subseteq [n] : |x| > n/2\}$ is a symmetric 2-wise intersecting family containing a positive fraction of all the sets in $2^{[n]}$. Ellis and Narayanan [12], verifying a conjecture of Frankl [19], proved that symmetric r-wise intersecting families must be significantly smaller when $r \geq 3$; more precisely, they showed the following.

Theorem 2.1.1. *If $\mathcal{A} \subseteq 2^{[n]}$ is a symmetric 3-wise intersecting family, then $|\mathcal{A}| = o(2^n)$.*

On the other hand, a projective-geometric construction of Frankl [19] shows that there exist regular 3-wise intersecting subfamilies of $2^{[n]}$ containing a positive fraction of all the sets in $2^{[n]}$, so the conclusion of Theorem 2.1.1 no longer holds when one considers regular families instead of symmetric ones.

Here, we investigate the middle ground between symmetric and regular families following Cameron, Frankl and Kantor [7]: they proved that if $\mathcal{A} \subseteq 2^{[n]}$ is a 4-wise intersecting family that is both regular *and increasing*, then $|\mathcal{A}| = o(2^n)$, and asked what one can say about regular 3-wise intersecting families. Our main result answers this question by showing that the conclusion of Theorem 2.1.1 does hold for regular families, provided again that they are increasing.

Theorem 2.1.2. *If $\mathcal{A} \subseteq 2^{[n]}$ is a 3-wise intersecting family that is both regular and increasing, then $|\mathcal{A}| = o(2^n)$.*

Of course, Theorem 2.1.2 implies Theorem 2.1.1; to see this, note that if $\mathcal{A} \subseteq 2^{[n]}$ is a symmetric 3-wise intersecting family, then $\{y : x \subseteq y \text{ for some } x \in \mathcal{A}\}$ is a 3-wise intersecting family containing $\mathcal{A}$ that is both regular and increasing.

It is worth highlighting that in both [12] and the present work, Fourier analysis plays a crucial, if invisible, role: indeed, the proof of Theorem 2.1.1 hinges on a sharp threshold result of Friedgut and Kalai [25], while here, to prove the stronger assertion of Theorem 2.1.2, we in turn rely on the somewhat heavier machinery of Friedgut's junta theorem [22]. The main new technical tool that we develop to prove Theorem 2.1.2 is a lemma demonstrating the existence of threshold-type behaviour under some rather mild conditions; this result (see Lemma 2.3.1) might be of some independent interest.

This chapter is organised as follows. We collect the various tools we require in Section 2.2. The proof of Theorem 2.1.2 follows in Section 2.3. We conclude in Section 2.4 with a brief discussion of open problems.

2.2 Preliminaries

In this section, we briefly describe the notions and tools we shall require for our arguments.

We abbreviate $\mu_{\frac{1}{2}}$ by μ, and note that this is just the normalised counting measure.

For a family $\mathcal{A} \subseteq 2^{[n]}$, we write $\mathcal{I}(\mathcal{A}) = \{x \cap y : x, y \in \mathcal{A}\}$ for the family of all possible intersections of pairs of sets from $\mathcal{A}$. We require the following proposition from [12]; we include a short proof for completeness.

Proposition 2.2.1. *For any $\mathcal{A} \subseteq 2^{[n]}$, if $\mu_p(\mathcal{A}) \geq \delta$, then $\mu_{p^2}(\mathcal{I}(\mathcal{A})) \geq \delta^2$.*

Proof. Let x and y be two random elements of $2^{[n]}$ drawn independently according to the distribution μ_p. It is then clear that $x \cap y$ has distribution μ_{p^2}, so we have

$$\mu_{p^2}(\mathcal{I}(\mathcal{A})) = \mathbb{P}(x \cap y \in \mathcal{I}(\mathcal{A})) \geq \mathbb{P}(x, y \in \mathcal{A}) = \mu_p(\mathcal{A})^2,$$

proving the proposition. $\square$

We shall require the notions of influences and juntas. First, given $\mathcal{A} \subseteq 2^{[n]}$, we say that an element $i \in [n]$ is *pivotal* for $\mathcal{A}$ at $x \in 2^{[n]}$ if exactly one of x and $x \triangle \{i\}$ lies in $\mathcal{A}$, and for $0 \leq p \leq 1$, we define the *total influence* $I_p(\mathcal{A})$ of $\mathcal{A}$ at p to be the expected number of pivotal elements for $\mathcal{A}$ at a random set $x \in 2^{[n]}$ drawn according to the distribution μ_p. The following fundamental formula was originally observed independently by Margulis [42] and Russo [48].

Proposition 2.2.2. *If $\mathcal{A} \subseteq 2^{[n]}$ is increasing, then*

$$\frac{d}{dp}\mu_p(\mathcal{A}) = I_p(\mathcal{A})$$

for all $0 < p < 1$. $\square$

Next, for $J \subseteq [n]$, a family $\mathcal{A} \subseteq 2^{[n]}$ is said to be a *J-junta* if the membership of a set in $\mathcal{A}$ is determined by its intersection with J, or in other words, if $x \in \mathcal{A}$ and $x \cap J = y \cap J$ for some $y \in 2^{[n]}$, then this implies that $y \in \mathcal{A}$. The following result due to Friedgut [22] will be our main tool.

Theorem 2.2.3. *For each $C > 0$ and $0 < \varepsilon < 1$, there exists $K > 0$ such that the following holds for all $\varepsilon \leq p \leq 1-\varepsilon$ and $n \in \mathbb{N}$. For any $\mathcal{A} \subseteq 2^{[n]}$ with $I_p(\mathcal{A}) \leq C$, there exists a set $J \subseteq [n]$ with $|J| \leq K$ and a J-junta $\mathcal{B} \subseteq 2^{[n]}$ such that $\mu_p(\mathcal{A} \triangle \mathcal{B}) \leq \varepsilon$.* $\square$

Finally, we say that two families $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$ are *cross-intersecting* if $x \cap y \neq \emptyset$ for all $x \in \mathcal{A}$ and $y \in \mathcal{B}$. We need the following simple fact also used in [12].

Proposition 2.2.4. *If $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$ are cross-intersecting, then*

$$\mu_p(\mathcal{A}) + \mu_{1-p}(\mathcal{B}) \leq 1$$

for any $0 \leq p \leq 1$.

Proof. Since $\mathcal{A}$ and $\mathcal{B}$ are cross-intersecting, it is clear that $\mathcal{A} \subseteq 2^{[n]} \setminus \tilde{\mathcal{B}}$, where $\tilde{\mathcal{B}} = \{[n] \setminus x : x \in \mathcal{B}\}$. Therefore,

$$\mu_p(\mathcal{A}) \leq \mu_p(2^{[n]} \setminus \tilde{\mathcal{B}}) = 1 - \mu_p(\tilde{\mathcal{B}}) = 1 - \mu_{1-p}(\mathcal{B}). \qquad \square$$

2.3 Proof of the main result

Our proof of Theorem 2.1.2 borrows ideas from both [7] and [12]. Before turning to the proof, let us briefly explain what is lost, relative to the argument in [12], by dropping the requirement of symmetry: for a family $\mathcal{A} \subseteq 2^{[n]}$ that is both symmetric and increasing, a result of Talagrand [49] guarantees that the total influence $I_p(\mathcal{A})$ is large whenever $\mu_p(\mathcal{A})$ is bounded away from both 0 and 1, which ensures, by Proposition 2.2.2, that the derivative of $\mu_p(\mathcal{A})$ with respect to p is also large under these circumstances; this is no longer the case when one considers regular families as opposed to symmetric ones. A replacement for this fact, the main new ingredient here, is the following lemma asserting a somewhat weaker version of this threshold behaviour under milder conditions.

Lemma 2.3.1. *For any $\varepsilon, \delta > 0$, the following holds for all sufficiently large $n \in \mathbb{N}$. If $\mathcal{A} \subseteq 2^{[n]}$ is both regular and increasing, and $\mu(\mathcal{A}) \geq \delta$, then $\mu_{\frac{1}{2}+\varepsilon}(\mathcal{A}) \geq 1 - \varepsilon$.*

Proof. In what follows, we fix $\eta = \varepsilon\delta/(2+\delta)$ and additionally suppose that n is large enough for all our estimates to hold; in particular, constants suppressed by the asymptotic notation may depend on ε and δ but, of course, not on n.

Since $\mu(\mathcal{A}) = \mu_{\frac{1}{2}}(\mathcal{A}) \geq \delta$ and $\mu_{\frac{1}{2}+\varepsilon}(\mathcal{A}) \leq 1$, it follows from Proposition 2.2.2 that there exists $q \in [1/2, 1/2+\varepsilon]$ such that $I_q(\mathcal{A}) \leq 1/\varepsilon$. Theorem 2.2.3 now implies that there exists $J \subseteq [n]$ with $|J| = K$ and a J-junta $\mathcal{B} \subseteq 2^{[n]}$ such that $\mu_q(\mathcal{A} \triangle \mathcal{B}) \leq \eta$, where K is a constant depending only on ε and δ.

Let us set up some notation before we proceed. For $i \in [n]$, let $\mathcal{A}_i$ denote the family of those sets in $\mathcal{A}$ containing i, and for $y \subseteq J$, define the *fibre* $\mathcal{A}(y)$ of $\mathcal{A}$ over y by

$$\mathcal{A}(y) = \{x \setminus y : x \in \mathcal{A} \text{ and } x \cap J = y\}.$$

Also, let $\mathcal{B}'$ be the family on J determining $\mathcal{B}$, i.e., $x \in 2^{[n]}$ belongs to $\mathcal{B}$ if and only if $x \cap J$ belongs to $\mathcal{B}'$.

We first note that, as $\mathcal{A}$ is regular, the sets $\mathcal{A}_i$ are all roughly half as large as $\mathcal{A}$; a similar observation is used in [7].

Claim 2.3.2. *For each $i \in [n]$, we have $\mu(\mathcal{A})/2 \le \mu(\mathcal{A}_i) \le \mu(\mathcal{A})/2 + O(1/\sqrt{n})$.*

Proof. The first inequality follows from the fact that $\mathcal{A}$ is increasing, so it suffices to verify the second. Let Z be a set drawn uniformly at random from $\mathcal{A}$, and for $i \in [n]$, let Z_i be the indicator of the event $\{i \in Z\}$. We shall rely on the properties of the binary entropy $H(\cdot)$ of a random variable; see [43] for the basic notions. It follows from the sub-additivity of entropy that $H(Z) \le \sum_{i=1}^{n} H(Z_i)$. Clearly, we have

$$H(Z) = \log|\mathcal{A}| = n + \log(\mu(\mathcal{A})) \ge n + \log\delta,$$

and, writing ϑ for the common value of $|\mathcal{A}_i|/|\mathcal{A}|$ for all $i \in [n]$, we also have

$$H(Z_i) = -\vartheta\log\vartheta - (1-\vartheta)\log(1-\vartheta)$$

for each $i \in [n]$. It is now easy to verify from the sub-additivity estimate above that $\vartheta = 1/2 + O(1/\sqrt{n})$, proving the claim. □

Next, we observe that all the fibres of $\mathcal{A}$ have roughly the same size as well. Let us write σ_p for the p-biased measure on the power set of J and τ_p for the p-biased measure on the power set of $[n] \setminus J$, so that $\mu_p = \sigma_p \times \tau_p$, and again, we abbreviate $\sigma_{\frac{1}{2}}$ and $\tau_{\frac{1}{2}}$ by σ and τ respectively.

Claim 2.3.3. *For all $y \subseteq J$, we have $\tau(\mathcal{A}(y)) = \mu(\mathcal{A}) + o(1)$.*

Proof. We note that

$$\mu(\mathcal{A}) = \sum_{y \subseteq J} \sigma(y)\tau(\mathcal{A}(y)),$$

and that $\sigma(y) = 2^{-K}$ for all $y \subseteq J$. For any $i \in y \subseteq J$, we have $\mathcal{A}(y \setminus \{i\}) \subseteq \mathcal{A}(y)$ because $\mathcal{A}$ is increasing, so

$$\tau(\mathcal{A}(y)) \ge \tau(\mathcal{A}(y \setminus \{i\})).$$

Since $|J| = K = O(1)$, to prove the claim, it clearly suffices to show that for any $i \in y \subseteq J$, we have

$$\tau(\mathcal{A}(y)) \le \tau(\mathcal{A}(y \setminus \{i\})) + o(1);$$

indeed, this would imply that

$$\tau(\mathcal{A}(y)) = \tau(\mathcal{A}(\emptyset)) + o(|y|) = \tau(\mathcal{A}(\emptyset)) + o(1)$$

for each $y \subseteq J$, and the claim would follow.

Fix $i \in J$, and note that

$$\mu(\mathcal{A}_i) = \sum_{i \in y \subseteq J} \sigma(y)\tau(\mathcal{A}(y)),$$

so we have

$$\mu(\mathcal{A}_i) - \mu(\mathcal{A})/2 = 2^{-K-1} \sum_{i \in y \subseteq J} \left(\tau(\mathcal{A}(y)) - \tau(\mathcal{A}(y \setminus \{i\}))\right).$$

We know from Claim 2.3.2 that $\mu(\mathcal{A}_i) - \mu(\mathcal{A})/2 = O(1/\sqrt{n})$, so for each $y \subseteq J$ containing i, we have

$$\tau(\mathcal{A}(y)) - \tau(\mathcal{A}(y \setminus \{i\})) = O(1/\sqrt{n}),$$

as required. □

We may now complete the proof of the lemma. Recall that we earlier fixed $q \in [1/2, 1/2 + \varepsilon]$ and a J-junta $\mathcal{B} \subseteq 2^{[n]}$ such that $\mu_q(\mathcal{A} \triangle \mathcal{B}) \leq \eta$, and defined $\mathcal{B}'$ to be the family on J determining $\mathcal{B}$.

First, note that

$$\mu_q(\mathcal{A} \triangle \mathcal{B}) = \sum_{y \in \mathcal{B}'} \sigma_q(y)\left(1 - \tau_q(\mathcal{A}(y))\right) + \sum_{y \notin \mathcal{B}'} \sigma_q(y)\left(\tau_q(\mathcal{A}(y))\right).$$

Since $\mathcal{A}$ is increasing, we see from Claim 2.3.3 that $\tau_q(\mathcal{A}(y)) \geq \tau(\mathcal{A}(y)) \geq \delta/2$ for all $y \subseteq J$. Therefore, since $\mu_q(\mathcal{A} \triangle \mathcal{B}) \leq \eta$, we see that

$$\sum_{y \notin \mathcal{B}'} \sigma_q(y) \leq 2\eta/\delta,$$

which implies that

$$\mu_q(\mathcal{B}) = \sum_{y \in \mathcal{B}'} \sigma_q(y) \geq 1 - 2\eta/\delta.$$

Again, since $\mu_q(\mathcal{A} \triangle \mathcal{B}) \leq \eta$ and $\eta = \varepsilon\delta/(2+\delta)$, it follows that

$$\mu_{\frac{1}{2}+\varepsilon}(\mathcal{A}) \geq \mu_q(\mathcal{A}) \geq 1 - 2\eta/\delta - \eta = 1 - \varepsilon,$$

proving the lemma. □

Armed with Lemma 2.3.1, we may now prove Theorem 2.1.2; the proof below by and large follows the argument in [12], with Lemma 2.3.1 serving as a substitute for the sharp threshold result used there.

Proof of Theorem 2.1.2. We need to show for any fixed $\delta > 0$, that for all but finitely many $n \in \mathbb{N}$, if $\mathcal{A} \subseteq 2^{[n]}$ is a 3-wise intersecting family that is both regular and increasing, then $\mu(\mathcal{A}) < \delta$; hence, suppose for a contradiction that n is sufficiently large and that $\mathcal{A} \subseteq 2^{[n]}$ is a family as just described with $\mu(\mathcal{A}) \geq \delta$.

Let us fix $\varepsilon = \min\{1/4, \delta^2/2\}$. First, since $\mathcal{A}$ is increasing, Lemma 2.3.1 tells us that

$$\mu_{\frac{3}{4}}(\mathcal{A}) \geq \mu_{\frac{1}{2}+\varepsilon}(\mathcal{A}) \geq 1 - \varepsilon > 1 - \delta^2.$$

Next, by Proposition 2.2.1, we have

$$\mu_{\frac{1}{4}}(\mathcal{I}(\mathcal{A})) \geq \delta^2.$$

Finally, since $\mathcal{A}$ is a 3-wise intersecting family, $\mathcal{A}$ and $\mathcal{I}(\mathcal{A})$ are cross-intersecting, so we conclude from Proposition 2.2.4 that

$$\mu_{\frac{3}{4}}(\mathcal{A}) \leq 1 - \mu_{\frac{1}{4}}(\mathcal{I}(\mathcal{A})) \leq 1 - \delta^2,$$

yielding a contradiction, and establishing the result. □

2.4 Conclusion

The best bound for Theorem 2.1.2 that we may read out of the argument here is rather poor on account of our reliance on the junta theorem; it would therefore be interesting to improve this. Concretely, it would be good to decide if any 3-wise intersecting family $\mathcal{A} \subseteq 2^{[n]}$ that is both regular and increasing must satisfy

$$\log |\mathcal{A}| \leq n - cn^{\delta},$$

where $c, \delta > 0$ are universal constants; as evidenced by the constructions in [12], a bound of this type would be the best one could hope for. We ought to point out that we do not yet know how to prove an estimate of the above form even for *symmetric* 3-wise intersecting families; what is known however is that such an estimate does hold for symmetric 4-wise intersecting families, as was shown by Cameron, Frankl and Kantor [7].

Chapter 3

Thresholds versus fractional expectation-thresholds

Joint work with Jeff Kahn, Bhargav Narayanan, and Jinyoung Park.

3.1 Preliminaries

Following [50, 51, 53], we say $\mathcal{F}$ is *p-small* if there is a $\mathcal{G} \subseteq 2^X$ such that $\mathcal{F} \subseteq \langle \mathcal{G} \rangle := \{T : \exists S \in \mathcal{G}, S \subseteq T\}$ and

$$\textstyle\sum_{S \in \mathcal{G}} p^{|S|} \leq 1/2. \tag{3.1}$$

Then $q(\mathcal{F}) := \max\{p : \mathcal{F} \text{ is } p\text{-small}\}$, which we call the *expectation-threshold* of $\mathcal{F}$ (note the term is used slightly differently in [33]), is a trivial lower bound on $p_c(\mathcal{F})$, since for $\mathcal{G}$ as above and T drawn from μ_p,

$$\mu_p(\mathcal{F}) \leq \mu_p(\langle \mathcal{G} \rangle) \leq \textstyle\sum_{S \in \mathcal{G}} \mu_p(T \supseteq S) = \sum_{S \in \mathcal{G}} p^{|S|} \quad (= \mathbb{E}[|\{S \in \mathcal{G} : S \subseteq T\}|]). \tag{3.2}$$

The following statement, the main conjecture (Conjecture 1) of [33], says that for *any* $\mathcal{F}$, this trivial lower bound on $p_c(\mathcal{F})$ is close to the truth.

Conjecture 3.1.1. *There is a universal K such that for every finite X and increasing $\mathcal{F} \subseteq 2^X$,*

$$p_c(\mathcal{F}) \leq K q(\mathcal{F}) \ln |X|.$$

We should emphasize how strong this is (from [33]: "It would probably be more sensible to conjecture that it is *not* true"). For example, it easily implies—and was largely motivated by—Erdős–Rényi thresholds for (a) perfect matchings in random r-uniform hypergraphs, and (b) appearance of a given bounded degree spanning tree in a random graph. These have since been resolved: the first—*Shamir's Problem*, circa 1980—in [31], and the second—a mid-90's suggestion of Kahn—in [44]. Both arguments are difficult and specific to the problems they address (e.g. they are utterly unrelated either to each other or to what we do here). See Section 3.7 for more on these and other consequences.

Talagrand [50, 53] suggests relaxing "p-small" by replacing the set system $\mathcal{G}$ above by what we may think of as a *fractional* set system, g: say $\mathcal{F}$ is *weakly p-small* if there is a $g : 2^X \to \mathbb{R}^+$ such that

$$\textstyle\sum_{S\subseteq T} g(S) \geq 1 \;\; \forall T \in \mathcal{F} \quad \text{and} \quad \sum_{S\subseteq X} g(S)p^{|S|} \leq 1/2.$$

Then $q_f(\mathcal{F}) := \max\{p : \mathcal{F} \text{ is weakly } p\text{-small}\}$, the *fractional expectation-threshold* for $\mathcal{F}$, satisfies

$$q(\mathcal{F}) \leq q_f(\mathcal{F}) \leq p_c(\mathcal{F}) \tag{3.3}$$

(the first inequality is trivial and the second is similar to (3.2)), and Talagrand [53, Conjectures 8.3 and 8.5] proposes a sort of LP relaxation of Conjecture 3.1.1, and then a strengthening thereof. The first of these, the following, replaces q by q_f in Conjecture 3.1.1; the second, which adds replacement of $|X|$ by the smaller $\ell(\mathcal{F})$, is our Theorem 1.0.1.

Conjecture 3.1.2. *There is a universal K such that for every finite X and increasing $\mathcal{F} \subseteq 2^X$,*

$$p_c(\mathcal{F}) \leq Kq_f(\mathcal{F}) \ln |X|.$$

Talagrand further suggests the following "very nice problem of combinatorics," which implies *equivalence of* Conjectures 3.1.1 and 3.1.2, as well as of Theorem 1.0.1 and the corresponding strengthening of Conjecture 3.1.1.

Conjecture 3.1.3. *There is a universal K such that, for any increasing $\mathcal{F}$ on a finite set X, $q(\mathcal{F}) \geq q_f(\mathcal{F})/K$.*

(That is, weakly p-small implies (p/K)-small.)

Note the interest here is in Conjecture 3.1.3 for its own sake and as the most likely route to Conjecture 3.1.1; all applications of the latter that we're aware of follow just as easily from Theorem 1.0.1.

Spread hypergraphs and spread measures. In this chapter, a *hypergraph* on the (*vertex*) set X is a collection $\mathcal{H}$ of subsets of X (*edges* of $\mathcal{H}$), with *repeats allowed.* For $S \subseteq X$, we use $\langle S\rangle$ for $\{T \subseteq X : T \supseteq S\}$, and for a hypergraph $\mathcal{H}$ on X, we write $\langle \mathcal{H}\rangle$ for $\cup_{S\in\mathcal{H}}\langle S\rangle$. We say $\mathcal{H}$ is *ℓ-bounded* (resp. *ℓ-uniform* or an *ℓ-graph*) if each of its members has size at most (resp. exactly) ℓ, and *κ-spread* if

$$|\mathcal{H} \cap \langle S\rangle| \le \kappa^{-|S|}|\mathcal{H}| \quad \forall S \subseteq X. \tag{3.4}$$

(Note that edges are counted with multiplicities on both sides of (3.4).)

A major advantage of the fractional versions (Conjecture 3.1.2 and Theorem 1.0.1) over Conjecture 3.1.1—and the source of the present relevance of [3]—is that they admit, via linear programming duality, reformulations in which the specification of $q_f(\mathcal{F})$ gives a usable starting point. Following [53], we say a probability measure ν on 2^X is *q-spread* if

$$\nu(\langle S\rangle) \le q^{|S|} \quad \forall S \subseteq X.$$

Thus a hypergraph $\mathcal{H}$ is κ-spread iff uniform measure on $\mathcal{H}$ is q-spread with $q = \kappa^{-1}$.

As observed by Talagrand [53], the following is an easy consequence of duality.

Proposition 3.1.4. *For an increasing family $\mathcal{F}$ on X, if $q_f(\mathcal{F}) \le q$, then there is a $(2q)$-spread probability measure on 2^X supported on $\mathcal{F}$.* □

This allows us to reduce Theorem 1.0.1 to the following alternate (actually, equivalent) statement. In this chapter, *with high probability* (w.h.p.) means with probability tending to 1 as $\ell \to \infty$.

Theorem 3.1.5. *There is a universal K such that for any ℓ-bounded, κ-spread hypergraph $\mathcal{H}$ on X, a uniformly random $((K\kappa^{-1}\ln \ell)|X|)$-element subset of X belongs to $\langle \mathcal{H}\rangle$ w.h.p.*

The easy reduction is given in Section 3.2.

Assignments. The second main result of this chpater provides upper bounds on the minima of a large class of hypergraph-based stochastic processes, somewhat in the spirit of [52] (see also [51, 54]), saying that in "smoother" settings, the logarithmic corrections of Conjectures 3.1.1 and 3.1.2 and Theorem 1.0.1 are not needed.

For a hypergraph $\mathcal{H}$ on X, let ξ_x $(x \in X)$ be independent random variables, each uniform from $[0,1]$, and set

$$\xi_{\mathcal{H}} = \min_{S\in\mathcal{H}} \sum_{x\in S} \xi_x \tag{3.5}$$

and $Z_{\mathcal{H}} = \mathbb{E}[\xi_{\mathcal{H}}]$.

Theorem 3.1.6. *There is a universal K such that for any ℓ-bounded, κ-spread hypergraph $\mathcal{H}$, we have $Z_{\mathcal{H}} \le K\ell/\kappa$, and $\xi_{\mathcal{H}} \le K\ell/\kappa$ w.h.p.*

These bounds are often tight (again up to the value of K). The distribution of the ξ_x's is not very important; e.g. it's easy to see that the same statement holds if they are Exp(1) random variables, as in the next example.

Theorem 3.1.6 was motivated by work of Frieze and Sorkin [26] on the "axial" version of the *random d-dimensional assignment problem*. This asks (for fixed d and large n) for estimation of

$$Z_d^A(n) = \mathbb{E}\left[\min \sum_{x\in S} \xi_x\right], \tag{3.6}$$

where the ξ_x's ($x \in X := [n]^d$) are independent Exp(1) weights and S ranges over "axial assignments," meaning $S \subseteq X$ meets each *axis-parallel hyperplane* ($\{x \in X : x_i = a\}$ for some $i \in [d]$ and $a \in [n]$) exactly once. For $d = 2$ this is classical; see [26] for its rather glorious history. For $d = 3$ the deterministic version was one of Karp's [34] original NP-complete problems. Progress on the random version has been limited; see [26] for a guide to the literature.

Frieze and Sorkin show (regarding bounds; they are also interested in algorithms) that for suitable $c_1 > 0$ and c_2

$$c_1 n^{-(d-2)} < Z_d^A(n) < c_2 n^{-(d-2)} \ln n. \tag{3.7}$$

(The lower bound is easy and the upper bound follows from the Shamir bound of [31].)

In present language, $Z_d^A(n)$ is essentially (that is, apart from the difference in the distributions of the ξ_x's) $Z_{\mathcal{H}}$, with $\mathcal{H}$ the set of perfect matchings of the complete, balanced d-uniform d-partite hypergraph on dn vertices (that is, the collection of d-sets meeting each of the pairwise disjoint n-sets $V_1, \ldots, V_d$). This is easily seen to be

κ-spread with $\kappa = (n/e)^{d-1}$ (apart from the nearly irrelevant d-particity, it is the $\mathcal{H}$ of Shamir's Problem), so the correct bound is an instance of Theorem 3.1.6:

Corollary 3.1.7. $Z_d^A(n) = \Theta(n^{-(d-2)})$.

Frieze and Sorkin also considered the "planar" version of the problem, in which S in (3.6) meets each *line* ($\{x \in X : x_j = y_j \ \forall j \neq i\}$ for some $i \in [d]$ and $y \in X$) exactly once; and one may of course generalise from hyperplanes/lines to k-dimensional "subspaces" for a given $k \in [d-1]$. It's easy to see what to expect here, and one may hope Theorem 3.1.6 will eventually apply, but we at present lack the technology to say the relevant $\mathcal{H}$'s are suitably spread (see Section 3.8).

Organisation. Section 3.2 includes minor preliminaries and the derivation of Theorem 1.0.1 from Theorem 3.1.5. The heart of our argument, Lemma 3.3.1, is proved in Section 3.3. Our approach here strengthens that of the recent breakthrough of Alweiss, Lovett, Wu and Zhang [3] on the Erdős–Rado "Sunflower Conjecture" [14]. Section 3.4 adds one small technical point (more or less repeated from [3]), and the proofs of Theorems 3.1.5 and 3.1.6 are given in Sections 3.5 and 3.6. Finally, Section 3.7 outlines a few applications and Section 3.8 discusses unresolved questions.

3.2 Little things

Usage. As is usual, we use $[n]$ for $\{1, 2, \ldots, n\}$, 2^X for the power set of X, $\binom{X}{r}$ for the family of r-element subsets of X, and $[S, T]$ for $\{R : S \subseteq R \subseteq T\}$. Our default universe is X, with $|X| = n$.

In what follows we assume ℓ and n are somewhat large (when there is an ℓ it will be at most n), as we may do since smaller values can by handled by adjusting the K's in Theorems 3.1.5 and 3.1.6. Asymptotic notation referring to some parameter λ (usually ℓ) is used in the natural way: implied constants in $O(\cdot)$ and $\Omega(\cdot)$ are independent of λ, and $f = o(g)$ (also written $f \ll g$) means f/g is smaller than any given $\varepsilon > 0$ for large enough values of λ. Following a standard abuse, we usually pretend large numbers are integers.

For $p \in [0,1]$ and $m \in [n]$, X_p and X_m are (respectively) a p-random subset of X (drawn from μ_p) and a uniformly random m-element subset of X. The latter is not entirely kosher, since we will also see *sequences* X_i; but we will never see both interpretations in close proximity, and the overlap should cause no confusion.

In a couple places it will be helpful to assume uniformity, which we will justify using the next little point.

Observation 3.2.1. *If $\mathcal{H}$ is ℓ-bounded and κ-spread, and we replace each $S \in \mathcal{H}$ by M new edges, each consisting of S plus $\ell - |S|$ new vertices (each used just once), then for large enough M, the resulting ℓ-graph $\mathcal{G}$ is again κ-spread.*

Derivation of Theorem 1.0.1 from Theorem 3.1.5. Let $\mathcal{F}$ be as in Theorem 1.0.1 with $\mathcal{G}$ its set of minimal elements, let ℓ with $\ell(\mathcal{F}) \le \ell = O(\ell(\mathcal{F}))$ be large enough that the exceptional probability in Theorem 3.1.5 is less than $1/4$ and let ν be the $(2q)$-spread probability measure promised by Proposition 3.1.4, where $q = q_f(\mathcal{F})$. We may assume ν is supported on $\mathcal{G}$ (since transferring weight from S to $T \subseteq S$ doesn't destroy the spread condition) and that ν takes values in $\mathbb{Q}$. We may then replace $\mathcal{G}$ by $\mathcal{H}$ whose edges are copies of edges of $\mathcal{G}$, and ν by uniform measure on $\mathcal{H}$.

Setting $m = ((2Kq \ln \ell)n)$ and $p = 2m/n$ (with $n = |X|$ and K as in Theorem 3.1.5), we then have (using Theorem 3.1.5 with $\kappa = 1/(2q)$)

$$\mu_p(\mathcal{F}) \ge \ \mathbb{P}(X_p \in \langle \mathcal{H} \rangle) \ge \mathbb{P}(|X_p| \ge m)\mathbb{P}(X_m \in \langle \mathcal{H} \rangle) \ge 3\mathbb{P}(|X_p| \ge m)/4 > 1/2,$$

implying $p_c(\mathcal{F}) < p = 4Kq \ln \ell$. (Note $\mathcal{H}$ q-spread with $\emptyset \notin \mathcal{H}$ implies $q \ge 1/n$, so that m is somewhat large and $\mathbb{P}(|X_p| \ge m) \approx 1$.) $\square$

Remark 3.2.2. *This was done fussily to cover smaller ℓ in Theorem 1.0.1; if $\ell \to \infty$, then it gives $\mathbb{P}(X_p \in \langle \mathcal{H} \rangle) \to 1$.*

3.3 Main Lemma

Let γ be a slightly small constant (e.g. $\gamma = 0.1$ suffices), and let C_0 be a constant large enough to support the estimates that follow. Let $\mathcal{H}$ be an r-bounded, κ-spread hypergraph on a set X of size n, with $r, \kappa \geq C_0^2$. Set $p = C/\kappa$ with $C_0 \leq C \leq \kappa/C_0$ (so $p \leq 1/C_0$), $r' = (1-\gamma)r$ and $N = \binom{n}{np}$. Finally, fix $\psi : \langle \mathcal{H} \rangle \to \mathcal{H}$ satisfying $\psi(Z) \subseteq Z$ for all $Z \in \langle \mathcal{H} \rangle$; set, for $W \subseteq X$ and $S \in \mathcal{H}$,

$$\chi(S, W) = \psi(S \cup W) \setminus W;$$

and say the pair (S, W) is *bad* if $|\chi(S, W)| > r'$ and *good* otherwise.

The heart of our argument is the following lemma (improving [3, Lemma 5.7]), regarding which a little orientation may be helpful. We will (in Theorems 3.1.5 and 3.1.6) be choosing a random subset of X in small increments and would like to say we are likely to be making good progress toward containing some $S \in \mathcal{H}$. Of course such progress is not to be expected for a *typical* S, but this is not the goal: having chosen a portion W of our eventual set, we just need the remainder to contain *some* $S \setminus W$, and may focus on those that are more likely (basically meaning small). The key idea (introduced in [3] and refined here) is that a general $S \setminus W$, while not itself small, will, in consequence of the spread assumption, typically *contain* some small $S' \setminus W$. (In fact $\chi(S, W)$ will usually be one of these: an $S' \setminus W$ contained in $S \setminus W$ will *typically* be small, so we don't need to steer this choice.) We then replace each "good" $S \setminus W$ by $\chi(S, W)$ and iterate, a second nice feature of the spread condition being that it is not much affected by this substitution.

Lemma 3.3.1. *For $\mathcal{H}$ as above, and W chosen uniformly from $\binom{X}{np}$,*

$$\mathbb{E}[|\{S \in \mathcal{H} : (S, W) \text{ is bad}\}|] \leq |\mathcal{H}| C^{-r/3}.$$

Proof. It is enough to show, for $s \in (r', r]$,

$$\mathbb{E}\left[|\{S \in \mathcal{H} : (S, W) \text{ is bad and } |S| = s\}|\right] \leq (\gamma r)^{-1} |\mathcal{H}| C^{-r/3}, \tag{3.8}$$

or, equivalently, that

$$|\{(S, W) : (S, W) \text{ is bad and } |S| = s\}| \leq (\gamma r)^{-1} N |\mathcal{H}| C^{-r/3}. \tag{3.9}$$

(Note $\gamma r = r - r'$ bounds the number of s for which the set in question can be nonempty, whence the negligible factors $(\gamma r)^{-1}$.)

We now use $\mathcal{H}_s = \{S \in \mathcal{H} : |S| = s\}$. Let $B = \sqrt{C}$ and for $Z \supseteq S \in \mathcal{H}_s$ say (S, Z) is *pathological* if there is $T \subseteq S$ with $t := |T| > r'$ and

$$|\{S' \in \mathcal{H}_s : S' \in [T, Z]\}| > B^r |\mathcal{H}| \kappa^{-t} p^{s-t}. \tag{3.10}$$

From now on we will always take $Z = W \cup S$ (with W as in Lemma 3.3.1); thus $|Z|$ is typically roughly np and, since $\mathcal{H}$ is κ-spread, $|\mathcal{H}|\kappa^{-t}p^{s-t}$ is a natural upper bound on what one might expect for the l.h.s. of (3.10).

Note that in proving (3.9) we may assume $s \le n/2$: we may of course assume $|\mathcal{H}_s|$ is at least the r.h.s. of (3.8); but then for an $S \in \mathcal{H}_s$ of largest multiplicity, say m, we have

$$m \le \kappa^{-s}|\mathcal{H}| \le \kappa^{-s}\gamma r C^{r/3}|\mathcal{H}_s| \le \kappa^{-s}\gamma r C^{r/3} m 2^n,$$

which is less than m if $s > n/2$ (since $\kappa > C$).

We bound the nonpathological and pathological parts of (3.9) separately; this (with the introduction of "pathological") is the source of our improvement over [3].

Nonpathological contributions. We first bound the number of (S, W) in (3.9) with (S, Z) nonpathological. This basically follows [3], but "nonpathological" allows us to bound the number of possibilities in Step 3 below by the r.h.s. of (3.10), where [3] settles for something like $|\mathcal{H}|\kappa^{-t}$.

Step 1. There are at most

$$\sum_{i=0}^{s} \binom{n}{np+i} \le \binom{n+s}{np+s} \le Np^{-s} \tag{3.11}$$

choices for $Z = W \cup S$.

Step 2. Given Z, let $S' = \psi(Z)$. Choose $T := S \cap S'$, for which there are at most $2^{|S'|} \le 2^r$ possibilities, and set $t = |T| > r'$. (If $t \le r'$ then (S, W) cannot be bad, as $\chi(S, W) = S' \setminus W \subseteq T$.)

Step 3. Since we are only interested in nonpathological choices, the number of possibilities for S is now at most

$$B^r |\mathcal{H}| \kappa^{-t} p^{s-t}.$$

Step 4. Complete the specification of (S,W) by choosing $W \cap S$, the number of possibilities for which is at most 2^s.

In sum, since $s \le r$ and $t > r' = (1-\gamma)r$, the number of nonpathological possibilities is at most

$$2^{r+s}N|\mathcal{H}|B^r(p\kappa)^{-t} \le N|\mathcal{H}|(4B)^rC^{-t} < N|\mathcal{H}|[4BC^{-(1-\gamma)}]^r. \tag{3.12}$$

Pathological contributions. We next bound the number of (S,W) as in (3.9) with (S,Z) pathological. The main point here is Step 4.

Step 1. There are at most $|\mathcal{H}|$ possibilities for S.

Step 2. Choose $T \subseteq S$ witnessing the pathology of (S,Z) (i.e. for which (3.10) holds); there are at most 2^s possibilities for T.

Step 3. Choose $U \in [T,S]$ for which

$$|\mathcal{H}_s \cap [U,(Z\setminus S)\cup U]| > 2^{-(s-t)}B^r|\mathcal{H}|\kappa^{-t}p^{s-t}. \tag{3.13}$$

(Here the left hand side counts members of $\mathcal{H}_s$ in Z whose intersection with S is precisely U. Of course, existence of U as in (3.13) follows from (3.10).) The number of possibilities for this choice is at most 2^{s-t}.

Step 4. Choose $Z \setminus S$, the number of choices for which is less than $N(2/B)^r$. To see this, write Φ for the r.h.s. of (3.13). Noting that $Z\setminus S$ must belong to $\binom{X\setminus S}{np}\cup\binom{X\setminus S}{np-1}\cup \cdots \cup \binom{X\setminus S}{np-s}$, we consider, for Y drawn uniformly from this set,

$$\mathbb{P}(|\mathcal{H}_s \cap [U, Y\cup U]| > \Phi). \tag{3.14}$$

Set $|U| = u$. We have

$$|\mathcal{H}_s \cap \langle U\rangle| \le |\mathcal{H} \cap \langle U\rangle| \le |\mathcal{H}|\kappa^{-u},$$

while, for any $S' \in \mathcal{H}_s \cap \langle U\rangle$,

$$\mathbb{P}(Y \supseteq S'\setminus U) \le \left(\frac{np}{n-s}\right)^{s-u}$$

(of course if $S' \cap S \ne U$ the probability is zero); so

$$\vartheta := \mathbb{E}\left[|\mathcal{H}_s \cap [U, Y\cup U]|\right] \le |\mathcal{H}|\kappa^{-u}\left(\frac{np}{n-s}\right)^{s-u} \le |\mathcal{H}|\kappa^{-u}\,(2p)^{s-u}$$

(since $n-s \geq n/2$). Markov's Inequality then bounds the probability in (3.14) by ϑ/Φ, and this bounds the number of possibilities for $Z \setminus S$ by $N(\vartheta/\Phi)$ (*cf.* (3.11)), which is easily seen to be less than $N(2/B)^r$.

Step 5. Complete the specification of (S, W) by choosing $S \cap W$, which can be done in at most 2^s ways.

Combining (and slightly simplifying), we find that the number of pathological possibilities is at most

$$|\mathcal{H}|N(16/B)^r. \tag{3.15}$$

Finally, the sum of the bounds in (3.12) and (3.15) is less than the $(\gamma r)^{-1}N|\mathcal{H}|C^{-r/3}$ of (3.9). □

3.4 Small uniformities

As in [3, Lemma 5.9], very small set sizes are handled by a simple Janson bound:

Lemma 3.4.1. *For an r-bounded, κ-spread $\mathcal{G}$ on Y, and $\alpha \in (0,1)$,*

$$\mathbb{P}(Y_\alpha \notin \langle \mathcal{G} \rangle) \leq \exp\left[-\left(\sum_{t=1}^{r}\binom{r}{t}(\alpha\kappa)^{-t}\right)^{-1}\right]. \tag{3.16}$$

Proof. We may assume $\mathcal{G}$ is r-uniform, since modifying it according to Observation 3.2.1 doesn't decrease the probability in (3.16). Denote members of $\mathcal{G}$ by S_i and set $\zeta_i = \mathbf{1}_{\{Y_\alpha \supseteq S_i\}}$. Then

$$\mu := \sum \mathbb{E}[\zeta_i] = |\mathcal{G}|\alpha^r$$

and

$$\Lambda := \sum\sum\{\mathbb{E}[\zeta_i\zeta_j] : S_i \cap S_j \neq \emptyset\} \leq |\mathcal{G}|\sum_{t=1}^{r}\binom{r}{t}\kappa^{-t}|\mathcal{G}|\alpha^{2r-t} = \mu^2\sum_{t=1}^{r}\binom{r}{t}(\alpha\kappa)^{-t}$$

(where the inequality holds because $\mathcal{G}$ is κ-spread), and Janson's Inequality (e.g. [30, Thm. 2.18(ii)]) bounds the probability in (3.16) by $\exp[-\mu^2/\Lambda]$. □

Corollary 3.4.2. *Let $\mathcal{G}$ be as in Lemma 3.4.1, let $t = \alpha|Y|$ be an integer with $\alpha\kappa \geq 2r$, and let $W = Y_t$. Then*

$$\mathbb{P}(W \notin \langle \mathcal{G} \rangle) \leq 2\exp[-\alpha\kappa/(2r)].$$

Proof. Lemma 3.4.1 gives

$$\exp[-\alpha\kappa/(2r)] \geq \mathbb{P}(Y_\alpha \notin \langle \mathcal{G} \rangle) \geq \mathbb{P}(|Y_\alpha| \leq t)\mathbb{P}(W \notin \langle \mathcal{G} \rangle) \geq \mathbb{P}(W \notin \langle \mathcal{G} \rangle)/2,$$

where we use the fact that any binomial ξ with $\mathbb{E}[\xi] \in \mathbb{Z}$ satisfies $\mathbb{P}(\xi \leq \mathbb{E}[\xi]) \geq 1/2$; see e.g. [40]. □

3.5 Proof of Theorem 3.1.5

It will be (very slightly) convenient to prove the theorem assuming $\mathcal{H}$ is (2κ)-spread. Let γ and C_0 be as in Section 3.3 and $\mathcal{H}$ as in the statement of Theorem 3.1.5, and recall that asymptotics refer to ℓ. We may of course assume that $\kappa \geq 2\gamma^{-1}C_0 \ln \ell$ (or the result is trivial with a suitably adjusted K).

Fix an ordering "$\prec$" of $\mathcal{H}$. In what follows we will have a sequence $\mathcal{H}_i$, with $\mathcal{H}_0 = \mathcal{H}$ and

$$\mathcal{H}_i \subseteq \{\chi_i(S, W_i) : S \in \mathcal{H}_{i-1}\},$$

where W_i and χ_i will be defined below (with χ_i a version of the χ of Section 3.3). We then order $\mathcal{H}_i$ by setting

$$\chi_i(S, W_i) \prec_i \chi_i(S', W_i) \Leftrightarrow S \prec_{i-1} S'.$$

(So each member of $\mathcal{H}_i$ ultimately inherits its position in $\prec_i$ from some member of $\mathcal{H}$. This is not very important: we will be applying Lemma 3.3.1 repeatedly, and the present convention just provides a concrete ψ for each stage of the iteration.)

Set $C = C_0$ and $p = C/\kappa$, define m by $(1-\gamma)^m = \sqrt{\ln \ell}/\ell$, and set $q = \ln \ell/\kappa$. Then $\gamma^{-1} \ln \ell \sim m \leq \gamma^{-1} \ln \ell$ and Theorem 3.1.5 will follow from the next assertion.

Claim 3.5.1. *If W is a uniform $((mp+q)n)$-subset of X, then $W \in \langle \mathcal{H} \rangle$ w.h.p.*

Proof. Set $\delta = 1/(2m)$. Let $r_0 = \ell$ and $r_i = (1-\gamma)r_{i-1} = (1-\gamma)^i r_0$ for $i \in [m]$. Let $X_0 = X$ and, for $i = 1, \ldots, m$, let W_i be uniform from $\binom{X_{i-1}}{np}$ and set $X_i = X_{i-1} \setminus W_i$. (Note the assumption $\kappa \geq 2\gamma^{-1}C_0 \ln \ell$ ensures $|X_m| \geq n/2$.)

For $S \in \mathcal{H}_{i-1}$ let $\chi_i(S, W_i) = S' \setminus W_i$, where S' is the first member of $\mathcal{H}_{i-1}$ contained in $W_i \cup S$ (with $\mathcal{H}_{i-1}$ ordered by $\prec_{i-1}$). Say S is *good* if $|\chi_i(S, W_i)| \leq r_i$ (and *bad* otherwise), and set

$$\mathcal{H}_i = \{\chi_i(S, W_i) : S \in \mathcal{H}_{i-1} \text{ is good}\}.$$

Thus $\mathcal{H}_i$ is an r_i-bounded collection of subsets of X_i and inherits the ordering $\prec_i$ as described above.

Finally, choose W_{m+1} uniformly from $\binom{X_m}{nq}$. Then $W := W_1 \cup \cdots \cup W_{m+1}$ is as in Claim 3.5.1. Note also that $W \in \langle \mathcal{H} \rangle$ whenever $W_{m+1} \in \langle \mathcal{H}_m \rangle$. (More generally, $W_1 \cup \cdots \cup W_i \cup Y \in \langle \mathcal{H} \rangle$ whenever $Y \subseteq X_i$ lies in $\langle \mathcal{H}_i \rangle$.)

So to prove the claim, we just need to show

$$\mathbb{P}(W_{m+1} \in \langle \mathcal{H}_m \rangle) = 1 - o(1) \tag{3.17}$$

(where the $\mathbb{P}$ refers to the entire sequence $W_1, \ldots, W_{m+1}$).

For $i \in [m]$ call W_i *successful* if $|\mathcal{H}_i| \geq (1-\delta)|\mathcal{H}_{i-1}|$, call W_{m+1} successful if it lies in $\langle \mathcal{H}_m \rangle$, and say a sequence of W_i's is successful if each of its entries is. We show a little more than (3.17):

$$\mathbb{P}(W_1, \ldots, W_{m+1} \text{ is successful}) = 1 - \exp\left[-\Omega(\sqrt{\ln \ell})\right]. \tag{3.18}$$

For $i \in [m]$, according to Lemma 3.3.1 (and Markov's Inequality),

$$\mathbb{P}(W_i \text{ is } \textit{not} \text{ successful} \,|\, W_1, \ldots, W_{i-1} \text{ is successful}) < \delta^{-1} C^{-r_{i-1}/3},$$

since $W_1, \ldots, W_{i-1}$ successful implies $|\mathcal{H}_{i-1}| > (1-\delta)^m |\mathcal{H}| > |\mathcal{H}|/2$, which, since $|\mathcal{H}_{i-1} \cap \langle I \rangle| \leq |\mathcal{H} \cap \langle I \rangle|$ and we assume $\mathcal{H}$ is (2κ)-spread), gives the spread condition (3.4) for $\mathcal{H}_{i-1}$. Thus

$$\mathbb{P}(W_1, \ldots, W_m \text{ is successful}) > 1 - \delta^{-1} \sum_{i=1}^{m} C^{-r_{i-1}/3} > 1 - \exp\left[-\sqrt{\ln \ell}\right] \tag{3.19}$$

(using $r_m = \sqrt{\ln \ell}$).

Finally, if $W_1, \ldots, W_m$ is successful, then Corollary 3.4.2 (applied with $\mathcal{G} = \mathcal{H}_m$, $Y = X_m$, $\alpha = nq/|Y| \geq q$, $r = r_m$, and $W = W_{m+1}$) gives

$$\mathbb{P}(W_{m+1} \notin \langle \mathcal{H}_m \rangle) \leq 2 \exp\left[-\sqrt{\ln \ell}/2\right], \tag{3.20}$$

and we have (3.18) and the claim. $\square$

3.6 Proof of Theorem 3.1.6

We assume the setup of Theorem 3.1.6 with γ and C_0 as in Section 3.3 and $\kappa \geq C_0^2$ (or there is nothing to prove). We may assume $\mathcal{H}$ is ℓ-uniform, since the construction of Observation 3.2.1 produces an ℓ-uniform, κ-spread $\mathcal{G}$ with $\xi_{\mathcal{G}} \geq \xi_{\mathcal{H}}$. In particular this gives

$$|\mathcal{H}|\ell = \sum_{x \in X} |\mathcal{H} \cap \langle x \rangle| \leq n\kappa^{-1}|\mathcal{H}|. \tag{3.21}$$

We first assume κ is *slightly* large, precisely

$$\kappa \geq \ln^3 \ell; \tag{3.22}$$

the similar but easier argument for smaller values will be given at the end. (The bound in (3.22) is convenient but there is nothing delicate about this choice.)

Claim 3.6.1. *For κ as in* (3.22) *and $C_0 \leq C \leq \gamma\kappa/(4\ln \ell)$,*

$$\mathbb{P}(\xi_{\mathcal{H}} > (3C/\gamma)\ell/\kappa) < \exp[-(\ln \ell \ln C)/4].$$

Proof of Theorem 3.1.6 in regime (3.22) *given Claim 3.6.1.* The "w.h.p." statement is immediate (take $C = C_0$). For the expectation, $Z_{\mathcal{H}}$, set $t = (3C_0/\gamma)\ell/\kappa$ and $T = 3\ell/(4\ln \ell)$. By Claim 3.6.1 we have, for all $x \in [t, T]$,

$$\mathbb{P}(\xi_{\mathcal{H}} > x) \leq f(x) := \exp\left[-\ln \ell \ln(\gamma\kappa x/3\ell)/4\right] = (bx)^a = b^a x^a,$$

where $a = -(\ln \ell)/4$ and $b = \gamma\kappa/3\ell$. Noting that $\xi_{\mathcal{H}} \leq \ell$, we then have

$$Z_{\mathcal{H}} \leq t + \int_t^T \mathbb{P}(\xi_{\mathcal{H}} > x)dx + \ell\mathbb{P}(\xi_{\mathcal{H}} > T) \leq t + \int_t^T f(x)dx + \ell f(T) = O(\ell/\kappa).$$

Here $t = O(\ell/\kappa)$ and the other terms are much smaller: the integral is less than $-1/(a+1)b^a t^{a+1} = O(1/\ln \ell)C_0^a t$, while (3.22) easily implies that $f(T) = (\gamma\kappa/(4\ln \ell))^a$ is $o(1/\kappa)$. $\square$

Proof of Claim 3.6.1. Terms not defined here (beginning with $p = C/\kappa$ and W_i; note C is now as in Claim 3.6.1, rather than set to C_0) are as in Section 3.5, but we (re)define m by $(1-\gamma)^m = \ln \ell/\ell$ and set $q = \ln C \ln^2 \ell/\kappa$, noting that (3.21) gives $p \geq C\ell/n$.

It's now convenient to generate the W_i's using the ξ_x's in the natural way: let

$$a_i = \begin{cases} (ip)n & \text{if } i \in \{0\} \cup [m], \\ (mp+q)n & \text{if } i = m+1, \end{cases}$$

and let W_i consist of the x's in positions $a_{i-1}+1, \ldots, a_i$ when X is ordered according to the ξ_x's.

Proposition 3.6.2. *With probability* $1 - e^{-\Omega(C\ell)}$,

$$\xi_x \le \varepsilon_i := \left\{ \begin{array}{ll} 2ip & \textit{if } i \in \{0\} \cup [m] \\ 2(mp+q) & \textit{if } i = m+1 \end{array} \right\} \textit{ for all } i \textit{ and } x \in W_i. \tag{3.23}$$

Proof. Failure at $i \ge 1$ implies

$$|\xi^{-1}[0, \varepsilon_i]| < a_i. \tag{3.24}$$

But $|\xi^{-1}[0, \varepsilon_i]|$ is binomial with mean $\varepsilon_i n = 2a_i \ge 2C\ell$, so the probability that (3.24) occurs for some i is less than $\exp[-\Omega(C\ell)]$ (see e.g. [30, Theorem 2.1]). $\square$

We now write $\overline{W}_i$ for $W_1 \cup \cdots \cup W_i$.

Proposition 3.6.3. *If* $W_{m+1} \in \langle \mathcal{H}_m \rangle$, *then* W *contains some* $S \in \mathcal{H}$ *with*

$$|S \setminus \overline{W}_i| \le r_i \ \forall i \in [m].$$

Proof. Suppose $W \supseteq S_m \in \mathcal{H}_m$. By construction (of the $\mathcal{H}_i$'s) there are $S_{m-1}, \ldots, S_0 =: S$ with $S_i \in \mathcal{H}_i$ and $S_i = S_{i-1} \setminus W_i$, whence $S_i = S \setminus \overline{W}_i$ for $i \in [m]$; and $S_i \in \mathcal{H}_i$ then gives the proposition. $\square$

We now define "success" for $(\xi_x : x \in X)$ to mean that $W_1, \ldots, W_{m+1}$ is successful in our earlier sense *and* (3.23) holds. Notice that with our current values of m and q (and $r_m = \ell(1-\gamma)^m = \ln \ell$), we can replace the error terms in (3.19) and (3.20) by essentially $\delta^{-1} C^{-\ln \ell/3}$ and $e^{-(\ln C \ln \ell)/2}$, which with Proposition 3.6.2 bounds the probability that $(\xi_x : x \in X)$ is *not* successful by (say) $\exp[-(\ln \ell \ln C)/4]$.

We finish with the following observation.

Proposition 3.6.4. *If* $(\xi_x : x \in X)$ *is successful then* $\xi_{\mathcal{H}} \le (3C/\gamma)\ell/\kappa$.

Proof. For S as in Proposition 3.6.3, we have (with $W_0 = \emptyset$ and $\varepsilon_0 = 0$)

$$\begin{aligned}
\xi_{\mathcal{H}} \le \sum_{i=1}^{m+1} \varepsilon_i |S \cap W_i| \; &= \; \sum_{i=1}^{m+1} (\varepsilon_i - \varepsilon_{i-1}) |S \setminus \overline{W}_{i-1}| \\
&\le 2 \left[\sum_{i=1}^{m} (1-\gamma)^{i-1} p + (1-\gamma)^m q \right] \ell \\
&\le 2[C/(\gamma\kappa) + (\ln \ell/\ell)(\ln C \ln^2 \ell/\kappa)]\ell < (3C/\gamma)\ell/\kappa. \qquad \square
\end{aligned}$$

This completes the proof of Claim 3.6.1 (and of Theorem 3.1.6 when κ satisfies (3.22)). $\square$

Finally, for κ below the bound in (3.22) (actually, for κ up to about $\ell/\ln \ell$), a subset of the preceding argument suffices. We proceed as before, but now only with $C = C_0$ (so $p = C_0/\kappa$), stopping at m defined by $(1-\gamma)^m = 1/\kappa$ (so $m \approx \gamma^{-1} \ln \kappa$). The main difference here is that there is no "Janson" phase: $W_1, \ldots, W_m$ is successful with probability $1 - \exp[-\Omega(\ell/\kappa)]$, and when it *is* successful we have (as in the proof of Proposition 3.6.4, now just taking $W_{m+1} = X \setminus \overline{W}_m$)

$$\xi_{\mathcal{H}} \le \sum_{i=1}^{m} (\varepsilon_i - \varepsilon_{i-1}) |S \setminus \overline{W}_{i-1}| + |S \cap W_{m+1}| < 2(C_0/(\gamma\kappa))\ell + \ell/\kappa$$

(so also $Z_{\mathcal{H}} \le O(\ell/\kappa) + \exp[-\Omega(\ell/\kappa)]\ell = O(\ell/\kappa)$).

3.7 Applications

Much of the significance of Theorem 1.0.1—and of the skepticism with which Conjecture 3.1.1 was viewed in [33]—derives from the strength of their consequences, a few of which we discuss (*briefly*) here.

For this discussion, $\mathcal{K}_n^r = \binom{V}{r}$ is the complete r-graph on $V = [n]$, and $\mathcal{H}_{n,p}^r$ is the r-uniform counterpart of the usual binomial random graph $G_{n,p}$. Given r, n and an r-graph H, we use $\mathcal{G}_H$ for the collection of (unlabeled) copies of H in $\mathcal{K}$ and $\mathcal{F}_H$ for $\langle \mathcal{G}_H \rangle$. As usual, Δ is maximum degree.

As noted earlier, Conjecture 3.1.1 was motivated especially by Shamir's Problem (since resolved in [31]), and the conjecture that became Montgomery's theorem [44]. Very briefly: for n running over multiples of a given (fixed) r, Shamir's Problem asks for estimation of $p_c(\mathcal{F}_H)$ when H is a perfect matching (n/r disjoint edges), and [31] proves the natural conjecture that this threshold is $\Theta(n^{-(r-1)}\ln n)$; and [44] shows that for fixed d, the threshold for $G_{n,p}$ to contain a given n-vertex tree with maximum degree d is $\Theta(n^{-1}\ln n)$, where the implied constant in the upper bound depends on d (though it probably shouldn't). In both of these—and in most of the other examples mentioned following Theorem 3.7.1 (all but the one from [37])—the lower bounds derive from the coupon-collectorish requirement that the (hyper)edges cover the vertices, and it is the upper bounds that are of interest. See [31, 44] for some account of the history of these problems.

In fact, Theorem 1.0.1 gives not just Montgomery's theorem, but its natural extension to r-graphs and more. (Strictly speaking, Montgomery proves more than the original conjecture—see Section 3.8—and we are not so far recovering this stronger result.) Say an r-graph F is a *forest* if it contains no *cycle*, meaning distinct vertices $v_1,\ldots,v_k$ and distinct edges $e_1,\ldots,e_k$ such that $v_{i-1},v_i\in e_i\ \forall i$ (with subscripts mod k). A *spanning tree* is then a forest of size $(n-1)/(r-1)$. For a (general) r-graph F, let $\rho(F)$ be the maximum size of a forest in F and set

$$\varphi(F)=\max\{1-\rho(F')/|F'|:\emptyset\neq F'\subseteq F\}.$$

Theorem 3.7.1. *For each r and c there is a K such that if H is an r-graph on $[n]$ with $\Delta(H)\le d$ and $\varphi(H)\le c/\ln n$, then*

$$p_c(\mathcal{F}_H)<Kdn^{-(r-1)}\ln|H|. \qquad \square$$

This gives $p_c(\mathcal{F}_H)=\Theta(n^{-(r-1)}\ln n)$ if H is a perfect matching (as in Shamir's Problem), or a "loose Hamiltonian cycle" (a result of [10], to which we refer for definitions and history of the problem), and $p_c(\mathcal{F}_H)<Kdn^{-(r-1)}\ln n$ if H is a spanning tree with $\Delta(H)\le d$. For fixed d the latter is the aforementioned r-graph generalization of [44] (or a slight improvement thereof in that the dependence on d—which, again, is probably unnecessary—is explicit), and for $d=n^{\Omega(1)}$ it is a result of Krivelevich [37, Theorem 1], which is again tight up to value of K see [37, Theorem 2]).

The last application we discuss here was suggested to us by Simon Griffiths and Rob Morris. Set $c_d=(d!)^{2/(d(d+1))}$ and $p^*(d,n)=c_dn^{-2/(d+1)}(\ln n)^{2/(d(d+1))}$.

Theorem 3.7.2. *For fixed d and H any graph on $[n]$ with $\Delta(H) \le d$,*

$$p_c(\mathcal{F}_H) < (1 + o(1))p^*(d, n). \tag{3.25}$$

When $(d+1) \mid n$ and H is a K_{d+1}-factor (that is, $n/(d+1)$ disjoint K_{d+1}'s), $p^*(d, n)$ is the asymptotic value of $p_c(\mathcal{F}_H)$. Here (3.25) with $O(1)$ in place of $1+o(1)$ was proved in [31], while the asymptotics are given by the combination of [32] and [47, 27]; we state this in a form convenient for use below:

Theorem 3.7.3. *For fixed d and $\varepsilon > 0$, and n ranging over multiples of $d + 1$, if $p > (1 + \varepsilon)p^*(d, n)$, then $G_{n,p}$ contains a K_{d+1}-factor w.h.p.* □

Interest in $p_c(\mathcal{F}_H)$ for H as in Theorem 3.7.2 dates to at least 1992, when Alon and Füredi [2] showed the upper bound $O(n^{-1/d}(\ln n)^{1/d})$, and has intensified since [31], motivated by the idea that K_{d+1}-factors should be the worst case. See [17, 18] for history and the most recent results; with $O(1)$ in place of $1 + o(1)$, Theorem 3.7.2 is conjectured in [18] and in the stronger "universal" form (see Section 3.8) in [17].

Theorem 3.7.3 probably extends to r-graphs and d of the form $\binom{s-1}{r-1}$. This just needs extension of Theorem 1 of [47] to r-graphs (suggested at the end of [47]), which (with [32]) would give asymptotics of the threshold for $\mathcal{H}^r_{n,p}$ to contain a K^r_s-factor (where K^r_s is the complete r graph on s vertices).

Each of Theorems 3.7.1 and 3.7.2 begins with the following easy observations. (The first, an approximate converse of Proposition 3.1.4, is the trivial direction of LP duality.)

Observation 3.7.4. *If an increasing $\mathcal{F}$ supports a q-spread measure, then $q_f(\mathcal{F}) < q$.*

(More precisely, $q_f(\mathcal{F})$ is the least q such that $\mathcal{F}$ supports a probability measure ν with $\nu(\langle S\rangle) \le 2q^{|S|}$ $\forall S$.)

Observation 3.7.5. *Uniform measure on $\mathcal{G}_H$ is q-spread if and only if: for $S \subseteq \mathcal{K}^r_n$ isomorphic to a subhypergraph of H, σ a uniformly random permutation of V and $H_0 \subseteq \mathcal{K}^r_n$ a given copy of H,*

$$\mathbb{P}(\sigma(S) \subseteq H_0) \le q^{|S|}. \tag{3.26}$$

Proving Theorem 3.7.1 is now just a matter of verifying (3.26) with $q = O(dn^{-(r-1)})$, which we leave to the reader. (It is similar to the proof of (3.28).)

Proof of Theorem 3.7.2. The next assertion is the main thing we need to check here.

Lemma 3.7.6. *There is $\varepsilon = \varepsilon_d > 0$ such that if H is as in Theorem 3.7.2 and has no component isomorphic to K_{d+1}, then*

$$q_f(\mathcal{F}_H) \le n^{-(2/(d+1)+\varepsilon)} =: q. \tag{3.27}$$

Proof. We just need to show (3.26) for q as in (3.27) and S, H_0 as in Observation 3.7.5, say with $W = V(S)$, $s = |S|$, and f the size of a spanning forest of S. We may of course assume S has no isolated vertices, so $w := |W| \le 2f$. We show

$$\mathbb{P}(\sigma(S) \subseteq H_0) < (e^2 d/n)^f \tag{3.28}$$

and

$$\frac{f}{s} \ge \frac{2(d+1)}{(d+2)d} = \frac{2}{d+1} + \varepsilon_0, \tag{3.29}$$

where $\varepsilon_0 = 1/((d+2)(d+1)d)$, implying that for any $\varepsilon < \varepsilon_0$, (3.26) holds for large enough n.

Proof of (3.28). Let $\alpha, \beta : W \to V$ be, respectively, a uniform injection and a uniform map. Then

$$\begin{aligned}(d/n)^f \ge \mathbb{P}(\beta(S) \subseteq H_0) \ &\ge\ \mathbb{P}(\beta \text{ is injective})\mathbb{P}(\beta(S) \subseteq H_0 | \beta \text{ is injective}) \\ = (n)_w n^{-w}\mathbb{P}(\alpha(S) \subseteq H_0) \ &>\ e^{-2f}\mathbb{P}(\sigma(S) \subseteq H_0).\end{aligned}$$

□

Proof of (3.29). We may of course assume S is connected, in which case we have $f = w - 1$ and upper bounds on s: $\binom{w}{2}$ if $w \le d$; $\binom{d+1}{2} - 1$ if $w = d+1$; and $wd/2$ if $w \ge d+2$. The corresponding lower bounds on f/s are $2/d$, $2d/((d+2)(d+1)-2)$ and $2(d+1)/((d+2)d)$, the smallest of which is the last. □

This completes the proof of Lemma 3.7.6. □

We are now ready for Theorem 3.7.2. Let $\varsigma = \varsigma_n$ be some slow $o(1)$ (e.g. $1/\ln n$). By Theorem 3.7.3 there is $p_1 \sim p^*(d,n)$ such that if $(d+1)\,|\,m > (1-\varsigma)n$ then G_{m,p_1}

contains a K_{d+1}-factor w.h.p., while by Lemma 3.7.6 and Theorem 1.0.1 (or, more precisely, Remark 3.2.2), there is p_2 with $p^*(d,n) \gg p_2 \gg n^{-(2/(d+1)+\varepsilon)}$ such that if $m \geq \varsigma n$ then for any given m-vertex H' with $\Delta(H') \leq d$, G_{m,p_2} contains (a copy of) H' w.h.p.

Let H_1 be the union of the copies of K_{d+1} in H (each of which must be a component of H), $H_2 = H - H_1$, and $n_i = |V(H_i)|$ (so $n_1 + n_2 = n$). Let $G_1 \sim G_{n,p_1}$ and $G_2 \sim G_{n,p_2}$ be independent on the common vertex set $V = [n]$ and $G = G_1 \cup G_2$. Then $G \sim G_{n,p}$ with $p = 1-(1-p_1)(1-p_2) \sim p^*(d,n)$, and we just need to show $G \supseteq H$ w.h.p. In fact we find each H_i in the corresponding G_i, in order depending on n_2: if $n_2 \geq \varsigma n$, then w.h.p. G_1 contains H_1, say on vertex set V_1, and w.h.p. $G_2[V \setminus V_1]$ contains H_2; and if $n_2 < \varsigma n$, then w.h.p. G_2 contains H_2 on some V_2, and w.h.p. $G_1[V \setminus V_2]$ contains H_1. $\square$

3.8 Concluding remarks

In closing we briefly mention (or recall) a few unresolved issues related to the present work.

A. First, of course, it would be nice to prove Conjecture 3.1.3, which now implies Conjecture 3.1.1. In Chapter 4 we present some progress towards this result.

B. It would be interesting to understand whether, in Shamir's and related problems, the $\ln \ell$ emerging from our argument somehow reflects the coupon-collector requirement (edges cover vertices) that drives the lower bounds. Partly as a way of testing this, one might try to see if the present machinery can be extended to apply directly (rather than *via* [47, 27]) to questions where coupon collector considerations (correctly) predict a smaller gap, as in the fractional powers of $\ln n$ in Theorem 3.7.3.

C. The arguments of [44] and [18] give stronger "universality" results; e.g. [44] says that the appropriate $G_{n,p}$ w.h.p. contains *every* tree respecting the degree bound. Whether this can be proved along present lines remains unclear; if so, it would seem to be more a question of managing *some* understanding of the class of universal graphs (with, of course, a view to the spread) than of extending Theorem 1.0.1.

D. As mentioned following Corollary 3.1.7, what prevents us from extending to other values of the dimension k is inadequate control of the spread. (Here it doesn't really matter whether we think of "assignments" or of the threshold for containing a member of the $\mathcal{H}$ in (3.5).) The difficulty is the same for the related problem of thresholds for existence of designs. We don't have anything to suggest in the way of a remedy and just indicate one issue, for simplicity sticking to Steiner triple systems (STS's; see [55] for background); thus $X = \mathcal{K}_n^3$ (with $n \equiv 1$ or $3 \pmod 6$), $\mathcal{H}$ is the hypergraph of STS's, and for the spread (which should be $\Theta(1/n)$), we may take

$$\kappa = \min_{S \subseteq X} \left(|\mathcal{H}|/|\mathcal{H} \cap \langle S \rangle|\right)^{1/|S|}. \tag{3.30}$$

Results of Linial and Luria [38] (upper bound) and Keevash [35] (lower bound) give

$$|\mathcal{H}| = ((1+o(1))n/e^2)^{n^2/6}. \tag{3.31}$$

Viewed enumeratively this is very satisfactory, having been an old conjecture of Wilson [56]. But for present purposes, even ignoring our weaker understanding of $|\mathcal{H} \cap \langle S \rangle|$ (the number of completions of a partial STS S), it is not enough: even if this quantity is, as one expects, roughly $(n/e^2)^{n^2/6-|S|}$, the r.h.s. of (3.30) can be dominated by the "error" factor $(1+o(1))^{n^2/(6|S|)}$ if S is slightly small and the $o(1)$ in (3.31) is negative.

E. Finally, we recall a related conjecture from [33] (stated there only for graphs, but this shouldn't matter). For $\mathcal{F} = \mathcal{F}_H$ as in Section 3.7, let $p_{\mathsf{E}}(\mathcal{F})$ be the least p such that for every $H' \subseteq H$ the expected number of (unlabeled) copies of H' in $\mathcal{H}^r_{n,p}$ is at least 1. Then $p_{\mathsf{E}}(\mathcal{F})/2$ is again a trivial lower bound on $p_c(\mathcal{F})$—and, where it makes sense, probably more intuitive than $q(\mathcal{F})$ or $q_f(\mathcal{F})$—and from [33, Conjecture 2] we have:

Conjecture 3.8.1. *There is a universal K such that for every $\mathcal{F} = \mathcal{F}_H$ as above,*

$$p_c(\mathcal{F}) \le K p_{\mathsf{E}}(\mathcal{F}) \ln |X|.$$

Again, we can presumably replace $\ln |X|$ by $\ln |H|$, as would now follow from a positive answer to the obvious question: do we always have $q_f(\mathcal{F}) = O(p_{\mathsf{E}}(\mathcal{F}))$?

Chapter 4

On a problem of M. Talagrand

Joint work with Jeff Kahn and Jinyoung Park.

4.1 Introduction

The main concern of this chapter is the relation between the following two notions of M. Talagrand [50, 51, 53]. (Our focus is Conjecture 4.1.4 and our main result is Theorem 4.1.6; we will come to these following some motivation.)

We assume throughout that $\mathcal{F} \subseteq 2^V$ is increasing and not equal to $2^V, \emptyset$. Say $\mathcal{F}$ is *p-small* if there is a $\mathcal{G} \subseteq 2^V$ such that

$$\langle \mathcal{G} \rangle \supseteq \mathcal{F} \tag{4.1}$$

(that is, each member of $\mathcal{F}$ contains a member of $\mathcal{G}$) and

$$\sum_{S \in \mathcal{G}} p^{|S|} \leq 1/2, \tag{4.2}$$

and set $q(\mathcal{F}) = \max\{p : \mathcal{F} \text{ is } p\text{-small}\}$. Say $\mathcal{F}$ is *weakly p-small* if there is a $\lambda : 2^V \setminus \{\emptyset\} \to [0,1]$ such that

$$\sum_{S \subseteq F} \lambda_S \geq 1 \quad \forall F \in \mathcal{F} \tag{4.3}$$

and

$$\sum_{S} \lambda_S p^{|S|} \leq 1/2, \tag{4.4}$$

and set $q_f(\mathcal{F}) = \max\{p : \mathcal{F} \text{ is weakly } p\text{-small}\}$. As in Chapter 3, we refer to $q(\mathcal{F})$ and $q_f(\mathcal{F})$ (respectively) as the *expectation-threshold* and *fractional expectation-threshold* of $\mathcal{F}$. Notice that

$$q(\mathcal{F}) \leq q_f(\mathcal{F}) \leq p_c(\mathcal{F}). \tag{4.5}$$

Here the first inequality is trivial and the second holds since for λ as in (4.3) and (4.4),

$$\mu_p(\mathcal{F}) \le \sum_{F\in\mathcal{F}} \mu_p(F) \sum_{S\subseteq F} \lambda_S \le \sum_S \lambda_S \mu_p(\langle S\rangle) = \sum_S \lambda_S p^{|S|} \le 1/2 \qquad (4.6)$$

(where $\langle S\rangle = \langle \{S\}\rangle$).

Thus each of q, q_f is a lower bound on p_c (and one that's easily understood in many cases of interest; see [21]). The next two conjectures—respectively the main conjecture (Conjecture 1) of [33] and a sort of LP relaxation thereof suggested by Talagrand [53, Conjecture 8.3]—say that these bounds are never far from the truth.

Conjecture 4.1.1. *There is a universal K such that for every finite V and increasing $\mathcal{F} \subseteq 2^V$,*

$$p_c(\mathcal{F}) < Kq(\mathcal{F}) \ln |V|.$$

Conjecture 4.1.2. *There is a universal K such that for every finite V and increasing $\mathcal{F} \subseteq 2^V$,*

$$p_c(\mathcal{F}) < Kq_f(\mathcal{F}) \ln |V|.$$

Talagrand [53, Conjecture 8.5] also proposes the following strengthening of Conjecture 4.1.2, in which $\ell(\mathcal{F})$ is the maximum size of a minimal member of $\mathcal{F}$.

Conjecture 4.1.3. *There is a universal K such that for every finite V and increasing $\mathcal{F} \subseteq 2^V$,*

$$p_c(\mathcal{F}) < Kq_f(\mathcal{F}) \ln \ell(\mathcal{F}).$$

Conjecture 4.1.3 is shown in Chapter 3, to which we also refer for discussion of the very strong consequences that originally motivated Conjecture 4.1.1, but follow just as easily from Conjecture 4.1.2.

Here we are interested in the following conjecture of M. Talagrand [53, Conjecture 6.3], which says that the parameters q and q_f are in fact not very different.

Conjecture 4.1.4. *There is a fixed L such that, for any $\mathcal{F}$, $q(\mathcal{F}) \ge q_f(\mathcal{F})/L$.*

(In other words, weakly p-small implies (p/L)-small.) This of course implies *equivalence of* Conjectures 4.1.2 and 4.1.1, as well as of Conjecture 4.1.3 and the corresponding strengthening of Conjecture 4.1.1, so in particular, in view of [21], would now supply a proof of Conjecture 4.1.1. (At present this implication is probably the strongest motivation for Conjecture 4.1.4, but we have long been interested in the conjecture for its own sake.)

The following mild reformulation of Conjecture 4.1.4 will be convenient.

Conjecture 4.1.5. *There is a fixed J such that for any V, p and $\lambda : 2^V \setminus \{\emptyset\} \to \mathbb{R}^+$,*

$$\{A \subseteq V : \sum_{S \subseteq A} \lambda_S \geq \sum_S \lambda_S (Jp)^{|S|}\}$$

is p-small.

As Talagrand observes, even simple instances of Conjecture 4.1.4 are not easy to establish. He suggests two test cases, which in the formulation of Conjecture 4.1.5 become:

(i) $V = \binom{[n]}{2} = E(K_n)$ and (for some k) λ is the indicator of {copies of K_k in K_n};

(ii) λ is supported on 2-element sets.

(He does show that Conjecture 4.1.4 holds if λ is supported on singletons; see Proposition 4.2.1 for a quantified version that will be useful in what follows.)

The very specific (i) above was treated in [8]. Here we dispose of the less structured (ii):

Theorem 4.1.6. *Conjecture 4.1.5 holds when* $\text{supp}(\lambda) \subseteq \binom{V}{2}$*; in other words, there is a J such that for any graph $G = (V, E)$, $p \in [0, 1]$ and $\lambda : E \to \mathbb{R}^+$,*

$$\{U \subseteq V : \lambda(G[U]) \geq J^2 \lambda(G) p^2\}$$

is p-small (where $G[U]$ is the subgraph induced by U).

It seems not out of the question that the ideas involved in proving Theorem 4.1.6 can be extended to give Conjecture 4.1.4 in full, but we don't yet see this.

The rest of this chapter is devoted to the proof of Theorem 4.1.6. The most important part of this turns out to be (a quantified version of) the "unweighted" case, where

λ takes values in $\{0,1\}$, though deriving Theorem 4.1.6 from this still needs some ideas. A precise statement is given at the end of Section 4.2, following a few preliminaries, and the unweighted and weighted arguments are then given in Sections 4.3 and 4.4 respectively.

4.2 Framework

We use $[n]$ for $\{1,2,\ldots,n\}$, 2^X for the power set of X, and $\binom{X}{r}$ for the family of r-element subsets of X, and recall from above that $\langle\mathcal{A}\rangle$ is the increasing family generated by $\mathcal{A}\subseteq 2^X$. For a set X and $p\in[0,1]$, X_p is the "p-random" subset of X in which each $x\in X$ appears with probability p independent of other choices. We assume throughout that p has been specified and usually omit it from our notation.

Graphs here are always simple, and are mainly thought of as sets of edges; thus $|G|$ is $|E(G)|$. We use $\nabla_G(v)$ or ∇_v for $\{e\in E(G): v\in e\}$; so the degree of v is $d_v=|\nabla_v|$. (We also use $N_G(v)$ for the neighborhood of v in G.)

The following convention will be helpful. Given a graph G on V, we associate with each $U\subseteq V$ a "weighted subset" $D(U)=D_G(U)$ of $E(G)$ that assigns to each e the weight $|e\cap U|/2$. (We also use D_v or $D_G(v)$ for $D(\{v\})$.) We then have, for any $\lambda: G\to\mathbb{R}^+$,

$$\lambda(D(U))=\tfrac{1}{2}\textstyle\sum_{v\in U}\lambda(\nabla_v)$$

(e.g. $|D(U)|:=\frac{1}{2}\sum_{v\in U}d_v$). To see why this is natural, notice that

$$\mathbb{E}\lambda(G[V_p])=\mathbb{E}\lambda(D(V_p))p$$

(e.g. $\mathbb{E}|G[V_p]|=\mathbb{E}|D(V_p)|p$), so that $\lambda(D(U))p$ is a natural benchmark against which to measure $\lambda(G[U])$.

For $\mathcal{A}\subseteq 2^V$, the *cost* of $\mathcal{A}$ (w.r.t. our given p) is $\mathrm{C}(\mathcal{A})=\sum_{S\in\mathcal{A}}p^{|S|}$. We say $\mathcal{A}$ *witnesses* $\mathcal{B}\subseteq 2^V$ if $\langle\mathcal{A}\rangle\supseteq\mathcal{B}$; set

$$\mathrm{C}^*(\mathcal{B})=\min\{\mathrm{C}(\mathcal{A}):\mathcal{A}\text{ witnesses }\mathcal{B}\},$$

and say $\mathcal{B}$ is *witnessable at cost* γ if $\mathrm{C}^*(\mathcal{B})\le\gamma$. Talagrand's observation that Conjecture 4.1.4 holds for λ supported on singletons may now be stated as:

Proposition 4.2.1. *For all* $\zeta : V \to \mathbb{R}^+$ *and* $J > 2e$,

$$\mathrm{C}^*(\{U \subseteq V : \zeta(U) \geq J\zeta(V)p\}) < 2e/(J-2e). \tag{4.7}$$

(The dependence on J is best possible up to constants.)

Proof. We may take $V = [n]$ and assume ζ is non-increasing (and positive) and $Jp \leq 1$ (since the proof is trivial when $Jp > 1$). Define R by

$$\frac{1}{Rp} = \lceil \frac{1}{Jp} \rceil =: a.$$

We claim that the collection

$$\mathcal{A} = \bigcup_{k \geq 1} \binom{[ak]}{k}$$

witnesses the family in (4.7); this gives the proposition since the l.h.s. of (4.7) is then at most

$$\mathrm{C}(\mathcal{A}) = \sum_{k \geq 1} \binom{ak}{k} p^k < \sum_{k \geq 1} \left(\frac{e}{R}\right)^k < \frac{e}{R-e} < \frac{2e}{J-2e}$$

(the last inequality holding since $Jp \leq 1$ implies $R > J/2$.)

To see that the claim holds, observe that its failure implies the existence of some $U = \{u_1 < u_2 < \cdots < u_\ell\} \subseteq [n]$ with $\zeta(U) \geq J\zeta(V)p$ such that $|U \cap [ak]| < k$ for all $k > 0$. But then $u_i > ia$ for all $i \in [\ell]$, yielding the contradiction

$$\zeta(V) > \sum_{i=0}^{\ell-1} \sum_{j \in [a]} \zeta(j + ia) \geq a\zeta(U) \geq \zeta(V). \qquad \square$$

We further define

$$\mathrm{C}^*_J(\mu, T)$$

to be the infimum of those γ's for which, for every p and (simple) graph G (on V) with $|G|p^2 \leq \mu$,

$$\{U \subseteq V : |G[U]| \geq \max\{T, J|D_G(U)|p\}\} \tag{4.8}$$

is witnessable at cost γ. In Section 4.4 we will need cost bounds that improve as T grows, even if T/μ does not, and this need not be the case without the extra condition involving $|D_G(U)|$ in (4.8). License to use this condition will be provided in Section 4.4 via the reduction of Theorem 4.1.6 to the following unweighted statement, which we regard as our main point.

Theorem 4.2.2. *For any* μ *and* $T = cJ^2\mu$ *with*

$$c \geq 256e/J \text{ and } J \geq 8e, \tag{4.9}$$

and $J_1 = J/(8e)$,

$$\mathrm{C}^*_J(\mu, T) \leq 32c^{-1} \min\{J_1^{-2}, J_1^{-\sqrt{T}/16}\}. \tag{4.10}$$

(Here and throughout we don't worry about getting good constants, and try instead to keep the argument fairly clean.)

4.3 Proof of Theorem 4.2.2

Aiming for simplicity, we first bound the cost in (4.10) assuming

$$T = 2^{2k+3}$$

for some positive integer k and

$$c = T/(\mu J^2) \geq 64e/J, \tag{4.11}$$

showing that in this case

$$\mathrm{C}^*_J(\mu, T) \leq 8c^{-1} J_1^{-2^{k-1}-1}. \tag{4.12}$$

Before proving this, we show that it implies Theorem 4.2.2, which, since $\mathrm{C}^*_J(\mu, t)$ is decreasing in t, just requires showing that the r.h.s. of (4.10) bounds $\mathrm{C}^*_J(\mu, T_0)$ for some $T_0 \leq T$. If $T < 32$ this follows from the trivial

$$\mathrm{C}^*_J(\mu, 1) \leq \mu \tag{4.13}$$

(take $\mathcal{G} = \{x, y\} : xy \in G\}$), since $\mu = T/(cJ^2) < 32c^{-1}J_1^{-2}$, matching the bound in (4.10). Suppose then that $T \geq 32$ and let $T_0 = c_0 J^2 \mu$ be the largest integer not greater than T of the form 2^{2k+3} (with positive integer k). We then have $c_0 > c/4$ (supporting (4.11)) and $2^{k-1} > \sqrt{T_0}/8 > \sqrt{T}/16$, and it follows that the bound on $\mathrm{C}^*_J(\mu, T_0)$ given by (4.12) is less than the bound in (4.10).

Proof of (4.12). We have $T = 2^{2k+3}$ ($= cJ^2\mu$ with J as in (4.9) and c as in (4.11)), and, with

$$\mathcal{U} := \{U \subseteq V : |G[U]| > \max\{T, J|D_G(U)|p\}\}, \tag{4.14}$$

want to show that $\mathrm{C}^*(\mathcal{U})$ is no more than the bound in (4.12).

A basic challenge for Conjecture 4.1.4 in general is identifying a suitable $\mathcal{G}$. In the present instance, each member of $\mathcal{G}$ will be a disjoint union of stars, where for present purposes a *star at v in W* ($\subseteq V$) is some $\{v\} \cup S \subseteq W$ with $S \subseteq N_G(v)$. (Where convenient we will also refer to this as the "star (v, S).") We say such a star is *good* if

$$|S| \geq Jd_v p/4. \tag{4.15}$$

Given L, we define

$$L^v = \max\{L, Jd_v p/4\} \tag{4.16}$$

and say a star $(v,\, S)$ is *L-special* if $|S| = L^v$.

For positive integers b and L, let $\mathcal{G}(b, L)$ ($\subseteq 2^V$) consist of all disjoint unions of b L-special stars in G. We will specify a particular collection $\mathcal{C}$ of pairs (b, L) and set

$$\mathcal{G} = \cup\{\mathcal{G}(b, L) : (b, L) \in \mathcal{C}\}.$$

Theorem 4.2.2 is then given by the following two assertions.

Claim 4.3.1. *$\mathcal{G}$ witnesses $\mathcal{U}$.*

Claim 4.3.2. $\mathrm{C}(\mathcal{G})$ *is at most the bound in* (4.12).

Set (with $i \in [k]$ throughout) $L_i = 2^{i-1}$ and

$$\delta_i = \max\{2^{-(i+2)}, 2^{i-k-3}\} \geq 1/(8L_i), \tag{4.17}$$

and notice that

$$\sum \delta_i \leq \sum 2^{-(i+2)} + \sum 2^{i-k-3} \leq 1/2. \tag{4.18}$$

Let

$$b_i = \delta_i 4^{-i} T \geq 2^{k-i}. \tag{4.19}$$

Finally, set

$$\mathcal{C} = \{(b_i, L_i) : i \in [k]\}.$$

Proof of Claim 4.3.1. We are given $U \in \mathcal{U}$ and must show it contains a member of $\mathcal{G}$. Let $U_0 = U$ and for $j = 1, \ldots$ until no longer possible do: let (v_j, S_j), with $S_j = N_G(v_j) \cap U_{j-1}$, be a largest good star in U_{j-1}, and set $d_j = |S_j|$ and $U_j = U_{j-1} \setminus (\{v_j\} \cup S_j)$.

The passage from U_{j-1} to U_j deletes at most d_j^2 edges containing vertices of S_j of U_{j-1}-degree at most d_j; any other edge deleted in this step contains $u \in S_j$ with U_{j-1}-degree less than $Jd_u p/4$ (or u, having U_{j-1}-degree greater than d_j, would have been chosen in place of v_j); and of course each vertex u of the final U_j has U_j-degree less than $Jd_u p/4$. We thus have

$$|G[U]| \leq \sum_j d_j^2 + \sum_{v \in U} Jd_v p/4 \leq \sum_j d_j^2 + |G[U]|/2$$

(using the second bound in (4.14)), so

$$\sum_j d_j^2 \geq T/2. \tag{4.20}$$

Set

$$B_i = \begin{cases} \{j : d_j \in [2^{i-1}, 2^i)\} & \text{if } i \in [k-1], \\ \{j : d_j \geq 2^{k-1}\} & \text{if } i = k. \end{cases}$$

(It may be worth noting that, while the d_j's are decreasing, the degrees corresponding to B_i increase with i.) In view of (4.20), either $|B_k| \geq 1$ or

$$\sum_{i \in [k-1]} |B_i| 4^i \geq T/2 \geq \sum_{i \in [k-1]} \delta_i T$$

(using (4.18)). Recalling that $b_k = 1$, it follows that for some $i \in [k]$ we have

$$|B_i| \geq b_i. \tag{4.21}$$

On the other hand, since $|S_j| \geq L_i^v$ for $j \in B_i$, the set $\bigcup\{S_j \cup \{v_j\} : j \in B_i\}$ contains some $W \in \mathcal{G}(b_i, L_i) (\subseteq \mathcal{G})$ whenever i is as in (4.21). This completes the proof of Claim 4.3.1. $\square$

Proof of Claim 4.3.2. We first bound the cost, say $\mathrm{C}(b, L)$, of the collection $\mathcal{G}(b, L)$. Set

$$q_v = p \left(\frac{ed_v p}{L^v} \right)^{L^v}.$$

Then q_v bounds the total cost of the set of L-special stars at v (as $\binom{d_v}{L^v} \le (ed_v/L^v)^{L^v}$), and it follows that

$$\mathrm{C}(b,L) \le \sum \Big\{ \prod_{v\in B} q_v : B \in \tbinom{V}{b} \Big\}. \tag{4.22}$$

For a given value of $\varphi := \sum_{v\in V} q_v$, the r.h.s. of (4.22) is largest when the q_v's are all equal (this just uses $xy \le [(x+y)/2]^2$), whence

$$\mathrm{C}(b,L) \le \binom{|V|}{b} \left(\frac{\varphi}{|V|}\right)^b \le \left(\frac{e\varphi}{b}\right)^b. \tag{4.23}$$

Recalling (4.16), we have

$$q_v \le d_v p^2 \cdot \frac{e}{L} \left(\frac{4e}{J}\right)^{L-1},$$

so

$$\varphi \le 2\mu \cdot \frac{e}{L} \left(\frac{4e}{J}\right)^{L-1}. \tag{4.24}$$

Now using (4.23) and (4.24), recalling that $T = cJ^2\mu$, $L_i = 2^{i-1}$, $b_i = \delta_i 4^{-i} T = \delta_i T/(4L_i^2)$ and $J_1 = J/(8e)$, and for the moment omitting the subscript i, we have (with the final inequality (4.25) justified below)

$$\begin{aligned} \mathrm{C}(b,L) &\le \left[\frac{2e^2\mu}{L}\frac{4L^2}{\delta T}\left(\frac{4e}{J}\right)^{L-1}\right]^b \\ &= \left[8e^2 L \cdot \frac{1}{cJ^2\delta}\left(\frac{4e}{J}\right)^{L-1}\right]^b \\ &= \left[c^{-1}\frac{L}{2\delta}\left(\frac{4e}{J}\right)^{L+1}\right]^b \\ &\le \left[\frac{c}{4}\cdot J_1^{L+1}\right]^{-b}. \end{aligned} \tag{4.25}$$

For (4.25), or the equivalent

$$2^{L+4}\delta \ge L, \tag{4.26}$$

it is enough to show $2^{L+1} \ge L^2$ (since $\delta \ge 1/(8L)$; see (4.17)), which is true for positive integer L.

Finally, returning to Claim 4.3.2 (and recalling that L and b in the display ending with (4.25) are really L_i and b_i), we have

$$\mathrm{C}(\mathcal{G}) = \sum_{i=1}^{k} \mathrm{C}(b_i, L_i) \le \sum_{i=1}^{k} \left[\frac{c}{4}\cdot J_1^{L_i+1}\right]^{-b_i}. \tag{4.27}$$

Since (4.11) implies that $cJ_1/4 \geq 2$, we can use $b_i \geq 2^{k-i}$ from (4.19) to bound the r.h.s. of (4.27) by

$$\begin{aligned}\sum_{i=1}^{k}\left[\frac{cJ_1^{2^{i-1}+1}}{4}\right]^{-2^{k-i}} &= \sum_{i=1}^{k} J_1^{-2^{k-1}}\left[\frac{cJ_1}{4}\right]^{-2^{k-i}} \\ &= \sum_{j=0}^{k-1}\left(\frac{c}{4}J_1^{2^{k-1}+1}\right)^{-1}\left[\frac{cJ_1}{4}\right]^{1-2^j} \\ &< 8c^{-1}J_1^{-2^{k-1}-1},\end{aligned}$$

matching (4.12) as desired. □

4.4 proof of Theorem 4.1.6

We prove the following quantified version of Theorem 4.1.6.

Theorem 4.4.1. *For any graph G on V, $\lambda : G \to \mathbb{R}^+$ and*

$$R \geq 4096\sqrt{2}e, \tag{4.28}$$

the set

$$\mathcal{U}_0 = \{U \subseteq V : \lambda(G[U]) \geq R^2\lambda(G)p^2\}$$

is witnessable at cost $O(1/R)$.

Proof. We take G, λ, R to be as in the theorem, use $D(U)$ for $D_G(U)$ (defined in Section 4.2), and assume throughout that

$$U \in \mathcal{U}_0.$$

We first observe that it is enough to prove the theorem assuming

$$\lambda \text{ takes only values } \theta_i := 2^{-i},\ \ i = 1, 2, \ldots\ \ , \tag{4.29}$$

with (4.28) slightly weakened to

$$R \geq 4096e. \tag{4.30}$$

Then for a general λ (which we may of course scale to take values in $[0, 1]$) and λ' given by

$$\lambda'_S = \max\{\theta_i : \theta_i \leq \lambda_S\},$$

$\mathcal{U}_0$ as in the theorem is contained in the corresponding collection with λ and R^2 replaced by λ' and $R^2/2$ (which supports (4.30)), since $U \in \mathcal{U}_0$ implies $2\lambda'(G[U]) > \lambda(G[U]) \geq R^2\lambda(G)p^2 \geq R^2\lambda'(G)p^2$. So we assume from now on that λ and R are as in (4.29) and (4.30) (respectively).

Note also that Proposition 4.2.1, with $\zeta(v) = \lambda(D_v)$ (for which we have $\zeta(V) = \sum \zeta(v) = \frac{1}{2}\sum \lambda(\nabla_v) = \lambda(G)$ and $\zeta(U) = \lambda(D(U))$), says that the set

$$\{U \subseteq V : \lambda(D(U)) \geq R\lambda(G)p\}$$

admits a witness of cost less than $6/R$. So we specify such a witness as a first installment on $\mathcal{G}$ and it then becomes enough to show that

$$\mathcal{U}^* := \{U \in \mathcal{U}_0 : \lambda(D(U)) < R\lambda(G)p\}$$

can be witnessed at cost $O(1/R)$; in fact we will show

$$\mathrm{C}^*(\mathcal{U}^*) = O(R^{-2}). \tag{4.31}$$

Set $G_i = \{e \in G : \lambda(e) = \theta_i\}$ and write $D_i(U)$ for $D_{G_i}(U)$. We then observe, for any $H \subseteq G$,

$$\lambda(H) = \sum_i \theta_i |H \cap G_i|,$$

and abbreviate

$$\mathsf{w}_i = \lambda(G_i) = \theta_i|G_i|, \quad \mathsf{w} = \lambda(G) = \sum \mathsf{w}_i.$$

Given U, define $L = L(U)$, $K = K(U)$, $L_i = L_i(U)$ and $K_i = K_i(U)$ by

$$\begin{aligned} \lambda(D(U)) &= L\mathsf{w}p, \\ \lambda(G[U]) &= KL\mathsf{w}p^2, \\ |D_i(U)| &= L_i|G_i|p, \end{aligned} \tag{4.32}$$

and

$$|G_i[U]| = K_iL_i|G_i|p^2. \tag{4.33}$$

Then

$$L\mathsf{w}p = \sum \theta_i|D_i(U)| = \sum L_i\mathsf{w}_ip \tag{4.34}$$

and

$$KL\mathsf{w}p^2 = \sum \theta_i |G_i[U]| = \sum K_i L_i \mathsf{w}_i p^2.$$

Since $U \in \mathcal{U}_0$, we have

$$\sum K_i L_i \mathsf{w}_i \geq R^2 \mathsf{w}, \tag{4.35}$$

while $U \in \mathcal{U}^*$ gives

$$L < R. \tag{4.36}$$

Note also that, with

$$I = I(U) = \{i : K_i > R/2\},$$

we have

$$\sum \{K_i L_i \mathsf{w}_i : i \in I\} > R^2 \mathsf{w}/2, \tag{4.37}$$

as follows from (4.35) and (using (4.34) and (4.36))

$$\sum \{K_i L_i \mathsf{w}_i : i \notin I\} \leq (R/2)L\mathsf{w} < R^2 \mathsf{w}/2.$$

Now let $\mathsf{E}_i = |G_i|p^2$ and, for integer α,

$$\mathcal{E}_\alpha = \{i : \mathsf{E}_i \in (2^{\alpha-1}, 2^\alpha]\}.$$

We arrange the i's in an array, with columns indexed by α's (in increasing order) and column α consisting of the indices in $\mathcal{E}_\alpha$, again in increasing order. (So w_i's within a column *decrease* as we go down. Note column lengths may vary.) Define $\mathcal{B}_\beta$ to be the set of indices in row β.

	$\cdots$ $\alpha - 1$	α	$\alpha + 1$ $\cdots$
1			
$\vdots$			
β		i	
$\vdots$			

Table 4.1: i is the βth smallest index in $\mathcal{E}_\alpha$ (when $|\mathcal{E}_\alpha| \geq \beta$).

Set $\mathsf{y}_i = \theta_i 2^\alpha / p^2$ (for $i \in \mathcal{E}_\alpha$) and $\mathsf{y} = \sum_{i \geq 1} \mathsf{y}_i$, noting that

$$\mathsf{y}_i/2 < \mathsf{w}_i \leq \mathsf{y}_i.$$

Set

$$c^*_\beta = (3/2)^{\beta-1}R^2/16 \quad (\beta \geq 1)$$

and $c_i = c^*_\beta$ if $i \in \mathcal{B}_\beta$. Let w^*_β and y^*_β be (respectively) the sums of the w_i's and y_i's over $i \in \mathcal{B}_\beta$, and notice that

$$\mathsf{y}^*_{\beta+1} \leq \mathsf{y}^*_\beta/2 \quad \text{for } \beta \geq 1$$

(since $i = \mathcal{B}_{\beta+1} \cap \mathcal{E}_\alpha$ (where we abusively use i for $\{i\}$) implies $i > j := \mathcal{B}_\beta \cap \mathcal{E}_\alpha$, whence $2\mathsf{y}_i \leq \mathsf{y}_j$).

Claim 4.4.2. *For each $U \in \mathcal{U}^*$ there is an $i \in I(U)$ with $K_i(U)L_i(U) > c_i$.*

Proof. With $\sum^\star$ denoting summation over I, we have (using (4.37) at the end)

$$\begin{aligned}\sum\nolimits^\star c_i\mathsf{w}_i \leq \sum c^*_\beta \mathsf{w}^*_\beta \leq \sum c^*_\beta \mathsf{y}^*_\beta &\\ &\leq \mathsf{y}^*_1(c^*_1 + c^*_2/2 + c^*_3/2^2 + \cdots) \\ &\leq \mathsf{y}(c^*_1 + c^*_2/2 + c^*_3/2^2 + \cdots) \\ &\leq (R^2/4)\mathsf{y} < (R^2/2)\mathsf{w} < \sum\nolimits^\star K_i(U)L_i(U)\mathsf{w}_i.\end{aligned}$$

□

It follows that if, for each i, $\mathcal{G}_i$ witnesses

$$\mathcal{U}_i := \{U \subseteq V : i \in I(U);\ K_i(U)L_i(U) > c_i\},$$

then $\cup\mathcal{G}_i$ witnesses $\mathcal{U}^*$; so we have

$$\mathrm{C}^*(\mathcal{U}^*) \leq \sum_i \mathrm{C}^*(\mathcal{U}_i). \tag{4.38}$$

On the other hand, if (α, β) is the pair corresponding to i (that is, i is the βth entry in column α of our array), then (see (4.8) for C^*_J)

$$\mathrm{C}^*(\mathcal{U}_i) \leq \mathrm{C}^*_{R/2}(2^\alpha, T_{\alpha,\beta}), \tag{4.39}$$

where $T_{\alpha,\beta} = \max\{c^*_\beta 2^{\alpha-1}, 1\}$; namely, $|G_i|p^2 = \mathsf{E}_i \leq 2^\alpha$, while $U \in \mathcal{U}_i$ implies (using (4.32), (4.33) and $i \in I(U)$)

$$|G_i[U]| = K_i(U)L_i(U)|G_i|p^2 \begin{cases} > c_i|G_i|p^2 > c^*_\beta 2^{\alpha-1} \\ = K_i|D_i(U)|p > (R/2)|D_i(U)|p. \end{cases}$$

So it is enough to show that the sum of the r.h.s. of (4.39) (over $\beta \geq 1$ and integer α) is $O(R^{-2})$.

For $T_{\alpha,\beta} = 1$ we bound the r.h.s. of (4.39) by the trivial (4.13), which—since $T_{\alpha,\beta} = 1$ iff $2^\alpha \leq 32R^{-2}(2/3)^{\beta-1}$—bounds the contribution of such pairs (to the sum in (4.38)) by

$$\sum_{\beta\geq 1} \sum_{\{\alpha: 2^\alpha \leq 2/c^*_\beta\}} 2^\alpha \leq 64R^{-2} \sum_{\beta\geq 1} (2/3)^{\beta-1} = 3 \cdot 64R^{-2}. \tag{4.40}$$

For $T_{\alpha,\beta} > 1$ we use Theorem 4.2.2 with $T = T_{\alpha,\beta} (= 2^{\alpha-1}c^*_\beta)$, $\mu = 2^\alpha$, $J = R/2$, and (thus)

$$c = T/(\mu J^2) = c^*_\beta/(2J^2) = (3/2)^{\beta-1}/8.$$

Note that (4.30) gives $J \geq 8e$ and $c \geq 256e/J$, so (4.9) holds.

For each integer $s \geq 0$ let $\mathcal{T}_s = \{(\alpha,\beta) : T_{\alpha,\beta} \in (2^s, 2^{s+1}]\}$. For each $\beta \geq 1$ there is a unique α such that $(\alpha,\beta) \in \mathcal{T}_s$, and every (α,β) with $T_{\alpha,\beta} > 1$ is in some $\mathcal{T}_s$. Let $f(s) = \min\{J_1^{-2}, J_1^{-2^{s/2-4}}\}$. Then for fixed s, we have (see (4.10))

$$\sum_{(\alpha,\beta)\in\mathcal{T}_s} \mathrm{C}^*_J(2^\alpha, T_{\alpha,\beta}) \leq \sum_{\beta\geq 1} 32c^{-1} f(s) = \sum_{\beta\geq 1} 256 \left(\frac{2}{3}\right)^{\beta-1} f(s) < 3 \cdot 256 f(s), \tag{4.41}$$

and summing over all s we get

$$\sum_{T_{\alpha,\beta}>1} \mathrm{C}^*_J(2^\alpha, T_{\alpha,\beta}) < \sum_{s\geq 0} 768 f(s) = \sum_{s\geq 0} 768 \min\{J_1^{-2}, J_1^{-2^{s/2-4}}\} = O(J_1^{-2}). \tag{4.42}$$

Finally, combining (4.42) and (4.40) gives (4.31). $\square$

Chapter 5

Automorphisms of Induced Subgraphs of $G_{n,p}$

Keith Frankston

5.1 Organization

We introduce our core machinery in Section 5.2 (leaving a couple small details to Section 5.5) and use it to prove a slight weakening of Theorem 1.0.3 as an orienting example. In Section 5.3 we prove our main result, Theorem 5.3.1, as well as the more general, but weaker, Theorem 5.3.2. Finally, in Section 5.4 we present some directions for future study, including possibilities for strengthening our current results.

5.2 Machinery

Conventions. Throughout this chapter, $G = G_{n,p}$ is taken to have vertex set $V = [n]$ unless otherwise specified; *all subgraphs are induced*, with $G[W]$ denoting the induced subgraph of G on $W \subseteq V$; and we view a graph H on V as a subset of $\binom{V}{2}$, so $H[W] = H \cap \binom{W}{2}$.

We say φ is a *subautomorphism* of G if $\varphi \in \mathrm{Aut}(G[W])$ for some $W \subseteq V$; φ is said to be of *size* $|\varphi| := |W|$.

We denote a partition of ℓ by $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r) \vdash \ell$ (meaning $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ are integers with $\sum_i \lambda_i = \ell$). We use *subpartition* of X to mean a collection $\mathcal{S} = \{S\}$ of disjoint subsets of X, *each of size at least* 2 and say $\mathcal{S}$ is of *type* $\lambda \vdash \ell$ if $\{\lambda_i\} = \{|S|\}_{S\in\mathcal{S}}$ as multisets. We use the shorthand $\cup\mathcal{S} = \cup_{S\in\mathcal{S}} S$ and $\ell(\mathcal{S}) = |\cup\mathcal{S}|$.

W.h.p. statements and asymptotic notation are with respect to n; so $f = o(g)$ and $f = \omega(g)$ (also written $f \ll g$ and $f \gg g$) mean f/g and g/f respectively are smaller than any given $\varepsilon > 0$ for large enough n.

For $\varphi \in \mathfrak{S}_X$ (the symmetric group on X), we use $\mathrm{Fix}(\varphi) = \{x \in X : \varphi(x) = x\}$ for the set of *fixed points* of φ, and $\mathcal{O}(\varphi) = \{\,\{\varphi^k(x) : k \in \mathbb{Z}\} : x \in X \setminus \mathrm{Fix}(\varphi)\}$ for the set of non-trivial orbits of φ. We refer to elements of $\cup\mathcal{O}(\varphi) = X \setminus \mathrm{Fix}(\varphi)$ as the *mobile points* of φ. We say φ is of type $\lambda \vdash \ell$ if $\mathcal{O}(\varphi)$ is of type λ (note that the type of φ depends only on its mobile points).

For a permutation φ of V, we denote by φ' the permutation it induces on $\binom{V}{2}$. Observe that φ induces a graph automorphism on $H \subseteq \binom{V}{2}$ iff $H \cap T \in \{\emptyset, T\}$ for every $T \in \mathcal{O}(\varphi')$. If $|T| = k$, we have

$$\mathbb{P}[G \cap T \in \{\emptyset, T\}] = p^k + (1-p)^k =: q_p(k) \tag{5.1}$$

(recall $G = G_{n,p}$; we will write q for q_p when p is known), with these events occurring independently for the various $T \in \mathcal{O}(\varphi')$ (since the orbits partition $\binom{V}{2}$).

Note that for $\varphi \in \mathrm{Aut}(H)$, $S \in \mathcal{O}(\varphi)$, and $x \in \mathrm{Fix}(\varphi)$,

$$\{\{x, y\} : y \in S\} \in \mathcal{O}(\varphi')$$

and therefore

$$N_H(x) \cap S \in \{\emptyset, S\} \tag{5.2}$$

(where $N_H(x) = \{y : \{x, y\} \in H\}$ denotes the neighborhood of x in H).

For $S \subseteq V$, we write

$$\mathrm{co}_G(S) = \mathrm{co}(S) = \{x \in V \setminus S : N_G(x) \cap S \in \{\emptyset, S\}\,\}, \tag{5.3}$$

and, more generally, for a subpartition $\mathcal{S}$ (of V),

$$\mathrm{co}(\mathcal{S}) = \bigcap_{S \in \mathcal{S}} \mathrm{co}(S). \tag{5.4}$$

Notice that φ is a subautomorphism of G iff

$$\mathrm{co}(\mathcal{O}(\varphi)) \supseteq \mathrm{Fix}(\varphi) \tag{5.5}$$

and $\varphi|_{\cup\mathcal{O}(\varphi)}$ is an automorphism of $G[\cup\mathcal{O}(\varphi)]$.

Hence, for a given subpartition $\mathcal{S}$ (of V), if φ is a subautomorphism of G with $\mathcal{O}(\varphi) = \mathcal{S}$, then

$$|\varphi| \leq |\mathrm{co}(\mathcal{S})| + \ell(\mathcal{S})$$

and, clearly, any $\sigma \in \mathrm{Aut}(G[\cup\mathcal{S}])$ with $\mathcal{O}(\sigma) = \mathcal{S}$ extends to an automorphism of $G[\mathrm{co}(\mathcal{S}) \cup \bigcup \mathcal{S}]$.

Therefore, G admits a size t subautomorphism of type $\lambda \vdash \ell$ iff there is some *derangement* (permutation with no fixed points) σ with $\mathcal{S} := \mathcal{O}(\sigma)$ a subpartition of type λ, satisfying:

$$\begin{aligned} &G \in \mathcal{A}(\mathcal{S}, t) := \{H : |\mathrm{co}_H(\mathcal{S})| \geq t - \ell(\mathcal{S})\};\ \text{and} \\ &G \in \mathcal{B}(\sigma) \quad := \{H : \sigma \text{ is a subautomorphism of } H\}. \end{aligned} \tag{5.6}$$

We would like to bound, for a given derangement σ, the probability that both these conditions hold. To do so, we bound their probabilities separately and use the fact that they are independent (since they are determined by $G \cap A$ and $G \cap B$ respectively for some disjoint $A, B \subseteq \binom{V}{2}$).

We first observe that $|\mathrm{co}(\mathcal{S})|$ is distributed binomially. For $S \subset V$ and $x \in V \setminus S$, we have

$$\mathbb{P}[x \in \mathrm{co}(S)] = q(S) := q(|S|) \left(= p^{|S|} + (1-p)^{|S|}\right) \tag{5.7}$$

and, more generally, for a subpartition $\mathcal{S}$ of type $\lambda \vdash \ell$ and $x \notin \cup\mathcal{S}$,

$$\begin{aligned} \mathbb{P}[x \in \mathrm{co}(\mathcal{S})] = q(\mathcal{S}) &:= \prod_{S \in \mathcal{S}} q(S) \\ &= \prod \left(p^{\lambda_i} + (1-p)^{\lambda_i}\right) =: q(\lambda) = q. \end{aligned} \tag{5.8}$$

Therefore, since $|V \setminus \cup\mathcal{S}| = n - \ell$,

$$|\mathrm{co}(\mathcal{S})| \sim \mathrm{Bin}(n - \ell, q). \tag{5.9}$$

This will allow us to bound

$$\mathbb{P}[G \in \mathcal{A}(\mathcal{S}, t)] = \mathbb{P}[\mathrm{Bin}(n - \ell, q) \geq t - \ell] \tag{5.10}$$

using a tail bound.

We then bound $\mathbb{P}[G \in \mathcal{B}(\sigma)]$ for a derangement σ. Let σ be a derangement with $\mathcal{O}(\sigma) = \mathcal{S} = \{S_i\}$ of type $\lambda \vdash \ell$ (where $|S_i| = \lambda_i$). By (5.1), we have

$$\mathbb{P}[G \in \mathcal{B}(\sigma)] = \mathbb{P}[\,\sigma \in \mathrm{Aut}(G[\cup\mathcal{S}])\,] = q(\mathcal{O}(\sigma')).$$

So to bound $\mathbb{P}[G \in \mathcal{B}(\sigma)]$, we want to understand the structure of $\mathcal{O}(\sigma')$ in terms of the $\lambda_i's$.

We observe that $\binom{S_i}{2}$ is partitioned in $\mathcal{O}(\sigma')$ according to cyclic distance. Let $S_i = \{v_1, \ldots, v_{\lambda_i}\}$ with $\sigma(v_i) = v_{i+1}$ (indices taken modulo λ_i). Then

$$E_d := \{\,\{v_k, v_{k+d}\} : k \in [\lambda_i]\} \in \mathcal{O}(\sigma')$$

for each $1 \le d \le \lambda/2$. Note that $|E_d| = \lambda_i$ for $d < \lambda/2$ and, if λ_i is even, $\left|E_{\lambda_i/2}\right| = \lambda_i/2$. Thus $\sqcup E_d = \binom{S_i}{2}$ is a partition into $\lfloor \frac{\lambda_i - 1}{2} \rfloor$ blocks of size λ_i, plus a block of size $\lambda_i/2$ when λ_i is even.

For any $i \ne j$, $\mathcal{O}(\sigma')$ partitions $\{\,\{x,y\} : x \in S_i,\, y \in S_j\}$ into (λ_i, λ_j) blocks of size $[\lambda_i, \lambda_j]$ (where $(\cdot,\cdot)$ and $[\cdot,\cdot]$ denote greatest common divisor and least common multiple respectively).

Therefore

$$q(\mathcal{O}(\sigma')) = \prod_i g(\lambda_i) \prod_j q([\lambda_i, \lambda_j])^{(\lambda_i,\lambda_j)/2} \tag{5.11}$$

where

$$g(s) = \begin{cases} \frac{q(s/2)}{q(s)} & \text{if } s \text{ is even,} \\ \frac{1}{\sqrt{q(s)}} & \text{if } s \text{ is odd.} \end{cases}$$

For any $s \ge 2$ and $r > 1$,

$$q(s \cdot r) = p^{sr} + (1-p)^{sr} < (p^s + (1-p)^s)^r = q(s)^r \tag{5.12}$$

(see 5.5.1).

Applying this to (5.11) with $r = \frac{[\lambda_i, \lambda_j]}{\lambda_i} \geq 1$, we get

$$\begin{aligned} q(\mathcal{O}(\sigma')) &\leq \prod_i g(\lambda_i) \prod_j q(\lambda_i)^{\lambda_j/2} \\ &= \prod_i g(\lambda_i) q(\lambda_i)^{\ell/2} \\ &\leq \prod_i q(\lambda_i)^{\ell/2-1} \\ &= q(\lambda)^{\frac{\ell-2}{2}} \end{aligned} \tag{5.13}$$

(since $g(s) \leq 1/q(s)$). Note that we have equality iff $\lambda_i = 2$ for all i (in which case $q(\mathcal{O}(\sigma')) = (1 - 2p(1-p))^{\ell(\ell-2)/4}$).

In fact, for any $s \geq 2$, (5.12) implies

$$q(s) \leq q(2)^{s/2},$$

and thus

$$q(\lambda) = \prod_i q(\lambda_i) \leq \prod q(2)^{\lambda_i/2} = \prod (1 - 2p(1-p))^{\lambda_i/2} = (1 - 2p(1-p))^{\ell/2}. \tag{5.14}$$

Combining (5.14) with (5.13), we get

$$\mathbb{P}[\, \sigma \in \mathrm{Aut}(G[\cup \mathcal{S}]) \,] \leq (1 - 2p(1-p))^{\frac{\ell(\ell-2)}{4}}. \tag{5.15}$$

Similarly, we use (5.11) to produce a lower bound for $\mathbb{P}[G \in \mathcal{B}(\sigma)]$:

$$q(\mathcal{O}(\sigma'))^2 \geq \prod_i g(\lambda_i)^2 \prod_j q(\lambda_i \lambda_j) \tag{5.16}$$

$$> \prod_i q(\lambda_i \ell) \tag{5.17}$$

$$> \prod_i (1-p)^{\lambda_i \ell} \left(= (1-p)^{\ell^2}\right)$$

$$\geq \prod_i q(\lambda_i)^{2\ell} \tag{5.18}$$

$$= q(\lambda)^{2\ell},$$

where (5.16) and (5.17) are implied by the trivial inequality $q(a)q(b) > q(ab)$, and (5.18) holds since $(1-p)^s \geq q(s)^2$ for $p \leq 1/2$ and $s \geq 2$ (see 5.5.2). (We may opt to use the lower bound $(1-p)^{\ell^2}$ when an expression in terms of p is useful.)

Putting everything together, we get the following two-sided bound for a fixed derangement σ of type $\lambda \vdash \ell$:

$$q(\lambda)^{\ell} \leq \mathbb{P}[G \in \mathcal{B}(\sigma)] \leq q(\lambda)^{\frac{\ell-2}{2}}. \tag{5.19}$$

As an easy application, we show the following slightly weaker version of Theorem 1.0.3.

Lemma 5.2.1. *For all* $1/2 \geq p = \frac{2\ln n + \omega(1)}{n}$*, w.h.p.* $G = G_{n,p}$ *is rigid.*

Proof. Let $p = \frac{2(\ln n + c)}{n}$ for $c = c(n) = \omega(1)$. Given $\mathcal{S}$ of type $\lambda \vdash \ell$, we can crudely bound

$$q(\mathcal{S}) = \prod_i q(\lambda_i) \leq q(2)^{\frac{\ell}{2}} \tag{5.20}$$

(see (5.14)), implying

$$\mathbb{P}[\,|\mathrm{co}(\mathcal{S})| = n - \ell\,] \leq (1 - 2p(1-p))^{\frac{\ell}{2}(n-\ell)}. \tag{5.21}$$

We can bound the number of derangements on ℓ vertices by $\binom{n}{\ell}\ell! < n^{\ell}$. Taking a union bound over all such derangements and combining (5.15) with (5.21), this gives

$$\begin{aligned} n^{\ell}(1-2p(1-p))^{\frac{\ell}{2}\left[(n-\ell)+\left(\frac{\ell-2}{2}\right)\right]} &\leq \exp\left[\ell\left(\ln n - p(1-p)(n-\ell/2-1)\right)\right] \\ &\leq \exp\left[\ell\left(\ln n - \frac{2(\ln n + c)}{n}\cdot\frac{n}{2}\right)\right] \\ &= \exp[-c\ell] \end{aligned} \tag{5.22}$$

(since $p(1-p)$ is increasing for $p \leq 1/2$).

Therefore the probability that G is rigid is

$$1 - \sum_{\ell \geq 2} \exp[-c\ell] = 1 - o(1). \qquad \square$$

This result is essentially best possible with the present approach, as

$$\mathbb{E}\,|\mathrm{Aut}(G)| > \mathbb{P}[G = \emptyset]\cdot n! = \exp[(n(\ln n - pn/2 - O(1))] \gg 1$$

for $p = \frac{2\ln n - \omega(1)}{n}$.

5.3 Main Results

We start with an essentially tight result in the special case $p = 1/2$. In this case, we define the *weight* of φ, a subautomorphism of type $\lambda \vdash \ell$, to be

$$\mathsf{w}(\varphi) = \mathsf{w}(\lambda) := \sum_i (\lambda_i - 1) \tag{5.23}$$

(i.e. the number of mobile points minus the number of non-trivial orbits) and observe that

$$q(\lambda) = 2^{-\mathsf{w}(\lambda)}. \tag{5.24}$$

Theorem 5.3.1. *W.h.p. every subautomorphism φ of $G = G_{n,1/2}$ satisfies $\mathsf{w}(\varphi) \leq (2 + o(1)) \log n$ and $|\varphi| \leq 2^{-\mathsf{w}} n + \sqrt{5n \log n}$.*

Proof. We will actually show that w.h.p. every φ satisfies

$$\mathsf{w}(\varphi) \leq 2 \log n + 5$$

and

$$|\varphi| < 2^{-\mathsf{w}(\varphi)} n + \Delta(n, \mathsf{w}(\varphi)), \tag{5.25}$$

where

$$\Delta(n, \mathsf{w}) = \max\{2\mathsf{w}(\log n - \mathsf{w}/2 + \log\log n),\ 3\sqrt{\mathsf{w} 2^{-\mathsf{w}} n \log n}\}. \tag{5.26}$$

Since there are fewer than n^ℓ derangements on ℓ vertices, it suffices to show, for every derangement σ of size ℓ, that the probability that σ extends to a subautomorphism of size $t = t(n, \mathsf{w}(\sigma)) := 2^{-\mathsf{w}(\sigma)} n + \Delta(n, \mathsf{w}(\sigma))$ is less than $\delta(\ell) n^{-\ell}$ where $\sum_{\ell \geq 2} \delta(\ell) = o(1)$ (here we take $\delta(\ell) = 2^{-\ell}/\log n$ for ease).

Given $\lambda \vdash \ell$, we set $\mathsf{w} = \mathsf{w}(\lambda)$, $\Delta = \Delta(n, \mathsf{w})$ and $t = t(n, \mathsf{w})$ and note that

$$\ell/2 \leq \mathsf{w} < \ell. \tag{5.27}$$

Thus, given a derangement σ with $\mathcal{S} = \mathcal{O}(\sigma)$ of type $\lambda \vdash \ell \geq 2$, we want to show (recall (5.6))

$$\mathbb{P}[G \in \mathcal{A}(\mathcal{S}, t)] \cdot \mathbb{P}[G \in \mathcal{B}(\sigma)] \cdot (2n)^\ell \log n \leq 1.$$

Combining (5.19) with (5.24) gives

$$\mathbb{P}[G \in \mathcal{B}(\sigma)] \leq 2^{-\mathsf{w}\frac{\ell-2}{2}}. \tag{5.28}$$

Since $\mathsf{w} < \ell$ (see (5.27)), we have

$$-\mathsf{w}\frac{\ell-2}{2} = -\ell\frac{\mathsf{w}}{2} + \mathsf{w} < -\ell(\frac{\mathsf{w}}{2} - 1).$$

Therefore

$$\begin{aligned} \log\left[\mathbb{P}[G \in \mathcal{B}(\sigma)] \cdot (2n)^{\ell} \log n\right] &\leq \log\left[2^{-\mathsf{w}\frac{\ell-2}{2}}(2n)^{\ell} \log n\right] \\ &= \ell(\log n + 1) - \mathsf{w}(\ell/2 - 1) + \log\log n \\ &< \ell(\log n + 2 - \mathsf{w}/2 + \log\log n/\ell), \end{aligned} \tag{5.29}$$

which is less than 0 for $\mathsf{w} > 2\log n + 5$. Hence, when $\mathsf{w} > 2\log n + 5$,

$$\mathbb{P}[G \in \mathcal{B}(\sigma)] \cdot (2n)^{\ell} \log n \leq 1,$$

so w.h.p. every subautomorphism of G has weight at most $2\log n + 5$.

We may therefore assume $\mathsf{w} \leq 2\log n + 5$, so (using $\ell \leq 2\mathsf{w}$ and (5.26))

$$\ell/\Delta \leq 1/(\log n - \mathsf{w}/2 + \log\log n) < 2/\log\log n =: \varepsilon(n) \ll 1.$$

Thus (see (5.10))

$$\begin{aligned} \mathbb{P}[G \in \mathcal{A}(\mathcal{S}, t)] &= \mathbb{P}[|\mathrm{co}(\mathcal{S})| \geq t - \ell] \\ &= \mathbb{P}[\mathrm{Bin}(n-\ell, q) \geq nq + \Delta - \ell] \\ &\leq \mathbb{P}[\mathrm{Bin}(n, q) \geq nq + (1-\varepsilon)\Delta]. \end{aligned} \tag{5.30}$$

We use Chernoff's inequality [30, Theorem 2.1],

$$\mathbb{P}[\mathrm{Bin}(n, q) \geq nq + \Delta] \leq \exp\left[\frac{-\Delta^2}{2(nq + \Delta/3)}\right], \tag{5.31}$$

the r.h.s. of which we bound by

$$B(n, q, \Delta) := \max\{\exp\left[\frac{-\Delta^2}{4nq}\right], \exp\left[\frac{-3\Delta}{4}\right]\}. \tag{5.32}$$

Using the two alternatives for Δ (see (5.26)), we show that either expression in (5.32) is strong enough to imply

$$B(n, q, (1-\varepsilon)\Delta) \cdot 2^{-\mathsf{w}\frac{\ell-2}{2}}(2n)^{\ell} \log n \leq 1.$$

We use $\Delta \geq 2\mathsf{w}(\log n - \mathsf{w}/2 + \log\log n)$ to show

$$\exp\left[\frac{-3(1-\varepsilon)\Delta}{4}\right] 2^{-\mathsf{w}\frac{\ell-2}{2}}(2n)^{\ell}\log n \leq 1,$$

which is equivalent to each of

$$\left(\frac{-3(1-\varepsilon)\Delta\log e}{4}\right) + \log\left[2^{-\mathsf{w}\frac{\ell-2}{2}}(2n)^{\ell}\log n\right] \leq 0$$

and

$$\frac{4}{3\log e}\log\left[2^{-\mathsf{w}\frac{\ell-2}{2}}(2n)^{\ell}\log n\right] \leq (1-\varepsilon)\Delta. \tag{5.33}$$

Using (5.29), we have

$$\begin{aligned}\frac{4}{3\log e}\log\left[2^{-\mathsf{w}\frac{\ell-2}{2}}(2n)^{\ell}\log n\right] &< \frac{4}{3\log e}\ell(\log n + 2 - \mathsf{w}/2 + \log\log n/\ell)\\ &< (1-\varepsilon)2\mathsf{w}(\log n - \mathsf{w}/2 + \log\log n) \qquad (5.34)\\ &\leq (1-\varepsilon)\Delta\end{aligned}$$

(where (5.34) holds since $(1-\varepsilon) > 4/(3\log e)$, $2\mathsf{w} \geq \ell$, and $\log\log n > 2 + \log\log n/\ell$), giving (5.33) as desired.

Similarly, we use $\Delta \geq 3\sqrt{\mathsf{w}2^{-\mathsf{w}}n\log n}$ to show

$$\exp\left[\frac{-(1-\varepsilon)^2\Delta^2}{4nq}\right](2n)^{\ell}\log n \leq 1,$$

which is implied by each of

$$\left(\frac{-(1-\varepsilon)^2\Delta^2\log e}{4nq}\right) + \log\left[(2n)^{\ell}\log n\right] \leq 0$$

and

$$\sqrt{\frac{4n2^{-\mathsf{w}}}{\log e}\ell\left(\log n + 1 + \log\log n/\ell\right)} \leq (1-\varepsilon)\Delta. \tag{5.35}$$

Since $\log n + 1 + \log\log n < (1+\varepsilon)\log n$ (very crudely) and $\ell \leq 2\mathsf{w}$,

$$\begin{aligned}\sqrt{\frac{4}{\log e}\ell 2^{-\mathsf{w}}n\left(\log n + 1 + \log\log n/\ell\right)} &< \sqrt{\frac{8(1+\varepsilon)}{\log e}\mathsf{w}2^{-\mathsf{w}}n\log n}\\ &< 3(1-\varepsilon)\sqrt{\mathsf{w}2^{-\mathsf{w}}n\log n} \qquad (5.36)\\ &\leq (1-\varepsilon)\Delta\end{aligned}$$

(where (5.36) holds since $3(1-\varepsilon) > \sqrt{8(1+\varepsilon)/\log e}$), giving (5.35). □

A couple remarks regarding Theorem 5.3.1:

- A relatively straightforward argument shows that Δ is essentially tight for fixed ℓ. In particular, for fixed $\lambda \vdash \ell$, there exists $C = C(\lambda)$ such that w.h.p. G admits a subautomorphism φ of type λ with $|\varphi| > n2^{-\mathsf{w}} + \Delta/C$. (Note $\Delta/C = \Theta(\sqrt{n \log n})$.)
- Similarly, we get $|\varphi| \leq (1+o(1))2\log^2 n$ for $\mathsf{w}(\varphi) \approx \log n$, which is off by at most $\log n$ since w.h.p. $\omega(G) \sim 2\log n$ (where $\omega(G)$ is the size of a largest clique in G).

We close this section with a weaker result for general $p = p(n)$.

Theorem 5.3.2. *Fix $\varepsilon > 0$ and let $G = G_{n,p}$ with $1/2 \geq p \gg \frac{\ln^2 n}{n}$. W.h.p.*

$$|\varphi| < (q(\mathcal{O}(\varphi)) + \varepsilon)n \tag{5.37}$$

for every subautomorphism φ of G.

Proof. We may assume $p > \frac{20\ln^2 n}{\varepsilon^2 n}$. We want to show that w.h.p. no φ violating (5.37) is a subautomorphism of G. It suffices to show, for any derangement σ on ℓ vertices and $t = t(n,\sigma) := (q(\mathcal{O}(\sigma)) + \varepsilon)n$, that $\mathbb{P}[G \in \mathcal{B}(\sigma)] \cdot \mathbb{P}[G \in \mathcal{A}(\mathcal{O}(\sigma), t)]$ is less than $\delta(\ell)n^{-\ell}$ for $\sum_{\ell \geq 2} \delta(\ell) = o(1)$ (here $\delta(\ell) = n^{-\ell}$ for simplicity).

We split into two cases depending on ℓ. If $\ell p > 10\ln n$, then (5.15) gives

$$\mathbb{P}[G \in \mathcal{B}(\sigma)] \leq \exp\left[-\ell \binom{\ell-2}{2} p(1-p)\right] < \exp\left[-\ell^2 p/5\right] < n^{-2\ell}.$$

We are left with cases where $\ell \leq 10\ln n/p < \varepsilon^2 n/(2\ln n)$. In such cases, we use

$$\mathbb{P}[G \in \mathcal{A}(\mathcal{O}(\sigma), t)] = \mathbb{P}[\mathrm{Bin}(n-\ell, q) \geq (q+\varepsilon)n - \ell] \tag{5.38}$$

(see (5.10)). We apply Hoeffding's inequality [28]

$$\mathbb{P}[\mathrm{Bin}(n,q) \geq k] \leq \exp\left[-2\frac{(k-nq)^2}{n}\right]$$

to bound (5.38) by

$$\exp\left[-2\frac{(\varepsilon n - (1-q)\ell)^2}{n}\right] < \exp[-\varepsilon^2 n] < n^{-2\ell}$$

(since $2\ell \ln n < \varepsilon^2 n$). □

We give the following as an immediate corollary.

Corollary 5.3.3. *For any $\varepsilon, p > 0$, w.h.p. $G_{n,p}$ is locally $((1-2p(1-p)+\varepsilon)n)$-rigid.*

Proof of Corollary 5.3.3. For p as in Theorem 5.3.2, we simply observe that

$$q(\mathcal{O}(\varphi)) \leq (1-2p(1-p))$$

for any non-trivial permutation φ. If $p = O(\frac{\ln^2 n}{n})$, then $n(1-2p(1-p)+\varepsilon) > n$ and the statement holds vacuously. □

5.4 Further Questions

A. While Theorem 5.3.2 holds vacuously for $p \ll 1/n$ (since then $q_p(\lambda) \geq q(n) = 1-o(1)$, so $(q+\varepsilon)n > n$), it is unclear what can be said for $\Omega(1/n) = p = O(\ln^2 n/n)$.

B. It would be interesting to extend our results to $G_{n,d}$ in analogy with Kim-Sudakov-Vu [36]. There has also been work on the rigidity of Preferential Attachment graphs (see e.g. [41]) and it would be interesting to see if an analogous result regarding local rigidity is possibly there.

C. It seems likely that existing results on the asymmetry of $G_{n,p}$ (e.g. Theorem 1.0.3 as well as the much stronger theorem on the rigidity of the 2-core of G due to Linial and Mosheiff [39]) can be extended to the following "hitting time" result.

Conjecture 5.4.1. *Let $e_1, \ldots, e_N$ be a uniformly random ordering of the pairs $\binom{V}{2}$ (where $|V| = n$ and $N = \binom{n}{2}$) and $G_t = \{e_1, \ldots, e_t\}$. Let*

$$T_0 = \min\{t : G_t \text{ has at most one isolated vertex}\}; \text{ and}$$

$$T_1 = \max\{t : \overline{G_t} \text{ has at most one isolated vertex}\}.$$

Then w.h.p. G_t is rigid for all $t \in [T_0, T_1]$.

Clearly G_t is non-rigid for all $t < T_0$. Although G_{T_0} is non-rigid when it has no isolates (as this implies the exposure process went from 2 isolates to 0 isolates at step T_0, so G_{T_0} contains vertices x, y adjacent only to each other), this occurs with probability $O(1/n)$. (Of course, the same comments apply to G_{T_1}.)

While our machinery seems promising for showing G_{T_0} is rigid, it seems not so easy to couple G_t and $G_{n,p}$ (for $p = t/N$) in a manner which respects rigidity (as rigidity is a highly non-monotone property).

5.5 Minor Inequalities

5.5.1

We want to show

$$p^{sr} + (1-p)^{sr} < (p^s + (1-p)^s)^r$$

for $s, r \geq 1$. It suffices to show

$$\frac{(a+b)^r}{a^r + b^r} = \frac{(1+b/a)^r}{1+(b/a)^r} = \frac{(1+\gamma)^r}{1+\gamma^r} > 1$$

for $\gamma = b/a > 0$ and $r > 1$. Taking derivatives of top and bottom w.r.t. γ and comparing, we get $r(1+\gamma)^{r-1} > r\gamma^{r-1}$ and are done as the ratio is 1 for $\gamma = 0$.

5.5.2

We want to show

$$(1-p)^{s/2} \geq p^s + (1-p)^s$$

for $s \geq 2$ and $0 \leq p \leq 1/2$. Setting $s = 2r$, this is equivalent to showing

$$1 \geq \left(\frac{p^2}{1-p}\right)^r + (1-p)^r \tag{5.39}$$

for $r \geq 1$. At $r = 1$, this becomes

$$\begin{aligned}(1-p) &\geq p^2 + (1-p)^2 \\ p(1-2p) &\geq 0,\end{aligned}$$

which holds for $p \leq 1/2$. Thus it suffices to show that the derivative of the r.h.s. of (5.39) with respect to r is negative:

$$\frac{\partial}{\partial r}\left(\frac{p^2}{1-p}\right)^r + (1-p)^r = \left(\frac{p^2}{1-p}\right)^r \ln\left(\frac{p^2}{1-p}\right) + (1-p)^r \ln(1-p) < 0.$$

This clearly holds since $p^2/(1-p) < 1$.

Bibliography

[1] R. Ahlswede and L. H. Khachatrian, *The complete intersection theorem for systems of finite sets*, European J. Combin. **18** (1997), 125–136.

[2] N. Alon and Z. Füredi, *Spanning subgraphs of random graphs*, Graphs and Combinatorics **8** (1992), no. 1, 91–94.

[3] R. Alweiss, S. Lovett, K. Wu, and J. Zhang, *Improved bounds for the sunflower lemma*, Preprint, arXiv:1908.08483.

[4] B. Bollobás, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

[5] B. Bollobás and A. Thomason, *Threshold functions*, Combinatorica **7** (1987), 35–38.

[6] L. Brick, P. Gao, and A. Southwell, *The threshold of symmetry in random graphs with specified degree sequences*, Preprint, arXiv:2004.01794, 2020.

[7] P. J. Cameron, P. Frankl, and W. M. Kantor, *Intersecting families of finite sets and fixed-point-free* 2*-elements*, European J. Combin. **10** (1989), 149–160.

[8] B. DeMarco and J. Kahn, *Note on a problem of M. Talagrand*, Random Structures & Algorithms **47** (2015), no. 4, 663–668.

[9] M. Deza and P. Frankl, *Erdős-Ko-Rado theorem*—22 *years later*, SIAM J. Algebraic Discrete Methods **4** (1983), 419–431.

[10] A. Dudek, A. Frieze, P.-S. Loh, and S. Speiss, *Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs*, the electronic journal of combinatorics (2012), P44–P44.

[11] D. Ellis, G. Kalai, and B. Narayanan, *On symmetric intersecting families*, European Journal of Combinatorics **86** (2020), 103094.

[12] D. Ellis and B. Narayanan, *On symmetric 3-wise intersecting families*, Proc. Amer. Math. Soc. **145** (2017), 2843–2847.

[13] P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, Quart. J. Math. Oxford **12** (1961), 313–320.

[14] P. Erdős and R. Rado, *Intersection theorems for systems of sets*, J. London Math. Soc. **35** (1960), 85–90.

[15] P. Erdős and A. Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61.

[16] P. Erdős and A. Rényi, *Asymmetric graphs*, Acta Mathematica Hungarica **14** (1963), no. 3-4, 295–315.

[17] A. Ferber, G. Kronenberg, and K. Luh, *Optimal threshold for a random graph to be 2-universal*, Transactions of the American Mathematical Society **372** (2019), no. 6, 4239–4262.

[18] A. Ferber, K. Luh, and O. Nguyen, *Embedding large graphs into a random graph*, Bulletin of the London Mathematical Society **49** (2017), no. 5, 784–797.

[19] P. Frankl, *Regularity conditions and intersecting hypergraphs*, Proc. Amer. Math. Soc. **82** (1981), 309–311.

[20] K. Frankston, J. Kahn, and B. Narayanan, *On regular 3-wise intersecting families*, Proceedings of the American Mathematical Society **146** (2018), no. 10, 4091–4097.

[21] K. Frankston, J. Kahn, B. Narayanan, and J. Park, *Thresholds versus fractional expectation-thresholds*, Preprint, arXiv:1908.08483.

[22] E. Friedgut, *Boolean functions with low average sensitivity depend on few coordinates*, Combinatorica **18** (1998), 27–35.

[23] ———, *Sharp thresholds of graph properties, and the k-sat problem*, J. Amer. Math. Soc. **12** (1999), 1017–1054, With an appendix by J. Bourgain.

[24] ———, *Hunting for sharp thresholds*, Random Structures Algorithms **26** (2005), 37–51.

[25] E. Friedgut and G. Kalai, *Every monotone graph property has a sharp threshold*, Proc. Amer. Math. Soc. **124** (1996), 2993–3002.

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

[27] A. Heckel, *Random triangles in random graphs*, Preprint, arXiv:1802.08472.

[28] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, The Collected Works of Wassily Hoeffding, Springer, 1994, pp. 409–426.

[29] F. Ihringer and A. Kupavskii, *Regular intersecting families*, Discrete Applied Mathematics **270** (2019), 142–152.

[30] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

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

[32] J. Kahn, *Asymptotics for Shamir's problem*, Preprint, arXiv:1909.06834.

[33] J. Kahn and G. Kalai, *Thresholds and expectation thresholds*, Combin. Probab. Comput. **16** (2007), 495–502.

[34] R. M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, New York.), 1972, pp. 85–103.

[35] P. Keevash, *Counting designs*, Preprint, arXiv:1504.02909.

[36] J. H. Kim, B. Sudakov, and V. H. Vu, *On the asymmetry of random regular graphs and random graphs*, Random Structures & Algorithms **21** (2002), no. 3-4, 216–224.

[37] M. Krivelevich, *Embedding spanning trees in random graphs*, SIAM Journal on Discrete Mathematics **24** (2010), no. 4, 1495–1500.

[38] N. Linial and Z. Luria, *An upper bound on the number of Steiner triple systems*, Preprint, arXiv:1108.5042, 2011.

[39] N. Linial and J. Mosheiff, *On the rigidity of sparse random graphs*, Journal of Graph Theory **85** (2017), no. 2, 466–480.

[40] N. Lord, *Binomial averages when the mean is an integer*, The Mathematical Gazette **94** (2010), no. 530, 331–332.

[41] A. Magner, S. Janson, G. Kollias, and W. Szpankowski, *On symmetry of uniform and preferential attachment graphs*, the electronic journal of combinatorics **21** (2014), no. 3.

[42] G. A. Margulis, *Probabilistic characteristics of graphs with large connectivity*, Problemy Peredači Informacii **10** (1974), 101–108.

[43] R. J. McEliece, *The theory of information and coding*, Addison-Wesley Publishing Co., Reading, MA, 1977.

[44] R. Montgomery, *Spanning trees in random graphs*, Adv. Math. **356** (2019).

[45] D. Mubayi and J. Verstraëte, *A survey of Turán problems for expansions*, Recent Trends in Combinatorics, Springer International Publishing, 2016, pp. 117–143.

[46] J. Nešetřil, *Graph theory and combinatorics*, Lecture Notes, Fields Institute, 2011.

[47] O. Riordan, *Random cliques in random graphs*, Preprint, arXiv:1802.01948.

[48] L. Russo, *An approximate zero-one law*, Z. Wahrsch. Verw. Gebiete **61** (1982), 129–139.

[49] M. Talagrand, *On Russo's approximate zero-one law*, Ann. Probab. **22** (1994), 1576–1587.

[50] ______, *Are all sets of positive measure essentially convex?*, Geometric Aspects of Functional Analysis, Springer, 1995, pp. 295–310.

[51] ______, *The generic chaining*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2005.

[52] ______, *Selector processes on classes of sets*, Probab. Theory Related Fields **135** (2006), 471–486.

[53] ______, *Are many small sets explicitly small?*, Proceedings of the 2010 ACM International Symposium on Theory of Computing, 2010, pp. 13–35.

[54] ______, *Upper and lower bounds for stochastic processes*, A Series of Modern Surveys in Mathematics, vol. 60, Springer, Heidelberg, 2014.

[55] J. H. van Lint and R. M. Wilson, *A course in combinatorics*, second ed., Cambridge University Press, Cambridge, 2001.

[56] R. M. Wilson, *Nonisomorphic Steiner triple systems*, Mathematische Zeitschrift **135** (1974), no. 4, 303–313.

[57] E. M. Wright, *Graphs on unlabelled nodes with a given number of edges*, Acta Mathematica **126** (1971), 1–9.

[58] ______, *Asymmetric and symmetric graphs*, Glasgow Mathematical Journal **15** (1974), no. 1, 69–73.