Methods in Extremal Combinatorics

Sam Spiro

February 5, 2023

Contents

Preface

This is a work in progress. There are typos, missing references, and so on scattered throughout. **Please let me know if you notices such errors, find anything confusing, or if you have any other suggestions!** If you prefer, you can let me know about any of this anonymously through this link.

The following is a set of lecture notes for a graduate level course in extremal combinatorics. These notes focus on standard methods that have been used to solve a large number of problems in extremal combinatorics. Throughout I assume basic knowledge of asymptotic analysis, probability theory, and linear algebra.

Due to the sheer scope of extremal combinatorics, there are many methods which I am not able to cover at all (and there is no topic which I am able to cover in complete depth). Below is a small list of methods and topics ***not*** covered by this text, as well as sources for a thorough treatment of the topics.

- Extremal Combinatorics in general: see books of Lovasz [45] or Bollobás [9]; surveys by Simonovits and Szemerédi [57] and Füredi and Simonovits [28]; and online courses by Morris and Gowers.

- Discrete geometry and the polynomial method: see the book by Sheffer [56], as well as the online minicourse on finite geometry and Ramsey theory by Bishnoi.

- Additive combinatorics and discrete Fourier analysis: see the book by Tao and Vu [61], as well as online courses by Prendiville and Zhao.

- Statistical mechanics: see notes by Will Perkins.

- The Linear algebra method: see the book by Babai and Frankl [4].

- The discharging method: see the survey by Cranston and West [16].

Part I

Basic Probabilistic Methods

This part is based heavily off of the book by Alon and Spencer [2] (which goes into much more depth on the topic), as well as lecture notes by Verstraëte.

1 Introduction

One of the most exciting developments in extremal combinatorics over the past century has been the incorporation of ideas and tools from probability theory into solving combinatorial problems. The first such use was by Erdős who proved an exponential lower bound for Ramsey numbers. We recall that the *Ramsey number* $R(s,t)$ is the smallest integer N such that any 2-coloring of the edges of K_N contains a monochromatic clique.

Theorem 1.1 ([20])**.** *For all n, we have*

$$R(n,n) \geq (1+o(1))\frac{n}{e\sqrt{2}}2^{n/2}.$$

This is essentially the best known lower bound (though we prove a slightly stronger bound in Theorem 3.3). The best known upper bound is roughly 4^n, so there's still quite a gap!

For this proof and throughout the text, we make heavy use of the union bound: if A, B are events in a probability space, then $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$. Often we will use an equivalent version: $\Pr[\overline{A} \cap \overline{B}] \leq 1 - \Pr[A] - \Pr[B]$, which follows from De Morgan's laws.

Proof. Let G be a ***random*** coloring of K_N with N to be determined later[1]. That is, for each edge of K_N, we independently and uniformly choose the edge to be colored either red or blue. The key observation is that if $\Pr[G \text{ contains no monochromatic } K_n] > 0$, then there exists a coloring of K_N with no monochromatic K_n (since otherwise the probability would be zero), proving the desired lower bound.

If S is a set of n vertices, we let A_S be the event that G contains a monochromatic K_n on S. With this we have

$$\Pr[G \text{ contains a monochromatic } K_n] = \Pr\Big[\bigcup_{S \in \binom{[N]}{n}} A_S\Big] \leq \sum_{S \in \binom{[N]}{n}} \Pr[A_S] = \binom{N}{n} \cdot 2^{1-\binom{n}{2}}.$$

If this quantity is less than 1, then we can conclude that $\Pr[G \text{ contains no monochromatic } K_n] > 0$, so our goal is to choose N as large as possible so that this happens. By using the bound $\binom{N}{n} \leq (eN/n)^n$ (which we will use many times throughout the text), we see that it suffices to have[2]

$$1 > (eN/n)2^{1-\binom{n}{2}} = 2(eN/n2^{(n-1)/2})^n.$$

Solving this shows that the desired bound holds if $N < 2^{1/n} \cdot \frac{n}{e\sqrt{2}}2^{n/2}$, proving the result[3]. □

[1] When trying to prove results in extremal and probabilistic combinatorics, one often uses a method that depends on some parameter such as N or p. Typically it is best to proceed through the argument without deciding what N, p is ahead of time, and only in the end do you optimize your parameter to give you the best bounds possible.

[2] Finding the "right" way to bound expressions like this takes time and practice. A reasonable strategy for these sorts of problems is try and get all of the main terms to have the same form (e.g. x^n in this example). Much more about the art of asymptotic analysis can be found in the book Asymptopia by Spencer [59].

[3] In fact, a closer analysis of this proof shows that asymptotically, almost every coloring of K_N with $N = (2-\epsilon)^{n/2}$ contains no monochromatic K_n. Despite almost every coloring working, we know of no explicit coloring that gives more than a polynomial lower bound for $R(n,n)$. Thus the probabilistic method gives us a way to find the hay in the haystack.

The proof of Theorem 1.1 implicitly used the following general principle, which is at the heart of the probabilistic method.

(*) Let T be an object chosen randomly from a set $\mathcal{T}$ (in some way) and P some property that objects in $\mathcal{T}$ could have. If $\Pr[T \text{ has property } P] > 0$, then there exists some $T' \in \mathcal{T}$ with this property.

We now turn to another classical extremal problem with a slick probabilistic proof. Recall that $\alpha(G)$ denotes the largest independent set of a graph G, i.e. the largest set of vertices I such that there exist no edge contained in I.

Theorem 1.2 (Caro-Wei Bound). *Let G be an n-vertex graph with degrees $d_1, \ldots, d_n$. Then*

$$\alpha(G) \geq \sum \frac{1}{d_i + 1}.$$

Moreover, equality holds if and only if G is a disjoint union of cliques.

Here and throughout the text we make heavy use of the principle of linearity of expectation: for two (possibly dependent) real-valued random variables, we have $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Proof. For π a bijection from $V(G)$ to $[n]$, we define

$$I(\pi) = \{v \in V(G) : \pi(v) < \pi(u)\ \forall u \in N(v)\}.$$

That is, $I(\pi)$ is the set of vertices which are smaller than all of their neighbors under π. Observe that $I(\pi)$ is an independent set (if u, v are adjacent we must have, say $\pi(v) < \pi(u)$, in which case $u \notin I(\pi)$), so in particular $\alpha(G) \geq |I(\pi)|$ for all π.

Let π be a random bijection chosen uniformly amongst all bijections from $V(G)$ to $[n]$, and let 1_v be the indicator variable which is 1 if $v \in I(\pi)$ and 0 otherwise. Note that regardless of what π is, we have $\alpha(G) \geq |I(\pi)| = \sum 1_v$, so by linearity of expectation we have

$$\alpha(G) \geq \mathbb{E}[I(\pi)] = \sum \mathbb{E}[1_v] = \sum \Pr[1_v = 1]. \tag{1}$$

Observe that $1_v = 1$ if and only if $\pi(v) = \min_{u \in \{v\} \cup N(v)} \pi(u)$. Since π was chosen uniformly at random, each $u \in \{v\} \cup N(v)$ is equally likely to achieve this minimum, so $\Pr[1_v = 1] = \frac{1}{d(v)+1}$, and plugging this into (1) gives the result.

Note that equality holds in (1) if and only if $I(\pi)$ is an independent set of maximum size for all bijections π. It is not too difficult to show that this holds if and only if G is a disjoint union of cliques, and we leave this as an exercise to the reader. $\square$

Theorem 1.2 implies Turán's theorem, which is essentially the result that jump started the entire field of extremal combinatorics[1] (though the original proof was not probabilistic).

[1] The first theorem in extremal combinatorics is typically attributed to Mantel, which is the $r = 3$ case of Turán's Theorem. However, it wasn't until Turán's result 30 years later that the field really took off.

To state this result, we define $\mathrm{ex}(n, F)$ to be the largest number of edges that an n-vertex F-free graph can have[1] which is called the *Turán number* or *extremal number* of F. We define the *Turán graph* $T_r(n)$ to be the complete n-vertex r-partite graph with parts of sizes as equal as possible. We let $t_r(n) = e(T_r(n))$. For example, $T_2(n) = K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ and $t_2(n) = \lfloor n/2\rfloor \cdot \lceil n/2\rceil = \lfloor n^2/4\rfloor$. More generally we have

$$t_r(n) \le \binom{r}{2}(n/r)^2 = \frac{r-1}{r}\cdot\frac{n^2}{2} = \left(1-\frac{1}{r}\right)\frac{n^2}{2},$$

with equality holding if $r|n$ and otherwise $t_r(n)$ is the floor of this upper bound.

Corollary 1.3 (Turán's Theorem)**.** *For all $r \le n$ we have*

$$\mathrm{ex}(n, K_r) = t_{r-1}(n).$$

Moreover, $T_{r-1}(n)$ is the unique n-vertex K_r-free graph with $t_{r-1}(n)$ edges.

Proof. The lower bound $\mathrm{ex}(n, K_r) \ge t_{r-1}(n)$ follows by considering $T_{r-1}(n)$. Let G be an n-vertex K_r-free graph with degrees $d_1, \ldots, d_n$. Observe that the complement $\overline{G}$ contains no independent set of size r, so by Theorem 1.2 we have

$$r - 1 \ge \alpha(\overline{G}) \ge \sum \frac{1}{n - d_i}. \tag{2}$$

Observe that if x, y are positive numbers, then[2]

$$\frac{1}{x} + \frac{1}{y} \ge \frac{1}{\frac{1}{2}(x+y)} + \frac{1}{\frac{1}{2}(x+y)}$$

with equality holding if and only if $x = y$. In view of this inequality, we see that (2) is minimized when all of the d_i are as close together as possible. Because $\sum d_i = 2e(G)$, we have

$$r - 1 \ge n \cdot \frac{1}{n - 2e(G)/n} = \frac{n^2}{n^2 - e(G)} \implies e(G) \le \left(1 - \frac{1}{r-1}\right) n^2/2,$$

so $e(G) \le t_{r-1}(n)$ as desired. Moreover, to have equality, $\overline{G}$ must be a union of cliques with sizes as close as possible to each other, i.e. G must be a complete r-partite graph with parts having sizes as close as possible to each other, i.e. G must be the Turán graph. ☐

In addition to using the probabilistic method to get an upper bound for $\mathrm{ex}(n, K_n)$ as in Corollary 1.3, one can also use it to give a general lower bound for $\mathrm{ex}(n, F)$.

Theorem 1.4. *Let F be a graph with v vertices and $e \ge 2$ edges. If $e \ge v$, then*

$$\mathrm{ex}(n, F) = \Omega_v(n^{2-\frac{v-2}{e-1}}).$$

[1] Throughout the text, a graph being F-free means that it contains no subgraph which is isomorphic to F (and we don't care whether this subgraph is induced or not).

[2] By multiplying both sides of the above expression by $xy(x+y)$, we see that this is equivalent to saying $y(x+y) + x(x+y) \ge 4xy$, which is equivalent to saying $x^2 - 2xy + y^2 = (x-y)^2 \ge 0$.

For this proof we use an object that is fundamental to probabilistic and extremal combinatorics. This is the *Erdös-Renyi random graph* $G_{n,p}$, which is the random graph on n vertices that contains each edge $e \in E(K_n)$ independently and with probability p. For example, $G_{n,1} = K_n$ and $G_{n,1/2}$ is equally likely to be any labeled graph on n vertices. The random graph is an incredibly fascinating object in its own right. We will not discuss it in too much depth in this text, see the book by Frieze and Karoński [26] for a thorough treatment of it.

Proof. Let $G_{n,p}$ be the random graph with p a quantity to be determined later. Let X denote the number of copies of F in $G_{n,p}$. For S a set of v vertices, let 1_S be the indicator variable which is 1 if S contains a copy of F in $G_{n,p}$ and which is 0 otherwise. With this,

$$\sum 1_S \le X \le v! \sum 1_S,$$

since each set of v vertices contains at most $v!$ copies of F. To have $1_S = 1$, we in particular need S to contain at least e edges, so

$$\Pr[1_S = 1] \le \sum_{k \ge e} \binom{\binom{v}{2}}{k} p^k (1-p)^{\binom{v}{2}-k} \le v^2 2^{v^2} p^e \le 4^{v^2} p^e.$$

In total this gives

$$\mathbb{E}[X] \le v! \binom{n}{v} \cdot 4^{v^2} p^e \le (4^v n)^v p^e.$$

Observe that when $p \gg n^{v/e}$, the calculation above suggests that $G_{n,p}$ will contain copies of F (at least in expectation), so $G_{n,p}$ will not work as an F-free graph for this range of p. However, we can get around this by using the following trick known as the method of alterations. Let G be any subgraph of $G_{n,p}$ obtained by deleting an edge from each copy of F in $G_{n,p}$. By definition G will be F-free. Moreover, the number of edges that G has is at least $e(G_{n,p}) - X$ since at most X of the original edges from $G_{n,p}$ are deleted. Using linearity of expectation gives

$$\mathbb{E}[e(G)] \ge \mathbb{E}[e(G_{n,p}) - X] \ge p\binom{n}{2} - (4^v n)^v p^e \ge \frac{1}{4} p n^2 - (4^v n)^v p^e. \tag{3}$$

At this point we want to choose p so that the above expression is roughly maximized. Intuitively this will happen when both terms on the rightside of (3) are roughly equal to each other, i.e. when $pn^2 \approx n^v p^e$. This suggests taking $p \approx n^{\frac{2-v}{e-1}}$. And indeed, after playing around for a bit, one sees that, for example, taking $p = \frac{1}{20 \cdot 16^v} n^{\frac{2-v}{e-1}}$ and plugging it into (3) gives[1] $\mathbb{E}[e(G)] \ge \frac{1}{160 \cdot 16^v} n^{2 - \frac{2-v}{e-1}}$. Because G is a (random) F-free graph, by (*) there exists some deterministic graph G' which is F-free with this many edges, proving the result. $\square$

For many F, there are known constructions which give much better lower bounds for $\mathrm{ex}(n, F)$ than Theorem 1.4. However, this is the best known lower bound which works for arbitrary F.

The method used in this proof is known as the method of alterations. Typically this works by defining some initial random set A (e.g. a set of edges of a graph) which contains some bad

[1] Here we use $4^{v^2} \le 4^{ve}$ and that $e \ge 2$.

subsets B (e.g. subsets of edges forming a forbidden graph F). We then define a random set A' by deleting an element from each bad subset B, giving that $|A'| \geq |A| - |B|$ and that A' has no bad subsets. At this point we win provided

$$\mathbb{E}[|A'|] = \mathbb{E}[|A|] - \mathbb{E}[|B|]$$

is large. Typically the expectations $\mathbb{E}[|A|], \mathbb{E}[|B|]$ depend on some common parameter p, and we often optimize this expression by finding p such that $\mathbb{E}[|A|] \approx \mathbb{E}[|B|]$, and then ultimately choosing p to be a bit smaller than this so that, say, $\mathbb{E}[|B|] \leq \frac{1}{2}\mathbb{E}[|A|]$.

(**) The method of alterations detailed above is often very useful.

The last core tenant of the probabilistic method that we have implicitly used throughout this section is the following.

(***) If one is trying to find a nice object, one should always try and see how well a random object does (possibly after applying alterations).

For example, the most straightforward random coloring gave the bound of Theorem 1.1, and the random graph together with alterations gave Theorem 1.4.

Lastly, we note that in principle many of these results could be proven without needing to use probability. However, for certain problems a probabilistic perspective is genuinely useful since it is allows one to use powerful tools from probability theory (e.g. martingales and concentration inequalities). Even when it isn't strictly needed, probability often provides for a much clearer perspective on a problem.

2 Some Random Examples

This section consists of an assorted collection of examples which provides both practice with the general principles of the probabilistic method, as well as proofs of many fundamental results from extremal combinatorics.

2.1 Graphs with Small and Large Chromatic Numbers

We start with a very simple example that will be used throughout the text (often without reference).

Lemma 2.1. *If G is an n-vertex graph, then there exists a bipartite subgraph $G' \subseteq G$ such that $e(G') \geq \frac{1}{2}e(G)$. Moreover, we can choose G' such that its partition classes U, V have sizes $\lfloor n/2 \rfloor, \lceil n/2 \rceil$.*

Given this lemma, if you want to prove a statement of the form "any graph G with $\Omega(m)$ edges has some monotone graph property", then you only need to consider graphs which are (balanced) bipartite.

Proof. The first part is very easy: let $U \subseteq V(G)$ be obtained by including each vertex independently and with probability $\frac{1}{2}$, and let $V = V(G) \setminus U$. Let G' be the graph which consists of every edge $e \in E(G)$ with one vertex in U and one vertex in V. It is easy to check that $\mathbb{E}[e(G')] = \frac{1}{2}e(G)$, so such a (bipartite) subgraph exists.

The second part is conceptually easy but computationally a little tedious. Let $U \subseteq V(G)$ be a set of size $\lfloor n/2 \rfloor$ chosen uniformly at random and let $V = V(G) \setminus U$. Let G' be the graph which consists of every edge $e \in E(G)$ with one vertex in U and one vertex in V. Observe that the probability that a given edge $xy \in E(G)$ is in G' is exactly

$$1 - \frac{\lfloor n/2 \rfloor \cdot (\lfloor n/2 \rfloor - 1)}{n(n-1)} - \frac{\lceil n/2 \rceil \cdot (\lceil n/2 \rceil - 1)}{n(n-1)} \geq \frac{1}{2},$$

with the last step following from a case analysis based on whether n is even or odd. Thus in expectation G' has at least $\frac{1}{2}e(G)$ edges, so such a balanced bipartite subgraph of G must exist. $\square$

A graph G is said to have *girth* ℓ if its smallest cycle is of size ℓ, and we say that it has infinite girth if G has no cycles. Observe that graphs of large girth locally look like a tree, i.e. if you pick any vertex v, then the graph induced by every vertex within distance ℓ of v is a tree. In particular, "locally" graphs of large girth can be properly colored using few colors, but does this necessarily hold globally as well? That is, does there exist graphs with girth at least ℓ and chromatic number at least k for all ℓ, k? A clever (random) argument of Erdős shows that such a graph does indeed exist.

Theorem 2.2 (Erdős). *For all ℓ, k there exist graphs of girth at least ℓ and chromatic number at least k.*

For this proof we use Markov's inequality: if X is a non-negative real-valued random variable, then $\Pr[X \geq x] \leq \mathbb{E}[X]/x$ for $x > 0$.

Proof. Consider $G_{n,p}$ with n, p to be determined later. Let $X_{\leq \ell}$ denote the number of cycles in $G_{n,p}$ of size at most ℓ. Linearity of expectation gives

$$\mathbb{E}[X_{\leq \ell}] \leq \sum_{t=3}^{\ell} n^t \cdot p^t \leq \ell (pn)^\ell.$$

Thus if we wanted $G_{n,p}$ to have girth smaller than ℓ with high probability, by Markov's inequality it would suffice to take $p \ll n^{-1}$. Unfortunately this naive approach is too weak since in this case $G_{n,p}$ will have very small chromatic number. To get around this, we will take p slightly larger than n^{-1} and then use alterations to delete a vertex from every small cycle of $G_{n,p}$. With some foresight[1] we will take $p = n^{-1+1/2\ell}$. With this we see that

$$\Pr[X_{\leq \ell} \geq n/2] \leq \mathbb{E}[X_{\leq \ell}]/(n/2) \leq 2\ell n^{-1/2}. \tag{4}$$

We now turn to the chromatic number of $G_{n,p}$, which is a slightly trickier quantity to get a handle on. To do this we use the inequality $\chi(G) \geq |V(G)|/\alpha(G)$, which follows from the fact that a k-coloring of G is a partition of $V(G)$ into independent sets. Thus for $G_{n,p}$ to have large chromatic number, it suffices to show that all of its independent sets are small . For m an integer we let Y_m be the number of independent sets of size m in $G_{n,p}$. Using linearity of expectation and $(1-x) \leq e^{-x}$ gives for $m \geq 2$

$$\mathbb{E}[Y_m] = \binom{n}{m} \cdot (1-p)^{\binom{m}{2}} \leq n^m \cdot \left(e^{-p(m-1)/2}\right)^m \leq \left(ne^{-pm/4}\right)^m.$$

By Markov's inequality and our choice of $p = n^{-1+1/2\ell}$, we find for $m = n/2k$ and n sufficiently large in terms of k, ℓ that

$$\Pr[Y_{n/2k} \geq 1] \leq (ne^{-n^{1/2\ell}/8k})^m < \frac{1}{2}. \tag{5}$$

By combining (4) and (5), we see for n suffiiciently large that $X_{\leq \ell} < n/2$ and $Y_{n/2k} = 0$ both occur with positive probability, i.e. there exists a graph G such that both of these events occur. Let G' be G after deleting a vertex from each cycle of length at most ℓ in G. This deletes at most half the vertices of G by assumption of $X_{\leq \ell}$, and we have $\alpha(G') \leq \alpha(G) \leq n/2k$. Thus

$$\chi(G') \geq |V(G')|/\alpha(G') \geq k,$$

proving the result. □

2.2 Random Permutations and Extremal Set Theory

In this subsection, we use random permutations (similar to the proof of Theorem 1.2) to prove two famous results from extremal set theory, which is roughly speaking the study of

[1] The exact choice of p doesn't matter here, the important thing is to take $p = n^{-1+\alpha}$ with $0 < \alpha < 1/\ell$.

extremal problems for hypergraphs. We only scratch the surface of this topic, see Frankl and Tokushige [24] for a more thorough treatment.

We start with the most fundamental theorem in extremal set theory: the Erdős-Ko-Rado theorem.

Theorem 2.3 (Erdős-Ko-Rado Theorem)**.** *Let $\mathcal{F} \subseteq \binom{[n]}{k}$ be an intersecting family, i.e. $F \cap F' \neq \emptyset$ for any $F, F' \in \mathcal{F}$. If $n \geq 2k$, then*

$$|\mathcal{F}| \leq \binom{n-1}{k-1}.$$

This bound is sharp by taking $\mathcal{F}$ to consist of every set containing the element 1 (and in fact, up to isomorphism this is the unique extremal construction when $n > 2k$). Note that if $n < 2k$, then $\mathcal{F} = \binom{[n]}{k}$ is an intersecting family, so we need $n \geq 2k$ for us to be able to prove a non-trivial bound.

Proof. The proof uses what is known as Katona's circle method, which involves choosing a random cyclic ordering $\pi : [n] \to \mathbb{Z}_n$, where $\mathbb{Z}_n$ is the integers mod n. Given such a π and a set $A \in \mathcal{F}$, we let 1_A be the indicator variable with $1_A = 1$ if $A = \{\pi(i), \pi(i)+1, \ldots, \pi(i)+k-1\}$ for some $i \in [n]$. We claim that $1_A = 1$ for at most k sets A.

Indeed, if $1_A = 0$ for all A then there is nothing to prove, so assume $1_A = 1$ for some A, say with $A = \{\pi(i), \pi(i)+1, \ldots, \pi(i)+k-1\}$. Let $S_j = \{\pi(i)+j, \pi(i)+j+1, \ldots, \pi(i)+j+k-1\}$, and observe that if $B \in \mathcal{F}$ has $1_B = 1$, then we must have[1] $B = S_j$ for some $-k < j < k$. Moreover, for each pair $\{S_{-k+\ell}, S_\ell\}$ with $0 \leq \ell < k$, at most one $B \in \mathcal{F}$ is equal to one of these sets since $S_{-k+\ell}, S_\ell$ are disjoint, so in total we conclude that $1_A = 1$ for at most k different $A \in \mathcal{F}$.

Observe[2] that $\Pr[1_A = 1] = n\binom{n}{k}^{-1}$, and this together with the claim above implies

$$k \geq \mathbb{E}[\sum_{A \in \mathcal{F}} 1_A] = \sum_{A \in \mathcal{F}} \Pr[1_A = 1] = |\mathcal{F}| \cdot n\binom{n}{k}^{-1},$$

and rearranging gives the desired bound. □

There are many, many proofs of the Erdős-Ko-Rado theorem, as well as many generalizations and applications. Again, we refer the reader to [24] for more on this. Our second result related to extremal set theory is the following.

Theorem 2.4 (Bollobás Set Pairs Inequality)**.** *Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ and $\mathcal{B} = \{B_1, \ldots, B_m\}$ be set systems such that $A_i \cap B_i = \emptyset$ for all i and $A_i \cap B_j \neq \emptyset$ for all $i \neq j$. Then*

$$\sum_{i=1}^{m} \binom{|A_i| + |B_i|}{|A_i|}^{-1} \leq 1$$

[1] Here we use that each $B \in \mathcal{F}$ intersects A and that $n \geq 2k$ implies S_k is disjoint from A

[2] This follows because for any cyclic ordering π there are exactly n sets S which have $1_S = 1$

Pairs of families as in Theorem 2.4 are called *cross-intersecting*.

Proof. Let π be a random permutation of the underlying ground set (the size of which is irrelevant for the conclusion/proof). Let 1_i be the indicator variable with $1_i = 1$ if $\pi(x) < \pi(y)$ for all $x \in A_i$ and $y \in B_i$. That is, 1_i is the indicator for the event that A_i appears completely before B_i under π. A simple counting argument shows that $\Pr[1_i = 1] = \binom{|A_i|+|B_i|}{|A_i|}^{-1}$ (where here we implicitly use that $A_i \cap B_i = \emptyset$, as otherwise $\Pr[1_i = 1] = 0$).

We claim that there is at most one i such that $1_i = 1$. Indeed, say $1_i = 1$. Then for any $j \neq i$, by hypothesis there is some $x \in A_j \cap B_i \subseteq A_j$ and $y \in A_i \cap B_j \subseteq B_j$, and since $1_i = 1$, we have $\pi(x) > \pi(y)$. Thus $1_j = 0$ for all $j \neq i$. With this claim we have

$$1 \geq \mathbb{E}\Big[\sum_i 1_i\Big] = \sum \Pr[1_i = 1] = \sum \binom{|A_i| + |B_i|}{|A_i|}^{-1}.$$

$\square$

Theorem 2.4 has many applications. One such application involves *antichains*, which are collections of sets $\mathcal{F}$ such that there exist no distinct $A, B \in \mathcal{F}$ with $A \subseteq B$.

Corollary 2.5 (LYM Inequality)**.** *If $\mathcal{F} \subseteq 2^{[n]}$ is an antichain, then*

$$\sum_{A \in F} \binom{n}{|A|}^{-1} \leq 1.$$

Proof. Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ and define $B_i = [n] \setminus A_i$. It is not difficult to check that since $\mathcal{F}$ is an antichain, $A_i \cap B_j = \emptyset$ if and only if $i = j$. The bound then follows from Theorem 2.4. $\square$

We note that the proof of Corollary 2.5 is a nice simplification of the proof of Theorem 2.4: now $1_i = 1$ if and only if $A_i = \{\pi(1), \ldots, \pi(|A_i|)\}$.

Corollary 2.6 (Sperner's Theorem)**.** *If $\mathcal{F} \subseteq 2^{[n]}$ is an antichain, then*

$$|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

This result is sharp, as can be seen by taking $\mathcal{F} = \binom{[n]}{\lfloor n/2 \rfloor}$ or $\binom{[n]}{\lceil n/2 \rceil}$.

Proof. We have $\binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor}$ for all k, so by the LYM inequality

$$1 \geq \sum_{A \in \mathcal{F}} \binom{n}{|A|}^{-1} \geq |\mathcal{F}| \binom{n}{\lfloor n/2 \rfloor}^{-1},$$

and moving things around gives the desired result. $\square$

2.3 The Crossing Lemma and Incidence Geometry

Our final result concerns drawings of graphs. Without being too precise with our definitions, we define the *crossing number* of a graph G to be the minimum number of crossings that an embedding $\phi(G)$ in the plane will have. For example, a graph is planar if and only if $cr(G) = 0$.

Lemma 2.7. *If G is an n-vertex graph with m edges, then $cr(G) \geq m - 3n$.*

Sketch of Proof. Let $\phi(G)$ be an embedding of G with $cr(G)$ crossings. By deleting an edge from each crossing, we obtain a planar graph G' with n vertices and at least $m - cr(G)$ edges. A simple consequence of Euler's formula shows that this means $m - cr(G) \leq 3n$, giving the result. $\square$

We will use the probabilistic method to "amplify" the elementary bound of Lemma 2.7 and give a bound that is effective for dense graphs.

Lemma 2.8 (Crossing Lemma)**.** *If G is an n-vertex graph with $m \geq 4n$ edges, then*

$$cr(G) \geq \frac{m^3}{64n^2}.$$

Proof. Let $\phi(G)$ be an embedding of G which has $cr(G)$ crossings. Let $V_p \subseteq V(G)$ be obtained by keeping each vertex of $V(G)$ independently and with probability p, and let $G_p = G[V_p]$. Observe that there is a natural embedding of G_p, namely the restriction of ϕ to G_p.

Let X denote the number of crossings in $\phi(G_p)$, and note that $\mathbb{E}[X] = p^4 cr(G)$ since a crossing survives if and only if all four of its relevant vertices lie in V_p. Using Lemma 2.7, we see that

$$p^4 cr(G) = \mathbb{E}[X] \geq \mathbb{E}[e(G') - 3|V_p|] = p^2 m - 3pn \implies cr(G) \geq p^{-2}m - 3p^{-3}n.$$

This lower bound will roughly be optimized when $p^{-2}m = p^{-3}n$, i.e. when $p = n/m$. More precisely, taking $p = 4n/m$ gives the desired bound. However, implicitly this argument requires that $0 \leq p \leq 1$, i.e. that $m \geq 4n$, and this holds by hypothesis. $\square$

In addition to being interesting in its own right, the crossing lemma gives a short proof of a fundamental result in incidence geometry.

Theorem 2.9 (Szemeredi-Trotter Theorem)**.** *Let $\mathcal{P}$ be a set of n points and $\mathcal{L}$ a set of m lines in the plane, and let $\mathcal{I} \subseteq \mathcal{P} \times \mathcal{L}$ denote their set of incidences, i.e. pairs (p, ℓ) with $p \in \ell$. Then*

$$|\mathcal{I}| = O(m^{2/3}n^{2/3} + m + n).$$

This bound is essentially best possible, though we omit the details of the (not too difficult) construction.

Proof (due to Székely). Without loss of generality, we can assume every point and line is in at least one incidence (otherwise we can delete these points/lines). Let G be the graph on $\mathcal{P}$ which makes two points p_1, p_2 adjacent if there exists a line $\ell \ni p_1, p_2$ such that there is no third point

q on the line segment p_1p_2. In other words, G is the graph obtained by drawing the points and lines on the plane, and then erasing the rays of lines which go off to infinity.

If $i(\ell)$ denotes the number of points incident to ℓ, then it is not difficult to see that $e(G) = \sum i(\ell) - 1 = |\mathcal{I}| - m$, where here we implicitly used that $i(\ell) \geq 1$ for all ℓ. If $|\mathcal{I}| \leq 2m$, then in particular $|\mathcal{I}| = O(m)$ and the result follows, so we can assume $e(G) \geq \frac{1}{2}|\mathcal{I}|$, and similarly we can assume $|\mathcal{I}| \geq 8m$ and hence $e(G) \geq 4n$. Thus by the crossing lemma we have

$$cr(G) \geq \frac{|\mathcal{I}|^3}{2^9 n^2}.$$

The critical observation is that $cr(G) \leq \binom{m}{2}$ since each crossing corresponds to two lines of $\mathcal{L}$ intersecting. Plugging this into the expression above gives the desired result. ☐

As a brief aside, we note that this idea of taking a weak result (Lemma 2.7) and amplifying it to a stronger result (Lemma 2.8) shows up in many other places in extremal combinatorics. For example, it is easy to prove a weak version of the Szemeredi-Trotter theorem with a bound of roughly $O(mn^{1/2} + n)$ by observing that there exist no points p_1, p_2 and ℓ_1, ℓ_2 such that all of the incidences (p_i, ℓ_j) are present, i.e. the "incidence graph" on $\mathcal{P} \cup \mathcal{L}$ contains no C_4. One can then use the method of polynomial partitioning to dissect $\mathbb{R}^2$ into small regions where this bound is effective. For much more on incidence geometry and polynomial partitioning, we refer the reader to the excellent book by Sheffer [56].

3 The Lovász Local Lemma

We say that an event A_i is mutually independent of a set of events $\{A_j : j \in J\}$ if for any $J' \subseteq J$, we have $\Pr[A_i \cap \bigcap_{j\in J'} A_j] = \Pr[A_i] \cdot \Pr[\bigcap_{j\in J'} A_j]$. We say that $A_1, \ldots, A_n$ are mutually independent events if A_i is mutually independent of $\{A_j : j \in [n] \setminus \{i\}\}$ for all i. Note that in this case we have $\Pr[\bigcap A_i] = \prod \Pr[A_i]$. In this section we consider a result which roughly says that if the A_i's are "almost independent", then we have $\Pr[\bigcap A_i] \approx \prod \Pr[A_i]$.

Theorem 3.1. *[Lovász Local Lemma] Let $A_1, \ldots, A_n$ be events and let $D_1, D_2, \ldots, D_n \subseteq [n]$ be such that A_i is mutually independent of $\{A_j : j \notin D_i \cup \{i\}\}$ for all i. If there exist real numbers $\gamma_i \in [0,1)$ such that $\Pr[A_i] \le \gamma_i \prod_{j\in D_i}(1-\gamma_j)$ for all i, then*

$$\Pr[\bigcap \overline{A_i}] \ge \prod (1-\gamma_i) > 0.$$

This result is often just referred to as "the local lemma". Note that if the A_i were all mutually independent, then we could take $D_i = \emptyset$ and $\gamma_i = \Pr[A_i]$ for all i and conclude from the local lemma that $\Pr[\bigcap \overline{A_i}] \ge \prod \Pr[\overline{A_i}]$.

Proof. We claim that for all i and $S \subseteq [n]$, we have

$$\Pr[\overline{A_i} | \bigcap_{j\in S} \overline{A_j}] \ge 1 - \gamma_i.$$

This will give the result since then

$$\Pr[\bigcap_i \overline{A_i}] = \prod_i \Pr[\overline{A_i} | \bigcap_{j\in[i-1]} \overline{A_j}] \ge \prod_i (1-\gamma_i).$$

We prove this claim by induction[1] on $|S|$. The base case $|S| = 0$ is equivalent to saying $\Pr[A_i] \le \gamma_i$ for all i, and this follows from $\Pr[A_i] \le \gamma_i \prod_{j\in D_i}(1-\gamma_j) \le \gamma_i$. Now consider any set S, and in particular assume we have proven the result for all $S' \subsetneq S$. If $i \in S$ then the result is trivial, so we can assume $i \notin S$. Observe that

$$\begin{aligned}\Pr[A_i | \bigcap_{j\in S} \overline{A_j}] = \frac{\Pr[A_i \cap \bigcap_{j\in S} \overline{A_j}]}{\Pr[\bigcap_{j\in S} \overline{A_j}]} &\le \frac{\Pr[A_i \cap \bigcap_{j\in S\setminus D_i} \overline{A_j}]}{\Pr[\bigcap_{j\in S\setminus D_i} \overline{A_j}] \cdot \Pr[\bigcap_{k\in S\cap D_i} \overline{A_k} | \bigcap_{j\in S\setminus D_i} \overline{A_j}]} \\ &= \frac{\Pr[A_i]}{\Pr[\bigcap_{k\in S\cap D_i} \overline{A_k} | \bigcap_{j\in S\setminus D_i} \overline{A_j}]}, \end{aligned} \tag{6}$$

where the first inequality used that we are taking a product over fewer events, and the second equality used that A_i is mutually independent of events not in D_i. Let $S \cap D_i = \{k_1, \ldots, k_p\}$. Then we can rewrite the probability in the denominator of (6) as

$$\prod_{q=1}^{p} \Pr[\overline{A_{k_q}} | \bigcap_{j\in(S\setminus D_i)\cup\{k_1,\ldots,k_{q-1}\}} \overline{A_j}] \ge \prod_{q=1}^{p} (1-\gamma_{k_q}) \ge \prod_{j\in D_i} (1-\gamma_j),$$

[1] It is perhaps more natural to try and prove the result by induction on n rather than on this somewhat weird looking claim. However, if one plays around with this problem, one quickly sees that one needs to prove something like the stated claim.

where the first inequality used the inductive hypothesis and the last step used $k_q \in S \cap D_i \subseteq D_i$ for all q. This together with (6) and the hypothesis $\Pr[A_i] \le \gamma_i \prod_{j \in D_i}(1-\gamma_j)$ implies that $\Pr[A_i | \bigcap_{j\in S} \overline{A_j}] \le \gamma_i$, which is equivalent to saying $\Pr[\overline{A_i} | \bigcap_{j \in S} \overline{A_j}] \ge 1-\gamma_i$. This proves the inductive hypothesis of our claim, and hence proves the result. ☐

The following version of the local lemma is often sufficient for most applications (and again this is often referred to as "the local lemma").

Corollary 3.2 (Symmetric Lovász Local Lemma). *Let $A_1, \ldots, A_n$ be events and let $D_1, D_2, \ldots, D_n \subseteq [n]$ be such that A_i is mutually independent of $\{A_j : j \notin D_i \cup \{i\}\}$ for all i. If $\Delta \ge 1$ is such that $|D_i| \le \Delta$ and $\Pr[A_i] \le \frac{1}{e(\Delta+1)}$ for all i, then $\Pr[\bigcap \overline{A_i}] > 0$.*

Proof. Observe that for all i we have

$$\frac{1}{\Delta+1}\prod_{j\in D_i}\left(1-\frac{1}{\Delta+1}\right) \ge \frac{1}{\Delta+1}\left(1-\frac{1}{\Delta+1}\right)^{\Delta} \ge \frac{1}{e(\Delta+1)} \ge \Pr[A_i],$$

where the second to last inequality used that $(1-1/x)^{x-1} > 1/e$ for $x \ge 2$. Thus the (asymmetric) local lemma applies with $\gamma_i = \frac{1}{\Delta+1}$ for all i, proving the result. ☐

We note that this result is essentially best possible. Indeed, consider rolling a fair $(\Delta+1)$-sided dice and let A_i be the event that the dice rolls i. In this case A_i is dependent on all of $D_i = [\Delta+1] \setminus \{i\}$ and we have $\Pr[A_i] = \frac{1}{\Delta+1} > \frac{1}{e(\Delta+1)}$, so the local lemma does not apply (which is good since we have $\Pr[\bigcap \overline{A}_i] = 0$). In particular, this example shows that we can not improve the requirement $\Pr[A_i] \ge \frac{1}{e(\Delta+1)}$ in the symmetric local lemma to $\Pr[A_i] \ge \frac{1}{\Delta+1}$ in general. Thus the hypothesis in the symmetric local lemma is sharp up to a factor of e, and in fact Shearer proved that this factor of e is necessary [55].

3.1 Applications

Our first application of the local lemma will be an asymptotic improvement to our lower bound for Ramsey numbers from Theorem 1.1.

Theorem 3.3 (Spencer [58]). *For all n we have*

$$R(n,n) \ge (1+o(1))\frac{\sqrt{2}n}{e}2^{n/2}.$$

Proof. Uniformly at random color the edges of K_N. For $S \in \binom{[N]}{n}$, let A_S be the event that G contains a monochromatic K_n on S, and as before we note that $\Pr[A_S] = 2^{1-\binom{n}{2}}$. Let D_S consist of all the sets $T \in \binom{[N]}{n} \setminus \{S\}$ such that $|S \cap T| \ge 2$. It is not difficult to see that A_S is mutually independent of $\{A_T : T \notin D_S \cup \{S\}\}$ since the color given to each pair of S is independent of these events. A weak bound gives $|D_S| \le \binom{n}{2}\binom{N}{n-2} - 1 \le n^2(eN/(n-2))^{n-2} - 1$, so by the (symmetric) local lemma we have that $\Pr[\bigcap \overline{A_S}] > 0$ provided $2^{1-\binom{n}{2}} < \frac{1}{en^2}(eN/n-2)^{2-n}$, i.e. if

$$(2en^2)^{1/n-2} \cdot 2^{\binom{n}{2}/n-2} \cdot \frac{n-2}{eN} = (2en^2)^{1/n-2} \cdot 2^{n/2+1/2-1/(n-2)} \cdot \frac{n-2}{eN} < 1,$$

and this happens if $N = (1-\epsilon)\frac{\sqrt{2}n}{e}2^{n/2}$ for any $\epsilon > 0$ provided n is sufficiently large, giving the desired result. $\square$

The local lemma works best if there are few dependencies between events. As such, it performs much better for off-diagonal Ramsey numbers.

Theorem 3.4. *For all n we have*

$$R(3,n) = \Omega(n^2/\log^2 n).$$

Proof. Randomly color each edge of K_N red with probability p and blue otherwise. Given a set $S \in \binom{[N]}{3}$, we let R_S be the event that the vertices of S form a red triangle, and similarly for $T \in \binom{[N]}{n}$ we define B_T. Observe that $\Pr[A_S] = p^3$ and $\Pr[B_T] = (1-p)^{\binom{n}{2}}$.

Given $S \in \binom{[N]}{3} \cup \binom{[N]}{n}$, we define D_S to be the sets of sizes 3 and n which intersect S in at least two vertices. Observe that if $|S| = 3$, then D_S contains at most $3N$ set of size 3 and at most $\binom{N}{n}$ sets of size n, and if $|S| = n$, we have that D_S contains at most $N\binom{n}{2}$ sets of size 3 and at most $\binom{N}{n}$ sets of size n. Our goal now is to choose some parameters γ_S, γ_T so that the (asymmetric) local lemma applies to the R_S, B_T events.

At this point there's a lot of undetermined variables floating around: N, p, γ_S. Let's think about reasonable guesses for how to optimize things. First of all, it seems clear that we probably want two parameters γ_3, γ_n such that we set $\gamma_S = \gamma_{|S|}$ when applying the local lemma. With this we in particular need

$$p^3 \le \gamma_3(1-\gamma_3)^{3N}(1-\gamma_n)^{\binom{N}{n}}. \tag{7}$$

In particular we need $\gamma_3 \ge p^3$, so let's naively take $\gamma_3 = Cp^3$ for some large constant C. Given this, we also need $\gamma_n \le c\binom{N}{n}^{-1}$ in order to have the $(1-\gamma_n)^{\binom{N}{n}}$ term be no larger than a constant. If we take $\gamma_n = c\binom{N}{n}^{-1}$, we see that (7) is satisfied provided $p = o(N^{-1/3})$ and c, C are chosen appropriately.

The other condition we need to satisfy is

$$(1-p)^{\binom{n}{2}} \le \gamma_n(1-\gamma_3)^{N\binom{n}{2}}(1-\gamma_n)^{\binom{N}{n}},$$

and by plugging in our choices for γ_3, γ_n and the assumption that p must be fairly small, we essentially need to have

$$e^{-p\binom{n}{2}} \le (n/N)^n \cdot e^{-p^3 N\binom{n}{2}},$$

and for this to hold we in particular need something like $p\binom{n}{2} \ge p^3 N\binom{n}{2}$, i.e. $p = O(N^{-1/2})$. Taking $p = c'N^{-1/2}$, we see that we also need roughly

$$p\binom{n}{2} \approx c'N^{-1/2}n^2 \ge n\log(N/n).$$

Assuming $N \ge n^{1+\epsilon}$ for some small $\epsilon > 0$, this reduces to $N^{1/2} \le n/\log n$, i.e. $N = n^2/(\log n)^2$.

Thus in total, a heuristic argument suggests that we can apply the local lemma with $N = \Theta(n^2/(\log n)^2)$ by taking $p = \Theta(N^{-1/2})$, $\gamma_3 = \Theta(N^{-3/2})$, and $\gamma_n = \Theta(\binom{N}{n})$. And indeed, a careful analysis shows that this will work out for n sufficiently large. $\square$

We note that the bound of $n^2/\log^2 n$ is the best one can do using this approach. However, it turns out that $R(3,n) = \Theta(n^2/\log n)$. This improved lower bound was originally proved by Kim [41]. The idea of their proof was to start with a K_N which is entirely colored blue, and then to iteratively randomly pick an edge of K_N and color it red if it does not create a red triangle. A careful analysis shows that with positive probability the final graph at the end contains no large blue clique, and it contains no red clique by construction. We will see a shorter proof of this lower bound in **a later section (expanders)**.

3.2 Related Lemmas

While the local lemma is very powerful, there are certain circumstances where it doesn't give you quite what you want. Fortunately there are many other lemmas which allow one to prove bounds on $\Pr[\bigcap \overline{A_i}]$ even when the A_i depend on each other in some way. For example, it is not too difficult to generalize the local lemma as follows (and as an exercise the reader should convince themselves that they can prove this result).

Theorem 3.5. *Let $A_1, \ldots, A_n$ be events. Assume there exists partitions $D_i \cup E_i = [n] \setminus \{i\}$ for all i and real numbers $0 \le \delta, \gamma \le 1$ such that $\gamma(1-\gamma)^{|D_i|} \ge \delta$ and for all $E \subseteq E_i$ we have $\Pr[A_i | \bigcap_{e \in E} \overline{A}_e] \le \delta$ and . Then*

$$\Pr[\bigcap \overline{A}_i] \ge (1-\gamma)^n > 0.$$

Note that when $\delta = \gamma$ we more or less recover Theorem 3.1 when $\gamma_i = \gamma$ for all i. The power here is that we allow each A_i to possible be dependent of every event, but it is not "very dependent" on the events of E_i.

Another result in a similar spirit as the local lemma is Janson's inequality. Given a set $S \subseteq X$ and $\vec{p}$, let $\mathcal{A}_S$ be the set containing

Theorem 3.6 (Janson's inequality)**.** *Let H be a hypergraph on a set V, and let V_p be the set obtained by including each vertex of V independently and with probability p. Let A_i denote the event that V_p contains the ith edge of H and define*

$$\mu = \sum \Pr[A_i], \;\; \Delta = \sum_{(S_i, S_j): S_i \cap S_j \neq \emptyset} \Pr[A_i \cap A_j].$$

Then

$$\prod_i \Pr[\overline{A_i}] \le \Pr[\bigcap_i \overline{A_i}] \le e^{-\mu + \frac{\Delta}{2}}.$$

Note that if all of the edges of H are disjoint, then these bounds are roughly $e^{-\mu} \le \Pr[\bigcap_i \overline{A_i}] \le e^{-\mu/2}$. Again there are many variants of Theorem 3.6 which are useful in different situations.

4 The Second Moment Method

In the previous section we saw several results that can be used to bound the probability of some event happening. In this section we look at what is perhaps the most broadly applicable result of this form: Chebyshev's inequality.

4.1 Concentration Inequalities

Roughly speaking, concentration inequalities are results which say that under reasonable circumstances, a random variable is likely to be close to its expectation.

Perhaps the most famous (one-sided) concentration inequality is Markov's inequality. We already saw this around the proof of Theorem 2.2, but for good measure we'll formally state it here.

Lemma 4.1 (Markov's inequality)**.** *If X is a non-negative real-valued random variable, then for all $\lambda > 0$ we have*

$$\Pr[X \geq \lambda] \leq \frac{\mathbb{E}[X]}{\lambda}.$$

In particular, if X is integer-valued, then

$$\Pr[X \neq 0] \leq \mathbb{E}[X].$$

Proof. For simplicity we only prove the result when X is integer valued. In this case we have

$$\Pr[X \geq \lambda] = \sum_{k \geq \lambda} \Pr[X = k] \leq \sum_{k \geq \lambda} \frac{k}{\lambda} \cdot \Pr[X = k] = \frac{\mathbb{E}[X]}{\lambda}.$$

The second statement follows by taking $\lambda = 1$. □

The "in particular" part of this lemma is probably the most common usage of Markov's inequality. To reiterate, this says that $\mathbb{E}[X] \to 0$ implies $X = 0$ with high probability, and this application of Markov's inequality is often known as *the first moment method*.

Unfortunately it is not true in general that $\mathbb{E}[X] \to \infty$ implies $X > 0$ with high probability (e.g. take $X = n$ with probability $n^{-1/2}$ and $X = 0$ otherwise). However, for many reasonable examples this implication does hold. Often one can show this by utilizing Chebyshev's inequality. We recall that the variance of a random variable is $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Lemma 4.2 (Chebyshev's inequality)**.** *Let X be a real-valued random variable with* $\mathrm{Var}(X) = \sigma^2$*. Then for all $\lambda > 0$, we have*

$$\Pr[|X - \mathbb{E}[X]| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

Proof. We have

$$\Pr[|X - \mathbb{E}[X]| \geq \lambda\sigma] = \Pr[(X - \mathbb{E}[X])^2 \geq \lambda^2\sigma^2] \leq \frac{1}{\lambda^2},$$

where this last step used Markov's inequality applied to the (non-negative) random variable $Y := (X - \mathbb{E}[X])^2$ after noting that $\mathbb{E}[Y] = \sigma^2$ by definition. $\square$

Morally speaking, Chebyshev's inequality says that if $\sigma = o(\mathbb{E}[|X|])$, then X is close to its expectation with high probability. The usage of Chebyshev's inequality is often referred to as *the second moment method.*

4.2 The Rödl Nibble

Our approach in this subsection borrows heavily from Alon and Spencer [2].

We say that a set of edges $E \subseteq E(K_n^r)$ is a *k-covering* if every k-set $S \subseteq [n]$ is contained in some edge of E, and we will simply call this a *covering* if $k = 1$. It isn't hard to see that every k-covering needs at least $\binom{n}{k}/\binom{r}{k}$ edges, with equality holding if and only if every k-set is contained in exactly one edge of E. S Erdös and Hannini asked whether one could find k-coverings with asymptotically this many edges, and this was answered positively by Rödl.

Theorem 4.3 (Rödl [52]). *For all fixed $k \le r$, there exists a k-covering $E \subseteq E(K_n^r)$ with*

$$|E| = (1 + o(1))\frac{\binom{n}{k}}{\binom{r}{k}},$$

where the $o(1)$ term tends to 0 as n tends towards infinity.

The first step of this argument is to reduce the problem of finding k-coverings to simply finding coverings. To this end we define $H_n^{r,k}$ to be the $\binom{r}{k}$-uniform hypergraph whose vertex set is $\binom{[n]}{k}$ and whose edge set is $\{\binom{S}{k} : S \in \binom{[n]}{r}\}$. That is, the hyperedges of $H_n^{r,k}$ are all the sets of size k covered by an edge $S \in E(H_n^r)$. It is not too difficult to check that Theorem 4.3 is equivalent to saying that there exists a covering of $H_n^{r,k}$ of the stated size.

Our approach for proving that $H_n^{r,k}$ has a small covering will more generally show that any "nice" r-uniform hypergraph H has a small covering, with the approach roughly being the following:

- Randomly choose $\epsilon n/r$ edges E_1 from H for some small $\epsilon > 0$. With high probability E_1 will cover about $e^{-\epsilon}n$ vertices.
- Delete vertices covered by E_1 to get a new hypergraph H_2. With high probability H_2 is also "nice", so we can iterate the procedure above and pick some random set of edges E_2.
- We keep doing this until ϵn vertices remain uncovered, and at this point we can trivially cover them using at most ϵn additional edges.

Broadly speaking this approach is known as the Rödl nibble or the semirandom method. To reiterate, the core idea is that you iteratively do something to a small chunk of vertices in such a way that the structure of the rest of your hypergraph remains roughly the same with high probability, which allows one to keep iterating this procedure until one is left with a very small problem to solve.

It turns out that for our purposes, "nice" hypergraphs are those which are roughly D-regular for some D and such that every pair of vertices is in $o(D)$ edges, i.e. the hypergraph is almost linear. These conditions are reasonable for trying to prove Theorem 4.3 since $H_n^{r,k}$ is $\binom{n-k}{r-k}$-regular and every pair of vertices is in at most $\binom{n-k-1}{r-k-1}$ edges. Our main technical lemma in this direction is the following, where throughout this section we write $c = 1 \pm \delta$ if $c \in [1-\delta, 1+\delta]$.

Lemma 4.4. *For every $r \geq 2$ and reals $K \geq 1$ and $\epsilon, \delta' > 0$, there are $\delta = \delta(r, K, \epsilon, \delta') > 0$ and $D_0 = D_0(r, K, \epsilon, \delta')$ such that for every $n \geq D \geq D_0$ the following holds.*

Let $H = (V, E)$ be an n-vertex r-graph such that

(i) For all but at most δn vertices $x \in V$, we have $d(x) = (1 \pm \delta)D$,

(ii) For all $x \in V$ we have $d(x) < KD$, and

(iii) For any two distinct $x, y \in V$, we have $d(x, y) < \delta D$.

In this case there exist a set of edges $E' \subseteq E$ such that

(a) $|E'| = (1 \pm \delta')(\epsilon n/r)$,

(b) The set $V' := V - \bigcup_{e \in E'} e$ has $|V'| = (1 \pm \delta')e^{-\epsilon}n$, and

(c) For all but at most $\delta'|V'|$ vertices $x \in V'$, the degree $d'(x)$ of x in the induced hypergraph $H[V']$ satisfies $d'(x) = (1 \pm \delta')De^{-\epsilon(r-1)}$.

Again, (i) and (ii) say that H is close to D-regular and (iii) says it has small codegrees. The main point of the conclusion is that the number of uncovered vertices and their degrees shrink in a predictable way.

Proof. We only give a sketch of the proof from [2] (their full version is complicated enough to involve 20 named constants!). Throughout the proof we'll introduce various constants δ_i which we always assume to be sufficiently small in terms of our relevant parameters.

Randomly choose a subset $E' \subseteq E$ such that each edge of E appears in E' independently and with probability $p = \epsilon/D$. Roughly speaking, our goal will be to show that with this choice of E', each of (a),(b),(c) occur in expectation, and then we will use Chebyshev to show that each of these occur with high probability.

To start, because H is essentially D-regular, we have $|E| = (1 \pm \delta_1)Dn/r$, so

$$\mathbb{E}[|E'|] = p|E| = (1 \pm \delta_1)\epsilon n/r.$$

We also have

$$\mathrm{Var}(|E'|) = p(1-p)|E| \leq 2\epsilon n/r.$$

Because $\mathrm{Var}(|E'|) = o(\mathbb{E}[|E'|])$, Chebyshev should be able to show that E' is close to $\mathbb{E}[E']$ with high probability. More precisely, Chebyshev's inequality implies

$$\Pr[||E'| - \mathbb{E}[|E'|] \geq \delta_1\sqrt{2\epsilon n/r} \cdot \sqrt{2\epsilon n/r}] \leq \frac{r}{2\delta_1^2 \epsilon n} \leq .01,$$

with this last step holding for n sufficiently large. Thus with probability at least .99,

$$|E'| = \mathbb{E}[|E'|] \pm 2\delta_1\epsilon n/r = (1 \pm 3\delta_1)\epsilon n/r.$$

This shows that (a) occurs with high probability. To deal with (b), let us first get a grasp on $\mathbb{E}[|V'|]$. For $x \in V$, let $1_x = 1$ if $x \notin \bigcup_{e\in E'} e$ and $1_x = 0$ otherwise. With this we see $|V'| = \sum 1_x$, so by linearity of expectation it suffices to bound each of $\mathbb{E}[1_x]$.

We will say that a vertex x is *good* if $d(x) = (1 \pm \delta)D$ and that it is *bad* otherwise. If x is bad we will simply use the trivial estimates $0 \leq \mathbb{E}[1_x] \leq 1$. If x is good we have

$$\mathbb{E}[1_x] = (1-p)^{d(x)} = (1 - \epsilon/D)^{(1\pm\delta)D} = (1 \pm \delta_3)e^{-\epsilon}, \tag{8}$$

where this last step used that $1-p$ is within a constant factor of e^{-p} for p sufficiently small and that δ is chosen to be sufficiently small in terms of ϵ (e.g. we can make sure that it's smaller than ϵ^{-1}).

Having at most δn bad vertices together with (8) implies $\mathbb{E}[|V|'] = (1 \pm \delta_4)ne^{-\epsilon}$. To compute the variance, we observe that

$$\mathrm{Var}[|V'|] = \sum_x \mathrm{Var}[1_x] + \sum_x \sum_{y\neq x} \mathbb{E}[1_x 1_y] - \mathbb{E}[1_x]\mathbb{E}[1_y]. \tag{9}$$

Because each 1_x is an indicator random variable, we have

$$\sum_x \mathrm{Var}[1_x] \leq \sum_x \mathbb{E}[1_x] = \mathbb{E}[|V'|].$$

For the mixed terms of (9), we have for any x, y that

$$\begin{aligned}\mathbb{E}[1_x 1_y] - \mathbb{E}[1_x]\mathbb{E}[1_y] &= (1-p)^{d(x)+d(y)-d(x,y)} - (1-p)^{d(x)+d(y)} \\ &\leq (1-p)^{-d(x,y)} - 1 \leq (1-\epsilon/D)^{-\delta D} - 1 \leq e^{\epsilon\delta} - 1 \leq \delta_5.\end{aligned}$$

In total we find

$$\mathrm{Var}[|V'|] \leq \mathbb{E}[V'] + \delta_5 n^2 \leq \delta_6(\mathbb{E}[V']|)^2,$$

where this last step used that $\mathbb{E}[|V'|] = \Theta_\epsilon(n)$. By Chebyshev we can guarantee with probability at least .99 that

$$|V'| = (1 \pm \delta_7)\mathbb{E}[|V'|] = (1 \pm \delta_8)ne^{-\epsilon}.$$

Proving that condition (c) holds with high probability is a little complicated, so we'll omit the full details[1] (which can be found in [2]). Let us instead give a heuristic argument as to why

[1] The argument is similar in spirit to that of (b): you define $1_e = 1$ if e survives in $H[V']$ and $1_e = 0$ otherwise. Then $d'(x)$ is just the sum of some of these indicator random variables, so one has to bound terms of the form $\mathbb{E}[1_e]$ and $\mathbb{E}[1_e 1_f]$. If e, f are "typical" edges then the computation of $\mathbb{E}[1_e]$ and $\mathbb{E}[1_e 1_f]$ are straightforward to estimate, and there are few terms involving e which are not typical.

(c) holds in expectation. We first condition on the event $x \in V'$, which means that no edge containing x is in E'. An edge $e \ni x$ survives in $H[V']$ only if every edge f with $e \cap f \neq \emptyset$ has $f \notin E'$. Because H is roughly linear and D-regular, there are about rD such edges f, but D of these (namely those containing x) are automatically not in E' since we conditioned on $x \in V'$. The remaining $(r-1)D$ edges are each included independently and with probability ϵ/D, so the probability that none are included is $(1-\epsilon/D)^{(r-1)D} \approx e^{-(r-1)\epsilon}$, and summing this over all of the roughly D edges containing x gives the result.

Once we have shown that each of (a),(b),(c) holds with probability at least .99, then the probability that all of them hold is at least .97, so in particular some choice of E' exists which satisfies these conditions. □

By repeatedly applying this lemma with carefully chosen values of ϵ, δ, one can prove the following result (and again, we omit the details of this, see [2]).

Theorem 4.5 (Pippenger). *For every $r \geq 2$ and reals $K \geq 1$ and $a > 0$, there are $\delta = \delta(r, K, a) > 0$ and $D_0 = D_0(r, K, a)$ such that for every $n \geq D \geq D_0$ the following holds.*

Let $H = (V, E)$ be an n-vertex r-graph such that

(i) For all but at most δn vertices $x \in V$, we have $d(x) = (1 \pm \delta)D$,

(ii) For all $x \in V$ we have $d(x) < KD$, and

(iii) For any two distinct $x, y \in V$, we have $d(x, y) < \delta D$.

Then there exists a cover of H using at most $(1 + a)(n/r)$ edges.

In particular, $H_n^{r,k}$ satisfies the conditions of the theorem, proving Theorem 4.3.

4.3 Steiner Systems

We say that a hypergraph is a Steiner system $S(n, r, k)$ if it has n-vertices, is r-uniform, and every k-set of its vertex set is contained in exactly one edge. Note that Steiner systems are k-coverings with exactly $\binom{n}{k}/\binom{r}{k}$ edges, so Theorem 4.3 shows that "approximate" Steiner systems exist, but when do actual Steiner systems exist?

The simplest non-trivial case is $S(n, 3, 2)$, which are also known as Steiner triple systems. It is not difficult to see that if a Steiner triple system on n vertices exists, then $3|\binom{n}{2}$ (each edge covers 3 pairs and each of the $\binom{n}{2}$ pairs are covered exactly once) and $2|(n-1)$ (for any given vertex v, each edge contains 2 pairs containing v and there are exactly $n-1$ such pairs). Equivalently, this argument says that if a Steiner triple system exists, then it is necessary that $n \equiv 1, 3 \mod 6$. It turns out that this condition is also sufficient due to certain constructions involving quasigropus and latin squares.

In general for an $S(n, r, k)$ to exist, there are certain "obvious" divisibility conditions that must be satisfied, but in general these are not sufficient. In fact, as of 2014, it wasn't even known if, say, any $S(n,k,6)$ Steiner systems existed, let alone if there were infinitely many n for which such a Steiner system existed. In a major breakthrough, it was shown by Keevash [39] and independently by Glock, Kühn, Lo, and Osthus [30] that if n is sufficiently large in terms of r, k, then $S(n,r,k)$ systems exist if and only if n satisfies the obvious divisibility conditions. The core of Keevash's proof was a variant of the Rödl nibble in an algebraic setting, but the proof is very, very complicated!

Part II

Further Probabilistic Methods

5 Dependent Random Choice

The following is all based off of the excellent survey by Fox and Sudakov [22]. Throughout this section we denote the common neighborhood of a set of vertices S by $N(S)$, i.e. $N(S) = \{u : u \in N(v)\ \forall v \in S\}$. Before we explain what dependent random choice is, let's first see an example of it in action.

Lemma 5.1. *Let G be an n-vertex graph with average degree at least d. For any choice of integers m, r, t, there exists a set $U \subseteq V(G)$ such that every r-subset of U has at least m common neighbors, and such that*

$$|U| \geq \frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

Proof. The statement of the result suggests how we should prove it: we'll randomly pick a set W which will have expected size at least d^t/n^{t-1}, and then we'll use the method of alterations to delete from W a set of a bad vertices, which in expectation will have size at most $\binom{n}{r}(m/n)^t$. The key twist is that we don't start by, say, defining W to include each vertex independently and with probability $p = d^t/n^t$, but instead W will end up depending on a different random set T.

To this end, let T be the random set obtained by uniformly at random selecting t vertices with repetition (i.e. each vertex is equally likely to be the ith vertex added to T, and in total T has size at most t), and define $W = N(T)$. The probability that a given vertex v is included in W is exactly $(d(v)/n)^t$, so by linearity of expectation and convexity we find that

$$\mathbb{E}[|W|] = \sum (d(v)/n)^t \geq d^t/n^{t-1}.$$

We say that a set of vertices $S \subseteq V(G)$ of size r is bad if $|N(S)| \leq m$. The probability that W contains a given bad set S is at most $(m/n)^t$ (since $S \subseteq W$ iff $T \subseteq N(S)$). Thus the expected number of bad sets of W is at most $\binom{n}{r}(m/n)^t$. If we let U be the set obtained by deleting a vertex from each bad set of W, then it has the desired properties by construction and in expectation it has the desired size, so such a choice of U exists. $\square$

Again, the key idea of this proof is that instead of defining W by including each vertex independently and with probability $p = d^t/n^t$, we instead formed it so that, on average, each vertex has probability at least p of being added, but the vertices are added in a very dependent way. In particular, the dependent way that W was generated made it more likely to have our desired property (i.e., we generated W by taking a common neighborhood, which made it less likely for W to contain sets of vertices with small common neighborhoods).

We can use Lemma 5.1 to prove some bounds on Turán numbers by using the following embedding lemma.

Lemma 5.2. *Let F be a bipartite graph on $A \cup B$ with $|A| = a,\ |B| = b$ such that the vertices in B all have degree at most r. If G is a graph which contains a set U such that $|U| = a$ and such that any subset of U of size r contains at least $a + b$ common neighbors, then G contains F as a subgraph.*

Proof. We define an injective homomorphism ϕ from $V(F)$ to $V(G)$ as follows. Choose $\phi|_A$ to be an arbitrary bijection onto U. For each $v \in B$ that has yet to be assigned, choose $\phi(v)$ to be any common neighbor of $\phi(N_F(v))$ which has yet to be assigned by ϕ. Note that there exist at least $a+b$ common neighbors of $\phi(N_F(v))$, so there certainly exists one which has yet to be assigned. This mapping gives the result. □

With this we can quickly prove the following.

Theorem 5.3 (Füredi [27]; Alon, Krivelevich, Sudakov [1]). *If F is a bipartite graph on $A \cup B$ such that the vertices of B all have degree at most r, then*

$$\mathrm{ex}(n, F) < 3(a+b)n^{2-1/r}.$$

Observe that this result generalizes Kővari-Sós-Turán, at least in terms of order of magnitude.

Proof. Assume G is an n-vertex F-free graph with average degree $d = 6(a+b)n^{1-1/r}$. By Lemma 5.2, we would be done if we could find a set U of size at least a such that every subset of size r had at least $m = a+b$ common neighbors. By Lemma 5.1, for any t we can find a set U with these properties of size at least

$$\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq (6a+6b)^t n^{1-t/r} - (e/r)^r (a+b)^t n^{r-t}.$$

We see that taking $t = r$ makes the powers of n on both sides equal, and in total this gives a set of size at least

$$(6a+6b)^r - (e(a+b)/r)^r.$$

Note that $6(a+b) \geq \frac{1}{2}(e(a+b)/r)$, so this is at least $\frac{1}{2}(6a+6b)^r \geq a$. We have thus found our desired set U, which together with Lemma 5.2 gives a copy of F in G, a contradiction. □

Another application of this method is to subdivisions. We define the 1-subdivision H^* of a graph H to be the graph obtained by replacing each edge of H by a P_2 (i.e. by inserting a new vertex in the middle of each edge). Note that subdivisions are bipartite graphs with all of its $e(H)$ new vertices having degree 2. Thus the previous theorem gives $\mathrm{ex}(n, K_a^*) = O(a^2 n^{3/2})$. It turns out that one can significantly improve upon this dependency of a.

Theorem 5.4 (Alon, Krivelevich, Sudakov [1]). *For all a we have*

$$\mathrm{ex}(n, K_a^*) = O(an^{3/2}).$$

Note that this only gives a reasonable bound when $a = O(n^{1/2})$, which makes sense since K_a^* has about a^2 vertices and thus can always be avoided by an n-vertex graph if $a \gg n^{1/2}$.

Unfortunately Lemma 5.1 on its own is not enough to prove Theorem 5.3, essentially because the size of U that we're guaranteed is too small. We can increase the size of U by demanding slightly weaker conditions for it to have, i.e. we only need that most pairs have many common neighbors[1]. More precisely, we use the following.

[1]This is a common situation that happens in applications of dependent random choice, though the exact way you weaken the conditions of Lemma 5.1 depends on the particular problem at hand.

Lemma 5.5. *Let G be an n-vertex graph with $an^{3/2}$ edges. Then G contains a subset of vertices U with $|U| = a$ such that for all $1 \le i \le \binom{a}{2}$, there are less than i pairs of vertices in U with fewer than i common neighbors in $V(G) \setminus U$.*

For example, this says that every pair of vertices of U has at least one common neighbor outside of U, and that there is at least one pair which has at least a common neighbors outside of U.

Add more intuition for the proof.

Proof. For simplicity we assume n is even, and by losing at most half of our edges we can assume that G is bipartite on $V_1 \cup V_2$ with $|V_1| = |V_2| = n/2$. Without loss of generality we can assume $\sum_{v \in V_1} d(v)^2 \le \sum_{v \in V_2} \sum d(v)^2$.

Let T be a random set obtained by including two vertices uniformly at random from V_1 with replacement. Let $W = N(T)$ and $X = |W|$. Similar to our computation before, we find

$$\mathbb{E}[X] = \sum_{v \in V_2} (d(v)/(n/2))^2 \ge 4n^{-2} \cdot (n/2)(an^{1/2})^2 = 2a^2.$$

Given distinct vertices $x, y \in V_2$, we define $f(x,y) = \frac{1}{|N_{V_1}(x,y)|}$ and we let $Y = \sum_{x,y \in W} f(x,y)$. Observe that

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{x,y \in V_2} f(x,y) \cdot \Pr[x,y \in W] = \sum_{x,y \in V_2} \frac{1}{|N_{V_1}(x,y)|} \cdot \left(\frac{|N_{V_1}(x,y)|}{n/2}\right)^2 = 4n^{-2} \sum_{x,y \in V_2} |N_{V_1}(x,y)| \\ &= 4n^{-2} \sum_{z \in V_1} \binom{d(z)}{2} \le 2n^{-2} \sum_{z \in V_1} d(z)^2 \le 2n^{-2} \sum_{z \in V_2} d(z)^2 = \frac{1}{2}\mathbb{E}[X].\end{aligned}$$

With this we see $\mathbb{E}[X - \mathbb{E}[X]/2 - Y] \ge 0$, and thus there exists a choice of T such that $X \ge Y$ and $X \ge \mathbb{E}[X]/2 \ge a^2$.

The trick now is to take $U \subseteq W$ a set of size exactly a uniformly at random, and let $Y' = \sum_{x,y \in U} f(x,y)$. In this case

$$\mathbb{E}[Y'] = \sum_{x,y \in W} f(x,y) \cdot \Pr[x,y \in U | x,y \in W] \le Y \cdot \frac{a(a-1)}{X(X-1)} \le X \cdot (a/X)^2 \le 1.$$

Thus there exists a choice of U such that $Y' \le 1$. We claim that such a U has the desired properties. Indeed, if there existed i pairs with fewer than i common neighbors, then this would immediately imply $Y' \ge i \cdot \frac{1}{i-1} > 1$, a contradiction. $\square$

Theorem 5.4 follows almost immediately from Lemma 5.5, and we omit its proof.

For our last result, we say that a graph F is *r-degenerate* if every subgraph of F contains a vertex of degree at most r. In this setting we can prove an embedding lemma analogous to Lemma 5.2.

Lemma 5.6. *Let G be a graph with vertex sets U_1, U_2 such that, for $k = 1, 2$, every subset of at most r vertices in U_k contains at least m common neighbors in U_{3-k}. Then G contains every r-degenerate bipartite graph H on m vertices.*

Proof. Let F_1 be an m-vertex r-degenerate bipartite graph on $V_1 \cup V_2$. By definition this means that there exists a vertex $v_1 \in F_1$ such that $d_{F_1}(v_1) \le r$, and that there is some $v_2 \in F_2 := F_1 - v_1$ with $d_{F_2}(v_2) \le r$ and so on. We now define a map $\phi : V_1 \cup V_2 \to U_1 \cup U_2$ with $\phi(V_i) \subseteq U_i$ as follows. Iteratively assume we have defined $\phi(v_m), \phi(v_{m-1}), \ldots, \phi(v_{q+1})$ and that $v_q \in V_i$. Since $S := N(v_q) \cap \{v_m, \ldots, v_{q+1}\}$ has at most r vertices by assumption, the set $\phi(S) \subseteq U_{3-i}$ has at least m common neighbors, so choose $\phi(v_q)$ to be any of these vertices that has yet to be assigned. It is not difficult to see that this gives the desired embedding. $\square$

Motivated by this lemma, we prove the following variant of Lemma 5.1.

Lemma 5.7. *Let $r, m \ge 2$ and let G be an n-vertex graph with at least $mn^{1-1/6r}$ edges. Then G contains two subsets U_1, U_2 such that, for $k = 1, 2$, every subset of r vertices in U_k has at least m common neighbors in U_{3-k}.*

Proof. The rough strategy of the proof is as follows. We will first apply Lemma 5.1 directly to obtain a large set U_1 such that every q-subset of U_1 (with $q > r$) has at least m common neighbors. We then mimic the proof of Lemma 5.1 by choosing a random set $T \subseteq U_1$ of size t and letting $U_2 = N(U_1)$. By choosing an appropriate value of t, the set U_2 will satisfy the condition. Moreover, if $q - t \ge r$, then for any r-subset $S \subseteq U_1$, the set $S \cup T$ has at least m common neighbors, all of which in particular lie in $N(T) = U_2$, so U_1 will also have the desired property.

We now being the formal argument. Apply Lemma 5.1 using $q = 3r$ instead of r, t to get a set U_1 such that every subset of size $3r$ has at least m common neighbors and such that

$$|U_1| \ge \frac{d^{3r}}{n^{3r-1}} - \binom{n}{3r}(m/n)^{3r} \ge m^{3r}n^{1/2} - m^r/(3r)! \ge mn^{1/2}.$$

Now let T be a set obtained by including $t = 2r$ vertices uniformly at random from U_1 with replacement, and let $U_2 = N(T)$. The probability that U_2 contains a set of r vertices which have fewer than m common neighbors in U_1 is at most

$$\binom{n}{r}(m/|U_1|)^{2r} \le \frac{1}{r!} < 1,$$

and in particular there exists a choice of T such that no r-subset of U_2 has fewer than m common neighbors. Note that for any r-subset $S \subseteq U_1$, the set $S \cup T$ has size at most $3r$ vertices, so by construction S has at least m common neighbors which lie in $N(T) = U_2$. Thus U_1, U_2 gives the desired result. $\square$

Combining these two lemmas immediately gives the following.

Theorem 5.8. *If F is an m-vertex r-degenerate graph, then*

$$\mathrm{ex}(n, F) < mn^{2-1/6r}.$$

We note that one can optimize the proof of Lemma 5.7 to improve the exponent of this theorem slightly (by using $(3 - 2\sqrt{2})r$ instead of $3r$ throughout). However, the end result is still weaker than the best known bound of $\mathrm{ex}(n, F) \le m^{1/2r} n^{2-1/4r}$ due to Alon, Krivelevich, and Sudakov [1], with their proof more or less being a slight refinement of the argument we gave.

As all of these examples illustrate: if you have a problem that could be magically solved if you had a large set of vertices U such that every r-set of U had many common neighbors, then a variant of dependent random choice might be worth trying out!

6 Strong Concentration Inequalities

As we have already seen in Section 4, there are many instances where one would like to show a random variable is concentrated around its expectation. Chebyshev's inequality is one tool which achieves this, and for arbitrary random variables this is essentially the best one can do. However, there are many specific kinds of random variables where one can get significantly stronger bounds. Here we present three results of this form in roughly increasing order of power.

We will omit the majority of the proofs of these results, focusing mostly on their applications. We refer the reader to the book of Dubhashi and Panconesi [19] for complete proofs. We note that the appendix of [19] consists of a very nice summary of these inequalities and many of their generalizations.

6.1 The Chernoff Bound

The Chernoff bound says that binomial random variables have exponential concentration around their means.

Theorem 6.1. *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables each with probability of success p, and let $X = \sum X_i$. Then for all $\lambda > 0$,*

$$\Pr[|X - pn| \ge \lambda pn] < 2e^{-\lambda^2 pn/2}.$$

Sketch of Proof. Observe that for all $\lambda, t > 0$, we have

$$\Pr[X \ge (1+\lambda)pn] = \Pr[e^{tX} \ge e^{t(1+\lambda)pn}] \le \mathbb{E}[e^{tX}]e^{-t(1+\lambda)pn},$$

with this last step using Markov's inequality. We note that $e^{tX} = \sum \frac{t^m \mathbb{E}[X^m]}{m!}$ is the *moment generating function* of X, and it is a common trick in probability to rephrase inequalities in terms of e^{tX}. And indeed, because the X_i are all independent, we have

$$\mathbb{E}[e^{tX}] = \prod \mathbb{E}[e^{tX_i}] = (e^t p + (1-p))^n.$$

Thus we are left with the problem of choosing t so that $\frac{e^t p + (1-p)}{e^{-t(1+\lambda)p}}$ is minimized. One can do this using calculus, and this will give $\Pr[X \ge (1+\lambda)pn] < e^{-\lambda^2 pn/2}$. The same argument gives $\Pr[X \le (1-\lambda)pn] < e^{-\lambda^2 pn/2}$, and combining these inequalities gives the desired result. $\square$

The Chernoff bound can be generalized, for example, by replacing the bernoulli random variables with any bounded random variable, see [19].

Given a hypergraph H and partition $V(H) = R \sqcup B$, define the discrepancy of the partition by

$$\operatorname{disc}(H, R, B) = \max_{e \in E(H)} ||e \cap R| - |e \cap B||,$$

and define the discrepancy of the hypergraph by $\operatorname{disc}(H) = \min_{R,B} \operatorname{disc}(H, R, B)$. In other words, $\operatorname{disc}(H)$ measures how well one can partition the vertex set so that each edge has about the same number of vertices from each part.

Theorem 6.2. *If H is an r-uniform hypergraph with m edges, then $\operatorname{disc}(H) \le 2\sqrt{r \log(2m)}$.*

If H is a clique on $2r - 1$ vertices, then $\operatorname{disc}(H) = r$ and $m \approx 4^r$, so this result is essentially best possible for general H.

Proof. Assign each vertex of H to R or B independently and with probability $\frac{1}{2}$. For $e \in E(H)$, let A_e be the event that

$$\left| |e \cap R| - \frac{1}{2}r \right| \ge \sqrt{r \log(2m)} = 2\sqrt{\frac{\log(2m)}{r}} \cdot \frac{1}{2}r.$$

Because $|e \cap R|$ has a binomial distribution, the Chernoff bound gives $\Pr[A_e] < 2e^{-\log(2m)} = m^{-1}$, and by a union bound we have $\Pr[\bigcup A_e] < 1$. Thus with positive probability, there exists a partition R, B such that none of the A_e occur. This means $\operatorname{disc}(H, R, B) \le 2\sqrt{r \log 2m}$, proving the result. □

Much more can be said about discrepancy problems, see [2, Chapter 13].

6.2 Martingales

We say that a sequence of real-valued random variables $X_0, X_1, \ldots$ is a *martingale* if $\mathbb{E}[X_{i+1}|X_i] = X_i$ for all i. One important class of Martingales, called Doob martingales, are defined as follows. Given random variables $Y_1, \ldots, Y_m$ and a real-valued function f, let

$$X_i = \mathbb{E}[f(Y_1, \ldots, Y_m)|Y_1, \ldots, Y_i].$$

It is not too difficult to show that any sequence of random variables X_i defined in this way is indeed a martingale.

One of the most common classes of (Doob) martingales in probabilistic combinatorics are the edge-exposure martingales. In this case, Y_i denotes the indicator random variable which is 1 if the ith pair of vertices in $G_{n,p}$ is an edge (where the pairs are ordered in some arbitrary way). Intuitively in this situation we think of revealing the edges of $G_{n,p}$ one at a time, and X_i denotes the value that we expect f to be after we reveal all of the remaining edges.

Let us look at the very concrete case of the edge-exposure martingale when $n = 3$ and f is the number of triangles in $G_{3,p}$. With this we have

$$X_0 = \mathbb{E}[f] = p^3, \; X_1 = \mathbb{E}[f|Y_1] = p^2 Y_1,$$

$$X_2 = \mathbb{E}[f|Y_1, Y_2] = pY_1Y_2,\ \ X_3 = \mathbb{E}[f|Y_1, Y_2, Y_3] = Y_1Y_2Y_3.$$

The main concentration result for martingales is Hoeffding's inequality.

Theorem 6.3 (Hoeffdings's inequality)**.** *Let $X_0, \ldots$ be a martingale with $|X_i - X_{i-1}| \leq \alpha_i$ for all i. Then for all $\lambda > 0$, we have*

$$\Pr[|X_m - X_0| \geq \lambda] < 2e^{\frac{-2\lambda^2}{\sum \alpha_i^2}}.$$

Sketch of Proof. Let $Y_i = X_i - X_{i-1}$. Similar to the proof of the Chernoff bound, we have

$$\Pr[X_m - X_0 \geq \lambda] = \Pr[e^{t(X_m - X_0)} \geq e^{t\lambda}] \leq \mathbb{E}[e^{t\sum_{i=1}^m Y_i}]e^{-t\lambda}.$$

We claim that this expectation is at most $e^{\frac{1}{8}t^2\sum_{i=1}^m \alpha_i^2}$. Indeed, we can use conditional expectations to write

$$\mathbb{E}[e^{t\sum_{i=1}^m Y_i}] = \mathbb{E}\left[\mathbb{E}[e^{t\sum_{i=1}^m Y_i}|X_0, \ldots, X_{m-1}]\right] = \mathbb{E}[e^{t\sum_{i=1}^{m-1} Y_i} \cdot \mathbb{E}\left[e^{tY_m}|X_0, \ldots, X_{m-1}]\right],$$

where this last step used that Y_i with $i < m$ is fixed given $X_0, \ldots, X_{m-1}$. Observe that conditional on $X_0, \ldots, X_{m-1}$, we have $\mathbb{E}[Y_m] = 0$ (due to the martingale property) and $|Y_m| \leq \alpha_m$ (due to the hypothesis of the theorem). One can show that for random variables of this form, the expected value of its moment generating function is at most $e^{\alpha_m^2 t^2/8}$. One gets the claim by repeating this argument inductively on the remaining terms.

In total, we have for any $t > 0$ that

$$\Pr[X_m - X_0 > \lambda\sqrt{m}] \leq e^{\frac{1}{8}t^2\sum \alpha_i^2 - t\lambda}.$$

Taking $t = 4\lambda/\sum \alpha_i^2$ gives $\Pr[X_m - X_0 \geq \lambda] < e^{\frac{-2\lambda^2}{\sum \alpha_i^2}}$. A symmetric argument shows $\Pr[X_m - X_0 \leq \lambda] < e^{\frac{-2\lambda^2}{\sum \alpha_i^2}}$ (this can also be seen by considering the martingale $X_i' := -X_i$ and applying the first inequality), which gives the result. $\square$

In the special case where $\alpha_i = 1$ for all i, this result is referred to as Azuma's inequality[1].

Corollary 6.4 (Azuma's inequality)**.** *Let $X_0, \ldots$ be a martingale which satisfies $|X_i - X_{i-1}| \leq 1$ for all i. Then for all $\lambda > 0$, we have*

$$\Pr[|X_m - X_0| \geq \lambda\sqrt{m}] < 2e^{-2\lambda^2}.$$

There are many generalizations of the Hoeffding's inequality which weakens the hypothesis that $|X_{i+1} - X_i| \leq \alpha_i$. For example, it suffices to have that this difference holds in expectation, or that it holds with high probability. Again, see [19] for details.

One application of Azuma's inequality is the following.

[1]The naming convention for these inequalities are all over the place: some people call these Hoeffding's inequalities, others Azuma (which is probably the most popular name in the combinaotrics community), some Azuma-Hoeffding, and yet others Hoeffding-Azuma.

Proposition 6.5. *We have*

$$\Pr[|\chi(G_{n,p}) - \mathbb{E}[\chi(G_{n,p})| \geq \lambda\sqrt{n}] < 2e^{-\lambda^2/2}.$$

We note that this result tells us that $\chi(G_{n,p})$ is concentrated around its expectation, but it gives no indication of what this expectation is. This is a common phenomenon when applying concentration inequalities.

Proof. We consider a vertex-exposure martingale, i.e. a Doob martingale $f(Y_1, \dots, Y_n)$ where Y_i is the set of vertices $j > i$ in $G_{n,p}$ which are adjacent to i. In particular, taking $f = \chi$ and $X_i = \mathbb{E}[f|Y_1, \dots, Y_i]$ gives $X_0 = \mathbb{E}[\chi(G_{n,p})]$ and $X_n = \chi(G_{n,p})$. It is clear that each time we reveal a set Y_i that the expected chromatic number changes by at most 1, i.e. $|X_{i+1} - X_i| \leq 1$ for all i. Thus Azuma's inequality applies, giving the result.

□

We note that one could try and prove this result using an edge-exposure martingale instead of a vertex-exposure martigale, but this approach gives essentially trivial bounds. In general, when using martingales you want to reveal information in as few rounds as possible, while also making it so that the information you reveal can't dramatically change your function each round.

While Proposition 6.5 says nothing about $\mathbb{E}[\chi(G_{n,p})]$, it is well known that this value is asymptotic to $\frac{n}{2\log_{1/(1-p)} n}$ for any fixed p. This was first proven by Bollobás using a clever martingale argument. Much more can be said about $\chi(G_{n,p})$, see for example the paper by Heckel and Riordan [33] which, in addition to surveying many of the known results on $\chi(G_{n,p})$, shows that the concentration in Proposition 6.5 is in some sense close to best possible.

There are a number of variants of all of the concentration inequalities stated in this chapter. One particular version of Azuma that we will need at some point is the following.

Lemma 6.6 ([44])**.** *Let $X_0, \dots$ be a martingale which satisfies $|X_i - X_{i-1}| \leq \alpha$ for all i. Then for all $\delta \in [0, 1]$, we have*

$$\Pr[|X_m - X_0| \geq \delta\alpha m] < e^{-\delta^2\alpha m/6c}.$$

Note that $|X_m - X_0| \leq \alpha m$ deterministically, so $\delta\alpha m$ is at least a δ fraction of the mean of $X_m - X_0$, and as such this is referred to as the "multiplicative Azuma inequality" (since its error term is multiplicative relative to the expectation as opposed to additive).

6.3 Talagrand's Inequality

Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$ be a product of probability spaces. For $\alpha = (\alpha_1, \dots, \alpha_n)$ a vector of non-negative real numbers, we define the *weighted Hamming distance* d_α on Ω by $d_\alpha(x, y) = \sum_{i:x_i \neq y_i} \alpha_i$. For example, $\alpha = (1, \dots, 1)$ gives the usual Hamming distance on product spaces.

Given α as above, a set $A \subseteq \Omega$, and a non-negative number t, we define

$$A_{\alpha,t} = \{x : \exists y \in A,\ d_\alpha(x, y) \leq t\}.$$

That is, $A_{\alpha,t}$ is the set of points in Ω which are within distance t of A. We also define $\overline{A} = \Omega \backslash A$. Our goal is to prove "isoperimetric" inequalities which state that, for any $A \subseteq \Omega$, we have

$$\Pr[A] \cdot \Pr[\overline{A_{\alpha,t}}] \leq f(t),$$

where f is some rapidly shrinking function. Isoperimetric inequalities are intimately related to concentration inequalities. For example, a corollary of an inequality as above is that if $\Pr[A] \geq \frac{1}{2}$, then $\Pr[\overline{A_{\alpha,t}}] \leq 2f(t)$ (i.e., most of Ω is concentrated around A). On the other hand, one can prove isoperimetric inequalities by using concentration inequalities.

Proposition 6.7. *For Ω a product space, $A \subseteq \Omega$, and α such that $\sum \alpha_i^2 = 1$, we have for all t that*

$$\Pr[A] \Pr[\overline{A_{\alpha,t}}] \leq 4e^{-t^2}$$

.

Proof. Define the function $f : \Omega \to \mathbb{R}$ by $f(y) = d_\alpha(y, A)$. Let $Y = (Y_1, \ldots, Y_n)$ be chosen according to the probability distribution on Ω and let $X_i = \mathbb{E}[f(Y)|Y_1, \ldots, Y_i]$. Observe that $X_n = 0$ iff $Y \in A$ and $X_n > t$ iff $Y \in \overline{A_{\alpha,t}}$, and also that $|X_i - X_{i-1}| \leq \alpha_i$. Thus Hoeffding's inequality implies

$$\begin{aligned}
\Pr[A] \Pr[\overline{A_{\alpha,t}}] &= \Pr[X_m = 0] \Pr[X_m > t] \\
&\leq \Pr[|X_m - X_0| \geq X_0] \Pr[|X_m - X_0| > t - X_0] \\
&\leq 4e^{-2X_0^2 - 2(t-X_0)^2} \leq 4e^{-t^2},
\end{aligned}$$

where this last step used that the exponent is maximized when $X_0 = \frac{1}{2}t$. □

A remarkable result of Talagrand shows that Proposition 6.7 essentially holds even when comparing A with the set of points which are at least distance t from A for *some* choice of α.

Theorem 6.8 (Talagrand's inequality)**.** *For all $A \subseteq \Omega$ and $t \geq 0$, we have*

$$\Pr[A] \Pr\left[\overline{\bigcap_\alpha A_{\alpha,t}}\right] \leq e^{-t^2/4},$$

where the intersection ranges over all α with $\sum \alpha_i^2 = 1$.

Again we emphasize that $\overline{\bigcap_\alpha A_{\alpha,t}}$ can be much larger than $\overline{A_{\alpha,t}}$ for any given α, but still essentially the same bound as in Proposition 6.7 holds. We omit the proof of Theorem 6.8, and we refer the reader to [2] for a direct proof, and to [19] for a longer, but perhaps more enlightening argument.

We note that Talagrand's inequality is often stated in the following equivalent form: Given $x \in \Omega$ and $A \subseteq \Omega$, define $d'(x, A) = \min_{y \in A} \max_\alpha d_\alpha(x, y)$, where the maximum ranges over all α with $\sum \alpha_i^2 = 1$. Note that having $d'(x, A) \leq t$ is equivalent to saying that for all α there exist $y \in A$ with $d_\alpha(x, y) \leq t$, which is equivalent to saying $x \in \bigcap_\alpha A_{\alpha,t}$. Thus Theorem 6.8 can be seen as an isoperemetric inequality with respect to the pseudo-distance d'.

Talagrand's inequality has a number of applications to concentration of random variables. One particular application is for certifiable functions. For a function $s : \mathbb{R} \to \mathbb{N}$, we say that a real-valued function f defined on a product space Ω is *s-certifiable* if having $f(x) \geq c$ implies that there exists a set $I \subseteq [n]$ of size $s(c)$ such that $f(y) \geq c$ whenever $y_i = x_i$ for all $i \in I$ (that is, the values in position I "certify" that $f(x) \geq c$).

For example, if $f(x) = |\{i : x_i \neq 0\}|$, then f is s-certifiable with $s(c) = c$, since $f(x) \geq c$ implies there exist c coordinates with $x_i \neq 0$, and any y which agrees with x on these coordinates satisfies $f(y) \geq c$. Lastly, we say that a function f is Lipschitz if $|f(x) - f(y)| \leq 1$ whenever x, y differ in at most one coordinate.

Corollary 6.9. *If f is an s-certifiable Lipschitz function on the product space Ω and X is chosen according to the probability space Ω, then for all m and $t > 0$ we have*

$$\Pr[f(X) < m - t\sqrt{s(m)}] \Pr[f(X) \geq m] \leq e^{-t^2/4}.$$

Proof. Let $A = \{x : f(x) < m - t\sqrt{s(m)}\}$. We claim that $\bigcap_\alpha A_{\alpha,t} \subseteq \{y : f(y) < m\}$.

Assume for contradiction that $y \in \bigcap_\alpha A_{\alpha,t}$ and $f(y) \geq m$. Because f is s-certifiable, there exists a set of $s(m)$ indices I which certifies $f(y) \geq m$. Let $\alpha'_i = \frac{1}{\sqrt{s(m)}}$ if $i \in I$ and $\alpha'_i = 0$ otherwise. Because $y \in \bigcap_\alpha A_{\alpha,t} \subseteq A_{\alpha',t}$, there exists some $x \in A$ such that $d_{\alpha'}(x, y) \leq t$, i.e. such that restricted to I, the vectors x, y differ in at most $t\sqrt{s(m)}$ coordinates. Let z be defined by $z_i = y_i$ if $i \in I$ and $z_i = x_i$ otherwise. Then $f(z) \geq m$ by definition of I, and f being Lipschitz implies $f(x) \geq f(z) - t\sqrt{s(m)} \geq m - t\sqrt{s(m)}$, contradicting $x \in A$. This proves the claim.

The contrapositive of the claim implies $\{y : f(y) \geq m\} \subseteq \overline{\bigcap_\alpha A_{\alpha,t}}$, so the first result follows from Talagrand's inequality. $\square$

We emphasize that Proposition 6.7 is too weak to prove Corollary 6.9: we genuinely have to make use of the fact that Talagrand's inequality allows us to choose a different distance function for each choice of y.

In most applications, one applies Corollary 6.9 where either m is a median, i.e. $\Pr[f(X) \geq m] = \frac{1}{2}$, or where $m - t\sqrt{s(m)}$ is a median. While medians are hard to estimate directly, a concentration result like that of Corollary 6.9 can usually be used to show that the median and expectation must be close to each other, see for example [19, Problem 11.4].

As an application, let $X = (X_1, \ldots, X_n)$ be a random vector with each X_i distributed uniformly on $[0, 1]$. Let $f(X)$ denote the length of a longest increasing subsequence, i.e. the largest k such that there exist indices with $X_{i_1} < X_{i_2} < \cdots < X_{i_k}$. Note that f is Lipschitz and is s-certifiable with $s(c) = c$, so if m is a median we conclude $\Pr[f(X) < m - t\sqrt{m}] \leq 2e^{-t^2/4}$. It is well known that $\mathbb{E}[f(X)] \sim 2\sqrt{n}$, so at least heuristically, this argument suggests $f(X)$ is highly concentrated around $2\sqrt{n} + \Theta(n^{1/4})$ (and it's not hard to make this more precise). In contrast, if one attempted to get concentration results for $f(X)$ by utilizing martingales, one would conclude that $f(X)$ is highly concentrated around $\Theta(n^{1/2})$, which is significantly weaker. One can literally dedicate an entire book to the longest increasing subsequence problem, see Romik [53] for more on this topic.

7 Coupling

It is often the case that one can understand a random variable X by comparing it to a "similar" random variable Y which is easier to do calculations for. One way to do this to form a *coupling*, i.e. a pair of random variables (X', Y') such that X', Y' have the same distribution as X, Y, respectively, and such that X', Y' have some (nice) relation between them. We start with a simple but non-trivial example.

Let S^n denote a simple random walk of length n, i.e. S^n is a random vector $(S_0^n, S_1^n, \ldots, S_n^n)$ where $S_0^n = 0$, and $\Pr[S_i^n = S_{i-1}^n + 1] = \Pr[S_i^n = S_{i-1}^n - 1] = \frac{1}{2}$. For n even, let T^n denote a random walk after conditioning on having $T_n^n = 0$ (i.e. we uniformly at random pick a walk which returns to to 0 at the end of the walk). It is easy to show via Chernoff boudns that S_t^n is likely to be within roughly $\sqrt{t}$ of 0 for any given value t. While Chernoff bounds don't apply to the random variables T_t^n, intuitively the same conclusion should also hold for T_t^n, since the condition of $T_n^n = 0$ should force T_t^n to be closer to the origin than S_t^n in general. It is possible to make this intuition rigorous, allowing one to bootstrap bounds of S_t^n to T_t^n.

Proposition 7.1. *For n even and all s, t, we have* $\Pr[|T_t^n| \geq s] \leq \Pr[|S_t^n| \geq s]$.

Proof. Our goal is to define a new random vector R^n such that (1) R^n has the same distribution as T^n, and (2) $|R_t^n| \leq |S_t^n|$ for all t. From this the result will quickly follow. Intuitively, we will define R_t^n in rounds by flipping biased coins. If the tth coin lands heads, then R_{t+1}^n moves towards 0, and if it lands tails, it moves towards/away from 0 if and only if S_{t+1}^n moves towards/away from 0. Such a process will always satisfy (2), and it will satisfy (1) by choosing the probability of our biased coins appropriately.

To this end, set $R_0^n = 0$. Given R_t^n, we define a random variable Y_t (which will be our biased coin flips) that equals 1 with probability $\frac{|R_t^n|}{n-t}$ and is 0 otherwise. If $Y_t = 1$, we set $R_{t+1}^n = R_t^n \pm 1$ such that $|R_t^n| > |R_{t+1}^n|$ (i.e. such that R^n moves towards 0; note that this is well defined since $Y_t = 1$ implies $R_t^n \neq 0$). If $Y_t = 0$ and $R_t^n \neq 0$, then we set $R_{t+1}^n = R_t^n \pm 1$ such that $|R_t^n| > |R_{t+1}^n|$ if and only if $|S_t^n| > |S_{t+1}^n|$ (i.e. R^n move away/towards 0 if S^n moves away/towards 0). If $R_t^n = 0$ then we set $R_t^n = \pm 1$ with equal probability.

It is straightforward to see that (2) is achieved from this process[1]. It is not difficult to prove that for T^n, we have that $|T_t^n| > |T_{t+1}^n|$ happens with probability[2] $\frac{\frac{1}{2}(n-t+|T_t^n|)}{n-t}$. One can check that R^n has $|R_t^n| > |R_{t+1}^n|$ with probability $\frac{\frac{1}{2}(n-t+|R_t^n|)}{n-t}$, so we conclude (1) and hence the result. □

Recall that an F*-factor* in a graph G is a collection of vertex disjoint copies of F such that every vertex is in one of these copies of F. By a similar argument as in Theorem 9.2, one can show that $G_{n,p}$ contains a K_r-factor provided $r|n$ and $p \gg n^{-1/\binom{r}{2}} \log n$. Intuitively, it seems reasonable that the set of K_r's in $G_{n,p}$ should be distributed like the hyperedges of $G_{n,\pi}^r$ where $\pi = p^{\binom{r}{2}}$ (at the very least, the expected number of K_r's in $G_{n,p}$ is equal to the expected number of

[1] Any time S^n moves towards 0, R^n does as well, except when $R_t^n = 0$. In this case S_t^n must be an even distance away from 0, so after one step R_t^n is still at least as close to 0.

[2] Given T_t^n, we still need to make $\frac{1}{2}(n - t + |T_t^n|)$ steps in the direction of 0 and $\frac{1}{2}(n - t - |T_t^n|)$ in the direction away from 0.

hyperedges in $G^r_{n,\pi}$). If this were true, then $G_{n,p}$ would contain a K_r-factor when $G^r_{n,\pi}$ contains a perfect matching, and Theorem 9.2 says this should happen when $p^{\binom{r}{2}} \approx \pi \gg n^{-1}\log n$, which implies that taking $p \gg n^{-1/\binom{r}{2}} \log^{1/\binom{r}{2}} n$ should suffice. And indeed, Johansson, Kahn, and Vu [36] proved that this is the threshold for K_r-factors in $G_{n,p}$ using a somewhat involved argument. A nice coupling result of Riordan [51] will allow us to conclude the result in a much easier way.

Let $(V_1, E_1), \ldots, (V_{\binom{n}{r}}, E_{\binom{n}{r}})$ be an arbitrary ordering of all the K_r's in K_n. To prove our desired coupling, we would like to construct a pair of random variables (G, H) such that (1) $G \sim G_{n,p}$ and $H \sim G^r_{n,\pi}$ with $\pi \approx p^{\binom{r}{2}}$, and such that (2) every hyperedge in H is a K_r in G. Note that (2) means that H containing a perfect matching implies that G has a K_r-factor. Let us first consider the following (very, very) naive attempt at this coupling.

Algorithm 1. Generate a random graph $G \sim G_{n,p}$. Let H be an initially empty r-graph on $[n]$. For each i with $E_i \subseteq G$, add V_i as a hyperedge to H. Output (G, H).

This algorithm definitely satisfies (2), but it completely fails at (1). Indeed, let A_i denote the event that $E_i \subseteq G$ (i.e. the event that V_i is a hyperedge in H), and assume V_1, V_2 have at least two vertices in common. Then $\Pr[A_2|A_1] \ge p^{\binom{r}{2}-1}$ and $\Pr[A_2|\overline{A_1}] < p^{\binom{r}{2}}$. But to have $H \sim G^r_{n,\pi}$ we would, in particular, need these two probabilities to equal each other. Thus we'll need to consider a somewhat more complicated algorithm. As before, let (V_i, E_i) be the K_r's in K_n, and let π be a parameter which will be approximately $p^{\binom{r}{2}}$.

Algorithm 2. Generate a random graph $G \sim G_{n,p}$ and an initially empty hypergraph H on $[n]$. We proceed in $\binom{n}{r}$ rounds as follows. For the ith round, let π_i be the conditional probability of having $E_i \subseteq G$ given all the information from the previous rounds.

- If $\pi_i < \pi$, then with probability π we add V_i to H.
- If $\pi_i \ge \pi$, then with probability $\frac{\pi}{\pi_i}$ we test whether $E_i \subseteq G$, and if so, we add V_i to H. Otherwise[1] we declare this hyperedge to be absent in H.

We note that the $\pi_i \ge \pi$ case of Algorithm 2 is similar in spirit to the proof of Proposition 7.1: each round we flip a coin which is biased based off of the current information we have. If the coin lands heads we do something to H independent of G, and otherwise we have H behave "in the same way" as G.

For Algorithm 2, it is not difficult to see that $H \sim G^r_{n,\pi}$. Unfortunately, if $\pi_i < \pi$, then it is possible that H contains edges which are not K_r's in G, i.e. the coupling could fail to satisfy (2). The key insight is that for applications, it suffices to have (2) be satisfied with high probability, which will turn out to be the case.

To try and convince ourselves that this algorithm has a chance of winning even when $\pi_i < \pi$, let's consider the most dangerous situation, namely that $\pi_i = 0$. It is not too hard to see that

[1]Note that with probability $1 - \frac{\pi}{\pi_i}$ we do not reveal any additional information about E_i. When working with random objects, it is usually best to reveal as little information as possible in order to "preserve" the randomness of your object.

$\pi_i = 0$ if and only if there exists some $j < i$ such that (a) we revealed that $E_j \not\subseteq E(G)$ and (b) every edge of $E_j \setminus E_i$ has been revealed to be in G. If this situation happens and if the algorithm adds V_i to H, then the coupling fails to satisfy (2). However, when this happens, every graph edge of E_j is contained in a hyperedge of H (by (b) and $V_i \in E(H)$) and V_j is not a hyperedge of H (by (a)). Thus the probability of this situation happening is at most the probability of $H \sim G^r_{n,\pi}$ containing such a configuration. These configurations can essentially be described as follows.

Lemma 7.2. *If H is an r-graph with $r \geq 4$ which contains a set of r vertices $V \notin E(H)$ such that every pair of V is contained in a hyperedge of H, then H contains a subgraph F which has $e(F) \leq \binom{r}{2}$ and $|V(F)| \leq (r-1)e(F) - 1$.*

This statement is false for $r = 3$. Indeed, one could take H to be the loose triangle with edges $\{1, 2, 4\}, \{2, 3, 5\}, \{1, 3, 6\}$ which satisfies the hypothesis of Lemma 7.2 with $V = \{1, 2, 3\}$ but which fails to satisfy the conclusion.

Proof. Let $V_1, \ldots, V_t$ be hyperedges such that every pair of V is contained in some V_i. By throwing away redundant hyperedges, we can assume that $|V_i \cap V| \geq 2$ for all i and that $t \leq \binom{r}{2}$. Let $F \subseteq H$ be the hypergraph with hyperedges $V_1, \ldots, V_t$

First assume $|V_1 \cap V_2| \geq 2$. Then $V_1 \cup V_2$ consists of at most $2r - 2$ vertices, and it is not difficult to see that it is possible to order the remaining sets so that $|V_i \setminus \bigcup_{j<i} V_j| \leq r - 1$ and that $|V_t \setminus \bigcup_{j<t} V_j| \leq r - 2$. In total this implies that F has the desired properties.

Thus we can assume that $|V_i \cap V_j| \leq 1$ for all i, j. This means every pair of V is covered by some unique V_i, so $t = e(F) = \binom{r}{2}$ and the number of vertices of F is at most $r + (r-2)e(F) = (r-1)e(F) - (e(F) - r) \leq (r-1)e(F) - 1$ since $\binom{r}{2} - r \geq 1$ for $r \geq 4$. □

Lemma 7.3. *For $r \geq 4$, if $H \sim G^r_{n,\pi}$ and $\pi \leq n^{-(r-1)+o(1)}$, then a.a.s. H does not contain a set V as in Lemma 7.2.*

Proof. If H did contain such a set V, then it must contain a subgraph F as in Lemma 7.2. Up to isomorphism, there are only finitely many subgraphs that F could be, and for each of these the expected number of copies of F in H is at most

$$O(\pi^{e(F)} n^{|V(F)|}) = O(\pi^{e(F)} n^{(r-1)e(F)-1}) = o(1).$$

We conclude the result by Markov's inequality. □

We note that for $\pi \approx p^{\binom{r}{2}}$ this lemma applies when $p \approx n^{-2/r}$. Thus when p is about this value, none of the "bad" configurations of Lemma 7.2 are likely to appear, and in this regime we have the following.

Theorem 7.4 ([51])**.** *For $r \geq 4$ and $p \leq n^{-2/r+o(1)}$, there exists some $\pi \sim p^{\binom{r}{2}}$ such that Algorithm 2 produces a pair (G, H) with $G \sim G_{n,p}$, $H \sim G^r_{n,\pi}$, and such that a.a.s. every hyperedge of H is the vertex set of a K_r in G.*

We emphasize that the theorem as stated does not cover the case $r = 3$. However, Heckel [32] showed that the same conclusion does hold for $r = 3$ by using a slightly different coupling.

For the proof of Theorem 7.4, we will need a standard result known as Harris' inequality (also referred to as Kleitman's inequality).

Lemma 7.5 (Harris' Inequality)**.** *Let $f, g, h : \mathbb{R}^n \to \mathbb{R}$ be functions such that f, g are non-decreasing and h is non-increasing. Let $X = (X_1, \ldots, X_n)$ be a random vector such that the X_i's are mutually independent. Then*

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)],$$

$$\mathbb{E}[f(X)h(X)] \leq \mathbb{E}[f(X)]\mathbb{E}[h(X)].$$

Proof. For $n = 1$, we deterministically have

$$(f(y) - f(z))(g(y) - g(z)) \geq 0 \geq (f(y) - f(z))(h(y) - h(z)).$$

Thus if Y, Z are independent random variables with the same distribution as $X = X_1$, the first inequality implies

$$0 \leq \mathbb{E}[f(Y)g(Y) + f(Z)g(Z) - f(Y)g(Z) - f(Z)g(Y)] = 2\mathbb{E}[f(X)g(X)] - 2\mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

This gives the first bound, and the second bound follows from an identical argument.

Assume the result has been proven up to some $n > 1$. By the inductive hypothesis and the $n = 1$ case applied to $f'(X_1) := \mathbb{E}[f(X)|X_1]$ and $g'(X_1) = \mathbb{E}[g(X)|X_1]$, we find

$$\begin{aligned}\mathbb{E}[f(X)g(X)] &= \mathbb{E}[\mathbb{E}[f(X)g(X)|X_1]] \geq \mathbb{E}[\mathbb{E}[f(X)|X_1] \cdot \mathbb{E}[g(X)|X_1]] \\ &= \mathbb{E}[f'(X_1)g'(X_1)] \geq \mathbb{E}[f'(X_1)]\mathbb{E}[g'(X_1)] = \mathbb{E}[f(X)]\mathbb{E}[g(X)].\end{aligned}$$

This proves the first inequality, and the second follows from an identical argument. □

The main application of Harris' inequality is when f is an indicator function. More precisely, we say that a set system $\mathcal{A} \subseteq 2^{[n]}$ is an *upset* if $A \in \mathcal{A}$ implies $B \in \mathcal{A}$ for all $B \supseteq A$, and we similarly define what it means for $\mathcal{A}$ to be a *downset*.

Corollary 7.6. *Let $\mathcal{A}, \mathcal{B}$ be upsets and $\mathcal{C}$ a downset of $[n]$, and let $S \subseteq [n]$ be obtained by including each element i independently and with probability p_i. Then*

$$\Pr[S \in \mathcal{A} \cap \mathcal{B}] \geq \Pr[S \in \mathcal{A}] \Pr[S \in \mathcal{B}],$$

$$\Pr[S \in \mathcal{A} \cap \mathcal{C}] \leq \Pr[S \in \mathcal{A}] \Pr[S \in \mathcal{C}].$$

Proof. Define $f : \mathbb{R}^n \to \mathbb{R}$ by having $f(x) = 1$ if $\{i : x_i > 0\} \in \mathcal{A}$ and $f(x) = 0$ otherwise. Similarly define g, h with respect to $\mathcal{B}, \mathcal{C}$. The result follows from Harris' inequality by letting $X = (X_1, \ldots, X_n)$ with the X_i being independent Bernoulli random with probability p_i. □

Proof of Theorem 7.4. Again, the only cases where the algorithm can fail is when π_i is small, so let us try and lower bound this quantity in terms of the (random) information we have at the ith step. Let Y be the set of "yes" indices j such that we have revealed that $E_j \subseteq E(G)$ and let N be the set of "no" indices such that we have revealed $E_j \not\subseteq E(G)$. Let $R = \bigcup_{j\in Y} E_j$ be the set of revealed edges, and let G' be the random graph which contains all of the edges of R, and which contains any $e \notin R$ independently and with probability p. Let $E'_j = E_j \setminus R$, and let A'_j be the event that $E'_j \subseteq G'$. It is not too hard to see that in total we have

$$\pi_i = \Pr[A'_i | \bigcap_{j\in N} \overline{A'_j}].$$

Define

$$D_0 = \bigcap_{j\in N,\ E_j\cap E_i=\emptyset} \overline{A'_j}, \qquad D_1 = \bigcap_{j\in N,\ E_j\cap E_i\neq\emptyset} \overline{A'_j}.$$

Intuitively D_0 shouldn't really influence π_i, and we can prove this using Harris' inequality. First note that

$$\pi_i = \Pr[A'_i|D_0\cap D_1] \ge \Pr[A'_i\cap D_1|D_0] = \Pr[A'_i|D_0] - \Pr[A'_i\cap\overline{D_1}|D_0] = \Pr[A'_i] - \Pr[A'_i\cap\overline{D_1}|D_0],$$

where this last step used that A'_i and D_0 are independent. Observe that A'_j is an upset for all j (i.e. A'_j is achieved precisely when the random set $E(G')$ is an element of an appropriately defined upset), D_0 is a downset (since complements of upsets are downsets, and downsets/upsets are preserved under intersection), and $A'_i\cap\overline{D_1}$ is an upset. Thus by Harris' inequality, we have[1]

$$\Pr[A'_i] - \Pr[A'_i\cap\overline{D_1}|D_0] = \Pr[A'_i] - \frac{\Pr[A'_i\cap\overline{D_1}\cap D_0]}{\Pr[D_0]} \ge \Pr[A'_i] - \Pr[A'_i\cap\overline{D_1}].$$

Now let $N_1 = \{j\in N : E_j\cap E_i \neq \emptyset\}$. Note that $\overline{D_1} = \bigcup_{j\in N_1} A'_j$, so by a union bound we have

$$\pi_i \ge \Pr[A'_i] - \sum_{j\in N_1}\Pr[A'_j\cap A'_i] = p^{|E'_i|} - \sum_{j\in N_1} p^{|E'_i\cup E'_j|} = p^{|E'_i|}(1-Q_i) \ge p^{\binom{r}{2}}(1-Q_i),$$

where

$$Q_i := \sum_{j\in N_1} p^{|E_j\setminus(E_i\cup R)|}.$$

Let Δ denote the maximum degree of G. We next prove a (somewhat imprecise) claim.

Claim 7.7. *Either $\Delta > n^{o(1)}$, or for all i, either $Q_i = o(1)$ or $V_i\in E(H)$ implies H contains a configuration as in Lemma 7.2.*

Proof. Assume $\Delta \le n^{o(1)}$ and consider some index i. Given j, let K_j denote the graph on V_j with edge set $E_i\cup R$, and let $C_1,\dots,C_{k+1}$ with $k\ge 0$ denote the connected components of K_j, say with $|V(C_\ell)| = r_\ell$ for all ℓ. Observe that $|E_j\setminus(E_i\cup R)|$ is at least the number of edges which aren't contained in any K_j component, i.e.

$$|E_j\setminus(E_i\cup R)| \ge \binom{r}{2} - \sum\binom{r_\ell}{2} \ge \binom{r}{2} - \binom{r-k}{2},$$

[1] This same sort of argument is essentially what you need to do prove Janson's inequality Theorem 3.6.

where this last inequality holds since if there are two terms with $r_\ell \geq 2$, then one can adjust these two terms to get a stronger bound. Because K_j is a graph using edges of $R \cup E_i \subseteq G$, we have that the number of $j \in N_1$ such that K_j has $k+1$ components is at most $rn^k\Delta^{r-k-1} = n^{k+o(1)}$, where the factor of r comes from the fact that $j \in N_1$ implies that K_j contains at least one vertex of V_i since E_i, E_j intersect in at least one edge.

In total then, the contribution to Q_i coming from j such that K_j has $k+1 \geq 2$ components is at most

$$\sum_{k=1}^{r-2} n^{k+o(1)} p^{\binom{r}{2}-\binom{r-k}{2}} = o(1),$$

where the equality follows from a simple calculation. Thus it remains to show that the contribution from terms with K_j connected is small. Because there are only $r\Delta^{r-1} = n^{o(1)}$ such terms, a similar argument shows that the contribution is negligible for terms with $e(K_j) < \binom{r}{2}$. The only non-trivial case then is when $e(K_j) = \binom{r}{2}$, i.e. when every edge of V_j is contained in $R \cup E_i$. In this case, $V_i \in E(H)$ implies that H contains a configuration as in Lemma 7.2. Thus for all i, either this happens or $Q_i = o(1)$, proving the result. □

Since $H \sim G^r_{n,\pi}$, we have that the expected degree of every vertex is roughly $\pi n^{r-1} = n^{o(1)}$. Thus if $\mathcal{B}_1$ is the "bad" event that $\Delta > n^{o(1)}$, then by the Chernoff bound we have $\Pr[\mathcal{B}_1] = o(1)$. Similarly if $\mathcal{B}_2$ is the event that H contains one of the configurations as in Lemma 7.2, then $\Pr[\mathcal{B}_2] = o(1)$ by Lemma 7.3.

We are now ready to complete the proof. Recall that the theorem claims the result holds for *some* $\pi \sim p^{\binom{r}{2}}$, so it suffices to prove it for $\pi = p^{\binom{r}{2}}(1 - o(1))$ where the $o(1)$ term is the upper bound for Q_i from the claim. Now all we have to do is verify that with this choice, a.a.s. every hyperedge of H is a K_r in G. The only way this can fail is if there exists an i with $\pi_i < \pi$ such that V_i is added as a hyperedge to H. By the previous claim and our choice of π, this is only possible if $\mathcal{B}_1 \cup \mathcal{B}_2$ occurs. As these occur with probability $o(1)$, we conclude the result. □

As noted previously, the proof of Theorem 7.4 does not go through for $r = 3$ due to the existence of loose triangles, but Heckel [32] managed to get around this issue. Essentially the idea of her proof is to first do a coupling on edges of H and G which are in loose triangles and then to run Riordan's argument.

It is also proven in [51] that one can to some extent generalize this approach to finding F-factors for sufficiently nice F. In this setting, H is not exactly a uniform hypergraph, but instead a collection of copies of F in K_n chosen with some probability π.

8 Random Algebraic Constructions

One can easily extend Theorem 1.4 to hypergraphs as follows.

Theorem 8.1. *Let F be an r-graph with v vertices and $e \geq r$ edges. If $e \geq v$, then*

$$\mathrm{ex}(n, F) = \Omega_v(n^{r-\frac{v-r}{e-1}}).$$

Sketch of Proof. Consider $G^r_{n,p}$, which in expectation has about pn^r edges and $p^{e(F)}n^{|V(F)|}$ copies of F. At $p = Cn^{-\frac{v-r}{e-1}}$ for some large constant C this first quantity is much larger than the second, so we can delete an edge from each copy of F to give the result. □

One way you could try and improve upon this argument is to delete edges which are in many copies of F. In $G^r_{n,p}$ this is too much to ask for, but it is possible to do this in other random hypergraph models. In particular, if our random model contains some algebraic structure, then it is often the case that edges will either be in many copies of F or almost none. We look at a few examples of this phenomenon.

8.1 Random Multilinear Maps

The problem of determining the Turán number of $K^r_{2,\ldots,2}$, the complete r-partite r-graph with each part having size 2, is called the Erdős box problem. Theorem 8.1 gives a lower bound of $n^{r-\frac{r}{2^r-1}}$, and for certain values of r this lower bound was improved by Gunderson, Rödl, and Sidorenko [31]. This result was significantly improved by Conlon, Pohoata, and Zakharov[15] who gave a polynomial improvement to the bound of Theorem 8.1 for all values of r.

Theorem 8.2 ([15])**.** *For all $r \geq 2$, we have*

$$\mathrm{ex}(n, K^r_{2,\ldots,2}) = \Omega(n^{r-\lceil\frac{2^r-1}{r}\rceil^{-1}}).$$

Note that r never[1] divides $2^r - 1$, so this does always give a polynomial improvemnt to Theorem 1.4.

We prove this result by considering a random hypergraph based off of multilinear maps. Recall that if $V_1, \ldots, V_r$ are vector spaces over $\mathbb{F}_q$, then a map $T : V_1 \times \cdots \times V_r \to \mathbb{F}_q$ is said to be *multilinear* if the one dimensional function $f(x) = T(v_1, \cdots, v_{i-1}, x, v_{i+1}, \cdots, v_r)$ is linear for all i and any choice of v_j. Note that there are only finitely many such maps over $\mathbb{F}_q$ if $V_1, \ldots, V_r$ are finite dimensional, so in this setting we can talk about choosing such a T uniformly at random.

Let $s = \lceil\frac{2^r-1}{r}\rceil$, and let $V_1, \ldots, V_r$ be copies of $\mathbb{F}^s_q$ with q a large prime power. Given a multilinear map T, let H_T denote the r-partite r-graph on $V_1 \cup \cdots \cup V_r$ with $\{v_1, \ldots, v_r\} \in E(H_T)$ if and only if $T(v_1, \ldots, v_r) = 1$ (here and throughout we assume $v_i \in V_i$ for all i). The proof relies on the following three results.

[1] If r is prime then $2^r - 1 \equiv 2 - 1 \mod r$ by Fermat's little theorem **though I don't see why this holds otherwise**

Lemma 8.3. *Let T be a uniformly random multilinear map and assume q is sufficiently large in terms of r. Then the following hold:*

(a) We have $\mathbb{E}[e(H_T)] = (q^s-1)^r q^{-1} \approx q^{rs-1}$.

(b) Let $\mathcal{F}$ denote the set of tuples $(v_1^0, v_1^1, \dots, v_r^0, v_r^1)$ with $v_i^j \in V_i$ and $v_i^0 \neq v_i^1$ such that $T(v_1^{j_1}, \dots, v_r^{j_r}) = 1$ (i.e. such that this forms a $K^r_{2,\dots,2}$ in H_T). Then $\mathbb{E}[|\mathcal{F}|] \sim q^{2rs-2^r}$.

(c) Let $\mathcal{B}$ denote the set of edges $\{v_1, \dots, v_r\}$ such that $(v_1, v_1', \dots, v_r, v_r') \in \mathcal{F}$ for some $\{v_1', \dots, v_r'\}$. Then $\mathbb{E}[|\mathcal{B}|] \le (1+o(1))q^{-r}\mathbb{E}[|\mathcal{F}|]$.

We note that $G^r_{n,p}$ with p, n chosen appropriately already roughly satisfy (a) and (b), so the crucial thing we gain here is (c), which says that there are not many edges that are contained in some $K^r_{2,\dots,2}$, i.e. the copies of $K^r_{2,\dots,2}$ are all clumped together. This is the key fact that we acquire from using a random algebraic construction.

Let us briefly observe that this lemma gives the result. Indeed, we can form a $K^r_{2,\dots,2}$-free hypergraph H_T' by deleting every edge of $\mathcal{B}$. The expected number of edges for this will be asymptotically at least $q^{rs-1} - q^{2rs-2^r-r}$. Because $s = \lceil \frac{2^r-1}{r} \rceil$, we have $s < \frac{2^r-1}{r} + 1$, which is equivalent to saying $rs - 1 > 2rs - 2^r - r$, and hence the number of edges is roughly q^{rs-1}. Since H_T' has $rq^s := n$ vertices, this gives $\mathrm{ex}(n, K^r_{2,\dots,2}) = \Omega(n^{r-1/s})$ as desired. Thus it remains to prove the lemma.

Proof of Lemma 8.3. For (a), note that $T(v_1, \dots, v_r) = 0$ if $v_i = 0$ for some i. For any other tuple, let $U_i \subseteq V_i$ be the one-dimensional subspace containing v_i and 0. It is not too hard to argue that T restricted to $U_1 \times \cdots \times U_r$ is still a uniform multilinear map. Further, every multilinear map on this space is uniquely determined by the value of $T(1, \dots, 1)$, and it is not hard to see that exactly one of these q maps has $T(v_1, \dots, v_r) = 1$. Thus such a tuple is an edge with probability q^{-1} and the result follows from linearity of expectation.

For (b), observe that the only tuples that can be in $\mathcal{F}$ are those such that $v_i^0 \neq \lambda v_i^1$ for any i, as otherwise

$$\lambda = \lambda T(v_1^0, \dots, v_i^0, \dots, v_r^0) = T(v_1^0, \dots, v_i^1, \dots, v_r^0) = 1,$$

which means $\lambda = 1$, contradicting $v_i^0 \neq v_i^1$. The number of such tuples with this property is asymptotic to q^{2rs}. For such a tuple, let U_i be the span of v_i^0, v_i^1, which is a 2-dimensional subspace. Again T restricted to $U_1 \times \cdots \times U_r$ is uniform, and it is not too hard to see that there are q^{2^r} choices for T with exactly one of these placing the tuple in $\mathcal{F}$. The result follows from linearity of expectation.

It remains to deal with (c). Given affine lines $\ell_1, \dots, \ell_r$ in $V_1, \dots, V_r$, let $P(\ell_1, \dots, \ell_r)$ denote the set of tuples $(v_1, v_1', \dots, v_r, v_r')$ such that $v_i, v_i' \in \ell_i$. It is not difficult to show the following:

- The sets $P(\ell_1, \dots, \ell_r)$ are disjoint for distinct choices of lines.
- We have $|P(\ell_1, \dots, \ell_r)| = q^r(q-1)^r$.
- Every element of $\mathcal{F}$ is contained in some $P(\ell_1, \dots, \ell_r)$.

- If $P(\ell_1, \ldots, \ell_r) \cap \mathcal{F} \neq \emptyset$ then $P(\ell_1, \ldots, \ell_r) \subseteq \mathcal{F}$, i.e. $T(u_1, \ldots, u_r) = 1$ for any $u_i \in \ell_i$.

If $\mathcal{L}$ denotes the set of tuples $(\ell_1, \ldots, \ell_r)$ with $P(\ell_1, \ldots, \ell_r) \cap \mathcal{F} \neq \emptyset$, then the above implies that

$$|\mathcal{L}|q^r(q-1)^r = |\mathcal{F}|.$$

Further, we have

$$|\mathcal{B}| = \left| \bigcup_{(\ell_1, \ldots, \ell_r) \in \mathcal{L}} \ell_1 \times \cdots \times \ell_r \right| \leq q^r |\mathcal{L}| = (q-1)^{-r} |\mathcal{F}|,$$

so taking expectations gives the result. □

We note that one can get a slightly stronger result by not just considering one multilinear map T, but a family of (random) multilinear maps $T_1, \ldots, T_\ell$ and then defining $H_{T_1, \ldots, T_\ell}$ by having a hyperedge if and only if $T_i(v_1, \ldots, v_r) = 1$ for all i. The analysis here is mostly the same, but for ease of presentation we only considered a single map.

8.2 Random Polynomial Graphs

Somewhat more complicated constructions can be made by utilizing random polynomials as opposed to random multilinear maps. This approach was first popularized by Bukh [10], and since then Bukh and Conlon have developed a lot of theory surrounding it.

To set things up, given a field $\mathbb{F}_q$, we define $\mathcal{P}_{d,b}$ to be the set of polynomials over $\mathbb{F}_q$ in t variables with degree at most d. We will say that f is a random polynomial from $\mathcal{P}_{d,b}$ if it is chosen uniformly at random from $\mathcal{P}_{d,b}$, which can be done, for example, by uniformly at random choosing the coefficient of each possible monomial. With a little bit of linear algebra one can show the following, which says that a random polynomial has the same distribution as a random function when evaluated on a few number of points.

Lemma 8.4 ([12] Lemma 2.3). *If $q > \binom{m}{2}$ and $d \geq m-1$, then if $f \in \mathcal{P}_{d,b}$ is uniformly random and $x_1, \ldots, x_m$ are m distinct points of $\mathbb{F}_q^b$, then*

$$\Pr[f(x_i) = 0 \ \ \forall i] = q^{-m}.$$

Maybe include proof.

The next lemma requires just a smidge of terminology from algebraic geometry. A *variety* is any set of the form $X = \{x \in \mathbb{F}_q^b : f_1(x) = \cdots = f_a(x) = 0\}$ where $f_1, \ldots, f_a : \mathbb{F}_q^t \to \mathbb{F}_q$ are polynomials. The variety X is said to have complexity at most M if a, b and the degrees of the f_i are bounded by M. One can prove the following using standard results from algebraic geometry.

Lemma 8.5 ([12] Lemma 2.7). *Let X, D be varieties over $\mathbb{F}_q$ of complexity at most M. If q is sufficiently large in terms of M, then either $|X \setminus D| \geq q/2$ or $|X \setminus D| \leq c$ for some c depending only on M.*

The actual lemma statement involves the algebraic closure, which I think is an artifact of the proof and isn't necessary in the statement. Please let me know if you think I'm wrong (or right) about this point.

One can think of this lemma as being analogous to the fact that if f is a degree d polynomial in one variable which is 0 on at least $d+1$ points, then it must in fact be 0 on an entire line. With these two results we can prove the following.

Theorem 8.6. *For all $s \geq 2$, there exists some $t_0 = t_0(s)$ such that for all $t \geq t_0$, we have*

$$\mathrm{ex}(n, K_{s,t}) = \Theta(n^{2-1/s}).$$

Proof. The upper bound follows from the Kővári-Sós-Túran theorem. For the lower bound, let q be a sufficiently large prime power, and with some foresight we define

$$r = s^2 + 1, \quad d = rs + 1, \quad N = q^s.$$

Let $f \in \mathcal{P}_{d,2s}$ be a polynomial chosen uniformly at random. Let G be the (random) graph with vertex set $\mathbb{F}_q^s \times \mathbb{F}_q^s$ where vertices $x^1 \in \mathbb{F}_q^s, x^2 \in \mathbb{F}_q^s$ form an edge of G if and only if $f(x^1, x^2) = 0$.

Fix vertices $x^1, \ldots, x^s \in \mathbb{F}_q^s \cup \mathbb{F}_q^s$. Let C be the set of vertices y such that $x^i \sim y$ for all i (noting that $C = \emptyset$ if the x^i don't all belong to the same copy of $\mathbb{F}_q^s$, and otherwise this means e.g. $f(x^i, y) = 0$ for all i). Observe that the number of $K_{s,r}$'s of G which has the x^i as its set of size s is equal to $\binom{|C|}{r}$, and motivated by this we will attempt to bound the rth moment $\mathbb{E}[|C|^r] = \mathbb{E}[|C^r|]$ (which will be slightly easier to work with compared to the rth falling moment). To this end, we observe that if a given tuple $(y^1, \ldots, y^r)$ with k distinct elements lies in C^r, then the corresponding copy of $K_{s,k}$ lies in G. By Lemma 8.4, the probability that any given copy of $K_{s,k}$ appears in G is exactly q^{-sk} (provided q is sufficiently large in terms of s, r). Moreover, the number of tuples with k distinct elements is $O_r(N^k)$. In total we conclude that **The exposition here can probably be cleaned up**

$$\mathbb{E}[|C^r|] \leq \sum_{k=1}^{r} q^{-sk} \cdot O_r(N^k) = O_r(1).$$

Note that C is an algebraic variety by definition. By Lemma 8.5, there exists some constant c such that either $|C| \leq c$ or $|C| \geq q/2$. Thus

$$\Pr[|C| > c] = \Pr[|C| \geq q/2] = \Pr[|C|^r \geq (q/2)^r] \leq \frac{\mathbb{E}[|C|^r]}{(q/2)^r} = O_r(q^{-r}),$$

with the last step using the previous inequality.

Call a sequence $(x^1, \ldots, x^s)$*bad* if there are more than c vertices y such that $x^i \sim y$ for all i, and let B_i denote the number of i-bad sequences. Our analysis above gives

$$\mathbb{E}[B_i] \leq 2N^s \cdot O_r(q^{-r}) = O_r(q^{s^2 - r}) = o_r(1). \tag{10}$$

Now let $G' \subseteq G$ be defined by deleting a vertex from each bad sequence. Because each vertex is in at most $N = q^s$ edges in G, by (**??**) and (10) we find

$$\mathbb{E}[e(G')] \geq \mathbb{E}[e(G)] - \mathbb{E}[B] \cdot q^s = \Omega(q^{2s-1}),$$

where this last step used the previous inequality and Lemma 8.4 to deduce $\mathbb{E}[e(G)] = q^{2s} \cdot q^{-1}$. By definition G' contains no copy of $K_{s,c}$, so for $t \geq t_0 := c$, we have shown that there exists a graph G' on at most q^{2s} vertices such that it contains at least $\Omega(q^{2s-1})$ edges and no ncopy of $K_{s,t}$. This gives the desired lower bound when n is a sufficiently large prime prime power. Using Bertrand's postulate gives the desired bound for all n. $\square$

We will admit that on its own Theorem 8.6 is not particularly groundbreaking. Indeed, there exist explicit constructions showing $\mathrm{ex}(n, K_{s,t}) = \Theta(n^{2-1/s})$ when $s > (t-1)!$, and this dependency on t is far better than the implicit constant t_0 one gets from the proof. However, the proof of Theorem 8.6 gives us a new idea for constructing F-free graphs.

And indeed, many developments have been made on this method since Bukh's original construction of this form. In particular, by using more sophisticated tools from algebraic geometry, Bukh [11] showed that Theorem 8.6 holds with $t_0 = C^s$ for some absolute constant C, which stands as the best known bounds for this problem. On the other hand, by carefully modifying the current proof of Theorem 8.6, Conlon [13] was able to construct large graphs avoiding theta graphs $\theta_{a,b}$, which we recall denotes the graph consisting of a internally disjoint paths of length b between two fixed vertices.

Theorem 8.7 ([13])**.** *For all $b \geq 2$, there exists some $a_0 = a_0(b)$ such that for all $a \geq a_0$, we have*

$$\mathrm{ex}(n, \theta_{a,b}) = \Theta(n^{1+1/b}).$$

Proof Sketch. The upper bound is a result of Faudree and Simonovits [21]. For the lower bound, the key idea is to consider *multiple* random polynomials $f_1, \ldots, f_a \in \mathcal{P}_{d,2b}$ chosen independently, with us defining our graph G on $\mathbb{F}_q^b \cup \mathbb{F}_q^b$ by having $x \sim y$ if and only if $f_i(x, y) = 0$ for all i.

Fix two vertices $x^1, x^{b+1} \in \mathbb{F}_q^b \cup \mathbb{F}_q^b$ and define C to be the tuples of distinct vertices $(x^2, \ldots, x^{b-1})$ such that $x^1 \cdots x^{b+1}$ is a path in G. As before (but with a somewhat more difficult analysis), one can show $\mathbb{E}[|C|^r] = O(1)$. However, in this case C is not quite an algebraic variety because of us requiring C to use distinct vertices. However, one can write C as $X \setminus D$ for two varieties X, D, so Lemma 8.5 still applies and the rest of the proof goes through. $\square$

One can further interpolate between this theorem and Theorem 8.6 to give effective lower bounds on $\mathrm{ex}(n, F)$ whenever F is a "large power of a rooted tree" (e.g. $K_{s,t}$ is just many copies of a star $K_{s,1}$, and $\theta_{a,b}$ is just many copies of a path P_b). This was done by Bukh and Conlon [12] in order to show that for every rational number $r \in [1, 2]$, there exists a finite set of graphs $\mathcal{F}$ such that $\mathrm{ex}(n, \mathcal{F}) = \Theta(n^r)$. The rational exponents conjecture, which says that one can achieve this with $\mathcal{F}$ consisting of a single graph, remains a major open problem.

8.3 Multicolor Ramsey Numbers

Let $r(t; \ell)$ denote the smallest number N such that every ℓ-coloring of $E(K_N)$ contains a monochromatic clique of size t. **Inset history and connection with the earlier Ramsey results proven in the text. Also sketch the proof of the bound you get with the naive method for comparison.**

The following observation will be the key towards going further. The initial idea for this lemma can be seen in Conlon and Ferber [14], though it was first really used by Wigderson [63] and then generalized by Sawin [54].

Lemma 8.8. *Let G be graph with no clique of size t, and let p be the probability that vertices $v_1, \ldots, v_t \in V(G)$ chosen independently and uniformly at random form an independent set. Then for all $\ell \geq 2$, we have*

$$r(t;\ell) \geq p^{-(\ell-2)/t} 2^{(t-1)/2}.$$

Note that when $\ell = 2$ this recovers the usual lower bound for Ramsey numbers from the random coloring.

Proof. Let N be an integer to be determined later, and let $f_1, \ldots, f_{\ell-2} : V(K_N) \to V(G)$ be chosen independently and uniformly at random. Define a coloring $\chi : E(K_N) \to [\ell]$ in the following way: for distinct $x, y \in V(K_n)$, if there exists i such that $f_i(x)f_i(y) \in E(G)$, then set $\chi(xy)$ to be the minimum i with this property. Otherwise, set $\chi(xy)$ to be $\ell - 1$ or ℓ with probability 1/2 each. That is (as Wigderson notes in his paper), this coloring comes from covering K_N with $\ell - 2$ randomly permuted blowups of G and then randomly using two colors to deal with any uncovered vertices.

We first observe that there is no monochromatic K_t in any color $i \leq \ell - 2$. Indeed, if $\{x_1, \ldots, x_t\}$ were such a clique then this would imply $\{f_i(x_1), \ldots, f_i(x_t)\}$ forms a clique in G (since $\chi(x_j x_k) = i$ implies $f_i(x_j)f_i(x_k) \in E(G)$). Thus it remains to show that, with positive probability, there is no monochromatic K_t in color $i \in \{\ell - 1, \ell\}$. Observe that a clique K_t in K_N has all of its edges colored by $\ell - 1$ or ℓ if and only if each f_i maps K_t to an independent set of G, and the probability that this happens is exactly $p^{\ell-2}$ by hypothesis, and from there this K_t will be monochromatic with probability $2^{1-\binom{t}{2}}$. In total then, the expected number of monochromatic cliques will equal $\binom{N}{t} p^{\ell-2} 2^{1-\binom{t}{2}}$, and this will be less than 1 provided $N \leq p^{-(\ell-2)/t} 2^{(t-1)/2}$. Thus there exists a coloring of this size with no monochromatic clique, giving the desired result. $\square$

Observe that the p in Lemma 8.8 roughly corresponds to the number of independent sets of size at most t in G, so we need to find a graph with small clique number and not too many small independent sets. To this end, let $V \subseteq \mathbb{F}_2^t$ be the set of vectors v with $v \cdot v = 0$ (i.e. vectors with even Hamming weight), and let G be the graph on V where two vectors u, v are adjacent if and only if $u \cdot v = 1$.

Lemma 8.9. *If t is even, then the graph G contains no clique of size t.*

Proof. Assume for contradictiont that there exist distinct vectors $v_1, \ldots, v_t \in V$ with $v_i \cdot v_j = 1$ for all $i \neq j$ (and $= 0$ for $i = j$ by definition of V). We claim that these vectors are linearly independent. Indeed, if there exists $\alpha_i \in \{0, 1\}$ with $\sum \alpha_i v_i = 0$, then by taking the dot product of v_j on both sides we find $\sum_{i \neq j} \alpha_i \equiv 0$ for all i, and it is not difficult to show that this implies $\alpha_i = 0$ for all i (here we need that t is even, else $\alpha_i = 1$ for all i would work). However, V is a $t - 1$ dimensional subspace, so it contains no set of t linearly independent vectors, proving the result. $\square$

Lemma 8.10. *The probability p that a uniformly random tuple $(v_1, \ldots, v_t) \in V^t$ is such that $\{v_1, \ldots, v_t\}$ is independent in G is at most $2^{-3t^2/8 + o(t^2)}$.*

Proof. Let X be the set of tuples $(v_1, \ldots, v_t) \in V^t$ such that $v_i \cdots v_j = 0$ for all i, j, so our goal is to upper bound $|X|/|V|^t = |X|2^{-t^2}$. Define the rank of a tuple in X to be the rank of the smallest subspaces containing every vertex of the tuple. We claim that the number of tuples in X of rank r is at most

$$t! \left(\prod_{i=0}^{r-1} 2^{t-i} \right) \cdot 2^{(t-r)r} = 2^{tr - \binom{r}{2} + tr - r^2}. \tag{11}$$

Indeed, possibly by reordering the tuple (giving us the factor of $t!$) we can assume the first r vectors are linearly independent, and given $v_1, \ldots, v_i$ with $0 \le i < r$, the number of choices for a v_{i+1} which is linearly independent of $v_1, \ldots, v_i$ is exactly q^{t-i}. After this every vector must lie in the span of $v_1, \ldots, v_r$, giving exactly q^r choices for the remaining $t - r$ vectors.

We next claim that there exists no tuple in X of dimension larger than $t/2$. Indeed, if S is the span of the vectors in a tuple of X, then note that $S \subseteq S^\perp$ since $v_i \cdot v_j = 0$ for all i, j. From linear algebra we have $t = \dim S + \dim S^\perp \ge 2 \dim S$, proving the claim.

It is not hard to prove that (11) is increasing for $r \le t/2$, so plugging in $r = t/2$ gives an upper bound for $|X|/(t/2)$ of the form $2^{5t^2/8 + o(t^2)}$, giving the desired bound on $|X|/|V|^t$. □

Putting all these lemmas together gives the following.

Corollary 8.11. *For $\ell \ge 3$ we have*

$$r(t; \ell) \ge \left(2^{\frac{3\ell}{8} - \frac{1}{4}} \right)^{t - o(t)}.$$

This bound stood as the best for about a year until Sawin [54] realized one could do somewhat better by replacing the algebraic graph G described above with a purely random graph, namely $G_{n,p}$ with $p \approx .455$. Thus, although the initial breakthrough for multicolor Ramsey numbers came from a random algebraic approach, the method was later subsumed by a simpler random model. This sort of thing happens somewhat often with proofs using the random algebraic method. Because of this, some mathematicians are of the opinion that any time the random algebraic method is used, there exists a simpler random model which gives better results. I don't personally believe that this is true, and even if it were, the fact that random algebraic methods consistently give initial breakthroughs to longstanding open problems makes them worth considering in my eyes.

Maybe comment on other ways random homomorphisms are useful.

Part III

Spread Hypergraphs and Thresholds

9 The Spreadness Theorem

Throughout this section we consider hypergraphs $\mathcal{H}$ which may have repeated edges, and we will typically denote the edges of $\mathcal{H}$ by S. We recall that $d(A)$ denotes the degree of a set of vertices A in $\mathcal{H}$, i.e. the number of edges of $\mathcal{H}$ containing A.

We say that a hypergraph $\mathcal{H}$ is r-bounded if all of its edges have size at most r. We say that a hypergraph $\mathcal{H}$ is q-spread[1] for some $0 < q < 1$ if $\mathcal{H}$ is non-empty and if $d(A) \le q^{|A|}|\mathcal{H}|$ for all sets of vertices A. The main result for q-spread hypergraphs is the following.

Theorem 9.1 ([3, 25])**.** *Let $\mathcal{H}$ be an r-bounded q-spread hypergraph on V. There exists an absolute constant K_0 such that if W is a set of size $Cq\log r \cdot |V|$ chosen uniformly at random from V with $C \ge K_0$, then*

$$\Pr[W \textit{ contains an edge of } \mathcal{H}] \ge 1 - 8C^{-1}.$$

We note that better quantitative versions of Theorem 9.1 exist, see e.g. **Tao's reformulation below?**, but as stated this theorem already does a lot. Let's start by looking at some applications before turning it's short (though very dense!) proof.

9.1 Applications

Our first application is the following.

Theorem 9.2. *Let $G^r_{n,m}$ be the r-graph chosen uniformly at random amongst all r-graphs with n vertices and m edges. Then there exists a constant C such that if $m \ge Cn\log n$ and n is a multiple of r, then $G^r_{n,m}$ contains a perfect matching a.a.s.* [2]

It is not too difficult to show that this bound on m is essentially best possible. We note that morally speaking, $G^2_{n,m}$ acts the same as $G_{n,p}$ where $p = m/\binom{n}{2}$. In particular, one can use Theorem 9.2 to prove that $G_{n,p}$ contains a perfect matching a.a.s. if $p = \Omega(\log n/n)$. Proving Theorem 9.2 for $r = 2$ is not hard, but the result for general r was thought to be very difficult, with its first proof due to Johansson, Kahn, and Vu [36] using a rather involved argument. We will prove Theorem 9.2 in just a few lines with Theorem 9.1.

Proof. Let $\mathcal{H}$ be the hypergraph on $E(K^r_n)$ where each hyperedge S is a perfect matching of

[1] Some texts would say that such an $\mathcal{H}$ is q^{-1}-spread.

[2] This means "asymptotically almost surely", i.e. the probability of this event happening tends to 1 as n tends to infinity.

K_n^r. Observe that for any set $A \subseteq E(K_n^r)$, we have

$$\begin{aligned} d(A) \cdot |\mathcal{H}|^{-1} &= \frac{(n - r|A|)!}{(r!)^{n/r-|A|}(n/r - |A|)!} \cdot \frac{(r!)^{n/r}(n/r)!}{n!} \\ &= (r!)^{|A|} \binom{n/r}{|A|} \binom{n}{r|A|}^{-1} \frac{|A|!}{(r|A|)!} \\ &\le (r!)^{|A|}(en/r|A|)^{|A|} \cdot (n/r|A|)^{-r|A|} \cdot (|A|)^{|A|} \cdot (r|A|/e)^{-r|A|} \\ &= (r!)^{|A|} e^{(r+1)|A|} n^{-(r-1)|A|} \le (n/re^3)^{-(r-1)|A|}. \end{aligned}$$

Thus $\mathcal{H}$ is $(n/re^3)^{-r+1}$-spread. It is also (n/r)-uniform and has a ground set $V = E(K_n^r)$ of size $\binom{n}{r}$. By Theorem 9.1, we see that if m is at least as large as in our hypothesis, then with high probability a random m-subset of $\mathcal{H}$ will contain a hyperedge, i.e. $H_{n,m}^r$ will contain a perfect matching with high probability. □

Another basic example is the following.

Proposition 9.3. *Let F be an r-graph and define $t(F) = \max\{|E(F')|/|V(F')| : F' \subseteq F\}$. Let $G_{n,m}^r$ be as in Theorem 9.2. There exists a constant $C(F)$ such that if $m \ge C(F)n^{r-1/t(F)}$, then $G_{n,m}^r$ contains a copy of F a.a.s.*

A simple first moment argument shows that this bound is tight. One can prove Proposition 9.3 using a standard but somewhat tedious second moment argument, but using Theorem 9.1 gives a shorter proof.

Proof. Let $\mathcal{H}$ be the hypergraph on $E(K_n^r)$ whose hyperedges correspond to copies of F. Observe that $\mathcal{H}$ being q-spread is equivalent to having $(d(A)/|\mathcal{H}|)^{1/|A|} \le q$ for all $A \subseteq V = E(K_n^r)$. Any set $A \subseteq E(K_n^r)$ of positive degree in $\mathcal{H}$ forms a subgraph $F' \subseteq F$ with $|E(F')| = |A|$, and in this case

$$\left(\frac{d(A)}{|\mathcal{H}|}\right)^{1/|A|} \le \left(\frac{n^{|V(F)|-|V(F')|}}{\binom{n}{|V(F)|}}\right)^{1/|A|} \le |V(F)|^{|V(F)|} \cdot n^{-|V(F')|/|E(F')|}.$$

Thus we see that $\mathcal{H}$ is q-spread with

$$q = \max\{|V(F)|^{|V(F)|} \cdot n^{-|V(F')|/|E(F')|} : F' \subseteq F\} = |V(F)|^{|V(F)|} \cdot n^{-1/t(F)}.$$

Plugging this into Theorem 9.1 gives the result. □

The study of q-spread hypergraphs was initiated by Alweiss, Lovett, Wu, and Zhang [3] where they proved a slightly weaker version of Theorem 9.1. Their motivation came from the Erdős sunflower conjecture. A *k-sunflower* is a hypergraph with edges $S_1, \dots, S_k$ such that there exists a set K called the *kernel* which has $S_i \cap S_j = K$ for all $i \neq j$.

Theorem 9.4. *There exists an absolute constant $C > 0$ such that if $\mathcal{H}$ is an r-graph with at least $(Ck \log r)^r$ edges, then $\mathcal{H}$ contains a k-sunflower.*

We note that [3] was the first to prove bounds of the form $(\log r)^{r+o(1)}$ for fixed k, with [49, 8] later giving better bounds in terms of k. Prior to [3], the best known bounds were of the form $r^{r-o(1)}$. It is a famous conjecture of Erdős that one can prove a bound of the form $c_k^{r+o(1)}$.

Proof. We prove the result by induction on r, the $r = 1$ case being trivial. Let $\mathcal{H}$ be an r-graph with at least $(Ck \log r)^r$ edges. If $\mathcal{H}$ is not q-spread with $q = (Ck \log r)^{-1}$, then there exists some $A \subseteq V(H)$ such that $d(A) \geq (Ck \log r)^{r-|A|}$. This means that the link hypergraph $\mathcal{H}_A = \{S \setminus A : S \in \mathcal{H}, A \subseteq S\}$ has size at least $(Ck \log r)^{r-|A|}$. Since $\mathcal{H}_A$ is an $(r - |A|)$-uniform hypergraph, by induction $\mathcal{H}_A$ contains a k-sunflower, say with edges $S_1 \setminus A, \ldots, S_k \setminus A \in \mathcal{H}_A$. It is not difficult to check that $S_1, \ldots, S_k \in \mathcal{H}$ forms a k-sunflower in $\mathcal{H}$. We conclude that any $\mathcal{H}$ with at least $(Ck \log r)^r$ edges which is not q-spread contains a k-sunflower, so from now on we may assume $\mathcal{H}$ is q-spread.

Possibly by adding isolated vertices to $\mathcal{H}$, we can assume that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2k$. Let $V_1, \ldots, V_{2k}$ be a random partition of V such that each $V_i \subseteq V$ has size $(2k)^{-1}|V|$. This means that each V_i is a uniformly chosen set of V of size $(2k)^{-1}|V| = \frac{1}{2}C(\log r)q|V|$. Let 1_i be the indicator variable for the event that V_i contains an edge of $\mathcal{H}$. By Theorem 9.1, we have $\Pr[1_i = 1] \geq \frac{1}{2}$ provided C is sufficiently large. In this case, $\mathbb{E}[\sum 1_i] \geq k$, and hence there exists some partition $V_1, \ldots, V_{2k}$ such that $\sum 1_i \geq k$, which in particular means there exist k disjoint edges of $\mathcal{H}$. This is a k-sunflower in $\mathcal{H}$, proving the result. $\square$

9.2 Proof of the Spreadness Theorem

There are by now a number of proofs of Theorem 9.1, though most of them maintain the same core set of ideas. The proof we present here is based off of a proof due to Rao [50] which gives weaker quantitative bounds. We emphasize that while the proof itself is very short, it is also very dense in content, so we'll spend some time trying to build up some intuition for it.

Recall that $\mathcal{H}$ is an r-bounded q-spread hypergraph on V, and that we want to show that a uniformly random set W of size $Cq \log r \cdot |V|$ contains an edge of $\mathcal{H}$ with high probability. In order to use an iterative approach, we consider a uniform random vector of disjoint sets $(W_1, \ldots, W_{\log r})$ each of size $Cq|V|$. Let $W_{\leq i} = \bigcup_{j \leq i} W_j$, and note that W has the same distribution as $W_{\log r}$, so it suffices to work with these random sets.

A super ideal situation for our iterative approach would be if for all $S \in \mathcal{H}$, we have $|S - W_{\leq i}| < 2^{-i}r$. Indeed, with this at $i = \log r$, we would get that every edge is contained in $W_{\leq \log r}$. Of course, this is far too much to hope for. However, since we only need $W_{\leq \log r}$ to contain a single edge, it would suffice to have this work out for *some* S. As such it perhaps make sense to say that an edge S "succeeds" at step i if $|S - W_{\leq i}| < 2^{-i}r$, and then to argue that with high probability some edge succeeds at each step. Unfortunately this notion of success is too restrictive to work. The key insight is that we can loosen our condition by saying that an edge S "succeeds" if there exists some edge $S' \subseteq S \cup W_{\leq i}$ (or equivalently $S' - W_{\leq i} \subseteq S - W_{\leq i}$) such that $|S' - W_{\leq i}| < 2^{-i}r$. The point is that (1) this condition is easier to achieve (and in particular will be achieved by "most" S at each step), and (2) if some S succeeds at each step, then the S' it points to for step $\log r$ will be contained in $W_{\leq \log r}$. Need to fact check that this is really what's going on, and in particular we maybe need S' also to have not failed at this point.

With this in mind, given $W_1, \ldots, W_i$, we iteratively define "failure" hypergraph $\mathcal{F}_i$ to be those $S \notin \mathcal{F}_{\leq i-1} := \bigcup_{j \leq i-1} \mathcal{F}_j$ (i.e. which haven't failed at any previous step) such that for all $S' \in \mathcal{H} - \mathcal{F}_{\leq i-1}$ satisfying $S' \subseteq S \cup W_{\leq i}$, we have $|S' - W_{\leq i}| \geq 2^{-i}r$. The key claim is the following.

Lemma 9.5. *Given $W_{\le i-1}$, we have $\mathbb{E}[|\mathcal{F}_i|] \le 2(C/4)^{-2^{-i}r}|\mathcal{H}|$.*

To upper bound $|\mathcal{F}_i|$, it will help to instead upper bound the size of an auxiliary hypergraph defined as follows. Given $W_1, \ldots, W_i$, we define the *fragment* $T(S, W_{\le i})$ of an edge $S \in \mathcal{H} - \mathcal{F}_{i-1}$ to be a set of minimum size in $\{S' - W_{\le i} : S' \in \mathcal{H} - \mathcal{F}_{\le i-1}, S' \subseteq S \cup W_{\le i}\}$ (say the lexicographically smallest set if there are multiple of minimum size). We let $\mathcal{G}_i$ be the hypergraph where T is an edge if $T = T(S, W_{\le i})$ for some $S \in \mathcal{F}_i$. Note that by definition this means $|T| \ge 2^{-i}r$, and that for every $S \in \mathcal{F}_i$, there exists some $T \in \mathcal{G}_i$ with $T \subseteq S$. This last condition says $\mathcal{G}_i$ is an *undercover* of $\mathcal{F}_i$, which will also be a key condition in our upcoming proof of the Park-Pham Theorem.

Proof. Let $w := Cq|V|$, which we recall is the size of W_i, and let $n_i = |V - W_{\le i-1}|$. Let $\mathcal{P}$ consist of all pairs (S, W) with $S \in \mathcal{H}$ and $W \in \binom{V - W_{\le i-1}}{w}$ such that $S \in \mathcal{F}_i$ whenever $W_i = W$. Similarly given an integer $a \ge 2^{-i}r$, let $\mathcal{P}_a$ consist of all pairs (T, W) with $|T| = a$ and $W \in \binom{V - W_{\le i-1}}{w}$ such that $T \in \mathcal{G}_i$ whenever $W_i = W$. We claim that. **Probably use t instead of a**

$$\mathbb{E}[|\mathcal{F}_i|] = |\mathcal{P}|\binom{n_i}{w}^{-1} \le \sum_{a \ge 2^{-i}r} q^a |\mathcal{H}||\mathcal{P}_a|\binom{n_i}{w}^{-1}. \tag{12}$$

Indeed, the equality is straightforward. Because $\mathcal{G}_i$ is an undercover of $\mathcal{F}_i$, for every pair $(S, W) \in \mathcal{P}$ there exists a pair $(T, W) \in \bigcup_a \mathcal{P}_a$ such that $T \subseteq S$. Moreover, for each set T of size a, the number of $S \in \mathcal{H}$ with $T \subseteq A$ is at most $q^a|\mathcal{H}|$ by the definition of $\mathcal{H}$ being q-spread. This gives the stated inequality

It remains to count the number of elements $(T, W) \in \mathcal{P}_a$. We will identify such a pair by first specifying the set $T \cup W$, and then specifying T (which uniquely determines W). We first note that $T \cup W$ is a set of size $a + w$, so the number of choices for this step is at most

$$\binom{n_i}{a+w} \le (n_i/w)^a \cdot \binom{n_i}{w} = (Cq)^{-a}\binom{n_i}{w}.$$

Given $T \cup W$, choose any $S' \in \mathcal{H} - \mathcal{F}_{\le i-1}$ with $S' - W_{\le i-1} \subseteq T \cup W$. Crucially, we must have $T \subseteq S' - W_{\le i-1}$, as otherwise if $T = T(S, W_{\le i-1} \cup W)$ for some S, then taking $T' = S' - (W_{\le i-1} \cup W) \subsetneq T$ (with the inclusion holding because $S' - W_{\le i-1} \subseteq T \cup W$, and the strictness holding if $T \not\subseteq S' - W_{\le i-1}$), we find that T cannot be the fragment of S (since T' is a smaller set than T satisfying the same properties). Note that $S' \notin \mathcal{F}_{\le i-1}$ implies $|S' - W_{\le i-1}| \le 2^{-i+1}r$, so the number of choices for T is at most $2^{2^{-i+1}r} = 4^{2^{-i}r}$.

In total we conclude that $|\mathcal{P}_a| \le (Cq)^{-a}4^{2^{-i}r}\binom{n_i}{w}$. Plugging this into (12), we find

$$\mathbb{E}[|\mathcal{F}_i|] \le \sum_{a \ge 2^{-i}r} q^a|\mathcal{H}| \cdot (Cq)^{-a}4^{2^{-i}r} = 4^{2^{-i}r}|\mathcal{H}| \sum_{a \ge 2^{-i}r} C^{-a} \le 2(C/4)^{-2^{-i}r}|\mathcal{H}|,$$

with this last step holding for C sufficiently small. $\square$

With this lemma, we have

$$\mathbb{E}[|\mathcal{F}_{\le \log r}|] \le \sum_{i=1}^{r} 2(C/4)^{-2^{-i}r}|\mathcal{H}| \le 16C|\mathcal{H}|.$$

By Markov, the probability that $|\mathcal{F}_{\le \log r}| = |\mathcal{H}|$ is at most $\frac{1}{16C}$, so $\mathcal{F}_{\le \log r} \neq \mathcal{H}$ is at least $1 - \frac{1}{16C}$. As noted above, if there exists $S \in \mathcal{H} - \mathcal{F}_{\le \log r}$ then S points to an edge which is contained in $W_{\le \log r}$, so we conclude that this random set $W_{\le \log r}$ of size $Cq \log(r)|V|$ contains an edge with probability at least $1 - \frac{1}{16C}$.

9.3 Losing Logarithms

As we noted earlier, the bound of Theorem 9.2 is best possible. In particular, the $\log r$ term of Theorem 9.1 is necessary in general. However, under certain conditions one can remove this logarithmic term. This was first observed by Kahn, Narayanan, and Park [38] where they found tight bounds on the threshold of a square of a Hamiltonian cycle in $G_{n,p}$. Here we briefly outline how, under special circumstances, one can modify the previous proof to get rid of the $\log r$ factor.

The main idea is that instead of setting our cutoff points for our fragments to be $r/2, r/4, \ldots$, we instead set them to be $k_1, k_2, \ldots$ for some suitable sequence k_i with significantly fewer than $\log r$ terms. In this setup, one could try to naively go through Lemma 9.5 and replace $2^{-i+1}r$ with k_{i-1} and $2^{-i}r$ with k_i, which will roughly give us

$$\mathbb{E}[|\mathcal{F}_i|] \le 2^{k_{i-1}} C^{-k_i} |\mathcal{H}|,$$

but this will be terrible unless k_{i-1} differs from k_i by a multiplicative constant depending on C.

One way we can get around this is if we impose that for all sets A and integers j with $k_{i-1} \ge |A| \ge j \ge k_i$, we have that the number of edges S' with $|A \cap S'| = j$ is at most $q^j |\mathcal{H}|$. Note that this is stronger than spreadness since, when taking $j = |A|$, the condition $|A \cap S'| = j$ just says S' is an edge containing A, so this bound exactly says $\deg_{\mathcal{H}}(A) \le q^{|A|}|\mathcal{H}|$, which is the spreadness condition. Assuming this condition holds, we will count the pairs $(S, W) \in \mathcal{P}$ in a more subtle way.

Let $T = T(S, W_{\le i-1} \cup W)$, which we note has $|T| \ge k_i$ and $T \subseteq S$. We first specify $T \cup W$, which as before can be done in roughly $(Cq)^{-a}\binom{n_i}{w}$ ways. We then pick some edge $S' \notin \mathcal{F}_{\le i-1}$ such that $A := S' - W_{\le i-1} \subseteq T \cup W$, where as before we have $T \subseteq A$. Since $T \subseteq A \cap S$, we have $j := |A \cap S| \ge a$. Given j, the number of S with $|A \cap S| = j$ is at most $q^j |\mathcal{H}|$ by our condition (and we have $|A| \le k_{i-1}$ as otherwise we would have $S' \in \mathcal{F}_{\le i-1}$). Since we now know S and $T \cup W$, we also know $S \cup W$ (since $T \subseteq S$), and hence T (since T is purely a function of the set $S \cup W$ and $W_{\le i-1}$ Need to double check this; in any case there are trivially at most 2^j choices for $T \subseteq A \cap S$), and hence $W = (T \cup W) \setminus T$ (since T is disjoint from W by definition). With this, we see that the total number of choices is at most

$$\sum_{a \ge k_i} C^{-a} \binom{n_i}{w} \sum_{j \ge a} q^j |\mathcal{H}| \approx C^{-k_i} \binom{n_i}{w} |\mathcal{H}|,$$

giving the desired result. A more formal theorem/proof can be found in [60], though the approach used there is an older and more complicated version of the one presented here. See also [26] for a proof in the specific case of getting rid of the logarithm for the square of a Hamiltonian cycle.

10 The Park-Pham Theorem TODO

11 Spread Measures and Absorption TODO

12 Spread Approximations TODO

Part IV

Approximating Graphs and Hypergraphs

Roughly speaking, this part is dedicated to the general idea that, sometimes, complicated objects S can be "approximated" by some other object A which is both simpler to analyze, and is such that properties of A lift to properties of S. **For example, we saw this with spread approximations in the previous chapter.**

13 The Delta-system Method

Recall that a *k-sunflower* (also called a delta-system) is a hypergraph $\mathcal{S}$ with edges $e_1, \ldots, e_k$ such that there exists a set K called the *kernel* which has $e_i \cap e_j = K$ for all $i \neq j$. Roughly speaking, the Delta-system method is any proof using the following observation, which is usually credited to Deza, Erdős, and Frankl [18].

Lemma 13.1. *If $\mathcal{H}$ is an r-graph which contains an $(r+1)$-sunflower with kernel K, then for every edge $e \in \mathcal{H}$, there exists an edge $f \in \mathcal{H}$ with $e \cap f \subseteq K$.*

Proof. Let $e_1, \ldots, e_{r+1}$ be the edges of the sunflower. Since each of the sets $e_1 \setminus K, \ldots, e_{r+1} \setminus K$ are non-empty disjoint sets, one of these $e_i \setminus K$ sets must be disjoint from e. Taking $f = e_i$ gives the result. $\square$

An effective tool to use in conjunction with this observation is Füredi's intersection semilattice lemma (which itself is proven using the Delta-system method). The full statement is a little intimidating, so we'll start by just stating a consequence of it.

Lemma 13.2 (Weak intersection semilattice lemma). *For all r, s, there exists $c = c(r,s) > 0$ such that for every r-graph $\mathcal{H}$, there exists a subgraph $\mathcal{H}' \subseteq \mathcal{H}$ with $e(\mathcal{H}') \geq ce(\mathcal{H})$ such that for all $e, f \in \mathcal{H}'$, $e \cap f$ is the kernel of an s-sunflower.*

That is, we can approximate $\mathcal{H}$ by a hypergraph $\mathcal{H}'$ such that any two edges of $\mathcal{H}'$ are petals of a large sunflower. This quickly gives the following strengthening of the Erdős-Rado sunflower lemma due to Mubayi and Zhao [48].

Corollary 13.3 ([48]). *For all r, s there exists a constant $C = C(r,s)$ such that if $\mathcal{H}$ is an n-vertex r-graph with $e(\mathcal{H}) \geq Cn^{r-t-1}$, then $\mathcal{H}$ contains an s-sunflower which has core of size at most t.*

Proof. We may assume n is sufficiently large in terms of r, as otherwise one can trivially find a sufficiently large C. Let $C = 2c^{-1}$ with c the constant from the previous lemma. Then $e(\mathcal{H}') \geq 2n^{r-t-1} > \binom{n-t-1}{r-t-1}$. By the Erdős-Ko-Rado theorem for $(t+1)$-intersecting hypergraphs (see Theorem 13.6), $\mathcal{H}'$ must contain two edges e, f which intersect in less than $t+1$ vertices. By assumption $e \cap f$ is the core of a sunflower with at least s petals, proving the result. $\square$

We'll now state the full intersection semilattice lemma. For this, if $\mathcal{H}$ is an r-partite r-graph with partition $\bigcup V_i$ and if S is a set of vertices, then we define $\mathrm{proj}(S) := \{i : S \cap V_i \neq \emptyset\}$. That is, $\mathrm{proj}(S)$ records which coordinates its vertices are in. Given a hypergraph $\mathcal{H}$ and a set of vertices S, define $d^*(S)$ to be the largest integer d such that there exist edges $e_1, \ldots, e_d \in \mathcal{H}$ with $e_i \cap e_j = S$ for all $i \neq j$. In other words, $d^*(K)$ is the size of the largest sunflower which contains K as its kernel.

Lemma 13.4 (Intersection semilattice lemma). *For all r, s, there exists $c = c(r,s) > 0$ such that for every r-graph $\mathcal{H}$, there exists an r-partite subgraph $\mathcal{H}' \subseteq H$ with $e(\mathcal{H}') \geq ce(\mathcal{H})$ and a hypergraph $\mathcal{J} \subseteq 2^{[r]}$ not containing the edge of size r such that:*

(1) $\mathcal{J}$ is intersection closed, i.e. $I, J \in \mathcal{J}$ implies $I \cap J \in \mathcal{J}$.

(2) For every $e \in \mathcal{H}'$, $\{\mathrm{proj}(e \cap f) : f \in \mathcal{H}' \setminus \{e\}\} = \mathcal{J}$.

(3) $d^*_{\mathcal{H}'}(e \cap f) \geq s$ for all $e, f \in \mathcal{H}'$.

Note that if we ignore (1) and (2) we get back Lemma 13.2. Roughly speaking, this result says that we can approximate a large chunk of $\mathcal{H}$, namely $\mathcal{H}'$, by a small hypergraph $\mathcal{J}$ such that for any edge $e \in \mathcal{H}'$, the hypergraph $\mathcal{J}$ tells you exactly how other edges can intersect e, and moreover, (3) guarantees that each possible intersection occurs at least s times. We say that an $\mathcal{H}'$ as in the conclusion of this lemma is $(s, \mathcal{J})$-homogeneous.

We postpone proving this result for the moment and instead look at some consequences. For $\mathcal{J} \subseteq 2^{[r]} \setminus [r]$, define the rank

$$\mathrm{rank}(\mathcal{J}) = \min\{|T| : T \subseteq [r],\ T \not\subseteq I\ \forall I \in \mathcal{J}\},$$

i.e. this is the smallest integer t such that there exists a t-set not contained in an edge of $\mathcal{J}$. For example, $\mathrm{rank}(\mathcal{J}) > 1$ if and only if every vertex of $[r]$ is contained in an edge of $\mathcal{J}$.

Lemma 13.5. *If $\mathcal{H}'$ is an n-vertex $(s, \mathcal{J})$-homogeneous r-graph, then $|\mathcal{H}'| \leq \binom{n}{\mathrm{rank}(\mathcal{J})}$.*

Note that s does not appear in this bound. Before looking at the proof, the reader may want to try proving this result for themselves when $\mathrm{rank}(\mathcal{J}) = 1$ in order to get a sense for the definitions.

Proof. Let $T \subseteq [r]$ be a set such that $|T| = \mathrm{rank}(\mathcal{J})$ and such that $T \not\subseteq J$ for all $J \in \mathcal{J}$. Given an edge $e \in \mathcal{H}'$, let $\phi(e) = e \cap \bigcup_{i \in T} V_i$. We claim that ϕ is injective. Indeed, if $\phi(e) = \phi(f)$, then $T \subseteq \mathrm{proj}(e \cap f) \in \mathcal{J}$, a contradiction to our assumption on T. Since ϕ maps edges of $\mathcal{H}'$ injectively to sets of size $\mathrm{rank}(\mathcal{J})$, we conclude the result. □

We can use this result to give a proof of the Erdős-Ko-Rado theorem for t-intersecting hypergraphs.

Theorem 13.6. *Let $\mathcal{H}$ be an n-vertex r-graph such that $|e \cap f| \geq t$ for all $e, f \in \mathcal{H}$ distinct. If n is sufficiently large in terms of r, then $e(\mathcal{H}) \leq \binom{n-t}{r-t}$ with equality holding if and only if $\mathcal{H}$ consists of every edge containing some fixed set T of size t.*

Proof. Apply Lemma 13.4 with $s = r + 1$ and let $\mathcal{H}', \mathcal{J}$ be the resulting hypergraphs with $\bigcup V_i$ the r-partition of $\mathcal{H}'$. First note that if $\mathrm{rank}(\mathcal{J}) < r - t$, then by Lemma 13.5 we have

$$e(\mathcal{H}) \leq c^{-1} e(\mathcal{H}') \leq c^{-1}\binom{n}{r-t-1} < \binom{n-t}{r-t},$$

where this last step holds for n sufficiently large in terms of $c = c(r, r+1)$. Thus if $e(\mathcal{H}) \geq \binom{n-t}{r-t}$, we must have $\mathrm{rank}(\mathcal{J}) \geq r - t$.

Let S be with $|S| = \mathrm{rank}(\mathcal{J})$ such that no edge of $\mathcal{J}$ contains S, and let $T = [r] \setminus S$. By the above we may assume $|S| \geq r - t$, and hence $|T| \leq t$. By definition of $\mathrm{rank}(\mathcal{J}) = |S|$, for every

$i \in S$, there exists an edge $J_i \in \mathcal{J}$ such that $S \setminus \{i\} \subseteq J_i$. Note that $i \notin J_i$ by assumption of S not being contained in any edge of $\mathcal{J}$, which means $J := \bigcap_{i \in S} J_i \subseteq T$. Because $\mathcal{J}$ is intersection closed, we have $J \in \mathcal{J}$.

We claim that $|J| \geq t$. Indeed, by definition of $J \in \mathcal{J}$, there exist two edges of $\mathcal{H}' \subseteq \mathcal{H}$ whose intersection is exactly J, and by the t-intersecting property we must have $|J| \geq t$. Because $|T| \leq t$ and $J \subseteq T$, we conclude that $J = T \in \mathcal{J}$.

Now let $e \in \mathcal{H}'$ and $K = e \cap \bigcup_{i \in T} V_i$, noting that $|K| = t$. Then Lemma 13.4 guarantees that there is a sunflower with at least $r+1$ petals and K as its kernel. This implies that for every edge $f \in \mathcal{H}$, there exists an edge $e' \in \mathcal{H}'$ which contains K and which is disjoint from $f \setminus K$. Thus to have $|e' \cap f| \geq t$, we must have $K \subseteq f$. In other words, every edge of $\mathcal{H}$ must contain the t-set K. This implies the result. $\square$

The above argument actually gives the following stability result: for all r, t there exists a constant $c' = c'(r, t)$ such that if $\mathcal{H}$ is t-intersecting with $e(\mathcal{H}) > c'\binom{n-t}{r-t}$, then there exists a set of size t which is contained in every edge of $\mathcal{H}$.

Remark 13.7. Stronger versions of Theorem 13.6 are known. For example, **REF**determined the maximum size of a t-intersecting family for *any* value n. In another direction, Frankl and Füredi [23] showed that the conclusion of the theorem holds if we only impose the hypothesis $|e \cap f| \neq t$ for $e \neq f$ provided $r \geq 2t+2$, with their proof using a somewhat more involved version of the Delta-system method. **We maybe prove this in the spread approximation chapter**

A similar argument works for more general kinds of intersection problems. Given a set $L \subseteq \{0, 1, \ldots, r-1\}$, we say that a hypergraph $\mathcal{H}$ is an (n, r, L)-system if it's an n-vertex r-graph such that $|e \cap f| \in L$ for all $e, f \in \mathcal{H}$ distinct. For example, $(n, r, \{t, t+1, \ldots, r\})$-systems are t-intersecting hypergraphs. Little is known about how large (n, r, L)-systems can be for general L, but one can get effective bounds in terms of ranks. To this end, for any $L \subseteq \{0, 1, \ldots, r-1\}$, define

$$\operatorname{rank}(r, L) = \max_{\mathcal{J}} \operatorname{rank}(\mathcal{J}),$$

where the maximum ranges over all $\mathcal{J} \subseteq 2^{[r]}$ containing no edges of size r which are intersection-closed with $|J| \in L$ for all $J \in \mathcal{J}$.

Theorem 13.8 (**Füredi; lost the reference**)**.** *If $\mathcal{H}$ is an (n, r, L)-system, then*

$$e(\mathcal{H}) = O(n^{\operatorname{rank}(r,L)}).$$

Proof. Let $\mathcal{H}', \mathcal{J}$ be as in Lemma 13.4. Observe that $\mathcal{J}$ is intersection closed and that $|J| \in L$ for all $J \in \mathcal{J}$ (since otherwise two edges of $\mathcal{H}' \subseteq \mathcal{H}$ would fail to have $|e \cap e'| \in L$). Thus letting c be the constant from Lemma 13.4, we have

$$e(\mathcal{H}) \leq c^{-1} e(H') \leq c^{-1}\binom{n}{\operatorname{rank}(\mathcal{J})} \leq c^{-1}\binom{n}{\operatorname{rank}(r, L)},$$

where this second inequality used Lemma 13.5 and the last inequality used the definition of $\operatorname{rank}(r, L)$. We conclude the result. $\square$

It's conjectured by Frankl that for all r, L there exist (n, r, L)-systems of size $\omega(n^{\mathrm{rank}(r,L)-1})$. This is unknown in general, but see e.g. [24, Theorem 16.6] for a construction of size $\Omega(n^{1+1/(r-1)})$ whenever $\mathrm{rank}(r, L) \geq 2$.

One of the best general bounds on the size of (n, r, L)-systems is the Deza-Erdős-Frankl theorem, which states that such a system $\mathcal{H}$ satisfies

$$e(\mathcal{H}) \leq \prod_{\ell \in L} \frac{n-\ell}{r-\ell} = O(n^{|L|}).$$

One can prove this result using a variant of the Delta-system method, though we omit doing so here. Instead, we give an easy proof of the asymptotic result by utilizing the following.

Lemma 13.9. *Every $L \subseteq \{0, 1, \ldots, r-1\}$ satisfies $\mathrm{rank}(r, L) \leq |L|$.*

Proof. We prove the result by induction on $|L|$, the case $|L| = 0$ being trivial. Assume we have proven the result for all L with $|L| < k$. We first consider the case $|L| = k$ and $0 \in L$. Let $\mathcal{J}$ be an intersection closed hypergraph on $2^{[r]} \setminus [r]$ with $|J| \in L$ for all $J \in \mathcal{J}$ and $\mathrm{rank}(\mathcal{J}) = \mathrm{rank}(r, L)$. For any vertex $x \in [r]$, let $\mathcal{J}_x = \{J - x : x \in J \in \mathcal{J}\}$ be the link hypergraph. If $L' = \{\ell - 1 : \ell \in L,\ \ell \neq 0\}$, then we see that $\mathcal{J}_x$ is intersection closed with $|J| \in L'$ for all $J \in \mathcal{J}_x$. Because $|L'| = |L| - 1$, our inductive hypothesis implies that $\mathrm{rank}(r-1, L') \leq |L| - 1$, i.e. there exists some set T of size $|L| - 1$ in $[r] \setminus \{x\}$ which is not contained in any edge of $\mathcal{J}_x$, which implies $T \cup \{x\}$ is a set of size $|L|$ not contained in any edge of $\mathcal{J}$. We conclude $\mathrm{rank}(r, L) = \mathrm{rank}(\mathcal{J}) \leq |L|$.

Now assume $0 \notin L$, and again let $\mathcal{J}$ be an intersection closed hypergraph on $2^{[r]} \setminus [r]$ with $|J| \in L$ for all $J \in \mathcal{J}$ with $\mathrm{rank}(\mathcal{J}) = \mathrm{rank}(r, L)$. Let $I = \bigcap_{J \in \mathcal{J}} J$. Note that by definition I is contained in every edge of $\mathcal{J}$, and by the intersection closed property we have $I \in \mathcal{J}$, and hence $|I| \in L$. Define $L' = \{\ell - |I| : \ell \geq |I|\}$, and note that the link hypergraph $\mathcal{J}_I = \{J \setminus I : J \in \mathcal{J}\}$ has all of its edge sizes lying in L'. Since $|L'| \leq |L|$ and $0 \in L'$, the previous case implies $\mathrm{rank}(\mathcal{J}_I) \leq |L|$, which implies there exists some set $J \subseteq [r] \setminus I$ of size L not contained in an edge of $\mathcal{J}_I$, and hence this set continues to not be contained in an edge of $\mathcal{J}$ (since every edge of $\mathcal{J}$ is the union of an edge of $\mathcal{J}_I$ with I). We conclude $\mathrm{rank}(r, L) = \mathrm{rank}(\mathcal{J}) \leq |L|$, proving the result. $\square$

This together with the previous theorem immediately gives the following.

Corollary 13.10. *If $\mathcal{H}$ is an (n, r, L)-system, then*

$$e(H) = O(n^{|L|}).$$

It is not difficult to show that this result is tight if $L = \{0, 1, \ldots, t-1\}$.

Before moving on, we note that while all of our applications here came from extremal set theory, the intersection semilattice lemma has application to other areas of extremal combinatorics as well. See for example [47], where this lemma is used to bound the Turán number of a class of linear hypergraphs called "expansions."

13.1 Proof of the Semilattice Intersection Lemma

We first recall that for every r-graph $\mathcal{H}$ there exists a subgraph $\mathcal{H}' \subseteq \mathcal{H}$ which is r-partite and which keeps a constant proportion of its edges. Thus it suffices to prove that if $\mathcal{H}$ has r-partition $\bigcup V_i$, then one can find a subgraph $\mathcal{H}' \subseteq \mathcal{H}$ and $\mathcal{J} \subseteq 2^{[r]} \setminus [r]$ such that $e(\mathcal{H}') \geq c(r,s)e(\mathcal{H})$ and

(1) $\mathcal{J}$ is intersection closed, i.e. $I, J \in J$ implies $I \cap J \in \mathcal{J}$.

(2) For every $e \in \mathcal{H}'$, $\{\mathrm{proj}(e \cap f) : f \in \mathcal{H}' \setminus \{e\}\} = \mathcal{J}$.

(3) $d^*_{\mathcal{H}'}(e \cap f) \geq s$ for all $e, f \in \mathcal{H}'$.

Claim 13.11. *If there exists $\mathcal{H}', \mathcal{J}$ satisfying (2) and (3) with $s \geq r+1$, then they automatically satisfy (1).*

Proof. Let $J_1, J_2 \in \mathcal{J}$. This means that for any edge $e \in \mathcal{H}'$, there exist edges e_1, e_2 such that e, e_i intersect exactly in the coordinates of J_i and that this intersection is the kernel of a sunflower in $\mathcal{H}'$ on at least $r+1$ petals. In particular, there must exist an edge $f \in \mathcal{H}'$ which contains $e \cap e_1$ and which is otherwise disjoint from e_2 (namely, f is one of the edges of the sunflower with core $e \cap e_1$). With this $\mathrm{proj}(f \cap e_2) = J_1 \cap J_2$, so necessarily $J_1 \cap J_2 \in \mathcal{J}$. $\square$

With this claim in mind, we only have to find $\mathcal{H}', \mathcal{J}$ satisfying (2) and (3) (this is immediate if $s \geq r+1$, and for all other values of s we can take $c(r,s) = c(r, r+1)$ and apply the $s = r+1$ result). For the rest of the proof, given a hypergraph $\mathcal{H}$, we define

$$\mathcal{I}(\mathcal{H}) = \{\mathrm{proj}(e \cap f) : e, f \in \mathcal{H},\ e \neq f\}.$$

Note that if (2) are (3) are satisfied for some $\mathcal{J}$, then it must be that $\mathcal{J} = \mathcal{I}(\mathcal{H})$.

Claim 13.12. *For any r-partite r-graph $\mathcal{H}$, one can decompose $\mathcal{H}$ as $\mathcal{H} = \mathcal{H}_0 \cup \bigcup_{I \in \mathcal{I}(\mathcal{H})} \mathcal{H}_I$ such that $\mathcal{H}_0$ satisfies (2) and (3) for some $\mathcal{J}$, and such that $I \neq \mathrm{proj}(K)$ for any set K which is the kernel of a sunflower on at least s petals in $\mathcal{H}_I$.*

Proof. Initially start with $\mathcal{H}_0 = \mathcal{H}$ and $\mathcal{H}_I = \emptyset$ for all $I \in \mathcal{I}$. Consider the following procedure. If at any point $\mathcal{H}_0$ satisfies (2) and (3) for some set $\mathcal{J}$, then we stop and output the current sets. Otherwise, it is not difficult to see that there must exist some edge $e \in \mathcal{H}_0$ and $K \subseteq e$ such that $\mathrm{proj}(K) \in \{\mathrm{proj}(f \cap g) : f, g \in \mathcal{H}_0,\ f \neq g\}$ but K is not the kernel of a sunflower with at least s petals in $\mathcal{H}_0$ (otherwise the conditions would be satisfied with $\mathcal{J} = \{\mathrm{proj}(f \cap g) : f, g \in \mathcal{H}_0,\ f \neq g\}$). Delete e from $\mathcal{H}_0$ and add it to $\mathcal{H}_{\mathrm{proj}(K)}$.

We claim that this procedure gives the desired result. Indeed, $\mathcal{H}_0$ satisfies (2) and (3) by construction. Assume for contradiction that there existed some I with $\mathcal{H}_I$ containing a sunflower on at least s petals with kernel K satisfying $\mathrm{proj}(K) = I$. Let e be the first edge of this sunflower that was added to $\mathcal{H}_I$ during the procedure. This implies that every edge of the sunflower was in $\mathcal{H}_0$ right before e was removed, i.e. that $\mathcal{H}_0$ contains a sunflower with at least s petals and kernel K. This contradicts us removing e from $\mathcal{H}_0$ at this step, so we include no such sunflower exists in any $\mathcal{H}_I$. $\square$

Claim 13.13. *Let $\mathcal{H}$ be an r-partite r-graph and I a set such that $I \neq \text{proj}(K)$ for any K which is the kernel of a sunflower with at least s petals. Then one can decompose $\mathcal{H}$ as $\mathcal{H} = \mathcal{H}^1 \cup \cdots \mathcal{H}^{r(s-1)}$ such that $I \notin \mathcal{I}(\mathcal{H}^j)$ for all j.*

Proof. Note that for each edge $e \in \mathcal{H}$, there exists a unique set $K \subseteq V$ with $\text{proj}(K) = I$. For each set of this form, let $\mathcal{H}(K)$ be its link hypergraph, i.e. $H(K) = \{e \setminus K : K \subseteq e \in H\}$. By hypothesis, none of the $\mathcal{H}(K)$ sets contain a matching of size s (since this translates to a sunflower of size s in $\mathcal{H}$ with kernel K). It is not difficult to see that one can decompose each $(r - |K|)$-graph $\mathcal{H}(K)$ into at most $(r - |K|)(s-1) \leq r(s-1)$ intersecting hypergrpahs $\mathcal{H}^1(K), \ldots, \mathcal{H}^{r(s-1)}(K)$ (e.g. by taking a largest matching M in $\mathcal{H}(K)$ and then assigning edges to $\mathcal{H}^i(K)$ if they contain the ith vertex which is contained in an edge of M). Let $\mathcal{H}^i[K] = \{e' \cup K : e' \in \mathcal{H}^i(K)\}$ and let $\mathcal{H}^i = \bigcup_K \mathcal{H}^i(K)$. Note that this decomposes $\mathcal{H}$, and that $K \neq e \cap f$ for any K with $\text{proj}(K) = I$ and $e, f \in H^i$ (as this would imply $e \setminus K, f \setminus K \in \mathcal{H}^i(K)$, and hence e, f contain an additional vertex since $\mathcal{H}^i(K)$ is intersecting). $\square$

By repeatedly applying the above two times a bounded number of times, one can decompose $\mathcal{H}$ as $\bigcup \mathcal{H}_i$ where each $\mathcal{H}_i$ satisfies (2) and (3). Taking the largest of these hypergraphs gives the desired result.

14 The Regularity Lemma TODO

To be done at some point. Until then, see, the lovely notes of Das which covers most of the content that I plan to write about.

15 Expansion and α-maximality

This almost surely needs to be rewritten to incorporate improvements to the bounds, as well as the claimed simplification mentioned at the end of "Towards the Erdős-Gallai Cycle Decomposition conjecture"

One of the nice features of random graphs is that they have good expansion properties; e.g. any set of vertices $B \subseteq V(G_{n,p})$ is likely to have about $pn|B|$ edges leaving B provided B is not too small. It is too much to ask that a graph has such strong expansion properties in general, but it is often the case that one can find subgraphs of arbitrary graphs which have reasonable expansion properties. There are many techniques in the field that achieve this end. The focus on this chapter will be an approached introduced by Tomon [62] which has the advantage of being both simple to state and powerful in applications.

Definition 1. Given a real number α, we say that a graph G is α-maximal if $e(G)/v(G)^{1+\alpha} = \max_{H \subseteq G} e(H)/v(H)^{1+\alpha}$. Equivalently, this says that if $e(G) = \gamma \cdot v(G)^{1+\alpha}$, then $e(H) \leq \gamma \cdot v(H)^{1+\alpha}$ for all $H \subseteq G$.

Observe that every graph has an α-maximal subgraph.

The motivation for this definition is that often in extremal graph theory, one wants to prove that graphs with $e(G) \geq \gamma v(G)^{1+\alpha}$ contain some desired structure. If such a result were true, then in particular any α-maximal subgraph of G must contain this structure, so being α-maximal is essentially the hardest case that one can consider. Moreover, it turns out that by reducing to α-maximal graphs, one gains a lot of nice expansion properties. Here and throughout this section we let $N(B)$ be the set of vertices $y \notin N(B)$ which are adjacent to a vertex in B, and we let $d(G)$ denote the average degree of G.

Proposition 15.1. *Let G be an n-vertex α-maximal graph with $\alpha \in (0,1]$ and $d(G) = \gamma n^\alpha$, and let $B \subseteq V(G)$ be such that $|B| \leq n/2$.*

(i) If G is non-empty, then $\gamma \geq \frac{1}{2}$.

(ii) The minimum degree of G is at least $\frac{1}{2}d(G) = \frac{1}{2}\gamma n^\alpha$.

(iii) We have $e(B, N(B)) \geq \frac{1}{4}\gamma n^\alpha |B|(1 + \alpha - (2|B|/n)^\alpha)$.

(iv) We have $|N(B)| > |B|((1 + \frac{1}{2}\alpha)(\frac{n}{2|B|})^{\alpha/(1+\alpha)} - 1)$.

The main benefit of (i) is that the bound is an absolute constant independent of α. Condition (ii) is obviously convenient to have. Note that in $G_{n,p}$ with $p = \gamma n^\alpha/n$, we have $\mathbb{E}[[e(B, N(B))] \approx \gamma n^\alpha |B|$ as long as $|B| \leq n/2$, so the level of expansion in (iii) is about as much as we could hope for. The bound for (iv) is roughly $(n^\alpha |B|)^{1/(1+\alpha)}$, which is best possible when $|B| \approx n$ **Though beyond this I don't have much intuition for why this is a reasonable condition to shoot for.**

Proof. For (i), taking $H \subseteq G$ to be a single edge implies $d(H)/v(H)^\alpha = 2^{-\alpha} \geq \frac{1}{2}$, so the same bound holds for G.

For (ii), let v be a vertex of minimum degree δ and let $H = G - v$. We have $d(H)/v(H)^\alpha \leq d(G)/v(G)^\alpha$ by definition of α-maximality, which is equivalent to

$$\frac{d(G)n - 2\delta}{(n-1)^{1+\alpha}} \leq \frac{d(G)}{n^\alpha}.$$

This implies

$$\frac{1}{2}d(G)(n - \frac{(n-1)^{1+\alpha}}{n^\alpha}) \leq \delta,$$

showing $\delta \geq \frac{1}{2}d(G)$.

For (iii) and (iv), let $C = V(G) \setminus B$. With this we have

$$\begin{aligned} e(B, N(B)) = e(G[B \cup C]) - e(G[B]) - e(G[C]) &= \frac{1}{2}\gamma(|B| + |C|)^{1+\alpha} - e(G[B]) - e(G[C]) \\ &\geq \frac{1}{2}\gamma|C|^{1+\alpha}((1 + |B|/|C|)^{1+\alpha} - \frac{1}{2}\gamma|B|^{1+\alpha} - \frac{1}{2}\gamma|C|^{1+\alpha}. \end{aligned}$$

Using $(1 + |B|/|C|)^{1+\alpha} \geq 1 + (1+\alpha)|B|/|C|$ and that $|C| \geq (n/2)^\alpha \geq \frac{1}{2}n^\alpha$, we find that this is at least

$$\frac{1}{2}\gamma(1+\alpha)|B||C|^\alpha - \frac{1}{2}\gamma|B|^{1+\alpha} \geq \frac{1}{2}\gamma|B|(\frac{1}{2}(1+\alpha)n^\alpha - |B|^\alpha),$$

giving (iii).

Similarly for (iv) we observe

$$e(G[B \cup N(B)]) \geq e(G[B \cup C]) - e(G[C]) \geq \frac{1}{2}\gamma(|B| + |C|)^{1+\alpha} - \frac{1}{2}\gamma|C|^{1+\alpha} \geq \frac{1}{2}\gamma(1+\alpha)|B||C|^\alpha.$$

However, by α-maximality we have $e(G[B \cup N(B)]) \leq \frac{1}{2}\gamma(|B| + |N(B)|)^{1+\alpha}$. Combining these inequalities gives

$$|N(B)| \geq ((1+\alpha)|B||C|^\alpha)^{1/(1+\alpha)} - |B|,$$

giving the result. □

Our main application of α-maximal graphs will be to something called rainbow Turán numbers, which were first introduced by Keevash, Mubayi, Sudakov, and Verstraëte [40]. We say that a colored graph F is *rainbow* if all of the colors of its edges are distinct. Given a set of graphs $\mathcal{F}$, we define $\mathrm{ex}^*(n, \mathcal{F})$ to be the maximum number of edges a *properly colored* n-vertex graph G can have without containing a rainbow copy of any $F \in \mathcal{F}$.

Note that $\mathrm{ex}(n, \mathcal{F}) \leq \mathrm{ex}^*(n, \mathcal{F})$ for all $\mathcal{F}$ (since we can take any extremal $\mathcal{F}$-free graph and give each edge a distinct color), and in general these two quantities can be somewhat far from each other. Indeed, let $\mathcal{C}$ denote the set of all cycles, which means $\mathrm{ex}(n, \mathcal{C}) = n - 1$. On the other hand, we have $\mathrm{ex}^*(n, \mathcal{C}) \geq n \log_2 n$ when n is a power of 2. This is because one can take G to be an n-vertex hypercube where an edge uv is colored i if u, v differ in the ith bit. It is not difficult to see that this is a proper coloring which contains no rainbow cycles.

Even though the problem of determining $\mathrm{ex}(n, \mathcal{C})$ is easy, determining $\mathrm{ex}^*(n, \mathcal{C})$ is an open and seemingly difficult problem. The first non-trivial upper bounds on $\mathrm{ex}^*(n, \mathcal{C})$ were established by Das, Lee, and Sudakov [17], and later O. Janzer [34] managed to prove $\mathrm{ex}^*(n, \mathcal{C}) = O((\log n)^4 n)$. Currently the best known upper bound is the following result due to Tomon [62].

Theorem 15.2 ([62]). *We have* $\mathrm{ex}^*(n,\mathcal{C}) = (\log n)^{2+o(1)}n$.

The main lemma we need to prove this is the following. Here a *Q-rainbow path* refers to a rainbow path which only uses colors in the set Q.

Lemma 15.3. *Given $p_c \in (0,1)$, there exists a constant C such that the following holds. Let $\lambda > C(\log\log n)^{10}$, and let G be an n-vertex α-maximal graph with proper coloring $c : E(G) \to R$ and $d(G) > C\lambda^2\alpha^{-2}n^\alpha$. If $Q \subseteq R$ is chosen by including each color independently and with probability p_c, then for every $v \in V(G)$, with probability at least $1-O(\alpha^{-1}e^{-\Omega(\lambda^{1/2})})$ at least $n/3$ vertices of G can be reached by a Q-rainbow path.*

Before proving this lemma, let us first show how this implies the main result.

Proof of Theorem 15.2. Let G be an n-vertex graph with $e(G) \geq 2(\log n)^{2+\epsilon}n$ and $c : E(G) \to R$ a proper coloring, and let $\alpha = 1/\log_2(G)$ and $\lambda = (\log n)^{\epsilon/10}$. Let H be a subgraph of G maximizing $d(H)/v(H)^\alpha$ and $m = v(H)$. Note that H is α-maximal and $d(H) \geq d(G)\cdot(v(H)/v(G))^\alpha \geq \frac{1}{2}d(G)$ due to our choice of α, and this quantity is at least $C\lambda^2\alpha^{-2}m^\alpha$ for n sufficiently large.

Pick some $v \in V(H)$. Partition R into four parts Q_1, Q_2, Q_3, Q_4 be independently and uniformly at random assigning each color to one of these sets, and let B_i be the set of vertices that can be reached by v with a Q_i-path. By Lemma 15.3 with $p_c = 1/4$, we see that with probability at least $4/5$ we have $|B_i| \geq n/3$, so there exists some partition $Q_1, \ldots, Q_4$ such that $|B_i| \geq n/3$ holds for all i.

Note that $B_i \cap B_j \neq \emptyset$ for some $i \neq j$, and let $w \in B_i \cap B_j$. By definition this means there exist rainbow paths P_i, P_j from v to w using colors in Q_i, Q_j. Thus the union of these two paths is a rainbow graph which contains a cycle, proving the result. $\square$

It remains to prove Lemma 15.3. Given a graph G and a proper coloring $c : E(G) \to R$, define $N_{Q,\phi}(v)$ with $\phi : V(G) \to 2^{V(G)\cup R}$ to be the set of vertices w with $vw \in E(G)$, $c(vw) \in Q \setminus \phi(v)$, and $w \notin \phi(v)$. That is, the is the neighborhood if we restrict to colors in Q and forbid some set of neighbors/colors for v to use. We define $N_{Q,\phi}(B) = \bigcup_{v\in B} N_{Q,\phi}(v) \setminus B$. To prove Lemma 15.3, we show that α-maximal graphs have vertex expansion about as strong as in Proposition 15.1 even when forbidding some colors/vertices.

Lemma 15.4. *Let $p_c, \alpha \in (0,1]$, let n be a positive integer and $\lambda > 10^{10}$. Let G be an n-vertex graph, $c : E(G) \to R$ a proper edge coloring, and $B \subseteq V(G)$ such that the following hold:*

- *G is α-maximal*
- *$d := d(G) \geq \lambda(p_c \cdot \alpha)^{-1}$,*
- *$\phi : V(G) \to 2^{V(G)\cup R}$ is such that $|\phi(v)| \leq d\alpha/32$ for all $v \in V(G)$, and*
- *$2\lambda^2 p_c^{-1} < |B| < n/2$.*

Let $Q \subseteq R$ be obtained by including each color independently and with probability p_c. Then with probability at least $1 - e^{-\Omega(\lambda^{1/2})}$ we have

$$|N_{Q,\phi}(B)| \geq \frac{1}{4}|B| \min\left\{\frac{d \cdot p_c \cdot \alpha}{64\lambda^{1/2}}, \left(\frac{n}{2|B|}\right)^{\alpha/(1+\alpha)} - 1\right\}$$

Proof. Let $d = \gamma n^\alpha$, and let H be the bipartite (uncolored) graph on $B \cup N(B)$ such that $x \in B$ and $y \in N(B)$ are adjacent if $xy \in E(G)$, $y \notin \phi(x)$ and $c(xy) \notin \phi(x)$. Let $H_Q \subseteq H$ be the (random) subgraph which only includes edges xy with $c(xy) \in Q$. Thus our problem is equivalent to showing $|N_{H_Q}(B)|$ is large with high probability.

Since we're aiming something comparable to that of Proposition 15.1 (namely, this is basically what we get when the right term in the lemma achieves the minimum), one might try to just naively replicate that proof. This almost works, but to get things to occur with high probability we need the vertices of $N_H(B)$ to have large degrees. To this end, let $S \subseteq N_H(B)$ be the vertices w such that $|N_H(w) \cap B| \geq \lambda^{1/2} p_c^{-1} =: \Delta$, and let $T = N_G(B) \setminus S$.

Claim 15.5. *If $e_G(B, T) \leq d\alpha|B|/16$, then the result follows.*

Note that the claim involves edges of the original graph G, not H.

Proof. Let $C = V(G) \setminus B$, noting that $|C|^\alpha \geq (\frac{1}{2}n)^\alpha \geq \frac{1}{2}n^\alpha$, and since $d \geq \frac{1}{2}\gamma n^\alpha$ by Proposition 15.1, we conclude

$$e_G(B, T) \leq \frac{1}{8}\alpha\gamma|B||C|^\alpha.$$

Note that

$$E(G) = E(G[B \cup S]) \cup E(G[C]) \cup E(G[B, T]),$$

where $E(G[B', T])$ denotes the set of edges of G with one end in B' and the other in T. To see this, we note that vertices of B can only be adjacent to vertices of $B \cup N_G(B) = B \cup S \cup T$. With this we have

$$\begin{aligned} e_G(B \cup S) &\geq e(G) - e(G[C]) - e_G(B, T) \\ &\geq \frac{1}{2}\gamma(|B| + |C|)^{1+\alpha} - \frac{1}{2}\gamma|C|^{1+\alpha} - \frac{1}{8}\alpha\gamma|B||C|^\alpha, \end{aligned}$$

where this inequality used $e(G) = \frac{1}{2}\gamma n^{1+\alpha} = \frac{1}{2}\gamma(|B'| + |C|)^{1+\alpha}$, α-maximality, and the inequality noted above. Note that

$$(|B| + |C|)^{1+\alpha} = |C|^{1+\alpha}(1 + |B|/|C|)^{1+\alpha} \geq |C|^{1+\alpha} + (1+\alpha)|B||C|^\alpha.$$

Using this gives

$$e_G(B \cup S) \geq \frac{1}{2}\gamma(1+\alpha)|B|C|^\alpha - \frac{1}{8}\alpha\gamma|B||C|^\alpha \geq \frac{1}{2}\gamma(1 + \frac{1}{2}\alpha)|B||C|^\alpha.$$

By α-maximality we have $e_G(B \cup S) \leq \frac{1}{2}\gamma(|B| + |S|)^{1+\alpha}$, so in total this implies

$$|S| \geq \left((1 + \frac{1}{2}\alpha)|B||C|^\alpha\right)^{1/(1+\alpha)} - |B|.$$

For $w \in S$, let X_w be the indicator random variable for the event $w \notin N_{H_Q}(B)$. Then

$$\Pr[X_w = 1] = (1-p_c)^{d_{H_Q}(w)} \le (1-p_c)^{\Delta} \le e^{-\lambda^{1/2}}.$$

Thus if $X = \sum_{w \in S} X_w$ then $\mathbb{E}[X] \le |S|e^{-\lambda^{1/2}}$, and by Markov's inequality this means $\Pr[X \ge |S|/2] \le 2e^{-\lambda^{1/2}}$. This gives the result. □

From now on we assume $e_G(B,T) > d\alpha|B|/16$. For each $w \in T$, let Y_w be the indicator random variable for the event $w \in N_{H_Q}(B)$. Then by the inequality $(1-a)^b \ge 1 - \frac{1}{2}ab$ for $ab < \frac{1}{2}$,

$$\mathbb{E}[Y_w] = 1 - (1-p_c)^{d_H(w)} \ge \frac{1}{2}\min\{1, p_c \cdot d_H(w)\}.$$

We partition T into two sets based on which term prevails in this minimum. Namely, let $T_1 = \{w \in T : d_H(w) \le p_c^{-1}\}$ and $T_2 = T \setminus T_1$. If $Y = \sum_{w \in T} Y_w$, then

$$\mathbb{E}[Y] \ge \sum_{w \in t_1} \frac{1}{2}p_c \cdot d_H(w) + \frac{1}{2}|T_2| = \frac{1}{2}p_c e_H(B,T_1) + \frac{1}{2}|T_2| \ge \frac{1}{2}p_c e_H(B,T_1) + \frac{1}{2}\lambda^{-1/2}p_c e_H(B,T_2),$$

where this last step used that each vertex of $T_2 \subseteq T$ has degree at most $\Delta = \lambda^{1/2}p_c^{-1}$ in H. Thus

$$\mathbb{E}[Y] \ge \frac{1}{2}\lambda^{-1/2}p_c e_H(B,T).$$

By hypothesis,

$$e_H(B,T) \ge e_G(B,T) - \sum_{v \in B} |\phi(v)| \ge \frac{1}{32}d\alpha|B|.$$

This together with the hypothesis $d \ge \lambda^{1/2}(p_c \cdot \alpha)^{-1}$ implies $\mathbb{E}[Y] \ge \frac{1}{64}\lambda^{1/2}|B|$.

Note that Y is a function of which colors survive in Q. Each color appears at most $|B|$ times since G is properly colored, so changing Q by a single element changes Y by at most $|B|$, i.e. Y is $|B|$-Lipschitz. By the multiplicative Azuma inequality (Lemma 6.6), we have $\Pr[Y \le \frac{1}{2}\mathbb{E}[Y]] \le e^{-\Omega(\lambda)}$. Since $Y = |N_{H_Q}(B)|$, we conclude the result.

□

Proof of Lemma 15.3. Similar to our proofs involving spread hypergraphs, we will iteratively generate random sets Q_i a total of $\ell = 100\alpha^{-1}\log\log(n)$ times and take $Q = \bigcup Q_i$, iteratively arguing that each Q_i is likely to have good properties.

Let q_c be the unique solution to $p_c = 1-(1-q_c)^{\ell}$; the main take away being that $q_c = \Omega(p_c/\ell)$. For $1 \le i \le \ell$, let Q_i be obtained by including each color of R independently and with probability q_c (and independent of any other Q_j set), noting that $\bigcup_{i=1}^{\ell} Q_i$ has the same distribution as Q.

Let B_i be the set of vertices x that can be reached from v by some $(Q_1 \cup \cdots \cup Q_i)$-rainbow path P_x of length at most i. Let $\phi_i : V(G) \to 2^{V(G) \cup R}$ be the function which maps to the vertices and colors of P_x if $x \in B_i$, and otherwise $\phi(x) = \emptyset$. Note that $|\phi_i(x)| \le 2i \le 2\ell$. We wish to show that $|B_i|$ is rapidly increasing with high probability.

First note that B_1 is just the set of neighbors of v with $c(vx) \in Q_1$. By Proposition 15.1 we have $d(v) \geq \frac{1}{2}d$, so $\mathbb{E}[|B_1|] \geq \frac{1}{2}dq_c$. Thus by the Chernoff bound we will have $|B_1| \geq \frac{1}{4}dq_c > 2\lambda^2 q_c^{-1}$ with high probability, so from now on we assume this is the case..

Note that $N_{Q_i,\phi_i}(B_i \cap U_i) \subseteq B_{i+1}$. Since $d \geq \lambda^2(\alpha \cdot q_c)^{-1}n^\alpha$ by hypothesis, and since $|B_i| \geq |B_1| > 2\lambda^2 q_c^{-1}$, as long as $|B_i| < n/3$ we can apply Lemma 15.4 to get

$$N_{Q_i,\phi_i}(B_i) \geq \frac{1}{4}|B_i| \min\left\{\frac{d \cdot q_c \cdot \alpha}{64\lambda^{1/2}}, \left(\frac{n}{2|B_i|}\right)^{\alpha/(1+\alpha)} - 1\right\}$$

with probability at least $1 - e^{-\Omega(\lambda^{1/2})}$. Note that the leftside of the minimum is always at least n^α, so the minimum is always achieved by the righthand side. Using this and $\alpha \leq 1$ gives

$$|N_{Q_i,\phi_i}(B_i)| \geq \frac{1}{4}|B_i|((n/2|B_i|)^{\alpha/2)} - 1)$$

with probability at least $1 - e^{-\Omega(\lambda^{1/2})}$. Thus with probability at least $1 - \ell e^{-\Omega(\lambda^{1/2})} = 1 - O(\alpha^{-1}e^{-\Omega(\lambda^{1/2})})$ this holds for all i (note that the $\log\log(n)$ gets absorbed by $e^{-\Omega(\lambda^{1/2})}$) since $\lambda \geq (\log\log n)^2$). We claim that this implies $|B_{\ell-1}| \geq n/3$. And indeed, using that $|B_{i+1}| \geq |B_i| + |N_{Q_i,\phi_i}(B_i)|$ (since B_i and $N(B_i) \supseteq N_{Q_i,\phi_i}(B_i)$ by definition), one can prove by induction that $|B_i| \geq (n/2)^{1-(1-\alpha/16)^i}$ provided $|B_{i-1}| \leq n/3$. This gives the result. □

Tomon [62] proved several other nice results using a result which extends Lemma 15.3 in two ways. The first way is by enforcing short paths from v provided we don't require v to reach as many vertices (and it is easy to adapt our current proof to achieve this end). The other extension is that it allows one to sample a random set of vertices $U \subseteq V(G)$ in addition to a random set of colors, and which guarantees short paths from v to a large set of vertices. To state such a result, we say that a path is a (U,Q)-rainbow path if it is a rainbow path whose internal vertices all lie in U and whose colors all lie in U.

Lemma 15.6. *There exists a sufficiently large constant C such that the following holds. Let $p, p_c, \alpha \in (0,1]$, n a positive integer, $\tau \in [1/\log_3 n, \frac{1}{2})$, and $\lambda > C(\log\log n)^1 0$. Let G be an n-vertex α-maximal graph with proper edge coloring $c : E(G) \to R$ with average degree $d = d(G)$ satisfying either $d > C\lambda^2(\alpha^2 \cdot p_c^2)^{-1}n^\alpha$ if $p = 1$, and otherwise $d > C\lambda^2(\alpha^3 \cdot p \cdot p_c^2)^{-1}n^\alpha$.*

Let $U \subseteq V(G)$ be obtained by including each vertex independently and with probability p, and similarly define $Q \subseteq R$. For each $v \in V(G)$, with probability at least $1 - O(\alpha^{-1}e^{-\Omega(\lambda^{1/2})})$, at least $n^{1-\tau}$ vertices of G can be reached from v by a (U,Q)-path of length at most $O(\alpha^{-1}\log(1/\tau))$.

To prove this, one needs to extend Lemma 15.4 to say that the same conclusion holds for $N_{Q,\phi}(U)$ with $U \subseteq B$ obtained by including each vertex independently and with probability p. This isn't too hard to prove if the vertices of B all have reasonable maximum degree, and one extra case deals with the situation where this doesn't happen.

With Lemma 15.6 it is possible to prove results about rainbow Turán numbers of subdivisions of K_t. To this end, let $\mathcal{K}_t$ denote the set of subdivisions of K_t (i.e. the graphs which can be obtained by subdividing each edge of K_t some number of times). Mader [46] showed $\mathrm{ex}(n, \mathcal{K}_t) = O_t(n)$, and again the hypercube shows $\mathrm{ex}^*(n, \mathcal{K}_t) = \Omega(n \log n)$ for $t \geq 3$. Jiang, Letzter, Methuku, and Yepremyan [35] showed $\mathrm{ex}^*(n, \mathcal{K}_t) = O((\log n)^{60} n)$. These bounds were improved significantly by Tomon [62].

Theorem 15.7 ([62]). *For all fixed t we have* $\mathrm{ex}^*(n, \mathcal{K}_t) \leq (\log n)^{6+o(1)} n$.

Proof Sketch. Let G be an n-vertex graph with $e(G) \geq 2(\log n)^{6+\epsilon} n$ and $c : E(G) \to R$ a proper coloring, and let $\alpha = 1/\log_2(G),\ s = (\log n)^{1+\epsilon/10},\ p = p_c = 1/s$, and $\lambda = (\log n)^{\epsilon/10}$. For some slight ease of notation we assume G is α-maximal (though it's easy for the rest of the proof to go through by reducing to an α-maximal subgraph). Note that by our choice of parameters, $d(G) \geq C\lambda(\alpha^3 p \cdot p_c^2)^{-1} n^\alpha$.

Define an auxiliary graph H where two vertices v, w are adjacent if there exist at least $s/6$ internally disjoint paths from v to w such that no color is used more than once int he union of the paths.

We claim that if H has minimum degree at least $n/6$, then G contains a rainbow K_t-subdivision. Indeed, by Theorem 5.4, H (easily) contains a 1-subdivision of K_t. One can then greedily replace each edge with a rainbow path which doesn't use any vertices or colors that have already been used.

To show that H has this minimum degree, we apply Lemma 15.6 with the stated parameters and $\tau = 1/\log_3(n)$ to any vertex v. By a similar argument to before, this implies there exist partitions $U_1, \ldots, U_s$ and $Q_1, \ldots, Q_s$ such that the sets B_i of vertices we can reach from v with a (U_i, Q_i)-rainbow path all have size at least $n/3$. This implies that there exist at least $n/6$ vertices w in at least s of the B_i sets, proving that $d_H(v) \geq n/6$. $\square$

Part V

Hypergraph Containers

This part is heavily based off of lecture notes by Balogh [5]. Throughout this section we let $\mathcal{I}(H)$ denote the set of independent sets of a hypergraph H and $\mathcal{I}_m(H)$ the set of independent sets of size m. We adopt the notation $\binom{n}{\leq k}$ to denote the number of subsets of $[n]$ of size at most k. Many of the bounds in this part will be rough approximations to the truth in order to emphasize the intuition of the results and techniques rather than the nitty gritty detail that is actually required.

16 Introduction

Many problems in extremal combinatorics can be stated in terms of independent sets of hypergraphs. For example, one can define $\mathcal{H}_n^{AP}$ to be the 3-graph on $[n]$ where every triple $S \subseteq [n]$ is a hyperedge if and only if S is a 3-term arithmetic progression. Thus Roth's theorem is equivalent to saying $\alpha(\mathcal{H}_n^{AP}) = o(n)$. Similarly one can define $\mathcal{H}_n^{\Delta}$ to be the 3-graph whose vertex set is $E(K_n)$ and whose hyperedges are triples of edges in K_n which form a triangle. Independent sets of $\mathcal{H}_n^{\Delta}$ are triangle-free subgraphs of K_n, so Mantel's theorem says $\alpha(H_n^{\Delta}) = \lfloor n^2/4 \rfloor$.

This part is dedicated to a powerful method of upper bounding the size of $\mathcal{I}(H)$. Observe that for any hypergraph H we have

$$2^{\alpha(H)} \leq |\mathcal{I}(H)| \leq \binom{n}{\alpha(H)} 2^{\alpha(H)} \leq (2n)^{\alpha(H)}.$$

In particular, the upper bound follows because every independent set is a subset of a set of size $\alpha(H)$. More generally, we say that a collection $\mathcal{C}$ of subsets $C \subseteq V(H)$ is a set of *containers* for H if every independent set $I \in \mathcal{I}(H)$ is a subset of some $C \in \mathcal{C}$. If such a set of containers exists, then

$$|\mathcal{I}(H)| \leq \sum_{C \in \mathcal{C}} 2^{|C|} \leq |\mathcal{C}| 2^{\max_{C \in \mathcal{C}} |C|}. \tag{13}$$

Thus we will get an effective upper bound on $|\mathcal{I}(H)|$ whenever we can find a small collection of containers, each of which are relatively small. Sometimes we will be interested in finding the number of independent sets of H of size m. The same reasoning as above gives

$$\binom{\alpha(H)}{m} \leq |\mathcal{I}_m(H)| \leq |\mathcal{C}| \binom{\max_{C \in \mathcal{C}} |C|}{m}. \tag{14}$$

The method of hypergraph containers gives a systematic way of obtaining such a collection of containers whenever H satisfies some fairly mild conditions. The main condition we need is that the codegrees of H to be relatively small, and in practice this often corresponds to having some notion of supersaturation.

17 Graph Containers

While the general method of containers involves bounding independent sets of *hypergraphs*, one can get pretty far by only considering independent sets of *graphs*. To this end we prove the following graph container lemma, which will be the main workhorse for the rest of this section. Recall that a collection $\mathcal{C}$ of subsets $C \subseteq V(G)$ is a set of *containers* for G if every independent set $I \in \mathcal{I}(G)$ is a subset of some $C \in \mathcal{C}$.

Lemma 17.1. *Let G be an n-vertex graph and $t > 0$ a positive number. There exists a collection $\mathcal{C}$ of containers such that*

(a) $|\mathcal{C}| \leq \binom{n}{\leq n/t}$.

(b) $\Delta(G[C]) < t - 1$ *for all* $C \in \mathcal{C}$.

In other words, there exists a small set of containers $\mathcal{C}$ such that each $C \in \mathcal{C}$ is "small" in the sense that it induces a graph with small maximum degree.

Proof. Our proof will be algorithmic: we construct a (deterministic) algorithm which takes as input a set $I \subseteq V(G)$ and which outputs a pair $(S(I), A(I))$ such that $S(I) \subseteq I \subseteq S(I) \cup A(I)$, and we will ultimately use $\{S(I) \cup A(I) : I \in \mathcal{I}(G)\}$ as our set of containers. We now describe the algorithm.

Fix an arbitrary ordering of $V(G)$. As input we take in an independent set $I \subseteq V(G)$. We initially set $S = \emptyset$ and $A = V(G)$ (the former corresponds to a set of "selected" vertices which are in I, and the latter to the set of "available" vertices which could possibly be in I given the current stage of the algorithm). The algorithm proceeds as follows:

Step 1 If $\Delta(G[A]) < t - 1$, output $(S(I), A(I))$. Otherwise proceed to Step 2.

Step 2 Let v be the vertex of maximum degree in $G[A]$, with ties being broken according to the fixed ordering of $V(G)$. If $v \notin I$, then set $A = A - v$ and repeat Step 1. Otherwise proceed to Step 2.

Step 3 Set $A = A - v - N_{G[A]}(v)$, $S = S \cup \{v\}$. Proceed to Step 1.

Let's reiterate what's going on here. It's not difficult to show inductively that we always have $I \subseteq S \cup A$, so $S \cup A$ serves as a container set for I, and we would like to trim this set down as much as possible. We do this by selecting a vertex $v \in A \cap I$ and adding it to S. If v has large degree in $G[A]$, then v being in the independent set I means that its many neighbors are not, so we get to remove all of these vertices from A while maintaining $I \subseteq S \cup A$. In particular, since we keep going so long as $G[A]$ has large maximum degree, we know at each step of this process that we're removing many vertices from A.

Define

$$\mathcal{C} = \{S(I) \cup A(I) : I \in \mathcal{I}(G)\},$$

which is a set of containers since $I \subseteq S(I) \cup A(I)$ at every step of the algorithm. Since we terminate the algorithm precisely when $\Delta(G[A(I)]) = \Delta(G[S(I) \cup A(I)]) < t-1$ (the equality holds since $S(I)$ has no neighbors in $S(I) \cup A(I)$), (b) holds. It thus remains to verify (a). To do this, we note the following which is easy to verify.

Claim 17.2. *Let I_1, I_2 be two independent sets and let $(S_1, A_1), (S_2, A_2)$ be their outputs from the algorithm. If $S_1 = S_2$, then $A_1 = A_2$.*

This claim implies that given $S(I)$, the container $S(I) \cup A(I)$ is uniquely determined[1]. In particular, if we always have $|S(I)| \leq n/t$, then the number of containers will be at most $\binom{n}{\leq n/t}$. And indeed, each round of the algorithm has $\Delta(G[A]) \geq t-1$, so every time a vertex is added to S at least $1 + (t-1) = t$ vertices are removed from A. In particular, at most n/t vertices can be added to S, giving the result. $\square$

Actually, a closer inspection of the proof gives the following.

Lemma 17.3. *Let G be a graph on n vertices and $t \in \mathbb{R}$. There is a collection $\mathcal{C}$ of containers and functions*

$$f : \mathcal{I}(G) \to \binom{V(G)}{\leq n/t}, \qquad g : \binom{V(G)}{\leq n/t} \to \mathcal{C}$$

such that the following hold.

(a) The function g is a surjection. In particular, $|\mathcal{C}| \leq \binom{n}{\leq n/t}$.

(b) We have $\Delta(G[C]) < t-1$ for all $C \in \mathcal{C}$.

(c) For every $I \in \mathcal{I}(G)$ we have

$$f(I) \subseteq I \subseteq g(f(I)).$$

Proof. Consider the exact same algorithm as before. Define $f(I) = S(I)$ and $g(S) = C(S)$ (if $S \neq S(I)$ for any I, then assign g arbitrarily). It's not hard to check that this works. $\square$

The extra source of power of this lemma is that for each $I \in \mathcal{I}$ we are given some set $S = f(I)$ contained in I. In many examples this extra information is needed to get tight upper bounds when counting independent sets, though for pedagogical purposes we will often work with the simpler Lemma 17.1 to get close to tight results.

In the coming subsections we'll show how to use Lemma 17.1 to solve several combinatorial problems. All of the proofs will be very similar to each other, though they'll become increasingly sophisticated as we go along.

Before going on, let us briefly note that there are many variants of Lemma 17.1 that one can prove using a similar approach. These variants of Lemma 17.1 are both a blessing and a curse since they give many options for how to solve a given problem (and it isn't always clear which is best).

[1]Because of this, S is often called a "certificate" or "fingerprint" of I.

17.1 Regular Graphs

Our first application of Lemma 17.1 will be to count the number of independent sets in a d-regular graph. As a point of reference, it is not difficult to show that if G consists of $n/2d$ disjoint copies of $K_{d,d}$, then

$$|\mathcal{I}(G)| = (2^{d+1} - 1)^{n/2d} = 2^{n/2+n/2d+o(n)}.$$

Thus for d-regular graphs, we can't possibly hope to prove an upper bound on $|\mathcal{I}(G)|$ stronger than roughly $2^{n/2}$ when d is large. We can prove that this is close to best possible using containers.

Theorem 17.4. *Let G be a d-regular n-vertex graph with $\log n \ll d \ll n/2$. Then*

$$|\mathcal{I}(G)| \le 2^{n/2+o(n)}.$$

In fact, it turns out that $|\mathcal{I}(G)| \le (2^{d+1} - 1)^{n/2d}$ for all d-regular n-vertex graphs. This was proven for bipartite graphs by Kahn [37] using entropy, and the problem was solved in full by Zhao [64]. As far as I'm aware, the proof of Theorem 17.4 presented here is due to Balogh.

As a first step to proving Theorem 17.4, we will apply Lemma 17.1 to our graph G to get a collection of containers $\mathcal{C}$. We would like to conclude the result by the observation from (13):

$$|\mathcal{I}(G)| \le \sum_{C \in \mathcal{C}} 2^{|C|} \le |\mathcal{C}| 2^{\max_{C \in \mathcal{C}} |C|},$$

but there's an issue with this. Namely, Lemma 17.1 only tells us that each $C \in \mathcal{C}$ induces a graph in G with small maximum degree. For a general graph this tells us nothing about $|C|$, but fortunately in d-regular graphs, $G[C]$ having small maximum degree is only possible if C is small. The following states a precise version of the contrapositive of the previous sentence.

Lemma 17.5. *For any $\epsilon > 0$, if G is a d-regular graph and $C \subseteq V(G)$ with $|C| = n/2 + \epsilon n$, then $\Delta(G[C]) \ge 2\epsilon d$.*

This lemma is a form of supersaturation: a d-regular graph can have a subset of size $n/2$ with $G[C]$ empty (e.g. if G is bipartite), but if C is just a bit larger than this, then it must have relatively high maximum degree. As we will see, supersaturation results are almost always a necessary ingredient for applying the method of containers.

Proof. Because the maximum degree is always at least the average degree, we have

$$\Delta(G[C]) \ge 2e(G[C])/|C| \ge 2e(G[C])/n$$

, so it will suffice to show that $e(G[C])$ is large. To do this, we let $\overline{C} = V(G) \setminus C$ and note that

$$d|C| = \sum_{v \in C} d(v) = 2e(G[C]) + e(C, \overline{C}) \le 2e(G[C]) + d|\overline{C}|.$$

Because $|C| = n/2 + \epsilon n$ and $|\overline{C}| = n/2 - \epsilon n$, in total this implies

$$2e(G[C]) \ge 2\epsilon dn.$$

Combining this with the observation at the start gives the result. $\square$

Corollary 17.6. *For all t, if G is an n-vertex d-regular graph, then there exists a set of containers $\mathcal{C}$ with $|\mathcal{C}| \le \binom{n}{\le n/t}$ and $|C| \le \frac{1}{2}n + \frac{t}{d}n$ for all $C \in \mathcal{C}$.*

Proof. Let $\mathcal{C}$ be a set of containers as guaranteed by Lemma 17.1. Because $\Delta(G[C]) < t - 1 \le t$, Lemma 17.5 implies that $|C| \le \frac{1}{2}n + \frac{t}{d}n$. $\square$

With this we can prove Theorem 17.4.

Proof of Theorem 17.4. At this point all we need to do is use (13) after applying Corollary 17.6 with a carefully chosen value of t. Note that

$$|\mathcal{C}| \approx \binom{n}{n/t} \approx 2^{n\log(t)/t},$$

and we already know $2^{\max|C|} \approx 2^{\frac{1}{2}n + \frac{t}{d}n}$. Thus to minimize $|\mathcal{C}| \cdot 2^{\max|C|}$ we should choose t so that $\frac{t}{d} \approx \log(t)/t$, and in particular $t = \sqrt{d\log n}$ is a reasonable choice. One can verify with a more formal argument that this does indeed give the desired result after applying (13). $\square$

We note that the statement of Corollary 17.6 and the optimization of t in the proof of Theorem 17.4 is in some sense independent[1] of the problem of determining $|\mathcal{I}(G)|$ for G a d-regular graph. That is, these results are effective for other problems which involve counting independent sets of d-regular graphs.

For example, recall that a q-coloring of a graph G is a map $\chi : G \to [q]$ such that $\chi(u) \neq \chi(v)$ whenever $uv \in E(G)$. Equivalently, a q-coloring is a partition of $V(G)$ into independent sets $I_1, \ldots, I_q$, With this latter formulation, we can use containers to get an effective bound on the number of q-colorings of G, which we'll denote by $X_q(G)$.

Again, let's consider a test case to figure out how strong of a bound we could possibly hope to prove. Let G be $n/2d$ disjoint copies of $K_{d,d}$. We know that G has close to as many independent sets as it could possible have, so it seems plausible that it would have many q-colorings as well. In particular, one can prove that $X_q(G) \approx (q/2)^n$, and once again we can prove that this is essentially best possible.

Theorem 17.7 ([29])**.** *Let G be an n-vertex d-regular graph and q an integer such that $q^2 \log n \ll d$. Then*

$$X_q(G) \le (q/2 + o(1))^n.$$

We note that a stronger result was proven by Galvin [29] with a somewhat more involved proof.

Proof. By the same reasoning as in Theorem 17.4, there exists a set of containers $\mathcal{C}$ for G such that $|\mathcal{C}| \approx 2^{\sqrt{\frac{\log n}{d}}n}$ and $|C| \approx \frac{1}{2}n$ for each $C \in \mathcal{C}$. Consider all vectors of the form $(C_1, \ldots, C_q)$ with $C_i \in \mathcal{C}$, noting that the number of such vectors is at most $|\mathcal{C}|^q = 2^{o(n)}$.

Observe that every q-coloring can be identified by a vector $(I_1, \ldots, I_q)$ where each I_j is an independent set and $\bigcup I_j = V(G)$. Each of these vectors is "contained" in some "container

[1] Ha.

vector" $(C_1, \dots, C_q)$ with $C_j \in \mathcal{C}$ in the sense that $I_j \subseteq C_j$ for all j. Thus it's enough to count how many q-colorings each container vector contains.

A naive upper bound for the number of q-colorings contained in $(C_1, \dots, C_q)$ is roughly $2^{qn/2}$ since this is the number of ways to choose an independent set from each C_i. This bound is too weak, so we have to utilize the extra information that the I_j partition $V(G)$.

To this end, assume $V(G) = \{v_1, \dots, v_n\}$. Given $(C_1, \dots, C_q)$, let a_i be the number of containers C_j with $v_i \in C_j$. It's not difficult to see that the number of q-colorings contained in this vector is then at most $\prod a_i$, and by the AMGM inequality this is at most $(\sum a_i/n)^n = (\sum |C_j|/n)^n$. Each of the q containers has size at most roughly $n/2 + \sqrt{\frac{\log n}{d}}n$, so this gives the desired result. $\square$

17.2 A Randomized Sperner's Theorem

Throughout this subsection we fix an integer n and define $N := 2^n$ and $m = \binom{n}{\lfloor n/2 \rfloor}$. An *antichain* of $[n]$ is a subset $S \subseteq 2^{[n]}$ such that $A \not\subseteq B$ for any distinct $A, B \in S$. For example, $\binom{[n]}{k}$ is an antichain for all k. A famous result of Sperner's says that an antichain of $[n]$ has size at most m.

Our first goal is to count the number of antichains of $[n]$. To do this, we form a graph where independent sets correspond to antichains. Let G_N denote the graph whose vertex set is $2^{[n]}$ and where A, B are adjacent to each other if either $A \subseteq B$ or $B \subseteq A$. Analogous to Lemma 17.5, we need a supersaturation lemma for G_N which says that any collection of vertices that is much larger than m induces many edges. In particular, the following suffices.

Lemma 17.8 ([42, 7])**.** *If $C \subseteq 2^{[n]}$ has $|C| > (1+\epsilon)m$ with $0 < \epsilon \le 1/3$, then $e(G_N[C]) \ge \epsilon mn/2$.*

We won't prove this, but we will briefly comment on some intuition for the result. Intuitively, if you want to build a set of size $(1+\epsilon)m$ which induces few edges, then a good place to start is with the middle layer $\binom{[n]}{\lfloor n/2 \rfloor}$ since this is a maximum independent set. From there one could greedily choose ϵm sets which have as few neighbors as possible in this middle layer, and in particular choosing them allfrom $\binom{[n]}{\lfloor n/2 \rfloor + 1}$ gives a total of $(\lfloor n/2 \rfloor + 1) \cdot \epsilon m \ge \epsilon mn/2$ edges. Kleitman [42] proved that this is indeed the best construction, and the exact numerical computation[1] was done by Balogh, Mycroft, and Treglown [7].

With our supersaturation lemma in hand, we can easily prove the following result of Kleitman.

Theorem 17.9 ([43])**.** *The number of antichains of $[n]$ is $2^{m+o(m)}$.*

Proof. We obtain a set of containers $\mathcal{C}$ for G_N by applying lemma 17.1 to G_N with a parameter t to be determined later. Let ϵ be such that $(1+\epsilon)m = \max_{C \in \mathcal{C}} |C|$ and let C be a container

[1]The numbers in Lemma 17.8 are slightly different from those that appear in [7], but it's not difficult to refine their proof to give this result.

achieving this bound. By Lemma 17.8 we have[1]

$$\Delta(G_N[C]) \geq 2e(G_N[C])/|C| \geq \epsilon mn/(1+\epsilon)m \approx \epsilon n.$$

By assumption this quantity is at most t, or equivalently we roughly have $\max_{|C|\in\mathcal{C}} |C| \leq (1+t/n)m$. By (13) we have an upper bound of roughly

$$\binom{N}{N/t} 2^{(1+t/n)m} \approx 2^{\frac{N}{t}\log(t)+(1+t/n)m}.$$

This quantity is optimized when $N\log(t)/t \approx tm/n$. We have $m \approx N/\sqrt{n}$, so in total we want $t \approx n^{3/4}/\sqrt{\log n}$, and one can verify that this choice of t gives the desired bound. □

We next prove a random version of Sperner's theorem. The setup is as follows. Choose a subset $R_p \subseteq 2^{[n]}$ by including each set in R_p independently and with probability p. How large is the size of a largest antichain in R_p (in expectation)? In terms of our graph G_N, this is equivalent to computing $\mathbb{E}[\alpha(G_N[R_p])]$.

Let's consider some simple cases first. If $p = 1$, then $G_N[R_p] = G_N$ and we know its independence number is m. Somewhat more generally, we always have the lower bound $\mathbb{E}[\alpha(G_N[R_p])] \geq pm$ since this is the expected size of the set $\binom{[n]}{\lfloor n/2 \rfloor} \cap R_p$. However (as will often be the case), the behavior of $\mathbb{E}[\alpha(G_N[R_p])]$ changes considerably when p is very small.

For example, if $p^2 3^n \ll p2^n$, then asymptotically we have $\mathbb{E}[\alpha(G_N[R_p])] \sim |R_p|$ by a simple deletion argument. Even above the deletion threshold it is possible to improve on the trivial lower bound. Indeed, construct an independent set I by keeping each vertex in $\binom{[n]}{\lfloor n/2 \rfloor} \cap R_p$ together with all the vertices in $\binom{[n]}{\lfloor n/2 \rfloor - 1} \cap R_p$ which are not contained in any of the vertices of $\binom{[n]}{\lfloor n/2 \rfloor} \cap R_p$. The expected number of vertices we get from this first part is $p\binom{n}{\lfloor n/2 \rfloor}$, and from the second is[2] $p(1-p)^{n+1-\lfloor n/2 \rfloor}\binom{n}{\lfloor n/2 \rfloor - 1}$. In particular, if $p = c/n$ for a fixed constant c, then this asymptotically gives $(1+\epsilon)pm$ for some $\epsilon > 0$.

It turns out that for larger p we do have $\mathbb{E}[\alpha(G_N[R_p])] \sim pm$. More precisely we have the following due to Balogh, Mycroft, and Treglown [7].

Theorem 17.10 ([7]). *For any $\epsilon > 0$, there exists a constant c so that if $p > c/n$, then a.a.s. $\alpha(G_N[R_p]) \leq (1+\epsilon)pm$.*

Roughly speaking the approach we would like to use is as follows. Observe that $\alpha(G_N[R_p]) \geq k$ if and only if $G_N[R_p]$ contains an independent set of size k. The expected number of such sets in $G_N[R_p]$ is exactly $p^k \mathcal{I}_k(G_N)$, and if this quantity is small then we can conclude the result by Markov's inequality. Thus to solve this problem (and in general to solve extremal problems in random sets), we need to get effective upper bounds on $\mathcal{I}_k(G_N)$.

Unfortunately a naive application of Lemmas 17.1 and 17.8 together with (14) turns out to be too weak. The bottleneck here is the supersaturation result from Lemma 17.8. While the

[1]Implicitly this assumes $\epsilon \leq 1/3$. One can get around this by taking $C' \subseteq C$ a set of size exactly $4m/3$, but ultimately this computation is just to obtain intuition for what value t should be.

[2]To keep a vertex it has to be in R_p and all its neighbors in $\binom{[n]}{\lfloor n/2 \rfloor - 1}$ have to be out.

stated bound is essentially tight for $|C| = (1+\epsilon)m$ with $0 < \epsilon \le 1/3$ and n sufficiently large, the bound is not tight when e.g. $|C| = (2+\epsilon)m$. Intuitively in this case the extremal example should come from taking two middle layers together with ϵm of the layer right above these two. In particular, each of the ϵm vertices will have degree about n^2, so we expect around $\epsilon m n^2$ edges in $G_N[C]$. And indeed, this is the case.

Lemma 17.11 ([42, 7]). *If $C \subseteq 2^{[n]}$ has $|C| > (2+\epsilon)m$ with $0 < \epsilon \le 1/3$, then $e(G_N[C]) \ge \epsilon m n^2/9$.*

With this we can prove the main result.

Proof of Theorem 17.10. The key idea is to apply the container lemma twice using the two different levels of supersaturation from Lemmas 17.8 and 17.11. In particular, let $\mathcal{C}_1$ be a set of containers coming from Lemma 17.1 using G_N and some t_1, and for each $C_1 \in \mathcal{C}_1$, let $\mathcal{C}_2(C_1)$ be a set of containers coming from Lemma 17.3 using $G_N[C_1]$ and some t_2.

We first want to choose t_1 so that each $C_1 \in \mathcal{C}_1$ has size roughly $\alpha(G_N) = m$. Observe that if $|C_1| > 3m$, then by Lemma 17.11, every $C_1' \subseteq C_1$ of size $3m$ has

$$\Delta(G_N[C_1]) \ge \Delta(G_N[C_1']) \ge 2e(G_N[C_1'])/|C_1'| \ge n^2/81.$$

Thus if we take $t_1 = n^{1.99}$ in Lemma 17.1, we find for n sufficiently large that $|C_1| \le 3m$ for all $C_1 \in \mathcal{C}_1$.

We now want to choose t_2 so that each $C_2 \in \mathcal{C}_2(C_1)$ has size very close to m. Let $G' = G_N[C_1]$. If $|C_2| > (1+\gamma)m$ with $\gamma \le 1/3$ we find that

$$\Delta(G'[C_2]) \le 2e(G'[C_2])/|C_2| \le \gamma n/3.$$

With some foresight we take $t_2 = \epsilon n/12$ to guarantee that $|C_2| \le (1+\epsilon/4)m$ for all $C_2 \in \mathcal{C}_2(C_1)$ by Lemma 17.8.

Recall that we want to show with high probability no independent set I of size $(1+\epsilon)pm$ lies in the random set R_p. To this end, we note that we can identify each I with a pair (C_1, S_2) where

- $C_1 \in \mathcal{C}_1$ contains I,
- S_2 is the set $f(I)$ from Lemma 17.3, i.e. $|S_2| \le |C_1|/t_2$, $S_2 \subseteq I$, and S_2 determines a set $C_2 \supseteq I$ in $\mathcal{C}_2(C_1)$.

Given this pair, if $I \subseteq R_p$ has size at least $(1+\epsilon)m$, then (1) $S_2 \subseteq R_p$ since $S_2 \subseteq I$, and (2)

$$|R_p \cap (C_2 \setminus S_2)| \ge (1+\epsilon)pm - |S_2|,$$

since $C_2 \setminus S_2$ contains $I \setminus S_2$. Observe that for $p = c/n$ with $c \gg \epsilon^{-2}$ we have

$$|S_2| \le \frac{|C_1|}{t_1} \approx \frac{m}{\epsilon n} \le \frac{\epsilon p m}{2}.$$

With this in mind, we define $A(S_2)$ to be the event that $S_2 \subseteq R_p$ and $B(S_2)$ to be the event that $|R_p \cap (C_2 \setminus S_2)| \geq (1+\epsilon/2)pm$. Observe that $A(S_2)$ and $B(S_2)$ are independent events, so by a union bound over all pairs (C_1, S_2), we see that the probability that $\alpha(G_N[R_p]) \geq (1+\epsilon)pm$ is at most

$$\sum_{C_1 \in \mathcal{C}_1} \sum_{S_2 : |S_2| \leq |C_1|/t_1} \Pr[A(S_2)] \cdot \Pr[B(S_2)]. \tag{15}$$

We have $\Pr[A(S_2)] = p^{|S_2|}$. By a Chernoff bound it is not difficult to show that $\Pr[B(S_2)] \leq e^{-c'\epsilon^2 pm}$ for some $c' > 0$. Thus if we fix some C_1 in the first term of the sum we get

$$\sum_{s \leq |C_1|/t_2} \binom{|C_1|}{s} p^s \cdot e^{-c'\epsilon^2 pm} \leq \sum_{s \leq |C_1|/t_2} (ep|C_1|/s)^s \cdot e^{-c'\epsilon^2 pm} \approx (p \cdot \epsilon n/12)^{12|C_1|/\epsilon n} \cdot e^{-c'\epsilon^2 pm} \leq (\epsilon c)^{36m/\epsilon n} \cdot e^{-c'\epsilon^2 cm/n},$$

where the approximation only looked at the term with $s = |C_1|/t_2$. Note that for $c \gg \epsilon^{-4}$ the second term dominates, so this bound is roughly

$$e^{-c'\epsilon^2 cm/n}. \tag{16}$$

Note that this is the critical place where we used the two applications of the container lemma: if we only applied the container lemma once to $V(G_N)$ instead of using C_1, then the first term here would be roughly $e^{N/\epsilon n}$ instead of $e^{m/\epsilon n}$, which would dominate the expression.

Returning to (15), we sum the bound of (16) for each element in C_1, which multiplies (16) by

$$\binom{N}{\leq N/t_1} \approx e^{N \log(t_1)/t_1} \approx e^{mn^{-1.49} \log(n)}.$$

This is much smaller than $e^{-c'\epsilon^2 cm/n}$, so the probability in (15) tends to 0 as n tends towards infinity as desired. □

We note that one can get almost as strong a result if one only uses Lemma 17.1, i.e. if one doesn't use the more refined Lemma 17.3. Indeed, the main consequence of using this refined lemma was the extra term $\Pr[A(S_2)] = p^s$ appearing in (15). If one omits this turn, then the same proof will go through provided $p = c \log n/n$. This is a common phenomenon in containers: if you don't use the fact you have certificates $S \subseteq I$, then you'll end up with a bound which is worse by a log factor.

17.3 Counting Sidon Sets

This is a neat example where you use roughly $\log n$ iterated supersaturation lemmas to get the right result. I may write this at some point, but in any case it will be nearly identical to Balogh's notes [5].

18 A Proof of an r-Uniform Container Lemma

In the previous section, we saw how Lemma 17.1 allowed us to effectively count the number of independent sets in "sufficiently nice" graphs. In this section we present a proof of a hypergraph container lemma which applies to "sufficiently nice" hypergraphs, but two things should be noted.

The first is that there are many different variants of hypergraph container lemmas, though most of them are quite similar and broadly speaking apply only to hypergraphs with small codegrees.

The second is that, quite frankly, one doesn't need to know the proof of the container lemma to use it or its variants. As such, the reader may just want to glance at the definition below, and then skip over to latter sections to see some nice applications before jumping back over here whenever they want to see the full proof.

18.1 An Informal Discussion

The hypergraph container lemma we prove comes from [**?**] (though we deviate somewhat from their notation). We'll formally state this as Theorem 18.5 below, but roughly our goal will be to prove the following. Recall that $\Delta_\ell(H)$ denotes the maximum ℓ-degree of H, i.e. the maximum number of edges containing a given set of ℓ vertices.

Proposition 18.1 (Informal). *If H is an r-graph such that for all $1 \le \ell \le r$ we have*

$$\Delta_\ell(H) \ll q^{\ell-1}\frac{e(H)}{v(H)},$$

then there exists $\delta > 0$, $\mathcal{S} \subseteq \binom{V(H)}{\ll q \cdot v(H)}$, and functions $f : \mathcal{S} \to \binom{V(H)}{\le (1-\delta)v(H)}$ and $g : \mathcal{I}(H) \to \mathcal{S}$ such that

$$g(I) \subseteq I \subseteq f(g(I)) \cup g(I).$$

In other words, if H has small codegrees, then one can find a set of small certificates $\mathcal{S}$ which are each associated with a container $f(S)$ which is of size at most $(1-\delta)v(H)$.

The proof of Proposition 18.1 will in essence be a proof by induction on the uniformity r, and the inductive step of the proof uses an algorithm which is similar to the one used in Lemma 17.1.

Definition 2 (Informal). The *Scythe Algorithm* takes as input a pair (H_{k+1}, I) with H_{k+1} a $(k+1)$-uniform hypergraph and $I \subseteq V(H_{k+1})$ an independent set. It then outputs a triple (H_k, A_k, S_k) with H_k a k-uniform hypergraph such that $I \subseteq V(H_k)$ is an independent set with $I \subseteq A_k \cup S_k$, and S_k is a small set which uniquely determines H_k and A_k given H_{k+1}. Moreover, if H_{k+1} is "nice", then either H_k will be "nice" or A_k will be small.

Given such an algorithm, we can start with any "nice" r-uniform hypergraph H_r and independent set I. We then repeatedly apply this algorithm until we get some H_k which is not "nice", at which point $\bigcup_{i \ge k} S_i$ is a small certificate which determines a small container $A_k \cup \bigcup_{i \ge k} S_i$ for I.

While it's not a priori clear what the "nice" conditions should be, they should in particular guarantee that H_{k+1} has few independent sets, as otherwise there's no hope of this method being effective. In particular, a reasonable set of conditions is to enforce that H_{k+1} has many edges and relatively low codegrees, and this will ultimately be the conditions that we use.

Now that we know what we want our algorithm to do, how should it actually work in practice? Perhaps the most naive approach is to do what we did in Lemma 17.1, where we iteratively select the vertex u of $I \cap H_{k+1}$ which has the largest degree and then adds this to S_k. Once we identify such a u, we know that I does not contain any k-set of the form $e \setminus \{u\}$ for any $e \ni u$ which is an edge in H_{k+1}, so it is natural to make all of these k-sets edges of H_k. If we do this repeatedly, then A_k will be relatively small and I will be an independent set of H_k. Moreover, if we enforce from the start that we'll run this procedure at most s times, then we will have $|S_k| \leq s$, giving a small certificate.

Unfortunately we have to be more careful than this. Namely, we need to ensure that H_k is "nice", and in particular that it has small codegrees. As it currently stands this might not work out, e.g. there may be some ℓ-set T which is in many edges containing vertices of S_k. To get around this, we define $D_\ell(H_k, \Delta)$ to be the set of "dangerous" ℓ-sets of $V(H_k)$ which have degree at least $\Delta/2$, where we think of Δ as being the maximum ℓ-degree we want H_k to have (which is analogous to the t parameter used in the algorithm of Lemma 17.1).

We now adjust our naive algorithm by making it so that whenever a set T gets added to $D_\ell(H_k, \Delta)$, we delete from H_{k+1} all of the edges that contain T. This ensures that T never passes over the enforced codegree threshold. With this it turns out that our algorithm will succeed.

18.2 A Formal Algorithm

Motivated by our discussion in the previous section, we make the following definitions. For any hypergraph H', we define the *max-degree order* on $V(H')$ as follows. Fix an arbitrary ordering of $V(H')$. For each integer j, recursively define u_j to be the maximum-degree vertex in the hypergraph $H'[V(H') \setminus \{u_1, \ldots, u_{j-1}\}]$ with ties broken based on the ordering of $V(H')$. The max-degree order is then the ordering $u_1, u_2, \ldots$, and for all j we define $W_{H'}(u_j) = \{u_1, \ldots, u_j\}$.

For any k-uniform hypergraph H', integer $\ell \leq k$, and real number Δ, we define

$$D_\ell(H', \Delta) = \left\{ T \in \binom{V(H')}{\ell} : \deg_{H'}(T) \geq \frac{1}{2}\Delta \right\}.$$

Definition 3. The Scythe Algorithm is defined as follows. It takes as input a $(k+1)$-uniform hypergraph H_{k+1} with $k \geq 1$, an independent set $I \subseteq V(H_{k+1})$, and parameters $s, \Delta_1^k, \ldots, \Delta_k^k$.

At the start of the algorithm, we set $H_{k+1}^{(0)} = H_{k+1}$, $S_k^{(0)} = \emptyset$, and we let $H_k^{(0)}$ be the empty hypergraph on $V(H_{k+1})$. For $j = 0, \ldots, s-1$, the algorithm proceeds as follows:

Step 1: If $I \cap V(H_{k+1}^{(j)}) = \emptyset$, then set $H_k = H_k^{(0)}$, $A_k = \emptyset$, $S_k = S_k^{(j)}$. If this happens, stop the algorithm and output (H_k, A_k, S_k).

Step 2: Let u_j be the vertex of $I \cap V(H_{k+1}^{(j)})$ which is first according to the max-degree ordering of $H_{k+1}^{(j)}$. Set $S_k^{(j+1)} = S_k^{(j)} \cup \{u_j\}$.

Step 3: Let $H_k^{(j+1)}$ by the hypergraph on $V(H)$ defined by

$$H_k^{(j+1)} \cup \{e \setminus \{u_j\} : e \in H_{k+1}^{(j)},\ u_j \in e\}.$$

Step 4: Let $H_{k+1}^{(j+1)}$ be the hypergraph on $V(H_{k+1}^{(j)}) \setminus W_{H_{k+1}^{(j)}}(u_j)$ with

$$H_{k+1}^{(j+1)} = \{e \in H_{k+1}^{(j)} : e \cap W_{H_{k+1}^{(j)}}(u_j) = \emptyset \text{ and } T \not\subseteq e \text{ for all } T \in \bigcup_{\ell=1}^{k} D_\ell(H_k^{(j+1)}, \Delta_\ell^k)\}.$$

After running through the above procedure, set $H_k = H_k^{(s)},\ A_k = V(H_{k+1}^{(s)})$, and $S_k = S_k^{(s)}$. Output (H_k, A_k, S_k). ■

We emphasize that this algorithm allows the hypergraphs H_{k+1} and H_k to have repeated edges.

We now analyze this algorithm through a series of lemmas. We omit many of the proofs since most are either straightforward or analogous to what was done in Lemma 17.1.

Lemma 18.2. *Assume one runs the Scythe Algorithm with parameters $s, \Delta_1^k, \ldots, \Delta_k^k$ on inputs $(H_{k+1}, I), (H_{k+1}, I')$ and that the algorithm outputs $(H_k, A_k, S_k), (H_k', A_k', S_k')$, respectively. If $S_k \subseteq I'$ and $S_k' \subseteq I$, then $(H_k, A_k, S_k) = (H_k', A_k', S_k')$.*

For the rest of this subsection we will assume that we have run the Scythe Algorithm with parameters $s, \Delta_1^k, \ldots, \Delta_k^k$ on (H_{k+1}, I) which outputs some (H_k, A_k, S_k). We observe the following basic properties.

Lemma 18.3. *The following hold:*

- *I is an independent set of H_k,*
- *$S_k \subseteq I \subseteq A_k \cup S_k$,*
- *Both H_k and A_k are determined by H_{k+1} and S_k,*
- *$|S_k| \le s$,*
- *For all $\ell \le k$ we have $\Delta_\ell(H_k) \le \frac{1}{2}\Delta_\ell^k + \Delta_{\ell+1}(H_{k+1})$.*

We now turn to the main lemma of this subsection, which roughly says that if H_{k+1} is "nice," then either H_k will also be "nice" or A_k will be small.

Lemma 18.4. *Either*

$$e(H_k) \ge \min\left\{\frac{s}{v(H)}, \frac{\Delta_1^k}{\Delta_1(H_{k+1})}, \ldots, \frac{\Delta_k^k}{\Delta_k(H_{k+1})}\right\} \cdot \frac{e(H_{k+1})}{(k+1)2^{k+2}},$$

or

$$|A_k| \leq v(H_{k+1}) - \frac{e(H_{k+1})}{4 \cdot \Delta_1(H_{k+1})}.$$

Note that this lemma will be most "effective" when each ratio of the minimum is roughly the same. And indeed, we will end up choosing our parameters so that these ratios are all at least $s/v(H)$.

Proof. If the algorithm ever stops at Step 1, then $|A_k| = 0$ and there is nothing to prove, so we can assume that Steps 2 through 4 are completed a total of s times. We observe that

$$|A_k| = v(H_{k+1}^{(s)}) = v(H_{k+1}) - \sum_{j=0}^{s-1} |W_j(u_j)|,$$

where for ease of notation we let $W_j = W_{H_{k+1}^{(j)}}$. Thus we can assume

$$\sum_{j=0}^{s-1} |W_j(u_j)| < \frac{e(H_{k+1})}{4\Delta_1(H_{k+1})}. \tag{17}$$

By construction, for all j we have

$$e(H_k^{(j+1)}) - e(H_k^{(j)}) = \deg_{H_{k+1}^{(j)}}(u_j).$$

Because u_j is the largest element of $I \cap V(H_{k+1}^{(j)})$ in the max-degree order, the degree of u_j is at least as large as the average degree of the subgraph of $H_{k+1}^{(j)}$ after deleting $W_j(u_j) \setminus \{u_j\}$, and again by definition of the max-degree order, this is at least as large as the average degree of $H_{k+1}^{(j)}$ (which has at most $v(H)$ vertices). In total then we find

$$e(H_k) = \sum_{j=0}^{s-1} e(H_k^{(j+1)}) - e(H_k^{(j)}) \geq \sum_{j=0}^{s-1} \frac{(k+1)e(H_{k+1}^{(j+1)})}{v(H)}.$$

If we have $(k+1)e(H_{k+1}^{(j+1)}) \geq e(H_{k+1})$ for all j, then the above sum is at least $(s/v(H)) \cdot e(H_{k+1})$, giving the desired result. Thus we can assume this fails for some j, and this implies

$$e(H_{k+1}^{(s)}) \leq e(H_{k+1}^{(j+1)}) < \frac{e(H_{k+1})}{k+1}. \tag{18}$$

This means that many edges of H_{k+1} were deleted in Step 4 of the algorithm. We claim that this implies that one of the sets $D_\ell(H_k, \Delta_\ell^k)$ is large. Indeed, observe that

$$e(H_{k+1}^{(j)}) - e(H_{k+1}^{(j+1)}) \leq |W_j(u_j)| \cdot \Delta_1(H_k) + \sum_\ell |D_\ell(H_k^{(j+1)}, \Delta_\ell^k) \setminus D_\ell(H_k^{(j)}, \Delta_\ell^k)| \cdot \Delta_\ell(H_{k+1}),$$

since edges are either deleted by deleting vertices in $W_j(u_j)$ or by deleting ℓ-sets which are in D_ℓ for $H_k^{(j+1)}$ but not $H_k^{(j)}$. Summing this over all j gives

$$\begin{aligned} e(H_{k+1}) - e(H_{k+1}^{(s)}) &\leq \sum_j |W_j(u_j)| \cdot \Delta_1(H_{k+1}) + \sum_\ell |D_\ell(H_k, \Delta_\ell^k)| \cdot \Delta_\ell(H_{k+1}) \\ &< \frac{k \cdot e(H_{k+1})}{2(k+1)} + \sum_\ell |D_\ell(H_k, \Delta_\ell^k)| \cdot \Delta_\ell(H_{k+1}), \end{aligned}$$

where this last step used (17) and $k \geq 1$. Using (18) shows that

$$\frac{k \cdot e(H_{k+1})}{2(k+1)} \geq \sum_{\ell} |D_\ell(H_k, \Delta_\ell^k)| \cdot \Delta_\ell(H_{k+1}),$$

so for some ℓ we must have

$$|D_\ell(H_k, \Delta_\ell^k)| \geq \frac{e(H_{k+1})}{2(k+1)\Delta_\ell(H_{k+1})}.$$

With ℓ as above, the handshaking lemma and definition of D_ℓ implies

$$e(H_k) = \binom{k}{\ell}^{-1} \sum_{T \in \binom{V(H_k)}{\ell}} \deg_{H_k}(T) \geq \binom{k}{\ell}^{-1} \cdot |D_\ell(H_k, \Delta_\ell^k)| \cdot \frac{1}{2}\Delta_\ell^k \geq \frac{e(H_{k+1}) \cdot \Delta_\ell^k}{(k+1)2^{k+2}\Delta_\ell(H_{k+1})},$$

giving the desired result. □

18.3 Proof of The Formal Result

We are now ready to prove the following formal statement.

Theorem 18.5 ([6]). *For every integer $r \geq 2$ and $c \geq 1$, there exists $\delta > 0$ such that the following holds. Let $q \in (0,1)$ and suppose H is an r-uniform hypergraph such that for every $1 \leq \ell \leq r$ we have*

$$\Delta_\ell(H) \leq cq^{\ell-1} \cdot \frac{e(H)}{v(H)}.$$

Then there exists $\mathcal{S} \subseteq \binom{V(H)}{\leq (r-1)q \cdot v(H)}$ and functions $f : \mathcal{S} \to \binom{V(H)}{\leq (1-\delta)v(H)}$ and $g : \mathcal{I}(H) \to \mathcal{S}$ such that for every $I \in \mathcal{I}(H)$ we have

$$g(I) \subseteq I \subseteq f(g(I)) \cup g(I).$$

Moreover, $S \cap f(S) = \emptyset$ for all $S \in \mathcal{S}$, and if $I, I' \in \mathcal{I}(H)$ satisfy $g(I) \subseteq I'$, $g(I') \subseteq I$, then $g(I) = g(I')$.

Proof. For all $\ell \leq r$ let $\Delta_\ell^r := \Delta_\ell(H)$, and inductively define

$$\Delta_\ell^k := \max\{2 \cdot \Delta_{\ell+1}^{k+1}, q \cdot \Delta_\ell^{k+1}\}.$$

The following is straightforward to prove given the hypothesis of the theorem.

Claim 18.6. *For all $k < r$ we have $\Delta_1^{k+1} \leq C2^r q^{r-k-1} \frac{e(H)}{v(H)}$.*

For each $I \in \mathcal{I}(H)$, iteratively run through the Scythe Algorithm with the parameters above and $s = q \cdot v(H)$, starting with $H_r = H$. Let $(H_{r-1}, A_{r-1}, S_{r-1}), \ldots, (H_1, A_1, S_1)$ denote the outputs of this algorithm.

It is straightforward to show that $\Delta_\ell(H_k) \leq \Delta_\ell^k$ for all ℓ, k by using induction and Lemma 18.3. For $k < r$ we define $c_k = (Cr2^{r+1})^{k-r}$. Let K be the smallest integer such that $|A_K| \leq$

$(1-c_K)v(H)$, and if no such $K \geq 1$ exists we set $K = 0$. It is straightforward to prove that $e(H_k) \geq c_k q^{k-r} e(H)$ for all $k > K$ by using induction, Lemma 18.4, and Claim 18.6.

Let $\delta := c_1$. Before we define our functions, for technical reasons it will be convenient to first define a function $f^* : \mathcal{I}(H) \to \binom{V(H)}{\leq (1-\delta)v(H)}$ before defining f. Pick some I and let the hypergraphs H_k and integer K be as defined above. Observe that if $K \geq 1$, then $|A_K| \leq (1-\delta)v(H)$, so in this case we will set

$$g(I) = \bigcup_{k \geq K} S_k, \quad f^*(I) = A_K.$$

If $K = 0$, then we set

$$g(I) = \bigcup_{k \geq 1} S_k, \quad f^*(I) = \{v \in V(H_1) : \{v\} \notin H_1\}.$$

Note that in this case $|f^*(I)| = v(H) - e(H_1)$, which is at most $(1-\delta)v(H)$ by our observations above. Lastly, we define $\mathcal{S} = \{g(I) : I \in \mathcal{I}(H)\}$ and $f(S) = f^*(I)$ for any $I \in g^{-1}(S)$. The fact that f is well defined is implied by the following claim, which itself follows from Lemma 18.3.

Claim 18.7. *If $I, I' \in \mathcal{I}(H)$ with $g(I) \subseteq I'$ and $g(I') \subseteq I$, then $g(I) = g(I')$ and $f^*(I) = f^*(I')$.*

The fact that these definitions give the desired result, except possibly the condition $S \cap f(S) = \emptyset$, can be checked by using the properties from Lemma 18.3. This last condition can be established by taking $f'(S) := f(S) \setminus S$ if needed. $\square$

The following weaker version of Theorem 18.5 is often good enough for most applications and is conceptually simpler.

Corollary 18.8. *For every integer $r \geq 2$ and $c \geq 1$, there exists $\delta > 0$ such that the following holds. Let $q \in (0,1)$ and suppose H is an r-uniform hypergraph such that for every $1 \leq \ell \leq r$ we have*

$$\Delta_\ell(H) \leq c q^{\ell-1} \cdot \frac{e(H)}{v(H)}.$$

Then there exists a collection of sets $\mathcal{C}$ such that every independent set of H is a subset of some $C \in \mathcal{C}$, and moreover, $|C| \leq (1-\delta)v(H)$ for all $C \in \mathcal{C}$ and $|\mathcal{C}| \leq \binom{v(H)}{\leq (r-1)q \cdot v(H)}$.

Proof. In the notation of Theorem 18.5, we let $\mathcal{C} = \{f(g(I)) \cup g(I) : I \in \mathcal{I}(H)\}$. $\square$

On its own, Corollary 18.8 (and even Theorem 18.5) isn't terribly useful since the containers it generates are rather large, and in practice one needs to reapply this lemma to each $C \in \mathcal{C}$ which is large, and to keep repeating this argument until the contains are sufficiently small. Because δ depends only on r and C, this only needs to be done a constant number of times. However, to reapply the lemma, each large $C \in \mathcal{C}$ must satisfy essentially the same hypothesis as H. While a generic hypergraph will fail to have this property, many nice hypergraphs will.

19 Hypergraph Containers and Triangle-Free Graphs

Let us restate our weak container theorem Corollary 18.8 for the special case of 3-uniform hypergraphs.

Theorem 19.1 ([6])**.** *For every $c \geq 1$, there exists $\delta > 0$ such that the following holds. Let $q \in (0,1)$ and suppose H is a 3-uniform hypergraph such that*

$$\Delta_1(H) \leq c\frac{e(H)}{v(H)},$$

$$\Delta_2(H) \leq cq\frac{e(H)}{v(H)},$$

$$\Delta_3(H) \leq cq^2\frac{e(H)}{v(H)}.$$

Then there exists a collection of sets $\mathcal{C}$ such that every independent set of H is a subset of some $C \in \mathcal{C}$, and moreover, $|C| \leq (1-\delta)v(H)$ for all $C \in \mathcal{C}$ and $|\mathcal{C}| \leq \binom{v(H)}{\leq 2q\cdot v(H)}$.

Note that for simple hypergraphs we have $\Delta_3(H) = 1$, so this last bound is equivalent to lower bounding the average degree by $c^{-1}q^{-2}$. One important consequence of Theorem 19.1 is the following.

Theorem 19.2. *For all $n, \epsilon > 0$, there exists a collection of n-vertex graphs $\mathcal{C}$ such that*

(a) Every triangle-free graph $G \subseteq K_n$ is a subgraph of some $C \in \mathcal{C}$,

(b) Every $C \in \mathcal{C}$ has less than ϵn^3 triangles, and

(c) We have $|\mathcal{C}| = n^{O_\epsilon(n^{3/2})}$.

That is, there exists a small set of nearly triangle-free graphs which contains every triangle-free graph.

Proof. Start with $\mathcal{C} = \{K_n\}$, and note that $\mathcal{C}$ trivially satisfies (a). Iteratively proceed as follows. If every $C \in \mathcal{C}$ has less than ϵn^3 triangles then output the current collection $\mathcal{C}$. Otherwise, let $C \in \mathcal{C}$ be such that it contains at least ϵn^3 triangles. Form a 3-graph H with vertex set $E(C)$ where three edges of C form a hyperedge in H if they form a triangle. Note that $e(H) \geq \epsilon n^3$ and $v(H) = e(C) \leq n^2$. Every edge is contained in at most n triangles, so $\Delta_1(H) \leq n \leq \epsilon^{-1}\frac{e(H)}{v(H)}$. We also have $\Delta_2(H) = \Delta_3(H) = 1 \leq \epsilon^{-1}(n^{-1/2})^2\frac{e(H)}{v(H)}$. With this we see that we can apply Theorem 19.1 with $q = n^{-1/2}$ and $c = \epsilon^{-1}$. This gives a collection of containers $\mathcal{C}'$ for C, i.e. subgraphs $C' \subseteq C$ such that every triangle-free subgraph of C is contained in some $C' \in \mathcal{C}$. Remove C from $\mathcal{C}$ and add every $C' \in \mathcal{C}'$ to $\mathcal{C}$. Repeat this process.

Let $\mathcal{C}$ be the final collection that this algorithm produces. It is straightforward to show that (a) holds inductively, and (b) holds by construction. To show that the final collection is small, first note that each time we apply the container lemma, the number of new graphs we create is at most $\binom{v(H)}{\le 2n^{-1/2}v(H)} = n^{O(n^{3/2})}$. Second, observe that each time we apply the container lemma to C, the graphs in $\mathcal{C}'$ have at most $(1-\delta)e(C)$ edges, where δ depends only on ϵ. Because we only iterate on C which have at least ϵn^2 edges (since they need at least ϵn^3 triangles), we iteratively apply the lemma at most some bounded number of times $b = b(\epsilon)$ to reach any element in the final collection $\mathcal{C}$. Thus the total number of containers we create is $\left(n^{O(n^{3/2})}\right)^b = n^{O_\epsilon(n^{3/2})}$ as desired. □

For Theorem 19.2 to be useful, we need to get a handle on graphs with at most ϵn^3 triangles. As is typical with containers, this will come from a supersaturation lemma.

Lemma 19.3. *For every $\delta > 0$ there exists an $\epsilon > 0$ such that if G is an n-vertex graph with $e(G) \ge (\frac{1}{2}+\delta)\binom{n}{2}$, then G contains at least ϵn^3 triangles.*

I'm not crazy about the ordering of δ, ϵ but I admit the final thing should be about ϵ...well actually in a lot of the applications it kind of makes more sense to do it the other way. Maybe do the K_r version in general depending on what I need later on.

Proof. TODO □

With this we can prove the following counting result.

Theorem 19.4. *The number of n-vertex triangle-free graphs is equal to*

$$2^{(1+o(1))n^2/4}.$$

Proof. The lower bound comes from considering all of the subgraphs of $K_{n/2,n/2}$. For the upper bound, fix some $\delta > 0$ and let ϵ be as in Lemma 19.3. Let $\mathcal{C}$ be the containers guaranteed by Theorem 19.2 with parameter ϵ. Because every triangle-free graph is a subgraph of some $C \in \mathcal{C}$, the number of triangle-free graphs is at most

$$\sum_{C\in\mathcal{C}} 2^{|C|} \le n^{O(n^{3/2})} \cdot 2^{\max_{C\in\mathcal{C}} e(C)},$$

Since each $C \in \mathcal{C}$ has less than ϵn^3 and at triangles, Lemma 19.3 implies $e(C) \le (\frac{1}{2}+\delta)\binom{n}{2}$ for all $C \in \mathcal{C}$. In total we get an upper bound of

$$2^{(\frac{1}{2}+\delta)\binom{n}{2}+O(n^{3/2}\log n)},$$

and letting δ tend towards 0 gives the result. □

In addition to proving counting results, containers also provide a useful framework for proving probabilistic analogs of classical extremal results. For example, the following result can be viewed as a random version of Mantel's theorem.

To this end, given two graphs G, F, we let $\mathrm{ex}(G, F)$ denote the largest F-free subgraph of G. For example, $\mathrm{ex}(K_n, F) = \mathrm{ex}(n, F)$. The following result can be viewed

Theorem 19.5. *Define* $\mathrm{ex}(G_{n,p}, K_3)$ *to be the largest triangle-free subgraph of* $G_{n,p}$*. We have* $\mathrm{ex}(G_{n,p}, K_3) = (1 + o(1))pn^2/4$ *whp provided* $p \gg n^{-1/2} \log n$*.*

Proof. The lower bound follows by considering $G_{n,p} \cap K_{n/2,n/2}$, which is always triangle-free and which has $(1 + o(1))pn^2/4$ edges whp. For the upper bound, fix $\delta > 0$, and let $\epsilon > 0$ be as in Lemma 19.3. Let $\mathcal{C}$ be the set of containers given by Theorem 19.2 with parameter ϵ, and as before we have $e(C) \le (1/2+\delta)\binom{n}{2}$ for all $C \in \mathcal{C}$. Because every triangle-free graph is contained in some $C \in \mathcal{C}$, in order to have $\mathrm{ex}(G_{n,p}, K_3) \ge (1 + 4\delta)pn^2/4$, there must exist some $C \in \mathcal{C}$ such that $|G_{n,p} \cap C| \ge (1+4\delta)pn^2/4$. Let E_C be the event that this bound holds. Observe that $|G_{n,p} \cap C|$ is a binomial random variable with probability p and at most $(1+2\delta)n^2/4$ trials. By the Chernoff bound, we find $\Pr[E_C] \le e^{-O_\delta(pn^2)}$. In total then, we have

$$\Pr[\mathrm{ex}(G_{n,p}, K_3) \ge (1 + 4\delta)pn^2/4] \le \Pr\left[\bigcup_{C\in\mathcal{C}} E_C\right] \le n^{O_\delta(n^{3/2})} \cdot e^{-O_\delta(pn^2)} \to 0,$$

with this last step holding by hypothesis on p. We conclude the reuslt by taking δ arbitrarily close to 0. $\square$

We note that for $p \ll n^{-1/2}$, a simple deletion argument shows that for $p \ll n^{-1/2}$ there exist triangle-free subgraphs with $(1 + o(1))p\binom{n}{2}$ edges, and this is certainly best possible since $G_{n,p}$ has at most this many edges asymptotically. Thus the bound for p in Theorem 19.5 is almost optimal. In fact, we can obtain the optimal bound in this theorem by using the strong container theorem Theorem 18.5, which in the case of 3-graphs can be written as follows.

Theorem 19.6. *For every* $c \ge 1$*, there exists* $\delta > 0$ *such that the following holds. Let* $q \in (0, 1)$ *and suppose* H *is a* 3*-uniform hypergraph such that*

$$\Delta_1(H) \le c\frac{e(H)}{v(H)},$$

$$\Delta_2(H) \le cq\frac{e(H)}{v(H)},$$

$$\Delta_3(H) \le cq^2\frac{e(H)}{v(H)}.$$

Then there exists $\mathcal{S} \subseteq \binom{V(H)}{\le 2q\cdot v(H)}$ *and functions* $f : \mathcal{S} \to \binom{V(H)}{\le(1-\delta)v(H)}$ *and* $g : \mathcal{I}(H) \to \mathcal{S}$ *such that for every* $I \in \mathcal{I}(H)$ *we have*

$$g(I) \subseteq I \subseteq f(g(I)) \cup g(I).$$

Moreover, $S \cap f(S) = \emptyset$ *for all* $S \in \mathcal{S}$*, and if* $I, I' \in \mathcal{I}(H)$ *satisfy* $g(I) \subseteq I'$*,* $g(I') \subseteq I$*, then* $g(I) = g(I')$*.*

This allows us to construct the following "strong" set of containers for triangle-free graphs.

Theorem 19.7. *Let $\mathcal{G}_n, \mathcal{T}_n$ denote the set of all n-vertex graphs and all n-vertex triangle-free graphs, respectively. For all $n, \epsilon > 0$, there exists a set of graphs $\mathcal{S}$ with at most $O_\epsilon(n^{3/2})$ edges, as well as functions $f : \mathcal{S} \to \mathcal{G}_n$ and $g : \mathcal{T}_n \to \mathcal{S}$ such that for every $G \in \mathcal{T}_n$, we have*

$$g(G) \subseteq G \subseteq f(g(G)) \cup g(G),$$

and such that $f(S)$ has less than ϵn^3 triangles for all $S \in \mathcal{S}$.

Proof. We start with $\mathcal{S}$ consisting only of the empty graph and define $g(G) = \emptyset$ and $f(\emptyset) = K_n$. Iteratively assume we have constructed some $\mathcal{S}, f, g$ satisfying all of the conditions except possibly that each $S \in \mathcal{S}$ has at most $O_\epsilon(n^{3/2})$ edges and that $f(S)$ has less than ϵn^3 triangles (which holds for our initial step). If $f(S)$ has less than ϵn^3 triangles for all $S \in \mathcal{S}$ then we end the procedure. Otherwise, let S be such that $C = f(S)$ has at least ϵn^3 triangles. By repeating our computations from the proof of Theorem 19.2, we see that we can apply Theorem 19.6 to the 3-graph H encoding triangles of C, and we let $\mathcal{S}_C, f_C, g_C$ be the output of this theorem.

Claim 19.8. *Let $\mathcal{S}' := (\mathcal{S} \setminus \{S\}) \cup \{S_C \cup S : S_C \in \mathcal{S}_C\}$, define $g'(G) = g(G)$ if $g(G) \neq S$ and $g'(G) = g_C(G - S)$ otherwise, and define $f'(S') = f(S')$ if $S' \in \mathcal{S} \setminus \{S\}$ and $f'(S') = f_C(S' - S)$ otherwise. These maps are well defined and satisfy the conditions of the theorem except possibly that each $S \in \mathcal{S}'$ has at most $O_\epsilon(n^{3/2})$ edges and that $f'(S)$ has less than ϵn^3 triangles.*

Proof. First observe that because $C \cap S = \emptyset$, each element of $\mathcal{S}_C$ (which is a subgraph of C) is disjoint from S. This implies that all of the elements $S_C \cup S$ for $S_C \in \mathcal{S}_C$ are distinct. Moreover, none of these elements are equal to any element of $\mathcal{S} \setminus \{S\}$. Indeed, if $S_C \cup S = S' \in \mathcal{S}$, then $\mathcal{S}$ would contain two elements with $S \subsetneq S'$. The last condition of Theorem 19.6 then implies that we must have $S = S'$. This all implies that g', f' are well defined maps, and it is not difficult to check that they inherit all of the other desired properties. $\square$

With this we can keep applying Theorem 19.6 until we get $\mathcal{S}, f, g$ which satisfies all of the conditions except possibly that $e(S)$ is small. As in the proof of Theorem 19.2, one can check that each $S \in \mathcal{S}$ is obtained by applying Theorem 19.6 at most $O_\epsilon(1)$ times, and each time its applied at most $O(n^{3/2})$ edges get added to S. With this we can conclude the result. $\square$

We note that there exists a somewhat stronger version of Theorem 19.6 (and more generally Theorem 18.5) which allows one to prove the previous result with less work. However, the theorem statement is somewhat more complicated conceptually (involving things called $(\mathcal{F}, \epsilon)$-dense families), so for this exposition we have opted to use the simpler version. In any case, with this enhanced version of Theorem 19.2, we can improve upon our threshold for the random Mantel theorem by dropping a logarithmic term.

Theorem 19.9. *Define* $\mathrm{ex}(G_{n,p}, K_3)$ *to be the largest triangle-free subgraph of $G_{n,p}$. We have* $\mathrm{ex}(G_{n,p}, K_3) = (1 + o(1))pn^2/4$ *whp provided $p \gg n^{-1/2}$.*

Proof. The lower bound follows by considering $G_{n,p} \cap K_{n/2,n/2}$, which is always triangle-free and which has $(1 + o(1))pn^2/4$ edges whp. For the upper bound, fix $\delta > 0$, and let $\epsilon > 0$ be as in Lemma 19.3. Let $\mathcal{S}, f, g$ be as in Theorem 19.7. Note that each $f(S)$ has at most $(1/4+2\delta)n^2$ edges by Lemma 19.3. For each $S \in \mathcal{S}$, let E_S be the event that $S \subseteq G_{n,p}$ and that

$|f(S) \cap G_{n,p}| \geq (1+4\delta)pn^2/4$. Note that in order to have $\text{ex}(G_{n,p}, K_3) \geq (1+4\delta)pn^2 + O_\epsilon(n^{3/2})$, some E_S event must occur, and moreover that $\Pr[E_S] = p^{|S|} \cdot e^{-O_\delta(pn^2)}$. With this we have

$$\Pr[\text{ex}(G_{n,p}, K_3) \geq (1+4\delta)pn^2/4 + O_\epsilon(n^{3/2})] \leq \Pr\left[\bigcup_{S \in \mathcal{S}} E_S\right] \leq \sum_{s=0}^{O_\epsilon(n^{3/2})} \sum_{S \in \mathcal{S}: |S|=s} p^s e^{-O_\delta(pn^2)}.$$

As the number of $S \in \mathcal{S}$ with $|S| = s$ is trivially at most $\binom{n^2}{s} \leq (en^2/s)^s$, we find that the above is at most

$$\sum_{s=0}^{O_\epsilon(n^{3/2})} (epn^2/s)^s e^{-O_\delta(pn^2)}.$$

One can check that the function $(epn^2/s)^s$ is increasing for $s \leq pn^2$. Since we know $s \leq C_\epsilon n^{3/2}$ for some suitable C_ϵ, we get that the sum above is at most

$$C_\epsilon n^{3/2} \cdot (eC_\epsilon^{-1} pn^{1/2})^{C_\epsilon n^{3/2}} e^{-O_\delta(pn^2)},$$

and this tends to 0 provided $pn^{1/2} \to \infty$ (since $pn^2 \gg n^{3/2} \log(pn^{1/2})$), proving the result. $\square$

Note that in this proof, the main extra power we gained by utilizing Theorem 19.7 is that S must be contained in our subgraph. This makes it so that the $S \in \mathcal{S}$ with many edges "cost more", allowing us to gain.

We note that in general, it is very common that by using the weak container lemma, one ends up getting the correct answer up to a logarithmic factor, which can usually be remedied by utilizing the strong container lemma in some straightforward (if slightly more tedious) way.

20 Maximal Triangle-Free Graphs TODO

21 Counting F-free graphs TODO

Ferber-McKinley-Samotij.

Part VI

Spectral Graph Theory TODO

Part VII

Bibliography

References

[1] N. Alon, M. Krivelevich, and B. Sudakov. Turán numbers of bipartite graphs and related ramsey-type questions. *Combinatorics, Probability and Computing*, 12(5-6):477–494, 2003.

[2] N. Alon and J. H. Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[3] R. Alweiss, S. Lovett, K. Wu, and J. Zhang. Improved bounds for the sunflower lemma. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 624–630, 2020.

[4] L. Babai and P. Frankl. *Linear algebra methods in combinatorics*. University of Chicago, 1988.

[5] J. Balogh. Container Method notes.

[6] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *Journal of the American Mathematical Society*, 28(3):669–709, 2015.

[7] J. Balogh, R. Mycroft, and A. Treglown. A random version of sperner's theorem. *Journal of Combinatorial Theory, Series A*, 128:104–110, 2014.

[8] T. Bell, S. Chueluecha, and L. Warnke. Note on sunflowers. *Discrete Mathematics*, 344(7):112367, 2021.

[9] B. Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.

[10] B. Bukh. Random algebraic construction of extremal graphs. *arXiv preprint arXiv:1409.3856*, 2014.

[11] B. Bukh. Extremal graphs without exponentially-small bicliques. *arXiv preprint arXiv:2107.04167*, 2021.

[12] B. Bukh and D. Conlon. Rational exponents in extremal graph theory. *Journal of the European Mathematical Society*, 20(7):1747–1757, 2018.

[13] D. Conlon. Graphs with few paths of prescribed length between any two vertices. *Bulletin of the London Mathematical Society*, 51(6):1015–1021, 2019.

[14] D. Conlon and A. Ferber. Lower bounds for multicolor ramsey numbers. *Advances in Mathematics*, 378:107528, 2021.

[15] D. Conlon, C. Pohoata, and D. Zakharov. Random multilinear maps and the erdos box problem. *arXiv preprint arXiv:2011.09024*, 2020.

[16] D. W. Cranston and D. B. West. An introduction to the discharging method via graph coloring. *Discrete Mathematics*, 340(4):766–793, 2017.

[17] S. Das, C. Lee, and B. Sudakov. Rainbow turán problem for even cycles. *European Journal of Combinatorics*, 34(5):905–915, 2013.

[18] M. Deza, P. Erdös, and P. Frankl. Intersection properties of systems of finite sets. *Proceedings of the London Mathematical Society*, 3(2):369–384, 1978.

[19] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[20] P. Erdös. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[21] R. J. Faudree and M. Simonovits. On a class of degenerate extremal graph problems. *Combinatorica*, 3:83–93, 1983.

[22] J. Fox and B. Sudakov. Dependent random choice. *Random Structures & Algorithms*, 38(1-2):68–99, 2011.

[23] P. Frankl and Z. Füredi. Forbidding just one intersection. *Journal of Combinatorial Theory, Series A*, 39(2):160–176, 1985.

[24] P. Frankl and N. Tokushige. *Extremal problems for finite sets*, volume 86. American Mathematical Soc., 2018.

[25] K. Frankston, J. Kahn, B. Narayanan, and J. Park. Thresholds versus fractional expectation-thresholds. *arXiv preprint arXiv:1910.13433*, 2019.

[26] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.

[27] Z. Füredi. On a turán type problem of erdös. *Combinatorica*, 11(1):75–79, 1991.

[28] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial*, pages 169–264. Springer, 2013.

[29] D. Galvin. Counting colorings of a regular graph. *Graphs and Combinatorics*, 31(3):629–638, 2015.

[30] S. Glock, D. Kühn, A. Lo, and D. Osthus. The existence of designs via iterative absorption: hypergraph f-designs for arbitrary f. *arXiv preprint arXiv:1611.06827*, 2016.

[31] D. S. Gunderson, V. Rödl, and A. Sidorenko. Extremal problems for sets forming boolean algebras and complete partite hypergraphs. *Journal of Combinatorial Theory, Series A*, 88(2):342–367, 1999.

[32] A. Heckel. Random triangles in random graphs. *Random Structures & Algorithms*, 2018.

[33] A. Heckel and O. Riordan. How does the chromatic number of a random graph vary? *arXiv preprint arXiv:2103.14014*, 2021.

[34] O. Janzer. Rainbow tur\'an number of even cycles, repeated patterns and blow-ups of cycles. *arXiv preprint arXiv:2006.01062*, 2020.

[35] T. Jiang, S. Letzter, A. Methuku, and L. Yepremyan. Rainbow clique subdivisions and blow-ups. *arXiv preprint arXiv:2108.08814*, 2021.

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

[37] J. Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combinatorics, Probability and Computing*, 10(3):219–237, 2001.

[38] J. Kahn, B. Narayanan, and J. Park. The threshold for the square of a hamilton cycle. *arXiv preprint arXiv:2010.08592*, 2020.

[39] P. Keevash. The existence of designs. *arXiv preprint arXiv:1401.3665*, 2014.

[40] P. Keevash, D. Mubayi, B. Sudakov, and J. Verstraëte. Rainbow turán problems. *Combinatorics, Probability and Computing*, 16(1):109–126, 2007.

[41] J. H. Kim. The ramsey number r (3, t) has order of magnitude t2/log t. *Random Structures & Algorithms*, 7(3):173–207, 1995.

[42] D. Kleitman. A conjecture of erdős-katona on commensurable pairs among subsets of an n-set. In *Theory of Graphs, Proc. Colloq. held at Tihany, Hungary*, pages 187–207, 1966.

[43] D. Kleitman. On dedekind's problem: the number of monotone boolean functions. *Proceedings of the American Mathematical Society*, 21(3):677–682, 1969.

[44] W. Kuszmaul and Q. Qi. The multiplicative version of azuma's inequality, with an application to contention analysis. *arXiv preprint arXiv:2102.05077*, 2021.

[45] L. Lovász. *Combinatorial problems and exercises*, volume 361. American Mathematical Soc., 2007.

[46] W. Mader. Homomorphieeigenschaften und mittlere kantendichte von graphen. *Mathematische Annalen*, 174(4):265–268, 1967.

[47] D. Mubayi and J. Verstraëte. A survey of turán problems for expansions. *Recent trends in combinatorics*, pages 117–143, 2016.

[48] D. Mubayi and Y. Zhao. Forbidding complete hypergraphs as traces. *Graphs and Combinatorics*, 23(6):667–679, 2007.

[49] A. Rao. Coding for sunflowers. *arXiv preprint arXiv:1909.04774*, 2019.

[50] A. Rao. Sunflowers: from soil to oil. *Bulletin of the American Mathematical Society*, 60(1):29–38, 2023.

[51] O. Riordan. Random cliques in random graphs. *arXiv preprint arXiv:1802.01948*, 2018.

[52] V. Rödl. On a packing and covering problem. *European Journal of Combinatorics*, 6(1):69–78, 1985.

[53] D. Romik. *The surprising mathematics of longest increasing subsequences*. Number 4. Cambridge University Press, 2015.

[54] W. Sawin. An improved lower bound for multicolor ramsey numbers and a problem of erdős. *Journal of Combinatorial Theory, Series A*, 188:105579, 2022.

[55] J. B. Shearer. On a problem of spencer. *Combinatorica*, 5(3):241–245, 1985.

[56] A. Sheffer. Polynomial methods and incidence theory.

[57] M. Simonovits and E. Szemerédi. Embedding graphs into larger graphs: results, methods, and problems. In *Building Bridges II*, pages 445–592. Springer, 2019.

[58] J. Spencer. Ramsey's theorem—a new lower bound. *Journal of Combinatorial Theory, Series A*, 18(1):108–115, 1975.

[59] J. Spencer. *Asymptopia*, volume 71. American Mathematical Soc., 2014.

[60] S. Spiro. A smoother notion of spread hypergraphs. *arXiv preprint arXiv:2106.11882*, 2021.

[61] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

[62] I. Tomon. Robust (rainbow) subdivisions and simplicial cycles. *arXiv preprint arXiv:2201.12309*, 2022.

[63] Y. Wigderson. An improved lower bound on multicolor ramsey numbers. *Proceedings of the American Mathematical Society*, 149(6):2371–2374, 2021.

[64] Y. Zhao. The number of independent sets in a regular graph. *Combinatorics, Probability and Computing*, 19(2):315–320, 2010.