Methods in Extremal Combinatorics

Sam Spiro*

July 22, 2021

Contents

*Dept. of Mathematics, UCSD sspiro@ucsd.edu. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650112.

Preface

This is a work in progress. There will be many typos, missing references, etc. Please let me know if you notices such errors, find anything confusing, or if you have any other suggestions!

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 are a 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 [19] or Bollobás [6]; surveys by Simonovits and Szemerédi [24] and Füredi and Simonovits [13]; and online courses by Morris and Gowers.

- Discrete geometry and the polynomial method: see the book by Sheffer [23], 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 [28], as well as online courses by Prendiville and Zhao.

- Statistical mechanics: see notes by Will Perkins.

Part I
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. The sections are loosely organized into three sub-parts: basic examples, methods for bounding the probability that a given event occurs, and other topics.

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 ([7]). *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]. $\square$

[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 [26].

[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

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

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 $r \leq n$ 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 $\mathrm{ex}(n, K_r) \geq t_{r-1}(n)$ 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

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

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

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

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

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

so $e(G) \leq t_{r-1}(n)$ 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 $e \geq 2$ edges. If $e \geq v$, 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 $y(x+y) + x(x+y) \geq 4xy$, which is equivalent to saying $x^2 - 2xy + y^2 = (x-y)^2 \geq 0$.

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 [11] 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

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

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 extremal

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

problems for hypergraphs. We only scratch the surface of this topic, see Frankl and Tokushige [9] 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 [9] 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

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

□

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. □

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. □

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. □

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. □

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. $\square$

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 [23].

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 $\Pr[A_i] \leq \gamma_i \prod_{j \in D_i}(1 - \gamma_j)$ for all i, then*

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

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 $\Pr[\bigcap \overline{A_i}] \geq \prod \Pr[\overline{A_i}]$.

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

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

This will give the result since then

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

We prove this claim by induction[1] on $|S|$. The base case $|S| = 0$ is equivalent to saying $\Pr[A_i] \leq \gamma_i$ for all i, and this follows from $\Pr[A_i] \leq \gamma_i \prod_{j \in D_i}(1 - \gamma_j) \leq \gamma_i$. 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

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

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

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

[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. $\square$

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. $\square$

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 [22].

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 [25])**.** *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

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

In particular we need $\gamma_3 \geq p^3$, so let's naively take $\gamma_3 = Cp^3$ for some large constant C. Given this, we also need $\gamma_n \leq c\binom{N}{n}^{-1}$ 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

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

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

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

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

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

Assuming $N \geq n^{1+\epsilon}$ for some small $\epsilon > 0$, this reduces to $N^{1/2} \leq n/\log n$, 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 [18]. 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**.

3.2 Other 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*

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

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 slightly 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*.

It often the case that one works with binomial random variables. In this case one can typically get much stronger concentration results by using the Chernoff bound.

Lemma 4.3 (Chernoff Bound)**.** *Let X be a random variable distributed according to a binomial distribution with n trials and probability p of success. Then for all $\lambda > 0$,*

$$\Pr[|X - pn| > \lambda pn] \le 2e^{-\lambda^2 pn/3}.$$

We omit the proof of this result, see e.g. the appendix of [2] for a proof.

There are many, many other nice concentration results that are useful in various circumstances when using the probabilistic method. We will omit discussing these further, but we suggest reading e.g. the chapter on martingales in [2].

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 4.4 (Rödl [21])**.** *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 4.4 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 4.4 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 4.5. *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

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

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

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

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 (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 4.6 (Pippenger). *For every $r \ge 2$ and reals $K \ge 1$ and $a > 0$, there are $\delta = \delta(r, K, a) > 0$ and $D_0 = D_0(r, K, a)$ such that for every $n \ge D \ge D_0$ 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 4.4.

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-

[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.

coverings with exactly $\binom{n}{k}/\binom{r}{k}$ edges, so Theorem 4.4 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 [17] and independently by Glock, Kühn, Lo, and Osthus [14] 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!

5 Spread Hypergraphs

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 5.1 ([3, 10])**.** *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 $Cq|V|\log r$ chosen uniformly at random from V with $C \ge K_0$, then*

$$\Pr[W \textit{ contains an edge of } \mathcal{H}] \ge 1 - \frac{K_0}{C\log r}.$$

5.1 Applications

Let's accept Theorem 5.1 for the moment and look at some basic applications.

Theorem 5.2. *Let $H^r_{n,m}$ be the r-graph chosen uniformly at random amongst all r-graphs with n vertices and m edges. Then there exists a constant K such that if $m \ge Kn\log n$ and n is a multiple of r, then $H^r_{n,m}$ 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, $H^2_{n,m}$ acts the same as $G_{n,p}$ where $p = m/\binom{n}{2}$. In particular, one can use Theorem 5.2 to prove that $G_{n,p}$ contains a perfect matching a.a.s. if $p = \Omega(\log n/n)$. Proving Theorem 5.2 for $r = 2$ is not that hard, but the result for general r was seemingly very difficult, with its first proof due to Johansson, Kahn, and Vu [15] which used very complicated arguments. We'll be able to prove it in a few lines with Theorem 5.1.

Proof. Let $\mathcal{H}$ be the hypergraph on $E(K^r_n)$ where each hyperedge S is a perfect matching of K^r_n. Observe that for any set $A \subseteq E(K_n)$, we have

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

Thus $\mathcal{H}$ is $(en/r)^{-r+1}$-spread. It is also (n/r)-uniform and has a ground set $V = E(K^r_n)$ of size $\binom{n}{r}$. By Theorem 5.1, we see that with high probability $H^r_{n,m}$ will contain a perfect matching, i.e. a hyperedge of $\mathcal{H}$, provided m is as least as large as claimed. $\square$

[1]Some texts would say that such an 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.

Another basic example is the following.

Proposition 5.3. *Let F be an r-graph and define $t(F) = \max\{|E(F')|/|V(F')| : F' \subseteq F\}$. Let $H^r_{n,m}$ be as in Theorem 5.2. There exists a constant $C(F)$ such that if $m \geq C(F)n^{r-1/t(F)}$, then $H^r_{n,m}$ contains a copy of F a.a.s.*

A simple first moment argument shows that this bound is tight. One can prove Proposition 5.3 using a somewhat tedious second moment argument, but using Theorem 5.1 gives a shorter proof.

Proof. Let $\mathcal{H}$ be the hypergraph on $E(K_n^r)$ whose hyperedges correspond to copies of F. 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

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

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 5.1 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 5.1. Their motivation came from the Erdős sunflower conjecture. A *k-sunflower* is a hypergraph with edges $S_1, \ldots, S_k$ such that there exists a set K called the *kernel* which has $S_i \cap S_j = K$ for all $i \neq j$.

Theorem 5.4 ([3, 20, 5])**.** *There exists a constant C such that if $\mathcal{H}$ is an r-graph with more than $(Ck \log r)^r$ edges, then $\mathcal{H}$ contains a k-sunflower.*

We note that [3] was the first to prove a theorem of this form, with [20, 5] later giving better bounds in terms of k. When k is fixed, Theorem 5.4 gives a bound of $(\log r)^{r+o(1)}$. Prior to [3], the best known bounds were of the form $r^{r-o(1)}$, and this was despite many years of people trying very hard to decrease this bound. It is a famous conjecture of Erdős that the best possible bound is 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 this many edges. If $\mathcal{H}$ is not q-spread with $q = (Ck \log r)^{-1}$, then there exists some $A \subseteq V(H)$ such that $d(A) \geq (Ck \log r)^{r-|A|}$, so by induction the link[1] of A contains a sunflower, which means $\mathcal{H}$ does as well. Thus we can 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 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 $(2k)^{-1}|V| = \frac{1}{2}Cq\log(r)|V|$. Let 1_i be the indicator variable for V_i containing an edge of $\mathcal{H}$. By Theorem 5.1, 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 $\mathcal{H}$. Thus $\mathcal{H}$ contains a sunflower, a contradiction. □

[1] By this we mean the $r - |A|$ uniform hypergraph consisting of all edges of the form $\{S \setminus A : S \in \mathcal{H}, A \subseteq S\}$

Soon after the work of [3], Frankston, Kahn, Narayanan, and Park [10] refined the q-spread method and proved Theorem 5.1 in order to prove an absolutely remarkable result regarding thresholds in random structures. **Which I may write up at some point in time, see notes by Das in the meantime.**

As we noted earlier, the bound of Theorem 5.2 is best possible. In particular, the $\log r$ term of Theorem 5.1 is necessary in general. However, under certain conditions one can remove this logarithmic term. This was first observed by Kahn, Narayanan, and Park [16] where they found tight bounds on the threshold of a square of a Hamiltonian cycle in $G_{n,p}$. A more general version of Theorem 5.1 which captures when it is possible to remove the logarithmic term is given in [27], and we'll briefly discuss the ideas of how to prove such a result after our proof of Theorem 5.1.

5.2 Proof of Theorem 5.1

The proof will closely follow this paper of mine, and I'll write this proof after the paper gets refereed.

6 Dependent Random Choice

The following is all based off of the excellent survey by Fox and Sudakov [8]. 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 6.1. *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 $S \subseteq V(G)$ 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. □

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 easier for vertices in W to have large common neighborhoods).

We can use Lemma 6.1 to prove some bounds on Turán numbers by using the following embedding lemma.

Lemma 6.2. *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. $\square$

With this we can quickly prove the following.

Theorem 6.3 (Füredi [12]; 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 Kovari-Sos-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 6.2, 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 6.1, 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 6.2 gives a copy of F in G, a contradiction. $\square$

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 6.4 (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 6.1 on its own is not enough to prove Theorem 6.3, 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 6.1 depends on the particular problem at hand.

Lemma 6.5. *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 6.4 follows almost immediately from Lemma 6.5, 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 6.2.

Lemma 6.6. *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 $\phi(v_m), \phi(v_{m-1}), \dots, \phi(v_{q+1})$ and that $v_q \in V_i$. Since $S := N(v_q) \cap \{v_m, \dots, v_{q+1}\}$ 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 6.1.

Lemma 6.7. *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 6.1 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 6.1 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 6.1 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 6.8. *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 6.7 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!

7 The Regularity Lemma

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.

Part II
Spectral Graph Theory

To be written.

Part III

Hypergraph Containers

This part is heavily based off of lecture notes by Balogh [4].

8 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 an edge if and only if S is a 3-term arithmetic progression. Thus Roth's theorem is equivalent to saying that $\alpha(\mathcal{H}_n^{AP}) = o(n)$. One can also 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 $\mathcal{H}_n^{\Delta}$ correspond to 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 $\mathcal{I}(H)$, the set of independent sets in a hypergraph 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 $C \in \mathcal{C}$. For such a set of containers, we have

$$|\mathcal{I}(H)| \leq \sum_{C \in \mathcal{C}} 2^{|C|} \leq |\mathcal{C}| 2^{\max_{C \in \mathcal{C}} |C|}. \tag{10}$$

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

$$\binom{\alpha(H)}{m} \leq |\mathcal{I}_m(H)| \leq |\mathcal{C}| \binom{\max_{C \in \mathcal{C}} |C|}{m}. \tag{11}$$

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

More notes forthcoming.

Part IV

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] J. Balogh. Container Method notes.

[5] T. Bell, S. Chueluecha, and L. Warnke. Note on sunflowers. *Discrete Mathematics*, 344(7):112367, 2021.

[6] B. Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.

[7] P. Erdös. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[8] J. Fox and B. Sudakov. Dependent random choice. *Random Structures & Algorithms*, 38(1-2):68–99, 2011.

[9] P. Frankl and N. Tokushige. *Extremal problems for finite sets*, volume 86. American Mathematical Soc., 2018.

[10] K. Frankston, J. Kahn, B. Narayanan, and J. Park. Thresholds versus fractional expectation-thresholds. *arXiv preprint arXiv:1910.13433*, 2019.

[11] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.

[12] Z. Füredi. On a turán type problem of erdös. *Combinatorica*, 11(1):75–79, 1991.

[13] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial*, pages 169–264. Springer, 2013.

[14] 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.

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

[16] J. Kahn, B. Narayanan, and J. Park. The threshold for the square of a hamilton cycle. *arXiv preprint arXiv:2010.08592*, 2020.

[17] P. Keevash. The existence of designs. *arXiv preprint arXiv:1401.3665*, 2014.

[18] J. H. Kim. The ramsey number r (3, t) has order of magnitude t2/log t. *Random Structures & Algorithms*, 7(3):173–207, 1995.

[19] L. Lovász. *Combinatorial problems and exercises*, volume 361. American Mathematical Soc., 2007.

[20] A. Rao. Coding for sunflowers. *arXiv preprint arXiv:1909.04774*, 2019.

[21] V. Rödl. On a packing and covering problem. *European Journal of Combinatorics*, 6(1):69–78, 1985.

[22] J. B. Shearer. On a problem of spencer. *Combinatorica*, 5(3):241–245, 1985.

[23] A. Sheffer. Polynomial methods and incidence theory.

[24] M. Simonovits and E. Szemerédi. Embedding graphs into larger graphs: results, methods, and problems. In *Building Bridges II*, pages 445–592. Springer, 2019.

[25] J. Spencer. Ramsey's theorem—a new lower bound. *Journal of Combinatorial Theory, Series A*, 18(1):108–115, 1975.

[26] J. Spencer. *Asymptopia*, volume 71. American Mathematical Soc., 2014.

[27] S. Spiro. A smoother notion of spread hypergraphs. *arXiv preprint arXiv:2106.11882*, 2021.

[28] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.