THE PROBABILISTIC PARADIGM: A COMBINATORIAL PERSPECTIVE

Niranjan Balachandran
Dept. of Mathematics, IIT Bombay.

Contents

1 Preface

One of the most popular motifs in mathematics in recent times, has been the study of the complementarity between the notions of 'Structure' and 'Randomness' in the sense that most mathematical structures seem to admit this broad dichotomy of characterisation. It is little wonder then, that probability theory has become a ubiquitous tool in as varied areas of mathematics as Differential equations, Number theory, and, Combinatorics, to name a few, as it brings forth the language to describe 'randomness'.

Combinatorics is one of the areas where this dichotomy has played a very critical role in the resolution of several very interesting problems. Combinatorics has long been held as an area which , unlike many other areas of mathematics, does not involve a great deal of theory building. However, it is of course not true that there is no general sense of 'theory in combinatorics; to quote Gowers from his iconic essay, 'The two Cultures of Mathematics', "The important ideas of combinatorics do not usually appear in the form of precisely stated theorems, but more often as general principles of wide applicability." What plays an equivalent role in Combinatorics for 'theory' would be something akin to 'general principles' which shape the manner in which the combinatorist generally forms his/her view. And one of the principal principles at work in combinatorics is the Probabilistic paradigm.

The word 'paradigm' as listed on dictionary.com describes '*a framework containing the basic assumptions, ways of thinking, and methodology that are commonly accepted by members of a scientific community*'. The probabilistic paradigm in combinatorics, was initiated largely due by the seminal work of Paul Erdős who ushered in the language,the viewpoint, and perspectives it offers to problems in combinatorics, and today, probabilistic tools are an indispensable part of the combinatorist's arsenal.

One of the main reasons for the ubiquity and all-pervasive nature of the method is that it provides a tool to deal with the 'local-global' problem. More specifically, many problems of a combinatorial nature ask for a existence/construction/enumeration of a finite set structure that satisfies a certain combinatorial structure locally at every element. The difficulty in many a combinatorial problem is to construct structures that are 'locally good', everywhere. A significant part of this difficulty arises from the fact that, often, there seem to be several possible choices for picking a local structure, but *no canonical*

choices, consequently, it is not clear which local choices are preferential. The probabilistic paradigm enables one to consider all these 'local' patches simultaneously and provide what one could call 'weak' conditions for building a global patch from the local data. In recent times, many impressive results settling several long standing open problems, via the use of the probabilistic method, essentially relying on this principle. But one thing that stands out in all these results is: the techniques involved are often quite subtle, ergo, one needs to understand *how to use these tools, and think probabilistically.*

Coming to existing literature on this subject, there are some truly wonderful monographs - the ones by Alon-Spencer [5], Spencer [29], Bollobás [6], Janson *et al* [18], Molloy-Reed[23] spring readily to mind - on the probabilistic method in combinatorics, specifically. In addition to this, one can find several lecture notes' compilations on the probabilistic method, on the internet. So, what would we seek to find in another book? What ought it to offer one that is, say missing, in all this plethora of material that is already available?

Most of the available material attempt to keep the proofs easy to follow and simple to verify. But that invariably makes the proof appear rather magical, and almost always obscures the thought processes behind them. Indeed, many interesting (probabilistic) arguments appear in situations that do not seem to involve any probability at all, so a certain *sprezzatura* is distinctly conveyed. So a new book could certainly do with a *deconstructionist perspective.*

This book arose as a result of lectures for a graduate course - first at Caltech, and then later at IIT Bombay - with the goal of providing that sense of perspective. Tim Gowers has on more than one occasion written about 'The Exposition Problem': "Solving an open exposition problem means explaining a mathematical subject in a way that renders it totally perspicuous. Every step should be motivated and clear; ideally, students should feel that they could have arrived at the results themselves." [1] An entire book in this spirit has not appeared before, and that is what this book really attempts to do.

This book could be criticsed as 'a bunch of deconstructions of some specific results that arise from the idiosyncrasies of the author's choices'. Indeed, while some of the results that appear are best known, there are others that are not best possible - even within the material that appears in the book. My counter to that would be - Yes....and No. The choice of material that has been included here is indeed a reflection of my own tastes and preferences. But the theorems and results that appear here also reflect an aspect of each of the techniques that are discussed, in a very specific sense; if for instance a result appears in the chapter on the Second moment method, then the second moment computation there *is key to the eventual result.* In that sense, the book is tightly structured.

[1] This is from his blog where he in turn was quoting Tim Chow.

I do not (deliberately) include proofs or detailed discussions of many important results from probability although I do state the ones in their full form for their utility within the confines of this book. The reason is twofold: this is basically a book on combinatorics, which forms and informs the topics of interest in the first place. Secondly, the imperative is to provide a perspective into probabilistic heuristics and reasoning and not get into the details and technicalities of probabilistic results in themselves. I list some sources as references throughout the text for related reading.

As mentioned earlier, principles in combinatorics play the role of theory in most other areas in mathematics. While most experts are aware of (or acquainted with) these principles, and have some other principles of their own[2] these never see an explicit mention in books (though some blogs like that of Tao or Gowers do a fabulous job there), and it is for a well-founded reason: these principles are *more akin to rules-of-thumb and a formal statement attempting to put this in words will inevitably be an oversimplification that amounts to an incorrect statement.* But my opinion is, these simplistic heuristics go a great way in laying a pathway, not just towards solving open conjectures, but also allow us to pose newer interesting questions. Towards that end, I put as an encapsulation, one core principle from each chapter; each chapter's title includes an epigram that attempts a heuristic describes the underlying principle.

I thank all my students who very actively and enthusiastically acted as scribes for the lectures over the years, and those scribed notes, formed the skeleton for this book.

[2] á la Groucho Marx, perhaps.

2 Notation with asymptotics

This is a brief primer on the Landau asymptotic notation. Given functions f, g, we write $f \gg g$ (resp. $f \ll g$) if $\lim_{n\to\infty} \frac{f(n)}{g(n)} \to \infty$ (resp. $\to 0$). We also write $f = o(g)$ to denote that $f \ll g$. We write $f = O(g)$ (resp. $f = \Omega(g)$) if there exists an absolute constant $C > 0$ and n_0 such that for all $n \geq n_0, |f(n)| \leq C|g(n)|$ (resp. if $|f(n)| \geq C|g(n)|$), and finally when we write $f = \Theta(g)$ then we mean $f = O(g)$ and $g = O(f)$. If $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$, then we write $f \sim g$.

Here is a simple proposition that is left as an exercise.

Proposition 1. *Suppose f, g, h are functions defined on the integers.*

1. *(Transitivity) $f = o(g)$ and $g = o(h)$ implies $f = o(h)$. A similar statement holds for O and Θ as well.*

2. *If $g = o(f)$ then $f + g = \Theta(f)$.*

The main advantage of using this notation is (besides making several statements look a lot neater than they would if written in their exact analytic form) that the asymptotic notation allows us to 'invert' some functions in an asymptotic sense when an exact inversion is not feasible. But before we make that precise, here is a simple proposition that follows easily from the definition of the notation.To make this precise, consider the following example. Suppose $n \leq Ck \log k$. Then how do we get a lower bound for k in terms on n?

Here is a simple trick (and this will be used repeatedly in the book). The given inequality gives us

$$k \geq \frac{cn}{\log k}$$

and since clearly $k \leq n$ (otherwise there is nothing to discuss further), this gives us

$$k \geq \frac{cn}{\log n}$$

which is best possible asymptotically since if $k = \frac{n}{\log n}$ then

$$k \log k = \frac{n}{\log n}(\log n - \log\log n) = n\left(1 - \frac{\log\log n}{\log n}\right) = n(1 - o(1)).$$

One can use this idea more iteratively as well as we shall see in the book.

Computations with asymptotics is an art, and also needs a bit of practice before one can get comfortable with it. We illustrate one case in a little more detail; this is the calculation from the second chapter (which is omitted there). Suppose for a fixed k, we wish to maximize $n - \binom{n}{k}2^{-\binom{k}{2}+1}$. Note that as n increases, this quantity eventually becomes negative, so one needs to find the optimal n for which this is large. Unfortunately, the usual maxima/minima methods of calculus are not directly applicable here, so the perspective is motivated more by an eye on the asymptotics.

Since the given quantity cannot exceed n, from an asymptotic perspective, we would be happy if this quantity is at least as large as $n/2$. To see if we can achieve that, since $k! \geq \left(\frac{k}{e}\right)^k$, we have $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$, so

$$\binom{n}{k}2^{-\binom{k}{2}} \leq \left(\frac{en}{k2^{(k-1)/2}}\right)^k.$$

Now set

$$4n = \left(\frac{en}{k2^{(k-1)/2}}\right)^k.$$

This gives

$$\begin{aligned} n &= \left(4^{1/k}\frac{k}{e}2^{(k-1)/2}\right)^{\frac{k}{k-1}} \\ &= 4^{1/(k-1)}(\frac{k}{e})^{k/(k-1)}2^{k/2} \\ &= (1+o(1))\frac{k2^{k/2}}{e} \end{aligned}$$

for k large.

Here is a second example, again from the second chapter. Again, the question is the following: Given d, maximize s such that $d > 4\binom{s^4}{s}\log\left(2\binom{s^4}{s}\right)$. Again, since we shall not be too concerned by constants, let us set $d = 8\binom{s^4}{s}\log\left(2\binom{s^4}{s}\right)$. Write $Y = 2\binom{s^4}{s}$; then this gives us $d = 4Y\log(Y)$, and so by our discussions from earlier, this gives us $Y = \Omega\left(\frac{d}{\log d}\right)$. As for the asymptotics of Y, since $\binom{s^4}{s} \leq (es^3)^s \leq s^{4s}$, we have $s^{4s} \geq \Omega(\frac{d}{\log d})$ so taking

log both sides gives us $4s \log s \geq \log d(1 - o(1))$. Once again, by the same argument as from earlier, this gives us $s \geq \Omega\left(\frac{\log d}{\log \log d}\right)$.

To the uninitiated, we close this chapter by recommending *Asymptopia* by J. Spencer [**?**] for a fascinating introduction to the world of asymptotics.

3 The Basic Idea

> If you cannot think of anything clever, or worse, cannot *think of anything*, roll the die and take your chances.
>
> T

he probabilistic paradigm or more simplistically, the probabilistic method, is based on the following premise: Given a (combinatorial) problem, one may set up a probability on the underlying set, and one may then compute (or estimate) the probability of an event not occurring. If this probability is less than one, then the corresponding set that describes the occurrence of the event is nonempty. This is a rather simple description of the method, and yet hardly anything that explains how one might go about this is revealed. Our goal in this book is to attempt a deconstruction of this process.

We shall assume a basic familiarity with the notions of probability theory and graph theory. As a good reference, we recommend [31] for probability theory, and [32] for graph theory.

3.1 Lower bounds on the Ramsey number $R(n, n)$

We start with the instance that ' started it all' - the famous Erdős lower bound on the Ramsey numbers. Ramsey theory, roughly stated, is the study of how "order" grows in systems as their size increases. In the language of graph theory, the first result that founded the basics of Ramsey theory is the following:

Theorem 2. *(Ramsey, Erdős-Szekeres) Given a pair of integers s, t, there is an integer $R(s, t)$ such that if $n \geq R(s, t)$, any 2-coloring of the edges of K_n using colors red and blue must yield either a red K_s or a blue K_t.*

A fairly simple recursive upper bound on $R(s, t)$ (proved inductively, and a good exercise if you haven't seen it before) is given by

$$R(s, t) \leq R(s, t-1) + R(s-1, t),$$

which gives us

$$R(s,t) \leq \binom{k+l-2}{k-1}$$

and thus, asymptotically, that

$$R(s,s) \leq C\frac{4^s}{\sqrt{s}}$$

for some constant C and for s sufficiently large.

A constructive lower bound on $R(s,s)$, discovered by Nágy, is the following:

$$R(s,s) \geq \binom{s}{3}.$$

Explicitly, his construction goes as follows: take any set S, and turn the collection of all 3-element subsets of S into a graph by connecting subsets iff their intersection is odd. The graph represents the red edges, and the non-edges are the blue edges. It is a not-entirely trivial exercise to show that this coloring admits no monochromatic clique of size s.

There is a rather large gap between these two bounds; one natural question to ask, then, is which of these two results is "closest" to the truth? Turán believed that the correct order was of the order s^2. Erdős, in 1947, in a tour-de-force paper "'' disproved Turán's conjecture in a rather strong form:

Theorem 3. *(Erdős) For $s \geq 3$,*

$$R(s,s) > \lfloor 2^{s/2} \rfloor.$$

Proof. A lower bound entails a coloring of the edges of K_n using colors red and blue with no monochromatic complete subgraph on s vertices. If one looks at Nágy's example, one is tempted to think of a 'global' recipe for coloring the edges in some manner that witnesses the lack of sufficiently large monochromatic complete subgraphs. The reason for this 'global' outlook is, if one starts with an ad-hoc coloring of the edges, there seems to be plenty of leeway to color each edge one way or the other before our color choices force our hand, and even on the occasions that they do, it is hard to see if earlier choices could have been altered to improve upon the number we have so far. And lastly, it is hard to see how this pattern (if one could use such a word here!) generalises for large s. And in this conundrum of a situation, where local choices for edge colorings do not seem clear, Erdős appealed to the principle explicated as the slogan of the chapter:

If you cannot think of anything clever, or worse, cannot think of anything, roll the die and take your chances.

Fix n and consider a random 2-coloring of the edges of K_n. In other words, let us work in the probability space $(\Omega, Pr) = (\text{all 2-colorings of } K_n\text{'s edges}, Pr(\omega) = 1/2^{\binom{n}{2}}.)$ An alternate way of describing this would be to consider a random 2-coloring to be one where each edge of K_n is independently colored red or blue with probability 1/2, since there is no reason to prefer one color over another.

For some fixed set R of s vertices in $V(K_n)$, let A_R be the event that the induced subgraph on R is monochromatic. Then, we have that

$$\mathbb{P}(A_R) = 2 \cdot \left(2^{\binom{n}{2}-\binom{s}{2}}\right)/2^{\binom{n}{2}} = 2^{1-\binom{s}{2}}.$$

Thus, we have that the probability of at least one of the A_R's occurring is bounded by

$$\mathbb{P}(\bigcup_{|R|=s} A_R) \leq \sum_{R\subset\Omega,|R|=s} \mathbb{P}(A_R) = \binom{n}{s} 2^{1-\binom{k}{2}}.$$

If we can show that $\binom{n}{s}2^{1-\binom{s}{2}}$ is less that 1, then we know that with nonzero probability there is a 2-coloring $\omega \in \Omega$ in which none of the bad events A_R's occur! In other words, we know that there is a 2-coloring of K_n that avoids both a red and a blue K_s, even though we do not have such a coloring explicitly!

Solving, we see that

$$\binom{n}{s} 2^{1-\binom{s}{2}} < \frac{n^s}{s!} \cdot 2^{1+(s/2)-(s^2/2)} = \frac{2^{1+s/2}}{s!} \cdot \frac{n^s}{2^{s^2/2}} < 1$$

whenever $n = \lfloor 2^{s/2} \rfloor, s \geq 3$. ■

3.2 Tournaments and the S_k Property

A **tournament** is simply an oriented K_n; in other words, it's a directed graph on n vertices where for every pair (i, j), there is either an edge from i to j or from j to i, but not both. A tournament T is said to have property S_k if for any set of k vertices in the tournament, there is some vertex that has a directed edge to each of those k vertices. One way to think of this is to imagine a tournament of some game where each pair of players play each other - there are no draws - and for players i, j we shall indicate by the directed edge (i, j) the outcome of the game between these players with i beating j. In these terms, the property S_k indicates that in this tournament, for every set of k players, there was always some player who beat them all.

One natural question to ask about the S_k property is the following:

Question 4. *For a given arbitrary k, is there always a tournament with property S_k? If yes, how small can such a tournament be?*

This problem again seeks an orientation of the edges that achieves S_k for starters, and when that is done, to see if this property has a universality to it. Note that unlike the Ramsey problem *it is not true that all sufficiently large tournaments have property* S_k. Indeed, a transitive tournament - a tournament where the players come seeded, and all the games between them respect their rankings - clearly does not possess S_k since no one beats the top ranked player.

For small k - $1, 2, 3$ - one can answer these questions to some degree of satisfaction; indeed, we can calculate values of S_k through ad-hoc arguments:

- If $k = 1$, a tournament will need at least 3 vertices to satisfy S_k (take a directed 3-cycle.)
- If $k = 2$, a tournament will need at least 5 vertices to satisfy S_k .
- If $k = 3$, a tournament will need at least 7 vertices to satisfy S_k (related to the Fano plane.)

For $k = 4$, constructive methods have yet to find an exact answer. Indeed, constructive methods have been fairly bad at finding asymptotics for how these values grow. And again, anyone who takes a stab at this problem, realises very quickly that the fundamental problem here is one of choice; there does not seem to be a canonical choice for orienting edges one way or another, for each edge, and again, as with the Ramsey problem, it is hard to unravel which choices lead to what outcomes. And so, we bring out the maxim once more.

Proposition 5. *(Erdős) There are tournaments that satisfy property* S_k *on* $O(k^2 2^k)$*-many vertices.*

Proof. Consider a random tournament: in other words, for every edge (i, j) of K_n direct the edge $i \to j$ with probability $1/2$ and from $j \to i$ with probability $1/2$. Again, this uniformity in choosing the edge orientation reflects our ambiguity for not preferring either direction.

Fix a set S of k vertices and some vertex $v \notin S$. What is the probability that v has an edge to every element of S? Relatively simple: in this case, it's just $1/2^k$, so that the probability that v fails to have a directed edge to each member of S is $1 - 1/2^k$. We shall notate this by event as $v \not\to S$.

For different vertices $v \notin S$, the events $v \not\to S$ are all independent since these events are determined by the edge orientations of disjoint sets of edges, so we know in fact that

$$\mathbb{P}(\text{for all } v \notin S, v \not\to S) = \left(1 - 1/2^k\right)^{n-k}.$$

There are $\binom{n}{k}$-many such possible sets S; so, by using the union bound again, we have

$$\mathbb{P}(\text{There exists } S \text{ such that for all } v \notin S, v \not\to S) \leq \binom{n}{k} \cdot \left(1 - 1/2^k\right)^{n-k}.$$

As before, it suffices to force the right-hand side to be less than 1 as this means that there is at least one orientation of the edges of K_n on which no such subsets S exist – i.e. that there is a tournament that satisfies S_k.

This takes us into a world of approximations. Using the approximations $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ and $1 - x \leq e^{-x}$, we calculate:

$$\begin{aligned} &\left(e^{-1/2^k}\right)^{n-k} < 1 \\ \Leftrightarrow &\left(\frac{en}{k}\right)^k < e^{(n-k)/2^k} \\ \Leftrightarrow &k(1 + \log(n/k)) \cdot 2^k + k < n \end{aligned}$$

Motivated by the above, take $n > 2^k \cdot k$; this allows us to make the upper bound

$$\begin{aligned} k(1 + \log(n/k)) \cdot 2^k + k &< k(1 + \log(k2^k/k)) \cdot 2^k + k \\ &= 2^k \cdot k^2 \cdot \log(2) \cdot \left(1 + \frac{1}{k \log(2)} + \frac{1}{k2^k \log(2)}\right) \\ &= k^2 2^k \log(2) \cdot (1 + O(1)); \end{aligned}$$

so, if $n > k^2 2^k \log(2) \cdot (1 + O(1))$ we know that a tournament on n vertices with property S_k exists. ■

Remark: The asymptotics of this problem are still not known exactly. However, it is known (as was shown by Szekeres) that a tournament on n players satisfying S_k needs $ck2^k$ vertices for some absolute constant $c > 0$.

As we move to our next few instances of the basic method, we introduce the basic tool that gets the probabilistic method going. For a real -valued random variable X on a finite probability space, the *Expectation* of X (denoted $\mathbb{E}(X)$) is defined as

$$\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x\mathbb{P}(X = x).$$

Note that since the sum is finite since the probability space is finite.

The expectation is the first important tool that one plays with, in this process, and one of the reasons it is a useful and simple quantity to play with, is that for random

variables X, Y on the same space, $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.[1] The main handle the expectation gives us is the following: If $\mathbb{E}(X) \geq \alpha$, then with positive probability, $X \geq \alpha$. A similar statement holds for other inequalities as well.

A philosophical point before we get into applications of the idea. Why is the expectation a useful tool? Here is a heuristic: An expectation computation, in some sense, enumerates ordered pairs, and the formal definition for the expectation, fixes one of the parameters of the ordered pair, and enumerates over the other parameter while fixing the former. But one of the oldest combinatorial insights is that one might *interchange the order of enumeration* and this principle allows us to reinterpret the same computation from another perspective. This has the advantage of making what seems a 'global' computation, the sum of 'local' computations.

3.3 Sum-Free Sets of Integers

This is another gem originally due to Erdős. A set $B \subset \mathbb{R}$ is called *sum-free* if the sum of any two elements in B does not lie in B.

Theorem 6. *Every set of n nonzero integers contains a sum-free subset of size $\geq n/3$.*

Proof. For ease of notation, let us write $B = \{b_1, \ldots b_n\}$. Firstly, (and this is now a standard idea in Additive Combinatorics), we note that it is easier to work over finite groups than the integers, so we may take p large so that all arithmetic in the set A (in $\mathbb{Z}$) may be assumed to be arithmetic in $\mathbb{Z}/p\mathbb{Z}$. Furthermore, if we assume that p is prime, we have the additional advantage that the set is now a field, which means we have access to the other field operations as well. Thus we pick some prime $p = 3k+2$ that's (for instance) larger than twice the maximum absolute value of elements in B, and look at B modulo p – i.e., look at B in $\mathbb{Z}/p\mathbb{Z}$. Because of our choice of p, all of the elements in B are distinct mod p.

Now, look at the sets

$$xB := \{xb : b \in B\} \text{ in } \mathbb{Z}/p\mathbb{Z},$$

and let

$$N(x) = |[k+1, 2k+1] \cap xB| .$$

We are then looking for an element x such that $N(x)$ is at least $n/3$. Why? Well, if this happens, then at least a third of xB's elements will lie between $p/3$ and $2p/3$; take those elements, and add any two of them to each other.This yields an element between $2p/3$ and p, and thus one that's not in our original third; consequently, this subset of

[1] In more fanciful terms, expectation as an operator on the space of random variables is linear operator with operator norm 1.

over a third of xB is sum-free. But this means that this subset is a sum-free subset of B, because p was a prime; so we would be done.

So, the question again is: Is there a clever way of choosing an x that would optimally bring a big chunk of xB into the middle? Not really. So, let's just roll the dice - pick x uniformly (there is no reason to pick one element more than another) at random, and examine the expectation of $N(x)$:

$$\mathbb{E}(N(x)) = \sum_{b \in B} \left(\mathbb{1}_{x \cdot b \in [k+1, 2k+1]} \right) = n \cdot \frac{k+1}{3k+1} \geq n/3.$$

Thus, some value of x must make $N(x)$ exceed $n/3$, and thus insure that a sum-free subset of size $n/3$ exists. ■

Remark: One can ask the same question more generally on an arbitrary abelian groups, and there, the corresponding constant is $2/7$ (see [2]). For the integers, it remained a hard problem to determine if the constant $1/3$ could be improved, and as it turns out, $1/3$ is indeed the best possible constant (see [14]).

3.4 The Distinguishing chromatic number of Levy graphs

For a graph $G = (V, E)$ let us denote by $Aut(G)$, its full automorphism group. A labelling of the vertices using the labels $\{1, \ldots, r\}$ is said to be *distinguishing* (or r-distinguishing) if no nontrivial automorphism of the graph preserves all of the vertex labels. The *Distinguishing Chromatic Number*, a variant of the usual chromatic number of the graph introduced by Collins and Trenk, is defined as the minimum number of colors r, needed to color the vertices of the graph so that the coloring is both proper (adjacent vertices receive different colors) and distinguishing. In other words, the distinguishing chromatic number of a graph G (denoted $\chi_D(G)$) is the least integer r such that the vertex set $V(G)$ can be partitioned into sets $V_1, V_2, \ldots, V_r$ such that each V_i is independent in G, and for every $I \neq \pi \in Aut(G)$ there exists some color class V_i such that $\pi(V_i) \neq V_i$. It is not hard to see that this variant is well defined, and in recent times, this variant has attracted a lot of attention.

Here, we shall consider the following specific problem. Let $\mathbb{F}_q$ denote the finite field of order q, and let us denote the vector space $\mathbb{F}_q^3$ over $\mathbb{F}_q$ by V. Let $\mathcal{P}$ be the set of 1-dimensional subspaces of V and $\mathcal{L}$, the set of 2-dimensional subspaces of V. We shall refer to the members of these sets by points and lines, respectively. The *Levi graph* of order q, denoted by LG_q, is a bipartite graph defined as follows: $V(LG_q) = \mathcal{P} \cup \mathcal{L}$, where this describes the partition of the vertex set; a point p is adjacent to a line ℓ if and only if $p \in \ell$.

The choice of terminology of 'lines', 'points' is because the pair $(\mathcal{P}, \mathcal{L})$ is a *Projective plane of order* q, so every pair of points lie on a unique line, and every pair of lines have

a unique common point. For more, we refer the reader to [17], for instance.

The fundamental theorem of projective geometry [17] states that the full group of automorphisms of the projective plane $PG(2, \mathfrak{B}_q)$ is induced by the group of all non-singular semi-linear transformations $P\Gamma L(\mathbb{F}_q^3)$. If $q = p^n$ for a prime number p, $P\Gamma L(\mathbb{F}_q^3) \cong PGL(\mathbb{F}_q^3) \rtimes Gal(\mathbb{F}_q/\mathbb{F}_p)$. In particular, if q is a prime, we have $P\Gamma L(\mathbb{F}_q^3) \cong PGL(\mathbb{F}_q^3)$, so $P\Gamma L(\mathbb{F}_q^3)$ is a subgroup of the full automorphism group of LG_q. The full group is larger since it also includes maps induced by isomorphism of the projective plane with its dual.

Proposition 7. $\chi_D(LG_q) = 3$ *for all prime powers* $q \geq 7$.

Proof. First, let us see why 2 colors will not do. It is easy to see that LG_q is connected, so the only proper colorings correspond to the vertex partition $(\mathcal{P}, \mathcal{L})$. But every non-trivial map $A \in PGL(\mathbb{F}_q^3)$ induces an automorphism of LG_q which keeps the two color classes intact, and that establishes that $\chi_D(LG_q) > 2$.

To get a proper distinguishing 3-coloring of LG_q, one may imagine partitioning $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ and have as the three color classes, $\mathcal{P}, \mathcal{L}_1, \mathcal{L}_2$. To see what makes this a distinguishing coloring (it is clearly proper), consider for starters, an automorphisms ϕ induced by $P\Gamma L(\mathbb{F}_q^3)$; call it ϕ. If ϕ preserves the other two color classes then for every line $\ell \in \mathcal{L}$, $\phi(\ell)$ is in the same color class. In particular, the orbit

$$Orb_\phi(\ell) := \{\ell, \phi(\ell), \ldots\}$$

is contained within the same color class. And this happens for every automorphism.

Thus, a partition of $\mathcal{L}$ such that for at least one of the automorphisms induced by $P\Gamma L(\mathbb{F}_q^3)$, this property of all its orbits being within the same part is violated. This suggests a random partition of $\mathcal{L}$, i.e., for each $\ell \in \mathcal{L}$, set it in $\mathcal{L}_1$ or $\mathcal{L}_2$ independently, and uniformly. A bad event in this context would be the presence of a nontrivial automorphism that maps both these partitions into themselves, or in terms of the observation above, a bad event E_ϕ is the event where for each $\ell \in \mathcal{L}$, the orbit of ℓ is contained entirely in the part $\mathcal{L}_i$ containing it. This idea can be captured more generally as follows.

Suppose a graph G is given a vertex coloring using $\chi(G)$ colors and suppose C_1 is one of its color classes. Let $\mathcal{G}$ be the subgroup of $Aut(G)$ consisting of all automorphisms that C_1 as a set. For each $A \in \mathcal{G}$, let θ_A denote the total number of distinct orbits induced by the automorphism A in C_1. Fix an integer $t \geq 2$, and partition C_1 randomly into t parts, i.e., for each $v \in C_1$, pick uniformly and independently, an element in $\{1, 2, \ldots, t\}$ and assign v to the corresponding part.

For $\phi \in \mathcal{G}$, let E_ϕ denote the event that ϕ fixes every color class. Observe that if ϕ fixes a color class containing a vertex v, then all other vertices in the set $orb_\phi(v)$ are also

in the same color class. Moreover the probability that $Orb_\phi(v)$ is in the same color class of v, equals $t^{1-|Orb_\phi(v)|}$. Then

$$\mathbb{P}(E_\phi) = \prod_{\theta_\phi} t^{1-|Orb_\phi(v)|} = t^{\theta_\phi - |C_1|}$$

Let $\mathcal{N} \subset \mathcal{G}$ denote the set of all automorphisms which fixes each of the t parts that were partitioned, and let $N = |\mathcal{N}|$. Then note that

$$\mathbb{E}(N) \leq \sum_{\phi \in \mathcal{G}} \frac{1}{t^{|C_1|-\theta_\phi}} \tag{3.1}$$

If $\mathbb{E}(N) \leq f(\mathcal{G}) := \sum_{A \in \mathcal{G}} t^{\theta_A - |C_1|} < r$, where r is the least prime dividing $|\mathcal{G}|$, then with positive probability $N < r$. Since $\mathcal{N}$ is in fact a subgroup of $\mathcal{G}$, N divides $|\mathcal{G}|$, so it follows that with positive probability, $N = 1$, which means, we have a distinguishing proper coloring using $\chi(G) + t - 1$ colors.

Define

$$\begin{aligned} Fix_\phi(S) &:= \{v \in S : \phi(v) = v\}, & (3.2)\\ F_\phi(S) &= |Fix_\phi(S)|, & (3.3)\\ F(S) &:= \max_{\substack{\phi \in \mathcal{G} \\ \phi \neq I}} F_\phi(S) & (3.4) \end{aligned}$$

. Since $\theta_\phi \leq F(C_1) + \frac{|C_1|-F(C_1)}{2}$

$$\mathbb{E}(N) \leq \sum_{A \in \mathcal{G}} t^{\frac{F(C_1)-|C_1|}{2}} = |\mathcal{G}| t^{\frac{F(C_1)-|C_1|}{2}}.$$

Thus, if $F(C_1) < |C_1| - 2\log_t |\mathcal{G}|$ then there exists a distinguishing proper $\chi(G) + t - 1$ coloring of the graph.

Let us return to our setting. Set $\mathcal{G} = PGL(\mathbb{F}_q^3)$. It is a simple exercise to check that every $A \in PGL(\mathbb{F}_q^3)$ which is not the identity fixes at most $q + 2$ points of LG_q. Hence

$$\theta_A \leq q + 2 + \frac{(q^2 + q + 1) - (q + 2)}{2} = \frac{q^2 + 2q + 3}{2}.$$

Consequently,

$$f(\mathcal{G}) < \frac{(q^8 - q^6 - q^5 + q^3)}{t^{(q^2+1)/2}} + 1 \tag{3.5}$$

For $q = 7, t = 2$, the right hand side of (3.5) is approximately 1.16. Since the right hand side of inequality (3.5) is monotonically decreasing in q, it follows that $f(\mathcal{G}) < 2$ for $q \geq 7$, hence $\chi_D(LG_q) \leq 3$. ■

Remark: It turns out that $\chi_D(LG_5) = 3$ as well, and this again follows the same argument. The only difference is that the analysis above does not work, and one needs to explicitly compute $f(\mathcal{G})$ In this case, for $t = 2$ one calculate $f(\mathcal{G})$ explicitly (using a computer program) to obtain $f(\mathcal{G}) \approx 1.2$ to see that $\chi_D(LG_5) = 3$. For more results on the distinguishing chromatic number, see [9].

3.5 Colored hats and a guessing game

There are n friends standing in a circle so that everyone can see everybody else. On each person's head a randomly chosen hat - either black or white - is placed. After they have had a look at each other, they must make a claim on their hat color, or declare that they are unable to determine the color from what they have seen. They cannot hear the answers of their friends, and cannot communicate with each other in any manner, but they may make a strategy prior to the placement of the hats. They are awarded a grand prize if at least one person gets the color of her hat correct, and *no one gets her hat color wrong*. In the latter case, they are all punished.

One easy strategy to achieve a 50% success is if one of the friends takes a random guess, and the others pass on their guess ("I don't know the color of my hat"). The friends wish to devise a strategy that increases the probability of their getting the reward. The question is, do they have a better strategy? Again, the question is to be viewed as one that deals with n large but fixed.

First, let us formalize the problem. Denoting white and black by 1 and 0 respectively, any configuration of hats (on the friends' heads) is a point in $\{0,1\}^n$. Thus the i^{th} member witnesses a vector $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,i-1}, *, x_{i,i+1}, \ldots, x_{i,n})$, and these vectors are compatible in the sense that for all distinct i, j and $k \neq i, j$, we have $x_{i,k} = x_{j,k}$.

Suppose there exists a set $L \subset \{0,1\}^n$ such that for every element $(x_1, x_2, \ldots, x_n) \in W = \{0,1\}^n \setminus L$ there is an element $(y_1, y_2, \ldots, y_n) \in L$ such that the set $\{i \mid x_i \neq y_i\}$ has size 1; call such a set L *desirable*. The upshot is the rather crucial observation that a desirable set allows us to strategize as follows. Person i knows x_j for all $j \neq i$, so if there is a unique value of x_i so that $(x_1, x_2, \ldots, x_n) \in W$ then person i declares that her hat color is x_i.

This allows the friends to argue as follows. If they have a desirable set on their hands, then the strategy outlined above works *unless the color choice profile of the hats corresponds to a point of* L. Consequently, the probability that the friends get the award following this strategy equals $1 - \frac{|L|}{2^n}$. Thus to maximize this probability, they need a 'small' desirable set L. And since there is no canonical choice here, we shall pick it randomly.

Pick a set X by choosing each element of $\{0,1\}^n$ independently. But this time, the probability distribution is not clear. Picking each element with probability $1/2$ as in the preceding examples would result in a very large set with high probability - this will become more formal in later chapters. So for the moment, let us pick each such element with probability p, where p is a parameter that is to be determined later.

For a fixed $\mathbf{x} \in \{0,1\}^n$, let $\mathbb{1}_{\mathbf{x}\in L}$ denote the random variable that takes value 1 if $\mathbf{x} \in L$ and 0 otherwise. Note that $\mathbb{E}(\mathbb{1}_{\mathbf{x}\in L}) = \mathbb{P}(\mathbb{1}_{\mathbf{x}\in L})$. By linearity of expectation,

$$\mathbb{E}(|X|) = \mathbb{E}(\sum_{\mathbf{x}\in\{0,1\}^n} \mathbb{1}_{\mathbf{x}\in L}) = \sum_{\mathbf{x}\in\{0,1\}^n} \mathbb{E}(\mathbb{1}_{\mathbf{x}\in L}) = \sum_{\mathbf{x}\in\{0,1\}^n} \mathbb{P}(\mathbb{1}_{\mathbf{x}\in L}) = 2^n p.$$

Let Y be the set of elements which differ from the chosen elements in at least two coordinates. For a fixed $\mathbf{x} \in \{0,1\}^n$, it is easy to see that $\mathbf{x} \in Y$ is equivalent to saying that no element in the 'ball'[2] $B(\mathbf{x})$ consisting of all the elements which differ from $\mathbf{x}$ in at most one coordinate are not chosen into X. Since $|B(\mathbf{x})| = n + 1$ (the element $\mathbf{x}$ and the n elements that differ from $\mathbf{x}$ in exactly one coordinate) we have

$$\mathbb{E}(|Y|) = \sum_{\mathbf{x}\in\{0,1\}^n} \mathbb{P}(\mathbf{x} \in Y) = 2^n(1-p)^{n+1}.$$

so But now consider the set $L = X \cup Y$; this is indeed a desirable set! Furthermore, $E(|L|) = E(|X| + |Y|) = 2^n(p + (1-p)^{n+1})$. Minimizing this over $p \in [0,1]$ (basic calculus), gives us $p = 1 - \frac{1}{(n+1)^{1/n}}$.

Plugging this back in the preceding expression gives us

$$\mathbb{E}(|L|) = 2^n \cdot \left(1 - \frac{nx}{n+1}\right)$$

where $x = (n+1)^{-1/n}$. Since this expression look cumbersome, we backtrack and work differently. Note that $p + (1-p)^{n+1} \leq p + e^{-(n+1)p}$ we minimise the latter function. That gives us $p = \frac{\log(n+1)}{n+1}$, and plugging this in gives

$$\mathbb{E}(|L|) \leq 2^n \left(\frac{1 + \log(n+1)}{n+1}\right) \leq 2^n \left(\frac{2\log n}{n}\right)$$

for all $n \geq 3$ (the last inequality is a simple exercise). By the probabilistic maxim, there then exists a set L whose size is at most $O(\frac{\log n}{n})$ fraction of the total size, which then means that the friends can achieve a success rate of $1 - \frac{2\log n}{n}$.

Remark: A concise (and low complexity) description of optimal sized desirable sets is possible for n of the form 2^k, and this is through what are known as Hamming codes. One can also prove similar results when the friends are assigned hats that may take any one of q different colors.

[2]This term is not a loose one. The Hamming distance $d_(\mathbf{x}, \mathbf{y})$ which counts the number of coordinates where $\mathbf{x}, \mathbf{y}$ differ is indeed a legitimate metric

3.6 The 1-2-3 theorem

The following question was first posed by Margulis: Given i.i.d random variables X, Y according to some distribution F, is there a constant C (independent of F; that is the important thing) such that

$$\mathbb{P}(|X - Y| \leq 2) \leq C\mathbb{P}(|X - Y| \leq 1)?$$

Note that it is far from obvious that such a $C < \infty$ must even exist. However, it is easy to see that such a C must be at least 3. Indeed, some X, Y are uniformly distributed on the even integers $\{2, 4, \ldots, 2n\}$ then it is easy to check that $\mathbb{P}(|X - Y| \leq 1) = 1/n$ and $\mathbb{P}(|X - Y| \leq 2) = \frac{3}{n} - \frac{2}{n^2}$. It was finally proved by Kozlov in the early 90s that the constant $C = 3$ actually works. Alon and Yuster shortly thereafter gave another proof which was simpler and had the advantage that it actually established

$$\mathbb{P}(|X - Y| \leq r) < (2r - 1)\mathbb{P}(|X - Y| \leq 1)$$

for any positive integer $r \geq 2$ which is also the best possible constant one can have for this inequality. We shall only show the weaker inequality with $\leq$ instead of the strict inequality. We shall later give mention briefly how one can improve the inequality to the strict inequality though we will not go over all the details.

Proof. The starting point for this investigation is based on one of the main tenets of Statistics: One can estimate (well enough) parametric information about a distribution from (large) finite samples from the same. In other words, if we wish to get more information about the unknown F, we could instead draw a large i.i.d sample $X_1, X_2, \ldots, X_m$ for a suitably large m and then the sample percentiles give information about F with high probability. This is in fact the basic premise of Non-parametric inference theory.

So, suppose we did draw such a large sample. Then a 'good' estimate for $\mathbb{P}(|X - Y| \leq 1)$ would be the ratio

$$\frac{\big|\{(i,j) : |X_i - X_j| \leq 1\}\big|}{m^2}.$$

A similar ratio, namely,

$$\frac{\big|\{(i,j) : |X_i - X_j| \leq r\}\big|}{m^2}$$

should give a 'good' estimate for $\mathbb{P}(|X - Y| \leq r)$. This suggests the following question.

Question 8. *Suppose $T = (x_1, x_2, \ldots, x_m)$ is a sequence of (not necessarily distinct) reals, and $T_r := \{(i,j) : |x_i - x_j| \leq r\}$. Is it true that $|T_r| \leq (2r - 1)|T_1|$?*

If this were false for some real sequence, one can consider F appropriately on the numbers in this sequence and maybe force a contradiction to the stated theorem. Thus,

it behooves us to consider this (combinatorial) question posed above.

Let us try to prove the above by induction on m. For $m = 1$ there is nothing to prove. In fact, for $m = 1$ one in fact has strict inequality. So suppose we have (strict) inequality for $r - 1$ and we wish to prove the same for r.

Fix an i and let $T' = T \setminus \{x_i\}$. Consider the interval $I := [x_i - 1, x_i + 1]$ and let $S_I = \{j | x_j \in I\}$, and let $|S_I| = s$. Then it is easy to see that

$$|T_1| = |T_1'| + (2s - 1).$$

Now in order to estimate $|T_r|$, note that we need to estimate the number of pairs (j, i) such that $|x_i - x_j| \leq r$. Suppose i was chosen such that $|S_I|$ is maximum among all choices for x_i. Then observe that if we partition

$$[x_i - r, x_i + r] = [x_i - r, x_i - (r-1)) \cdots, [x_i - 2, x_1 - 1), [\mathbf{x_i - 1, x_i + 1}], (x_i + 1, x_i + 2], \cdots, (x_i + (r-1), x_i + r]$$

as indicated above, then in each of the intervals in this partition there are at most s values of j such that x_j is in that corresponding interval. This follows by the maximality assumption about x_i.

In fact, a moment's thought suggests a way in which this estimate can be improved. Indeed, if we also choose x_i to be the largest among all x_k that satisfy the previous criterion, then note that each of the intervals $(x_i + l, x_i + (l + 1)]$ can in fact contain at most $s - 1$ x_j's. Thus it follows (by induction) that

$$|T_r| \leq |T_r'| + 2(r-1)s + (2s-1) + 2(r-1)(s-1) < (2r-1)|T_1'| + (2r-1)(2s-1) = (2r-1)|T_1|.$$

This completes the induction and answers the question above, in the affirmative, with strict inequality.

Now, we are almost through. Suppose we do sample i.i.d observations $X_1, X_2, \ldots, X_m$ from the distribution F, and define the random variables $T_1 := \big|\{(i, j) : |X_i - X_j| \leq 1\}\big|$ and $T_r := \big|\{(i, j) : |X_i - X_j| \leq r\}\big|$, then note that

$$\mathbb{E}(T_1) = \sum_{i \neq j} \mathbb{P}(|X_i - X_j| \leq 1) + m = (m^2 - m)p_1 + m,$$

where $p_1 = \mathbb{P}(|X_i - X_j| \leq 1)$. Similarly, we have

$$\mathbb{E}(T_r) = (m^2 - m)p_r + m$$

with $p_r = \mathbb{P}(|X_i - X_j| \leq r)$. By the inequality

$$T_r < (2r - 1)T_1$$

we have

$$(m^2 - m)p_r + m = \mathbb{E}(T_r) < (2r - 1)\mathbb{E}(T_1) = (2r - 1)((m^2 - m)p_1 + m).$$

This simplifies to $p_r < (2r - 1)p_1 + \frac{2r-2}{m-1}$. As $m \to \infty$, the desired result follows. ■

As mentioned at the beginning, Alon and Yuster in fact obtain strict inequality. We shall briefly describe how they go about achieving that. They first prove that if $p_r = (2r-1)p_1$, then if we define $p_r(a) = \mathbb{P}(|X-a| \leq r)$ there exists some $a \in \mathbb{R}$ such that $p_r(a) > (2r-1)p_1(a)$. Once this is achieved, one can tweak the distribution F as follows.

Let X be a random variable that draws according to the distribution F with probability $1-\alpha$ and picks the number a (the one satisfying the inequality $p_r(a) > (2r-1)p_1(a)$) with probability α for a suitable α. Let us call this distribution G. Then from what we just proved above, it follows that $p_r^{(G)} \leq (2r-1)p_1^{(G)}$. Here $p_r^{(G)}$ denotes the probability $p_r = \mathbb{P}(|X-Y| \leq r)$ if X, Y are picked i.i.d from the distribution G instead. However, if we calculate these terms, we see that $p_r^{(G)} = p_r(1-\alpha)^2 + 2\alpha(1-\alpha)p_r(a) + \alpha^2$, so the above inequality reads

$$p_r(1-\alpha)^2 + 2\alpha(1-\alpha)p_r(a) + \alpha^2 \leq (2r-1)\left(p_1(1-\alpha)^2 + 2\alpha(1-\alpha)p_1(a) + \alpha^2\right)$$

which holds if and only if

$$\alpha \geq \frac{\beta}{r-1+\beta} \text{ where } \beta = p_r(a) - (2r-1)p_1(a) > 0.$$

But since choosing α is our prerogative, picking α smaller than this bound yields a contradiction and completes the proof.

As we complete this chapter, we leave the reader with an important caveat. The power of the probabilistic method becomes more evident *only when one runs out of ideas*, so to speak. Where one has a deterministic argument that seems to work, the probabilistic method is *not to be unsheathed as it will be suboptimal*. To illustrate this point, suppose $[n] := \{1, 2, \ldots, n\}$, and suppose we wish to show that for $n \leq m$, there is an injective function from $[n]$ to $[m]$. We pick a random function ϕ which maps for each $x \in [n]$ a uniformly random member of $[m]$ as its image, and independently for $x \in [n]$. For a fixed pair $x, y \in [n]$, the probability that x and y are mapped to the same element in $[m]$ by ϕ is $1/m$. Hence the union bound tells us that if $\frac{\binom{n}{2}}{m} < 1$ then with positive probability, the random function ϕ is injective. In other words, our methods of this chapter only testify towards the existence of an injective function for $m \geq cn^2$, and the suboptimality of this conclusion is evident to all. One may ask if this suboptimality is due to the union bound, or if the reason is something subtler. We shall return to this point in a later chapter.

4 Managing Expectations and the Markov bound

If the expected value of a *non-negative* random variable is small, then the random variable is not very likely to take large values.

As we saw in the latter half of the preceding chapter, Expectation of a random variable is one of the most useful tools within the Probabilistic paradigm. One of the reasons the expectation appears a very natural tool is because most combinatorially relevant functions can be regarded as random variables that tend to get robust with a larger population, so the expected value gives an idea of where a 'typical' observation of the random variable lies and that is often a very useful start. For instance, suppose our random variable in question counts the number of undesirable events. Having its expected value less than one is good for our cause. But even if that is not the case, having a low value of its expectation has useful consequences - it instantiates the existence of a configuration with very few undesirable events.

There is another important feature that the expectation computation provides. While the expectation tells us that the random variable in question can take small/large values relative to its expected value, it does not tell us *how likely such an outcome might be.* Here is a concrete instantiation of the same. Consider a random variable X that takes the value n^2 with probability $1/n$ and 0 with probability $1 - 1/n$ (for n large). The expected value is n and yet the random variable itself is non-zero with very low probability. Of course, this example also illustrates that one needs a *relative perspective on what large/small ought to be.* But *if the expected value of a non-negative random variable is small, then the random variable is not very likely to take large values.*

In this chapter, we shall expound upon these two principles.

4.1 Revisiting the Ramsey Number $R(n, n)$

Let us revisit the problem of the lower bounds for $R(n, n)$. As usual, color the edges of the complete graph K_n red or blue with equal probability, and independently for distinct edges. Then the expected number of monochrome copies of K_k is $m := \binom{n}{k}2^{-\binom{k}{2}+1}$. Thus there is a coloring of the edges in which there are at most m monochrome copies of K_k. Now, from each such monochrome copy, delete a vertex; then the resulting graph on $n-m$ vertices has no monochrome K_k! Thus we get $R(k, k) > n - \binom{n}{k}2^{-\binom{k}{2}+1}$.

Now, to see if this improves upon our earlier bound, we need to do some calculus. If $m = n/2$, then we get $R(k, k) > n/2$. Some routine asymptotics (see the chapter on asymptotics for the detail) give us

$$R(k, k) > (1 + o(1))\frac{k}{e}2^{k/2}$$

for large k.

4.2 An approximate form of Caratheodory's theorem

One of the most powerful consequences of the probabilistic paradigm is that it allows for an 'approximate' version that allows for more efficient (albeit randomized) algorithms, and here we present one such instance.

The Caratheodory theorem is a well known and fundamental theorem in discrete geometry. It states: Every point in the convex hull of a set $T \subset \mathbb{R}^n$ can be written as a convex combination[1] of at most $n+1$ points of T, and the number $n+1$ is best possible. Now, consider an *approximate* version of this statement: Suppose we wish to approximate a point in the convex hull of T. Do we still need close to n points to be able to make the approximation? More precisely, suppose x is a point in the convex hull. How small a number k of points of T are needed so that some convex combination of those points is close to x?

Theorem 9. *Suppose T is a subset of $\mathbb{R}^n$ of bounded diameter[2] d, and suppose $\varepsilon > 0$. Then one can find points $x_1, \ldots, x_t$ with $t \leq \frac{d^2}{\varepsilon^2})$ such that there is a convex combination y of the x_i such that*

$$\|x - y\| \leq \varepsilon$$

where the norm is the usual L^2-norm.

The interesting feature of this theorem is that the ambient dimension n does not feature at all! The only thing that matters is how good an approximation we are hoping

[1] A convex combination of $x_1, \ldots, x_t$ is a sum of the form $\lambda_1 x_1 + \cdots + \lambda_t x_t$ with $\lambda_i \geq 0$ for all i and $\sum_i \lambda_i = 1$.

[2] The diameter of a set A is defined as the supremum of the distances between pairs of points on A.

to find. This is again a feature that appears on more than one occasion when one encounters randomized methods.

Proof. Without loss of generality, by translating T, we assume that the radius of T is at most 1, i.e., for all $x \in T$, we assume that $\|x\| \leq 1$. Let x be a point in the convex hull of T. By Caratheodory's theorem, there exist $x_1, \ldots, x_m$ with $m \leq n+1$ such that $x = \lambda_i x_i$ where $\lambda_i > 0$ and $\sum_i \lambda_i = 1$. The λ_i summing to one suggests a natural random variable. Let y be the (vector-valued) random variable with $\mathbb{P}(y = x_i) = \lambda_i$ for $i = 1, \ldots, m$. Then $\mathbb{E}(y) = \sum_i \lambda_i x_i = x$. Here the expectation is taken coordinate-wise.

Picking from the general principle that an average of i.i.d (independent and identically distributed) random variables converges to its expected value, it is somewhat natural to consider $y_1, \ldots, y_k$ independent and distributed as y above, and look at the average $z := \frac{1}{k}(y_1 + \cdots + y_k)$. Clearly, $\mathbb{E}(z) = x$. To see how well this fares, let us compute (or estimate) $\mathbb{E}(\|z - x\|^2$. For the random variable y,

$$\begin{aligned}
\mathbb{E}(\|y - x\|^2) &= \mathbb{E}\left(\|y\|^2 + \|x\|^2 - 2\langle y, x\rangle\right) \\
&= \mathbb{E}\left(\|y\|^2\right) - \|x\|^2 \\
&= \sum_{i=1}^{m} \lambda_i \left(\|x_i\|^2 - \|x\|^2\right) \\
&\leq 1.
\end{aligned}$$

Set $z_i = y_i - x$ so that $z - y = \frac{1}{k}\sum_{i=1}^{k} z_i$. Then

$$\begin{aligned}
\mathbb{E}\|z - x\|^2 &= \frac{\mathbb{E}\|\sum_i z_i\|^2}{k^2} \\
&= \frac{1}{k^2}\left(\sum_{i=1}^{k} \mathbb{E}\|z_i\|^2 + 2\sum_{i<j} \mathbb{E}\langle z_i, z_j\rangle\right) \\
&\leq \frac{1}{k}
\end{aligned}$$

where the last inequality comes from the preceding calculation and the fact that z_i, z_j are independent which consequently gives $\mathbb{E}\langle z_i, z_j\rangle = 0$ for all $i < j$. In particular, this computation tells us that there exist k vectors y_i for which $\|z - x\|^2 \leq 1/k$ and z as above. The rest is a routine consequence. ■

Remark: The theme of approximations is closely tied with the probabilistic paradigm, and we will encounter the motif several times in the book.

4.3 Graphs with many edges and high girth

The *girth* of a graph G is the size of its smallest cycle (should a cycle exist) and if the graph is acyclic, then its girth is infinite. It is both intuitively and mathematically clear

that as the number of edges in a graph increases (proportional to the total possible number of edges) then its girth can go down dramatically. So, for a fixed parameter k, the following extremal problem is both natural and of great interest to the extremal combinatorist: What is the maximum possible number of edges in a graph on n vertices with girth at least k?

The following simple argument gives an upper bound. Suppose the graph G has minimum degree at most d. Set $\ell = \frac{k-2}{2}$. If the girth of G is at least k, then for any vertex v, the subgraph induced on the *ℓ-fold neighborhood centered at v*, i.e., the set of vertices at a distance of at most ℓ from v, is a tree This follows since if this was not a tree, then there is a cycle contained in this graph. However, since any vertex w is at a distance of at most ℓ from v, the size of the cycle is at most $2\ell + 1 < k$ contrary to the assumption. Since each vertex has degree at least d, this induced subgraph has at least

$$1 + d + d(d-1) + \cdots + d(d-1)^{\ell-1} = 1 + \frac{d\left((d-1)^{\ell} - 1\right)}{d-2}$$

vertices.

Now for a given graph G with *average* degree $\overline{d}$, note that if we remove a vertex of degree at most k (for some k - we'll see what k to set) then the modified graph has average degree at least $\frac{n\overline{d}-2k}{n-1}$. If we set $k = \overline{d}2$ then this last expression is at least $\overline{d}$. In other words, if we delete a vertex of degree at most $\overline{d}/2$ then the average degree of the graph does not decrease in this process. Consequently, this process must eventually terminate and when it does, every vertex has degree at least $\overline{d}/2$. It is a simple exercise to show that if the average degree of a graph is at least $Cn^{2/(k-2)}$ for a suitably large C then G must have a cycle of size at most $k-1$. In other words, the maximum number of edges in a graph with girth at least k is $O(n^{1+\frac{2}{k-2}})$.

A lower bound was established by Erdős.

Theorem 10. *For a given integer $k \geq 4$ and n somewhat large, there exist graphs on n vertices with girth at least k with $\Omega(n^{1+\frac{1}{k-2}})$ edges.*

Proof. To establish a lower bound, one needs to construct a graph with girth at least k and as many edges as possible. Like in the Ramsey problem, the small cases for k ($k = 3, 4$ for starters) are well studied; indeed one knows the best possible bounds in these cases. But in the general case, the specificity of the examples in the smaller cases makes it harder to generalize. And so Erdős did what came to him naturally; he picked the graph at random.

Let us construct a random graph where each edge is chosen independently with probability p where p (as in a previous example) will be determined later. The 'bad instances' here are incidences of small cycles. Indeed, for $3 \leq t \leq k-1$ let N_t denote the number

of t-cycles in G. Then

$$\mathbb{E}(N_t) = \frac{n(n-1)\cdots(n-t+1)}{2t}p^t < \frac{(np)^t}{2t}$$

since every cyclic permutation of size t counts a particular t-cycle exactly $2t$ times - the first vertex is picked in one of t ways, and the orientation in one of two possible ways. This gives

$$\mathbb{E}(\sum_{3\leq t\leq k-1} N_t) \leq \frac{(np)^3 + \cdots + (np)^{k-1}}{6} \leq \frac{(np)^{k-1}}{3}.$$

On the other hand, the expected number of edges of G is $\mathbb{E}(e(G)) = \binom{n}{2}\frac{pn^2}{3}$ for n large enough.

The key insight again of Erdős was: If the number of small cycles is not that large, say, it is at most half the total number of edges, then one may throw away one edge from each of the small cycles, thus eliminating all small cycles, and yet, retaining at least *half the total number of edges*. This suggests, that if

$$(np)^{k-1} \leq \frac{n^2 p}{2}$$

then by linearity of expectation[3]

$$\mathbb{E}\left(e(G) - \sum_{t<k} N_t\right) \geq \frac{\mathbb{E}(e(G))}{2}$$

so that there is an instance with this inequality holding. That gives us a lower bound on the number of edges of a graph with girth at least k.

The computation now is straightforward. We leave it to the reader to see that the inequality we have forced gives us $p = \frac{c}{n^{(k-3)/(k-2)}}$ (for a small constant $c > 0$) which in turn gives us a bound of $e(G) = \Omega(n^{1+\frac{1}{k-2}})$. ■

Remark: As it turns out, the random construction here is not best possible, and the considered opinion of the experts in extremal combinatorics, is that the exponent that appears in the upper bound is the truth. However that remains an open problem. The best known bound is a remarkable algebraic construction by Lazebnik, Ustimenko and Woldar [21] which gives a lower bound of $\Omega(n^{1+\frac{4}{3(k-2)-+\varepsilon}})$ which, in terms of the exponent of n is 'halfway' between the randomized construction and the simple upper bound. This again reinforces the caveat: *If one can, one should strive for non-random constructions.*

[3] Again, the linearity is key here.

4.4 List Chromatic Number and minimum degree

The list chromatic number is a notion introduced by Erdős, Rubin and Taylor in their seminal paper that sought to address what was called the 'Dinitz problem'. This variant of the usual chromatic number goes as follows. For a graph G, let $\mathcal{L} = \{L_v | v \in V(G)\}$ be a collection of subsets of some set $\mathcal{C}$ indexed by the vertices of G. These are to be interpreted as lists of colors assigned to each vertex. An *$\mathcal{L}$-coloring* of G is an assignment of an element $\chi(v) \in L_v$ for each $v \in V$ such that if u and v are adjacent vertices in G, then $\chi(u) \neq \chi(v)$. In the parlance of colorings, this is a choice of color assignments to each vertex such that no two adjacent vertices are assigned the same color. The *list chromatic number* of G, denoted $\chi_l(G)$, is the smallest k such that for *any* family $\mathcal{L}$ with $|L_v| \geq k$ for all v, G is $\mathcal{L}$-colorable. It is not hard to see that this is well defined, and in fact, the usual chromatic number of G corresponds to the case where all the vertex lists are identical. The next result shows that the reverse inequality need not hold.

The list chromatic number is a very interesting invariant for a host of reasons. One natural way to motivate this notion is the following. Suppose we attempt to properly color the vertices of a graph using colors, say, $1, \ldots, k$ and to suppose we are given a partial coloring. For each uncolored vertex v, let D_v denote the set of colors that appear among any of its neighbors from the partial coloring. Then the partial coloring extends to a proper k coloring of the graph if and only if the induced graph on the remaining uncolored vertices is $\mathcal{L}$-colorable where $L_v := [k] \setminus D_v$. So in that sense, list colorings arise quite naturally in connection with proper colorings.

As was observed in the seminal paper of Erdős, Rubin, and Taylor, there are bipartite graphs with arbitrarily large list chromatic number.

Theorem 11. *(Erdős, Rubin, Taylor)* $\chi_l(K_{n,n}) > k$ *if* $n \geq \binom{2k-1}{k}$.

Proof. We wish to show there is some $\mathcal{L} = \{L_v | v \in V(G)\}$ with $|L_v| = k$ for each $v \in V(G)$ such that $K_{n,n}$ is not $\mathcal{L}$-colorable. Let A and B denote the two partition classes of $K_{n,n}$, i.e., the two sets of vertices determined by the natural division of the complete bipartite graph $K_{n,n}$ into two independent subgraphs.

Now we construct $\mathcal{L}$. Take the set of all colors from which we can construct L_v's to be $\{1, 2, ..., 2k-1\}$. Since $n \geq \binom{2k-1}{k}$, which is the number of possible k-subsets of $\{1, 2, ..., 2k-1\}$, we can choose our L_v's for the v's in B so that each k-subset of $\{1, 2, ..., 2k-1\}$ is L_v for some $v \in B$, and similarly we choose lists for vertices of A.

If S is the set of all colors that appear in some L_v with $v \in B$, then S intersects every k-element subset of $\{1, 2, ..., 2k-1\}$. Then we must have that $|S| \geq k$ (since otherwise its complement has size $\geq k$ and thus contains a subset of size k disjoint from S). But then since $|S| \geq k$, by choice of lists there exists $a \in A$ with $L_a \subset S$. Since a is adjacent to every vertex in B, so no $\mathcal{L}$-coloring is possible. ∎

Another interesting feature of the list chromatic number is the following result due to Alon.

Theorem 12. *(Alon) Suppose d denotes the minimum degree of G. Then*

$$\chi_l(G) = \Omega\left(\frac{\log d}{\log\log d}\right).$$

This is quite at variance with the usual chromatic number since one has bipartite graphs with arbitrarily large minimum degree.

Proof. If the result holds, then it also holds in the case the chromatic number is two (that is the first nontrivial case), so let us first assume that the graph is bipartite with partition classes A and B, and $|A| \geq |B|$.

We shall assume that the minimum degree is sufficiently large (for asymptotic reasons). In order to establish a lower bound of the form $\chi_l(G) > s$, we need to assign lists of size s to each vertex and ensure that from these lists, a list coloring is not possible. Suppose $\mathcal{C}$ is a set of colors from which we shall allocate lists to each of the vertices of G. Without loss of generality, let $\mathcal{C} := \{1, 2, ..., L\}$ for some L to be fixed/determined later.

How do we show that a vertex $a \in A$ cannot be colored from its list? Let us take a cue from the previous result: Suppose a vertex $a \in A$ has, among the lists assigned to its neighbors in B, all the possible s-subsets of $\mathcal{C}$. Consider a choice of colors assigned to the vertices of B from their respective lists, and let W be the set of colors that are witnessed by this choice. Note that since the neighbors of a witness all possible s-subsets of $\mathcal{C}$, $W \cap S \neq$ for all $S \subset \mathcal{C}$ of size s, so that in particular, $|W| \geq L - s + 1$. If this choice extends successfully to a choice for a, then L_a must contain an element from a very small set, viz., $\mathcal{C} \setminus W$, which has at most $s-1$ colors of $\mathcal{C}$. Now, if there are several such vertices $a \in A$ (i.e., that witness every s-subset as the list of one of its neighbors) then this same criterion must be met by each of these vertices. And that is not very likely to happen if *we were to allot random lists to the vertices of A*! This potentially sets up a contradiction.

Let us set this in motion. Call a vertex $a \in A$ *critical* if among its neighbors, all possible s-subsets of $\mathcal{C}$ appear. To achieve this, assign for each $b \in B$, the set L_b to be an s-subset of $\mathcal{C}$ uniformly at random and independently over different vertices. Then the probability that a is not critical is equal to the probability that there exists some s-subset T of $\mathcal{C}$ such that no neighbor of a is assigned T as its list. Since there are $\binom{L}{s}$ possible T's it follows by the union bound that

$$P(a \text{ is not critical}) \leq \binom{L}{s}\left(1 - \frac{1}{\binom{L}{s}}\right)^d \leq \binom{L}{s} e^{-d/\binom{L}{s}}.$$

Now assume that $d \gg \binom{L}{s}$. Then by the above, $P(a \text{ is not critical}) < \frac{1}{2}$. So if N denotes the number of critical vertices of A,

$$\mathbb{E}(N) = \sum_{a \in A} \mathbb{P}(a \text{ is critical}) > \frac{|A|}{2}.$$

Thus there exists an assignment of lists for vertices in B, $\{L_v | v \in B\}$, such that the number of critical vertices is greater than $\frac{|A|}{2}$. Fix these choices for the lists for the vertices of B.

Fix a color palette w from these assigned lists, i.e., a choice of an element each from the collection $\{L_v | v \in B\}$. Denote as $W = W(w)$ the set of colors that appear among the vertices on B from the palette w.

Since there exists critical $a \in A$, W has nonempty intersection with all s-subsets of $[L]$, so $|W| \geq L - s + 1$. If an extension of w to a coloring to a exists for a critical vertex a, then as we observed earlier, exists, $L_a \cap \overline{W} \neq \emptyset$.

Since we haven't yet dealt with the color lists for A, let us pick color lists for the vertices of A uniformly at random from the s-subsets of $\mathcal{C}$. Then for a critical $a \in A$

$$P(w \text{ extends to } a \text{ exists}) \leq \frac{(s-1)\binom{L-1}{s-1}}{\binom{L}{s}} < \frac{s^2}{L}.$$

For an extension of w to G to exist, we need an extension of w to all critical vertices of A. Since there are $s^{|B|}$ possible w's and the number of critical vertices is greater than $\frac{|A|}{2}$, we have(since the color lists for the vertices of a are picked independently)

$$P(\text{an extension to a coloring of } G \text{ exists}) \leq s^{|B|} \left(\frac{s^2}{L}\right)^{|A|/2} \leq \left(s\left(\frac{s^2}{L}\right)^{\frac{1}{2}}\right)^{|B|}$$

which is less than 1 if $s\sqrt{\frac{s^2}{L}} < 1$, or equivalently, if $L > s^4$, so set $L = 2s^4$.. Recall the assumption made earlier that $d \gg \binom{L}{s}$. We needed this to make $\binom{L}{s} e^{-d/\binom{L}{s}} < \frac{1}{2}$, which is equivalent to $d > \binom{L}{s} \log(2\binom{L}{s})$.

In summary, if

$$d \geq 4\binom{2s^4}{s} \log\left(2\binom{2s^4}{s}\right),$$

then there is a collection of lists $\mathcal{L} = \{L_v | v \in G\}$ with $|\{L_v\}| = s$ for all $v \in G$ such that no $\mathcal{L}$-coloring of G exists, i.e., $\chi_l(G) > s$. Again, arriving at the lower bound as stated in the theorem is a good exercise in asymptotics. For the precise details, see the first

chapter on asymptotics.

This is all under the assumption that G was bipartite. But it is a very simple fact that every graph contains a 'large' bipartite subgraph:

Lemma 13. *For any graph G, there exists a subgraph H of G with $V(H) = V(G)$ such that H is bipartite and $d_H(v) \geq \frac{1}{2} d_G(v)$ for all $v \in V(G)$.*

To see why, consider a partition of the vertex set into two parts so that the number of crossing edges (across the partition) is maximized. It is now a straightforward observation to see that for every vertex at least half of its neighbors must be in the other part since otherwise we can move that vertex to the other part. This partition in particular produces a bipartite subgraph H as stated. This completes the proof of the lemma and the theorem as well. ■

Alon later improved his bound to $\chi_l(G) > (\frac{1}{2} - o(1)) \log d$ with $d = \delta(G)$. We shall show a proof of a slightly weaker form of this result ($\chi_l(G) \geq c \log d$ for some constant $c > 0$ in a later chapter by a different probabilistic paradigm. Alon also conjectured that $\chi_l(G) \leq O(\Delta(G))$ where $\Delta(G)$ denotes the maximum degree of G. That remains an open problem.

We now move on to the next principle outlined in the introduction of this chapter, and that is this fairly easy inequality.

Theorem 14. *(The Markov Inequality) IF X is a non-negative value random variable then*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

As seen in the example earlier, the expectation of a random variable being 'large' does not guarantee that its takes large values with high probability. But if the random variable is bounded, then it must take 'somewhat large' values with ' not-too-little' probability:

Proposition 15. *Suppose X is a non-negative values random variable and suppose $X \leq M$ with probability one. Then for $a < \mathbb{E}(X)$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X) - a}{M - a}.$$

The idea of proof is straightforward. Write

$$\begin{aligned} \mathbb{E}(X) &= \int_{x<a} X \, d\mathbb{P} + \int_{x \geq a} X \, d\mathbb{P} \\ &\leq a\mathbb{P}(X \leq a) + M(1 - \mathbb{P}(X \geq a)) \end{aligned}$$

and now the conclusion follows by a straightforward computation.

4.5 The Varnavides' averaging argument

An old conjecture due to Erdős and Turán was the following: Given $\varepsilon > 0$ and an integer $k \geq 3$, there exists $N_0 := N_0(\varepsilon, k)$ such that the following holds: If $N \geq N_0$ and $A \subset \{1, \ldots, N\}$ with $|A| \geq \varepsilon N$ then A contains a k-term arithmetic progression (k-AP for short). This conjecture came on the heels of the theorem due to van der Waerden that states that given positive integers, r, k there exists an integer $W(r, k)$ such that if $N \geq W$, then any r-coloring of the integers in $[N] := \{1, \ldots, N\}$ necessarily contains a monochromatic k-AP. The Erdős-Turán conjecture basically captures the intuitive idea that van der Waerden's theorem holds because it does so on *the most popular color class*. This conjecture was settled in the affirmative for $k = 3$ by Roth, and then later in its full generality by Szemerédi.

What we are after (following Varnavides) is a generalization of this result. The statement we aim to prove is that, if N is sufficiently large, and $|A| \geq \varepsilon N$ then A in fact contains as many k-APs as there can be (upto a constant multiple):

Theorem 16. *(Varnavides) Given $\varepsilon > 0$ and $k \geq 3$, there exists $\delta := \delta(\varepsilon, k)$ such that the following holds. Any subset $A \subset [N]$ of size at least εN contains at least δN^2 APs of length k.*

Note that the total number of possible k term APs in $[N]$ is determined by specifying the first term, and the common difference, so there are at most N^2 k-APs in all. Thus, this theorem is best possible up to the constant.

Proof. By Szemerédi's result, we know that there is an $N_0 := N_0(\varepsilon, k)$ such that any subset $A \subset [N_0]$ of size at least εN_0 contains at least one k-AP. The simplest thing one can imagine doing is, cutting the set $[N]$ into linear chunks of length N_0; clearly at least one of these chunks must meet A in at least an ε proportion of its size. That unfortunately does not give us anything new. But one thing is does suggest is, the number of such chunks that have a reasonable proportion of A in them will each give us one distinct k-AP.

But this breaking into chunks is a little too wasteful. Every AP of size N_0 is again a model for the set $[N_0]$ so one might want to consider all possible APs of length N_0 and see how many among them meet A in a significantly large portion. Of course, the same k-AP might be a part of several such N_0-APs, so there is a double counting issue to sort out anyway. But that immediately suggests the following.

Pick an N_0-AP at random, i.e., pick x_0, d uniformly and independently from $[N]$ with $d \neq 0$ and let $K := K(x_0, d) = \{x_0, x_0 + d, \ldots, x_0 + (N_0 - 1)d\}$. One problem this immediately poses is that not all possible such pairs give rise to $K \subset [N]$. To overcome this nuisance, let us work in $\mathbb{Z}/N\mathbb{Z}$. The relevant random variable now is $|A \cap K|$. To compute the expectation using the linearity of expectation, we need to compute $\mathbb{P}(a \in K)$ for an arbitrary $a \in [N]$. To count the number of pairs (x_0, d) such that $K(x_0, d)$ contains a,

observe that if $a = x_0$ is the first term, then all possible choices for d count such valid K, otherwise, there are $N-1$ choices for x_0, and for each of those, if a is the i^{th} term of $K(x_0, d)$ (with $N_0 - 1$ choices for i) this determines d uniquely, *provided we can solve the corresponding linear equation* $a = x_0 + id$ *for* d. Again, this makes things messy, but as we have observed earlier, this works provided N is prime.

So, let us start again. Instead of working in $\mathbb{Z}/N\mathbb{Z}$, pick a prime $p \in (2N, 4N)$ and let us then work in $\mathbb{Z}/p\mathbb{Z}$. We choose $p > 2N$ since that ensures that addition of two elements in $\{1, \ldots, N\}$ is the same as addition in $\mathbb{Z}/p\mathbb{Z}$. We also don't want p to be too large because we need to estimate p in terms of N. Now pick x_0, d uniformly from $\mathbb{Z}/p\mathbb{Z}$ as outlined before, and let $K := K(x_0, d)$. Then

$$\mathbb{E}(|K \cap A| = \sum_{a \in A} \mathbb{P}(a \in K) = \frac{|A|}{p} N_0 \geq \frac{\varepsilon}{4} N_0$$

by the assumption on the size of A. We would ideally like it to be the case that $|K \cap A|$ takes somewhat large values with not-too-small probability, and by Proposition 15

$$\mathbb{P}\left(|K \cap A| \geq \frac{\varepsilon}{8} N_0\right) > \frac{\varepsilon}{8}$$

since $|K \cap A| \leq N_0$. This suggests the following: Let $N_0 = N_0(\varepsilon/8, k)$ from Szemerédi's theorem. Then it follows by everything seen above that with probability at least $\varepsilon/8$, $K \cap A$ contains a k-AP. But on the other hand, if $\mathcal{A}$ is the set of all k-APS contained in A, then

$$\begin{aligned} \mathbb{P}(K \cap A \text{ contains a member of } \mathcal{A}) &\leq \sum_{P \in \mathcal{A}} \mathbb{P}(P \subset K) \\ &\leq |\mathcal{A}| \frac{N_0(N_0 - 1)}{N(N-1)} \end{aligned}$$

since to determine if $K(x_0, d)$ contains P, we have at most $N_0(N_0 - 1)$ choices for determining the first and second elements of P in K. Rearranging terms, this gives

$$|\mathcal{A}| \geq \frac{\varepsilon}{16 N_0(N_0 - 1)} N^2$$

and that completes the proof. ■

Remark: The constants $N_0(\varepsilon, k)$ that Szemerédi's proof offer are extremely large and are of a tower type.

4.6 A conjecture of Daykin-Erdős and its resolution

Suppose $\mathcal{H} \subset \mathcal{P}([n])$ is a hypergraph. One can construct a graph $G_{\mathcal{H}}$ with vertex set $E(\mathcal{H})$, and $E \neq F$ are adjacent in $G_{\mathcal{H}}$ iff $E \cap F = \emptyset$.

Why would one define this particular graph? This is actually one of the most well-studied instances of a graph arising naturally from hypergraphs. The Kneser graphs are instances of this when the underlying hypergraph is the complete uniform hypergraph of order r, i.e., $V(\mathcal{H}) = [n]$ and $E(\mathcal{H}) = \binom{[n]}{r}$ for some $r < n/2$. One of the other motivations for studying this graph comes from the problem of explicit constructions for Ramsey graphs: We say that a graph on n vertices is *k-Ramsey* (if the k is implicitly clear, we simply say Ramsey graph) if it neither contains an independent set[4] nor a clique[5] of order k. Note that if the edges of a k-Ramsey G are colored red, and the remaining edges of K_n are colored blue then by the definition of G being k-Ramsey, this coloring does not contain a monochromatic K_k.

As seen earlier, the first explicit construction of a k-Ramsey graph due to Nágy was by considering the graph $G_{\mathcal{H}}$ where $\mathcal{H}$ was the complete uniform hypergraph of order 3. As seen before, Erdős proved that $R(k,k) > \Omega(k2^{k/2})$ but his proof did not provide an explicit deterministic construction. This also suggests the following question: Suppose $e(\mathcal{H}) = 2^{(\frac{1}{2}+\delta)n}$ for some $\delta > 0$. Is there a hypergraph $\mathcal{H}$ on n vertices such that the graph $G_{\mathcal{H}}$ is n-Ramsey? If true, this would in one stroke improve upon the probabilistic lower bound, and also provide an explicit construction for n-Ramsey graphs.

While this would indeed be nice, it seems like asking for too much. And to see why that would be the case, suppose G is a Ramsey graph on n vertices, i.e., suppose G contains neither an independent set, nor a clique of order $2\log_2 n$ (Why?!). We claim that $e(G_n)$ must be rather large. Indeed, a celebrated theorem in extremal graph theory due to Turán (in an alternate formulation) states that a graph on n vertices admits an independent subset of size at least $\frac{n}{d+1}$, where d is the average degree of the vertices of the graph. If the Ramsey graph G_n satisfies $e(G_n) \leq cn^{2-\delta}$ for some constants $c, \delta > 0$ then by Turán's theorem, [6]$\alpha(G_n) \geq \frac{n}{d+1} = \Omega(n^{\delta})$, so such graphs could not be Ramsey.

Since examples of Ramsey graphs of size $2^{(\frac{1}{2}+\delta)n}$ seemed difficult to contruct (they still are!) this line of argument possibly convinced Daykin and Erdős to conjecture the following:

Conjecture 17 (Daykin-Erdős)**.** *If $|\mathcal{H}| = m = 2^{(\frac{1}{2}+\delta)n}$, then*

$$d(\mathcal{H}) := \#\{\{E,F\} \in \mathcal{H} \mid E \cap F = \emptyset\} = o(m^2).$$

Note that if $m = 2^{n/2}$ then in fact there do exist hypergraphs $\mathcal{H}$ for which the graph $G_{\mathcal{H}}$ are dense (though not Ramsey graphs). For instance, take the set $[n]$ and partition it into two sets A, B of size $n/2$ each, and consider $\mathcal{H}$ to consist of all subsets of A along with all subsets of B. Since A, B are disjoint, $G_{\mathcal{H}}$ has all edges of the type (E, F) where

[4] A subset S of vertices is called independent if no two vertices of S are adjacent.

[5] A clique in a graph is a set of pairwise adjacent vertices.

[6] $\alpha(G)$ denotes the size of a largest independent subset of vertices in G,

$A \subset A, F \subset F$. The conjecture of Daykin and Erdős says that this cannot be improved upon if the exponent were strictly greater than $1/2$.

The Daykin-Erdős conjecture was settled by Alon and Füredi in 1985.

Theorem 18. *(Alon-Füredi) Suppose $0 < \delta < 1$ is fixed, and suppose n is sufficiently large. If $|\mathcal{H}| = m = 2^{(1/2+\delta)n}$ then $d(\mathcal{H}) < cm^{2-\delta^2/2}$ for some positive constant c.*

Proof. Let us see how we could go about this. If the graph $G_{\mathcal{H}}$ is dense, one should expect to find two large disjoint subsets $\mathcal{S}, \mathcal{T}$ of $V(G_{\mathcal{H}})$ which constitute a dense pair, i.e., one ought to expect to see lots of edges between $\mathcal{S}$ and $\mathcal{T}$. If this pair witnesses all possible edges, one has

$$\left(\bigcup_{S\in\mathcal{S}} S\right) \bigcap \left(\bigcup_{T\in\mathcal{T}} T\right) = \emptyset.$$

So if the pair $\mathcal{S}, \mathcal{T}$ constitute a dense pair, it would not be a stretch to expect that $\bigcup_{S\in\mathcal{S}} S$ and $\bigcup_{T\in\mathcal{T}} T$ are *almost* disjoint from each other. But if these sets are also 'large' subsets of $[n]$, this appears unlikely and would probably give a contradiction.

To see if we can pull this off, let us try something simpler first: Suppose there exists an $\mathcal{S}$ such that $A(\mathcal{S}) := \bigcup_{S\in\mathcal{S}} S$ is large, and further, that the number of sets $E \in \mathcal{H}$ satisfying $E \cap A(\mathcal{S}) = \emptyset$ is also large. For the sake of simplicity, suppose $|A(\mathcal{S})| > \frac{n}{2}$. Then

$$\#\{E \subset \mathcal{H} \,|\, E \cap A(\mathcal{S}) = \emptyset\} < 2^{n/2}$$

so if there exists $\mathcal{S}$ such that

- $A(\mathcal{S}) = \bigcup_{S\in\mathcal{S}} S$ satisfies $|A(\mathcal{S})| > \frac{n}{2}$ (which is likely), *and*
- $\#\{E \in \mathcal{H} \,|\, E \cap A(\mathcal{S}) = \emptyset\} \geq 2^{n/2}$,

then we have a contradiction.

Let us begin formally now. We seek a collection $\mathcal{S}$ with $A(\mathcal{S})$ being very large. Since we are bereft of specific choices, pick $S_1, S_2, \cdots, S_t \in \mathcal{H}$ uniformly and independently for some t that shall be determined later. If $A(\mathcal{S})$ has at most $n/2$ elements, then there exists $T \subset [n]$ such that $|T| = n/2$ and each $S_i \subset T$. Fix such a choice for T.

$$\mathbb{P}(S_1 \subset T) = \frac{\#\{E \in \mathcal{H} \,|\, E \subset T\}}{|\mathcal{H}|} = \frac{\#\{E \in \mathcal{H} \,|\, E \subset T\}}{2^{(\frac{1}{2}+\delta)n}} \leq \frac{2^{\frac{n}{2}}}{2^{(\frac{1}{2}+\delta)n}} = \frac{1}{2^{\delta n}}.$$

Therefore by the union bound,

$$\mathbb{P}(|A(\mathcal{S})| \leq n/2) \leq \binom{n}{n/2}\left(\frac{1}{2^{\delta n}}\right)^t < \frac{2^n}{2^{t\delta n}} = \frac{1}{2^{(t\delta-1)n}}.$$

Thus, to ensure that this is a low probability event, we need $t\delta - 1 > 0$, or equivalently, $t > \frac{1}{\delta}$.

For the second part, we want $X := \#\{E \in \mathcal{H} \mid E \cap A(\mathcal{S}) = \emptyset\}$ to be at least $2^{n/2}$. Writing $X = \sum_{E\in\mathcal{H}} \mathbb{1}_{\{E\cap A(\mathcal{S})=\emptyset\}}$ we have

$$\mathbb{E}(X) = \sum_{E\in\mathcal{H}} \mathbb{P}(E \cap A(\mathcal{S}) = \emptyset).$$

Fix $E \in \mathcal{H}$.

$$\mathbb{P}(E \cap A(\mathcal{S}) = \emptyset) = \mathbb{P}(E \cap S_i = \emptyset \text{ for all } i = 1, \ldots, t) = \left(\frac{d(E)}{m}\right)^t$$

where $d(E)$ is the degree of E in $G_{\mathcal{H}}$. Denoting $e(G_{\mathcal{H}}) = M$, we have

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{E\in\mathcal{H}} \mathbb{P}(E \cap A(\mathcal{S}) = \emptyset) = \sum_{E\in\mathcal{H}} \left(\frac{d(E)}{m}\right)^t \\
&= \frac{1}{m^{t-1}}\left(\frac{1}{m}\sum_{E\in\mathcal{H}} (d(E))^t\right) \\
&\geq \frac{2^t M^t}{m^{2t-1}}.
\end{aligned}$$

By proposition 15 we have (setting $a = \frac{\mathbb{E}[X]}{2M}$)

$$\mathbb{P}\Big(X \geq aM\Big) \geq \frac{\frac{1}{M}\mathbb{E}[X] - \frac{\mathbb{E}[X]}{2M}}{1 - \frac{\mathcal{E}[X]}{2M}} = \frac{\frac{1}{2}\left(\frac{2^t M^{t-1}}{m^{2t-1}}\right)}{1 - \left(\frac{2^{t-1}M^{t-1}}{m^{2t-1}}\right)}$$

which gives us

$$\mathbb{P}(|A(\mathcal{S})| > \frac{n}{2}) \geq 1 - \frac{1}{2^{(t\delta-1)n}}.$$

If

$$\frac{\frac{2^t M^{t-1}}{m^{2t-1}}}{1 - \frac{2^{t-1}M^{t-1}}{m^{2t-1}}} > \frac{1}{2^{(t\delta-1)n}},$$

then both events as outlined in the sketch happen simultaneously and our contradiction is achieved. Choose $t = \frac{2}{\delta}$. If $M = cm^2$, then this forced inequality is feasible for a suitable t that depends only on δ and c. To determine the upper bound for M as in the statement of the theorem is a straightforward exercise. ■

4.7 Graphs with high girth and large chromatic number

While it is easy to ensure that a graph constructed has a high chromatic number (make a clique of that size as a subgraph), it became a considerably harder task of ensuring that the same holds if we forbid large cliques. The first such question that arose was the following:

Question 19. *Do there exist graphs with chromatic number k (for any given k) and which are also triangle free?*

This was settled with the 'Mycielski construction' in the affirmative. This led to the next natural question: What if we also forbid 4 cycles? Tutte produced a sequence of graphs with girth 6 and arbitrarily large chromatic number, but the bigger question loomed large: Do there exist graphs with arbitrarily large chromatic number and also arbitrarily large girth? It took the ingenuity of Erdős to settle this in the affirmative.

To see why this is a little surprising, note that insisting on large girth g, simply implies that for each vertex v, the induced subgraph on the set of vertices at a distance at most $g/2$ is a tree, which can be 2-colored. Yet, it is indeed conceivable that the chromatic number of the entire graph varies vastly from the chromatic number of small induced subgraphs.

This again fits the general template we have discussed. We need a graph G in which locally small induced subgraphs are trees, and yet, the graph itself has large chromatic number. A random graph appears a sound candidate for such a possibility.

Theorem 20. *(Erdős) There are graphs with arbitrarily high girth and chromatic number.*

Proof. Let $G_{n,p}$ denote a random graph on n vertices, where each pair of vertices $\{x, y\}$ is added independently with probability p. Let n be sufficiently large. For the random graph to give us what we seek we want:

- $G_{n,p}$ will have relatively few small cycles *with reasonably high probability.*
- G has large chromatic number *with reasonably high probability.*

Fix a number ℓ, and let N_ℓ denote the number of cycles of length at most ℓ in $G_{n,p}$. As seen before,

$$\mathbb{E}(N_\ell) \leq \sum_{j=3}^{\ell} \frac{n^j p^j}{2j} \leq \frac{(np)^3}{6} \cdot \frac{(np)^{\ell-2} - 1}{(np) - 1} \leq \frac{(np)^\ell}{2}.$$

Hence by the Markov inequality,

$$Pr(|N_\ell| \geq (np)^\ell) \leq 1/2$$

in other words, with *probability at least* $1/2$ $G_{n,p}$ has at most $(np)^\ell$ cycles of size at most ℓ. This is the first step.

To show that the chromatic number of our random graphs is large, we need to understand how one might bound the chromatic number from below. Doing this directly, by working with the chromatic number itself, would be rather ponderous. But a simple observation based on the definition tells us that since each color class is an *independent set*, we have

$$\chi(G) \geq \frac{n}{\alpha(G)}$$

where $\alpha(G)$ is the independence number of the graph. How does this help?

Let us examine $\alpha(G)$ in $G_{n,p}$. We need this to be small since we want the chromatic number to be large. Then

$$\begin{aligned}\mathbb{P}(\alpha(G) \geq m) &= \mathbb{P}(\text{there exists an independent subset of } G \text{ of size } m) \\ &\leq \sum_{S \subset V, |S|=m} \mathbb{P}(\text{there are no edges inside } S) \\ &= \binom{n}{m}(1-p)^{\binom{m}{2}} < \left(n\exp(-p(m-1)/2)\right)^m.\end{aligned}$$

To get a handle on these parameters we have freely introduced, note that the term inside the parenthesis in the last inequality above can be expressed as $\exp\left(\log n - (\frac{m-1}{2}p)\right)$, so if $m = \left\lceil \frac{3\log n}{p} \right\rceil$, then the probability that $G_{n,p}$ contains an independent set of size at least m goes to zero as $n \to \infty$.

So, what do we have on our hands now? If p is chosen in some manner, and then we set $m = \left\lceil \frac{3\log n}{p} \right\rceil$ then with positive probability (certainly) we have that G has at most $9np)^\ell$ cycles of size at most ℓ AND that its independence number is at most m. Pick such a G.

As before, we shall perform some deletions to G to rid it of all small cycles. But unlike the earlier instance, if we deleted edges, we run the risk of pumping up its independence number, so this time let us delete vertices instead. The advantage is that vertex deletions result in an induced subgraph of the original graph, so its independence number remains the same.

This suggests that we set $(np)^\ell \leq n/2$, or equivalently, $p < \frac{Cn^{1/g}}{n}$ for some constant C. So, set $p = \frac{n^\lambda}{n}$ for some $\lambda \in (0, 1/\ell)$, and m as suggested. Then remove an arbitrary vertex from each small cycle from G, and call the resulting graph G'. Then G' has girth $\geq \ell$ and at least $n/2$ vertices. Finally, since deleting vertices doesn't decrease the independence number of a graph,

$$\chi(G') \geq \frac{|V(G')|}{\alpha(G')} \geq \frac{n/2}{\alpha(G)} \geq \frac{np}{6\log n} = \frac{n^\lambda}{6\log n},$$

which goes to infinity as n grows large. ■

Remark: There have been subsequently many constructive forms of this result, with the first one by Lovász, and then subsequently by many others. Many of those constructions actually construct hypergraphs with the same property. The nicest description of such graphs however are the Ramanujan graphs constructed by Lubotzky-Phillips-Sarnak ([22]). But the proof involves some sophisticated number theory.

5 Dependent Random Choice

Sometimes, the desired object is not the random object itself, but an associate of it.

In this chapter, we consider another aspect of tweaking a randomized construction: *Sometimes it pays off to pick the object of desire not be picking it directly as a random object, but rather pick another object randomly and then pick a relevant associated object to the randomly picked object, to be our desired object.* This sounds a bit roundabout but on quite a few occasions, it turns out to be the correct thing to do.

The premise for some of the investigations in this chapter is motivated by the following question: Given a 'small' graph H, how many edges must a graph G have in order that $H \subset G$? We denote by $ex(H;n)$ the maximum number of edges in an n vertex graph G which is H-free. If H is not bipartite then theorem of Erdős-Stone-Simonovits settles this upto a multiplicative factor of $1 + o(1)$. But if H is bipartite then the Erdős-Stone-Simonovits theorem only tells us that $ex(n;H) = o(n^2)$. This begs the following question:

Question 21. *For H bipartite, what is the correct value of α with $1 \leq \alpha < 2$ such that $ex(n;H) = \Theta(n^{\alpha(H)})$?*

Suppose $H = (A \cup B, E)$ with $|A| = a, |B| = b$. One constructive way to find a copy of H in a large graph G is to try and embed H into G, one vertex at a time. Suppose there is a large subset A_0 of G into which the vertices of A have already been embedded in some fashion. Let $B = \{v_1, v_2, \cdots v_b\}$ and suppose that we have embedded $v_1, \cdots v_{i-1}$ into $V(G)$. The new idea that provides a scheme by which this inductive procedure extends to embedding v_i as well is the following: Suppose v_i has degree r, and *suppose that every r-subset of A_0 has many common neighbors in G.* One elementary bound here is that the number of common neighbors is at least $a + b$. Since A has been embedded into A_0, this gives a set $U \subset A_0$ of size $\leq r$ which should be the neighbor set for v_i. Since $|U| \leq r$ and it has at least $a + b$ common neighbors in G there is some available choice for v_i in $V(G)$ which is not a vertex that has already been taken! In short, we have the following

Proposition 22. *Let H be bipartite, $H = (A \cup B, E)$ with $|A| = a, |B| = b$, any vertex in B has degree at most r. Suppose there is a set $A_0 \subset V(G)$ of size at least a such that*

every r-subset of A_0 has at least $a+b$ common neighbors in G. Then H can be embedded in G.

This proposition presents a technique which allows one to establish that a graph H can be embedded into a bigger graph G and the technical criterion in the proposition leads to the following question: Given a graph G, under what conditions can one ensure that there exists a subset of vertices A_0 of size at least a such that every r-subset of A has at least $a+b$ common neighbors?

Since we do not choose our actual objects of interest by the random method but rather in this *dependent manner*, this method is referred to as the method of Dependent Random Choice.

5.1 A graph embedding lemma

Let $V(H) = A \cup B, |A| = a, |B| = b$, let A_0 be subset of $V(G)$ containing all the vertices of A. We seek to embed the graph H in G as described in the preceding section, and this brings us concretely to the following question: How do we determine a set A_0 such that every r-subset of A_0 has many common neighbors in G?

Before we launch into how we might prove this proposition, let us see if picking the desired set randomly would work. A moment's reflection will tell us that it may not be feasible at all. Suppose G is bipartite with both parts of considerable size. Then a random set is very likely to pick at least one vertex from each of the parts and then the condition cannot be satisfied. But even if we were aware of the graph being bipartite there is yet another issue. Indeed, suppose we pick each vertex to be in independently with probability p, then $\mathbb{E}(|U|) = np$ and the expected number of r-subsets of U that do not have, say, m common neighbors, is at most $\binom{n}{r}r^p$. If we were to try and alter this set by removing one element from each bad r-subset, we are left with a set whose size is only guaranteed to be at least $np - \binom{n}{r}r^p$ and this can only certify that the set has size at least $\Omega_r(1)$, a far cry from what we want.

The main key idea to overcome this problem is to invert this search technique: instead of picking the set A_0 randomly, *pick a set T and let A_0 be the set of those vertices which contain T among their neighbors.* This is a healthy heuristic since by fiat, we know that all the chosen vertices have the vertices of T among their neighbors.

Indeed, over t rounds, pick a vertex v_i uniformly at random and independently across the rounds. Call this set T and consider the set of common neighbors of T - we shall

denote that by $N^*(T)$. Then

$$\begin{aligned}\mathbb{E}(|N^*(T)|) &= \sum_{v\in V}\mathbb{P}(v\in N^*(T))\\ &= \sum_{v\in V}\left(\frac{d(v)}{n}\right)^t\\ &\geq \frac{1}{n^{t-1}}\left(\frac{1}{n}\sum_{v\in V}d(v)\right)^t\\ &= \frac{(\bar{d})^t}{n^{t-1}}\end{aligned}$$

where $\overline{d}$ denotes the average degree of the vertices of G. The inequality above follows from Jensen's inequality for convex functions.

Let Y denote the number of r-subsets U of $N^*(T)$ such that U has fewer than m common neighbors. Then

$$\mathbb{E}(Y)\leq \sum_{\substack{U\subset V(G),|U|=r\\ |N^*(T)|<m}}\mathbb{P}(U\subset N^*(T)).$$

If $U\subset N^*(T)$, it means that every choice for T was picked from among the common neighbors of U, so $\mathbb{P}(U\subset N^*(T))\leq \left(\frac{m}{n}\right)^t$. Consequently,

$$\mathbb{E}(Y)\leq \binom{n}{r}(\frac{m}{n})^t$$

which implies

$$\mathbb{E}(|N^*(T)|-Y)\geq \frac{(\bar{d})^t}{n^{t-1}}-\binom{n}{r}\left(\frac{m}{n}\right)^t$$

so that there exists (by the method of alterations as seen in the preceding chapter) $A_0\subset N^*(T)$ of size at least $\frac{(\bar{d})^t}{n^{t-1}}-\binom{n}{r}\left(\frac{m}{n}\right)^t$ such that every r-subset of A_0 has at least m common neighbors. This gives the following

Theorem 23. *(Alon, Krivelevich, Sudakov) H is bipartite with vertex partition (A,B), and if every vertex of B has degree $\leq r$, then $ex(n;H)=O_H(n^{2-\frac{1}{r}})$.*

Proof. We only need to fill in the gaps now. Note that

$$e(G)\geq Cn^{2-\frac{1}{r}}\implies \bar{d}\geq 2Cn^{1-\frac{1}{r}}$$

where $C=C_H$ is a constant depending on H. To complete the proof, we need

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

Now plugging in the lower bound for $\overline{d}$ from before, we have

$$\frac{(\bar{d})^t}{n^{t-1}} - \binom{n}{r}(\frac{a+b}{n})^t \geq \frac{((2C)^t n^{t-\frac{t}{r}})^t}{n^{t-1}} - \frac{n^r}{r!}\left(\frac{a+b}{n}\right)^t.$$

Now, setting $r = t$ gives that the last expression is at least

$$(2C)^r - \frac{(a+b)^r}{r!} > a$$

with $C > \frac{1}{2}\left(a + \frac{(a+b)^r}{r!}\right)^{1/r}$ and that completes the proof. ■

Before we move on, we highlight the inequality obtained earlier to the status of an observation.

Observation 24. *Suppose the average degree of a graph G is d. Then there exists a subset A_0 of size at least $\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t$ such that every r-subset of A_0 has at least m common neighbors.*

The peculiar aspect of this observation is that the parameter t which appears is not present in the consequence, so it is more of a driving parameter that gives a condition to make a conclusion.

5.2 An old problem of Erdős

We now take a look at another old problem of Erdős that was settled in the affirmative following the Dependent Random Choice line. But first, we need a definition.

Definition 25. *A topological copy of a graph H is formed by replacing every edge of H by a path such that paths corresponding to distinct edges are internally disjoint, i.e., have no common internal vertices.*

Erdős conjectured that if $e(G_n) \geq cp^2 n$, then there is a topological copy of K_p in G. This was proved in 1998 by Bollobás and Hind. Erdős' conjecture implies that there is a topological copy of $K_{\sqrt{n}}$ in G_n if $e(G_n) \geq cn^2$.

Definition 26. *A t-subdivision where each edge is replaced by a path with $\leq t$ internal vertices.*

Erdős also asked if ε-dense graphs, i.e., graphs G_n with $e(G_n) \geq \varepsilon n^2$ admit a 1-subdivision of $K_{\Omega(\sqrt{n})}$. More formally, is there a 1-subdivision of $K_{\delta\sqrt{n}}$ in an ε-dense graph for some absolute $\delta = \delta(\varepsilon) > 0$? Note that the Bollobás-Hind result does not establish a 1-subdivision, since the paths in the topological copy could involve some long paths.

The following perspective is key:

If one seeks to embed a fixed bipartite graph into another graph, and if all the vertices on one side of the bipartite graph have somewhat small degree, then the Dependent Choice method gives a handle - effectively reducing the problem to a calculation - on proving sufficiency results.

In the Erdős problem above, note that a *strict* 1-subdivision of the complete graph K_a, i.e., one where each edge of K_a is subdivided to get a path of length two, corresponds to a bipartite graph with parts of size $a, \binom{a}{2}$, respectively. Observe that every vertex in the part of size $\binom{a}{2}$ has degree 2 since each of these vertices is placed in an edge of the original K_a, and hence has degree 2. Thus, the Dependent Random Choice technique appears a likely tool.

Theorem 27. *(Alon, Krivelevich, Sudakov) If $e(G_n) \geq \varepsilon n^2$, then G has a 1-subdivision of $K_{\varepsilon^{3/2}\sqrt{n}}$.*

Proof. If we think along the lines of the embedding procedure that we discussed in the previous sections, then as remarked above, we have a sufficiency condition provided we make a back calculation. Indeed, we would have the result we seek if

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

Here $r = 2, m = a + \binom{a}{2} < 2\binom{a}{2} < a^2,$ and $\bar{d} \geq 2\varepsilon n$.

Consequently,

$$LHS > (2\varepsilon)^t n - \frac{n^2}{2}\frac{a^{2t}}{n^t}.$$

For $a = \delta\sqrt{n}$ and $\delta = \varepsilon^{3/2}$, we have

$$LHS > \varepsilon^t\left(2^t n - \frac{n^2}{2}\varepsilon^{2t}\right)$$

so if the second term in the square bracket equals n then we may factor out n from both these terms. This basically boils down to setting $t = \frac{\log n}{2\log(1/\varepsilon)}$ so that

$$LHS > \frac{\sqrt{n}}{2}(2^{t+1} - 1) > \frac{\sqrt{n}}{2}2^t = \frac{\sqrt{n}}{2}n^{\frac{\log 2}{2\log(1/\varepsilon)}}$$

As n goes large, this beats $a = \delta\sqrt{n}$ and settles the conjecture. ■

5.3 A special case of Sidorenko's conjecture

One of the most beautiful conjectures in extremal graph theory is Sidorenko's conjecture. To get to it, we need a definition first:

Definition 28. *A graph homomorphism between graphs H, G is a map $\phi : V(H) \to V(G)$ such that whenever uv is an edge in H, $\phi(u)\phi(v)$ is an edge in G.*

Homomorphism capture graph adjacencies *at the local level*, i.e., for each vertex u the neighbors of u are mapped to the neighbors of the image of u. The map ϕ is not required to be injective, so for instance, there is a homomorphism from K_3 to any odd cycle. It is usually of greater interest to consider *isomorphisms* between graphs, i.e., injective maps ϕ such that ϕ also preserves non-adjacencies. But if H is small compared to G, then the non-injective maps are asymptotically far fewer than the injective ones, so homomorphisms are easier to study, since to count the number of homomorphisms, one can think of it in terms of embeddings where for each vertex $u \in H$, the neighbors of u in H are mapped into neighbors of its image in G.

Let $h_H(G)$ denote the number of homomorphisms from H to G. The *homomorphism density* of H in G denoted $t_H(G)$ is defined as $t_H(G) := \frac{h_H(G)}{|V(G)|^{V(H)|}}$.

Sidorenko's conjecture (also attributed to Erdős and Simonovits) states that for any bipartite graph H, among all graphs G with edge density p, the random graph $G(n,p)$ has asymptotically the least number of copies of H. More formally,

Conjecture 29. *(Sidorenko): Suppose $H = (A, B, E)$ is bipartite and suppose G is a graph with edge density p. Then $t_H(G) \geq p^{e(H)}$.*

One way to intuit this is that a random graph tends to 'spread out' all the copies of H so that no conglomeration of the copies of H is possible. While there have been several attacks on this problem with beautiful results by several researchers, the problem still remains open. In this section we shall see a beautiful result due to Conlon, Fox and Sudakov (2010). But before we get to that result, let us quickly see how Sidorenko's conjecture may also be stated in terms of counting homomorphisms instead of dealing with homomorphism densities. Let $|V(H)| = n, e(H) = m$, and suppose $h_H(G) \geq c_H p^m N^n$ holds for all graphs G on N vertices with $pN^2/2$ edges. Here c_H is a constant that depends only on H. We first observe that this establishes the Sidorenko conjecture for H.

Indeed, suppose $t_H(G) < p^m$ for some graph G with edge density p. The idea is to 'boost' up the edge density of H in another related graph, by what is called the 'tensoring trick'. For graphs G_1, G_2, the (weak) product $G_1 \times G_2$ is the graph on the vertex set $V(G_1) \times V(G_2)$ and (u, v) is adjacent to (u', v') if and only if $uu' \in E(G_1)$ and $vv' \in E(G_2)$. The simplicity of this definition leads to many things, but here, the relevant point is that for any H, $t_H(G_1 \times G_2) = t_H(G_1)t_H(G_2)$. We shall denote by $G^{\otimes r}$ the r-fold product $G \times \cdots \times G$.

Consider $0 \leq c = \frac{t_H(G)}{p^m} < 1$, if possible. Then for any integer $r \geq 1$

$$c_H p^{rm} \leq t_H(G^{\otimes r}) = t_H(G)^r = c^r p^{mr} = c^r (p^r)^m$$

since p^r is the edge density of $G^{\otimes r}$ and the assumption about the number of homomorphisms of H in any graph. But since $c < 1$, this yields a contradiction if r is sufficiently large.

Theorem 30. *(Conlon, Fox, Sudakov) Suppose $H = (A, B, E)$ is bipartite with $n = a+b$ vertices and m edges, and suppose there is a vertex in A (which shall be referred to as a special vertex) which is adjacent to all the vertices of B. Then for any graph G_N with at least $pN^2/2$ edges, the number of homomorphisms from $H \to G$ is at least $(2n)^{-n^2}p^m N^n$. Consequently, Sidorenko's conjecture holds for H.*

Proof. Let $A = \{u_1, \ldots, u_a\}$ and $B = \{w_1, \ldots, w_b\}$. Let us start with a scheme for constructing homomorphisms from H to G. One curious aspect of the hypothesis of the theorem is the assumption about the existence of a special vertex in A. But if we were to approach this from a constructive perspective, it makes certain things more rigid and natural: Suppose $u_1 \in A$ is special. To construct a homomorphism ϕ one vertex at a time, we shall first fix the image $\phi(u_1) = x_1$ in G, and then (since we are interested in homomorphisms which need not be injective) fix a *sequence* $\mathfrak{B} = (y_1, \ldots, y_b)$ with $y_i \in N(x_1)$ which would act as the image of B under ϕ, and then finally, choose the images of the other $u_i \in A$.

Suppose we have chosen $y_i \in N(x_1)$ that act as $\phi(w_i)$. To see if this sequence $\mathfrak{B}$ is a good extension to the choice for x_1, let us examine how well this extends to the other u_i as a homomorphism. Each u_i picks a subsequence $\mathfrak{B}' := (y_{i_1}, \ldots, y_{i_k})$ that corresponds to the neighbors of u_i, so one is guaranteed many homomorphism extensions for defining $\phi(u_i)$ if $N^*(\mathfrak{B}')$ is large. Since the neighbors of the u_i may be arbitrary subsets of B, a naturally good choice $\mathfrak{B}$ is one for which for *every subsequence* $\mathfrak{B}' := (y_{i_1}, \ldots, y_{i_k})$, the set $N^*(\mathfrak{B}')$ is 'large'. This begs the question: How large is 'large'? Since Sidorenko's conjecture posits that the random graph generates the least number of homomorphisms, it is reasonable to compare the size of the $N^*(\mathfrak{B})$ with $p^{d(u_i)}N$ since in a random graph of edge density p, the expected number of common neighbors for a set of size k is $p^k N$.

So, we formally postulate: A sequence $\mathfrak{B} = (y_1, \ldots, y_b)$ is *desirable* if for each $1 \le k \le b$ the subsequence $\mathfrak{B}' = (y_{i_1}, \ldots, y_{i_k})$ has $|N^*(\mathfrak{B}')| \ge \alpha p^k N$ for some small α that we will pin down later. We define a vertex $x \in V(G)$ to be *good* if the number of desirable sequences $\mathfrak{B} = (y_1, \ldots, y_b)$ with $y_i \in N(x)$ is at least $\frac{d(x)^b}{2}$. Again, this is not out of the blue; if x were to act as $\phi(u_1)$ the number of possible sequences of neighbors of x is at most $d(x)^b$, and we require that at least half of those are desirable sequences. We denote by Good, the set of good vertices in G.

This gives us a back-of-the-envelope estimate on the number of homomorphisms from H into G to be at least

$$\sum_{x_1 \in \text{ Good}} \frac{d(x_1)^b}{2} \left(\prod_{i=2}^{a} \alpha p^{d(u_i)} N \right).$$

To explain this, we first pick x_1 to be a good vertex, and fix x_1 as the image of the special u_1. For each desirable sequence $(y_1, \ldots, y_b)$ which will serve as $(\phi(w_1), \ldots, \phi(w_b))$, we pick choices for the remaining u_i. Since $\mathfrak{B}$ is desirable, there are at least $\alpha p^{d(u_i)} N$ choices for each u_i.

Hence

$$\begin{aligned}
\sum_{x_1 \in \text{ Good}} \frac{d(x_1)^b}{2} \left(\prod_{i=2}^{a} \alpha p^{d(u_i)} N \right) &= \frac{\alpha^{a-1} p^{m-b} N^{a-1}}{2} \sum_{x_1 \in \text{ Good}} d(x_1)^b \\
&\geq \frac{\alpha^{a-1} p^{m-b} N^{a-1}}{2} N \left(\frac{\sum_{x_1 \in \text{ Good}} d(x_1)}{N} \right)^b \\
&= \frac{\alpha^{n-1} p^{m-b} N^a}{2N^b} \left(\sum_{x_1 \in \text{ Good}} d(x_1) \right)^b
\end{aligned}$$

so if we can prove a lower bound of the form

$$\sum_{x_1 \in \text{ Good}} d(x_1) \geq \Omega(pN^2)$$

then we are through.

This sets us a goal, but there is still a lot packed tightly into the notion of what it means for a vertex to be good. To unspool this a bit, suppose that a vertex $x \notin$ Good. An alternate way to state this is: Suppose $\mathfrak{B}$ is picked uniformly at random from $(N(x))^b$. Then the probability that $\mathfrak{B}$ is not desirable is at most $1/2$.

Let $\mathfrak{B} = (y_1, \ldots, y_b)$, and let us fix $1 \leq i_1 < \cdots < i_k \leq b$, and set $\mathfrak{B}' = (y_{i_1}, \ldots, y_{i_k})$. If there exists $\beta > 0$ such that $\mathbb{P}(|N^*(\mathfrak{B}')| < \alpha p^k N) < \beta$ for all k and choices of $(i_1, \ldots, i_k)$ the

$$\mathbb{P}(\mathfrak{B} \text{ is not desirable}) < 2^b \beta \leq 2^{n-1} \beta$$

and if we choose $\beta = 2^{-n}$ then this contradicts that $x \notin$ Good.

This motivates the following definition. For $1 \leq k \leq b$, we say that a vertex x is *problematic for* k (which we shall denote by $x \sim k$) for a positive integer $k \leq b$ if the number of subsequences $\mathfrak{B}' = (y_1, \ldots, y_k) \in N(x)^k$ with $|N^*(\mathfrak{B}')| < \alpha p^k N$ is somewhat large, say at least $\beta d(x)^k$. By our terminology, $x \in$ Good if and only if x is not problematic for k for each $1 \leq k \leq b$.

We now invoke the Dependent Random Choice principle: *Small subsets of the common neighborhood of small random sets admit many common neighbors.* Let $\mathfrak{B}'$ be a random k-sequence of vertices from V, and let COUNT_k be the number of vertices $x \in V$ such

that $\mathfrak{B}' \in N(x)^k$ and $|N^*(\mathfrak{B}')| < \alpha p^k N$. Then

$$\begin{aligned}
\mathbb{E}(\mathrm{COUNT}_k) &= \sum_{x \in V} \mathbb{P}(\mathfrak{B}' \in N(x)^k \text{ and } |N^*(\mathfrak{B}')| < \alpha p^k N) && (5.1)\\
&\geq \sum_{x \sim k} \mathbb{P}(\mathfrak{B}' \in N(x)^k) \geq \beta \sum_{x \sim k} \left(\frac{d(x)}{N}\right)^k && (5.2)\\
&\geq \beta N^{1-2k} \left(\sum_{x \sim k} d(x)\right)^k && (5.3)
\end{aligned}$$

by the convexity of the function $f(x) = x^k$ and the definition of x being problematic for k. On the other hand,

$$\mathbb{E}(\mathrm{COUNT}_k) < \alpha p^k N \tag{5.4}$$

since any such x that is counted in this which admits $\mathfrak{B}' \in N(x)^k$ with $|N^*(\mathfrak{B}')| < \alpha p^k N$ must necessarily satisfy $x \in N^*(\mathfrak{B}')$. Thus we have

$$\sum_{x \sim k} d(x) < \left(\frac{\alpha}{\beta}\right)^{1/k} pN^2.$$

Now, if $\alpha < \beta$, then summing this over all $k \leq b$ gives

$$\sum_{x \notin \text{ Good}} d(x) < b \left(\frac{\alpha}{\beta}\right)^{1/b} pN^2$$

so, if we take $\beta = \frac{1}{2^n}, \alpha = \frac{1}{4^n n^n}$ we have $\sum_{x \in \text{ Good}} d(x) \geq \frac{pN^2}{2}$ and the proof is complete. ■

5.4 The Balog-Szemerédi-Gowers Theorem

The last section of this chapter deals with a deep result in Additive Combinatorics, originally due to Balog and Szemerédi, and then was reproved by Gowers with stronger estimates than in the original. To motivate the statement, we set up some terminology. For sets $A, B \subset Z$ for an ambient abelian group, we mean by $A + B$ (called the *sum set*) the set of all elements of the form $a + b$ with $a \in A, b \in B$. One of the principal questions in Additive Combinatorics studies the size of sum sets in interesting abelian groups. One of the foundational theorems in this direction is Freiman's theorem which states that if $|A + A| \leq K|A|$ for some bounded constant K, then A is a large subset of a well-structured set, called a Generalized Arithmetic progression, so in that sense the size of $|A + A|$ measures how structured the set A is.

But suppose we only have access to consider pairs of sums for a restricted number of pairs of A. More precisely, consider a bipartite graph $G = G_A$ with both vertex partitions corresponding to the set A and we only have access to the pairs of sums $a + b$ whenever $ab \in E(G)$. We shall denote by $A +_G A$ the set $\{a + b : a, b \in A, ab \in E(G)\}$. How much information does $A +_G A$ capture about $|A + A|$?

Let $|A| = n$. If the graph G is sparse, then there is not much hope of gathering much information about A, so suppose that G is dense, i.e., $e(G) \geq \alpha n^2$ for some fixed $0 < \alpha < 1$. If $|A +_G A| \leq cn$ for some absolute constant c then could we conclude that $|A + A| \leq c_1 n$ for some c_1?

A moment's thought tells us that such a conclusion is too good to be true. Indeed, suppose R is a random subset of $[1, n^2]$ where each $x \in [1, n^2]$ is picked uniformly and independently with probability $\frac{1}{n}$, say, and let $A = [1, n] \cup R$. Let G be the graph corresponding to the pairs ab with $a, b \in [1, n]$. Then $|A +_G A| = 2n - 1$ and it is not hard to see that $\mathbb{P}(|R + R| \geq \Omega(n^2)) = \Omega(1)$.

However, in this example the set A had a large subset $A' = [1, n]$ for which $|A' + A'| = 2|A'|$. This motivates the following modification: If $|A +_G A| = O(n)$, then is there a large structured subset of A? The answer in the affirmative is the substance of the Balog-Szemerédi-Gowers theorem:

Theorem 31. *Suppose $0 < \alpha < 1, c > 0$ are reals then there exist c', c'' (depending on c, α) and $n_0(c, k)$ such that the following holds. Suppose $n \geq n_0$ and A is a subset of the integers of size n, and suppose that for a graph $G = G_A$ with $e(G) \geq \alpha n^2$ we have $|A +_G A| \leq cn$ then there exists $A' \subseteq A$ with $|A'| \geq c'|A|$ with $|A' + A'| \leq c''n$.*

We will prove a slight generalization of this for sets A, B and $A +_G B$.

There is a natural injection from paths of length 3 in G to elements in $A+B$: Suppose $x, x', x'' \in A +_G B$ and $x = a + b', x' = a' + b', x'' = a' + b$. Then $a + b = x - x' + x''$, so one can associate to the path $ab'a'b$ in G, the element $a + b \in A + B$. Hence *every* path of length 3 in G corresponds to a unique pair (a, b) which corresponds to an element in $A + B$. If we can lower bound the number of paths of length 3 corresponding to each element in $A+B$ then we have an upper bound on the size of $A+B$. Since G is dense it is reasonable to expect many paths of length 3 between most pairs of vertices (but perhaps not all pairs) so we might need to restrict to subsets A' and B' so that this property holds.

More precisely, suppose we can find $A' \subseteq A, B' \subseteq B$ such that $|A'| \geq c'n, |B'| \geq c''n$ and between every pair $(a, b) \in\in A' \times B'$ there are $\Omega(n^2)$ paths of length 3. Then there

are at least $\Omega(n^2)$ triples $(x, x', x'') \in (A +_G B)^3$ such that

$$\begin{aligned} x - x' + x'' &= a + b, \\ x &= a + b' \\ x'' &= a' + b \end{aligned}$$

holds for some $(a', b') \in A \times B$. Since by assumption $|A +_G B| \leq cn$ the number of triples $(x, x', x'') \in (A +_G B)^3 \leq c^3 n^3$, so

$$\begin{aligned} c^3 n^3 &\geq \sum_{y \in A'+B'} \#\{(x, x', x'' | y = x - x' + x'')\} \quad (5.5) \\ &\geq \Omega(n^2)|A' + B'| \quad (5.6) \end{aligned}$$

which gives $|A' + B'| = O(n)$.

So we would be done if we can answer the following question affirmatively:

Question 32. *If $G = G[A, B]$ is bipartite, $|A| = |B| = n$ and $e(G) \geq \kappa n^2$ can we find subsets $A' \subseteq A, B' \subseteq B$ with $|A'| \geq c'n, |B'| \geq c''n$ such that for all $a \in A', b \in B'$ there are $\Omega(n^2)$ paths $ab'a'b$?*

Let $A_1 = \{a \in A : d(a) \geq \frac{\kappa n}{2}\}$. Then $e(A \setminus A_1, B) \leq \frac{\kappa n^2}{2}$, so

$$n|A_1| = |A_1| \cdot |B| \geq e(A_1, B) \geq \frac{\kappa n^2}{2}$$

which gives

$$|A_1| \geq \frac{\kappa n}{2}.$$

Now let us speculate a bit.

Conjecture 33. *Suppose $e(A, B) \geq \kappa n^2$. There exist absolute constants $\alpha, \delta > 0$ depending only on κ such that the following holds: There exists $A' \subseteq A$ with $|A'| \geq \alpha|A|$ such that every pair $\{a, a'\}$ in A' has at least $\delta|B|$ common neighbors in B.*

This conjecture is a natural first-line-of-attack. If conjecture 33 holds, then get $U \subseteq A_1, |U| \geq \alpha|A_1|$ such that every pair $\{a, a'\}$ in U have at least $\delta|B|$ common neighbors; by the arguments outlined earlier, this gives us that $|U| \geq \alpha\kappa n/2$.

Now we choose $B_1 \subset B$ to consist of those vertices with large degree into U - that would provide many choices for the 3rd edge. If $|B_1| = \Omega(n)$, we are through.

Let $\mu > 0$ be a parameter that we shall fix later. Set

$$B_1 := \{b \in B : d(b, U) \geq \mu|U|\} \quad (5.7)$$

and again exactly as before, since $e(B \setminus B_1, U) \leq \mu|U|n$ and $e(U, B) \geq |U|(\frac{\kappa n}{2})$, we have

$$|U| \cdot |B_1| \geq e(U, B_1) \geq (\frac{\kappa}{2} - \mu)|U|n \tag{5.8}$$

so that

$$|B_1| \geq \frac{\kappa n}{4}. \tag{5.9}$$

We now claim that (U, B_1) will do the job. Indeed, since each $b \in B_1$ has $\geq \frac{\kappa|U|}{4}$ neighbors in U, b has at least $\frac{\kappa|U|}{4} - 1$ neighbors in $U \setminus \{a\}$. For each $a' \in N(b) \setminus \{a\}$, there exist at least $\delta n - 1$ common neighbors for a, a' in $B \setminus \{b\}$, so that there are at least

$$(\frac{\kappa}{4}|U| - 1)(\delta n - 1) \geq \frac{\kappa\delta}{16}|U|n \geq \frac{\alpha\kappa\delta}{32}n^2 \tag{5.10}$$

paths of length 3 from a to b.

Unfortunately, here is the bombshell; Conjecture 33 is FALSE! For an explicit counter example, see [20].

How does one salvage this? If this line of argument can still be exploited, the next natural question in place of Conjecture 33 would be

Question 34. *Does conjecture 33 hold if the words 'every pair' are replaced by 'most pairs'? More precisely, suppose $\kappa > 0$, and $G = G[A, B]$ is bipartite with edge density κ. Does there exist subset $A' \subseteq A$ such that $|A'| \geq \alpha|A|$ such that there are at least $(1 - \varepsilon)|A'|^2$ ordered pairs of A', each of which have at least $\delta|B|$ common neighbours in B, for some suitable $\alpha, \delta, \varepsilon$ depending only on κ?*

First, let us see if this weakening still yields the desired outcome. Let U, B be chosen as before, but with U being the set guaranteed by an affirmative answer to question 34 rather than the erroneous Conjecture 33. To be precise, call (a, a') a pair *bad* if they have fewer than $\delta|B|$ common neighbors in B. Let $U \subset A_1$ such that $|U| \geq \alpha|A_1|$ and with at most $\varepsilon|U|^2$ bad pairs. Then again as before, $b \in B_1$ implies that $d(b, U) \geq \frac{\kappa}{4}|U|$. But this time, instead of using U, we refine it further. Let

$$A' := \{a \in U : a \text{ is in at most } \frac{\kappa}{8}|U| \text{ bad pairs}\}. \tag{5.11}$$

Since the total number of bad pairs in U is at most $\varepsilon|U|^2$, the total number of bad pairs featuring a vertex in $U \setminus A'$ is at least $\frac{\kappa}{8}|U|(|U| - |A'|)$, So

$$\varepsilon|U|^2 \geq \frac{\kappa}{8}|U|(|U| - |A'|)$$

which gives

$$|A'| \geq (1 - \frac{8\varepsilon}{\kappa})|U| \tag{5.12}$$

So for instance if $\varepsilon = \frac{\kappa}{16}$ then $|A'| \geq \frac{1}{2}|U|$.

So for $(a, b) \in A' \times B_1$, the number of paths of length 3 from a to b is at least

$$(\frac{\kappa}{4}|U| - 1 - \frac{\kappa}{8}|U|)(\delta n - 1) \geq \frac{\kappa\delta}{32}|U|n \geq \frac{\alpha\kappa\delta}{64}n^2 = \Omega(n^2) \tag{5.13}$$

as before with only slightly worse constants. Thus we are just left with proving settling question (34) in the affirmative. And the good news is

Theorem 35. *Suppose $0 < \varepsilon < 1$ and $G = G[A, B]$ is bipartite with $e(G) = \kappa|A||B|$. There exist constants $\alpha, \delta > 0$ that depend only on κ, ε such that the following holds: There exists $A' \subseteq A$ satisfying*

- $|A'| \geq \alpha|A|$,
- *All pairs (a, a') in A' except for at most $\varepsilon|A'|^2$ admit at least $\delta|B|$ common neighbors in B.*

Proof. We go back to the Dependent Random Choice heuristic: *Small subsets of the common neighborhood of small random sets admit many common neighbors.* Pick $b \in B$ at random and let $A' = N(b)$. Then

$$\mathbb{E}[|A'|] = \frac{1}{|B|}\sum_{b\in B} d(B) = \kappa|A|. \tag{5.14}$$

As before, call a pair (a, a') bad if the number of common neighbors is at most $\delta|B|$, and let BAD denote the number of bad pairs (a, a') in A. Then

$$\begin{aligned} \mathbb{E}[|\text{BAD}|] &= \sum_{(a,a')\in \text{BAD}} P(\text{b is chosen from a set of size at most } \delta|B|) && (5.15)\\ &\leq \delta|A|^2 && (5.16) \end{aligned}$$

Our goal is to find a b such that $|N(b)| = \Omega(|A|)$ and $|BAD| \leq \varepsilon|N(b)|^2$. Note that by Cauchy-Schwarz,

$$\mathbb{E}(|A'|^2 - \frac{1}{\varepsilon}|BAD|) \geq \kappa^2|A|^2 - \frac{\delta}{\varepsilon}|A|^2 \tag{5.17}$$

Setting $\delta = \frac{\varepsilon\kappa^2}{2}$, we have

$$\mathbb{E}[|A'|^2 - \frac{1}{\varepsilon}|BAD|] \geq \frac{\kappa^2|A|^2}{2}$$

so that there exists $b \in B$ such that

$$|N(b)|^2 - \frac{1}{\varepsilon}|BAD| \geq \frac{\kappa^2|A|^2}{2}$$

Thus we have

$$|BAD| \leq \varepsilon|N(b)|^2 = \varepsilon|A'|^2 \tag{5.18}$$

$$|A'| = |N(b)| \geq \frac{\kappa|A|}{\sqrt{2}} \tag{5.19}$$

which proves theorem(35) and consequently, the Balog-Szemerédi-Gowers theorem. ■

5.5 A Ramsey bound for sparse bipartite graphs

We shall conclude this chapter with one final application - a beautiful result due to Conlon. Before we state the result, we need to recall the definition of the Ramsey number $r(H)$ for an arbitrary graph H. Since the (usual) Ramsey number concerns two colorings of the edges of a complete graph, it can also be restated as follows. For a graph H, the Ramsey number of an arbitrary graph, $r(H)$ is the least integer N such that any G_N with at least $\frac{1}{2}\binom{N}{2}$ edges contains H as a subgraph. (It is a simple exercise to see why this is well defined).

As we now know, the case for $H = K_n$ gives $N = 2^{\Theta(n)}$. An old problem of Erdős also considered this problem for sparse H, and conjectured that for r-sparse graphs (graphs with maximum degree at most r), the bound for $r(H)$ is linear in n. This was proved by Chvátal *et al* via the regularity lemma, and hence the constant (depending on r) was a tower-type bound. Later, this constant was improved to an exponential in r. The result of Conlon that we shall see in this section is a simpler and slightly weaker result that the original. More precisely, if H is bipartite and with maximum degree r, then $N = O_r(n)$. We will soon see a more explicit version of this, with an explicit bound in the $O_r(1)$ term there.

But before we get there, recall the Dependent Random Choice principle: If G_N has average degree d then there is a subset U with $|U| > \frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t$ such that every r-subset of U admits at least m common neighbors. Let us see if this version of the Dependent Random Choice trick gives us what we want. If N were to be linear in n, then we need a t such that

$$(2\varepsilon)^t N - \left(\frac{eN}{r}\right)^r \cdot \left(\frac{n}{N}\right)^t \geq \Omega(N)$$

and that is plainly infeasible.

Is there a workaround? Let us take a cue from the proof of the Balog-Szemerédi-Gowers theorem: instead of asking for all r-subsets of U to have many common neighbors, *we could ask for a U for which many (most?) of its r-subsets have many common neighbors.* Our current embedding lemma is rather loose in that one of the parts of H can

be embedded arbitrarily, so if we have to be more careful in embedding the first vertex set of H, then such an argument is bound to produce a sharper result.

Here is a rough sketch of a possible strengthening of the result. Suppose $H = (A, B, E)$ and let $A = \{u_1, \ldots, u_a\}$ and $B = \{v_1, \ldots, v_b\}$, and suppose we have embedded $u_i \to x_i$ in G. To be able to extend this embedding to a complete embedding of H, it would suffice if for each $v \in B$ the image of the set $N_H(v)$ among the x_i's is contained in at least one r-subset with, say, at least n common neighbors. In that case, then for each v_j, pick a 'good' r-subset containing the image of $N_H(v_j)$; since there are at least n common neighbors for that r-subset, there is at least one (still) available option that can serve as the image of v_j, thereby extending the embedding to all of B as well.

First, let us attempt a weakening of the conclusion of a Dependent-random-Choice type lemma. Suppose G_N has average degree at least εN. Call an r-subset *good* if $|N^*(R)| \geq n$. If we want a U such that the fraction of bad subsets of U is 'small', then as before let T be chosen randomly by picking $v_1, \ldots, v_r$ independently and uniformly, and let $U = N^*(T)$. Then as before, we have $\mathbb{E}(|U|) \geq \varepsilon^r N$ and if BAD denotes the set of bad subsets of U, then $\mathbb{E}|BAD| < \binom{N}{r}\left(\frac{n}{N}\right)^r$.

Here is a nice trick that allows us to combine these two into a single inequality to get two inequalities to be satisfied simultaneously. Since $\mathbb{E}|U|^r \geq (\mathbb{E}|U|)^r$ (Jensen's inequality), we rewrite this as

$$\mathbb{E}\left(|U|^r - \frac{1}{2}\cdot(\mathbb{E}(|U|)^r - \frac{1}{2}\cdot\frac{|BAD|}{\mathbb{E}(|BAD|)}\cdot(\mathbb{E}|U|)^r\right) \geq 0$$

and as a consequence there is a U for which this inequality holds, which then also guarantees that

$$\begin{aligned} |U|^r &\geq \frac{1}{2}(\mathbb{E}(|U|)^r \\ |U|^r &\geq \frac{1}{2}\frac{|BAD|}{\mathbb{E}(|BAD|)}(\mathbb{E}|U|)^r \end{aligned}$$

Some simple rearranging of these gives us the following: There exists U such that

- $|U| \geq 2^{-1/r}\varepsilon^r N \geq \dfrac{\varepsilon^r N}{2}$,
- The proportion of bad subsets of U is at most

$$\theta_0 := \frac{|BAD|}{\binom{|U|}{r}} < \frac{2^r}{\varepsilon^{r^2}}\cdot\left(\frac{n}{N}\right)^r$$

where in the last inequality, we use $\binom{|U|}{r} \geq \dfrac{1}{2^{r-1}}\cdot\dfrac{|U|^r}{r!}$.

Let us try an inductive procedure to embed A as described in the sketch above. For each $1 \leq i \leq a$, we will embed u_i into distinct x_i so that for each $v \in B$ the image of $N_H(v) \cap \{u_1, \ldots, u_i\}$ is contained in several good sets. We will along the way, make these as precise as our requirements force us.

To first embed u_1 into U, we need to find a *desirable* vertex x_1 that will serve as the image of u_1 in the embedding. While that word is yet to acquire a precise meaning, it certainly means one thing: *the proportion of bad subsets of U containing x_1 must not blow up too much.* Let $Und(u_1) := \{x \in U : x \text{ is in more than } \theta_1 \text{ proportion of bad sets}\}$ for a θ_1 that will be determined later. Then

$$r \cdot \theta_0 \binom{|U|}{r} \geq \#\Big\{(x, R) : x \in Und(u_1), R \in BAD, x \in R\Big\} > \theta_1 \binom{|U|-1}{r-1} \cdot |Und(u_1)|$$

which gives

$$|Und(u_1)| < \frac{\theta_0}{\theta_1}|U|.$$

As long as $\theta_0 \ll \theta_1$ there is a desirable x_1 which can serve as the image of u_1.

More generally, for a subset $S \subset U$ with $|S| = s < r$ define

$Und(S) := \{x \in U \setminus S : x \text{ is in more than } \theta_s \text{ proportion of bad sets containing } S\}$

for a suitable θ_s. Then, exactly arguing as above gives us

$$|Und(S)| < \frac{\theta_{s-1}}{\theta_s}|U|.$$

In this schematic, we have not fixed the θ_i yet, but we merely observe for the time being that we must have $\theta_0 \ll \theta_1 \ll \cdots \ll \theta_r \leq 1$.

Suppose inductively, that we have embedded $u_1, \ldots, u_i$ into U successfully. Suppose u_{i+1} has neighbors $v_{i_1}, \ldots, v_{i_t}$ with $t \leq r$; set $S_j := N_H(v_{i_j}) \cap \{u_1, \ldots, u_{i+1}\}$. The total number of bad options as a possible image of u_{i+1} is less than $\displaystyle\sum_j \left(\frac{\theta_{|S_j|-1}}{\theta_{|S_j|}}\right)|U|$.

Now, let us make some choices that will quantitate some of the previous statements we have made qualitatively. Since there are at most r terms in the summand above, if there is a uniform bound $f(r)$ for θ_s/θ_{s-1} for all s, then the sum above is at most $rf(r)|U|$. Since we need this to be at least n, $f(r) = \frac{1}{2r}$, and $|U| > 2n$ is naturally, the simplest setting. This loops back to giving $\theta_s = (1/2r)^{r-s}$, so that $\theta_0 = (1/2r)^r$, and this (a simple back-calculation) gives $N \geq 4r\varepsilon^{-r}n$.

We are now in a position to give the finishing touches to the result of Conlon, and thereby, the Ramsey bound stated earlier, by writing all our discoveries in a formal sequence.

Suppose $N \geq 4r\varepsilon^{-r}n$. If G_N has at least $\varepsilon N^2/2$ edges, then its average degree is at least εN, so there is a subset U of size at least $\frac{\varepsilon^r N}{2} \geq 2rn$ such that the proportion of bad r-subsets of U is at most $(2r)^{-r}$. Now, call a subset $S \subset U$ with $|S| \leq r$ to be undesirable if the proportion of bad r-subsets of U that contain S is at least $(2r)^{|S|-r}$, and for a desirable set S, let $Und(S)$ be the set of x in $U \setminus S$ such that $S \cup x$ is undesirable. Then $|Und(S)| \leq \frac{|U|}{2r}$. Thus, we have the following embedding procedure: embed each $u_i \in A$ in such a way that at the stage where we have embedded $u_1, \ldots, u_i$ we have that for every $v \in B$ the image of $N_H(v) \cap \{u_1, \ldots, u_i\}$ is not an undesirable set. It then follows that there is always an option to embed all the u_i into U satisfying this property. Finally, embed each $v \in B$; this is feasible because for each $v \in B$ the embedded image of $N_H(v)$ into U is not undesirable, so there is at least one good r-subset of U containing this set, thereby admitting at least n common neighbors.

To get the Ramsey bound, note that if $N = r2^{r+3}n$ then in any 2-coloring of the edges of K_N the more abundant color class has at least $\frac{1}{2}\binom{N}{2} = (1 - \frac{1}{N})N^2/2$ edges. Set $\varepsilon' = \frac{N-1}{2N}$. Since $2 \geq (\frac{N}{N-1})^r$ it follows by the previous result that G_N contains H.

6 The Second Moment Method

When you have a random variable of interest, and you can compute its expectation, you should try to compute the variance next.

The method of using expectation of random variables is a very useful and powerful tool, and its strength lies in its ease of computing the expectation. However, in order to prove stronger results, one needs to obtain results which prove that the random variable in concern takes values close to its expected value, with sufficient (high) probability. The method of the second moment, as we shall study here gives one such result which is due to Chebyshev. We shall outline the method, and illustrate a couple of examples. The last section covers one of the most impressive applications of the second moment method - Pippenger and Spencer's theorem on coverings in uniform almost regular hypergraphs.

6.1 Variance of a Random Variable and Chebyshev's theorem

For a real random variable X, we define $\text{Var}(X) := \mathbb{E}(X - \mathbb{E}(X))^2$ whenever it exists. It is easy to see that if $\text{Var}(X)$ exists, then $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Theorem 36 (Chebyshev's Inequality)**.** *Suppose X is a random variable, and suppose $\mathbb{E}(X^2) < \infty$. The for any positive λ,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{Var(X)}{\lambda^2}.$$

Proof. $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \geq \lambda^2 \mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda)$. ■

The use of Chebyshev's inequality, also called the Second Moment Method, applies in a very wide context, and it provides a very basic kind of 'concentration about the mean' inequality. The applicability of the method is most pronounced when the variance is of the order of the mean, or smaller. We shall see in some forthcoming chapters that

concentration about the mean can be achieved with much greater precision in many situations. What, however still makes Chebyshev's inequality useful is the universality of its applicability.

If $X = X_1 + \cdots + X_n$, then the following simple formula calculates $\mathrm{Var}(X)$ in terms of the $\mathrm{Var}(X_i)$. For random variables X,Y, define the Covariance of X and Y as

$$Cov(X, Y) := \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

For $X = X_1 + \cdots + X_n$, we have

$$\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} Cov(X_i, X_j).$$

This is a simple consequence of the definition of Variance and Covariance. In particular, if the X_i's are pairwise independent, then $\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i)$.

The (usually) difficult part of using the second moment method arises from the difficulty of calculating/estimating $Cov(X, Y)$ for random variables X, Y. One particularly pleasing aspect of the second moment method is that this calculation becomes much simpler if for instance we have *pairwise independence* of the random variables which is much weaker than the joint independence of all the random variables.

The preceding example illustrates one important aspect of the applicability of the second moment method: If $\mathrm{Var}(X_n) = O(\mathbb{E}(X_n))$ and $\mathbb{E}(X_n) \to \infty$ then Chebyshev's inequality gives

$$\mathbb{P}(|X_n - \mathbb{E}(X_n)| > \varepsilon \mathbb{E}(X_n)) = o(1).$$

In particular, X_n is 'close to' $\mathbb{E}(X)$ with high probability.

6.2 The Erdős-Ginzburg-Ziv theorem: When do we need long sequences?

Our first application in this section that arises more as an outcome of curiosity, and is in fact a probabilistic statement.

The Erdős-Ginzburg-Ziv theorem states that every sequence of length $2n-1$ of elements of $\mathbb{Z}_n$ contains a subsequence of size n whose sum equals zero. This is best possible in the sense that the sequence $(0^{n-1}1^{n-1})$ admits no such zero-sum subsequence. But a more natural question that arises is: *How necessary* is the length $2n-1$? In other words, are there other sequences that look nothing like these and yet need to be significantly long to witness a zero-sum subsequence of length n? If you had a *typical* sequence of elements from $\mathbb{Z}_n$, then how long does it need to be to contain a zero-sum subsequence of length n?

The answer, perhaps surprisingly, is that one typically needs much shorter sequences.

Theorem 37. *Suppose $\mathcal{X} := (X_1, \ldots, X_{n+2})$ be a random $\mathbb{Z}_n$-sequence, i.e., suppose X_i are chosen uniformly and independently from $\mathbb{Z}_n$. Then with high probability, $\mathcal{X}$ contains a zero-sum subsequence of length n.*

Proof. We first set up some notation. For a subset $I \subset [1, n+2]$ we shall denote by $\mathcal{X}_I$ the sum $\mathcal{X}_I := \sum_{i \in I} X_i$. Consider the indicator random variables $\mathbb{I}(\mathcal{X}_I) := 1$ if $\mathcal{X}_I = 0$ and zero otherwise. Let

$$\begin{aligned} \mathcal{H} &:= \{I \subset [n+2] : |I| = n\}, \\ N &:= \sum_{I \in \mathcal{H}} \mathbb{I}(\mathcal{X}_I). \end{aligned}$$

Then

$$\mathbb{E}(N) = \sum_{I \in \mathcal{H}} \mathbb{P}(\mathcal{X}_I = 0) = \frac{1}{n} \binom{n+2}{n} = \frac{(n+2)(n+1)}{2n},$$

and,

$$Var(N) = \sum_{I \in \mathcal{H}} Var(\mathbb{I}(\mathcal{X}_I)) + \sum_{\substack{I \neq J \\ I,J \in \mathcal{H}}} Cov(\mathbb{I}(\mathcal{X}_I), \mathbb{I}(\mathcal{X}_J)).$$

The main observation is that since X_i's are i.i.d, it follows that the $\mathcal{X}_I$ are *pairwise* independent. Indeed pick $i \in I \setminus J$ and $j \in J \setminus I$ and condition on the values of the random variables $\{X_\ell\}_{\ell \neq i,j}$; this determines X_i, X_j uniquely, so the conditional (and hence also the unconditional probability) of $\mathcal{X}_I = \mathcal{X}_J = 0$ is $1/n^2 = \mathbb{P}(\mathcal{X}_I = 0) \cdot \mathbb{P}(\mathcal{X}_J = 0)$.

Consequently, $Cov(\mathbb{I}(\mathcal{X}_I), \mathbb{I}(X_J)) = 0$ for $I \neq J \in \mathcal{H}$. Also, $\mathrm{Var}(\mathbb{I}(\mathcal{X}_I)) = \frac{1}{n}(1 - \frac{1}{n})$, so

$$Var(N) = \sum_{I \in \mathcal{H}} \mathrm{Var}(\mathbb{I}(X_I)) = \frac{1}{n}\left(1 - \frac{1}{n}\right) \frac{(n+2)(n+1)}{2}.$$

Therefore, by Chebyshev's inequality we have,

$$\mathbb{P}(N = 0) \leq \mathbb{P}(|N - \mathbb{E}(N)| \geq \mathbb{E}(N)) \leq \frac{\mathrm{Var}(N)}{(\mathbb{E}(N))^2} = \frac{\frac{1}{2}(1 - \frac{1}{n})}{\frac{1}{4}(1 + \frac{2}{n})(n+1)} = O\left(\frac{1}{n}\right)$$

which implies that $\mathbb{P}(N > 0) \to 1$. This completes the proof. ∎

Remark: The theory of zero-sum problems considers various instances where one is interested in sequences of elements of an abelian group admitting zero-sum subsequences with other characteristics. Interestingly, many of these group invariants behave marked differently for random sequences. For instance, the *Davenport* constant of a group is the minimum m such that every sequence of m elements from the group admits a non-trivial zero sum subsequence. The Davenport constant of the cyclic group $\mathbb{Z}/n\mathbb{Z}$ is n (easy to see) whereas its random analogue (as described above) is of the order $(1 + o(1)) \log_2 n$. One also has analogues of these invariants which admit weights, and the random analogues

of these analogues are typically much smaller. The only interesting instance where this is not the case is when we allow weighted sums for the subsequences with weights in $\{-1, 1\}$. For $\mathbb{Z}/n\mathbb{Z}$ the corresponding weighted Davenport constant is $\log_2 n$ whereas the random analogue is $(1/2 + o(1)) \log_2 n$. For more such results, see [10].

6.3 Distinct subset sums

For the next application, we need a definition.

Definition 38. *We say a set of positive integers $\{x_1, x_2, \ldots, x_k\}$ is said to have distinct sums if $\sum_{x_i \in S} x_i$ are all distinct for all subsets $S \subseteq [k]$.*

For instance, if $x_k = 2^k$, then we see that $\{x_1, x_2, \ldots, x_k\}$ has distinct sums. Erdős posed the question of estimating the maximum size $f(n)$ of a set $\{x_1, x_2, \ldots, x_k\}$ with distinct sums and $x_k \leq n$ for a given integer n. The preceding example shows that $f(n) \geq \lfloor \log_2 n \rfloor + 1$.

Erdős conjectured that $f(n) \leq \lfloor \log_2 n \rfloor + C$ for some absolute constant C. He was able to prove that $f(n) \leq \log_2 n + \log_2 \log_2 n + O(1)$ by a simple counting argument. Indeed, there are $2^{f(n)}$ distinct sums from a maximal set $\{x_1, x_2, \ldots, x_k\}$. On the other hand, since each x_i is at most n, the maximum such sum is at most $nf(n)$. Hence $2^{f(n)} < nf(n)$. Taking logarithms and simplifying gives us the aforementioned result.
As before, here is a probabilistic spin. Suppose $\{x_1, x_2, \ldots, x_k\}$ has distinct sums. Pick a random subset S of $[k]$ by picking each element of $[k]$ with equal probability and independently. This random subset gives the random sum $X_S := \sum_{x_i \in S} x_i$. Now $\mathcal{E}(X_S) = \frac{1}{2}(x_1 + x_2 + \cdots + x_k)$. Similarly, $\mathrm{Var}(X_S) = \frac{1}{4}(x_1^2 + x_2^2 + \cdots + x_k^2) \leq \frac{n^2 k}{4}$, so by Chebyshev we have

$$\mathbb{P}(|X_S - \mathcal{E}(X_S)| < \lambda) \geq 1 - \frac{n^2 k}{4\lambda^2}.$$

Now the key point is this: since the set has distinct sums and there are 2^k distinct subsets of $\{x_1, x_2, \ldots, x_k\}$, for any integer r we have that $\mathbb{P}(X_S = r) \leq \frac{1}{2^k}$; in fact it is either 0 or $\frac{1}{2^k}$. This observation coupled with Chebyshev's inequality gives us

$$1 - \frac{n^2 k}{4\lambda^2} \leq \mathbb{P}(|X_S - \mathcal{E}(X_S)| < \lambda) \leq \frac{2\lambda + 1}{2^k}.$$

Optimizing for λ we get

Proposition 39. $f(n) \leq \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1)$.

6.4 The space complexity of approximating frequency moments

One of the paradigmatic features of the probabilistic method is that it suggests different perspectives to many problems, and one of the features of probabilistic thinking is to be more accepting of approximate solutions, provided we have a control on the errors that accrue. This section features one such result due to Alon, Matias, and Szegedy.

One of the features of the Theory of Complexity is to study efficient handling of resources in various algorithmic computational problems (see [33] for a fantastic overview of the subject). Usually, the resource that is optimized is run time of an algorithm. In this section, we look at an optimization for *space constraints*.

Suppose $\mathfrak{A} = \{a_1, \ldots, a_m\}$ is a sequence of elements from $[N] := \{1, \ldots, N\}$, and for each $1 \leq i \leq m$ let m_i denote the number of occurrences of the element i in $\mathfrak{A}$. Define for each $k \geq 0$

$$F_k := \sum_{i=1}^{N} m_i^k$$

which are referred to as the *frequency moments* of the sequence. In particular, F_0 denotes the number of distinct members of the sequence, F_1 is the number of elements of the sequence (which is always m) and F_2 is called the *repeat rate* of the sequence and so on. We also define

$$F_\infty^* := \max_{1 \leq i \leq n} m_i$$

the most popular element of the sequence $\mathfrak{A}$[1]. For various reasons, one wishes to compute/estimate these statistics of a given sequence as the provide useful information about $\mathfrak{A}$.

It is straightforward to see that the frequency moments can be efficiently computed if we maintain a full histogram - keep a counter for each m_i as we scan over the data once) of the data - which requires memory space of size $\Omega(N)$. But the problem of interest here is to if it can be done efficiently with lesser *memory space* (i.e. with $o(N)$) for storage and processing. More precisely, suppose we are allowed to scan the data *once* and we have limited memory. One of the first interesting results is that for accurate computation of the frequency moments, one cannot improve upon the memory allocation (see [3]). So, we relax our requirements a little bit. We allow for a relative error in computing the F_i up to a factor of $1 - \lambda$ for some fixed $0 < \lambda < 1$, provided we have control on the error probability. The feature of this section is the following theorem of Alon, Matias, and Szegedy [3].

Theorem 40. *Suppose $k > 0$ and $0 <, \lambda, \varepsilon < 1$. There is a randomized algorithm that, given a sequence $\mathfrak{A} = (a_1, \ldots, a_m)$ of elements from $[N]$ computes after scanning the*

[1]The reason for the $*$ in the definition is that if we adopt the usual ℓ_p notation, then $F_\infty = \lim_{k \to \infty} (F_k)^{1/k}$ whereas the F_i are not defined with a k^{th} power.

sequence in one pass, a number Y such that the probability that Y deviates from F_k by more than λF_k is at most ε. Most importantly, the algorithm only uses

$$O\left(\frac{k\log(1/\varepsilon)}{\lambda^2}N^{1-1/k}(\log N+\log m)\right)$$

memory bits.

Proof. First, note that the statement does not require that we know the size of the sequence $\mathfrak{A}$ in advance. But for starters, let us assume that m is known. Since we seek a randomized algorithm, the key first step is to *identify a random variable whose expected value is the parameter of interest*, viz., F_k for each k. A first natural guess is to do the following. Pick p uniformly from $[m]$ and consider $R := |\{q \geq p : a_q = a_p\}|$. Since we seek to estimate F_k, the first natural choice is the random variable $X := mR^k$. But a quick check reveals why it is not good enough, and also how one can fix it. Indeed, suppose the element $i \in [N]$ occurs in positions $i_1 < \cdots < i_u$ for some u. The contribution from the element i towards F_k is u^k. However, in computing the expected value of X, the contribution instead is $u^k + (u-1)^k + \cdots + 1^k$, so a fix for this would be to let $X = m(R^k - (R-1)^k)$. Then

$$\begin{aligned}
\mathbb{E}(X) &= (m_1^k - (m_1-1)^k) + ((m_1-1)^k - (m_1-2)^k) + \cdots + (2^k - 1^k) \\
&+ ((m_2)^k - (m_2-1)^k) + \cdots + (2^k - 1^k) + \cdots \\
&+ (m_N^k - (m_N-1)^k) + ((m_N-1)^k - (m_N-2)^k) + \cdots + (2^k - 1^k) \\
&= F_k
\end{aligned}$$

as desired. Also, if we pre-process the random choice prior to the one pass, then the number of storage bits needed is at most $O(\log N + \log m)$ bits that are needed to keep track of the element a_p and the number of occurrences of a_p starting from position p in $\mathfrak{A}$.

To see how good an estimate X is, we follow the maxim in the epigraph of this chapter: *After you have computed the expectation of a random variable, you should try to compute the variance.* Towards that end, we see

$$\begin{aligned}
\mathbb{E}(X^2) &= \frac{m^2}{m}\sum_{i=1}^{N}\left((m_i^k - (m_i-1)^k)^2 + \cdots + (2^k-1^k)^2 + 1^{2k}\right) \\
&\leq m\sum_{i=1}^{N}\left(km_i^{k-1}(m_i^k - (m_i-1)^k) + \cdots + k2^{k-1}(2^k-1^k) + k1^{2k-1}\right) \\
&\leq km\sum_{i=1}^{N} m_i^{2k-1} \\
&= kF_1F_{2k-1}
\end{aligned}$$

where we basically use the fact that

$$a^k - b^k = (a-b)\sum_{i=0}^{k-1} a^{k-i}b^i \leq (a-b)ka^{k-1}$$

for positive reals $a > b > 0$. To bound this further, let $M = \max_{1\leq i\leq N} m_i$. Then

$$\begin{aligned} F_1 F_{2k-1} &\leq F_1 M^{k-1} F_k \\ &\leq F_1 \left(\sum_{i=1}^{N} m_i^k\right)^{(k-1)/k} F_k \\ &\leq N^{1-1/k} F_k^{1/k} F_k^{2-1/k} \\ &= N^{1-1/k} F_k^2 \end{aligned}$$

where in the penultimate line we use the power mean inequality:

$$\frac{1}{N}\sum_i m_i \leq \left(\frac{1}{N}\sum_i m_i^k\right)^{1/k}.$$

Hence if we were to sample $X_1, \ldots, X_s$ independently as above (for some s to be determined) and then take X to be their mean, then by Chebyshev,

$$\mathbb{P}(|X - F_k| \geq \lambda F_k) \leq \frac{\mathrm{Var}(X)}{\lambda^2 F_k^2} \leq \frac{N}{\lambda^2 s N^{1/k}}$$

so that if $s = N^{1-1/(2k)}$ we already have a saving in the number of memory bits since we still only require $O(s(\log N + \log m))$ bits of memory. But one can do better; let $(X_1, \ldots, X_s)$ be independently sampled according to X as above, and let Y be their mean - only this time, we take $s = \frac{Cn^{1-1/k}}{\lambda^2}$ for some constant C, which does not quite give the high probability estimate we want but rather $\mathbb{P}(|Y - F_k| > \lambda F_k) \leq 1/C$. But repeat this process r times (for some r to be determined), and then report the value $Z = \mathrm{Median}(Y_1, \ldots, Y_r)$.

Define the random variable $\tilde{Y_i}$ to equal 1 if $Y_i \in [F_k - \lambda F_k, F_k + \lambda F_k]$ and zero otherwise and let $\tilde{Z} := \sum_{i=1}^{r} \tilde{Y_i}$, so that we can bound tail probabilities of $\tilde{Z}$ by the distribution of the Binomial variable $\mathrm{Bin}(r, 1 - 1/C)$. If Z lies outside $[F_k - \lambda F_k, F_k + \lambda F_k]$ then $\tilde{Z}$ is less than $r/2$. If $C = 8$, say, one can again use the Chebyshev bound (we omit these details) to show that for $r = O(1/\sqrt{\varepsilon})$ one has $\mathbb{P}(|Z - F_k| > \lambda F_k) \leq \varepsilon$.

But again, this is not optimal; the Binomial distribution approximates the Gaussian random variable for sufficiently large r, so deviation from the expectation is an exponentially unlikely event. A more precise form of this appears in the next chapter which establishes exponential decay away from the expected value for the Binomial distribution. Thus, (and these details will be more clear in the next chapter) one can take

$r = O(\log(1/\varepsilon))$ and the proof is complete.

The last point is to deal with this when m is not known *apriori*. In that case, start with $m = 1$, and choose a_p as in the randomized algorithm stated above. If m is not one, we update $m = 2$, and replace $p = 1$ with $p = 2$ with probability $1/2$. More generally, having reached m', if $m > m'$ then we replace p with $m' + 1$ with probability $\frac{1}{m'+1}$. It is not hard to see that this keeps the argument intact and the implementation still only needs $O(\log m + \log N)$ bits. ■

Remark: We have throughout assumed implicitly, that m is not much larger than a polynomial in N, but if m grows, say exponentially with N, then there are older results that give a similar saving in memory.
The paper [3] includes several other interesting results. For instance, they show that the space complexity results here are almost best possible: for $k \geq 6$, randomized algorithms need at least $\Omega(n^{1-5/k})$ memory bits, and that the estimation of F_∞^* requires $\Omega(N)$ bits. Another beautiful result is that the estimation of F_2 can actually be done with only $O(\log N)$ memory bits. This uses the fact that there is a simple deterministic construction of a set (using what are known as BCH codes) of $O(N^2)$ {-1,1}-valued vectors of length N which are *four-wise independent*, i.e., any 4 of the coordinates are independently distributed amongst the possible 4-tuples of $\{-1, 1\}$. If $\mathfrak{v} := (v_1, \ldots, v_N)$ is randomly picked from the above set, define $X := \langle \mathfrak{v}, \mathfrak{A} \rangle = \sum_{i=1}^{N} v_i a_i$. For the details and other interesting results, we refer the reader to the paper [3].

6.5 Uniform Dilations

We start with some definitions. By $\mathbb{T}$ we mean the 1-dimensional torus, i.e., $\mathbb{T} := \mathbb{R}/\mathbb{Z}$. For $0 < \varepsilon < 1$, a subset $X \subset \mathbb{T}$ is called *ε-dense* if it meets every interval in $\mathbb{T}$ of length ε. A *dilation* of $X \subset \mathbb{T}$ is $nX := \{nx : x \in X\}$ for some integer n where the multiplication is performed modulo one, i.e., in $\mathbb{T}$. For any $\varepsilon > 0$, Berend and Peres defined the integer $k(\varepsilon)$ to be the least integer k such that for any $X \subset \mathbb{T}$ of size at least k, there is some dilation nX which is ε-dense. Moreover, they showed that

$$\Omega(1/\varepsilon^2) \leq k(\varepsilon) \leq O(1/\varepsilon)^{O(1/\varepsilon)}$$

and posed the question of determining the correct order of $k(\varepsilon)$. This was achieved by Alon and Peres, who proved the following theorem.

Theorem 41. *Given $\gamma > 0$, there exists $\varepsilon_0 = \varepsilon_0(\gamma)$ such that for $\varepsilon < \varepsilon_0$, every set $T \subset \mathbb{T}$ of cardinality at least $\varepsilon^{-(2+\gamma)}$ has an ε-dense dilation nX. In other words,*

$$\Omega(\varepsilon^{-2}) \leq k(\varepsilon) \leq \varepsilon^{-(2+\gamma)}$$

for all $\gamma > 0$.

We will see a proof of this in the special case when X consists entirely of rationals with the same prime denominator p so that every element in X is of the form x/p. Under this assumption, we first observe that the problem reduces to considering dilations of subsets in the finite field $\mathbb{F}_p$. In this case, the lower bound is of th eright order upto a multiplicative constant.

To state the precise form of the stronger result, suppose p is a prime and $\varepsilon > 0$. Define $k(\varepsilon, p)$ to be the minimum integer k such that the following holds. For every subset $X \subset \mathbb{F}_p$ of size at least k, there exists n such that nX intersects every interval of size at least εp. Here, by an interval of length r we mean a set of the form $\{a, a+1, \ldots, a+r-1\}$. In words, this states that when a set is of size then some dilate of this set is fairly well-spread out, and so touches all the intervals of length εp.

Here is the theorem of Alon and Peres [4] in its exact form.

Theorem 42. *For every prime p and $0 < \varepsilon < 1$ for which εp is an integer,*

$$k(\varepsilon, p) \leq \frac{4}{\varepsilon^2}.$$

Proof. We shall omit floors and ceilings in our presentation below for convenience. Let $X := \{x_1, \ldots, x_k\} \subset \mathbb{F}_p^*$. At first glance, the main issue is that there are p different intervals of length εp while any dilate aX contains only $O(\varepsilon^{-2})$ elements, so it seems quite a task for the dilate to meet every interval of size εp. But a simple trick makes this task realistically feasible. Set $s = 2/\varepsilon$ and let $I_1, \ldots, I_s$ be disjoint intervals of length $\frac{\varepsilon p}{2}$ that partition $\mathbb{F}_p$. Since each interval I of length εp *necessarily contains one of the* I_i, it suffices to show that there exists $a \in \mathbb{F}_p$ such that $|aX \cap I_i| > 0$ for each of these intervals.

As with our previous rules of thumb, it is quite natural to try a random dilate of X. Let $a \in \mathbb{F}_p$ be a random element and for a fixed interval I_i in the partition above, observe that

$$\mathbb{E}(|aX \cap I_i|) = \sum_{x \in I_i} \mathbb{P}(a \in \{x/x_1, \ldots, x/x_k\}) = \frac{\varepsilon k}{2}.$$

As before, to compute the variance, we have

$$\mathbb{E}(|aX \cap I_i|^2) = \sum_{x,y \in I_i} \mathbb{P}(x, y \in aX) = \frac{\varepsilon k}{2} + \sum_{\substack{x \neq y \\ x,y \in I_i}} \mathbb{P}(x, y \in aX)$$

and the last double sum poses a bit of a problem. The brilliant idea of Alon and Peres was to modify the random process so as to make *this sum computable*. And to do that, they consider affine translates of the set as well.

Indeed, pick $a, b \in \mathbb{F}_p$ independently and consider the set $aX + b$. The key point is that if there are a, b such that $aX + b$ meets every interval of size εp, then since translates

of intervals are again intervals, the same holds for aX as well! Furthermore, for $x \neq y$ the events $x \in aX + b$ and $y \in aX + b$ are *pairwise independent*, so we have

$$\mathbb{E}(|(aX+b) \cap I_i|) = \sum_{x \in I_i} \mathbb{P}(x, y \in aX + b) = \frac{\varepsilon k}{2}$$

$$\begin{aligned} \mathrm{Var}(|(aX+b) \cap I_i|) &= \sum_{x \in I_i} \mathrm{Var}(\mathbf{1}_{x \in aX+b}) \\ &\leq \frac{\varepsilon k}{2}. \end{aligned}$$

Hence by Chebyshev,

$$\mathbb{P}((aX+b) \cap I_i = \emptyset) < \frac{2}{\varepsilon k}$$

which implies that

$$\mathbb{P}(aX + b \cap I_i = \emptyset \text{ for some } 1 \leq i \leq s) \leq \frac{4}{\varepsilon^2 k} < 1$$

for k as in the theorem. ■

Remark: To move from this special case to all intervals in $\mathbb{T}$ as in the main theorem stated at the beginning of this section, the idea is to pick a large enough A, and pick $a \in \{1, \ldots, A\}$ randomly, and $b \in \mathbb{T}$ uniformly, and consider the affine translate $aX + b$ as before. As before, we shall fix a partition of $\mathbb{T}$ into intervals of size $\varepsilon/2$ and fix such an interval I. Again, $\mathbb{E}(|(aX + b) \cap I| = \varepsilon k/2$, but now computing the variance is a little more complicated. If we write $X = \{x_1, \ldots, x_k\}$ then it is not hard to show that computing the variance of the aforementioned random variable can be estimated in terms of the differences $x_i - x_j$. The technical ideas deal with how one can control the covariance terms, and this involves a few more subtleties than the simple result above.
The paper [4] contains several other general results on when one can find dilations that are ε-dense. For instance, they also show that one can pick a dilate n which is prime for which nX is ε-dense, and further, if $k \geq \varepsilon^{-(4+\gamma)}$ then there is an ε-dense dilation of the form n^2X as well. We direct the interested reader to the paper [4] for other results.

6.6 Resolution of the Erdős-Hanani Conjecture: The Rödl 'Nibble'

One of the most effective proof techniques in Combinatorics is the method of induction. How would the method of induction blend with the probabilistic method? How does one effectively carry out an 'inductive probabilistic method proof'?

The *Rödl 'Nibble'* refers to a probabilistic paradigm (pioneered by Vojtech Rödl) for a host of applications in which a desirable combinatorial object is constructed via a random process, through a series of several small steps, with a certain amount of control over each step. Subsequently, researchers realized that Rödl's method can be extended as a

paradigm to a host of other constructions, particularly for coloring problems in graphs, and matchings/coverings problems in hypergraphs. Indeed, the proof of Erdős-Hanani conjecture - the result that launched the Rödl Nibble - is an instance of a covering problem of a specific hypergraph. In this section, we shall see a resolution of the Erdős-Hanani conjecture following a latter simplification by Pippenger and Spencer [24].

We start with a definition. As always, $[n]$ denotes the set $\{1, \ldots, n\}$. Suppose $r, t \in \mathbb{N}$. An *r-uniform covering* for $\binom{[n]}{t}$ is a collection $\mathcal{A}$ of r-subsets of $[n]$ such that for each t-subset $T \in \binom{[n]}{t}$, there exists an $A \in \mathcal{A}$ such that $T \subset A$. An *r-uniform packing* for $\binom{[n]}{t}$ is a collection $\mathcal{A}$ of r-subsets of $[n]$ such that for each t-subset $T \in \binom{[n]}{t}$, there exists at most one $A \in \mathcal{A}$ such that $T \subset A$.

When $t = 1$, if r divides n, then there obviously exists a collection $\mathcal{A}$ of r-subsets of $[n]$, $|\mathcal{A}| = n/r$, such that $\mathcal{A}$ is both an r-uniform covering and packing for $\binom{[n]}{1} = [n]$. In general, there exists a covering of size $\lceil n/r \rceil$ and a packing of size $\lfloor n/r \rfloor$.

Let $M(n, r, t)$ be the size of a minimum covering, and $m(n, r, t)$ be the size of a maximum packing. A simple combinatorial counting argument shows that

$$m(n, r, t) \leq \frac{\binom{n}{t}}{\binom{r}{t}} \leq M(n, r, t).$$

Indeed, if one were to consider a covering, then for each t-subset, there is at least one r-subset containing it; conversely, each r-subset contains $\binom{r}{t}$ t-subsets, so the first inequality is obtained. The argument is similar for maximal packings. It then seems natural to ask if there exists a collection $\mathcal{A}$ of r-subsets of $[n]$ with size $|\mathcal{A}| = \binom{n}{t}/\binom{r}{t}$ such that $\mathcal{A}$ is both an r-uniform covering and packing for $\binom{[n]}{t}$. This is called a $t - (n, r, 1)$ *design* and is also referred to as a *Steiner t-design*.

In the 60s, Erdős and Hanani proved that

$$\lim_{n\to\infty} \frac{M(n, r, 2)}{\binom{n}{2}/\binom{r}{2}} = \lim_{n\to\infty} \frac{m(n, r, 2)}{\binom{n}{2}/\binom{r}{2}} = 1.$$

and further conjectured that this is true for all positive integers $r \geq t$. In a sense, the conjecture posits that as n grows large, one gets more room to attempt to fit these r-subsets to cover all t-subsets, so as n gets larger, one ought to be able to get closer to as tight a packing (or covering) as one can. This conjecture was settled affirmatively by Vojtech Rödl in 1985.

Here, we consider a more general problem. Suppose $r \geq 2$ is a fixed integer. By an r-uniform hypergraph we mean a hypergraph $\mathcal{H} = (V, \mathcal{E})$ on a set V of size n such that

each $e \in \mathcal{E}$ has size r. The *degree* of a vertex in a hypergraph is the same as the one we have encountered in the graph case, i.e., $d(x) = |\{E \in \mathcal{E} : E \ni x\}|$. Given an r-uniform hypergraph $\mathcal{H}$ on n vertices which is D-regular for some D, i.e. $d(x) = D$ for all $x \in V$, we seek a covering (resp. a packing) of $\mathcal{H}$ which is as tight as possible, i.e., a covering (resp. packing) of size approximately n/r. This more general question subsumes the Erdős-Hanani question: consider the hypergraph $\mathcal{H} = (V, \mathcal{E})$ where $V = \binom{[n]}{t}$ and the edges of $\mathcal{H}$ correspond to r-subsets of $[n]$ with each such r-subset E containing all the vertices x that correspond to t-subsets of E. It is easy to see that this is an $\binom{r}{t}$-uniform regular hypergraph with degree $D = \binom{n-t}{r-t}$.

Let $\varepsilon > 0$. Note that in this new formulation, if we can find a packing of size $\frac{(1-\varepsilon)n}{r}$, then there are at most εn vertices uncovered. Hence, we can find a covering of size $\frac{(1-\varepsilon)n}{r} + \varepsilon n = (1 - (r-1)\varepsilon)\frac{n}{r}$. On the other hand, if we can find a covering $\mathcal{A}$ of size $\frac{(1+\varepsilon)n}{r}$, then for every x which is covered by $d(x)$ hyperedges, we delete $d(x) - 1$ of them. The number of deleted edges is at most

$$\sum_{x \in V}(d(x) - 1) = \sum_{x \in V} d(x) - \sum_{x \in V} 1 = |\{(x, E) : E \in \mathcal{A}\}| - n = \left(\frac{(1+\varepsilon)n}{r}\right) \cdot r - n = \varepsilon n$$

so there is a packing of size at least $\frac{(1+\varepsilon)n}{r} - \varepsilon n = (1 - (r-1)\varepsilon)\frac{n}{r}$. The upshot is:

Finding a covering of size approximately $\frac{n}{r}$ is equivalent to finding a packing of size approximately $\frac{n}{r}$.

We shall try to obtain a covering $\mathcal{A}$ of size $\leq (1+\varepsilon)\frac{n}{r}$ for n sufficiently large. Since we seek a covering, we do not have to worry if some vertex is covered more than once.

Let us try a simple probabilistic idea first to see what its shortcomings are. Suppose we decide to pick each edge of the hypergraph $\mathcal{H}$ independently with probability p. We seek a collection $\mathcal{E}^*$ with $|\mathcal{E}^*| \approx \frac{n}{r}$; if we start with an almost regular graph of degree D, then $r|\mathcal{E}| \approx nD$, so that implies that we need $p \approx \frac{1}{D}$. Let us see how many vertices get left behind by this probabilistic scheme. A vertex x is uncovered only if every edge containing it is discarded. In other words, the probability that a vertex x gets left behind is approximately $(1 - 1/D)^D \approx 1/e$. This is now a problem because this implies that the expected number of vertices that go uncovered is approximately n/e which is a *constant proportion* of the total number of vertices.

Rödl's idea was to, as we described in the beginning of this section, attempt an inductive procedure: We pick a small number of edges, so that the rest of $\mathcal{H}$ is as 'close as possible' to the original one. If the inductive procedure were to take over for the modified hypergraph, then after several "nibbles" into the hypergraph we are left with a very small proportion of the vertex set that is yet uncovered. But for these, we pick an arbitrary edge for each of these vertices to get a covering for the entire vertex set.

However, note that after each step, some of the regularity conditions of the hypergraph are bound to be violated, so for the inductive procedure to apply to the smaller hypergraph the hypotheses would have to be milder. We will get to this point momentarily.

As $\mathcal{H}$ is D-regular, $r|\mathcal{E}| = |\{(x, E) : x \in E \in \mathcal{E}\}| = nD \Rightarrow |\mathcal{E}| = \frac{nD}{r}$. Since our modified hypergraph ought to be as close as possible to the original hypergraph, we need to pick a few - *but not too few!* - edges in the first step. If we want to pick about $\frac{\varepsilon n}{r}$ edges, we will need $\mathbb{P}(E \text{ is picked}) = \frac{\varepsilon}{D}$.

This sets our paradigm into motion. In the first 'step', each edge $E \in \mathcal{E}$ is picked independently with probability $p = \frac{\varepsilon}{D}$. If $\mathcal{E}^*$ is the set of chosen edges then we have

$$\mathbb{E}[|\mathcal{E}^*|] = \frac{\varepsilon n}{r}.$$

Also, the probability that a vertex x is not covered in this process is $(1 - \varepsilon/D)^{d(x)} \approx e^{-\varepsilon}$.

In the rest of this section, and also in subsequent chapters, we shall adopt Pippenger and Spencer's wonderfully terse notation. We shall write $x = a \pm b$ to mean $x \in (a - b, a + b)$. Also, since there will be many constants that keep popping up, we shall throw in new variables to denote various small quantities which can all be tied down eventually, if need be.

Getting back to our first step, after a 'nibble', the rest of the hypergraph is no longer regular, so as mentioned earlier, we need to make the hypotheses milder, so we propose: Given an r-uniform hypergraph $\mathcal{H}$ on n vertices such that $d(x) = D(1 \pm \delta)$ for all $x \in V$ for some small $\delta > 0$, we want to find a covering of size $\approx \frac{n}{r}$. For this to work, our first step necessarily has to reduce the degrees of *all* the vertices that are not covered during round one, and that is still a little too strong. So, the hypothesis needs to be milder:

Given an r-uniform hypergraph $\mathcal{H}$ on n vertices such that except at most δn vertices, $d(x) = D(1 \pm \delta)$ for other vertices $x \in V$.

So under the milder hypothesis we wish to find a collection of edges $\mathcal{E}^*$ such that

1. $|\mathcal{E}^*| = \frac{\varepsilon n}{r}(1 + \delta')$,

2. $|V^*| = ne^{-\varepsilon}(1 \pm \delta'')$ where $V^* := V \backslash \Big(\bigcup_{E \in \mathcal{E}^*} E \Big)$,

3. For all $x \in V^*$ except at most $\delta'''|V^*|$ of the vertices, if $d^*(x)$ denotes its degree in the residual hypergraph then $d^*(x) = D(1 \pm \mu)$.

To explain this requirement, let $\mathbf{1}_x = \mathbf{1}_{\{x \notin \text{ any edge of } \mathcal{E}^*\}}$. If each edge is picked independently with probability ε/D then

$$\mathbb{E}(|V^*|) = \sum_{x \in V} \Big(1 - \frac{\varepsilon}{D}\Big)^{d(x)} \approx n(1-\delta)\Big(1 - \frac{\varepsilon}{D}\Big)^{D(1\pm\delta)} \approx n(1-\delta)e^{-\varepsilon(1\pm\delta)} \approx ne^{-\varepsilon}(1 \pm \delta'').$$

Furthermore,

$$\begin{aligned} \mathrm{Var}(|V^*|) &= \mathrm{Var}\Big(\sum_{x \in V} \mathbf{1}_x\Big) \\ &= \sum_{x \in V} \mathrm{Var}(\mathbf{1}_x) + \sum_{x \neq y} \mathrm{Cov}(\mathbf{1}_x, \mathbf{1}_y) \end{aligned}$$

If $d(x) = D(1 \pm \delta)$ and $d(y) = D(1 \pm \delta)$, then

$$\begin{aligned} \mathrm{Cov}(\mathbf{1}_x, \mathbf{1}_y) &= \mathbb{E}[\mathbf{1}_{x,y}] - \mathbb{E}[\mathbf{1}_x]\mathbb{E}[\mathbf{1}_y] \\ &= \Big(1 - \frac{\varepsilon}{D}\Big)^{d(x)+d(y)-d(x,y)} - \Big(1 - \frac{\varepsilon}{D}\Big)^{d(x)+d(y)} \\ &= \Big(1 - \frac{\varepsilon}{D}\Big)^{d(x)+d(y)} \left(\Big(1 - \frac{\varepsilon}{D}\Big)^{-d(x,y)} - 1\right) \\ &\approx e^{-2\varepsilon(1\pm\delta)}(e^{-\frac{\varepsilon}{D}d(x,y)} - 1) \end{aligned}$$

where $d(x,y)$ denotes the *codegree* of x and y, i.e., $d(x,y) = |\{E : x, y \in E\}|$.

Note that $e^{-\frac{\varepsilon}{D}d(x,y)} - 1$ is very small provided that $d(x,y) \ll D$. This is true in the original Erdős-Hanani problem, where $V = \binom{[n]}{t}$, since $D = \binom{n-t}{r-t} = O(n^{r-t})$, while $d(x,y) = \binom{n-|T_1 \cup T_2|}{r-|T_1 \cup T_2|} \leq \binom{n-t-1}{r-t-1} = O(n^{r-t-1}) \ll D$, where x and y corresponds to t-subsets T_1 and T_2 respectively.

Before we make our speculation outright, there is one more aspect that suggests that the hypothesis should be made milder. Suppose that for some vertex x the degree of x is super large, i.e., $D = o(d(x))$. Since we wish to retain the r-uniformity of the hypergraphs, our process would entail throwing away all edges that intersect some vertex of $V \setminus V^*$ to get to the modified hypergraph. But if $d(x)$ is very large, then it is somewhat likely that some edge containing x is chosen, and since x would get picked, all edges that contain x will have to be discarded to get to the residual hypergraph, and we may lose too many edges in this process. So, to prevent this, we may want $d(x) = O(D)$ for all x. This motivates the following tentative statement:

Lemma 43. *('Nibble' lemma) Suppose $r \geq 2$ is a positive integer, and $k, \varepsilon, \delta^* > 0$ are given. Then there exist $\delta_0(r,k,\varepsilon,\delta^*) > 0$ and $D_0(r,k,\varepsilon,\delta^*)$ such that for all $n \geq D \geq D_0$ and $0 < \delta \leq \delta_0$, if $\mathcal{H}$ is an r-uniform hypergraph on n vertices satisfying*

(i) except at most δn vertices, $d(x) = D(1 \pm \delta)$ for other vertices $x \in V$,

(ii) $d(x) < kD$ *for all* $x \in V$,

(iii) $d(x,y) < \delta D$,

then there exists $\mathcal{E}^* \subset \mathcal{E}$ *such that*

(a) $|\mathcal{E}^*| = \frac{\varepsilon n}{r}(1 \pm \delta^*)$;

(b) $|V^*| = ne^{-\varepsilon}(1 \pm \delta^*)$, *where* $V^* = V \backslash \Big(\bigcup_{E \in \mathcal{E}^*} E \Big)$;

(c) Except at most $\delta^*|V^*|$ *vertices,* $d^*(x) = De^{-\varepsilon(r-1)}(1 \pm \delta^*)$, *where* d^* *is the degree on the induced hypergraph on* V^*.

We say that $\mathcal{H}$ is an (n,k,D,δ)-hypergraph when (i), (ii) and (iii) hold for $\mathcal{H}$. This lemma says that if $\mathcal{H}$ is an (n,k,D,δ)-hypergraph then it contains an induced (n^*,k^*,D^*,δ^*) -hypergraph $\mathcal{H}^*$ where

$$\begin{aligned} \delta^* &= \delta e^{\varepsilon(r-1)}, \\ n^* &= ne^{-\varepsilon}(1 \pm \delta^*), \\ k^* &= ke^{\varepsilon(r-1)}, \\ D^* &= De^{-\varepsilon(r-1)}(1 \pm \delta^*) \end{aligned}$$

To see why these are the relevant new parameters, consider for instance, the parameter δ;

$$d^*(x,y) \le d^*(x,y) < \delta D = \delta^* D^* \text{ forces } \delta = \frac{\delta^* D e^{-\varepsilon(r-1)}}{D} = \delta^* e^{-\varepsilon(r-1)}$$

and that gives $\delta^* = \delta e^{\varepsilon(r-1)}$. Similarly for the parameter k, $d^*(x) \le d(x) < kD = k^* D^*$ forces $k^* = \frac{kD}{De^{-\varepsilon(r-1)}} = ke^{\varepsilon(r-1)}$.

Let us see if this lemma is good enough to take us through. If we repeat the nibble t times (where t shall shortly be determined) then we have $\delta = \delta_0 < \delta_1 < \cdots < \delta_t$ with $\delta_i = \delta_{i-1}e^{\varepsilon(r-1)}$, and $\mathcal{H} = \mathcal{H}_0 \supset \mathcal{H}_1 \supset \cdots \supset \mathcal{H}_t$. Note that this establishes a cover of size $\sum_{i=1}^{t-1} |\mathcal{E}_i| + |V_t|$ where

$$|V_i| = |V_{i-1}|e^{-\varepsilon}(1 \pm \delta_i) \le ne^{-\varepsilon i} \prod_{j=1}^{i} (1 + \delta_j)$$

and

$$|\mathcal{E}_i| = \frac{\varepsilon |V_{i-1}|}{r}(1 \pm \delta_i) \le \frac{\varepsilon n e^{-\varepsilon(i-1)}}{r} \prod_{j=1}^{i} (1 + \delta_j),$$

so the size of the cover is

$$\begin{aligned}\sum_{i=1}^{t-1}|\mathcal{E}_i|+|V_t| &\leq \left(\sum_{i=1}^{t-1}\frac{\varepsilon n e^{-\varepsilon(i-1)}}{r}\right)\prod_{i=1}^{t}(1+\delta_i)+ne^{-\varepsilon t}\prod_{i=1}^{t}(1+\delta_i)\\ &=\left(\prod_{i=1}^{t}(1+\delta_i)\right)\frac{n}{r}\left(\sum_{i=1}^{t}\varepsilon e^{-\varepsilon(i-1)}+re^{-\varepsilon t}\right)\\ &\leq\left(\prod_{i=1}^{t}(1+\delta_i)\right)\frac{n}{r}\left(\frac{\varepsilon}{1-e^{-\varepsilon}}+re^{-\varepsilon t}\right).\end{aligned}$$

Pick t such that $e^{-\varepsilon t}<\varepsilon$ - for instance take $t=2\varepsilon^{-1}\log(1/\varepsilon)$. For this t, pick δ small enough such that $\prod_{i=1}^{t}(1+\delta_i)\leq 1+\varepsilon$. Since $\lim_{\varepsilon\to 0}\frac{\varepsilon}{1-e^{-\varepsilon}}=1$, the limit of this expression goes to $\frac{n}{r}$ as $\varepsilon\to 0$. Therefore, all that remains is to prove the 'Nibble' Lemma 43.

Proof. (Proof of Lemma 43) We will use subscripts $\delta_{(i)}$ to denote various small constants. Keeping with the probabilistic paradigm, we pick each edge of $\mathcal{H}$ independently with probability $\frac{\varepsilon}{D}$. Let $\mathcal{E}^*$ be the set of picked edges.

We say $x\in V$ is *good* if $d(x)=(1\pm\delta)D$, else we say that x is *bad*. Note that

$$|\{(x,E):x\in E\}|\geq|\{(x,E):x\text{ good}\}|>(1-\delta)D\cdot(1-\delta)n=(1-\delta)^2Dn.$$

On the other hand,

$$\begin{aligned}|\{(x,E):x\in E\}| &= |\{(x,E):x\text{ good}\}|+|\{(x,E):x\text{ bad}\}|\\ &\leq(1+\delta)D\cdot n+kD\cdot\delta n.\end{aligned}$$

So

$$\begin{aligned}\frac{(1-\delta)^2Dn}{r} &\leq |\mathcal{E}|\leq\frac{Dn}{r}(1+(k+1)\delta)\text{ which gives}\\ |\mathcal{E}| &= \frac{Dn}{r}(1\pm\delta_{(1)}).\end{aligned}$$

Hence

$$\mathbb{E}[|\mathcal{E}^*|]=\sum_{E\in\mathcal{E}}\mathbb{P}(E\text{ is picked})=\frac{\varepsilon}{D}\frac{Dn}{r}(1\pm\delta_{(1)})=\frac{\varepsilon n}{r}(1\pm\delta_{(1)}).$$

Let $\mathbf{1}_E=\mathbf{1}_{\{E\text{ is picked}\}}$. By independence, $\mathrm{Var}(|\mathcal{E}^*|)=\sum_{E\in\mathcal{E}}\mathrm{Var}(\mathbf{1}_E)\leq\mathbb{E}[|\mathcal{E}^*|]$. By Chebyshev's inequality, we get $\mathbb{P}\left(\left||\mathcal{E}^*|-\mathbb{E}[|\mathcal{E}^*|]\right|>\delta_{(2)}\mathbb{E}[|\mathcal{E}^*|]\right)\leq\frac{\mathrm{Var}(|\mathcal{E}^*|)}{\delta_{(2)}^2\mathbb{E}[|\mathcal{E}^*|]^2}$. So if $n\gg 0$, then

$$|\mathcal{E}^*|=\frac{\varepsilon n}{r}(1\pm\delta_{(1)})(1\pm\delta_{(2)})=\frac{\varepsilon n}{r}(1\pm\delta^*)$$

with high probability, yielding (a).

Let $\mathbf{1}_x = \mathbf{1}_{\{x \notin \text{ any edge of } \mathcal{E}^*\}}$. Note that

$$\mathbb{E}[|V^*|] = \sum_{x \in V} \left(1 - \frac{\varepsilon}{D}\right)^{d(x)} \geq \sum_{x \text{ good}} \left(1 - \frac{\varepsilon}{D}\right)^{D(1+\delta)} \geq e^{-\varepsilon}(1 - \delta_{(3)}) \cdot (1 - \delta)n.$$

On the other hand,

$$\begin{aligned}
\mathbb{E}[|V^*|] &= \sum_{x \text{ good}} \left(1 - \frac{\varepsilon}{D}\right)^{d(x)} + \sum_{x \text{ bad}} \left(1 - \frac{\varepsilon}{D}\right)^{d(x)} \\
&\leq \sum_{x \text{ good}} \left(1 - \frac{\varepsilon}{D}\right)^{D(1-\delta)} + \delta n \\
&\leq e^{-\varepsilon}(1 + \delta_{(4)}) \cdot n + \delta n
\end{aligned}$$

So

$$ne^{-\varepsilon}(1 - \delta_{(3)})(1 - \delta) \leq \mathbb{E}[|V^*|] \leq ne^{-\varepsilon}(1 + \delta_{(4)} + \delta e^{\varepsilon})$$

implying

$$\mathbb{E}[|V^*|] = ne^{-\varepsilon}(1 \pm \delta_{(5)}).$$

Again, as we compute the variance, we see

$$\mathrm{Var}(|V^*|) = \sum_{x \in V} \mathrm{Var}(\mathbf{1}_x) + \sum_{x \neq y} \mathrm{Cor}(\mathbf{1}_x, \mathbf{1}_y) \leq \mathbb{E}[|V^*|] + \sum_{x \neq y} \mathrm{Cor}(\mathbf{1}_x, \mathbf{1}_y)$$

where

$$\begin{aligned}
\mathrm{Cov}(\mathbf{1}_x, \mathbf{1}_y) &= \mathbb{E}[\mathbf{1}_{x,y}] - \mathbb{E}[\mathbf{1}_x]\mathbb{E}[\mathbf{1}_y] \\
&= \left(1 - \frac{\varepsilon}{D}\right)^{d(x)+d(y)-d(x,y)} - \left(1 - \frac{\varepsilon}{D}\right)^{d(x)+d(y)} \\
&= \left(1 - \frac{\varepsilon}{D}\right)^{d(x)+d(y)} \left(\left(1 - \frac{\varepsilon}{D}\right)^{-d(x,y)} - 1\right) \\
&\leq 1 \cdot \left(\left(1 - \frac{\varepsilon}{D}\right)^{-\delta D} - 1\right) \leq e^{\varepsilon\delta} - 1 \text{ which is small.}
\end{aligned}$$

This implies $\mathrm{Var}(|V^*|) = o(\mathbb{E}[|V^*|]^2)$. By Chebyshev's inequality, we get

$$\mathbb{P}\left(\left||V^*| - \mathbb{E}[|V^*|]\right| > \delta_{(6)}\mathbb{E}[|V^*|]\right) \leq \frac{\mathrm{Var}(|V^*|)}{\delta_{(6)}^2 \mathbb{E}[|V^*|]^2}.$$

So if $n \gg 0$, $|V^*| = ne^{-\varepsilon}(1 \pm \delta_{(5)})(1 \pm \delta_{(6)}) = ne^{-\varepsilon}(1 \pm \delta^*)$ with high probability, yielding (b).

To prove (c), suppose x survives after the removal of $\mathcal{E}^*$. Fix an $E \in \mathcal{E}$ such that $E \ni x$. We wish to estimate the probability that E also survives conditioned on the assumption that x survives. Let $\mathcal{F}_E = \{F \in \mathcal{E} : x \notin F, F \cap E \neq \emptyset\}$. Then E survives if and only if $\mathcal{F}_E \cap \mathcal{E}^* = \emptyset$.

Call $E \in \mathcal{E}$ *bad* if E contains at least one bad vertex. Suppose x is good, and E is good. Then

$$\begin{aligned}\mathbb{P}(E \text{ survives} \mid x \text{ survives}) &= \left(1 - \frac{\varepsilon}{D}\right)^{(r-1)(1\pm\delta)D - \binom{r-1}{2}\delta D} (1 \pm \delta_{(7)}) \\ &= \left(1 - \frac{\varepsilon}{D}\right)^{(r-1)D} (1 \pm \delta_{(8)}).\end{aligned}$$

Let $\mathrm{Bad}(x) := \{E : E \text{ is bad and does not contain } x\}$. If $|\mathrm{Bad}(x)| < \delta_{(9)}D$, then $\mathbb{E}[d^*(x)] = De^{-\varepsilon(r-1)}(1 \pm \delta_{(10)})$.

Now, the question is: how many x have $|\mathrm{Bad}(x)| \geq \delta_{(9)}D$? Call x *Incorrigible* if x is good but $|\mathrm{Bad}(x)| \geq \delta_{(9)}D$. We now want to bound the size of $V_{\mathrm{INCOR}} := \{x \in V : x \text{ is incorrigible}\}$. Note that

$$|\{(x, E) : |\mathrm{Bad}(x)| \geq \delta_{(9)}D\}| \geq \delta_{(9)}D \cdot |V_{\mathrm{INCOR}}|.$$

On the other hand,

$$\begin{aligned}|\{(x, E) : |\mathrm{Bad}(x)| \geq \delta_{(9)}D\}| &\leq |\{(x, E) : E \text{ is bad}\}| \\ &\leq r|\{(x, E) : x \text{ is bad}\}| \\ &\leq r(kD)(\delta n).\end{aligned}$$

Hence, $|V_{\mathrm{INCOR}}| := \delta^* n = \frac{r(\delta n)k}{\delta_{(9)}}$. Therefore, except at most $\delta^* n$ vertices, the remaining vertices x satisfy $\mathbb{E}[d^*(x)] = De^{-\varepsilon(r-1)}(1 \pm \delta_{(10)})$.

Let $\mathbf{1}_E = \mathbf{1}_{\{E \text{ survives}\}}$. For those x that are neither incorrigible nor bad,

$$\begin{aligned}\mathrm{Var}(d^*(x)) &= \sum_{E \in \mathcal{E}} \mathrm{Var}(\mathbf{1}_E) + \sum_{E \neq F} \mathrm{Cov}(\mathbf{1}_E, \mathbf{1}_F) \\ &\leq \mathbb{E}[d^*(x)] + \sum_{E \neq F \text{ good}} \mathrm{Cov}(\mathbf{1}_E, \mathbf{1}_F) + \delta_{(9)}D \cdot (1+\delta)D \cdot 1 \\ &\leq \mathbb{E}[d^*(x)] + \sum_{\substack{E \neq F \text{ good} \\ E \cap F = \{x\}}} \mathrm{Cov}(\mathbf{1}_E, \mathbf{1}_F) + \sum_{\substack{E \neq F \text{ good} \\ |E \cap F| > 1}} \mathrm{Cov}(\mathbf{1}_E, \mathbf{1}_F) \\ &\quad + \delta_{(9)}(1+\delta)D^2 \\ &\leq \mathbb{E}[d^*(x)] + \sum_{\substack{E \neq F \text{ good} \\ E \cap F = \{x\}}} \mathrm{Cov}(\mathbf{1}_E, \mathbf{1}_F) + (r-1)\delta D \cdot (1+\delta)D \cdot 1 \\ &\quad + \delta_{(9)}(1+\delta)D^2.\end{aligned}$$

Now, denote by $\mathcal{F}_E$ the collection of those edges that intersect E non-trivially. Then,

$$\begin{aligned}
\mathrm{Cov}(\mathbf{1}_E, \mathbf{1}_F) &= \mathbb{E}[\mathbf{1}_{E,F}] - \mathbb{E}[\mathbf{1}_E]\mathbb{E}[\mathbf{1}_F] \\
&= \left(1 - \frac{\varepsilon}{D}\right)^{|\mathcal{F}_E \cup \mathcal{F}_F|} - \left(1 - \frac{\varepsilon}{D}\right)^{|\mathcal{F}_E| + |\mathcal{F}_F|} \\
&= \left(1 - \frac{\varepsilon}{D}\right)^{|\mathcal{F}_E| + |\mathcal{F}_F|} \left(\left(1 - \frac{\varepsilon}{D}\right)^{-|\mathcal{F}_E \cap \mathcal{F}_F|} - 1\right) \\
&\leq \left(1 - \frac{\varepsilon}{D}\right)^{-|\mathcal{F}_E \cap \mathcal{F}_F|} - 1 \\
&\leq \left(1 - \frac{\varepsilon}{D}\right)^{-(r-1)^2 \delta D} - 1 \leq e^{\varepsilon (r-1)^2 \delta} \text{ which is small.}
\end{aligned}$$

All these together imply $\mathrm{Var}(d^*(x)) = o(\mathbb{E}[d^*(x)]^2)$. By Chebyshev's inequality, $d^*(x) = De^{-\varepsilon(r-1)}(1 \pm \delta^*)$ with high probability.

Let $N = |\{x \text{ good} : d^*(x) \neq e^{-\varepsilon(r-1)} D(1 \pm \delta^*)\}|$. Now Markov's inequality gives $\mathbb{E}[N] < \delta_{(11)} n$, so all except $\delta^* n$ vertices satisfy (c). This completes the proof of the Nibble lemma, and hence of the proof of the Erdős-Hanani conjecture as well. ■

Remark: The theory of Steiner designs is one of the oldest problems in Design theory. The existence and explicit constructions of Steiner 2-designs for all feasible parameters (parameters (n, r) for which the corresponding numbers are integers) and very large set sizes ($n \gg 0$) is the pioneering work of R. Wilson [34, 35, 36] beginning in the early 70s. The problem of existence of Steiner t-designs for $t \geq 6$ was open completely until P. Keevash [19] in 2014 settled this by a tour-de-force algebraic probabilistic argument. Keevash's proof is a little too involved for us to include in this book, but we will see some relevant ideas in later chapters.

7 Concentration Inequalities

7.1 A Simple Random Walk

Consider a simple random walk on integers. At each time unit, the walker flips a coin to decide whether to go one step to the right or to the left, depending on whether heads or tails shows up.

Formally, let X_t denote a random variable which takes the value 1 if heads shows up and -1 if tails shows up (i.e. takes $+1$ or -1 with probability $\frac{1}{2}$ each) at time t.

$$X_t \overset{\text{iid}}{\sim} \begin{pmatrix} 1 & -1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \forall t \geq 0$$

Assuming that it starts in the origin, It is quite obvious that in expectation it'll be at the origin. But a good question to ask would be how far it can go. Although it can reach arbitrarily large distances from the origin (with some tiny, non-zero probability as it'll mostly concentrate around it's expectation), we would like to establish some bounds on the chances that the walker is a certain distance away after some time.

Let S_n denote the sum of the random variable $\{X_i\}$ upto time n:

$$S_n = \sum_{i=0}^{n-1} X_i$$

S_n effectively measures the distance of the walker from the origin. Now, our question can be interpreted as bounds on the following probability:

$$\mathbb{P}[|S_n| \geq a] = ?$$

Theorem 44 (Chernoff Bound)**.** *Given a collection of random variables* $\{X_i\}$ *where each* $X_i \overset{\text{iid}}{\sim} \left(\begin{smallmatrix} 1 & -1 \\ \frac{1}{2} & \frac{1}{2} \end{smallmatrix} \right)$

$$\mathbb{P}[|S_n| > a] \leq 2e^{-\frac{a^2}{2n}} \tag{7.1}$$

Proof. It is enough to show the bound for $\mathbb{P}[S_n \geq a]$ as the other case will follow identically. Since $\{X_i\}$ are independent, $\mathbb{E}[S_n] = \sum_{i=0}^{n-1} \mathbb{E}[X_i] = 0$ by linearity of expectation. Now,

$$\mathbb{P}[S_n \geq a] = \mathbb{P}[e^{tS_n} \geq e^{ta}] \leq \frac{\mathbb{E}[e^{tS_n}]}{e^{ta}}$$

by Markov's Inequality. Using the fact that $\mathbb{E}[e^{tS_n}]$ is the moment generating function for S_n and since S_n is the sum of n independent random variables $\{X_i\}_{i=0}^{n-1}$

$$\begin{aligned}\mathbb{E}[e^{tS_n}] &= \prod_{i=0}^{n-1} \mathbb{E}[e^{tX_i}] = \mathbb{E}[e^{tX_0}]^n \\ &= \left(\frac{e^t + e^{-t}}{2}\right)^n \leq e^{\frac{nt^2}{2}}\end{aligned}$$

Hence,

$$\mathbb{P}[S_n \geq a] \leq e^{\frac{nt^2}{2} - ta}$$

Minimizing the RHS with respect to t, we get the following

$$\mathbb{P}[S_n > a] \leq e^{-\frac{a^2}{2n}}$$

■

This bound helps us characterise how well a random variable is concentrated around it's expected value without worrying about the asymptotics.

7.2 Why are these bounds so important?

The Central Limit Theorem (CLT) states that under appropriate conditions, the distribution of a normalized version of the empirical mean converges to a standard normal distribution, even though the underlying data might not have been from a normal distribution.

Theorem 45 (Central Limit Theorem)**.** *Suppose that $\{X_i\}$ is a collection of i.i.d samples with mean μ and variance σ^2, then*

$$\mathbb{P}\left(\frac{\sum_{i=0}^{n-1} X_i - n\mu}{\sigma\sqrt{n}} \leq k\right) \xrightarrow{n\to\infty} \int_{-\infty}^{k} e^{-\frac{t^2}{2}} dt \tag{7.2}$$

i.e.

$$\frac{\sum_{i=0}^{n-1} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Now, even though we now have an idea about the limiting case, it is difficult to say when we can say that this normalized empirical distribution is 'close enough' to the normal distribution.

In the case of $X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, the normalized empirical distribution approaches the normal distribution pretty rapidly, only requiring a small numbers of samples to approximate it. However, in order to obtain the same estimates for something like the power law ($F_X(x) = \frac{1}{x^\alpha}, \alpha > 2$), we might require absurdly large number of samples to do so. Hence, we can't say anything about a general random variables, except for the asymptotic case, without getting (possibly significant) error. Chernoff bound and other tail estimates therefore help us provide estimates for a broad class of random variables without worrying about the asymptotics and associated error terms.

Observation 46. *A core idea to get better tail estimates is that the more we know about the higher moment of the random variable, the more control we have on these tail estimates.*

7.3 The Johnson-Lindenstrauss Lemma

This is very useful tool which has its most important application in dimensionality reduction. Dimensionality reduction is a technique of transforming high dimensional data into some low dimensional subspace in the way that preserves much of the properties of original data. In many cases, original data have lot of sparsity; and is computationally difficult to analyze. In such cases, dimensionality reduction helps to project this data onto some lower dimensional subspace which is easier to analyze and have less sparsity. Coming to the lemma, although it was initially proven by Johnson-Lindenstrauss in 1984 ([**?**]), we will be following a more streamlined proof, developed by Frankl-Maehara ([**?**])

Definition 47. *Suppose $S \subset \mathbb{R}^d$ and $|S| = n$. A map $\phi : \mathbb{R}^d \to \mathbb{R}^m$ is said to be an ε-**approximation isometry** on S if for all $u, u' \in S$,*

$$(1-\varepsilon)\left\|u-u'\right\|_2 \leq \left\|\phi(u)-\phi(u')\right\|_2 \leq (1+\varepsilon)\left\|u-u'\right\|_2$$

Theorem 48 (The Johnson-Lindenstrauss Lemma)**.** *Given $\varepsilon > 0$ there exists $m = O_\varepsilon\left(\frac{\log n}{\varepsilon^2}\right)$ and an map $\phi : \mathbb{R}^d \to \mathbb{R}^m$ that is an ε-approximation isometry.*

A small digression: In [**?**] showed the existence of a set of n vectors in $\mathbb{R}^d$ so that any ε-approximation scheme must have at least $m = \Omega\left(\frac{\log(n)}{\varepsilon^2}\right)$. Hence the bounds obtained are tight up to a constant factor.

Observation 49. *A random projection should do the job !*

But how do we get one? Note that randomly choosing a unit vector is equivalent to uniformly sampling a point on a $(d-1)$-sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$.

So here's an idea. Consider the matrix

$$R = \frac{1}{\sqrt{m}} \begin{pmatrix} \xi_{1,1} & \cdots & \xi_{1,d} \\ \vdots & & \vdots \\ \xi_{m,1} & \cdots & \xi_{m,d} \end{pmatrix}$$

where $\xi_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \ \ \forall i \in [m], j \in [d]$ This matrix essentially 'projects' a d-dimensional point to a m-dimensional point.

Now consider any d-dimensional vector $\mathbf{u} \in \mathbb{R}^d$. Then, if

$$\mathbf{v} = R\mathbf{u}$$

Then,

$$v_i = \frac{1}{\sqrt{m}} \sum_{j=1}^{d} \xi_{i,j} u_j$$

where v_i and u_j are the ith and jth coordinates of vectors $\mathbf{v}$ and $\mathbf{u}$ respectively.

Proposition 50. *If $x_i \sim \mathcal{N}(0,1) \ \ \forall i \in [d]$, then*

$$\sum_{i=1}^{d} a_i x_i \sim \mathcal{N}(0, \sum_{i=1}^{d} a_i^2)$$

Proof of Theorem 48 . Using Proposition 50, we see that

$$v_i^2 = \frac{1}{m} \left(\sum_{j=1}^{d} u_j \xi_{i,j} \right)^2$$

and

$$v_i = \frac{1}{\sqrt{m}} \sum_{j=1}^{d} u_j \xi_{i,j} \sim \mathcal{N}(0, \frac{\|\mathbf{u}\|^2}{m})$$

and hence

$$\frac{v_i}{\|\mathbf{u}\|} \sim \mathcal{N}(0, \frac{1}{m})$$

We want to establish bounds on the following:

$$\begin{aligned} \mathbb{P}\left(\|v\|^2 > (1+\varepsilon) \|\mathbf{u}\|^2 \right) &= \mathbb{P}\left(\sum_{i=1}^{m} v_i^2 > (1+\varepsilon) \|\mathbf{u}\|^2 \right) \\ &= \mathbb{P}\left(\sum_{i=1}^{m} \left(\frac{v_i}{\|\mathbf{u}\|} \right)^2 > (1+\varepsilon) \right) \end{aligned}$$

To make our analysis easier, we define a vector $\mathbf{X}$ as $\mathbf{X} = (X_1, X_2 \dots X_m)$ where

$$X_i = \sqrt{m}\frac{v_i}{\|\mathbf{u}\|} \sim \mathcal{N}(0,1)$$

We also use the following theorem to understand $\mathbb{E}[\theta X_i^2]$

Theorem 51. *If $X_i \sim \mathcal{N}(0,1)$ then m.g.f. of X_i^2*

$$\mathbb{E}[e^{\theta X_i^2}] = \frac{1}{\sqrt{1-2\theta}}$$

Proof.

$$\begin{aligned}\mathbb{E}[e^{\theta X_i^2}] &= \int_{\mathbb{R}} e^{\theta y^2}\frac{1}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}dy \\ &= \frac{1}{\sqrt{1-2\theta}}\end{aligned}$$

■

Therefore, using Theorem 51

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{m}\left(\frac{v_i}{\|\mathbf{u}\|}\right)^2 > (1+\varepsilon)\right) &= \mathbb{P}\left(\sum_{i=1}^{m} X_i^2 > m(1+\varepsilon)\right) \\ &= \mathbb{P}\left(\Pi_{i=1}^{m} e^{\theta X_i^2} > e^{\theta m(1+\varepsilon)}\right) \\ &\leq \frac{\Pi_{i=1}^{m}\mathbb{E}[e^{\theta X_i^2}]}{e^{\theta m(1+\varepsilon)}} \\ &= \frac{\left(\frac{1}{\sqrt{1-2\theta}}\right)^m}{e^{\theta m(1+\varepsilon)}} \\ &= \left(\frac{1}{\sqrt{1-2\theta}\cdot e^{\theta(1+\varepsilon)}}\right)^m\end{aligned}$$

Now,

$$\frac{1}{\sqrt{1-2\theta}\cdot e^{\theta(1+\varepsilon)}} = e^{-\frac{\log(1-2\theta)}{2}-\theta(1+\varepsilon)}$$

Setting $\lambda = \frac{\varepsilon}{2(1+\varepsilon)}$, we get,

$$\mathbb{P}\left(\sum_{i=1}^{m}\left(\frac{v_i}{\|\mathbf{u}\|}\right)^2 > (1+\varepsilon)\right) \leq \left((1+\varepsilon)e^{-\varepsilon}\right)^{\frac{k}{2}}$$

Now, we use the following inequality (for all $\varepsilon > 0$)

$$\log(1+\varepsilon) < \varepsilon - \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3}$$

we get,

$$\mathbb{P}\left(\sum_{i=1}^{m}\left(\frac{v_i}{\|\mathbf{u}\|}\right)^2 > (1+\varepsilon)\right) < \left(e^{\frac{-\varepsilon^2}{2}+\frac{\varepsilon^3}{3}}\right)^{\frac{k}{2}}$$

Choosing $k \geq \frac{24}{3\varepsilon^2-2\varepsilon^3}\log n$, we get,

$$\mathbb{P}\left(\sum_{i=1}^{m}\left(\frac{v_i}{\|\mathbf{u}\|}\right)^2 > (1+\varepsilon)\right) < \frac{1}{n^2}$$

Therefore,

$$\mathbb{P}\left(\|v\|^2 \leq (1+\varepsilon)\|\mathbf{u}\|^2\right) > 1 - \frac{1}{n^2}$$

A similar bound can be established in the other case

$$\mathbb{P}\left(\|v\|^2 \geq (1-\varepsilon)\|\mathbf{u}\|^2\right) > 1 - \frac{1}{n^2}$$

and

$$\mathbb{P}\left(\|v\|^2 \notin \left[(1-\varepsilon)\|\mathbf{u}\|^2, (1+\varepsilon)\|\mathbf{u}\|^2\right]\right) < \frac{2}{n^2}$$

Therefore,

$$\mathbb{P}\left(\exists u, u' \text{ such that our hypothesis doesn't hold}\right) < \binom{n}{2} \times \frac{2}{n^2} = 1 - \frac{1}{n} < 1$$

and hence, with positive probability, there exists a projection map that is an ε-approximation identity.

■

7.4 The Azuma-Höeffding Inequality

Definition 52. *Suppose, we have a filtration of $\sigma-$algebras, $\mathbb{A}_0 \subseteq \mathbb{A}_1 \subseteq \mathbb{A}_2 \subseteq \mathbb{A}_3 \subseteq \dots$ in probability space. A sequence of random variables $(X_0, X_1, X_2, \dots)$ is said to be* ***Martingale*** *if,*

1. *X_i is $\mathbb{A}_i-$measurable for each i.*
2. *$\mathbb{E}\left[X_i|\mathbb{A}_{i-1}\right] = X_{i-1}$ almost surely $\forall i \geq 1$.*

Definition 53 (Doob Martingale)**.** *Suppose, we have a filtration of $\sigma-$algebras, $\mathbb{A}_0 \subseteq \mathbb{A}_1 \subseteq \mathbb{A}_2 \subseteq \mathbb{A}_3 \subseteq \cdots \subseteq \mathbb{A}_n$ in probability space. Let, a random variable Y be $\mathbb{A}_n-$measurable and $\mathbb{E}[|Y|] < \infty$ or $Y \geq 0$. Define $X_k := \mathbb{E}[Y|\mathbb{A}_k]$. Then, the sequence $\{X_k\}$ is a martingale, commonly known as the* ***Doob Martingale****.*

Proof. First condition for Martingale is obvious. For second condition,

$$\begin{aligned}\mathbb{E}[X_k|\mathbb{A}_{k-1}] &= \mathbb{E}[\mathbb{E}[Y|\mathbb{A}_k]|\mathbb{A}_{k-1}] \\ &= \mathbb{E}[Y|\mathbb{A}_{k-1}] \\ &= X_{k-1} \text{ almost surely.}\end{aligned}$$

■

Theorem 54 (Azuma-Hoeffding Inequality)**.** *Suppose, we have a Martingale, $(X_0, X_1, X_2 \dots)$ and suppose there are constants C_i such that $|X_i - X_{i-1}| \leq C_i$ almost surely for all i. Then, given any $K > 0$ and $n \in \mathbb{N}$,*

$$\mathbb{P}(X_n - X_0 \geq K) \leq e^{-\frac{K^2}{2\sum_{i=1}^n C_i^2}}$$

In particular, if you take $K = \lambda\sqrt{n}$ for some $\lambda > 0$ and for all $i, C_i = 1$,

$$\mathbb{P}\left(X_n - X_0 \geq \lambda\sqrt{n}\right) \leq e^{-\frac{\lambda^2}{2}}$$

Proof. To prove Azuma-Hoeffding Inequality, we will need the following result:

Proposition 55. *Let, Y be a random variable with mean 0 such that $|Y| \leq d$ almost surely, then $\mathbb{E}\left[e^{\theta Y}\right] \leq e^{\frac{\theta^2 d^2}{2}}$ for any $\theta \in \mathbb{R}$.*

Proof. For any y with $|y| \leq d$, Write

$$\theta y = \frac{(y+d)}{2d}\theta d + \frac{(d-y)}{2d}\theta(-d)$$

As $e^{\theta y}$ is a convex function of y, we get,

$$e^{\theta y} \leq \frac{(y+d)}{2d}e^{\theta d} + \frac{(d-y)}{2d}e^{-\theta d}$$

Thus,

$$\begin{aligned}\mathbb{E}\left[e^{\theta Y}\right] &\leq \frac{(\mathbb{E}[Y]+d)}{2d}e^{\theta d} + \frac{(d-\mathbb{E}[Y])}{2d}e^{-\theta d} && \text{(using linearity of expectation)} \\ &= \frac{1}{2}e^{\theta d} + \frac{1}{2}e^{-\theta d} && \text{(as } \mathbb{E}[Y] = 0\text{)} \\ &= cosh(\theta d) \\ &\leq e^{\frac{\theta^2 d^2}{2}}\end{aligned}$$

■

Now, for $\theta > 0$,

$$\begin{aligned}\mathbb{P}\left(X_n - X_0 \geq K\right) &= \mathbb{P}\left(e^{\theta(X_n - X_0)} \geq e^{\theta K}\right) && \text{(as } e^{\theta x} \text{ is monotonic increasing if } \theta > 0\text{)} \\ &\leq e^{-\theta K}\mathbb{E}\left[e^{\theta(X_n - X_0)}\right] && \text{(Markov's Inequality)} \end{aligned} \tag{7.3}$$

For $\mathbb{E}\left[e^{\theta(X_n - X_0)}\right]$, we have,

$$\begin{aligned}\mathbb{E}\left[e^{\theta(X_n - X_0)}\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{\theta(X_n - X_{n-1})}e^{\theta(X_{n-1} - X_0)}\middle|\mathbb{A}_{n-1}\right]\right] \\ &= \mathbb{E}\left[e^{\theta(X_{n-1} - X_0)}\mathbb{E}\left[e^{\theta(X_n - X_{n-1})}\middle|\mathbb{A}_{n-1}\right]\right] \\ &\qquad\qquad \text{(as } e^{\theta(X_{n-1} - X_0)} \text{ is } \mathbb{A}_{n-1}\text{−measurable.)}\end{aligned}$$

As $\{X_i\}$ is Martingale, $\mathbb{E}\left[X_n - X_{n-1}|\mathbb{A}_{n-1}\right] = X_{n-1} - X_{n-1} = 0$ almost surely. Also, from assumptions in Azuma's Inequality, we have $|X_n - X_{n-1}| \leq C_n$ almost surely. Thus, using Theorem 55, we can bound $\mathbb{E}\left[e^{\theta(X_n - X_{n-1})}\middle|\mathbb{A}_{n-1}\right]$ as:

$$\begin{aligned}\mathbb{E}\left[e^{\theta(X_n - X_0)}\right] &= \mathbb{E}\left[e^{\theta(X_{n-1} - X_0)}\mathbb{E}\left[e^{\theta(X_n - X_{n-1})}\middle|\mathbb{A}_{n-1}\right]\right] \\ &\leq \mathbb{E}\left[e^{\theta(X_{n-1} - X_0)}e^{\frac{\theta^2 C_n^2}{2}}\right] \\ &\leq e^{\frac{\theta^2 C_n^2}{2}}\mathbb{E}\left[e^{\theta(X_{n-1} - X_0)}\right]\end{aligned}$$

Now, one can use the same argument repetitively to get,

$$\mathbb{E}\left[e^{\theta(X_n - X_0)}\right] \leq e^{\frac{\theta^2 \sum_{i=1}^n C_i^2}{2}}$$

Thus,

$$\begin{aligned}\mathbb{P}\left(X_n - X_0 \geq K\right) &\leq e^{-\theta K}\mathbb{E}\left[e^{\theta(X_n - X_0)}\right] && \text{(from Inequality 7.3)} \\ &\leq e^{\frac{\theta^2 \sum_{i=1}^n C_i^2}{2} - \theta K}\end{aligned}$$

As this is true for any $\theta > 0$, we can choose optimal θ for which RHS is minimum. Such optimal θ turns out to be:

$$\theta_* = \frac{K}{\sum_{i=1}^n C_i^2}$$

and thus, the inequality:

$$\mathbb{P}\left(X_n - X_0 \geq K\right) \leq e^{-\frac{K^2}{2\sum_{i=1}^n C_i^2}}$$

■

7.5 McDiarmid's Inequality

Theorem 56. *Suppose $\xi_1, \xi_2, \xi_3, \ldots, \xi_n$ are independent random variables and let,*

$$Y = f(\xi_1, \xi_2, \ldots, \xi_n)$$

where function f is $d-$Lipschitz i.e. for any x_i, x_i',

$$|f(x_1, x_2, \ldots, x_i, \ldots, x_n) - f(x_1, x_2, \ldots, x_i', \ldots, x_n)| \leq d$$

Then, for any $t > 0$,

$$\mathbb{P}\left(|Y - \mathbb{E}[Y]| > t\right) \leq 2e^{-\frac{t^2}{2d^2n}}$$

(In the above theorem, one can take $t = O\left(\sqrt{n \log n}\right)$ for maximum utility.)

Proof. Define $\mathbb{A}_k := \sigma(\xi_1, \xi_2, \ldots, \xi_k)$ and $Y_k := \mathbb{E}[Y|\mathbb{A}_k]$. Then, $\{Y_k\}$ by definition, forms a Doob Martingale Process. Let ξ_k' be an independent copy of ξ_k. Observe that

$$\begin{aligned}
Y_k' :&= \mathbb{E}[f(\xi_1, \xi_2, \ldots, \xi_k', \ldots, \xi_n)|\sigma(\xi_1, \xi_2, \ldots, \xi_k)] \\
&= \mathbb{E}[f(\xi_1, \xi_2, \ldots, \xi_k', \ldots, \xi_n)|\sigma(\xi_1, \xi_2, \ldots, \xi_{k-1})] \\
&\qquad\qquad \text{(as all entries in } f \text{ are independent of } \xi_k\text{)} \\
&= \mathbb{E}[Y|\mathbb{A}_{k-1}] \\
&= Y_{k-1}
\end{aligned}$$

Thus,

$$\begin{aligned}
&|Y_k - Y_{k-1}| \\
&= |\mathbb{E}[Y|\mathbb{A}_k] - Y_k'| \\
&= |\mathbb{E}[f(\xi_1, \xi_2, \ldots, \xi_k, \ldots, \xi_n)|\sigma(\xi_1, \xi_2, \ldots, \xi_k)] - \mathbb{E}[f(\xi_1, \xi_2, \ldots, \xi_k', \ldots, \xi_n)|\sigma(\xi_1, \xi_2, \ldots, \xi_k)]| \\
&\leq \mathbb{E}[|f(\xi_1, \xi_2, \ldots, \xi_k, \ldots, \xi_n) - f(\xi_1, \xi_2, \ldots, \xi_k', \ldots, \xi_n)||\sigma(\xi_1, \xi_2, \ldots, \xi_k)] \\
&\leq d \qquad\qquad \text{(almost surely as } f \text{ is } d\text{−Lipschitz)}
\end{aligned}$$

Hence, by applying Azuma-Hoeffding's Inequality (Theorem 54),

$$\mathbb{P}\left(|Y - \mathbb{E}[Y]| > t\right) \leq 2e^{-\frac{t^2}{2d^2n}}$$

for any $t > 0$. ■

7.6 Janson's Inequality

Suppose, $H = (V, \mathcal{E})$ is given hypergraph. Consider the following random experiment: Each $v \in V$ has some probability $0 \leq p_v \leq 1$ assigned to it. We want to sample vertex set V as per these probabilities i.e., we are going to choose randomly a subset $R \subseteq V$ such that independently, $\forall v \in V, \mathbb{P}(v \in R) = p_v$.
Now, the question is what is the probability that R contains some $E \in \mathcal{E}$?

Theorem 57 (Janson's Inequality)**.** *Let,* $N := \#\{E \in \mathcal{E} | E \subseteq R\}$. *Define,*

$$\mu := \mathbb{E}[N] = \sum_{E \in \mathcal{E}} \mathbb{P}(E \subseteq R)$$

For $B_1, B_2 \in \mathcal{E}$, *let* $B_1 \sim B_2$ *if* $B_1 \neq B_2$ *and* $B_1 \cap B_2 \neq \emptyset$. *Define,*

$$\Delta := \sum_{B_1 \sim B_2} \mathbb{P}(B_1 \subseteq R \wedge B_2 \subseteq R)$$

Then,

1. $\mathbb{P}(N = 0) \leq e^{-\mu + \frac{\Delta}{2}}$
2. *If* $\Delta \geq \mu, \mathbb{P}(N = 0) \leq e^{-\frac{\mu^2}{2\Delta}}$

A simple application

Definition 58. *Suppose* p *is prime and let,* $X = (x_1, x_2, \ldots, x_k)$ *be a finite sequence in* $\mathbb{F}_p^*$. *Then,* X *is said to have a* $0-$***sum subsequence*** *if* $\exists S \subseteq [k]$ *such that* $\sum_{s \in S} x_s = 0$ *in* $\mathbb{F}_p$.

Now, let $(G, +)$ be finite abelian group. Define,

$$D(G) := \min\{m : \text{Any sequence of } m \text{ elements of G has } 0-\text{sum subsequence.}\}$$

For example, $D(\mathbb{Z}/n\mathbb{Z}) = n$.
We can also allow elements in sum to have $\pm$ signs.

$D_{\pm}(G) := \min\{m : \text{Any sequence of } m \text{ elements of G has } 0-\text{sum weighted subsequence with weights } =$

In the example above, $D_{\pm}(\mathbb{Z}/n\mathbb{Z}) = \lfloor \log_2 n \rfloor + 1$. To see why this bound can't be reduced further, take $n = 2^k$ and $k-$sequence to be $(1, 2, 2^2, \ldots, 2^{k-1})$. For this sequence, one can't find any $0-$sum subsequence even with weights ± 1.
This idea can be more generalized for $G = \mathbb{F}_p$ as follows:
Let, $A \subseteq \mathbb{F}_p^*$. Define $D_A(\mathbb{F}_p)$ as:

$$D_A(\mathbb{F}_p) := \min\{m : \text{Any sequence } (x_1, x_2, \ldots, x_m) \text{ of } \mathbb{F}_p \text{ admits } j_1, j_2, \ldots, j_r$$
$$\text{and } a_1, a_2, \ldots, a_r \in A \text{ such that } \sum_{l=1}^{r} a_l x_{j_l} = 0 \text{ in } \mathbb{F}_p.\}$$

Examples:

1. $D_{\pm}(\mathbb{Z}/n\mathbb{Z}) = \lfloor \log_2 n \rfloor + 1$ where $A = \{1, -1\}$
2. $D_A(\mathbb{Z}/n\mathbb{Z}) = \left\lceil \frac{n}{k} \right\rceil$ where $A = \{1, 2, \ldots, k\}$

3. $D_A(\mathbb{Z}/n\mathbb{Z}) = \lceil \log_2 n \rceil$ where $A = \left\{1, 2, \ldots, \left\lfloor \frac{n}{\log_2 n} \right\rfloor \right\}$

Now, for $k \geq 2$, let

$$f^{(0)}(p,k) := \min\{|A| : A \subseteq \mathbb{F}_p^* \text{ and every sequence } (x_1, x_2, \ldots, x_k) \in \left(\mathbb{F}_p^*\right)^k \\ \text{admits an } A-\text{weighted zero-sum subsequence in } \mathbb{F}_p\}$$

Proposition 59. $f^{(0)}(p,k) \geq p^{1/k} - 1$

Proof. Suppose $A \subseteq \mathbb{F}_p^*$ is optimal. Let $A_0 := A \cup \{0\}$. Consider the bipartite graph $G = (\mathbb{A}, \mathbb{B}, \mathcal{E})$ where $\mathbb{A} = \left(\mathbb{F}_p^*\right)^k, \mathbb{B} = (A_0)^k \setminus \{\mathbf{0}\}$ and $\mathcal{E}$ is defined such that:

$$\mathcal{E} = \left\{ \{\boldsymbol{x}, \boldsymbol{a}\} \middle| \boldsymbol{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{A}, \boldsymbol{a} = (a_1, a_2, \ldots, a_k) \in \mathbb{B} \text{ and } \sum_{i=1}^{k} a_i x_i = 0 \text{ in } \mathbb{F}_p \right\}$$

Note that this bipartite graph is capturing the idea of "subsequence" of sum zero as one can simply put the a_i's corresponding to terms not participating in $0-$sum subsequence to be zero. Also, $\mathbf{0}$ being removed from $\mathbb{B}$ essentially ensures that there is a non-empty $0-$sum subsequence taking coefficients from A. Thus, by our assumption about A, $d(\boldsymbol{x}) \geq 1 \ \forall \boldsymbol{x} \in \mathbb{A}$

$$\implies |\mathcal{E}| \geq (p-1)^k \tag{7.4}$$

On the other hand, for fixed $\boldsymbol{a} \in \mathbb{B}$, $d(\boldsymbol{a}) \leq (p-1)^{k-1}$. Because, one can take any $a_i \neq 0$ in $\boldsymbol{a}$ and define $\boldsymbol{x}$ like this: Pick x_j's arbitrarily for $j \neq i$ and let, $x_i = -a_i^{-1} \sum_{j \neq i} a_j x_j$. Note that this is unique choice for x_i so that there is an edge between $\boldsymbol{x}$ and $\boldsymbol{a}$. This implies $d(\boldsymbol{a}) \leq (p-1)^{k-1}$ and thus,

$$|\mathcal{E}| \leq (p-1)^{k-1} \left((|A|+1)^k - 1 \right) \tag{7.5}$$

Thus, from inequalities 7.4 and 7.5,

$$\begin{aligned} & (p-1)^k \leq (p-1)^{k-1} \left((|A|+1)^k - 1 \right) \\ \implies \ & p - 1 \leq (|A|+1)^k - 1 \\ \implies \ & p^{1/k} - 1 \leq |A| = f^{(0)}(p,k) \end{aligned}$$

■

Theorem 60. *For $k \geq 2$,*

$$f^{(0)}(p,k) \leq \mathcal{O}\left((p \log p)^{1/k} \right)$$

Proof. (We will prove it for case when $k = 2$.)
Let $A \subseteq \mathbb{F}_p^*$ be a $\theta-$random subset i.e. independently $\forall a \in \mathbb{F}_p^*$, $\mathbb{P}(a \in A) = \theta$.
Note that for given $x, y \in \mathbb{F}_p^*$, A annihilates it i.e. $\exists a, b \in A$ such that,

$$\begin{aligned} & ax + by = 0 \\ \iff \quad & a.1 + b(yx^{-1}) = 0 \end{aligned}$$

that is, it is enough to check if A annihilates $\{1, x\}$ for any $x \in \mathbb{F}_p^*$.
This is trivial for $x = 1$, so let $x \in \mathbb{F}_p^* \setminus \{1\}$. If $\exists a, b \in A$ such that either $a + bx = 0$ or $ax + b = 0$,

$$\implies x = -ab^{-1} \text{ or } x = -ba^{-1}$$

Using this, lets define a graph G_x where $V(G_x) = \mathbb{F}_p^*$ and edges are given by:

$$\mathcal{E}(G_x) = \left\{ \{u, v\} \middle| x = -\frac{u}{v} \text{ or } x = -\frac{v}{u} \right\}$$

G_x thus formed, is $2-$regular as each vertex v has exactly two neighbours $-vx$ and $-vx^{-1}$.
So, in terms of graph, A annihilates $\{1, x\}$ iff A is not independent in G_x or in other words, there exists an edge between some pair of points in A. This is where we can use **Janson's inequality**. Define $N := \#\{E \in \mathcal{E} | E \subseteq A\}$. Thus,

$$\begin{aligned} \mu := \mathbb{E}[N] &= \sum_{E \in \mathcal{E}} \mathbb{P}(E \subseteq A) \\ &= \theta^2(p-1) \qquad \left(|\mathcal{E}| = \frac{1}{2} \sum_{v \in V} d(v) = \frac{2(p-1)}{2} = p - 1 \right) \end{aligned}$$

and

$$\begin{aligned} \Delta :&= \sum_{B_1 \sim B_2} \mathbb{P}(B_1 \subseteq A \wedge B_2 \subseteq A) \qquad \text{where } B_1 \sim B_2 \text{ if } B_1 \neq B_2 \text{ and } B_1 \cap B_2 \neq \emptyset \\ &= \theta^3(p-1) \end{aligned}$$

Thus, by Janson's inequality (Theorem 57),

$$\mathbb{P}(N = 0) \leq e^{\frac{(p-1)\theta^3}{2} - (p-1)\theta^2}$$

Therefore,

$$\mathbb{P}(\{1, x\} \text{ is not annihilated by } A) \leq e^{\frac{(p-1)\theta^3}{2} - (p-1)\theta^2}$$

Hence,

$$\mathbb{P}\left(\vee_{x \in \mathbb{F}_p^*} \{1, x\} \text{ is not annihilated by } A \right) \leq (p-1) e^{\frac{(p-1)\theta^3}{2} - (p-1)\theta^2} \qquad \text{(union bound)}$$

We want this probability to be strictly less than 1 so that there will be A which annihilates $\{1, x\}\ \forall x \in \mathbb{F}_p^*$. Thus,

$$\begin{aligned}
&(p-1)e^{\frac{(p-1)\theta^3}{2}-(p-1)\theta^2} < 1 \\
\implies &\log(p-1) + \frac{(p-1)\theta^3}{2} - (p-1)\theta^2 < 0 \\
\implies &\frac{\log(p-1)}{p-1} < \theta^2 - \frac{\theta^3}{2} \approx \mathcal{O}(\theta^2) \qquad \text{(as we can safely assume } \theta << 1)
\end{aligned}$$

This implies that $\theta = \Omega\left(\sqrt{\frac{\log p}{p}}\right) \implies \mathbb{E}[|A|] = (p-1)\theta = \Omega\left(\sqrt{p \log p}\right)$ which further implies:

$$f^{(0)}(p, 2) \leq \mathcal{O}\left((p \log p)^{1/2}\right)$$

■

If p is sufficiently large and k fixed, $f^{(0)}(p, k) \leq 4^{k^2} p^{1/k}$.

8 Some applications of the basic concentration inequalities

It is often the case that the random variable of interest is a sum of independent random variables. In many of those cases, the theorem of Chebyshev is much weaker than what can be proven. Under reasonably mild conditions, one can prove that the random variable is *tightly concentrated* about its mean, i.e., the probability that the random variable is 'far' from the mean decays exponentially, and this exponential decay is crucial in several probabilistic applications.

The distribution of the sum of i.i.d random variables, suitably normalized, behaves like the Standard Gaussian; that is the import of the Central Limit Theorem (CLT for short) in Probability, so in that sense, the Chernoff bound has its antecedents from much earlier - indeed this goes back to Laplace. But the CLT is a limiting theorem, whereas the Chernoff bounds are not. This qualitative difference is also very useful from an algorithmic point of view.

In this chapter, we consider a few prototypes[1] of such results along with some combinatorial applications.

8.1 The Chernoff Bound

We first recall (a version of) the Chernoff[2] bound. We have seen a general template of this theorem in a previous chapter. Any such exponential type decay bound will be called a *Chernoff type bound* or simply, a Chernoff bound:

[1] Actually it is a misnomer to call them Chernoff bounds because these also date back to Chebyshev. But they were independently discovered by Chernoff, and the name has stuck since.

[2] Though as Spencer explicitly has stated, this bound was already there in Chebyshev; the (somewhat inaccurate) attribution to H. Chernoff was mostly popularized by Spencer. But since that is the name everyone uses, we shall do the same.

Proposition 61 (Chernoff Bound). *Let $X_i \in \{-1, 1\}$ be independent random variables, with $\mathbb{P}[X_i = -1] = \mathbb{P}[X_i = 1] = \frac{1}{2}$, and let $S_n = \sum_{i=1}^{n} X_i$. For any $a > 0$ and any n, $\mathbb{P}[S_n > a] < e^{-a^2/2n}$.*

Proposition 61 can be generalized and specialized in various ways. We state two such modifications here.

Proposition 62 (Chernoff Bound (Generalized Version)). *Let $p_1, \ldots, p_n \in [0, 1]$, and let X_i be independent random variables such that $\mathbb{P}[X_i = 1 - p_i] = p_i$ and $\mathbb{P}[X_i = -p_i] = 1 - p_i$, so that $\mathbb{E}[X_i] = 0$ for all i. Let $S_n = \sum_{i=1}^{n} X_i$. Then*

$$\mathbb{P}[S_n > a] < e^{-2a^2/n} \quad \text{and} \quad \mathbb{P}[S_n < -a] < 2e^{-2a^2/n}$$

Letting $p = \frac{1}{n}(p_1 + \ldots + p_n)$, this can be improved to

$$\mathbb{P}[S_n > a] < e^{-a^2/pn + a^3/2(pn)^2}$$

Proposition 63 (Chernoff Bound (Binomial Version), see [18]). *Let $X \sim$ Binomial(n, p), and let $0 \leq t \leq np$. Then*

$$\mathbb{P}[|X - np| \geq t] \leq \left(\frac{-t^2}{2(np + t/3)} \right)$$

and the last expression is at most $2e^{-t^2/3np}$ if $t \leq np$.

In all three cases, the independence assumption can be removed while preserving the exponential decay (although with a worse constant).

Before we move on to some applications, we make a quick remark. While the aforementioned version of the Chernoff bound holds always, its efficacy, especially when we wish to establish that some event occurs with high probability only works when $np \to \infty$. If $p = O(1/n)$ so that $np = O(1)$ then this bound does not work as well. And this is again an observation that goes back to Poisson; the Binomial distribution, suitably normalized, can be well approximated by the standard Gaussian when the expected value goes to infinity with n, and if the expected value is bounded by a constant, then for large n, the behavior is more like the Poisson. We will return to this point in a later chapter.

8.2 First applications of the Chernoff bound

We start with a return to a result from a previous chapter. Recall the randomized algorithm to determine the frequency moments using a sub-linear number of bits. We had a sequence of random variables $(Y_1, \ldots, Y_r)$ with $\mathbb{E}(Y_i) = F_k$ and $\mathbb{P}(|Y - F_k| > \lambda F_k) \leq 1/8$. Our final report was the random variable $Z = \text{Median}(Y_1, \ldots, Y_r)$ and our interest was in obtaining a bound for r such that $\mathbb{P}(|Z - F_k| > \lambda F_k) \leq \varepsilon$. Towards that end, we

had defined the random variable $\tilde{Y}_i$ to equal 1 if $Y_i \in [F_k - \lambda F_k, F_k + \lambda F_k]$ and zero otherwise, so that $\tilde{Y}_i$ is distributed as Ber(7/8). Setting $\tilde{Z} := \sum_{i=1}^{r} \tilde{Y}_i$ allows us to estimate this probability by the bounds of a Binomial random variable Bin$(r, 7/8)$. If $Z \notin [F_k - \lambda F_k, F_k + \lambda F_k]$, then $\tilde{Z}$ is less than $r/2$. The second moment method gives $\mathbb{P}(|Z - F_k| > \lambda F_k) \leq \varepsilon$ for $r = O(1/\sqrt{\varepsilon})$. Instead, if we use the Chernoff bound, then we have

$$\mathbb{P}(|Z - F_k| > \lambda F_k) \leq \mathbb{P}(\tilde{Z} < r/2) \leq \exp\left(\frac{-9r}{126}\right)$$

so that we can, as claimed earlier, take $r = O\left(\log(\frac{1}{\varepsilon})\right)$ to get the same probability bound as stated.

8.3 Discrepancy in hypergraphs

The notion of *Discrepancy* is the topic of very deep study and there are even a few monographs dedicated to this (see []). Here, we shall contend ourselves with a very basic result. Given a hypergraph $\mathcal{H} = (V, \mathcal{E})$ and a 2-coloring $c : V(\mathcal{H}) \to \{-1, 1\}$, the *discrepancy of an edge E for the coloring c* is simply $\text{disc}_c(E) := |\sum_{v \in E} c(v)|$. The *discrepancy of the coloring c* is the maximum discrepancy that c produces amongst the edges of $\mathcal{H}$, i.e., $\text{disc}(c)) := \max_{E \in \mathcal{E}} \text{disc}_c(E)\}$. The *discrepancy of the hypergraph* is the minimum discrepancy amongst all two colorings of $\mathcal{H}$, i.e., $\text{disc}(\mathcal{H}) := \min_c \text{disc}(c)$. In words, the discrepancy of a hypergraph attempts to measure how equitably one can partition the vertex set into two parts in the sense that that *every* edge of the hypergraph gets partitioned as equitably as possible. One of the most celebrated results in Discrepancy theory is the theorem of Spencer that states that a hypergraph on n vertices and $O(n)$ edges has a discrepancy of order $O(\sqrt{n})$. This was subsequently given an algorithmic proof by N. Bansal [11] and several others, with the current 'book proof'[3] due to T. Rothvoß [27].

But here, we shall prove a much more modest statement, which is also the starting point for many of the improved results in this direction.

Proposition 64. *Suppose $\mathcal{H}$ is a hypergraph on n vertices and m edges. Then disc$(\mathcal{H}) = O(\sqrt{n \log m})$.*

Proof. To show an upper bound on the discrepancy, we need to exhibit a coloring c for which the discrepancy is small. Pick a random coloring, i.e., for each v assign it the color 1 or -1 independently. For an edge E with $|E| = k$, let $X_E := \sum v \in Ec(v)$. Then $\mathbb{E}(X_E) = 0$, so using the Chernoff bound to decree $\mathbb{P}(|X_E| > t) \leq 2\exp^{-t^2/2k} < 1/m$ suggests $t = O(\sqrt{n \log m})$ since $k \leq n$. By choice, this implies that the expected number of edges with discrepancy greater than t is less than one, so again by the method of expectations(!), there is a coloring c such that all the edges discrepancies are at most t. This completes the proof. ∎

[3] This was Erdős' notion of the best/cleanest possible proof of any result. For more on this, see [1].

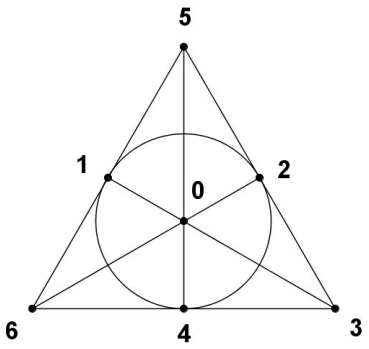
Remark: It is not hard to show that there are hypergraphs with n vertices and n edges with a discrepancy of $\Omega(\sqrt{n})$, so the result of Spencer is asymptotically tight. But as we have already seen with the Rödl nibble method, some probabilistic constructions cannot achieve the desired goal in a single step process, and the proofs for the sharp discrepancy follow a random process. We will address random processes in a later chapter.

8.4 Projective Planes and Property B

Given a hypergraph $\mathcal{H} = (V, \mathcal{E})$, we say that $\mathcal{H}$ has *property B* if there exists $S \subseteq V$ such that for all $E \in \mathcal{E}$ both S and $\overline{S}$ intersect E. This is an extension of the notion of graph colorings to hypergraphs, where the notion of proper coloring now (considering that for hypergraphs one has more than one interpretation of what a proper coloring ought to be) is that there is no monochromatic edge.

Unlike the graph case where there is a very simple algorithmic characterization of 2-colorability, the problem of deciding when a hypergraph has property B is far from well understood. Indeed, one of Erdős/ oldest and returning motifs was to determine $m(n)$ the minimum number of edges in an n-uniform hypergraph that does not have property B.

One of the earliest observations regarding property B was the following due to Lovász, which effectively comes from an algorithmic procedure to attempt to 2-color the vertices: If $\mathcal{H}$ is such that $|E_1 \cap E_2| \neq 1$ for all pairs of distinct edges E_1, E_2, then $\mathcal{H}$ is 2-colorable, and therefore has property B. Indeed, number the vertices $1, \ldots, n$. Color each vertex, in order, avoiding monochromatic edges. It is easily seen that by the assumptions on $\mathcal{H}$, this must yield a valid coloring. So for now, let us work with a situation where the hypergraph violates this condition in the extreme, i.e., suppose that $\mathcal{H}$ has the property that every pair of edges meet at exactly 1 vertex. Examples of such hypergraphs arise from the projective planes which we have encountered in Chapter 1. The *Fano Plane*, shown here with each edge represented as a line, shows that such hypergraphs are not necessarily 2-colorable. Following Erdős, we now define a stronger version of property B,



which we will refer to as Property B(s).

Definition 65 (Property B(s)). *A hypergraph $\mathcal{H} = (V, \mathcal{E})$ has* property B(s) *if there exists $S \subseteq V$ such that for every edge E, $0 < |E \cap S| \leq s$.*

If we set $s = n - 1$, then for n-uniform hypergraphs, property B(s) is the same as the usual property B.

Recall the notion of a *Projective Plane of order n*, denoted by $\P_n$: It is an $(n+1)$-uniform hypergraph $\P_n := (\mathcal{P}, \mathcal{L})$ with $|\mathcal{P}| = n^2 + n + 1$ (called points), $|\mathcal{L}| = n^2 + n + 1$ (called lines) such that every pair of points determines a unique edge, and every pair of lines intersect in a unique point.

Theorem 66 (Erdős, Silverman, Steinberg). *There exist constants k, K such that for all n there exists a subset S of the points of $\P_n$ with $k \log n \leq |L \cap S| \leq K \log n$ for all $L \in \mathcal{L}$.*

Proof. Choose S at random, with each point x placed in S with probability $p = \frac{f(n)}{n+1}$, for some $f(n)$ to be determined later.

Fix a line L, and let $S_L = |S \cap L|$. Note that $\mathbb{E}[S_L] = (n+1)p = f(n)$. By the Chernoff Bound, $\mathbb{P}[|S_L - f(n)| > f(n)/2] < 2e^{-f(n)/12}$. Since $\P_n$ contains n^2+n+1 lines,

$$\mathbb{P}[\text{There exists } L \text{ such that } |S_L - f(n)| > f(n)/2] < 4e^{-Cf(n)}n^2$$

for some absolute constant C. Therefore, if $e^{Cf(n)} > \Omega(n^2)$, a set S with the desired property exists. This in turn tells us that setting $f(n) = O(\log n)$ gives us the stated result for sufficiently large n. ■

Remark: Erdős conjectured that for the projective planes, a much stronger statement holds: There exists an absolute constant s such that for all sufficiently large n, the projective plane of order n has property B(s).

The problem of determining $m(n)$ which was alluded to earlier remains one of the most elusive problems in extremal combinatorics. We will, later in this book, see a proof of the statement

$$\Omega\left(\sqrt{\frac{n}{\log n}}2^n\right) \leq m(n) \leq O(n^2 2^n)$$

which still marks the best known result to date.

8.5 Graph Coloring and Hadwiger's Conjecture

IN this section we see a counterexample to a conjecture of Hájos in an attempt to solve the famous Hadwiger conjecture. To get there, we first need a couple of definitions. For an edge $e = uv$ in a graph G, the *contraction of e* is a graph denoted G/e obtained by deleting the vertices u, v and replacing it with a new vertex v_e which is adjacent to all

the neighbors of u and v counting with multiplicity. In other words, if a vertex w was adjacent to both u, v in G, then v_e has two edges to w in G/e.

Definition 67 (Graph Minor)**.** *Given a graph G, H is a* minor *of G if H can be obtained from G by a sequence of edge deletions edges and edge contractions and deletions of isolated vertices.*

Definition 68 (Subdivision)**.** *A graph H is a subdivision of G if H can be made isomorphic to a subgraph of G by inserting vertices of degree 2 along the edges of H.*

One can think of H as a subgraph of G in which disjoint paths are allowed to act as edges. Note that if H is a subdivision of G, then H is also a minor of G; however, the converse is false in general.

The deep conjecture of Hadwiger is the following

Conjecture 69 (Hadwiger's Conjecture)**.** *Let G be a graph with $\chi(G) \geq p$. Then G contains K_p as a minor.*

Paraphrasing, Hadwiger's conjecture states that for a graph G to be p-colorable, the clique on p vertices is forbidden as a minor. The best known result towards settling Hadwiger's conjecture is the celebrated Robertson-Seymour Theorem on graph minors [26]which shows that p-colorability is characterized by a finite set of forbidden minors.

Hadwiger's Conjecture is notoriously diffcult. Indeed, the special case of $p = 5$is equivalent to the four color theorem for planar graphs[4]. One way is straightforward: if $\chi(G) \geq 5$ then by the conjecture G contains K_5 as a minor and is therefore nonplanar as a consequence of Kuratowski's theorem (see [32]). But the other way needs more work. The conjecture is currently open for $p > 6$. In fact, all known proofs of Hadwiger's conjecture for $p = 5, 6$ use the four-color theorem.

Due to the apparent difficulty of Hadwiger's Conjecture, Hajós strengthened the conjecture to state that if $\chi(G) \geq p$, then G contains K_p as a *subdivision*. But as it is usually the case, this strengthened conjecture turned out to be false as was shown by Catlin via an explicit counterexample. However, the motivating question really is: How good a conjecture is the strengthened version? If the counterexamples were freak instances, then maybe one at least had an asymptotically strong statement since subdivisions are easier to understand from a verification perspective unlike minors. But later, Erdős and Fajtlowicz put this possibility to rest showing that the conjecture is almost never true.

Theorem 70 (Erdős, Fajtlowicz)**.** *There exist graphs G such that $\chi(G) \geq \frac{n}{3 \log n}$ and G has no $K_{3\sqrt{n}}$ subdivision.*

[4]The four color theorem states that ever planar graph, i.e., graph that can be embedded on the plane, is 4-colorable.

Proof. Let $G = (V, E)$ be a random graph on n vertices, with each edge placed in the graph with probability $1/2$. We first show that with high probability, G has large chromatic number, and then also that G has no large K_p subdivision.

Since $\chi(G) \geq n/\alpha(G)$, let us examine an upper bound for $\alpha(G)$. We have

$$\begin{aligned}\mathbb{P}[\alpha(G) \geq x] &= \mathbb{P}[\text{there exists a set of } x \text{ vertices which form an independent set}] \\ &\leq \binom{n}{x} 2^{-\binom{x}{2}} \leq \left(\frac{n}{2^{\frac{x-1}{2}}}\right)^x\end{aligned}$$

Set $x = 2\log n + 3$ so that $2^{(x-1)/2} = 2n$; then

$$\mathbb{P}[\alpha(G) \geq x] \leq \left(\frac{1}{2}\right)^{2\lg n+3} = \frac{1}{8n^2}$$

so with high probability, $\alpha(G) \leq 2\log n + 3 < 3\log n$.

Now suppose that G contains K_t as a subdivision. Since K_t contains $\binom{t}{2}$ edges, G must contain as many disjoint paths. Now, each vertex of G must either be a vertex of the K_t subdivision, or else be contained in at most one of the paths. Since there are n vertices in G, we end up forcing many of these paths to be single edges if $\binom{t}{2} = \Omega(n)$. Setting $t = 3\sqrt{n}$ the argument outlined gives us that at least $3n$ of the paths in the subdivision of K_t must be single edges of G.

Fix a set $U \subset V$, $|U| = 3\sqrt{n}$. If U forms the vertices of a $K_{3\sqrt{n}}$ subdivision, then $e(U) \geq 3n$. By the Chernoff Bound we have

$$\mathbb{P}[|e(U) - \mathbb{E}[e(U)]| \geq \frac{1}{4}\mathbb{E}[e(U)]] \leq 2e^{-\mathbb{E}[e(U)]/48}$$

so that

$$\mathbb{P}[e(U) \geq 3n] \leq 2e^{-(9n-3\sqrt{n})/192} < e^{-n/25}$$

which implies that

$$\mathbb{P}[U \text{ forms the vertices of a } K_{3\sqrt{n}} \text{ subdivision}] < e^{-n/25}$$

Hence by the union bound

$$\mathbb{P}[G \text{ has a } K_{3\sqrt{n}} \text{ subdivision}] < \binom{n}{3\sqrt{n}} e^{-n/25} < \left(\frac{e\sqrt{n}}{3}\right)^{3\sqrt{n}} e^{-n/25} = o(1)$$

as $n \to \infty$. So with high probability, G does not contain a $K_{3\sqrt{n}}$ subdivision.

Thus, it follows that with high probability, $\chi(G) \geq \frac{n}{3\log n}$ and G has no $K_{3\sqrt{n}}$ subdivision, as desired. ■

Remark: This result shows that the chromatic number of a graph is a more esoteric global feature of the graph. In fact, the determination of the chromatic number of the random graph $G_{n,p}$ is an interesting problem which still has many unresolved facets, and we will examine some related results in the forthcoming chapters.
The fact that the chromatic number of a graph is a somewhat enigmatic invariant is further evidenced by the following theorem due to Erdős: Given $\varepsilon > 0$, and an integer k, there exist graphs $G = G_n$ (for n sufficiently large) such that $\chi(G) > k$, while every induced subgraph H on εn vertices satisfies $\chi(H) \leq 3$. This was again based on a random graph construction, and the interested reader can see this result in [5].

8.6 Why the Regularity lemma needs many parts

One of the most fundamental results in extremal graph theory is the celebrated Regularity lemma of Szemerédi, and the result is essentially a probabilistic paradigmatic statement for dense graphs[5]: Every graph can be decomposed into a bounded number of parts such that the graph between these parts is basically 'random-like'.

To make this more precise, we need the notion of an ε *regular partition*. For a pair of sets (not necessarily disjoint) U, W of vertices of a graph G, we denote by $e(U, W)$ the number of pairs $(u, w) \in U \times W$ such that $uw \in E(G)$and by teh *density* of the pair (U, W) we means $d(U, W) := \frac{e(U,W)}{|U||W|}$. A pair (u, W) is called ε-regular if whenever $A \subset U, B \subset W$ with $|A| \geq \varepsilon|U|$ and $|B| \geq \varepsilon|W|$ then the densities of the pairs (A, B) and (U, W) differ by at most ε, i.e., $|d(A, B) - d(U, W)| \leq \varepsilon$. A partition of the vertex set $V = \cup_{i=1}^{k} V_i$ is called an ε-regular partition if the number of irregular pairs (V_i, V_j) from among these sets is at most εk^2. The regularity lemma then states the following: Given $\varepsilon > 0$, there exists $M = O_\varepsilon(1)$ such that *every* graph admits an ε-regular partition into at most M parts. We will see this in greater detail in a subsequent chapter (Chapter).

The regularity lemma has been found to be of deep consequence in extremal graph theory, and since the proof procedure has an algorithmic flow to it, the regularity lemma also finds applications in Theoretical Computer Science (Property Testing). However, one drawback in the algorithmic application of the Regularity lemma is that the number M that is obtained through the proof is a tower of 2's with the height of the tower is $\Omega(1/\varepsilon^5)$. This makes the result immensely interesting and useful *theoretically*, but completely useless from a practical point of view. The natural question that arises is: Do we really need such a large M? Gowers [15]) settled this question in the affirmative. More precisely, there exist graphs G for which every ε-regular partition has partition size a tower of 2's with height $1/\varepsilon^c$ where $0 < c < 1$ is an absolute constant.

[5]If the graph is not dense, then the statement in the usual version is completely tautological.

While we shall not prove Gowers' result here, we shall give an easier and a weaker version of his result which also appears in the same paper.

Theorem 71. *Given* $1/1024 > \varepsilon > 0$*, there exist graphs such that every* ε*-regular partition has size a tower of* 2*'s with the height of the tower being of the order* $\log_2(1/\varepsilon)$*.*

9 Property B: Lower and Upper bounds

We have seen a brief glimpse of Property B in hypergraphs in the preceding chapter. As with many other problems and notions in this book, the main study into this notion was pioneered by Erdős and some of the problems introduced then are still open. In this brief digression of a chapter, we shall look at the current bounds (lower and upper) in the corresponding problem.

9.1 Introduction

For an integer $n \geq 2$, an $n-$uniform hypergraph $\mathcal{H}$ is an ordered pair $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a finite non-empty set of vertices and $\mathcal{E}$ is a family of distinct $n-$subsets of $\mathcal{V}$. A 2-coloring of $\mathcal{H}$ is a partition of its vertex set hv into two color classes, R and B (for red, blue), so that no edge in $\mathcal{E}$ is monochromatic. A hypergraph is 2-colorable if it admits a 2-coloring. For an $n-$uniform hypergraph, we define

$$m(n) := \arg\min_{|\mathcal{E}|} \{\mathcal{H} = (\mathcal{V}, \mathcal{E}) \text{ is 2-colorable}\} \tag{9.1}$$

2-colorability of finite hypergraphs is also known as "Property B". In [**?**], Erdős showed that $2^{n-1} < m(n) < O(n^2 2^n)$.

Let us start with a brief look at these results. The first of these, namely, that any n-uniform hypergraph with at most 2^{n-1} edges is 2-colorable is an immediate consequence of considering a random coloring and computing the expected number of monochromatic edges. The upper bound for $m(n)$ too follows from a simple randomized construction, and here is the gist.

In [12], Beck proved that $m(n) = \Omega(n^{\frac{1}{3}} 2^n)$ and this was improved to $m(n) = \Omega\left(2^n \sqrt{\frac{n}{\log n}}\right)$ by Radhakrishnan et al in [25]. In fact, Erdős-Lovaśz conjecture that $m(n) = \Theta(n2^n)$. Here, we outline the proofs of both Beck's and Radhakrishnan's results.

We will begin with some notation, if an edge $S \in \mathcal{H}$ is monochromatic, we will denote it as $S \in \mathbb{M},$, and in addition, if it is red (blue), we write $S \in RED$ ($S \in BLUE$). Also for a vertex $v \in \mathcal{V}$, $v \in RED$ and $v \in BLUE$ have a similar meaning. We shall freely abuse notation and denote by RED (resp. $BLUE$) both, the set of points colored RED as well as the set of edges of $\mathcal{H}$ that are colored RED and this should not create any confusion, hopefully.

9.2 Beck's result

Theorem 72 ([12])**.**

$$m(n) = \Omega(n^{\frac{1}{3}}2^n)$$

Proof. We will show that $m(n) > cn^{\frac{1}{3}-o(1)}2^n$, getting rid of $o(1)$ will need some asymptotic analysis which is not relevant to the class and hence is not presented here. Let $m := |\mathcal{E}| = k2^{n-1}$, we will show that $k > cn^{\frac{1}{3}-o(1)}$. The hypergraph will be colored in two steps.
Step 1: Randomly color all vertices with red or blue with probability 1/2 and independently.
Step 2: Randomly re-color vertices that belong to monochromatic edges independently with probability p.
For an edge S, $S(1)$ denotes its status after step 1 and $S(2)$ its status after step 2. For a vertex $v \in \mathcal{V}$, $v(1)$ and $v(2)$ have similar meanings. Let N_1 denote the number of monochromatic edges after step 1, then note that $\mathbb{E}(N_1) = k$. Also let N denote the number of monochromatic edges after step 2. For an appropriately chosen p, we will show that $\mathbb{E}(N) < 1$.

$$\begin{aligned}\mathbb{E}(N) &= \sum_{S\in\mathcal{E}} P(S(2) \in \mathbb{M}) = \sum_{S\in\mathcal{E}} (P(S(2) \in RED) + P(S(2) \in BLUE)) \\ &= 2\sum_{S\in\mathcal{E}} P(S(2) \in RED) \\ P(S(2) \in RED) &= \underbrace{P(S(1) \in \mathbb{M}, S(2) \in RED)}_{P_1} + \underbrace{P(S(1) \notin \mathbb{M}, S(2) \in RED)}_{P_2}\end{aligned}$$

It is easy to bound P_1

$$\begin{aligned}P_1 = P(S(1) \in RED, S(2) \in RED) + P(S(1) \in BLUE, S(2) \in RED) &= \frac{p^n + (1-p)^n}{2^n} \\ &\leq \frac{2(1-p)^n}{2^n} \qquad (9.2)\end{aligned}$$

In (9.2), we used the fact that p is small, in particular $p < 0.5$, this will be validated in the following analysis. Towards analyzing P_2, note that, for the vertices that were blue

after step 1 to have turned red, they must belong to blue monochromatic edges, i.e., for each $v \in S$ that is blue, there is an edge T such that $T \cap S \neq \Phi$ and $T \in BLUE$. Define

$$E_{ST} := \text{event } S(1) \notin \mathbb{M},\ T(1) \in BLUE,\ S \cap T \neq \Phi \text{ and } S(2) \in RED$$

Then we have

$$P_2 \leq \sum_{T \neq S} P(E_{ST}) \tag{9.3}$$

Let $U := \{v \in S \setminus T \,|\, v(1) \in BLUE\}$ and $E_{STU} :=$ event $S \cap T \neq \Phi$, $T(1) \in BLUE$, $U \in BLUE$ and $S(2) \in RED$, then

$$P(E_{ST}) = P\left(\bigvee_{U \subseteq S \setminus T} E_{STU}\right) \leq \sum_{U \subseteq S \setminus T} P(E_{STU})$$

For a fixed triple (S, T, U), for U to even flip it must belong to some other edge which is blue after step 1. But for an upper bound, let is just flip to red.

$$\begin{aligned}\mathbb{P}(E_{STU}) &\leq \frac{1}{2^{2n-|S\cap T|}} p^{|S\cap T|+|U|} = \frac{p}{2^{2n-1}} (2p)^{|S\cap T|-1} p^{|U|} \\ &\leq \frac{p}{2^{2n-1}} p^{|U|}\end{aligned}$$

Using this in (9.3), we have

$$\begin{aligned}\mathbb{P}(E_{ST}) &\leq \sum_{U \subseteq S\setminus T} \frac{p}{2^{2n-1}} p^{|U|} \leq \frac{p}{2^{2n-1}} \sum_{|U|=0}^{n-1} \binom{n-1}{|U|} p^{|U|} \\ &= \frac{(1+p)^{n-1} p}{2^{2n-1}} \leq \frac{2p(1+p)^n}{2^{2n}} \leq \frac{2p\exp(np)}{2^{2n}} \\ \implies \sum_{S \neq T} \mathbb{P}(E_{ST}) &\leq \frac{2mp\exp(np)}{2^{2n}}\end{aligned} \tag{9.4}$$

Using (9.2),(9.3),(9.4), we get (recall that $m = k2^n$)

$$\begin{aligned}\mathbb{E}(N) &\leq 2 \sum_S \left(\frac{m^2 p \exp(np)}{2^{2n}} + \frac{(1-p)^n}{2^n}\right) \\ &= 2\left(k^2 p \exp(np) + k(1-p)^n\right)\end{aligned} \tag{9.5}$$

For an arbitrary $\varepsilon > 0$, let $p = \frac{(1+\varepsilon)\log k}{n}$, then $k(1-p)^n \leq k\exp(-np) = k^{-\varepsilon}$ and $k^2 p \exp(np) = \frac{k^{3+\varepsilon}(1+\varepsilon)\log k}{n}$. So, (9.5) gives

$$\mathbb{E}(N) \leq 2k^{-\varepsilon} + \frac{2k^{3+\varepsilon}(1+\varepsilon)\log k}{n} \tag{9.6}$$

So, if $k \sim n^{1/3-2\varepsilon/3}$, then (9.6) will be less than 1, so that $\mathbb{P}(N = 0) > 0$. ■

9.3 The Radhakrishnan-Srinivasan (R-S) improvement

Theorem 73 ([25])**.**

$$m(n) = \Omega\left(2^n \sqrt{\frac{n}{\log n}}\right) \tag{9.7}$$

(R-S) take Beck's recoloring idea and improve it. Their technique is motivated by the following observation

Observation 74. *Suppose S is monochrome after step 1, then it suffices to re-color just one vertex in S, the rest can stay as is. So, after the first vertex in S changes color, the remaining vertices can stay put unless they belong to other monochromatic edges.*

This motivates the following modification, do not re-color all vertices simultaneously, put them in an ordered list and re-color one vertex at a time. Here is the modified step 2.

Step 2: For a given ordering, if the first vertex lies in a monochromatic edge, flip its color with probability p. After having colored vertices $1, \ldots, i-1$, if vertex i is in a monochromatic edge after having modified the first $i-1$ vertices, then flip its color with probability p.

The analysis proceeds along similar to that in the previous section until (9.2). Consider P_2. The last blue vertex v of S changes color to red because there is some $T \neq S$ such that T was blue after step 1 and $|S \cap T| = 1$. We shall say that S *blames* T (which we shall denote by $S \longmapsto T$) if this happens. Also, none of the vertices in T that were considered before v change their color to red. To summarize,

Lemma 75. *$S \longmapsto T$ iff*

1. *$|S \cap T| = 1$, call this vertex v.*
2. *$T(1) \in BLUE$ and v is the last blue vertex in S.*
3. *All vertices before v in S change color to red.*
4. *No vertex of T before v changes color to red.*

Then,

$$P_2 \leq \mathbb{P}\left(\bigvee_{T \neq S} S \longmapsto T\right) \leq \sum_{T \neq S} \mathbb{P}(S \longmapsto T) \tag{9.8}$$

Fix an ordering π on the vertices. With respect to this ordering, let v be the $(i_\pi + 1)^{th}$ vertex in S and the $(j_\pi + 1)^{th}$ vertex in T. If the index of w is less than that of v, we

write is as $\pi(w) < \pi(v)$. Also define,

$$S_v^- := \{w \in S \,|\, \pi(w) < \pi(v)\}$$
$$S_v^+ := \{w \in S \,|\, \pi(w) > \pi(v)\}$$

T_v^- and T_v^+ have similar meanings. To compute $\mathbb{P}(S \longmapsto T)$, we will need to list some probabilities

1. $\mathbb{P}(v(1) \in BLUE, v(2) \in RED) = \dfrac{p}{2}$
2. $\mathbb{P}\left((T \setminus v)(1) \in BLUE\right) = \dfrac{1}{2^{n-1}}$
3. $\mathbb{P}(S_v^+(1) \in RED) = \dfrac{1}{2^{n-i_\pi-1}}$
4. $\mathbb{P}(T_v^-(2) \notin RED \,|\, T(1) \in BLUE) = (1-p)^{j_\pi}$
5. For $w \in S$ with $\pi(w) < \pi(v)$,

$$\mathbb{P}((w(1) \in RED) \text{ or } (w(1) \in BLUE, w(2) \in RED) \,|\, S \notin \mathbb{M}) = \frac{1+p}{2}$$

So, subject to this ordering π,

$$\begin{aligned}\mathbb{P}(S \longmapsto T) &\leq \frac{p}{2} \cdot \frac{1}{2^{n-1}} \cdot \frac{1}{2^{n-i_\pi-1}} \cdot (1-p)^{j_\pi} \cdot \left(\frac{1+p}{2}\right)^{i_\pi} \\ &= \frac{p}{2^{2n-1}}(1-p)^{j_\pi}(1+p)^{i_\pi} \end{aligned} \tag{9.9}$$

Let the ordering π be random. Then $\mathbb{P}(S \longmapsto T) = E_\pi \mathbb{P}(S \longmapsto T \,|\, \pi)$. A random ordering is determined as follows. Each vertex picks a real number uniformly at random from the interval $(0, 1)$, this real number is called its delay. Then the ordering is determined by the increasing order of the delays.

Lemma 76.

$$\mathbb{P}(S \longmapsto T) = \mathbb{E}\left(\mathbb{P}(S \longmapsto T \,|\, \pi)\right) \leq \frac{p}{2^{2n-1}} \tag{9.10}$$

Proof. Let the delay of a vertex w be denoted by $\ell(w)$. Let $U := \{w \in S \setminus v | w(1) \in BLUE\}$, then $\ell(w) \leq \ell(v)$, since v, by definition, is the last blue vertex in S. Also for each $w \in T$, either $\ell(w) > \ell(v)$ or w did not flip its color in step 2. So, for $w \in T$ $P(\ell(w) \leq \ell(v), \text{w flips color}) = px$, so $\mathbb{P}(\ell(w) > \ell(v) \text{ or w did not flip}) = (1 - px)$. Now,

conditioning on $\ell(v) \in (x, x+dx)$ and with some abuse of notation, we can write

$$\mathbb{P}\left(S \longmapsto T, |U| = u \,|\, \ell(v) = x\right) = \underbrace{\frac{1}{2^{2n-1}}}_{\text{coloring after step 1}} \underbrace{x^u}_{\ell(U)\leq x} \underbrace{p^{1+u}}_{U\cup\{v\} \text{ flip to red}} (1-px)^{n-1}$$

$$\begin{aligned}
\implies \mathbb{P}(S \longmapsto T) &\leq \sum_{u=0}^{n-1} \binom{n-1}{u} \int_0^1 \frac{1}{2^{2n-1}} p^{1+u} x^u (1-px)^{n-1} dx \\
&= \frac{p}{2^{2n-1}} \int_0^1 \left(\sum_{u=0}^{n-1} \binom{n-1}{u} (px)^u \right) (1-px)^{n-1} dx \\
&= \frac{p}{2^{2n-1}} \int_0^1 (1-p^2x^2)^{n-1} dx \\
&\leq \frac{p}{2^{2n-1}} \qquad (9.11)
\end{aligned}$$

■

Proof of theorem 73. Using (9.11) in (9.8), we get $P_2 \leq \frac{mp}{2^{2n-1}}$. Recall that $P_1 \leq \frac{2(1-p)^n}{2^n}$, summing over all edges S, we get

$$\mathbb{E}(N) \leq \frac{k(1-p)^n}{2} + \frac{k^2 p}{2} \qquad (9.12)$$

Compare (9.12) with (9.5) and note that $\exp(np)$ is not present in (9.12). For an arbitrary $\varepsilon > 0$, setting $p = \frac{(1+\varepsilon)\log k}{n}$ and approximating $(1-p)^n \approx \exp(-np)$, we get

$$\mathbb{E}(N) \leq 0.5 \left(k^{-\varepsilon} + (1+\varepsilon)\frac{k^2 \log k}{n} \right) \qquad (9.13)$$

Clearly $k \sim \sqrt{\frac{n}{\log n}}$ makes $\mathbb{E}(N) < 1$ giving the result. ■

Spencer's proof of lemma 76. Aided by hindsight, Spencer gives an elegant combinatorial argument to arrive at (9.11). Given the pair of edges S, T with $|S \cap T| = 1$, fix a matching between the vertices $S \setminus \{v\}$ and $T \setminus \{v\}$. Call the matching $\mu := \{\mu(1), \ldots, \mu(n-1)\}$, where each $\mu(i)$ is an ordered pair (a, b), $a \in S \setminus \{v\}$ and $b \in \setminus\{v\}$, define $\mu_s(i) := a$ and $\mu_t(i) := b$. We condition on whether none, one or both vertices of $\mu(i)$ appear in $S_v^- \cup T_v^-$, for each $1 \leq i \leq n-1$. Let $X_i = |\mu(i) \cap (S_v^- \cup T_v^-)|$. Since the ordering is uniformly

random, X_i and X_j are independent for $i \neq j$. From (9.9), consider $\mathbb{E}\left((1-p)^{j_\pi}(1+p)^{i_\pi}\right)$.

$$\begin{aligned}
\mathbb{E}\left((1-p)^{j_\pi}(1+p)^{i_\pi} \mid \mu \cap S_v^- \cup T_v^-\right) &= \mathbb{E}\left((1-p)^{\sum_{i=1}^{n-1}\mathbb{I}(\mu(i)\cap S_v^- \neq \Phi)}(1+p)^{\sum_{i=1}^{n-1}\mathbb{I}(\mu(i)\cap T_v^- \neq \Phi)}\right) \\
&= \mathbb{E}\left(\prod_{i=1}^{n-1}(1-p)^{\mathbb{I}(\mu_s(i)\in S_v^-)}(1+P)^{\mathbb{I}(\mu_t(i)\in T_v^-)}\right) \\
&= \prod_{i=1}^{n-1}\mathbb{E}\left((1-p)^{\mathbb{I}(\mu_s(i)\in S_v^-)}(1+P)^{\mathbb{I}(\mu_t(i)\in T_v^-)}\right) \\
&= \prod_{i=1}^{n-1}\left(\frac{1}{4}(1-p+1+p+1+1-p^2)\right) \\
&= \prod_{i=1}^{n-1}\left(1-\frac{p^2}{4}\right) < 1
\end{aligned}$$

which implies that

$$\mathbb{E}\left((1-p)^{j_\pi}(1+p)^{i_\pi}\right) = \mathbb{E}\left(\mathbb{E}\left((1-p)^{j_\pi}(1+p)^{i_\pi} \mid \mu \cap S_v^- \cup T_v^-\right)\right) < 1$$

and that implies

$$\mathbb{P}(S \longmapsto T) < \frac{p}{2^{2n-1}}$$

and completes the proof. ■

9.4 And then came Cherkashin and Kozik...

A nice coda to this chapter is a beautiful argument due to Cherkashin-Kozik (2015)[13] (a 'Book Proof') which greatly simplifies the R-S argument (though it gives the same bound) to essentially ridding the argument of the recoloring as well. But, it builds upon the ideas from the previous results, and all the previous hard work now gives payoff in a very satisfactory manner.

As before, suppose $e(\mathcal{H}) = k2^{n-1}$. The coloring algorithm puts all the vertices in a (random) order, and processes one vertex at a time. A vertex is give a default color of *BLUE unless it ends up coloring some edge BLUE* in which case, we color the vertex *RED*. Note that the only monochromatic edges are all *RED* at the end of this procedure. The ordering of the vertices is decided in the same manner as in the $R-S$ algorithm. Each vertex v picks independently and uniformly, $X_v \in [0,1]$ at random. As observed in the R-S algorithm, if an edge is colored *RED* at the end of this procedure, there is some edge T such that $|S \cap T| = 1$, and the common vertex of these edges is *the last vertex of* S *and the first vertex of* T. We shall, following Cherkashin and Kozik sat that in this case (S,T) is a *conflicting pair*. We shall estimate the probability that the coloring produces no *RED*

edges, and to do that we shall estimate the probability that there exists a conflicting pair.

Let $0 < p < 1$ be a parameter. Call an edge S an *p-extreme* edge if for each $v \in S$, $X_v \leq \frac{1-p}{2}$ or $X_v \geq \frac{1+p}{2}$. To estimate the probability that there is a conflicting pair, we consider the two possibilities: One of the pair of edges is an extreme edge, and the other case, when neither of the edges is extreme. The probability of the former is at most $2 \cdot (k2^{n-1}) \cdot \left(\frac{1-p}{2}\right)^n = k(1-p)^n$. In the other case, note that if $S \cap T = \{v\}$ then we must have $X_v \in (\frac{1-p}{2}, \frac{1+p}{2})$ and for all the other $u \in S, w \in T$ we have $X_u < X_v$ and $X_v < X_w$ and the probability of this is $(k2^{n-1})^2 \cdot pX_v^{n-1}(1-X_v)^{n-1} < k^2 4^{n-1} \cdot p\left(\frac{1}{4}\right)^{n-1} = pk^2$.

Hence, if $pk^2 + k(1-p)^n < 1$ then we are done, and the asymptotics for this are the same as seen in the discussion following the R-S algorithm.

10 The Lovász Local Lemma and Applications

Not all events in a probability space occur reasonably often; some events are a once-in-a-lifetime events, and yet, they do occur.

Most of the applications of probabilistic methods we have thus far encountered in fact prove that an overwhelming majority of 'instances' from the corresponding probability spaces satisfy the criteria that we sought, so that in effect, one could say that 'almost all' of those instances would gives examples (or counterexamples) for the problem at hand. While this makes it very useful from an algorithmic point of view - one could envisage a randomized algorithm that would contruct the desired object - it may not always be the case that the 'good' or 'desirable' configurations we seek are plenty. For instance, suppose we have two large finite sets A, B of equal size, then we know that there is an injection from A to B but almost all random maps are bound to be bad. The so-called Lovász Local Lemma - discovered by Erdős and Lovász - gives us a very useful and important tool that allows us to show that certain probabilities are non-zero, even though they might be extremely small. In this chapter, we shall consider the lemma,and see some applications.

10.1 The Lemma and its proof

We know that given a set of *independent* events, $A_1, A_2, ..., A_n$, each with nonzero probability, then $P(A_1 \cup A_2 \cup ... \cup A_n) > 0$. The idea behind the Lovasz Local Lemma (LLL) is that in certain cases we can relax the assumption that the A_i be mutually independent, as long as each A_i is only dependent on a small number of the rest. We can visualize this by imagining a graph with vertices labeled by the A_i, and edge set $\{\{A_i, A_)\} \ : \ A_i \text{ and } A_j \text{ are dependent}\}$. Call this the *dependency graph*. Then the degree of vertex A_i is the number of other events with which A_i is dependent. We call this degree the *dependence degree* of A_i. Intuitively, if the maximum dependence degree is small, then we should still have nonzero probability of all the events occurring. The LLL formalizes this.

We now state the LLL formally, in its most general form:

Theorem 77 (Local Lemma)**.** *Suppose ξ_i $(i = 1, 2, \ldots, N)$ are events in a probability space $(\Omega, \mathbb{P})$ and suppose D is the dependence graph of $\{\xi_i\}$ which is constructed such that ξ_i is jointly independent of $\{\xi_j \mid (i,j) \notin E(D)\}$. Suppose there exist reals $0 \leq x < 1, i = 1, \ldots, N$ such that $\mathbb{P}(\xi_i) \leq x_i \prod_{ij \in E(D)}(1 - x_j)$ for each i. Then,*

$$\mathbb{P}\Big(\bigcap_{i=1}^{N} \overline{\xi_i}\Big) \geq \prod_{i=1}^{N}(1 - x_i) > 0.$$

In particular, with positive probability, none of the events ξ_i occur.

We will present the proof shortly. As an immediate corollary, we have:

Corollary 78 (Symmetric form of LLL)**.** *Let ξ_i be events in $(\Omega, \mathbb{P})$. Suppose $\mathbb{P}(\xi_i) \leq p$, and if the maximum degree in the dependence graph is d, and suppose $ep(d+1) \leq 1$, then with positive probability none of the ξ_i's occur.*

Proof. Follows from Chapter 77, take $x_i = 1/(d+1)$. Then note that

$$x_i \prod_{j \leftrightarrow i}(1 - x_j) \geq \frac{1}{\Delta + 1}\left(1 - \frac{1}{\Delta + 1}\right)^{\Delta} = \frac{1}{\Delta + 1}\left(\frac{\Delta}{\Delta + 1}\right)^{\Delta}.$$

Note that $\left(\frac{\Delta+1}{\Delta}\right)^{\Delta} = \left(1 + \frac{1}{\Delta}\right)^{\Delta} \leq e$, so for each i,

$$x_i \prod_{j \leftrightarrow i}(1 - x_j) \geq \frac{1}{e(\Delta + 1)} \geq p \geq P(A_i).$$

Then applying the general version of the local lemma yields the result. ■

The above corollary allows for a more useful symmetric version of the local lemma, which will be the version that will also be referred to as the local lemma. In most situations, this is the version that is easy to apply directly.
We now present a proof of the local lemma. As we can see, this is a simple induction argument, coupled with an elementary conditioning argument.

Proof of the Local lemma. We have,

$$\begin{aligned} &\mathbb{P}(\overline{\xi_1} \cap \overline{\xi_2} \cap \cdots \cap \overline{\xi_N}) \\ &= \mathbb{P}(\overline{\xi_1}).\mathbb{P}(\overline{\xi_2} \mid \overline{\xi_1}) \cdots \mathbb{P}(\overline{\xi_N} \mid \overline{\xi_1} \cap \cdots \cap \overline{\xi_{N-1}}) \end{aligned}$$

Therefore, it suffices to show, if $S \subseteq \mathbb{N}$ and $i \notin S$,

$$\mathbb{P}(\xi_i \mid \bigcap_{j \in S} \overline{\xi_j}) \leq x_i$$

We will show this by induction on $|S|$. Clearly the statement holds for $|S| = 0$. Let $S_1 = \{j \in S : ij \in E(D)\}$ and $S_2 = S \setminus S_1$. We have,

$$\begin{aligned}
&\mathbb{P}\big(\xi_i \mid \bigcap_{j\in S_1} \overline{\xi_j} \cap \bigcap_{l\in S_2} \overline{\xi_l}\big) \\
=&\frac{\mathbb{P}\big(\xi_i \cap \bigcap_{j\in S_1} \overline{\xi_l} \mid \bigcap_{l\in S_2} \overline{\xi_l}\big)}{\mathbb{P}\big(\bigcap_{j\in S_1} \overline{\xi_l} \mid \bigcap_{l\in S_2} \overline{\xi_l}\big)} \\
\leq&\frac{x_i . \prod_{ij\in E(G)}(1-x_j)}{\mathbb{P}\big(\bigcap_{j\in S_1} \overline{\xi_l} \mid \bigcap_{l\in S_2} \overline{\xi_l}\big)}
\end{aligned}$$

It suffices to show that,

$$\mathbb{P}\big(\bigcap_{j\in S_1} \overline{\xi_j} \mid \bigcap_{l\in S_2} \overline{\xi_l}\big) \geq \prod_{ij\in E(G)} (1-x_j)$$

Suppose $S_1 = \{j_1, j_2, \ldots, j_k\}$. We have,

$$\begin{aligned}
&\mathbb{P}\big(\bigcap_{i=1}^{k} \overline{\xi_{j_i}} \mid \bigcap_{l\in S_2} \overline{\xi_l}\big) \\
=&\prod_{i=1}^{k} \Big(1 - \mathbb{P}\big(\xi_{j_i} \mid \bigcap_{l=1}^{i-1} \overline{\xi_{j_l}} \cap \bigcap_{l\in S_2} \overline{\xi_l}\big)\Big) \\
\geq&\prod_{i=1}^{k}(1-x_{j_i}) \geq \prod_{ij\in E(D)} (1-x_j).
\end{aligned}$$

■

10.2 Applications of the Lovász Local Lemma

We now illustrate several applications of the symmetric version of the Local Lemma.

Example: Property B
Recall that a hypergraph has Property B, or is 2-colorable, if there is a coloring of its vertices using two colors such that no edge is monochromatic. We call a hypergraph *k-uniform* if each of its edge sets contains k elements. We call it *d-regular* if each vertex is involved in exactly d edges.

Question: Suppose $\mathcal{H}$ is a k-uniform, d-regular hypergraph. What conditions on $\mathcal{H}$ will ensure that Property B is satisfied?

Let each vertex toss a fair coin. If the toss reads heads, we color the vertex red. If tails, we color it blue. For each edge A, consider the event E_A that A is monochrome. Then 2-colorability of $\mathcal{H}$ is equivalent the case that none of the events E_A occur, that is, the event $\wedge_{A\in\mathcal{H}}\overline{E_A}$. Now,

$$P(A) = P(A \text{ is monochrome}) = \frac{2}{2^k} = \frac{1}{2^{k-1}}.$$

Now, E_A is dependent with E_B if $A \cap B \neq \emptyset$. Since edge A contains k vertices, each of which is contained in $d-1$ other edges, we obtain an upper bound for the dependence degree as $|\{B \in \mathcal{H} | B \cap A \neq \emptyset\}| \leq (d-1)k$. Thus, by the Local Lemma, if

$$e\frac{1}{2^{k-1}}[(d-1)k+1] \leq 1,$$

then we can guarantee that

$$P(\wedge_{A\in\mathcal{H}}\overline{E_A}) > 0,$$

so in particular, we have the following:

Theorem: *If $\mathcal{H}$ is k-regular and k-uniform, then for $k \geq 9$, $\mathcal{H}$ has Property B.*

Remark: It turns out that this result is true even for $k \geq 7$. Another aspect of the proof of this theorem is that if n (the number of edges) is large, then this probability goes to zero, but it is nonetheless strictly greater than zero. Also, the Lovász Local Lemma does not extend if there are infinitely many events.

Example: A Substitute for the Pigeonhole Principle
We know from the Pigeonhole Principle that if S and T are finite sets, with $|T| \geq |S|$, then we can find a function $f : S \to T$ such that f is injective (one-to-one).

But suppose we didn't know the Pigeonhole Principal (!). Then we could try picking a function f at random by selecting, uniformly and independently, the images of the elements of S in T. Then,

$$E(|f(S)|) = \sum_{t\in T} P(t \text{ is selected by } f) = |T| \cdot \left[1 - \left(1 - \frac{1}{|T|}\right)^{|S|}\right],$$

so by the first moment method, there exists an injection f if this is greater than $|S|$.

Alternatively, we could let N be the number of pairs of distinct members of S which have the same image in T under a randomly chosen function f. f will be injective provided that $N = 0$. Again using the first moment method,

$$E(N) = \sum_{\{x,y\}\in\binom{S}{2}} P(f(x) = f(y)) = \frac{1}{|T|}\binom{|S|}{2},$$

so we see that if $|T| > \binom{|S|}{2}$, there exists an injection.

We can get a remarkable improvement, however, if we use the Local Lemma. On this note, for any edge $E = \{x, y\}$, let A_E be the event that both x and y have the same image in T under the chosen function f. Then,

$$P(A_E) = \frac{1}{|T|}.$$

Since A_E is independent of $A_{E'}$ if $E \cap E' = \emptyset$, the dependence degree of these events can be at most $2(|S| - 2)$ (we can get a dependent edge by replacing either x or y with one of the remaining $|S| - 2$ elements). Thus, by the Local Lemma, if $\frac{e(2|S|-3)}{|T|} \leq 1$, then with nonzero probability, f is injective. Thus, using the Local Lemma, we see that we need only have that $|T| \geq e(2|S| - 3)$ in order to endure the existence of an injection $S \to T$.

Example: Cycles in digraphs of specific sizes

Alon and Linial consider the following general question: Given a graph, when can we guarantee the existence of 'special' types of cycles? In the case of directed graphs, questions as simple as those concerning even directed cycles are difficult. However, there is a positive result for the case of a directed graph $\mathcal{D}$. If $\deg(\mathcal{D}) \geq 7$ and $\mathcal{D}$ is regular, then the answer is yes.

Theorem 79. *Suppose $\mathcal{D}$ is a directed graph with maximum in degree Δ and minimum outdegree δ. Then, for $k > 0$, if $e(\delta\Delta + 1)\left(1 - \frac{1}{k}\right)^{\delta} \leq 1$, then there exists a directed cycle in $\mathcal{D}$ of length $0(\mod k)$.*

First, consider the following observations. Let c be a $k-$coloring of $V(\mathcal{D})$. Let the colors be $\{0, 1, \ldots, k-1\}$. If from a vertex x, colored i, there exists an edge from x to a vertex of color $i + 1 (\mod k)$ for every $x \in V(\mathcal{D})$, then there exists a directed cycle in $\mathcal{D}$ of length $0(\mod k)$. Thus, Theorem 79 is true if there is a coloring such that at each x, the aforementioned local condition is satisfied.

Proof. Let us randomly color V using k-colors with each vertex colored independently by a color in $\{0, 1, \ldots, k-1\}$. We may assume that $d^+(v) = \delta$ for any $v \in V$, because if not, we can throw away certain edges without tweaking the problem, until this condition is satisfied. Define the following event for each $v \in V$,

$$E_v := \text{There is no vertex } u \text{ in } N^+(v) \text{ such that } \text{color}(u) = \text{color}(v) + 1 (\mod k).$$

Notice that $\mathbb{P}(E_v) = \left(1 - \frac{1}{k}\right)^{\delta}$. We need to show that $\mathbb{P}(\wedge_v \overline{E_v}) > 0$. Moreover, E_u and E_v are dependent if $u \in N^+(v)$. Also, E_u and E_v are dependent if they share a common out-neighbor. Also, E_v is determined by the color choices of v and $N^+(v)$. Therefore, $d = \delta\Delta$ in the Lovász Local Lemma. Hence, if $e(\delta\Delta + 1)\left(1 - \frac{1}{k}\right)^{\delta} \leq 1$ then there exists an oriented cycle $\mathcal{D}$ of length $0(\mod k)$. ∎

10.3 A Theorem of Erdós and Lovaśz on a problem of Straus

The following question was proposed by Straus[1]: Given $S \subseteq \mathbb{R}$ such that $|S| < \infty$, is there a k-coloring of $\mathbb{R}$ such that EVERY translate of S is MULTICOLORED?

Definition 80 (MULTICOLORED)**.** *A set $S \subseteq \mathbb{R}$ is multicolored if all k colors appear in the set.*

Definition 81 (k-coloring of $\mathbb{R}$)**.** *A k-coloring of $\mathbb{R}$ is a function $c : \mathbb{R} \to k$.*

Theorem 82 (Erdós-Lovaśz)**.** *If $|S| \geq (3 + o_k(1))k \log(k)$, then the answer is YES.*

Remark. *This bound is optimal upto a constant. There exists a set S of size $k \log(k)$, for which it isn't possible to obtain a k-coloring of $\mathbb{R}$ such that every translate of S is multicolored.*

Proof. First let us fix a finite set X corresponding to the translations (i.e. we will consider the translates $x + S$ for $x \in X$). Fix a large finite set Ω such that $x + S \subseteq \Omega$, $\forall x \in X$. Write $|S| = m$ for simplicity. Color each $w \in \Omega$ independently + randomly in $[k]$. Define the events,

$$\xi_x := x + S \text{ isn't multicolored, } \forall x \in X.$$

We have,

$$\mathbb{P}(\xi_x) \leq k(1 - 1/k)^m$$

which follows as the probability of some color being missing from the set $(x + S)$ is $(1 - 1/k)^m$. The dependence graph D looks like,

$$\xi_x \leftrightarrow \xi_y \text{ if and only if } (x + S) \cap (y + S) \neq \emptyset.$$

For fixed x, we want to compute,

$$\#\{y \,:\, \xi_x \leftrightarrow \xi_y\}$$

where $y \neq x$. If $\exists s_1, s_2 \in S$ such that $x + s_1 = y + s_2$, then $y = x + s_1 - s_2$. Therefore, maximum dependence degree $\leq m(m-1)$. So if

$$ek(1 - 1/k)^m(m(m-1) + 1) \leq 1, \tag{10.1}$$

holds, then from Chapter 78 it follows that with positive probability none of the events occur i.e. $\mathbb{P}\big(\bigcap_{x \in X} \overline{\xi_x}\big) > 0$. Notice that Chapter 10.1 holds for $m = (3 + o(1)).k \log(k)$. To see why, note that

$$\begin{aligned}
eke^{-m/k}m^2 \leq 1 \quad &\implies \\
e^{m/k} \geq ekm^2 \quad &\implies \\
m \geq k \log k + 2k \log m + c \quad &\implies \\
\geq k \log k + 2k \log(k \log k) \quad &\text{(substituing the above)} \implies \\
\geq 3k \log k + 2k \log \log k + \ldots &
\end{aligned}$$

[1] collaborated with both Erdós and Einstein!

We have thus shown that for every finite set of translates X, the theorem holds. To prove the theorem fully (i.e. for the case of the set of translates being infinite), we'll use Tychonoff's theorem! The space of $[k]$-colorings of $\mathbb{R}$ is simply $[k]^{\mathbb{R}}$. Endow each component $[k]$ of χ with the discrete topology. Since each component $[k]$ is a finite set, it is compact under discrete toplogy. Since arbitrary product of compact spaces is compact, it follows that $\chi = [k]^{\mathbb{R}}$ is compact under the product topology. Let

$$\mathcal{C}_x = \{c \in [k]^{\mathbb{R}} \ : \ x + S \text{ is multicolored wrt } c\}.$$

Note that $\mathcal{C}_x$ is closed in χ with respect to the product topology. We have shown that $\bigcap_{x \in X} \mathcal{C}_x \neq \emptyset$, for every finite subset X of $\mathbb{R}$. Since $\mathcal{C}_x$ are closed subspaces of the compact space χ and satisy the finite intersection property, it follows that $\bigcap_{x \in \mathbb{R}} \mathcal{C}_x \neq \emptyset$. ■

Remark: It turns out that the $k \log k$ term in the above expression is not only sufficient, but also necessary.

10.3.1 Linear Arboricity Conjecture of Harary

Definition 83 (Arboricity of a graph)**.** *The arboricity of a graph G is the minimum number of edge-disjoint forests needed to partition $E(G)$.*

As an example, arboricity of the 5-cycle is 2. Note that arboricity of a graph G is 1 if and only if it is a forest.

Definition 84 (Linear)**.** *Each tree in the forest decomposition of G must be a path.*

For a given graph G, we denote its linear arboricity by $la(G)$. Every graph G can be embedded in a d-regular subgraph by adding more vertices and edges. Let G be a graph on n vertices such that $\Delta(G) \leq d$. Let $F_1, F_2, \ldots, F_r$ be a (linear) forest decomposition of G. Since each F_i is a forest, $e(F_i) \leq (n-1)$. We have,

$$(n-1)r \leq \sum_{i=1}^{r} e(F_i) = \sum_{i=1}^{r} d_i/2 \leq dn/2$$

This gives (taking $r = la(G)$),

$$d/2 < dn/2(n-1) \leq la(G).$$

We have the following conjecture by Harary[**?**] which essentially says that this bound is tight!

Conjecture 85 (Harary,1980)**.** *$la(G) \leq \lceil (d+1)/2 \rceil$, where $d =$ maximum degree of the graph G.*

Following is a directed version of the conjecture, which if true, will imply the undirected version.

Conjecture 86 (Directed version)**.** *Suppose D is a directed d-regular digraph. For each v, if $N^+(v) = \{u : (v,u) \in E(D)\}$, then $d^+(v) = |N^+(v)| = d$ and similarly, $d^-(v) = d$. If D is directed and d-regular, then $dla(D) = d + 1$.*

Theorem 87 (Alon)**.** *If G is directed and d-regular, then*

$$dla(G) \leq d + O(d^{3/4} \log^{1/2}(d)) = d(1 + o_d(1)).$$

In particular, the Linear arboricity conjecture holds asymptotically.

Given D directed and d-regular, create two copies V and V' of the vertex set of D such that $(u, v') \in E(T_D)$ if and only if $(u, v) \in E(D)$, where $u \in V$ and $v' \in V'$. By construction, T_D is a d-regular, bipartite graph. From Hall's theorem, it follows that

$$E(T_D) = M_1 \uplus M_2 \uplus \cdots \uplus M_d$$

where M_i's are perfect matchings in T_D. Each perfect matching in the bipartite graph T_D corresponds to a union of disjoint cycles in the graph D. Therefore,

$$E(D) = F_1 \uplus F_2 \uplus \cdots \uplus F_d$$

where each F_i is a union of disjoint cycles in the graph D. So clearly, $\text{dla}(D) \leq 2d$.
Idea: If it possible to choose one edge from each cycle such that the resulting edges form a matching, then we have $\text{dla}(D) \leq d + 1$. Look at the line graph, we want an independent set there! More generally, suppose we have

$$V(G) = V_1 \uplus \cdots \uplus V_r$$

Can one pick a TRANSVERSAL INDEPENDENT set with respect to this partition? Pick one vertex $v_i \in V_i$ such that the resulting set is independent!

Theorem 88 (Alon)**.** *Suppose*

$$V(G) = V_1 \uplus \cdots \uplus V_r$$

If $\Delta(G) \leq d$ and $|V_i| \geq 2ed$, then G admits an independent transversal for this partition.

Proof. Pick $u \in V_i$ uniformally at random. For $1 \leq i < j \leq r$, $\xi_{ij} \equiv v_i v_j \in E(G)$. Notice, $\mathbb{P}(\xi_{ij}) = e(V_i, V_j)/4e^2d^2 = 1/2e$. This doesn't work out! ■

Remark. *In the above attempt, things didn't work out because we had too few bad events!*

10.3.2 Directed Linear Arboricity Conjecture

Suppose D is a d-regular directed graph ($d^+(v) = d^-(v) = d, \forall v$). Then $\text{dla}(D) \leq (d+1)$. Recall that we defined $\text{dla}(D)$ as the minimum number of colors needed to color $E(D)$ such that each color class induces a LINEAR FOREST i.e. each connected component is a directed path. We have already seen that $\text{dla}(D) > d$. To avoid the issue faced before in finding the transversal independent set, we will sparsify the bad events!

Theorem 89 (Alon,[?, ?]). *Suppose $V(G) = V_1 \uplus \cdots \uplus V_r$ with $\Delta(G) \leq d$, and $|V_i| \geq \lceil 2ed \rceil$. Then the collection $\{V_i\}$ admits an independent transversal, i.e. $\exists v_i \in V_i$ such that $I = \{v_1, \ldots, v_r\}$ is independent in G.*

Proof. WLOG $|V_i| = \lceil 2ed \rceil$ (throw vertices out!). Pick $v_i \in V_i$ independently + uniformly, i.e. one v_i is picked randomly from each V_i. For each edge $e = \{u, v\}$, let $\xi_e =$ both u, v are picked, where $e = (u, v)$. We have, 1+ dependence degree $\leq 2.2ed.d = 4ed^2$. This follows as $|V_i| = |V_j| = 2ed, \forall\, i \neq j$ and degree of each vertex is d. If $u \in V_i$ and $v \in V_j$ $(i \neq j)$,

$$\mathbb{P}(\xi_e) \leq 1/4e^2d^2 \Rightarrow e(1/4e^2d^2)(4ed^2) = 1.$$

The local lemma applies. ■

Remark. *The best constant c such that if $|V_i| \geq cd$, then there is an independent transversal is ≤ 2 (best constant has to be > 1).*

Given D, construct H bipartite as $H = (V, V', E)$, where $V' \simeq V = V(D)$ and $(u, v') \in E(H)$ if and only if $(u, v) \in E(D)$. Since D is d-regular, H is also d-regular. We have,

$$E(H) = M_1 \uplus \cdots \uplus M_d$$

where each M_i is a matching in H. We obtain this by applying Hall's theorem repeatedly and reducing the graph by removing perfect matchings at each step. Note that such a reduction preserves the regularity of graph. Each matching M_i induces a partition of $V(D)$ into vertex-disjoint cycles.

Observation 90. *Consider Chapter 10.1, if one can pick one edge from each of these cycles (in these cycle decompositions) such that the chosen edges form a matching, then we can color each $\mathcal{M}_i$ with color i and color the matching formed by the 'chosen' edges with one extra color. Total $(d+1)$ colors will be used in this process.*

Consider the line graph L of the given graph D. Let $V(L) = E(D) = \biguplus_{i=1}^{N} V_i$, where each V_i is a set of edges forming a cycle in some matching $\mathcal{M}_j$. So in terms of the line graph, we have seen that if each $|V_i| \geq 2ed$, then $\{V_i\}$ admits an independent transversal. Small cycles are a nuisance! If a small cycle pops up, then it might not be possible to choose an independent transversal among the sets $\{V_i\}$. Observe that the degree of each vertex in the line graph L is $\leq (4d-2)$. This follows by looking at all the possible edges incident on one of u or v, given the edge $e = (u, v)$. Since $2e(4d-2) \leq 8ed$, it is an

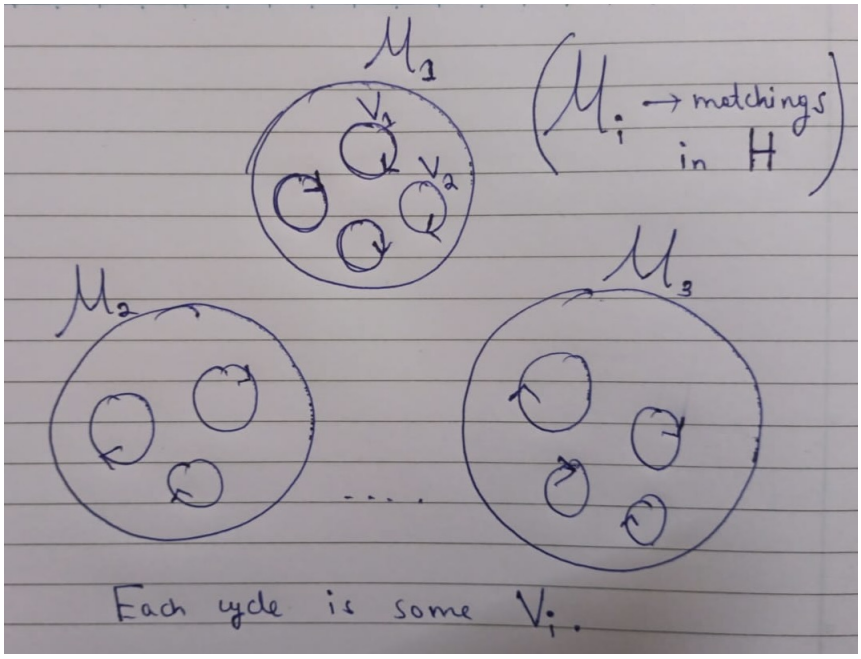


Figure 10.1: We work with the line graph of D to check if the chosen edges from each cycle form a matching

easy consequence of Chapter 88 to see that if the girth of the graph D is $\geq 8ed$, then $\text{dla}(D) = (d+1)$. The conjecture holds for graphs of large girth! How do we proceed from here? Given D, if we can partition

$$E(D) = D_0 \uplus D_2 \uplus \cdots \uplus D_{p-1}$$

for some integer p such that each D_i has large girth and "proportionally small" degree, then could apply this result on each D_i. Let degree of each vertex in some D_i be $\approx d/p$. Then we have,

$$\text{dla}(D) \leq (d/p+1)p = d+p$$

If $p = o(d)$, then this gives a bound $\text{dla}(D) = d + o(D)$.

The main idea: Pick a p (shall see how to do this!) and if $E(D)$ can be partitioned into digraphs $D_0, D_1, \ldots, D_{p-1}$ such that,

- $\Delta^+(D_i), \Delta^-(D_i) \approx d/p$
- $\text{girth}(D_i) \gg (d/p)$
- $p = o(d)$,

then one can repeatedly use the result for digraphs of "large girth" to get $\text{dla}(D) \leq (d/p+1).p = d + o(d)$. Let $p \gg \sqrt{d}$. We will show that the minimum cycle length is $> p$

by constructing D_i's and using $p \geq \text{girth}(D_i) \geq d/p$. The inequality $p > d/p$ holds as we chose $p >> \sqrt{d}$.

Remark. *It is possible to attain (1) by locally splitting at each vertex to obtain (d/p) degree for each split and then using Chernoff type bound to obtain a global split.*

Here is the strategy: First we pick a large prime p (of order $\sqrt{d}$, as one back of the envelope calculation). We shall color the edges of D using colors $\{0, 1, \ldots, p-1\}$ such that $\forall v \in V(D)$ and each $i \in \{0, 1, \ldots, p-1\}$, we have

$$N^+(v, i) := \#\{u \ : \ (v, u) \in E(D) \text{ and } (v, u) \text{ has color } i\}.$$

Similarly, we define,

$$N^-(v, i) := \#\{u \ : \ (u, v) \in E(D) \text{ and } (u, v) \text{ has color } i\}.$$

Suppose

$$N^+(v, i) = N^-(v, i) = d/p \pm O(\sqrt{d/p \log(d)})$$

has been achieved (Chernoff). Define the digraph $D_i = (V, E_i)$ (for $1 \leq i \leq (p-1)$), where $(u, v) \in E_i$ if and only if $\chi(v) = \chi(u) + i \mod (p)$. Chapter 10.2 gives a representation of the D_i's and the coloring idea, exploiting the fact that p is a prime. Choosing p to be prime is integral to ensuring that each split has 'large enough' girth.

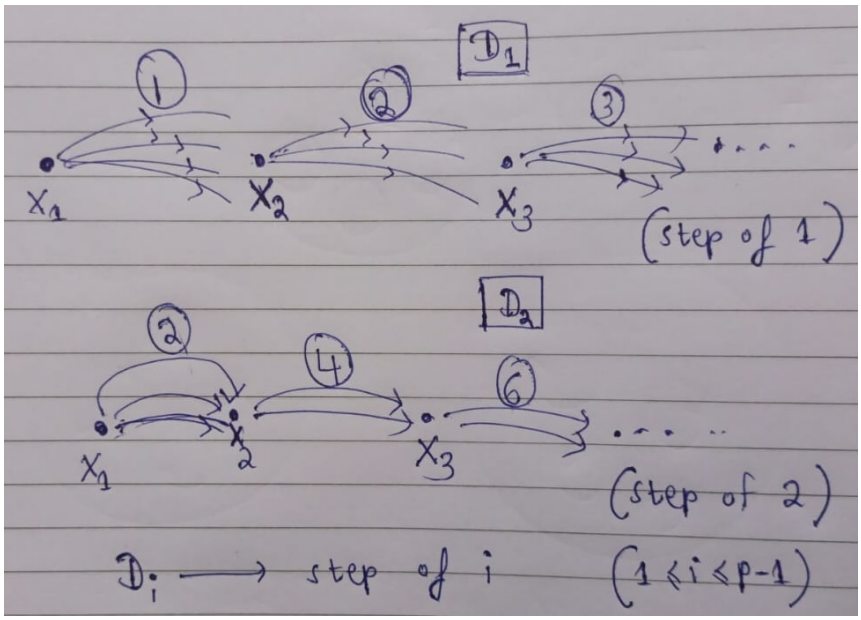


Figure 10.2: We define each equitable split D_i in such a way as to ensure $\text{girth}(D_i) > p$

Recap: If $\text{girth}(D) \geq 8ed$, then $\text{dla}(D) = d + 1$. Given D, we want to partition D into $D_0, D_1, \ldots, D_{p-1}$ such that

- girth(D_i) is large ($> 9e(d/p)$, where p is a parameter to be determined)
- $\Delta^+(D_i), \Delta^-(D_i) = (1 \pm o(1))d/p$.

In particular, each D_i can be partitioned into $(1 \pm o(1))d/p$ linear forests. So, D can be partitioned into $(1 \pm o(1))d/p.p = (d + o(d))$ linear forests.

Remark. *Note that any d-regular digraph has $dla(D) \leq 2d$.*

As before, construct T_D bipartite such that $(u, v') \in E(T_D)$ if and only if $(u, v) \in E(D)$. By Hall's theorem, the graph T_D can be partitioned into d matchings, where each matching of T_D gives rise to a disjoint union of cycles among the edges of D. We want to obtain D_i as an (almost) equitable split of every vertex. To get this, let p be a prime (sufficiently large?!) and let $\chi : V(D) \to \{0, 1, \ldots, p-1\}$ be a uniformly random map, i.e. $\chi(v) = i$ with probability $1/p$ for each i and independently for $v \in V(D)$.

Claim 91. *With + probability for each v, and each $i \in \{0, 1, \ldots, p-1\}$, if*

$$N^+(v, i) = \#\{u : (v, u) \in E(D) \text{ and } \chi(u) = i\}$$
$$\text{and } N^-(v, i) = \#\{u : (u, v) \in E(D) \text{ and } \chi(u) = i\},$$

then $N^+(v, i), N^-(v, i) = (1 \pm o(1))d/p$.

For $0 \leq i \leq (p-1)$, let D_i be those edges (u, v) such that $\chi(v) = \chi(u) + i$. Then note that for $1 \leq i \leq p-1$, girth$(D_i) \geq p$. So we want, $p^2 \geq 9ed$, which reduces to $p \geq \Omega(\sqrt{d})$. So, armed with these observations, presciently(!) pick a prime p with,

$$10\sqrt{d} < p < 20\sqrt{d}.$$

Assume the claim holds with error term $t = 10d^{1/4}(log(d))^{1/2}$. This gets us that for $1 \leq i \leq p-1$,

$$\text{dla}(D_i) \leq 1 + d/p + 10d^{1/4}log(d)^{1/2} \tag{10.2}$$

Summing Chapter 10.2 over $1 \leq i \leq p-1$,

$$\text{dla}\Big(\bigcup_{i=1}^{n} D_i\Big) \leq 20\sqrt{d} + d + O(d^{3/4}log(d)^{1/2}) \qquad (p < 20\sqrt{d})$$

Further, we have

$$\text{dla}(D_0) \leq 2(d/p + O(d^{1/4}log(d)^{1/2})) = O(\sqrt{d}).$$

This gives that, $\text{dla}(D) \leq d + O(d^{3/4}log(d)^{1/2})$. Note that $N^+(v, i) \sim \text{Bin}(d, 1/p)$ i.e. it is a sum of independent Bernoulli indicator variables. Fix v and i. We shall use the following version of the Chernoff bound,

Theorem 92 (Chernoff). *Let $X \sim Bin(n,p)$. For any $0 \leq t \leq np$,*

$$\mathbb{P}(|X - np| \geq t) \leq 2e^{-t^2/3np}$$
$$\Rightarrow \mathbb{P}(|N^+(v,i) - d/p| \geq t) \leq O\big(e^{-O(t^2/\sqrt{d})}\big)$$

So we may take $t = 10d^{1/4}(log(d))^{1/2}$ as the error term.

Remark. *Note that Chernoff works only for a single vertex equitable split. If we apply Chernoff for each vertex and then union over all, we bring n into the picture (not good!). Since we only want +ve probability, local lemma serves the purpose.*

So it suffices to prove Chapter 91. We will do this by the local lemma!

Proof of Chapter 91. Let $A^+(v,i)$ and $B^-(v,i)$ be the BAD events where,

$$A^+(v,i) : |N^+(v,i) - d/p| > 10d^{1/4}(log(d))^{1/2}$$
$$B^-(v,j) : |N^-(v,j) - d/p| > 10d^{1/4}(log(d))^{1/2}$$

By Chernoff, $\mathbb{P}(A^+(v,i))$, $\mathbb{P}(B^-(v,j)) \leq O(1/d^{10})$. To ensure indpendence of the bad events corresponding to vertices u and v, we want $\text{dist}(u,v) \geq 3$. It follows that the maximum dependence degree is $\leq O(d^{5/2}) = O(d^2 p)$. Local lemma applies with room to spare! ■

Alon and others[**?**] brought down the error term in the conjecture to $O(d^{2/3})$. Following is the best known result[**?**],

Theorem 93 (Ferber, Fox, Jain). *We can get a sharper bound $dla(D) \leq d + O(d^{2/3-\alpha})$, for some $\alpha > 0$.*

Remark: Alon's original proof of the asymptotic directed linear arboricity conjecture used the general version of the local lemma, where he proved that if $\mathcal{D}$ has 'large girth,' then the conjecture is true. In general, he asked: given a regular graph G of degree d, can we find $H \subset G$ of large relative girth? We want to prove something like this:

Conjecture 94. *There exists an $H \subset G$ such that for any $v \in V$,*

1. $f(d) - g(d) \leq d_H(v) \leq f(d) + g(d)$
2. $Girth(H) \geq h(d)$.

Proof. To show this, pick an edge of G to be in H independently with probability $\frac{f(d)}{d}$. Therefore, Chernoff gives

$$\mathbb{P}(|d_H(v) - f(d)| > g(d)) \leq 2\exp\left[\frac{-g(d)^2}{3f(d)}\right].$$

An optimal choice of $g(d)$ is approximately $\sqrt{Cf(d)\log g(d)}$ so that

$$\mathbb{P}(|d_H(v) - f(d)| > \sqrt{Cf(d)\log g(d)}) \leq \frac{1}{\sqrt{Cf(d)\log g(d)}}.$$

Suppose C is a cycle in G of size $k \leq n(d)$. Then,

$$\mathbb{P}(C \text{ is retained in } H) = \left(\frac{f(d)}{d}\right)^k$$

for $3 \leq k \leq n(d)$. Let $A_v := \{|d_H(v) - f(d)| > C)\}$, so that $\mathbb{P}(A_v) \leq \frac{1}{g(d)}$ and let $B_C := \{C \text{ is retained in } H\}$ so that $\mathbb{P}(B_c) = \left(\frac{f(d)}{d}\right)^k$. Moreover

$$\begin{aligned} A_v \leftrightarrow A_w \quad &\text{if and only if} \quad v \leftrightarrow w \\ A_v \leftrightarrow B_c \quad &\text{if and only if} \quad v \in C \\ B_c \leftrightarrow B_{c'} \quad &\text{if and only if} \quad E(c) \cap E(c') \neq \emptyset \end{aligned}$$

We now need to find the number of cycles of size k containing v.We use induction to see that the number of k cycles containing v is less than d^{k-1}. Similarly, for any edge $\hat{e}$, the number of cycles of length k containing $\hat{e}$ is less than d^{k-2}. Using the general form of the Lovász Local Lemma tells us that

$$\mathbb{P}(A_v) \leq \frac{1}{g(d)} \leq x(1-x)^d \prod_{k=3}^{h(d)}(1-y_k) \tag{10.3}$$

where $1-x$ corresponds to adjacent vertices and $1-y_k$ corresponds to adjacent cycles. We also obtain

$$\mathbb{P}_{|c|\geq k}(B_C) \leq \left(\frac{f(d)}{d}\right)^k \leq y_k(1-x)^k \prod_{l\geq 3}(1-y_l)^{kd^{l-2}}. \tag{10.4}$$

If there exists x and y_k such that (10.3) and (10.4) hold then the Lovász Local Lemma works. A nice start is to try $y_k = \frac{1}{d^{k-1}}$.

Alon actually proved the corresponding theorem with

$$h(d) = \frac{\log d}{2d\log\log d} \qquad f(d) = \log^{10} d \qquad g(d) = \log^6 d.$$

■

10.4 Another 'twist' to the Lovász Local Lemma

Erdös and Spencer proved the following result. Suppose A is an $n \times n$ matrix filled with integers such that each integer occurs at most $k = \frac{n-1}{4e}$ times. Then A admits a *latin transversal*.

Definition 95 (Latin Transversal)**.** *Let $A = [a_{ij}]$ be an $n \times n$ matrix whose entries are integers. A latin transversal is a permutation $\pi \in S_n$ such that the cells $\{a_{i\pi(i)} | i = 1, \ldots, n\}$ are all distinct integers.*

Furthermore, let $BAD = \{(c_1, c_2) | c_1, c_2$ are cells of A and are the same integer$\}$. This is simply the set of all pairs of coordinates which take the same value. Also, let $\mathcal{D}$ be a directed graph with maximum degree d. Let $V(\mathcal{D}) = BAD$ and let there be an edge between (c_1, c_2), and (c_1', c_2') if both of these pairs are in $V(\mathcal{D})$. Next say that $(c_1, c_2) \leftrightarrow (c_1', c_2')$ if and only if $(\{i_1, i_2\} \cap \{i_1', i_2'\}) \cup (\{j_1, j_2\} \cap \{j_1', j_2'\}) \neq \emptyset$. This condition says that two pairs of cells are adjacent if there is a common column or row. Thus, the dependence of degree of $(c_1, c_2) < 4nk$. Notice that this is not tight, and could be improved upon, but is sufficient for our purposes.

Proof. Pick a $\pi \in S_n$ at random. We want $\mathbb{P}(\wedge_{T \in BAD} \overline{A}_T) > 0$, where A_T is the event that the chosen permutation picks cells in T. Observe that the Lovász Local Lemma actually proves that if we have events $\{A_1, \ldots, A_n\}$ and a directed graph $\mathcal{D}$ with maximum degree d, such that $\mathbb{P}(A_i \text{ Var}ert \wedge_{j \in S,\, i \not\to j} \overline{A}_j) \leq p$. Then if $pe(d+1) \leq 1$ we have $\mathbb{P}(\wedge_{i=1}^n \overline{A}_i) > 0$.

Also, without loss of generality we can take $c_1 = (1, 1)$ and $c_2 = (2, 2)$ and consider $\mathbb{P}(A_{(c_1,c_2)} | \wedge_{T \in S} \overline{A}_T) \leq p$ where $S \subset ([3, n] \times [3, n]) \cap BAD$. We need $e\frac{1}{n-1}4k \leq 1$. In other words $k \leq \frac{n-1}{4e}$. Hence, it is enough the show $\mathbb{P}(A_{(c_1,c_2)} \text{ Var}ert \wedge_{T \in S} \overline{A}_T) \leq \frac{1}{n(n-1)}$ where S is fixed.

Call a permutation π eligible if it picks no bad pairs from S. Further, let

$$S_{12} = \{\pi | \pi \text{ is eligible, } \pi(1) = 1, \pi(2) = 2\}.$$

Therefore

$$\mathbb{P}\left(A_{(c_1,c_2)} | \bigwedge_{T \in \mathcal{S}} \overline{A}_T\right) = \frac{k!}{\#\text{ eligible sets}}$$

where $\mathcal{S}$ is the set of $S_{ij} = \{\pi | \pi \text{ is eligible}, \pi(1) = i, \pi(2) = j\}$. We know $|S_{12}|n(n-1)| \leq \sum_{i \neq j} |S_{ij}| = \#$ eligible sets. We also see that $|S_{12}| \leq |S_{ij}|$ for all $i \neq j$. This is one of those (rare!) cases where the Lovász Local Lemma works nicely in conditional probability. ■

10.5 2-point concentration of $\chi(G(n,p))$ for $p = n^{-1/2-\delta}$

The chromatic number of the random graph $G(n, p)$ has been the topic of serious interest for quite a well, and continues to be so. There have been two kids of results in this general direction - either determine the threshold function for the concentration of $G(n, p)$ or describe a 'small' interval where $\chi(G(n, p))$ concentrates. One of the most interesting results of the latter type is for sparse graphs, and here, it turns out that $\chi(G(n, p)$ concentrates on an interval of constant size. The sharpest result in this direction is the following theorem of Alon and Krivelevich, which shows that with high probability, the chromatic number concentrates on two consecutive integers for $p = n^{-1/2-\delta}$ for any $\delta > 0$.

We present a special case of this result when $\delta < 1/6$. In a sense this is the densest among the graphs of this sparse regime, but the essential idea here, surprisingly, goes via the local lemma.

Theorem 96 (Alon-Krivelevich, 1997)**.** *Suppose $\delta > 0$ and $\varepsilon > 0$. Then for $n \gg 0$, there exists $t = t(\delta, \varepsilon, n)$ such that*

$$\mathbb{P}[\chi(G(n,p)) \in \{t, t+1\}] \geq 1 - \varepsilon$$

for $p = n^{-1/2-\delta}$.

The proof of the theorem considers three different cases, viz., $0 < \delta < 1/6$, $1/6 < \delta < 3/8$ and $3/8 < \delta < 1$. We shall only see a proof of the first case, which is the most interesting among the three. This is because for lower values of δ, the graph is 'relatively' more dense, making the result more impressive.

Before getting to the proof, we take a short detour and recap the ideas from the Shamir-Spencer theorem ([**?**]) seen earlier.

Theorem 97 (Shamir-Spencer, 1987)**.** *Given $\varepsilon > 0$, there exists $u = u(\varepsilon, p)$ such that the following holds: Let $\alpha > 5/6$ and consider $G = G(n,p)$ with $p = n^{-\alpha}$, then*

$$\mathbb{P}[\chi(G) \in \{u, u+1, u+2, u+3\}] \geq 1 - \varepsilon.$$

The proof of the above was done in two steps:

1. Find the threshold t such that $\chi(G) \leq t$ with probability at least ε. Then, using the Azuma-Hoeffding inequality, show that there is a subset $X \subset V$ that is t-colorable and $R = V \setminus X$ has size at most $c\sqrt{n}$.

2. The remainder of the proof used the fact that for "small enough" values of p, the probability that for a set of $O(\sqrt{n})$ vertices, some subgraph of size i induces more than $3i/2$ edges is almost zero.

The first step of the proof follows the framework of Shamir-Spencer, but to prove this stronger result, we make use of the Local Lemma.

Lemma 98. *If $G = G(n,p)$ with $p = n^{-1/2-\delta}$ and if $r = \lceil \frac{1}{\delta} \rceil$, then for any $c > 0$, $i \leq c\sqrt{n}$ vertices induce fewer than ri edges a.a.s.*

Proof. Let $\mathcal{E}_{\text{small}}$ be the event that a collection of $i \leq c\sqrt{n}$ vertices induces less than ri edges. We want this to be a high probability event. So,

$$\begin{aligned}
\mathbb{P}(\overline{\mathcal{E}_{\text{small}}}) &\leq \sum_{i=r}^{c\sqrt{n}} \binom{n}{i} \binom{\binom{i}{2}}{ri} p^{ri} \\
&\leq \sum_{i=r}^{C\sqrt{n}} \left[O(1) \frac{n}{i} i^r n \left(-\frac{1}{2} - \delta \right) r \right]^i \\
&\leq \sum_{i=r}^{C\sqrt{n}} \left[O(1)^r n^{1+\frac{1}{2}(r-1)-\left(\frac{1}{2}+\delta\right)r} \right]^i \\
&= \sum_{i=r}^{C\sqrt{n}} \left[O(1)^r n^{\frac{1}{2}-\delta r} \right]^i = o(1).
\end{aligned}$$

■

We now introduce some definitions and results about *list-colorings*.

Definition 99. *A graph $G = (V, E)$ is said to be k-choosable if for every collection of lists $\{\ell(v) : |\ell(v)| = k\}$ of possible colors for each vertex $v \in V$, there is a proper coloring of the vertices of the graph G such that each vertex receives a color from the list associated with it.*

A coloring as in the above definition is called a *list coloring* and we denote the *list Chromatic number* of the graph by $\chi_\ell(G)$. One can directly observe that for a graph G with maximum degree Δ, $\chi_\ell(G) \leq \Delta + 1$. In fact, we can make a stronger statement.

Definition 100. *The degeneracy of a graph G is defined as:*

$$\text{degeneracy}(G) := \max_{H \subseteq_{induced} G} \delta(H)$$

and G is said to be d-degenerate if $\text{degeneracy}(G) \leq d$.

An equivalent definition of the above may be stated thus: *A graph G is said to be d-degenerate if every subgraph contains a vertex of degree at most d*'. The following is a familiar result:

Proposition 101. *If a graph is d-degenerate, then it is $(d+1)$-choosable.*

Lemma 102. *Any subgraph G' of G induced by $i \leq c\sqrt{n}$ vertices is $2r$-choosable a.a.s.*

Proof. From 98, we know that for any $c > 0$ and with $n >> 0$, the induced graph on any $i \leq c\sqrt{n}$ vertices of G has fewer than ri edges whp, and so, each subgraph of the induced graph G' on any i vertices almost surely is such that each subgraph of G' has a vertex of degree at most $2r - 1$, and thus G' is $2r$-choosable. ■

A sketch of the main idea is the following: From the initial stages of the Shamir-Spencer analysis, find $t = t(\delta, \varepsilon, n)$ such that all but $c\sqrt{n}$ vertices are t-colorable with probability at least $1 - \varepsilon$, and denote the uncolored set of size at most $c\sqrt{n}$ by R. Now, we want to find a set $U \supset R$, with $|U| \leq 2c\sqrt{n}$ such that each $v \in V \setminus U$ has fewer than $4r$ neighbours in U. If each subset of U has a large independent neighbourhood in $V \setminus U$, then we can color the graph efficiently.

Before getting into the above, we want to get an idea of how large t should be.

Proposition 103. *With high probability(asymptotically almost surely),* $\chi(G(n,p)) \geq \frac{np}{2\log n}$.

Proof. To show that the chromatic number is 'large', it suffices to show that there are no 'large' independent sets as if α denotes the size of the largest independent set in G, then

$$\chi(G) \geq \frac{n}{\alpha}.$$

So, we want to show that the probability that $\alpha(G) > \ell$ is small, where $\ell = (2\log n)/p$.

$$\begin{aligned}
\mathbb{P}[\alpha(G) > \ell] &\leq \binom{n}{l}(1-p)^{\binom{l}{2}} \\
&\leq \left[\left(\frac{en}{\ell}\right)e^{-p(\ell-1)/2}\right]^{\ell} \\
&= o(1) \quad \text{for } \ell = (2\log n)/p.
\end{aligned}$$

Thus, a.a.s $\chi(G(n,p)) \geq \frac{np}{2\log n}$. ■

Now, we get back to the main idea. From the first step of the Shamir-Spencer proof, find the set $R \subseteq V$ of size at most $c\sqrt{n}$ such that the induced graph on $V \setminus R$ is t-colorable. Now, we claim that we can find a set $U \supset R$ of size at most $2c\sqrt{n}$ such that each vertex in $V \setminus U$ has at most $4r$ neighbours in U. This set can be constructed as follows: Start from the set $U = R$. If there is a vertex $v \in V \setminus R$ with more than $4r$ neighbours in U, then update $U = U \cup \{v\}$. This process stops at with $|U| < 2c\sqrt{n}$, as otherwise, we have a vertex set U of size $i = 2c\sqrt{n}$ which contains at least $(c\sqrt{n})(4r) > ir$ induced edges, contradicting Lemma 98.

Now, as U has size $O(\sqrt{n})$ and induces very few edges, by Lemma 102, U is $2r$-choosable with high probability.

Let $U = \{u_1, \ldots, u_k\}$ and $N_i = N(u_i) \cap V(V \setminus U)$. Fix a t-coloring of $V \setminus U$, call it f. By the previous observations, $G[U]$ is $2r$-choosable. Therefore if we can make $2r$ colors "available" at each u_i, we can extend the coloring on $V \setminus U$ to U. Let us adjust all constants and state a reformulation.

A reformulation

Note that, WHP, $G(n,p)$ for $p = n^{-\frac{1}{2}-\delta}$ satisfies:

1. Any $i \leq c\sqrt{n}$ vertices induce $< ri$ edges.
2. Any vertex has degree $\leq 3np$. (This is because, probability of some vertex having degree $> 2np$ is $\leq n \cdot \binom{n}{3np} \cdot p^{3np} = o(1)$.)
3. $\chi(G) \geq t > \frac{np}{2\log n}$.
4. There exists a $U \subset V$, with $|U| \leq c\sqrt{n}$ such that every vertex in $V \setminus U$ has $\leq 4r$ neighbours in U, and $G[V \setminus U]$ is t-colorable.
5. There are $\leq f(n,p)$ paths of length 3 between any two distinct vertices of G.

Here $f(n,p)$ is a function that will be chosen later.

Proposition 104. *For n sufficiently large, any graph G with n vertices satisfying the above conditions is $(t+1)$-colorable.*

The idea is, for any $u_i \in U$, pick a set I_i from $[t]$ randomly, made of $2r$ randomly chosen elements (possibly with repetition). Let $J_i = \{x \in N_i \mid f(x) \in I_i\}$, and let I be the union of all J_i. We want to show that I is an independent set with positive probability. Indeed, then we will be done, as we can color I with a new $(t+1)^{st}$ color, and use the I_i to color the vertices of U using list coloring.

Rephrase in terms of an auxiliary graph. Let H be a graph with $V(H) = \sqcup_{i=1}^{k} W_i$ where each W_i is a copy of the corresponding N_i. For any edge in $G[V \setminus U]$, say between x, y, put an edge between each copy of x and each copy of y in H. The t-coloring on $G[V \setminus U]$ extends to a coloring on H: Indeed, color all copies with the same color as the original. Call this coloring f as well. It is enough to prove that I forms an independent set in H.

We now define the bad events for pairs. Suppose $e = w_1w_2$ where $w_1 \in W_i$ and $w_2 \in W_j$ ($i = j$ is possible). Then A_e is the event that $f(w_1) \in I_i$ and $f(w_2) \in I_j$.

$$\implies \mathbb{P}(A_e) \leq (\frac{2r}{t})^2 < \frac{(4r\log n)^2}{n^2p^2} = O(\frac{\log^2 n}{n^2p^2}).$$

What is the dependence degree? Note that A_e only depends on edges with one vertex in $W_i \cup W_j$. Also,

$$|W_i| = |N_i| \leq \deg(u_i) \leq 3np = o(\sqrt{n}),$$

so by condition (1) in the reformulation, number of edges in W_i is $< 3rnp = O(np)$. What about edges between W_i and W_l for $l \neq i$? Note that any vertex in $V \setminus U$ has at most $4r$ copies in H; in fact, number of copies of x in H is equal to the number of

neighbours of x in U, which is at most $4r$ by condition (4). Hence every edge has at most $16r^2$ copies in H. Further, any edge xy in the original graph where $x \in N_i$ and $y \in N_l$ corresponds to a path $u_i x y u_l$ of length 3 between u_i and u_l, of which there are at most $f(n,p)$. Adjusting for the blow-up factor, number of edges between W_i and W_l is at most $16r^2 f(n,p) = O(f(n,p))$. Since l can take $k \le c\sqrt{n} = O(\sqrt{n})$ different values, we finally get the dependence degree as $d = O(np) + O(\sqrt{n} f(n,p))$.

To use Lovasz Local Lemma, it is sufficient to prove that $\mathbb{P}(A_e) \cdot d = o(1)$, equivalent to $O(\frac{\log^2 n}{np}) + O(\frac{f(n,p)\log^2 n}{n^{\frac{3}{2}}p^2}) = o(1)$. Since the first term is obviously $o(1)$, it is enough to focus on the second term. We will prove:

Lemma 105. *For $0 < \delta < \frac{1}{6}$ and $p = n^{-\frac{1}{2}-\delta}$, WHP, $G(n,p)$ satisfies the property that there are $O(n^2p^3 \log n)$ paths of length 3 between any two distinct vertices.*

This will finish the problem because

$$\frac{f(n,p)\log^2 n}{n^{\frac{3}{2}}p^2} = O(n^{\frac{1}{2}} p \log n) = o(1).$$

So now we prove the lemma.

Proof. Let $c_0 = \lceil \frac{2}{\delta} \rceil$. First we show that, WHP, any two vertices are connected by $\le c_0$ paths of length 2. Indeed, fixing two vertices u, v, we have to choose c_0 middle vertices, so

$$\mathbb{P}(\text{more than } c_0 \text{ paths between } u,v) \le \binom{n}{c_0} p^{2c_0} \le (\frac{en\delta}{2})^{\frac{2}{\delta}+1} \cdot \frac{1}{n^{\frac{2}{\delta}+4}} = O(\frac{1}{n^3}).$$

So the probability of there existing a pair of vertices containing with more than c_0 paths of length two between them is $O(\frac{1}{n}) = o(1)$.

Similarly, fixing u, v, the probability that there are at least d_0 internally-disjoint paths between u, v is at most

$$\binom{n}{d_0}^2 \cdot d_0! \cdot p^{3d_0} \le (\frac{en}{d_0})^{2d_0} \cdot d_0^{d_0} \cdot p^{3d_0} = \frac{e^{2d_0}}{d_0^{d_0}} \cdot n^{(\frac{1}{2}-3\delta)d_0}.$$

Explanation of the first term: We choose d_0 second vertices and d_0 third vertices in the path between u and v, and we can match the second vertices to the third vertices in $d_0!$ ways. In the above expression, if we put $d_0 = n^2p^3 \log n = n^{\frac{1}{2}-3\delta} \log n$, we get the above probability is $\le (\frac{e^2}{\log n})^{d_0}$. Since $\delta < \frac{1}{6}$, the exponent of n in d_0 is positive, so the above probability is exponentially small; in particular it is $o(\frac{1}{n^2})$. Therefore by union bound, WHP, there are less than d_0 internally-disjoint paths of length 3 between any two vertices.

Therefore it is enough to find a constant α such that, WHP, for any vertices u, v, if there at least αd_0 paths of length 3 between u, v, then there are at least d_0 internally-disjoint paths between u, v. Indeed, WHP any two distinct vertices are joined by $\leq c_0$ paths of length 2. Hence any vertex $x \neq u, v$ lies in $\leq 2c_0$ paths of length 3 joining u, v. Construct an auxiliary graph H_0 with vertices being paths of length 3 between u, v, and two paths are adjacent only if they share an internal vertex. Then by the previous observation, the degree of each path in H_0 is at most $2 \cdot 2c_0 = 4c_0$ (since there are two internal vertices). This implies that H_0 is $(4c_0 + 1)$-colorable, and hence has an independent set of size at least $\frac{|V(H_0)|}{4c_0+1}$. Hence taking the constant $\alpha = 4c_0 + 1$ works, and we are done. ■

10.6 Graph connectivity codes

We now look at one instance where the general version of the Local lemma is used.

The notion of a *graph-code* is the following. Suppose we have a family $\mathcal{G}$ of graphs all of them defined on the same vertex set - say $[n]$, for simplicity. It is natural to think of these graphs as 0-1 incidence vectors of length $\binom{n}{2}$. A very natural coding theoretic perspective then considers the *sums* of these graph-vectors: the 'sum' of graphs G_1 and G_2, denoted $G_1 \oplus G_2$, with edge sets E_1, E_2, respectively, is the graph with edge set $E_1 \Delta E_2$, the symmetric difference of the sets E_1 and E_2.

This creates a natural extremal problem. Define the graph-code $\mathcal{F} = \mathcal{F}_{\mathcal{G}}$ to be a collection of graphs on $[n]$ such that for $G_1 \neq G_2$ in $\mathcal{F}$ the graph $G_1 \oplus G_2 \in \mathcal{G}$. The natural extremal problem considers the maximum size of a graph-code for 'natural' families $\mathcal{G}$. In fact this formulation covers many extremal problems about how large a family of graphs on the same vertex set could be, subject to certain properties being preserved.

We restrict ourselves here to notion called connectivity-codes. For a connected graph G on $[n]$, let $\mathcal{G}$ be the collection of all connected subgraphs of G, and we shall denote by $\mathcal{F}_{conn} := \mathcal{F}_{\mathcal{G}}$ the maximum sized graph-code for this collection $\mathcal{G}$. One easy observation is the following: If G has minimum degree d, then $|\mathcal{F}_{conn}| \leq 2^d$. Indeed, otherwise, there must exist distinct graphs G_1, G_2 such that the set of edges of G_i incident at a vertex v of minimum degree are identical, so in $G_1 \oplus G_2$ the vertex v is isolated. The same argument actually shows that $|\mathcal{F}_{conn}| \leq 2^{\lambda(G)}$ where $\lambda(G)$ denotes the edge-connectivity of G. The main result that we shall present in this section is a result of Alon that describes a fairly large family of d-regular graphs that attain this bound.

In order to show this bound, suppose G_n is d-regular and has some 'nice' properties (our ideas towards a proof - or rather, Alon's ideas - will reveal what they must be). One way to construct a family $\mathcal{F}$ of 2^d graphs on $[n]$, all of them being subgraphs of G is to turn towards a standard coding theoretic trick and try to construct a 'linear' graph-code.

One natural way to do that is to associate a distinct graph $G_{\mathbf{u}}$ for each $\mathbf{u} \in \mathbb{F}_2^d$ so that for any vectors $\mathbf{u} \neq \mathbf{v}$ the graph $G_{\mathbf{u}} \oplus G_{\mathbf{v}}$ is a connected subgraph of G. Our idea is to describe these graphs in such a way that $G_{\mathbf{u}} \oplus G_{\mathbf{v}} = G_{\mathbf{u}+\mathbf{v}}$, so that the graph-code is indeed linear.

A very natural way to attempt this is to first assign to each edge $e \in E = E(G)$, a vector $\mathbf{v}(e) \in \mathbb{F}_2^d$. If the edge set $E_{\mathbf{u}}$ of $G_{\mathbf{u}}$ is described in 'some linear fashion' (depending on the assignments of the vectors $\mathbf{v}(e)$ and the vector $\mathbf{u}$ then this description makes the code linear. A natural such description for $G_{\mathbf{u}}$ is given by

$$E_{\mathbf{u}} := \{e \in E : \langle \mathbf{v}(e), \mathbf{u} \rangle = 1\}$$

where $\langle \cdot \rangle$ is the usual bilinear form on $\mathbb{F}_2^d$.

Why is *this* the natural choice? Or rather, why not set $\langle \mathbf{v}(e), \mathbf{u} \rangle = 0$ as the condition to determine $\mathcal{E}_{\mathbf{u}}$?

11 The Entropy Method

Let F be a probability distribution over a finite set, i.e.

$$F \sim \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}$$

Then we define the *entropy* of F to be:

$$H(F) := \sum_{i=1}^{n} p_i \log_2 \left(\frac{1}{p_i} \right)$$

Note that the entropy of F depends only on the probability values, and not on the set over which it is defined.
We first state an easy proposition about entropy:

Lemma 106. *$H(F) \leq \log_2(n)$, and equality is achieved for the uniform distribution.*

Proof. Write $\varphi(x) := -\log_2(x)$. Then φ is convex. Consequently, by Jensen's inequality, we have $\mathbb{E}[\varphi(X)] \leq \varphi(\mathbb{E}[X])$. Now, let X be a random variable which takes the value $-\log_2(p_i)$ with probability p_i. Then

$$\mathbb{E}[\varphi(X)] = \sum_{i=1}^{n} p_i \cdot (-\log_2(p_i)) \leq -\log_2 \left(\sum_{i=1}^{n} p_i^2 \right)$$

By the RMS-AM inequality (using the fact that $\sum_{i=1}^{n} p_i = 1$),

$$\sum_{i=1}^{n} p_i^2 \geq \frac{1}{n}$$

Finally, since $x \mapsto -\log_2(x)$ is decreasing, we get that:

$$\sum_{i=1}^{n} p_i \cdot (-\log_2(p_i)) = H(F) \leq -\log_2(1/n) = \log_2(n)$$

as desired. Finally, equality is achieved only when all the p_i's are equal, i.e. the distribution is uniform. ∎

We now make a few more definitions:

Definition 107. *Let X, Y be jointly distributed random variables. Let the distribution of X be F_X, Y be F_Y, and the joint distribution be $F_{X,Y}$. Then we define:*

$$H(X) := H(F_X)$$

$$H(X, Y) := H(F_{X,Y})$$

$$H(X|Y = b) := H(F(X|Y = b))$$

$$H(X|Y) := \mathbb{E}_b[H(X|Y = b)]$$

We also state a few facts about the entropy function:

1. $H(X, Y) = H(X) + H(Y|X)$. The 'intuitive' explanation is that the information content of the joint random variable (X, Y) is equal to the information content of the random variable X, added to the information content of the variable Y, but conditioned on the fact that we already know what X is, i.e. X has been 'exposed'.

2. From the first point, we have $H(X, Y) \geq H(X)$.

3. $H(X|Y) \geq H(X|Y, Z)$: The 'intuitive' explanation for this is that there is more information remaining to be known about X given that we just know Y, as opposed to if we know both Y and Z.

We will now see some applications of the entropy method.

11.1 Shearer's Lemma

Lemma 108 (Shearer's Lemma)**.** *Let $S \subseteq \mathbb{N}^3$ be a finite set, i.e. $|S| < \infty$. Define the projection of S onto the XY plane as:*

$$S_{XY} := \{(x, y) : (x, y, z) \in S\}$$

Similarly, we define S_{YZ}, S_{ZX} as:

$$S_{YZ} := \{(y, z) : (x, y, z) \in S\}$$

$$S_{XZ} := \{(x, z) : (x, y, z) \in S\}$$

Then

$$|S| \leq \sqrt{|S_{XY}| \cdot |S_{YZ}| \cdot |S_{XZ}|}$$

Proof. Note that Shearer's lemma is equivalent to proving that:

$$2\log_2|S| \leq \log_2|S_{XY}| + \log_2|S_{YZ}| + \log_2|S_{XZ}|$$

Pick $(x, y, z) \in S$ uniformly at random. Then $H(X, Y, Z) = \log_2|S|$. On the other hand, we also have:

$$\begin{aligned}
H(X,Y,Z) &= H(X) + H(Y|X) + H(Z|X,Y) \\
H(X,Y) &= H(X) + H(Y|X) \\
H(Y,Z) &= H(Y) + H(Z|Y) \\
H(X,Z) &= H(X) + H(Z|X)
\end{aligned}$$

Thus,

$$H(X,Y) + H(Y,Z) + H(X,Z) - 2H(X,Y,Z)$$

$$= (H(Y) - H(Y|X)) + (H(Z|Y) - H(Z|X,Y)) + (H(Z|X) - H(Z|X,Y)) \geq 0$$

where all 3 terms are non-negative due to the fact that conditioning decreases entropy. ■

Similar techniques allow us to prove a more general version of Shearer's lemma. Before that, we set up some notation:
Suppose $X = (X_1, \ldots, X_n) \in S_1 \times \cdots \times S_n$. For any $I \subseteq [n]$, define:

$$X_I := (X_i)_{i \in I}$$

Lemma 109 (Generalized Shearer's Lemma)**.** *Suppose $\mathcal{F} \subseteq 2^{[n]}$ is a set system satisfying that each $i \in [n]$ is in $\geq k$ members of $\mathcal{F}$. Then*

$$kH(X_1, \ldots, X_n) \leq \sum_{I \in \mathcal{F}} H(X_I)$$

Proof. We will induct on k.
Thus, let $k = 1$. Write $\mathcal{F} = \{F_1, \ldots, F_m\}$. Now, if $F_1, \ldots, F_m$ are pairwise disjoint, then the above lemma is immediate: Indeed,

$$H(X_1, \ldots, X_n) = H(X_{F_1}) + H(X_{[n]\setminus F_1}|X_{F_1}) \leq H(X_{F_1}) + H(X_{[n]\setminus F_1}) \leq \cdots \leq H(X_{F_1}) + \cdots + H(X_{F_m})$$

If the sets aren't pairwise disjoint, then define:

$$F_1' := F_1, F_i' := F_i \setminus \bigcup_{j=1}^{i-1} F_j, i \geq 2$$

Since $F_i' \subseteq F_i$, we have $H(X_{F_i}) \geq H(X_{F_i'})$, and thus $\sum_{i=1}^n H(X_{F_i}) \geq \sum_{i=1}^n H(X_{F_i'}) \geq H(X_1, \ldots, X_n)$, where the last inequality follows since $F_1', \ldots, F_m'$ do indeed partition $[n]$.

For the induction step, we claim that it suffices to show $H(X_E) + H(X_F) \geq H(X_{E\cup F}) + H(X_{E\cap F})$ for any $E, F \in \mathcal{F}$. Indeed, if this holds, then we have

$$\sum_{I\in\mathcal{F}} H(X_I) \geq H(X_{F_1\cap F_2}) + H(X_{(F_1\cup F_2)\cap F_3}) + \cdots + H(X_{(F_1\cup\cdots\cup F_{m-1})\cap F_m}) + H\left(X_{\cup_{i=1}^m F_i}\right)$$

Since every element occurs in the F_i's at least $k \geq 1$ times, $\cup_{i=1}^m F_i = [n]$. Consequently, it suffices to show that

$$H(X_{F_1\cap F_2}) + H(X_{(F_1\cup F_2)\cap F_3}) + \cdots + H(X_{(F_1\cup\cdots\cup F_{m-1})\cap F_m}) \geq (k-1)H(X_1, \ldots, X_n)$$

But note that if we had $i \in F_{t_1}, \ldots, F_{t_\ell}$ (where $t_1 < t_2 < \cdots < t_\ell$), then we have

$$i \in \left(\cup_{j=1}^{t_2-1} F_j\right) \cap F_{t_2}, \ldots, \left(\cup_{j=1}^{t_\ell-1} F_j\right) \cap F_{t_\ell}$$

In other words, if i occurred in ℓ sets in $\mathcal{F}$, then it occurs in $\ell - 1$ sets in $\{F_1 \cap F_2, (F_1 \cup F_2) \cap F_3, \ldots, (F_1 \cup \cdots \cup F_{m-1}) \cap F_m\}$. Consequently, since i occurred in $\geq k$ sets in $\mathcal{F}$, i occurs in $\geq k-1$ sets in $\{F_1 \cap F_2, (F_1 \cup F_2) \cap F_3, \ldots, (F_1 \cup \cdots \cup F_{m-1}) \cap F_m\}$, and then we're done by the induction hypothesis.
Finally, to finish the proof, we prove the following:

Claim 110. *Suppose $X = (X_1, \cdots, X_n) \in S_1 \times \cdots \times S_n$, and $\mathcal{F} := 2^{[n]}$. Then , $\forall E, F \in \mathcal{F}$, we have: $H(X_E) + H(X_F) \geq H(X_{E\cap F}) + H(X_{E\cup F})$.*

Proof. We note the following inequatity due to conditioning:

$$H(X_{F\setminus E}|X_{E\setminus F}, X_{E\cap F}) \leq H(X_{F\setminus E}|X_{E\cap F}) \tag{11.1}$$

Now, we have:

$$\begin{aligned} H(X_{E\cup F}) &= H(X_{E\setminus F}, X_{E\cap F}, X_{F\setminus E}) \\ &= H(X_{E\cap F}) + H(X_{E\setminus F}|X_{E\cap F}) + H(X_{F\setminus E}|X_{E\setminus F}, X_{E\cap F}) \\ &\leq H(X_{E\cap F}) + H(X_{E\setminus F}|X_{E\cap F}) + H(X_{F\setminus E}|X_{E\cap F}), (\text{from eq (1)}) \\ &= H(X_{E\cap F}) + H(X_{E\setminus F}, X_{E\cap F}) - H(X_{E\cap F}) + H(X_{F\setminus E}, X_{E\cap F}) - H(X_{E\cap F}) \\ &= H(X_E) + H(X_F) - H(X_{E\cap F}) \end{aligned} \tag{11.2}$$

whew the last two equalities follow since $H(X, Y) = H(X) + H(Y|X)$, and $H(X_{A\cup X}) := H(X_A, X_B), \forall$
$A \cap B = \emptyset$. Thus, proof of our claim is complete. ■

■

The above lemma can be rephrased slightly differently, as stated in the corollary below. It immediately follows from generalized Shearer's lemma.

Corollary 111. *Suppose $\mathcal{F}$ is a set of vectors taking values in $S_1 \times \cdots \times S_n$. Also suppose $\mathcal{F} \subseteq 2^{[n]}$ such that each $i \in [n]$ is in $\geq k$ members of $\mathcal{F}$. If*

$$\mathcal{F}_I := \{f|_I : f \in \mathcal{F}\}$$

then

$$|\mathcal{F}| \leq \left(\prod_{F \in \mathcal{F}} |\mathcal{F}_F|\right)^{1/k}$$

11.2 Union-Closed Conjecture of Frankl

Let $\mathcal{F} \subseteq 2^{[n]}$ be a set family. We say that $\mathcal{F}$ is *union closed* if for $A, B \in \mathcal{F}$, we have $A \cup B \in \mathcal{F}$.
Frankl [**?**] conjectured in 1979 that if $\mathcal{F}$ was a union-closed family, then there was some $i \in \bigcup_{F \in \mathcal{F}} F$ such that i belonged to $\geq |\mathcal{F}|/2$ members of $\mathcal{F}$.

This conjecture remained open for a very long time until Gilmer [**?**] resolved it upto constant factors. More precisely, he proved the existence of $i \in \bigcup_{F \in \mathcal{F}} F$ such that i appeared in $\geq |\mathcal{F}|/100$ elements of $\mathcal{F}$. His argument was immediately refined by a variety of authors. We follow the presentation of Chase and Lovett [**?**] in our scribe. Chase and Lovett tightened Gilmer's arguments and improved the 0.01 factor to 0.38, which is captured in the following theorem (whose proof we'll now see).

Theorem 112 (Gilmer, Chase, Lovett)**.** *Let $\mathcal{F}$ be a union-closed family of subsets of $[n]$, ie $\mathcal{F} \subseteq 2^{[n]}$ such that $\forall A, B \in \mathcal{F}, A \cup B \in \mathcal{F}$. Then, there is some element of $[n]$ that occurs in atleast $0.38|\mathcal{F}|$ members of $\mathcal{F}$.*

Before we present their argument, we prove an analytic lemma.

Lemma 113. *Define $H(x) := -(x \log_2 x + (1-x)\log_2(1-x))$ for $x \in [0,1]$. Also define the function $f : [0,1]^2 \mapsto [0,1]$, where:*

$$f(x,y) := \begin{cases} \frac{H(xy)}{xH(y)+yH(x)}, & (x,y) \in (0,1)^2 \\ 1, & \text{otherwise} \end{cases}$$

Then f is minimized at (φ, φ), where $\varphi := (\sqrt{5}-1)/2$. Furthermore, $f(\varphi,\varphi) = 1/(2\varphi)$.

Proof Sketch of Lemma. By routine calculations, we can verify that, f is continuous in $[0,1]^2$, and $f(x,y) < 1, \forall (x,y) \in (0,1)^2$. Thus, minimum is attained in $(0,1)^2$. We define $g : (0,1) \to RED$ as $g(x) = \frac{H(x)}{x}$, then $f(x,y) = \frac{g(xy)}{g(x)+g(y)}$. We first show that, f is minimized on a diagonal point.

Assume that, f is minimized at some (x^*, y^*), and $\alpha = f(x^*, y^*)$. Define $G(x, y) = g(x, y) - \alpha(g(x) + g(y))$. Then, $G(x, y) \geq 0, \forall (x, y) \in (0, 1)^2$, and $G(x^*, y^*) = 0$. Thus (x^*, y^*) is a stationary point of G, thus $\nabla G(x^*, y^*) = 0$. Thus, we have:

$$0 = \frac{\partial G}{\partial x}(x^*, y^*) = g'(x^*y^*).y^* - \alpha g'(x^*) = \frac{\partial G}{\partial y}(x^*, y^*) = g'(x^*y^*).x^* - \alpha g'(y^*)$$

Defining $F(x) = xg'(x)$, we see that the above condition gives $F(x^*) = F(y^*)$. A simple calculation gives $F(x) = \frac{\log(1-x)}{x}$. Thus, F is strictly increasing and thus, $x^* = y^*$. Hence, the minima lies on the diagonal. Restricting to points of the form (x, x), we get that, $f(x, x) = \frac{H(x^2)}{2xH(x)}$. Using numerical simulations, it can be verified that, the minima occurs at (φ, φ), and $f(\varphi, \varphi) = \frac{1}{2\varphi}$. Thus, $f(x, y) \geq \frac{1}{2\varphi}, \forall (x, y) \in [0, 1]^2$. ■

Proof of Union-Closed Conjecture. Sample A, B uniformly, and independently, from $\mathcal{F}$. Then $H(A) = \log_2 |\mathcal{F}|$. Now, since $\mathcal{F}$ is union closed, $A \cup B \in \mathcal{F}$. Furthermore, $H(A \cup B) \leq H(A)$. Since $\mathcal{F}$ is union closed, $A \cup B$ is some distribution over $\mathcal{F}$, whose enrtropy is less than equal to entropy of uniform distribution ovewr $\mathcal{F}$, which is the entropy $H(A)$, (since A is sampled uniformly at random from $\mathcal{F}$).
Now from chain rule,

$$H(A \cup B) = H((A \cup B)_1, \ldots, (A \cup B)_n) = \sum_{i=1}^{n} H((A \cup B)_i | (A \cup B)_{<i})$$

where $(A \cup B)_i$ is the indicator random variable which denotes if $i \in A \cup B$. Also, from data processing inequality,

$$\sum_{i=1}^{n} H((A \cup B)_i | (A \cup B)_{<i}) \geq \sum_{i=1}^{n} H((A \cup B)_i | A_{<i}, B_{<i})$$

since revealing $A_{<i}, B_{<i}$ reveals more information than merely revealing $(A \cup B)_{<i}$. Now, fix some i, and for any $x, y \in \{0, 1\}^{i-1}$, define:

$$p(x) := \Pr(A_i = 0 | A_{<i} = x), q(y) := \Pr(B_i = 0 | B_{<i} = y)$$

Then observe that

$$H((A \cup B)_i | A_{<i} = x, B_{<i} = y) = H(p(x)q(y))$$

Indeed, the above equation follows from the fact that $(A \cup B)_i$ is a $\{0, 1\}$ valued Bernoulli random variable, and A, B are independent. From the Lemma 6, it follows that:

$$H(p(x)q(y)) \geq \frac{1}{2\varphi}(p(x)H(q(y)) + q(y)H(p(x)))$$

Taking expectations,

$$H((A \cup B)_i | A_{<i}, B_{<i}) \geq \frac{1}{2\varphi}\mathbb{E}_{x,y}\left[p(x)H(q(y)) + q(y)H(p(x))\right]$$

Since x, y are independent, the RHS expression equals

$$\mathbb{E}_x\left[p(x)\right] \cdot \mathbb{E}_y\left[H(q(y))\right] + \mathbb{E}_y\left[q(y)\right] \cdot \mathbb{E}_x\left[H(p(x))\right]$$

Now, suppose $p = \min_{i\in[n]} \Pr_{A\in\mathcal{F}}(A_i = 0)$, then our goal is to lower bound $1 - p$. Then, the above expressions imply that,

$$H((A \cup B)_i | A_{<i}, B_{<i}) \geq \frac{p}{2\varphi}(\mathbb{E}_{B_{<i}}[H(q(B_{<i})] + \mathbb{E}_{A_{<i}}[H(p(A_{<i}))])$$

The avove expression follows from the definition that $q(y) = \Pr(B_i = 0 | B_{<i} = y)$, and similarly for $p(x)$. Now, we note that, $\mathbb{E}_{B_{<i}}[H(q(B_{<i})] = H(B_i|B_{<i})$. Now, we once again have from theb definition of conditiopnal entropy that,

$$H(B_i|B_{<i}) = \mathbb{E}_y[H(B_i|B_{<i} = y)] = \sum_y \Pr(B_{<i} = y) H(B_i|B_{<i} = y) = \mathbb{E}_{B_{<i}}[H(q(B_{<i}))]$$

So, from the above expressions, we have that,

$$\sum_i H((A \cup B)_i | A_{<i}, B_{<i}) \geq \frac{p}{2\varphi}(\sum_i H(A_i|A_{<i}) + H(B_i|B_{<i})) = \frac{p}{2\varphi}(H(A) + H(B))$$

. Thus, we have,

$$\log_2 |\mathcal{F}| \geq H(A \cup B) \geq \frac{p}{\varphi} \log_2 |\mathcal{F}|$$

This implies that,

$$p \leq \varphi \implies \exists i \in [n] \text{ st } \Pr(A_i = 0) \leq \varphi \implies \exists i \in [n] \text{ st } \Pr(A_i = 1) \geq 1 - \varphi = \frac{3 - \sqrt{5}}{2} \approx 0.38$$

Thus, $\exists i \in [n]$, such that i occurs in atleast 0.38 fraction of the sets in the family $\mathcal{F}$. This completes Chase and Lovett's strengthened version of Gilmer's proof for the union closed conjecture. ■

11.3 A Theorem of Brégman on permanents

Theorem 114. *Suppose, $M_{n\times n}$ is a $\{0, 1\}$ valued matrix, and let the sum of ith row of the matrix be d_i. Then,* $\mathrm{per}(M) \leq \prod_{i=1}^{n} (d_i!)^{\frac{1}{d_i}}$, *where* $\mathrm{per}(M)$ *denotes the permanant of the matrix M.*

Proof. (Radhakrishnan) We discuss the proof of Brégman's theorem by Jaikumar Radhakrishnan [**?**] which uses the entropic method.
Note that, the binary matrix M is equivalent to the bipartite graph $G(U, V, E)$, where $|U| = |V| = n$, and $(u_i, v_j) \in E \iff M_{ij} = 1, \forall u_i \in U, v_j \in V$. Thus, $\mathrm{per}(M)$ is the number of perfect matchings in G. Let $\mathbb{M}$ be the set of perfect matchings in G. We

sample a $\sigma \in \mathbb{M}$ uniformly at random.

The entropy of the sampled perfect matching is given by $H(\sigma) = H(\sigma_1, \cdots, \sigma_n) = \log_2 |\mathbb{M}|$ (since σ has uniform distribution over $\mathbb{M}$). Thus, it's equivalent to show that, $H(\sigma) \leq \sum_{i=1}^{n} \frac{1}{d_i} \log(d_i!)$

Decomposing H as sum of conditional entropies gives

$$H(\sigma_1, \cdots, \sigma_n) = H(\sigma_1) + \sum_{i=2}^{n} H(\sigma_i | \sigma_1, \cdots, \sigma_{i-1}).$$

Suppose $\tau \in \mathcal{S}_n$ is a permutation of $[n]$. We process the u_i's according to the order of τ, and we shall "reveal" the matching mates of $u_1, u_2, \cdots, u_n$ in the order $\sigma(\tau(1)), \sigma(\tau(2)), \cdots, \sigma(\tau(n))$ to compute the entropy $H(\sigma_1, \cdots, \sigma_n)$ in terms of the successive conditional entropies and then average on τ. Since a good choice of τ is not clear, let us pick $\tau \in \mathcal{S}_n$ uniformly at random!

Thus, the entropy calculation becomes

$$H(\sigma) = H(\sigma(\tau(1)) + \sum_{k=2}^{n} H(\sigma(\tau(k)) | \sigma(\tau(1)), \cdots, \sigma(\tau(k-1))).$$

For the kth summand, if $\tau(k) = i$, then each summand is of the form

$$H(\sigma(\tau(k)) | \sigma(\tau(1)), \cdots, \sigma(\tau(k-1))) = H(\sigma(i) | \sigma(\tau(1)), \cdots, \sigma(\tau(k_i - 1))).$$

Now for any given random variables X, Y we have: $H(Y|X) = \sum_a H(Y_a) \Pr(X = a)$, where Y_a is the conditional distribution of Y given $X = a$. Moreover, if $|\text{supp}(Y_a)| \leq j$, then $H(Y_a) \leq \log_2 j$, since entropy is maximum for uniform distribution (here, $\text{supp}(X)$ denotes support of the distribution of X). This suggests that we work with the sets $E_j := \{a \in \text{supp}(X) : |\text{supp}(Y_a)| = j\}$. So, $H(Y|X) \leq \sum_j \Pr(x \in E_j) \log_2 j$. Applying this to our setup, we consider $Y = \sigma(i)$ and $X = (\sigma(\tau(1)), \cdots, \sigma(\tau(k_i - 1)))$. Define $R_i(\sigma, \tau) := \text{Nbr}(u_i) \setminus \{\sigma(\tau(1)), \cdots, \sigma(\tau(k_i - 1))\}$. Then,

$$H(\sigma(i) | \sigma(\tau(1)), \cdots, \sigma(\tau(k_i - 1))) \leq \sum_{j=1}^{d_i} \Pr(|R_i(\sigma, \tau)| = j) \log_2 j (\text{where } d_i \text{ is the degree of } u_i)$$

$$\implies H(\sigma) \leq \frac{1}{n!} \sum_{\tau \in \mathcal{S}_n} \sum_{i=1}^{n} \sum_{j=1}^{d_i} \Pr_{\sigma,\tau}(|R_i(\sigma, \tau)| = j) \log_2 j (\text{summing across } i, \text{ averaging across } \tau) \tag{11.3}$$

We now make the following crucial observation: for a fixed matching σ, $\Pr_\tau(|R_i(\sigma, \tau)| = j) = \frac{1}{d_i}, \forall j \in \{1, \cdots, d_i\}$. Note that, the observation is followed by the proof of Brégman's

theorem since from equation (3), we get:

$$\begin{aligned} H(\sigma) &\leq \frac{1}{n!}\sum_{\tau\in\mathcal{S}_n}\sum_{i=1}^{n}\sum_{j=1}^{d_i}\frac{1}{d_i}\log_2 j \\ &= \frac{1}{n!}\sum_{\tau\in\mathcal{S}_n}\sum_{i=1}^{n}\frac{1}{d_i}\log_2(d_j!) \\ &= \sum_{i=1}^{n}\frac{1}{d_i}\log_2(d_j!)\frac{1}{n!}\sum_{\tau\in\mathcal{S}_n}1 = \sum_{i=1}^{n}\frac{1}{d_i}\log_2(d_j!) \end{aligned} \tag{11.4}$$

which is exactly what we wanted to prove. Now, we provide proof of our observation to complete the proof.

Lemma 115. *For a fixed matching σ, $\Pr_\tau(|R_i(\sigma,\tau)| = j) = \frac{1}{d_i}, \forall j \in \{1,\cdots,d_i\}$*

Proof. The cardinality of $R_i(\sigma,\tau)$ depends on the permutation τ. Visibly, $|R_i(\sigma,\tau)| = j$ occurs according to the order of the set $\{\sigma^{-1}(v)\}_{v\in\mathrm{Nbr}(u_i)}$ under the permutation τ, where the nodes $\tau(1),\cdots,\tau(k_i-1)$ 'eat up' some of the neighbours of u_i, leaving exactly j choices. Since $\tau \in \mathcal{S}_n$ is distributed uniformly, the "rank" of u_i in the list of it's 2-hop neighbours being $d_i - j + 1$ is equidistributed over $j = 1, 2, \cdots d_i$, making the probability uniform and equal to $\frac{1}{d_i}$ for all $j \in [d_i]$. ■

■

11.4 Algorithmizing the Local Lemma: The Moser-Tárdos algorithm

A significant challenge associated with the Lovász Local Lemma (LLL) is its non-constructive nature. While the lemma asserts the existence of low-probability events, it is quite difficult to construct or identify these events. Moreover, due to their inherently low probability, randomly sampling these events is often an inefficient approach for their discovery. In the breakthrough work of Moser and later by Moser and Tárdos, the authors made the local lemma go algorithmic. We shall not prove the Algorithmic LLL in all its glory; Rather, we try to convince the reader that it simply works. The technique is dubbed the 'Entropy Compression Method'.

11.5 The k-SAT Problem

Let $x_1, x_2 \ldots x_n$ be boolean variables. A *literal* is a boolean variable x_i or it's negation $\neg x_i$. A k-CNF (Conjuctive Normal Form) Formula is an AND of clauses, each clauses being an OR of k-literals. Such a CNF formula f is said to be satisfiable if there exists an $a \in \{0,1\}^n$ such that $f(a) = 1$. We use the following notation for a k-CNF:

$$f(x_1, x_2 \ldots, x_n) = \bigwedge_{i=1}^{t} C_i$$

where each clause C_i is of the form

$$C_i = \bigvee_{j=1}^{k} Y_{i_j}$$

and

$$Y_{i_j} = x_{i_j} \text{ or } \neg x_{i_j}$$

We also define the *Support* of a clause C as the collection of variables whose literals are present in the clause. It is denoted by $\text{Supp}(C)$

Theorem 116 (Folklore)**.** *Given a CNF f with clauses $\{C_i\}_{i\in[t]}$ and suppose for each i,*

$$\text{Supp}(C_i) \cap \text{Supp}(C_j) \neq \emptyset$$

for at most 2^{k-2} clauses, then f is satisfiable.

Proof. This is a rather straightforward application of LLL. Consider a sequence of n coin tosses and each variable is assigned the a binary value on the basis of these tosses. If some clause C_k fails, then each of it's literal must be assigned the wrong value and therefore the probability that clause evaluates to 0 is $1/2^k$. Furthermore, if C_i and C_j have disjoint supports, the event of them failing are independent of each other. Therefore, the dependency graph as degree at most 2^{k-2}. Since

$$ep(d+1) = e \times \frac{2^{k-2}+1}{2^k} < 1$$

for $k \geq 4$ and therefore, by the symmetric form of LLL, none of the clauses fail with some positive probability and hence our CNF formula f evaluates to 1. ■

This theorem is 'non-constructive' in nature. It doesn't give us a clue on how to find such input in polynomial time.

Theorem 117 ([**?**])**.** *There exists a constant c such that given a k-CNF formula f with t clauses, none of which overlaps with more than $r = 2^{k-c}$ other clauses, one can find a satisfying assignment for f in expected time polynomial in t.*

In the original paper, the authors develop an algorithm that could handle a dependence degree of 2^{k-5}. However, for the sake of convenience, we show a bound of 2^{k-c} where $c > 0$ is some constant?

11.6 The Fix-It Algorithm

A priori, it's not obvious that this algorithm would even terminate.

Algorithm 1: Moser's Fix-It Algorithm

Input: k-CNF Formula $f = f(x_1, x_2, \ldots, x_n)$ with clauses $C_1, C_2, C_3 \ldots C_k$ on n variables with $d \leq 2^{k-c}$

Output: Satisfying assignment $s \in \{0,1\}^n$

$R \leftarrow$ An assignment of $(x_1, x_2 \ldots, x_n)$ from $\{0,1\}^n$, sampled u.a.r;
$\mathcal{S} \leftarrow$ Set of all clauses in f;
$\mathcal{T} \leftarrow$ Set of all unsatisfied clauses in f;
Function `FIX(`C`)`
    Update the values of the k literal present in C, u.a.r, and update the value of R;
    while $\exists D \in \mathcal{T}$ *such that* $\mathrm{Supp}(C) \cap \mathrm{Supp}(D) \neq \emptyset$ **do**
        `FIX(`D`)`
end
while $\exists C \in \mathcal{T}, C \neq \emptyset$ **do**
    `FIX(`C`);`
return R;

11.6.1 Correctness and Termination of the Fix-It Algorithm

It is not hard to see the following observation:

Observation 118. *If* `FIX(`C`)` *is called then it terminates with an assignment in which* C *and all clauses sharing a variable with* C *are satisfied.*

Armed with this observation, we prove that *if* the algorithm terminates, the resulting assignment must satisfy our k-CNF formula.

Lemma 119 (Correctness)**.** *A call to* `FIX` *that terminates cannot change some already satisfied clause to an unsatisfied clause.*

Informal Proof. Suppose that `FIX(`C`)` was called and after it's termination and some clause A switches from satisfied from unsatisfied. Then

- $\mathrm{Supp}(C) \cap \mathrm{Supp}(A) = \emptyset$ as otherwise by Observation 118, A must have ended up as satisfied. Therefore, any reassignments of literals in clause C wouldn't affect the literals in clause A.
- Hence, while `FIX(`C`)` is being called, clause A cannot have any change as otherwise $\mathrm{Supp}(C) \cap \mathrm{Supp}(A) \neq \emptyset$.

■

Therefore, subsequent applications of `FIX` would eventually lead to a satisfiable assignment and we obtain the following theorem,

Theorem 120 (Correctness)**.** *If the Fix-It Algorithm terminates, it must output a satisfiable assignment.*

11.6.2 Moser's Entropy Compression Argument

We only give a high level demonstration of the argument. At the start, we used n random bits, 1 for each boolean variable, to initialize our random assignment R. Now imagine that we start off with a huge reserve string, S, of uniformly sampled random bits. At any stage of algorithm, if we call `FIX`, we splice the first k bits of our reserve string S to randomly assign k bits to the clause that the algorithm is working with. Using this idea, fix a random string S of length $n + Tk$, where T denotes the number of time the algorithm runs, is large enough, and assume the algorithm uses the first n bits as the initial assignment A, and k bits each to replace the variables in each `FIX` call.

The random string S is used in Step 1 of `FIX(`C`)` to replace the values of variables in C by freshly sampled random values, and each time next k bits of S are used. If we know which clause is being fixed, we know the clause is violated so we know all the bits of this clause and thus we learn k bits of S (recall that assignments used by an algorithm are from the string S). We then replace those bits with another part of S. So we can describe S by the list of clauses we fix plus the remaining n bits of the final assignment. We can describe each clause C such that `FIX(`C`)` is called by the Fix-It algorithm using $\mathcal{O}(m \log m)$ bits.

If we can somehow encode the information about the neighbors of clause C in the dependency graph, say in $k-c$ bits, then having this information, we can recover our original string. However, randomly sampled string cannot be uniformly retrieved from a drastically shorter string. This argument requires some heuristics from Kolmogorov Complexity, but a rough idea is as follows: Pr[some string of length k can be represented using a string of lengt
$c] = 1/2^{c-1} - 1/2^k \ll 1$

Therefore, length of original string must be greater than the recovered string and hence

$$n + Tk \leq n + T(k - c + \mathcal{O}(n)) + \mathcal{O}(m \log m)$$

Therefore,

$$T = \mathcal{O}(m \log m)$$

11.7 A List-Coloring Generalisation of Thue's Theorem

A list is said to be non-repetitive if it doesn't contain any adjacent identical blocks. In 1906, Thue proved that we can create an arbitrarily long non-repetitive list with numbers $\{1, 2, 3\}$. We consider a List-Coloring Generalization of this theorem i.e.

Question 121. *Given n and collection of n lists $\{\mathcal{L}_i\}_{i\in[n]}$, does there exists a sequence $\tilde{s} = (s_1, s_2, \ldots, s_n)$, $s_i \in \mathcal{L}_i$ $\forall i \in [n]$ such that it contains* ***no Consecutive Repeating Blocks (CRBs)****?*

Before proceeding, we first define Consecutive Repeating Blocks

Definition 122. *A list $\tilde{s} = (s_1, s_2, \ldots, s_n)$ is said to contain a Consecutive Repeating Block if $\exists i \in [n]$ and $h \in \mathbb{N}$ such that:*

$$[s_{i-h}, s_{i-h+1} \ldots s_{i-1}] = [s_i, s_{i+1} \ldots s_{i+h-1}]$$

Theorem 123 ([?])**.** *Given $|\mathcal{L}_i| \geq 4 \; \forall i \in [n]$, there exists a list containing no Consecutive Repeating Blocks.*

11.7.1 The Algorithm

Consider the following randomized algorithm. The input is a sequence of lists $\{\mathcal{L}_i\}_{i\in[n]}$. Random elements are chosen independently with uniform distribution.

Algorithm 2: Choosing a non-repetitive sequence from lists of size 4

```
1 i ← 1; while i ≤ n do
2 |   s_i ← l_i, a randomly sampled element from L_i;
3 |   if s_1, s_2, ..., s_i is non-repetitive then
4 |   |   i ← i + 1;
5 |   else
6 |   |   There is only one repetition, say,
  |   |   [s_{i-2h}, s_{i-2h+1} ... s_{i-h-1}] = [s_{i-h}, s_{i-h+1} ... s_i] i ← i − h + 1;
```

The idea behind this algorithm is same as the one behind the Fix-It algorithm, if the algorithm works long enough for all evaluations of the random experiments, then a lot of repetitions occur, based on which we can compress a random string to a better extent than is actually possible

Proof of Theorem 123. Suppose for a contradiction that it is not possible to obtain a non-repetitive sequence from $\{\mathcal{L}_i\}_{i\in[n]}$. Set M to be a sufficiently large integer. We are going to record, in two different ways, the possible scenarios of what algorithm does in the first M steps.

Order the elements of $\mathcal{L}_i$ for each i randomly. Then, the algorithm picks a random element from a list of size 4. Let r_j denote the position of the j'th element in our sequence, with respect to the original list. Then, $r_1, r_2, \ldots, r_M$ is a sequence of random variables with 4^M possible values. Fix a random sequence of length M from one of 4^M possible values

Let $\tilde{d} = (d_1, d_2, \ldots d_M)$ be a running log of the algorithm where

$$d_i = \begin{cases} 1 & \text{if the length of string increases 1} \\ -(h-1) & \text{if a repeated block of length } h \text{ is found} \end{cases}$$

Let $\tilde{r}_i$ denote the tuple of first i elements of $\tilde{r}$ and $\tilde{d}_j$ denote the tuple of first j elements of $\tilde{d}$. $\tilde{s}_M = (s_1, s_2, \ldots s_l)$ denotes our resulting string after M steps.

Proposition 124. *$(\tilde{d}_M, s_M)$ uniquely recovers $\tilde{r}_M$*

It is enough to show that we can recover $(\tilde{d}_{M-1}, \tilde{s}_{M-1})$ from $(\tilde{d}_M, \tilde{s}_M)$.

- If $d_M = 1$, then, the last element of $\tilde{s}_M (= s_l)$ must be equal to r_M and therefore, $\tilde{s}_{M-1} = \tilde{s}_M \setminus r_M$
- If $d_M = -(h-1) \leq 0$, then, we must have had a repeating block of length h after the addition of r_M to s_M. Therefore, $\tilde{s}_{M-1} = \tilde{s}_M + [r_{M-h}, r_{M-h+1} \dots r_{M-1}]$

Now note that

- $\forall 1 \leq i \leq M, d_i \leq 1$
- $\forall 1 < k \leq M,\ \sum_{i=1}^{k} d_i \geq 1$ as the string can never have negative length.

The number of M-tuple such that (i) and (ii) holds, along with an additional condition that $\sum_{i=1}^{M} d_i = 1$ is known to be the very well known 'Catalan Opening'. Let D_M denote the total possible number of tuples $\tilde{d}_M$. Then, $D_M = \frac{1}{N}\binom{2N-2}{N-1} = o(4^M)$. The number of sequence satisfying conditions (i) and (ii) and $\sum_{i=1}^{M} d_i = k \leq n$ is certainly less than T_M. Note that since our sequence doesn't terminate after M steps, $d_M \leq n$. Therefore, the total possible number of tuple $\tilde{d}_M$ possible is $\leq n \cdot T_M$. For each instance of $\tilde{d}_M$ corresponds to at most 4^n non-repetitive strings but since the number of tuples $\tilde{r}_M$ is exactly 4^M, and each $\tilde{r}_M$ is uniquely determined by our logs,

$$4^M \leq 4^n \cdot n \cdot T_M \leq o(4^M)$$

which certainly doesn't hold for a large enough M. ■

12 More sophisticated concentration: Talagrand's Inequality

A relatively recent, extremely powerful, and by now well utilized technique in probabilistic methods, was discovered by Michel Talagrand and was published around 1996. Talagrand's inequality is an instance of what is refered to as the phenomenon of 'Concentration of Measure in Product Spaces' (his paper was titled almost exactly this). Roughly speaking, if we have several probability spaces, we many consider the product measure on the product space. Talagrand showed a very sharp concentration of measure phenomenon when the probability spaces were also metrics with some other properties. One of the main reasons this inequality is so powerful is its relatively wide applicability. In this chapter, we briefly study the inequality, and a couple of simple applications. Suppose $(\Omega_i, \mathbb{P}_i, \rho_i)$ are metric spaces, where ρ_i are metrics. We have, $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \rho_i, \prod_{i=1}^n \mathbb{P}_i)$ is the product (METRIC) probability space. Recall McDiarmid's inequality,

Theorem 125 (McDiarmid)**.** *Let $f : \prod_{i=1}^n \Omega_i \to \mathbb{R}$ be Lipschitz, i.e.*

$$|f(\underset{\sim}{x}) - f(\underset{\sim}{y})| \leq \rho(\underset{\sim}{x}, \underset{\sim}{y}), \text{for any } \underset{\sim}{x}, \underset{\sim}{y} \in \prod_{i=1}^n \Omega_i.$$

If $\underset{\sim}{x} \in \prod_{i=1}^n \Omega_i$ is picked according to $\prod_{i=1}^n \mathbb{P}_i$, and if f is bounded, then

$$\mathbb{P}(|f(x) - \mathbb{E}(f(x))| > t) \leq 2e^{-t^2/2n}.$$

Remark. *The above says that sufficiently smooth functions are heavily concentrated around their mean in these product spaces.*

Suppose $(\Omega_i, \mathbb{P}_i)$ are probability spaces, consider the product space $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mathbb{P}_i)$. A "natural" metric for this comes from the Hamming metric (counting coordinates where they differ),

$$d_H(\underset{\sim}{x}, \underset{\sim}{y}) = \#\{i : x_i \neq y_i\}.$$

This may not always be a smart choice! Suppose $\Omega_i = \{0, 1\}$, $\forall\ 1 \leq i \leq n$. Let $\Omega = \prod_i \Omega_i = \{0, 1\}^n$ and $\mathbb{A} = \{\underset{\sim}{x} \in \{0, 1\}^n : |\underset{\sim}{x}| \leq (n/2)\}$, where $|\underset{\sim}{x}| = \#\{i : x_i \neq 0\}$.

We have, $\mathbb{P}(\mathbb{A}) = 1/2$. Pick $\underset{\sim}{x}$ at random according to $\mathbb{P}$. Since the random variable $|\underset{\sim}{x}|$ is binomial, an application of Chernoff gives,

$$\mathbb{P}(|\underset{\sim}{x}| > n/2 + t) \leq e^{-t^2/2n}.$$

The above bound doesn't take into account that there might be a LOT of points with $|\underset{\sim}{x}| = (n/2 + t)$! We want a notion of distance which takes this information into account. This motivates the following,

Definition 126 (Talagrand convex distance)**.** *Given $(\Omega_i, \mathbb{P}_i)$, let $\mathbb{A} \subseteq \Omega = \prod_{i=1}^n \Omega_i$ and $\underset{\sim}{x} \in \Omega$. Let $r \in \mathbb{R}^n$ such that $r \geq 0$ and $\|r\|_2 = 1$. We define,*

$$\rho_0(\underset{\sim}{x}, \mathbb{A}) := \max_r \min_{y \in \mathbb{A}} \langle r, h(x, y) \rangle.$$

Here $r = (r_1, r_2, \ldots, r_n)$ is the 'cost' vector.

We define the set $A_t = \{\underset{\sim}{y} \in \Omega : \rho_0(\underset{\sim}{y}, \mathbb{A}) \leq t\}$, for $t > 0$. Following is the main version of Talagrand's inequality [**?**],

Theorem 127 (Talagrand (95))**.** *Let $\mathbb{P} = \prod_{i=1}^n \mathbb{P}_i$. We have,*

$$\mathbb{P}(\mathbb{A}) \cdot (1 - \mathbb{P}(A_t)) \leq e^{-t^2/4}.$$

12.0.1 A Combinatorialist's Version of Talagrand's inequality

For the purpose of applications, we will look at a different formulation of the Talagrand bound.

Definition 128. *A random variable $X : \Omega \to \mathbb{R}$ is f-certifiable (for a function f) if whenever $X \geq s$, then there exists,*

- *$X(w_1, w_2, \ldots, w_n) \geq s$*
- *$I \subseteq [n]$ with $|I| \leq f(s)$ s.t. for any w' with $w'_i = w_i \; \forall \, i \in I$, $X(w') \geq s$.*

Remark. *The notion of f-certifiability becomes weak if the function f attains large values.*

Following is the widely used combinatorialist version of Talagrand,

Theorem 129 (Talagrand)**.** *Let $\Omega = \prod_{i=1}^n \Omega_i$ and $\mathbb{P} = \prod_{i=1}^n \mathbb{P}_i$. If X is lipschitz and r-certifiable (i.e. $f(s) = rs$), then*

$$\mathbb{P}(|X - \mathbb{E}X| > t + 60\sqrt{r\mathbb{E}X}) \leq e^{-t^2/8r\mathbb{E}X}.$$

If the expectation $\mathbb{E}X$ is linear in n, the above bound is similar to McDiarmid. If it isn't linear in n, the above bound is better! As some summary of our discussion so far, we have the following remark,

Remark. *The shortcoming of Hamming is that it cannot tell if there are lots of points at same distance away from the set. Talagrand takes this account and averages it out, exploiting more information and hence giving a better bound.*

Fix $\mathbb{A} \subseteq \Omega$, $x \in \Omega$ and vector $r = (r_1, \ldots, r_n) \geq 0$. The Hamming difference vector (as seen earlier) is the binary vector $h(x, y) = (h_1, h_2, \ldots, h_n)$ such that $h_i = 1$ if $x_i \neq y_i$ and 0 otherwise. For the rest of this section, we denote $\|x\| = \|x\|_2$ in $\mathbb{R}^n$.

Definition 130. *Define the set,*

$$\mathcal{U}'_A(x) = \{h(x, y) \in \{0, 1\}^n \mid y \in A\}.$$

Recall Talagrand's notion of distance,

Definition 131 (Talagrand convex distance)**.**

$$\begin{aligned}\rho_r(x, A) &:= \min\{\langle r, h(x, y)\rangle : y \in A\} \\ \rho_0(x, A) &:= \max_r\{\rho_r(x, A) : r \geq 0,\ \|r\| = 1\}\end{aligned}$$

The following theorem gives an equivalent characterization for the Talagrand distance $\rho_0(x, A)$.

Theorem 132.

$$\begin{aligned}\rho_0(x, A) &= \min\{\|z\| \ : \ z \in CH(\mathcal{U}'_A(x))\} \\ &= \min\{\|z\| \ : \ z \in CH(\mathcal{U}_A(x))\}\end{aligned}$$

where $\mathcal{U}_A(x)$ denotes the UPSET[1] *generated by $\mathcal{U}'_A(x)$ and CH denotes the convex hull of a set.*

Proof. First we have the following claim,

Claim 133.

$$\rho_r(x, A) = \min\{\langle r, z\rangle \ : \ z \in CH(\mathcal{U}'_A(x))\} \tag{12.1}$$
$$= \min\{\langle r, z\rangle \ : \ z \in CH(\mathcal{U}_A(x))\}. \tag{12.2}$$

Proof. Suppose minimum of RHS in Chapter 12.1 is attained at z. Then $z = \sum_i \lambda_i h(x, y_i)$, for $y_i \in A$, $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. So $\langle r, z\rangle = \sum_i \lambda_i \langle r, h(x, y_i)\rangle$, which gives that $\langle r, z\rangle \geq \langle r, h(x, y_i)\rangle$ for some i. This gives Chapter 12.1. Similarly, suppose minimum of RHS in Chapter 12.2 is attained at $z_0 = \sum_i \lambda_i z_i$, where $z_i \in \mathcal{U}_A(x)$. We have,

$$\langle r, z_0\rangle = \sum_i \lambda_i \langle r, z_i\rangle \geq \sum_i \lambda_i \langle r, h(x, y_i)\rangle$$

where $\langle r, z_i\rangle \in \mathcal{U}_A(x)$ and $\langle r, h(x, y_i)\rangle \in \mathcal{U}'_A(x)$. We are done as before, proving the claim. ∎

[1] $\mathcal{U}_A(x) = \{(z_1, \ldots, z_n) : \exists (w_1, \ldots, w_n) \in \mathcal{U}'_A(x) \text{ s.t. } z_i \geq w_i \ \forall i\}$

We have the following claim which completes the proof of the theorem,

Claim 134.

$$\rho_0(x, A) = \max_r \{\rho_r(x, A) \ : \ r \geq 0, \|r\| = 1\} \tag{12.3}$$
$$= \min\{\|z\| \ : \ z \in CH(\mathcal{U}_A(x))\}. \tag{12.4}$$

Proof. Chapter 12.3 follows from the definition of $\rho_0(x, A)$. For $r \geq 0$, $\|r\| = 1$ and $z \in \mathrm{CH}(\mathcal{U}_A(x))$, we have

$$\langle r, z\rangle \leq \|z\|$$

from Cauchy-Schwarz. This gives $\rho_0(x, A) \leq \min\{\|z\| \ : \ z \in \mathrm{CH}(\mathcal{U}_A(x))\}$. Suppose the minimum in Chapter 12.4 is attained at $z_0 \in \mathrm{CH}(\mathcal{U}_A(x))$. Then for any $z \in \mathrm{CH}(\mathcal{U}_A(x))$ and $\forall\, 0 \leq \lambda \leq 1$, we have

$$\lambda z + (1 - \lambda)z_0 \in \mathrm{CH}(\mathcal{U}_A(x)).$$

Since $\|z_0\|$ is a minimum,

$$\|z_0\|^2 \leq \|\lambda z + (1 - \lambda)z_0\|^2.$$

Following this, set $p(\lambda) = \|\lambda z + (1 - \lambda)z_0\|^2$ which is a quadratic in λ. Differentiating and evaluating at the minimum of $p(\lambda)$ gives,

$$\|z_0\|^2 \leq \langle z, z_0\rangle.$$

Since $\rho_0(x, A) \leq z_0$, $z_0 = 0$ gives $\rho_0(x, A) = 0$ and the claim is proved. Therefore, WLOG let $z_0 \neq 0$. Set $r = z_0/\|z_0\|$. Since $\rho_r(x, A) = \min\{\langle r, z\rangle : z \in \mathrm{CH}(\mathcal{U}_A(x))\}$, we have

$$\frac{\langle z_0, z\rangle}{\|z_0\|} \geq \|z_0\|$$
$$\Rightarrow \rho_r(x, A) \geq \|z_0\|$$
$$\Rightarrow \rho_0(x, A) \geq \|z_0\|.$$

This completes the proof of the claim. ■

From the two claims above, proof of Chapter 132 follows. ■

12.0.2 Talagrand's inequality

For any $A \subseteq \Omega$ and $t \geq 0$, let $A_t = \{w \in \Omega : \rho_0(w, A) \leq t\}$. Recall Talagrand's inequality,

Theorem 135 (Talagrand)**.** *For A, A_t as above, we have*

$$\mathbb{P}(A).\mathbb{P}(\overline{A_t}) \leq e^{-t^2/4}.$$

Let $w \in \Omega$ and $f : \mathbb{R} \to \mathbb{R}^+$ be some function. We have the following definition,

Definition 136 (f-certifiability)**.** *A random variable X $(= X(w))$ is said to be f-certifiable if $X(w) \geq s \Rightarrow \exists I \subseteq [n]$ with $|I| \leq f(s)$ s.t. for ANY w' that agrees with w on I, $X(w') \geq s$ as well.*

The following is another version of the inequality,

Theorem 137. *If X is 1-lipschitz and f-certifiable, then for any b, we have*

$$\mathbb{P}\big(X < b - t\sqrt{f(b)}\big).\mathbb{P}\big(X \geq b\big) \leq e^{-t^2/4}.$$

Proof. Set $A = \{w : X(w) < b - t\sqrt{f(b)}\}$. We shall show that $\{w : X(w) \geq b\} \subseteq \overline{A_t}$. Then the conclusion will follow from Chapter 135. Suppose $X(w) \geq b$, we need to show that $w \notin A_t$. Suppose not *i.e.* $w \in A_t$, or equivalently $\rho_0(w, A) \leq t$. Since X is f-certifiable, $\exists I \subseteq [n]$ with $|I| = f(b)$ s.t. any w' agreeing with w on I must also have $X(w') \geq b$. Set

$$r = \frac{\mathbb{1}_{i \in I}}{\sqrt{|I|}}_{i=1\ldots n}.$$

By our assumption that $w \in A_t$, there exists $y \in A$ such that $\langle r, h(w, y)\rangle \leq t$. Then the number of coordinates (in I) on which y and w disagree is no more than $t\sqrt{|I|} \leq t\sqrt{f(b)}$. Now pick $z \in \Omega$ such that $z_i = y_i$ for all $i \notin I$ and $z_i = w_i$ for $i \in I$. Since z disagrees with y on no more than $t\sqrt{f(b)}$ coordinates and X is 1-lipschitz, we have $|X(z) - X(y)| \leq t\sqrt{f(b)}$. But since $y \in A$, we have $X(y) < b - t\sqrt{f(b)}$, so by the closeness of $X(y)$ and $X(z)$ we have $|X(z)| < b$. But since z agrees with w on the coordinates of I, f-certifiability guarantees that $X(z) \geq b$, and we have a contradiction. ■

Remark. *In particular, if $b = Med[X]$ in Chapter 137,*

$$\mathbb{P}\big(X < Med[X] - t\sqrt{f(Med[X])}\big) \leq O(e^{-t^2/4}) \text{ and}$$
$$\mathbb{P}\big(X > Med[X] + t\sqrt{f(Med[X])}\big) \leq O(e^{-t^2/4})$$

which essentially gives the concentration of the random variable X around its median. Note that $\mathbb{P}(X \leq Med[X]) = \mathbb{P}(X \geq Med[X]) = 1/2$.

We have the following Corollary to Chapter 137,

Corollary 138. *If X is Lipschitz and r-certifiable (i.e. $f(s) = rs$), then*

$$\mathbb{P}\big(|X - \mathbb{E}X| \geq t + 60\sqrt{r\mathbb{E}X}\big) \leq e^{-t^2/8r\mathbb{E}X}.$$

FACT: If X is r-certifiable and Lipschitz, then $|\mathbb{E}X - \text{Med}[X]| \leq O(\sqrt{\mathbb{E}X})$.

12.1 First examples

1. Non-isolated vertices in random graphs

Suppose G is a d-regular graph on n vertices. Let H be a random subgraph of G with each edge of G being retained in H with probability p. Let X denote the number of non-isolated vertices in H. By linearity of expectation,

$$\mathbb{E}[X] = \sum_{v \in V} \mathbb{P}[d_H(v) > 0] = n(1 - (1-p)^d).$$

The probability space in question is a product of the $nd/2$ binary probability spaces corresponding to retaining each edge, so that the events are tuples representing the outcomes for each edge. Changing the outcome of a single edge can isolate or un-isolate at most two vertices, so X is 2-Lipschitz. Furthermore, for any value of H with $X(H) \geq s$, we can choose one edge adjacent to each of s non-isolated vertices whose existence in another subgraph H' of G will ensure that the same s vertices are not isolated in H', i.e. $X(H') \geq s$. Thus X is also 1-certifiable, and Talagrand gives us

$$\mathbb{P}\left[|X - \mathbb{E}[X]| > (60+k)\sqrt{\mathbb{E}[X]}\right] \leq e^{-k^2/32}$$

so with high probability the number of non-isolated vertices is within an interval of length $O(\sqrt{\mathbb{E}[X]}) = O(\sqrt{n})$ about the mean. Compare this to the result using Azuma on the edge-exposure martingale, which would only give an interval of size $O\left(\sqrt{\binom{n}{2}}\right) = O(n)$ about the mean.

12.1.1 An Application: Longest Increasing Subsequences in random permutations

Suppose $\pi \in S_n$ is chosen at random, let $X(\pi)$ = length of a longest monotone subsequence in π. The following theorem gives a lower bound on the longest monotone subsequence in any sequence.

Theorem 139 (Erdós-Szekeres)**.** *Any real sequence of length* (n^2+1) *has a monotone subsequence of length* $\geq (n+1)$*.*

From Chapter 139, we have $X(\pi) \geq \sqrt{n-1}+1$. Also, we will show that $X(\pi) \leq 3\sqrt{n}$ holds WHP. Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{U}[0,1]$, this gives an uniform $\pi \in S_n$. Note that,

$$\mathbb{P}(X \geq k) \leq \binom{n}{k} \cdot \frac{1}{k!} \leq \frac{(en)^k}{k^k} \cdot \frac{2^k}{k^k} = \left(\frac{2en}{k^2}\right)^k.$$

Setting $k = 3\sqrt{n}$ gives an upper bound of $O((2e/9)^{\sqrt{n}})$, which tends to zero exponentially quickly with increasing n. Continuing, we also have $\mathbb{E}X$ is $O(\sqrt{n})$ since $\sqrt{n} \leq X \leq 3\sqrt{n}$ holds WHP. The random variable X is 1-Lipschitz as changing the position of any one

coordinate in the permutation π makes the length of the longest monotone subsequence go up or down by atmost 1. We also have that X is 1-certifiable, which follows from definition. Now we apply Talagrand bound on the random variable X. This will show that X lies in an interval of length $O(n^{1/4})$ around $\mathbb{E}X$, WHP. Set $t = Cn^{1/4}\sqrt{\log(n)}$. We have,

$$\mathbb{P}(|X - \mathbb{E}X| \geq t + 60\sqrt{r\mathbb{E}X}) \leq e^{-t^2/8r\mathbb{E}X}.$$

where $r = 1$. Since $\mathbb{E}X$ is $O(\sqrt{n})$, we have the result.

Remark. *Notice that McDiarmid is weak! Applying McDiarmid gives,*

$$\mathbb{P}(|X - \mathbb{E}X| > t) \leq e^{-t^2/2n}.$$

To ensure concentration, we are forced to choose $t >> \sqrt{n}$.

12.2 An Improvement of Brooks's Theorem

Let us recall Brook's Theorem: For a connected graph G is neither K_n nor C_{2k+1} then $\chi(G) \leq \Delta(G)$. But one might ask if this can be improved, especially if the maximum degree is rather large. Indeed, that has been the subject of quite some interest, and here, we note down two of the most well-known results (in chronological order):

- (J. H. Kim, 2002): For G with girth at least 5, $\chi(G) \leq (1 + o(1))\frac{D}{\log D}$.
- (Johansson, 2004): For G which is triangle free,, $\chi(G) \leq O(\frac{D}{\log D})$.

Both these results (especially Johansson's result) are quite non-trivial and involved, but we can get a slight improvement on Brooks's theorem for triangle free graphs, which is what we shall see now.

Theorem 140. *If G is triangle free and has maximim degree D, then $\chi(G) \leq (1 - \alpha)D$ for some $\alpha > 0$.*

Proof. Without loss of generality, let G be D-regular.
Scheme - We shall color the vertices uniformly at random from $[c]$. If two adjacent vertices are colored the same, uncolor both.
WTS - With positive probability, each vertex v has $\geq \alpha D + 1$ colors that are retained on ≥ 2 neighbors of v. If this is done, color each vertex greedily. The greedy algorithm will complete the proof.

Let A_v be the event that vertex v has $\leq \alpha D$ colors retained on ≥ 2 neighbors of v. $A_v \leftrightarrow A_w$ are dependent for $< D^4$ choices of w. Therefore, if $\mathbb{P}(A_v) = O(\frac{1}{D^5})$, then we are through.
Let X_v be the number of colors retained on ≥ 2 neighbors of v,

X'_v be the number of colors retained on exactly 2 neighbors of v, and
X''_v be the number of colors assigned on 2 neighbors of v and retained from the start. Note that $X_v \geq X'_v \geq X''_v$.
$\mathbb{E}(X''_v) \geq \binom{D}{2} \frac{1}{c}(1-\frac{1}{c})^{3D-3}$. If $u, w \in N(v)$ are assigned RED, then no vertext in V is assigned RED, where $V \biguplus (N(v) \setminus \{u,w\} \bigcup N(u) \bigcup N(w))$.
Now let $C = \beta D \implies \mathbb{E}(X''_v) \geq \frac{D(D-1)}{2} \frac{1}{\beta D}[(1-\frac{1}{\beta D})^{D-1}]^3 \geq \frac{D-1}{2} e^{-\frac{3}{\beta}}$, $D \gg 0$.

Let us note that X_v is 1-Lipschitz and certifiable for $X_v \geq s$.
Let us write $X''_v = Ass_v - Del_v$ where Ass_v is the number of colors assigned to 2 neighbors of v and Del_v is the number of colors asssigned to 2 neighbors but deleted from at least one of these two. We can see that Ass_v is 1-Lipschitz. If $Del_v \geq s$, then $\exists$ 2s vertices making color choices in pairs picking the same color and another $\leq$s neighbors of at least one of each of these pairs that witnesses G discoloration. Therefore, $Del_v \geq s$ and Del_v is s-certifiable.

Lets us recall the following inequalities:
If X is 1-Lipschitz and determined by independent trials $\{T_1, \ldots, T_m\}$, then $\mathbb{P}(|X - \mathbb{E}X| > t) \leq e^{-\frac{t^2}{2m}}$. If X is also r-certifiable, then Talagrand tells us that $\mathbb{P}(|X - \mathbb{E}X| > t + 60\sqrt{r\mathbb{E}X}) \leq e^{-\frac{t^2}{8r\mathbb{E}X}}$
This implies that for $t = C\sqrt{D \log D}$ (for a suitable constant C) we have

$$\mathbb{P}(|Ass_v - \mathbb{E}(Ass_v)| > t) \leq 2e^{-\frac{t^2}{D}} = 2e^{-\frac{C^2 \log D}{2}}.$$

Similarly,

$$\mathbb{P}(|Del_v - \mathbb{E}(Del_v)| > t + 60\sqrt{3\mathbb{E}(Del_v)}) \leq 2e^{-\frac{t^2}{24\mathbb{E}(Del_v)}}$$

so we simply take $\beta = \frac{1}{2}$and that gives $\alpha = 2e^{-6}$. ■

12.3 Almost Steiner Designs

In this section, we shall look now at a result due to Hod, Ferber, Krivelevich, and Sudakov [16], which achieves something very close to a Steiner design. Recall that a Steiner t-design with parameters (k, n) (and denoted $S(t, k, n)$) is a k-uniform hypergraph on n vertices such that every t-subset of the vertices is contained in exactly of the edges of the hypergraph. A simple counting argument shows that the number of edges of a Steiner t-design $S(t, k, n)$ is $\frac{\binom{n}{t}}{\binom{k}{t}}$.

We shall following [16] prove that, for n sufficiently large, there exists a k-uniform hypergraph such that every t-subset of the vertex set is in at least one edge, and at most 2 edges, and also, the number of edges is asymptotically close to the correct number. More precisely,

Theorem 141. *For n sufficiently large, and given fixed integers $k > t \geq 2$ there exist k-uniform hypergraphs $\mathcal{H}$ on the vertex set V satisfying*

- $e(\mathcal{H}) = (1 + o(1))\frac{\binom{n}{t}}{\binom{k}{t}}$.
- *Every t-subset of V is contained in at least one $E \in \mathcal{E}(\mathcal{H})$ and is contained in at most two edges.*

This rather neat looking theorem has a relatively short proof.

Proof. For starters, one might want to start with an almost tight packing $\mathcal{H}$ and then for each t-subset T that was not covered by the packing, we would like to pick another k-subset that accounts for covering T. This motivates the following

Definition 142. *For a k-uniform hypergraph $\mathcal{H}$ on $[n]$ the Leave hypergraph associated with $\mathcal{H}$ is the t-uniform hypergraph*
$\mathcal{L}_{\mathcal{H}} := \{T \subset [n] : |T| = t, T \not\subset E \text{ for any } E \in \mathcal{H}\}$.

Thus for every T in the Leave Hypergraph we wish to choose another k edge from the complete k-uniform hypergraph in order to cover every t-subset of $[n]$. In particular, one would like that the size of $\mathcal{L}_{\mathcal{H}}$ is small in comparison to the size of $\mathcal{H}$. This was already achieved by Grable; in fact he proved

Theorem 143. *(Grable, 1999) Let $k > t \geq 2$ be integers. There exists a constant $\varepsilon = \varepsilon(k, t) > 0$ such that for sufficiently large n there exists a partial Steiner design $\mathcal{H} = ([n], \mathcal{E})$ satisfying the following:*
For every $0 \leq l < t$ every set $S \subset [n]$ with $|S| = l$ is contained in $O(n^{t-l-\varepsilon})$ edges of the leave hypergraph $\mathcal{L}_{\mathcal{H}}$.

In particular, the size of $\mathcal{L}_{\mathcal{H}}$ is at most $O(n^{t-\varepsilon})$. But by picking one edge arbitrarily to cover each $T \in \mathcal{L}_{\mathcal{H}}$ we run the risk of having some t subset covered more than twice - something we do not want. Thus we need to be a bit *choosy* in picking edges to cover the edges of the leave hypergraph.

For each $A \in \mathcal{L}_{\mathcal{H}}$ define $\mathcal{T}_A := \{E : |C| = k, A \subset C\}$. Firstly, note that we can form a refinement of $\mathcal{T}_A$ as follows:

$$\mathcal{S}_A := \mathcal{T}_A \setminus \left(\bigcup_{B \in \mathcal{L}_{\mathcal{H}}, B \neq A} \mathcal{T}_B \right).$$

In other words, $\mathcal{S}_A$ consists of all $E \in \mathcal{T}_A$ such that no other t-subset (other than A) of the leave hypergraph is also in E. Suppose $B \in \mathcal{L}_{\mathcal{H}}$ and $|A \cap B| = i$. Then the number of sets $E \in \mathcal{T}_A$ that are **not** in $\mathcal{S}_A$ (on account of B) is $\binom{n-2t+i}{k-2t+i}$. Let

$$n_i(A) := \left|\{B \in \mathcal{L}_{\mathcal{H}} : B \neq A, |B \cap A| = i\}\right|.$$

If we fix $S = |A \cap B|$ is a subset of size i, it follows by the result of Grable that there are at most $O(n^{t-i-\varepsilon})$ distinct $B \in \mathcal{L}_{\mathcal{H}}$ such that $A \cap B = S$. Since there are $\binom{t}{i}$ choices for S, it follows that $n_i(A) \leq \binom{t}{i} n^{t-i-\varepsilon}$. Thus,

$$|\mathcal{S}_A| \geq \binom{n-t}{k-t} - \sum_{i=0}^{t-1} n_i(A) \binom{n-2t+i}{k-2t+i} = \Theta(n^{k-t}) - O(n^{k-t-\varepsilon}) = \Theta(n^{k-t}).$$

So, the sets $\mathcal{S}_A$ are all quite large.

Note also that by definition, the collections $\mathcal{S}_A$ are pairwise disjoint for different $A \in \mathcal{L}_{\mathcal{H}}$. Thus we have plenty of choice for picking $E \in \mathcal{S}_A$ for distinct $A \in \mathcal{L}_{\mathcal{H}}$. The standard probabilistic instinct is to now pick $E_A \in \mathcal{S}_A$ uniformly at random. But as we will see, it is not a good idea. Note that if E_A, E_B have been picked from $\mathcal{S}_A, \mathcal{S}_B$ respectively, and suppose $|E_A \cap E_B| \geq t$. Then it is highly likely that a t-subset of the intersection is contained in at least 3 of the sets, and that is not what we seek. So, it seems imperative that the choices E_A must satisfy $|E_A \cap E_B| < t$. In fact, if such choices can be made, then it is easy to see that the collection $\{E_A : A \in \mathcal{L}_{\mathcal{H}}\}$ along with $\mathcal{H}$ will serve as the 'almost' Steiner design we seek.

But this casts a new problem vis-á-vis the choice for E_A. Suppose $E_A \in \mathcal{S}_A$ has been chosen. Now if $B \in \mathcal{L}_{\mathcal{H}}$ satisfies $|A \cap B| = t - 1$, then the probability that $F = E_B$ will have $|E_A \cap F_B| < t$ is at most $\Theta\left(\frac{\binom{n-k-1}{k-t}}{\binom{n-t}{k-t}}\right) = \Theta_{k,t}(1)$. In particular, if there are several such B, then it is unlikely the choices for E_B made for all those will satisfy the intersection size criterion that we have zoomed into. Hence, we need an alternate idea about choosing the set E_A.

One of the interesting new perspectives of the probabilistic method that this proof suggests is the following principle:

Instead of picking exactly one set E_A, we can instead offer a small set of choices for E_A.

The heuristic idea here is that rather than one choice, if we are given a reasonable set $\mathcal{R}_A$ of choices for E_A for each $A \in \mathcal{L}_{\mathcal{H}}$ then if for each pair (A, B) in $\mathcal{L}_{\mathcal{H}}$ there are many compatible choices for (E_A, E_B) then is likely that we are able to choose $E_A \in \mathcal{R}_A$ so that all these choices are pairwise compatible. And to pick $\mathcal{R}_A$, we again resort to randomness. In other words, let us pick a random collection $\mathcal{R}_A \subset \mathcal{S}_A$ as follows. For each $E \in \mathcal{S}_A$, pick it as a member of $\mathcal{R}_A$ independently and with probability p (for some suitably small p).

Now, if for each A, we decide to make the pick $E_A \in \mathcal{R}_A$, we wish to show that $|E_A \cap E_B| < t$ for all $A \neq B$ in the leave hypergraph. Showing that $|E_A \cap E| < t$ for all

$E \in \mathcal{R}$ where

$$\mathcal{R} = \bigcup_{B \neq A, B \in \mathcal{L}_{\mathcal{H}}} \mathcal{R}_B$$

is more uniform, so let us aim to do that.

Fix $A \in \mathcal{L}_{\mathcal{H}}$, and suppose $\mathcal{R}_A$ has been determined but suppose $\mathcal{R}_B$ for the other sets of $\mathcal{L}_{\mathcal{H}}$ are not yet made. Knowing $\mathcal{R}_B$ for all $B \in \mathcal{L}_{\mathcal{H}} \setminus \{A\}$ amounts to independent trials made by the members of

$$\mathcal{S} = \bigcup_{B \neq A, B \in \mathcal{L}_{\mathcal{H}}} \mathcal{S}_B.$$

To say that we can make a choice $E_A \in \mathcal{R}_A$, we need good bounds on how many elements of $\mathcal{R}_A$ are poor choices, i.e., we need an estimate on

$$\mathfrak{N}_A := \big|\{E \in \mathcal{R}_A : |E \cap F| \geq t \text{ for some } F \in \mathcal{R}\}\big|.$$

Note that if we assume that $\mathcal{R}_A$ has already been chosen, then $\mathfrak{N}_A$ is determined by the outcome of $|\mathcal{S}|$ independent Bernoulli trials. Moreover, it is clear from the definition that $\mathfrak{N}_A$ is 1-certifiable. Indeed, if $\mathfrak{N}_A \geq s$, then there are $E_1, E_2, \ldots, E_s \in \mathcal{R}_A$ and at most s sets $F_1, F_2, \ldots, F_s \in \mathcal{S}$ such that $|E_i \cap F_i| \geq t$. In order to obtain good concentration, it would help if $\mathfrak{N}_A$ were also Lipschitz.

But unfortunately, that may not be the case. Suppose $B \in \mathcal{L}_{\mathcal{H}}$ and $|A \cap B| = t - 1$. Then for any $F \in \mathcal{R}_B$ and $E \in \mathcal{R}_A$, we would have $|E \cap F| \geq t - 1$, so the only way the intersection has size strictly less than t is if these sets are disjoint. Thus, it is conceivable that a single trial $F \in \mathcal{S}_B$ can affect $\mathfrak{N}_A$ substantially.

But now, we use an old trick of Bollobas, which 'Lipschitzises' this random variable, i.e., considers another related random variable which is Lipschitz, and in addition is very close to the random variable in question.

More precisely, suppose for each A, we pick a large enough sub collection $\mathcal{Q}_A \subset \mathcal{R}_A$ by adding an element of $\mathcal{R}_A$ into $\mathcal{Q}_A$ as long as it does not intersect any of the members already picked outside of A. Thus, $\mathcal{Q}_A$ is a subfamily of $\mathcal{R}_A$ in which any two sets are pairwise disjoint outside of A itself. If $\mathcal{R}_A$ is large enough, then perhaps one can imagine obtaining a large enough $\mathcal{Q}_A \subset \mathcal{R}_A$ by this process.

If we set

$$\mathfrak{N}_{\mathcal{Q}}(A) := \big|\{E \in \mathcal{Q}_A : |E \cap F| \geq t \text{ for some } F \in \mathcal{R}\}\big|$$

then note that the same argument for $\mathfrak{N}_A$ also works here, so $\mathfrak{N}_{\mathcal{Q}}(A)$ is 1-certifiable. But now, this is also Lipschitz. Indeed, if a certain choice $F \in \mathcal{R}$ is altered, then since the sets in $\mathcal{Q}_A$ are pairwise disjoint outside of A, it follows that $\mathfrak{N}_{\mathcal{Q}}(A)$ changes by at most $k - t$, so $\mathfrak{N}_{\mathcal{Q}}(A)$ is $k - t$-Lipschitz. Hence by Talagrand, we have

$$\mathbb{P}(\mathfrak{N}_{\mathcal{Q}}(A) > t) < 2e^{-t/16k^2} \text{ where } t \geq 2\mathbb{E}(\mathfrak{N}_{\mathcal{Q}}(A)) + 80k\sqrt{\mathbb{E}(\mathfrak{N}_{\mathcal{Q}}(A))}.$$

Let us estimate $\mathbb{E}(\mathfrak{N}_{\mathcal{Q}}(A))$ first. Note that (recall that we are assuming that $\mathcal{R}_A$, and $\mathcal{Q}_A$ are fixed)

$$\mathfrak{N}_{\mathcal{Q}}(A) = \sum_{E \in \mathcal{Q}_A} \mathbb{1}_E$$

where $\mathbb{1}_E$ counts the set E if there exists $F \in \mathcal{S}$ such that $|E \cap F| \geq t$. Let us first fix $E \in \mathcal{Q}_A$. Write

$$\mathcal{L}_{\mathcal{H}} \setminus \{A\} = \bigcup_{l=0}^{t-1} \mathcal{B}_l$$

where

$$\mathcal{B}_l := \{B \in \mathcal{L}_{\mathcal{H}} : |B \cap E| = l\}.$$

We wish to count the number of $F \in \mathcal{S}$ that trigger E and count in among $\mathfrak{N}_{\mathcal{Q}}(A)$.

If $B \in \mathcal{B}_l$ we have

$$\begin{aligned} \big|\{F \in \mathcal{S}_B : |E \cap F| \geq t\}\big| &\leq \big|\{F \in \mathcal{T}_B : |E \cap F| \geq t\}\big| \\ &= \big|\{F \in \mathcal{T}_B : |(E \cap F) \setminus B| \geq t - l\}\big| \end{aligned}$$

Consequently,

$$\big|\{F \in \mathcal{S} : B \subset F \text{ for some } B \in \mathcal{B}_l, |E \cap F| \geq t\}\big| \leq \sum_{i=t-l}^{k-t} \binom{k-l}{i} \binom{n-k-t+l}{k-t-i} = O(n^{k-2t+l}).$$

Indeed, pick a subset of $E \setminus B$ of size i, where $t - l \leq i \leq k - t$, then to get a choice for $F \in \mathcal{S}_B$, we need to pick the remaining $k - (t + i)$ elements from the set $[n] \setminus (E \cup B)$. Now, for fixed l with $0 \leq l \leq t-1$, we have $|\mathcal{B}_l| \leq \binom{k}{l} O(n^{t-l-\varepsilon}) = O(n^{t-l-\varepsilon})$. This is seen by first fixing a set of E of size l and then by the result of Grable stated earlier, there are at most $O(n^{k-l-\varepsilon})$ elements $B \in \mathcal{L}_{\mathcal{H}}$ that contains a set of size l. Hence, by a very generous amount, we have

$$\mathbb{E}(\mathbb{1}_E) = \mathbb{P}(E \text{ leads to increment of } \mathfrak{N}_{\mathcal{Q}}(A)) \leq pO(n^{k-2t+l})O(n^{t-l-\varepsilon}) = pO(n^{k-t-\varepsilon})$$

so

$$\mathbb{E}(\mathfrak{N}_{\mathcal{Q}}(A)) \leq |\mathcal{Q}_A| pO(n^{k-t-\varepsilon}).$$

Now suppose we had $p = n^{t-k+\varepsilon/2}$; then the estimate above gives us that

$$\mathbb{E}(\mathfrak{N}_{\mathcal{Q}}(A)) \leq |\mathcal{Q}_A| O(n^{-\varepsilon/2}).$$

Note that for this value of p we have with high probability $|\mathcal{R}_A| \approx \Theta(n^{\varepsilon/2})$ **for all** A (standard Chernoff bounds). We shall now argue that the greedy process produces $|\mathcal{Q}_A| \geq (n^{\varepsilon/3})$ for all A with high probability. We can then choose to stop at around this stage while constructing $\mathcal{Q}_A$, so that we indeed do have $|\mathcal{Q}_A| = \Theta(n^{\varepsilon/3})$ This completes the proof.

Suppose that the greedy process stops after m steps, with $m < n^{\varepsilon/3}$. Then there exist sets $E_1, E_2, \ldots, E_m$ such that every set in $\mathcal{S}_A$ that is 'disjoint' from $\bigcup E_i$ (i.e., disjoint outside of A) is not picked into $\mathcal{R}_A$. Now, if we set $X = \bigcup E_i$ then $|X| < kn^{\varepsilon/3}$. We now need to ensure that the number of sets of $\mathcal{S}_A$ that do not intersect X outside of A is of the right order. In other words, the number of sets of $\mathcal{T}_A$ that meet X non-trivially is at most

$$\sum_{i=1}^{k-t} \binom{|X|-t}{i}\binom{n-|X|}{k-t-i} = \sum_{i=1}^{k-t} \Theta(n^{i\varepsilon/3})n^{k-t-i} = o(n^{k-t})$$

which implies that the number of sets in $\mathcal{S}_A$ that are disjoint from X is $M = \Theta(n^{k-t})$. Thus, the probability that there exists some set X of size at most $kn^{\varepsilon/3}$ that satisfies this condition above is at most

$$\binom{n}{kn^{\varepsilon/3}}(1-p)^M < O(n^{n^{\varepsilon/3}})\exp(-n^{t-k+\varepsilon/2}\Theta(n^{k-t})) = \exp(n^{\varepsilon/3}\log n - \Theta(n^{\varepsilon/2})) < \exp(-n^{\varepsilon/7})$$

for n sufficiently large, so the result follows. ■

12.4 Chromatic number of graph powers

Recall, for $k \geq 1$, the k^{th} **Graph Power** G^k is defined as follows:

- $V(G^k) = V(G)$.
- For $u \neq v$, $u \leftrightarrow v$ iff $dist(u,v)_G \leq k$.

In other words, two vertices are adjacent in G^k if they are at most a distance k apart in G. Let $\Delta(G) = d$. One would like bounds on $\chi(G^k)$. The greedy algorithm tells us $\chi(G^k) \leq d^k + 1$.

Johansson improved Brooks' theorem for triangle free graphs by showing that $\chi(G) = O(\frac{\Delta}{\log \Delta})$. The following theorem below is a generalization of this extending to graphs where the neighborhood of any vertex is sparse.

Theorem 144. *(Alon-Krivelevich-Sudakov, 2002): If G has at most $\frac{d^2}{t}$ edges in the induced subgraph on $N(v)$ for each $v \in V(G)$ then $\chi(G) \leq \frac{d}{\log(t)}$.*

This implies (follows easily) that for G with girth at least $3k+1$, $\chi(G^k) \leq O\left(\frac{d^k}{\log d}\right)$.

In particular one is interested to see if the above result is asymptotically best possible. The following result of Alon and Mohar settles this in the affirmative.

Theorem 145. *(Alon-Mohar 2001): For large d and any fixed $g \geq 3$ there exist graphs with max degree $\Delta \leq d$, girth at least g, and $\chi(G^k) \geq \Omega\left(\frac{d^k}{\log d}\right)$.*

Proof: First, we shall bound Δ and Γ. We want to pick $G = G_{n,p}$ such that for all $v \in V(G)$, $\mathbb{E}[deg(v)] = (n-1)p < np$. Let $p = \frac{d}{2n}$. Because this process is a binomial distribution, we can bound the number of vertices with degree at least d using Chernoff.

$$\mathbb{P}[deg(v) \geq d] < \mathbb{P}[(deg(v) - \mathbb{E}(d(v)) > \frac{d}{2})] \leq e^{\frac{-(d/2)^2}{3(d/2)}} = e^{-d/6}$$

Now, let $N_{bad} = |\{v \in V | deg(v) > d\}| \Longrightarrow$

$$\mathbb{E}[N_{bad}] < ne^{-d/6}$$

By the Markov inequality

$$\mathbb{P}[N_{bad} > 10ne^{-d/6}] < .1$$

Similarly, let $N_{<g} = |\{C_k \subseteq G | k < g\}| \Longrightarrow$

$$\mathbb{E}[N_{<g}] = \sum_{i=3}^{g-1} \frac{(n)_i}{2i} \left(\frac{d}{2n}\right)^i < d^g$$

Again, Markov tells us that

$$\mathbb{P}[N_{<g} > 10d^g] < .1$$

This implies that with probability at least .8, G satisfies $N_{bad} \leq 10ne^{-d/6}$ and $N_{<g} \leq 10d^g$. We shall assume $n \gg d^g + ne^{-d/6}$ so that we can remove an arbitrary vertex from all small cycles and remove all vertices of degree more than d. If we want to ensure $\Delta = d$, it is simple enough to add some cycles of length g. Thus in order to get a condition on the maximum degree and girth, all we need to do is delete a small number of vertices from such a G.

To complete the proof we wish to show that a maximum independent set is not too large. More precisely, we wish to show that $\alpha(G) = O(\frac{n \log d}{d^k}$. This amounts to saying that whp, every set U of this size is NOT independent in G^k.

IN order to achieve this, what we shall do is this. If we could show that for any such set U, there are several paths of length $k$between some two vertices u, v in U, then in order to make the pair $\{u, v\}$ a non-edge in G^k, we should have deleted a vertex from each of those paths between u, v. But if the number of such paths is way more, then u, v is an edge in G^k giving us what we want. But showing that the number of paths is concentrated is a difficult task, so we shall try to show that there are several internally

disjoint paths between two such vertices. This is again another instance of the same trick that was mentioned in the previous section.

Let us get to the details. Let the path P be a U-path if the end vertices of P lie in U and the internal vertices lie outside of U.Set $U \subseteq V(G)$ such that

$$|U| = \frac{c_k n \log(d)}{d^k} = x$$

Now, to show $\chi(G^k) \geq \Omega\left(\frac{d^k}{\log(d)}\right)$, we will show that $\alpha(G^k) \leq c_k \frac{n \log(d)}{d^k}$ for some c_k (as outlined above) To do this, we will show that with high probability, for every U, $\Pi(G)$, the number of internally disjoint U-paths of length k, is large. Specifically, we will show that there are still many of these paths after we make vertex deletions for girth and maximum degree considerations. This will bound independent sets in G^k.
Let μ be the number of U-paths of length k. It is easy to show that

$$\mathbb{E}[\mu] = \binom{x}{2}(n-X)_{k-1}p^k > \frac{c_k^2 n^2 \log^2(d)}{2d^{2k}} \frac{n^{k-1}}{2} \frac{d^k}{2^k n^k} = \frac{c_k^2 n \log^2(d)}{2^{k+2} d^k}$$

Now, we need to say that $\mathbb{E}[\nu]$, the expected number of non-internally disjoint U-paths, is much smaller than $\mathbb{E}[\mu]$. For $n \gg d \gg k$, the expected number of U-paths which share one endpoint and the unique neighbor is at most

$$\mu n^{k-2} x p^{k-1} = \frac{\mu c_k \log d}{2^{k-1} d} \ll \mu$$

It is easy to see that the number of other types of intersecting U-paths is smaller, implying that

$$\mathbb{E}[\Pi] = \frac{c_k^2 n \log^2(d)}{2^{k+2} d^k}$$

Let us note that, because $\Pi(G)$ counts the number internally disjoint U-paths, removing one edge can change $\Pi(G)$ by at most one. Therefore, $\Pi(G)$ is a 1-Lipschitz function. Let us also note that $\Pi(G)$ is f-certifiable. That is, for $f(s) = ks$, when $\Pi(G) \geq s$, G contains a set of at most ks edges so that $\forall G'$ which agree with G on these edges, $\Pi(G') \geq s$. We can now use Talagrand's inequality to bound the number of graphs with insufficiently many U-paths.
For any b and t, Talagrand's tells us that

$$\mathbb{P}[|X - \mathbb{E}[X]| > t] \leq e^{-\frac{\beta t^2}{\mathbb{E}[X]}}$$

for some $\beta > 0$. This implies that for $t = \varepsilon \mathbb{E}[\Pi]$, $\varepsilon > 0$,

$$\mathbb{P}[\Pi < \frac{(1-\varepsilon) c_k^2 n \log^2(d)}{2^{k+2} d^k}] \leq e^{-\beta \varepsilon^2 \frac{c_k^2 n \log^2(d)}{2^{k+2} d^k}} = o(1)$$

Now, because the maximum number of sets U is at most

$$\binom{n}{x} \leq \left(\frac{en}{x}\right)^x \leq \left(\frac{ed^k}{c_k \log d}\right)^{c_k \frac{n}{d^k} \log d} \leq exp\left(c_k k \frac{n}{d^k} \log^2 d\right)$$

So, if

$$\frac{\beta \varepsilon^2 c_k^2}{2^{k+2}} > 2kc_k$$

then, with probability $1-o(1)$, for every set U, there are at least $\frac{\varepsilon n \log^2 d}{2^{k+2} d^k}$ pairwise internally disjoint U-paths.
Now, for $n \gg d \gg k$

$$10n2^{-d/10} + 10d^g < \frac{\varepsilon n \log^2 d}{2^{k+2} d^k}$$

so we can remove all small cycles and high-degree vertices without destroying all U-paths and therefore

$$\alpha(G^k) \leq c_k \frac{n \log(d)}{d^k} \Longrightarrow \chi(G^k) \geq \Omega\left(\frac{d^k}{\log(d)}\right)$$

as desired, and this completes our proof.

13 Martingales and Concentration Inequalities

The theory of Martingales and concentration inequalities were first used spectacularly by Janson, and then later by Bollobás in the determination of the chromatic number of a random graph. Ever since, concentration inequalities Azuma's inequality and its corollaries in particular, have become a very important aspect of the theory of probabilistic techniques. What makes these such an integral component is the relatively mild conditions under which they apply and the surprisingly strong results they can prove which might be near impossible to achieve otherwise. In this chapter, we shall review Azuma's inequality and as a consequence prove the Spencer-Shamir theorem for the chromatic number for sparse graphs and later, study the Pippenger-Spencer theorem for the chromatic index of uniform hypergraphs. Kahn extended some of these ideas to give an asymptotic version of the yet-open Erdős-Faber-Lovász conjecture for nearly disjoint hypergraphs.

13.1 Martingales

Suppose $\Omega, \mathcal{B}, \mathcal{P}$ is underlying probability space. $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq ...\mathcal{F}_n \subseteq ...$ where $\mathcal{F}_i$ is σ-algebra in $\mathcal{B}$.

$$\mathcal{F} = \bigcup_i \mathcal{F}_i$$

X_i is a martingale if X_i is $\mathcal{F}_i$ measurable and $\mathbb{E}(X_{i+1}|\mathcal{F}_i) = X_i$.
In general, if X is $\mathcal{F}$-measurable and $\mathbb{E}(X) < \infty$, then $X_i = \mathbb{E}(X|\mathcal{F}_i)$ always gives a martingale. This is called Doob's Martingale Process.

13.2 Examples

- **Edge Exposure Martingale**
  Let the random graph $G(n,p)$ be the underlying probability space. Label the potential edges $\{i,j\} \subseteq [n]$ by $e_1, e_2, ..e_m$ where $m = \binom{n}{2}$. Let f be any graph theoretic function. Then we can define martingale $X_0, X_1, X_2, ...X_m$ where:

$$X_i = \mathbb{E}(f(G)|e_j \; is \; revealed \; \forall 1 \leq j \leq i)$$

In other words to find X_i we first expose $e_1, e_2, ..., e_i$ and see if they are in G. Then X_i will be expectation of $f(G)$ with this information. Note that X_0 is constant.

- **Vertex Exposure Martingale**
  Again $G(n,p)$ is underlying probability space and f is any function of G. Define $X_1, X_2, ..., X_n$ by:

$$X_i = \mathbb{E}(f(G)|\forall x, y \leq i \ e_{x,y} \ is \ exposed)$$

  In words, to find X_i, we expose all edges between first i vertices (i.e. expose subgraph induced by $v_1, v_2, ..., v_i$) and look at the conditional expectation given this information.

13.3 Azuma's Inequality

Definition 146 (Lipshitz). *A function f is $K-Lipschitz$ if $\forall x, y \quad |f(x) - f(y)| \leq K|x-y|$. A martingale $X_0, X_1, ...$ is $K-Lipschitz$ if $\forall i \quad |X_i - X_{i+1}| \leq K$*

Theorem 147 (Azuma's Inequality). *Let $0 = X_0, X_1, ...X_m$ be a martingale with*

$$|X_{i+1} - X_i| \leq 1 \quad (i.e. 1-Lipschitz)$$

$\forall 0 \leq i < m$. Let $\lambda > 0$ be arbitrary. Then

$$\mathbb{P}(X_m > \lambda\sqrt{m}) < e^{-\lambda^2/2}$$

Proof. Set $\alpha = \lambda/\sqrt{m}$. Set $Y_i = X_{i+1} - X_i$ so that $|Y_i| \leq 1$ and $E(Y_i|X_{i-1}) = 0$. Then similar to argument used for proving Chernoff bound, we have:

$$\mathbb{E}(e^{\alpha Y_i}|X_{i-1}) \leq cosh(\alpha) \leq e^{\alpha^2/2}$$

Hence:

$$\begin{aligned}
\mathbb{E}(e^{\alpha X_m}) &= \mathbb{E}(\prod_{i=1}^{m} e^{\alpha Y_i}) \\
&= \mathbb{E}((\prod_{i=1}^{m-1} e^{\alpha Y_i})\mathbb{E}(e^{\alpha Y_m}|X_{m-1})) \\
&\leq \mathbb{E}(\prod_{i=1}^{m-1} e^{\alpha Y_i})e^{\alpha^2/2} \leq e^{\alpha^2 m/2} \quad (by \ induction)
\end{aligned}$$

and using this result we get:

$$\mathbb{P}(X_m > \lambda\sqrt{m}) = \mathbb{P}(e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}})$$

$$\leq \mathbb{E}(e^{\alpha X_m})e^{-\alpha\lambda\sqrt{m}}$$
$$\leq e^{\alpha^2 m/2-\alpha\lambda\sqrt{m}}$$
$$= e^{-\lambda^2/2} \quad (since\ \alpha = \lambda/\sqrt{m})$$

■

Corollary 148. *Let $c = X_0, X_1, ...X_m$ be a martingale with*

$$|X_{i+1} - X_i| \leq 1$$

$\forall 0 \leq i < m$. Let $\lambda > 0$ be arbitrary. Then

$$\mathbb{P}(|X_m - c| > \lambda\sqrt{m}) < 2e^{-\lambda^2/2}$$

13.4 The Shamir-Spencer Theorem for Sparse Graphs

Theorem 149 ([?])**.** *Consider $G = G(n,p)$ with $p = n^{-\alpha})$ for $\alpha > \frac{5}{6}$. Then for any $\varepsilon > 0$ there exists $u = u(\varepsilon, p)$ such that*

$$\mathbb{P}(\chi(G) \in \{u, u+1, u+2, u+3\}) \geq 1 - \varepsilon$$

Proof. Pick $\mu = \mu\,(p, \varepsilon)$ such that

$$\mathbb{P}(\chi(G) \leq \mu) > \varepsilon$$
$$\mathbb{P}(\chi(G) \leq \mu - 1) \leq \varepsilon$$

i.e. $\mu = min\{k \in \mathbb{Z}_{\geq 0} \mid \mathbb{P}(\chi(G) \leq \mu) > \varepsilon\}$

Let Y be the **largest** sub-graph that is μ-colorable (in terms of, say, Lexicographic Order). Let $R = V(G) \setminus Y$. $|R|$ denotes the number of vertices in R and note that $|R| = 0 \Longrightarrow$ Graph was μ-colorable.

Consider the **vertex exposure martingale**, that is the expected value of $|R|$, conditioned on what vertices are 'exposed' to the us. Since any vertex will be either in R or in Y, we have

$$X_i = \mathcal{E}\,(|R| \mid G\,[1, 2 \ldots i]) \rightsquigarrow \text{ a Doob Martingale!}$$

Now, $X_0 = \mathcal{E}(R)$ and $X_n = R$. As addition of any vertex can change the value of X_i by 1 for all $i \in [n]$,

$$|X_i - X_{i-1}| \leq 1 \ \ \forall i \in [n]$$

and hence by the Azuma-Hoeffding Inequality (Theorem 54), we have,

$$\mathbb{P}(||R| - \mathcal{E}[|R|]| \geq \lambda\sqrt{n}) \leq 2e^{-\frac{\lambda^2}{2}} \ \forall \lambda > 0$$

Pick λ such that $2e^{-\frac{\lambda^2}{2}} < \varepsilon$. Then, since $|R| = 0 \implies G$ is μ-colorable, $\mathbb{P}(|R| \leq 0) > \varepsilon$ and $\mathbb{P}(||R| - \mathcal{E}[|R|]| \leq \lambda\sqrt{n}) > 1 - \varepsilon$, these events cannot be disjoint and hence

$$0 \in \left(\mathcal{E}[|R|] - \lambda\sqrt{n}, \mathcal{E}[|R|] + \lambda\sqrt{n}\right)$$

and hence,

$$\mathbb{P}(|R| > 2\lambda\sqrt{n}) \leq \mathbb{P}(|\mathcal{E}[|R|] - |R|| > \lambda\sqrt{n}) < \varepsilon$$

Therefore, $|R| \leq 2\lambda\sqrt{n}$ w.h.p. Now we make the following claim:

Proposition 150. *Any $2\lambda\sqrt{n}$-sized graph is 3-colorable w.h.p.*

Before proving this, we make the following remark:

Proposition 151. *If H is graph that is not k-colorable. Then, there exist an induced subgraph H' such that H' is not k-colorable and $\delta(H') \geq k$*

We will not provide the proof for this assertion; however, interested readers can refer to [**?**, Theorem 2, Chapter 5] for further details. Now we can start proving Proposition 150.

Proof. If $G[T]$ is not 3-colorable, then there exists a $T' \subset T$ such that $\delta(T') \geq 3$. Therefore,

$$\begin{aligned}\mathbb{P}(\exists T' \subset G, |T'| \leq t, G[T'] \text{ is not 3-colorable}) &\leq \sum_{k=3}^{t} \binom{n}{k} \binom{\binom{k}{2}}{\frac{3k}{2}} p^{\frac{3k}{2}} \\ &\leq \sum_{k=3}^{t} \left(\frac{ne}{k}\right)^k \left(\frac{ke}{3}\right)^{\frac{3k}{2}} p^{\frac{3k}{2}}\end{aligned}$$

as $e(G[T']) \geq \frac{3k}{2}$. Using $k \leq t \leq 2\lambda\sqrt{n}$, each term would be of the form

$$\mathcal{O}(n\sqrt{k}p^{\frac{3k}{2}})^k = \mathcal{O}(n^{\frac{5}{4} - \frac{3\alpha}{2}})^k = o(1)$$

if $\alpha > \frac{5}{6}$. Hence, w.h.p, $G[T]$ is at most 3-colorable. ■

Combining Theorem 151 and fact that $Y = V(G) \setminus R$ is μ-colorable, we obtain that, w.h.p, $\chi(G) \in \{\mu, \mu + 1, \mu + 2, \mu + 3\}$ ■

13.5 The Pippenger-Spencer Theorem

Let $\mathcal{H}$ be a hypergraph. We say that $\mathcal{E}(\mathcal{H})$ can be properly $N-colored$ if $\mathcal{E}(\mathcal{H})$ can be partitioned into N matchings in $\mathbb{H}$. By a matching, we mean a set of mutually non-intersecting hyper-edges.
The smallest N for which $\mathcal{E}(\mathcal{H})$ can be $N-colored$ is called chromatic index of $\mathcal{H}$, denoted by $\chi'(\mathcal{H})$.
If G is a graph, we know that $\Delta(G) \leq \chi(G)$ where $\Delta(G)$ is max vertex degree.
Also from **Vizing-Gupta Theorem** we have $\chi'(G) \leq \Delta(G)+1$. Overall we know:

$$\Delta(G) \leq \chi'(G) \leq \Delta(G)+1$$

for graphs.
However it is computationally hard to figure out if $\chi'(G) = \Delta(G)$ *or* $\Delta(G)+1$.

For $\mathcal{H}$ note that $\chi'(\mathcal{H}) \geq \Delta(\mathcal{H})$ where Δ still denotes max degree in H i.e.:
$\Delta(\mathcal{H}) = \max\{d(x)|x \in V(\mathcal{H})\}, \quad d(x) = \#\ of\ hyperedges\ containing\ x$

Theorem 152 (The Pippenger-Spencer Theorem)**.** *Given $\varepsilon > 0$, $\exists$ a $\delta > 0$ and $D_0(\varepsilon)$ s.t. the following holds if $n \geq D \geq D_0$ and:*

- $D > d(x) > (1-\delta)D$
- $d(x,y) < \delta D \forall x, y \in V \quad where\ D = \Delta(\mathcal{H})$ $\qquad\qquad \star$

Then $\chi'(\mathcal{H}) < (1+\varepsilon)D$

Note: *$d(x,y)$ is codegree of x, y i.e. $d(x,y) = |\{E \in \mathcal{E}(\mathcal{H})\ s.t.\ \{x,y\} \subseteq E\}|$*

The proof of this theorem due to Pippenger-Spencer follows the paradigm of the 'pseudo-random method' pioneered by Vojtech Rödl and the 'Nibble'.

Proof of the P-S theorem:
Idea: Pick each edge of $\mathcal{E}$ with probability $\frac{\varepsilon}{D}$ independent of each other. Form the sub-collection that is obtained, $\mathcal{E}_1$, throw away these edges and other incident edges to $\mathcal{E}_1$. The resulting hypergraph is $\mathcal{H}_1$. Then with high probability $\mathcal{H}_1$ also satisfies the same the same 2 conditions $\star$ of Pippenger-Spencer for a different D.
From $\mathcal{E}_1$ extract a matching $\mathcal{M}_1$, i.e. pick those edges of $\mathcal{E}_1$ that do not intersect any other edges of $\mathcal{E}_1$. By repeating this procedure we have:

$$\mathcal{H} = \mathcal{H}_0 \xrightarrow{\mathcal{E}_1} \mathcal{H}_1 \xrightarrow{\mathcal{E}_2} \mathcal{H}_2 \ldots \xrightarrow{\mathcal{E}_t} \mathcal{H}_t$$

$D_1 \approx De^{-\varepsilon k}$ (where $\mathcal{H}$ is k-uniform) since

$$\mathbb{P}(edge\ surviving) \approx \left[(1-\frac{\varepsilon}{D})^D\right]^k = e^{-\varepsilon k}$$

asymptotically. Now let:

$$\mathcal{M}^{(1)} = \bigcup_{i=1}^{t} \mathcal{M}_i \quad (M_i \textit{ are disjoint by construction})$$

For an edge A:

$$\mathbb{P}(A \in \mathcal{M}^{(1)}) = \sum_{i=1}^{t} P(A \in M_i) \quad \textit{and}$$

$$\mathbb{P}(A \in \mathcal{M}_1) \approx \frac{\varepsilon}{D}, \;\; \mathbb{P}(A \in \mathcal{M}_2) \approx \frac{\varepsilon}{D_1}(1 - \frac{\varepsilon}{D})^{k(D-1)} \approx \frac{\varepsilon}{D_1} e^{-\varepsilon k} \;\; \textit{in general}:$$

$$\mathbb{P}\big(A \in \mathcal{M}_i\big) \approx \frac{\varepsilon}{D} e^{-\varepsilon k + \varepsilon(i-1)}$$

$$\implies \mathbb{P}(A \in \mathcal{M}^{(1)}) = e^{-\varepsilon k} \left(\frac{\varepsilon}{D}\right) \sum_{i=1}^{t} e^{\varepsilon(i-1)} = e^{-\varepsilon\varepsilon k} \left(\frac{\varepsilon}{D}\right) \left(\frac{1 - e^{\varepsilon t}}{1 - e^{\varepsilon}}\right) \approx \frac{\alpha}{D}$$

where $\alpha = \alpha(\varepsilon, t, k) = \varepsilon e^{-\varepsilon k \frac{(1-e^{\varepsilon t})}{1-e^{\varepsilon}}}$. Now, we can generate a second independent matching $M^{(2)}$ by repeating the same process and so on.

Just like the Rödl's nibble start by picking a 'small' number of 'independent' matchings from $\mathcal{H}$. Let $0 < \theta < 1$ and $\mu = \lfloor \theta D \rfloor$ and generate independent matchings $\mathcal{M}^{(1)}, \mathcal{M}^{(2)}, \mathcal{M}^{(3)} \ldots \mathcal{M}^{(\mu)}$ with each $\mathcal{M}^{(i)}$ having:

$$\mathbb{P}(A \in \mathcal{M}^{(i)}) \approx \frac{\alpha}{D}$$

Let $\mathcal{P}^{(1)} = \mathcal{M}^{(1)} \cup \mathcal{M}^{(2)} \cup \mathcal{M}^{(3)} \cup \cdots \cup \mathcal{M}^{(\mu)}$.

$$\mathcal{H} = \mathcal{H}^{(0)} \xrightarrow{\mathcal{P}^{(1)}} \mathcal{H}^{(1)} \xrightarrow{\mathcal{P}^{(2)}} \mathcal{H}^{(2)} \ldots \xrightarrow{\mathcal{P}^{(s)}} \mathcal{H}^{(s)}$$

Here first 'packing' $\mathcal{P}^{(1)}$ is $\mu = \theta D$-colorable since we can assign each matching $\mathcal{M}^{(i)}$ a separate color. Note that $\chi'(\mathcal{H}^{(0)}) \leq \mu + \chi'(\mathcal{H}^{(1)})$ (since chromatic number is subadditive). Similarly $\mathcal{P}^{(2)}$ is $\theta D^{(1)} - \textit{colorable}$ and so on.

Hence so far we need $\theta D + \theta D^{(1)} + \cdots + \theta D^{(s-1)}$ colors. After removing colored edges (i.e. $\textit{edges} \in \textit{some } \mathcal{P}^{(i)}$), very few edges will be left in $\mathcal{H}^{(s)}$.

Bounding $\chi'(\mathcal{H}^{(s)})$: For any $k - uniform$ hypergraph $\mathcal{H}$ with max degree D, we have: $\chi'(\mathcal{H}) \leq k(D-1) + 1 \implies \chi'(\mathcal{H}^{(s)}) \leq k(D^{(s)} - 1) + 1$

Hence:

$$\textit{total \# of colors we used} = \theta \sum_{i=1}^{s-1} D^{(i)} + \theta D + k(D^{(s)} - 1) + 1 \approx D$$

s will be chosen as large as possible. Here we need to make sure that $\mathcal{H}^{(i)}$ is similar to $\mathcal{H}^{(i-1)}$ (i.e. all degrees are almost equal and the co-degree is small). (In particular we'll be interested in $i = 1$ case).

Fix any $x \in \mathcal{H}$, what is the $\mathbb{E}(d^{(1)}(x))$?

$$d^{(1)}(x) = \sum_{A: x \in A \in \mathcal{H}^{(0)}} \mathbb{1}_{A \notin \mathcal{P}^{(1)}}$$

$$\implies \mathbb{E}(d^{(1)}(x)) = \sum_{A: x \in A \in \mathcal{H}^{(0)}} (1 - \frac{\alpha}{D})^{\mu} \approx D(1 - \frac{\alpha}{D})^{\mu} \approx D(1 - \frac{\alpha}{D})^{\theta D} \approx De^{-\alpha\theta} = D^{(1)}$$

Hence $\mathbb{E}(d^{(1)}(x)) \approx D^{(1)} = De^{-\alpha\theta}$
Use Azuma's inequality to get a concentration inequality for $d^{(1)}(x)$. The art is to pick the right filtration.

(We will consider the following martingale $X_i = \mathbb{E}[d^{(1)}(x) \mid \mathcal{M}^{(1)}, \mathcal{M}^{(2)}, \ldots, \mathcal{M}^{(i)}]$)

Let $\mathcal{F}_i = \{\mathcal{M}^{(1)}, \mathcal{M}^{(2)}, \ldots, \mathcal{M}^{(i)}\}$ since $\mathcal{M}^{(i)}$ is a matching $\implies$ at most one edge containing x is exposed.

Then $\mathbb{E}[d^{(1)}(x)|\mathcal{F}_i] := X_i$ is a $1 - Lipschitz$ martingale. So by Azuma's inequality:
$\mathbb{P}(|d^{(1)}(x) - D^{(1)}| > \lambda\sqrt{\mu}) \leq e^{-\lambda^2/2}$ (Here x is fixed and $\mu \approx \theta D = o(1)D$)

Now question is: "How to guarantee this for all vertices?". Use Lovasz Local Lemma (LLL):

$$A_x := |d^{(1)}(x) - D^{(1)}| > \lambda\sqrt{o(1)D^{(1)}}$$

Want to show that:

$$\mathbb{P}\left(\bigwedge_{x \in V} \overline{A}_x\right) > 0$$

We know: $\mathbb{P}(A_x) \leq 2e^{-\lambda^2/2}$. To compute the dependence degree among $\{A_x | x \in V(\mathcal{H})\}$:
$\mathcal{M}^{(i)} = \mathcal{M}_1^{(i)} \cup \mathcal{M}_2^{(i)} \cup \ldots \mathcal{M}_t^{(i)}$

(Distance between two vertices is the shortest number of edges one needs to go from x to y.)
Note that each matching $\mathcal{M}^{(i)}$ is generated by atoms $\mathbb{1}_E$ where each $E \in \mathcal{H}^{(0)}$ and whose 'distance' from $x \leq t$. So if *distance between x and y* $\geq 2t+1$, A_x *and* A_y are independent.

$\implies$ *Dependence degree*

$\leq (k-1)D^{(0)} + 2(k-1)^2(D-1)D + \cdots + r(k-1)^r(D-1)^r + \cdots + 2t(k-1)^{2t}(D-1)^{2t}$

$\leq (2t+1)(kD^{(0)})^{2t+1}$

So for LLL, we need:

$$e2e^{-\lambda^2/2}(2t+1)(kD^{(0)})^{2t+1} < 1$$

Put $\lambda = \sqrt{o(1)D^{(1)}}$ to get: $\iff \frac{e(2t+1)(kD^{(0)})^{2t+1}}{e^{o(1)D^{(1)}/2}} < 1$.

Asymptotically $D^{(1)}$ beats t (big time), so condition for LLL will hold hence we are in business.

Finally repeating the previous argument:
$\chi'(\mathcal{H}) \leq \mu^{(0)} + \mu^{(0)} + \cdots + \mu^{(s-1)} + \chi'(\mathcal{H}^{(s)})$
where $\mu^{(i)} = \theta D^{(i)}$ and $D^{(i)} = e^{-\alpha\theta i}D$ and $\chi'(\mathcal{H}^{(s)})$ bounded above as before. Then we get:

$$\chi'(\mathcal{H}) \leq \theta D(1+e^{-\alpha\theta}+e^{-2\alpha\theta}+\cdots+e^{-(s-1)\alpha\theta})+k\theta De^{-s\alpha\theta}$$

$$\leq \frac{\theta D}{1-e^{-\alpha\theta}}+k\theta De^{-s\alpha\theta} \to D(1+o(1))$$

as $t \to \infty$, $s \to \infty$, $\varepsilon \to \infty$, etc. Thus we'll have the desired result.

When we do the calculations, everything works out nicely.

13.6 A Conjecture of Erdős-Faber-Lovász (EFL) and a theorem of Kahn

Definition 153. *A hypergraph $\mathcal{H}$ is nearly-disjoint or linear if*

$$\forall A \neq B \in \mathcal{E}(\mathcal{H}), \quad |A \cap B| \leq 1$$

.

Conjecture 154. *If $\mathcal{H}$ is nearly-disjoint on n vertices, then $\chi'(H) \leq n$*

Theorem 155 (Erdos-de Bruijn Theorem)**.** *If $\mathcal{H}$ is a hypergraph on n vertices with*

$$|A \cap B| = 1 \quad \forall A \neq B$$

then $|\mathcal{E}(\mathcal{H})| \leq n$.

As an aside, $|\mathcal{E}(\mathcal{H})| \leq n \implies \chi'(\mathcal{H}) \leq n$. This theorem is tight in the sense that if it is a projective plane of order n, then n^2+n+1 colors are needed $\implies \chi'(\mathcal{H}) = |\mathcal{E}(\mathcal{H})|$. ($\P_n$ = *projective plane of order n*)

Theorem 156 (Theorem - Jeff Kahn (1992)). *The EFL conjecture is asymptotically true, i.e.* $\chi'(\mathcal{H}) \leq n(1+o(1))$ *for* $\mathcal{H}$ *nearly-disjoint on n-vertices.*

Note that in this general situation, the edge sizes need not be the same; in fact they need not even be absolutely bounded, and as we shall see, that causes some of the trouble.

Firstly, we start with a simple observation. If there is an integer k such that for each edge E in a nearly disjoint hypergraph $\mathcal{H}$ we have $|E| \leq k$, then we can 'uniformize' the edge sizes. This is a standard trick, so we will not describe it in detail. One may form a bipartite graph $\mathcal{G}$ whose vertex sets are the vertices and edges of $\mathcal{H}$, and (v, E) is an incident pair iff $v \in E$. Then the uniformization described earlier is equivalent to embedding $\mathcal{G}$ into a bipartite graph with uniform degree over all the vertices $E \in \mathcal{E}$ such that the graph is C_4-free. This is a fairly standard exercise in graph theory.

If all the edges are of bounded size, i.e., if $3 \leq b \leq |E| \leq a$ for all edges E then the Pippenger-Spencer theorem of the preceding section proves the result claimed by the aforementioned theorem. Indeed, for any x count the number of pairs (y, E) where $y \neq x$, and $x, y \in E$. Since $\mathcal{H}$ is nearly disjoint, any two vertices of $\mathcal{H}$ are in at most one edge so this is at most $n-1$. On the other hand, this is precisely $\sum_{x \in E}(|E|-1)$, so we have $(b-1)d(x) \leq n-1 \Rightarrow d(x) \leq \frac{n-1}{b-1} < \frac{n}{2}$.

Here is a general algorithm for trying to color the edges of $\mathcal{H}$ using C colors: Arrange the edges of $\mathcal{H}$ in decreasing order of size and color them greedily. If the edges are $E_1, E_2, \ldots, E_m$ with $|E_i| \geq |E_{i+1}|$ for all i then when E_i is considered for coloring, we may do so provided there is a color not already assigned to one of the edges $E_j, j < i$ for which $E_i \cap E_j \neq \emptyset$. To estimate $|\{1 \leq j < i | E_j \cap E_i \neq \emptyset\}|$, let us count the number of triples (x, y, j) where $x \in E_i \cap E_j, y \in E_j \setminus E_i$. Write $|E_i| = k$ for simplicity. Again, since $\mathcal{H}$ is nearly disjoint, any two vertices of $\mathcal{H}$ are in at most one edge, hence the number of such triples is at most the number of pairs (x, y) with $x \in E_i, y \notin E_i$, which is $k(n-k)$. On the other hand, for each fixed E_j such that $1 \leq j < i, E_j \cap E_i \neq \emptyset$, $E_i \cap E_j$ is uniquely determined, so the number of such triples is $|E_j| - 1$. Hence denoting $\mathcal{I} = \{1 \leq j < i | E_j \cap E_i \neq \emptyset\}$ and noting that for each $j \in \mathcal{I}$ $|E_j| \geq k$, we get

$$(k-1)|\mathcal{I}| \leq \sum_{j \in \mathcal{I}}(|E_j|-1) \leq k(n-k) \Rightarrow |\mathcal{I}| \leq \frac{k(n-k)}{k-1}.$$

In particular, if $C > \frac{|E|(n-|E|)}{|E|-1}$ for every edge E, the greedy algorithm properly colors $\mathcal{H}$.

Upshot: For any nearly disjoint hypergraph $\mathcal{H}$ on n vertices $\chi'(\mathcal{H}) \leq 2n-3$.

The previous argument actually shows a little more. Since $\frac{k(n-k)}{k-1}$ is decreasing in k if $|E| > a$ for some (large) constant a, then $|\mathcal{I}| < (1+\frac{1}{a})n$. So, for a given $\varepsilon > 0$ if we

$a > 1/\varepsilon$, say, then for $C = (1+2\varepsilon)n$, following the same greedy algorithm will properly color all edges of size greater than a. This motivates us to consider

- $\mathcal{E}_s := \{E \in \mathcal{E} : |E| \leq b\}$.
- $\mathcal{E}_m := \{E \in \mathcal{E} : b < |E| \leq a\}$.
- $\mathcal{E}_l := \{E \in \mathcal{E} : |E| > a\}$

for some absolute constants a, b which we shall define later. We have seen that $\chi'(\mathcal{H}_l) \leq (1+2\varepsilon)n$; also by a preceding remark, if we pick $b > O(1)/\varepsilon$ we have $\chi'(\mathcal{H}_m) \leq \varepsilon n$. Thus, let us do the following.

Let $C = \lfloor (1+4\varepsilon)n \rfloor$; we shall color the edges of $\mathcal{H}$ using the colors $\{1, 2 \ldots, C\}$. Let $C_1 = \{1, 2 \ldots, \lfloor (1+3\varepsilon)m \rfloor\}$; $C_2 := C \setminus C_1$. Fix a coloring f_1 of $\mathcal{H}_l$ using the colors of C_1, and a coloring f_2 of $\mathcal{H}_m$ using the colors of C_2. We now wish to color $\mathcal{H}_s$. We shall attempt to do that using the colors of C_1. For each $E \in \mathcal{H}_s$ let

$$\text{Forb}(E) := \{c \in C_1 | E \cap A \neq \emptyset \text{ for some } A \in \mathcal{H}_l, f_1(A) = c\}.$$

Then as before, $|\text{Forb}(E)| \leq |\{A \in \mathcal{H}_l | A \cap E \neq \emptyset\}| \leq \frac{a(n-a)}{b} < \eta D$ for $\eta = a/b, D = n$. In other words, every edge of $\mathcal{H}_s$ also has a (small) list of forbidden colors for it. If we can prove a theorem that guarantees a proper coloring of the edges with no edge given a forbidden color, we have an asymptotic version of the EFL.

At this point, we are motivated enough (as was Kahn) to state the following

Conjecture 157. *Let $k \geq 2$, $\nu > 0$, $0 \leq \eta < 1$. Let C be a set of colors of size at least $(1+\nu)D$. There exists $\beta > 0$ such that if $\mathcal{H}$ is a k-uniform hypergraph satisfying*

- *$(1-\beta)D < d(x) \leq D$ for all vertices x of $\mathcal{H}$,*
- *$d(x, y) < \beta D$ for all distinct pairs of vertices x, y,*
- *For each $A \in \mathcal{E}$, there is a subset $Forb(A) \subset C$ with $|Forb(A)| < \eta D$.*

then there is a proper coloring f of $\mathcal{E}$ such that for every edge A, $f(A) \notin Forb(A)$.

Note that the first two conditions are identical to those of the PS theorem. Also, it is important to note that there might be some additional constraints on η, ν which indeed is the case. We will see what those are as we proceed with the proof.

To prove this conjecture, let us again recall the idea of the proof of the PS theorem. The i^{th} step/iteration in the proof of the PS theorem does the following: Fix $0 < \theta < 1$, and let t, s be large integers. Starting with the hypergraph $\mathcal{H}^{(i)} (1 \leq i \leq s)$ which satisfies conditions (1), (2) above with $D^{(i)} := e^{-\alpha\theta i} D$ with $\alpha = \alpha(\varepsilon, t, k) = \varepsilon e^{-\varepsilon k} \frac{(1-e^{\varepsilon t})}{1-e^{\varepsilon}}$, with positive probability there is a random packing $\mathcal{P}^{(i+1)} := \mathcal{M}_{i+1}^{(1)} \cup \mathcal{M}_{i+1}^{(2)} \cup \cdots \cup \mathcal{M}_{i+1}^{(\mu_i)} \in \mathcal{H}^{(i)}$ with $\mu_i = \lfloor \theta D^{(i)} \rfloor$, such that

- $\mathbb{P}(A \in \mathcal{P}^{(i+1)}) \approx \frac{\alpha}{D^{(i)}}$.
- For all $A \in \mathcal{H}^{(i)}$ the event "$A \in \mathcal{P}^{(i+1)}$" is independent of all events "$B \in \mathcal{P}^{(i+1)}$" if distance between A, B is at least $2t$. Here, the distance is in the hypergraph $\mathcal{H}^{(i)}$.

The idea is to try to give every edge its 'default color' as and when we form the packings $\mathcal{P}^{(i)}$. Since each such packing consists of up to μ_i different matchings, $\mathcal{P}^{(i)}$ can be (by default) colored using μ_i colors, so that when we complete s iterations we have used $\sum_i \mu_i$ different colors to color all the edges except those of $\mathcal{H}^{(s)}$. The PS theorem finishes off by coloring these edges greedily using a fresh set and colors by observing that the number of edges in $\mathcal{H}^{(s)}$ is 'small'.

To keep track of these let us write

$$\mathcal{C} := \bigcup_{1 \le j \le \mu_i, 1 \le i \le s} \mathcal{C}_{ij} \cup \mathcal{C}^*, \quad \text{with } \mathcal{C}_{ij} := \{c_{i1}, c_{i2}, \ldots, c_{i\mu_i}\},$$

where these sets $\mathcal{C}_{ij}$ are mutually disjoint and the matching $\mathcal{M}_{i+1}^{(j)}$ is by default allocated color c_{ij}.

In our present situation, the default colors allocated to some of the edges may be forbidden at those edges. More specifically, define

$$\mathcal{B}^{(i)} := \{A \in \mathcal{H}^{(i)} | A \in \mathcal{M}_{i+1}^{(j)} \text{ for some j and } c_{ij} \in \text{Forb}(A)\}.$$

For each vertex v, let $B_v^{(i)} := |\{A \in \mathcal{B}^{(i)} | v \in A\}|$.

At each stage, remove the 'bad edges' from the packings, i.e., the ones assigned a forbidden color. After s iterations the edges that need to be (re)colored are the ones in $\mathcal{H}' := \mathcal{H}^{(s)} \bigcup_{i=1}^{s} \mathcal{B}^{(i)}$ and the colors that are left to be used are those in $\mathcal{C}^*$. Note that for each vertex v we have $d_{\mathcal{H}'}(v) \le D_v^{(s)} + B_v$. The first term is $o(D)$; if the second term is also $o(D)$ then we may finish the coloring greedily. Thus, if we can show that we can pick our random packing at stage i in such a way that apart from the criteria in the PS-theorem, we can also ensure that $B_v^{(i)}$ is 'small' (compared to the order of $D^{(i)}$) then we are through (there is still some technicality but we will come to that later).

Hence to start with, we need to show that at each step i of the iteration, we can get a random packing $\mathcal{P}^{(i+1)}$ such that

- $|d^{(i)}(v) - D^{(i)}| < o(D^{(i)})$ for all v.
- $B_v^{(i)} < \mathcal{E}(B_v^{(i)}) + o(D)$

The proof of this part is identical to that of the PS theorem; use the same martingale, the same filtration, and use Azuma's inequality.

To complete the proof, we need to get an (over)estimate of $\mathcal{E}(B_v^{(i)})$. For each $A \in \mathcal{H}^{(i)}$, A is **not** in $B^{(i)}$ if and only if for each $c_{ij} \in \mathrm{Forb}(A)$ we have $A \notin \mathcal{M}_{i=1}^{(j)}$. Denoting $\mathrm{Forb}^{(i)}(A) := \{j | c_{ij} \in \mathrm{Forb}(A)\}$ we have

$$\mathbb{P}(A \in B^{(i)}) = 1 - \left(1 - \frac{\alpha}{D^{(i)}}\right)^{|\mathrm{Forb}^{(i)}(A)|} < \frac{\alpha |\mathrm{Forb}^{(i)}(A)|}{D^{(i)}}.$$

Hence,

$$\begin{aligned}\mathbb{E}(B_v^{(i)}) &= \sum_{v \in A \in \mathcal{H}^{(i)}} \mathbb{P}(A \in B^{(i)}) \\ &\lesssim \frac{\alpha}{D^{(i)}} \sum_{v \in A \in \mathcal{H}^{(i)}} |\mathrm{Forb}^{(i)}(A)|\end{aligned}$$

Let $i(A) := \max\{0 \leq i \leq s | A \in \mathcal{H}^{(i)}\}$. Note that for any fixed i,

$$|\{A \in \mathcal{H} | v \in A, i(A) = i\}| \leq \theta e^{-\alpha\theta i} D.$$

Hence we have

$$\begin{aligned}\sum_{i=0}^{s} \mathbb{E}(B_v^{(i)}) &\lesssim \alpha \sum_{i=0}^{s} \frac{1}{D^{(i)}} \sum_{v \in A \in \mathcal{H}^{(i)}} |\mathrm{Forb}^{(i)}(A)| \\ &= \alpha \sum_{v \in A} \sum_{i} \frac{|\mathrm{Forb}^{(i)}(A)|}{D^{(i)}} \left(\mathbf{1}_{A \in \mathcal{H}^{(i)}}\right) \\ &\leq \alpha \sum_{v \in A} \frac{1}{D^{(i(A))}} \left(\sum_{i} |\mathrm{Forb}^{(i)}(A)|\right) \\ &\leq \alpha \sum_{v \in A} \frac{|\mathrm{Forb}(A)|}{D} e^{\alpha\theta i(A)} \\ &< \alpha o(1) \sum_{i=0}^{s} e^{\alpha\theta i} |\{A | v \in A, i(A) = i\}|\end{aligned}$$

The last term in the above expression can be made 'small'. This completes the proof of Kahn's theorem.

14 Algebraic Rigidity versus Randomness

Algebraic constructions have a form of rigidity that allow us to define distributions with sharply curtailed distributions.

In 2016, Bukh and Conlon [8] proved the long-standing open problem posed by Erdős (albeit in a slightly weaker form): Given $1 \leq r \leq 2$ with $r \in \mathbb{Q}$ there exists a family $\mathcal{H}$ of graphs such that $\mathrm{ex}(n; \mathcal{H}) = \Theta(n^r)$ for sufficiently large n. The original conjecture of Erdős posited the same for a single graph H instead of a family $\mathcal{H}$.

The main new idea here involves a randomized construction with an algebraic twist. This chapter illustrates this principle with two instances of this principle. Both these are results in extremal graph theory.

14.1 The Turán number for $K_{s,t}$

An old (though elementary) argument of Kövari-Sós-Turán establishes the still-best-known bound for the Zarankiewicz problem, which is basically the question of determining the Turán number for complete bipartite graphs. For fixed constants $s \leq t$, and n sufficiently large,

$$\mathrm{ex}(n; K_{s,t}) \leq O_{s,t}(n^{2-1/s}).$$

The hard problem concerns a matching lower bound[1] for $\mathrm{ex}(n; K_{s,t})$. To see why this is a much harder problem, let us first attempt a straightforward randomized construction. Suppose $V = L \cup R$ with $|L| = |R| = n$ and consider the random graph with edge probability $p = n^{-1/s}$. Then for any set $U \subset L$ of size s, and a fixed $v \in R$ $|N(U)| \sim \mathrm{Bin}(n, 1/n)$ so the Poisson approximation for the Binomial tells us that the probability that there is no U with $|N(U)|$ at least t is $o(1)$ only if $t = \Omega\left(\frac{\log n}{\log \log n}\right)$, and this is unfortunately too large for us to be of any use. This is in some sense a fundamental obstruction; since the

[1] It is universally believed, and for good reason too, that this bound is asymptotically tight, upto constants that depend on s, t.

Poisson distribution has a smooth tail,[2] this construction is fundamentally doomed.

Another interesting open problem of a similar spirit is the Faudree-Simonovits conjecture: Let $\Theta_{k,\ell}$ denote the graph which consists of ℓ internally vertex-disjoint paths of length k each, between two fixed vertices. Then $\mathrm{ex}(n;\Theta_{k,\ell}) \leq O_{k,\ell}(n^{1+1/k})$as was shown by Faudree & Simonovits. But a construction of a graph that matches a corresponding lower bound (at least asymptotically) is still open.

14.2 Algebraic Rigidity

We are often interested in studying a finite subset of $\mathbb{F}^n$ (where $\mathbb{F}$ is a finite field) as a zero set of some (small degree) polynomial $f \in \mathbb{F}[X_1, \ldots, X_n]$. In this regard, we have two very important facts:

Lemma 158 (Parameter Counting)**.** *Let $\mathbb{F}[X_1, \ldots, X_n]$ be the ring of polynomials in n variables over $\mathbb{F}$. For any $f \in \mathcal{F}[X_1, \ldots, X_n]$, denote by $Z(f) := \{a \in \mathbb{F}^n : f(a) = 0\}$ the zero-set of the polynomial f. Then:*

1. *Suppose $S \subseteq \mathbb{F}^n$, then there exists a non-zero polynomial $f \in \mathbb{F}[X_1, \ldots, X_n]$ of degree $\leq D$ such that $S \subseteq Z(f)$ if $|S| < \binom{n+D}{n}$. In particular, $|S| < \binom{n+D}{n}$ holds for $D = n \cdot |S|^{1/n}$.*
2. *Let $\mathcal{L}$ be a collection of lines in $\mathbb{F}^n$. Then there exists a non-zero polynomial $f \in \mathbb{F}[X_1, \ldots, X_n]$ of degree $\leq D$ such that $\ell \subseteq Z(f)$ for all $\ell \in \mathcal{L}$ if $(D+1)|\mathcal{L}| < \binom{n+D}{n}$. In particular, $(D+1)|\mathcal{L}| < \binom{n+D}{n}$ holds for $D = (2n+1) \cdot |\mathcal{L}|^{1/(n-1)}$.*

Proof. Let $f = \sum a_{\mathbf{i}} X^{\mathbf{i}}$ be a generic polynomial of degree $\leq D$, where $\mathbf{i} = (i_1, \ldots, i_n)$ is a multi-index with $i_j \geq 0, \sum_{j=1}^{n} i_j \leq D$, and $X^{\mathbf{i}}$ denoting $X_1^{i_1} \cdots X_n^{i_n}$. f has $\binom{n+D}{n}$ coefficients, and the equalities $f(s) = 0$ for each $s \in S$ lead to $|S|$ linear equations in those coefficients. Since the number of variables ($\binom{n+D}{n}$) is greater than the number of equations ($|S|$), there exists a non-trivial solution to those equations, which corresponds to a non-zero polynomial of degree $\leq D$ whose zero set contains S, as desired.
Now, note that every line ℓ in $\mathbb{F}^n$ is of the form $\ell(t) := a_0 + a_1 t$ for some $a_0, a_1 \in \mathbb{F}^n$. Thus, for any $f \in \mathbb{F}[X_1, \ldots, X_n]$ of degree $\leq D$, $f|_\ell$ is a univariate polynomial of degree $\leq D$. Thus, if $f|_\ell$ is 0 on $D+1$ points on ℓ, then $f|_\ell = 0$, since a non-zero univariate polynomial of degree $\leq D$ over a field has $\leq D$ roots.
Thus, for every $\ell \in \mathcal{L}$, arbitrarily choose $D+1$ points on ℓ, and set f to 0 on those points. Then we have $(D+1)|\mathcal{L}|$ linear equations, and once again if $(D+1)|\mathcal{L}| < \binom{n+D}{n}$, we're done. ∎

Note the dichotomy that a degree D polynomial can either vanish on at most D points on a line, or it must vanish everywhere on the line. We shall exploit this dichotomy to prove some interesting results now. But before that, some definitions:

[2] The Poisson distribution is a discrete distribution, but the distribution function goes down smoothly

Definition 159. *Let $\mathcal{L}$ be a collection of lines in $\mathbb{F}^n$. Define*

$$\deg(\mathcal{L}) := \min\{\deg(f) : Z(f) \supseteq \ell \text{ for all } \ell \in \mathcal{L}\}$$

Also, for $r \geq 2$, define:

$$\mathcal{P}_r(\mathcal{L}) := \{p \in \mathbb{F}^n : \geq r \text{ lines from } \mathcal{L} \text{ pass through } p\}$$

The points in $\mathcal{P}_r(\mathcal{L})$ are also known as the r-rich points of $\mathcal{L}$.

Theorem 160. *Suppose $\mathcal{L}$ is a collection of lines in $\mathbb{F}^3$, and write $|\mathcal{L}| = L$. Suppose each $\ell \in \mathcal{L}$ has $\geq A$ 2-rich points. Then $\deg(\mathcal{L}) = O(L/A)$.*

Remark 1: Note that planes are degree 1 curves. Thus, the above result should be interpreted as imposing some sort of 'approximate planarity' (especially if $A \sim L$) on $\mathcal{L}$ purely from local, combinatorial information.
Remark 2: If $A = \Omega(\sqrt{L})$, then the above result is asymptotically the best possible.

Proof. Note that if $A = O(\sqrt{L})$, then $L/A = \Omega(\sqrt{L})$, in which case we're done, since by 158 we have a polynomial in $\mathbb{F}[X_1, X_2, X_3]$ of degree $O(\sqrt{L})$ whose zero set contains all lines in $\mathcal{L}$. Thus WLOG assume $A \geq 100\sqrt{L}$. Also set $D = 10^3 L/A$.
The key idea is that although 158 will only allow us to cover $O(D^2)$ lines with a polynomial of degree D, the prevalence of 2-rich points on other lines will 'force' them to lie in the zero set of f: More precisely, suppose f covers (i.e., contains in its zero set) a sub-collection $\mathcal{L}' \subseteq \mathcal{L}$ of lines, which we call 'RED'. All points on red lines are called red. Now, if some non-red line contains $D + 1$ red points, then that line also lies in the zero set of f!
To formalize the above idea, set $p := D^2/(100L) = 10^4 L/A^2$, and pick a p-random subset $\mathcal{L}'$ of $\mathcal{L}$. Then $\mathbb{E}|\mathcal{L}'| = D^2/100$, and by the Chernoff bound [3],

$$\Pr(|\mathcal{L}'| \geq D^2/50) \leq \exp\left(-\frac{D^2}{150}\right)$$

Thus, with probability $\geq 1 - \exp\left(-\frac{D^2}{150}\right)$, $|\mathcal{L}'| \leq D^2/50 \implies 7\sqrt{|\mathcal{L}'|} < D$. Consequently, by 158, there exists a non-zero polynomial $f \in \mathbb{F}[X_1, X_2, X_3]$ of degree $\leq 7\sqrt{|\mathcal{L}'|} < D$ such that $\bigcup_{\ell \in \mathcal{L}'} \ell \subseteq Z(f)$.
Now, for every $\ell \in \mathcal{L} \setminus \mathcal{L}'$, define:

$$\mathrm{RED}_\ell := \{x \in \ell : x \text{ is red}\}$$

Now, let x be a 2-rich point on ℓ. Then $x \in \ell_x$ for some $\ell_x \in \mathcal{L} \setminus \{\ell\}$. Furthermore, if $x \neq y \in \ell$, then $\ell_x \neq \ell_y$, since two distinct lines can have at most one point in common. Furthermore, ℓ_x is red with probability p. Thus,

$$\mathbb{E}|\,\mathrm{RED}_\ell\,| \geq Ap = 10^4 L/A$$

[3] we use the following version: If X is a binomial RV with mean μ, then $\Pr(X \geq (1+\delta)\mu) \leq \exp(-\delta^2\mu/(2+\delta))$

Once again, by Chernoff [4],

$$\Pr(|\operatorname{RED}_\ell| \geq 5000L/A) \geq 1 - \exp\left(-\frac{1250L}{A}\right)$$

Thus, by an union bound over $\mathcal{L} \setminus \mathcal{L}'$, we get:

$$\Pr(|\operatorname{RED}_\ell| \geq 5000L/A \text{ for all } \ell \in \mathcal{L} \setminus \mathcal{L}') \geq 1 - L\exp\left(-\frac{1250L}{A}\right)$$

Finally,

$$\Pr\left(|\mathcal{L}'| \leq D^2/50 \bigwedge |\operatorname{RED}_\ell| \geq 5000L/A \text{ for all } \ell \in \mathcal{L} \setminus \mathcal{L}'\right) \geq 1 - L\exp\left(-\frac{1250L}{A}\right) - \exp\left(-\frac{D^2}{150}\right)$$

$$> \frac{1}{2} - L\exp\left(-\frac{1250L}{A}\right)$$

where the last inequality holds for $D \gg 0$.
Thus, if $A < 1250L/\ln(2L)$, then with positive probability all lines in $\mathcal{L} \setminus \mathcal{L}'$ have $\geq 5000L/A > D + 1$ red points, and thus belong to the zero set of f, as desired.
For the remaining A, we're going to prove the statement by induction. Assume $A \geq 1000L/\ln(L)$ [5]. Let $\mathcal{L}', f$ be as defined above, and define $\mathcal{L}_2 := \{\ell \in \mathcal{L} : \ell \not\subset Z(f)\}$. Now, note that $\ell \not\subset Z(f)$ can only happen if $|\operatorname{RED}_\ell| \leq D$. Thus, (keeping in mind that $A \leq L$)

$$\Pr(\ell \not\subset Z(f)) \leq \Pr(|\operatorname{RED}_\ell| \leq D) = \Pr(|\operatorname{RED}_\ell| \leq 10^3 L/A) \overset{\text{Chernoff}}{\leq} \exp\left(-\frac{4050L}{A}\right)$$

$$\leq \exp(-4050) < \frac{1}{1000}$$

Thus $\mathbb{E}|\mathcal{L}_2| \leq L/1000$, and thus by Markov's inequality with probability ≥ 0.9, $|\mathcal{L}_2| \leq L/100$. Now, for each $\ell \in \mathcal{L}_2$, the number of 2-rich points arising solely due to intersection of lines within $\mathcal{L}_2$ is $\geq A - D$ (since the other red lines intersect any line in $\mathcal{L}_2$ at most D times). Thus by setting $L' := L/100, A' := A - D$ and invoking the induction hypothesis on $\mathcal{L}_2$, we get a function $f' \in \mathbb{F}[X_1, X_2, X_3]$ such that $\deg(f') \leq O(L'/A')$ and $\bigcup_{\ell \in \mathcal{L}_2} \ell \subseteq Z(f')$. But then note that $\bigcup_{\ell \in \mathcal{L}} \ell \subseteq Z(f \cdot f')$, and $\deg(f \cdot f') \leq \deg(f) + \deg(f') = O(L/A) + O(L'/A') = O(L/A) + O(L/A) = O(L/A)$, as desired. ■

[4] note that the the occurrence of each 2-rich point x in RED_ℓ is an independent Bernoulli random variable with probability p, once we have fixed a choice of ℓ_x. Also, we use the following version of Chernoff: If X is a binomial RV with mean μ, then $\Pr(X \leq (1-\delta)\mu) \leq \exp(-\delta^2\mu/2)$

[5] Note that $1000L/\ln(L) < 1250L/\ln(2L)$ for $L \gg 0$

14.3 A brief tour into results from Algebraic Geometry

Definition 161. *A variety(affine) over a field* $\mathbb{F}$ *defined by polynomials* $f_1, \ldots, f_s \in \mathbb{F}[X_1, \ldots, X_n]$ *is given as* $\mathbb{V} = V(f_1, \ldots, f_s) = \{\tilde{\mathbf{x}} \in \mathbb{F}^n : f_i(\tilde{\mathbf{x}}) = 0, i \in [s]\}$.

Definition 162. *A variety* $\mathbb{V} = V(f_1, \ldots, f_s), f_i \in \mathbb{F}^n, \forall i \in [s]$ *is said to have complexity at most* M *if* $s, n, \deg(f_i) \leq M, \forall i \in [s]$.

Definition 163 (Dimension of a Variety $\mathbb{V}$)**.** $\dim(\mathbb{V})$ *is defined as the longest chain of distinct nonempty (irreducible) subvarieties in* $\mathbb{V}$. *If* $\dim(\mathbb{V}) = r$, *then* $\mathbb{V}_0 \subsetneq \mathbb{V}_1 \subsetneq \mathbb{V}_2 \subsetneq \cdots \subsetneq \mathbb{V}_r = \mathbb{V}$. *We note that if* $|\mathbb{V}| = O_q(1)$, *then* $\dim(\mathbb{V}) = 0$.

Note: It should be mentioned that, any variety $\mathbb{V}$ corresponds to an ideal $\mathcal{I} \subseteq \mathbb{F}[X_1, \ldots, X_n]$, $\mathcal{I}(V) := \{f : f(x) = 0 \text{ for all } x \in \mathbb{V}\}$.
Note: $\mathcal{R} := \mathbb{F}[X_1, \ldots, X_n]/\mathcal{I}(V)$ is known as the co-ordinate ring of the variety $\mathbb{V}$.
But, what actually is $\dim(\mathbb{V})$**, anyway?**
For a $\mathbb{V} \in \mathbb{F}^n$, $\mathbb{V} = \langle f \rangle$ means that $\mathbb{V} = \{\tilde{\mathbf{x}} \in \mathbb{F}^n : f(\tilde{\mathbf{x}}) = 0\}$. Note that, $V(f)$ can be represented as some kind of a "surface" in $\mathbb{F}^n$. Dimension of a variety $\mathbb{V}$ is then, the minimum number of "co-ordinate functions" that will specify any point of the variety.
Now, we take a look at the "General Algebraic Geometry chart", which provides correspondence between algebraic and geometric interpretations.

Algebraic view	**Geometric view**
Ideals in $k[X_1, \ldots, X_n]$	subsets of k^n
$\cup\mid$	
Radical Ideals, $\mathcal{I} = \sqrt{\mathcal{I}}$	Affine varieties
$\cup\mid$	
Prime ideals	Irreducible varieties

Note: V is a reducible variety if $V = V_1 \cup V_2$, where V_1 and V_2 are non-trivial sub-varieties of V.

Theorem 164 (**Lang-Weil Bound**)**.** *Suppose* W *be a variety of complexity* M *defined over* $\overline{\mathbb{F}_q}$*(algebraic closure of* $\mathbb{F}_q$*). Then:*

- $|W(\mathbb{F}_q)| \leq O_M(q^{\dim(W)})$
- *If* f *is absolutely irreducible (i.e irreducible over the algebraic closure* $\overline{\mathbb{F}}_q$*, for example* $F(x, y, z) = x^3 + xy + z$ *is irreducible over both RED and* $\mathbb{C}$*), then* $|W(\mathbb{F}_q)| = q^{\dim(W)}(1 + O_M(\frac{1}{\sqrt{q}})) = q^{\dim(W)} + O_M(q^{\dim(W)-\frac{1}{2}})$

We now give the following proposition:

Proposition 165. W *be an affine variety over* $\overline{\mathbb{F}}_q$ *of complexity at most* M. $W \subset (\overline{\mathbb{F}}_q)^n, W = \langle f_1, \ldots, f_m \rangle$ *for* $f_i \in \mathbb{F}_q[X_1, \ldots, X_n]$ *Then,* $\exists$ *absolutely irreducible varieties* $Y_1, \ldots, Y_s$ *such that* $W(\mathbb{F}_q) = \bigcup_{i=1}^{s} Y_i(\mathbb{F}_q)$. *Moreover, both* s *and the complexity of* $Y_i, \forall i \in [s]$ *is* $O_M(1)$.

Proof. Write W as the union of its $\mathbb{F}_q$-irreducible components, let $W(\mathbb{F}_q) = X_1 \cup X_2 \cup \cdots \cup X_r$, where each X_i is $\mathbb{F}_q$ irreducible variety. Pick such a $\mathbb{F}_q$ irred. component X. If X is not absolutely ireducible, let $Y_1, \ldots, Y_t$ be the absolutely irreducible components, i.e $X(\overline{\mathbb{F}}_q) = \bigcup_{i=1}^{t} Y_i$. Now, the Frobenius map $\phi : \overline{\mathbb{F}}_q \to \overline{\mathbb{F}}_q : x \mapsto x^q$ preserves each of the $f_i \in \mathbb{F}[X_1, \ldots, X_n]$ that defines W, hence preserves the expressions $W(\mathbb{F})$ and $W(\overline{\mathbb{F}}_q)$ as well. (Note: $\phi : \overline{\mathbb{F}}_q \to \overline{\mathbb{F}}_q$ fixes elements of $\mathbb{F}_q$ pointwise, and only those of $\mathbb{F}_q$).

Now, we observe that, $\phi(Y_i) = Y_j$ for some j(this follows by absolute irreducibility, otherwise if there was some intersection or $\phi(Y_i) \subsetneq Y_j$, then the Y_i's won't be absolutely irreducible). Also note that, ϕ acts transitively on the Y_i's. If not, we take an orbit $\mathcal{O}$ of this action. But then, $\bigcup_{Y_j \in \mathcal{O}} Y_j$ is fixed by ϕ(by the definition of orbit under the action of ϕ). Since this is fixed by ϕ, we must have $\bigcup_{Y_j \in \mathcal{O}} Y_j \in \mathbb{F}_q$, which contradicts the irreducibility of X, unless the orbit includes all the Y_i's that constitute X.

Since $X \in \mathbb{F}_q$ is fixed by ϕ, by the previous argument, we have $X(\mathbb{F}_q) \subseteq Y_1(\mathbb{F}_q) \cap Y_2(\mathbb{F}_q) \cap \cdots \cap Y_t(\mathbb{F}_q)$. Hence, we may thus replace $X(\mathbb{F}_q)$ by $\bigcap_i Y_i(\mathbb{F}_q)$. But, $\bigcap_i Y_i(\mathbb{F}_q)$ has lower dimension than $W(\mathbb{F}_q)$. Thus, we may iterate this process again by now starting with $\bigcap_i Y_i(\mathbb{F}_q)$, and moreover the initial complexity of $W(\mathbb{F}_q)$, and each of $Y_i(\mathbb{F}_q)$ is $O_M(1)$. Thus, this process must necessarily terminate in $O_M(1)$ steps, and hence the statement of the proposition follows. ■

The Lang -Weil bound together with the above proposition gives the following dichotomy theorem:

Theorem 166. *Suppose, W and D are varieties defined over $\mathbb{F}_q$ of complexity at most M. Then, if $q \gg_M 0$ then one of the following holds:*

1. $|W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| \leq c$, *where c depends only on M*
2. $|W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| \geq \frac{q}{2}$

Proof. The proof requires the following fact from algebraic geometry:
Fact: Suppose, W is absolutely irreducible and defined over $\mathbb{F}_q$, and suppose $\dim(W) \geq 1$. Then, for any other variety D, wither $W \subseteq D$ or $W \cap D$ as an $\mathbb{F}_q$ variety has lower dimension than $W(\mathbb{F}_q)$
Now, following 165, we write $W(\mathbb{F}_q)$ as union of absolutely irreducible varieties: $W(\mathbb{F}_q) = \bigcup_i Y_i(\mathbb{F}_q)$, where $Y_i(\mathbb{F}_q)$'s are absolutely irreducible of complexity $O_M(1)$. Now, suppose $\dim(Y_i) \geq 1$ for some i. By 164, we get that $|Y_i(\mathbb{F}_q)| = q^{\dim(Y_i)} - O_M(q^{\dim(Y_i)-\frac{1}{2}})$. By the fact above, $Y_i \subseteq D$, or $\dim(Y_i \cap D) \leq \dim(Y_i) - 1$. We now consider the following cases:
Case-1: If $\dim(Y_i \cap D) \leq \dim(Y_i) - 1$ (i.e $Y_i \not\subseteq D$), then:

$$\begin{aligned} |Y_i(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| &\geq q^{dim(Y_i)} - O_M(q^{dim(Y_i)-\frac{1}{2}}) - O_M(q^{dim(Y_i)-1}) \\ &\geq \frac{q}{2}, \forall q \gg 0 \end{aligned} \tag{14.1}$$

Since $Y_i(\mathbb{F}_q) \subseteq W(\mathbb{F}_q)$, we have that, $|W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| \geq |Y_i(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| \geq \frac{q}{2}$, and statement-2 of the theorem follows.

Case-2: If $Y_i \subseteq D$, the contribution of Y_i to $W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)$ is nill. Since complexity of W is at most M, if all Y_i's fall in this case, statement-1 of the theorem would follow. Thus, proof of the theorem is complete. ■

14.4 Bukh's construction for $\mathrm{ex}(n; K_{s,t})$ for $t \gg_s 0$

What we really seek is a distribution (i.e. a random graph model) that does not admit a slowly vanishing tail distribution. The connection with algebraic structures arises thus: *Since geometric objects often display some forms of rigidity, we need a geometric perspective in the required random model.* We illustrate this point with a simple example: If $f(\mathbf{x})$ is a polynomial of degree d in n variables, and suppose f vanishes on more than d points on some line (over some field $\mathbb{F}$) then f vanishes on the whole of the line. This sharp dichotomy - either the polynomial has at most d zeroes on the line or vanishes identically on the line - is *an algebraic consequence.*

Let us explore this a bit more, with an eye on the lower bound for $\mathrm{ex}(n; K_{s,t})$. Incidentally, the lower bound asymptotically matches the upper bound (at least, the exponent) for $t > (s-1)!$, and that is a result of Alon-Ronyái-Szabo, which incidentally comes from what are called 'norm graphs'.

Suppose $f(\mathbf{X}, \mathbf{Y})$ be a polynomial in $2s$ variables over $\mathbb{F}_q$. Since polynomials also describe functions, we may think of f as a function $f : \mathbb{F}_q^s \times \mathbb{F}_q^s \to \mathbb{F}_q$. Define the bipartite graph $G_f = (L, R, E)$ with $L = R = \mathbb{F}_q^s$, and $(u, v) \in L \times R$ is an edge in G_f iff $f(u, v) = 0$. In other words, the affine variety defined by f describes the graph G_f.

We now turn to note some facts from Algebraic Geometry. Some of these results are simple (not necessarily elementary!) but some other facts (the Lang-Weil bound) are quite non-trivial and we shall not get into their proofs.

Now we are in a position to see some applications of this dichotomy theorem.

Theorem 167 (Algebraic Dichotomy)**.** *Let q be a prime power and let W, D be affine varieties defined over $\mathbb{F}_q$ having complexity at most M. Then, if $q \gg_M 0$, exactly one of the following holds:*

1. $|W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| \leq C = C(M)$.
2. $|W(\mathbb{F}_q) \setminus D(\mathbb{F}_q)| \geq q/2$.

Now, we shall see some applications of the above theorem in extremal combinatorics. For a positive integer n and a graph H on $m \leq n$ vertices, let $\mathrm{ex}(n, H)$ denote the maximum number of edges an n-vertex graph can have so that it contains no copy of H. A classic extremal problem posed by Zarankiewicz is to compute $\mathrm{ex}(n, K_{s,t})$ where $K_{s,t}$ denotes the complete bipartite graph where one part has s vertices and the other has t.

Theorem 168 (Kövari-Sós-Turán). *Given n and $s \leq t$,*

$$\mathrm{ex}(n, K_{s,t}) \leq O_{s,t}(n^{2-\frac{1}{s}}).$$

Although the above upper bound has been known for more than 60 years now, we do not yet know whether or not the widely believed and conjectured lower bound

$$\mathrm{ex}(n, K_{s,t}) = \Omega_{s,t}(n^{2-\frac{1}{s}}) \tag{14.2}$$

is true. However, there exist proofs for this in some very special cases. One such case is when $(s,t) \in \{(2,2),(2,3),(3,3)\}$. Another was proved in 1999 by [**?**] and is stated below.

Theorem 169 (Alon-Rónyai-Szabó). *If $t \geq (s-1)! + 1$, then 14.2 holds.*

The proof of this result goes through some highly involved geometric constructions that use '*projective norm graphs*' which we shall not address here. We instead study the following result from [**?**] which uses the probabilistic method and 167 to give a much simpler proof of a similar result.

Theorem 170 (Bukh). *Given $s \in \mathcal{N}$, there exists a constant $C = C(s)$ such that for any $t > C$, $\mathrm{ex}(n, K_{s,t}) = \Omega_{s,t}(n^{2-\frac{1}{s}})$.*

Remark: In the process of simplifying the proof, Bukh loses the precision that comes with explicitly stating $C(s)$ in the statement (as a matter of fact, this constant is the C from the statement of 167), but as long as it is a constant depending only on s, this is not an issue.

Now, we proceed with the proof of 170. Let q be a prime power. Construct a graph $G = G(L \cup R, E)$ with $L = R = \mathbb{F}_q^s$. Let the number of vertices in this graph be $N = \Theta(q^s)$. For a polynomial $p \in \mathbb{F}_q[X_1, X_2, \ldots X_s, Y_1, Y_2, \ldots, Y_s]$, the edges of $G = G_p$ are defined as $(\tilde{u}, \tilde{v}) \in E$ if and only if $p(\tilde{u}, \tilde{v}) = 0$. Pick a uniformly random polynomial of degree at most $d = d(s)$ (d will be fixed later). By this, we mean that the coefficients of p are picked uniformly at random and independently from $\mathbb{F}_q$. Let

$$p = \sum_{\tilde{\alpha},\tilde{\beta},|\alpha|+|\beta|\leq d} a_{\tilde{\alpha},\tilde{\beta}} X_1^{\alpha_1} \cdots X_s^{\alpha_s} \cdot Y_1^{\beta_1} \cdots Y_s^{\beta_s}$$

We want to estimate the number of edges in this graph G_p. To do so, we try and compute

$$\mathcal{E}(e(G_p)) = \sum_{\tilde{u},\tilde{v} \in \mathbb{F}_q^s} \Pr[p(\tilde{u}, \tilde{v}) = 0].$$

Proposition 171. *For any $\tilde{u}, \tilde{v} \in \mathbb{F}_q^t$, $\Pr[p(\tilde{u}, \tilde{v}) = 0] = 1/q$.*

Proof. Sampling a random polynomial (satisfying the degree bound) can be thought of as sampling a random polynomial $p = p_0 + \alpha$ where p_0 is such that $p_0(\tilde{0}, \tilde{0}) = 0$ (in other words, p_0 does not have a constant term). Now, once such a p_0 has been sampled, there is exactly one choice of $\alpha \in \mathbb{F}_q$ for which the polynomial vanishes, completing the proof. ■

So, we have

$$\mathbb{E}(e(G_p)) = \frac{1}{q} \cdot q^{2s} = q^{2s-1} = \Omega(N^{2-\frac{1}{s}}),$$

which matches the extremal number we are going for. We are still left with the task to verify that this graph does not contain a $K_{s,t}$.

Let $U \subseteq L$ be a fixed set of s vertices. Define the common neighbour set $N(U) = \{\tilde{v} \in R : p(\tilde{u}, \tilde{v}) = 0 \text{ for all } \tilde{u} \in U\}$. We want to calculate the expected size of this set $N(U)$. Observe that if $|N(U)| \geq t$ for some integer t, this implies the existence of a $K_{s,t}$ inside the graph G. Now,

$$|N(U)| = \sum_{\tilde{v} \in \mathbb{F}_q^s} \mathbf{1}_{p(\tilde{u}_i, \tilde{v}) = 0 \ \forall i \in [s]} \tag{14.3}$$

$$\implies |N(U)|^r = \sum_{\tilde{v}_1, \ldots, \tilde{v}_r \in \mathbb{F}_q^s} \mathbf{1}_{p(\tilde{u}_i, \tilde{v}_j) = 0 \ \forall i \in [s], j \in [r]} \quad \text{(for some integer } r\text{)} \tag{14.4}$$

$$\implies \mathbb{E}|N(U)|^r = \sum_{\tilde{v}_1, \ldots, \tilde{v}_r \in \mathbb{F}_q^s} \Pr[p(\tilde{u}_i, \tilde{v}_j) = 0 \ \forall i \in [s], j \in [r]]. \tag{14.5}$$

Assume for the time being that the events $p(\tilde{u}_i, \tilde{v}_j) = 0$ are jointly independent, then the above probability ($\Pr[p(\tilde{u}_i, \tilde{v}_j) = 0 \ \forall i \in [s], j \in [r]]$) is q^{-ls}, where l is the number of distinct v_js.
Now, we try and view $\mathbb{E}|N(U)|$ from a different lens. For a fixed U (as above), let $f_i(Y_1, \ldots, Y_s) = p(\tilde{u}_i, Y_1, \ldots, Y_s)$ and W be the variety $\langle f_1, \ldots, f_s \rangle$. This $W = W(\mathbb{F}_q)$ is a variety defined over $\mathbb{F}_q$ of complexity at most $M = M(s)$. Then, by 167, either

1. $|W| \leq C = C(s)$ or
2. $|W| \geq q/2$.

For $V = \{\tilde{v}_1 \ldots, \tilde{v}_r\}$ as in 14.4 (a set where $p(\tilde{u}_i, \tilde{v}_j) = 0 \ \forall i \in [s], j \in [r]$), clearly $V \subseteq W$. Now, we use a neat trick which follows as a simple application of Markov's inequality. For a suitably chosen $\lambda > 0$,

$$\mathbb{P}(|N(U)| \geq \lambda) = \Pr(|N(U)|^r \geq \lambda^r) \leq \frac{\mathbb{E}|N(U)|^r}{\lambda^r}.$$

Setting $\lambda = C(s)$ as in our formulation of the algebraic dichotomy, this gives us that

$$\mathbb{P}[|N(U)| > C(s)] = \mathbb{P}[|N(U)| \geq q/2] \leq \frac{\mathbb{E}|N(U)|^r}{(q/2)^r}. \tag{14.6}$$

Fix $1 \leq l \leq r$ and let $V = \{\tilde{v}_1, \ldots, \tilde{v}_l\}$. Modulo the assumption that the events $\{p(\tilde{u}_i, \tilde{v}_j) = 0\}_{i,j}$ are jointly independent, we have that $\mathbb{P}[p(\tilde{u}_i, \tilde{v}_j) = 0 \ \forall i \in [s], j \in [l]] = q^{-ls}$. Now, returning to 14.5, we have

$$\mathbb{E}|N(U)|^r \leq \sum_{l=1}^{r} \binom{N}{l} M_{l,r} \frac{1}{q^{ls}} \leq \sum_{l=1}^{r} \frac{M_{l,r}}{l!} =: M_{s,r}$$

where $M_{l,r}$ denotes the number of surjections from $[r] \to [l]$. Thus $\mathbb{E}|N(U)|^r \leq M_{s,r}$ (this is fine as $r = r(s)$), and so from 14.6, we have

$$\mathbb{P}[|N(U)| > C] \leq \frac{2^r M_{s,r}}{q^r} \quad \text{(where } C = C(s) \text{ is the constant from the dichotomy threshold).}$$

Call a set $U \subseteq L$ of size s 'BAD' if $|N(U)| > C$.

$$\mathbb{E}(\text{number of BAD sets}) \leq \binom{q^s}{s} \cdot \frac{2^r M_{s,r}}{q^r} \leq \frac{q^{s^2-r} M_{s,r} 2^r}{s!} =: B$$

So, there exists a polynomial p such that G_p has at most $B = O_{s,r}(q^{s^2-r})$ BAD subsets(of L) of vertices. From each BAD subset of such a graph, delete one vertex, this leaves the remaining graph with no BAD subsets. Additionally, such a graph will have at least $q^{2s-1} - O_{s,r}(q^{s^2-r}) \cdot q^s$ edges[6].
Now, choosing $s^2 + s - r < 2s - 1$ is sufficient to achieve the desired bound. For simplicity, set $s^2 + s - r = 2s - 2$, that is $r = s^2 - s + 2$.

Now, it remains to choose the degree $d = d(s)$ suitably and argue the joint independence of the events $\{p(\tilde{u}_i, \tilde{v}_j) = 0\}_{i,j}$. The following lemma shows that $d = rs = s(s^2 - s + 2)$ is enough:

Lemma 172. *Suppose f is a uniformly randomly chosen polynomial of degree at most d in $\mathbb{F}_q[X_1, \ldots, X_t]$, and suppose $\tilde{y}_1, \ldots, \tilde{y}_m$ are distinct points in $\mathbb{F}_q^t$ such that $d \geq m - 1$ and $q > \binom{m}{2}$. Then*

$$\mathbb{P}[f(\tilde{y}_i) = 0 \ \forall i] = \frac{1}{q^m}.$$

Proof. Suppose first that $y_{1,1}, y_{2,1}, \ldots, y_{m,1}$ are pairwise distinct (i.e., the first coordinates are pairwise distinct). Write $f(\tilde{X}) = g(\tilde{X}) + h(\tilde{X})$, where $g(\tilde{X}) = \sum_{i=0}^{m-1} a_i X_1^i$ consists of all terms with only X_1, and $h(\tilde{X})$ consists of all other terms. Hence, if coefficients of h are fixed, then by Lagrange interpolation, there is a unique value of a_i that satisfies $f(\tilde{y}_j) = 0$ for all j. Since there are m coefficients that are fixed, the probability of a random f satisfying the condition is $\frac{1}{q^m}$.

[6]This can be argued by viewing $[e(G) - (\text{number of BAD subsets})]$ as a random variable and computing its expectation.

Now consider the general case. The idea is to find an invertible linear transform T such that $(T\tilde{y}_i)_1$ are pairwise distinct. Then, f chosen uniformly at random is equivalent to $f_1 = f \odot T^{-1}$ chosen uniformly at random, and $f_1(T\tilde{y}_i) = f(\tilde{y}_i)$, so we'll be done by the first part.

To that end, consider $\alpha_i \in \mathbb{F}_q$ to be chosen later, and let $(T\tilde{X})_1 = X_1 + \sum_{i=2}^{t} \alpha_i X_i$ and $(T\tilde{X})_j = X_j$ for all $j \geq 2$. Since determinant of T is 1, T is invertible.

If $(T\tilde{y}_i)_1 = (T\tilde{y}_j)_1$, then

$$y_{i,1} + \sum_{l=2}^{t} \alpha_l y_{i,l} = y_{j,1} + \sum_{l=2}^{t} \alpha_l y_{j,l}.$$

Note that if $y_{i,l} = y_{j,l}$ for all $l \geq 2$, then we get $y_{i,1} = y_{j,1}$, which is impossible since $\tilde{y}_i \neq \tilde{y}_j$. Hence $y_{i,l_0} \neq y_{j,l_0}$ for some $l_0 \geq 2$. So if α_l is fixed for $l \neq l_0$, then there is a unique value of l_0 that makes $(T\tilde{y}_i)_1 = (T\tilde{y}_j)_1$. Therefore there are at most q^{t-2} tuples $(\alpha_2, \ldots, \alpha_t)$ for which $(T\tilde{y}_i)_1 = (T\tilde{y}_j)_1$. Summing over all pairs i, j, the number of bad tuples is at most $\binom{m}{2} q^{t-2} < q^{t-1}$, so there is a tuple $(\alpha_2, \ldots, \alpha_t)$ for which $(T\tilde{y}_i)_1$ are pairwise distinct, as required. ■

Bibliography

[1] M Aigner, G. M. Ziegler, *Proofs from The Book*, Springer.

[2] N. Alon, D. J. Kleitman, Sum-free subsets, *A tribute to Paul Erdős (A. Baker, B. Bollobás, A. Hajnál eds)*, Cambridge University Press, pp. 13-26.

[3] N. Alon, Y. Matias, M. Szegedy, The space complexity of approximating the frequency moments, *J. Comp. Sys. Sci.* **58** (1999) (1), 137-147

[4] N. Alon, Y. Peres, Uniform Dilations, *Geom. Func. Anal.* **2** (1992), No. 1, 1-28.

[5] N. Alon, J. H. Spencer, *The Probabilistic Method,* 4^{th} *ed.*, Wiley International, 2016.

[6] B. Bollobás. *Random Graphs*,

[7] B. Bukh, Random algebraic construction of extremal graphs, *Bull. Lond. Math. Soc.* **47** (2015), no. 6, 939–945.

[8] B. Bukh, D. Conlon, Rational exponents in extremal graph theory, *J. Eur. Math. Soc. (JEMS)*, **20** (2018), no. 7, 1747-1757.

[9] N. Balachandran, S. Padinhatteeri, $\chi_D(G), |Aut(G)|$, and a variant of the motion lemma, *Ars Math. Contemp.* **12** (2017), no. 1, 89-109.

[10] N. Balachandran, E. Mazumdar, Zero sums in restricted sequences.

[11] N. Bansal,

[12] J. Beck,

[13] D. D. Cherkashin, J. Kozik, A note on random greedy coloring of uniform hypergraphs, *Random Struct. Algorithms.*

[14] S. Eberhart, B. Green, F. Manners, Sets of integers with no large sum-free subset, *Ann. Math.* **180**(2014), 621-652.

[15] W. T. Gowers, Lower Bounds of Tower Type for Szemerédi's Uniformity Lemma, *GAFA, Geom. Funct. Anal.*, **7**(1997), 322-337.

[16] Hod, A. Ferber, M. Krivelevich, B Sudakov,

[17] D.R. Hughes and F. C. Piper, *Projective Planes*, Graduate Texts in Mathematics 6, Springer-Verlag, New York, 1973.

[18] S. Janson, Łuczak, D. Rucinski, *Random Graphs*

[19] P. Keevash, The existence of Steiner designs.

[20] A. Kostockha, B. Sudakov,

[21] F. Lazebnik, V.A. Ustimenko and A.J. Woldar, A New Series of Dense Graphs of High Girth', *Bulletin of the AMS*, **32**(1995), Number 1, 73–79.

[22] A. Lubotzky, Phillips, P. Sarnak,

[23] Molloy, B. Reed, *Graph Colouring and the Probabilistic Method*,

[24] N. Pippenger, J.H. Spencer,

[25] J. Radhakrishnan, A. Srinivasan,

[26] N. Robertson, P. Seymour,

[27] T. Rothvoß,

[28] Shafarevich, *Basic algebraic geometry*, Springer-Verlag, New York-Heidelberg, 1974.

[29] J. H. Spencer, *Ten lectures on the Probabilistic Method*

[30] J. H. Spencer (with L. Florescu), *Asymptopia*, Student Mathematical Library, Volume 71, American Mathematical Society, 2014.

[31] S. Venkatesh, *The Theory of Probability*, Cambridge University Press, 2013.

[32] D. B. West, *Introduction to Graph Theory*,

[33] A. Wigderson, *Mathematics and Computation*, Princeton University Press, 2019.

[34] R. M. Wilson,

[35] R. M. Wilson,

[36] R. M. Wilson,