THE PROBABILISTIC METHOD IN COMBINATORICS

Lectures by Niranjan Balachandran.

Contents

Preface

These notes arose as a result of lectures I gave, starting with a course at Cal Tech in Spring, 2010, and 2011, and later with lectures at IIT Bombay. The material is very much restricted to material that interested me, to start with. The Probabilistic Method has now become one of the most important and indispensable tools for the Combinatorist. There have been several hundred papers written which employ probabilistic ideas and some wonderful monographs - the ones by Alon-Spencer, Spencer, Bollobás, Janson *et al*, and the more recent one by Molloy-Reed, spring readily to mind.

One of the main reasons this technique seems ubiquitous is because it provides a simpler tool to deal with the 'local-global' problem. More specifically, most 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 fundamental problem of design theory, for instance, considers the 'local' condition at every t-element subset of a given set. The difficulty of a combinatorial problem is to ensure this property of 'locally good' everywhere. The probabilistic paradigm considers all these 'local' information simultaneously, and provides what one could call 'weaker' conditions for building a global patch from all the local data. Over the past 2 decades, the explosion of research material, along with the wide array of very impressive results demonstrates another important aspect of the Probabilistic method; some of the techniques involved are subtle, one needs to know *how to use those tools, more so than simply understand the theoretical underpinnings.*

Keeping that in mind, I decided to emphasize more on the *methods* involved. Another very important feature of probabilistic arguments is that they sometimes seem to crop up in situations that do not (outwardly) seem to involve any probability arguments *per se.* Most of the material covered in these notes primarily discuss combinatorial results which do not involve the notions of probability at the outset. Hence, gadgets such as Random graphs appear as tools required to prove/disprove a deterministic statement; these notes do not study random graphs and related phenomena such as threshold functions, and the Erdős-Rényi phase transitions.

One possible difference between these notes and existing literature on this subject is in the treatment of the material. I have attempted to address some well-known results (following probabilistic methods) along the lines of what Tim Gowers calls 'The Exposition Problem'. Many results that prove or disprove a (probably hastily stated) conjecture often evoke a sense of awe at the ad-hoc nature of the proof and the brilliance of the solver. One thing I wished to outline in my lectures was a sense of 'naturality' in thinking of such a proof. I do not claim that my deconstruction explains the inner workings of some brilliant minds; rather the emphasis is upon the notion,"*What if I had to come up with this result? What would be a 'valid' point of view to address this problem?*

Another aspect of the Probabilistic Method is that these techniques often end up proving *asymptotic results* rather than exact ones, and so, when one adopts the probabilistic paradigm, one is essentially looking for the *best possible result* through such a line of argument. Hence, the *precise nature of the result proven is only known once the result is obtained.* While this is true of all of mathematical research, the very nature of some of these results using probabilistic techniques, makes this a stand-out feature. In trying to pass on the spirit of the principle, I worked most of these results *backwards.* i.e., trying to work with certain ideas, and then shave off the best asymptotics one can manage under such considerations. This approach is not completely new in these notes. The monograph by Spencer (which is again a transcript of lectures) too does something similar. But these notes contain certain results which have not appeared outside of the original paper, in a deconstructed form. So at least that much is new.

I thank all my students of this course who very actively and enthusiastically acted as scribes for the lectures for this course over the two years.

1 The Probabilistic Method: Some First Examples

1.1 Lower Bounds on the Ramsey Number $R(n, n)$

Ramsey theory, roughly stated, is the study of how "order" grows in systems as their size increases. In the language of graph theory, the central result of Ramsey theory is the following:

Theorem 1. *(Ramsey, Erdös-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 K_n's edges must yield a red K_s or a blue K_t.*

A fairly simple recursive upper bound on $R(s, t)$ 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

$$R(s, s) \leq 2^{2s} \cdot s^{-1/2}.$$

A constructive lower bound on $R(s, s)$, discovered by Nagy, 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.)

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? Erdös, in 1947, introduced probabilistic methods in his paper "Some Remarks on the Theory of Graphs" to answer this very question:

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

Proof. Fix some value of n, and consider a random uniformly-chosen 2-coloring of K_n's edges: in other words, let us work in the probability space $(\Omega, Pr) =$ (all 2-colorings of K_n's edges, $Pr(\omega) = 1/2^{\binom{n}{2}}$.)

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 monochrome. 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 occuring is bounded by

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

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 will be some 2-coloring $\omega \in \Omega$ in which none of the 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.

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

1.2 Tournaments and the S_k Property

Definition 3. *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.*

Definition 4. *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 natural question to ask about the S_k property is the following:

Question 5. *How small can a tournament be if it satisfies the S_k property, for some k?*

We can calculate values of S_k for the first three values by hand:

- 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; as well, constructive methods have been fairly bad at finding asymptotics for how these values grow. Probabilistic methods, however, give us the following useful bound:

Proposition 6. *(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, take some graph K_n, and for every edge (i, j) direct the edge $i \to j$ with probability $1/2$ and from $j \to i$ with probability $1/2$.

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 $1/2^k$.

Consequently, the probability that v fails to have a directed edge to each member of S is $1 - 1/2^k$. For different vertices, these events are all independent; 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 a naïve union upper bound, we have that

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

Thus, it suffices to force the right-hand side to be less than 1, as this means that there is at least one graph on which no such subsets S exist – i.e. that there is a graph that satisfies the S_k property.

So, using the approximation $\binom{n}{k} \cdot \left(1 - 1/2^k\right)^{n-k} \le \left(\frac{en}{k}\right)^k$, 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. $\square$

1.3 Dominating Sets

Definition 7. *Let $G = (V, E)$ be a graph. A set of vertices $D \subseteq V$ is called **dominating** with respect to G if every vertex in $V \setminus D$ is adjacent to a vertex in D.*

Theorem 8. *Suppose that $G = (V, E)$ is a graph with n vertices, and that $\delta(G) = \delta$, the minimum degree amongst G's vertices, is strictly positive. Then G contains a dominating set of size less than or equal to*

$$\frac{n \cdot (1 + \log(1 + \delta))}{1 + \delta}$$

Proof. Create a subset of G's vertices by choosing each $v \in V$ independently with probability p; call this subset X. Let Y be the collection of vertices in $V \setminus X$ without any neighbors in X; then, by definition, $X \cup Y$ is a dominating set for G.

What is the expected size of $|X \cup Y|$? Well; because they are disjoint subsets, we can calculate $|X \cup Y|$ by simply adding $|X|$ to $|Y|$:

$$\begin{aligned}
\mathbb{E}(|X|) &= \sum_{v \in V} \mathbb{E}\left(\mathbb{1}_{\{v \text{ is chosen}\}}\right) \\
&= p \cdot n, \text{ while} \\
\mathbb{E}(|Y|) &= \sum_{v \in V} \mathbb{E}\left(\mathbb{1}_{\{v \text{ is in } Y\}}\right) \\
&= \sum_{v \in V} \mathbb{E}\left(\mathbb{1}_{\{v \text{ isn't in } X, \text{ nor are any of its neighbors}\}}\right) \\
&= \sum_{v \in V} \mathbb{E}\left(1 - p\right)^{\deg(v)+1}, \text{(b/c we've made } \deg(v) + 1 \text{ choices independently)} \\
&\leq \sum_{v \in V} (1 - p)^{\delta+1} \\
&= n(1 - p)^{\delta+1}.
\end{aligned}$$

This tells us that

$$\begin{aligned}
\mathbb{E}(|X \cup Y|) &\leq np + n(1 - p)^{\delta+1} \\
&\leq np + ne^{-p(\delta+1)},
\end{aligned}$$

which has a minimum at

$$p = \frac{\log(1+\delta)}{1+\delta}.$$

Thus, for such p, we can find a dominating set of size at most

$$\frac{n \cdot (1 + \log(1+\delta))}{1+\delta},$$

as claimed. □

1.4 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 gave (at around the same time) 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

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

A similar ratio, namely,

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

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

Question 9. *Suppose $T = (x_1, x_2, \dots, x_m)$ is a sequence of (not necessarily distinct) reals, and $T_r := \{(i,j) : |x_i - x_j| \le r\}$. Is it true that $|T_r| \le (2r-1)|T_1|$?*

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 $|x_i - x_j| \le r$. Suppose i was chosen such that $|S_I|$ is maximum among all choices for x_i. Then observe that if we partition

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

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

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

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 $X_1, X_2, \dots, X_m$ from the distribution F, and define the random variables $T_1 := \big|\{(i,j) : |X_i - X_j| \le 1\}\big|$ and $T_r := \big|\{(i,j) : |X_i - X_j| \le r\}\big|$, then note that

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

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

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

with $p_r = \mathbb{P}(|X_i - X_j| \le r)$. 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 $p_r(a) = \mathbb{P}(|X - a| \le r)$ 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 $p_r^{(G)} \le (2r - 1)p_1^{(G)}$. Here $p_r^{(G)}$ denotes the probability $p_r = \mathbb{P}(|X - Y| \le r)$ 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

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

which holds if and only if $\alpha \ge \frac{\beta}{r-1+\beta}$ with $\beta = p_r(a) - (2r - 1)p_1(a) > 0$. But since α is our choice, picking α smaller than this bound yields a contradiction.

1.5 Sum-Free Sets of Integers

This is another gem originally due to Erdős. Another interpretation of this was due to Alon and Kleitman.

Proposition 10. *Every set of $B = \{b_1, \dots b_n\}$ of n nonzero integers contains a sum-free[1] subset of size $\ge n/3$.*

[1]A set $B \subset \mathbb{R}$ is called sum-free if the sum of any two elements in B does not lie in B.

Proof. 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 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: using the probabilistic method, we 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.

□

2 The Linearity of Expectation and small tweaks

One of the most useful tools in the Probabilistic method is the Expectation. One of the prime 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 a good idea of where a 'typical' observation of the random variable lies. And that is often a very useful start. Later, we will see results where certain random variables tend to concentrate around the expectation. Computationally, this is a very useful tool since it is easy to calculate. In the examples that we shall consider here, the linearity of expectation will turn out to be key when it comes to calculations and estimations.

2.1 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 calculations from setting $m = n/2$ gives us $R(k, k) > \frac{(1+o(1))k2^{k/2}}{e}$ for large k.

2.2 List Chromatic Number and minimum degree

Definition 11. *Let G be a graph, and let $\mathcal{L} = \{L_v | v \in V(G)\}$, where L_v is a set of colors for vertex $v \in G$. An $\mathcal{L}$-coloring of G is an assignment of a color in L_v to v for each $v \in G$, such that no two adjacent vertices are assigned the same color.*

Definition 12. *Let G be a graph. The list chromatic number of G, denoted $\chi_l(G)$, is the smallest k such that there exists $\mathcal{L}$ with $|L_v| \geq k$ and G is $\mathcal{L}$-colorable.*

If $\chi(G)$ is the chromatic number of G, then $\chi(G) \leq \chi_l(G)$ by taking $\mathcal{L}$ as $L_v = [\chi(G)]$ for all $v \in G$. The next result shows that the reverse inequality need not hold.

Theorem 13 (Erdős, Rubin, Taylor, 1978). *$\chi_l(K_{n,n}) > k$ if $n \geq \binom{2k-1}{k}$.*

Proof. We want 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. □

Now we state and prove another result due to Alon that provides a bound on the growth of $\chi_l(G)$. But first we introduce some notation and a lemma.

Definition 14. *We say "f is Big Omega of g" and write $f = \Omega(g)$ if there is a constant $K > 0$ such that for n sufficiently large, $|f(n)| \geq K|g(n)|$.*

Lemma 15. *For any graph G, there exists a subgraph H of G with $V(H) = V(G)$ such that H is bipartite and $\deg_H(v) \geq \frac{1}{2}\deg_G(v)$ for all $v \in V(G)$.*

Proof. Partition G into two subsets A_0 and B_0, and let J_0 denote the corresponding bipartite subgraph of G, i.e. the subgraph containing all vertices of G and edges between vertices in different partition classes. Pick any vertex $v \in G$ with $\deg_{J_0}(v) < \frac{1}{2}\deg_G(v)$, and move v to the other partition class of J_0. Let A_1 and B_1 be the resulting partition classes and J_1 the corresponding bipartite subgraph, and repeat this process.

Since for each i, J_{i+1} has strictly more edges than J_i, this process eventually terminates at some finite stage n, and the resulting graph $H = J_n$ is a bipartite subgraph with $\deg_H(v) \geq \frac{1}{2}\deg_G(v)$ for all $v \in V(G)$. □

Theorem 16 (Alon).

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

where $d = \delta(G)$.

Proof. By the previous lemma, we can assume without loss of generality that G is bipartite with partition classes A and B, and $|A| \geq |B|$.

Before we get into the proof, let us, perhaps vaguely, outline our scheme of proof. We may and will assume that the minimum degree is sufficiently large, and the final calculation will show how large we need it to be. We will try to assign lists of size s to each vertex and ensure that from these lists, a list coloring is not possible. Suppose C is the set of colors from which we shall allocate lists to each of the vertices of G.

Firstly, if each vertex a in A say has, among the lists of its neighbors in B, all the possible s-subsets of C then in particular, for any choice of colors assigned to the vertices of B, since we end up picking one from each s-subset of C, it follows that in order that the particular choices from B extend successfully to a choice for this vertex a, then there must be at least one color in the list of a that is picked from a very small set (namely, some set of at most $s-1$ colors of C. Now, if there are several such vertices like 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. Now, if we were to allot random lists to the vertices of A, then it becomes extremely unlikely that each of these lists actually intersects a very small set non trivially. This potentially sets up a contradiction.

Let the set of available colors be $C := \{1, 2, ..., L\}$. We want to show that there is a collection of lists $\mathcal{L} = \{L_v\}$, $|\{L_v\}| = s$, for which there is no $\mathcal{L}$-coloring of G, given certain conditions on the number s.

Call a vertex $a \in A$ *critical* if among its neighbors, all possible s-subsets of $\{1, 2, ..., L\}$ appear. For each $b \in B$, assign L_b to be an s-subset of $[L]$ uniformly at random and independently. Note that the probability that a is not critical is equal to the probability that there exists some s-subset T of $[L]$ such that no neighbor of a is assigned T as its list. Since there are $\binom{L}{s}$ possible T's and each neighbor of a is assigned any given one of them with probability $\frac{1}{\binom{L}{s}}$, it follows that

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

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,

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

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 $\frac{|A|}{2}$.

Now pick an actual color choice for each $b \in B$ from these assigned lists, i.e., a choice function for $\{L_v | v \in B\}$. There are $s^{|B|}$ different ways of coloring the vertices in B. Fix one such coloring, and denote it as w. Denote as W the set of all colors chosen by w for some vertex in B, i.e., the set of all colors in the range of w. Since there exists a critical $a \in A$, W has nonempty intersection with all s-subsets of $[L]$.

Now note that $|W| \geq L - s + 1$. Otherwise, $|W^c| \geq s$, and since W intersects all s-subsets of [L], this would imply that $W \cap W^c \neq \emptyset$, a contradiction. So if (an extension of) a coloring exists, L_a must contain an element of W^c, with $|W^c| \leq s - 1$. Now let's pick color lists for vertices of A uniformly at random from the s-subsets of $[L]$. Then we have the following upper bound on the probability that we can extend w to $B \cup \{a\}$:

$$P(\text{an extension to } a \text{ exists}) \le \frac{(s-1)\binom{L-1}{s-1}}{\binom{L}{s}} < \frac{s^2}{L}.$$

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 that

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

which is less than 1 if $s\sqrt{\frac{s^2}{L}} < 1$, if $\frac{s^2}{L} < 1$. So take $L > s^4$. 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 $d > \binom{L}{s}\log(2\binom{L}{s})$. It follows that if

$$d > 4\binom{s^4}{s}\log(2\binom{s^4}{s}),$$

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., $\chi_l(G) > s$. □

Alon later improved his bound to $\chi_l(G) > (\frac{1}{2} - o(1))\log d$ with $d = \delta(G)$. It is not known if this is best possible.

2.3 A conjecture of Daykin and Erdős and its resolution

Suppose $\mathcal{H} \subset \mathcal{P}([n])$ is a hypergraph. One can construct the following graph $G_{\mathcal{H}}$:

- $V(G_{\mathcal{H}}) = E(\mathcal{H})$, and
- $E \leftrightarrow F$ in $G_{\mathcal{H}}$ iff $E \cap F = \emptyset$.

Suppose $e(\mathcal{H}) = 2^{(\frac{1}{2}+\delta)n}$ for some $\delta > 0$.

- $\{E, F\}$ is a disjoint pair if $E \cap F \neq \emptyset$.

Erdős had already obtained $R(k,k) > 2^{k/2}$ but his example(s) were not deterministic constructions. One wishes to check if the graphs $G_{\mathcal{H}}$ could be examples for certain collection $\mathcal{H}$.

Now, suppose G_n is Ramsey for K_k, i.e., $\alpha(G_n) < 2\log_2 n$ and $w(G_n) < 2\log_2 n$, where $\alpha(G_n)$ and $w(G_n)$ denote the independence number and maximum clique size of G_n, respectively. Note that the theorem of Turán tells us that $\alpha(G) \ge \frac{n}{\overline{d}+1}$, where $\overline{d}$ is the average degree of the vertices of G, so in particular, if $e(G_n) \le cn^{2-\delta}$ for some constant $c, \delta > 0$, then $\alpha(G_n) \ge \frac{n}{d+1} = \Omega(n^\delta)$. Consequently, these graphs would **not** be examples of graphs that can achieve lower bounds for $R(k, k)$.

Daykin and Erdős, perhaps in an attempt to get a better upper bound for $R(k, k)$ with graphs $G_{\mathcal{H}}$ and perhaps after due failure, were then forced to consider the following possibility:

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 $A \subset A, F \subset F$. 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. In fact they proved: $d(\mathcal{H}) < cm^{2-\frac{\delta^2}{2}}$ if $|\mathcal{H}| = 2^{(1/2+\delta)n}$. This is the result we shall consider in this section.

Let us see an idea for this proof first. If the underlying 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., there are lots of edges between these two pairs. In the extreme case of this pair witnesses all possible edges, one would have

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

So if the pair $\mathcal{S}, \mathcal{T}$ constitute a dense pair, we would expect that $\bigcup\limits_{S\in\mathcal{S}} S$ and $\bigcup\limits_{T\in\mathcal{T}} T$ are *almost* disjoint from each other; yet both these are 'large' subsets of $[n]$. This appears unlikely and is probably a manner to pull off a contradiction.

To see if we can pull this off, one might want to try something simpler first: If we find a $\mathcal{S}$ such that $A(\mathcal{S}) = \cup_{S\in\mathcal{S}} S$ is large, and also, that the number of sets $E \in \mathcal{H}$ such that $E \cap A(\mathcal{S}) = \emptyset$, is also large, then too one has a contradiction.

For instance, if $|A(\mathcal{S})| > \frac{n}{2}$, then in particular, $\#\{E \subset \mathcal{H} \mid E \cap A(\mathcal{S}) = \emptyset\} < 2^{\frac{n}{2}}$. So if there exists $\mathcal{S}$ such that

- $A(\mathcal{S}) = \cup_{S\in\mathcal{S}} S$ has $|A(\mathcal{S})| > \frac{n}{2}$ (which is likely), and

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

then we are through.

Proof. (i) Over t rounds, pick $S_1, S_2, \cdots, S_t \in \mathcal{H}$ uniformly and independently, t is to be determined.

$$\mathbb{P}(|A(\mathcal{S})| \leq \frac{n}{2}) = \mathbb{P}(\exists\ T \subset [n],\ |T| = \frac{n}{2} \text{ and } S_i \subset T\ \forall\, i)$$

Fix T of size $\frac{n}{2}$.

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

Therefore,

$$\begin{aligned}\mathbb{P}(|A(\mathcal{S})| \leq \frac{n}{2}) \leq \binom{n}{\frac{n}{2}} \frac{1}{2^{t\delta n}} &< \frac{2^n}{2^{t\delta n}} \\ &= \frac{1}{2^{(t\delta-1)n}}.\end{aligned}$$

Remark 18. As long as $t\delta - 1 > 0$, it is likely. This implies $t > \frac{1}{\delta}$. We may choose $t = \frac{2}{\delta}$.

(ii) We want $X := \#\{E \in \mathcal{H} \,|\, E \cap A(\mathcal{S}) = \emptyset\}$ to be 'large'.
X can be rewritten as $X = \sum_{E \in \mathcal{H}} \mathbb{1}_{\{E \cap A(\mathcal{S}) = \emptyset\}}$. This implies $\mathbb{E}[X] = \sum_{E \in \mathcal{H}} \mathbb{P}(E \cap A(\mathcal{S}) = \emptyset)$.
Fix $E \in \mathcal{H}$.

$$\begin{aligned}\mathbb{P}(E \cap A(\mathcal{S}) = \emptyset) &= \mathbb{P}(E\ \leftrightarrow S_i\ \forall\, i = 1, \cdots, t) \\ &= \left(\frac{d(E)}{m}\right)^t,\end{aligned}$$

where $d(E)$ is the degree of E in $G_{\mathcal{H}}$.
So,

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

where $e(G_{\mathcal{H}}) = M$.

Note: $0 \le X \le M$ with probability 1. This implies $0 \le \frac{X}{M} \le 1$ with probability 1. Therefore,

$$\mathbb{E}\left[\frac{X}{M}\right] = \frac{1}{M}\sum_{E\in\mathcal{H}} \mathbb{P}(E \cap A(\mathcal{S}) = \emptyset).$$

Now, $\mathbb{P}(X < a) \le \frac{1-\mathbb{E}[X]}{1-a}$. Therefore,

$$\mathbb{P}(X \ge a) \ge 1 - \frac{1-\mathbb{E}[X]}{1-1} = \frac{\mathbb{E}[X]-a}{1-a}.$$

Also,

$$\mathbb{P}\left(\frac{X}{M} \ge a\right) \ge \frac{\frac{1}{M}\mathbb{E}[X]-a}{1-a}.$$

Let $a = \frac{\mathbb{E}[X]}{2M}$. Therefore,

$$\mathbb{P}\left(\frac{X}{M} \ge a\right) \ge \frac{\frac{1}{M}\mathbb{E}[X] - \frac{\mathbb{E}[X]}{2M}}{1-\frac{\mathbb{E}[X]}{2M}} = \frac{\frac{E[X]}{2M}}{1-\frac{\mathbb{E}[X]}{2M}} = \frac{\frac{1}{2}\mathbb{E}\left[\frac{X}{M}\right]}{1-\frac{1}{2}\mathbb{E}\left[\frac{X}{M}\right]}$$

$$= \frac{\frac{1}{2}\left(\frac{2^t M^{t-1}}{m^{2t-1}}\right)}{1-\left(\frac{2^{t-1}M^{t-1}}{m^{2t-1}}\right)}.$$

Getting back to (i),

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

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.

Remark 19. Suppose $M = cm^2$. We have that t depends only on δ and c. Question is at what value of M it satisfies a contradiction. One can show by routine calculus that for $M \ge m^{2-\frac{\delta^2}{2}}$, we have a contradiction.

□

2.4 An example in Combinatorial Geometry

During the 1950s Paul Erdős conjectured that every set of more than 2^n points in $\mathbb{R}$ determines at least one obtuse angle (angle strictly greater than $\pi/2$.). In 1962 Ludwig

Danzer and Branko Grünbaum proved this conjecture and in their paper they conjectured the following: "Any configuration of $2n$ points in $\mathbb{R}^n$ contains some three points that form a non-acute angle". In fact, they managed a proof of the same for dimensions $2, 3$.

The conjecture needed about another 20 years to see it settled. In 1983 Erdős and Füredi proved that the Danzer-Grünbaum conjecture is outrageously false, for n large.

Theorem 20. *There exist point set $S \subset \mathbb{R}^n$ with only acute angles with*

$$|S| \geq c(1+\epsilon)^n$$

for some absolute constants c and ϵ.

To start off, note that if we included right angles, then there is indeed an exponentially large set, namely, the n dimensional hypercube consisting of all points $(x_1, x_2, \ldots, x_n)$ with $x_i \in \{0, 1\}$ for all i - a fact that Erdős knew quite well. Hence a staring point would be to see if one could pick a subset of this set of vertices of the n-dimensional hypercube, and make sure we haven't included any right angles.

Note that each vertex of the cube can also be thought of as a subset of $[n]$, i.e., let $P = \{i : P(i) \neq 0\}$ is a subset of $[n]$. Let $X = \#\{\{P, Q, R\} : \angle PQR = \frac{\pi}{2}\}$. Suppose the sets P, Q, R form a right angle, then $\langle V_P - V_Q, V_R - V_Q \rangle = 0$ (where $\langle \rangle$ is the usual dot product in $\mathbb{R}^n$ and V_P denote the vertex P.)
For example, see the case $n = 2$ in the figure 2.1:

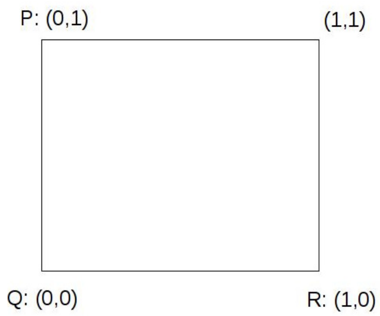


Figure 2.1: $n = 2$

Here, $V_P - V_Q = (0, 1)$ and $V_R - V_Q = (1, 0)$. In general zeros of $V_P - V_Q$ corresponds to $P \cap Q$ or $\bar{P} \cap \bar{Q}$, ones of $V_P - V_Q$ corresponds to $P \cap \bar{Q}$ and -1s of $V_P - V_Q$ corresponds to $\bar{P} \cap Q$. Similarly,the same holds for $V_R - V_Q$. The dot product is zero means

$$\emptyset = \bar{P} \cap Q \cap \bar{R}$$

and

$$\emptyset = P \cap \bar{Q} \cap R.$$

Observe that $\emptyset = \bar{P} \cap Q \cap \bar{R}$ implies $Q \subset P \cup R$ and $\emptyset = \bar{P} \cap Q \cap \bar{R}$ implies $P \cap R \subset Q$. Therefore $\{P, Q, R\}$ is a right angle triple if and only if

$$P \cap R \subset Q \subset P \cup R.$$

Since we are looking for a 'large' subset of $\mathcal{P}([n])$ in which there are no triples (P, Q, R) with $P \cap R \subset Q \subset P \cup R$, the probabilistic method beckons. Indeed, let each vertex of the $n-$cube be picked independently with probability p for some p to be determined. Denote this set by S. Then $\mathbb{E}(|S|) = 2^n p$, and furthermore,

$$\mathbb{E}(\text{Right angle triples}) = p^3(\text{Number of right angle triples}).$$

Now, the number of triples (P, Q, R) satisfying $P \cap R \subset Q \subset P \cup R$ is precisely 6^n. Indeed, every $x \in [n]$ is in exactly one of the following pairwise disjoint sets: $\overline{P \cup Q}, P \setminus (Q \cup R), (P \cap R) \setminus Q, P \cap Q \cap R, (Q \cap R) \setminus P, Q \setminus (P \cup R)$. Hence the expected number of right angles that are picked into the set S is

$$\mathbb{E}(\text{Right angle triple}) = \frac{p^3 6^n}{2}.$$

Hence there is a subset S' of size at least

$$2^n p - \frac{p^3 6^n}{2} \tag{2.1}$$

which admits no right angles. To optimize this, let

$$f(p) = 2^n - \frac{1}{2} 6^n p^3.$$

Then,

Optimizing this over p, we see that this attains a maximum for

$$p = \frac{\sqrt{2}}{(\sqrt{3})^{n+1}}.$$

Plugging this in (2.1), it follows that there is a set of size at least

$$\sqrt{\frac{2}{3}}\left(\frac{2}{\sqrt{3}}\right)^n - \frac{2}{3}\sqrt{\frac{2}{3}}\left(\frac{2}{\sqrt{3}}\right)^n = \frac{1}{3}\sqrt{\frac{2}{3}}\left(\frac{2}{\sqrt{3}}\right)^n$$

in which every angle formed by any three points is in fact acute. This concludes the proof of the theorem.

2.5 Graphs with High Girth and High 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 21. *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 22. *There are graphs with arbitrarily high girth and chromatic number.*

Proof. So: let $G_{n,p}$ denote a random graph on n vertices, formed by doing the following:

- Start with n vertices.
- For every pair of vertices $\{x, y\}$, flip a biased coin that comes up heads with probability p and tails with probability $1 - p$. If the coin is heads, add the edge $\{x, y\}$ to our graph; if it's tails, don't.

Our roadmap, then, is the following:

- For large n and well-chosen p, we will show that $G_{n,p}$ will have relatively "few" short cycles at least half of the time.
- For large n, we can also show that G will have high chromatic number at least half the time.
- Finally, by combining these two results and deleting some vertices from our graph, we'll get that graphs with both high chromatic number and no short cycles exist in our graph.

To do the first: fix a number l, and let X be the random variable defined by $X(G_{n,p}) =$ the number of cycles of length $\leq l$ in $G_{n,p}$.

We then have that

$$X(G_{n,p}) \leq \sum_{j=3}^{l} \sum_{all j-tuples x_1 \ldots x_j} N_{x_1 \ldots x_j},$$

where $N_{x_1 \ldots x_j}$ is the event that the vertices $x_1 \ldots x_j$ form a cycle.

Then, we have that

$$\begin{aligned}
\mathbb{E}(X) &\leq \sum_{j=3}^{l} \sum_{j-tuples\ x_1 \ldots x_j} Pr(N_{x_1 \ldots x_j}) \\
&= \sum_{j=3}^{l} \sum_{j-tuples\ x_1 \ldots x_j} p^j \\
&= \sum_{j=3}^{l} n^j p^j.
\end{aligned}$$

To make our sum easier, let $p = n^{\lambda - 1}$, for some $\lambda \in (0, 1/l)$; then, we have that

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{j=3}^{l} n^j p^j \\
&= \sum_{j=3}^{l} n^j n^{j\lambda - j} \\
&= \sum_{j=3}^{l} n^{j\lambda} \\
&< \frac{n^{\lambda(l+1)}}{n^{\lambda} - 1} \\
&= \frac{n^{\lambda l}}{1 - n^{-\lambda}}
\end{aligned}$$

We claim that this is smaller than n/c, for any c and sufficiently large n. To see this, simply multiply through; this gives you that

$$\begin{aligned}
&\frac{n^{\lambda l}}{1 - n^{-\lambda}} < n/c, \\
&\Leftrightarrow n^{\lambda l} < n/c - n^{1-\lambda}/c, \\
&\Leftrightarrow n^{\lambda l} + n^{1-\lambda}/c < n/c,
\end{aligned}$$

which, because both λl and $1-\lambda$ are less than 1, we know holds for large n.

So: to recap: we've shown that

$$\mathbb{E}(|X|) < n/4.$$

So: what happens if we apply Markov's inequality? Well: we get that

$$Pr(|X| \geq n/2) \leq \frac{\mathbb{E}(|X|)}{n/2} < \frac{n/4}{n/2} = 1/2;$$

in other words, that more than half of the time we have relatively "few" short cycles! So this is the first stage of our theorem.

Now: we seek to show that the chromatic number of our random graphs will be "large," on average. Doing this directly, by working with the chromatic number itself, would be rather ponderous. Rather, we will work with the **independence number** $\alpha(G)$ of our graph, the size of the independent set of vertices[1] in our graph. Why do we do this? Well, in a proper k-coloring of a graph, each of the colors necessarily defines an independent set of vertices, as there are no edges between vertices of the same color; ergo, we have that

$$\chi(G) \geq \frac{|V(G)|}{\alpha(G)},$$

for any graph G.

So: to make the chromatic number large, it suffices to make $\alpha(G)$ small! So: look at $Pr(\alpha(G) \geq m)$, for some value m. We then have the following:

$$\begin{aligned}
Pr(\alpha(G) \geq m) &= Pr(\text{there is a subset of } G \text{ of size } m \text{ with no edges in it})\\
&\leq \sum_{S \subset V, |S| = m} Pr(\text{there are no edges in } S\text{'s induced subgraph})\\
&= \binom{n}{m} \cdot (1-p)^{\binom{m}{2}}\\
&< n^m \cdot e^{-p \cdot \binom{m}{2}}\\
&= n^m \cdot e^{-p \cdot m \cdot (m-1)/2}.
\end{aligned}$$

So: motivated by a desire to make the above simple, let $m = \left\lceil \frac{3}{p} \ln(n) \right\rceil$. This then gives us that

$$\begin{aligned}
Pr(\alpha(G) \geq m) &< n^m \cdot e^{-p \cdot \left\lceil \frac{3}{p} \ln(n) \right\rceil \cdot (m-1)/2}\\
&= n^m \cdot n^{-3(m-1)/2},
\end{aligned}$$

[1] A set of vertices is called **independent** if their induced subgraph has no edges. in it

which goes to 0 as n gets large. So, in particular, we know that for large values of n and any m, we have

$$Pr(\alpha(G) \geq m) < 1/2.$$

So: let's combine our results! In other words, we've successfully shown that for large n,

$$Pr(G \text{ has more than } (n/2)\text{-many cycles of length } \leq l, \text{ or } \alpha(G) \geq m) < 1.$$

So: for large n, there is a graph G so that neither of these things happen! Let G be such a graph. G has less than $n/2$-many cycles of length $\leq l$; so, from each such cycle, delete a vertex. Call the resulting graph G'.

Then, we have the following:

- By construction, G' has girth $\geq l$.
- Also by construction, G' has at least $n/2$ many vertices, as it started with n and we deleted $\leq n/2$.
- Because deleting vertices doesn't decrease the independence number of a graph, we have that

$$\begin{aligned}
\chi(G') &\geq \frac{|V(G')|}{\alpha(G')} \\
&\geq \frac{n/2}{\alpha(G)} \\
&\geq \frac{n/2}{3\ln(n)/p} \\
&= \frac{n/2}{3n^{1-\lambda}\ln(n)} \\
&= \frac{n^{\lambda}}{6\ln(n)},
\end{aligned}$$

  which goes to infinity as n grows large.

Thus, for large n, this graph has arbitrarily large girth and chromatic number. □

3 2-colorability of Uniform Hypergraphs

Colorability of hypergraphs is a much more difficult problem in comparison to the corresponding problem for graphs. In fact, even determining if a hypergraph can be 2-colored is a very hard problem. In this chapter, we look at two celebrated theorems due to Jozsef Beck, and Radhakrishnan and Srinivasan, respectively on 2-colorability of hypergraphs which improve upon an earlier result of Erdős. The general situation is still wide open. Both these theorems illustrate the subtle technique of making alterations to a probabilistic argument.

3.1 Introduction

Definition 23. *A **hypergraph** is a pair $\mathcal{H} = (V, \mathcal{E})$, where V denotes a collection of vertices and $\mathcal{E}$ denotes a collection of subsets of V (the "hyperedges" of H.) If all of $\mathcal{E}$'s elements have cardinality n, we say that $\mathcal{H}$ is a n-uniform hypergraph.*

Definition 24. *A k-coloring of a hypergraph $\mathcal{H}$ is a way of assigning k distinct colors to all of $\mathcal{H}$'s vertices, so that every edge has at least 2 colors.*

Given these definitions, a natural question we can ask is the following:

Question 25. *When is a given hypergraph $\mathcal{H}$ 2-colorable?*

(Hypergraphs that are 2-colorable are often said to possess "Property B" – this bit of nomenclature is in honor of Felix Bernstein, who was one of the first people to investigate this property of hypergraphs.)

Given the current trend of these notes, it should come as no surprise that Erdős found a way to use the probabilistic method to answer some of this question:

Theorem 26. *(Erdős) If $\mathcal{H} = (V, \mathcal{E})$ is a n-regular hypergraph and $|\mathcal{E}| \leq 2^{n-1}$, then $\mathcal{H}$ is 2-colorable.*

Proof. Independently and randomly color every vertex $v \in V$ either red or blue with probability 1/2. Then, if we let N denote the number of monochrome edges created under this coloring, we have that

$$\begin{aligned}
\mathbb{E}(N) &= \sum_{e \in \mathcal{E}} \mathbb{E}\left(\mathbb{1}_{\{e \text{ is monochrome}\}}\right) \\
&= \sum_{e \in \mathcal{E}} \mathbb{E}\left(\mathbb{1}_{\{e \text{ is entirely red or } e \text{ is entirely blue}\}}\right) \\
&= \sum_{e \in \mathcal{E}} \mathbb{E}\left(\mathbb{1}_{\{e \text{ is entirely red}} + \mathbb{1}_{\{e \text{ is entirely blue}\}}\right) \\
&= \sum_{e \in \mathcal{E}} \frac{1}{2^n} + \frac{1}{2^n} \\
&= \frac{|\mathcal{E}|}{2^{n-1}}.
\end{aligned}$$

Thus, if $|\mathcal{E}| \leq 2^{n-1}$, we have that $\mathbb{E}(N) \leq 1$. Then there are two potential cases:

- $\mathbb{P}(N = 0) = 0$. In this case, we have that $N = 1$ with probability 1; but this is clearly impossible, as in any graph with more than one edge there are colorings under which multiple monochrome edges exist. So this cannot occur.

- $\mathbb{P}(N = 0) > 0$. In this case, there is a nonzero probability that $N = 0$; thus, we can simply take some coloring that witnesses this event. This gives us a 2-coloring of $\mathcal{H}$, which is what we sought to find.

□

One quick corollary of the above result is the following:

Corollary 27. *Let $m(n)$ denote the smallest number of edges needed to form a n-uniform hypergraph that is* ***not*** *2-colorable. Then $m(n) \geq 2^{n-1}$.*

So: our above results have given us a lower bound on the quantity $m(n)$. Can we find an upper bound?

3.2 Upper bound for $m(n)$

One rather trivial upper bound on $m(n)$ we can get is the following:

Proposition 28.

$$m(n) \leq \binom{2n}{n}.$$

Proof. Let $V = \{1, \dots 2n\}$, and let $\mathcal{E} = \binom{V}{n}$, all of the n-element subsets of V. Because any 2-coloring of V must yield at least n vertices all of the same color, there is always a monochrome edge in this coloring. □

Using constructive methods, it is hard to realistically improve on the rather simple bound above, which is asymptotically growing somewhat like 4^n – a long ways from our lower bound of $2^{n-1}!$. Probabilistic methods, as it turns out, can offer a much better estimate as to what $m(n)$ actually is:

Theorem 29. *(Erdős)*

$$m(n) \leq O(n^2 2^n).$$

Proof. Let V be some set of vertices, with $|V| = v$.

Choose m edges from the collection $\binom{V}{n}$ independently at random: specifically, pick an edge E with probability $1/\binom{v}{n}$, and repeat this process m times. (Notice that we are allowing repetitions.) Take this chosen collection of edges, $\mathcal{E}$, to be the edges of our hypergraph.

Pick a 2-coloring χ of V, and let a denote the number of red vertices and b denote the number of blue vertices. Then, for any edge $E \in \mathcal{E}$, we have that

$$\mathbb{P}(E \text{ is monochrome under } \chi) = \frac{\binom{a}{n} + \binom{b}{n}}{\binom{v}{n}} \geq \frac{\binom{v/2}{n} + \binom{v/2}{n}}{\binom{v}{n}}.$$

For simplicity's sake, denote the quantity $\frac{\binom{v/2}{n}+\binom{v/2}{n}}{\binom{v}{n}}$ by P: then we have that

$$\begin{aligned} & \mathbb{P}(E \text{ is monochrome under } \chi) \geq P \\ \Rightarrow\; & \mathbb{P}(E \text{ is not monochrome under } \chi) \leq 1 - P \\ \Rightarrow\; & \mathbb{P}(\chi \text{ is a proper 2-coloring}) \leq (1 - P)^m. \end{aligned}$$

Therefore, if we look at the collection of all possible colorings of our graph, we have that

$$\begin{aligned} \mathbb{P}\left(\bigvee_{\chi \text{ a 2-coloring}} \chi \text{ is a proper 2-coloring} \right) &\leq \sum_{\chi \text{ a 2-coloring}} \mathbb{P}\left(\chi \text{ is a proper 2-coloring}\right) \\ &\leq (\# \text{ of 2-colorings of } V) \cdot (1 - P)^m \\ &= 2^v \cdot (1 - P)^m \\ &\leq 2^v \cdot e^{-Pm} \\ &= e^{v \ln(2) - Pm}. \end{aligned}$$

So: this quantity is less than 1 – and thus a choice of edge-set exists for which the associated graph is not 2-colorable – if

$$m \geq \frac{v \ln(2)}{P}.$$

Finding the optimal choice of v to maximize this lower bound on m is then just a matter of manipulating a few inequalities. First, let's create an upper bound for P:

$$\begin{aligned}
P &= \frac{2 \cdot \binom{v/2}{n}}{\binom{v}{n}} \\
&= 2\frac{(v/2) \cdot \ldots \cdot (v/2 - (n-1))}{v \cdot \ldots \cdot (v - (n-1))} \\
&= \frac{2}{2^n} \cdot \frac{v \cdot (v-2) \cdot \ldots \cdot (v - 2(n-1))}{v \cdot \ldots \cdot (v-(n-1))} \\
&= \frac{1}{2^{n-1}} \cdot \prod_{i=0}^{n-1} \left(\frac{v-2i}{v-i} \right) \\
&= \frac{1}{2^{n-1}} \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{v-i} \right) \\
&= \frac{1}{2^{n-1}} \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{v} \cdot \frac{1}{1-i/v} \right) \\
&= \frac{1}{2^{n-1}} \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{v} \left(1 + \frac{i}{v} + \frac{i^2}{v^2} + \ldots \right) \right) \\
&= \frac{1}{2^{n-1}} \cdot \prod_{i=0}^{n-1} \left(1 - \frac{i}{v} + O\left(\left(\frac{i}{v} \right)^2 \right) \right) \\
&\leq \frac{1}{2^{n-1}} \cdot \prod_{i=0}^{n-1} \exp\left(-\frac{i}{v} + O\left(\left(\frac{i}{v} \right)^2 \right) \right) \\
&\leq \frac{1}{2^{n-1}} \exp\left(-\frac{n(n-1)}{v} + O\left(\frac{n^3}{v^2} \right) \right)
\end{aligned}$$

Using this, we then have that

$$\begin{aligned} m &\geq \frac{v \ln(2)}{P} \\ &\geq v \ln(2) \cdot 2^{n-1} \cdot \exp\left(\frac{n(n-1)}{v} + O\left(\frac{n^3}{v^2}\right)\right). \end{aligned}$$

If we pick $v = n^2/2$ and n sufficiently large such that the $O(n^3/v^2)$-portion above is negligible, we then have that there is a non-2-colorable n-uniform hypergraph on m edges if

$$m \geq \frac{n^2}{2} \ln(2) \cdot 2^{n-1} \cdot e^{1+o(1)}.$$

In other words, we've shown that

$$m(n) \leq O(n^2 2^n),$$

which is what we claimed. □

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

$$m(n) := \arg\min_{|\mathcal{E}|} \{\mathcal{H} = (\mathcal{V}, \mathcal{E}) \text{ is 2-colorable}\} \tag{3.1}$$

2-colorability of finite hypergraphs is also known as "Property B". In [?], Erdős showed that $2^{n-1} < m(n) < O(n^2 2^n)$. In [?], 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 and Srinivasan in [?]. In fact, Erdős-Lovaśz conjecture that $m(n) = \Theta(n 2^n)$. Here, we outline the proofs of both Beck's and Radhakrishnan's results. We will begin with some notation, if an edge $S \in \mathcal{E}$ is monochromatic, we will denote it as $S \in \mathbb{M}$, 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.

3.3 Beck's result

Theorem 30 ([?]).

$$m(n) = \Omega(n^{\frac{1}{3}} 2^n)$$

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 $\mathbb{E}(N) < 1$.

$$\begin{aligned}
\mathbb{E}(N) &= \sum_{S\in\mathcal{E}} \mathbb{P}(S(2) \in \mathbb{M}) = \sum_{S\in\mathcal{E}} (\mathbb{P}(S(2) \in RED) + \mathbb{P}(S(2) \in BLUE)) \\
&= 2\sum_{S\in\mathcal{E}} \mathbb{P}(S(2) \in RED) \\
\mathbb{P}(S(2) \in RED) &= \underbrace{\mathbb{P}(S(1) \in \mathbb{M}, S(2) \in RED)}_{P_1} + \underbrace{\mathbb{P}(S(1) \notin \mathbb{M}, S(2) \in RED)}_{P_2}
\end{aligned}$$

It is easy to bound P_1

$$\begin{aligned}
P_1 = \mathbb{P}(S(1) \in RED, S(2) \in RED) + \mathbb{P}(S(1) \in BLUE, S(2) \in RED) &= \frac{p^n + (1-p)^n}{2^n} \\
&\le \frac{2(1-p)^n}{2^n}
\end{aligned} \tag{3.2}$$

In (3.2), 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 $T \cap S \ne \Phi$ and $T \in BLUE$. Define

$$E_{ST} := \text{event } S(1) \notin \mathbb{M},\ T(1) \in BLUE,\ S \cap T \ne \Phi \text{ and } S(2) \in RED$$

Then we have

$$P_2 \le \sum_{T\ne S} \mathbb{P}(E_{ST}) \tag{3.3}$$

Let $U := \{v \in S \setminus T \,|\, v(1) \in BLUE\}$ and E_{STU} := event $S \cap T \ne \Phi$, $T(1) \in BLUE$, $U \in BLUE$ and $S(2) \in RED$, then

$$\mathbb{P}(E_{ST}) = \mathbb{P}\left(\bigvee_{U\subseteq S\setminus T} E_{STU}\right) \le \sum_{U\subseteq S\setminus T} \mathbb{P}(E_{STU})$$

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 (3.3), we have

$$\begin{aligned}\mathbb{P}(E_{ST}) &\leq \sum_{U\subseteq S\setminus T} \frac{p}{2^{2n-1}} p^{|U|} \leq \frac{p}{2^{2n-1}} \sum_{|U|=0}^{n-1} \binom{n-1}{|U|} p^{|U|} \\ &= \frac{(1+p)^{n-1}p}{2^{2n-1}} \leq \frac{2p(1+p)^n}{2^{2n}} \leq \frac{2p\exp(np)}{2^{2n}} \\ \implies \sum_{S\neq T} \mathbb{P}(E_{ST}) &\leq \frac{2mp\exp(np)}{2^{2n}} \end{aligned} \tag{3.4}$$

Using (3.2),(3.3),(3.4), we get (recall that $m = k2^n$)

$$\begin{aligned}\mathbb{E}(N) &\leq 2\sum_S \left(\frac{m^2 p\exp(np)}{2^{2n}} + \frac{(1-p)^n}{2^n}\right) \\ &= 2\left(k^2 p\exp(np) + k(1-p)^n\right) \end{aligned} \tag{3.5}$$

For an arbitrary $\epsilon > 0$, let $p = \frac{(1+\epsilon)\log k}{n}$, then $k(1-p)^n \leq k\exp(-np) = k^{-\epsilon}$ and $k^2 p\exp(np) = \frac{k^{3+\epsilon}(1+\epsilon)\log k}{n}$. So, (3.5) gives

$$\mathbb{E}(N) \leq 2k^{-\epsilon} + \frac{2k^{3+\epsilon}(1+\epsilon)\log k}{n} \tag{3.6}$$

So, if $k \sim n^{1/3-2\epsilon/3}$, then (3.6) will be less than 1, so that $\mathbb{P}(N = 0) > 0$. □

3.4 An improvement by Radhakrishnan-Srinivasan

Theorem 31 ([?]).

$$m(n) = \Omega\left(2^n\sqrt{\frac{n}{\log n}}\right) \tag{3.7}$$

(R-S) take Beck's recoloring idea and improve it. Their technique is motivated by the following observation

Observation 32. *Suppose S is monochrome after step 1, then it suffices to recolor 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 recolor all vertices simultaneously, put them in an ordered list and recolor 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 (3.2). 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, i.e., $S \longmapsto T$, if this happens. Also, none of the vertices in T that were considered before v change their color to red. To summarize,

Lemma 33. *$S \longmapsto T$ 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,

$$P_2 \leq \mathbb{P}\left(\bigvee_{T \neq S} S \longmapsto T\right) \leq \sum_{T \neq S} \mathbb{P}(S \longmapsto T) \tag{3.8}$$

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 $w \in S \,|\, \pi(w) < \pi(v)$, $\mathbb{P}((w(1) \in RED)$ or $(w(1) \in BLUE, w(2) \in RED) \,|\, S \notin \mathbb{M}) = \dfrac{1+p}{2}$

So,

$$\begin{aligned}\mathbb{P}(S \longmapsto T \,|\, \pi) &\leq \frac{p}{2}\frac{1}{2^{n-1}}\frac{1}{2^{n-i_\pi-1}}(1-p)^{j_\pi}\left(\frac{1+p}{2}\right)^{i_\pi} \\ &= \frac{p}{2^{2n-1}}(1-p)^{j_\pi}(1+p)^{i_\pi} \end{aligned} \tag{3.9}$$

Let the ordering be random, then $\mathbb{P}(S \longmapsto T) = E_\pi P(S \longmapsto T \,|\, \pi)$. 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 34.

$$\mathbb{P}(S \longmapsto T) = \mathbb{E}\left(\mathbb{P}(S \longmapsto T \,|\, \pi)\right) \leq \frac{p}{2^{2n-1}} \tag{3.10}$$

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 $w \in T$ $\mathbb{P}(\ell(w) \leq \ell(v), \text{w flips color}) = px$, 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

$$\begin{aligned}\mathbb{P}\left(S \longmapsto T, |U| = u \,|\, \ell(v) = x\right) &= \underbrace{\frac{1}{2^{2n-1}}}_{\text{coloring after step 1}} \quad \underbrace{x^u}_{\ell(U)\leq x} \quad \underbrace{p^{1+u}}_{U \cup \{v\} \text{ flip to red}} \quad (1-px)^{n-1} \\ \implies \mathbb{P}(S \longmapsto T) &\leq \sum_{u=0}^{n-1}\binom{n-1}{u}\int_0^1 \frac{1}{2^{2n-1}}p^{1+u}x^u(1-px)^{n-1}dx \\ &= \frac{p}{2^{2n-1}}\int_0^1\left(\sum_{u=0}^{n-1}\binom{n-1}{u}(px)^u\right)(1-px)^{n-1}dx \\ &= \frac{p}{2^{2n-1}}\int_0^1 (1-p^2x^2)^{n-1}dx \\ &\leq \frac{p}{2^{2n-1}} \end{aligned} \tag{3.11}$$

□

Proof of theorem 31. Using (3.11) in (3.8), 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

$$\mathbb{E}(N) \leq \frac{k(1-p)^n}{2} + \frac{k^2p}{2} \tag{3.12}$$

Compare (3.12) with (3.5) and note that $\exp(np)$ is not present in (3.12). For an arbitrary $\epsilon > 0$, setting $p = \frac{(1+\epsilon)\log k}{n}$ and approximating $(1-p)^n \approx \exp(-np)$, we get

$$\mathbb{E}(N) \leq 0.5\left(k^{-\epsilon} + (1+\epsilon)\frac{k^2 \log k}{n}\right) \tag{3.13}$$

Clearly $k \sim \sqrt{\frac{n}{\log n}}$ makes $\mathbb{E}(N) < 1$ giving the result. □

Spencer's proof of lemma 34. Aided by hindsight, Alon-Spencer give an elegant combinatorial argument to arrive at (3.11). 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 (3.9), consider $\mathbb{E}\left((1-p)^{j_\pi}(1+p)^{i_\pi}\right)$.

$$\begin{aligned}
\mathbb{E}\left((1-p)^{j_\pi}(1+p)^{i_\pi} \mid \mu \cap S_v^- \cup T_v^-\right) &= \mathbb{E}\left((1-p)^{\sum_{i=1}^{n-1}\mathbb{I}(\mu(i)\cap S_v^- \neq \Phi)}(1+p)^{\sum_{i=1}^{n-1}\mathbb{I}(\mu(i)\cap T_v^- \neq \Phi)}\right)\\
&= \mathbb{E}\left(\prod_{i=1}^{n-1}(1-p)^{\mathbb{I}(\mu_s(i)\in S_v^-)}(1+p)^{\mathbb{I}(\mu_t(i)\in T_v^-)}\right)\\
&= \prod_{i=1}^{n-1}\mathbb{E}\left((1-p)^{\mathbb{I}(\mu_s(i)\in S_v^-)}(1+p)^{\mathbb{I}(\mu_t(i)\in T_v^-)}\right)\\
&= \prod_{i=1}^{n-1}\left(\frac{1}{4}(1-p+1+p+1+1-p^2)\right)\\
&= \prod_{i=1}^{n-1}\left(1-\frac{p^2}{4}\right) < 1\\
\implies E\left((1-p)^{j_\pi}(1+p)^{i_\pi}\right) &= \mathbb{E}\left(\mathbb{E}\left((1-p)^{j_\pi}(1+p)^{i_\pi} \mid \mu \cap S_v^- \cup T_v^-\right)\right) < 1\\
\implies \mathbb{P}(S \longmapsto T) &< \frac{p}{2^{2n-1}}
\end{aligned}$$

□

4 Dependent Random Choice

4.1 Introduction

There is another aspect to the manner of tweaks which were discussed in the preceding chapter. Sometimes, (unlike what we have done thus far) 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 has the merit of achieving a desired effect. 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.

The premise for some of the investigations in this chapter is the following question: Given a 'small' graph H, if a graph G is dense on a large vertex set, can we find a copy of H in G? We say G is dense if $e(G_n) \geq \epsilon n^2$ for some $\epsilon > 0$, where ϵ is independent of n. More formally, we make the following definition.

Definition 35. *For any graph H, by $ex(n; H)$ we mean the maximal number of edges in a graph on n vertices without a copy of H.*

As it turns out, this need not always be true. Indeed, if $H = K_3, G = K_{n,n}, e(K_{n,n}) = n^2$, then clearly, there is no copy of H in G. Turán's theorem which is considered the foundation of extremal graph theory, says that if G has no copy of K_{r+1}, then $e(G_n) \leq (1 - \frac{1}{r})\frac{n^2}{2}$, or more precisely, $\text{ex}(n, K_{r+1}) = (1 - \frac{1}{r})\frac{n^2}{2}$. This was extended enormously by what is considered the fundamental theorem in extremal graph theory, in the following theorem of Erdős-Stone-Simonovits:

Theorem 36. *(Erdős-Stone-Simonovits) Suppose H is a 'small' graph, and $\chi(H) = r+1, r \geq 2$. For any $\epsilon > 0, \exists\ n(\epsilon)$, s.t. $\forall\ n \geq n(\epsilon)$, if $e(G_n) \geq (1 - \frac{1}{r})\frac{n^2}{2} + \epsilon n^2$, then $H \subset G_n$.*

In particular, if a graph H has chromatic number at least 3, then $\text{ex}(n, H)$ is determined asymptotically to a factor by its chromatic number.

Erdős-Stone-Simonovits: If $\chi(H) \geq 3$, then $ex(n; H) \approx (1 - \frac{1}{r})\frac{n^2}{2}$.

But note that if $r = 1$, i.e., if the graph H is bipartite then the Erdő-Stone-Simonovits theorem tells us the following: If $e(G_n) \geq \epsilon n^2$, then H $\subset G_n$. Thus we only know that in such cases, $ex(n; H) = o(n^2)$. This begs the following question:

Question 37. *For H bipartite, what is the correct value of α s.t. $ex(n; H) = \Theta(n^{\alpha(H)}), 1 \leq \alpha < 2$?*

The asymptotics for $ex(n; H)$ are known for very few bipartite graphs. Thus, one is curious to know if we can find a 'better power' α such that $ex(n; H) = O(n^\alpha)$? More generally, given a bipartite graph H, and a large graph G, how may we find a copy of H in G?

4.2 A graph embedding lemma and Dependent Random Choice

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. Consider the following scenario: Every vertex in B has degree at most r. If every r-subset of A_0 has 'many' common neighbors in G, then heuristically, our chances of embedding each vertex of B get better.

Proposition 38. *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 $A_0 \subset V(G), |A_0| \geq a$, s.t. every r-subset of A_0 has at least $a + b$ common neighbors in G. Then H can be embedded in G.*

Proof: Suppose we embed the vertices of A into A_0 arbitrarily. We shall now try to embed each vertex of B into G, one at a time. Let $B = \{v_1, v_2, \cdots v_b\}$. Suppose we have already embedded $v_1, \cdots v_{i-1}$. Now let V= neighbors of v_i in A (in H). 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 $|U| \leq r$, it has $\geq a + b$ common neighbors in G $\implies$ there is some available choice for v_i. $\square$

So, how does one find such an A_0?

Here is a heuristic: Pick a small subset T of V(G) over t rounds, at random (with repetition). Consider the set of common neighbors of T. Heuristically, denote the common neighbors of T by $N^*(T)$.

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

The inequality follows from Jensen's inequality for convex functions.

Let Y= number of r-subsets U of $N^*(T)$ s.t. 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.

$$\begin{aligned}
&\mathbb{P}(U \subset N^*(T)) \leq (\frac{m}{n})^t \\
\Longrightarrow &\mathbb{E}(Y) \leq \binom{n}{r}(\frac{m}{n})^t \\
\Longrightarrow &\mathbb{E}(|N^*(T)| - Y) \geq \frac{(\bar{d})^t}{n^{t-1}} - \binom{n}{r}(\frac{m}{n})^t
\end{aligned}$$

$\Longrightarrow \exists\ A_0 \subset N^*(T)$ of size $\geq \frac{(\bar{d})^t}{n^{t-1}} - \binom{n}{r}(\frac{m}{n})^t$, s.t. every r-subset of A_0 has $\geq m$ common neighbors.

Theorem 39. *(Alon, Krivelevich, Sudakov; 2003) H is bipartite with vertex partition $(A \cup B$, and if every vertex of B has degree $\leq r$, then $ex(n;H) = O(n^{2-\frac{1}{r}})$.*

Proof: Note that

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

where $C = C_H$ is some constant depending on H. To apply the lemma, we need that

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

$$\begin{aligned}
LHS &\geq \frac{((2C)^t n^{t-\frac{t}{r}})^t}{n^{t-1}} - \frac{n^r}{r!}(\frac{a+b}{n})^t \\
&= (2C)^r - \frac{(a+b)^r}{r!} \quad (\textit{if } t = r) \\
&> a \\
\Longrightarrow C &> \frac{1}{2}(a + \frac{(a+b)^r}{r!})^{\frac{1}{r}}
\end{aligned}$$

4.3 An old problem of Erdős

Firstly we need a definition.

Definition 40. *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 $e(G_n) \geq cp^2n$, then there is a topological copy of K_p in G. This was proved in 1998 by Bollabás 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 41. *A t-subdivision where each edge is replaced by a path with $\leq t$ internal vertices.*

Erdős had in fact asked the following question as well: Is there a 1-subdivision of $K_{\delta\sqrt{n}}$ in a dense graph for some absolute $\delta > 0$? This was settled in the affirmative by Alon, Krivelevich, and Sudakov.

Theorem 42. *(Alon, Krivelevich, Sudakov) If $e(G_n) \geq \epsilon n^2$, then G has a 1-subdivision of $K_{\epsilon^{3/2}\sqrt{n}}$.*

Proof: Firstly note that a 1-subdivision of the complete graph K_a with every edge getting subdivided corresponds to a bipartite graph with parts of size a, $\binom{a}{2}$, respectively. Furthermore, 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.

If we think along the lines of the embedding lemma we have proved we want the following: $\frac{(\bar{d})^t}{n^{t-1}} - \binom{n}{r}(\frac{m}{n})^t \geq a$, here $r = 2, m = a + \binom{a}{2} < 2\binom{a}{2} < a^2, \bar{d} \geq 2\epsilon n$

$$LHS > (2\epsilon)^t n - \frac{n^2}{2}\frac{a^{2t}}{n^t} \geq a$$

If $a = \delta\sqrt{n}$, and set $\delta = \epsilon^{3/2}$, then

$$LHS > \epsilon^t[(2)^t n - \frac{n^2}{2}\epsilon^{2t}]$$

If the second term in the square bracket becomes n then we may factor out n from both these terms. This basically means setting $t = \frac{\log n}{2\log(\frac{1}{\epsilon}}$).

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

As n goes large, this beats $a = \delta\sqrt{n}$. $\square$

4.4 The Balog-Szemerédi-Gowers Theorem

Let A be an arithmetic progression of length n then $|A + A| = (2 - \frac{1}{n})|A|$. Now let A pick any $\frac{n}{2}$ terms of an arithmetic progression then $|A + A| \leq 4|A|$. If we add an arbitrary number to an arithmetic progression we get $|A + A| \leq 4|A|$. There are results in the converse direction too like Freiman's theorem which states that if $|A + A| < c|A|$ for some constant $'c'$ then A lies in a generalized arithmetic progression$(a + n_1d_1 + n_2d_2 + ... + n_kd_k)$ of length and dimension which only depend on c.

So the size of $|A + A|$ in a way measures how structured the set is. Now say we do not have access to look at all possible sums but only a restricted sums (say denoted by a bipartite graph 'G on two vertex sets each containing elements from A and (a, a') is an edge in G if we have access to the corresponding sum.)

Then how much information does this partial sum set capture about $|A + A|$?

If G were sparse it would not contain much information so lets look at the case when G is dense so $e(G) \geq kn^2$. Now we can ask the following question.

Question 43. *Suppose $|A| = n$, $G = G[A, A]$ is dense that is $e(G) \geq kn^2$ for some $K > 0$ and suppose $|A +_G A| \leq cn$ for some absolute constant c.then can we conclude that $|A + A| \leq c_1 n$ for some c_1*

NO! Consider $A = [1, n] \cup R$ where R is a random subset of $[1, n^2]$ where each $x \in [1, n^2]$ is picked uniformly and independently with probability $\frac{1}{n}$. Now $E[|A|] \leq 2n$. Let G be the graph corresponding only to the sub sum set $[1, n] + [1, n]$, then $|A +_G A| = 2n - 1$. Also

$$\mathbb{P}(i \in R + R) = \frac{1}{n^2}(\lfloor \frac{k}{2} \rfloor - 1) + \frac{1}{n}.$$

So $E|R + R| \geq O(n^2)$. This random subset dense partial sum is small where as the overall sum set size is large.

But observe that A had a subset of half its size which was structured that is $|A' + A'| = 2|A'|$. So this motivates the following question.

Question 44. *If $|A +_G A| = O(n)$, then is there a large structured subset of A?*

YES! In fact the following theorem is true.

Theorem 45. *Suppose $0 < k < 1, c > 0$ are reals then there exist c', c'' (depending on c,k) such that the following holds: $\exists n_0(c, k)$ such that if $n \geq n_0$ and suppose $|A| = n, e(G) \geq kn^2$ and $|A +_G A| \leq cn$ then $\exists A' \subseteq A$ with $|A'| \geq c'|A|$ and $|A' + A'| \leq c''n$*

We will prove a generalization of this in this lecture. Its true even for two sets A, B and $A +_G B$.

Proof. Since we are looking for a property which holds generally for all additive sets A and does not depend on which numbers are chosen in A, so we cant use much of the number theoretic structure other than the basic properties of numbers.

The following combinatorial observation gives some hope of relating the structure of the graph G to the sumset $|A + B|$.

Observation 46. *We can construct an injection from 3-length paths in G to elements in A+B. If $x, x', x'' \in A +_G B$ and $x = a + b', x' = a' + b, x'' = a' + b'$ then $a + b = x - x' + x''$. So now for every 3 length path in G there exists a unique (a, b) which corresponds to a unique element in $A + B$. We know an upper bound on the number of 3 length paths in G. So if we can lower bound the number of 3 length paths corresponding to each element in A+B then we can get an upper bound on the size of $A + B$.*

So it would be good to have a lower bound on the number of paths of length 3 between every two vertices. Since we have a dense bipartite graph we expect many paths of length 3 between most pairs of vertices but not all pairs so we might need to refine our vertex sets to A' and B' so that this property holds.

To be precise say we can find $A' \subseteq A, B' \subseteq B$ such that $|A'| \geq c'n, |B'| \geq c''n$ and between every pair $a \in A', b \in B'$ there are $\Omega(n^2)$ paths of length 3. then for fixed a, b,

$$|\{(a', b') \in A \times B : (a, b'), (a', b'), (a', b) \in G\}| = \Omega(n^2) \tag{4.1}$$

$$|\{(x,x',x'') \in (A+_G B)^3 : x-x'+x'' = a+b, \exists b' \in B \text{ st } x = a+b', \exists a' \in A \text{ st } x'' = a'+b\}| = \Omega(n^2) \tag{4.2}$$

Now we forget a,b and just look at a+b,

$$|\{(x,x',x'') \in (A +_G B)^3 : x - x' + x'' = a + b\}| = \Omega(n^2) \tag{4.3}$$

Total number of triples $(x, x', x'') \in (A +_G B)^3 \le c^3 n^3$.
Thus

$$c^3 n^3 \ge \sum_{y \in A'+B'} \#\{(x, x', x''|y = x - x' + x'')\} \tag{4.4}$$

$$\ge \Omega(n^2)|A' + B'| \tag{4.5}$$

Thus we get $|A' + B'| = O(n)$.

Note : In equation(4.2) we are not looking at triples of edges of G but at triples of elements from $A +_G B$. This is providing us the freedom to look at only $a + b$ and not (a, b).
So the following question now arises:

Question 47. *If $G = G[A, B]$ is bipartite, $|A| = |B| = n$ and $e(G) \ge kn^2$ does there exist $A' \subseteq A, B' \subseteq B, |A'| \ge c'n, |B'| \ge c''n$ such that $\forall a \in A', b \in B'$ there are $\Omega(n^2)$paths $a-b'-a'-b$ from a to b where $a' \in A, b' \in B$?*

As we need more paths first lets enrich A by removing all the vertices which have degree $\le \frac{1}{2}$avg degree. So let

$$A_1 = \{a \in A|d(a) \ge \frac{kn}{2}\} \tag{4.6}$$

So

$$e(A\ A_1, B) \le \frac{kn^2}{2} \tag{4.7}$$

$$e(A, B) \ge \frac{kn^2}{2} \tag{4.8}$$

$$e(A_1, B) \le |A_1||B| = n|A_1| \tag{4.9}$$

$$|A_1| \ge \frac{e(A_1, B)}{n} \ge \frac{kn}{2} \tag{4.10}$$

Now say we have the following theorem,

Theorem 48. *$\exists A' \subseteq A, |A'| \ge \alpha k|A|$ and every pair $\{a, a'\}$ in A' has $\ge \delta|B|$ common neighbors.*

Lets apply this theorem on $A_1$to get $U \subseteq A_1, |U| \ge \alpha|A_1|$ such that every pair $\{a, a'\}$ in U have $\ge \delta|B|$ common neighbors. So $|U| \ge \frac{\alpha k}{2} n$.

Now we choose B_1 to have the vertices with large degree w.r.t 'U' so that we can have many choices for the 3rd edge. If $|B_1| = \Omega(n)$, we are through.
So for some μ TBD, let

$$B_1 = \{b \in B|d_U(B) \ge \mu|U|\} \tag{4.11}$$

$$\begin{aligned} e(B\ B_1, U) &\leq \mu|U|n & (4.12)\\ e(U,B) &\geq |U|(\frac{kn}{2}) & (4.13)\\ e(U,B_1) &\geq (\frac{k}{2}-\mu)|U|n & (4.14)\end{aligned}$$

For instance if $\mu = \frac{k}{4}$ then we have

$$\begin{aligned} |B_1||U| &\geq e(U,B_1) \geq (\frac{k}{4})|U|n & (4.15)\\ |B_1| &\geq \frac{kn}{4} & (4.16)\end{aligned}$$

Claim : (U, B_1) will do the job

Each $b \in B_1$, b has $\geq \frac{k|U|}{4}$ neighbors in U. So b has $\geq \frac{k|U|}{4} - 1$ neighbors in U that are not a.For each $a' \in N(b), a' \neq a$, there exist $\geq \delta n$ common neighbors of $a, a' \in B$, so $\delta n - 1$ common neighbors not including b. So the total number of 3 length paths from a to b are

$$\geq (\frac{k}{4}|U| - 1)(\delta n - 1) \geq \frac{k\delta}{16}|U|n \geq \frac{\alpha k^2 \delta}{32} n^2 \qquad (4.17)$$

thus we are done if we prove theorem (48).

BOMBSHELL: UNFORTUNATELY "THEOREM (48) is FALSE"

An explicit counter example due to Kostochka-Sudakov(2003). We ask the next natural question.

Question 49. *However if we replcae the words 'EVERY PAIR' by almost all pairs can we salvage the theorem?*

That is

Theorem 50. *$G = G[A,B], e(G) = k|A||B|. \exists A' \subseteq A$ such that $|A'| \geq \alpha k|A|$ and $\geq (1-\epsilon)|A'|^2$ ordered pairs of A' have $\geq \delta|B|$ common neighbours.*

If thorem(50) holds then how do we prove the statement we want ?

Let U, B be chosen as before. U this time is guaranteed by theorem (50). So

$$b \in B_1 \Longrightarrow d_U(b) \geq \frac{k}{4}|U| \qquad (4.18)$$

Now we need to refine U so let

$$A' = \{a \in U | a \text{ is in } \leq \frac{k}{8}|U| \text{ BAD pairs }\}. \qquad (4.19)$$

Now we want to show $|A'| \geq \lambda|U|$for some $\lambda > 0$, so that instead of U we use A'.

Total number of BAD ordered pairs in U is $\leq \epsilon|U|^2$. So the total number of BAD pairs featuring $U\ A'$ is $\frac{k}{8}|U|(|U|-|A'|)$, So

$$\begin{aligned}
\epsilon|U|^2 &\geq \frac{k}{8}|U|(|U|-|A'|) && (4.20)\\
\frac{8\epsilon}{k}|U| &\geq |U|-|A'| && (4.21)\\
|A'| &\geq (1-\frac{8\epsilon}{k})|U| && (4.22)
\end{aligned}$$

So for instance if we choose $\epsilon = \frac{k}{16}$ then $|A'| \geq \frac{1}{2}|U|$.

So now total number of 3 length paths from a to b is

$$\geq (\frac{k}{4}|U| - 1 - \frac{k}{8}|U|)(\delta n - 1) \geq \frac{k\delta}{32}|U|n \geq \frac{\alpha k^2\delta}{64}n^2 = \Omega(n^2) \tag{4.23}$$

Thus we are just left with proving theorem (50).

Note: theorem (50) is true.

Lets restate the theorem.

Theorem 51. *$G = G[A,B], e(G) = k|A||B|$. Given $0 < \epsilon < 1, \exists A' \subseteq A$ such that*

- *$|A'| \geq \alpha k|A|$ and (for some fixed $\alpha > 0$)*
- *Except $\epsilon|A'|^2$ ordered pairs of A, the rest of them each have $\geq \delta|B|$ common neighbors in B for some $\delta > 0$.*

The basic heuristic is that you need to sample in such a way that the properties force the object we to get sampled. So here we know that all $a \in N(b)$ have b as a common neighbor. So lets try and see the following

So pick $b \in B$ uniformly and let $X = |N(b)|, A' = N(b)$. Then

$$E[X] = \frac{1}{|B|}\sum_{b\in B} d(B) = k|A| \tag{4.24}$$

Using Cauchy- Schwarz we also have

$$E[X^2] \geq E[X]^2 = k^2|A|^2 \tag{4.25}$$

$$\text{BAD } = \#\{(a,a')|(a,a') \text{ is a BAD pair }\} \tag{4.26}$$

$$\begin{aligned}
E[|\text{ BAD }|] &= \sum_{(a,a')\in \text{ BAD(A)}} P(\text{b is chosen from a set of size } \leq \delta|B|) && (4.27)\\
&\leq \delta|A|^2 && (4.28)
\end{aligned}$$

Our goal is to find a b such that $|BAD| \leq \epsilon N(b)^2$ and $|N(b)| = \Omega(|A|)$. So lets look at

$$E[X^2 - \frac{1}{\epsilon}|BAD|] \geq k^2|A|^2 - \frac{\delta}{\epsilon}|A|^2 \tag{4.29}$$

Lets choose $\delta = \frac{\epsilon k^2}{2}$, so that

$$E[X^2 - \frac{1}{\epsilon}|BAD|] \geq \frac{k^2|A|^2}{2} \tag{4.30}$$

So $\exists$ b such that

$$N(b)^2 - \frac{1}{\epsilon}|BAD| \geq \frac{k^2|A|^2}{2} \tag{4.31}$$

Thus we have

$$|BAD| \leq \epsilon|N(b)|^2 = \epsilon|A'|^2 \tag{4.32}$$

$$|A'| = |N(b)| \geq \frac{k|A|}{\sqrt{2}} \tag{4.33}$$

this proves theorem(51) and thus proves the Balog-Szemerédi-Gowers theorem. □

5 The Second Moment Method

The method of using expectation of random variablesis a very useful and powerful tool, and its strength lies in its ease. 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.

5.1 Variance of a Random Variable and Chebyshev's theorem

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

Theorem 52 (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. $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \geq \lambda^2 \mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda)$. □

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 simplicity of its applicability.

If $X = X_1 + X_2 + \cdots + X_n$, 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 $X = X_1 + X_2 + \cdots + X_n$, we have

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

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

5.2 Applications

We first present a theorem due to Hardy-Ramanujan in Number theory, with a probabilistic proof by Turán.

Theorem 53. *Let $\omega(n) \to \infty$. Let $\nu(n)$ denote the number of primes dividing n not counting multiplicity. Then the number of positive integers x in $\{1, 2, \ldots, n\}$ such that*

$$|\nu(x) - \log\log n| > \omega(n)\sqrt{\log\log n}$$

is $o(n)$.

Proof. The idea of the proof germinates from the following simple device. Suppose we pick an integer x uniformly at random from $\{1, 2, .., n\}$. For a prime p, denote by X_p the following random variable:

$$\begin{aligned} X_p &:= 1 \qquad \text{if } p|x \\ &:= 0 \qquad \text{otherwise.} \end{aligned}$$

Note that $\nu(X) = \sum_p X_p$, where the sum is over all primes p less than n. Now, note that $\mathbb{E}(X_p) = \frac{\lfloor n/p \rfloor}{n} = 1/p + O(1/n)$. Hence, $\mathbb{E}(\nu(X)) = \sum_{p \le n} \left(\frac{1}{p} + O(\frac{1}{n}\right) = \log\log n + O(1)$. The last equality follows from a standard exercise in analytic number theory and is an application of the Abel Summation formula.

Thus, to 'estimate' $\nu(x)$ we could approximate $\nu(x)$ by its average. Chebyshev's inequality now gives a 'reasonable' interval of estimation. In order to do that, it remains to first calculate $\text{Var}(X_p), Cov(X_p, X_q)$.
It is easy to see that $\text{Var}(X_p) = \frac{1}{p}(1 - \frac{1}{p}) + O(\frac{1}{n})$, so $\sum_p \text{Var}(X_p) = \log\log n + O(1)$.

Now in order to calculate the covariances, note that for distinct primes p, q, we have $X_p X_q = 1$ if and only if $pq|x$. So,

$$\begin{aligned} Cov(X_p, X_q) &= \mathbb{E}(X_p X_q) - \mathbb{E}(X_p)\,\mathbb{E}(X_q) \\ &= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n}\frac{\lfloor n/q \rfloor}{n} \\ &\le \frac{1}{pq} - (\frac{1}{p} - \frac{1}{n})(\frac{1}{q} - \frac{1}{n}) \le \frac{1}{n}(\frac{1}{p} + \frac{1}{q}). \end{aligned}$$

Hence

$$\sum_{p\neq q} Cov(X_p, X_q) \leq \frac{1}{n}\sum_{p\neq q}\frac{1}{n}(\frac{1}{p}+\frac{1}{q}) \leq \frac{O(1)}{\log n}\sum_{p\leq n}\frac{1}{p} \leq \frac{O(1)\log\log n}{\log n} = o(1).$$

Here, we use a well known (but relatively non-trivial) result that $\pi(x) = O(\frac{x}{\log x})$, where $\pi(x)$ denotes the arithmetic function that counts the number of primes less than or equal to x. This result of Chebyshev, is much weaker than the Prime number theorem which actually says that the constant in question is asymptotically equal to 1. A similar calculation also gives $\sum_{p\neq q} Cov(X_p, X_q) \geq -o(1)$.

Hence, Chebyshev's inequality gives

$$\mathbb{P}(|\nu(X) - \log\log n| > \omega(n)\sqrt{\log\log n}) \leq \frac{1}{\omega^2(n)} = o(1).$$

□

Turán's proof makes the argument simpler by letting x be chosen at random in $\{1, 2, .., M\}$ where $M = n^{1/10}$, say. This avoids using that $\pi(x) = O(\frac{x}{\log x})$; on the other hand it gives an interval estimate for $\nu(n) - 10$. But asymptotically this result is the same as that of the statement in the Hardy-Ramanujan theorem.

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 moot 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 $\mathbb{E}(X_n)$ goes to infinity then Chebyshev's inequality gives us

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

In particular, X_n is of the 'around $\mathbb{E}(X)$' with very high probability.

For the next application, we need a definition.

Definition 54. *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 $\mathbb{E}(X_S) = \frac{1}{2}(x_1 + x_2 + \cdots + x_k)$. Similarly, $\text{Var}(X_S) = \frac{1}{4}(x_1^2 + x_2^2 + \cdots + x_k^2) \leq \frac{n^2 k}{4}$, so by Chebyshev we have

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

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

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

Optimizing for λ we get

Proposition 55. $f(n) \leq \log_2 n + \frac{1}{2}\log_2 \log_2 n + O(1)$.

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

The Rödl 'Nibble' refers to a probabilistic paradigm (discovered by Vojtech Rödl) for a host of applications in which a desirable is constructed by random means, not in one shot, but rather, by several small steps, with a certain amount of 'control' over each step. Subsequent researchers realized that Rödl's methodology extended to a host of applications, particularly for coloring problems in graphs and matchings/coverings in hypergraphs. Indeed, the Erdős-Hanani conjecture is an instance of a covering problem of a specific hypergraph. We present here, a version of a theorem that resolves the conjecture proved by Pippenger and Spencer.

Definition. Consider the set $[n] = \{1, \ldots, n\}$, and let $n \geq r \geq t$ be positive integers.
1. An r-uniform covering of $[n]$ 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 $T \subset A$.
2. An r-uniform packing of $[n]$ 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 $r|n$, then there obviously exists a collection $\mathcal{A}$ of r-subsets of $[n]$, $|\mathcal{A}| = \frac{n}{r}$, 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 $\lceil \frac{n}{r} \rceil$ and a packing of size $\lfloor \frac{n}{r} \rfloor$.

Let $M(n, r, t)$ be the size of a minimum covering, and $m(n, r, t)$ be the size of a maximum packing. Idealistically, there exists a collection $\mathcal{A}$ of r-subsets of $[n]$, $|\mathcal{A}| = \binom{n}{t}/\binom{r}{t}$, 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. The number $\binom{n}{t}/\binom{r}{t}$ comes from the observation that for each t-subset, there is a unique r-subset containing it, and for each r-subset, it contains $\binom{r}{t}$ t-subsets. Hence, we have the inequality $m(n,r,t) \leq \binom{n}{t}/\binom{r}{t} \leq M(n,r,t)$.

When $t = 2$, Erdős-Hanani (1960's) proved that

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

They further conjectured that this is true for all positive integers $r \geq t$. This conjecture was settled affirmatively by Vojtech Rödl in 1985.

Here, we consider a more general problem. Given an r-uniform hypergraph $\mathcal{H}$ on n vertices which is D-regular for some D, i.e. $\deg(x) = D\ \forall x \in V$, where $\deg(x) = |\{E \in \mathcal{E} : E \ni x\}|$, and $\mathcal{E}$ = set of hyperedges in $\mathcal{H}$, we want to find a covering and a packing out of $\mathcal{H}$, both of sizes $\approx \frac{n}{r}$. This is more general since we can treat the t-subsets in the conjecture as vertices in this problem, and all the degrees of these vertices are equal.

Note that in this new problem, if we can find a packing of size $\frac{(1-\epsilon)n}{r}$, then there are at most ϵn vertices uncovered. Hence, we can find a covering of size $\leq \frac{(1-\epsilon)n}{r} + \epsilon n = (1-\epsilon+r\epsilon)\frac{n}{r} \xrightarrow{\epsilon\to 0} \frac{n}{r}$. On the other hand, if we can find a covering $\mathcal{A}$ of size $\frac{(1+\epsilon)n}{r}$, then for every x which is covered by $\deg(x)$ hyperedges, we delete $\deg(x)-1$ of them. The number of deleted hyperedges $\leq \sum_{x\in V}(\deg(x)-1) = \sum_{x\in V}\deg(x) - \sum_{x\in V} 1 = |\{(x,E) : E \in \mathcal{A}\}| - n = \frac{(1+\epsilon)n}{r} \times r - n = \epsilon n$. Hence, we can find a packing of size $\geq \frac{(1+\epsilon)n}{r} - \epsilon n = (1+\epsilon-r\epsilon)\frac{n}{r} \xrightarrow{\epsilon\to 0} \frac{n}{r}$. Therefore, finding a covering of size $\approx \frac{n}{r}$ is equivalent to finding a packing of size $\approx \frac{n}{r}$.

In the following, we try to obtain a covering $\mathcal{A}$ of size $\leq (1+\epsilon)\frac{n}{r}$ for all large n when an $\epsilon > 0$ is given. Since we seek a covering, we do not have to worry if some vertex is covered more than once, so this is a freer thing to attempt than to obtain a packing.

Let us try a bare probabilistic idea first to see what shortcomings it has. 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 $p \approx \frac{1}{D}$. Now let us see how many vertices get left behind by this probabilistic scheme. Again, a vertex x is uncovered, only if every edge containing it is discarded. In other words, the probability that a vertex x goes covered 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. So, it is clear that a single shot pick for $\mathcal{E}^*$ would not work.

Rödl's idea. First, pick a small number of hyperedges, so that the rest of $\mathcal{H}$ is as close as possible to the original one. If, by taking T such "nibbles", we are left with δn vertices for some small δ, we cover each of these vertices with a hyperedge to finish the problem.

As $\mathcal{H}$ is D-regular, $r|\mathcal{E}| = |\{(x,E) : x \in E \in \mathcal{E}\}| = nD \Rightarrow |\mathcal{E}| = \frac{nD}{r}$. So if we want to pick around $\frac{\epsilon n}{r}$ edges in each nibble, then we can use $\mathbb{P}(E \text{ is picked}) = \frac{\epsilon}{D}$.

Probability Paradigm. Each edge $E \in \mathcal{E}$ is picked independently with probability $p = \frac{\epsilon}{D}$. Let $\mathcal{E}^*$ be the set of picked edges. By the choice of p, $\mathbb{E}[|\mathcal{E}^*|] = \frac{\epsilon n}{r}$.

Notation. $x = a \pm b \Leftrightarrow x \in (a-b, a+b)$.

After a "nibble", the rest of the hypergraph is no longer regular, so we modify the problem. Given an r-uniform hypergraph $\mathcal{H}$ on n vertices such that $\deg(x) = D(1 \pm \delta)$ $\forall x \in V$ for some small $\delta > 0$, we want to find a covering of size $\approx \frac{n}{r}$.

We further modify the problem. Given an r-uniform hypergraph $\mathcal{H}$ on n vertices such that except at most δn vertices, $\deg(x) = D(1 \pm \delta)$ for other vertices $x \in V$, suppose that $\forall x \in V$, $\deg(x) < kD$ for some constant k. We want to find a collection of hyperedges $\mathcal{E}^*$ such that
(i) $|\mathcal{E}^*| = \frac{\epsilon n}{r}(1 + \delta')$ for some $\delta' > 0$;
(ii) $|V^*| = ne^{-\epsilon}(1 \pm \delta'')$ for some $\delta'' > 0$, where $V^* := V \backslash \Big(\bigcup\limits_{E \in \mathcal{E}^*} E \Big)$.

Let $\mathbf{1}_x = \mathbf{1}_{\{x \notin \text{ any edge of } \mathcal{E}^*\}}$. The motivation of (ii) is $|V^*| = \sum\limits_{x \in V} \mathbf{1}_x \Rightarrow \mathbb{E}[|V^*|] = \sum\limits_{x \in V} (1 - \frac{\epsilon}{D})^{\deg(x)} \approx n(1-\delta)(1-\frac{\epsilon}{D})^{D(1\pm\delta)} \approx n(1-\delta)e^{-\epsilon(1\pm\delta)} \approx ne^{-\epsilon}(1 \pm \delta'')$. $\mathrm{Var}(|V^*|) = \mathrm{Var}\Big(\sum\limits_{x \in V} \mathbf{1}_x\Big) = \sum\limits_{x \in V} \mathrm{Var}(\mathbf{1}_x) + \sum\limits_{x \neq y} \mathrm{Cov}(\mathbf{1}_x, \mathbf{1}_y)$. If $\deg(x) = D(1 \pm \delta)$ and $\deg(y) = D(1 \pm \delta)$, then

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

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

Theorem. (Pippenger-Spencer) Suppose $r \geq 2$ is a positive integer, and $k, \epsilon, \delta^* > 0$ are given. Then there exist $\delta_0(r,k,\epsilon,\delta^*) > 0$ and $D_0(r,k,\epsilon,\delta^*)$ 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, $\deg(x) = D(1 \pm \delta)$ for other vertices $x \in V$;
(ii) $\forall x \in V$, $\deg(x) < kD$;
(iii) $\deg(x,y) < \delta D$,
then there exists $\mathcal{E}^* \subset \mathcal{E}$ such that
(a) $|\mathcal{E}^*| = \frac{\epsilon n}{r}(1 \pm \delta^*)$;

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

(c) except at most $\delta^*|V^*|$ vertices, $\deg^*(x) = De^{-\epsilon(r-1)}(1 \pm \delta^*)$, where $\deg^*$ 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) are true. This lemma, in short, says that $\mathcal{H}$ is $(n, k, D, \delta) \Rightarrow$ there exists an induced hypergraph $\mathcal{H}^*$ which is $(ne^{-\epsilon}(1 \pm \delta^*), ke^{\epsilon(r-1)}, De^{-\epsilon(r-1)}, \delta^*)$. The parameter k^* is due to $\deg^*(x) \le \deg(x) < kD = k^*D^* \Rightarrow k^* = \frac{kD}{De^{-\epsilon(r-1)}} = ke^{\epsilon(r-1)}$, and the parameter δ follows $\deg^*(x,y) \le \deg^*(x,y) < \delta D = \delta^* D^* \Rightarrow \delta = \frac{\delta^* De^{-\epsilon(r-1)}}{D} = \delta^* e^{-\epsilon(r-1)}$. In other words, $\delta^* = \delta e^{\epsilon(r-1)}$.

Repeat the "nibble" t times (t to be determined) will give $\delta = \delta_0 < \delta_1 < \cdots < \delta_t$ with $\delta_i = \delta_{i-1}e^{\epsilon(r-1)}$, and $\mathcal{H} = \mathcal{H}_0 \supset \mathcal{H}_1 \supset \cdots \supset \mathcal{H}_t$. Note that this establishes a cover of size $\sum\limits_{i=1}^{t-1} |\mathcal{E}_i| + |V_t|$. By

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

and

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

the size of this cover is

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

Pick t such that $e^{-\epsilon t} < \epsilon$. For this t, pick δ small enough such that $\prod\limits_{i=1}^{t} (1 + \delta_i) \le 1 + \epsilon$. Then the size of the cover $\xrightarrow{\epsilon \to 0} \frac{n}{r}$ from above. Therefore, all that remains is to prove the "Nibble" Lemma.

Proof of "Nibble" Lemma.

Probability Paradigm. Each edge of $\mathcal{H}$ is picked independently with probability $\frac{\epsilon}{D}$. Let $\mathcal{E}^*$ be the set of picked edges.

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

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

On the other hand,

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

So $\frac{(1-\delta)^2Dn}{r}\le|\mathcal{E}|\le\frac{Dn}{r}(1+(k+1)\delta)\Rightarrow|\mathcal{E}|=\frac{Dn}{r}(1\pm\delta_{(1)})$.

$\mathbb{E}[|\mathcal{E}^*|]=\sum_{E\in\mathcal{E}}\mathbb{P}(E\text{ is picked})=\frac{\epsilon}{D}\frac{Dn}{r}(1\pm\delta_{(1)})=\frac{\epsilon n}{r}(1\pm\delta_{(1)})$. Let $\mathbf{1}_E=\mathbf{1}_{\{E\text{ is picked}\}}$. By independence, $\mathrm{Var}(|\mathcal{E}^*|)=\sum_{E\in\mathcal{E}}\mathrm{Var}(\mathbf{1}_E)\le\mathbb{E}[|\mathcal{E}^*|]$. By Chebyshev's inequality, we get $\mathbb{P}\left(\Big||\mathcal{E}^*|-\mathbb{E}[|\mathcal{E}^*|]\Big|>\delta_{(2)}\mathbb{E}[|\mathcal{E}^*|]\right)\le\frac{\mathrm{Var}(|\mathcal{E}^*|)}{\delta_{(2)}^2\mathbb{E}[|\mathcal{E}^*|]^2}$. So if $n\gg0$, then $|\mathcal{E}^*|=\frac{\epsilon n}{r}(1\pm\delta_{(1)})(1\pm\delta_{(2)})=\frac{\epsilon n}{r}(1\pm\delta^*)$ with high probability, yielding (a).

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

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

On the other hand,

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

So $ne^{-\epsilon}(1-\delta_{(3)})(1-\delta)\le\mathbb{E}[|V^*|]\le ne^{-\epsilon}(1+\delta_{(4)}+\delta e^{\epsilon})$, implying $\mathbb{E}[|V^*|]=ne^{-\epsilon}(1\pm\delta_{(5)})$.

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

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

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

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

So if $n\gg0$, $|V^*|=ne^{-\epsilon}(1\pm\delta_{(5)})(1\pm\delta_{(6)})=ne^{-\epsilon}(1\pm\delta^*)$ with high probability, yielding (b).

Suppose x survives after the removal of $\mathcal{E}^*$, and let $E \ni x$. We want to estimate

$$\mathbb{P}(E \text{ survives} \mid x \text{ survives}).$$

Let $\mathcal{F}_E = \{F \in \mathcal{E} : x \notin F, F \cap E \neq \emptyset\}$. Then E survives if and only if $\mathcal{F}_E \cap \mathcal{E}^* = \emptyset$.

Recall that x is good if $\deg(x) = (1 \pm \delta)D$, else x is bad. Call $E \in \mathcal{E}$ bad if E contains some bad vertices. Suppose x is good, and E is good. Then

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

Let $\mathrm{Bad}(x) = \{E \ni x : E \text{ is bad}\}$. If $|\mathrm{Bad}(x)| < \delta_{(9)} D$, then $\mathbb{E}[\deg^*(x)] = De^{-\epsilon(r-1)}(1 \pm \delta_{(10)})$.

So 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 find 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, $|V_{\mathrm{INCOR}}| \leq \frac{r(\delta n)k}{\delta_{(9)}} = \delta^* n$. Therefore, except at most $|\delta^* n|$ vertices, the remaining very good vertices x satisfy $\mathbb{E}[\deg^*(x)] = De^{-\epsilon(r-1)}(1 \pm \delta_{(10)})$.

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

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

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

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

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

Let $N = |\{x \text{ good} : \deg^*(x) \neq e^{-\epsilon(r-1)} D(1 \pm \delta^*)\}|$. Then we can use Markov's inequality to show that $\mathbb{E}[N] < \delta_{(11)} n$, so all except $\delta^* n$ vertices satisfy (c).

6 Bounding Large Deviations - The Chernoff Bounds

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. We consider a few prototypes of such results and a few combinatorial applications. One of the first such results is the following:

Proposition 56 (Chernoff Bound). *Let $X_i \in \{\pm 1\}$ 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, $\mathbb{P}[S_n > a] < e^{-a^2/2n}$.*

Proof. Consider $e^{\lambda X_i}$, with λ to be optimized. Then $\mathbb{E}[e^{\lambda X_i}] = (e^{\lambda} + e^{-\lambda})/2 = \cosh(\lambda)$. Taking the Taylor expansion, we see that

$$\mathbb{E}\left[e^{\lambda X_i}\right] = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} < \sum_{k=0}^{\infty} \frac{(\lambda^2/2)^k}{k!} = e^{\lambda^2/2}$$

Since the X_i are independent,

$$\mathbb{E}\left[e^{\lambda S_n}\right] = \mathbb{E}\left[e^{\sum \lambda X_i}\right] = \prod_i \mathbb{E}[e^{\lambda X_i}] = \cosh(\lambda)^n < e^{\lambda^2 n/2}$$

By Markov's Inequality,

$$\mathbb{P}\left[e^{\lambda S_n} > e^{\lambda a}\right] \leq \frac{\mathbb{E}[e^{\lambda S_n}]}{e^{\lambda a}} < e^{\lambda^2 n/2 - \lambda a}$$

Since $\mathbb{P}[S_n > a] = \mathbb{P}[e^{\lambda S_n} > e^{\lambda a}]$, we see that $\mathbb{P}[S_n > a] < e^{\lambda^2 n/2 - \lambda a}$. Optimizing this bound by setting $\lambda = a/n$, we see that $\mathbb{P}[S_n > a] < e^{\lambda^2 n/2}$, as desired. $\square$

Proposition 56 can be generalized and specialized in various ways. We state two such modifications here.

Proposition 57 (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*

$$\mathbb{P}[S_n > a] < e^{-2a^2/n} \quad \textit{and} \quad \mathbb{P}[S_n < -a] < 2e^{-2a^2/n}$$

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 58 (Chernoff Bound (Binomial Version))**.** *Let* $X \sim$ *Binomial*(n, p)*, and let* $0 \leq t \leq np$*. Then* $\mathbb{P}[|X - np| \geq t] < 2e^{-t^2/3np}$*.*

In all three cases, the independence assumption can be removed while preserving the exponential decay (although with a worse constant).

6.1 Projective Planes and Property B

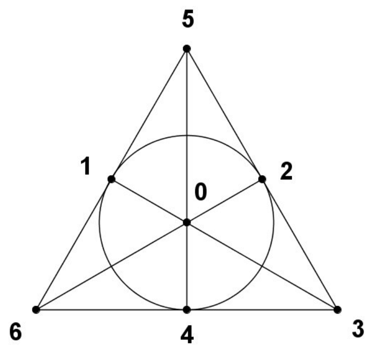
Given a hypergraph $\mathcal{H} = (V, \mathcal{E})$, we say that $\mathcal{H}$ has *property B* if there exists $S \subseteq V$ such that for all $E \in \mathcal{E}$ (i) $S \cap E \neq \emptyset$ and (ii) $E \not\subseteq S$.

Note that any 2-colorable hypergraph has property B, by letting S be the set of blue vertices.

Observation 59 (Lovász)**.** *If* $\mathcal{H}$ *is such that* $|E_1 \cap E_2| \neq 1$ *for all* $E_1, E_2 \in \mathcal{E}$*, then* $\mathcal{H}$ *is 2-colorable, and therefore has property B.*

Proof. 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. □

We now consider the opposite assumption. Suppose that every pair of edges meet at exactly 1 vertex. The *Fano Plane*, shown here with each edge represented as a line, shows that such a hypergraph need not be 2-colorable. The Fano Plane is an example of a projective plane.



Definition 60 (Projective Plane $\P_n$)**.** *The* projective plane of order n*, denoted* $\P_n = (\mathcal{P}, \mathcal{L})$*, is an* $n+1$*-uniform hypergraph such that*

1. *For all* $x \in \mathcal{P}$*,* $\#\{L \in \mathcal{L} \,|\, x \in L\} = n + 1$
2. *For all* $L_1, L_2 \in \mathcal{L}$*,* $L_1 \neq L_1$*,* $|L_1 \cap L_2| = 1$
3. *For all* $x, y \in \mathcal{P}$*,* $x \neq y$*, there exists a unique* $L \in \mathcal{L}$ *such that* $x \in L$ *and* $y \in L$*.*

The elements of $\mathcal{P}$ are referred to as "points" and the elements of $\mathcal{L}$ are referred to as "lines". Using this terminology, we see that requirements (2) and (3) in the definition can be rephrased as (2) Any two lines meet at unique point and (3) Any two points define a line.

We now define a strengthening of Property B, which we will refer to as Property B(s).

Definition 61 (Property B(s))**.** *A hypergraph $\mathcal{H} = (V, \mathcal{E})$ has* property B(s) *if there exists $S \subseteq V$ such that for every $E \in \mathcal{E}$, $0 < |E \cap S| \leq s$.*

For n-uniform hypergraphs, we obtain the original property B by letting $s = n - 1$.

Conjecture 62 (Erdős)**.** *There exists a constant c such that for all n, $\P_n$ has property $B(s)$.*

Theorem 63 (Erdős, Silverman, Steinberg)**.** *There exist constants k, K such that for all n there exists $S \subseteq \P_n$ 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, $\mathbb{P}[|S_L - f(n)| > f(n)] < 2e^{-f(n)/3}$. Since $\P_n$ contains $n^2 + n + 1$ lines,

$$\mathbb{P}[\exists L \text{ st } |S_l - f(n)| > f(n)] < (2e^{-f(n)/3})(n^2 + n + 1) < 4e^{-f(n)/3}n^2$$

Therefore, if $e^{f(n)/3} > 4n^2$, the desired S exists.

Solving for $f(n)$, we see that $f(n) = 3 \log 4n^2 \approx 6 \log n$, as desired. □

6.2 Graph Coloring and Hadwiger's Conjecture

Definition 64 (Graph Minor)**.** *Given a graph G, H is a* minor *of G if H can be obtained from G by*

1. *Deleting edges and vertices*
2. *Contracting an edge*

Definition 65 (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.

Conjecture 66 (Hadwiger's Conjecture)**.** *Let G be a graph with $\chi(G) \geq p$. Then G contains K_p as a minor.*

By the Robertson-Seymour Theorem, the property of being p-colorable is exactly characterized by a finite set of forbidden minors. Hadwiger's Conjecture therefore says that the sole forbidden minor is K_p.

Hadwiger's Conjecture is notoriously diffcult. Indeed, the special case of $p = 5$ implies the four-color theorem. To see this, suppose that $\chi(G) \geq 5$. Then by the conjecture, G contains K_5

and is therefore nonplanar. With more work, the case of $p = 5$ can be shown to be equivalent to the four-color theorem.

The conjecture is currently open for $p > 6$. Although $p \leq 4$ can be proven directly, all known proofs 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 G contains K_p as a subdivision. This strengthened conjecture was shown to be false by Catlin via an explicit counterexample. Erdös and Fajtlowicz then showed that for large n, with probability approaching 1, a random graph has chromatic number almost quadraticly larger than the size of its largest complete minor.

Theorem 67 (Erdős, Fajtlowicz)**.** *There exist graphs G such that $\chi(G) \geq \frac{n}{3\lg n}$ and G has no $K_{3\sqrt{n}}$ subdivision.*

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 G has large chromatic number, and then show that G has no large K_p subdivision.

Bounding the Chromatic Number It is known that $\chi(G) \geq n/\alpha(G)$, where $\alpha(G)$ is the size of the largest independent set in G. As a result, it suffices to upper-bound $\alpha(G)$ in order to lower-bound $\chi(G)$.

We have

$$\begin{aligned}\mathbb{P}[\alpha(G) \geq x] &= \mathbb{P}[\exists \text{ a set of } x \text{ vertices which form an independant set}] \\ &\leq \binom{n}{x}\frac{1}{2^{\binom{x}{2}}} \leq \left(\frac{n}{2^{\frac{x-1}{2}}}\right)^x\end{aligned}$$

Let $x = 2\lg n + 3$. Then $2^{\frac{x-1}{2}} = 2n$, and so

$$\mathbb{P}[\alpha(G) \geq x] \leq \left(\frac{1}{2}\right)^{2\lg n + 3} = \frac{1}{8n^2}$$

So with high probability, $\alpha(G) \leq 2\lg(n) + 3 < 3\lg n$.

Bounding the Complete Subdivisions Now suppose that G contains $K_{c\sqrt{n}}$ as a subdivision. Since $K_{c\sqrt{n}}$ contains $\frac{(3\sqrt{n})(3\sqrt{n}-1)}{2} > 4n$ edges, G must contain that many disjoint paths. However, a vertex of G must either be a vertex of the $K_{3\sqrt{n}}$ subdivision, or else be contained in at most one of the paths. Since there are only n vertices in G, at least $3n$ of the paths must be single edges of G.

Fix a set $U \subset V$, $|U| = 3\sqrt{n}$, and let $e(U) = \sum_{u,v \in U} [\![(u, v) \in E]\!]$. If U forms the vertices of a $K_{3\sqrt{n}}$ subdivision, then $e(U) \geq 3n$. Since

$$\mathbb{E}[e(U)] = \frac{1}{2}\frac{(3\sqrt{n})(3\sqrt{n}-1)}{2} = \frac{9n - 3\sqrt{n}}{4}$$

by the Chernoff Bound we have

$$\mathbb{P}[|e(U) - \mathbb{E}[e(U)]| \geq \frac{1}{4}\mathbb{E}[e(U)]] \leq 2e^{-\mathbb{E}[e(U)]/48}$$

Therefore

$$\mathbb{P}[e(U) \geq 3n] \leq 2e^{-(9n-3\sqrt{n})/192} < e^{-n/25}$$

which means that

$$\mathbb{P}[U \text{ forms the vertices of a } K_{3\sqrt{n}} \text{ subdivision}] < e^{-n/25}$$

Since there are $\binom{n}{3\sqrt{n}}$ choices of U,

$$\mathbb{P}[\exists \text{ a } K_{3\sqrt{n}} \text{ subdivision}] < \binom{n}{3\sqrt{n}} e^{-n/25} < \left(\tfrac{e\sqrt{n}}{3}\right)^{3\sqrt{n}} e^{-n/25}$$

which approaches 0 as $n \to \infty$. So with high probability, G does not contain a $K_{3\sqrt{n}}$ subdivision.

Taking a union bound, we see that with high probability, $\chi(G) \geq \frac{n}{3 \lg n}$ and G has no $K_{3\sqrt{n}}$ subdivision, as desired. □

Note that the counterexamples constructed here require more than 10^5 vertices, highlighting the difficulty in constructing such counterexamples explicitly by a computer search. Perhaps also justifying why these conjectures were made since it is find such graphs with a small number of vertices.

This result shows that the chromatic number of a graph is a global property of the graph. As further evidence of this fact, Erdős also proved the following.

Theorem 68 (Erdős). *Given $\epsilon > 0$, 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 $\chi(H) \leq 3$.*

We will however not get into those details here.

7 Bounds for $R(3,k)$: 2 results

The first non-trivial problem of evaluating the Ramsey numbers is that of evaluating $R(3,k)$ for a fixed integer k. Far from determining these numbers precisely, it was apparent that even determining the asymptotic order of these integers was going to be a highly non-trivial problem. In this chapter, we shall consider two results - an upper bound and a lower bound. The upper bound is of the 'correct' order - a fact that was established later and is considerably more difficult. The lower bound we shall discuss here is weaker than the best result known, and was proved by Erdős in 1960.

7.1 Upper bound for $R(3,k)$

Note that $R(3,k)$ is the smallest n such that any red-blue coloring of K_n contains a red triangle or a blue K_k. If we retain only the red edges say, then $R(3,k)$ can be viewed as the minimum n such that every graph on n vertices has a triangle or an independent set of size k.

In general, $R(l,k) \leq^{k+l-2} C_{l-1} \leq ck^2$ for $l=3$. Erdős showed that $R(3,k) = \Omega((\frac{k}{\log k})^{3/2})$. Ajtai, Komlos, and Szemerédi proved that $R(3,k) = O(\frac{k^2}{\log k})$. The lower bound, namely, $R(3,k) = \Omega(\frac{k^2}{\log k})$ which turns out to be much harder to prove was first furnished by J-H Kim involving several deep inequalities, and an extremely involved proof. More recently, Bohman gave a different proof by considering a graph evolution process.

Turán's theorem states that $K_{r+1} \not\subset G \Rightarrow e(G) \leq (1-\frac{1}{r})\frac{n^2}{2}$. In terms of complements, $\alpha(G) \geq \frac{n}{\overline{d}+1}$ where $\overline{d}$ is the average degree of G. This can be proved as follows:

$$\frac{n\overline{d}}{2} = e(G) \geq \frac{n(n-1)}{2} - (1-\frac{1}{r})\frac{n^2}{2} = \frac{1}{2}(\frac{n^2}{r} - n)$$

$$\Rightarrow \overline{d} \geq \frac{n}{r} - 1 \Rightarrow r \geq \frac{n}{\overline{d}+1} \Rightarrow \alpha(G) \geq \frac{n}{\overline{d}+1}$$

Theorem 69 (Ajtai-Komlos-Szemeŕedi). *A triangle-free graph in $\frac{Ck^2}{\log k}$ vertices has an independent set of size k*

Proof. Note that if the average degree is small, Turán's bound should help. If not, we delete a vertex to try and reduce the average degree.

If G is triangle-free, then define $G' = G \backslash \{v\}$, $e' = e - \sum_{w \leftrightarrow v} d(w)$. There will be a reduction in the average degree of G' compared to G if v is chosen such that $\sum_{w \leftrightarrow v} d(w) \geq d.d(v)$. Call a vertex v a *groupie* if

$$\frac{1}{d(v)}\sum_{w\leftrightarrow v} d(w) \geq d$$

v is a groupie $\Leftrightarrow$

$$\sum_{w\leftrightarrow v} d(w) - d.d(v) \geq 0$$

$$g(v) = \sum_{w\in V} d(w)\mathbb{I}_{w\leftrightarrow v} - d.d(v) \geq 0$$

Pick v uniformly at random

$$E[g(v)] = \sum_{w\in V} d(w)\frac{d(w)}{n} - d\sum_{v\in V}\frac{d(v)}{n}$$

$$= \frac{1}{n}\sum_{w\in V} d^2(w) - d^2 \geq 0$$

by Cauchy-Schwarz Inequality. Hence, there is a groupie in every graph. Pick such a groupie. $G' = G\backslash\{v\}$. Define $f(n,d)$ as the size of the maximum independent set in a triangle free G with n vertices and average degree d. Note that we have $f(n,d) \geq 1 + f(n-1-d(v),d)$. Heuristically, suppose we denote the remaining number of vertices after t steps of this by $nR(t)$, we have $R(t+1) - R(t) \approx -\frac{d}{n}R(t)$. If we parametrize time so that the next step after time t occurs at $t + \frac{d}{n}$, then $R(t+\frac{d}{n}) - R(t) \approx -\frac{d}{n}R(t)$ giving us $R'(t) \approx -R(t)$.

We will start with a G. Define $f(n,d) = c\frac{n}{d}\log d$. Pick a groupie v, then delete v and all its neighbors. In the resulting graph G' we have $n' - 1 - d(v)$ and $e' = \frac{nd}{2} - \sum_{w\leftrightarrow v} d(w) \leq \frac{nd}{2} - d(v)d = (\frac{n}{2} - d(v))d$

$$d' = \frac{2e'}{n-1-d(v)} \geq \frac{n-2d(v)}{n-1-d(v)}d$$

By induction, we have

$$f(n,d) \geq 1 + f(n',d').$$

The authors check this for $c = 1/100$ to say that $R(3,k) < n \Leftrightarrow$ every triangle free graph on n vertices has an independent set of size k. If G is triangle free and for some v, $\Delta(v) \geq k$, then we are done. Suppose $\Delta(G_n) \leq k$, The authors also showed that $\alpha(G) > c\frac{n}{d}\log d$. If this bound is less than k we have

$$\frac{1}{100}\frac{n}{d}\log d < k \Rightarrow n < 100\frac{k^2}{\log k}.$$

□

7.2 Lower bound for $R(3,k)$

We know that $R(3,k) \geq n \Rightarrow \exists G_n$, a triangle free graph such that $\alpha(G_n) < k$. Pick edges of G at random with probability p. Let N_3 denote the number of triangles in G.

$$\mathbb{E}[N_3] = \binom{n}{3}p^3 < (np)^3/6,$$

so, we have

$$\mathbb{P}(\alpha(G) \geq x) \leq \binom{n}{x}(1-p)^{\binom{x}{2}} \leq \frac{n^x}{x!}e^{-p\binom{x}{2}} \leq \frac{e^{-px^2/3+x\log n}}{x!} \approx (e^{-px/3+\log n})^x.$$

Choose $px > 3\log n$, so that $(np)^3 = \Theta(n)$. Setting $p = \frac{1}{n^{2/3}}$ gives $\mathbb{E}[N_3] < n/6$. Now applying the Markov inequality gives $N_3 < n/3$ with probability greater than 1/2. Hence, with positive probability the following hold simultaneously:

- $N_3 < n/3$
- $\alpha(G) \leq 3n^{2/3}\log n$

Pick an assignment of edges such that these conditions both hold. From the resulting graph, throw away one vertex from each of the triangles it contains. We then have a graph G with $|V(G)| > \frac{2}{3}n$, no triangles and $\alpha(G) \leq 3n^{2/3}\log n = k$.

$$n = (\frac{k}{\log n})^{3/2} = (\frac{k}{\log(\frac{k}{\log n})^{3/2}})^{3/2} = \Omega((\frac{k}{\log k})^{3/2}).$$

Erdős improved the bound in a remarkable manner following the same asymptotics for $p = \frac{\epsilon}{\sqrt{n}}$ and $x = A\sqrt{n}\log n$ for a small ϵ and a large A. The fact that this paper was written as early as 1960 shows how ahead Erdős was in this game!

Theorem 70 (Erdős)**.** *There exists G_n, which is triangle free such that $\alpha(G) < A\sqrt{n}(\log n)$ for some large A. Consequently,*

$$R(3,k) = \Omega((\frac{k}{\log k})^2).$$

Proof. Let $G = G(n,p)$ with $p = \frac{\epsilon}{\sqrt{n}}$ and $x = A\sqrt{n}\log n$. Put all the edges in some order. Delete an edge only if it forms a triangle with the preapproved edges, to get a triangle free graph. Suppose $G(n,p)$ satisfies that every x-subset of V has an edge e_X which is not a part of any crossing triangle. Consider $G' \subseteq G$ which consists only of the edges e_X as X varies over all x-subsets of V. Now run the triangle freeing process. The resulting G'' has no independent sets of size x.

Suppose not, then the edge e_X is involved in some triangle. $\{e_X, e_{X'}, e_{X''}\} \subseteq X$. Fix an X of size x. We wish to calculate $P(X$ has no nice edge $e_X)$. Suppose there is a pair $\{u,v\} \subseteq X$ such that no $w \notin X$ is adjacent to both u and v. Then putting the edge $\{u,v\}$ makes it the choice of e_X. Call a pair $\{u,v\} \subseteq X$ bad for X if they are both adjacent to some $w \notin X$. The number of such bad pairs is at most $\sum_{y\notin X}\binom{d_X(y)}{2}$.

Define for $i \geq 0$

$$Y^{(i)} = \{y | \frac{\epsilon\sqrt{n}}{2^i} \leq d_X(y)\}$$

$$P(|Y^{(i)}| > \beta n \text{ for some } X \subseteq V(G) \text{ with } |X| = x)$$

$$\leq \binom{n}{A\sqrt{n}(\log n)}\binom{n}{\beta n}(\binom{A\sqrt{n}\log n}{\frac{\epsilon\sqrt{n}}{2^i}}(\frac{\epsilon}{\sqrt{n}})^{\frac{\epsilon\sqrt{n}}{2^i}})^{\beta n}$$

$$\leq e^{A\sqrt{n}(\log n)^2 + 2\beta n\log(\frac{1}{\beta}) + \frac{\epsilon\sqrt{n}\beta n}{2^i}(i\log 2 + \log\log n + O(1)) - \frac{\epsilon\sqrt{n}\beta n\log n}{2.2^i}}.$$

Thus, for $0 \leq i \leq \frac{\log_2 n}{4}$, note that if we set $\frac{\epsilon\sqrt{n}\beta n\log n}{2.2^i} = 8A\sqrt{n}(\log n)^2$ we have

- $\sqrt{n}(\log n)^2 >> \beta n \log(\frac{1}{\beta})$
- $\frac{\epsilon\sqrt{n}\beta n}{2^i}(i\log 2 + \log\log n + O(1)) < 5A\sqrt{n}(\log n)^2$

This gives us

$$\mathbb{P}(|Y^{(i)}| \geq \frac{16A}{\epsilon}2^i \log n) \leq e^{-(1/2)A\sqrt{n}(\log n)^2}.$$

For higher i, we make a different kind of estimate. Let

$$Y_{(i)} = \{y | 2^i A\epsilon \log n \leq d_X(y)\}, i \leq \frac{\log_2 n}{4}$$

This is again motivated by the observation that $E[d_X(y)] \approx A\epsilon \log n$. Clearly, almost surely $Y = (\bigcup_i Y^{(i)}) \cup (\bigcup_i Y_{(i)})$. Note that

$$\sum_{y \in Y^{(i)} \setminus Y^{(i-1)}} \frac{d_X^2(y)}{2} \leq \frac{|Y^{(i)}|}{2}(\frac{\epsilon\sqrt{n}}{2^{i-1}})^2 \leq \frac{16}{A}\frac{\epsilon^2 n \log n}{2^{i-2}}.$$

Hence,

$$\mathbb{P}(|Y_{(i)}| \geq \frac{n}{i^2 4^i} \text{ for some } X) \leq \binom{n}{A\sqrt{n}(\log n)}\binom{n}{\frac{n}{i^2 4^i}}(\binom{A\sqrt{n}(\log n)}{2^i\sqrt{A}(\log n)}(\frac{1}{\sqrt{An}})^{2^i\sqrt{A}(\log n)})^{\frac{n}{i^2 4^i}}$$

$$\leq \exp\left(A\sqrt{n}(\log n)^2 + \frac{n}{i^2 4^i}\log 8 - \frac{\sqrt{A}n\log n}{i2^i}\log 2\right)$$

which clearly goes to zero as $n \to \infty$. We see that $|Y_{(i)}| < \frac{n}{i^2 4^i}$ for all $0 \leq i \leq \frac{\log_2 n}{4}$, and hence $\sum_y \frac{d_X^2(y)}{2}$ is of the order of $(A\epsilon)^2 n(\log n)^2$. Thus if $\epsilon = \frac{1}{\sqrt{A}}$, for large A, this sum is bounded by $(\sum_i \frac{1}{i^2})An(\log n)^2$ and for large A, this is much smaller than $\frac{A^2}{5}n(\log n)^2$.

$$\mathbb{P}(X \text{has no } e_X) \leq \mathbb{P}(\text{No good pair in } X \text{ is an edge})$$

$$\leq (1-p)^{\frac{1}{2}\binom{x}{2}} \leq e^{-\frac{\epsilon}{\sqrt{n}}\frac{A^2 n(logn)^2}{5}}$$

$$\mathbb{P}(\exists X \text{ with no } e_X) \leq \binom{n}{x} e^{-\frac{\epsilon A^2\sqrt{n}(\log n)^2}{5}}$$

$$\leq e^{-\frac{\epsilon A^2\sqrt{n}(\log n)^2}{5} + A\sqrt{n}(\log n)^2 + o(1)A\sqrt{n}\log n},$$

which completes the proof. □

8 The Lovász Local Lemma and Applications

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.

8.1 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 $\{\{A_i, A_)\} \ : \ A_i$ and A_j 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:

The Lovasz Local Lemma:
Suppose we have events, $A_1, A_2, ..., A_n$, and real numbers $x_1, ..., x_n$ such that for each i satisfying $0 \le x_i < 1$, and

$$P(A_i) \le x_i \prod_{j \leftrightarrow i} (1 - x_j),$$

where the product is taken over all neighbors A_j of A_i in the dependency graph. Then,

$$P(\wedge_{i=1}^{n} \overline{A_i}) \ge \prod_{i=1}^{n} (1 - x_i) > 0,$$

where $\overline{A_i}$ denotes the complement of A_i—the event that A_i does not occur. So in particular, there is nonzero probability that none of the events A_i occur.

We will present the proof shortly. As an immediate corollary, we have:

Corollary:
Suppose there is some x, $0 \leq x < 1$, such that for each i,

$$P(A_i) \leq x(1-x)^{d(i)},$$

where $d(i)$ denotes the degree of A_i in the dependence graph, ie, the number of $A_j, j \neq i$ with which A_i is dependent. Then with nonzero probability, none of the events A_i occur.

Finally, we state a more useful symmetric version of the LLL, which we will most often apply in solving our problems:

Lovasz Local Lemma (Symmetric Version):
Suppose we have events $A_1, ..., A_n$, and that there exists some p such that $P(A_i) \leq p$ for each i, and $ep(\Delta+1) \leq 1$, where Δ is the maximum degree of the dependence graph. Then $P(\wedge_{i=1}^n \overline{A_i}) > 0$ (that is, with nonzero probability, none of the events occur).

Proof of Symmetric Version, using General Version of Local Lemma:

Take $x_i = 1/(\Delta+1)$, $\forall i$. 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. □

Proof of General Version of Local Lemma:
Let $S \subseteq \{1, ..., n\} \setminus \{i\}$. We will show by induction on $|S|$ that

$$P(A_i | \wedge_{j \in S} \overline{A_j}) \leq x_i.$$

If $|S| = 0$, then we are done, since by assumption $P(A_i) \leq x_i \prod_{j \leftrightarrow i}(1-x_j) \leq x_i$. Now take $|S| > 0$, and suppose we have proven the result for all smaller sizes of S. Let $N_S(i)$ be the set of neighbors of i in S (in the dependency graph), and let $NN_S(i)$ be the set of "non-neighbors" of i in S, $NN_S(i) = S \setminus N_S(i)$. Then

$$P(A_i | \wedge_{j \in S} \overline{A_j}) = P(A_i | \left(\wedge_{j \in N_S(i)} \overline{A_j}\right) \wedge \left(\wedge_{j \in NN_S(i)} \overline{A_j}\right))$$

$$= \frac{P(A_i \wedge \left(\wedge_{j\in N_S(i)}\overline{A_j}\right) \mid \wedge_{j\in NN_S(i)} \overline{A_j})}{P(\wedge_{j\in N_S(i)}\overline{A_j} | \wedge_{j\in NN_S(i)} \overline{A_j})}$$

$$\leq \frac{P(A_i | \wedge_{j\in NN_S(i)} \overline{A_j})}{P(\wedge_{j\in N_S(i)}\overline{A_j} | \wedge_{j\in NN_S(i)} \overline{A_j})}$$

$$= \frac{P(A_i)}{P(\wedge_{j\in N_S(i)}\overline{A_j} | \wedge_{j\in NN_S(i)} \overline{A_j})}.$$

Now, note that if $\alpha \in N_S(i)$, we can write

$$P(\wedge_{j\in N_S(i)}\overline{A_j} | \wedge_{j\in NN_S(i)} \overline{A_j}) = P(\overline{A_\alpha} \wedge \left(\wedge_{j\in N_S(i)\setminus\{\alpha\}}\overline{A_j}\right) \mid \wedge_{j\in NN_S(i)} \overline{A_j})$$

$$= P(\overline{A_\alpha} | \left(\wedge_{j\in N_S(i)\setminus\{\alpha\}}\overline{A_j}\right) \wedge \left(\wedge_{j\in NN_S(i)}\overline{A_j}\right)) \cdot P(\wedge_{j\in N_S(i)\setminus\{\alpha\}}\overline{A_j} | \wedge_{j\in NN_S(i)} \overline{A_j})$$

$$= P(\overline{A_\alpha} | \wedge_{j\in S\setminus\{\alpha\}} \overline{A_j}) \cdot P(\wedge_{j\in N_S(i)\setminus\{\alpha\}}\overline{A_j} | \wedge_{j\in NN_S(i)} \overline{A_j}).$$

Now, by our inductive hypothesis,

$$P(\overline{A_\alpha} | \wedge_{j\in S\setminus\{\alpha\}} \overline{A_j}) \geq 1 - x_\alpha,$$

and by another inductive argument, we have

$$P(\wedge_{j\in N_S(i)}\overline{A_j} | \wedge_{j\in NN_S(i)} \overline{A_j}) = P(\overline{A_\alpha} | \wedge_{j\in S\setminus\{\alpha\}} \overline{A_j}) \cdot P(\wedge_{j\in N_S(i)\setminus\{\alpha\}}\overline{A_j} | \wedge_{j\in NN_S(i)} \overline{A_j})$$

$$\geq (1 - x_\alpha) \cdot \prod_{\beta\in N_S(i)\setminus\{\alpha\}} (1 - x_\beta)$$

$$= \prod_{\alpha\in N_S(i)} (1 - x_\alpha).$$

Thus, from our work above,

$$P(A_i | \wedge_{j\in S} \overline{A_j}) \leq \frac{P(A_i)}{\prod_{\alpha\in N_S(i)}(1 - x_\alpha)} \leq x_i,$$

where the last inequality follows from the hypothesis of the theorem.

Finally, we have

$$P(\wedge_{i=1}^n\overline{A_i}) = P(\overline{A_n} | \wedge_{i=1}^{n-1} \overline{A_i}) \cdot P(\wedge_{i=1}^{n-1}\overline{A_i}) \geq (1 - x_n) \cdot P(\wedge_{i=1}^{n-1}\overline{A_i}),$$

where we have taken $S = \{1, ..., n-1\}$. Thus, we see by induction that

$$P(\wedge_{i=1}^n\overline{A_i}) \geq (1 - x_n) \cdot \prod_{i=1}^{n-1}(1 - x_i) = \prod_{i=1}^{n}(1 - x_i) > 0,$$

which completes the proof. □

8.2 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 71. *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 71 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 $v \in V$,

$$E_v := \text{There is no vertex } u \text{ in } N^+(v) \text{ such that color}(u) = \text{color}(v) + 1 \ (\mod k).$$

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 $0(\mod k)$. □

We now come to another appliction of the Lovász Local Lemma. Strauss proposed the following conjecture.

Conjecture 72 (Strauss)**.** *Fix $k \geq 2$. Is there an $m(k) > 0$ such that for any given fixed S of size m, every translate of S is multicolored? Here multicolored means all colors appear.*

Before returning to this conjecture, we introduce a theorem due to Van der Waerden.

Theorem 73 (Van der Waerden)**.** *Given $k, r \in \mathbb{N}$ there exists $W(k, r)$ such that coloring $\mathbb{Z}$ by k colors implies there exists a monocrhomatic arithmetic progression of length at least r.*

Erdős and Lovász showed that this is true when $|S| = (3 + o(1))k \log k$ is sufficient. They also proved this in the affirmative for another case. Fix S of size m and let $|X| = n$. If $ek\left(1 - \frac{1}{k}\right)^m \{m(m-1) + 1\} \leq 1$, then every translate $x + s$ is multicolored for $x \in X$.

Proof. Let $\mathcal{X} = \cup_{x \in X}(x + S)$ and notice that $|\mathcal{X}| < \infty$. Now, color every element of $\mathcal{X}$ from $\{1, 2, \ldots, k\}$ independently and uniformly. We want $\mathbb{P}(\wedge_{x \in X}((x + S)$ is not multicolored$)) > 0$. Fix x. Therefore,

$$\begin{aligned}
\mathbb{P}((x + S) \text{ is not multicolored}) &= \mathbb{P}(\exists\, i \in \{1, \ldots, k\} \text{ such that color } i \text{ is missing in } (x + S)) \\
&\leq k\left(1 - \frac{1}{k}\right)^m
\end{aligned}$$

Moreover, $(x+S)$ and $(y+S)$ are co-dependent if $(x+S) \cap (y+S) \neq \emptyset$, where $x + s = y + t$ and $y = x + (s - t)$ for some $y \in X$ and $t \in S$. This shows us that we are done if $ek\left(1 - \frac{1}{k}\right)^m m^2 \leq 1$. Therefore, we can loosen this to see that we are done if $eke^{-m/k}m^2 \leq 1$. We now use this expression to obtain a bound on m.

$$\begin{aligned}
eke^{-m/k}m^2 \leq 1 \quad &\Longrightarrow \\
e^{m/k} \geq ekm^2 \quad &\Longrightarrow \\
m \geq k\log k + 2k\log m + c \quad &\Longrightarrow \\
\geq k\log k + 2k\log(k\log k) \quad &\text{(substituing the above)} \Longrightarrow \\
\geq 3k\log k + 2k\log\log k + \ldots &
\end{aligned}$$

□

We now move on to Erdős and Lovász' resolution of Strauss' conjecture. Any coloring of $\mathbb{R}$ by k-colors is an element of $[k]^{\mathbb{R}}$. Where, $[k] = [1, 2, \ldots, k]$ with the discrete topology. Tychonoff's theorem implies $[k]^{\mathbb{R}}$ is compact. For each $x \in \mathbb{R}$, let $\mathcal{C}_x = \{c \in [k]^{\mathbb{R}} | x + S \text{ is multicolored}\}$. Observe that each $\mathcal{C}_x$ is closed in $[k]^{\mathbb{R}}$. Therefore $\mathcal{C} = \{\mathcal{C}_x | x \in \mathbb{R}\}$ is a family of closed sets in $[k]^{\mathbb{R}}$ with the finite intersection property. We have just shown that for any x finite, $\cap_{x \in X}\mathcal{C}_x \neq \emptyset$ by the compactness of $[k]^{\mathbb{R}}$. We remark that this is equivalent to the Rado selection principle. Thus we have proved

Theorem 74. *For any fixed subset $T \subseteq RED$ of size at least $m = (3 + o(1))k \log k$, there is a k coloring of RED such that every translate of T contains all k colors.*

Remark: it turns out that the $k \log k$ term in the above expression is not only sufficient, but also necessary.

Example: Independent Transversals in Graphs

We next give an example which demonstrates how the Local Lemma often works more effectively when *more* "bad" events are involved, because intuitively, this brings down the dependence degree of each event.

Definition: Suppose G is a graph, with vertex set $V = V_1 \cup V_2 \cup ... \cup V_r$ (a disjoint partition). We say a set $\mathcal{I}$ is a *transversal* for this partition if it contains exactly one element from each V_i. An *independent transversal* is a transversal which is also an independent set (containing no edges).

Theorem: *Suppose $V = V_1 \cup ... \cup V_r$, and for each i, $|V_i| \geq 2e\Delta$, where Δ is the maximum degree of G. Then $\{V_i\}$ admits an independent transversal.*

Proof: We may assume that $|V_i| = \lceil 2e\Delta \rceil$, for each i (by just ignoring any extra vertices in each set V_i). Let us pick $v_i \in V_i$ independently and uniformly for each i. Seeking to apply the Local Lemma, we must consider how to define our "bad" events. We could, for instance, define a bad event to indicate that v_i and v_j are adjacent. But this turns out to have a large dependence degree, and is difficult to work with.

Instead, we define our events as follows. For an edge E, let A_E denote the event that both vertices in E are selected, so that E is involved in the transversal. This will significantly increase the number of bad events, but this is of no matter, since the lemma depends only on the local properties of each event, namely that the dependence degree of each event is relatively low.

Now, the probability that both vertices of E are included in the transversal is at most $1/(2e\Delta)^2$. Furthermore, since each vertex in E has degree at most Δ, the dependence degree is at most 2Δ. Thus, by the Local Lemma, since

$$\frac{e(2\Delta + 1)}{(2e\Delta)^2} \leq 1, \forall \Delta$$

then there is positive probability that no edge is contained in our transversal. Thus, there exists an independent transversal for our graph G. $\square$

8.3 The Linear Arboricity Conjecture

Definition 75 (Forest). *A forest is a an undirected cycle-free graph.*

For any graph G, $E(G)$ can be partitioned into disjoint forests. If we insist that every connected component of each of these forests is a path, then the forest is called a linear forest.

Definition 76 (Linear Forest). *A graph is a linear forest if all of its components are paths.*

A natural question to ask is, how many linear forests are needed for this partition? Let X' be an edge coloring. Recall that the Vizing-Gupta theorem gives $X'(G) \leq \Delta(G)+1$. Therefore $\delta(G)+1$ linear forests suffice to cover G.

On the other hand, if G is a regular graph of degree d, and if $E(G)$ is partitioned into m linear forests with $|V(G)| = n$, then each linear forest has at most $n-1$ edges and all the linear forests toegether can cover at most $(n-1)d$ edges. Finally, $|E(G)| = \frac{dn}{2} \leq (n-1)m$ implies that $m \geq \frac{d}{2}\frac{n}{n-1} > \frac{d}{2}$.

Example 77. *Consider K_n, the complete graph on n vertices. If n is even, then $E(K_n)$ can be partitioned into Hamiltonian paths.*

Definition 78 (Linear Arboricity)**.** *Let G be a graph with maximum degree Δ. Define the linear arboricity, $la(G)$:= the minimum number of linear forests needed to partition the edge set, $E(G)$.*

Arboricity is closely related to the density of edges in a graph and linear arboricity to the maximum degree of a graph.

Conjecture 79 (Harary's Conjecture)**.** *The minimum number of linear forests needed to partition $E(G)$ is the linear arboricity of G, where $la(G) \leq \lceil \frac{d+1}{2} \rceil$, where d is the degree of G.*

This has been proven for $d = \{3, 4, 5, 6, 8, 9, 10\}$. Alon showed that for any $\epsilon > 0$, and d 'sufficiently large', that $la(G) \leq \left(\frac{1}{2}+\epsilon\right)(d+1)$. We now introduce equivalent arboricity definitions for directed graphs.

Definition 80 (Directed Linear Forest)**.** *A directed graph D is a directed linear forest if all components are directed paths.*

Definition 81 (Directed Linear Arboricity)**.** *The directed linear arboricity, $dla(\mathcal{D})$:= the minimum number of directed linear forests.*

Conjecture 82 (Directed Linear Arboricity)**.** *Let $\mathcal{D}$ be a regular directed graph of degree d, then $dla(\mathcal{D}) = d+1$.*

This issue is not yet settled, but the result can be shown under slightly weaker assumptions.

Theorem 83. *Let $\mathcal{D}_r$ be a regular directed graph of degree d, and assume that $\mathcal{D}_r$ has directed girth $\geq 8ed$. Then, $dla(\mathcal{D}_r) = d+1$.*

Alon showed in 1986 that the directed linear arboricity conjecture is asymptotically true.

Theorem 84. *If d is sufficiently large, then given any $\epsilon > 0$, we have $dla(\mathcal{D}) \leq d(1+\epsilon)+1$.*

We can prove Theorem 84 with the following idea. Break $\mathcal{D}$ into 'many' subdigraphs each having large girth. Think of coloring $V(\mathcal{D})$ using p colors. We construct these subdigraphs as follows:

$$\begin{aligned}
\mathcal{D}_1 \subset \mathcal{D}: &\text{ if } c(w) = c(v)+1, \text{ then edge } (v,w) \in \mathcal{D}_1 \\
\mathcal{D}_2 \subset \mathcal{D}: &\text{ if } c(w) = c(v)+2, \text{ then edge } (v,w) \in \mathcal{D}_2 \\
\vdots \qquad & \qquad \vdots \\
\mathcal{D}_i \subset \mathcal{D}: &\text{ if } c(w) = c(v)+i, \text{ then edge } (v,w) \in \mathcal{D}_i
\end{aligned}$$

Next, notice that picking p prime implies that all the $\mathcal{D}_i$'s have girth at least p. Thus, if p is a 'large' prime, then all the $\mathcal{D}_i$'s ($i = 1, 2, \ldots, p-1$) have girth at least p. If further, each $\mathcal{D}_i$ is 'almost regular,' then we can embed $\mathcal{D}_i$ into a regular directed graph, with the girth condition also intact. We do this by iteratively placing vertices in a direction that does not mess up the girth.

We now introduce the probabilistic paradigm we are using for this proof. Randomly color vertices using colors in $[1, 2, \ldots, p]$. If each resulting $\mathcal{D}_i$ is almost regular and has large girth, then we can apply Theorem 83 to each $\mathcal{D}_i$. We introduce further notation, this time regarding the number of in and out edges. Let

$$\begin{aligned}
d_i^+(v) &= \#\{w \in N^+(v) | c(w) = c(v) + i\} \\
d_i^-(v) &= \#\{w \in N - +(v) | c(v) = c(w) + i\}
\end{aligned}$$

Next, use Chernoff bounds to obtain

$$\mathbb{P}\left(\left|d_i^+(v) - \frac{d}{p}\right| > t\right) \leq 2\exp\left[-\frac{t^2}{3d/p}\right].$$

if $0 \leq t \leq d/p$. Next, choose $t = \sqrt{3\frac{d}{p}\log f(d)}$ for some $f(d)$. We see that

$$\mathbb{P}\left(\left|d_i^+(v) - \frac{d}{p}\right| > \sqrt{C\frac{d}{p}\log f(d)}\right) \leq \frac{1}{f(d)}$$

for some constant C. Similarly,

$$\mathbb{P}\left(\left|d_i^-(v) - \frac{d}{p}\right| > \sqrt{K\frac{d}{p}\log f(d)}\right) \leq \frac{1}{f(d)}$$

for some constant K. Thus, we need to show there is a coloring c such that

$$\left|d_i^+(v) - \frac{d}{p}\right| \leq \sqrt{C\frac{d}{p}\log f(d)} \quad \text{and} \quad \left|d_i^-(v) - \frac{d}{p}\right| \leq \sqrt{K\frac{d}{p}\log f(d)}$$

then we have a partition of $E(\mathcal{D})$ into $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_{p-1}$ such that for any $1 \leq i \leq p-1$ such that the directed girth of $\mathcal{D}_i \geq **** \, d_i = d_i^+ + d_i^- \, ****$. Furthermore, we now define events that will be useful in obtaining the dependence degree. Let

$$A_{i,v}^{+} := \left\{ \left| d_i^{+}(v) - \frac{d}{p} \right| > \sqrt{C\frac{d}{p}\log f(d)} \right\}$$

$$A_{i,v}^{-} := \left\{ \left| d_i^{-}(v) - \frac{d}{p} \right| > \sqrt{K\frac{d}{p}\log f(d)} \right\}$$

Furthermore, observe that

$$\mathbb{P}(A_{i,v}^{+}) \leq \frac{1}{f(d)} \quad \text{and} \quad \mathbb{P}(A_{i,v}^{-}) \leq \frac{1}{f(d)}.$$

We now want to show that

$$\mathbb{P}\left(\left(\wedge_{i,v} \overline{A_{i,v}^{+}} \right) \wedge \left(\wedge_{i,v} \overline{A_{i,v}^{-}} \right) \right) > 0. \tag{8.1}$$

Next, realize that the dependence degree is ***** ********. Moreover, the above definitions arguments show that Dependence degree $+ 1 \leq (2d)(2d)p = 4d^2p$. Therefore, if $e(4d^2p)\frac{1}{f(d)} \leq 1$ then 8.1 is satisfied. Now, suppose $p = \Theta(\sqrt{d})$. Hence, $f(d) = \Omega(d)$ satisfied 8.1. We need to guarantee that $p \geq 8ed\left(\frac{d}{p} + \sqrt{c\frac{d}{p}\log d}\right)$. Therefore, consider p greater than the constant κ. We want $\kappa\sqrt{d} \geq \frac{8e\sqrt{d}}{\kappa} + \sqrt{\frac{c}{\kappa}\sqrt{d}\log d}$. This is clearly satisfied if $\kappa^2 \geq 16e$, so take $\kappa \geq 4\sqrt{e}$. Now, pick p such that

$$4\sqrt{e}\sqrt{d} \leq p \leq 8\sqrt{e}\sqrt{d}.$$

Notice that we are guaranteed the existence of a prime in this interval by the Bertrand-Chebyshev Theorem.

Therefore, we have proved that there exists a coloring c of $V(\mathcal{D})$ such that for each color, and for each vertex v,

$$\left| d_i^{+}(v) - \frac{d}{p} \right| \leq \sqrt{C\frac{d}{p}\log f(d)} \quad \text{and} \quad \left| d_i^{-}(v) - \frac{d}{p} \right| \leq \sqrt{K\frac{d}{p}\log f(d)},$$

for $i = 1, \ldots, p-1$. Hence, for each $\mathcal{D}_i$, $dla(\mathcal{D}_i) \leq \left(\frac{d}{p} + \sqrt{c\frac{d}{p}\log d}\right)$. Hence,

$$\begin{aligned} dla(\mathcal{D}) &\leq dla(\mathcal{D}_0) + \sum_{i=1}^{n-1} dla(\mathcal{D}_i) \\ &\leq dla(\mathcal{D}_0) + \left(\frac{d}{p} + \sqrt{C\frac{d}{p}\log d}\right)(p-1) \end{aligned}$$

Now we consider the asymptotics. Observe that

$$\begin{aligned} dla(\mathcal{D}) &\leq O(\sqrt{d}) + d + \sqrt{C\frac{d}{p}\log d} \\ &= d + O\left(d^{3/4}\log^{1/2} d\right) \end{aligned}$$

Where this quantity can be made smaller than any $\epsilon > 0$ by taking d large. Therefore $dla(\mathcal{D}) \leq d + O\left(d^{3/4} \log^{1/2} d\right)$.

Alon's proof of the asymptotic directed linear arboricity conjecture proves 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 85. *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

$$\mathbb{P}(A_v) \leq \frac{1}{g(d)} \leq x(1-x)^d \prod_{k=3}^{h(d)} (1 - y_k) \tag{8.2}$$

where $1 - x$ corresponds to adjacent vertices and $1 - y_k$ corresponds to adjacent cycles. We also obtain

$$\mathbb{P}_{|c| \geq k}(B_C) \leq \left(\frac{f(d)}{d}\right)^k \leq y_k(1-x)^k \prod_{l \geq 3} (1 - y_l)^{kd^{l-2}}. \tag{8.3}$$

If there exists x and y_k such that (8.2) and (8.3) hold then the Lovász Local Lemma works. A nice start is to try $y_k = \frac{1}{d^{k-1}}$.

Alon actually proves the corresponding theorem with

$$h(d) = \frac{\log d}{2d \log \log d} \qquad f(d) = \log^{10} d \qquad g(d) = \log^6 d.$$

□

8.4 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 86 (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 $\mathbb{P}(A_i \text{ Var}ert \wedge_{j \in S,\, i \not\leftrightarrow j} \overline{A}_j) \leq p$. 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. □

8.5 The Moser-Tárdos algorithm

9 Martingales and Concentration Inequalities

The theory of Martingales and concentration inequalities were first used spectacularly by Janson, and then later by Bollobas 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.

9.1 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 Doobs' Martingale Process.

9.2 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:

  $$X_i = \mathbb{E}(f(G)|e_j \textit{ is revealed } \forall 1 \leq j \leq i)$$

  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.

9.3 Azuma's Inequality

Definition 87 (Lipshitz)**.** *A function f is $K-Lipschitz$ if $\forall x, y \ \ |f(x) - f(y)| \leq K|x-y|$. A martingale $X_0, X_1, ...$ is $K-Lipschitz$ if $\forall i \ \ |X_i - X_{i+1}| \leq K$*

Theorem 88 (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:

$$\begin{aligned}
\mathbb{P}(X_m > \lambda\sqrt{m}) &= \mathbb{P}(e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}}) \\
&\leq \mathbb{E}(e^{\alpha X_m})e^{-\alpha\lambda\sqrt{m}} \\
&\leq e^{\alpha^2 m/2 - \alpha\lambda\sqrt{m}} \\
&= e^{-\lambda^2/2} \quad (since\ \alpha = \lambda/\sqrt{m})
\end{aligned}$$

□

Corollary 89. *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*

$$\mathbb{P}(|X_m - c| > \lambda\sqrt{m}) < 2e^{-\lambda^2/2}$$

9.4 The Shamir-Spencer Theorem for Sparse Graphs

Theorem 90 (Theorem of Shamir-Spencer)**.** *If $G = G(n,p)$ with $p = n^{-\alpha}$ for some α then there exists an integer $\mu = \mu(n)$ such that*

$$\mathbb{P}(\mu \leq \chi(G) \leq \mu + 3) \to 1 \text{ as } n \to 1$$

(i.e., $\chi(G)$ get concentrated over only 4 values.)
(Almost every graph parameter has a behavior similar to chromatic number.)

Proof. Let $\epsilon > 0$ be arbitrarily small and let μ be defined as follows:

$$\mu = \inf\{v \mid \mathbb{P}(\chi(G) > v) < 1 - \epsilon\}$$

i.e. with probability $\geq \epsilon$, $\chi(G) \leq \mu$ however $\mathbb{P}(\chi(G) \leq \mu - 1) < \epsilon$.
Let Y be the vertex set of largest subgraph of G which is $\mu - colorable$. Let $R = V \setminus Y$ where $V = vertex(G)$, consider $|R|$. Consider a vertex exposure martingale i.e., we know if the vertex is in R *or* Y one at a time.

$$X_i = \mathbb{E}(|R| \mid exposed\ till\ i'th\ vertex);\ clearly\ |X_{i+1} - X_i| \leq 1$$

By Azuma's inequality we have:

$$\mathbb{P}(||R| - \mathbb{E}(|R|)| > \lambda\sqrt{n-1}) \leq 2e^{-\lambda^2/2} \quad \forall\lambda > 0$$

Pick λ s.t. $2e^{-\lambda^2/2} < \epsilon$. $R = 0 \implies G\ is\ \mu\ colorable$ and this happens with prob $\geq \epsilon$ i.e.

$$0 \in (\mathbb{E}(R) - \lambda\sqrt{n}, \mathbb{E}(R) + \lambda\sqrt{n}) \implies |R| \approx c\sqrt{n} \quad w.p. \geq 1 - \epsilon$$

But any induced subgraph on $c\sqrt{n}$ vertices can be 3-colored with high probability, i.e. $\mathbb{P}(\chi(G(R)) > 3) < \epsilon$ if n is large enough. Here $G(R)$ is the graph induced by R.

Claim: Let S be s.t. $|S| \leq c\sqrt{n}$, w.h.p. S is 3-colorable.

Proof. Suppose not. Then $\exists$ a smallest subgraph of size $\leq c\sqrt{n}$ that is not 3-colorable. Let T be smallest such set. Note that every vertex in T has degree $\geq 3 \implies e(T) \geq \frac{3}{2}|T|$.
But in a graph $G_{n,p}$ the probability that $\exists$ some set T of size $\leq c\sqrt{n}$ which has $\geq \frac{3t}{2}$ edges is $o(1)$. Because:

$$\mathbb{P}(\exists T\ of\ size\ t\ and\ with\ \frac{3t}{2}\ edges) \approx \binom{\binom{t}{2}}{\frac{3t}{2}} p^{3t/2} \quad where\ p \sim n^{-\alpha}$$

Because:

$$\mathbb{P}(\exists T\ of\ size \leq c\sqrt{n}\ and\ with\ \frac{3t}{2}\ edges) \leq \sum_{t=0}^{t=c\sqrt{n}} \binom{n}{t}\binom{\binom{t}{2}}{\frac{3t}{2}} p^{3t/2} \to o(1)$$

if $p \sim n^{-\alpha}$ and $\alpha > 5/6$. □

This concludes the proof of Shamir-Spencer as $\mu \leq \chi(G) \leq \mu + 3$ with high probability. □

9.5 The Pippenger-Spencer (PS) 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 91 (The Pippenger-Spencer Theorem)**.** *Given $\epsilon > 0$, $\exists$ a $\delta > 0$ and $D_0(\epsilon)$ s.t. the following holds if $n \geq D \geq D_0$ and:*

- $D > d(x) > (1 - \delta)D$
- $d(x, y) < \delta D \forall x, y \in V$ *where* $D = \Delta(\mathcal{H})$ $\star$

Then $\chi'(\mathcal{H}) < (1 + \epsilon)D$

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 $\frac{\epsilon}{D}$ independent of each other. Form the subcollection 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:

$$\mathcal{H} = \mathcal{H}_0 \xrightarrow{\mathcal{E}_1} \mathcal{H}_1 \xrightarrow{\mathcal{E}_2} \mathcal{H}_2 \dots \xrightarrow{\mathcal{E}_t} \mathcal{H}_t$$

$D_1 \approx De^{-\epsilon k}$ (where $\mathcal{H}$ is k-uniform) since

$$\mathbb{P}(edge\ surviving) \approx \left[(1 - \frac{\epsilon}{D})^D\right]^k = e^{-\epsilon k}$$

asymptotically. Now let:

$$\mathcal{M}^{(1)} = \bigcup_{i=1}^{t} \mathcal{M}_i \quad (M_i\ are\ disjoint\ by\ construction)$$

For an edge A:

$$\mathbb{P}(A \in \mathcal{M}^{(1)}) = \sum_{i=1}^{t} P(A \in M_i) \quad and$$

$$\mathbb{P}(A \in \mathcal{M}_1) \approx \frac{\epsilon}{D}, \quad \mathbb{P}(A \in \mathcal{M}_2) \approx \frac{\epsilon}{D_1}(1 - \frac{\epsilon}{D})^{k(D-1)} \approx \frac{\epsilon}{D_1} e^{-\epsilon k} \quad in\ general:$$

$$\mathbb{P}\big(A \in \mathcal{M}_i\big) \approx \frac{\epsilon}{D} e^{-\epsilon k + \epsilon(i-1)}$$

$$\implies \mathbb{P}(A \in \mathcal{M}^{(1)}) = e^{-\epsilon k} \left(\frac{\epsilon}{D}\right) \sum_{i=1}^{t} e^{\epsilon(i-1)} = e^{-\epsilon k} \left(\frac{\epsilon}{D}\right) \left(\frac{1 - e^{\epsilon t}}{1 - e^{\epsilon}}\right) \approx \frac{\alpha}{D}$$

where $\alpha = \alpha(\epsilon, t, k) = \epsilon e^{-\epsilon k} \frac{(1-e^{\epsilon t})}{1-e^{\epsilon}}$. 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 $\mathcal{M}^{(1)}, \mathcal{M}^{(2)}, \mathcal{M}^{(3)} \dots \mathcal{M}^{(\mu)}$ with each $\mathcal{M}^{(i)}$ having:

$$\mathbb{P}(A \in \mathcal{M}^{(i)}) \approx \frac{\alpha}{D}$$

Let $\mathcal{P}^{(1)} = \mathcal{M}^{(1)} \cup \mathcal{M}^{(2)} \cup \mathcal{M}^{(3)} \cup \dots \cup \mathcal{M}^{(\mu)}$.

$$\mathcal{H} = \mathcal{H}^{(0)} \xrightarrow{\mathcal{P}^{(1)}} \mathcal{H}^{(1)} \xrightarrow{\mathcal{P}^{(2)}} \mathcal{H}^{(2)} \dots \xrightarrow{\mathcal{P}^{(s)}} \mathcal{H}^{(s)}$$

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 $\theta D^{(1)} - colorable$ and so on.

Hence so far we need $\theta D + \theta D^{(1)} + \dots + \theta D^{(s-1)}$ colors. After removing colored edges (i.e. $edges \in some\ \mathcal{P}^{(i)}$), 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:

$\chi'(\mathcal{H}) \leq k(D-1)+1 \Longrightarrow \chi'(\mathcal{H}^{(s)}) \leq k(D^{(s)}-1)+1$

Hence:

$$total\ \#\ of\ colors\ we\ used = \theta\sum_{i=1}^{s-1} D^{(i)} + \theta D + k(D^{(s)}-1)+1 \approx D$$

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 $\mathbb{E}(d^{(1)}(x))$?

$$d^{(1)}(x) = \sum_{A: x\in A\in \mathcal{H}^{(0)}} \mathbb{1}_{A \not\subseteq \mathcal{P}^{(1)}}$$

$$\Longrightarrow \mathbb{E}(d^{(1)}(x)) = \sum_{A: x\in A\in \mathcal{H}^{(0)}} (1-\frac{\alpha}{D})^{\mu} \approx D(1-\frac{\alpha}{D})^{\mu} \approx D(1-\frac{\alpha}{D})^{\theta D} \approx De^{-\alpha\theta} = D^{(1)}$$

Hence $\quad \mathbb{E}(d^{(1)}(x)) \approx D^{(1)} = De^{-\alpha\theta}$

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 $\Longrightarrow$ 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.

$\Longrightarrow$ *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 $t \to \infty$, $s \to \infty$, $\epsilon \to \infty$, etc. Thus we'll have the desired result.

When we do the calculations, everything works out nicely.

9.6 A Conjecture of Erdős-Faber-Lovász (EFL) and a theorem of Kahn

Definition 92. *A hypergraph $\mathcal{H}$ is nearly-disjoint (linear) if*

$$\forall A \neq B \in \mathcal{E}(\mathcal{H}), \quad |A \cap B| \leq 1$$

.

Conjecture 93. *If $\mathcal{H}$ is nearly-disjoint on n vertices, then $\chi'(H) \leq n$*

Theorem 94 (Erdos-de Bruijn Theorem)**.** *If $\mathcal{H}$ is a hypergraph on n vertices with*

$$|A \cap B| = 1 \ \ \forall A \neq B$$

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 95 (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 hav $|E| \le k$, 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 $3 \le b \le |E| \le a$ 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 $y \ne x$, 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 $(b-1)d(x) \le n-1 \Rightarrow d(x) \le \frac{n-1}{b-1} < \frac{n}{2}$.

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 $|E_i| \ge |E_{i+1}|$ 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 $E_i \cap E_j \ne \emptyset$. To estimate $|\{1 \le j < i | E_j \cap E_i \ne \emptyset\}|$, 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 $1 \le j < i, E_j \cap E_i \ne \emptyset$, $E_i \cap E_j$ is uniquely determined, so the number of such triples is $|E_j| - 1$. Hence denoting $\mathcal{I} = \{1 \le j < i | E_j \cap E_i \ne \emptyset\}$ and noting that for each $j \in \mathcal{I}$ $|E_j| \ge k$, we get

$$(k-1)|\mathcal{I}| \le \sum_{j\in\mathcal{I}}(|E_j| - 1) \le k(n-k) \Rightarrow |\mathcal{I}| \le \frac{k(n-k)}{k-1}.$$

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 $\chi'(\mathcal{H}) \le 2n - 3$.

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 $\epsilon > 0$ if we $a > 1/\epsilon$, say, then for $C = (1 + 2\epsilon)n$, following the same greedy algorithm will properly color all edges of size greater than a. This motivates us to consider

- $\mathcal{E}_s := \{E \in \mathcal{E} | |E| \le b\}$.
- $\mathcal{E}_m := \{E \in \mathcal{E} | b < |E| \le a\}$.
- $\mathcal{E}_l := \{E \in \mathcal{E} | |E| > a\}$

for some absolute constants a, b which we shall define later. We have seen that $\chi'(\mathcal{H}_l) \le (1+2\epsilon)n$; also by a preceding remark, if we pick $b > O(1)/\epsilon$ we have $\chi'(\mathcal{H}_m) \le \epsilon n$. Thus, let us do the following.

Let $C = \lfloor(1+4\epsilon)n\rfloor$; we shall color the edges of $\mathcal{H}$ using the colors $\{1, 2 \ldots, C\}$. Let $C_1 = \{1, 2 \ldots, \lfloor(1+3\epsilon)m\rfloor\}$; $C_2 := C \setminus C_1$. 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 96. *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 $\alpha = \alpha(\epsilon, t, k) = \epsilon e^{-\epsilon k}\frac{(1-e^{\epsilon t})}{1-e^{\epsilon}}$, 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.
- $B_v^{(i)} < \mathbb{E}(B_v^{(i)}) + o(D)$

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 $\mathbb{E}(B_v^{(i)})$. For each $A \in \mathcal{H}^{(i)}$, A is **not** in $B^{(i)}$ if and only if for each $c_{ij} \in \text{Forb}(A)$ we have $A \notin \mathcal{M}_{i=1}^{(j)}$. Denoting $\text{Forb}^{(i)}(A) := \{j | c_{ij} \in \text{Forb}(A)\}$ we have

$$\mathbb{P}(A \in B^{(i)}) = 1 - \left(1 - \frac{\alpha}{D^{(i)}}\right)^{|\text{Forb}^{(i)}(A)|} < \frac{\alpha |\text{Forb}^{(i)}(A)|}{D^{(i)}}.$$

Hence,

$$\begin{aligned} \mathbb{E}(B_v^{(i)}) &= \sum_{v \in A \in \mathcal{H}^{(i)}} \mathbb{P}(A \in B^{(i)}) \\ &\lesssim \frac{\alpha}{D^{(i)}} \sum_{v \in A \in \mathcal{H}^{(i)}} |\text{Forb}^{(i)}(A)| \end{aligned}$$

Let $i(A) := \max\{0 \le i \le s | A \in \mathcal{H}^{(i)}\}$. Note that for any fixed i,

$$|\{A \in \mathcal{H} | v \in A, i(A) = i\}| \le \theta e^{-\alpha\theta i} D.$$

Hence we have

$$\begin{aligned}
\sum_{i=0}^{s} \mathbb{E}(B_v^{(i)}) &\lesssim \alpha \sum_{i=0}^{s} \frac{1}{D^{(i)}} \sum_{v \in A \in \mathcal{H}^{(i)}} |\mathrm{Forb}^{(i)}(A)| \\
&= \alpha \sum_{v \in A} \sum_{i} \frac{|\mathrm{Forb}^{(i)}(A)|}{D^{(i)}} \left(\mathbf{1}_{A \in \mathcal{H}^{(i)}}\right) \\
&\le \alpha \sum_{v \in A} \frac{1}{D^{(i(A))}} \left(\sum_{i} |\mathrm{Forb}^{(i)}(A)| \right) \\
&\le \alpha \sum_{v \in A} \frac{|\mathrm{Forb}(A)|}{D} e^{\alpha\theta i(A)} \\
&< \alpha o(1) \sum_{i=0}^{s} e^{\alpha\theta i} |\{A | v \in A, i(A) = i\}|
\end{aligned}$$

The last term in the above expression can be made 'small'. This completes the proof of Kahn's theorem.

10 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 that one can prove the concentration of measure phenomenon holds on the product space as well. One of the main reasons this inequality is so powerful is the its relatively wide applicability. In this chapter, we briefly study the inequality, and a couple of simple applications.

10.1 Talagrand's Inequality

Let (Ω, P, ρ) be a metric probability space, and let $A \subseteq \Omega$ with $\mathbb{P}(A) \geq 1/2$. For fixed t, let $A_t = \{\omega \in \Omega \mid \rho(\omega, A) \leq t\}$. What is $\mathbb{P}[A_t]$? That is, what can we say about the probability of an outcome close to one in A?

Definition 97. *Suppose $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$, the product of n (not necessarily metric) probability spaces (Ω_i, P_i). Then we can define a measure ρ on Ω by*

$$\rho(x, A) := \sup_{\|\alpha\|=1} \inf_{y \in A} \sum_{x_i \neq y_i} \alpha_i$$

Here α can be thought of as a cost (set by an adversary) for changing each coordinate of x to get to some event $y \in A$. Then we can intuitively think of ρ as the worst-case cost necessary to get from x to some element in A by changing coordinates.

Now for any probability space we can define $A_t = \{x \in \Omega \mid \rho(x, A) \leq t\}$, as above.

Theorem 98. *(Talagrand's Inequality)*

$$\mathbb{P}[A](1 - \mathbb{P}[A_t]) \leq e^{-t^2/4}$$

For the proof see p. 55 of Talagrand's paper "Concentration of Measure and Isoperimetric Inequalities in Product Spaces."

We can also define the measure ρ in another, perhaps more intuitive way. For a given $x \in \Omega$ and $A \subseteq \Omega$ let

$$\text{Path}(x, A) = \{s \in \{0, 1\}^n \mid \exists y \in A \text{ with } x_i \neq y_i \Leftrightarrow s_i = 1\}$$

and let $V(x, A)$ be the convex hull of Path(x, A) (in $[0,1]^n$). We can think of Path(x, A) as the set of all possible paths from x to some element $y \in A$, that is, the set of choices given some cost vector.

Theorem 99. $\rho(x, A) = \min_{v \in V(x,A)} \|v\|$.

Note that it is now clear that we can use min instead of sup and inf, since the convex hull is closed. It is also clear now that $\rho(x, A) = 0$ iff $(0, 0) \in V(x, A)$ iff $x \in A$.

Concentration of Measure about the Mean

Recall the definition of a Lipschitz function from earlier:

Definition 100. *A random variable $X : \Omega \to RED$ is c-Lipschitz if for any $\omega_1, \omega_2 \in \Omega$ differing in one coordinate $|X(\omega_1) - X(\omega_2)| \leq c$.*

We will also need to define another similar notion.

Definition 101. *A random variable $X : \Omega \to RED$ is f-certifiable for $f : RED \to RED$ if the following holds: If some $\omega \in \Omega$ satisfies $X(\omega) \geq s$, then there is a set $I \in [n]$ of size $\leq f(s)$ such that $X(\omega') \geq s$ for any $\omega' \in \Omega$ with $\omega'_i = \omega_i$ for all $i \in I$.*

We can now state a useful consequence of Talagrand's Inequality:

Corollary 102. *If X is 1-Lipschitz and f-certifiable then*

$$\mathbb{P}[X \geq b]\mathbb{P}[X \leq b - t\sqrt{f(b)}] \leq e^{-t^2/4}.$$

In particular, if b is the median of X, i.e. $b = \inf\{t \in RED \,|\, \mathbb{P}[X \geq t] \leq 1/2\}$, we have

$$\mathbb{P}[X \leq b - t\sqrt{f(b)}] \leq 2e^{-t^2/4}.$$

Proof. Let $A = \left\{\omega \,|\, X(\omega) < b - \sqrt{f(b)}\right\}$. We want to show that $\{\omega \,|\, X(\omega) < b\} \supseteq A_t$ so that $\mathbb{P}[X \geq b] \leq 1 - \mathbb{P}[A_t]$. That is, we want to show that for any ω' with $X(\omega) \geq b$, $\omega' \notin A_t$. Suppose otherwise. Since X is f-certifiable, there is a set $I \subseteq [n]$ of size no more than $f(b)$ such that if x agrees with ω' on I then $X(x) \geq b$. Now consider the penalty function $\alpha_i = 1_{\{i \in I\}}(|I|)^{-1/2}$. By our assumption that $\omega' \in A_t$, there exists $y \in A$ such that $\sum_{y_i \neq \omega'_i} \alpha_i \leq t$. Then the number of coordinates in which y and ω' 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 $z_i = \omega'_i$ 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 ω' on the coordinates of I, f-certifiability guarantees that $X(z) \geq b$, and we have a contradiction.

This phenomenon is known as concentration of measure about the median. The median tends to be difficult to compute, but fortunately it is often close to the mean. The conversion from median to mean is responsible for the constant factors in the following corollary.

Corollary 103. *(Talagrand's Inequality About the Mean) Suppose X is c-Lipschitz and r-certifiable (i.e. f-certifiable with $f(s) = rs$). Then*

$$\mathbb{P}[|X - \mathbb{E}[X]| > t + 60c\sqrt{r\,\mathbb{E}[X]} \leq e^{-t^2/8c^2 r\,\mathbb{E}[X]}.$$

Here we tend to think of t as some large multiple of $\sqrt{\mathbb{E}[X]}$, so that we can rewrite this as

$$\mathbb{P}[|X - \mathbb{E}[X]| > k\sqrt{\mathbb{E}[X]}] \leq e^{-\Omega(1)}$$

or

$$\mathbb{P}[|X - \mathbb{E}[X]| > \mathbb{E}[X]^{\frac{1}{2}+\epsilon}] \leq e^{-\mathbb{E}[X]^{\epsilon}}.$$

10.2 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

$$X = |\{v \mid d_H(v) > 0\}| = \sum_{v \in V} 1_{d_H(v)>0}$$

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.

2. Longest increasing subsequence

Suppose $x_1, \ldots, x_n \in [0,1]$ are picked uniformly and independently at random, and put them in increasing order to generate a permutation of $[n]$. Let X be the length of the longest increasing subsequence, and note that X is 1-Lipschitz (as changing a certain value could only either add it to a long increasing subsequence or remove it from one) and 1-certifiable (as any choice of the x_i with a particular increasing subsequence of length s always has $X \geq s$).

It is also easy to show that $X \leq 3\sqrt{n}$ with high probability. For any $i_1 < \cdots < i_k$, $\mathbb{P}[x_{i_1} \leq \cdots \leq x_{i_k}] = \frac{1}{k!}$ so

$$\mathbb{P}[X \geq k] \leq \binom{n}{k}\frac{1}{k!} \leq \left(\frac{en}{k}^k\right)\frac{e^k}{k^k}$$

and thus $\mathbb{P}[X \geq 3n] \leq \left(\frac{e}{3}\right)^{6\sqrt{n}} \to 0$. On the other hand, there is always an increasing or decreasing subsequence of length $\sqrt{n-1}$, so we actually find that with high probability

$$\frac{1}{3}\sqrt{n} \leq X \leq 3\sqrt{n}$$

so $\mathbb{E}[X] = O(\sqrt{n})$.

Talagrand's inequality now tells us that X is with high probability in an interval of length $O(\sqrt{\mathbb{E}[X]}) = O(n^{1/4})$. Note that Azuma would only give an interval of length $O(\sqrt{n})$, since the corresponding martingale would be of length n. The strength of Talagrand is that unlike Azuma it does not depend on the dimension of the product space.

10.3 An Improvement of Brook's Theorem

Let us recall Brook's Theorem: If a graph G is not K_n or C_{2k+1} then $\chi(G) \leq \triangle(G)$.

Here are two improvements:

Kim (2002): For G with girth ≥ 5, $\chi(G) \leq (1+o(1))\frac{D}{\log D}$.

Johansson (2004): For $\triangle$-free G, $\chi(G) \leq O(\frac{D}{\log D})$.

Theorem: If G is $\triangle$-free with max. 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 $C = \beta D \Longrightarrow \mathbb{E}(X''_v) \geq \frac{D(D-1)}{2}\frac{1}{\beta D}[(1-\frac{1}{\beta D})^{D-1}]^3 \geq \frac{D-1}{2}e^{-\frac{3}{\beta}}$, $D \gg 0$.

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 $t = C\sqrt{D \log D}$ then $\mathbb{P}(|Ass_v - \mathbb{E}(Ass_v)| > t) \leq 2e^{-\frac{t^2}{D}} = 2e^{-\frac{C^2 \log D}{2}} = \frac{2}{D^{C^2/2}}$.
Also, $\mathbb{P}(|Del_v - \mathbb{E}(Del_v)| > t + 60\sqrt{3\mathbb{E}(Del_v)}) \leq 2e^{-\frac{t^2}{24\mathbb{E}(Del_v)}} = \frac{2}{D^{\frac{C^2 D^2}{24\mathbb{E}(Del_v)}}}$.
We may now take $\beta = \frac{1}{2}$ so that $\alpha = 2e^{-6}$.

10.4 Almost Steiner Designs

We shall look now at a recent (2013) result due to Hod, Ferber, Krivelevich, and Sudakov, which achieves something very close to a Steiner design; they produce a uniform k-hypergraph in which every t-subset of the underlying set is contained in at least one, but at most two edges of the hypergraph, and one that has asymptotically the 'right' number of edges as well. More precisely,

Theorem 104. *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 $E \in \mathbb{E}(\mathcal{H})$ and at most two.*

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 105. *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 106. *(Grable, 1999) Let $k > t \geq 2$ be integers. There exists a constant $\epsilon = \epsilon(k, t) > 0$ 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 $O(n^{t-l-\epsilon})$ edges of the leave hypergraph $\mathcal{L}_{\mathcal{H}}$.

In particular, the size of $\mathcal{L}_{\mathcal{H}}$ is at most $O(n^{t-\epsilon})$. 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 $n_i(A); = \big|\{B \in \mathcal{L}_{\mathcal{H}} : B \neq A, |B \cap A| = i\}\big|$. 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 $O(n^{t-i-\epsilon})$ 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 $n_i(A) \leq \binom{t}{i} n^{t-i-\epsilon}$. Thus,

$$|\mathcal{S}_A| \geq \binom{n-t}{k-t} - \sum_{i=0}^{t-1} n_i(A) \binom{n-2t+i}{k-2t+i} = \Theta(n^{k-t}) - O(n^{k-t-\epsilon}) = \Theta(n^{k-t}).$$

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 $A \in \mathcal{L}_{\mathcal{H}}$. This however, will not be good enough. This distillation does ensure that different t-subsets of the leave hypergraph are covered exactly once, but it may happen that t-subsets that were initially covered by $\mathcal{H}$ may now be covered more than twice. Thus we need to be choosier.

One idea to deal with this is the following. If we can choose the edges E covering the t-subsets of the leave hypergraph in such a way that for distinct $A, B \in \mathcal{L}_{\mathcal{H}}$ if we have picked $E \in \mathcal{S}_A, F \in \mathcal{S}_B$ and we also have $|E \cap F| < t$ then the issue addressed above will not happen and then we can be sure that our second sub collection along with $\mathcal{H}$ will satisfy the conditions of the theorem. Thus the sense of choosiness that we want may be stated exactly as this:

For each t-subset A of the leave hypergraph we need to pick $E_A \in \mathcal{S}_A$ such that for $A \neq B$ we have $|E_A \cap E_B| < t$.

One of the interesting new perspectives of the probabilistic method that this proof suggests is the following principle:

The notion of being choosy can be interpreted probabilistically.

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 $|\mathcal{B}_l| \leq \binom{k}{l} O(n^{t-l-\epsilon}) = O(n^{t-l-\epsilon})$. 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 $O(n^{k-l-\epsilon})$ elements $B \in \mathcal{L}_\mathcal{H}$ that contains a set of size l. Hence, by a very generous amount, we have

$$\mathbb{E}(\mathbb{1}_E) = \mathbb{P}(E \text{ leads to increment of } \mathfrak{N}_\mathcal{Q}(A)) \leq pO(n^{k-2t+l})O(n^{t-l-\epsilon}) = pO(n^{k-t-\epsilon})$$

so

$$\mathbb{E}(\mathfrak{N}_\mathcal{Q}(A)) \leq |\mathcal{Q}_A| pO(n^{k-t-\epsilon}).$$

Now suppose we had $p = n^{t-k+\epsilon/2}$; then the estimate above gives us that

$$\mathbb{E}(\mathfrak{N}_\mathcal{Q}(A)) \leq |\mathcal{Q}_A| O(n^{-\epsilon/2}).$$

Note that for this value of p we have *whp* $|\mathcal{R}_A| \approx \Theta(n^{\epsilon/2})$ **for all** A (standard Chernoff bounds). We shall now argue that the greedy process produces $|\mathcal{Q}_A| \geq (n^{\epsilon/3})$ for all A *whp*. We can then choose to stop at around this stage while constructing $\mathcal{Q}_A$, so that we indeed do have $|\mathcal{Q}_A| = \Theta(n^{\epsilon/3})$ This completes the proof.

Suppose that the greedy process stops after m steps, with $m < n^{\epsilon/3}$. 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 $|X| < kn^{\epsilon/3}$. 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

$$\sum_{i=1}^{k-t} \binom{|X|-t}{i} \binom{n-|X|}{k-t-i} = \sum_{i=1}^{k-t} \Theta(n^{i\epsilon/3}) n^{k-t-i} = o(n^{k-t})$$

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 $kn^{\epsilon/3}$ that satisfies this condition above is at most

$$\binom{n}{kn^{\epsilon/3}}(1-p)^M < O(n^{n^{\epsilon/3}})\exp(-n^{t-k+\epsilon/2}\Theta(n^{k-t})) = \exp(n^{\epsilon/3}\log n - \Theta(n^{\epsilon/2})) < \exp(-n^{\epsilon/7})$$

for n sufficiently large, so the result follows. □

10.5 Chromatic number of graph powers: A result of Alon and Mohar

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 107. *(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 108. *(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 $N_{bad} = |\{v \in V | deg(v) > d\}| \implies$

$$\mathbb{E}[N_{bad}] < ne^{-d/6}$$

By the Markov inequality

$$\mathbb{P}[N_{bad} > 10ne^{-d/6}] < .1$$

Similarly, let $N_{<g} = |\{C_k \subseteq G | k < g\}| \implies$

$$\mathbb{E}[N_{<g}] = \sum_{i=3}^{g-1} \frac{(n)_i}{2i} \left(\frac{d}{2n}\right)^i < d^g$$

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 $t = \epsilon \mathbb{E}[\Pi]$, $\epsilon > 0$,

$$\mathbb{P}[\Pi < \frac{(1-\epsilon) c_k^2 n \log^2(d)}{2^{k+2} d^k}] \leq e^{-\beta \epsilon^2 \frac{c_k^2 n \log^2(d)}{2^{k+2} d^k}} = o(1)$$

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

$$\frac{\beta \epsilon^2 c_k^2}{2^{k+2}} > 2kc_k$$

then, with probability $1 - o(1)$, for every set U, there are at least $\frac{\epsilon n \log^2 d}{2^{k+2} d^k}$ pairwise internally disjoint U-paths.
Now, for $n \gg d \gg k$

$$10n2^{-d/10} + 10d^g < \frac{\epsilon n \log^2 d}{2^{k+2} d^k}$$

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.

11 The Janson Inequalities

The perspective on the local lemma is based on the following prototype: Suppose $\{A_i\}$ are 'bad' events, then under certain suitable conditions (each bad event has a relatively low probability of occurrence, and has a commonality with only a small number of other bad events, i.e., each bad event is independent of a large set of other bad events) then with positive probability, no bad event occurs. The counterpart of this prototype is to consider $\{A_i\}$ as 'good' events; here the good events are by themselves each low probability events, but we are interested in the situation where at least one of these good events occur. The Janson inequalities bound the probability that no good event occurs. More precisely, under certain conditions, the number of good events that can occur behaves almost like a Poisson random variable with the appropriate mean. Here, we shall explore this paradigm with a few instances.

11.1 Restricted Integer Partition functions

Suppose $A, M \subset \mathbb{N}$ are infinite sets, and suppose that for $n \in \mathbb{N}$, nn sufficiently large, we are interested in expressing

$$n = \sum_{a\in A} m_a a$$

where $m_a \in M \cup \{0\}$. We may think of the set M as providing multiplicities for expressing n as a sum of members of A. Let $p(n, A, M)$ denote the number of such representations for n. Canfield and Wilf proposed the following question:

Question 109. *Do there exist A, M such that for all n sufficiently large, $p(n, A, M) > 0$ and yet $p(n, A, M)$ has polynomial growth (in n)?*

Firstly, note that it is important that A is infinite. Otherwise for any n the number of expressions of the type $\sum_{a\in A, a\le n} m_a a$ is at most $n^{|A|}$.

Suppose we denote $A_n := A \cap [n]$. Ljujić and Nathanson observed that if $|A_n| > \delta \log n$ for any $\delta > 0$ then the above question has a negative answer for any such M. However, if $|A_n|$ were of lesser order, they did not quite settle the question. In particular, they posed the following:

Question 110. *Let $A_1 := \{k^k | k \in \mathbb{N}\}, A_2 := \{k! | k \in \mathbb{N}\}$. Does there exist M such that $p(n, A_i, M) > 0$ for all large n and yet has polynomial growth, for either $i = 1, 2$?*

Note that for these sets A, $|A_n| = (1 + o(1))\frac{\log n}{\log\log n}$. Alon settled this question in the affirmative as we shall see now.

Theorem 111. *(Alon) Suppose $1 \in A$ and $|A_n| = (1+o(1))\frac{\log n}{\log\log n}$ for all sufficiently large n. Then there exists M such that the answer to the question above is affirmative.*

Proof. Suppose we decide to pick M at random, i.e., for each $x \in \mathbb{N}$ we let $\mathbb{P}(x \in M) = p(x)$ and independently for distinct $x \in \mathbb{N}$. Now taking a cue from the earlier observation, the number of expressions of the form

$$\sum_{\substack{a\in A, m_a \in M\\ a\le n\\ m_a \le n}} m_a a$$

is at most $(|M_n|+1)^{|A_n|}$. Hence we need

$$\mathbb{E}(|M_n|)^{(1+o(1))\frac{\log n}{\log\log n}} = n^k$$

for some integer k, which suggests $\mathbb{E}(|M_n|) = O(\log^k n)$.

To ensure that there are expressions of the type $n = \sum_{a\in A} m_a a$, we shall need to restrict ourselves to certain specific types of such sums. In order to make our probability calculations amenable, let us structure the set $M = \bigcup_{a\in A} M_a$. We can then look for sums of the aforementioned type with the restriction that $m_a \in M_a$.

We may now posit that each $x \in M_a$ with probability $p_a(x)$. Then in order that $\mathbb{E}(|M_n|) = O(\log^k n)$ we need $\sum_{x\le n}\sum_{a\in A} p_a(x) = O(\log^k n)$. Observe that if we can show

$$\sum_{a\in A} p_a(x) \le \frac{(\log x)^{k-1}}{x}$$

then

$$\sum_{x\le n}\sum_{a\in A} p_a(x) \le \sum_{x\le n} \frac{\log^{k-1} n}{x} \le \log^k n.$$

Now, let us revert to examining what it entails to show that $p(n, A, M) > 0$. To make matters a little easier to analyze, let us only focus on expressions

$$\sum_{\substack{a\in A\\ a\le N}} m_a a$$

where N is some number to be appropriately chosen later. Suppose $1 = a_1 < a_2 < \cdots < a_q \le N$; we wish to show that whp there is at least one good q tuple $(m_1, m_2, \ldots, m_q)$ satisfying $\sum_{i=1}^q m_i a_i = n$ that is picked into the random set M. This fits the template for Janson's inequality, so let us set it up appropriately. For a good q tuple, the probability that this q tuple has been picked is then $\prod_{i=1}^q p_{a_i}(m_i)$.

This expression is obviously cumbersome; moreover, as we are in quest of a suitable function $p_a(x)$ that makes this calculation accessible, the form above is not very conducive to making a good guess/choice for $p_a(x)$.

To deal with it, here is a 'structural' idea: Suppose we instead decide to add further structure into each M_a as follows. Write $M_a = \bigcup_{i\in I_a} M_{a,i}$, where $M_{a,i} \subset [2^i]$ with each element of $[2^i]$ in $M_{a,i}$ with probability p_i for an appropriate p_i. For simplicity we may set $I_a = [f(a), \infty)$ for some function $f(\cdot)$.

At first glance, this appears like we are merely shifting our problem elsewhere. However, this idea has a couple of merits. For instance, we can further only consider expressions $\sum_{j=1}^{q} m_j a_j$ with $m_j \in M_{a_j,i_j}$ for some appropriate i_j. Under this restriction, the probability that $(m_1, m_2, \ldots, m_q)$ is picked is $\prod_{j=1}^{q} p_j(i_j)$.

Now reverting to the earlier expression, we need $\sum_{a\in A} p_a(x) \le \frac{(\log x)^{k-1}}{x}$ for each x. Fix an x and suppose it is sufficiently large. Note that

$$p_a(x) \le \sum_{\substack{i\ge f(a)\\ 2^i\ge x}} p(i),$$

so in particular, we wish to obtain p such that

$$\sum_{a\in A}\sum_{\substack{i\ge f(a)\\ 2^i\ge x}} p(i) = \sum_{a\le x}\sum_{\substack{i\ge f(a)\\ 2^i\ge x}} p(i) + \sum_{a>x}\sum_{\substack{i\ge f(a)\\ 2^i\ge x}} p(i) \le \frac{(\log x)^{k-1}}{x}. \qquad (\dagger)$$

Let us now make a simplifying assumption. Suppose $p(i+1) \le \frac{2}{3}p(i)$, say, for each i, a. This is easily achieved if say, p decays exponentially. Then

$$\sum_{i\ge f(a)\vee \log_2 x} p(i) \le 3p(f(a)\vee \log_2 x).$$

Now, taking a cue from this expression, suppose we set $f(a) = \log_2 a$; then the first sum in $(\dagger)$ is at most $O(\log x)p_a(\log_2 x)$ (the $O(\log x)$ is in fact $o(\log x)$ and this comes from the assumption about $|A_x|$ for each x); we need this to be at most $\frac{(\log x)^{k-1}}{x}$, so this suggests

$$p(i) = \frac{i^{k-2}}{2^i}.$$

Note that this also satisfies the assumption made earlier, namely, $p(i) < p(i+1)$ for i sufficiently large. If this ratio is greater than 1, then we set $p(i) = 1$.

Let consider the second sum in $(\dagger)$. This simplifies to

$$\sum_{a>x}\sum_{i\ge \log_2 a} \frac{i^{k-2}}{2^i} < 3\sum_{a>x} \frac{(\log a)^{k-2}}{a} = 3\sum_{j\ge 0}\sum_{2^{j-1}x<a\le 2^j x} \frac{(\log a)^{k-2}}{a}. \qquad (\dagger\dagger)$$

For each j, the sum $\sum_{2^{j-1}x<a\le 2^j x} \frac{(\log a)^{k-2}}{a}$ is at most $O(\log(2^j x)\frac{\log^{k-2} 2^j x}{2^j x}$, so when this is summed over j we see that the last term in $(\dagger\dagger)$ is at most

$$\sum_{j\ge 0} \frac{\log^{k-1} 2^j x}{2^j x} = \sum_{j\ge 0} \frac{(\log x + j\log 2)^{k-2}}{2^j x} \le \frac{\log^{k-1} x}{x}\sum_{j\ge 0} \frac{\left(1 + j\frac{\log 2}{\log x}\right)^k}{2^j} = O\left(\frac{\log^{k-1} x}{x}\right).$$

Now, with the choice for p all ready, let us tackle the requirement that $p(n, A, M) > 0$. As indicated earlier, we shall now restrict $m_j \in M_{a_j, i_j}$ for some suitable i_j.

To determine what should constitute the right i_j's the idea is the following. Given $m_2, m_3, \dots, m_q$ such that the sum $m_2 a_2 + \cdots + m_q a_q < n$ we can then pick m_1 in order to get the sum to equal n, so m_1 is then uniquely determined. In order to ensure this always, and keeping in mind that $M_{a,i}$ is a random subset of $[2^i]$ we shall impose that $n \leq 2^{i_1} < 2n$ (so that i_1 is minimally chosen). We shall seek similar bounds for i_j for each $2 \leq j \leq q$.

Now since $q = (1/3 + o(1)) \frac{\log n}{\log \log n}$, and we need $m_2 a_2 + \cdots + m_q a_q < n$ let us impose that

$$2^{i_j} < \frac{2n}{a_j \log n}.$$

With all this at hand we now make a formal definition:

We say that $(m_1, m_2, \dots, m_q)$ is a good q-tuple if

- $n = \sum_{i=1}^{q} m_i a_i$ where $1 = a_1 < a_2 < \cdots < a_q \leq n^{1/3}$.
- $m_j \in [2^{i_j}]$.

The ii_j's in the previous clause shall be determined exactly in a little while. For the moment we note that $2^{i_j} < \frac{2n}{a_j \log n}$.

We say that a given good q-tuple is *special* if $m_j \in M_{a_j, i_j}$. Our focus now shall be to show that with high probability there is a special q-tuple for each n sufficiently large.

For a given good q-tuple, the probability that it is a special tuple is precisely $\prod_{j=1}^{q} \frac{(i_j)^{k-2}}{2^{i_j}}$. Since we wish to use Janson's inequality, we first need to calculate

$$\mu := \sum_{(m_1, m_2 \dots m_q)} \mathbb{P}((m_1, m_2, \dots M_q) \in M^q) = \prod_{j=2}^{q} 2^{i_j} \prod_{j=1}^{q} \frac{(i_j)^{k-2}}{2^{i_j}} = \frac{\prod_{j=1}^{q} i_j^{k-2}}{2^{i_1}}.$$

Since we seek a lower bound for this (as we would like μ to be large) let us now retroactively impose that $\frac{n}{a_j \log n} \leq 2^{i_j} < \frac{2n}{a_j \log n}$; by our choices for i_j if we set $k = 8$ it follows that $\mu = n^{1-o(1)}$. Note that this choice is not necessarily the tightest possible. In fact it is possible to do better but we shall leave this as it is.

The other entity that we need to compute is Δ. In order to make this clear, first note that $(m_1, m_2, \dots, m_q)$ and $(m'_1, m'_2, \dots, m'_q)$ feature independent choices unless $m_i = m'_i$ for some i. Denote $\mathfrak{m} = (m_1, m_2, \dots, m_q), \mathfrak{m}' = (m'_1, m'_2, \dots, m'_q)$. We say $\mathfrak{m} \sim \mathfrak{m}'$ if $m_i = m'_i$ for some i and $\mathfrak{m} \neq \mathfrak{m}'$. Then

$$\Delta = \sum_{\mathfrak{m} \sim \mathfrak{m}'} \mathbb{P}(\mathfrak{m} \wedge \mathfrak{m}' \text{ are special}).$$

Denote by Δ_1 the sub-sum over all pairs $(\mathfrak{m}, \mathfrak{m}')$ for which $m_1 \neq m_1'$. Then we claim that

$$\Delta_1 \leq (\frac{i_1^6}{2^{i_1}})^2 \sum_{\substack{I \subset [2,q] \\ \emptyset \neq I}} \prod_{j \in I} 2^{i_j} \prod_{j \notin I} \{2^{i_j}(2^{i_j} - 1)\} \prod_{j \in I} \frac{i_j^6}{2^{i_j}} \prod_{j \notin I} \left(\frac{i_j^6}{2^{i_j}} \right)^2 . \qquad (\ddagger)$$

Let us quickly see how this comes about. For each such pair $(\mathfrak{m}, \mathfrak{m}')$ contributing to Δ_1, there is some non-empty set I of indices j where $m_j = m_j'$. For a fixed such I, the number of choices for $m_i (= m_i') \in I)$ equals $\prod_{j \in I} 2^{i_j}$. For the other indices the number of such choices for m_j, m_j' equals $prod_{j \notin I} \{2^{i_j}(2^{i_j} - 1)\}$. Finally, once these choices are fixed, m_1, m_1' are determined. It is of course possible that such free choices might result in $m_1 = m_1'$ but then this expression is only an upper bound, so we are not missing anything. Finally, the last two products in (‡) give the respective probabilities. A simple computation (and using the estimates we have made so far) shows that we get

$$\Delta_1 \leq \mu^2 \sum_{\substack{I \subset [2,q] \\ \emptyset \neq I}} \prod_{j \in I} \frac{1}{i_j^6} = \mu^2 \left[\prod_{j=1}^{q} \left(1 + \frac{1}{i_j^6} \right) - 1 \right] \leq O \left(\frac{\mu^2}{\log^5 n} \right).$$

More generally, we may now define for each $r \geq 1$ the sub-sum Δ_r over all pairs $(\mathfrak{m}, \mathfrak{m}')$ such that $m_1 = m_1', m_2 = m_2', \ldots, m_{r-1} = m_{r-1}', m_r \neq m_r'$. In a similar manner we have

$$\Delta_r \leq 2^{i_r} \left(\frac{i_r^6}{2^{i_r}} \right)^2 \prod_{j=2}^{r-1} 2^{i_j} \prod_{j=2}^{r-1} \frac{i_j^6}{2^{i_j}} \sum_{I \subset (r,q]} \prod_{j \in I} 2^{i_j} \prod_{j \in (r,q] \setminus I} (2^{i_j}(2^{i_j} - 1)) \prod_{j \in I} \frac{i_j^6}{2^{i_j}} \prod_{j \in (r,q] \setminus I} \left(\frac{i_j^6}{2^{i_j}} \right)^2 \left(\frac{i_1^6}{2^{i_1}} \right).$$

Again, to give a quick explanation; the first term in (‡‡) corresponds to the number of choices of m_r followed by the probability of picking m_r, m_r'; the next product term give the number of choices for m_i for $2 \leq i \leq r-1$ followed by the next product term that gives the probability of seeing any of those chosen $r-1$ tuples $(m_1, m_2, \ldots, m_{r-1})$. As for the terms in the summation, again, as before, fix an $I \subset [r+1, q]$ which determines the subset of other indices where $m_j = m_j'$. Then the first product term inside the summation gives the number of choices for m_j for $j \in I$; the next gives the number of ordered pairs (m_j, m_j') for $j \notin I, j > r$, and finally the last two terms correspond to the probabilities of picking those appropriate terms m_j. Finally, note that since all the m_i are chosen for $i \geq 2$, m_1 is uniquely determined, and the last term is the probability of picking that m_1. This then forces at most one value for m_r'. Thus the expression in (‡‡) give an upper bound.

Thus,

$$\Delta_r < \mu^2 \frac{2^{i_i}}{2^{i_r}} \prod_{j=1}^{r-1} \frac{1}{i_j^6} \sum_{I \subset [r+1,q]} \prod_{j \in I} \frac{1}{i_j^6}.$$

By some more calculations, one can show that

$$\Delta_r \leq O \left(\frac{\mu^2}{\log^7 n} \right) \text{ for } r \geq 3 \text{ and}$$
$$\Delta_2 \leq O \left(\frac{\mu^2}{\log^5 n} \right).$$

Since $q = o(\log n)$ it follows that

$$\Delta = \sum_{r=1}^{q} \Delta_r = O\left(\frac{\mu^2}{\log^5 n}\right).$$

Now Janson's inequality applies with room to spare. The details are left to the reader. $\square$