Probabilistic method

Benny Sudakov

September 6, 2020

Acknowledgement

Most of the material in these notes is based on the book: *Probabilistic method* by Noga Alon and Joel Spencer.

Contents

1 The Basics

The probabilistic method is a powerful technique used in combinatorics and other branches of mathematics, as well as in computer science. Basically, the method is used as follows: In order to show that a structure with a certain property exists, one shows that an appropriately defined random structure has the property with positive probability. The best way of learning the method is through examples, and we will see a lot of them in this course.

1.1 Ramsey numbers

We start with a classic example. For positive integers k and ℓ, let $R(k,\ell)$ be the smallest integer N such that for every red-blue coloring of the edges of K_N there exists a red clique of size k or a blue clique of size ℓ. Determining $R(k,\ell)$ for various k,ℓ is one of the major problems in combinatorics.

Remark 1.1. It is an easy exercise to show $R(3,3)=6$, which means that however we color K_6 with two colors, there is a clique of size 3 in one of the colors, and that there exists a coloring of K_5 with no monochromatic clique of size 3. One can also show that $R(4,4)=18$. It is known that $43 \leq R(5,5) \leq 48$, but we are far from knowing the actual value of $R(5,5)$.

Proposition 1.2. $R(k,\ell) \leq 2^{k+\ell}$

Proof. We want to show that every red-blue coloring of the edges of K_n (for $n=2^{k+\ell}$) contains a red K_k or a blue K_ℓ. The proof uses induction on $k+\ell$. The base case when $k+\ell=2$ is trivial. Let $n=2^{k+\ell}$. Let v be an arbitrary vertex in K_n. Then v has at least $\lceil \frac{n-1}{2} \rceil = 2^{k+\ell-1}$ incident edges of the same color. W.l.o.g., assume these edges are blue. By induction, the neighbours of v which touch these blue edges, contain either a red K_k (which means we are done), or a blue $K_{\ell-1}$. In the latter case v extends this blue $K_{\ell-1}$ to a blue K_ℓ, so we are done. □

Proposition 1.3. $(k-1)^2 \leq R(k,k) \leq 2^{2k}$

Proof. For the lower bound, partition $K_{(k-1)^2}$ into $k-1$ sets of vertices of equal size and color the edges inside of the sets red, and all other edges blue. There is no red or blue clique of size k in such a coloring of $K_{(k-1)^2}$, which proves that $(k-1)^2 \leq R(k,k)$. The upper bound follows from Proposition 1.2. □

Now let us show our first example of the usage of the probabilistic method; we substantially improve the lower bound from the previous proposition.

Theorem 1.4 (Erdős, 1947)**.** $2^{k/2} \leq R(k,k)$

Proof. Let $n=2^{k/2}$. We color each of the edges of K_n randomly and independently with probability 1/2 in color blue, and red otherwise. Observe the following:

- For a fixed subset $S \subseteq V(K_n)$ with $|S| = k$, let A_S denote the event that S forms a monochromatic clique; then $P(A_S) = 2 \cdot 2^{-\binom{k}{2}}$. This is because each of the $2^{\binom{k}{2}}$ possible colorings of edges in S happens with the same probability $2^{-\binom{k}{2}}$, while there are only 2 outcomes which give a monochromatic coloring.
- For any two events A, B it holds that $P(A \cup B) \leq P(A) + P(B)$. (This is called the union bound and we will use it often in this course.)

Now we have that the probability that at least one A_S happens is bounded as follows:

$$P(\exists S\colon A_S) = P\left(\bigcup_S A_S\right) \leq \sum_S P(A_S) = \binom{n}{k} 2^{-\binom{k}{2}+1} \leq \frac{n^k}{k!} 2^{1+k/2-k^2/2} = \frac{2^{1+k/2}}{k!} \cdot \frac{n^k}{2^{k^2/2}},$$

because the number of sets S of size k in $V(K_n)$ is $\binom{n}{k}$. By plugging in $n = 2^{k/2}$, we see that the probability that at least one A_S happens is less than 1 (when $k \geq 3$). Consequently, the probability that none of the A_S happens is positive, which means that there exists a coloring where none of the sets of size k forms a monochromatic clique. □

Remark 1.5. Finding explicit constructions of such colorings is an open problem. Furthermore, the bounds $2^{k/2} \leq R(k,k) \leq 2^{2k}$ had only minor improvements since the 1940's.

1.2 Domination property in tournaments

Definition 1.6. A *tournament* T_n is a directed graph obtained by orienting the edges of the complete graph K_n. So a tournament has exactly one edge between every pair of vertices u, v: either $u \to v$ or $v \to u$. We say that a tournament T_n has the *S_k-property* if for all $S \subseteq V(T_n)$ with $|S| \leq k$, there is a vertex in T_n which dominates all vertices in S. (We say that vertex u dominates vertex v if the directed edge uv points to v.)

Theorem 1.7 (Erdős). *If $n \geq 3k^2 2^k$, then there exists a tournament T_n with the S_k-property.*

Proof. Look at K_n and orient its edges independently and uniformly at random, i.e. for each pair of vertices u, v, orient the edge $\{u, v\}$ independently with $P(u \to v) = P(v \to u) = 1/2$. In the resulting graph, for a fixed set of vertices S of size $|S| = k$, the probability that a fixed vertex v outside of S dominates every vertex in S is:

$$P(v \text{ dominates } S) = 2^{-k}$$

Since for a fixed S those events are independent for different v, the probability that a fixed S does not have a vertex which dominates it, is exactly $(1-2^{-k})^{n-k}$. Since there are $\binom{n}{k}$ such sets S, using the estimates $1 - x \leq e^{-x}$ for all x, and $\binom{n}{k} \leq n^k$, by a union bound we get that:

$$P(\text{No set } S \text{ of size } k \text{ is dominated}) \leq \binom{n}{k}(1-2^{-k})^{n-k} \leq n^k e^{-\frac{n-k}{2^k}} = e^{k \log n - \frac{n-k}{2^k}} < 1$$

where the last inequality comes from $n \geq 3k^2 2^k$. □

Remark 1.8. Let $f(k)$ be smallest n such that there exists a tournament on n vertices with the S_k property. Szekeres proved a lower bound $ck2^k \leq f(k)$ for a sufficiently large constant c, while Theorem 1.7 shows that $f(k) \leq 3k^2 2^k$. In the homework assignment we will show $2^{k-1} \leq f(k)$.

1.3 Hypergraph coloring

Definition 1.9. A hypergraph $G = (V, E)$ is a set of vertices V together with a set of (hyper)edges E, where each (hyper)edge is a subset of V. A hypergraph is k-uniform if every edge contains exactly k vertices.

Example 1.10. A 2-uniform hypergraph is simply a graph.

We define the following:

- A k-coloring of the vertices of a hypergraph G is called *proper*, if no edge in G is monochromatic.
- $m(k)$ is the smallest number m such that there exists a k-uniform hypergraph with m edges for which there is no proper 2-coloring.

If $k = 2$, then we are working with graphs, so $m(2) = 3$, because a triangle cannot be properly colored with 2 colors.

Proposition 1.11 (Erdős)**.** *$2^{k-1} < m(k) \leq ck^2 2^k$ for a large enough constant c.*

Proof. First we prove the lower bound. For this we find a proper 2-coloring for each k-uniform hypergraph with $m \leq 2^{k-1}$ edges. Let $G = (V, E)$ be a k-uniform hypergraph with m edges, and color its vertices uniformly at random with two colors. For an edge $e \in E$, let A_e be the event that e is monochromatic. Then we have that $P(A_e) = 2 \cdot 2^{-k}$. Now it holds that

$$P(\text{some edge is monochromatic}) = P(\bigcup_{e \in E} A_e) \leq \sum_{e \in E} P(A_e) = \frac{m}{2^{k-1}}$$

where, in fact, the last inequality is strict, as some of the events are not disjoint (for example, the outcome where each vertex in G gets the same color is in each A_e). Thus the probability that some edge is monochromatic is strictly less than 1, hence with positive probability the coloring is proper. This proves the existence of the desired coloring, and our lower bound.

For the upper bound, we may assume that $k \geq 10$, because we can easily choose a constant c so that $m(k) \leq ck^2 2^k$ holds for all $k < 10$. Now we find a k-uniform hypergraph G with $m = 3k^2 2^k$ edges, which is not 2-colorable. In order to do this, we in some sense turn the proof for the lower bound on its head – we take a randomly generated hypergraph and use a union bound over all possible colorings to show that no coloring is proper with positive probability. For the vertices of G take $V(G) = [k^2]$, and for the edges choose m subsets of size k from $V(G)$, uniformly at random with repetitions (if an edge is chosen more times, we still count it once in the resulting hypergraph).

Now, if we fix any 2-coloring of $V(G)$, then there is a subset of vertices of size at least $k^2/2$ which have the same color. The probability that a randomly chosen subset of size k is contained in this subset is

$$\frac{\binom{k^2/2}{k}}{\binom{k^2}{k}} = \frac{\frac{k^2}{2}\dots(\frac{k^2}{2}-k+1)}{k^2\dots(k^2-k+1)} = 2^{-k}\prod_{i=0}^{k-1}\frac{k^2-2i}{k^2-i} = 2^{-k}\prod_{i=0}^{k-1}\left(1-\frac{i}{k^2-i}\right)$$
$$> 2^{-k}\prod_{i=0}^{k-1}\exp\left(-2\frac{i}{k^2-i}\right) > 2^{-k}\exp\left(-2\frac{\sum_{i=0}^{k-1} i}{k^2-k}\right) = 2^{-k}e^{-1}$$

(Here, in the first inequality we have used that $1-x > e^{-2x}$ for $x < 1/4$.) Therefore, the probability that a fixed edge is not monochromatic is at most $1-2^{-k}e^{-1}$, and since the edges are chosen independently, the probability that none of them is monochromatic is at most $(1-2^{-k}e^{-1})^m \leq \exp\left(-\frac{m}{e2^k}\right) < e^{-k^2}$. Using this, and a union bound over all colorings of $V(G)$ (there are $2^{|V(G)|} = 2^{k^2}$ of them), we get that the probability that there is a coloring which is proper is at most $e^{-k^2}2^{k^2} < 1$, so there exists a hypergraph with m edges with no proper 2-coloring. □

Remark 1.12. The best known lower bound is $c\sqrt{k/\log k}\cdot 2^k \leq m(k)$.

Theorem 1.13 (Bollobás)**.** *Suppose that a family of pairs of sets (A_i, B_i) for $i \in [m]$ satisfies the following properties:*

- *$|A_i| = a$, for all $i \in [m]$,*
- *$|B_i| = b$, for all $i \in [m]$,*
- *$A_i \cap B_i = \emptyset$, and $A_i \cap B_j \neq \emptyset$ if $i \neq j$.*

Then $m \leq \binom{a+b}{a}$.

Proof. Let $X = \bigcup_i (A_i \cup B_i)$ be the set of all elements which appear in at least one of the sets. Take a permutation $\pi\colon X \to X$ uniformly at random from the set of all permutations of X. Let E_i be the event that all elements of A_i are before the elements of B_i in the obtained permutation. Then

$$P(E_i) = \frac{a!b!}{(a+b)!}.$$

Note that all events E_i are disjoint. Indeed, if E_i happens, then all elements of A_i are before B_i in the permutation. Therefore, for all $j \neq i$, A_j cannot be before B_j, because $A_j \cap B_i \neq \emptyset$ and $B_j \cap A_i \neq \emptyset$, so E_j cannot happen.

Finally, since all events are disjoint, it holds that

$$1 \geq P(\bigcup_i E_i) = \sum_i P(E_i) = m \cdot \frac{a!b!}{(a+b)!},$$

hence our claim follows. □

Remark 1.14. This bound is tight. Consider $[a+b] = \{1, \dots, a+b\}$ and let each A_i be one of the subsets of size a of $[a+b]$, and let B_i be its complement $[a+b] \setminus A_i$. Then there are $\binom{a+b}{a}$ pairs of sets and the conditions in the theorem hold.

2 Linearity of Expectation

Let $X_1, \ldots, X_n$ be random variables with $|E[X_i]| < \infty$ for each i, and let $X = X_1 + \ldots + X_n$. Then $E[X] = E[X_1] + \ldots + E[X_n]$. This identity is a very useful tool often used in probabilistic arguments. Its power lies in the fact that there are no restrictions on the dependencies between the X_i. In many cases this principle is used to calculate $E[X]$ where X is the sum of very simple (indicator) random variables. One common way that expectation is used is the trivial but very useful fact that if $E[X] = M$, then there is a point in the probability space where $X \geq M$.

Example 2.1. Let π be a permutation on $[n]$, chosen uniformly at random. Let X be the random variable which counts the number of fixed points of π, i.e. the elements $x \in [n]$ for which $\pi(x) = x$. Then $X = X_1 + \ldots + X_n$, where each X_i is the indicator random variable of the event $\pi(i) = i$. Then we have:

$$E[X_i] = P(X_i = 1) = 1/n$$

which implies that $E[X] = n \cdot \frac{1}{n} = 1$.

2.1 Maximum Cut

Given a graph $G = (V, E)$, we call a partition of its vertices into disjoint sets $A \cup B$ a *cut*, and the set of edges which have an endpoint in both A and B, a *cut-set.* Given G, how large can a cut-set be? Or in other words, what is the largest bipartite subgraph of G?

Claim 2.2. Every graph G with m edges has a cut-set of size at least $m/2$.

Proof. Take a random partition $V = A \cup B$ such that for each vertex $v \in V(G)$:

$$P(v \in A) = P(v \in B) = 1/2$$

(and the choices for different vertices are independent). Let X be the number of edges in the cut-set induced by A and B. Then it holds that $X = \sum_{e \in E(G)} X_e$ where

$$X_e = \begin{cases} 1, \, e \in E(A,B) \\ 0, \text{ otherwise} \end{cases}$$

If $e = (u, v) \in E(A, B)$ then either $u \in A$ and $v \in B$, or $v \in A$ and $u \in B$. Therefore, $E[X_e] = 2 \cdot (1/2)^2$. Hence, $E[X] = \sum_{e \in E(G)} \frac{1}{2} = m/2$. Since the expected size of this random cut-set is $m/2$, there exists a cut-set of at least this size. □

2.2 Tournaments and Hamilton Paths

Recall that a tournament is an orientation of the complete graph K_n. A Hamiltonian path in a directed graph is a directed path which contains all the vertices of the graph. It is not hard to show that every tournament has a Hamiltonian path. Furthermore, it is easy to construct a tournament with just one Hamiltonian path: take an arbitrary ordering of the vertices of K_n, and orient the edges to go from smaller to larger vertices in the ordering (this is called the transitive tournament). A natural question is what is the largest number of Hamiltonian paths a tournament can have.

Theorem 2.3 (Szelle, 1943)**.** *There exists a tournament T_n on n vertices with at least $\frac{n!}{2^{n-1}}$ Hamiltonian paths.*

Proof. We orient the edges of K_n independently and uniformly at random to obtain a tournament T_n. Let X be the random variable which counts the number of Hamiltonian paths in T_n. Then, $E[X] = \sum_\pi E[X_\pi]$, where π goes over all orderings of the vertices in T_n, and X_π is an indicator random variable which is equal to 1 if π is a Hamiltonian path in T_n, and 0 otherwise.

Now we have $E[X_\pi] = P(X_\pi) = (1/2)^{n-1}$, as each of the $n-1$ edges in the path must be oriented accordingly. This gives $E[X] = \sum_\pi E[X_\pi] = \sum_\pi \frac{1}{2^{n-1}} = \frac{n!}{2^{n-1}}$. □

2.3 Combinatorial Number Theory

Let $A \subset \mathbb{Z} \setminus \{0\}$. We say that a subset $B \subseteq A$ is *sum free* if $b_1 + b_2 \neq b_3$ for all $b_1, b_2, b_3 \in B$. Given $|A| = n$, how large can a sum free subset of A be?

Theorem 2.4 (Erdős)**.** *Let $A \subset \mathbb{Z} \setminus \{0\}$ with $|A| = n$. Then there exists a sum free $B \subseteq A$ with $|B| > n/3$.*

Proof. Let $p = 3k+2$ be a prime number such that $p > 2\max\{|a| : a \in A\}$. Then every two elements of A have different residues modulo p, and none of them is zero, so we can think of A as a subset of $\mathbb{Z}_p \setminus \{0\} = \{1, \ldots, p-1\}$. Let $I = [k+1, 2k+1]$, and observe that I is sum free modulo p. Let $z \in \mathbb{Z}_p \setminus \{0\}$ be chosen uniformly at random, and let $B \subseteq A$ be the set of elements $a \in A$ for which $a \cdot z \in I$, where multiplication is done modulo p. Then B is sum free modulo p (and hence also over the integers), because if $b_1 + b_2 = b_3$ (modulo p) then also $zb_1 + zb_2 = zb_3$ (modulo p), which is impossible if $zb_i \in I$ for $1 \leq i \leq 3$.

Now we estimate $E[|B|]$. Note that for a fixed $a \in A$, $z \cdot a$ ranges over all non-zero elements of Z_p, when z ranges over the same set. Therefore, for a fixed $a \in A$, we have

$$P(z \cdot a \in I) = \frac{|I|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}.$$

Hence $E[|B|] > n/3$, because of linearity of expectation, so we are done. □

3 Permanents of Matrices

Definition 3.1. Let A be an $n \times n$ matrix. The *determinant* of A is

$$\det(A) = \sum_{\sigma \in S_n} (-1)^{\mathrm{sign}(\sigma)} \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

The *permanent* of A is

$$\mathrm{Per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

Remark 3.2. Computing the permanent of a matrix is $\#P$ (Read: Sharp P), and it is even harder than NP-complete problems.

3.1 Counting Hamiltonian Paths

Recall that in Section 2.2 we have shown that a random tournament on n vertices has in expectation $\frac{n!}{2^{n-1}}$ Hamiltonian paths. Given a tournament T, denote by $P(T)$ the number of Hamiltonian paths in T. Denote by $P(n)$ the maximum number of Hamiltonian paths an n-vertex tournament can have. In this section we are going to use Brégman's theorem (which we will state shortly) to prove that no tournament can have too many more Hamiltonian paths than a random tournament:

Theorem 3.3 (Alon)**.** $P(n) \leq cn^{3/2}\frac{n!}{2^{n-1}}$.

Given a tournament T, define

$$a_{ij} = \begin{cases} 1 & i \to j \\ 0 & j \to i \text{ or } j = i \end{cases}$$

Then $A_T = (a_{ij})_{i,j}$ is a matrix corresponding to this tournament, and there are precisely $\binom{n}{2}$ 1's and $\binom{n}{2} + n$ 0's. Note that whenever $\sigma \in S_n$ is a permutation, we have

$$\prod_i a_{i\sigma(i)} = \begin{cases} 1 & \forall i.\ a_{i\sigma(i)} = 1 \iff \forall i.\ i \to \sigma(i) \\ 0 & \text{otherwise.} \end{cases}$$

Note that for a Hamilton cycle $v_1 v_2 \ldots v_n$ we can define a permutation σ by $\sigma(v_i) = v_{i+1}$, and then we have $\prod_i a_{i\sigma(i)} = 1$. Different Hamilton cycles lead to different permutations, so any upper bound on $\mathrm{Per}(A_T)$ gives an upper bound on the number of Hamiltonian cycles in the tournament T.

Consider $r_i = \sum_{j=1}^{n} a_{ij}$, i.e. the sum of entries in the i-th row. Then $\mathrm{Per}(A_T) \leq \prod_{i=1}^{n} r_i$ is a trivial upper bound on $\mathrm{Per}(A_T)$, but we can do better.

Theorem 3.4 (Brégman's theorem, Minc conjecture)**.** *Let A be an n by n matrix with 0-1 entries and row sums r_i. Then*

$$\mathrm{Per}(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i} \approx \prod_{i=1}^{n} \frac{r_i}{e}.$$

(Note: $n! \approx c\left(\frac{n}{e}\right)^n \sqrt{n}$)

Proof. We will prove this later. □

For every tournament T, we know that the row sums r_i of A_T satisfy $\sum_i r_i = \frac{n(n-1)}{2}$. Using Brégman's theorem, we want to estimate $\prod_i (r_i!)^{1/r_i}$. We claim that this is maximized when $r_i = \frac{n-1}{2}$, for each $i \in [n]$. To prove this, we will use the following technical lemma.

Lemma 3.5. *Let $b \geq a+2 > a \geq 1$. Then $(a!)^{1/a}(b!)^{1/b} < ((a+1)!)^{1/(a+1)})((b-1)!)^{1/(b-1)}$.*

Proof sketch: Define $f(a) = \frac{(a!)^{1/a}}{((a+1)!)^{1/(a+1)}}$. Then we wish to show that $f(a) < f(b-1)$. To do this, we will show that f is an increasing function. Note that

$$f(a) = (a+1)^{-\frac{1}{a+1}}(a!)^{\frac{1}{a(a+1)}},$$

and the rest is easy computation. □

Corollary 3.6. *The solution to $\sum_i r_i = \frac{n(n-1)}{2}$ that maximizes $\prod_i (r_i!)^{1/r_i}$ is the one when the r_i-s are all as equal as possible, i.e. they all differ by at most one.*

Proof. Suppose there are i, j such that $r_i \geq r_j + 2$. By the previous lemma, the product increases if we decrease r_i by one and increase r_j by one. □

Given a tournament T, denote by $\mathrm{HC}(T)$ the number of Hamiltonian cycles in T.

Corollary 3.7. *In any n-vertex tournament T, we have*

$$\mathrm{HC}(T) \leq \prod_{i=1}^{n} \left(\left(\frac{n-1}{2}\right)!\right)^{\frac{2}{n-1}} \approx n^{1/2}\frac{n!}{2^{n-1}}.$$

Proof sketch: The first inequality follows from Brégman's theorem (and the discussion before it). For the second estimate, note that $x! \approx c\left(\frac{x}{e}\right)^x \sqrt{x}$ and so $\left(\frac{n-1}{2}\right)! \approx c\left(\frac{n-1}{2e}\right)^{\frac{n-1}{2}} \sqrt{n}$. □

Lemma 3.8. *Given any tournament T on n vertices, there exists a tournament T' on $n+1$ vertices such that $\mathrm{HC}(T') \geq \frac{1}{4}P(T)$.*

Remark 3.9. This shows that $P(T) \leq 4\,\mathrm{HC}(T') \leq c(n+1)^{1/2}\frac{(n+1)!}{2^n} \approx c'n^{3/2}\frac{n!}{2^{n-1}}$.

Proof. The tournament T' will be obtained from T by attaching a new vertex v to T. Orient the edges from v to the vertices of T independently at random. Given any Hamiltonian path $P = v_1v_2\ldots v_n$ in T, it gives a Hamiltonian cycle in T' if $v \to v_1$ and $v_n \to v$. Hence the probability that any Hamiltonian path in T extends to a Hamiltonian cycle in T' is $1/4$. Every Hamiltonian path gives rise to a different Hamiltonian cycle. Hence, using the linearity of expectation, the expected number of Hamiltonian cycles in T' is $P(T)/4$. Thus there must exist an orientation of the edges touching v for which the resulting tournament T' satisfies $HC(T') \geq P(T)/4$. □

This concludes the upper bound proof for $P(T)$, it only remains to prove Brégman's theorem.

Example. Let A be an $n \times n$ $\{0,1\}$ matrix, $t_1, \ldots, t_m$ such that $\sum t_i = n$, with A being a block matrix; the ith block of A has size $t_i \times t_i$, each entry in the blocks is 1 and each entry outside is 0. Then $\prod_{i=1}^{n} (r_i!)^{1/r_i} = \prod_{i=1}^{m} t_i!$. Moreover, $\mathrm{Per}(A) = \prod_{B:B \text{ is a block of } A} \mathrm{Per}(B)$, and if B is a block of size $t_i \times t_i$ then $\mathrm{Per}(B) = t_i!$. Hence for such block matrices A, we have $\mathrm{Per}(A) = \prod_{i=1}^{n} (r_i!)^{1/r_i}$, showing that Brégman's theorem is tight.

Proof of Brégman's theorem. Let S be the set of permutations $\sigma \in S_n$ with $a_{i,\sigma(i)} = 1$ for every $1 \leq i \leq n$. Then the permanent of A is simply $|S|$. Pick $\sigma \in S$ and $\tau \in S_n$ independently and uniformly at random. Set $A^{(1)} = A$. Let $R_{\tau(1)}$ be the number of ones in row $\tau(1)$ in $A^{(1)}$. Delete row $\tau(1)$ and column $\sigma\tau(1)$ from $A^{(1)}$ to give $A^{(2)}$. In general, let $A^{(i)}$ denote A with rows $\tau(1), \ldots, \tau(i-1)$ and columns $\sigma\tau(1), \ldots, \sigma\tau(i-1)$ deleted and let $R_{\tau(i)}$ denote the number of ones of row $\tau(i)$ in $A^{(i)}$. (This is nonzero as the $\sigma\tau(i)$th column has a one.) Set

$$L = L(\sigma, \tau) = \prod_{i=1}^{n} R_{\tau(i)}.$$

We think, roughly, of L as Lazyman's permanent calculation. There are $R_{\tau(1)}$ choices for a one in row $\tau(1)$, each of which leads to a different subpermanent calculation. Instead, Lazyman takes the factor $R_{\tau(1)}$, takes the one from permutation σ, and examines $A^{(2)}$. As $\sigma \in S$ is chosen uniformly, Lazyman tends toward the high subpermanents and so it should not be surprising that he tends to overestimate the permanent. To make this precise we define the geometric mean $G[Y]$. If $Y > 0$ takes values $a_1, \ldots, a_s$ with probabilities $p_1, \ldots, p_s$, respectively, then $G[Y] = \prod a_i^{p_i}$. Equivalently, $G[Y] = e^{E[\ln Y]}$. Linearity of expectation translates into the geometric mean of a product being the product of the geometric means.

Claim 3.10. $\mathrm{Per}(A) \leq G[L]$.

Proof. We show this for any fixed τ. Set $\tau(1) = 1$ for convenience of notation. We use induction on the size of the matrix. Reorder, for convenience, so that the first row has ones in the first r columns, where $r = r_1$. For $1 \leq j \leq r$ let t_j be the permanent of A with the first row and jth column removed or, eqivalently, the number of $\sigma \in S$ with $\sigma(1) = j$. Set

$$t = \frac{t_1 + \ldots t_r}{r}$$

so that $\mathrm{Per}(A) = rt$. Conditioning on $\sigma(1) = j$, $R_2 \cdots R_n$ is Lazyman's calculation of $\mathrm{Per}(A^{(2)})$, where $A^{(2)}$ is A with the first row and jth column removed. By induction

$$G[R_2 \cdots R_n | \sigma(1) = j] \geq t_j$$

and so

$$G[L] \geq \prod_{j=1}^{r} (rt_j)^{t_j/\mathrm{Per}(A)} = r \prod_{j=1}^{r} t_j^{t_j/rt}.$$

Lemma 3.11. $\left(\prod_{j=1}^{r} t_j^{t_j}\right)^{1/r} \geq t^t$.

Proof. Taking logarithms, this is equivalent to

$$\frac{1}{r}\sum_{j=1}^{r} t_j \ln t_j \geq t \ln t,$$

which follows from the convexity of the function $f(x) = x \ln x$. □

Applying the lemma,

$$G[L] \geq r \prod_{j=1}^{r} t_j^{t_j/rt} \geq r(t^t)^{1/t} = rt = \mathrm{Per}(A).$$

□

Now we calculate $G[L]$ conditional on a fixed σ. For convenience of notation reorder so that $\sigma(i) = i$ for all i, and assume that the first row has ones in precisely first r_1 columns. With τ selected uniformly, the columns $1, \ldots, r_1$ are deleted in order uniform over all $r_1!$ possibilities. R_1 is the number of those columns remaining when the first column is to be deleted. Ad the first column is equally likely to be in any position among those r_1 columns R_1 is uniformly distributed from 1 to r_1 and $G[R_1] = (r_1!)^{1/r_1}$. "Linearity" then gives

$$G[L] = G\left[\prod_{i=1}^{n} R_i\right] = \prod_{i=1}^{n} G[R_i] = \prod_{i=1}^{n} (r_i!)^{1/r_i}.$$

The overall $G[L]$ is the geometric mean of the conditional $G[L]$ and hence has the same value. That is,

$$\mathrm{Per}(A) \leq G[L] = \prod_{i=1}^{n} (r_i!)^{1/r_i},$$

as desired. □

4 Method of Alterations

The basic probabilistic method was described in Section 1 as follows: Trying to prove that a structure with certain desired properties exists, one defines an appropriate probabilistic space of structures and then shows that the desired properties hold in this space with positive probability. In this section we consider situations where the "random" structure does not have all the desired properties but may have a few "blemishes". With a small alteration we remove the blemishes, giving the desired structure.

4.1 High girth and high chromatic number

Many consider this one of the most pleasing uses of the probabilistic method, as the result is surprising and does not appear to call for nonconstructive techniques. The *girth* $g(G)$ of a graph G is the size of its shortest cycle, $\alpha(G)$ is the size of the largest independent set in G and $\chi(G)$ denotes its chromatic number.

Theorem 4.1 (Erdős (1959))**.** *For all k, ℓ there exists a graph G with $g(G) > \ell$ and $\chi(G) > k$.*

Proof. Fix $\theta < 1/\ell$ and let $G \sim G(n,p)$ with $p = n^{\theta-1}$; that is, G is a random graph on n vertices chosen by picking each pair of vertices as an edge randomly and independently with probability p. Let X be the number of cycles of size at most ℓ. The total possible number of cycles of length i is $\frac{n(n-1)\dots(n-i+1)}{2i}$ (this is the number of such cycles in K_n). It is convenient to use the notation $(n)_i := n(n-1)\dots(n-i+1)$. The reason that we divide by $2i$ is that $(n)_i$ counts each cycle $2i$ times: there are i ways to choose a starting point and 2 directions in which we can traverse the cycle. Each of the $\frac{(n)_i}{2i}$ possible cycles appears in G with probability p^i. Therefore:

$$\mathbb{E}[X] = \sum_{i=3}^{\ell} \frac{(n)_i}{2i} p^i \leq \sum_{i=3}^{\ell} \frac{n^{\theta i}}{2i} = o(n)$$

as $\theta\ell < 1$. In particular (by Markov's inequality),

$$\mathbb{P}[X \geq n/2] = o(1).$$

Set $x = \lceil \frac{3 \ln n}{p} \rceil > \frac{2 \ln n}{p} + 1$, so that

$$\mathbb{P}[\alpha(G) \geq x] \leq \binom{n}{x}(1-p)^{\binom{x}{2}} < \frac{1}{x!}\left[ne^{-p(x-1)/2}\right]^x = o(1).$$

Let n be sufficiently large so that both of these events have probability less than 0.5. Then there is a specific G with less than $n/2$ cycles of length at most ℓ and with $\alpha(G) < 3n^{1-\theta}\ln n$. Remove from G a vertex from each cycle of length at most ℓ. This gives a graph G^* with at least $n/2$ vertices. G^* has girth greater than ℓ and $\alpha(G^*) \leq \alpha(G)$. Thus

$$\chi(G^*) \geq \frac{|V(G^*)|}{\alpha(G^*)} \geq \frac{n/2}{3n^{1-\theta}\ln n} = \frac{n^\theta}{6 \ln n}.$$

To complete the proof, choose n to be sufficiently large so that this is greater than k.

4.2 Non 2-colourable n-uniform hypergraphs

Recall that in an earlier part of the course, we denoted by $m(n)$ the minimum number of edges in an n-uniform non 2-(vertex)-colourable hypergraph, and showed that $2^{n-1} < m(n) < cn^2 2^n$. We would like to now improve the lower bound and show that

$$2^{n-2}\sqrt{\frac{n}{\log n}} \leq m(n).$$

If an n-uniform hypergraph H has m edges, then a random red/blue colouring has

$$\mathbb{E}[\text{Number of Monochromatic Edges}] = m2^{1-n}.$$

Suppose H has $m = k2^{n-1}$ edges. Keep in mind that in the end, we will choose $k \approx \sqrt{\frac{n}{\log n}}$.

For every $v \in V(H)$, choose a $x_v \in [0,1]$ uniformly at random. Order the vertices of H in the increasing order of x_v. Colour the vertices one at a time in this order with the following rule:

- Colour v red whenever possible.
- Colour v blue if colouring v red would make some edge incident to v monochromatic red.

Observe that there are no red edges using the algorithm, since we would have coloured the last vertex appearing in the ordering blue. Thus if there is a monochromatic edge, it has to be a blue edge. If $f \in E(H)$ is a blue edge and v is the first vertex of f according to the ordering, then there must be some edge e such that $v \in e$, and all vertices of $e \setminus \{v\}$ are colored red and appear before v; in particular $e \cap f = \{v\}$.

Let us call such a pair of edges (e, f) a *conflicting pair*; that is, a conflicting pair satisfies that $e \cap f$ is a single vertex v, and the vertices of e appear before the vertices of f in the ordering that we chose (with v being the last vertex of e and the first vertex of f). We saw that if the ordering does not have any conflicting pairs, then the colouring has no monochromatic edges.

We now estimate the probability that conflicting pairs exist. To do this, let us partition the interval $[0, 1]$ into three parts; let

$$R := \left[0, \frac{1-p}{2}\right], \qquad P = \left[\frac{1-p}{2}, \frac{1+p}{2}\right], \qquad B = \left[\frac{1+p}{2}, 1\right],$$

where p will be chosen later. If (e, f) is a conflicting pair and $e \cap f = \{v\}$, then $x_u \leq x_v$ for every $u \in e$ and $x_w \geq x_v$ for every $w \in f$. In particular, either $e \subseteq R$ or $f \subseteq B$ or $v \in P$. We have

- $\mathbb{P}(e \subseteq R) = \left(\frac{1-p}{2}\right)^n = \mathbb{P}(f \subseteq B)$,
- $\mathbb{P}((\forall u \in e,\ x_u \leq x_v)\ \wedge\ (\forall w \in f,\ x_w \geq x_v) \wedge (x_v \in P)) \leq \frac{p}{2^{2(n-1)}}$.

The second inequality above needs some explanation. Let's call the left hand side of the inequality Q. Let us condition on the value of x_v, say $x_v = x \in P$. Then the conditional probability that $x_u \leq x_v$ for every $u \in e \setminus \{v\}$ and $x_w \geq x_v$ for every $w \in f \setminus \{v\}$ is $x^{n-1}(1-x)^{n-1}$. Therefore, using the law of total probability and $x(1-x) \leq 1/4$, we get

$$Q = \int_{\frac{1-p}{2}}^{\frac{1+p}{2}} x^{n-1}(1-x)^{n-1} dx \leq \int_{\frac{1-p}{2}}^{\frac{1+p}{2}} (1/4)^{n-1} dx = \frac{p}{2^{2(n-1)}}.$$

Therefore,

$$\begin{aligned}\mathbb{P}(H \text{ has conflicting pairs}) &\leq m\left(\frac{1-p}{2}\right)^n + m\left(\frac{1-p}{2}\right)^n + m^2 \frac{p}{2^{2(n-1)}} \\ &= \frac{k(1-p)^n}{2} + \frac{k(1-p)^n}{2} + k^2 p = k(1-p)^n + k^2 p \leq k \cdot e^{-np} + k^2 p.\end{aligned}$$

Now choosing $p = \frac{\log n}{n}$ and $k = \frac{1}{2}\sqrt{\frac{1}{p}} = \frac{1}{2}\sqrt{\frac{n}{\log n}}$ gives that this probability is at most $o(1) + 1/4 < 1$, as required. We conclude that every H with at most $k2^{n-1} = 2^{n-2}\sqrt{\frac{n}{\log n}}$ edges is 2-colourable. □

4.3 Dependent random choice

There is a technique that is very useful in studying Ramsey numbers, called dependent random choice. We will start with an example involving the hypercube. We denote the hypercube of dimension k by $Q_k := \{0,1\}^k$; the vertices are the k-bit binary strings, and $x \sim y$ if and only if $d_H(x,y) = 1$, where $d_H(\cdot,\cdot)$ is the Hamming distance. Then there are $n = 2^k$ vertices, it is a k-regular bipartite graph; bipartiteness comes from partitioning the vertices $V(Q_k)$ into vertices with odd number of 1's and even number of 1's.

What is the Ramsey number $R(Q_k)$, i.e. the minimum $N \in \mathbb{N}$ such that any 2-colouring of the edges of K_N contains monochromatic copy of Q_k? Obviously $R(Q_k) \leq R(K_n) \leq 2^{2n}$, as proved before in lectures. We will use the convention that an even vector is a vertex with even number of 1's and odd vector is a vertex with odd number of 1's. We will first prove a lemma, and then a main result.

Lemma 4.2. *Let a, d, m, n, r be positive integers. Let $G = (V, E)$ be a graph with $|V| = n$ vertices and average degree $d = 2|E(G)|/n$. If there is a positive integer t such that*

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

then G contains a subset U of at least a vertices such that every r vertices in U have at least m common neighbors.

Proof. For a vertex set T, we will denote by $N(T)$ the set of all common neighbours of T, that is, all $x \in V$ such that $\{x, y\} \in E$ for every $y \in T$. Pick a set T of t vertices of V uniformly at random with repetition. Set $A = N(T)$, and let X denote the cardinality of A. By linearity of expectation,

$$\mathbb{E}[X] = \sum_{v \in V(G)} \left(\frac{|N(v)|}{n}\right)^t = n^{-t} \sum_{v \in V(G)} |N(v)|^t \geq n^{1-t}\left(\frac{\sum_{v \in V(G)} |N(v)|}{n}\right)^t = \frac{d^t}{n^{t-1}},$$

where the inequality is by the convexity of the function $f(z) = z^t$.

Let Y denote the random variable counting the number of subsets $S \subset A$ of size r with fewer than m common neighbors. For a given such $S \subseteq V$, the probability that it is a subset of A equals $\left(\frac{|N(S)|}{n}\right)^t$. Since there are at most $\binom{n}{r}$ subsets $S \subseteq V$ of size r for which $|N(S)| < m$, it follows that

$$\mathbb{E}[Y] < \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

By linearity of expectation,

$$\mathbb{E}[X - Y] \geq \frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq a.$$

Hence there exists a choice of T for which the corresponding set $A = N(T)$ satisfies $X - Y \geq a$. Delete one vertex from each subset S of A of size r with fewer than m common neighbors. We let U be the remaining subset of A. The set U has at least $X - Y \geq a$ vertices and all subsets of size r have at least m common neighbors. □

Lemma 4.3. *Let $H = (A \cup B, F)$ be a bipartite graph in which $|A| = a$, $|B| = b$, and the vertices in B have degree at most r. If G is a graph with a vertex subset U with $|U| = a$ such that all subsets of U of size r have at least $a + b$ common neighbors, then H is a subgraph of G.*

Proof. We find an embedding of H in G given by an injective function $f : A \cup B \to V(G)$. Start by defining an injection $f : A \to U$ arbitrarily. Label the vertices of B as $v_1, \ldots, v_b$. We embed the vertices of B in this order one vertex at a time. Suppose that the current vertex to embed is $v_i \in B$. Let $N_i \subset A$ be those vertices of H adjacent to v_i, so $|N_i| \leq r$. Since $f(N_i)$ is a subset of U of cardinality at most r, there are at least $a + b$ vertices adjacent to all vertices in $f(N_i)$. As the total number of vertices already embedded is less than $a+b$, there is a vertex $w \in V(G)$ which is not yet used in the embedding and is adjacent to all vertices in $f(N_i)$. Set $f(v_i) = w$. It is immediate from the above description that f provides an embedding of H as a subgraph of G. □

Theorem 4.4. $r(Q_r) \leq 2^{3r} = n^3$.

Proof. In any two-coloring of the edges of the complete graph on $N = 2^{3r}$ vertices, the denser of the two colors has at least $\frac{1}{2}\binom{N}{2} \geq 2^{-7/3}N^2$ edges. Let G be the graph of the densest color, so the average degree d of G is at least $2^{-4/3}N$. Let $t = \frac{3}{2}r$, $m = 2^r$ and $a = 2^{r-1}$. We have

$$\frac{d^t}{N^{t-1}} - \binom{N}{r}\left(\frac{m}{N}\right)^t \geq 2^{-\frac{4}{3}t}N - N^{r-t}m^t = 2^r - 1 \geq 2^{r-1}.$$

Therefore, applying Lemma 4.2 we find in G a subset U of size 2^{r-1} such that every set of r vertices in U has at least 2^r common neighbors. Since Q_r is an r-regular bipartite graph with 2^r vertices and parts of size 2^{r-1}, Lemma 4.3 demonstrates that Q_r is a subgraph of G. □

Using the same type of proof, one can actually show that $R(Q_r) \leq 2^{2r+o(r)} \approx n^2$. More than thirty years ago, Burr and Erdős conjectured that $R(Q_r)$ is in fact linear in the number of vertices. Although this conjecture has drawn a lot of attention, it is still open.

5 Second moment method

In a typical application of the probabilistic method, we are given a random variable X that is a function of many Bernoulli random variables, e.g. X is a graph parameter, which is the function of the indicator random variables of the edges. In many such cases, we can observe that X is highly concentrated around its mean. How can one measure this?

We have already seen one such measure, namely Markov's inequality. To recall, Markov's inequality tells us that if X is non-negative, then for every $t > 0$, we have $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$. Unfortunately, this inequality does not tell us anything about the probability of X being much smaller than $\mathbb{E}(X)$.

However, if we can also calculate the variance of X, $Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, we get a stronger concentration inequality. (The expectation of the square, $\mathbb{E}(X^2)$, is also called the *second moment* of X.)

Theorem 5.1. *(Chebyshev's inequality) Let X be a random variable with finite mean and variance. Then for every $t > 0$,*

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

Proof. Let $Y = (X - \mathbb{E}(X))^2$. Then

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) = \mathbb{P}(Y \geq t^2) \leq \frac{\mathbb{E}(Y)}{t^2} = \frac{\mathrm{Var}(X)}{t^2},$$

where the inequality is due to Markov's inequality (noting that Y is indeed non-negative). □

The following immediate consequence of Chebyshev's inequality is useful to bound the probability that a given random variable is 0.

Theorem 5.2. *Let X be a random variable. Then*

$$\mathbb{P}(X = 0) \leq \frac{Var(X)}{(\mathbb{E}(X))^2}.$$

Proof. We have

$$\mathbb{P}(X = 0) \leq \mathbb{P}(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)) \leq \frac{Var(X)}{(\mathbb{E}(X))^2},$$

where the last inequality is the application of Chebyshev's inequality with $t = \mathbb{E}(X)$. □

5.1 Sum-free sets

As a first application of Chebyshev's inequality, or equivalently the second moment method, we will consider the following problem of Erdős.

Let $A \subset [n]$. Then A is called *sum-free* if the $2^{|A|}$ sums $\sum_{a \in S} a$ $(S \subseteq A)$, are all distinct. What is the largest possible size of a sum-free subset of $[n]$? Let $f(n)$ denote this number.

Clearly, we have $1 \leq \sum_{a \in S} a \leq |A| \cdot n$ for every $S \subseteq A$, so if A is sum-free, i.e. if all of these $2^{|A|}$ sums are different, then $2^{|A|} \leq |A| \cdot n$. This gives $f(n) < \log_2 n + \log_2 \log_2 n + C$ for some absolute constant C. On the other hand $A = \{2^i : i = 0, 1 \ldots, \lfloor \log_2 n \rfloor\}$ is sum-free, so $f(n) \geq \lfloor \log_2 n \rfloor$. It is an open question whether $f(n) - \log_2 n \to \infty$, but the upper bound can be slightly improved.

Theorem 5.3. $f(n) < \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1)$.

Proof. Let $A = \{a_1, \ldots, a_k\}$ be a sum-free set of size $k = f(n)$, and let $\epsilon_1, \ldots, \epsilon_k$ be independent random variables taking 0 or 1 with probability 1/2. Let $X = \sum_{i=1}^{k} \epsilon_i a_i$. (This is the same as looking at $\sum_{a \in S} a$ for a randomly chosen subset $S \subseteq A$.) Then

$$\mathbb{E}(X) = \sum_{i=1}^{k} \mathbb{E}(\epsilon_i a_i) = \frac{1}{2} \sum_{i=1}^{k} a_i$$

and

$$\mathrm{Var}(X) = \sum_{i=1}^{k} \mathrm{Var}(\epsilon_i a_i) = \frac{1}{4} \sum_{i=1}^{k} a_i^2 \leq \frac{1}{4} k n^2.$$

(Recall that $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$. In particular, if X, Y are independent then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.)

Therefore, by Chebyshev's inequality, we can write

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq n\sqrt{k}) \leq \frac{1}{4}.$$

Note that $\mathbb{P}(|X - \mathbb{E}(X)| \geq n\sqrt{k}) = \frac{N}{2^k}$, where N is the number of sums $s = \sum_{a \in S} a$ such that $|s - \mathbb{E}(X)| \geq n\sqrt{k}$. Therefore, we get that the number of sums $s = \sum_{a \in S} a$ such that $|s - \mathbb{E}(X)| < n\sqrt{k}$ is at least $\frac{3}{4} 2^k$. But each of these sums is different and concentrated in an interval of size $2n\sqrt{k}$, so we get $\frac{3}{4} 2^k \leq 2n\sqrt{k}$. Solving this inequality, we get $k \leq \log_2 n + \frac{1}{2} \log_2 \log_2 n + O(1)$. □

5.2 Number of prime divisors

Let $\nu(n)$ denote the number of different prime divisors of n. Clearly, the function ν fluctuates: we have $\nu(n) = 1$ for infinitely many n, but $\nu(n)$ can also be roughly $\log n / \log \log n$ for infinitely many n. However, somewhat surprisingly, it turns out that for most values n, $\nu(n)$ behaves in an ordered fashion, and its value is very close to $\log \log n$.

Theorem 5.4. *(Hardy-Ramanujan 1920) Let $w(n)$ be an arbitrary function tending to infinity. The number of positive integers $x \leq n$ such that*

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

is $o(n)$.

Proof. The proof we present here is due to Turán. Pick $x \in [n]$ randomly from the uniform distribution. For every prime p, define the random variable

$$x_p = \begin{cases} 1 & \text{if } p|x \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, we have $\nu(x) = \sum_{p \leq n} x_p$, where the sum iterates through all primes less than n. For technical reasons, which will become clear later, let us estimate $\nu(x)$ with the random variable $Z = \sum_{p<N} x_p$, where $N = n^{1/3}$. Note that any number at most n can have at most 2 prime divisors which are at least $n^{1/3}$, so $|Z - \nu(x)| \leq 2$. To prove the theorem, it is enough to show that

$$\mathbb{P}(|Z - \log\log n| > \omega(n)\sqrt{\log\log n}) \to 0.$$

Similarly as before, we would like to use Chebyshev's inequality to show that Z is concentrated around its expected value. In order to do so, we need to calculate the expectation and the variance.

We have

$$\mathbb{E}(x_p) = \frac{\lfloor n/p \rfloor}{n} = \frac{1}{p} + O\left(\frac{1}{n}\right),$$

hence

$$\mathbb{E}(Z) = \sum_{p<N} \left[\frac{1}{p} + O\left(\frac{1}{n}\right)\right] = \log\log N + O(1).$$

The second equality is a well known result in number theory, which we shall use without proof. Now let us calculate the variance.

$$\mathrm{Var}(Z) = \sum_{p<N} \mathrm{Var}(x_p) + 2\sum_{p<q<N} \mathrm{Cov}(x_p, x_q).$$

Here,

$$\mathrm{Var}(x_p) = \mathbb{E}(x_p^2) - (\mathbb{E}(x_p))^2 = \frac{1}{p} - \frac{1}{p^2} + O\left(\frac{1}{n}\right),$$

and if $p \neq q$, then

$$\begin{aligned} \mathrm{Cov}(x_p, x_q) =& \mathbb{E}(x_p x_q) - \mathbb{E}(x_p)\mathbb{E}(x_q) \\ =& \frac{\lfloor n/pq \rfloor}{n} - \left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right)\left(\frac{1}{q} + O\left(\frac{1}{n}\right)\right) \\ =& O\left(\frac{1}{n}\right) - O\left(\frac{1}{pn} + \frac{1}{qn} + \frac{1}{n^2}\right) \\ =& O\left(\frac{1}{n}\right). \end{aligned}$$

Therefore,

$$2\sum_{p<q<N} \mathrm{Cov}(x_p, x_q) = O\left(\frac{N^2}{n}\right) = o(1).$$

Hence,

$$\mathrm{Var}(Z) = o(1) + \sum_{p<N} \mathrm{Var}(x_p) \leq o(1) + \sum_{p<N} \left(\frac{1}{p} + O\left(\frac{1}{n}\right)\right) = o(1) + \sum_{p<N} \frac{1}{p} = \log\log N + O(1).$$

Applying Chebyshev's inequality, we get

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > \omega(n)\sqrt{\log\log n}) \leq \frac{\mathrm{Var}(Z)}{\omega(n)^2 \log\log n} = O\left(\frac{1}{\omega(n)^2}\right) = o(1).$$

This finishes the proof. □

5.3 Threshold functions

The Erdős-Rényi random graph $G(n, p)$ is the graph on n vertices in which each pair of vertices is joined by an edge independently with probability p. Observe that for a fixed graph G the probability that $G(n, p) \equiv G$ is $p^{|E(G)|}(1-p)^{\binom{n}{2}-|E(G)|}$. If $p = 1/2$, $G(n, p)$ is the uniform distribution on all graphs with n vertices.

Properties of the Erdős-Rényi random graph model are extensively studied. The problem we are going to consider here is to understand the probability that $G(n, p)$ contains a clique of size 4 (which we denote by K_4). Let $f(p)$ denote this probability. Clearly, $f(0) = 0$, $f(1) = 1$ and f is monotone increasing. But how does this function look like between 0 and 1? It turns out that there is some p_0 such that $f(p) < \epsilon$ if p is slightly below p_0, and $f(p) > 1 - \epsilon$ if p is slightly above p_0. We call such a p_0 a threshold. Let us state this more precisely.

Definition 5.5. $p_0 = p_0(n)$ is a (weak) threshold for containing K_4 if for every $\varepsilon > 0$ there exists $C > 1$ such that

1. for all $p < \frac{p_0}{C}$, $f(p) < \varepsilon$,
2. for all $p > Cp_0$, $f(p) > 1 - \varepsilon$.

In what follows, we calculate a weak threshold for containing K_4. Note that the threshold is not unique; if p_0 is a threshold then so is every constant multiple of p_0.

Theorem 5.6. *$p_0 = n^{-2/3}$ is a weak threshold for containing K_4.*

Proof. Let X be the number of K_4's in $G(n, p)$, and for a given 4-element set S of vertices, let X_S be the indicator random variable of the event that S spans a clique; namely, $X_S = 1$ if S spans a clique, and $X_S = 0$ otherwise. Then $X = \sum_S X_S$, where the sum ranges through all 4-element subsets of the vertex set. Clearly, we have $\mathbb{E}(X_S) = p^6$, and so $\mathbb{E}(X) = \binom{n}{4}p^6$. Note that $G(n, p)$ contains a K_4 if and only if $X \geq 1$.

Now let us verify that Items 1 and 2 in the definition of a threshold are satisfied. First, let us show that Item 1 holds, so let $p < \frac{p_0}{C}$ with some constant C. By Markov's inequality, we have

$$f(p) = \mathbb{P}(X \geq 1) \leq \mathbb{E}(X) = \binom{n}{4}p^6 < n^4p^6 < C^{-6}.$$

Therefore, by choosing C sufficiently large, we have $f(p) < \epsilon$, satisfying Item 1.

Now let $p > Cp_0$. We can write $f(p) = 1 - \mathbb{P}(X = 0)$. Now by Theorem 5.2, we bound the probability $\mathbb{P}(X = 0)$ as follows:

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2}.$$

Now our task is to calculate $\mathrm{Var}(X)$. We have

$$\mathrm{Var}(X) = \sum_{S,T} \mathrm{Cov}(X_S, X_T).$$

(Here, the sum is over all ordered pairs (S,T), including $S = T$.)

Clearly, $\mathrm{Cov}(X_S, X_T)$ depends only on the size of the intersection of S and T. In particular,

$$\mathrm{Cov}(X_S, X_T) = \begin{cases} p^6 - p^{12} & \text{if } S = T \\ p^9 - p^{12} & \text{if } |S \cap T| = 3 \\ p^{11} - p^{12} & \text{if } |S \cap T| = 2 \\ 0 & \text{ otherwise} \end{cases}$$

This is because $\mathrm{Cov}(X_S, X_T) = \mathbb{E}(X_S X_T) - \mathbb{E}(X_S)\mathbb{E}(X_T) = \mathbb{E}(X_S X_T) - p^{12}$, and $\mathbb{E}(X_S X_T) = p^{g(S,T)}$, where $g(S,T)$ is the number of edges contained entirely in S or entirely in T.

Also, for $k \in \{0, 1, 2, 3, 4\}$, the number of pairs (S,T) such that $|S \cap T| = k$ is less than n^{8-k} as $S \cup T$ occupies $8 - k$ vertices. Therefore, we can write

$$\mathrm{Var}(X) < p^6 n^4 + p^9 n^5 + p^{11} n^6.$$

Also, $\mathbb{E}(X) = \binom{n}{4} p^6 = \left(\frac{1}{24} - o(1)\right) n^4 p^6 = \Omega(n^4 p^6)$. So we get

$$\begin{aligned} P(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}(X))^2} < \frac{p^6 n^4 + p^9 n^5 + p^{11} n^6}{\Omega(n^8 p^{12})} &= O(p^{-6} n^{-4} + p^{-3} n^{-3} + p^{-1} n^{-2}) \\ &= O(C^{-6} + C^{-3} n^{-1} + C^{-1} n^{-4/3}). \end{aligned}$$

Therefore, if C is sufficiently large, then $P(X = 0) \leq \epsilon$, so Item 2 is also satisfied. In conclusion, p_0 is truly a threshold. □

5.4 Clique number of random graphs

As a last application of the second moment method, we will consider the clique number $\omega(G)$ of $G(n, 1/2)$. As a reminder, $\omega(G)$ denotes the size of the largest clique of a graph G. As we have seen in previous sections, $G(n, 1/2)$ is a good *Ramsey graph*, meaning it typically has cliques and independent sets of at most logarithmic size. Therefore, this leads to the question: what is the clique number of $G(n, 1/2)$, i.e. what is the distribution of this random parameter? Surprisingly, it turns out that the clique number is concentrated on two consecutive integers.

Let $G \equiv G(n, 1/2)$. Let k be a positive integer, and let X_k be the number of cliques of size k in G. Then $X_k = \sum_S X_S$, where S iterates over all k-element subsets of the vertices, and X_S is the indicator random variable of the event that S spans a clique. Therefore, we have

$$\mu_k := \mathbb{E}(X_k) = \binom{n}{k}\left(\frac{1}{2}\right)^{\binom{k}{2}}.$$

Note that μ_k drops below 1 when $k \sim 2\log_2 n$. Also,

$$\frac{\mu_k}{\mu_{k+1}} = 2^k \frac{k+1}{n-k},$$

so $\mu_k/\mu_{k+1} = n^{1+o(1)}$ if $k \sim 2\log_2 n$. Let k_0 be the minimal $k \sim 2\log_2 n$ such that $\mu_k \leq n^{1/2}$ (say). Then $\mu_{k_0+1} \leq n^{-1+o(1)} \cdot \mu_{k_0} \leq n^{-1/2+o(1)}$, and therefore, by Markov's inequality, G has no clique of size k_0+1 with probability tending to 1. Also, $\mu_{k_0-1} > n^{1/2}$ by our choice of k_0. We show that this implies that G contains a clique of size k_0-1 with high probability.

Claim 5.7. Let $k = k(n)$ satisfy $k \sim 2\log_2 n$ and $\mu_k \to \infty$. Then $\mathbb{P}(X_k = 0) \to 0$.

Proof. Using Theorem 5.2, we have

$$\mathbb{P}(X_k = 0) \leq \frac{\mathrm{Var}(X_k)}{(\mathbb{E}(X_k))^2}.$$

Here,

$$\mathrm{Var}(X_k) = \sum_{S,T} \mathrm{Cov}(X_S, X_T) < \sum_{|S\cap T|\geq 2} \mathbb{E}(X_S X_T).$$

This is because if $|S \cap T| \leq 1$ then X_S and X_T are independent, so $\mathrm{Cov}(X_S, X_T) = 0$.

Note that if $r = |S \cap T|$, then $\mathbb{E}(X_S X_T) = 2^{\binom{r}{2}-2\binom{k}{2}}$, and the number of such pairs (S,T) is $\binom{n}{k}\binom{k}{r}\binom{n-k}{k-r}$. Therefore,

$$\mathrm{Var}(X_k) < \sum_{r=2}^{k} 2^{\binom{r}{2}-2\binom{k}{2}} \binom{n}{k}\binom{k}{r}\binom{n-k}{k-r} =: g(k).$$

Let us estimate $g(k)/\mu_k^2$. We have

$$\frac{g(k)}{\mu_k^2} = \sum_{r=2}^{k} 2^{\binom{r}{2}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} = \frac{1}{\mu_k} + \sum_{r=2}^{k-1} 2^{\binom{r}{2}} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}}.$$

Let us consider the term $r = 2$, which equals:

$$2 \cdot \frac{\binom{k}{2}\binom{n-k}{k-2}}{\binom{n}{k}} = \Theta\left(\frac{k^4}{n^2}\right) = n^{-2+o(1)}.$$

On the other extreme, we have the term $r = k-1$, which equals

$$2^{\binom{k-1}{2}} \frac{k(n-k)}{\binom{n}{k}} = 2^{-k+1} \frac{k(n-k)}{\mu_k} = \frac{n^{-1+o(1)}}{\mu_k} < n^{-1+o(1)},$$

where the second equality holds noting that $2^{-k+1} \sim n^{-2}$ and $k(n-k) \sim 2n \log_2 n$. With somewhat more complicated calculations, one can show that the other terms are also of order $n^{-1+o(1)}$, so

$$\frac{g(k)}{\mu_k^2} = \frac{1}{\mu_k} + n^{-1+o(1)}.$$

Therefore, as $\mu_k \to \infty$ and $\mathrm{Var}(X_k) \leq g(k)$, we get $\mathbb{P}(X_k = 0) \leq g(k)/\mu_k^2 \to 0$. □

From this claim and the explanation before the claim, one can conclude the following.

Theorem 5.8. *There exists $k_0 = k_0(n)$ such that with high probability, $\omega(G) \in \{k_0 - 1, k_0\}$.*

6 Concentration

In many applications of the probabilistic method, our task is to estimate the distribution of a variable X that is the sum of n independent indicator random variables. In this case, Chebyshev's inequality tells us that a positive proportion of the mass is concentrated in an interval of radius $\approx \sqrt{n}$ around the mean. We show that outside of this interval, the mass starts to decay exponentially fast.

Theorem 6.1. *(Chernoff's inequality) Let $X_1, \dots, X_n$ be independent random variables taking the values -1 and +1 with probability $1/2$. Let $X = \sum_{i=1}^n X_i$. Then for every $t > 0$,*

$$\mathbb{P}(X \leq -t) = \mathbb{P}(X \geq t) \leq e^{-t^2/2n}.$$

Proof. By symmetry, it is enough to prove the statement for $\mathbb{P}(X \geq t)$. By Markov's inequality, for every $c > 0$, we have

$$\mathbb{P}(X \geq t) = \mathbb{P}(e^{cX} \geq e^{ct}) \leq \frac{\mathbb{E}(e^{cX})}{e^{ct}}.$$

As $X_1, ..., X_n$ are independent and identically distributed, we can write

$$\mathbb{E}(e^{cX}) = \mathbb{E}\left(\prod_{i=1}^n e^{cX_i}\right) = \prod_{i=1}^n \mathbb{E}(e^{cX_i}) = (\mathbb{E}(e^{cX_1}))^n = \left(\frac{e^{-c} + e^c}{2}\right)^n.$$

(Recall that if X and Y are independent random variables then $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$; so the second equality above follows from the independence of $X_1, \dots, X_n$.)

By considering the Taylor expansion of e^x, it is easy to show that $\frac{e^{-c}+e^c}{2} \leq \exp(c^2/2)$. Therefore, we get

$$\mathbb{P}(X \geq t) \leq \exp\left(\frac{c^2 n}{2} - ct\right).$$

Choosing $c = \frac{t}{n}$ (to minimize the quadratic function $\frac{c^2 n}{2} - ct$), we get the desired inequality. □

Theorem 6.1 is about random variables taking values -1 and 1, but we can easily transform this statement to refer to indicator random variables.

Corollary 6.2. *Let $X_1, \dots, X_n$ be independent random variables taking the values 0 and 1 with probability $1/2$. Let $X = \sum_{i=1}^n X_i$. Then for every $t > 0$,*

$$\mathbb{P}\Big(X \le \frac{n}{2} - t\Big) = \mathbb{P}\Big(X \ge \frac{n}{2} + t\Big) \le e^{-2t^2/n}.$$

Proof. Put $Y_i := 2X_i - 1$. Then Y_i takes the values -1 and 1, with probability 1/2 each. Now put $Y := \sum_{i=1}^n Y_i = 2X - n = 2(X - \frac{n}{2})$. Then $X \ge \frac{n}{2} + t$ iff $Y \ge 2t$ and the same for the other direction. So the corollary follows by applying Theorem 6.1 to Y with parameter $2t$. □

6.1 Random tournaments

Let T be a tournament (i.e. orientation of a complete graph) on the vertex set $[n]$. Given a permutation $\pi : [n] \to [n]$, say that an edge $i \to j$ is *consistent* if $\pi(i) < \pi(j)$, otherwise *inconsistent.* We think of π as the ordering $\pi(1), \dots, \pi(n)$ of $V(T) = [n]$. We will denote by π' the "reverse ordering" $\pi(n), \dots, \pi(1)$.

Observe that if $i \to j$ is consistent for π then it is inconsistent for π' and vice versa, so for any tournament there exist a permutation such that at least half of the edges are consistent. We show that for a random tournament, one cannot hope for a much larger number of consistent edges.

Theorem 6.3. *Let T be a random tournament on n vertices. Then with high probability, for every permutation π the number of consistent edges is at most*

$$\frac{1}{2}\binom{n}{2} + O(n^{3/2}\sqrt{\log n}).$$

Proof. Fix a permutation π. Let X_{ij} denote the indicator random variable that the edge ij is consistent in T with π. Then $\{X_{ij} : 1 \le i < j \le n\}$ is a system of $\binom{n}{2}$ independent indicator random variables. Let $X = \sum_{1 \le i < j \le n} X_{ij}$ be the number of edges consistent with π. Then by Corollary 6.2, we have

$$\mathbb{P}\bigg(X > \frac{1}{2}\binom{n}{2} + n^{3/2}\sqrt{\log n}\bigg) \le e^{-2n^3 \log n/\binom{n}{2}} \le n^{-n}.$$

Say that a permutation π is *bad* if the number of consistent edges is more than $\frac{1}{2}\binom{n}{2} + n^{3/2}\sqrt{\log n}$. Then $\mathbb{P}(\pi \text{ is bad}) \le n^{-n}$ by the previous inequality. There are $n!$ permutations, so by the union bound the probability that there exists a bad permutation is at most $n!n^{-n} \to 0$. This finishes the proof. □

Next, we show that the previous theorem remains true with the $\sqrt{\log n}$ factor removed.

Theorem 6.4. *(de la Vega) Let T be a random tournament on n vertices. Then with high probability, for every permutation the number of consistent edges is at most $\frac{1}{2}\binom{n}{2} + O(n^{3/2})$.*

Proof. For simplicity, assume that $n = 2^k$, and let π be a permutation. Then π defines a binary tree-like partition of $[n]$: for $i = 0, \dots, k$, let $V_{i,1}, \dots, V_{i,2^i}$ be a partition of $[n]$ into 2^i sets such that $V_{i,j} = \{\pi(a) : 2^{k-i}(j-1) + 1 \le a \le 2^{k-i}j\}$ for each $1 \le j \le 2^i$. (That is, $V_{i,j}$ is an interval of length

2^{k-i} with respect to π.) For example, $V_{0,1} = V$, $V_{1,1} = \{\pi(a) : 1 \leq a \leq \frac{n}{2}\}$ is the "first half" of the vertices, and $V_{1,2} = \{\pi(a) : \frac{n}{2} + 1 \leq a \leq n\}$ is the "second half" of the vertices. In the other extreme, $V_{k,j}$ is a singleton for every $1 \leq j \leq 2^k$.

Another way of thinking about this construction is to think of the corresponding complete binary tree, where $V_{i,1}, \ldots, V_{i,2^i}$ are the vertices at level i of the tree, so in particular $V_{0,1} = V$ is the root, and the children of each $V_{i,j}$ ($0 \leq i \leq k-1$ and $1 \leq j \leq 2^i$) are the two sets in the $(i+1)$th level which are contained in $V_{i,j}$; it is easy to check that these are $V_{i+1,2j-1}$ and $V_{i+1,2j}$.

Also, for $i = 1, \ldots, k$ and $j = 1, \ldots, 2^{i-1}$, let $E_{i,j}$ be the set of edges between $V_{i,2j-1}$ and $V_{i,2j}$. Note that the sets $E_{i,j}$ ($1 \leq i \leq k$ and $1 \leq j \leq 2^{i-1}$) partition the edge set of T, and $|E_{i,j}| = 2^{2(k-i)}$. Let $E_i = \bigcup_{j=1}^{2^{i-1}} E_{i,j}$, then $|E_i| = 2^{2k-i-1} = n^2 2^{-i-1}$. (In the corresponding tree, E_i is the set of edges between pairs of siblings in the ith level.)

Let X_i be the number of consistent edges in E_i. Then by Corollary 6.2, we have

$$\mathbb{P}\left(X_i \geq \frac{1}{2}|E_i| + t_i\right) \leq \exp\left(-\frac{2t_i^2}{|E_i|}\right) = \exp\left(-\frac{2^{i+2}t_i^2}{n^2}\right).$$

Choosing $t_i = n^{3/2}2^{-i/2}\sqrt{i}$, the right-hand-side equals e^{-4ni}. Note that the number of different partitions of $[n]$ into $V_{i,1}, \ldots, V_{i,2^i}$ is at most 2^{ni} (since for each of the n vertices $v \in V(T)$, we have at most 2^i choices for the set $V_{i,j}$ to which v is assigned). Therefore, the probability that there exists a permutation π for which $X_i > \frac{1}{2}|E_i| + t_i$ is at most $2^{ni}e^{-4ni} < e^{-2ni}$.

Furthermore, as $\sum_{i=1}^{k} e^{-2ni} < ke^{-n} \to 0$, with high probability for *every* permutation and *every* $1 \leq i \leq k$, we have $X_i \leq \frac{1}{2}|E_i| + t_i$. But if this happens then the number of consistent edges is

$$X = \sum_{i=1}^{k} X_i \leq \sum_{i=1}^{k} \left[\frac{1}{2}|E_i| + n^{3/2}2^{-i/2}\sqrt{i}\right] = \frac{1}{2}\binom{n}{2} + n^{3/2}\sum_{i=1}^{k} 2^{-i/2}\sqrt{i}.$$

The sum on the right-hand-side converges, so we get $X \leq \frac{1}{2}\binom{n}{2} + O(n^{3/2})$ with high probability. □

7 Lovász Local Lemma

The basic probabilistic method can be summarised as follows. There is a collection of "bad" events that we want to avoid. The strategy we employ is to use the union bound to deduce that the probability that one of these bad events occurs is small, in particular smaller than 1, which tells us there is an outcome which avoids all these events. The bound we get is simply the sum of probabilities of bad events and unfortunately, very often the number of bad events is too high compared with the probabilities of bad events happening. However, imagine that all of the bad events are independent. Then, provided none of them occurs with probability 1, we know we can avoid all of them, despite this probability being potentially tiny. (Indeed, if events $B_1, \dots, B_m$ are independent, the probability that none of them occurs is $\prod_{i=1}^{m}(1 - \mathbb{P}(B_i))$.) What happens if we know that *most* of our bad events are independent but not all? Can we do something better than the standard union bound argument? The answer is yes in a certain sense and is precisely the content of the Lovász local lemma.

Theorem 7.1. *Let $B_1, \dots, B_m$ be a collection of events in a probability space $(\Omega, \mathbb{P})$. Suppose that*

1. *$\mathbb{P}(B_i) \leq p < 1$,*
2. *Every event B_i is mutually independent from all but at most d other events B_j and*
3. *$ep(d + 1) \leq 1$.*

then we have $\mathbb{P}(\cap_{i=1}^{m} \bar{B}_i) > 0$.

We think of the events B_i as bad events; the first condition requires a bound on the probability of any single bad event happening, while the second condition requires that every bad event is mutually independent from all but a small collection of other bad events. The last condition quantifies how good the bounds should be in first two conditions.

This lemma is surprisingly powerful and often allows one to prove much stronger results than with the standard probabilistic method. A good indication that the local lemma might be useful is if the standard probabilistic method fails due to a union bound being too weak, and if it seems that there is a lot of independence between events; i.e., if the bad events are, in a sense, local (hence the name of the lemma). The main conceptual difficulty when using the local lemma, as with the standard probabilistic method, is to identify the correct probability space and events. Before proving it let us give some examples of how to use it.

7.1 2-colorability of hypergraphs

The first example we have already encountered in one of the first lectures. Recall that a hypergraph is said to be 2-colorable if there exists a red-blue coloring of its vertices such that no edge is monochromatic. We have seen in Section 1.3 that any k-uniform hypergraph with at most 2^{k-1} edges is 2-colorable and that this is not too far from best possible. The local lemma allows us to improve this quite substantially, by replacing the condition of having not-too-many edges with the (weaker) condition that each edge intersects not-too-many other edges.

Theorem 7.2. *Let $k \geq 2$. If H is a k-uniform hypergraph and any edge $e \in E(H)$ intersects at most $d \leq \frac{2^{k-1}}{e} - 1$ other edges then H is 2-colorable.*

Proof. As usual with probabilistic arguments, we begin with defining the probability space. Here we are going to color every vertex of H red or blue uniformly at random and independently of all other vertices. Now let us proceed to identify our "bad" events. For any $e \in E(H)$, let B_e denote the event that edge e is monochromatic. If none of the events B_e occurs, then in the resulting coloring there are no monochromatic edges, so H is 2-colorable. Therefore, it is enough to show that $\mathbb{P}(\cap_{e \in E(H)} \bar{B}_e) > 0$.

Observe that $\mathbb{P}(B_e) = \frac{1}{2^{k-1}} =: p$, since there are exactly two ways in which e can be colored monochromatically (out of a total of 2^k colorings). Note that an event B_e only depends on how we colored vertices belonging to e. In particular, it is independent of how we color vertices outside of e. This means that B_e is mutually independent of the collection of all events indexed by an edge disjoint from e. (Intuitively, the only case where an event $B_{e'}$ (for some $e' \neq e$) can tell us something about the outcome of the event B_e, is if it tells us something about the color of one of the vertices of e, which can only happen if e and e' share a vertex.)

Since B_e intersects at most d other edges, we have shown that B_e is independent of all but d other events $B_{e'}$. Note that our assumed bound on d precisely implies $ep(d+1) \leq 1$ which means that the Lovász local lemma applies and completes the proof. □

Note that if we used the union bound in the above proof, we would only get the bound $\mathbb{P}(\cup_{e \in E(H)} B_e) \leq \frac{|E(H)|}{2^{k-1}}$, and so we would need to assume that the number of edges is not-too-large (at most 2^{k-1}) in order to get a meaningful result (i.e. a bound of less than 1 on the probability). By using the local lemma, however, we don't need to pose any restriction on the number of edges, but only need to restrict it *locally*, in a sense.

7.2 Finding directed cycles of certain length

In the above example the probability space and the events were fairly straightforward given what we wanted to prove. The next one shows that they can be quite clever as well.

Theorem 7.3. *Let k be a positive integer. Let $D = D(V, E)$ be a directed graph with maximum in-degree Δ^- and minimum out-degree δ^+ such that*

$$e((\delta^+ + 1)\Delta^- + 1)\left(1 - \frac{1}{k}\right)^{\delta^+} \leq 1.$$

Then D contains a directed cycle of length divisible by k.

Proof. First let us remove edges until every vertex in D has *exactly* δ^+ out-neighbours. Note that the new digraph has unchanged δ^+, and Δ^- might only have gone down so the condition remains satisfied. Also our new digraph is a subgraph of the original one, so any cycle inside it is also present in the original digraph. Let us hence assume that every vertex in D has exactly δ^+ out-neighbours. This is only going to be important for certain technical reasons later.

Let us begin with the probability space. We will color every vertex in one of k colors uniformly at random and independently between vertices. Let us denote by $c : V \to [k]$ the outcome coloring. Let us now proceed to define the "bad" events. For any vertex $v \in V$, we define B_v to be the event that there is no $u \in V$ such that $vu \in E$ and $c(u) \equiv c(v) + 1 \pmod{k}$.

Now let us explain why if none of the events B_v happens then we find our desired cycle. For any $v \in V$, let us denote by $u(v)$ its out-neighbour with $c(u(v)) \equiv c(v) + 1 \pmod k$ whose existence is guaranteed by the fact that B_v did not occur (if there is more than one choice for $u(v)$ we pick one of them arbitrarily). Let $n = |V|$ and let us fix some $v \in V$. If we consider $v, u(v), \ldots, u^n(v)$ (where $u^i(v) = u(u(\ldots u(v)\ldots))$ is a composition taken i times), we know that two vertices need to repeat since there are only n in total. This means that $u^i(v) = u^{i+j}(v) =: v'$ for some $i \geq 0$ and $j \geq 1$. If we choose such v' with minimum possible value of $j \geq 1$, then we get that $v', u(v') \ldots, u^j(v')$ make a directed cycle in D (minimality of j ensures it is a simple cycle and not a closed walk which intersects itself). In addition, by our definition of u we know that $c(v') = c(u^j(v')) \equiv c(v') + j \pmod k$, which in particular means that $k \mid j$, or in other words that the cycle length is divisible by k as desired.

(Another way of saying the above argument is to consider the digraph D' obtained from D by only keeping the edges $vu \in E$ with $c(u) = c(v) + 1$, and to note that if none of the events B_v happened then $\delta^+(D') \geq 1$. This implies that D' contains a directed cycle (the argument for showing that every digraph with $\delta^+ \geq 1$ contains a directed cycle is essentially the same as given in the previous paragraph), and the length of this cycle must be divisible by k by the definition of D'.)

What remains to be shown is that $\mathbb{P}(\cap_{v\in V} \bar{B}_v) > 0$, which we prove by using the local lemma. First note that $\mathbb{P}(B_v) \leq \left(\frac{k-1}{k}\right)^{\delta^+} = \left(1 - \frac{1}{k}\right)^{\delta^+}$ since v has at least δ^+ out-neighbours and regardless of the value of $c(v)$ there is exactly one choice for any of its out-neighbours which would violate B_v. In terms of dependencies, note that B_v only depends on the assignment of colors to v and its out-neighbours, namely $\{v\} \cup N^+(v)$. This means that B_v is mutually independent of all events B_u for which $\{u\} \cup N^+(u)$ is disjoint from $\{v\} \cup N^+(v)$. If this fails to happen then u needs to be equal to a vertex in $\{v\} \cup N^+(v)$ or be an in-neighbour of such a vertex. Of course, we don't count $u = v$. For each $x \in N^+(v)$, there are at most Δ^- vertices $u \neq v$ such that $u = x$ or $u \in N^-(x)$ (because there are at most $\Delta^- - 1$ in-neighbours of x other than v). So the total number of vertices $u \neq v$ such that $\{u\} \cup N^+(u)$ intersects $\{v\} \cup N^+(v)$ is at most $(\delta^+ + 1)\Delta^-$ (here we use the assumption that the out-degree of v is exactly δ^+, not larger). We conclude that B_v is mutually independent of all but at most $(\delta^+ + 1)\Delta^-$ other events. This means that our condition is precisely what the local lemma requires and this completes the proof. □

7.3 Proof of the Lovász local lemma

Let us now prove the local lemma. We will prove it with condition 3 replaced with $4pd \leq 1$. This makes the proof less technical and by being a bit more careful with the inequalities one can obtain the claimed condition (good exercise!).

Proof of Theorem 7.1. Let us first prove by induction on k that $\mathbb{P}(B_{i_1} \mid \bar{B}_{i_2} \cap \ldots \cap \bar{B}_{i_k}) \leq 2p$ for any $k \leq m$ and distinct $B_{i_1}, \ldots, B_{i_k}$. The base case $k = 1$ follows from condition 1 with an even stronger bound $\mathbb{P}(B_{i_1}) \leq p$. Throughout the proof, we will use the convention that an empty intersection of events (such as $\bar{B}_{i_2} \cap \ldots \cap \bar{B}_{i_k}$ if $k = 1$) equals the whole probability space, so conditioning on it doesn't change the probability of events.

Let us now assume this claim holds for $k - 1$ and any choice of $k - 1$ of our bad events.

It is enough to show $\mathbb{P}(B_1 \mid \bar{B}_2 \cap \ldots \cap \bar{B}_k) \leq 2p$, since we may reorder the events as we like. For the same reason, we may assume that B_1 is mutually independent from events $B_{\ell+1}, \ldots, B_k$ (for some

$1 \leq \ell \leq k$). By condition 2 we know $\ell - 1 \leq d$. Now by definition of conditional probability we have

$$\begin{aligned}\mathbb{P}(B_1 \mid \bar{B}_2 \cap \ldots \cap \bar{B}_k) = \frac{\mathbb{P}(B_1 \cap \bar{B}_2 \cap \ldots \cap \bar{B}_k)}{\mathbb{P}(\bar{B}_2 \cap \ldots \cap \bar{B}_k)} &= \frac{\mathbb{P}(B_1 \cap \bar{B}_2 \cap \ldots \cap \bar{B}_k)}{\mathbb{P}(\bar{B}_{\ell+1} \cap \ldots \cap \bar{B}_k)} \cdot \frac{\mathbb{P}(\bar{B}_{\ell+1} \cap \ldots \cap \bar{B}_k)}{\mathbb{P}(\bar{B}_2 \cap \ldots \cap \bar{B}_k)} \\ &= \frac{\mathbb{P}(B_1 \cap \bar{B}_2 \cap \ldots \cap \bar{B}_\ell \mid \bar{B}_{\ell+1}, \ldots, \bar{B}_k)}{\mathbb{P}(\bar{B}_2 \cap \ldots \cap \bar{B}_\ell \mid \bar{B}_{\ell+1}, \ldots, \bar{B}_k)}\end{aligned}$$

Consider the last expression above. The numerator of this expression is

$$\mathbb{P}(B_1 \cap \bar{B}_2 \ldots \cap \bar{B}_\ell \mid \bar{B}_{\ell+1} \cap \ldots \cap \bar{B}_k) \leq \mathbb{P}(B_1 \mid \bar{B}_{\ell+1} \cap \ldots \cap \bar{B}_k) = \mathbb{P}(B_1) \leq p,$$

where the equality is because B_1 is mutually independent from $B_{\ell+1}, \ldots, B_k$. On the other hand, the denominator is

$$\begin{aligned}\mathbb{P}(\bar{B}_2 \cap \ldots \cap \bar{B}_\ell \mid \bar{B}_{\ell+1} \cap \ldots \cap \bar{B}_k) &= 1 - \mathbb{P}(B_2 \cup \ldots \cup B_\ell \mid \bar{B}_{\ell+1} \cap \ldots \cap \bar{B}_k) \\ &\geq 1 - \sum_{i=2}^{\ell} \mathbb{P}(B_i \mid \bar{B}_{\ell+1} \ldots, \bar{B}_k) \\ &\geq 1 - (\ell - 1)2p \geq 1 - 2pd \geq \frac{1}{2}\end{aligned}$$

where we used union bound in the first inequality and induction hypothesis in the second. Combining these inequalities we get $\mathbb{P}(B_1 \mid \bar{B}_2, \ldots, \bar{B}_k) \leq 2p$, completing the proof of the claim. Finally, repeatedly using the claim we obtain the local lemma as follows

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \bar{B}_i\right) = \prod_{i=1}^{n} \mathbb{P}(\bar{B}_i \mid \bar{B}_1, \ldots, \bar{B}_{i-1}) \geq (1 - 2p)^n > 0.$$

□

7.4 Multicolored sets of real numbers

For a final example we will show a certain coloring result on the real line. The really surprising part is that one can use the local lemma to prove a result involving uncountable objects.

Definition 7.4. Given $f : \mathbb{R} \to [k]$, a subset $T \subseteq \mathbb{R}$ is said to be multicolored if $Im(f(T)) = [k]$, i.e. $f \mid_T$ is a surjection.

Theorem 7.5. *Let $m, k \in \mathbb{N}$ and $m > 11k \log k > 0$. Then for any $S \subseteq \mathbb{R}$ of size $|S| = m$ there is an $f : \mathbb{R} \to [k]$ such that for any $x \in \mathbb{R}$, $x + S = \{x + s \mid s \in S\}$ is multicolored w.r.t. f.*

Proof. Instead of proving the statement for all $x \in \mathbb{R}$, we will prove it for every finite subset $X \subseteq \mathbb{R}$ and every $x \in X$. The general result then follows by a standard compactness argument based on Tikhonov's theorem (which is not part of this course). So our goal is to show that for every $S \subseteq \mathbb{R}$ of size $|S| = m$ and every finite $X \subseteq \mathbb{R}$, there is a k-coloring of $T := \cup_{x \in X}(x + S)$ such that $x + S$ is multicolored for every $x \in X$.

Let us color every $x \in T$ uniformly at random in a color from $[k]$, independently. For each $x \in X$, we have a bad event B_x defined to occur if $x + S$ is not multicolored. In particular, we have

$\mathbb{P}(B_x) \leq k(1-1/k)^m$. Turning to the dependencies, note that every B_x only depends on the coloring of points in $x+S$. In particular, B_x is independent of all B_y for which $(x+S)\cap(y+S)=\emptyset$. There are in total at most $m(m-1)$ elements $y \neq x$ for which $(x+S)\cap(y+S)\neq\emptyset$, because every such y is of the form $x+s-s'$ for $s, s' \in S$, $s \neq s'$. So we see that every bad event is independent from all but at most $d := m(m-1)$ other events. Finally,

$$ep(d+1) \leq e \cdot k(1-1/k)^m \cdot m^2 \leq e^{1+\log k - m/k + 2\log m} < 1$$

where the last inequality follows from $m > 11k\log k$. Hence, the local lemma applies and completes the proof. □

Remark. Note that in general one can not apply the local lemma to an infinite set of events (recall that we used induction in the proof!) so the compactness argument is necessary in the above proof (see Alon-Spencer for more details about compactness).

8 Martingales

A martingale is a sequence $X_0, X_1, \ldots, X_n$ of random variables such that for all $0 \leq i < n$, we have

$$\mathbb{E}[X_{i+1} \mid X_i, X_{i-1}, \ldots, X_0] = X_i.$$

For example, assume $Y_1, \ldots, Y_n$ are independent variables with $\mathbb{P}[Y_i = 1] = \mathbb{P}[Y_i = -1] = 1/2$. Define $X_i = \sum_{j=1}^{i} Y_j$. Then we have

$$\mathbb{E}[X_{i+1} \mid X_i, X_{i-1}, \ldots, X_0] = \mathbb{E}[X_i + Y_{i+1} \mid X_i, X_{i-1}, \ldots, X_0] = X_i + \mathbb{E}[Y_{i+1}] = X_i,$$

so the sequence $X_0, X_1, \ldots, X_n$ is a martingale. Recall that Chernoff's inequality tells us that

$$\mathbb{P}[|X_n| \geq \lambda] \leq 2\exp(-\lambda^2/2n).$$

A similar concentration inequality holds for more general martingales. We will not consider martingales explicitly here, but only discuss a corollary of this martingale inequality. For more information on martingales, consult the book 'The Probabilistic Method' by Alon and Spencer.

8.1 Azuma-Hoeffding inequality

The setting is the following. Let $\Omega = \prod_{i=1}^{n} \Omega_i$ be a product probability space, and suppose that $f \colon \Omega \to \mathbb{R}$ is a random variable. For instance, in the above example, $\Omega_i = \{-1, +1\}$ with the uniform distribution, and for $x = (x_1, \ldots, x_n) \in \Omega$, we had $f(x) = \sum_{i=1}^{n} x_i$. One crucial property of this function is that changing one coordinate of x changes the value of f by at most 2. This is a so-called *Lipschitz property.* More generally, we say that f as above is C-Lipschitz if changing any one coordinate affects f by at most C. More formally, for all $x = (x_1, \ldots, x_n) \in \Omega$, every $i \in [n]$ and every $x'_i \in \Omega_i$, we have

$$|f(x) - f(x')| \leq C,$$

where x' is obtained from x by replacing x_i with x'_i (with all other coordinates unchaged).

Theorem 8.1 (Azuma-Hoeffding inequality). *Let $\Omega = \prod_{i=1}^{n} \Omega_i$ be a product probability space, and assume that $f \colon \Omega \to \mathbb{R}$ is C-Lipschitz. Then, for any $\lambda \geq 0$, we have*

$$\mathbb{P}[|f - \mathbb{E}[f]| \geq \lambda] \leq 2\exp\left(-\frac{\lambda^2}{2C^2 n}\right).$$

This concentration inequality is extremely powerful. Here, we are mainly interested in applications for random graphs, so we close this subsection by discussing the random graph model a bit more carefully and giving examples of Lipschitz functions.

Consider the random graph $G \sim G(n, p)$, that is, the random graph on n vertices where every possible edge is included independently with probability p. More formally, we consider the product space $\Omega = \prod_e \Omega_e$, where e runs over all unordered pairs of vertices, and Ω_e has two elements, let's call them 'edge' and 'non-edge', where $\mathbb{P}[\text{edge}] = p$ and $\mathbb{P}[\text{non-edge}] = 1 - p$. Given a graph parameter f, we see that f is C-Lipschitz if for any given graph G, adding or removing one edge changes the value of f by at most C. For example, the size of a maximum matching $\nu(G)$ and the edge-connectivity $\kappa'(G)$ are both 1-Lipschitz.

For any such graph parameter, we can apply the Azuma-Hoeffding inequality. Note that the parameter n in the Azuma-Hoeffding inequality is in this case $\binom{n}{2}$, since we consider the product space over $\binom{n}{2}$ potential edges. This has the drawback that we need a large λ to obtain a good probability bound; oftentimes the result would be useless.

However, we can represent $G(n,p)$ a bit differently. For concreteness, let the vertex set be $[n]$. For $i \in [n]$, let Ω_i be the set of all possible outcomes for the set of edges between i and the vertices $1, \dots, i-1$. (Note that the probability of one such specific outcome with $0 \leq k \leq i-1$ edges is $p^k(1-p)^{i-1-k}$.) We can now write $\Omega = \prod_{i=1}^n \Omega_i$ to represent $G(n,p)$, which is still a product space since the Ω_i are determined by disjoint sets of edges.

Given a graph parameter f, we see that f is C-Lipschitz (with respect to this "vertex-exposure" representation) if for any given graph G, adding or removing edges at any fixed vertex changes the value of f by at most C. For example, the independence number $\alpha(G)$ and the chromatic number $\chi(G)$ are both 1-Lipschitz.

To distinguish the two representations, we say that a random variable Y defined for $G(n,p)$ is C-Lipschitz with respect to edge exposure, or with respect to vertex exposure.

8.2 Chromatic number of random graphs

The chromatic number is a very complicated graph parameter. Nevertheless, the Azuma-Hoeffding inequality immediately gives us that the chromatic number of a random graph is concentrated around its mean.

Corollary 8.2 (Shamir-Spencer, 1987)**.** *For the random graph $G \sim G(n,p)$, we have for any $\lambda > 0$ that $\mathbb{P}[|\chi(G) - \mathbb{E}[\chi(G)]| \geq \lambda] \leq 2\exp(-\lambda^2/2n)$.*

Proof. This follows from Theorem 8.1, as the chromatic number is 1-Lipschitz with respect to vertex exposure. □

This means that the chromatic number of the random graph $G(n, \frac{1}{2})$ is concentrated in an interval of length $\mathcal{O}(\sqrt{n})$ around its expectation. However, the proof gives no clue as to what the expectation is.

We will now show that for sparse random graphs, we have an even stronger concentration of the chromatic number, namely it will with high probability be one of 4(!) consecutive values (but again, we don't know what these values are).

Theorem 8.3. *Fix $\alpha > 5/6$ and consider $G \sim G(n,p)$, where $p = p(n) = n^{-\alpha}$. Then there exists $u := u(n,p)$ such that with probability $1 - o(1)$, we have $u \leq \chi(G) \leq u + 3$.*

Proof. Let $\epsilon > 0$ be given arbitrarily small. Define $u = u(n,p,\epsilon)$ as the smallest integer for which $\mathbb{P}[\chi(G) \leq u] > \epsilon$. This choice ensures that $\mathbb{P}[\chi(G) < u] < \epsilon$. It remains to show that $\mathbb{P}[\chi(G) > u + 3] < 2\epsilon$, as then we have $\mathbb{P}[u \leq \chi(G) \leq u + 3] \geq 1 - 3\epsilon$.

Now comes the crucial definition: define the random variable Z to be the smallest size of a set of vertices X with the property that $\chi(G - X) \leq u$.

We will prove two claims. First, we claim there exists a constant $C = C(\epsilon) > 0$ such that with probability at least $1 - \epsilon$, we have $Z \le C\sqrt{n}$.

Secondly, we claim that with probability at least $1 - \epsilon$, for every set W of vertices of size $\le C\sqrt{n}$, we have $\chi(G[W]) \le 3$.

Assume that the two claims hold true. Then, with probability at least $1 - 2\epsilon$, there exists a set X of vertices of size at most $C\sqrt{n}$ such that $\chi(G - X) \le u$, and we know that $\chi(G[X]) \le 3$, which implies that $\chi(G) \le \chi(G - X) + \chi(G[X]) \le u + 3$.

It remains to prove the two claims. For the first claim, note that the variable Z is 1-Lipschitz with respect to vertex exposure (convince yourselves!). Thus, the Azuma-Hoeffding inequality implies that for any $\lambda > 0$, we have

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \ge \lambda] \le 2\exp(-\lambda^2/2n)$$

We choose $C = C(\epsilon) > 0$ such that $2\exp(-C^2/2) \le \epsilon$. Thus, setting $\lambda = C\sqrt{n}$, we have

$$\mathbb{P}\big[|Z - \mathbb{E}[Z]| \ge C\sqrt{n}\big] \le \epsilon. \tag{1}$$

Note that by definition of Z, we have $Z = 0$ if only if $\chi(G) \le u$, and by definition of u, the probability of this is $> \epsilon$. Observe that this implies that $\mathbb{E}[Z] \le C\sqrt{n}$. Indeed, if we had $\mathbb{E}[Z] > C\sqrt{n}$, then $Z = 0$ would be one example for which $|Z - \mathbb{E}[Z]| \ge C\sqrt{n}$, the probability of which is at most ϵ by (1), and this would contradict our observation that $\mathbb{P}[Z = 0] > \epsilon$.

Now, having established that $\mathbb{E}[Z] \le C\sqrt{n}$, we can apply (1) again to see that

$$\mathbb{P}\big[Z \ge 2C\sqrt{n}\big] \le \mathbb{P}\big[Z - \mathbb{E}[Z] \ge C\sqrt{n}\big] \le \epsilon.$$

We now turn to the second claim. Note that if for some set W of vertices, $G[W]$ is not 3-colorable, then there exists in W a minimal subset T which is not 3-colorable. This means that every vertex in $G[T]$ has degree at least 3, and thus $G[T]$ has at least $3|T|/2$ edges. We now want to show that such a set of size at most $C\sqrt{n}$ does not exist with high probability. For this, we apply a union bound over all sets T. For a given value $4 \le t \le C\sqrt{n}$, there are $\binom{n}{t}$ possibilities to choose T of size t, and for any such choice, the probability that T induces at least $3t/2$ edges is at most $\binom{\binom{t}{2}}{3t/2}p^{3t/2}$. Altogether, the probability that our desired event fails is at most

$$\sum_{t=4}^{C\sqrt{n}} \binom{n}{t}\binom{\binom{t}{2}}{3t/2}p^{3t/2} \le \sum_{t=4}^{C\sqrt{n}} \left(\frac{en}{t}\right)^t \left(\frac{et}{3}\right)^{3t/2} n^{-3\alpha t/2} \le \sum_{t=4}^{C\sqrt{n}} \left(3t^{1/2}n^{1-3\alpha/2}\right)^t \le \sum_{t=4}^{C\sqrt{n}} \left(3C^{1/2}n^{5/4-3\alpha/2}\right)^t$$

Here, we used the general estimate $\binom{a}{b} \le \left(\frac{ea}{b}\right)^b$, the definition of p and the upper bound on t. By our assumption that $\alpha > 5/6$, the exponent of n is negative, so we have $q := 3C^{1/2}n^{5/4-3\alpha/2} = o(1)$. We can estimate the last sum above with the geometric series $\sum_{t=1}^{\infty} q^t = q/(1-q) = o(1)$, which completes the proof. □

As the final result of this subsection, we want to prove that the chromatic number of $G(n, 1/2)$ is roughly $\frac{n}{2\log_2 n}$ with high probabiltity. This problem was open for a long time until finally settled by Bollobás in the 80s.

Define the function $f(k) := \binom{n}{k}2^{-\binom{k}{2}}$, which is the expected number of cliques of order k in $G \sim G(n, 1/2)$. Since the complement of G has the same distribution as G, it is also the expected number of independent sets of size k.

In Section 5.4 we saw that $f(k)$ gets below 1 at a point $k \sim 2\log_2 n$, and that $f(k)/f(k+1) = n^{1+o(1)}$ for k in the range $k \sim 2\log_2 n$ (these are simple calculations). Define k_0 such that $f(k_0 - 1) > 1 > f(k_0)$, and let $k = k_0 - 4$. Then $k \sim 2\log_2 n$ and $f(k) > n^{3-o(1)}$.

Our goal is to show that $\chi(G(n, 1/2)) = (1 + o(1))n/k$. The lower bound can be seen relatively easily (as we will show soon). The upper bound is much more involved, as here we need to partition the graph into independent sets of average size $(1 + o(1))k$. The crucial step is to show that with high probability $G(n, 1/2)$ contains large independent sets 'everywhere', (where large means of size roughly k). The key ingredient is the following theorem. Recall that $\omega(G)$ and $\alpha(G)$ denote the clique number (the largest number of vertices in a clique) and the independence number of G, respectively.

Theorem 8.4. *For $G \sim G(n, 1/2)$ and k as above, we have $\mathbb{P}[\omega(G) < k] \leq \exp(-n^{2-o(1)})$.*

Assuming this, we first show how it can be used to prove our main result.

Theorem 8.5 (Bollobás, 1988)**.** *For $G \sim G(n, 1/2)$, we have $\chi(G) = (1+o(1))\frac{n}{2\log_2 n}$ with probability $1 - o(1)$.*

Proof. A simple first moment calculation shows that with high probability, $\alpha(G) \leq (1+o(1))2\log_2 n$, which implies the lower bound since $\chi(G) \geq \frac{n}{\alpha(G)}$.

We now turn to the upper bound. Define $m = \lfloor n/\ln^2 n \rfloor$. For any set S of m vertices, the induced graph $G[S]$ has the distribution of $G(m, 1/2)$. We apply Theorem 8.4 (to the complement of $G[S]$) to see that $\mathbb{P}[\alpha(G[S]) < k'] \leq \exp(-m^{2-o(1)})$, where k' is defined in the same way as k by replacing n with m. Since m is only smaller than n by a polylog factor, k' is essentially of the same order as k. More precisely, $k' = (1 + o(1))2\log_2 m = (1 + o(1))2\log_2 n = (1 + o(1))k$.

Now, we can use the union bound over the less than 2^n possibilities for S to conclude that with probability at least $1 - 2^n \exp(-m^{2-o(1)}) = 1 - o(1)$, every set S of size m contains an independent set of size k'.

Assume that this property holds. Repeatedly pull out independent sets of size k' from G and give each a new color, until less than m vertices remain. Give each remaining vertex a unique new color. Clearly, this produces a proper coloring. Moreover, the number of colors used is at most $\frac{n}{k'} + m$. Since $m = o(\frac{n}{\log n})$ and $k' = (1 + o(1))2\log_2 n$, we have $\frac{n}{k'} + m = (1 + o(1))\frac{n}{2\log_2 n}$, implying the theorem. □

Finally, we prove Theorem 8.4. This is where we use the Azuma-Hoeffding inequality. As in the proof of Theorem 8.3, a somewhat unusual random variable is defined.

Proof of Theorem 8.4. Let Y be the random variable denoting the size of a maximal collection of edge-disjoint k-cliques in G. Clearly, $\omega(G) < k$ if and only if $Y = 0$. Thus, our aim is to prove that $\mathbb{P}[Y = 0] \leq \exp(-n^{2-o(1)})$.[1] Below, we will see that the expectation of Y is sufficiently large. More precisely, we will show that $\mathbb{E}[Y] \geq (1 - o(1))n^2/2k^4$.

The crucial observation is that Y is 1-Lipschitz with respect to edge exposure. Indeed, if we remove one edge from G, this destroys at most one clique in a collection of edge-disjoint k-cliques. Similarly, adding an edge can only help to add at most one further clique.

Hence, we are in a position to apply the Azuma-Hoeffding inequality and get

$$\mathbb{P}[|Y - \mathbb{E}[Y]| \geq \lambda] \leq 2\exp\left(-\frac{\lambda^2}{2\binom{n}{2}}\right) \leq 2\exp\left(-\frac{\lambda^2}{n^2}\right).$$

Since $Y = 0$ implies $|Y - \mathbb{E}[Y]| \geq \mathbb{E}[Y]$, we can substitute $\lambda = \mathbb{E}[Y] \geq (1 - o(1))n^2/2k^4 \geq n^2/3k^4$ and infer

$$\mathbb{P}[Y = 0] \leq \mathbb{P}[|Y - \mathbb{E}[Y]| \geq \mathbb{E}[Y]] \leq 2\exp\left(-\frac{(n^2/3k^4)^2}{n^2}\right) \leq 2\exp\left(-\frac{n^2}{9k^8}\right).$$

Since $k = n^{o(1)}$, this implies the claim.

It remains to prove the lower bound on $\mathbb{E}[Y]$. Recall that $\mu := f(k) = \binom{n}{k}2^{-\binom{k}{2}}$ is the expected number of k-cliques in G. Let W be the number of unordered pairs $\{A, B\}$ of k-cliques in G such that $2 \leq |A \cap B| \leq k - 1$, that is, they have non-trivial edge intersection. Then

$$\mathbb{E}[W] = \frac{1}{2} \cdot \binom{n}{k} 2^{-\binom{k}{2}} \sum_{i=2}^{k-1} \binom{k}{i}\binom{n-k}{k-i} 2^{\binom{i}{2}-\binom{k}{2}} = \frac{1}{2} \cdot \mu^2 \sum_{i=2}^{k-1} \binom{k}{i} \frac{\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}. \tag{2}$$

(The $\frac{1}{2}$-factor is because we are counting unordered pairs.)

One can check that $\frac{\binom{n-k}{k-i}}{\binom{n}{k}} = (1 + o(1))\frac{k(k-1)\cdots(k-i+1)}{n^i}$. Therefore, for $i = 2$, the summand becomes

$$(1 + o(1))\binom{k}{2}\frac{k(k-1)}{n^2}2^{\binom{2}{2}} = (1 + o(1))\frac{k^4}{n^2}.$$

On the other extreme, for $i = k - 1$ the summand equals

$$k \cdot \frac{n-k}{\binom{n}{k}} \cdot 2^{\binom{k-1}{2}} = \frac{k(n-k)}{f(k)} \cdot 2^{-k+1} = \frac{n^{-1+o(1)}}{f(k)} < n^{-4+o(1)},$$

which is negligible compared to the term $i = 2$. (Here we used that $k \sim 2\log_2 n$ and $\mu > n^{3-o(1)}$.) One can check that the other summands are also negligible compared to the term $i = 2$, meaning that the sum in (2) is dominated[2] by this term, and hence $\mathbb{E}[W] \sim \mu^2 k^4/2n^2$. We omit the details of these calculations.

[1] So far, the same would be true if we omit 'edge-disjoint' in the definition of Y. However, this trick (of considering edge-disjoint cliques) is crucial to obtain a Lipschitz condition for Y.

[2] The calculations done here are essentially the same as those done in Section 5.4. There the dominating term was $i = k - 1$, while here it is $i = 2$. The reason is that there we only knew that $f(k) \to \infty$, while here we know that $f(k) > n^{3-o(1)}$.

Now fix some $q \in [0,1]$, to be chosen later. We claim that $Y \geq q \cdot X_k - q^2 \cdot W$, where X_k is the number of k-cliques in G. To see this, we use the alteration method as follows. First, "activate" every k-clique of G independently with probability q. (In other words, we toss a coin with success probability q for each k-clique, and call the k-clique activated if the toss was successful.) Let W' be the number of pairs counted by W (i.e., pairs of cliques which share edges) for which both cliques were activated. Clearly, $\mathbb{E}[W'] = q^2 \cdot W$ (here the graph G is considered fixed and the expectation is with respect to the random coin tosses). Then, remove one clique from each such pair. This yields a set of edge-disjoint k-cliques, whose expected size is at least $q \cdot X_k - q^2 \cdot W$, as required. Taking expectations (now with respect to the choice of G), we see that

$$\mathbb{E}[Y] \geq q \cdot \mathbb{E}[X_k] - q^2 \cdot \mathbb{E}[W].$$

We choose $q = \frac{\mathbb{E}[X_k]}{2\mathbb{E}[W]} \sim (\mu k^4/n^2)^{-1}$ to maximize this expression, and obtain that $\mathbb{E}[Y] \geq \frac{\mathbb{E}[X_k]^2}{4\mathbb{E}[W]} \geq (1-o(1))n^2/2k^4$, as desired. Note here that for this to work, we need $q \leq 1$ as a probability. This is satisfied since $\mu = f(k) > n^{3-o(1)}$ by our choice of k. □

8.3 Isoperimetric inequality for the Hamming cube

As a final illustration, we use the Azuma-Hoeffding inequality to prove an isoperimetric inequality for the Hamming cube.

Consider $\Omega = \{0,1\}^n$ equipped with the Hamming metric, that is, the distance of two points $x, y \in \Omega$ is defined as $\rho(x,y) = \sum_{i=1}^n |x_i - y_i|$, i.e. the number of coordinates in which they differ.

The ball of radius r around a vertex x is $B(x,r) = \{y \colon \rho(x,y) \leq r\}$. Moreover, for a subset $A \subset \Omega$, we define $B(A,r) = \cup_{x \in A} B(x,r)$ as the set of all points which have distance at most r from A.

The following theorem shows that for any given set A of positive density, say ϵ, almost all points are very close to A.

Theorem 8.6. *Let $\lambda > 0$ and $\epsilon = 2\exp(-\lambda^2/2)$. Then for any $A \subset \Omega$ with $|A| > \epsilon 2^n$, we have $|B(A, 2\lambda\sqrt{n})| \geq (1-\epsilon)2^n$.*

Proof. Define the probability space by choosing a random point $y \in \Omega$ as follows: for each coordinate i independently, let $y_i = 0$ with probability $1/2$ and $y_i = 1$ otherwise. Note that this implies that every point of Ω is equally likely to be chosen (i.e., this is just the uniform distribution on Ω).

Define the random variable Y as the distance of the randomly chosen point y to A, that is $Y = \min_{x \in A} \rho(x,y)$.

Observe that $Y = 0$ if and only if $y \in A$, thus, since the distribution is uniform, $\mathbb{P}[Y = 0] = \frac{|A|}{|\Omega|} > \epsilon$.

Moreover, we have $Y \leq r$ if and only if $y \in B(A,r)$, so $\mathbb{P}[Y \leq r] = \frac{|B(A,r)|}{|\Omega|}$. Hence, our goal is to show that $\mathbb{P}[Y \leq r] \geq 1 - \epsilon$, or equivalently, $\mathbb{P}[Y > r] \leq \epsilon$, for $r = 2\lambda\sqrt{n}$.

Since changing one coordinate of y can change its distance to any point by at most 1, we have that Y is 1-Lipschitz. Thus, by the Azuma-Hoeffding inequality,

$$\mathbb{P}\big[|Y - \mathbb{E}[Y]| \geq \lambda\sqrt{n}\big] \leq 2\exp\big(-(\lambda\sqrt{n})^2/2n\big) = 2\exp(-\lambda^2/2) = \epsilon.$$

We claim that $\mathbb{E}[Y] \leq \lambda\sqrt{n}$. Indeed, if we had $\mathbb{E}[Y] \geq \lambda\sqrt{n}$, then $Y = 0$ would be a special case of $|Y - \mathbb{E}[Y]| \geq \lambda\sqrt{n}$, but then we would have $\mathbb{P}[Y = 0] \leq \mathbb{P}[|Y - \mathbb{E}[Y]| \geq \lambda\sqrt{n}] \leq \epsilon$, in contradiction to our earlier observation that $\mathbb{P}[Y = 0] > \epsilon$.

Now, having $\mathbb{E}[Y] \leq \lambda\sqrt{n}$, we obtain that $\mathbb{P}[Y \geq 2\lambda\sqrt{n}] \leq \epsilon$, as desired. □

A nice corollary of the above theorem is the following result that given a set A as above, there will be two points in A such that one is almost identical to the complement of the other.

Corollary 8.7. *Let ϵ, λ be as above. Then for any $A \subset \Omega$ with $|A| > \epsilon 2^n$, there are two points $x, y \in A$ such that $\rho(x, \bar{y}) \leq 2\lambda\sqrt{n}$, where $\bar{y}_i = 1 - y_i$ for all $i \in [n]$.*

Proof. Applying the previous theorem, $|B(A, 2\lambda\sqrt{n})| \geq (1 - \epsilon)2^n$. Let $A' = \{\bar{y} : y \in A\}$. Clearly $|A'| = |A| > \epsilon 2^n$. Hence, $B(A, 2\lambda\sqrt{n}) \cap A' \neq \emptyset$. Let $z \in B(A, 2\lambda\sqrt{n}) \cap A'$. Since $z \in B(A, 2\lambda\sqrt{n})$, there exists $x \in A$ with $\rho(x, z) \leq 2\lambda\sqrt{n}$. On the other hand, $z \in A'$ means that there exists $y \in A$ with $z = \bar{y}$. Now x and y satisfy the required conditions. □

9 Correlation inequalities

Many results in this section trace their origins to the study of percolation and statistical physics, but have proved to be quite useful in combinatorics as well.

Example 9.1. Let $G \sim \mathcal{G}(n,p)$. Consider the following properties:

- P_1 is the property of having a triangle, let $p_1 = \mathbb{P}(P_1)$;
- P_2 is the property of containing a Hamilton cycle, let $p_2 = \mathbb{P}(P_2)$;
- P_3 is the property of containing *both* a triangle and a Hamilton cycle, let $p_3 = \mathbb{P}(P_3)$.

Is there a relation between p_1, p_2 and p_3? Intuitively, knowing that G is Hamiltonian suggests that it should have many edges, hence making it more likely to contain a triangle. In other words, we may suspect that:

$$\mathbb{P}(\mathcal{G}(n,p) \text{ has a triangle} \mid \mathcal{G}(n,p) \text{ has a Hamilton cycle}) \geq p_1.$$

Substituting the definition of conditional probability, we obtain the equivalent inequality $p_3 \geq p_1 p_2$. This intuition turns out to be correct, even in far greater generality. However, it is quite non-obvious how to actually prove this type of statement.

Our main goal in this section is to prove this claim for arbitrary increasing properties P_1, P_2 and $P_3 = P_1 \cap P_2$. Let us first define what we mean by an increasing or decreasing property.

Definition 9.2. A property P of graphs is said to be increasing/decreasing if it is preserved under addition/deletion of edges.

For example, being Hamiltonian, having a triangle, being connected, or having a perfect matching are all examples of increasing properties. On the other hand, containing an independent set of some fixed size k, being k-colourable, or being planar are decreasing properties.

We will prove the claim in even more generality, namely for product measures on $\Omega = \{0,1\}^N$. In this setting, we are given $p_1, \ldots, p_N \in [0,1]$, and the probability of $\omega \in \Omega = \{0,1\}^N$ is $\mathbb{P}(\omega) = \prod_{i:w_i=1} p_i \cdot \prod_{i:w_i=0}(1-p_i)$ (in other words, the ith coordinate is 1 with probability p_i and 0 otherwise, and the coordinates are independent). If all $p_i = p$ and $N = \binom{n}{2}$, then we recover $\mathcal{G}(n,p)$.

For $x, y \in \{0,1\}^N$, let us write $x \leq y$ to mean that $x_i \leq y_i$ for every $1 \leq i \leq N$. An event $A \subseteq \Omega$ is called increasing if for all $x, y \in \{0,1\}^N$ with $x \leq y$, if x is in A then y is also in A. We define a decreasing event analogously (i.e., A is decreasing if $y \in A$ whenever $x \in A$ and $y \leq x$). Since there is a natural correspondence between $\{0,1\}^N$ and subsets of $[N]$, where $x \in \{0,1\}^N$ corresponds to the set $\{1 \leq i \leq N : x_i = 1\}$, one can think of $A \subseteq \{0,1\}^N$ as a family of subsets of $[N]$.

Theorem 9.3. *Let $A, B \subseteq \{0,1\}^N$. If both A and B are increasing or both are decreasing, then*

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A)\mathbb{P}(B).$$

If one of A and B is increasing and the other is decreasing, then

$$\mathbb{P}(A \cap B) \leq \mathbb{P}(A)\mathbb{P}(B).$$

Proof. Let us first prove the theorem in the case when both A and B are assumed to be increasing.

The proof is by induction on N. If $N = 1$ there are 3 different options for $A \subseteq \{0,1\}$ to be increasing. If A is empty then both sides of the inequality are 0 so it holds. If $A = \{0,1\}$ then $\mathbb{P}(A \cap B) = \mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B)$. Same holds for B by symmetry, so the only remaining case is $A = B = \{1\}$, and then $\mathbb{P}(A \cap B) = p_1 \geq p_1^2 = \mathbb{P}(A)\mathbb{P}(B)$.

Now suppose $N > 1$. For $i = 0, 1$, define $A_i := \{x \in \{0,1\}^{N-1} : (x, i) \in A\} \subseteq \{0,1\}^{N-1}$. We define B_i analogously. Observe that since A and B are increasing, the A_i's and B_i's are increasing as well, and $A_0 \subseteq A_1$ and $B_0 \subseteq B_1$ (check!). Let us put $a := \mathbb{P}(A), a_0 := \mathbb{P}(A_0), a_1 := \mathbb{P}(A_1)$ and $b := \mathbb{P}(B), b_0 := \mathbb{P}(B_0), b_1 := \mathbb{P}(B_1)$. In particular $a_0 \leq a_1$ and $b_0 \leq b_1$.

Observe now that we have $a = a_0(1-p_N) + a_1 p_N$ (recall that p_N is the probability that the Nth coordinate is 1). Analogously we have $b = b_0(1-p_N) + b_1 p_N$. We also have

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(A_0 \cap B_0) \cdot (1-p_N) + \mathbb{P}(A_1 \cap B_1) \cdot p_N \\
&\geq a_0 b_0 (1-p_N) + a_1 b_1 p_N =: X, \\
\mathbb{P}(A)\mathbb{P}(B) &= (a_0(1-p_N) + a_1 p_N)(b_0(1-p_N) + b_1 p_N) =: Y,
\end{aligned}$$

where in the only inequality we used the induction hypothesis. Finally,

$$\begin{aligned}
X - Y &= a_0 b_0 - a_0 b_0 p_N + a_1 b_1 p_N - a_0 b_0 (1-p_N)^2 - (a_0 b_1 + a_1 b_0) p_N (1-p_N) - a_1 b_1 p_N^2 \\
&= a_0 b_0 (p_N - p_N^2) + a_1 b_1 (p_N - p_N^2) - (a_0 b_1 + a_1 b_0) p_N (1-p_N) \\
&= (a_1 - a_0)(b_1 - b_0) p_N (1-p_N) \\
&\geq 0,
\end{aligned}$$

which completes the proof for the increasing case. The remaining cases follow by taking complements and observing this changes the type of monotonicity (exercise!). □

This inequality is called FKG in the literature and it holds in greater generality (see e.g. Alon-Spencer for more details).

Let us show an example. Let $\mathcal{F}$ be a collection of subsets of $[N]$. We say that $\mathcal{F}$ is intersecting if any two sets in $\mathcal{F}$ have a non-empty intersection. For example, the family consisting of all subsets containing 1 is intersecting and has size 2^{N-1}. Another example is the family of all sets of size larger than $N/2$, which if N is odd also has size 2^{N-1}. In fact this is the largest possible size of an intersecting family. Denote by $\mathcal{P}([N])$ the set of all subsets of $[N]$.

Proposition 9.4. *If $\mathcal{F} \subseteq \mathcal{P}([N])$ is intersecting then $|\mathcal{F}| \leq 2^{N-1}$.*

Proof. Note that the pairs $\{J, [N] \setminus J\}$ partition $\mathcal{P}([N])$. Note also that since J and $[N] \setminus J$ are disjoint, only one of them can be in $\mathcal{F}$ without violating the intersecting condition. Therefore, $|\mathcal{F}| \leq 2^{N-1}$. □

Another closely related notion is that of a family where no two sets contain the whole ground set $[N]$. An example here is the family of all sets not containing element 1. Observe that if $\mathcal{F}$ satisfies this condition (i.e., $X \cup Y \neq [N]$ for every pair $X, Y \in \mathcal{F}$), then $\{[N] \setminus X : X \in \mathcal{F}\}$ is intersecting, and vice versa. So by the above proposition, the size of such a family is at most 2^{N-1} as well.

What happens if $\mathcal{F}$ satisfies both these conditions? In other words, if it is both intersecting and there are no two sets in the family whose union is $[N]$. One example is the $\mathcal{F}$ which consists of all sets containing element 1 but not containing element 2. Then $|\mathcal{F}| = 2^{N-2}$. It turns out that this is the largest possible size of such a family.

Theorem 9.5. *Any intersecting family $\mathcal{F}$ of subsets of $[N]$ with no two $F, F' \in \mathcal{F}$ having $F \cup F' = [N]$ has size at most 2^{N-2}.*

Proof. For $F \subseteq [N]$, let $x_F \in \{0,1\}^N$ denote the characteristic vector of F (namely $(x_F)_i = 1$ if $i \in F$ and $(x_F)_i = 0$ otherwise).

Let $\mathcal{A}$ be the *up-closure* of $\mathcal{F}$, defined as the family of all subsets containing a set in $\mathcal{F}$. Observe that if we let $A = \{x_F \mid F \in \mathcal{A}\} \subseteq \{0,1\}^N$, then A is increasing. Note also that $\mathcal{A}$ is still intersecting so $|A| = |\mathcal{A}| \leq 2^{n-1}$.

Similarly, we can define the *down-closure* $\mathcal{B}$ of $\mathcal{F}$ to be the family of all subsets of $[N]$ contained in a set in $\mathcal{F}$. Observe that if we let $B = \{x_F \mid F \in \mathcal{B}\} \subseteq \{0,1\}^N$, then B is decreasing. Note also that $\mathcal{B}$ still contains no two sets with union $[N]$, so $|B| = |\mathcal{B}| \leq 2^{n-1}$.

Observe now that $\mathcal{F} \subseteq \mathcal{A} \cap \mathcal{B}$ so $|\mathcal{F}| \leq |A \cap B|$. If we apply the FKG inequality to A and B, w.r.t. the uniform measure (i.e., $p_i = 1/2$ for all $1 \leq i \leq N$), we obtain

$$\frac{|A \cap B|}{2^n} = \mathbb{P}(A \cap B) \leq \mathbb{P}(A)\mathbb{P}(B) = \frac{|A|}{2^n} \cdot \frac{|B|}{2^n} \leq \frac{1}{4},$$

since A and B are monotone in opposite ways (A is increasing and B is decreasing). Rearranging gives the desired bound. □

10 Probabilistic gems

In this last section, we will see some beautiful probabilistic proofs of famous results in extremal graph theory, discrete geometry and additive number theory. Previously, many of these results had long and complicated proofs.

10.1 Independence number of triangle-free graphs

Let G be a graph with n vertices and average degree d. Recall that $\alpha(G)$ denotes the independence number of G, that is, the size of the largest independent set. By applying Turán's theorem to the complement of G, we can see that $\alpha(G) \geq n/(d+1)$, and this bound is tight in general as can be seen by taking G to be the disjoint union of cliques of size $d+1$.

The question is whether one can improve this bound if one has additional information about G, for instance if we know that G is triangle-free. Ajtai, Komlós and Szemerédi famously proved that any triangle-free graph G with average degree d has an independent set of size $c\frac{n}{d}\log d$ for some universal constant $c > 0$.[3]

Here, we present a beautiful proof of this theorem due to Shearer. We just pick an independent set I from the set of all independent sets uniformly at random, and then show that the expected size is as desired. For the proof to work, we need to assume that the maximum degree of G is at most d. Note that this still implies the result of Ajtai, Komlós and Szemerédi since if G has average degree at most d, then there are at most $n/2$ vertices of degree larger than $2d$, so we can delete them and find a large independent set in the remaining graph.

Theorem 10.1 (Shearer)**.** *Let G be a triangle-free graph on n vertices and maximum degree at most d. Then there exists an independent set of size at least $c\frac{n}{d}\log d$, for some universal constant $c > 0$.*

Proof. We want to assume $4 \leq \log d \leq \sqrt{d}$. This clearly holds if d is sufficiently large. By choosing c small enough, we can ensure that the statement also holds for small d by the aforementioned bound from Turán's theorem.

Pick an independent set I uniformly at random from the set of all independent sets. For a vertex v, define the random variable

$$x_v := d\mathbf{1}_{v\in I} + |N(v) \cap I|.$$

Here $\mathbf{1}_{v\in I}$ is 1 if $v \in I$ and 0 otherwise. Note that

$$\sum_{v\in V(G)} x_v = d \sum_{v\in V(G)} \mathbf{1}_{v\in I} + \sum_{v\in V(G)} |N(v) \cap I| = d|I| + \sum_{v\in I} d_G(v) \leq 2d|I|,$$

where the last inequality holds by the maximum degree assumption.

The crucial claim is that, for some universal constant $c > 0$, we have $\mathbb{E}[x_v] \geq c\log d$ for all vertices v. If this holds true, then by linearity of expectation, we have $2d\mathbb{E}[|I|] \geq \sum_v \mathbb{E}[x_v] \geq cn\log d$, and rearranging yields the theorem.

[3] In this section, log denotes the logarithm to base 2, but since we do not optimize the constant c, the base does not really matter.

To prove the claim, fix any vertex v, and define $W = N(v) \cup \{v\}$. We now condition on the outcome of $I \setminus W$. Assume $I \setminus W = I_0$. Clearly this implies that I_0 itself is independent, and hence the event has positive probability and the conditional probability space is well-defined. We will show that for any such I_0, we have $\mathbb{E}[x_v \mid I \setminus W = I_0] \geq c \log d$, which, by the law of total probability, implies the claim.

Henceforth, fix I_0. Let U denote the set of vertices in $N(v)$ which do not have a neighbour in I_0, and let $u = |U|$. The crucial observation is the following: Since G is triangle-free, U is an independent set. Thus, conditioning on $I \setminus W = I_0$, there are exactly $1 + 2^u$ possibilities for I: one in which $I = I_0 \cup \{v\}$, and 2^u in which I is the union of I_0 and some subset of U. In the first case, the first term of x_v contributes d, and in the second case, a subset of U of size i contributes i to the second term of x_v. We deduce

$$\mathbb{E}[x_v \mid I \setminus W = I_0] = \frac{d}{1+2^u} + \frac{\sum_{i=0}^{u} i\binom{u}{i}}{1+2^u} = \frac{d}{1+2^u} + \frac{u2^{u-1}}{1+2^u} \geq \frac{d}{2^{u+1}} + \frac{u}{4},$$

where we used the identity $\sum_{i=0}^{u} i\binom{u}{i} = \sum_{i=1}^{u} u\binom{u-1}{i-1} = u2^{u-1}$ and $1 + 2^u \leq 2^{u+1}$.

Finally, we simply have to check that either of the terms $\frac{d}{2^{u+1}}$ and $\frac{u}{4}$ is large enough. Clearly, if $u \geq \frac{1}{4}\log d$, then we are done since the second term is large enough. So assume that $u \leq \frac{1}{4}\log d$. Then (recall our assumption $4 \leq \log d \leq \sqrt{d}$) we have $u + 1 \leq \frac{1}{2}\log d$ and hence $2^{u+1} \leq \sqrt{d}$, which implies that the first term is at least $\frac{d}{\sqrt{d}} = \sqrt{d} \geq \log d$. □

10.2 The crossing number of a graph

Let $G = (V, E)$ be a graph. A *planar drawing* of G is a representation of G in the plane so that vertices of V are represented by distinct points in the plane, and the edges are represented by simple continuous curves that connect the two endpoints. We assume that no curve contains any other vertices but its endpoints, any two curves have a finite number of intersections, and no point in the plane, other than the vertices, is incident with three or more curves.[4]

The *crossing number* of G is the minimum number of crossings in any planar drawing of G, we denote it by $cr(G)$. We remark that in an optimal drawing of G (i.e. in a drawing with minimal number of crossings) no edge crosses itself. Clearly, $cr(G) = 0$ if and only if G is planar. We use the notation $e(G) = |E(G)|$ and $v(G) = |V(G)|$.

Recall that Euler's formula implies that any planar graph G has at most $3v(G) - 6$ edges. This easily implies the following lower bound on the crossing number: for any graph G,

$$cr(G) \geq e(G) - 3v(G) + 6 > e(G) - 3v(G). \tag{3}$$

(Delete edges in crossings one by one.) However, if the number of edges is large, then this bound is not very good. We will now give a much better estimate.

Theorem 10.2 (Ajtai-Chvátal-Newborn-Szemerédi; Leighton, 1982-83)**.** *Let G be a graph with $e(G) \geq 4v(G)$. Then*

$$cr(G) \geq \frac{e(G)^3}{64v(G)^2}.$$

[4] Every (finite) graph has a planar drawing: for example, choose points for the vertices randomly on the unit circle, the edges are represented by straight line segments. This is a planar drawing with probability equal to 1.

Proof. Take a planar drawing of G that has $cr(G)$ crossings. Let $0 < p \leq 1$ be a parameter to be specified later. Let G' be the induced subgraph of G obtained by choosing each vertex (independently) with probability p. Clearly, we have $\mathbb{E}v(G') = pv(G)$ and $\mathbb{E}e(G') = p^2e(G)$. The sub-drawing obtained from the optimal drawing of G by keeping only the vertices and edges corresponding to G' is a planar drawing of G'. Let us denote the number of crossings of G' in this sub-drawing by X. Clearly, $cr(G') \leq X \leq cr(G)$. (The sub-drawing is not necessarily an optimal drawing of G'.)

Note that each crossing survives if and only if the four endpoints of the two edges in the crossing are chosen. Hence, $\mathbb{E}X = p^4cr(G)$.

By (3), we have $X \geq cr(G') \geq e(G') - 3v(G')$. This holds for any outcome G', so we must also have that

$$\mathbb{E}X \geq \mathbb{E}e(G') - 3\mathbb{E}v(G').$$

Substituting the calculated expectations yields $p^4cr(G) \geq p^2e(G) - 3pv(G)$, and after rearranging we get

$$cr(G) \geq \frac{p^2e(G) - 3pv(G)}{p^4}.$$

Substituting $p = 4v(G)/e(G)$ (then $0 < p \leq 1$ since $e(G) \geq 4v(G)$) gives the required bound for $cr(G)$. □

10.3 The Szemerédi-Trotter theorem

The crossing number theorem can be used to prove the Szemerédi-Trotter theorem; this was discovered by Székely. The original proof was far more complicated.

Let P be a set of n distinct points in the plane, and let L be a set of m distinct lines in the plane. Let $I(P, L) = |\{(p, l)|\ p \in P, l \in L, p \in l\}|$, the number of point-line incidences.

Theorem 10.3 (Szemerédi-Trotter, 1983)**.** *$I(P, L) \leq 4(m^{2/3}n^{2/3} + m + n)$.*

Proof. We define a graph $G = (V, E)$. The vertices of V are the points of P. We have an xy edge in E if x, y lie on the same line, say $l \in L$, and there is no other point of P on l between x and y. We have the following lower bound for $e(G)$:

$$e(G) = \sum_{l \in L} \max\{|P \cap l| - 1, 0\} \geq \sum_{l \in L} |P \cap l| - m = I(P, L) - m.$$

Consider the natural drawing of G which is defined by the points P and line segments coming from the set of lines L. Since every crossing in G is an intersection point of two lines in L, and any two lines intersect in at most one point, we have that $cr(G) \leq \binom{m}{2} < m^2/2$. We consider two cases. In the first case $e(G) \geq 4v(G)$. Then the crossing number theorem implies that

$$m^2/2 \geq cr(G) \geq \frac{e(G)^3}{64n^2} \geq \frac{(I(P, L) - m)^3}{64n^2},$$

hence, in this case $I(P, L) \leq (32m^2n^2)^{1/3} + m$. In the second case $e(G) < 4v(G)$, so $I(P, L) - m < 4v(G)$. Thus, $I(P, L) < 4n + m$. In both cases $I(P, L) \leq 4(m^{2/3}n^{2/3} + m + n)$. □

10.4 The sum-product theorem

Let $A \subset \mathbb{R} \setminus \{0\}$ be a finite set. Let $A+A = \{a+b \mid a, b \in A\}$ and let $A \cdot A = \{ab \mid a, b \in A\}$. Observe that $A+A$ is comparatively small if A consists of an arithmetic progression. Similarly, $A \cdot A$ is comparatively small if A consists of a geometric progression. One has the intuition that at the same time $|A+A|$ and $|A \cdot A|$ cannot both be small. Erdős conjectured that $\max\{|A+A|, |A \cdot A|\} \geq |A|^{2-o(1)}$. This conjecture is still wide open. Erdős and Szemerédi proved that there exists $\varepsilon > 0$ such that $\max\{|A+A|, |A \cdot A|\} \geq |A|^{1+\varepsilon}$. Elekes observed that the Szemerédi-Trotter theorem can be used to obtain a much better bound, with a simpler proof.

Theorem 10.4 (Elekes, 1997)**.** *There exists a positive constant c such that*

$$\max\{|A+A|, |A \cdot A|\} \geq c|A|^{5/4}.$$

Proof. We will use the Szemerédi-Trotter theorem, for that we need points and lines. Let $P = (A+A) \times (A \cdot A)$ be the Cartesian product of the sum set $A+A$ and the product set $A \cdot A$. This is a point set in the plane. Clearly, $n = |P| = |A+A||A \cdot A|$, and hence $\max\{|A+A|, |A \cdot A|\} \geq \sqrt{n}$. We define a set L of lines as follows: For every $a, b \in A$ we have a line $l_{ab} \in L$, where $l_{ab} = \{(x, y) \mid y = a(x-b)\}$. Clearly, $|L| = |A|^2$. By the Szemerédi-Trotter theorem we get that

$$I(P, L) \leq 4(|A|^{4/3}n^{2/3} + n + |A|^2).$$

By the definition of P, we have $|A|^2 \leq n \leq |A|^4$. This implies that $n \leq |A|^{4/3}n^{2/3}$ and $|A|^2 \leq |A|^{4/3}n^{2/3}$, so absorbing the minor order terms we have $I(P, L) \leq 12|A|^{4/3}n^{2/3}$.

On the other hand, it is easy to see that for every pair $a, b \in A$ and arbitrary $c \in A$ the point $(b+c, ac)$ belongs to P, hence, every line of L contains at least $|A|$ points from P. This implies that $I(P, L) \geq |L||A| = |A|^3$.

Combining the lower and upper bound for $I(P, L)$, we obtain

$$|A|^3 \leq I(P, L) \leq 12|A|^{4/3}n^{2/3}.$$

Rearranging yields $\sqrt{n} \geq 12^{-3/4}|A|^{5/4}$, completing the proof. □