Probabilistic Methods in Combinatorics

Joshua Erde

Department of Mathematics,
Universität Hamburg.

Contents

Preface

These notes were used to lecture a course at UHH for masters level students in the winter semester of 2015, and again in the Summer Semester of 2018. A large part of the course material was heavily based on the excellent set of lecture notes "The Probabilistic Method" by Matoušek and Vondrák, which themselves follow the book of the same name by Alon and Spencer. The latter cannot be recommended highly enough as an introductory text to the subject and covers most of the material in these lectures (and much more) in depth.

There are a few reasons for writing yet another set of notes on the same subject. Firstly there is simply the matter of length. German semesters are slightly longer than some other university terms, and so the lecture course was to run for 26 lectures of 90 minutes each, which would easily have exhausted the material in the previous notes with plenty of time to spare. This extra time has allowed us to include more examples of applications of the method, such as the beautiful result on the threshold for Hamiltonicity in Section 7, as well as chapters on wider topics not usually covered in an introductory course.

Secondly the course was designed with graph theorists in mind, with examples chosen as much as possible with an emphasis on graph theoretical problems. This was not always possible, and indeed in some cases not always preferable as there are many beautiful and instructive examples arising naturally from set systems or hypergraphs, but hopefully the result was that the students were able to easily understand the statement of the results and motivations for them.

The material from Section 5 on the method of dependent random choice comes from the survey paper "Dependent Random Choice" by Fox and Sudakov, which is a good introductory text if you wish to know more about the technique. Similarly the material from Section 15 on pseudorandomness follows in part the survey paper "Pseudo-random graphs" by Krivelevich and Sudakov. Section 11.3 comes from the paper "A stronger bound for the strong chromatic index" by Bruhn and Joos and Section 14 follows in part a lecture of Alistair Sinclair.

1 Preliminaries

1.1 Probability Theory

This section is intended as a short introduction to the very basics of probability theory, covering only the basic facts about finite probability spaces that we will need to use in this course.

Definition. A *probability space* is a triple $(\Omega, \Sigma, \mathbb{P})$, where Ω is a set, $\Sigma \subseteq 2^\Omega$ is a σ-algebra (A non-empty collection of subsets of Ω which is closed under taking complements and countable unions/intersections), and $\mathbb{P}$ is a measure on Σ with $\mathbb{P}(\Omega) = 1$. That means:

- $\mathbb{P}$ is non-negative;
- $\mathbb{P}(\emptyset) = 0$;
- For all countable collections of disjoint sets $\{A_i\}_{i=1}^n$ in Σ,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The elements of Σ are called *events* and the elements of Ω are called *elementary events*. For an event A, $\mathbb{P}(A)$ is called the *probability of* A.

During this course we will mostly consider *finite probability spaces*, those where Ω is finite and $\Sigma = 2^\Omega$. In this case the probability measure $\mathbb{P}$ is determined by the value it takes on elementary events. That is, given any function $p : \Omega \to [0, 1]$ that satisfies $\sum_{\omega\in\Omega} p(\omega) = 1$, then the function on Σ given by $\mathbb{P}(A) = \sum_{\omega\in A} p(\omega)$ is a probability measure.

In a finite probability space, the most basic example of a probability measure is the *uniform distribution* on Ω, where

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \text{ for all } A \subseteq \Omega.$$

An important example of a probability space, which we will return to throughout the course, is that of a *random graph*.

Definition. The probability space of *random graphs* $\mathcal{G}(n, p)$ is a finite probability space whose elementary events are all graphs on a fixed set of n vertices, and where the probability of each graph with m edges is

$$p(G) = p^m(1 - p)^{\binom{n}{2}-m}$$

This corresponds to the usual notion of picking a random graph by including every potential edge independently with probability p. We will often denote an arbitary event from this space by $G(n, p)$. We note that p can, and often will be, a function of n.

Given a property of graphs P we say that $G(n, p)$ satisfies P *almost surely* or *with high probability* if

$$\lim_{n\to\infty} \mathbb{P}(G(n, p) \text{ satisfies } P) = 1.$$

Another natural, and often utilized, probability space related to graphs is that of $\mathcal{G}(n, M)$ where the elementary events are all graphs on n vertices with M edges, with the uniform probability measure. $\mathcal{G}(n, M)$ is, except in certain cases, much less nice to work with than $\mathcal{G}(n, p)$. We won't work with this space in the course, and mention that, quite often in applications you can avoid working with $\mathcal{G}(n, M)$ at all by proving the result you want in $\mathcal{G}(n, p)$ for an appropriate p and using some standard theorems that translate results between the spaces.

One elementary fact that we will use often is the following, often referred to as the union bound:

Lemma 1.1 (Union bound)**.** *For any $A_1, A_2, \ldots, A_n \in \Sigma$,*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \Sigma_{i=1}^{n} \mathbb{P}(A_i)$$

Proof. For each $1 \leq i \leq n$ let us define

$$B_i = A_i \setminus (A_1 \cup A_2 \cup \ldots \cup A_{i-1}).$$

Then $B_i \subset A_i$, and so $\mathbb{P}(B_i) \leq \mathbb{P}(A_i)$, and also $\bigcup B_i = \bigcup A_i$. Therefore, since the events $B_1, B_2, \ldots, B_n$ are disjoint, by the additivity of $\mathbb{P}$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \Sigma_{i=1}^{n} \mathbb{P}(B_i) \leq \Sigma_{i=1}^{n} \mathbb{P}(A_i)$$

□

Definition. Two events $A, B \in \Sigma$ are *independent* if

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

More generally, a set of events $\{A_1, A_2, \ldots, A_n\}$ is *mutually independent* if, for any subset of indices $I \subseteq [n]$,

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

It is important to note that the notion of mutual independence is stronger than simply having pairwise independence of all the pairs A_i, A_j. Intuitively, the property of independence of two events, A and B, should mean that knowledge about whether or not A occurs should not influence the likelihood of B occurring. This intuition is made formal with the idea of *conditional probability.*

Definition. Given two events $A, B \in \Sigma$ such that $\mathbb{P}(B) \neq 0$, we define the *conditional probability of A, given that B occurs*, as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Note that, as expected, A and B are independent if and only if $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Definition. A *real random variable* on probability space $(\Omega, \Sigma, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ that is $\mathbb{P}$-measurable. (That is, for any $a \in \mathbb{R}, \{\omega \in \Omega : X(\omega) \leq a\} \in \Sigma$.)

For the most part all the random variables we consider will be real valued, so for convenience throughout the course we will suppress the prefix "real". However we will in some contexts later in the course consider random variables which are functions from Ω to more general spaces.

In the case of a finite probability space, any function $X : \Omega \to \mathbb{R}$ will define a random variable. Given a measurable set $A \subseteq \mathbb{R}$ the probability that the value X takes lies in A is $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$ which we will write as $\mathbb{P}(X \in A)$.

Definition. The *expectation* of a random variable X is

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P}(\omega).$$

In the case of a finite probability space this can be expressed more clearly as

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p(\omega) X(\omega).$$

The set of random variables forms an algebra over $\mathbb{R}$ with addition and multiplication defined pointwise. For example the random variable $X + Y$ is the function from Ω to $\mathbb{R}$ defined by $(X + Y)(\omega) = X(\omega) + Y(\omega)$.

Lemma 1.2 (Linearity of expectation)**.** *For any two random variables X and Y*

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Proof.

$$\begin{aligned}\mathbb{E}(X + Y) &= \int_\Omega (X + Y)(\omega)\, d\mathbb{P}(\omega) = \int_\Omega X(\omega) + Y(\omega)\, d\mathbb{P}(\omega) \\ &= \int_\Omega X(\omega)\, d\mathbb{P}(\omega) + \int_\Omega Y(\omega)\, d\mathbb{P}(\omega) = \mathbb{E}(X) + \mathbb{E}(Y).\end{aligned}$$

□

So expectation is linear, however in general it is not multiplicative. Indeed $\mathbb{E}(XY)$ can be quite different to $\mathbb{E}(X)\mathbb{E}(Y)$, however if the two random variable are independent the two will coincide.

Definition. Two random variables X, Y are *independent* if, for any two measurable sets $A, B \subseteq \mathbb{R}$ we have

$$\mathbb{P}(X \in A \text{ and } Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

More generally, a set of random variables $\{X_1, X_2, \ldots, X_n\}$ is *mutually independent* if, for any subset of indices $I \subseteq [n]$ and any set of measurable sets $\{A_i \subseteq \mathbb{R} : i \in I\}$ we have

$$\mathbb{P}\left(X_i \in A_i \text{ for all } i \in I\right) = \prod_{i \in I} \mathbb{P}(X_i \in A_i).$$

Lemma 1.3. *For any two independent random variables, X and Y,*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Proof. We will just go through the case where X and Y are random variables on a finite probability space. Let V_X and V_Y be the set of values attained by X and Y respectively. Given any $a \in V_X$ and $b \in V_Y$ we have by independence that $\mathbb{P}(X = a \text{ and } Y = b) = \mathbb{P}(X = a)\mathbb{P}(Y = b)$. So

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{a\in V_X, b\in V_Y} ab.\mathbb{P}(X = a \text{ and } Y = b) \\ &= \sum_{a\in V_X, b\in V_Y} ab.\mathbb{P}(X = a)\mathbb{P}(Y = b) \\ &= \left(\sum_{a\in V_X} a.\mathbb{P}(X = a)\right)\left(\sum_{b\in V_Y} b.\mathbb{P}(Y = b)\right) = \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

For general probability spaces the proof is idea is the same, but is formally a little more complicated. □

1.2 Useful Estimates

Many proofs using the probabilistic method will reduce to calculating certain probabilities, for example showing they are less than 1 or tend to 0. For this purpose we will often need to estimate some quite complicated combinatorial expressions. In this section we will note down some useful estimates to apply later, both weak and strong.

Firstly the factorial function $n!$. We can bound this above weakly as $n! \leq n^n$. A more careful estimate of

$$\left(\frac{n}{e}\right)^n \leq n! \leq en\left(\frac{n}{e}\right)^n$$

can be proved by induction. Stirlings formula, $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, is more precise, but we won't need to use it's full strength.

For the binomial co-efficient $\binom{n}{k}$ we have a weak upper bound of $\binom{n}{k} \leq n^k$ (or if we're being even more imprecise $\binom{n}{k} \leq 2^n$). A more careful estimation gives

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k.$$

Sometimes it will be necessary to bound more precisely the middle binomial co-efficient and for this purpose we have

$$\frac{2^{2m}}{2\sqrt{m}} \leq \binom{2m}{m} \leq \frac{2^{2m}}{\sqrt{2m}}$$

Finally for bounding expressions of the type $(1-p)^m$ with $p > 0$ small we use the inequality $1 + x \leq e^x$, valid for all real x, which give us

$$(1-p)^m \leq e^{-mp}.$$

For bounding such expressions from below, which is usually more delicate, we often use

$$1 - p \geq e^{-2p},$$

which holds for all $0 \leq p \leq \frac{1}{2}$.

We will also use throughout the notes the following notation for comparing growth rates of functions, which it will be useful to be familiar with. Given two functions $f, g : \mathbb{N} \to \mathbb{R}$ we say that:

- $f = O(g)$ if there exists $C > 0$ such that for all sufficiently large n, $f(n) \leq Cg(n)$;
- $f = \Omega(g)$ if there exists $C > 0$ such that for all sufficiently large n, $f(n) \geq Cg(n)$;
- $f = o(g)$ if for sufficiently large n, $f(n) \leq Cg(n)$, for any fixed $C > 0$;
- $f = \omega(g)$ if for sufficiently large n, $f(n) \geq Cg(n)$, for any fixed $C > 0$;

2 The Probabilistic Method

In its most basic form the probabilistic method can be described as follows: In order to prove the existence of a combinatorial object with certain properties we pick a random object from a suitable probability space and calculate the probability that it satisfies these conditions. If we can prove that this probability is strictly positive, then we conclude that such an object must exist, since if none of the objects satisfied the conditions, the probability of a random object doing so would be zero.

The probabilistic method is useful in cases when an explicit construction of such an object does not seem feasible, and when we're more interested in the existence of such an object than in a specific example. We will start by giving some examples of the method in this form.

2.1 Ramsey Numbers

Definition. Given a graph G a *clique* is a set of vertices inducing a complete subgraph and an *independent set* is a set of vertices inducing a subgraph containing no edges. The *clique number*, $\omega(G)$, is the size of the largest clique in G and the *independence number*, $\alpha(G)$, is the size of the largest independent set.

Theorem 2.1 (Ramsey's Theorem)**.** *For any k, l there exists an n such that every $|G| \geq n$ either has $\omega(G) \geq k$ or $\alpha(G) \geq l$.*

We define the *Ramsey number*, $R(k, l)$, to be the smallest such n, that is

$$R(k,l) := \min\{n \,:\, \text{any graph on } n \text{ vertices contains a clique of size } k \text{ or an independent set of size } l\}.$$

Theorem 2.1 asserts that such a number exists. However the precise values of $R(k, l)$ are not known beyond a small number of cases, and it is an active area of research to provide good bounds for such numbers. An early use of the probabilistic method was the prove a lower bound on the diagonal Ramsey numbers $R(k, k)$.

Theorem 2.2. *For any $k \geq 3$*

$$R(k,k) > 2^{k/2-1}$$

Proof. We consider a random graph $G(n, 1/2)$, so that every pair of vertices forms an edge with probability $\frac{1}{2}$, independently of the other edges. We want to calculate the probability that this graph contains no clique or independent set of size k. Given any fixed set of k vertices, $A \subset V(G)$, we have

$$\mathbb{P}(A \text{ is a clique}) = 2^{-\binom{k}{2}},$$

since for this to be the case each of the $\binom{k}{2}$ possible edges must have been chosen, and the same holds for the probability that A is an independent set. This holds for every such $A \subset V$, of which there are $\binom{n}{k}$, and so we can use the union bound to say that

$$\mathbb{P}(G(n,1/2) \text{ contains a clique or independent set of size } k) \leq 2\binom{n}{k}2^{-\binom{k}{2}}.$$

If this probability is less than 1 then

$$\begin{aligned}&\mathbb{P}(G(n,\frac{1}{2}) \text{ doesn't contain a clique or independent set of size } k)\\ &= 1-\mathbb{P}(G(n,\frac{1}{2}) \text{ contains a clique or independent set of size }) > 0,\end{aligned}$$

and so there exists at least one graph in the probability space $\mathcal{G}(n,1/2)$, which contains neither a clique, nor an independent set of size k.

It remains to choose n such that the expression $2\binom{n}{k}2^{-\binom{k}{2}} < 1$. Using the simple estimate that $\binom{n}{k} \leq n^k$ we see that it is sufficient to choose n such that $2n^k < 2^{k(k-1)/2}$, which is certainly true if $n \leq 2^{k/2-1}$. Therefore there is a graph on $2^{k/2-1}$ vertices (ignoring floor signs for convenience) with neither a clique nor an independent set of size k, and hence $R(k,k) > 2^{k/2-1}$. □

Clearly in the last step we were slightly careless with our estimates, and in fact if we had been more careful we could improve this lower bound to $2^{k/2}$, however this is in some sense close to the best possible bound known today. In particular no lower bound of the form c^k for $c > \sqrt{2}$ is known, although, by comparison, the best known upper bound is around 4^k. Note that, whilst this proof tells us that there is a graph on $2^{k/2-1}$ vertices which contains no clique or independent set of size k, the proof is not constructive. The best known explicit constructions of graphs with no large clique or independent set give much worse bounds.

In some sense the use of a probability space here is a little forced, what is really happening in the proof is we are estimating the number of graphs on n vertices which contain either a clique or independent set of size k, and comparing this to the total number of graphs on n vertices. Showing the probability is strictly positive is equivalent to showing the first number is smaller than the second.

However, not only is the probabilistic formulation often simpler than the counting arguments, phrasing such arguments in this framework allows us to use many results from probability theory, which it would be impossible to replace with direct counting arguments (or at the very least, much more difficult).

2.2 Set Systems

Definition. A family $\mathcal{F}$ of sets is *intersecting* if for all $A, B \in \mathcal{F}, A \cap B \neq \emptyset$

For a given n, how large can an intersecting family contained in $2^{[n]}$ be? One can note that if A and B are large enough then they must intersect, in particular if both $|A|$ and $|B| > n/2$. So for example when n is odd we can construct an intersecting family of size 2^{n-1}, exactly half of the possible sets, by taking all the subsets of size $> n/2$. In the case where n is even the same can be achieved if we take a bit more care with the sets of size exactly $n/2$. Note that the only way $A, B \in [n]^{(n/2)}$ can fail to intersect is if $A = B^c$. Therefore we can form an intersecting family by arbitrarily picking one of each pair $A, A^c \in [n]^{(n/2)}$ together with all sets of larger size. Finally we can see that 2^{n-1} is best possible, since for any pair $A, A^c \in 2^{[n]}$ we can only ever pick at most one to be in our family.

The Erdős-Ko-Rado Theorem considers how large an intersecting family can be if we restrict to sets of fixed sizes, that is, given n and k, what is the size of the largest intersecting family $\mathcal{F} \subset [n]^{(k)}$?

We note first that if $k \geq n/2$ by the previous comments the question is not interesting. Also we note that there is a natural way to construct a family of sets which is intersecting, simply fix some $I \subset [n]$ and consider the family $\mathcal{F}_I = \{A \in [n]^{(k)} : I \subseteq A\}$. The size of this family is $\binom{n-|I|}{k-|I|}$, and so the largest such family is when I is just a single element. The conclusion of the Erdős-Ko-Rado Theorem is that this is best possible.

Theorem 2.3 (The Erdős-Ko-Rado Theorem)**.** *For any n and $k < n/2$, if $\mathcal{F} \subset [n]^{(k)}$ is an intersecting family, then*

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

Proof. The following probabilistic proof is due to Katona. Suppose we pick a set A uniformly at random from $[n]^{(k)}$ then clearly the probability of picking a member of our family $\mathcal{F}$ is

$$\mathbb{P}(A \in \mathcal{F}) = \frac{|\mathcal{F}|}{\binom{n}{k}}.$$

Let us estimate this probability in a different way. Suppose instead of picking A uniformly at random we first pick a random permutation $\sigma \in S_n$, and then pick a number $i \in [n]$ uniformly at random. We can associate with the pair (σ, i) a set

$$A(\sigma, i) = \{\sigma(i), \sigma(i+1), \ldots, \sigma(i+(k-1))\} \in [n]^{(k)},$$

where addition is performed modulo n. It is simple to check that picking $A = A(\sigma, i)$ by picking both σ and i uniformly at random is equivalent to picking A uniformly at random from $[n]^{(k)}$. However, we note that, for a fixed choice of σ, at most k of the sets $\{A(\sigma, i) : i \in [n]\}$ are in $\mathcal{F}$.

Indeed, suppose $A(\sigma, i) \in \mathcal{F}$. The only $A(\sigma, j)$ which intersect $A(\sigma, i)$ are those such that $j = i+t$ for $1 \leq |t| \leq k-1$. There are $2(k-1)$ such sets, but for any pair $\{A(\sigma, i+t), A(\sigma, i+t-k)\}$ at most one of the sets can be in $\mathcal{F}$, since they are disjoint. Therefore at most k are in $\mathcal{F}$.

Since this holds for each σ we see that, when we pick A in this manner,

$$\mathbb{P}(A \in \mathcal{F}) \leq \frac{k}{n}$$

Combining these two inequalities we see that

$$|\mathcal{F}| = \mathbb{P}(A \in \mathcal{F})\binom{n}{k} \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$$

□

This may seem a different idea to the preceding proof. Rather than picking an object at random and estimating the probability that it satisfies certain properties we have started with an unknown object (in this case an intersecting family), and estimated the probability that a randomly picked set is a member of this family. If we can bound above this probability this will

give us an upper bound on the size of the family, since the larger the family is, the more likely a randomly picked set is a member. However it can be phrased as another application of the same method. As another example we have the following simple proof using the same idea.

Let us suppose we have two families of sets $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ and $\mathcal{B} = \{B_1, B_2, \ldots, B_n\}$ satisfying the following properties

- $|A_i| = k$, $|B_i| = l$ for all $1 \leq i \leq n$;
- $A_i \cap B_i = \emptyset$ for all $1 \leq i \leq n$;
- $A_i \cap B_j \neq \emptyset$ for all $i \neq j$, $1 \leq i, j \leq n$.

How large can such a family be? This question is related to the *transversal number* of a set system.

Definition. A set T is a *transversal* of a set system $\mathcal{F}$ if $F \cap T \neq \emptyset$ for all $F \in \mathcal{F}$. The *transversal number* $\tau(\mathcal{F})$ is the size of the smallest transversal of $\mathcal{F}$.

Often in combinatorics when studying objects with certain properties it is useful to consider critical objects, those such that the removal of an element results in an object which no longer satisfies this property. If $\tau(\mathcal{A}) = l + 1$ then it is critically so if and only if there exists such a corresponding family $\mathcal{B}$, since if we remove any $A_i \in \mathcal{A}$, then B_i is a transversal of size l of the remaining set system. So a bound on the size of such families would give us a bound on the largest such critical $\mathcal{F}$.

As before there is a natural example to consider. If we let $\mathcal{A} = [k+l]^{(k)}$ and let $\mathcal{B}$ be the set of complements then it is simple to verify that that satisfy the two conditions, hence if we let $n(k, l)$ be the size of the largest n such that there exists such $\mathcal{A}$ and $\mathcal{B}$ we have that $n(k, l) \geq \binom{k+1}{k}$. The following proof of Bollobás shows this is best possible.

Theorem 2.4. *For any $k, l \geq 1$, $n(k, l) = \binom{k+l}{k}$*

Proof. Suppose $\mathcal{A}$ and $\mathcal{B}$ as above exist. Let $X = \bigcup_{i=1}^{n} (A_i \cup B_i)$ be the ground set. We pick an ordering of X uniformly at random and let U_i be the event "each element of A_i is smaller than every element of B_i". Since $|A_i| = k$ and $|B_i| = l$ we have that

$$\mathbb{P}(U_i) = \frac{1}{\binom{k+l}{k}}.$$

Also we note that U_i and U_j cannot happen simultaneously for $i \neq j$. Indeed, suppose it did happen. Then in the chosen ordering A_i totally preceded B_i and A_j totally precedes B_j. However, since $A_i \cap B_j \neq \emptyset \neq A_j \cap B_i$ we have there exists some element $x \in A_i \cap B_j$ and $y \in A_j \cap B_i$. No we see that x is not less than y, since A_j precedes B_j, but also y is not less than x, since A_i precedes B_i, a contradiction.

Therefore we see that, since the events are disjoint

$$\mathbb{P}\left(\bigcup_{i=1}^{n} U_i\right) = \sum_{i=1}^{n} \mathbb{P}(U_i) = \frac{n}{\binom{k+l}{k}},$$

and so

$$\frac{n}{\binom{k+l}{k}} \leq 1 \Rightarrow n \leq \binom{k+l}{k}.$$

□

3 The First Moment Method

Suppose we have a collection of random variables $X_1, \ldots, X_n$ and we are considering the random variable $X = \sum_i \lambda_i X_i$. The following is immediate from the linearity of integration.

Lemma 3.1. *Let $X_1, \ldots, X_n$ be random variables, $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, and $X = \sum_{i=1}^n \lambda_i X_i$. Then*

$$\mathbb{E}(X) = \sum_{i=1}^{n} \lambda_i \mathbb{E}(X_i).$$

This is particularly powerful because it places no restriction on dependence between the X_i. This makes it simple to calculate the expected value of many random variables by first decomposing them into the sum of simpler (often indicator) random variables.

Definition. For an event A in a probability space the*indicator random variable* I_A is defined by

- $I_A(\omega) = 1$ if $\omega \in A$
- $I_A(\omega) = 0$ if $\omega \notin A$

Note that for any indicator variables I_A, $\mathbb{E}(I_A) = \mathbb{P}(A)$. In many cases we calculate the expectation of a random variable X by expressing it as a sum of indicator variables, since if

$$X = \sum_{i=1}^{n} I_{A_i} \text{ then } \mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

by linearity of expectation.

We make use of this by using the fact that there must always exist a point in the probability space such that $X \geq \mathbb{E}(X)$ and another where $X \leq \mathbb{E}(X)$. This is essentially the same idea as in Section 2, since it is using the fact that, for a randomly picked object, $\mathbb{P}(X \geq \mathbb{E}(X)) > 0$ and $\mathbb{P}(X \leq \mathbb{E}(X)) > 0$, which is apparent from the definition of expectation.

Therefore, showing that the expected value of a random variable is small or large will guarantee that there is an object in the probability space on which this variable is small or large. Sometimes we will require a more general version of this, to show that it is unlikely that a random variable exceeds its expectation significantly.

Lemma 3.2. *[Markov's Inequality] Let X be a non-negative random variable and $a > 0$, then*

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

Proof. Consider the indicator random variable of the event that $X \geq a$, let us denote it by I. Since $I = 0$ if $X < a$ and X is non-negative we have that $aI \leq X$. Therefore

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

□

As simple example let us look again at our proof of a lower bound on the Ramsey number $R(k,k)$. We took a random graph $G(n,1/2)$ and tried to estimate the probability that it contained a clique of size k. If we denote by X the random variable that counts the number of cliques of size k in G then we see that $X = \sum_{A \subset V(G)\,:\,|A|=k} X_A$, where X_A is the indicator function of the event that the set $A \subset V(G)$ forms a clique. So by the linearity of expectation

$$\mathbb{E}(X) = \sum_{A \subset V(G)\,:\,|A|=k} \mathbb{E}(X_A) = \sum_{A \subset V(G)\,:\,|A|=k} \mathbb{P}(A \text{ is a clique}).$$

Since, as before, this probability is just $2^{-\binom{k}{2}}$, we have that $\mathbb{E}(X) = \binom{n}{k}2^{-\binom{k}{2}}$. Similarly if we let Y be the number of independent sets of size k in G we have that $\mathbb{E}(Y) = \binom{n}{k}2^{-\binom{k}{2}}$ as well. Hence, again by the linearity of expectation

$$\mathbb{E}(X+Y) = 2\binom{n}{k}2^{-\binom{k}{2}}.$$

If we choose n sufficiently large that $\mathbb{E}(X+Y) < 1$ then we know that there must exist some graph $G(n,1/2)$ in $\mathcal{G}(n,1/2)$ such that $X+Y \leq \mathbb{E}(X+Y) < 1$, however $X+Y$ is non-negative and takes only integer values, and so we must have that $X+Y=0$, that is the number of cliques or independent sets of size k is 0.

3.1 Hamiltonian Paths in a Tournament

Definition. A *tournament* is an orientation of a complete graph, that is a directed graph such that for every $u,v \in V(G)$ exactly one of (u,v) or $(v,u) \in E(G)$. A *Hamiltonian path* in a tournament is a directed path that meets every vertex.

This following result of Szele, from 1943, is often considered the first use of the probabilistic method.

Theorem 3.3. *There exists a tournament T on n vertices which has at least $n!/2^{n-1}$ Hamiltonian paths.*

Proof. We construct a random tournament T by choosing independently, for every pair of vertices $x,y \in [n]$, the edge (x,y) or (y,x) with probability $\frac{1}{2}$ each. Let X be the random variable which counts the number of Hamiltonian paths in T. For any given ordering of the permutation σ of $[n]$ let X_σ be the indicator variable of the event that $\sigma(1),\sigma(2),\ldots,\sigma(n)$ is a Hamiltonian path in T. We have that, since the orientation of each edge is chosen independently,

$$\mathbb{E}(X_\sigma) = \mathbb{P}\left((\sigma(i),\sigma(i+1)) \in T \text{ for } 1 \leq i \leq n-1\right) = \prod_{i=1}^{n-1} \mathbb{P}\left((\sigma(i),\sigma(i+1)) \in T\right) = \frac{1}{2^{n-1}}.$$

However $X = \sum_\sigma X_\sigma$, since σ ranges over all possible paths. Therefore

$$\mathbb{E}(X) = \sum_\sigma \mathbb{E}(X_\sigma) = \frac{n!}{2^{n-1}}.$$

Hence, the exists some tournament T where the value of X is a least $\mathbb{E}(X)$, that is, in this tournament, there are at least $n!/2^{n-1}$ Hamiltonian paths. □

Szele also showed that the maximum number of Hamiltonian paths in a tournament is $O(n!/2^{\frac{3}{4}n})$. These bounds were not improved for almost 60 years, until Alon showed that the upper bound could be improved to $O(n^{\frac{3}{2}}n!/2^{2n})$, also using probabilistic methods.

3.2 Turán's Theorem

We start with another application of the first moment method to bound the size of the largest independent set in a graph G.

Lemma 3.4. *Let G be a graph and, for each $v \in V(G)$, let $d(v)$ be the degree of v. Then G contains an independent set of size at least*

$$\sum_{x \in V} \frac{1}{d(v)+1}$$

Proof. We pick an ordering, $<$, uniformly at random from the $|V|!$ orderings of the set V. Let

$$I_< = \{x \in V \,:\, x < y \text{ for all } y \in N(v)\}$$

be the set of all vertices which are the smallest in their neighbourhoods. We first note that $I_<$ is an independent set. Indeed, if $v_1, v_2 \in I$ and $(v_1, v_2) \in E(G)$ then, by the definition of $I_<$ we must have $v_1 < v_2$ and $v_2 < v_1$, a contradiction. We let $X = |I_<|$ and for each $v \in V$ let X_v be the indicator random variable of the event that $v \in I_<$, and note that, as before, $X = \sum_{v \in V} X_v$.

We have that $\mathbb{E}(X_v) = \mathbb{P}(v \in I)$ and, since for a randomly picked ordering each vertex in $v \cup N(v)$ is equally as likely to be the smallest,

$$\mathbb{P}(v \in I) = \frac{1}{d(v)+1}.$$

Hence

$$\mathbb{E}(X) = \sum_{v \in V} \mathbb{E}(X_v) = \sum_{v \in V} \frac{1}{d(v)+1}.$$

So there must exist some ordering $<$ such that

$$|I_<| \geq \sum_{x \in V} \frac{1}{d(v)+1},$$

which is the desired independent set. □

Corollary 3.5. *Let G be a graph on n vertices. Then G contains a clique of size at least*

$$\sum_{x \in V} \frac{1}{n - d(v)}$$

Proof. Apply Lemma 3.4 to the complement of G. □

It is simple to deduce (a weak form of) Turán's Theorem from this corollary.

Theorem 3.6 (Turán's Theorem)**.** *Let G be a graph on n vertices such that $K_r \not\subset G$, then*

$$e(G) \leq \frac{(r-2)n^2}{2(r-1)}.$$

Remark 3.7. *Note that when $(r-1)|n$ we have that the Turán graph $T(n,r)$ has $(r-1)$ classes of size $n/(r-1)$ and so has exactly*

$$\frac{1}{2} \times n \times (r-2) \times \frac{n}{r-1} = \frac{(r-2)n^2}{2(r-1)}$$

edges.

Proof. Let k be the size of the largest clique in G, we have by corollary 3.5

$$r - 1 \geq k \geq \sum_{x \in V} \frac{1}{n - d(v)}.$$

Since $1/(n-x)$ is concave (as a function of x), we have by Jenson's inequality that, if d is the average degree in G,

$$\sum_{x \in V} \frac{1}{n - d(v)} \geq \frac{n}{n-d} = \frac{n^2}{n^2 - 2e(G)}.$$

So, by combining the two inequality, we see that

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

Re-arranging for $e(G)$ gives that

$$e(G) \leq \frac{(r-2)n^2}{2(r-1)}.$$

□

3.3 Crossing Number of Graphs

Given a graph $G = (V, E)$ an *embedding* of G into the plane is a planar representation of G, where each vertex is represented by a point, and each edge from u to v is represented by a curve between the points represented by u and v. The *crossing number* of an embedding of a graph is the number of pairs of curves which intersect, which do not share endpoints. The *crossing number* of G, $\mathrm{cr}(G)$ is the minimal crossing number of a planar embedding of G.

For example, G is a planar graph if and only if it has crossing number $\mathrm{cr}(G) = 0$. We can use the probabilistic method to get a good lower bound on the crossing number of the graph.

Theorem 3.8. *Let G be a graph such that $|E| \geq 4|V|$, then*

$$cr(G) \geq \frac{|E|^3}{64|V|^2}.$$

Proof. Suppose we take an planar embedding of a graph with n vertices and m edges. The set of edges which don't cross any other edges forms a planar graph, and so by Euler's formula there are at most $3n - 6 \leq 3n$ such edges. Therefore, it follows that the crossing number of G is at least $m - 3n$.

Let us take an embedding of G with exactly $\mathrm{cr}(G) = t$ crossing, and let's take a random subgraph H of G formed by picking each vertex of G to be in H independently with probability p, where we will pick p later. The expected number of vertices in H is $p|V|$ and the expected number of edges is $p^2|E|$.

However the embedding of G induces an embedding of H in the plane, and the only crossing points of H in this embedding will be crossing points of G. Moreso, the expected number of crossing points in this embedding of H is exactly $p^4 t$, since for a crossing point to appear we need all four endpoints to be in H.

However, since for each subgraph H we have that $\mathrm{cr}(H) \geq |E(H)| - 3|V(H)|$, and the number of crossings in the embedding given by G is at least the crossing number, we have that

$$p^4 t \geq p^2|E| - 3p|V|.$$

We let $p = 4|V|/|E|(\leq 1)$ and see that

$$\begin{aligned} t \geq \frac{1}{p^4}(p^2|E| - 3p|V|) &= \frac{|E|}{p^2} - 3\frac{|V|}{p^3} \\ &= \frac{|E|^3}{16|V|^2} - 3\frac{|E|^3}{64|V|^2} = \frac{|E|^3}{64|V|^2}. \end{aligned}$$

□

4 Alterations

Sometimes it is the case that the first attempt to find a 'good' object via a random construction fails, but what we find is an object which *almost* satisfies our conditions. Often it is possible to then deterministically modify this object to get what we need.

4.1 Ramsey Numbers Again

Let us again give an example of this by estimating a lower bound for the diagonal Ramsey numbers, $R(k,k)$

Theorem 4.1. *For any integer n,*

$$R(k,k) > n - 2.\binom{n}{k}2^{-\binom{k}{2}}.$$

Proof. Let us choose a graph from $\mathcal{G}(n,1/2)$. As before let X be the random variable which counts the number of cliques of size k and Y be the random variable which counts the number of independent sets of size k. We want to estimate the number of 'bad' k-sets, $X+Y$.

We know that $\mathbb{E}(X+Y) = 2.\binom{n}{k}2^{-\binom{k}{2}}$. In Theorem 2.1 we chose n such that $\mathbb{E}(X+Y) < 1$, to guarantee the existence of a graph with no clique or independent set of size k.

However, for any n, we know that there must exists a graph $G \in G(n,p)$ with $X+Y \leq \mathbb{E}(X+Y)$, that is with at most the expected number of 'bad' k-sets. We define a graph $H \subset G$ as follows: for each k-clique and independent set of size k in G pick a vertex and delete it. Clearly H does not contains any clique or independent set of size k. However H has at least $n - (X+Y) \geq n - \mathbb{E}(X+Y) = n - 2.\binom{n}{k}2^{-\binom{k}{2}}$ vertices. □

So by picking an appropriate n we might hope to get a better bound then before (note that when we apply Theorem 4.1 to the smallest n such that $2.\binom{n}{k}2^{-\binom{k}{2}} < 1$ we get the same bound as before, so we can only do better). We have that

$$n - 2.\binom{n}{k}2^{-\binom{k}{2}} \geq n - 2.\left(\frac{en}{k}\right)^k 2^{-\frac{k(k-1)}{2}},$$

and so by differentiation the maximum should appear when

$$1 = 2.kn^{k-1}\frac{e^k 2^{-\frac{k(k-1)}{2}}}{k^k},$$

which is when

$$n \sim \frac{k}{e}2^{\frac{k}{2}}.$$

In fact, if we're careful and use Stirling's approximation we can show this optimal. Plugging this back into Theorem 4.1 gives

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

This is an improvement of our bound from Theorem 2.1 of $2^{\frac{k}{2}-1}$, however a more careful analysis of that proof would give a bound that only differs from this one in a factor of $\frac{1}{\sqrt{2}}$, not much of an improvement at all.

We can also get similar bounds for the off diagonal Ramsey numbers in the same way.

Theorem 4.2. *For all integers n and $p \in [0,1]$*

$$R(k,l) > n - \binom{n}{k}p^{\binom{k}{2}} - \binom{n}{l}(1-p)^{\binom{l}{2}}.$$

The proof is the same as Theorem 4.1, however instead of picking a graph from $\mathcal{G}(n,1/2)$ we do so from $\mathcal{G}(n,p)$. The two terms in the theorem are now the expected number of cliques of size k and independent sets of size l, and hence by the alteration method there exists a graph with at most the expected number of such sets. Deleting a vertex from each such gives the desired graph witnessing the bound on $R(k,l)$. Let us compare this to the bound we get from a straightforward application of the random method.

Theorem 4.3. *If there exists an integer n and $p \in [0,1]$ such that*

$$\binom{n}{k}p^{\binom{k}{2}} + \binom{n}{l}(1-p)^{\binom{l}{2}} < 1$$

Then $R(k,l) > n$.

As before, any n and p which give a bound in Theorem 4.3 give the same bound in Theorem 4.2. The precise asymptotics of these two expressions can get fairly complex, we will see an example in the exercise class that Theorem 4.2 gives the bound

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

4.2 Graphs of High Girth and High Chromatic Numbers

A natural question to ask about many graph parameters that are defined globally, is to what extent local restrictions can affect them. For example let us consider the chromatic number of a graph, $\chi(G)$.

Definition. A *k-colouring* of a graph G is a function $f : V(G) \to [k]$ such that $(v,w) \in E(G) \Rightarrow f(v) \neq f(w)$. The *chromatic number* of G, $\chi(G)$, is the smallest k such that a k-colouring exists.

Suppose we know that in a local sense the chromatic number is small, can the graph still have large chromatic number?

Recall that the *girth* of a graph G, $g(G)$, is the size of the shortest cycle in G. If a graph has no short cycles of odd length (which is certainly true if it contains no short cycles) then locally the graph looks bipartite. In particular if the girth of G is larger than g then we can properly 2-colour all the vertices within distance g of any specific vertex. Therefore if graphs exists with arbitrarily high girth and chromatic number, we know that we can't deduce anything about the chromatic number of a graph from the chromatic numbers of its subgraphs of a fixed size.

Theorem 4.4. *For any $k, l > 0$, there exists a graph G such that $\chi(G) > k$ and $g(G) > l$.*

Proof. We want to prove this using the random method. So we want to pick a graph $G(n,p)$ for some appropriately chosen n, p and show that with positive probability it has high girth and cromatic number. The chromatic number seems difficult to calculate for a random graph, but we can bound it very naturally by the independence number of the graph, $\chi(G) \geq n/\alpha(G)$, since each colour class must form an independent set.

So our idea is to look at the random variables X, which counts the number of cycles in G with length $\leq l$, and Y, which counts the number of independent sets of size n/k. Note that if we have no independent sets of size n/k, we also don't have any larger ones.

As before if we could show that, for some n and p the expected value of $X + Y$ is less than 1, then we would know there existed a graph on n vertices satisfying our conditions.

However a bit of thought suggests this is unlikely to work. A random graph is likely to have lots of triangles if $p >> 1/n$. Indeed the expected number of triangles is $\binom{n}{3}p^3 \sim (np)^3$, so if we want $\mathbb{E}(X) < 1$ we need $p = O(1/n)$. However in this range of probability we expect many big independent sets. Indeed the expected number of independent sets of size a is, when $p < 1/2$

$$\binom{n}{a}(1-p)^{\binom{a}{2}} \geq \left(\frac{n}{a}\right)^a (1-p)^{a^2} \geq \left(\frac{n}{a}\right)^a e^{-2pa^2}.$$

So if $p = O(1/n)$ and we look at independent sets of size n/k we see that the expected number is at least

$$(k)^{\frac{n}{k}} e^{-\frac{2n}{k^2}},$$

which for general k will tend to infinity.

However if we pick p to be just more than $1/n$ we will be able to show that this term tends to 0, and whilst the first term will no longer be small, we will be able to show it is still small compared to n.

We can then use Markov's inequality to say that the probability that both X and Y are far away from their averages is small, and so conclude there is a graph G with no large independent sets (and so large chromatic number) and less than $n/2$ short cycles. We would then like to remove a vertex from each of the short cycles to get a graph with large girth and chromatic number. However we have to be a little careful here, since removing vertices from the graph will change the bound we get on the chromatic number from the independence number. With this in mind we will be slightly overzealous and instead of Y consider the random variable Y' which counts the number of independent sets of size $n/2k$.

Now suppose we let $p = n^{\epsilon - 1}$ for some small $\epsilon > 0$, how many cycles of length i do we expect for each i? Well there are at most n^i potential (ordered) vertex sets for the cycles, and for each of those vertex sets the probability that they form a cycle is p^i. Therefore the expected number of cycles of length $\leq l$

$$\mathbb{E}(X) \leq \sum_{i=3}^{l} n^i p^i = \sum_{i=3}^{l} (n.n^{\epsilon-1})^i = \sum_{i=3}^{l} n^{\epsilon i}$$

So if we let $\epsilon < 1/l$ we have that $\mathbb{E}(X) = o(n)$ and so, by Markov's inequality we have that, for

large enough n,

$$\mathbb{P}(X \geq n/2) \leq \frac{\mathbb{E}(X)}{\frac{n}{2}} < \frac{1}{2}.$$

How does this improvement of ϵ affect $\mathbb{E}(Y')$? Well, with $a = n/2k$ we have that

$$\begin{aligned}\mathbb{E}(Y') &= \binom{n}{a}(1-p)^{\binom{a}{2}} \leq n^a e^{-p\frac{a(a-1)}{2}} \\ &= e^{a\log(n)} e^{-p\frac{a(a-1)}{2}} = e^{a(\log(n) - p\frac{(a-1)}{2})}\end{aligned}$$

However, since $a = n/2k$ and $p = n^{\epsilon-1}$ for some $\epsilon > 0$, we have that $\log(n) - p(a-1)/2 \to -\infty$ and hence, as $n \to \infty$, $\mathbb{E}(Y') \to 0$. Therefore, for large enough n, $\mathbb{E}(Y') < 1/2$ and so by Markov's inequality

$$\mathbb{P}(Y' \geq 1) < 1/2.$$

Therefore, by the union bound, there is a non-zero probability that there exists a graph $G(n,p)$ such that $Y' = 0$ and $X < n/2$. If we remove one vertex from each of the cycles of length less than l in G we produce a graph G' with the following properties

- $|G'| \geq n/2 > 0$;
- $g(G') > l$;
- $\alpha(G') < n/2k$ and so $\chi(G') \geq |G'|/\alpha(G') \geq k$.

□

5 Dependent random choice

Recently a technique which is based on a simple application of the alteration method has been used in various contexts, normally to do with embedding sparse graphs. In this chapter we give a short overview of the method, which is known as dependent random choice, and a few examples of applications.

The basic idea can be summarised as follows: We would like to find, in a dense graph (that is a graph with large minimum/average degree), a set of vertices U such that every small subset of U has many common neighbours. To do this, we first pick a small set of vertices T at random from the graph and let U' be the set of common neighbours of T. Intuitively, if we have some subset of G with not many common neighbours, then it is unlikely that all the members of T will lie in this set of common neighbours, and hence it is unlikely to be a subset of U'. Therefore the expected number of 'bad' subsets in U' will be small and so by removing a small number of vertices, one from each 'bad' set, we should find a set U with the desired properties.

Lemma 5.1. *Let G be a graph with $|G| = n$ and let $d = 2|E(G)|/n$ be the average degree of G. If there exist positive integers t, a, m, r 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 subset $R \subset U$ of size $|R| = r$ has at least m common neighbours.

Proof. For any set $X \subset V$ let $\Gamma(X) = \{v \in V : v \in N(x) \text{ for all } x \in X\}$ be the set of common neighbours of X. We pick a set of vertices T uniformly at random from V^t, that is with repetition. Let $A = \Gamma(T)$ be the set of common neighbours of T, and let X be the random variable with counts the size of A. for any vertex v, the probability that $v \in A$ is the probability that every element of T is a neighbour of v and so, by linearity of expectation

$$\mathbb{E}(X) = \sum_{v \in V}\left(\frac{|N(v)|}{n}\right)^t = n^{-t}\sum_{v \in V}|N(v)|^t.$$

Since the function $f(x) = x^t$ is convex, we can use Jensen's inequality to say

$$\mathbb{E}(X) \geq n^{-t}.n\left(\frac{\sum_{v \in V}|N(v)|}{n}\right)^t = n^{1-t}\left(\frac{2|E(G)|}{n}\right)^t = \frac{d^t}{n^{t-1}}.$$

Let Y be the random variable which counts the number of subset $R \subset A$ of size r with fewer than m common neighbours. For any $R \subset V$, let $\Gamma(R)$ be the set of common neighbours of R, then the probability that R is a subset of A is just

$$\left(\frac{|\Gamma(R)|}{n}\right)^t.$$

Therefore, if we let $\mathcal{R} = \{R \subset G : |R| = r \text{ and } |\Gamma(R)| < m\}$ be the set of 'bad' subset of V, we have that

$$\mathbb{E}(Y) = \sum_{R \in \mathcal{R}}\mathbb{P}(R \subset A) = \sum_{R \in \mathcal{R}}\left(\frac{|\Gamma(R)|}{n}\right)^t < \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

Therefore we have that

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

Therefore there exists a choice of T for which $X - Y \geq a$. We delete one vertex from each subset R of $\Gamma(T)$ of size r with fewer than m common neighbours. Let U be the remaining subset of $\Gamma(T)$. We have that $|U| = X - Y \geq a$ and by construction every subset of U of size r has at least m common neighbours. □

Once we have a large set U such that every small subset has many common neighbours we can embed bipartite graphs in it in the following way

Lemma 5.2. *Let G be a graph, a, m, r be positive integers and suppose there exists a subset $U \subset V(G)$ of at least a vertices such that every subset $R \subset U$ of size r has at least m common neighbours.*

If H is a bipartite graph on vertex sets A and B such that $|H| \leq m$, $|A| \leq a$ and every vertex in B has degree at most r, then H is a subgraph of G.

Proof. We wish to find an embedding of H in G given by an injective function $\phi : V(H) \to V(G)$. We start by picking an injective function $\phi : A \to U$ arbitrarily, which is possible since $|U| \geq a \geq |A|$.

We label the vertices of B as $v_1, v_2, \ldots, v_b$ and try to embed them in this order one at a time. Suppose we have already defined $\phi(v_i)$ for all $i < j$ and we wish to embed v_j. Let $N_j \subset A$ be the neighbourhood of v_j, so $|N_j| \leq r$. Since $\phi(N_j)$ is a subset of U of size at most r, there are at least m vertices in G adjacent to all the vertices in $\phi(N_j)$. Since the total number of vertices embedded already is less than $|H| \leq m$, there is at least one vertex $w \in G$ which has not been used in the embedding and is adjacent to all the vertices in $\phi(N_j)$. We set $\phi(v_j) = w$.

After we have embedded every v_b it follows that ϕ is the desired embedding of H as a subgraph of G. □

5.1 Turán Numbers of Bipartite Graphs

The abundance of variables in Lemma 5.1 make it difficult to understand exactly what's going on, so let's look at an example of an application. For a graph H and an integer n, the Turán number $\mathrm{ex}(n, H)$ denotes the maximum number of edges in a graph on n vertices which does not contain H as a subgraph. Turán's theorem determines this number precisely for complete graphs $H = K_r$, and the asymptotic behaviour for graphs of chromatic number at least 3 is given by the well known result of Erdős and Stone

Theorem 5.3 (The Erdős-Stone Theorem)**.** *For any graph H with $\chi(H) \geq 3$*

$$ex(n, H) = \left(1 - \frac{1}{\chi(H)-1} + o(1)\right)\binom{n}{2}$$

For bipartite graphs the situation is much more complicated, and there are relatively few non-trivial bipartite H for which the order of magnitude of $\mathrm{ex}(n, H)$ is known. The following

result gives a bound for the Turán number of bipartite graphs in which one vertex class has bounded degree.

Theorem 5.4. *Let H be a bipartite graph on vertex sets A and B such that all vertices in B have degree at most r. Then there exists some constant $c = C(H)$ such that*

$$ex(n, H) \leq cn^{2-\frac{1}{r}}.$$

Proof. Let $a = |A|$ and $b = |B|$. The idea is, given a graph G with $|V(G)| = n$ and $e(G) \geq cn^{2-1/r}$, to use Lemma 5.1 to find a subset $U \subset V(G)$ of size at least a in which all the subsets of size r have at least $a + b$ common neighbours.

So let us check that the required bound holds in Lemma 5.1. We let $m = a + b$, $t = r$ and (for reasons which will become clear) let $c = \max\left(a^{1/r}, \frac{e(a+b)}{r}\right)$, note that c depends only on H. Given a graph G with $|V(G)| = n$ and $e(G) \geq cn^{2-1/r}$, the average degree of G satisfies $d \geq 2cn^{1-1/r}$. Therefore

$$\begin{aligned}\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t &\geq (2c)^r - \left(\frac{en}{r}\right)^r \left(\frac{a+b}{n}\right)^r \\ &\geq (2c)^r - \left(\frac{e(a+b)}{r}\right)^r \geq c^r \\ &\geq a.\end{aligned}$$

Therefore by Lemma 5.1 there exists a subset U of $V(G)$ of size at least a in which all the subsets of size r have at least $a + b$ common neighbours. Hence, by Lemma 5.2 H is a subgraph of G. □

These bounds are best possible in terms of their dependence on r. Indeed it is known the the Turán number of the complete bipartite graphs $K_{t,r}$ when $t \geq (r-1)!$ is $\Omega(n^{2-\frac{1}{t}})$.

5.2 The Ramsey Number of the Cube

Definition. The *Ramsey number* of an arbitrary graph H is

$$r(H) = \min\{n : \text{ Every 2 colouring of } K_n \text{ contains a monochromatic copy of } H\}.$$

For example in this language we have that $r(K_k) = r(k, k)$. The *r-dimensional Hypercube*, $\mathcal{Q}_r$, is a graph with vertex set $\{0, 1\}^r$ where two vertices are adjacent if and only if they differ in exactly one coordinate.

An old conjecture of Burr and Erdős is that the Ramsey number of the cube is linear in the number of vertices, that is there exists some constant C such that $r(\mathcal{Q}_r) \leq C2^r$. Early bounds were much worse than this, for example Beck showed that $r(\mathcal{Q}_r) \leq 2^{Cr^2}$. More recently Shi obtained the first bound which was polynomial in the number of vertices, showing that $r(\mathcal{Q}_r) \leq 2^{Cr+o(r)}$ for some $C \sim 2.618$. Lemma 5.1 easily implies a, slightly worse, polynomial bound on $r(\mathcal{Q}_r)$.

Theorem 5.5.

$$r(\mathcal{Q}_r) \leq 2^{3r}$$

Proof. Let $n = 2^{3r}$. Given any two colouring of K_n, one of the colour classes contains at least half the edges. Let G be the graph of this colour.

Since $\mathcal{Q}_r$ is a bipartite graph, with vertex sets of size 2^{r-1} and maximum degree r, we would like to use Lemma 5.1 as before to find a set U of size at least 2^{r-1} such that every set of r vertices has at least 2^r common neighbours.

So let us set $a = 2^{r-1}$, $m = 2^r$ and $r = r$. We want to choose an appropriate t such that

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

where d is the average degree of G. Note that, since G has at least half of the edges of K_n we have that

$$d \geq 2\frac{e(G)}{n} \geq \frac{1}{2}(n-1) \geq 2^{-c}n,$$

for an appropriately chosen $c > 1$. Now

$$\begin{aligned}\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t &\geq 2^{-ct}n - \frac{n^r}{r!}\frac{m^t}{n^t} \\ &= 2^{-ct}n - \frac{n^{r-t}m^t}{r!} \\ &= 2^{3r-ct} - \frac{2^{3r^2-2rt}}{r!}.\end{aligned}$$

Setting $t = \frac{3}{2}r$ makes the second term negligible and so, to make the first term large we need $3r - c\frac{3}{2}r \geq r$, so for example we can take $c = 4/3$. All together this gives us

$$\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq 2^r - \frac{1}{r!} \geq 2^{r-1} = a.$$

Hence, as before, we can use Lemma 5.2 to say that $\mathcal{Q}_r$ is a subgraph of G, that is, in any 2-colouring of K_n, the largest colour class will contain a subgraph isomorphic to $\mathcal{Q}_r$. Therefore $r(\mathcal{Q}_r) \geq n = 2^{3r}$. □

5.3 Improvements

Lemma 5.1 tells us that in any sufficiently dense graph on n vertices we can find a large set of vertices U such that every small subset has many common neighbours. For many applications it would be useful to have both the size of U and the number of common neighbours to be linear in n, for example if we wished to prove that the Ramsey number of the cube was linear in the number of vertices using the same method.

However one can construct graphs with average degree just less than $n/2$ such that any linear size subset of the vertices contains a small subset (in fact even a pair of vertices) with $o(n)$ common neighbours.

However using a similar proof based on alterations one can prove that in every dense graph there exists a subset U of linear size in which almost every small subset has linearly many common neighbours.

Lemma 5.6. *Let $\epsilon > 0$, $r \leq n$ be positive integers, and G a graph on $N > 4r\epsilon^{-r}n$ vertices with at least $\epsilon\frac{N^2}{2}$ edges. Then there is a subset $U \subset V(G)$ with $|U| > 2n$ such that number of subsets $S \subset U$ with $|S| = r$ and less than n common neighbours is at most*

$$\frac{1}{(2r)^r}\binom{|U|}{r}.$$

How might this be useful? Well if we think about that proof of Lemma 5.2 given a set U such that every small subset had many common neighbours, we embedded a bipartite graph H by arbitrarily embedding the left hand side in U, and that verifying that we can always extend that to an embedding of H, using the fact that when we want to embed a vertex v on the right hand side, the image of it's neighbourhood is a small set in U, and so has many common neighbours which are all candidates for the image of v.

If we were more careful in how we embedded the left hand side of H into U at the beginning, then if sufficiently few of the small sets in U don't have many common neighbours, we could try to embed the left hand side in such a way that none of these 'bad' small sets appear as neighbourhoods of things in the right hand side of H. We could then extend this to an embedding of H as before.

Obviously this will require some slightly stronger conditions on the graphs H we consider. The specific numbers in this lemma have been chosen so that an analogy of the embedding lemma (Lemma 5.2) carries over in this way for graphs with $\Delta(H) \leq r$. Using this one can improve on the previous bound to

Theorem 5.7.

$$r(\mathcal{Q}_r) \leq r2^{2r+3} \leq 2^{2r+o(r)}.$$

6 The Second Moment Method

6.1 Variance and Chebyshev's Inequality

Markov's inequality tells us that, for a non-negative random variable X, if the expectation of X gets small, then it's very likely that X is small. What about if the expectation of X is large, can we say that it's very likely that X is large? Clearly in general this is not true, we can take a random variable that is almost always 0, except it takes the values N^2 with probability $1/N$ for some large N. The expectation is arbitrarily large, and yet it's very likely that X is small.

With this in mind we introduce the concept of *variance* which can be thought of as a measure of how close to its expectation we expect a random variable to be.

Definition. The *variance* of a random variable X is

$$\mathrm{Var}(X) := \mathbb{E}\big((X - \mathbb{E}(X))^2 \big) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2,$$

where the first equality is the definition, and the second follows from linearity of expectation.

So $\mathrm{Var}(X)$ is the expected value of the square of the difference between X and its average. Here we take the square since we only care about the magnitude of the difference between X and its average, not the sign, but working with absolute values is much more difficult.

Unlike $\mathbb{E}(X)$, the variance is not a linear operator. If we want to calculate the variance of a sum of random variables we need to know something about their pairwise dependence. As a example suppose we have two random variables X and Y, we can calculate the variance of $X + Y$ in terms of X and Y directly from the definition using the linearity of expectation.

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}\big((X+Y)^2\big) - \big(\mathbb{E}(X+Y)\big)^2 \\
&= \mathbb{E}(X^2 + 2XY + Y^2) - \big(\mathbb{E}(X) + \mathbb{E}(Y)\big)^2 \\
&= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \big(\mathbb{E}(X)\big)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - \big(\mathbb{E}(Y)\big)^2 \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))
\end{aligned}$$

This motivates the following definition.

Definition. The *covariance* of two random variables X and Y is

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

Lemma 6.1. *Given a sequence of random variables $X_1, X_2, \ldots, X_n$, let $X = \sum_i X_i$. Then*

$$Var(X) = \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq k} Cov(X_i, X_j).$$

Proof.

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \mathbb{E}\left(\left(\sum_i X_i\right)^2\right) - \left(\mathbb{E}\left(\sum_i X_i\right)\right)^2 \\ &= \sum_i \mathbb{E}\left(X_i^2\right) + \sum_{i\neq j} \mathbb{E}(X_i X_j) - \sum_i \left(\mathbb{E}(X_i)\right)^2 - \sum_{i\neq j} \mathbb{E}(X_i)\mathbb{E}(X_j) \\ &= \sum_i \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j).\end{aligned}$$

□

Note that if X and Y are independent then $\mathrm{Cov}(X, Y) = 0$, but be careful to remember that the opposite is not true. If the variance of the random variable is small we will expect it to be quite likely that the random variable takes values near it's mean, since the expected deviation is low. The following inequality of Chebyshev formalises this idea.

Lemma 6.2 (Chebyshev's Inequality)**.** *Let X be a random variable with $Var(X) < \infty$. Then for any $t > 0$*

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

Proof. We apply Markov's inequality to the non-negative random variable $(X - \mathbb{E}(X))^2$. Since $\mathbb{E}((X - \mathbb{E}(X))^2) = \mathrm{Var}(X)$ we have that

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

□

It is not too hard to produce simple random variables where Lemma 6.2 is best possible.

6.2 Threshold Functions

Given a random graph $G(n, p)$ and an arbitrary graph H, a natural question to consider is what is the probability that H appears as a subgraph of G. For example let us consider the triangle, K_3. We note that the property of containing a triangle as a subgraph is a *increasing property* of graphs; that means, if it holds for a graph G and $G \subset G'$, it also holds for G'. We can similarly define a *decreasing property* of graphs, and we call a property that is either decreasing or increasing a *monotone property*.

It seems likely that, for very small p, $G(n, p)$ will almost surely contain no triangles, whereas for large p, the appearance of a triangle is very likely.

Let T be the random variable which counts the number of triangles in $G(n, p)$. For any given triple of vertices, the probability that they form a triangle is p^3. Hence, by linearity of expectation, the expected number of triangles is

$$\mathbb{E}(T) = \binom{n}{3} p^3 \sim n^3 p^3.$$

This will tend to 0 if $p(n) = o(1/n)$. Hence, by Markov's inequality, if $p(n) = o(1/n)$ we have that

$$\mathbb{P}(K_3 \subset G(n,p)) = \mathbb{P}(T \geq 1) \leq \mathbb{E}(T) \to 0,$$

and so the probability that $G(n,p)$ contains a triangle tends to zero as well.

On the other hand suppose that $p = \omega(1/n)$. We have that $\mathbb{E}(T) \to \infty$ and so we would like to be able to conclude that G will almost surely contain at least 1 triangle. As we mentioned before, the fact that the expected number of triangles grows arbitrarily large is not sufficient to conclude the probability of containing one is large, to do so we must prove that the variance is also small (compared to the expectation). The standard method to do so, for a non-negative random variable X, is to apply Chebyshev's inequality with $t = \mathbb{E}(X)$.

Lemma 6.3. *Let $X_1, X_2, \ldots$ be a sequence of non-negative random variables such that*

$$\frac{Var(X_n)}{\mathbb{E}(X_n)^2} \to 0.$$

Then

$$\mathbb{P}(X_n > 0) \to 1.$$

Proof. We let $t = \mathbb{E}(X_n)$ and apply Chebyshev's inequality to see that

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

Hence have that

$$\mathbb{P}(X_n = 0) \to 0,$$

from which the claim follows. □

So if we can show that, when $p = \omega(1/n)$, $\mathrm{Var}(T)/(\mathbb{E}(T))^2 \to 0$, we would know that the probability that $G(n,p)$ doesn't contain a triangle will tend to 0. (Strictly here we are thinking of T as a sequence of random variables, one for each $n \in \mathbb{N}$).

So we need to calculate $\mathrm{Var}(T)$. When a random variable X can be written as a sum of indicator random variable of a set of events $\mathcal{A} \subset \Sigma$, that is

$$X = \sum_{A \in \mathcal{A}} I_A,$$

then it is possible to express the variance of X in a simple way as a sum. We note that $\mathbb{E}(I_A) = \mathbb{P}(A)$, and that $(I_A)^2 = I_A$ for all A, and so

$$\mathrm{Var}(I_A) = \mathbb{E}((I_A)^2) - (\mathbb{E}(I_A))^2 = \mathbb{P}(A) - \mathbb{P}(A)^2 = \mathbb{P}(A)(1 - \mathbb{P}(A)).$$

Also, it is a simple check that

$$\mathrm{Cov}(I_A, I_B) = \mathbb{E}(I_A I_B) - \mathbb{E}(I_A)\mathbb{E}(I_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(\mathbb{P}(B|A) - \mathbb{P}(B)).$$

Therefore by Lemma 6.1 we see that

$$\mathrm{Var}(X) = \sum_{A \in A} \mathbb{P}(A)(1 - \mathbb{P}(A)) + \sum_{A \neq B} \mathbb{P}(A)(\mathbb{P}(B|A) - \mathbb{P}(B)) = \sum_{A \in \mathcal{A}} \mathbb{P}(A) \left(\sum_{B \in \mathcal{A}} \mathbb{P}(B|A) - \mathbb{P}(B) \right).$$

This can be a useful equality to remember. For example T is the sum of the indicator random variables $I_{\{A \text{ is a triangle}\}}$ over $A \subset [n]$ such that $|A| = 3$. So to evaluate $\mathrm{Var}(T)$ we can use the above sum.

For any particular A we have to split the inner term into cases depending on $|A \cap B|$. When $|A \cap B| \leq 1$, then A and B do not share an edge, and so we have that $\mathbb{P}(B|A) = \mathbb{P}(B)$. Hence, there is no contribution to the sum from these pairs.

When $|A \cap B| = 2$ the two triangles share an edge, and so $\mathbb{P}(B|A) - \mathbb{P}(B) = p^2 - p^3$. For any particular A there are $3(n-3)$ such B, since for each of the 3 edges in A we have $n-3$ remaining choices for the third vertex in B.

Finally if $|A \cap B| = 3$ (that is, $A = B$) we have that $\mathbb{P}(B|A) - \mathbb{P}(B) = 1 - p^3$, and for each A there is exactly 1 such B. Hence we see that

$$\begin{aligned}
\mathrm{Var}(T) &= \sum_{A \in [n]^{(3)}} \mathbb{P}(A \text{ is a triangle}) \left(\sum_{B \in [n]^{(3)}} \mathbb{P}(B|A) - \mathbb{P}(B) \right) \\
&= \sum_{A \in \binom{[n]}{3}} \mathbb{P}(A \text{ is a triangle}) \left(3(n-3)(p^2 - p^3) + (1 - p^3) \right) \\
&= \binom{n}{3} p^3 \left(3(n-3)(p^2 - p^3) + (1 - p^3) \right).
\end{aligned}$$

So we can now say that

$$\begin{aligned}
\frac{\mathrm{Var}(T)}{\mathbb{E}(T)^2} &= \frac{\binom{n}{3} p^3 \left(3(n-3)(p^2 - p^3) + (1 - p^3) \right)}{\left(\binom{n}{3} p^3 \right)^2} \\
&= \frac{\left(3(n-3)(p^2 - p^3) + (1 - p^3) \right)}{\binom{n}{3} p^3} \\
&\leq \frac{\left(3n(p^2) + 1 \right)}{\frac{(np)^3}{12}} \leq \frac{\left(36(np)^2 + 12 \right)}{(np)^3} \\
&= \frac{36}{np} + \frac{12}{(np)^3}.
\end{aligned}$$

Therefore if $p = \omega(1/n)$ we have that $1/(np) = o(1)$ and so

$$\frac{\mathrm{Var}(T)}{\mathbb{E}(T)^2} \to 0.$$

Therefore, by Lemma 6.3, if $p = \omega(1/n)$ we have that $\mathbb{P}(T \geq 1) \to 1$, and so almost surely $G(n,p)$ contains a triangle.

So we have that there is a transition between random graphs that contain a triangle almost never or almost always, when p is around $1/n$. In order to decsribe this phenomenon more generally, Erdős and Rényi introduce the notion of a *threshold function*.

Definition. A function $r : \mathbb{N} \to \mathbb{R}$ is a *threshold function* for a monotone graph property A, if for any $p : \mathbb{N} \to [0,1]$

- $p(n) = o(r(n)) \Rightarrow \mathbb{P}(A \text{ holds for } G(n,p)) \to 0$;

- $p(n) = \omega(r(n)) \Rightarrow \mathbb{P}(A \text{ holds for } G(n,p)) \to 1$.

Note that for a general property A a threshold function might not exist, and, if it does exists, it is not unique. For example for the property of containing a triangle we saw that a threshold function was $r(n) = 1/n$, but $r(n) = c/n$ (for any $c > 0$) could serve as well.

6.3 Balanced Subgraphs

We could ask the same question for any graph H: Is there a threshold function for the property of containing H as a subgraph, and if so, what is an example of one? It turns out that the approach given above for triangle can be extended to any graph H which is *balanced*.

Definition. Let H be a graph with $|H| = v$ and $e(H) = e$. The *density* of H is defined to be

$$\rho(H) = \frac{e}{v}.$$

H is said to be *balanced* if no subgraph of H has strictly greater density than H itself.

Theorem 6.4. *Let H be a balanced graph with density ρ. Then*

$$r(n) = n^{-\frac{1}{\rho}}$$

is a threshold function for the event that H is a subgraph of $G(n,p)$.

Proof. Let $|V(H)| = v$ and $|E(H)| = e$, and so $\rho = e/v$. We denote the vertices of H by $\{h_1, h_2, \ldots, h_v\}$. Given an ordered v-tuple $\beta = (b_1, b_2, \ldots, b_n) \in [n]\big(= V(G(n,p))\big)$ we let A_β denote the event that $G(n,p)$ contains an appropriately ordered copy of H on β. That is, A_β occurs if the mapping $h_i \mapsto b_i$ is a graph homomorphism.

Let X_β be the indicator random variable of the event A_β and let $X = \sum_\beta X_\beta$, where the sum is over all ordered v-tuples. Note that X does not count precisely the number of copies of H in $G(n,p)$ since, due to potential symmetries in H, some copies may be counted more than once. However the two conditions $X = 0$ and $X > 0$ are still equivalent to the events that $G(n,p)$ doesn't contain a copy of H and $G(n,p)$ does contain a copy of H respectively.

For any β we have that $\mathbb{P}(A_\beta) = p^e$. Hence, by linearity of expectation

$$\mathbb{E}(X) = \sum_\beta \mathbb{P}(A_\beta) = v!\binom{n}{v}p^e \leq n^v p^e.$$

Therefore, if $p(n) = o(n^{-v/e})$, then

$$\mathbb{E}(X) \leq n^v p^e = o(1) \to 0.$$

Hence by Markov's inequality we have that $\mathbb{P}(X = 0) \to 1$. So it remains to show that, when $p(n) = \omega(n^{-v/e})$ we have that $\mathbb{P}(X \geq 1) \to 1$. For this we will need to calculate $\mathrm{Var}(X)$.

We have, as before, that

$$\mathrm{Var}(X) = \sum_\beta \mathbb{P}(A_\beta)\left(\sum_\gamma \mathbb{P}(A_\gamma | A_\beta) - \mathbb{P}(A_\gamma)\right).$$

As before we split the sum into cases according to $|\beta \cap \gamma|$ (when understood as sets). If $|\beta \cap \gamma| \leq 1$, the two vertex sets share no edges, and so the two events are independent. Therefore there is no contribution to the sum from such pairs. If $|\beta \cap \gamma| = t \geq 2$ then, since H is balanced, the two copies of H have at most $t\rho$ edges in common. Hence

$$\mathbb{P}(A_\gamma | A_\beta) - \mathbb{P}(A_\gamma) \leq p^{e-t\rho}.$$

For any given β there are $O(n^{v-t})$ different γ which share t vertices with β (since the size of H is a fixed constant). Therefore we have,

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_\beta \mathbb{P}(A_\beta) \left(\sum_\gamma \mathbb{P}(A_\gamma | A_\beta) - \mathbb{P}(A_\gamma) \right) \\
&= \sum_\beta \mathbb{P}(A_\beta) \left(\sum_{t=2}^{v} O(n^{v-t}) p^{e-t\rho} \right) \\
&\leq \mathbb{E}(X) \left(\sum_{t=2}^{v} O((n^v p^e)^{1-\frac{t}{v}}) \right).
\end{aligned}$$

Therefore, noting that $\mathbb{E}(X) \geq 1/2(n^v p^e)$ for large enough n, we have that

$$\begin{aligned}
\frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2} &\leq 2 \frac{\sum_{t=2}^{v} O((n^v p^e)^{1-\frac{t}{v}})}{n^v p^e} \\
&= O\left(\sum_{t=2}^{v} (n^v p^e)^{-\frac{t}{v}} \right) \to 0.
\end{aligned}$$

The last equality holds since $p = \omega(n^{-v/e})$ implies that $n^v p^e \to \infty$. Hence, by Lemma 6.3, we have that $\mathbb{P}(X \geq 1) \to 1$, and so almost surely $G(n,p)$ contains a copy of H. □

We note that, if H is a graph with $|H| = v$ and $e(H) = e$ which is not balanced, then $r(n) = n^{-v/e}$ is not a threshold function the existence of H as a subgraph. Indeed, suppose $H_1 \subset H$ is a subgraph with $\rho(H) < \rho(H_1)$. Let $|H_1| = v_1$ and $e(H_1) = e_1$. Then we can pick some α such that $v_1/e_1 < \alpha < v/e$ and set $p = n^{-\alpha}$. However now the expected number of copies of H_1 in $G(n,p)$ is $o(1)$, and so by Markov's inequality almost surely $G(n,p)$ contains no copy of H_1, and thus no copy of H.

So for general H, if a threshold function of the form $n^{-\alpha}$ exists, α can't be any bigger than $1/\rho(H')$ for any subgraph H'. It turns out that the converse of this is true. We don't include the proof, but note that it can be proved using the same methods.

Theorem 6.5. *Let H be a graph and $H' \subset H$ a subgraph of H with the maximum density. Then*

$$r(n) = n^{-\frac{1}{\rho(H')}}$$

is a threshold function for the event that H is a subgraph of $G(n,p)$.

7 The Hamiltonicity Threshold

In the following chapter we present a proof of the threshold function for the existence of a Hamiltonian cycle in $G(n,p)$. We will first find the threshold function for connectivity. It is clear that, since a Hamiltonian graph must be connected, the threshold for Hamiltonicity must be at least as large as the threshold for connectivity. It will turn out that, up to a constant factor, the two are the same.

7.1 The Connectivity Threshold

In this section we will show that $r(n) = \log(n)/n$ is a threshold function for the event that $G(n,p)$ is connected. The proof can be split into two smaller parts, we first show that if $p(n) = o(r(n))$ then in fact almost surely there exists an isolated vertex (and so in particular G is disconnected). Then we show that if $p(n) = \omega(r(n))$ then with high probability G has no component of size $\leq n/2$ (and so, in particular G must be connected).

It is a general feature of many probabilistic proofs that, if we want to estimate the probability of an event happening it is often more convenient to estimate the probability of a simpler event which will imply the former.

In fact we will show more for the first part, that in fact $r(n)$ is also a threshold function for the event that $G(n,p)$ contains an isolated vertex.

Theorem 7.1. *$r(n) = \log(n)/n$ is a threshold function for the event that $G(n,p)$ contains an isolated vertex*

Proof. We let X_1 be the random variable which counts the number of isolated vertices in $G(n,p)$. We note that $X_1 = \sum_{v \in V} X_v$ where X_v is the indicator random variable of the event that v is isolated, or equivalently that $d(v) = 0$. We have that $\mathbb{E}(X_v) = \mathbb{P}(d(v) = 0) = (1-p)^{n-1}$. Hence by linearity of expectation

$$\mathbb{E}(X_1) = \sum_{v \in V} \mathbb{E}(X_v) = n(1-p)^{n-1}.$$

If $p(n) = \omega(\log(n)/n)$ then we see that

$$\mathbb{E}(X_1) = n(1-p)^{n-1} \leq ne^{-p(n-1)} = ne^{-\omega(\log(n))} = o(1).$$

(In fact, as we will use later, this is true as long as $p(n) \geq c(\log(n)/n)$ for some fixed large c.)

So by Markov's inequality we have that $\mathbb{P}(X_1 = 0) \to 1$ and so, almost surely, G contains no isolated vertices. If $p(n) = o(\log(n)/n)$ we wish to apply the second moment method, so we have to estimate $\mathrm{Var}(X_1)$. We have that

$$\begin{aligned}\mathrm{Var}(X_1) &= \sum_v \mathbb{P}(d(v)=0)\left(\sum_w \mathbb{P}(d(w)=0|d(v)=0) - \mathbb{P}(d(w)=0)\right)\\ &= \sum_v \mathbb{P}(d(v)=0)\left((1-(1-p)^{n-1}) + \sum_{w\neq v}(1-p)^{n-2} - (1-p)^{n-1}\right)\\ &= \mathbb{E}(X_1)\left((n-1)p(1-p)^{n-2} + (1-(1-p)^{n-1})\right).\end{aligned}$$

Hence

$$\frac{\mathrm{Var}(X_1)}{\mathbb{E}(X_1)^2} \leq \frac{(n-1)p(1-p)^{n-2}+1}{n(1-p)^{n-1}} \leq \frac{p}{1-p} + \frac{1}{n(1-p)^{n-1}} = o(1).$$

Therefore, by the usual arguments, $\mathbb{P}(X_1 \geq 1) \to 1$. □

If a graph is connected, it cannot have any isolated vertices and therefore we have that, if $p(n) = o(\log(n)/n)$, then

$$\mathbb{P}(G(n,p) \text{ is connected}) \leq \mathbb{P}(X_1 = 0) \to 0$$

and so almost surely $G(n,p)$ isn't connected. To show the converse, that if $p(n) = \omega(\log(n)/n)$ then almost surely $G(n,p)$ is connected, we estimate the expected number of small components. Let X_k be the random variable which counts the number of components of size k in $G(n,p)$.

Lemma 7.2. *If $p(n) = c\log(n)/n$, for $c \geq 1$, then*

$$\sum_{k=2}^{\frac{n}{2}} \mathbb{P}(X_k > 0) = o(1).$$

Proof. By Markov's inequality we have that

$$\sum_{k=2}^{\frac{n}{2}} \mathbb{P}(X_k > 0) \leq \sum_{k=2}^{\frac{n}{2}} \mathbb{E}(X_k).$$

We would like to bound the expected number of components of order k. For any set $|A| = k$ of vertices we let I_A be the indicator random variable that A is a component in $G(n,p)$.

Since $\mathbb{E}(X) = \sum \mathbb{P}(A \text{ is a component})$ we would like to bound the probability that A is a component. This can be split into two independent events.

Firstly it is necessary that there are no edges between A and $[n] \setminus A$. This happens with probability $(1-p)^{k(n-k)}$, since there are $k(n-k)$ edges to consider. Secondly we need to estimate the probability that A is connected. To do this we use a result of Cayley (without proof) that the number of spanning trees on k vertices is k^{k-2}. For each tree we have that the probability that it is contained in $G(n,p)$ is p^{k-1} and so, by a simple union bound, we have the probability that A is connected is at most $k^{k-2}p^{k-1}$.

Combining this we see that

$$\mathbb{E}(X_k) = \sum_{|A|=k} \mathbb{P}(A \text{ is a component}) \leq k^{k-2} p^{k-1} (1-p)^{k(n-k)},$$

and so,

$$\sum_{k=2}^{\frac{n}{2}} \mathbb{E}(X_k) \leq \sum_{k=2}^{\frac{n}{2}} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} = \sum_{k=2}^{\frac{n}{2}} \mu_k.$$

What can we say about the terms μ_k. Well, using our standard inequalities $\binom{n}{k} \leq (en/k)^k$ and $(1-p) \leq e^{-p}$ we see that

$$\begin{aligned}
\mu_k &\leq \left(\frac{en}{k}\right)^k k^{k-2} \left(\frac{c\log(n)}{n}\right)^{k-1} \left(1 - \frac{c\log(n)}{n}\right)^{k(n-k)} \\
&\leq ne^k (c\log(n))^{k-1} e^{-\frac{c\log(n)}{n} k(n-k)} \\
&\leq \exp\left(\log(n) + k + (k-1)\log(c) + (k-1)\log(\log(n)) - kc\log(n) + k^2 \frac{c\log(n)}{n}\right).
\end{aligned}$$

For small k, say in the range $2 \leq k \leq 10$ we can crudely bound the expression by

$$\mu_k \leq \exp\big((k-1)(\log(\log(n)) - \log(n)) + O(1)\big) \leq O\left(\left(\frac{\log(n)}{n}\right)^{(k-1)}\right).$$

For larger k we have that,

$$\begin{aligned}
\mu_k &\leq \exp\left(\log(n) + k + (k-1)\log(c) + k\log(\log(n)) - kc\log(n) + k\frac{c\log(n)}{2}\right) \\
&\leq \exp\left(\log(n) + k\log(\log(n)) - k\frac{\log(n)}{4}\right) \\
&= O\left(n\left(\frac{\log(n)}{n^{\frac{1}{4}}}\right)^k\right).
\end{aligned}$$

So we can conclude that

$$\begin{aligned}
\sum_{k=2}^{\frac{n}{2}} \mu_k &\leq \sum_{k=2}^{10} O\left(\left(\frac{\log(n)}{n}\right)^{(k-1)}\right) + \sum_{k=10}^{\frac{n}{2}} O\left(n\left(\frac{\log(n)}{n^{\frac{1}{4}}}\right)^k\right) \\
&\leq O\left(\frac{\log(n)}{n}\right) + \frac{n}{2} O\left(n\frac{\log(n)}{n^{\frac{10}{4}}}\right) \\
&= o(1).
\end{aligned}$$

Hence

$$\sum_{k=2}^{\frac{n}{2}} \mathbb{P}(X_k > 0) = o(1)$$

as claimed. □

Theorem 7.3. *$r(n) = \log(n)/n$ is a threshold function for the event that $G(n,p)$ is connected.*

Proof. As noted if $p(n) = o(\log(n)/n)$ then by Theorem 7.1

$$\mathbb{P}(G(n,p) \text{ is connected}) \leq \mathbb{P}(X_1 = 0) \to 0$$

and so almost surely $G(n,p)$ isn't connected. Conversely if $p(n) = c(\log(n)/n)$, for large enough C, then by Lemma 7.2

$$\sum_{k=2}^{\frac{n}{2}} \mathbb{P}(X_k > 0) = o(1).$$

Also note that, by Theorem 7.1 $\mathbb{P}(X_1 > 0) = o(1)$ for $p(n) = c(\log(n)/n)$ as well, for large enough c. Now we note that $G(n,p)$ is disconnected if and only if it has some component of size less than $n/2$. Hence by a union bound

$$\mathbb{P}(G(n,p) \text{ is not connected}) \leq \sum_{k=1}^{\frac{n}{2}} \mathbb{P}(X_k > 0) = o(1).$$

Therefore, almost surely $G(n,p)$ is connected. Since being connected is a monotone property, and since if a monotone property holds almost surely for $G(n,p)$ it also holds almost surely for $G(n,p')$ for all $p' \geq p$, we have that, if $p(n) = \omega(\log(n)/n$, then almost surely $G(n,p)$ is connected. □

We note that a more careful analysis of the proof would show that in fact there is a much sharper threshold. More specifically if $p(n) = c\log(n)/n$ then $G(n,p)$ is almost surely connected when $p > 1$ and almost surely not connected when $p < 1$. Much more precise results are know for the threshold for connectivity, specifically

- $p(n) = (\log(n) - \omega(n))/n$ then almost surely $G(n,p)$ contains an isolated vertex;
- if $p(n) = (\log(n) + \omega(n))/n$ then almost surely $G(n,p)$ is connected.

We now have that if $p(n) = c\log(n)/n$ with $c < 1$ then $G(n,p)$ is almost surely disconnected and so definitely almost surely not Hamiltonian. In order to show that $G(n,p)$ is almost surely Hamiltonian for large enough p we might hope to calculate the expected number of Hamiltonian cycles in $G(n,p)$, show that it is large, and also calculate the variance and show that is it (relatively) small. However, it turns out that the variance in this case is too large for Chebyshev's inequality to give us the result we want. Instead we will have to use a more subtle argument.

7.2 Posá's Rotation-Extension Technique

We introduce in this section a useful technique due to Posá that he used to prove the threshold for Hamiltonicity in $G(n,p)$.

Given a graph G and a vertex $x_0 \in V(G)$ suppose that $P = x_0x_1 \dots x_k$ is a longest path in G starting at x_0. Given an edge $(x_k, x_i) \in E(G)$ a *rotation* of P is a new path $P' = x_0x_1 \dots x_ix_kx_{k-1} \dots x_{i+1}$. We say that a path P' is a *transform* of P if it can be obatined from P by a sequence of rotations. Let U be the set of endvertices of all possible transforms of P and let

$$N = \{x_i : \{x_{i+1}, x_{i-1}\} \cap U \neq \emptyset\}$$

be the set of neighbours of this set in P. Finally let $R = V(P) \setminus (U \cup N)$ be the rest of the vertices in P.

Note that, since by assumption P is a longest path, the neighbourhood of U is a subset of $V(P)$. Posá observed that in fact it had to be a subset of $U \cup N$

Lemma 7.4. *Let G be a graph, $x_0 \in V(G)$ and P a longest path in G starting at x_0. If U, N and R are defined as above then there are no edges in G between U and R.*

Proof. Suppose $x \in U$ and $(x,y) \in E(G)$. Then there is some transform of P_x of P ending at x. Since we have $(x,y) \in E(G)$, there is some transform P_z ending at a vertex z which is a neighbour of y on P_x.

If (y,z) were in $E(P)$, then $y \in N$. Otherwise one of the rotations that formed the transformation from P to P_x must have deleted an edge $(y,w) \in E(P)$. If we look at the first time this happened during the transformation from P to P_x, then either y or w will have become the endpoint of one of the intermediate transformations. Therefore $y \in U$ or N. Hence for every $x \in U$, $N(x) \subset U \cup N$. □

We note then that,

$$U \cup N(U) \subset U \cup N = U \cup \{x_{k-1}\} \bigcup_{x_j \in U, j \neq k} \{x_{j-1}, x_{j+1}\}.$$

Hence $|U \cup N(U)| \leq 3|U| - 1$. That is, in some ways U has small 'expansion'. We will show that in a sufficiently dense random graph, all 'small' subsets have large expansion and so, in particular, U must be large.

Lemma 7.5. *Suppose c is sufficiently large and $p = c\log(n)/n$. Then almost surely in $G(n,p)$ every subset $U \subset V(G)$ with $|U| \leq n/4$ satisfies*

$$|U \cup N(U)| \geq 3|U|.$$

In particular the property that some set U fails to satisfy this property is less then n^{-2}.

Proof. We consider the probability q that there exists a set U of k vertices and a set W of $n-3k+1$ vertices, with $1 \leq k \leq n/4$ such that there are no edges between U and W. Note that

if there is a subset $U \subset V(G)$ with $|U| \leq n/4$ and $|U \cup N(U)| < 3|U| - 1$ then there is a subset W of $V(G) \setminus (U \cup N(U))$ of size $n - 3|U| + 1$ such that there are no edges between U and W.

For each such pair (U, W) the probability there are no $U - W$ edges is

$$(1-p)^{k(n-3k+1)}.$$

Therefore by the union bound, the probability that such a pair of sets exist is at most

$$\begin{aligned} q &= \sum_{k=1}^{\frac{n}{4}} \binom{n}{k} \binom{n-k}{n-3k+1} \left(1 - \frac{c\log(n)}{n}\right)^{k(n-3k+1)} \\ &= \sum_{k=1}^{\frac{n}{4}} \binom{n}{k} \binom{n-k}{2k-1} \left(1 - \frac{c\log(n)}{n}\right)^{k(n-3k+1)} \\ &\leq \sum_{k=1}^{\frac{n}{4}} n^{3k} \exp\left(-c\frac{\log(n)}{n} k(n-3k+1)\right). \end{aligned}$$

. Since $n - 3k + 1 \geq n/4$ we can conclude that, if $c \geq 24$

$$\begin{aligned} q &\leq \sum_{k=1}^{\frac{n}{4}} n^{3k} n^{-c\frac{k}{4}} \\ &\leq \sum_{k=1}^{\frac{n}{4}} n^{3k} n^{-6k} \\ &\leq \frac{n}{4} n^{-3} \leq n^{-2} \to 0. \end{aligned}$$

.

Therefore if c is large enough then almost surely no such pair exists. Therefore almost surely $G(n,p)$ satisfies the required property. □

Finally we can use this to show that, if $p = c\log(n)/n$ with c sufficiently large then, almost surely G contains a Hamiltonian path.

Lemma 7.6. *Suppose c is sufficiently large and $p = c\log(n)/n$. Then almost surely in $G(n,p)$ contains a Hamiltonian path.*

Proof. For a fixed vertex x we want to estimate the probability that every longest path in $G(n,p)$ contains x. If the probability that this doesn't happen for any particular vertex is $o(1/n)$ then the probability that it doesn't happen for any vertex is $o(1)$ and so almost surely every longest path in $G(n,p)$ contains every vertex. In particular every maximal path would be a Hamiltonian path.

So, given x we let $G(x)$ be the graph on $n-1$ vertices obtained by deleting x from $G(n,p)$. Note that $G(x)$ has the same distribution as $G(n-1,p)$. We choose a longest path in $G(x)$, $P = x_0x_1 \dots x_k$.

We note that, if x can extend some transform of P to a longer path then there is a longer path in $G(n,p)$ than in $G(x)$. However if that is the case then every longest path in $G(n,p)$

must contain x, since otherwise it would be in $G(x)$. Therefore the probability that x is not contained in every longest path in G is less than the probability that x does not extend some transform of the path P.

Note that by Lemma 7.5 almost surely $G(x)$ is such that $U \subset V(G(x))$ with $|U| \leq (n-1)/4$ satisfies

$$|U \cup N(U)| \geq 3|U|.$$

Let us call a graph which does *good*. If this is the case then, since we have shown that in a longest path P the set of endpoints of possible rotations U satisfies $|U \cup N(U)| \leq 3|U| - 1$, it follows that $|U| > (n-1)/4$

Therefore, given that $G(x)$ is good, the probability that x does not extend some transform of the path P is less than the probability of the non-existence of at least $(n-1)/4$ edges between x and U, each of which is independent of the event that $G(x)$ is good, which is

$$(1-p)^{\frac{n-1}{4}} \leq e^{-\frac{n-1}{4}\frac{c\log(n)}{n}} \leq n^{-\frac{c}{5}}.$$

Therefore we see that, if A_x is the event that x is contained in every longest path,

$$\begin{aligned}\mathbb{P}(A_x \text{ doesn't happen}) &\leq \mathbb{P}(A_x \text{ doesn't happen} \,|\, G(x)\text{is good}) + \mathbb{P}(G(x) \text{ isn't good}) \\ &\leq n^{-\frac{c}{5}} + n^{-2} = o(n^{-1}).\end{aligned}$$

for c large enough. Hence, by the union bound,

$$\mathbb{P}(\text{ Some } A_x \text{ doesn't happen}) \leq \sum_x \mathbb{P}(A_x \text{ doesn't happen}) \leq \sum_x o(n^{-1}) = o(1).$$

Therefore almost surely every longest path is a Hamiltonian path, and so $G(n,p)$ contains a Hamiltonian path as claimed. □

SO, almost surely G will have a Hamiltonian path, but how can we change this path into a cycle? The following lemma shows that, if U (for the Hamiltonian path) is large, there are many edges we could add to G to form a Hamiltonian cycle.

Lemma 7.7. *Let G be a connected graph. Suppose that the longest path in G has length $k \geq 2$, G contains no cycles of length $k+1$ and for some $u \in \mathbb{N}$ we have that for every subset $U \subset V(G)$ with $|U| < u$*

$$|U \cup N(U)| \geq 3|U|.$$

Then there are at least $u^2/2$ non-edges in G whose addition forms a $k+1$ cycle in G.

Proof. Let $P = x_0x_1 \dots x_k$ be a longest path and let U, N, R be defined as above. Then every $x_0 - U$ edge creates a cycle of length $k+1$. Note also that, as before

$$|U \cup N(U)| \leq 3|U| - 1$$

and so, by assumption, $|U| \geq u$. Consider a subset $\{y_1, y_2, \ldots, y_u\} \subset U$. For each y_i there is some path P_i of length k from x_0 to y_i. Considering these paths as starting at y_i, we let Y_i be the set of endvertices of transforms of P_i. By the same argument as before we have that $|Y_i| \geq u$ for each i, and also any edge between Y_i and y_i forms a cycle of length $k+1$. Hence none of these $u^2/2$ edges are in G, and they are the claimed set of edges whose additions forms a $k+1$ cycle in G. □

7.3 Hamiltonicty Threshold

We now have the tools necessary to prove the main result of this chapter. The main part of the proof remaining is to show the existence of a Hamiltonian cycle when $p(n) = \omega(\log(n)/n)$. We already know that when $p(n) = c\log(n)/n$ for large enough c then $G(n,p)$ almost surely contains a Hamiltonian path and satisfies the condition of Lemma 7.7 for a large u, and so, if $G(n,p)$ does not contain a Hamiltonian cycle then there are a large number of non-edges whose addition would form a Hamiltonian cycle in $G(n,p)$

One useful idea in the proof, which appears in many proofs, is to pick two random graphs $G(n,p_1)$ and $G(n,p_2)$ and let $H = G(n,p_1) \cup G(n,p_2)$ be the union of the two graphs. It is a simple a exercise to show that H is distributed as $G(n,q)$ for $q = p_1 + p_2 - p_1p_2$. We can think of the process as a two step exposure of the edges in $G(n,q)$.

In this case we pick $p_1 = c\log(n)/n$ to guarantee that $G(n,p_1)$ contains a Hamiltonian path and there exists a large number of edges not in $G(n,p_1)$ whose addition would form a Hamiltonian cycle. We then pick p_2 large enough that almost surely we must pick one of those edges. Hence almost surely in $G(n,p_1) \cup G(n,p_2)$ there is a Hamiltonian cycle, and so the same holds true for $G(n,q)$.

This technique is sometimes called *sprinkling*, we can think of it as picking $G(n,p_1)$ and then 'sprinkling' some extra edges on top, with a fixed probability.

Theorem 7.8. *$r(n) = \log(n)/n$ is a threshold function for the event that $G(n,p)$ is Hamiltonian.*

Proof. We note that if $p(n) = o(r(n))$ then by Theorem 7.3 $G(n,p)$ is almost surely not connected and so almost surely not Hamiltonian.

We also note that, since the property of containing a Hamiltonian cycle is monotone, if $G(n,p)$ almost surely contains a Hamiltonian cycle for some $p(n) = O(\log(n)/n)$, then the same will be true for all $p(n) = \omega(\log(n)/n)$.

So let us take $p(n) = 25\log n/n$. We choose $G(n,p)$ by picking $G(n,p_1)$ and $G(n,p_2)$ as before, with $p_1 = 24\log n/n$ and

$$p_2 = \frac{p - p_1}{1 - p_1}.$$

Here we have just chosen p_2 so that $p = p_1 + p_2 - p_1p_2$, all we will use is that $p_2 \geq \log(n)/n$.

We have by Lemma 7.6 and Lemma 7.5 that almost surely $G(n,p_1)$ contains a Hamiltonian path P and also almost surely $G(n,p_1)$ satisfies the conditions of Lemma 7.7 with $u = n/4$.

Hence almost surely there is either a Hamiltonian cycle in $G(n,p_1)$ already, or there exists a set of $n^2/32$ non-edges $F \subset E(G(n,p_1))^c$ whose addition would form a Hamiltonian cycle in $G(n,p_1)$.

We now expose the edges in $G(n,p_2)$. The probability that none of these $n^2/32$ edges appear in $G(n,p_2)$ is

$$\mathbb{P}(F \cap E(G(n,p_2)) = \emptyset) = (1-p_2)^{|F|} \leq e^{-p\frac{n^2}{32}} \leq n^{-\frac{n}{32}} \to 0$$

.

So almost surely $G(n,p_2)$ will contain at least one edge from F, and so almost surely $G(n,p)$ will contain a Hamiltonian cycle. □

As with Theorem 7.3 one can prove that much sharper threshold exists for Hamiltonicity in $G(n,p)$. It has been shown, by Bollobás and independently Komlós and Szemerédi that there is a sharp additive threshold for Hamiltonicty, that is, if $\omega(n)$ is a function tending to infinity arbitrarily slowly then

- $p(n) = (\log(n) + \log(\log(n)) - \omega(n))/n$ then almost surely $G(n,p)$ satisfies $\delta(G) \leq 1$;
- if $p(n) = (\log(n) + \log(\log(n)) + \omega(n))/n$ then almost surely $G(n,p)$ contains a Hamiltonian cycle.

We note that, quite amazingly, similarly to how the threshold for connectivity coincides with the threshold for having minimal degree 1, the threshold for Hamiltonicity coincides with the threshold for having minimal degree 2, clearly a necessary condition.

8 Strong Concentration

8.1 Motivation

Let us consider, as a motivating question, what we expect the maximum degree of the random graph $G(n, 1/2)$ to be. It appears to be a complicated variable, it is not even clear how to compute it's expectation. For any particular vertex, the expected degree is $d = (n-1)/2$, but this doesn't tell us much about the maximum.

Suppose however we could show that the probability that the degree of any given vertex exceeds d by at least t is $o(1/n)$, for some suitably chosen t. Then we could conclude by the union bound that the probability that the maximum degree is more than $d + t$ is $o(1)$, that is, almost surely the maximum degree is less that $d + t$.

In this case, and in many application of the probabilistic method, we want to bound probabilities of the for $\mathbb{P}(X \geq \mathbb{E}(X) + t)$ for some random variable X, or often more generally $\mathbb{P}(|X - \mathbb{E}(X)| \geq t)$. We call bounds for such probabilities *tail estimates*, since we think of them as the probability that X lies in the tail of it's distribution. If we can show that with high probability $|X - \mathbb{E}(X)| \leq t$, where t is considerably smaller than $\mathbb{E}(X)$, we say that X is *concentrated* about its expectation.

Chebyshev's inequality is a very general result of this type, which holds for all random variables with finite mean. However it is quite weak, especially if we want to deal with many random variables simultaneously. For example Chebyshev's inequality tells us that

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

If we let X be the degree of a fixed vertex in $G(n, 1/2)$ we have that X is distributed as a binomial random variable $B(n-1, 1/2)$, and so $\mathrm{Var}(X) = (n-1)/4$. But then for the tail probability to be $o(1/n)$ we would need to apply Chebyshev's with $t >> n$, and so we would only get a bound on the probability of deviations of size $\omega(n)$ which isn't useful at all.

However we will see for this particular X, and in fact a large class of similar random variables, a much better inequality holds with $1/t^2$ replaced by the exponentially small bound $2e^{-t^2/2}$. This would already by sufficient to conclude that the maximum degree of $G(n, 1/2)$ is quite tightly concentrated about its expectation (in fact, as we shall see, that it almost never exceeds $n/2 + O(\sqrt{n \log(n)})$).

8.2 The Chernoff Bound

In the previous example the degree of a vertex in $G(n, 1/2)$ was the sum of $n-1$ mutually independent random variables which took the values 0 and 1, each with probability $1/2$. For ease of presentation we will consider a similar situation, except the variables will take values 1 and -1, so that the expectation is 0. Results for the original setting can be recovered by a simple re-scaling. The following result is often called Chernoff's inequality or the Chernoff bound

(although Chernoff proved a different inequality that is less useful and the following theorem (and the ingenious proof) is due to Bernstein).

Theorem 8.1. *Let $X_1, X_2, \ldots X_n$ be independent random variables taking the values 1 and -1, each with probability $1/2$. Let $X = X_1 + X_2 + \ldots + X_n$. Then, for any $t \geq 0$,*

$$\mathbb{P}(X \geq t) < e^{-\frac{t^2}{2n}} \text{ and } \mathbb{P}(X \leq -t) < e^{-\frac{t^2}{2n}}.$$

Proof. Note that we can write down an explicit formula for $\mathbb{P}(X \geq t)$ in terms of binomial coefficients, and it would be possible to prove the result by a careful estimation of these quantities. However the idea in this proof is not only much simpler, but is applicable in a lots of other cases where explicit formulas are not available.

We will just prove the first inequality, the second follows by symmetry. The key idea is to consider instead of X the random variable $Y = e^{\lambda X}$ where $\lambda > 0$ is some real number that we will pick later. We have that $\mathbb{P}(X \geq t) = \mathbb{P}(Y \geq e^{\lambda t})$ and by Markov's inequality

$$\mathbb{P}(Y \geq e^{\lambda t}) \leq \frac{\mathbb{E}(Y)}{e^{\lambda t}}.$$

However we can calculate

$$\mathbb{E}(Y) = \mathbb{E}\left(E^{\lambda(\sum_i X_i)}\right) = \mathbb{E}\left(\prod_i e^{\lambda X_i}\right) = \prod_i \mathbb{E}\left(e^{\lambda X_i}\right)$$

by the independence of the X_i. Hence

$$\mathbb{E}(Y) \leq \left(\frac{e^{\lambda} + e^{-\lambda}}{2}\right)^n = (\cosh(\lambda))^2$$

However

$$\cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots \leq 1 + \frac{x^2}{2} + \frac{(\frac{x^2}{2})^2}{2!} + \ldots = e^{\frac{x^2}{2}}$$

Hence,

$$\mathbb{E}(Y) < e^{\frac{n\lambda^2}{2}}$$

So we have

$$\mathbb{P}(Y \geq e^{\lambda t}) < \frac{e^{\frac{n\lambda^2}{2}}}{e^{\lambda t}} = e^{\frac{n\lambda^2}{2} - \lambda t}.$$

Since this holds for all $\lambda > 0$ we can in particular take $\lambda = t/n$ (which also happens to be the value minimising the above quantity) to see that

$$\mathbb{P}(X \geq t) < e^{-\frac{t^2}{2n}}.$$

□

We see that it follows immediately from the Chernoff bound that the maximum degree of $G(n, 1/2)$ is tightly concentrated around it's expectation.

Theorem 8.2. *The maximum degree of $G(n, 1/2)$ is almost surely $n/2 + O(\sqrt{n \log(n)})$.*

Proof. For each $v \in V$ let D_v be the random variable which counts the degree of v. If we let X_e be the indicator random variable of the event that e is an edge, we have that $D_v = \sum_{v \in e} X_e$. We define $Y_e = 2X_e - 1$, so that Y_e takes the values 1 and -1 each with probability $1/2$, and let $\hat{D}_v = \sum_{v \in e} Y_e$ so that $\hat{D}_v = 2D_v - (n-1)$. We apply Theorem 8.1 to the random variables $\{Y_e : v \in e\}$ with $t = 2\sqrt{n \log(n)}$ to see that

$$\mathbb{P}(|\hat{D}_v|) > 2\sqrt{n \log(n)}) < 2e^{-\frac{4n \log(n)}{2(n-1)}} < 2e^{-2 \log(n)} = \frac{2}{n^2}.$$

Therefore, for each v we have that

$$\mathbb{P}\big(|D_v - \frac{n-1}{2}| > \sqrt{n \log(n)}\big) = \mathbb{P}(|\hat{D}_v| > 2\sqrt{n \log(n)}) < \frac{2}{n^2}.$$

Hence by the union bound we have that

$$\mathbb{P}\big(|D_v - \frac{n-1}{2}| \leq \sqrt{n \log(n)} \text{ for all } v\big) \geq 1 - \frac{2}{n} \to 1$$

and so almost surely the maximum degree of $G(n, 1/2)$ is $n/2 + O\big(\sqrt{n \log(n)}\big)$ □

In general we often want to apply this sort of reasoning to slightly more general class of random variables. A very similar proof to that as in Theorem 8.1 gives the following theorem for bounded random variables.

Theorem 8.3. *Let $X_1, X_2, \ldots, X_n$ be independent random variables, each taking values in $[0, 1]$, let $X = X_1 + X_2 + \ldots + X_n$ and let $\sigma^2 = Var(X) = \sum_i Var(X_i)$. Then for any $t \geq 0$*

$$\mathbb{P}(X \geq \mathbb{E}(X) + t) < e^{-\frac{t^2}{2(\sigma^2 + \frac{t}{3})}} \text{ and } \mathbb{P}(X \leq \mathbb{E}(X) - t) < e^{-\frac{t^2}{2(\sigma^2 + \frac{t}{3})}}.$$

In particular we can take X to a binomial random variable $B(n, p)$ for $p \neq 1/2$, where $\mathrm{Var}(X) = np(1-p)$, which is often useful in applications. For example we can use this to prove good lower bounds on the diameter of $G(n, p)$ for certain ranges of p.

Definition. For any two vertices x and y in a connected graph G we define the *distance* between x and y, $\mathrm{dist}(x, y)$, to be the length of the shortest path between x and y. The *diameter* of a graph G is the maximum distance between any pair of vertices, and the *radius* is the minimum distance r such that there exists a vertex x such that $\mathrm{dist}(x, y) \leq r$ for all $y \in G$.

Theorem 8.4. *Let $d \geq 2$ be fixed. Suppose $c > 0$ and $p^d n^{d-1} = \log(n^2/c)$. Then almost surely $G(n, p)$ has diameter at least d.*

Proof. For some $v \in V$ let us denote by $N_k(v)$ the set of vertices whose distance in G is exactly k from v. We will show that, almost surely for $0 \leq k < d$, the size of the $N_k(v)$ is $o(n)$, and so there must be $(1 - o(1))n$ vertices of distance at least d away from v. Specifically we want to show that almost surely $|N_k(v)| \leq (2np)^k$.

If we know the size of $N_1(v), N_2(v), \ldots, N_{k-1}(v)$, then $|N_k(v)|$ is distributed as a binomial random variable, since each point in $V \setminus \bigcup_{i=1}^{k-1} N_i(v)$ is in $N_k(v)$ with probability $1 - (1-p)^{|N_{k-1}(v)|}$. Let us define A_i to be the event that $|N_i(v)| \leq (2np)^i$ and condition on the events $A_1, A_2, \ldots, A_{k-1}$. For each possible value of $|N_{k-1}(v)|$ conditioned on A_{k-1} we have that $|N_k(v)| \sim B(m, q)$ where $m < n$ and $q = 1 - (1-p)^{|N_{k-1}(v)|} \leq p|N_{k-1}(v)| \leq p(2pn)^{k-1}$.

Therefore we have that, conditioned on $A_1, A_2, \ldots, A_{k-1}$, $N_k(v)$ as a random variable is bounded above by some random variable Y distributed as $Bin(n, p(2pn)^{k-1})$ on the same probability space. Therefore the probability that $N_k(v)$ is large can be bounded above by the probability that Y is large, which we can estimate using the generalised Chernoff's bound. We have that

$$\mu = \mathbb{E}(Y) = np(2pn)^{k-1} \text{ and } \mathrm{Var}(Y) \leq \mu.$$

So we can say that

$$\begin{aligned}
\mathbb{P}\left(|N_k(v)| \geq (2np)^k \mid A_1, A_2, \ldots, A_{k-1}\right) &\leq \mathbb{P}\left(|Y - \mathbb{E}(Y)| \geq \mu\right) \\
&\leq e^{-\frac{\mu^2}{2(\mu + \frac{\mu}{3})}} = e^{-\frac{3}{8}\mu} \\
&\leq e^{-\frac{3}{8}2^{k-1}(np)^k}.
\end{aligned}$$

Note that we have that

$$np = n^{\frac{1}{d}} \left(\log\left(\frac{n^2}{c}\right)\right)^{\frac{1}{d}}$$

and so,

$$(np)^k \geq np = \omega(n^{\frac{1}{d}})$$

Hence

$$\mathbb{P}\left(|N_k(v)| \geq (2np)^k \mid A_1, A_2, \ldots, A_{k-1}\right) = o(n^{-1}).$$

The claim then follows by observing that for any v,

$$\mathbb{P}\left(\bigcup_{k=1}^{d-1} N_k(v) = [n]\right) \leq \sum_{k=1}^{d-1} \mathbb{P}(\overline{A_k} \mid A_1, A_2, \ldots, A_{k-1}) = o(n^{-1}),$$

and so with high probability not only is the diameter of G at least d, but, by the union bound, for every vertex v there is some vertex w at distance at least d from v, that is, even the radius of G is at least d.

□

We note that the above theorem is close to tight, and mention for the sake of completeness how the parameter c affects the diameter of $G(n,p)$ (and to explain why we mentioned it at all).

Theorem 8.5. *Let $d \geq 2$ be fixed. Suppose $c > 0$ and $p^d n^{d-1} = \log(n^2/c)$. Then (as $n \to \infty$) with probability $e^{-c/2}$ the diameter of $G(n,p)$ is d and with probability $1 - e^{-c/2}$ the diameter is $d+1$.*

8.3 Combinatorial Discrepancy

Consider a general set system $(V, \mathcal{F})$ where $V = [n]$ is the ground set and $\mathcal{F}$ is a family of subsets of $[n]$. Suppose we want to 2-colour $\mathcal{F}$, say in colours red and blue, such that colouring is 'balanced' in each member of $\mathcal{F}$, that is, each $F \in \mathcal{F}$ contains about as many blue points as red points.

This is clearly not always possible, since if we let $\mathcal{F} = 2^{[n]}$ then some set of size $\geq n/2$ will be monochromatic (as will all its subsets). In what situations can we say more? It will be convient to think of the two colouring, rather than as a map $\chi : V \to \{0,1\}$, as instead a map $\chi : V \to \{-1,+1\}$. Then for any $F \in \mathcal{F}$ we can define

$$\chi(F) = \sum_{i \in F} \chi(i)$$

and see that $\chi(F)$ is the difference between the number of 'red' and 'blue' points in F. We define the *discrepancy of $\mathcal{F}$ with respect to χ* by

$$\mathrm{disc}_\chi(\mathcal{F}) = \max_{F \in \mathcal{F}} |\chi(F)|$$

and the *discrepancy* of $\mathcal{F}$ to be

$$\mathrm{disc}(\mathcal{F}) = \min_\chi \mathrm{disc}_\chi(\mathcal{F}).$$

Using Chernoff's bound we can say that, if $|\mathcal{F}|$ is not too large, then the discrepancy is not much larger than $\sqrt{n}$.

Theorem 8.6. *Let $(V, \mathcal{F})$ be a set system such that $|V| = n$ and $|\mathcal{F}| = m$. Then if the maximum size of a set $F \in \mathcal{F}$ is s*

$$\mathit{disc}(\mathcal{F}) \leq \sqrt{2s \log(2m)}.$$

In particular, for any $\mathcal{F}$

$$\mathit{disc}(\mathcal{F}) \leq \sqrt{2n \log(2m)}.$$

Proof. We choose a colouring $\chi : V \to \{-1, +1\}$ randomly by choosing $\chi(v)$ for each $v \in V$ independently and uniformly at random. For $F \in \mathcal{F}$ we have that $\chi(F)$ is the sum of $|F|$ independent random ± 1 variables. Therefore, by Theorem 8.1, for any $t \geq 0$

$$\mathbb{P}\left(|\chi(F)| \geq t\right) < 2e^{-\frac{t^2}{2|F|}} \leq 2e^{-\frac{t^2}{2s}}.$$

If we let $t = \sqrt{2s \log(2m)}$ we see that

$$\mathbb{P}\left(|\chi(F)| \geq t\right) < 2e^{-\frac{2s \log(2m)}{2s}} = 2e^{-\log(2m)} = \frac{1}{m}.$$

Therefore

$$\mathbb{P}(\mathrm{disc}_\chi(\mathcal{F}) \leq t) = \mathbb{P}\left(\bigcup_{F \in \mathcal{F}} |\chi(F)| \leq t\right) \geq 1 - \sum_{F \in \mathcal{F}} \mathbb{P}\left(|\chi(F)| \geq t\right) > 1 - \frac{m}{m} = 0$$

and so with a positive probability a random colouring is a witness that $\mathrm{disc}(\mathcal{F}) \leq \sqrt{2s \log(2m)}$. □

8.4 A Lower Bound for the Binomial Distribution

Sometimes we also need a lower bound for probabilities like $\mathbb{P}(X \geq \mathbb{E}(X) + t)$, that is we need to know that the probability of a deviation of at least t is not *too* small. One can think of theorems like Theorem 8.1 in terms of the central limit theorem. That is, a sum of many independent random variables should approach a normal distribution. So the bounds in Theorem 8.1 and Theorem 8.3 should not be too far from the actual probability.

Let us first prove a counterpart to the weaker Theorem 8.9, the proof of which is just an exercise in elementary estimates.

Theorem 8.7. *For n even, let $X_1, X_2, \ldots, X_n$ be independent random variables taking the values 1 and -1, each with probability $1/2$. Let $X = X_1 + X_2 + \ldots + X_n$. Then we have, for any integer $t \in [0, n/8]$,*

$$\mathbb{P}\left(X \geq 2t\right) \geq \frac{1}{15} e^{-\frac{16t^2}{n}},$$

Proof. Let us write $n = 2m$. We can estimate the probability that $X \geq t$ by counting the number of events in the probability space (of size 2^n) where $X \geq t$, and so

$$\begin{aligned}
\mathbb{P}(X \geq 2t) &= 2^{-2m} \sum_{j=t}^{m} \binom{2m}{m+j} \\
&\geq 2^{-2m} \sum_{j=t}^{2t-1} \binom{2m}{m+j} \\
&= 2^{-2m} \sum_{j=t}^{2t-1} \binom{2m}{m} \frac{m}{m+j} \cdot \frac{m-1}{m+j-1} \cdots \frac{m-j+1}{m+1} \\
&\geq \frac{1}{2\sqrt{m}} \sum_{j=t}^{2t-1} \prod_{i=1}^{j} \left(1 - \frac{j}{m+i}\right) \text{ (using } \binom{2m}{m} \geq \frac{2^{2m}}{2\sqrt{m}}) \\
&\geq \frac{t}{2\sqrt{m}} \left(1 - \frac{2t}{m}\right)^{2t} \\
&\geq \frac{t}{2\sqrt{m}} e^{-\frac{8t^2}{m}} \text{ (since } 1 - x \geq e^{-2x} \text{ for } 0 \leq x \leq \frac{1}{2})
\end{aligned}$$

Where the last inequality is only true for $t \leq n/8$. When $t \geq \sqrt{m}/4$, the last expression is at least

$$\frac{1}{8} e^{-\frac{16t^2}{n}}.$$

For $0 \leq t < \sqrt{m}/4$, we have that

$$\mathbb{P}(X \geq 2t) \geq \mathbb{P}\left(X \geq 2\left(\frac{\sqrt{m}}{4}\right)\right) \geq \frac{1}{8} e^{-\frac{1}{2}} \geq \frac{1}{15}.$$

Thus the bound holds for all $t \leq n/8$. We note that no real effort has been made to optimise the constants. □

We can use this lower bound to show that the upper bound of $O(\sqrt{n \log(2m)})$ for the discrepancy of a family of size m on a set $|V| = n$ is nearly the best possible for a wide range of m.

Theorem 8.8. *For any $n, m \in \mathbb{N}$ such that $15n \leq m \leq 2^{\frac{n}{8}}$ there exists a set system $(V, \mathcal{F})$ such that $|V| = n$ and $|\mathcal{F}| = m$ such that*

$$\mathit{disc}(\mathcal{F}) \geq \Omega\left(\sqrt{n \log\left(\frac{m}{15n}\right)}\right).$$

Proof. We consider a random set system $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ on the ground set $[n]$ where we choose each F_i by including each $x \in [n]$ in F_i independently with probability $1/2$. Note that $\mathbb{P}(F_i = F_j) = 2^{-n}$ and so, for $m \leq 2^{\frac{n}{8}}$ we have by a union bound that almost surely this produces m distinct sets. We will assume n is even to apply Theorem 8.7, obviously in that lemma we only restrict to the case that n is even for ease of presentation, and a similar bound there (and so here) will hold with n odd.

Let us consider a fixed colouring $\chi : [n] \to \{-1, +1\}$. Suppose that $\chi([n]) = n - 2a$, and so the number of $x \in [n]$ such that $\chi(x) = 1$ is $n - a$. A point x such that $\chi(x) = 1$ contributes 1

to $\chi(F_i)$ if $x \in F_i$ and 0 otherwise. Since $x \in F_i$ with probability $1/2$ the contribution of x to $\chi(F_i)$ is a random variable taking the values 0 and 1 each with probability $1/2$. Similarly the number of $x \in [n]$ such that $\chi(x) = -1$ is a and the contribution of each such x to $\chi(F_i)$ is a random taking the values 0 and -1 each with probability $1/2$.

Therefore $\chi(F_i)$, as a random variable, can be expressed of the sum of n independent random variables $X_1, X_2, \ldots, X_{n-a}$ and $Y_1, Y_2, \ldots, Y_a$ where the X_i take values 0 and 1 with probability $1/2$ and the Y_i takes values -1 and 0 with probability $1/2$. So if we re-scale we see that $(2X_i - 1)$ and $(2Y_i + 1)$ are independent random variables taking the values 1 and -1 each with probability $1/2$. Hence if we define

$$X := 2\chi(F_i) - (n - 2a) = \sum_{i=1}^{n-a}(2X_i - 1) + \sum_{j=1}^{a}(2Y_i + 1)$$

we can apply Theorem 8.7 to see that, for $a \leq n/2$

$$\mathbb{P}(|\chi(F_i)| \geq t) \geq \mathbb{P}(X \geq 2t - (n - 2a)) \geq \mathbb{P}(X \geq 2t) \geq \frac{1}{15} e^{-\frac{16t^2}{n}}$$

provided that $t < n/8$. By symmetry we get a similar bound for $a > n/2$ (for example by considering $-\chi$). Therefore for any of the possible 2^n different colouring χ we have, since the sets F_i were picked independently,

$$\mathbb{P}(\mathrm{disc}_\chi(\mathcal{F}) \leq t) \leq \left(1 - \frac{1}{15} e^{-\frac{16t^2}{n}}\right)^m \leq e^{-\frac{m}{15} e^{-\frac{16t^2}{n}}}$$

If we take

$$t = \sqrt{\frac{n}{16} \log\left(\frac{m}{15n}\right)},$$

which satisfies $t \leq n/8$ as long as $m \leq 2^{n/8}$, then this last expression becomes $e^{-n} < 2^{-n}$. Hence by the union bound with positive probability the discrepancy of $\mathcal{F}$ is at least t for every colouring χ. □

We note that for $m \geq n^2$ say, the lower and upper bounds are the same up to a constant. When m and n are close there is a gap, and it turns out that it is the upper bound that can be improved (by a complicated probabilistic argument) to match the order of the lower bound.

A general result, which we state without proof, in the other direction is

Theorem 8.9. *Let X be a sum of independent random variables, each taking values in $[0, 1]$, and let $\sigma = \sqrt{Var(X)} \geq 200$. Then for all $t \in [0, \sigma^2/100]$, we have*

$$\mathbb{P}(X \geq \mathbb{E}(X) + t) \geq c e^{-\frac{t^2}{3\sigma^2}}$$

for a suitable constant $c > 0$.

9 The Lovás Local Lemma

9.1 The Local Lemma

In a typical probabilistic proof of a combinatorial result, one has to show that the probability of a certain event is positive. As we have seen the in the last section, many of these proofs tend to show more, that the probability is not only positive but very large, often tending to 1 as the 'dimension' of the problem considered grows.

On the other hand, there is a trivial case in which one can show that a certain event holds with positive, even though very small, probability. Suppose we have n mutually independent events A_i, each of which hold with probability $p > 0$, then the probability that they all hold simultaneously is at least p^n, which is positive, but may be exponentially small in n.

It is natural to expect that something similar will be true if the events are not entirely independent, but only 'mostly independent', for some sensible definition of 'mostly independent'. One way to define it is as follows.

Definition. Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. A directed graph $D = ([n], E)$is called a *dependency digraph* for the events $A_1, A_2, \ldots, A_n$ if for all i the event A_i is mutually independent of all the events $\{A_j : (i,j) \notin D\}$.

So for example when $A_1, A_2, \ldots, A_n$ are all mutually independent a dependency graph is the empty graph E_n. Note that we are not simply insisting that A_i is independent of A_j if $(i,j) \notin E$, since then we could take an undirected graph, the property we are checking is stronger, and so it's not sufficient to simply put an edge in D between every pair of dependent events.

We might expect that there are some natural conditions which tell us that when the graph is sparse enough, there is some positive probability that all the events hold. In the following, to follow standard notations, we will think of the events which all happen with small probability as being the negation of a set of events A_i, which we will denote by $\overline{A_i}$. The condition we ask for the graph to satisfy will seem complicated, but will be informed by the proof.

We will use the following elementary result twice in the proof

Lemma 9.1.

$$\mathbb{P}(A|B \cap C) = \frac{\mathbb{P}(A \cap B|C}{\mathbb{P}(B|C)}.$$

Proof. By definition of conditional probability

$$\begin{aligned}\mathbb{P}(A|B \cap C) &= \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C} \\ &= \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C} \cdot \frac{\mathbb{P}(C)}{\mathbb{P}(B \cap C} \\ &= \frac{\mathbb{P}(A \cap B|C}{\mathbb{P}(B|C)}.\end{aligned}$$

□

Lemma 9.2 (The Lovás Local Lemma)**.** *Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. Suppose that $D = ([n], E)$ is a dependency digraph for the events $\{A_i\}_{i=1}^n$ and there exists $x_1, x_2, \ldots x_n \in [0, 1)$ such that*

$$\mathbb{P}(A_i) \leq x_i \prod_{(i,j)\in E} (1 - x_j)$$

for all $i \in [n]$. Then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) \geq \prod_{i=1}^{n} (1 - x_i).$$

In particular, with positive probability no event A_i holds.

Proof. We want to show that with positive probability none of the events A_i occur. This will be impossible if an occurence of some combination of the $\overline{A_j}$ made it very likely that some other A_i held. So we first try to bound the probability of each A_i, given that some of the other events do not occur.

We claim that for any $S \subset [n]$ and $i \notin S$

$$\mathbb{P}\left(A_i \mid \bigcap_{j\in S} \overline{A_j}\right) \leq x_i.$$

We do so by inducting on $|S|$. When $S = \emptyset$ we have from the assumption of the lemma that

$$\mathbb{P}(A_i) \leq x_i \prod_{(i,j)\in E} (1 - x_j) \leq x_i.$$

Now let us suppose that the claim holds for every S' with $|S'| < |S|$ and set $S_1 = \{j \in S : (i, j) \in E\}$, $S_2 = S \setminus S_1$. In particular A_i is mutually independent of the set $\{A_j : j \in S_2\}$. We first note that if $S_1 = \emptyset$ then A_i is independent of $\cap_{j\in S}\overline{A_j}$ and so the claim holds as before. So we can assume $S_1 \neq \emptyset$. We have by Lemma 9.1

$$\mathbb{P}\left(A_i \mid \bigcap_{j\in S} \overline{A_j}\right) = \frac{\mathbb{P}\left(A_i \cap \bigcap_{j\in S_1} \overline{A_j} \mid \bigcap_{l\in S_2} \overline{A_l}\right)}{\mathbb{P}\left(\bigcap_{j\in S_1} \overline{A_j} \mid \bigcap_{l\in S_2} \overline{A_l}\right)}.$$

We bound the numerator by the fact that A_i is mutually independent of the events $\{A_l : l \in S_2\}$ and so

$$\mathbb{P}\left(A_i \cap \bigcap_{j\in S_1} \overline{A_j} \,|\, \bigcap_{l\in S_2} \overline{A_l}\right) \leq \mathbb{P}\left(A_i \,|\, \bigcap_{l\in S_2} \overline{A_l}\right)$$
$$= \mathbb{P}(A_i) \leq x_i \prod_{(i,j)\in E} (1-x_j).$$

The denominator on the other hand can be bounded by the induction hypothesis. Indeed, let $S_1 = \{j_1, j_2, \ldots, j_r\}$ then, again using Lemma 9.1

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{j\in S_1} \overline{A_j} \,|\, \bigcap_{l\in S_2} \overline{A_l}\right) &= \mathbb{P}\left(\overline{A_{j_1}} \cap \ldots \cap \overline{A_{j_r}} \,|\, \bigcap_{l\in S_2} \overline{A_l}\right) \\
&= \mathbb{P}\left(\overline{A_{j_1}} \,|\, \bigcap_{l\in S_2} \overline{A_l}\right) \times \mathbb{P}\left(\overline{A_{j_2}} \,|\, \overline{A_{j_1}} \cap \bigcap_{l\in S_2} \overline{A_l}\right) \times \ldots \\
&\ldots \times \mathbb{P}\left(\overline{A_{j_r}} \,|\, \overline{A_{j_1}} \cap \ldots \cap \overline{A_{j_{r-1}}} \cap \bigcap_{l\in S_2} \overline{A_l}\right) \\
&\geq (1-x_{j_1})(1-x_{j_2})\ldots(1-x_{j_r}) \\
&\geq \prod_{(i,j)\in E} (1-x_j).
\end{aligned}$$

Combining these two inequalities we see that, as claimed

$$\mathbb{P}\left(A_i \,|\, \bigcap_{j\in S} \overline{A_j}\right) \leq x_i.$$

The assertion of the lemma now follows easily, as

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) &= \left(1-\mathbb{P}\left(A_1\right)\right)\left(1-\mathbb{P}\left(A_2|\overline{A_1}\right)\right)\ldots\left(1-\mathbb{P}\left(A_n|\bigcap_{i=1}^{n-1}\overline{A_i}\right)\right) \\
&\geq \prod_{i=1}^{n}(1-x_i).
\end{aligned}$$

□

Often in application the sets A_i satisfy certain symmetric conditions which allow us to simplify the (rather complicated looking) conditions in Lemma 9.2.

Corollary 9.3. *[Symmetric Local Lemma] Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all but at most d of the other A_j (equivalently there is a dependency digraph with all outdegrees less than d), and that $\mathbb{P}(A_i) \leq p$ for all i. If $ep(d+1) \leq 1$ then*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) > 0.$$

Proof. If $d = 0$ then the events are mutually independent and the result follows trivially. Otherwise let $x_i = 1/(d+1) < 1$. There is a dependency digraph $D = (V, E)$ such that all outdegrees are less than d and so

$$x_i \prod_{(i,j)\in E} (1 - x_j) \geq \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d.$$

Note that it is a simple check that

$$\left(1 - \frac{1}{d+1}\right)^d = \left(1 + \frac{1}{d}\right)^{-d} > e^{-1}$$

and so

$$x_i \prod_{(i,j)\in E} (1 - x_j) > \frac{1}{e(d+1)} \geq p \geq \mathbb{P}(A_i).$$

Therefore by Lemma 9.2 the conclusion holds. □

9.2 Ramsey Bounds for the last time

We can use the Local Lemma to get a slight (multiplicative) improvement to the lower bound on the diagonal Ramsey numbers from Section 4.

Theorem 9.4. *If*

$$e\binom{k}{2}\binom{n-2}{k-2}2^{1-\binom{k}{2}} < 1$$

then $R(k, k) > n$.

Proof. We consider a random two colouring of the edges of the graph K_n and for each $S \subset [n]$ with $|S| = k$ let A_S be the event that S is monochromatic. We have that $\mathbb{P}(A_S) = 2^{1-\binom{k}{2}}$ for each S. We also have A_S is mutually independent of the set of A_T where S and T don't share an edge, and we can crudely bound above the number of those remaining by $\binom{k}{2}\binom{n-2}{k-2} - 1$. We then apply Corollary 9.3 with $p = 2^{1-\binom{k}{2}}$ and $d = \binom{k}{2}\binom{n-2}{k-2} - 1$ to see that the claim holds. □

A short computation shows that this gives $R(k, k) > \left(\sqrt{2}/e\right)(1 + o(1))k2^{k/2}$, which is only an improvement of $\sqrt{2}$ from our previous bound. This is not too surprising, since we expect the Local Lemma to be at it's most powerful when the dependencies between the events are rare,

which is not the case here. Indeed, there are a total number of $K = \binom{n}{k}$ events considered and the maximum outdegree d of the dependency digraph is $\sim \binom{k}{2}\binom{n-2}{k-2}$. For large k and larger n, which is the case of interest for us, we have that $d > K^{1-o(1/k)}$.

In contrast if we consider small sets, for example $k = 3$, we see that out of the $K = \binom{n}{3}$ 3-sets, each shares an edge with only $3(n-2) \sim K^{1/3}$ of the other sets. Therefore, we might expect the Local Lemma to give better improvements for the off-diagonal Ramsey numbers $R(k, l)$ for small fixed l. Indeed, by a straightforward application of Lemma 9.2 it is possible to show that for any k, n, if there exists $p \in (0, 1)$ and $x, y \in [0, 1)$ such that

$$p^3 \leq x(1-x)^{3n}(1-y)^{\binom{n}{k}} \text{ and } (1-p)^{\binom{k}{2}} \leq y(1-x)^{\frac{k^2 n}{2}}(1-y)^{\binom{n}{k}},$$

then $R(k, 3) > n$. It is a tedious (although elementary) computation to find the largest possible k such that there is a choice of p, x and y, eventually showing that

$$R(k, 3) \geq c\frac{k^2}{(\log(k))^2}$$

for some constant c, which matches a lower bound of Erdős proved with a highly complicated probabilistic argument, and is within a log factor of being best possible. A similar argument shows that

$$R(k, 4) \geq k^{\frac{5}{2}+o(1)}$$

which is better than any known bound for $R(k, 4)$ known without using the Local Lemma.

9.3 Directed Cycles

In this section we present another application of Corollary 9.3

Theorem 9.5. *Let $D = (V, E)$ be a directed graph with minimum outdegree δ and maximum indegree Δ. Then for any $k \in \mathbb{N}$ such that*

$$k \leq \frac{\delta}{1 + \log(1 + \delta\Delta)},$$

D contains a directed cycle of length divisible by k.

Proof. In order to simplify the computations we first delete all but δ outgoing edges from each vertex in V to find a subgragh $D' = (V, E')$ of D where each outdegree is exactly δ, and the maximum indegree is still less than Δ. Note that any cycle in D' is also in D.

We take a random k-colouring of the vertices of D, that is we pick a function $f : V \to [k]$ by choosing $f(v)$ for each $v \in V$ independently and uniformly from $[k]$. For each $v \in V$ we let

$$N^+(v) = \{w : (v, w) \in E'\}$$

be the set of outneighbours of v and let A_v be the event that no vertex in $N^+(v)$ is coloured by $f(v)+1 \pmod k$.

Note that the probability of A_v is $p = (1-1/k)^\delta$ for each $v \in V$. We claim that each A_v is mutually independent of the set of events A_w such that $N^+(v) \cap (N^+(w) \cup \{w\}) = \emptyset$. That is, A_v is mutually independent of the set eof events A_w such that w is not an outneighbour of v and w and v have no common outneighbour. Note that v may still be a successor of w. Indeed, if we condition on a specific value of f on $V \setminus N^+(v)$ (even on v itself), this doesn't change the probability of A_v, but it determines A_w for all such w.

The number d of vertices w not satisfying the above conditions is at most $\delta + \delta(\Delta - 1) = \delta\Delta$. Therefore

$$ep(d+1) \le e(1-\frac{1}{k})^\delta(\delta\Delta+1) \le e^{1-\frac{\delta}{k}}(\delta\Delta+1) \le 1,$$

and so by the Local Lemma there is a colouring such that no event A_v holds, that is for every vertex $v \in V$ there is a $w \in N^+(v)$ such that $f(w) = f(v)+1 \bmod(k)$. So, starting at any vertex let us generate a sequence of vertices $v_0, v_1, \ldots$ such that $(v_i, v_{i+1}) \in E'$ and $f(v_{i+1}) = f(v_i)+1 \bmod(k)$. Such a sequence must eventually contain a directed cycle, and the colouring guarantees that the length of the cycle will be divisible by k. □

9.4 The Linear Aboricity of Graphs

Definition. Given a graph G the *aboricity* of G, $\mathrm{a}(G)$, is the minimum number of forests into which the edge set $E(G)$ can be partitioned. A *linear forest* is a forest in which every component is a path, and the *linear aboricity* of a graph, $\mathrm{la}(G)$, is the minimum number of linear forests into which the edge set $E(G)$ can be partitioned.

The following simple conjecture is longstanding.

Conjecture 9.6 (The Linear Aboricity Conjecture)**.** *Let G be a d-regular graph. Then*

$$la(G) = \left\lceil \frac{d+1}{2} \right\rceil.$$

Note that since every d-regular graph on n vertices has $nd/2$ edges and every linear forest has at most $n-1$ edges we have that $\mathrm{la}(G) > d/2$, and so the content of the conjecture is to show that every d-regular graph can indeed by decomposed into a small number of forests. Also, since every graph of maximum degree Δ can be embedded into a Δ-regular graph, the conjecture is equivalent to the statement that every G with satisfies $\mathrm{la}(G) \le \lceil (\Delta(G)+1)/2 \rceil$.

Much work has been done towards the conjecture and the best known bound without a probabilistic argument was that $\mathrm{la}(G) \lesssim 3\Delta(G)/5$. .

It will be convenient to deduce the result from a corresponding result for directed graphs. A *d-regular digragph* is a directed graph in which the indegree and outdegree of every vertex is precisely d. A *linear directed forest* is a directed graph in which every connected component is

a directed path and the *dilinear aboricity* of a directed graph D, which we denote by $\mathrm{dla}(D)$, is the minimum number of linear directed forests into which the edge set $E(G)$ can be partitioned. We then have the directed version of the Linear Aboricity Conjecture.

Conjecture 9.7. *Let D be a d-regular digraph. Then*

$$dla(D) = d + 1.$$

Note that since the edges of any connected undirected $2d$-regular graph can be oriented along an Euler cycle, so the the resulting digraph is d-regular, Conjecture 9.7 for d implies Conjecture 9.6 for $2d$.

It is a simple exercise to show that a graph G contains an independent set of size at least $n/(\Delta(G)+1)$. We will require for our proof a lemma that tells us that, at the price of decreasing the size by a constant factor, we can find a large independent set with additional structure

Lemma 9.8. *Let $H = (V, E)$ be a graph with maximum degree Δ, and let $V = V_1 \cup V_2 \cup \ldots \cup V_r$ be a partition of V into r pairwise disjoint sets. Suppose that $|V_i| \geq 2e\Delta$ for each $i \in [r]$. Then there is an independent set $W \subset V$ that contains a vertex from each V_i.*

Proof. Without loss of generality we may assume that $|V_i| = \lceil 2e\Delta \rceil = g$ for each i. We pick a single vertex from each V_i independently and uniformly at random and let W be the union of these vertices. We will show that with positive probability W is independent.

For each edge $f \in H$ let A_f be the event that both ends of f are contained in W. Clearly $\mathbb{P}(A_f) \leq \frac{1}{g^2}$. Moreover, if the endpoints of f are in V_i and V_j respectively then A_f is mutually independent of all the events A_e such that the endpoints of e do not lie in $V_i \cup V_j$.

Therefore there is a dependency digraph for the events A_f in which the maximum degree is $< 2g\Delta$. Since $e.2g\Delta.(1/g^2) = 2e\Delta/g < 1$ we have by Corollary 9.3 that with positive probability none of the events A_f hold. However this means that W is an independent set containing a vertex from each V_i. □

We note at this point that Lemma 9.8 enables us to prove Conjecture 9.7 for digraphs with no short directed cycle. The *directed girth* of a graph is the minimum length of a directed cycle in that graph.

Theorem 9.9. *Let $D = (V, E)$ be a d-regular directed graph with directed girth $g \geq 8ed$. Then*

$$dla(D) = d + 1.$$

Proof. We first use Hall's theorem to decompose D into d-pairwise disjoint 1-regular spanning subgraphs, $D_1, D_2, \ldots, D_d$. Strictly we form a bipartite graph on (A, B) where $A = B = V$ and let (a, b) be an edge if and only if (a, b) is a directed edge in D. Then this bipartite graph is d-regular and so we can decompose it into d pefect matchings, each of which correspond to a 1-regular spanning directed subgraph.

Each D_i is a union of vertex disjoint directed cycles, $C_{i_1}, C_{i_2}, \dots C_{i_{r_i}}$. Let $E_1, E_2, \dots, E_r$ the edge sets of each of these cycles, taken over each $1 \leq i \leq d$. We have that $\{E_i\}$ is a partition of the edge set of D and by the girth condition each $|E_i| \geq g \geq 8ed$.

We consider the (undirected) line graph L of D, that is the graph whose vertex set is E and two edges are adjacent if and only if they share a vertex. Note that L is $4d-2$ regular and $\{E_i\}$ is now a partition of the vertex set of L. Since $|E_i| \geq 8ed \geq 2e(4d-2)$ we can apply Lemma 9.8 to L to find an independent set in L containing an element of each E_i. However this corresponds to a matching M in D containing at least one edge from each cycle C_{i_j}.

Hence if we consider the subgraphs $D_1 \setminus M, D_2 \setminus M, \dots, D_d \setminus M, M$ we see that each $D_i \setminus M$ is a linear directed forest, and M is a matching, and between them they cover the edges of D. Hence

$$\mathrm{dla}(D) \leq d + 1.$$

Finally we note that, as before, D has $|V|d$ edges and each directed linear forest can have at most $|V| - 1$ edges, and so

$$\mathrm{dla}(D) \geq \frac{|V|d}{|V| - 1} > d.$$

Therefore $\mathrm{dla}(D) = d + 1$ as claimed.

□

In order to prove the theorem for general digraphs we show that we can decompose almost all of the edges of a regular digraph into a relatively small number of regular digraphs with large girth. To do so we need the following technical lemma, which also uses the Local Lemma in its proof.

Lemma 9.10. *Let $D = (V, E)$ be a d-regular directed graph, where d is sufficiently large, and let p be an integer such that $10\sqrt{d} \leq p \leq 20\sqrt{d}$. Then there is a p-colouring of V, $f : V \to [p]$, such that, for each $v \in V$ and each $i \in [p]$ the number*

$$N^+(v, i) = |\{u \in V \,:\, (v, u) \in E \text{ and } f(u) = i\}|$$

and

$$N^-(v, i) = |\{u \in V \,:\, (u, v) \in E \text{ and } f(u) = i\}|$$

satisfy

$$\left|N^+(v, i) - \frac{d}{p}\right|, \left|N^-(v, i) - \frac{d}{p}\right| \leq 3\sqrt{\frac{d}{p}\log(d)}.$$

Proof. We pick a random p-colouring $f : V \to [p]$ by choosing $f(v)$ for each $v \in V$ independently and uniformly at random from $[p]$. For each $v \in V$ and $i \in [p]$ let $A^+_{v,i}$ be the event that

$$\left|N^+(v, i) - \frac{d}{p}\right| > 3\sqrt{\frac{d}{p}\log(d)}$$

and similarly for $A^-_{v,i}$. We have that $N^+(v,i)$ is a binomial random variable with expectation d/p, so if we let $t = 3\sqrt{\frac{d}{p}\log(d)}$ then, by Theorem 8.3 we have that

$$\mathbb{P}(A^+_{v,i}) < e^{-\frac{t^2}{2(\frac{d}{p}+\frac{t}{3})}} \leq e^{-\frac{9\frac{d}{p}\log(d)}{3\frac{d}{p}}} \leq d^{-3}$$

and similarly for $A^-_{v,i}$. Clearly each event $A^+_{v,i}, A^-_{v,i}$ is mutually independent of all the events $A^+_{u,i}, A^-_{u,i}$ such that u and v do not have a common neighbour. Therefore there is a dependency digraph for these events with maximum degree $\leq (2d^2)p$. Since

$$e\frac{1}{d^3}((2d)^2p+1) \leq 1$$

we have that by Corollary 9.3 there is a non-zero probability that none of the events $A^+_{v,i}, A^-_{v,i}$ happen. Therefore there is a colouring f satisfying the required properties.

□

We are now ready to argue the general case.

Theorem 9.11. *There exists a constant $c > 0$ such that for every d-regular digraph D*

$$dla(D) \leq d + cd^{\frac{3}{4}}(\log(d))^{\frac{1}{2}}.$$

Proof. Let $D = (V,E)$ be an arbitrary d-regular digraph. Let p be a prime satisfying $10\sqrt{d} \leq p \leq 20\sqrt{d}$ (which exists by Bertrand's postulate). By Lemma 9.10 there is a p-colouring of V, f, such that the conclusions of the lemma hold. For each $i \in [p]$ let $D_i = (V, E_i)$ be the spanning subgraph of D defined by

$$E_i = \{(u,v) \in E : , f(v) \equiv f(u) + i \bmod(p)\}.$$

By assumption we have that the maximum outdegree Δ^+_i and the maximum indegree Δ^-_i of G_i are at most

$$\frac{d}{p} + 3\sqrt{\frac{d}{p}\log(d)}.$$

Moreover, for each $i \neq p$, the length of every directed cycle in G_i is divisible by p, and so each G_i has directed girth $g_i \geq p$. It is a simple exercise to see that G_i can be completed, by adding vertices and edges, to a Δ_i-regular digraph with $\Delta_i = \max(\Delta^+_i, \Delta^-_i)$, which has the same directed girth g_i. Since $g_i > 8e\Delta_i$ (for all sufficiently large d) we have that by Theorem 9.9 that, for each $i \neq p$

$$\mathrm{dla}(G_i) \leq \Delta_i + 1 \leq \frac{d}{p} + 3\sqrt{\frac{d}{p}\log(d)} + 1.$$

To bound the size of G_p we see that, since G_p can be partitioned into Δ_p disjoint 1-regular spanning subgraphs, we can split each of these into two linear directed forest to get that

$$\mathrm{dla}(G_p) \leq 2\Delta_p \leq 2\frac{d}{p} + 6\sqrt{\frac{d}{p}\log(d)}.$$

These last two inequalities together with the fact that $10\sqrt{d} \leq p \leq 20\sqrt{d}$ imply that

$$\begin{aligned}
\mathrm{dla}(G) &\leq (p-1)\left(\frac{d}{p} + 3\sqrt{\frac{d}{p}\log(d)} + 1\right) + 2\frac{d}{p} + 6\sqrt{\frac{d}{p}\log(d)} \\
&= d + \frac{d}{p} + (p-1) + 3(p-1)\sqrt{\frac{d}{p}\log(d)} + 6\sqrt{\frac{d}{p}\log(d)} \\
&\leq d + cp\sqrt{\frac{d}{p}\log(d)} \\
&\leq d + cd^{\frac{3}{4}}(\log(d))^{\frac{1}{2}}
\end{aligned}$$

□

Since any $2d$-regular graph G can be oriented so the resulting digraph is d-regular (and since every $2d-1$-regular G is a subgraph of a $2d$-regular graph), we have as an immediate corollary of Theorem 9.11.

Corollary 9.12. *There exists a constant $c > 0$ such that for every d-regular graph G*

$$la(G) \leq \frac{d}{2} + cd^{\frac{3}{4}}(\log(d))^{\frac{1}{2}}.$$

10 Martingales and Strong Concentration

10.1 The Azuma-Hoeffding Inequality

The strong concentration bounds we obtained in Section 8 only applied to sums of random variables which were independent. This is quite a strong condition to ask for, and in this section we will prove strong concentration results of a similar nature that do not need to rely on independence. Suppose we play a game where a fair coin is continually flipped and each time it comes up heads you win Â£1 and each times it comes up tails you lose Â£1. The expected change in your bankroll for each flip is 0 and, by the chernoff bounds, we know that as the number of flips gets very large it is exponentially unlikely that you end up with a large difference between the money you start with and the money you ended with.

Now suppose that instead that instead of a fair coin flip at each stage an arbitrary opponent can choose a bet to make, that can depend on the result of the previous bets, with only the added condition that the bet is fair, that is the expected change in your bankroll at each stage is 0.

Let Z_i denote the outcome of the ith bet and let X_i be your bankroll after that bet. Let us make the following definition:

Definition. Let Z and X be random variables on the same probability space. We define the random variable $\mathbb{E}(X|\sigma(Z))$ by

$$\mathbb{E}(X|\sigma(Z))(\omega) = \mathbb{E}(X|Z = Z(\omega)).$$

Then, the preceding property can be written as $\mathbb{E}(X_i \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1})) = X_{i-1}$. Although the variable X_i is not independent of the previous variables, it is still true that, whatever their values were, we don't expect our bankroll to change from each bet. If the value of each bet is also small, say bounded by some constant number, then we should expect a similar bound to hold.

Motivated by this we define:

Definition. Let $Z_1, Z_2, \ldots, Z_n$ and $X_0, X_1, \ldots, X_n$ be sequences of random variables on the same probability space such that X_i is determined by $\{Z_1, Z_2, \ldots, Z_i\}$ and, for all i,

$$\mathbb{E}(X_i \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1})) = X_{i-1}.$$

Then (X_i) is called a *martingale* with respect to (Z_i).

The actual definition of a martingale is slightly more complicated. One normally takes a sequence of σ-fields $\{\emptyset, \Sigma\} = \Sigma_0 \subseteq \Sigma_1 \subseteq \ldots \subseteq \Sigma_n$, which we call a *filtration* on a probability space and a sequence (X_i) of random variables such that X_i is measurable with respect to Σ_i. Then we call (X_i) a martingale with respect to (Σ_i) if $\mathbb{E}(X_i \,|\, \Sigma_{i-1}) = X_{i-1}$.

In the finite case these Σ_is correspond to partitions of Ω. We can think of a martingale as being a gradual exposure of a random variable X_n on the probability space $(\Omega, \Sigma_n, \mathbb{P})$ by taking finer and finer estimates. Each X_i tells me what I expect X_n to be on average over a set of

events A which lies in Σ_i and partition Ω. In the next step Σ_{i+1} splits each A into some smaller sets of events, and the values of X_{i+1} on these smaller sets will tell me more precisely what X_n looks likes. However in order for this to be consistent we want that the average of X_{i+1} over the sets which partition A should be the same as $X_i(A)$, since we're thinking of both as representing in some way the expectation we have of X_n.

In the first step we have that X_0 is just a constant function whose value is the expected value of X_n over the entire space, and in the last step we just have X_n.

In applications, which motivates the previous definition, we usually take these filtrations as being defined by a set of random variables (Z_i), and so Σ_i corresponds to the partition of Ω into sets which agree on all $Z_1, Z_2, \ldots, Z_i$. We are then thinking about the space $(\Omega, \Sigma_n, \mathbb{P})$ almost as a product space over the smaller spaces defined by the random variables (Z_i). If the Z_i were independent this would be a product space in a more traditional sense.

One reason martingale methods can be applied to so many different cases is that we can form a martingale from essentially any random variable.

Lemma 10.1. *Let A and (Z_i) be random variables on the same probability space. Then $X_i = \mathbb{E}(A \,|\, \sigma(Z_1, Z_2, \ldots, Z_i))$ is a martingale with respect to (Z_i).*

Proof. Note firstly that X_i is determined by $\{Z_1, Z_2, \ldots, Z_i\}$. Also, for all i we have that

$$\mathbb{E}(X_i \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1})) = \mathbb{E}(\mathbb{E}(A \,|\, \sigma(Z_1, Z_2, \ldots, Z_i)) \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1})).$$

However it is clear that the above expectation is, for given $Z_1, Z_2, \ldots, Z_{i-1}$, averaging over all possible values of Z_i the expected value of $(A|Z_1, Z_2, \ldots, Z_i)$. Hence

$$\mathbb{E}(X_i \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1})) = \mathbb{E}(A \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1})) = X_{i-1}.$$

□

There are two very natural examples of martingales on graphs that we will want to apply the results in this section to, the *edge exposure martingale* and the *vertex exposure martingale*. For the first we take our underlying probability space to be $\mathcal{G}(n,p)$ and order the set of potential edges $(i,j) \in [n]^{(2)}$ arbitrarily as $e_1, e_2, \ldots, e_m$, where $m = \binom{n}{2}$. We define the random variables Z_i to be the indicator function of the event that e_i is an edge in $G(n,p)$. Any graph theoretic function f is some function of the random variables (Z_i). Therefore if we consider the martingale $X_i = \mathbb{E}(f \,|\, \sigma(Z_1, Z_2, \ldots, Z_i))$ with respect to (Z_i) we have that X_i is a martingale, $X_0 = \mathbb{E}(f)$ and $X_n = f$. We can think of this martingale as revealing the edges of $G(n,p)$ one by one, for any graph theoretic function the martingale at step i is the expected value of that function on a random $G(n,p)$, that agrees with the first i exposed edges (either being in or out).

For the vertex exposure martingale we take $Z_i \in \{0,1\}^{i-1}$ to be the vector of indicators of whether the edge between the vertex i and $j < i$ is in $G(n,p)$. Again for any graph theoretic function f we can consider the corresponding martingale $X_i = \mathbb{E}(f \,|\, \sigma(Z_1, Z_2, \ldots, Z_i))$. Similar to before we can think of this martingale as revealing the vertices (and edges adjacent to them) of $G(n,p)$ one by one, for any graph theoretic function the martingale at step i is the expected value of that function on a random $G(n,p)$, that agrees with the first i exposed vertices. Note that here we have $X_1 = \mathbb{E}(f)$ in contrast to the previous example, since exposing the first vertex gives no information, so the length of the martingale sequence here is really $(n-1)$.

Our main tool will be the following concentration result for martingales known as Azuma's inequality, similar versions of which were proved concurrently by multiple authors including Azuma, Hoeffding and Steiger.

Theorem 10.2 (The Azuma-Hoeffding Inequality)**.** *Let $c_1, \ldots, c_n > 0$ and let $(X_i)_0^n$ be a martingale with respect to $(Z_i)_1^n$ such $|X_i - X_{i-1}| \leq c_i$ for all $1 \leq i \leq n$ then*

$$\mathbb{P}(X_n \geq X_0 + t) \leq e^{-\frac{t^2}{2\sigma^2}} \text{ and } \mathbb{P}(X_n \leq X_0 - t) \leq e^{-\frac{t^2}{2\sigma^2}}$$

where $\sigma^2 = \sum_{i=1}^n c_i^2$.

Proof. We will need a simple technical lemma

Lemma 10.3. *Let Y be a random variable which takes values in $[-1, +1]$ such that $\mathbb{E}(Y) = 0$. Then for any $t \geq 0$*

$$\mathbb{E}(e^{tY}) \leq e^{\frac{t^2}{2}}.$$

Proof. We first note that for $t \geq 0$ the function e^{tx} is convex and so, for any $x \in [-1, +1]$

$$e^{tx} \leq \frac{1}{2}(1+x)e^t + \frac{1}{2}(1-x)e^{-t}.$$

Therefore by taking expectations of both sides we see

$$\begin{aligned}
\mathbb{E}(e^{tY}) &\leq \frac{1}{2}e^t + \frac{1}{2}e^{-t} \\
&= 1 + \frac{t^2}{4} + \frac{t^4}{4!} + \ldots \\
&= \sum \frac{t^2 n}{(2n)!} \leq \sum \frac{t^2 n}{2^n (n)!} \\
&= \sum \frac{\left(\frac{t^2}{2}\right)^n}{(n)!} = e^{\frac{t^2}{2}}.
\end{aligned}$$

□

From this point the proof is similar in nature to that of the Chernoff bound. For any $\lambda > 0$ we have

$$\mathbb{P}(X_n - X_0 \geq t) = \mathbb{P}\left(e^{\lambda(X_n - X_0)} \geq e^{\lambda t}\right).$$

Therefore by Markov's inequality we see

$$\begin{aligned}\mathbb{P}\left(e^{\lambda(X_n-X_0)} \geq e^{\lambda t}\right) &\leq e^{-\lambda t}\mathbb{E}\left(e^{\lambda(X_n-X_0)}\right)\\ &= e^{-\lambda t}\mathbb{E}\left(e^{\lambda\left(\sum_{i=1}^{n} X_i - X_{i-1}\right)}\right)\\ &= e^{-\lambda t}\mathbb{E}\left(\mathbb{E}\left(e^{\lambda\left(\sum_{i=1}^{n} X_i - X_{i-1}\right)} \mid \sigma(Z_1, Z_2, \dots Z_{n-1})\right)\right)\end{aligned}$$

However $X_1, \dots X_{n-1}$ are determined by $Z_1, Z_2, \dots, Z_{n-1}$ and so $(\sum_{i=1}^{n-1} X_i - X_{i-1} | \sigma(Z_1, Z_2, \dots Z_{n-1}))$ is constant. Therefore we can take that term outside the inner expectation to see that

$$\mathbb{P}(X_n - X_0 \geq t) \leq e^{-\lambda t}\mathbb{E}\left(e^{\lambda\left(\sum_{i=1}^{n-1} X_i - X_{i-1}\right)}\mathbb{E}\left(e^{\lambda(X_n - X_{n-1})} \mid \sigma(Z_1, Z_2, \dots Z_{n-1})\right)\right)$$

However now we can bound the inner expectation by noting that, $|X_n - X_{n-1}| \leq c_n$ and $\mathbb{E}(X_n - X_{n-1} \mid \sigma(Z_1, Z_2, \dots Z_{n-1})) = 0$. Therefore by the technical lemma, applied to $(X_n - X_0)/c_n$, we have that

$$\mathbb{P}(X_n - X_0 \geq t) \leq e^{-\lambda t} e^{\frac{\lambda^2 c_n^2}{2}} \mathbb{E}\left(e^{\lambda\left(\left(\sum_{i=1}^{n-1} X_i - X_{i-1}\right)\right)}\right) = e^{-\lambda t} e^{\frac{\lambda^2 c_n^2}{2}} \mathbb{E}\left(e^{\lambda(X_{n-1} - X_0)}\right)$$

However we can now handle the term in the expectation inductively in the same way to get

$$\mathbb{P}(X_n - X_0 \geq t) \leq e^{-\lambda t} e^{\frac{\lambda^2}{2}\sum_{i=1}^{n} c_i^2}.$$

Since the above holds for any $\lambda > 0$ we optimise by taking $\lambda = t/(\sum c_i^2)$, which gives

$$\mathbb{P}(X_n - X_0 \geq t) \leq e^{-\frac{t^2}{2\sum c_i^2}}.$$

□

Often we want to be able to deduce the boundedness condition $|X_i - X_{i-1}| \leq c_i$ from more local conditions. That is, since we can think of X_n as being a function $f(Z_1, Z_2, \dots, Z_n)$, then if changing any single co-ordinate doesn't affect the value of X very much, we would expect the differences to be bounded, since the range of $X_i - X_{i-1}$ will depend on how much our expectation of X_n changes once we learn what Z_i is (as long the later Z_j are not too dependent on this value).

In the particular case of graphs we say a graph theoretic functions f is *edge Lipschitz* if whenever H and H' differ in only one edge then $|f(H) - f(H')| \leq 1$. Equivalently, if we consider f as a function of the variables $f(Z_1, Z_2, \dots, Z_n)$ then we require that changing one coordinate does not change f by more than 1. Similarly it is *vertex Lipschitz* if whenever H and H' differ at only one vertex $|f(H) - f(H')| \leq 1$.

Lemma 10.4. *For any graph theoretic function f, if f is edge Lipschitz then the corresponding edge exposure martingale satisfies $|X_i - X_{i-1}| \leq 1$ and similarly if f is vertex Lipschtiz.*

Proof. We first note that in both cases the corresponding variables (Z_i) are all mutually independent. The proof of this lemma involves a clever trick called the duplication trick. For any i

let us consider a random variable Z_i' which has the same distribution as Z_i, but is independent of it.

$$\begin{aligned} X_{i-1} &= \mathbb{E}(f(Z_1, \ldots, Z_i, \ldots, Z_n) \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1})) \\ &= \mathbb{E}(f(Z_1, \ldots, Z_i', \ldots, Z_n) \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1})) \\ &= \mathbb{E}(f(Z_1, \ldots, Z_i', \ldots, Z_n) \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1}, Z_i)) \end{aligned}$$

Here we have replaced Z_i by Z_i' in the calculation of X_{i-1} so that we can condition on Z_i. The reason for this becomes clear when we look at

$$\begin{aligned} |X_i - X_{i-1}| &= |\mathbb{E}(f(Z_1, \ldots, Z_i, \ldots, Z_n) \,|\, \sigma(Z_1, Z_2, \ldots, Z_i)) - \mathbb{E}(f(Z_1, \ldots, Z_i, \ldots, Z_n) \,|\, \sigma(Z_1, Z_2, \ldots, Z_{i-1}))| \\ &= \mathbb{E}(f(Z_1, \ldots, Z_i, \ldots, Z_n) - f(Z_1, \ldots, Z_i', \ldots, Z_n) \,|\, \sigma(Z_1, Z_2, \ldots, Z_i))| \\ &\leq 1 \end{aligned}$$

Where the last inequality holds by the Lipshitz condition on any particular event, and so holds in the average case.

□

Combining Lemma 10.4 and Theorem 10.2 for a Lipschitz graph function f (and it's associated martingale) tells us that the difference between X_n and X_0, which are f and $\mathbb{E}(f)$ respectively, is quite small. In particular we can often say that a function is quite tightly concentrated around it's mean (even when we don't know what the mean is).

10.2 The Chromatic Number of a Dense Random Graph

We note that the chromatic number of a graph $\chi(G)$ is an edge Lipschitz function, it is a simple check that if we add or remove an edge from a graph it can change the chromatic number by at most one. However applying Theorem 10.2 to the edge exposure martingale will not tell us much. Indeed the Azuma-Hoeffding inequality, applied with $c_i = 1$, tells us that

$$\mathbb{P}(|\chi(G(n,p)) - \mathbb{E}(\chi(G(n,p))| \geq t) \leq e^{-\frac{t^2}{2m}}$$

where $m = \binom{n}{2} \sim n^2$ is the number of possible edges in $G(n,p)$. So,nto have the right hand side $\to 0$ we need $t = \omega(n)$. However χ only takes values in $[n]$, so in this case saying it is almost surely within $\omega(n)$ of it's expectation is not very useful.

However we can instead use the vertex exposure martingale. Again it is a simple check that χ is also a vertex Lipschitz function. In this case combining Lemma 10.4 and Theorem 10.2 gives us a better bound, first due to Shamir and Spencer

Theorem 10.5. *For any n and p and for all $t \geq 0$*

$$\mathbb{P}(|\chi(G(n,p)) - \mathbb{E}(\chi(G(n,p))| \geq t) \leq e^{-\frac{t^2}{2(n-1)}}.$$

Proof. As noted, χ is a vertex Lipschitz function. So by Lemma 10.4 we can apply Theorem 10.2 to the associated vertex exposure martingale with $c_i = 1$ to conclude that, for all $t \geq 0$

$$\mathbb{P}(|\chi(G(n,p)) - \mathbb{E}(\chi(G(n,p))| \geq t) \leq e^{-\frac{t^2}{2(n-1)}}.$$

□

Now we can take $t = \omega(\sqrt{n})$ to see that $\chi(G(n,p))$ is 'tightly' concentrated about its expectation, although we still don't know what it's expectation is.

Let us restrict attention to the case $p = 1/2$ for ease of presentation, although it is not too hard to adapt the following arguments to cover the case of any fixed p.

Since any colour class must be an independent set, for any graph G,

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

It is possible to use the second moment method to show that for any $\epsilon < 0$ almost surely $\alpha(G(n,1/2))$ is between $(2-\epsilon)\log_2(n)$ and $2\log_2(n)$. However the calculation of the second moment is particular lengthy and messy, so we will just state it. For those interested a proof can be found in Alon and Spencer. Therefore we have that almost surely

$$\chi(G(n,1/2)) \geq (1+o(1))\frac{n}{2\log_2(n)}.$$

The corresponding upper bound was an open question for almost 25 years.

Note that in $G(n,1/2)$ there is a natural equivalence between cliques and independent sets. In what follows we will consider cliques instead of independent sets, but only because "k-clique" is easier to type than "an independent set of size k". Let

$$f(k) = \binom{n}{k} 2^{-\binom{k}{2}}$$

and let k_0 be such that $f(k_0 - 1) > 1 > f(k_0)$, so that k_0 is the smallest integer such that the expected number of k_0 cliques in $G(n,1/2)$ is less than 1. This will be around $2\log_2(n)$ and, as we just stated without proof, almost surely $G(n,1/2)$ will have a clique of almost this size.

However, in the course of the proof, we will need an explicit bound on the probability that $G(n,1/2)$ contains a clique of size around k_0. So, we will first show that with very high probability the size of the largest clique in $G(n,1/2)$ is at least $k_1 = k_0 - 4$. It is not too hard to check that, as a function of n, $f(k_1) \to \infty$, and in fact more precisely that $f(k_1) = n^{3+o(1)}$, which we shall use later.

In order to show that with high probability $G(n, 1/2)$ contains a clique of size around k_1 we could look at the random variable X which counts the number of cliques of that size. Since we know the expected number of cliques of this size is around $f(k_1) = n^{3+o(1)}$ we might hope to use a tight concentration result to show that the probability that X is non-zero is exponentially small. However there is a problem in that the function X is not edge or vertex lipschitz, and so we can't apply Theorem 10.2.

We get around this in a clever way by first defining an auxilliary random varibale Y, which counts the maximal size of a family of edge disjoint cliques of size k_1 in $G(n, 1/2)$ (that is, the number of cliques it contains, not the union of their sizes). G will have no k_1-clique if and only if $Y = 0$ and since Y only counts families of edge disjoint k_1-cliques, changing one edge can only change the value of Y by one, and so it is edge Lipschitz. This will let us use Azuma-Hoeffding to say that Y is tightly concentrated about it's expectation. In order to then deduce that $Y \neq 0$ with high probability we will need to bound below its expectation, which is still slightly messy.

Lemma 10.6.

$$\mathbb{E}(Y) \geq (1 + o(1))\frac{n^2}{4k_1^4}.$$

Proof. Let $\mathcal{K}$ be the collection of all k_1-cliques in $G(n, 1/2)$, so that $\mathbb{E}(|\mathcal{K}|) = f(k_1) = \mu$. We will need to count the number of pairs of cliques $A, B \in \mathcal{K}$ whose edge sets intersect. If we let W denote the number of such pairs we see that

$$\mathbb{E}(W) = \mu \sum_{i=2}^{k_1-2} \binom{k_1}{i}\binom{n-k_1}{k_1-i} 2^{-(\binom{k_1}{2}-\binom{i}{2})}.$$

It is an involved, but elementary exercise, to show that this sum in dominated by the leading term $i = 2$. In this case we have, since $k_1 << n$,

$$\begin{aligned}
\frac{\mathbb{E}(W)}{\mu^2} &\sim \frac{1}{\mu}\binom{k_1}{2}\binom{n-k_1}{k_1-2}2^{-(\binom{k_1}{2}-1)} \\
&\sim 2\frac{\binom{k_1}{2}\binom{n-k_1}{k_1-2}}{\binom{n}{k_1}} \\
&\sim k_1^2\frac{\binom{n}{k_1-2}}{\binom{n}{k_1}} \\
&\sim k_1^2\frac{k_1^2}{n^2} \\
&\sim \frac{k_1^4}{n^2}
\end{aligned}$$

up to constant factors.

We choose a random set X from $\mathcal{K}$ by including each k_1-clique with probability q, where we will optimise q later. The expected size of X is just $q\mathbb{E}(|\mathcal{K}|)$ and the expected number of pairs in X which have non empty intersection is $q^2\mathbb{E}(W)$. Hence by the alteration method, if we remove one clique from each bad pair we are left with a set X' such that X' is a union of disjoint cliques and

$$\begin{aligned}\mathbb{E}(|X'|) &\geq q\mathbb{E}(|\mathcal{K}|) - q^2\mathbb{E}(W) \\ &\geq q\mu - q^2\mu^2\frac{k_1^4}{n^2}\end{aligned}$$

We apply the above with $q \sim n^2/2\mu k_1^4$. Note that, since $\mu = f(k_1) \sim n^{3+o(1)}$, $q < 1$. Hence

$$\mathbb{E}(Y) \geq \mathbb{E}(|X'|) \geq \frac{n^2}{4k_1^4}$$

□

Lemma 10.7.

$$\mathbb{P}(\omega(G(n,1/2)) < k_1) = e^{-n^{2+o(1)}}.$$

Proof. We apply the Azuma-Hoeffding inequality to the edge exposure martingale associated with Y. Since Y is edge Lipschitz we have that

$$\begin{aligned}\mathbb{P}(Y - \mathbb{E}(Y) \leq -\mathbb{E}(Y)) &\leq e^{-\frac{\mathbb{E}(Y)^2}{2\binom{n}{2}}} \\ &\leq e^{-(1+o(1))\frac{n^4}{16k_1^8}\frac{1}{n^2}} \\ &\leq e^{-(1+o(1))\frac{1}{16}\frac{n^2}{k_1^8}} \\ &= e^{-n^{2+o(1)}}\end{aligned}$$

□

We now have all the ingredients to present the following proof, due to Bollbás

Theorem 10.8. *Almost surely*

$$\chi(G(n,1/2)) \leq (1+o(1))\frac{n}{2\log_2(n)}.$$

Proof. The idea of the proof is as follows. If we consider the restriction of $G(n,1/2)$ to a subset S of a smaller, but still quite large size, say $|S| = m$, then this subgraph will look like $G(m,1/2)$.

We can use Lemma 10.7 to say that almost surely this subgraph will contain a $k_1(m)$-clique. If the probability is sufficiently large, then we can conclude the same holds true for *all* subsets of that size.

We can then produce a colouring greedily by picking independent sets of size $k_1(m)$ in $G(n,1/2)$, and letting each be a colour classes, which we can do until there are less than m vertices left.

If m can be chosen to be large enough that $k_1(m) \sim k_1(n)$, then we will use about the right amount of colours in this process. Similarly, if m can be chosen to be small enough that

$m << n/k_1(n)$, then if we colour the remaining vertices each by a different colour, it still won't affect the total number of colours we've used. Let's get down to the details.

We set $m = n/(\log(n))^2$ (natural, not base 2). We let $k_1 = k_1(m) = k_0(m) - 4$ as above and note that $k_1(m) \sim 2\log_2(m) \sim 2\log_2(n)$. For any set S of size m we have the restriction of $G(n, 1/2)$ onto S is distributed as $G(m, 1/2)$.

Therefore by Lemma 10.7 we have that

$$\mathbb{P}(\alpha(G(n,1/2)|_S) < k_1) < e^{-m^{2+o(1)}}.$$

There are $\binom{n}{m} < 2^n = 2^{m^{1+o(1)}}$ such sets, and so by a union bound

$$\mathbb{P}(\alpha(G(n,1/2)|_S) < k_1 \text{ for some } m\text{-set } S) < 2^{m^{1+o(1)}} e^{-m^{2+o(1)}} = o(1).$$

So almost surely every set of m vertices contains an independent set of size k_1. We then greedily colour $G(n, 1/2)$ by the process described above. We inductively choose independent subsets of size k_1 from the set of remaining vertices, by looking at a set of m vertices inside which almost surely there is an independent set of size k_1,, and give them a different colour until we have less than m vertices left. We then colour each of the remaining vertices with a different colour each. Since this produces a proper colouring we have that almost surely

$$\begin{aligned}\chi(G(n,1/2)) &\leq \frac{n}{k_1} + m \\ &\leq (1+o(1))\frac{n}{2\log_2(n)} + \frac{n}{\log(n)^2} \\ &\leq (1+o(1))\frac{n}{2\log_2(n)}.\end{aligned}$$

□

Therefore, combining this with the claimed lower bound we have that almost surely

$$\chi(G(n,1/2)) = (1+o(1))\frac{n}{2\log_2(n)}.$$

Recall that we in fact know, by Theorem 10.5, that almost surely $\chi(G(n, 1/2))$ is within $\sqrt{n}$ of it's mean, which lies somewhere in this range.

10.3 The Chromatic Number of Sparse Random Graphs

In this section we prove a rather remarkable concentration result for the chromatic number of $G(n, p)$ where $p = n^{-\alpha}$ for some fixed $\alpha > 5/6$. In the previous section we showed the chromatic number of a graph is almost always concentrated in a range of about $\sqrt{n}$ values. When p is sufficiently small we can in fact say much more, that almost always $\chi(G)$ takes one of at most four values.

Theorem 10.9. *Let $p = n^{-\alpha}$ for some fixed $\alpha > 5/6$ then there exists some $u = u(n,p)$ such that almost surely*

$$u \leq \chi(G(n,p)) \leq u + 3.$$

The proof of the theorem uses a result that was previously known, that asserts that, in the range of p we're looking at, if we look at a small subset of the vertices in $G(n,p)$ then we can always 3-colour the graph restricted to that subset.

Lemma 10.10. *Let $p = n^{-\alpha}$ for some fixed $\alpha > 5/6$ and let $c > 0$. Then almost surely for every subset $S \subset [n]$ with $|S| = c\sqrt{n}$*

$$\chi(G(n,p)|_S) \leq 3.$$

Proof. Suppose there exists some set S such that $|S| = s \leq c\sqrt{n}$ which cannot be 3-coloured. We can take without loss of generality such an S which is vertex minimal, and therefore we must have that the minimum degree of $G(n,p)|_S) \geq 3$. Therefore the S must contain at least $3s/2$ edges. We will show that the probability of the existence of such a subset is $o(1)$, from which the result will follow.

For each $|S| = s \leq c\sqrt{n}$ the probability that the subgraph contains at least $3s/2$ edges is at most

$$\binom{\binom{s}{2}}{\frac{3s}{2}} p^{\frac{3s}{2}}$$

by the union bound. Therefore, again by the union bound, the probability that some 'bad' subset of size $\leq c\sqrt{n}$ exists is at most

$$\sum_{s=4}^{c\sqrt{n}} \binom{n}{s} \binom{\binom{s}{2}}{\frac{3s}{2}} p^{\frac{3s}{2}} = \sum_{s=4}^{c\sqrt{n}} f(s).$$

Using the bound that $\binom{n}{k} \leq (en/k)^k$ we see that,

$$\begin{aligned}
f(s) &\leq \left(\frac{en}{s}\right)^s \left(\frac{e\binom{s}{2}}{\frac{3s}{2}}\right)^{\frac{3s}{2}} p^{\frac{3s}{2}} \\
&\leq \left(\frac{en}{s}\right)^s \left(\frac{es}{3}\right)^{\frac{3s}{2}} n^{-\frac{3\alpha s}{2}} \\
&\leq \left(\frac{en}{s}\left(\frac{es}{3}\right)^{\frac{3}{2}} n^{-\frac{3\alpha}{2}}\right)^s \\
&\leq \left(K_1 s^{\frac{1}{2}} n^{1-\frac{3\alpha}{2}}\right)^s \\
&\leq \left(K_2 n^{\frac{5}{4}-\frac{3\alpha}{2}}\right)^s,
\end{aligned}$$

for some constants K_1 and K_2. However, since $\alpha > 5/6$ we have that $\frac{5}{4} - \frac{3\alpha}{2} < 0$ and so the last term is $((K_2 n^{-\epsilon})^s$ for some $\epsilon > 0$. Therefore

$$\sum_{s=4}^{c\sqrt{n}} f(s) \leq \sum_{s=4}^{c\sqrt{n}} (K_2 n^{-\epsilon})^s = o(1).$$

□

This Lemma tells us that, in any random graph in our range of p, if we can colour almost all it's vertices with u colours, then we can colour the rest of the graph using at most 3 more colours. Using another clever choice of random variable, which will give us a vertex Lipschitz martingale, we will see that, for a certain choice of u, we can almost surely colour all but $c\sqrt{n}$ of the vertices in $G(n,p)$ with u colours.

Proof of Theorem 10.9. Let $\epsilon > 0$ be arbitrarily small and let $u = u(n,p,\epsilon)$ be the smallest integer such that

$$\mathbb{P}(\chi(G(n,p)) \leq u) > \epsilon.$$

We define a random variable Y to be the size of the smallest set of vertices S such that $G(n,p)\setminus S$ can be u-coloured. We first note that changing a single vertex of $G(n,p)$ can only change the value of Y by at most one, and so Y is vertex Lipschitz. Therefore we can apply Theorem 10.2 to see that, for any $t \geq 0$

$$\mathbb{P}(Y \geq \mathbb{E}(Y) + t\sqrt{n-1}) < e^{-\frac{t^2}{2}} \text{ and } \mathbb{P}(Y \leq \mathbb{E}(Y) - t\sqrt{n-1}) < e^{-\frac{t^2}{2}}$$

We pick t such that $e^{-t^2/2} = \epsilon$, and so both the tail events have probability less than ϵ. However we note that $Y = 0$ if and only if $G(n,p)$ itself can be u-coloured, however by definition of u this happens with probability $> \epsilon$. So

$$\mathbb{P}(Y \leq \mathbb{E}(Y) - \mathbb{E}(Y)) \geq \mathbb{P}(Y = 0) > \epsilon.$$

Since the tail events are nested we much have that $\mathbb{E}(Y) \leq t\sqrt{n-1}$. But then, again by the nestedness of the tail events,

$$\mathbb{P}(Y \geq 2t\sqrt{n-1}) \leq \mathbb{P}(Y \geq \mathbb{E}(Y) + t\sqrt{n-1}) < \epsilon$$

and so with probability at least $1-\epsilon$ we can remove a set of vertices of size $2t\sqrt{n-1}$ from $G(n,p)$ such that the remaining graph is u-colourable.

By Lemma 10.10, almost surely this set of vertices can be coloured using at most 3 new colours, giving a colouring of $G(n,p)$ using at most $u+3$ colours. Therefore with probability at least $1-\epsilon$, $G(n,p)$ can be $(u+3)$-coloured.

However by definition of u, with probability at least $1-\epsilon$ at least u colours are needed to colour G, therefore

$$\mathbb{P}(u \leq \chi(G(n,p)) \leq u+3) \geq 1-2\epsilon.$$

We might be a little concerned that, u depended on ϵ and so we can't simply let $\epsilon \to 0$ to deduce the result. However it is clear that, once ϵ gets small enough it's impossible that the ranges for different ϵ don't overlap, since it can't happen almost surely that $G(n,p)$ lies in two disjoint sets. So after a certain point there are at most 3 possible values for u, one of which is taken infinitely often, from which the result then follows. □

Using similar techniques it as been shown that in fact, for any $\alpha > 1/2$, $\chi(G(n,n^{-\alpha}))$ is concentrated on at most two values.

11 Talagrand's Inequality

Another useful result that asserts strong concentration of a random variable is Talagrand's inequality. Our previous results on strong concentration gave us an exponentially small bound on the probability of deviations from the mean of suitably well behaved random variables. Our notion of suitably well behaved was basically that it was close, or bounded by, some random variable on a product space where in each co-ordinate the random variable was bounded. However the Azuma-Hoeffding inequality required that the deviations we considered be at least as large as the square root of the dimension of this product space. Talagrands inequality gives us a much stronger result for a similar class of random variables.

We will state just a special case of the equality, without proof, in a form which is useful for applications. We first need to define formally the notion of a product space, which we will do only for finite probability spaces to avoid using too much measure theory. The case for general probability spaces is similar.

Definition. Let $\{(\Omega_i, 2^{\Omega_i}, \mathbb{P}_i) : i \in [n]\}$ be finite probability spaces. We let

$$\Omega = \prod_{i=1}^{n} \Omega_i = \{(\omega_1, \omega_2, \ldots, \omega_i) : \omega_i \in \Omega_i \text{ for all } i \in [n]\}$$

be the product of the sets Ω_i and define a probability measure $\mathbb{P}$ on 2^Ω by defining the probability of elementary events to be

$$\mathbb{P}((\omega_1, \omega_2, \ldots, \omega_i)) = \prod_{i=1}^{n} \mathbb{P}_i(\omega_i)$$

and extending it to 2^Ω in the obvious way. Then the *product space* (of $\{(\Omega_i, \Sigma_i, \mathbb{P}_i) : i \in [n]\}$) is the probability space $(\Omega, 2^\Omega, \mathbb{P})$.

For example $\mathcal{G}(n,p)$ is the product of $\binom{n}{2}$ identical probability spaces, each of which corresponds to a possible edge of $G(n,p)$. We note that, given a random variable on a product space, there is a a natural martingale associated with this random variable given by 'exposing' each co-ordinate in turn, the edge and vertex exposure martingales being examples of this.

Talagrand's inequality is a very broad inequality about the concentration of measure in product spaces. We consider a generalisation of the *Hamming distance*. This counts the number

of co-ordinates in which two $\omega, \omega' \in \Omega$ differ. In other words, the distance is $\sum_{w_i \neq w_i'} 1$. We can take a weighted version of this and consider, for any unit vector $\alpha \in \mathbb{R}^n$ the *α-Hamming distance* between ω and ω' to be

$$d_\alpha(\omega, \omega') = \sum_{\omega_i \neq \omega_i'} \alpha_i.$$

Given a set $A \subset \Omega$ and a point ω for any α we can consider the α-Hamming distance between ω and A.

$$d_\alpha(\omega, A) = \inf\{d(\omega, \omega') \,:\, \omega' \in A\}.$$

We will think of ω as being far from A if it's far in *some* α-Hamming distance, with α a unit vector. That is, we define

$$d(\omega, A) = \sup_{|\alpha|=1} d_\alpha(\omega, A).$$

Talagrand's inequality is then the following

Theorem 11.1 (Talagrand's Inequality)**.** *Let $\{(\Omega_i, \Sigma_i, \mathbb{P}_i) \,:\, i \in [n]\}$ be probability spaces and let $(\Omega, \Sigma, \mathbb{P})$ be their product. If $A, B \in \Sigma$ are such that $d(\omega, A) \geq \tau$ for all $\omega \in B$, then*

$$\mathbb{P}(A)\mathbb{P}(B) \leq e^{-\frac{\tau^2}{4}}.$$

Note that we can vary α for different points $\omega \in B$, so it doesn't even have to be 'uniformly' far from A. An equivalent formulation is to say that we are bounding $\mathbb{P}(A)\mathbb{P}(A_t)$ where $A_\tau = \{\omega \,:\, d(\omega, A) \geq \tau\}$.

This doesn't look a lot like the tail estimates we have from earlier in the course, however with not too much work one can deduce the following as a corollary, which we may sometimes refer to also as Talagrand's inequality.

First let us make a few definitions

Definition. A random variable $X : \Omega \to \mathbb{R}$ is *c-Lipschitz* if changing just one co-ordinate can change the value of X by at most c. Given some function $f : \mathbb{N} \to \mathbb{N}$ we say that X *is f-certifiable* if whenever $X(\omega_1, \omega_2, \ldots, \omega_n) \geq s$ there is a subset $I \subset [n]$ of size $|I| = f(s)$ such that X is greater than s on the entire subspace

$$\{(\omega_1', \omega_2', \ldots, \omega_n') \,:\, \omega_i' = \omega_i \text{ for all } i \in I\}.$$

To put it in words, X is f-certifiable if, whenever X takes a value bigger than s, you can verify this by looking at just $f(s)$ of it's co-ordinates. Talagrand's inequality tells us that, when a c-Lipschitz random variable is f-certifiable for a suitably small f, then it is highly concentrated, sometimes more concentrated than the Azuma-Hoeffding inequality implies.

Corollary 11.2. *Let X be a c-Lipschitz random variable which is f-certifiable and let m be the median of X (that is m is the unique real number such that $\mathbb{P}(X > m) \leq 1/2$ and $\mathbb{P}(X < m) \leq 1/2$). Then for any $t \geq 0$*

$$\mathbb{P}(X \leq m - t) \leq 2e^{-\frac{t^2}{4c^2 f(m)}} \text{ and } \mathbb{P}(X \geq m + t) \leq 2e^{-\frac{t^2}{4c^2 f(m+t)}}.$$

Proof. Let us consider the two sets

$$A = \{\omega \,:\, X(\omega) \leq m - t\} \text{ and } B = \{\omega \,:\, X(\omega) \geq m\}$$

Since $\mathbb{P}(B) \geq 1/2$ by definition, if we can show that $d(\omega, A) \geq \tau$ for all $\omega \in B$ for some τ, then we can use Theorem 11.1 to get a bound on the probability of A.

However, since X is f-certifiable, for every $\omega \in B$ there is some set $I \subset [n]$ of size $f(m)$ such that $X(\omega') \geq m$ for every ω' which agrees with ω on I. Then, since every $\omega' \in A$ has $X(\omega) \leq m - t$ and changing the value of one co-ordinate can change the value of X by at most c, it follows that every $\omega' \in A$ must disagree with ω in at least t/c of the co-ordinates in I.

Hence, if we take α to be the unit vector with $\alpha_i = 1/\sqrt{|I|}$ for all $i \in I$ and 0 otherwise, it follows that every ω has α-Hamming distance at least $t/c\sqrt{f(m)}$ to every $\omega' \in A$. Hence, $d(\omega, A) \geq t/\sqrt{f(m)}$.

Hence, applying Theorem 11.1 tells us that

$$\mathbb{P}(X \leq m - t)\mathbb{P}(X \geq m) \leq e^{-\frac{t^2}{4c^2 f(m)}}$$

and so, since by definition $\mathbb{P}(X \geq m) \geq 1/2$ we have that

$$\mathbb{P}(X \leq m - t) \leq 2e^{-\frac{t^2}{4c^2 f(m)}}.$$

The other inequality follows in a similar fashion. □

Note that the two tail estimates are not necessarily symmetric. Looking at Corollary 11.2, we see that we can get exponentially good bounds for the tail estimates when $t >> \sqrt{f(m)}$. Most importantly this doesn't necessarily depend on the dimension of the product space we live in, and so, when $f(m)$ is small compared to n, we will get much better bounds than the Azuma-Hoeffding inequality would give us.

Also, this theorem talks about a variable being concentrated about its median rather than its mean, however, as the following lemma shows, in a lot of cases one can show the median must be close to the expectation.

Lemma 11.3. *Let $\{(\Omega_i, 2^{\Omega_i}, \mathbb{P}_i) : i \in [n]\}$ be probability spaces and let $(\Omega, 2^{\Omega}, \mathbb{P})$ be their product. Let X be a c-Lipschitz, $f(s) = rs$-certifiable random variable and let m be the median of X. Then*

$$|\mathbb{E}(X) - m| \leq 20c\sqrt{rm}.$$

11.1 Longest Increasing Subsequence

Suppose we pick a sequence $x_1, x_2, \ldots, x_n \in [0, 1]$ independently and uniformly at random. If we put the sequence in increasing order, $x_{i_1} < x_{i_2} < \ldots < x_{i_n}$, this defines a permutation of $[n]$, and it is not hard to check that the distribution we get by picking a permutation in this way is also uniform.

Let us consider the random variable X which counts the longest increasing subsequence from this sequence. What can we say about this random variable?

Well, given some ordered subset of the x_i, $\{x_{i_1}, x_{i_2}, \dots x_{i_k}\}$, the probability that this forms an ordered subsequence is simply $1/k!$ by symmetry. So the expected number of increasing sequences of length k is just $\binom{n}{k}/k!$ and so by Markov's inequality , if we let X_k be the number of increasing sequences of length k, we have that

$$\mathbb{P}(X \geq k) = \mathbb{P}(X_k \geq 1) \leq \mathbb{E}(X_k) = \frac{\binom{n}{k}}{k!} \leq \left(\frac{en}{k}\right)^k \left(\frac{e}{k}\right)^k = \left(\frac{e\sqrt{n}}{k}\right)^{2k}.$$

Therefore if the median is the number m such that $\mathbb{P}(X \geq m) = 1/2$, we have that $m \leq 3\sqrt{n}$, since we know that

$$\mathbb{P}(X \geq 3\sqrt{n}) \leq \left(\frac{e}{3}\right)^{6\sqrt{n}} < 1/2.$$

However, a classical theorem of Erdős and Szekeres tells us that if we let Y be the length of the largest decreasing subsequence then we always have that $XY \geq n$ (we have to be a little careful since this talks about non-decreasing and non-increasing subsequences, but with high probability no two x_i take the same value). However, by symmetry we have that X and Y have the same distribution and so, since $\mathbb{P}(X \leq 3\sqrt{n}) \geq 1/2$ we must have that $\mathbb{P}(Y \geq \frac{1}{3}\sqrt{n}) \geq 1/2$ and so the median of X must satisfy

$$\frac{1}{3}\sqrt{n} \leq m \leq 3\sqrt{n}.$$

Now, the random variable X is clearly 1-Lipschitz, changing the value of any one x_i can only change the length of the largest increasing subsequence by 1, and also we have that X is f-certifiable with $f(s) = s$. Indeed, to verify that $X \geq s$ we can simply look at the values of the x_i in an increasing subsequence of length s. Therefore by Theorem 11.1 we have

$$\mathbb{P}(X \leq m - t) \leq 2e^{-\frac{t^2}{4m}} \text{ and} \mathbb{P}(X \geq m + t) \leq 2e^{-\frac{t^2}{4(m+t)}}.$$

Since $m \sim \sqrt{n}$ we can take t to be slighty larger than $n^{1/4}$, say $t = n^{1/4}\log(n)$, to see that with high probability X must lie in the interval $[m - t, m + t]$.

Let us compare this to the bound we would get from applying Theorem 10.2. Here we would be considering the exposure martingale associated with X by exposing the values of each x_i in turn. Since the function is 1-Lipschitz we have by a similar argument to Lemma 10.4 that the associated martingale satisfies $|X_i - X_{i-1}| \leq 1$ and so by the Azuma-Hoeffding inequality we have that

$$\mathbb{P}(X \geq \mathbb{E}(X) + t) \leq e^{-\frac{t^2}{2n}} \text{ and } \mathbb{P}(X \leq \mathbb{E}(X) - t) \leq e^{-\frac{t^2}{n}}.$$

So in order to get tight concentration we would need to take $t >> n^{1/2}$, which is not only much worse than what Talagrand can give, it's not especially useful since it follows from Lemma 11.3 that $|\mathbb{E}(X) - m| \leq O(n^{\frac{1}{4}})$ and not $\mathbb{E}(X) \sim m \sim \sqrt{n}$.

11.2 Chromatic Number of Graph Powers

Definition. Given a graph G and $x, y \in V(G)$ let us define the *distance* between x and y, $\text{dist}_G(x, y)$, to be the length of the shortest path between them. Given a graph G the *kth power*

of G, G^k is defined to be the graph with

$$V(G^k) = V(G) \text{ and } E(G^k) = \{(x, y) : \text{dist}_G(x, y) \leq k\}.$$

Suppose we have a graph G with maximum degree $\Delta(G) = d$. What can we say about the chromatic number of G^k?

A simple application of the greedy algorithm tells us that we can colour any graph H with $\Delta(H) + 1$ colours, and since $\Delta(G^k) \leq d^k$ we have that $\chi(G^k) \leq d^k + 1$. Brook's theorem tells us that for any graph which is not a cycle or complete, $\chi(H)$ beats this naive bound and $\chi(H) \leq \Delta(H)$. Reasonably recently this result have been improved, first by Kim to show that

Theorem 11.4. *Let* H *be such that* $g(H) \geq 5$*, then* $\chi(G) \leq (1 + o(1))\Delta(H)/\log(\Delta(H))$

and later Johansson showed

Theorem 11.5. *Let* H *be such that* $g(H) \geq 4$ *(that is, triangle-free), then* $\chi(g) \leq O\big(\Delta(H)/\log(\Delta(H))\big)$

Applying this to graph powers we see that, as long as G still has reasonably large girth compared to k, then we get an improvement of a log factor over the naive bound for G^k as well. The following result of Alon and Mohar tells us that if we fix $g \geq 3$ and k, and allow the maximum degree to be arbitrarily large, then there exist graphs achieving this bound.

Theorem 11.6. *Let* $g \geq 3$ *and* k *be fixed. Then for large enough* d *there exist graphs* G *with* $g(G) \geq g$ *and* $\Delta(G) \leq d$ *such that*

$$\chi(G^k) \geq \Omega\left(\frac{d^k}{\log(d)}\right).$$

Proof. We want to construct such a graph by picking a random graph $G(n, p)$ for suitable n and p. Let $p = \frac{d}{2n}$ so that the expected degree of each vertex is $\sim d/2$. We first want to make sure that our graph satisfies the conditions claimed on $\Delta(G)$ and $g(G)$, to do that we will use the alteration method.

It is a simple application of the chernoff bound that, for each vertex $v \in V$

$$\mathbb{P}(d(v) \geq d) < e^{-\frac{d}{100}}.$$

(where no attempt has been made to optimise the constant, none at all). So, if we let N_{bad} be the number of vertices with degree larger than d we see that

$$\mathbb{E}(N_{\text{bad}}) < ne^{-\frac{d}{100}}.$$

Hence, by Markov's inequality

$$\mathbb{P}(N_{\text{bad}} > 100ne^{-\frac{d}{100}}) \leq \frac{1}{100}.$$

(that is, is very small). Similarly if we look at the random variable $C_{<g}$ which counts the number of cycles of size $< g$, we have, by a similar calculation as that in Section 4, that

$$\mathbb{E}(C_{<g}) = \sum_{i=1}^{g-1}(np)^i = \sum_{i=1}^{g-1}(\frac{d}{2})^i < d^g.$$

So again an application of Markov's inequality tells us that

$$\mathbb{P}(C_{<g} > 100d^g) < \frac{1}{100}.$$

Combining these two estimate we see that with probability at least 98/100, $G(n,p)$ is such that

$$N_{\text{bad}} < 100ne^{-\frac{d}{6}} \text{ and } C_{<g} < 100d^g$$

and so, since we are free to take $d >> g$ and $n >> d$, we can remove a vertex from each small cycles and delete all vertices of degree more than d to get a graph G' with neither, which still has $(1+o(1))n$ vertices.

We want to get a bound on the chromatic number of G'^k, and similar to Section 4 we will do so by bounding above the size of the largest independent set in G'^k. So, since $\chi(G'^k) \geq |G'|/\alpha(G'^k) \geq n/2\alpha(G'^k)$ we will need to show that, with positive probability, even after the alterations we made

$$\alpha(G'^k) \leq c_k n\frac{\log(d)}{d^k}$$

for some constant c_k. For each subset U of that size we want to show that, even after we make our alterations, there is still some edge in G'^k inside U. To guarantee this we will show that with a high probability we can find many vertex disjoint paths of length k between pairs of vertices in U. Since each vertex we removed from the graph $G(n,p)$ can be in at most one of these paths, if there are sufficiently many we can conclude that U is still not independent in G'. So we will prove the following auxiliary lemma.

Lemma 11.7. *Let $G(n,p)$ be chosen with p as above, the for an appropriate choice of constant c_k the following holds. For every subset $U \subset V(G)$ of size*

$$|U| = c_k n\frac{\log(d)}{d^k} = x$$

let P be the random variable which counts the maximum size of a family of paths of length k which lie in $G(n,p)$ such that both endpoints lie in U, all the internal vertices of the paths lie outside of U, and no two paths share a vertex except in U. Then almost surely

$$P \geq \frac{c_k^2 n\log(d)^2}{2^{k+6}d^k}.$$

Proof. If we let P' be the random variable which just counts the number of paths of length k satisfying the first two conditions we see that

$$\begin{aligned}\mathbb{E}(P') &= \binom{x}{2}(n-x)(n-x-1)\dots(n-x-k+1)p^k \\ &> c_k^2 n^2\frac{\log(d)^2}{d^{2k}}\frac{n^{k-1}}{4}\frac{d^k}{2^k n^k} \\ &= \frac{c_k^2 n\log(d)^2}{2^{k+2}d^k}\end{aligned}$$

As in Theorem 10.5 we want to say that the expected value of P is close to that of P' by showing that the expected number of pairs of paths which are not vertex disjoint in this way is small. If we let Q be the random variable which counts the pairs of paths which share an internal vertex outside of U it is again a rather tedious calculation to show that the largest contribution to Q comes from pairs of paths which share an endpoint and the neighbour of that endpoint. The expected number of such pairs is at most

$$\mathbb{E}(P')n^{k-2}xp^{k-1} = \mathbb{E}(P')\frac{c_k \log(d)}{2^{k-1}d} << \mathbb{E}(P').$$

Therefore if we pick a random set of such paths, including each one with some fixed probability q, the expected size of our collection is just $q\mathbb{E}(P')$ and the expected number of bad pairs is $q^2\mathbb{E}(Q) < q^2\mathbb{E}(P')$. Therefore by the alteration method we can find a collection such that when we remove 1 path from each bad pair we still have at least $(q-q^2)\mathbb{E}(P')$ left, and so we have that, with $q = 1/2$

$$\mathbb{E}(P) \geq \frac{1}{4}\mathbb{E}(P') > \frac{c_k^2 n \log(d)^2}{2^{k+4}d^k} := b.$$

If we define m to be the median of P we claim that $m \geq b/2$. Indeed, if not, then both m and $20\sqrt{km}$ are less than $b/2$. However, since P counts the size of the largest set of paths of length k satisfying certain conditions, including being edge disjoint, we have that P is 1-Lipschitz, that is, changing any one edge can change P by at most 1. We also note that P is f-certifiable for $f(s) = ks$. Hence, by Lemma 11.3

$$\mathbb{E}(P) \leq m + 20c\sqrt{km} < b$$

a contradiction.

Hence, since $m \geq b/2$ it follows from Talagrands inequality that

$$\mathbb{P}(P \leq b/4) \leq \mathbb{P}(P \leq m/2) \leq 2e^{-\frac{m}{16k}} \leq 2e^{-\frac{b}{32k}}.$$

Therefore,

$$\mathbb{P}(P \leq \frac{b}{4}) \leq 2e^{-\frac{b}{32k}} = 2\exp\left(-\frac{c_k^2 n \log(d)^2}{2^{k+9}kd^k}\right).$$

Therefore, for any fixed U, we've showed that the probability that U has less than $b/4$ appropriate paths (which we note is the claimed number in the statement of the lemma), can be bounded above by this quantity. If we can show that this quantity is small compared to the total number of U of size x, the we could conclude the result of the lemma by the union bound. Now the total number of such sets is

$$\begin{aligned}\binom{n}{x} \leq \left(\frac{en}{x}\right)^x &= \left(\frac{ed^k}{c_k \log(d)}\right)^{\frac{c_k n \log(d)}{d^k}} \\ &\leq \exp\left(\frac{c_k k n \log(d)^2}{d^k}\right).\end{aligned}$$

Therefore if we choose c_k such that

$$\frac{c_k^2}{2^{k+9}k} > 2kc_k$$

then with high probability for every such set U there will be at least the claimed number of paths. □

So, as we showed before, with probability at least 98/100, $G(n,p)$ is such that

$$N_{\text{bad}} < 100ne^{-\frac{d}{6}} \text{ and } C_{<g} < 10d^g$$

and we also know that with high probability $G(n,p)$ satisfies the conclusion of Lemma 11.7, and hence with positive probability $G(n,p)$ satisfies all three. Therefore there exists some graph G' satisfying all three conditions. From such a graph, let us delete a vertex from each cycle of length less than g and delete each vertex of degree $\geq d$. We therefore obtain a graph G such that $g(G) \geq g$ and $\Delta(G) \leq d$ and also

$$|G| \geq n - 100(ne^{-\frac{d}{6}} + d^g) \geq \frac{n}{2}.$$

We also claim that $\alpha(G^k) \leq x$. Indeed, since G' satisfied the conclusion of Lemma 11.7 we know that each subset of $V(G)$ of size x, when considered as a subset of $V(G')$, contained at least

$$\frac{c_k n \log(d)^2}{2^{k+6}d^k}$$

paths of length k, in G', between vertices of U such that no internal vertices were inside U, or in more than 1 path. Since we removed at most

$$100(ne^{-\frac{d}{6}} + d^g) < \frac{c_k n \log(d)^2}{2^{k+6}d^k}$$

vertices of G' to form G, at least one of these paths is contained in G, and hence in G^k, U is not independent.

Therefore G is a graph with $g(G) \geq g$ and $\Delta(G) \leq d$ and

$$\chi(G^k) \geq \frac{|G^k|}{\alpha(G^k)} \geq \frac{n}{2x} = \Omega\left(\frac{d^k}{\log(d)}\right).$$

□

11.3 Exceptional outcomes

Our previous concentration results have all relied on our random variables being c-Lipschitz for some small enough c, when considered as functions on an underlying product probability space. Whilst they can't tell us anything if the function is not well behaved in this way, sometimes it will be the case that the random variable is still be tightly concentrated.

Let us consider, as a motivating example, the random variable T which counts the number of triangles in $G(n,p)$. By a simple application of the first and second moment methods we know that $r(n) = 1/n$ is a threshold function for the event that $G(n,p)$ contains a triangle. We will consider the range where $p = n^{-1+\beta}$ for some small $\beta > 0$. Here, the expected number of

triangles $n^{3\beta}$ will get arbitrarily large, we would like to be able to say that with high probability the number of triangles is 'close' to that expectation.

As before we can think of $G(n,p)$ as a product of $\binom{n}{2}$ probability spaces on $\{0,1\}$, and T as a function on this space. However T is not very Lipschitz, indeed, adding or deleting a single edge can change the number of triangles by as much as $n-2$. Thinking in terms of Theorem 11.1, we have that T is f-certifiable where $f(s)=3s$ and we expect the median to be close to the average which is $\mathbb{E}(T)=\binom{n}{3}p^3 \sim \frac{1}{6}n^{3\beta}$ (although this doesn't follow from Lemma 11.3).

So to get a good bound on deviations of size smaller than $\mathbb{E}(T)$, we would need to choose $t \leq \mathbb{E}$ such that

$$2e^{-\frac{t^2}{12n^2\mathbb{E}(T)}}$$

is small. Since $t \leq \mathbb{E}(T) \sim n^{3\beta}$, we would need β to be strictly larger than $2/3$.

However it turns out that T will be tightly concentrated about it's mean even when β is much smaller. The intuitive reason that this should be true is that, whilst an edge can be in many triangle, it's very unlikely that any particular edge will be. For example the expected number of triangles containing any particular edge is only $n^{-1+2\beta}$, which is very small.

It might seem possible that a condition of this sort could be sufficient to prove concentration, if the *expected* effect of each co-ordinate was small, however the following example shows us that this is not the case.

Example. Let $m=4k$, we will consider a probability space on $\{0,1\}^m$ where each event $\omega=(\omega_1,\omega_2,\ldots,\omega_m)$ is such that the probability that $\omega_i=1$ is, independently, $p=m^{-\frac{1}{2}}$ for each i. We will consider the following function

$$f(\omega_1,\omega_2,\ldots,\omega_m)=(\omega_1\omega_2+\omega_2\omega_3+\ldots+\omega_{2k-1}\omega_{2k})(\omega_{2k+1}+\omega_{2k+2}+\ldots+\omega_{4k}).$$

Since the functions in the brackets are independent, by the linearity of expectation we see that

$$\mathbb{E}(f)=2k^2p^3=\frac{1}{8}m^{\frac{1}{2}}.$$

What is the expected effect of each co-ordinate? Well, for ω_1 we have that

$$|\mathbb{E}(f|\omega_1=0)-\mathbb{E}(f|\omega_1=1)|=|\mathbb{E}(\omega_2((\omega_{2k+1}+\omega_{2k+2}+\ldots+\omega_{4k}))|=2kp^2=\frac{1}{2},$$

and similarly for $\omega_2,\ldots,\omega_{2k}$. Doing a similar calculation for $\omega_{2k+1},\ldots,\omega_{4k}$ gives an expected effect of $\frac{1}{4}$.

However we note that the sum inside the second set of brackets is just a binomial random variables and so by the Chernoff bound we know that with high probability it will close to it's expected value, of $m^{\frac{1}{2}}/2$. The second bracket is also some binomial random variable, but on top of that it's always an integer and so with high probability $f(t)$ is either 0, or larger than $m^{\frac{1}{2}}/2=4\mathbb{E}(f)$. Hence f is not tightly concentrated about it's mean.

Instead, we will prove a generalisation of Talagrand's inequality that excludes a small set Ω^* of *exceptional outcomes*. Since these exceptional outcomes will (hopefully) allow us to reduce

our dependence on how 'Lipschitz' T is, we will want to modify our definition of certifiable somewhat.

Definition. Given an exceptional set $\Omega^* \subset \Omega$ and $s, c > 0$ we say that a random variable X has (s,c)-*certificates* if for every $t > 0$ and every $\omega \in \Omega \setminus \Omega^*$ there is an index set I of size at most s so that $X(\omega') > X(\omega) - t$ for any $\omega' \in \Omega \setminus \Omega^*$ for which the ω and ω' differ in less than t/c co-ordinates.

Note that, if X is f-certifiable and c-Lipschitz and s is the maximum value of f over the range of X, then it is easy to check that X has (s,c)-certificates. We will want to show the following version of Talagrand's inequality.

Theorem 11.8. *Let $\{(\Omega_i, 2^{\Omega_i}, \mathbb{P}_i) : i \in [n]\}$ be probability spaces and let $(\Omega, 2^\Omega, \mathbb{P})$ be their product. Let $\Omega^* \subseteq \Omega$ be a set of exceptional events. Suppose X is a random variable which has (s,c)-certificates, let m be the median of X and let $t \geq 0$. Then*

$$\mathbb{P}(|X - m| \geq t) \leq 4e^{-\frac{t^2}{4c^2 s}} + 4\mathbb{P}(\Omega^*).$$

Proof. The same as the proof of Corollary 11.1. We consider the sets

$$A = \{\omega \in \Omega \setminus \Omega^* : X(\omega) \leq m - t\}, \text{ and}$$

$$B = \{\omega \in \Omega \setminus \Omega^* : X(\omega) \geq m\}.$$

and use that fact that

$$\mathbb{P}(X \leq m - t)\mathbb{P}(X \geq m) \leq \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(\Omega^*).$$

□

By the same argument as Lemma 11.3, we can also deduce in this sitautation (as long as X is bounded) that m is quite close to $\mathbb{E}(X)$.

Lemma 11.9. *Let $\{(\Omega_i, 2^{\Omega_i}, \mathbb{P}_i) : i \in [n]\}$ be probability spaces and let $(\Omega, 2^\Omega, \mathbb{P})$ be their product. Let $\Omega^* \subseteq \Omega$ be a set of exceptional events. Let X be a random variable which has (s,c)-certificates, let m be the median of X and let $M = \max\{ \sup |X|, 1\}$. Then*

$$|\mathbb{E}(X) - m| \leq 20c\sqrt{s} + 20M^2\mathbb{P}(\Omega^*).$$

Combining the two we get the following corollary.

Corollary 11.10. *Let $\{(\Omega_i, 2^{\Omega_i}, \mathbb{P}_i) : i \in [n]\}$ be probability spaces and let $(\Omega, 2^\Omega, \mathbb{P})$ be their product. Let $\Omega^* \subseteq \Omega$ be a set of exceptional events. Let X be a random variable which has (s,c)-certificates, let m be the median of X and let $M = \max\{ \sup |X|, 1\}$. If $\mathbb{P}(\Omega^*) \leq M^{-2}$ then for $t > 50c\sqrt{s}$*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq 4e^{-\frac{t^2}{16c^2 s}} + 4\mathbb{P}(\Omega^*).$$

Proof. By Theorem 11.8 and 11.9 it follows that

$$\begin{aligned}\mathbb{P}(|X - \mathbb{E}(X)| \geq t/2 + 20c\sqrt{s} + 20M^2\mathbb{P}(\Omega^*)) &\leq \mathbb{P}(|X - m| \geq t/2) \\ &\leq 4e^{-\frac{t^2}{16c^2 s}} + 4\mathbb{P}(\Omega^*).\end{aligned}$$

If $t > 50c\sqrt{s}$ and $\mathbb{P}(\Omega^*) \leq M^{-2}$ then

$$t \geq t/2 + (20c\sqrt{s} + 20M^2\mathbb{P}(\Omega^*)$$

and hence

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq 4e^{-\frac{t^2}{16c^2 s}} + 4\mathbb{P}(\Omega^*).$$

□

Let us return to the question of the number of triangles in a random graph $G(n,p)$, which we denote by T. Recall that, using Talagrands inequality we can deduce that T is tightly concentrated around its mean when $p = n^{-1+\beta}$ as long as $\beta > 2/3$. Recently multiple methods have been developed which allow one to deduce concentration in such cases (where the 'effect' of changing a single coordinate may be large, but is expected to be quite small). For example Warnke recently developed a method for dealing with such cases that can show concentration when $\beta > \frac{1}{5}$. There is also a highly technical method of Kim and Vu which can deal with an even larger range.

We will use how the previous results results to deduce concentration of T about its mean for β even smaller than $2/3$. For example, let's start with $\beta = 2/3 + \gamma$ for some small $\gamma > 0$ small, so that $p = n^{-\frac{1}{3}+\gamma}$.

We have firstly that $\mathbb{E}(T) = \binom{n}{3}p^3 = \Omega(n^{2+3\gamma})$. Also, T is $f(s) = 3s$ certifiable and $n-2$ edge-Lipschitz, so we can apply Talagrands inequality (technically using Lemma 11.3 to relate the mean to the median and arguing as in Corollary 11.10).

$$\mathbb{P}(T \geq \frac{1}{3}n^{2+3\gamma}) = \mathbb{P}(T \geq 2\mathbb{E}(T)) = \mathbb{P}(T - \mathbb{E}(T) \geq \mathbb{E}(T)) \leq e^{-\frac{\mathbb{E}(T)^2}{6(n-2)^2 3\mathbb{E}(T)}}$$

So it's incredibly unlikely that $T \geq 2\mathbb{E}(T) \geq \frac{1}{3}n^{2+3\gamma}$. We also note that, by the Chernoff bounds, for any $\delta > 0$ it's exponentially unlikely that any edge is in more that $\max\{n^\delta, 2np^2\}$ triangles, since the expected number is np^2 and it is a binomial random variable.

(We need the n^δ in case the expected number gets too small, so that our concentration is no longer exponentially good. For this first step $np^2 = n^{\frac{1}{3}+2\gamma}$, and so it's not important)

However, both T and the number of triangles at an edge are monotone, and so if we decrease p, they only get more unlikely. Hence if we let $\beta = \frac{4}{9} + \gamma'$ for some small enough $\gamma' > 0$, then it's still true that

- The probability that $T \geq \frac{1}{3}n^{2+3\gamma}$ is exponentially small;
- The probability that any edge is in more than $n^{\frac{1}{3}+2\gamma}$ triangles is exponentially small.

So, let's let Ω^* be our exceptional set of events, and consider $\beta = \frac{4}{9}+\gamma'$. Note that $p = n^{-\frac{5}{9}+\gamma'}$ and $\mathbb{E}(T) = \Omega(n^{\frac{4}{3}+3\gamma'})$.

Now, in $\Omega \setminus \Omega^*$ T will have (s,c) certificates for $s = \frac{1}{3}n^{2+3\gamma}$ and $c = n^{\frac{1}{3}+2\gamma}$. Indeed, we can take for any ω the set I to be the set of *all* edges lying in any triangle, and then $|I| \leq s$. Similarly, since every edge lies in at most $n^{\frac{1}{3}+2\gamma}$, if ω and ω' differ on I in less than t/c coordinate, then $T(\omega)$ and $T(\omega')$ (which don't depend at all on coordinates outside of I) can differ by at most t.

Hence, we can use Corollary 11.10 to say that

$$\begin{aligned}\mathbb{P}(|T - \mathbb{E}(T)| \geq t) &\leq 4e^{-\frac{t^2}{16c^2 s}} + 4\mathbb{P}(\Omega^*) \\ &= 4e^{-\frac{t^2}{O(n^{\frac{8}{3}+5\gamma})}} + 4\mathbb{P}(\Omega^*)\end{aligned}$$

and this probability will be exponentially small as long as $t^2 \geq n^{\frac{8}{3}+5\gamma}$. So, as long as $t \geq n^{\frac{4}{3}+\frac{5}{2}\gamma}$, we get an exponentially small probability of deviations of size t.

So, if $3\gamma' > \frac{5}{2}\gamma$ then $\mathbb{E}(T) = n^{\frac{4}{3}+3\gamma'}$ and

$$\begin{aligned}\mathbb{P}(|T - \mathbb{E}(T)| \geq \mathbb{E}(T)/2) &\leq 4e^{-\frac{\mathbb{E}(T)^2}{64c^2 s}} + 4\mathbb{P}(\Omega^*) \\ &= 4e^{-\frac{n^{\frac{8}{3}+6\gamma'}}{O(n^{\frac{8}{3}+5\gamma})}} + 4\mathbb{P}(\Omega^*) \\ &= o(1).\end{aligned}$$

So, we've managed to deduce concentration as low down as $\beta > \frac{4}{9}$, but we can then just bootstrap this result as above to get an improvement.

That is to say, if we take some $\beta < \frac{4}{9}$ then by monotonicity we know that the probability that $T \geq \frac{1}{3}n^{\frac{4}{3}+\gamma'}$ is exponentially small Ã°and, as before, the probability that any edge is in more that $\max\{np^2, n^\delta\}$ triangles is also exponentially small. Since $\beta < 4/9 < 1/2$, the maximum will be n^δ, where we'll choose $\delta > 0$ to be very small.

Hence, if we add all these events to our set of exceptional outcomes Ω^*, we now have that T now has (s,c)-certificates for $s = 2n^{\frac{4}{3}+\gamma'}$ and $c = n^\delta$ by the same arguments as before.

Then we can conclude that by Corollary 11.10 that

$$\begin{aligned}\mathbb{P}(|X - \mathbb{E}(X)| \geq t) &\leq 4e^{-\frac{t^2}{16c^2 s}} + 4\mathbb{P}(\Omega^*) \\ &= 4e^{-\frac{t^2}{O(n^{\frac{4}{3}+\gamma'+2\delta})}} + 4\mathbb{P}(\Omega^*).\end{aligned}$$

Again, this will allow us to conclude exponentially tight concentration as long as $t \geq n^{\frac{2}{3}+\gamma'/2+\delta}$. Hence, if we choose β such that $(n^{3\beta} \sim)\mathbb{E}(T) \geq n^{\frac{2}{3}+\gamma'/2+\delta}$ we can say that with high probability T is concentrated about its mean.

Unpacking everything, as long as $\beta > \frac{2}{9}+\gamma''$ with $3\gamma'' > \gamma'/2+\delta$ it follows that $\mathbb{E}(T) \geq n^{\frac{2}{3}+3\gamma''}$, and hence T will be tightly concentrated about its mean.

(And so, if we want to deduce this concentration for a fixed γ'', we have to go back and choose γ' and δ sufficiently small in the previous part, which means choosing γ sufficiently small in the first part...)

However, there's not reason we can't keep bootstrapping the result up like this. By choosing the sequence of constants δ, γ, γ', ... appropriately small, we can eventually show that T is tightly concentrated about its mean for β arbitrarily small.

With a bit more care one can use these methods to show the following result.

Theorem 11.11. *Let $p = n^{-1+\beta}$ for $\beta > 0$ and let $\delta > 0$. Then with high probability*

$$|T - \mathbb{E}(T)| \leq n^{\delta}\sqrt{\mathbb{E}(T)}.$$

12 Entropy Methods

12.1 Basic Results

Given a discrete random variable X let us denote by $p(x) := \mathbb{P}(X = x)$ for each x in the range of X. We define the *entropy* of the random variable X to be

$$H(X) = \sum_x p(x) \log \left(\frac{1}{p(x)} \right).$$

Note that this quantity is always positive.

We want to think of entropy, at least heuristically, as a measure of the expected amount of 'surprise' we have upon discovering the value of X. We then have the following heuristic argument for why $H(X)$ should be defined as above.

If we have an event A, such as the event $X = x$ for some x, the amount of 'surprise' we have at the event A happening should just be some function$f(p)$ of $p := \mathbb{P}(A)$. There are a number of reasonable conditions we should expect f to satisfy:

- $f(1) = 0$, since a certain event is no surprise;
- f should be decreasing, since rarer events are more surprising;
- f is continuous;
- $f(pq) = f(p) + f(q)$, which can be motivated by considering independent events happening with probability p and q;
- finally, for normalisation we may as well assume $f(1/2) = 1$.

It turns out that $f(p) = \frac{1}{\log p}$ is the unique function satisfying these constraints. Then, $H(X)$ is the expected value, taken over the range of X, of the surprise of the event that X takes a certain value, and so $H(X)$ is the only 'reasonable' function representing the idea following these heuristics.

For example if X only takes two values, 0 and 1 with probability p and $1 - p$ respectively, then

$$H(X) = p \log \left(\frac{1}{p} \right) + (1 - p) \log \left(\frac{1}{1 - p} \right),$$

and so as $p \to 1$ or 0, $H(X) \to 0$. It is not hard to see that the entropy of X is maximised when $p = 1/2$, in general we have that:

Lemma 12.1. *Let X be a discrete random variable and let R be the range of X.*

$$H(X) \leq \log (|R|).$$

Proof. We will use the following form of Jensen's inequality. Let f be concave on $[a, b]$, $\lambda_i \geq 0$ such that $\sum_{i=1}^n \lambda_i = 1$ and let $x_1, \ldots x_n \in [a, b]$. Then

$$\sum_{i=1}^n \lambda_i f(x_i) \leq f \left(\sum_{i=1}^n \lambda_i x_i \right).$$

We note that $\log(x)$ is a concave function on $(0,\infty)$, and so

$$H(X) = \sum_{x\in R} p(x)\log\left(\frac{1}{p(x)}\right) \leq \log\left(\sum_{x\in R}\frac{p(x)}{p(x)}\right) = \log(|R|).$$

□

Given two discrete random variables, X and Y, we define the *joint entropy* (X,Y) to be

$$H(X,Y) = \sum_x\sum_y p(x,y)\log\left(\frac{1}{p(x,y)}\right),$$

where, as before, $p(x,y) := \mathbb{P}(X=y, Y=y)$. We also define the *conditional entropy*, of Y given X, to be

$$H(Y|X) = \sum_x p(x)H(Y|X=x) = \mathbb{E}_x\big(H(Y|X)\big).$$

Note the difference between $H(Y|X=x)$, which is the entropy of the random variable $(Y|X=x)$, and $H(Y|X)$. We can think of the conditional entropy as being the expected surprise in learning the value of Y, given that the value of X is known. We might expect, heuristically, that having extra knowledge should only decrease how surprised we are, and indeed that turns out to be the case

Lemma 12.2. *Let X and Y be discrete random variables. Then*

$$H(Y|X) \leq H(Y).$$

Proof. Let us write, as before $p(y|x) = \mathbb{P}(Y=y|X=x)$. Noting that $p(y)p(x|y) = p(x)p(y|x) = p(x,y)$, we see that

$$\begin{aligned}
H(Y|X) &= \sum_x p(x)\sum_y p(y|x)\log\left(\frac{1}{p(y|x)}\right)\\
&= \sum_y p(y)\sum_x p(x|y)\log\left(\frac{1}{p(y|x)}\right)\\
&\leq \sum_y p(y)\log\left(\sum_x\frac{p(x|y)}{p(y|x)}\right)\\
&= \sum_y p(y)\log\left(\sum_x\frac{p(x)}{p(y)}\right)\\
&= \sum_y p(y)\log\left(\sum_x\frac{1}{p(y)}\right)\\
&= H(Y).
\end{aligned}$$

Where in the above we make repeated use of the fact that, if we sum the probabilities of a random taking a specific value over it's entire range the result is 1. □

Lemma 12.3 (Chain rule)**.** *Let X and Y be discrete random variables. Then*

$$H(X,Y) = H(X) + H(Y|X).$$

Proof.

$$\begin{aligned}
H(X,Y) &= \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x,y)}\right) \\
&= \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x)p(y|x)}\right) \\
&= \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(x)}\right) - \sum_x \sum_y p(x,y) \log\left(\frac{1}{p(y|x)}\right) \\
&= \sum_x p(x) \log\left(\frac{1}{p(x)}\right) + \sum_x \sum_y p(x)p(y|x) \log\left(\frac{1}{p(y|x)}\right) \\
&= \sum_x p(x) \log\left(\frac{1}{p(x)}\right) + \sum_x p(x) \sum_y p(y|x) \log\left(\frac{1}{p(y|x)}\right) \\
&= H(X) + \sum_x p(x) H(Y|X=x) \\
&= H(X) + H(Y|X).
\end{aligned}$$

□

One can also define the joint entropy of a sequence of discrete random variables $X_1, X_2, \ldots, X_n$ in a similar way and by induction it follows that

$$H(X_1, X_2, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \ldots H(X_n|X_1, X_2, \ldots, X_{n-1}).$$

We shall sometimes refer to this as the *chain rule*. Note that, by Lemma 12.2 and Lemma 12.3 we have that $H(X_1, X_2, \ldots, X_n) \leq \sum_i H(X_i)$. Finally it is a simple check (left as an exercise) that if the random variable Y completely determines the random variable Z then $H(X|Y) = H(X|Y,Z)$, which again agrees with our intuition.

12.2 Brégman's Theorem

We now give an example of an application of entropy methods to graph theory. Brégman's Theorem was originally a statement about the permanent of a square $(0,1)$ matrix, however it can easily be re-formulated to talk about the number of perfect matchings in a graph. Let G be a graph, $\Phi(G)$ be the set of perfect matchings of G and $\phi(G) = |\Phi(G)|$. The following proof of Brégman's Theorem using entropy methods is due to Radhakrishnan.

Theorem 12.4 (Brégman's Theorem)**.** *Let G be a bipartite graph on vertex classes A and B such that $|A| = |B| = n$. Then*

$$\phi(G) \leq \prod_{v \in A} (d(v)!)^{\frac{1}{d(v)}}.$$

Proof. Let M be a perfect matching of G chosen uniformly at random from $\Phi(G)$. For convenience we will associate A with the set $[n]$ in the natural way, and denote by d_i the degree of the vertex i. For each $i \in [n]$ let X_i be the neighbour of i in M and we identify M with $X = (X_i)_{i=1}^n$. Since we chose M uniformly at random from a set of $\phi(G)$ possibilities we have

that $H(X) = \log(\phi(G))$. Hence if we can bound $H(X)$ from above, we can also bound $\phi(G)$. Note that to get the stated bound we would need to show that

$$H(X) \leq \sum_{i=1}^{n} \frac{\log(d_i!)}{d_i}.$$

A naive first approach might be the use the sub-additivity of entropy to say

$$H(X) \leq \sum_{i=1}^{n} H(X_i),$$

and since there are at most d_i possibilities for the random variable X_i we have that

$$H(X) \leq \sum_{i=1}^{n} H(X_i) \leq \sum_{v \in A} \log(d_i).$$

However, by Stirling's approximation, $\log(d_i!)/d_i \sim \log(d_i/e)$, and so this bound is not enough. However perhaps we can improve this bound by using the chain rule, since we have

$$H(X) = \sum_{i=1}^{n} H(X_i | X_1, X_2, \dots X_{i-1}).$$

We can think of this as revealing the matching one edge at a time, and working out the remaining entropy at each step given what we know. Now instead of just using the naive bound for each X_i we can hopefully take into account the fact that, if we already know $X_1, , X_2, \dots X_{i-1}$ this may reduce the number of possibilities for X_i, since some of the vertices $1, 2, \dots, i-1$ may be matched to neighbours of i in M, reducing the range of X_i.

However, since the matching M was random and the ordering of A were arbitrary, we don't know how many neighbours of i have already been used in M by vertices $j < i$. However, given any permutation σ of $[n]$ we can apply the chain rule with respect to the ordering given by σ to see

$$H(X) = \sum_{i=1}^{n} H(X_{\sigma(i)} | X_{\sigma(1)}, X_{\sigma(2)}, \dots X_{\sigma(i-1)}).$$

For each matching M there are some orderings that will give a significant improvement on the bound above, so if we average over all possible choices of σ

$$H(X) \leq \frac{1}{n!} \sum_{\sigma} \sum_{i=1}^{n} H(X_{\sigma(i)} | X_{\sigma(1)}, X_{\sigma(2)}, \dots X_{\sigma(i-1)}),$$

we might hope to get a reasonable improvement in our bound.

Now, for each X_i there are $n!$ different terms of the form $H(X_i | X_j : j \in J_{\sigma,i})$, for some $J_{\sigma,i} \subset [n] \setminus \{i\}$, appearing in the sum, where $J_{\sigma,i}$ is the set of indices preceding i in the order defined by σ. So we can re-write the sum as

$$\begin{aligned} H(X) &\leq \frac{1}{n!} \sum_{\sigma} \sum_{i=1}^{n} H(X_{\sigma(i)} | X_{\sigma(1)}, X_{\sigma(2)}, \dots X_{\sigma(i-1)}) \\ &= \frac{1}{n!} \sum_{i=1}^{n} \sum_{\sigma} \frac{1}{d_i} H(X_i | X_j : j \in J_{\sigma,i}) \end{aligned}$$

For each of these terms we've reduced the range of X_i by how many of the previously exposed X_j lie in $N(i)$, the neighbourhood of i. Let us denote by $N_\sigma(i) = N(i) \setminus \{X_j : j \in J_{\sigma,i}\}$ the vertices in the neighbourhood of i without those already chosen by some X_j, that is, the range of $(X_i | X_j : j \in J_{\sigma,i})$. It follows that, for any fixed σ and i

$$\begin{aligned} H(X_i | X_j : j \in J_{\sigma,i}) &= \sum_{x_j} \mathbb{P}(X_j = x_j \text{ for } j \in J_{\sigma_i}) H(X_i | X_j = x_j \text{ for } j \in J_{\sigma_i}) \\ &\leq \sum_{j=1}^{d_i} \mathbb{P}(|N_\sigma(i)| = j) \log j \end{aligned}$$

Where we used the defiition of conditional entropy, and then Lemma 12.1. However, since we're picking a random matching, it doesn't seem like we have any control over this improvement, since we don't know how much this will reduce the range of X_i.

However, for any given matching M, if we pick a random ordering σ, we claim that the size of the neighbourhood $|N_\sigma(i)|$ is uniformly distributed between 1 and d_i. Indeed, for a given matching we only care about the order in which we pick i and the vertices matched in M to the neighbours of i. Since i is equally likely to be chosen in any position in this list, the claim follows. In other words, for a fixed matching M, the proportion of σ such that $|N_\sigma(i)| = k$ is $\frac{1}{d_i}$ for each $1 \leq k \leq d_i$.

Since this is true separately for each particular matching, then it is also true when we pick a random matching. So, even though we can't bound any of the terms $\mathbb{P}(|N_\sigma(i)| = j)$ for a fixed σ, we can bound their average.

That is to say, if we pick M and σ both uniformly at random then

$$\mathbb{P}_{\sigma,M}(|N_\sigma(i)| = j) = 1/d_i$$

and hence, by definition

$$\frac{1}{n!} \sum_\sigma \mathbb{P}(|N_\sigma(i)| = j) = \frac{1}{d_i}$$

Hence,

$$\begin{aligned} H(X) &= \frac{1}{n!} \sum_\sigma \sum_{i=1}^n H(X_{\sigma(i)} | X_{\sigma(1)}, X_{\sigma(2)}, \dots X_{\sigma(i-1)}) \\ &= \sum_{i=1}^n \frac{1}{n!} \sum_\sigma H(X_i | X_j : j \in J_{\sigma,i}) \\ &\leq \sum_{i=1}^n \frac{1}{n!} \sum_\sigma \sum_{j=1}^{d_i} \mathbb{P}(|N_\sigma(i)| = j) \log(j) \\ &= \sum_{i=1}^n \sum_{j=1}^{d_i} \left(\sum_\sigma \frac{1}{n!} \mathbb{P}(|N_\sigma(i)| = j) \right) \log(j) \\ &= \sum_{i=1}^n \sum_{j=1}^{d_i} \frac{\log j}{d_i} = \sum_{i=1}^n \frac{\log(d_i!)}{d_i} \end{aligned}$$

giving the bound as claimed. □

Note that this bound is tight. If we take G to be $\frac{n}{d}$ copies of $K_{d,d}$ then we have that $d(v) = d$ for all $v \in A$ and every matching consists of picking one from the $d!$ possible matchings on each $K_{d,d}$. Therefore.

$$\phi(G) = \prod_{i=1}^{\frac{n}{d}} d! = \prod_{v \in A} (d(v)!)^{\frac{1}{d(v)}}.$$

12.3 Shearer's lemma and the Box theorem

Given a sequence of discrete random variables random variables $X_1, X_2, \ldots, X_n$ and some subset $A \subseteq [n]$ let define $X_A := (X_i : i \in A)$.

Lemma 12.5 (Shearer's inequality)**.** *Let $X_1, X_2, \ldots, X_n$ be discrete random variables and $\mathcal{A}$ a collection (not necessarily distinct) of subsets of $[n]$, such that each $i \in [n]$ is in at least m members of $\mathcal{A}$. Then*

$$H(X_1, X_2, \ldots, X_n) \leq \frac{1}{m} \sum_{A \in \mathcal{A}} H(X_A).$$

Proof. Let $A = \{a_1, a_2, \ldots, a_k\}$ with $a_1 < a_2 < \ldots < a_k$. We have that

$$\begin{aligned} H(X_A) &= H(X_{a_1}) + H(X_{a_2}|X_{a_1}) + \ldots + H(X_{a_k}|X_{a_1}, X_{a_2}, \ldots, X_{a_{k-1}}) \\ &\geq H(X_{a_1}|X_{<a_1}) + H(X_{a_2}|X_{<a_2}) + \ldots + H(X_{a_k}|X_{<a_k}), \end{aligned}$$

where $X_{<i} = (X_1, X_2, \ldots X_{i-1})$. This follows from repeated applications of the chain rule, and the fact that entropy only decreases if we condition on more variables. Therefore

$$\begin{aligned} \sum_{A \in \mathcal{A}} H(X_A) &\geq m. \sum_{i \in [n]} H(X_i|X_{<i}) \\ &= m.H(X_1, X_2, \ldots, X_n) \end{aligned}$$

□

Shearer's Lemma is closely related to notions of isoperimetry, the relation between the volume of a shape and it's 'perimeter' in the following way. If we think about a shape $S \subseteq \mathbb{R}^n$ with area $|S|$ then we can think about the process of picking a random point inside of S. This determines a vector $X = (X_1, \ldots, X_n)$ where the X_i are dependent on each other, depending on what the shape S is.

Suppose we take a very fine grid approximating $\mathbb{R}^n$, we can then think of A as being a discrete subset of this grid, with $\sim |A|$ many points. Since this vector $X = (X_1, \ldots, X_n)$ now has some finite range, we can relate the volume of S directly to the entropy of X. That is

$$H(X) = \log |S|.$$

How can we interpret the random variable X_A for $A \subset [n]$? Well in this case, this is relatively clear, these correspond to the projections on the shape A onto the subspace spanned by the coordinates in A. That is, if we let S_A be the projection of S onto the subspace

$$\{(x_1, \ldots, x_n) : x_i = 0 \text{ for all } i \in A\}$$

Then the range of X_A is the 'volume' (in the $n - |A|$-dimensional sense) of S_A. We will write S_j for $S_{\{j\}}$.

In this way, Shearer's inequality gives us a way to relate the volume of a shape to it's lower dimensional projections. For example, if we just consider the 1-dimensional projections, we have the famous Loomis Whitney inequality:

Theorem 12.6 (The Loomis-Whitney inequality)**.** *Let $S \subset \mathbb{Z}^n$ then,*

$$|S|^{n-1} \leq \prod_{i=1}^{n} |S_{[n]\setminus\{i\}}|$$

For example, in two dimensions this simply says that the area of a shape can be bounded above by the product of its one-dimensional projections, a relatively trivial fact. But even in three-dimensions it is not clear what the relationship should be between the volume of a shape and its projections onto two dimensional subspaces.

Notice that, this theorem is tight when $|S|$ is a 'box', that is, a set of the form $[1, m_1] \times [1, m_2] \times \ldots \times [1, m_n]$. Indeed, the volume of $|S|$ is $\prod_{i=1}^{n} m_i$ and the volume of the projection of S onto the hyperplane where $x_i = 0$ is just $\prod_{j \neq i} m_i$. This is perhaps not surprise, as a box represents the case where the X_is are independent, where we get equality in the argument for Shearer's inequality.

In fact, we will show a more general theorem, and deduce the Loomis-Whitney theorem as a corollary. We say a collection of sets $\mathcal{C} = \{C_1, \ldots, C_m\} \subset 2^{[n]}$ is a *k-uniform cover* if each $i \in [n]$ belongs to exactly k many of the C_j.

Theorem 12.7 (Uniform covers theorem)**.** *Let $S \subset \mathbb{Z}^n$ and let $\mathcal{C} \subset 2^{[n]}$ be a k-uniform cover, then*

$$|S|^k \leq \prod_{C \in \mathcal{C}} |S_C|$$

Remark 12.8. *Note that $\mathcal{C} = \{[n] \setminus \{i\} : i \in [n]\}$ is an $(n-1)$-uniform cover of $[n]$, and so Theorem 12.6 follows from Theorem 12.7.*

Proof. Let us choose a points $X = (X_1, \ldots, X_n)$ uniformly at random from S. Then, $H(X) = \log |S|$. By Lemma 12.5 it follows that

$$H(X) \leq \frac{1}{k} \sum_{C \in \mathcal{C}} H(X_C).$$

However, the range of X_C is $|S_C|$ and so it follows that

$$H(X) \leq \frac{1}{k} \sum_{C \in \mathcal{C}} \log |S_C|.$$

Combining the two equatons we see that

$$\log |S| \leq \frac{1}{k} \sum_{C \in \mathcal{C}} \log |S_C|$$

and so

$$|S|^k \leq \prod_{C \in \mathcal{C}} |S_C|,$$

as claimed. □

As before, if we consider the 1-uniform cover $\{\{i\} : i \in [n]\}$, Theorem 12.7 tells us the elementary fact the volume of a shape can be bounded by the product of its one-dimensional projections.

If we go back to our original continuous setting, it is relativelys simple to show that Theorem 12.7 still holds for any 'reasonable' (say, measurable) shape $S \subset \mathbb{R}^n$, where $|.|$ now denotes the normal volume, by taking a limit of finer and finer grids. Bollobás and Thomason used Theorem 12.7 to prove quite a surprising theorem:

Theorem 12.9 (Bollobás-Thomason Box Theorem)**.** *Let $S \subset \mathbb{R}^n$. Then there is a box $B \subset \mathbb{R}^n$ such that $|B| = |S|$ and $|B_A| \leq |S_A|$ for all $A \subseteq [n]$.*

That is, for any shape we can find a box of the same area such that *every* lower dimensional projection of this box has smaller volume than the corresponding projection of S. This immediately tells us that for any upper bound we might want to prove for the volume of a set given the volumes of its projection, we only have to check that it holds for boxes.

Indeed, if we know that for every box B, $|B| \leq f(B_A : A \subset [n])$ for some function f which is increasing in each coordinate, then for any S we have that $|S| = |B| \leq f(B_A : A \subset [n]) \leq f(S_A : A \subset [n])$.

We will need the following definition and lemma. Let $\mathcal{C}$ be a uniform cover of $[n]$ we say $\mathcal{C}$ is *irreducible* if we cannot write it as the disjoint union $\mathcal{C} = \mathcal{C}' \cup \mathcal{C}''$ of two uniform covers. Note, this is equivalent to saying there is no non-trivial subset $\mathcal{C}' \subset \mathcal{C}$ which is a uniform cover.

Lemma 12.10. *There are only finitely many irreducible uniform covers of $[n]$*

Proof. Suppose we have a sequence $\mathcal{C}_1, \mathcal{C}_2, \ldots$ of distinct irreducible uniform covers of $[n]$. Let us list the elements of $2^{[n]}$ as $E_1, E_2, \ldots, E_{2^n}$. We pick a subsequence $\mathcal{C}_{i_1}, \mathcal{C}_{i_2}, \ldots$ of the covers such that the number of copies of E_1 in $\mathcal{C}_{i_j}$ is increasing (not necessarily strictly). Then we choose a subsequence of this sequence on which the number of copies of E_2 is increasing. We do this for each E_j with $j \leq 2^n$, so that we have a subsequence $\mathcal{C}_{k_1}, \mathcal{C}_{k_2}, \ldots$ on which the number of copies of each E_j is increasing for all j. However, then $\mathcal{C}_{k_1} \subset \mathcal{C}_{k_2}$, contradicting the assumption that $\mathcal{C}_{k_2}$ is irreducible. □

Proof of Theorem 12.9. We may assume that $n \geq 2$. We want to pick real numbers x_A for each $A \subseteq [n]$ with $A \neq \emptyset, [n]$ such that:

- $0 \leq x_A \leq |S_A|$ for all $A \subseteq [n]$;
- $x_A \leq \prod_{i \in A} x_i$ for all $A \subseteq [n]$;
- $|S|^k \leq \prod_{C \in \mathcal{C}} x_C$ for every k-uniform irreducible cover $\mathcal{C} \neq \{[n]\}$.

Note that if the third condition is satisfied for irreducible uniform covers, it is satisfied for all uniform covers.

By Theorem 12.7 these conditions hold if we let $x_A = |S_A|$, so at least one choice exists. Furthermore its not hard to see that our solution set is closed and bounded, and so compact,

and hence we can choose one which minimises $\sum_A x_A$. Note further that $x_A > 0$ for every A, since every A appears in some uniform cover.

Claim. For every $i \in [n]$, there is some irreducible k-uniform cover $\{i\} \in \mathcal{C}_i$ such that $|S|^k = \prod_{C \in \mathcal{C}_i} x_C$.

Proof of claim. There are only finitely many inequalities we have to satisfy. Clearly if for every inequality where x_i appeared the inequality was strict, then we could reduce x_i by some small amount such that all the inequalities still held, contradicting the minimality of $\sum_A x_A$.

Hence there is some inequality in which x_i appears which is in fact an equality. Suppose it is one of the form $x_A = \prod_{i \in A} x_i$ for $i \in A \subseteq [n]$. Then, by the same argument applied to A, there is some irreducible cover $A \in \mathcal{C}$ such that $|S|^k = \prod_{C \in \mathcal{C}} x_C$.

However we can consider the uniform cover given by $\mathcal{C}' = \mathcal{C} \setminus \{A\} \cup \{\{j\} : j \in A\}$. This is not irreducible, but it's still true that

$$|S|^k = \prod_{C \in \mathcal{C}'} x_C.$$

Now, there is some irreducible subcover of $\mathcal{C}'$ which contains $\{i\}$, in which equality must also hold, proving the claim. □

Consider the uniform cover given by $\mathcal{C} := \bigcup_i \mathcal{C}_i$. This is also a $(k'+1)$-uniform cover such that

$$|S|^{k'+1} = \prod_{C \in \mathcal{C}_i} x_C,$$

and furthermore $\{1\}, \{2\}, \ldots, \{n\} \in \mathcal{C}$. Hence $\mathcal{C}' := \mathcal{C} \setminus \{\{1\}, \{2\}, \ldots, \{n\}\}$ is a k'-uniform cover and we know that

$$|S|^{k'} \leq \prod_{C \in \mathcal{C}'} x_C \qquad \text{and} \qquad |S|^{k'+1} = \prod_{C \in \mathcal{C}} x_C = \prod_{C \in \mathcal{C}'} x_C \times \prod_{i=1}^{n} x_i.$$

It follows that $|S| = \prod_{i=1}^n x_i$, and so this is a good choice of a box whose volume is S. Furthermore, for any $A \subseteq [n]$, $A \neq \emptyset, [n]$ we can consider the 1-uniform cover $\{A, A^c\}$ of $[n]$. It follows that

$$|S| \leq x_A x_{A^c} \leq \left(\prod_{i \in A} x_i\right)\left(\prod_{i \notin A} x_i\right) = |S|,$$

and so $x_A = \prod_{i \in A} x_i$. Hence, if we consider the box $B = [0, x_1] \times [0, x_2] \times [0, x_n]$, we have for every $\emptyset \neq A \subset [n]$

$$|B_A| = \prod_{i \in A} x_i = x_A \leq |S_A|.$$

□

12.4 Independent Sets in a Regular Bipartite Graph

Let G be a d-regular bipartite graph on $2n$ vertices with vertex classes A and B, and let $\mathcal{I}(G)$ be the class of independent subsets of $V(G)$. We would like to bound this number from above. As in this previous case, letting G be a disjoint union of $K_{d,d}$'s seems a natural guess for a best possible graph. Indeed in G it is clear that any independent set in G consists of an arbitrary subset taken from one side of each $K_{d,d}$. Therefore we have that

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

The following proof of a corresponding upper bound on $|\mathcal{I}(G)|$ using entropy methods is due to Kahn.

Theorem 12.11. *Let G be a d-regular bipartite graph on $2n$ vertices with vertex classes A and B, and let $\mathcal{I}(G)$ be the class of independent subsets of $V(G)$. Then*

$$|\mathcal{I}(G)| \leq (2^{d+1} - 1)^{\frac{n}{d}}$$

Proof. The basic idea of the proof is the same as in Theorem 12.4, we pick a random independent set I and estimate the entropy $H(I)$. As before we have that $H(I) = \log(|\mathcal{I}|)$.

We identify I with it's characteristic vector $(X_v : v \in A \cup B)$, note that I is determined by (X_A, X_B). The idea is that, rather than splitting X into X_v for each v, we can use the neighbourhoods of each $v \in A$ as a d-uniform cover of the vertices of B, and so use Shearer's Lemma to express X_B in terms of $X_{N(v)}$.

For each $v \in A$ let $N(v)$ be the neighbourhood of v in B. Each $w \in B$ is in exactly d of the sets $N(v)$ and so we have

$$\begin{aligned} H(I) &= H(X_A|X_B) + H(X_B) \\ &\leq \sum_{v \in A} H(X_v|X_B) + \frac{1}{d}\sum_{v \in A} H(X_{N(v)}) \\ &\leq \sum_{v \in A} \big(H(X_v|X_{N(v)}) + \frac{1}{d}H(X_{N(v)}\big), \end{aligned}$$

where the second line follows from Shearer's inequality, and the third since $N(v) \subset B$.

Fix some $v \in A$. Let χ_v be the indicator random variable of the event that $I \cap N(v) \neq \emptyset$, and let $p := \mathbb{P}(\chi_v = 0)$, that is the probability that $I \cap N(v) = \emptyset$. The nice thing about this random variable is that it contains all the information about $X_{N(v)}$ that we need to determine $H(X_v|X_{N(v)})$.

Hence ,

$$\begin{aligned} H(X_v|X_{N(v)}) &\leq H(X_v|\chi_v) \\ &= \mathbb{P}(\chi_v = 0)H(X_v|\chi_v = 0) + \mathbb{P}(\chi_v = 1)H(X_v|\chi_v = 1) \\ &= \mathbb{P}(\chi_v = 0)H(X_v|\chi_v = 0) \leq p, \end{aligned}$$

since the event $\chi_v = 1$ determines that $X_v = 0$, and since $H(X_v) \leq \log(|\text{range}(X_v)|) = 1$.

Also,

$$\begin{aligned} H(X_{N(v)}) &= H(X_{N(v)}, \chi_v) \\ &= H(\chi_v) + H(X_{N(v)}|\chi_v) \\ &\leq H(p) + (1-p)\log(2^d - 1), \end{aligned}$$

where $H(p) = p\log(1/p) + (1-p)\log(1/(1-p))$. Putting these inequalities together gives us

$$H(I) \leq \sum_{v \in A} (p + \frac{1}{d}\left(H(p) + (1-p)\log\left(2^d - 1\right)\right).$$

All that remains is to maximise the quantity on the right hand side according to p. It is a simple exercise to check that the function is convex, and to calculate its derivative, giving that the maximum is attained at $p = 2^d/(2^{d+1} - 1)$, and so $(1-p) = 2^d - 1/(2^{d+1} - 1)$ giving that:

$$\begin{aligned} H(I) &\leq \sum_{v \in A} (p + \frac{1}{d}\left(H(p) + (1-p)\log\left(2^d - 1\right)\right) \\ &= n\left(p + \frac{1}{d}\left(p\log(1/p) + (1-p)\log(1/(1-p)) + (1-p)\log\left(2^d - 1\right)\right)\right) \\ &= n\left(p + \frac{1}{d}\left(p\log\left(\frac{2^{d+1}-1}{2^d}\right) + (1-p)\log\left(\frac{2^{d+1}-1}{2^d - 1}\right) + (1-p)\log\left(2^d - 1\right)\right)\right) \\ &= n\left(p + \frac{1}{d}\left(p\log\left(2^{d+1} - 1\right) - pd + (1-p)\log\left(2^{d+1} - 1\right)\right)\right) \\ &= n\left(p - p + \frac{1}{d}\left((p + (1-p))\log\left(2^{d+1} - 1\right)\right)\right) \\ &= n\frac{1}{d}\log\left(2^{d+1} - 1\right) \end{aligned}$$

$$\log(|\mathcal{I}|) = H(I) \leq n.\frac{1}{d}\log\left(2^{d+1} - 1\right),$$

from which the result follows. □

12.5 Bipartite Double Cover

Theorem 12.11 tells us the maximum number of independent sets in a d-regular graph when it is bipartite, however it also makes sense to ask that question for general graphs. After a little thought it may seem that the same example from the previous sections, a disjoint union of $K_{d,d}$s, is still a sensible graph to consider. In fact it had been conjectured that this was the best possible case for general d-regular graphs for around 10 years, even before Kahn proved it for bipartite graphs.

Then, another 10 years after Kahn's result, whilst still an undergraduate, Zhao used an ingenious, and simple, argument to show that the bipartite case actually implies the general case, using the idea of a bipartite double cover of a graph.

Given any graph G we define the *bipartite double cover* of G to be the tensor product, $G \times K_2$. Recall that this is a graph with vertex set

$$\{(v, i) : v \in V(G) \text{ and } i \in \{0, 1\}\}$$

and edge set

$$\{\big((v,0),(w,1)\big) \,:\, (v,w) \in E(G)\}.$$

If G is d-regular then it is clear that $G \times K_2$ is also. Any independent set in $G \times K_2$ corresponds to a pair of subsets of V, (A, B), such that there is no edge between A and B in G, we would like to relate the number of such pairs in some way to the number of independent sets in G.

We say that A is *independent from* B if there are no edges between A and B. If we let $\mathcal{J}(G)$ be the set of (A, B) such that A is independent from B and $G|_{A\cup B}$ is bipartite, then we have that $\mathcal{J}(G) \subset \mathcal{I}(G \times K_2)$. We can relate $\mathcal{J}(G)$ to $\mathcal{I}(G)$ in the following way. .

Lemma 12.12. *For any graph G, there exists a bijection between $\mathcal{I}(G) \times \mathcal{I}(G)$ and $\mathcal{J}(G)$.*

Proof. For every $X \subset V$ such that $G|_X$ is bipartite let us fix a particular bipartition $X = X_1 \cup X_2$, such that X_1 and X_2 are independent in G.

We let $\mathcal{K}(G)$ be the set of pairs of subsets (A, B) such that $G|_{A\cup B}$ is bipartite. Note that, in particular $\mathcal{I}(G) \times \mathcal{I}(G) \subset \mathcal{K}(G)$ and $\mathcal{J}(G) \subset \mathcal{K}(G)$.

Given some $(A, B) \in \mathcal{K}(G)$, since $G|_{A\cup B}$ is bipartite, we fixed some particular bipartition $X_1 \cup X_2$ of $A \cup B$ at the beginning of the proof. We note that

$$\phi(A, B) := \big((A \cap X_1) \cup (B \cap X_2), (A \cap X_2) \cup (B \cap X_1)\big)$$

is also an element of $\mathcal{K}(G)$, and that $\phi^2(A, B) = (A, B)$. So ϕ is an involution on $\mathcal{K}(G)$, and it is clear that ϕ is size preserving.

However we claim that ϕ maps $\mathcal{I}(G) \times \mathcal{I}(G)$ to $\mathcal{J}(G)$, and vice versa. Indeed, given $(A, B) \in \mathcal{I}(G) \times \mathcal{I}(G)$ let $(C, D) = \phi(A, B)$. Since A, B, X_1 and X_2 are independent, it is clear that C is independent from D, and $G|_{C\cup D} = G|_{A\cup B}$ is bipartite. Therefore $(C, D) \in \mathcal{J}(G)$.

Conversely, given $(C, D) \in \mathcal{J}(G)$ let $(A, B) = \phi(C, D)$. Since C is independent from D and X_1 and X_2 are independent, we have that both A and B are independent, and so $(A, B) \in \mathcal{I}(G) \times \mathcal{I}(G)$.

However, since ϕ is an involution, it must therefore be a bijection between $\mathcal{I}(G) \times \mathcal{I}(G)$ and $\mathcal{J}(G)$. □

Theorem 12.13. *Let G be a d-regular graph on n vertices, and let $\mathcal{I}(G)$ be the class of independent subsets of $V(G)$. Then*

$$|\mathcal{I}(G)| \leq (2^{d+1} - 1)^{\frac{n}{2d}}$$

Proof. We consider the bipartite double cover of G. Recall that an independent set in $G \times K_2$ is a pair of subsets (A, B) such that A is independent from B. We have by Lemma 12.12 that

$$|\mathcal{I}(G \times K_2)| \geq |\mathcal{J}(G)| = |\mathcal{I}(G) \times \mathcal{I}(G)| = |\mathcal{I}(G)|^2$$

However, since $G \times K_2$ is bipartite and d-regular, we can apply Theorem 12.11 to see that

$$|\mathcal{I}(G \times K_2)| \leq (2^{d+1} - 1)^{\frac{n}{d}}.$$

Combining the two inequalities gives the desired result. □

13 Derandomization and Combinatorial Games

13.1 Maximum Cuts in Graphs

Let us consider with the following algorithmic problem, known as MAXCUT: given a graph G, paritition V into two classes A and $B = V \setminus A$ such that the number of edges going between A and B is maximised. This problem is in fact NP-complete. The following application of the probabilistic method tell us that we can always achieve at least half of the edges in the graph.

Theorem 13.1. *Any graph G, with $e(G) = m$, contains a bipartite subgraph with at least $m/2$ edges.*

Proof. Given $G = (V, E)$ let us choose a subset $A \subseteq V$ randomly by including each vertex in A independently with probability $1/2$. We consider the partition of V into A and $B = V \setminus A$. For any given edge $e = (u, v)$ let X_e be the indicator variable of the event that e is in the bipartite subgraph between A and B. So we have that

$$\mathbb{E}(X_e) = \mathbb{P}\big((u \in A \text{ and } v \in B) \text{ or } (u \in B \text{ and } v \in B)\big) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Let X be the number of edges in the bipartite subgraph between A and B, it is apparent that $X = \sum_{e \in E(G)} X_e$, and so

$$\mathbb{E}(X) = \sum_{e \in E(G)} \mathbb{E}(X_e) = \frac{m}{2}.$$

Therefore for some choice of A the random variable X has value $\geq m/2$, and so the bipartite subgraph of G between A and B contains at least $m/2$ edges. □

We note that while the theorem asserts the existence of a large cut, it gives no indication of how to find one deterministically. However in many cases, including this one in particular, we can "derandomize" such an argument to produce a fully deterministic algorithm.

Example. Suppose we wish to construct a partition of the vertex set with a large number of edges across it. First we order the vertices, let us suppose they are labelled $1, \ldots, n$ and we are going to go through them in turn and decide if they should be in A or B.

To do this we define a *potential function* w which is defined on the set of partial bipartitions, that is pairs of subsets $(A, B) \subset V^2$ such that $A \cap B = \emptyset$. Given (A, B) we extend it randomly to a bipartiton of V, that is independently for each vertex $v \in V \setminus (A \cup B)$ we add v to A with probability $1/2$ or B with probability $1/2$. We end up with some bipartition $V = A' \cup B'$ such that $A \subset A'$ and $B \subset B'$. We define the potential of (A, B), $w(A, B)$, to be the expected number of edges across this partition.

The algorithm can then be described as follows, we start with $(A_0, B_0) = (\emptyset, \emptyset)$. We will construct a sequence of partial bipartitions $\{(A_i, B_i : i \in [n]\}$ such that $A_0 \subset A_1 \subset \ldots \subset A_n$, $B_0 \subset B_1 \subset \ldots \subset B_n$ and $A_i \cup B_i = \{1, 2, \ldots, i\}$, that is (A_i, B_i) is a bipartition of the first i vertices of the graph. Suppose we have constructed (A_{i-1}, B_{i-1}), to construct A_i, B_i we have to choose between the two bipartitions $(A_{i-1} \cup i, B_{i-1})$ and $(A_{i-1}, B_{i-1} \cup i)$, to do so we pick the one with the largest potential. Clearly this defines algorithmically a bipartition (A_n, B_n) of the graph G.

Let us note a few things about the algorithm. Firstly we see that $w(A_0, B_0)$ is just the expected number of edges across a random bipartition, which we calculated to be $m/2$ in the previous theorem. Also we note that, for any i

$$w(A_i, B_i) = \frac{1}{2}\big(w(A_{i-1} \cup i, B_{i-1}) + w(A_{i-1}, B_{i-1} \cup i)\big),$$

since any random bipartition extending (A_{i-1}, B_{i-1}) must extend one of $(A_{i-1} \cup i, B_{i-1})$ or $(A_{i-1}, B_{i-1} \cup i)$, and it has an equal probability (of $1/2$) of extending either. Hence

$$\max\big(w(A_{i-1} \cup i, B_{i-1}), w(A_{i-1}, B_{i-1} \cup i)\big) \geq w(A_i, B_i),$$

and so at each stage of the algorithm the potential of the partial bipartition is non-decreasing. Therefore we can conclude that

$$w(A_n, B_n) \geq w(A_0, B_0) = \frac{m}{2}.$$

However, $w(A_n, B_n)$ is just the number of edges across the bipartition (A_n, B_n), and hence this algorithm produces a bipartition with at least $m/2$ edges going across it.

13.2 Ramsey graphs

Let us applying the same reasoning to a similar problem. In Section 2 we used the basic probabilistic method to show that, if

$$2\binom{n}{k}2^{-\binom{k}{2}} < 1$$

then $R(k, k) > n$. That is, there is a graph on n vertices which contains no clique or independent set of size k. Again, the proof was non-constructive, but we can use the idea of de-randomization to find an algorithm, which will in fact be polynomial in the size of n (for k fixed), to construct such a graph. It will be simpler to think about finding a 2-colouring of K_n without a monochromatic K_k.

We argue in a similar way to as before. We define a potential function w on the set of all partial colourings of the edge set of K_n, which will be the expected number of monochromatic K_k's in a random extension of that colouring. Explicitly if we have a colouring $c : E \to \{\text{red}, \text{blue}\}$ on some subset $E \subset E(G)$ we choose a random colouring $c' : E(G) \to \{\text{red}, \text{blue}\}$ such that $c' = c$ on E by choosing for each edge in $E(G) \setminus E$ the colour red or blue independently with probability $1/2$. We define the potential of a partial colouring c, $w(c)$, to be the expected number of monochromatic K_ks in such a c'.

So, we pick an arbitrary ordering $\{e_0, e_1, \ldots, e_m\}$ of the edges of K_n and we define a colouring algorithmically as follows: we start with the empty partial colouring, c_0. Suppose we have a colouring c_{i-1} of the set $E_{i-1} = \{e_1, e_2, \ldots, e_{i-1}\}$ we consider the two colourings of E_i, c_i^r, which equals c_{i-1} on E_{i-1} and $c_i^r(e_i) = \text{red}$, and c_i^b, which equals c on E_{i-1} and $c_i^b(e_i) = \text{blue}$. We let c_i be the partial colouring of minimal weight. Again, this will define algorithmically a colouring of $E_m = E(G)$.

We note that, for every i

$$w(c_{i-1}) = \frac{1}{2}\big(w(c_i^r) + w(c_i^b)\big)$$

and so

$$w(c_i) \leq w(c_{i-1}).$$

Therefore $w(c_m) \leq w(c_0)$. However $w(c_0)$ is just the expected number of monochromatic k-sets in a random 2-colouring of K_n. Hence if

$$w(c_0) = 2\binom{n}{k}2^{-\binom{k}{2}} < 1,$$

then we can conclude the $w(c_m) < 1$ as well. However, $w(c_m)$ is the number of monochromatic k-sets in the colouring c_m, and so, since it is an integer, we can conclude the $w(c_m) = 0$. That is, c_m is a 2-colouring of K_n which contains no monochromatic set of size k, as claimed.

We note that this polynomial has running time $O(n^{k+2})$ for fixed k. Indeed, to compute the potential of any partial colouring we need to calculate the expected number of monochromatic k-sets in an extension of that colouring. However we can split this up into the sum of the indicator functions of each specific k-set being monochromatic. There are $\binom{n}{k} = O(n^k)$ such sets, and for each k-set to calculate the probability it will be monochromatic we only need to look at the colouring restricted to the $\binom{k}{2}$ edges contained in it, since the probability is a function of that colouring. Hence for each partial colouring the potential can be calculated in time $O(n^k)$, and so, since the algorithm has $m = \binom{n}{2}$ steps, each taking time at most $O(n^k)$, we have that the total algorithm has run-time $O(n^{k+2})$.

13.3 Positional Games

Definition. A *strong positional game* consists of a pair $(X, \mathcal{F})$ where X is a set, called the *board*, and $\mathcal{F} \subset 2^X$ is a family of *winning lines*. The game is played by two players, sometimes referred to as Red and Blue, who take turns claiming points of the board (with Red going first), which we may think of as colouring some point $x \in X$ as either red or blue. Given a particular play of the game, that is a sequence of moves $(r_1, b_2, r_3, b_4 \ldots)$, the winner is the first player to claim all points in some winning set $F \in \mathcal{F}$. If at no point during the game either player achieves this, the game is a draw.

For an example, a well known strong positional game is the game Tic-Tac-Toe, or noughts and crosses. If Red has a strategy to win a game $(X, \mathcal{F})$ we call the game *Red-win*, and similarly for Blue. If both players have a drawing strategy we call the game a draw.

Lemma 13.2. *If X is finite then all strong positional games $(X, \mathcal{F})$ are either Red-win, Blue-win, or a draw.*

Proof. If we think of any particular play of the game as being a sequence of moves $(r_1, b_2, r_3, b_4 \ldots)$, then if the game is Red-win we know that

$$\exists r_1 \, \forall b_2 \, \exists r_3 \, \forall b_4 \ldots \text{ such that } \{r_1, b_2, r_3, b_4 \ldots\} \text{ is a win for Red,}$$

and if the game is Blue win

$$\forall r_1 \, \exists b_2 \, \forall r_3 \, \exists b_4 \ldots \text{ such that } \{r_1, b_2, r_3, b_4 \ldots\} \text{ is a win for Blue.}$$

So, if neither happens, by De Morgan's laws both

$$\forall r_1 \, \exists b_2 \, \forall r_3 \, \exists b_4 \ldots \text{ such that } \{r_1, b_2, r_3, b_4 \ldots\} \text{ is not a win for Red,}$$

and

$$\exists r_1\, \forall b_2\, \exists r_3\, \forall b_4 \ldots \text{ such that } \{r_1, b_2, r_3, b_4 \ldots\} \text{ is not a win for Blue,}$$

and so both players have a drawing strategy. □

It is a folklore theorem that in fact every strong positional game is either a Red-win or a draw. The following argument is usually referred to as strategy stealing.

Theorem 13.3. *Let $(X, \mathcal{F})$ be a strong positional game with X finite. Then $(X, \mathcal{F})$ is either a Red-win or a draw.*

Proof. Let us assume that Blue has a winning strategy Φ. We use Φ to describe a winning strategy for Red as follows. Red claims his first point arbitrarily, and from this point onwards in the game he ignores the point and pretends to be Blue.

That is, he responds to each of Blue's moves according to the strategy Φ as if he had not taken the first point and Blue was the first player. If at any point the strategy calls for him to claim a point that he has already taken, but is ignoring, then he claims another points arbitrarily and ignores that one instead.

Since the arbitrary extra point game only help Red, and since Φ was a winning strategy for blue, against any strategy of Blue's Red will claim a winning set first, including when Blue plays according to the strategy Φ. However this contradicts the assumption that Φ was a winning strategy. □

We note that both of the preceeding theorems also hold in the case where X is infinite, but $\mathcal{F}$ is a set of finite subsets of X.

If we consider $(X, \mathcal{F})$ as a hypergraph, then we note that at the end of a game which is a draw, the resulting colouring of the hypergraph contains no monochromatic winning set by definition. Therefore if $(X, \mathcal{F})$ is such that *every* 2-colouring of X contains a monochromatic $F \in \mathcal{F}$, then we know that no play of the game can end in a draw, and so $(X, \mathcal{F})$ must be Red-win.

Conversely there is a very natural condition, arising from a probabilistic proof, that ensures that such a 2-colouring does exist:

Lemma 13.4. *Suppose $(X, \mathcal{F})$ is a hypergraph. If*

$$\sum_{F \in \mathcal{F}} 2^{-|F|} < 1/2$$

then there exists a 2-colouring of X containing no monochromatic $F \in \mathcal{F}$.

Proof. This is a simple application of the probabilistic method. We choose a 2-colouring uniformly at random by picking, for each $x \in X$ independently, the colour of x to be red or blue with probability $1/2$. If we let X be the number of monochromatic F in this colouring, we can split

X into the sum of the indicator random variables of the event that each F is monochromatic, which happens with probability $2^{1-|F|}$. Therefore

$$\mathbb{E}(X) = \sum_{F \in \mathcal{F}} 2^{1-|F|} < 1.$$

Hence there exists some colouring for which the number of monochromatic F is less than 1, that is, there are no monochromatic F. □

So, if the condition above is satisfied, if both players played randomly we would expect the game to be a draw. We would like to derandomize this argument to find a strategy for Blue that will show that the game in fact is a draw. We note that, by Theorem 13.3, it would be sufficient to find a strategy for Blue which stops Red from winning.

Theorem 13.5. *[The Erdős-Selfridge Theorem] Suppose $(X, \mathcal{F})$ is a strong positional game. If*

$$\sum_{F \in \mathcal{F}} 2^{-|F|} < 1/2$$

then $(X, \mathcal{F})$ is a draw.

Proof. Let us first enumerate $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$. We will a define a potential function ϕ on the set of partial colourings of the board. Given some partial colouring $c : S \to \{\text{red,blue}\}$, let us define for each $F_i \in \mathcal{F}$

$$f_i = \text{the number of } x \in F_i \text{ coloured red}$$

We can then define the potential of each F_i to be

$$\phi_c(F_i) = \begin{cases} 0 & \text{if there is some } x \in F_i \text{ coloured blue} \\ 2^{f_i - |F_i|} & \text{otherwise} \end{cases}$$

We note that the potential of a set is just the probability that that set would be coloured all red in a random extension of c'. We can think of the potential as being a weighted measure of how 'dangerous' a set F_i is for Blue. If there are many red points in the set, then the set is quite close to making Red win, and so Blue should want to colour a point in that set.

We then define

$$\phi(c) = \sum_{i=1}^{m} \phi_c(F_i)$$

and note that $\phi(c)$ is the expected number of red winning sets in a random extension of c'. We note that the potential of the empty colouring is less than $1/2$.

So let us consider what can happen to the potential when Red moves. When Red colours a point $x \in X$ red, changing the partial colouring c to c', what is the effect on the potential? For every $x \in F_i \in \mathcal{F}$ which doesn't already have a blue point in we have that

$$\phi_{c'}(F_i) = 2\phi_c(F_i),$$

and for every $x \in F_i \in \mathcal{F}$ which does already have a blue point in, this equation is also true, since both sides are zero. Therefore the change in potential after Red's move is to increase the potential by

$$\sum_{x \in F_i} \phi_c(F_i).$$

In particular, after Red's first move the total potential is less than 1. We would like to argue that Blue has a strategy that will always decrease this potential. A natural strategy to investigate would be the one in which Blue always chooses the point minimising the potential. Suppose Blue plays according to this strategy Let us consider, after the first move, some pair of moves, consisting of a move of Blue's and a move of Red's. At this point we have some partial colour c on a subset $S \subset X$.

What happens to the potential when Blue colours a point $x \in X \setminus S$ blue, again changing the partial colouring c to c'? Well, for every $x \in F_i \in \mathcal{F}$, colouring x blue changes the potential of F_i to zero and so the total change in potential is

$$-\sum_{x \in F_i} \phi_c(F_i).$$

So, if Blue chooses the x which minimises this quantity we have that, after Red has coloured some point $y \in X \setminus (S \cup x)$ red, changing c' to c'',

$$\begin{aligned}\phi(c'') &= \phi(c') + \sum_{y \in F_i} \phi_{c'}(F_i) = \phi(c) - \sum_{x \in F_i} \phi_c(F_i) + \sum_{y \in F_i} \phi_{c'}(F_i) \\ &= \phi(c) - \max_{x \in X \setminus S} \sum_{x \in F_i} \phi_c(F_i) + \sum_{y \in F_i} \phi_{c'}(F_i) \\ &\leq \phi(c) - \max_{x \in X \setminus S} \sum_{x \in F_i} \phi_c(F_i) + \max_{y \in X \setminus (S \cup x)} \sum_{y \in F_i} \phi_{c'}(F_i).\end{aligned}$$

However we note that, since the difference between c and c' is a singe point being coloured blue, we have that $\phi_{c'}(F_i) \leq \phi_c(F_i)$ for all $F_i \in \mathcal{F}$ and so

$$\phi(c'') \leq \phi(c) - \max_{x \in X \setminus S} \sum_{x \in F_i} \phi_c(F_i) + \max_{y \in X \setminus (S \cup x)} \sum_{y \in F_i} \phi_c(F_i) \leq \phi(c).$$

Therefore, after each of Red's moves we have that the potential is < 1, and so after the last move of Red's this is still true. If this is the last move of the game, then the potential at the end of the game is < 1, if not, then the last move of blue can still only decrease the potential.

Therefore, against any strategy of Red, this gives a strategy for blue that ensures that, at the end of the game, the value of the potential function on the colouring obtained is < 1. However, given a complete colouring $c : X \to \{\text{red,blue}\}$, it is clear that $\phi(c)$ is equal to the number of winning sets which are totally coloured red, and since this must be an integer, if it is < 1 it must be 0.

Therefore Red cannot have a winning strategy (since it would lose against this strategy of Blue's), and so by Theorem 13.3 both players have a drawing strategy, and hence the game is a draw. □

Let us consider as an application of these ideas the *Ramsey Game*. This is played with the board being the edge set of a complete graph K_n and the winning sets being the edge sets of any complete subgraph K_k. So there are $\binom{n}{k}$ winning sets, each of size $\binom{k}{2}$.

If $n \geq R(k,k)$, then every 2-colouring of K_n contains a monochromatic K_k, and so neither player can have a drawing strategy, and so the game is Red-win.

Conversely, if

$$\sum_{F \in \mathcal{F}} 2^{-|F|} = \binom{n}{k} 2^{-\binom{k}{2}} < 1/2$$

then we have by Theorem 13.5 that both players have a drawing strategy. We note that this gives an slightly different proof that if the above condition is satisfied then we can algorithmically find a colouring of K_n with no monochromatic K_k.

Since both players have a drawing strategy, if they both play using that drawing strategy the resulting colouring will contain no monochromatic K_k. However the proof above gives an explicit description of such a strategy for both players, each move of which can be computed in algorithmic time. In fact, this shows slightly more than the previous section, that we can construct such a colouring which is *balanced*, that is, an equal number of edges are coloured red and blue.

As another example consider the following game, a natural extension of Tic-Tac-Toe, called the n-in-a-row game. The board is $\mathbb{Z}^2$, and the winning lines are any consecutive line of n points in a row, either horizontally, vertically or diagonally.

For very small n, it is not too hard to find a winning strategy as Red, however, using the Theorem 13.5 we can show that, for large enough n, the game is a draw.

Theorem 13.6. *For $n \geq 40$ the n-in-a-row game is a draw.*

Proof. Since the number of winning sets is infinite, we can't apply the Erdős-Selfridge theorem directly. However, suppose we split $\mathbb{Z}^2$ into squares, and on each of the squares pretend we're playing a smaller n'-in-a-row game inside there. That is, whenever Red plays in a square, our strategy will be the respond inside the same square.

Suppose we're playing in an $m \times m$ square. The number of winning sets is at most $4m^2$, and each winning set has n' points in it. So by Theorem 13.5, as long as

$$4m^2 2^{-n'} < 1/2,$$

then Blue can stop Red from forming a line of line of length n' inside this $m \times m$ square. Note that we can take $m >> n'$ and still have this hold. However, if we split $\mathbb{Z}^2$ into $m \times m$ squares with $m \geq n$ then it is not to hard to check that any winning line meets at most 3 different squares, and so, if Blue has a strategy to stop Red from forming a line of length $n/3$ in any $m \times m$ square, then he can stop Red from forming a line of length n in the entire plane, since any such line would contain an interval of length $\geq n/3$ in some square.

It just remains to pick a suitable n and m. If we take, for simplicity's sake, $n = m$, then we need to pick n such that

$$n^2 2^{-\lceil n/3 \rceil} < 1/2,$$

and it is a simple check that $n = 40$ will do. □

We note that no attempt has been made the optimise the number 40 here, with the same method it can be made lower, but there is a much simpler method, called a pairing strategy,

with which one can show that in fact the 8-in-a-row game is a draw. Conversely it is known that 4-in-a-row is Red-win, the cases for $n = 5, 6, 7$ are still unknown.

13.4 Weak Games

Suppose we wanted to prove a converse to Theorem 13.5, to say that if the initial value of the potential function was large, that Red has a strategy to keep it large, and thus to win at the end.

After a bit of thought, one sees that this isn't too likely. The problem is that, for Red to win, he needs not only to guarantee that he forms a winning set, but also stop Blue from claiming one first. This seems like too complicated an aim to express with the maximisation of a single quantity. In fact, in general, this fact often makes winning strategies for strong positional games quite difficult to analyze. For this reason sometimes a simpler game is considered, called a *weak positional game*, or *Maker-Breaker game*.

Definition. Given a strong positional game $(X, \mathcal{F})$ the corresponding *weak positional game*, or *Maker-Breaker game*, $MB(X, \mathcal{F})$ is played as follows. The two players, Maker and Breaker, take turns claiming points of the board X, with Maker going first. We will still sometimes think of Maker as colouring his points red, and Breaker blue. Maker wins if at some point in the game he can claim all the points in some winning set $F \in \mathcal{F}$, and Breaker wins otherwise. If Maker has a winning strategy we call the game *Maker-win*, and similarly if Breaker has a winning strategy.

Similar to the proof of Lemma 13.2, every Maker-Breaker game is either Maker-win or Breaker-win. We note that if $MB(X, \mathcal{F})$ is Breaker-win, then the second player in $(X, \mathcal{F})$ has a drawing strategy and so the game is a draw. Similarly if $(X, \mathcal{F})$ is Red-win, then the first player has a winning strategy, and so $MB(X, \mathcal{F})$ is Maker-win. However neither of the converses are true in general.

Broadly speaking the Maker-Breaker game is easier for Maker than for Red, since he does not have to bother with stopping Blue from winning, just with forming a winning set himself. Our aim is to prove a converse of the Erdős-Selfridge theorem to find a sufficient condition for a game to be Maker win. Let us restrict our attention to the case where all winning sets have the same size n.

Suppose we wished to find some converse to the Erdős-Selfridge condition, which says that if

$$|F| < 2^{n-1}$$

then Breaker has a winning strategy, a natural thing to hope for is some lower bound on the number of winning sets which guarantees a Maker win. However, a little thought shows this is not possible without some extra conditions.

For example, if there are many winning sets, but they are all disjoint, then Breaker has a winning strategy by simply playing in the same set that Maker does. Hence, we will need some control on $|X|$, the size of the ground set.

Similarly, it could be that the ground set is small, but the winning sets all contain the same two points x and y. Then Breaker has a strategy to claim at least one of x and y, at which

point Maker cannot when. Therefore, we will need some sort of control on

$$\Delta_2(\mathcal{F}) = \max_{x,y \in X} |\{F \in \mathcal{F} : x, y \in F\}|,$$

that is, the maximum number of winnings sets that any pair is in. As the next theorem shows however, these two considerations are then sufficient.

Theorem 13.7. *Suppose $MB(X, \mathcal{F})$ is a weak positional game, with $\mathcal{F}$ n-uniform. If*

$$|\mathcal{F}| > 2^{n-3}|X|\Delta_2(\mathcal{F})$$

then $MB(X, \mathcal{F})$ is Maker-win.

Proof. The proof follows the same idea as that of Theorem 13.5. We want to define a potential function, on the set of partial colourings, which will start large, and which we will be able to prevent from decreasing too much over any two turns of the game.

Let us try the same function as before, that is, for each $F \in \mathcal{F}$, $\phi_c(F)$ can be thought of as the probability that a randomly chosen extension of c will make F monochromatically red, and then $\phi(c) = \sum_F \phi_c(F)$ is the expected number of monochromatically Red sets in a random extension of c.

So, a natural strategy for Maker would be, at each stage, to pick the point x which will maximise the new value of the potential function. So, let us estimate what can happen in a pair of moves. Suppose we are at some stage of the game and have some partial colouring c on a subset $S \subset X$.

What happens to the potential when Maker colours a point $x \in X \setminus S$ red, changing the partial colouring c to c'? Well, for every $x \in F_i \in \mathcal{F}$, colouring x red doubles the contribution of F_i to ϕ and so the total change in potential is

$$\sum_{x \in F_i} \phi_c(F_i).$$

Now what happens if Breaker colours a point $y \in X \setminus (S \cup \{x\})$ blue, changing the colouring to c''? As before, for every $y \in F_i \in \mathcal{F}$, colouring y blue removes the contribution of F_i to ϕ and so the total change in potential is

$$-\sum_{y \in F_i} \phi_{c'}(F_i).$$

For most of the F_i, ϕ will not have changed between c and c', however for some of the F_i we will have doubled the contribution of F_i to ϕ by colouring x red, before Breaker removed the contribution of F_i by colouring y blue. This will be for precisely the set of F_i such that both x and $y \in F_i$.

Hence the total change in potential is

$$\begin{aligned}-\sum_{y\in F_i}\phi_{c'}(F_i) &= -\sum_{y\in F_i, x\notin F_i}\phi_c(F_i) - \sum_{y\in F_i, x\in F_i}\phi_{c'}(F_i)\\ &= -\sum_{y\in F_i, x\notin F_i}\phi_c(F_i) - \sum_{y\in F_i, x\in F_i}2\phi_c(F_i)\\ &= -\sum_{y\in F_i}\phi_c(F_i) - \sum_{y\in F_i, x\in F_i}\phi_c(F_i).\end{aligned}$$

Therefore, if Maker chooses x to maximise the first quantity, the total change in potential from c to c'' will be

$$\begin{aligned}\phi(c'') &= \phi(c') - \sum_{y\in F_i}\phi_{c'}(F_i) = \phi(c) + \sum_{x\in F_i}\phi_c(F_i) - \sum_{y\in F_i}\phi_{c'}(F_i)\\ &= \phi(c) + \max_{x\in X\setminus S}\sum_{x\in F_i}\phi_c(F_i) - \sum_{y\in F_i}\phi_c(F_i) - \sum_{y\in F_i, x\in F_i}\phi_c(F_i)\\ &\geq \phi(c) + \big(\max_{x\in X\setminus S}\sum_{x\in F_i}\phi_c(F_i) - \max_{y\in X\setminus(S\cup\{x\})}\sum_{y\in F_i}\phi_c(F_i)\big) - \sum_{y\in F_i, x\in F_i}\phi_c(F_i)\\ &\geq \phi(c) - \sum_{y\in F_i, x\in F_i}\phi_c(F_i) \geq \phi(c) - \frac{1}{4}\Delta_2.\end{aligned}$$

Where the last line follows since there are at most Δ_2 such F_i containing both x and y and, since x and y were uncoloured by c by assumption, the potential of all these sets is at most $\frac{1}{4}$.

We note that the potential at the start is equal to

$$2^{-n}|\mathcal{F}| > \frac{1}{8}|X|\Delta_2(\mathcal{F})$$

and in every pair of moves the potential decreases by at most $\frac{1}{4}\Delta_2$. So at the end of the game we have some colouring c, and there have been $|X|/2$ pairs of moves, and so the potential is still

$$\phi(c) > \frac{1}{8}|X|\Delta_2(\mathcal{F}) - \frac{|X|}{2}\frac{1}{4}\Delta_2 > 0.$$

However, as before, the potential of a complete colouring counts the number of monochromatically red winning sets, and hence, at the end of the game, Maker must have completely claimed a winning sets. □

For example, applying this to the Ramsey game gives a criterion for Maker win, however it is asymptotically worse than the bound which follows from Ramsey theoretical arguments. As a more useful example, one can use this theorem to show that the n-in-a-row game is a Maker win if we allow winning lines of arbitrary rational slope.

13.5 The Neighbourhood Conjecture

The Erdős-Selfridge theorem is quite a broad global condition on a game, that there are not 'too many' winning sets, which guarantees that the game is a draw. There is quite a natural, albeit

weak, local condition under which we can conclude the same, sometimes known as a draw by pairing strategy.

For example, suppose we play a positional game on a set X such that the winning sets are a collection of disjoint pairs, then Blue can force a draw by, whenever Red claims a point in a winning set, claiming the second point from that pair. More generally, if Blue can choose a disjoint collection of pairs $\{\{x_i, y_i\} : i \in [m]\}$ from X, such that every transversal of this collection contains a point from every $F \in \mathcal{F}$, then $(X, \mathcal{F})$ is a draw.

Indeed, if whenever Red claims x_i or y_i Blue claims the remaining point then, at the end of the game, Blue has claimed some transversal of the set of pairs and so, by assumption, he has claimed a point in every winning set. The following Lemma gives a natural condition arising from Hall's theorem when we can find such a collection of pairs.

Lemma 13.8. *Let $(X, \mathcal{F})$ be a positional game. Suppose that for every $\mathcal{G} \subset \mathcal{F}$*

$$\left| \bigcup_{G \in \mathcal{G}} G \right| \geq 2|\mathcal{G}|$$

(That is, the total number of points in X contained in some member of $\mathcal{G}$ is at least twice as large as the number of sets in $\mathcal{G}$). Then $(X, \mathcal{F})$ is a draw.

Proof. Let us a form a bipartite graph G on vertex sets

$$A = X \text{ and } B = \mathcal{F} \times \{0, 1\},$$

where we join every point $x \in A$ to each $(F, i) \in B$ such that $x \in F$. We first note that a matching from B to A in G gives us a disjoint collection of pairs as above. Indeed, if we let $\{x_F, y_F\}$ be the two vertices joined to $(F, 1)$ and $(F, 0)$ in such a matching, we have that the collection of pairs

$$\{\{x_F, y_F\} : F \in \mathcal{F}\}$$

is disjoint, and every transversal of this collection contains a point from each $F \in \mathcal{F}$.

However it is a simple check that the graph satisfies Hall's condition. Indeed, suppose we have some subset $C \subset B$ such that the size of the neighbourhood of C is smaller than the size of C.

Let us choose the smallest subset $\mathcal{G} \subset \mathcal{F}$ such that

$$C \subset \mathcal{G} \times \{0, 1\},$$

and note that $2|\mathcal{G}| \geq |C|$. Since for every $G \in \mathcal{G}$ we have at least one of $(G, 0), (G, 1) \in C$, we have that

$$|N(C)| \geq \left| \bigcup_{G \in \mathcal{G}} G \right| \geq 2|\mathcal{G}| \geq |C|,$$

and hence a matching from B to A exists by Hall's theorem. □

Corollary 13.9. *Let $(X, \mathcal{F})$ be a positional game in which every winning set has size $\geq n$. If every point $x \in X$ is in at most $n/2$ winning sets, then the game is a draw.*

Proof. We note that for any $\mathcal{G} \subset \mathcal{F}$, by double counting,

$$\left| \bigcup_{G \in \mathcal{G}} G \right| \geq |\mathcal{G}| n \frac{2}{n} = 2|\mathcal{G}|.$$

Hence by Lemma 13.8 $(X, \mathcal{F})$ is a draw. □

So, Corollary 13.9 tells us that, if each point isn't in 'too many' winning sets, then the game is a draw, where 'too many' is linear in n. However, there is a sensible heuristic reason why we might think Corollary 13.9 can be improved.

If we consider $(X, \mathcal{F})$ as a hypergraph, then we know from the previous section that in order for $(X, \mathcal{F})$ to be a draw there must exist a 2-colouring of $\mathcal{F}$ with no monochromatic edge, since this is the gamestate at the end of any drawn game.

Using the probabilistic method we were able to show that such a colouring must exist when $\mathcal{F}$ satisfied some global condition, and by de-randomizing that argument we were able to show that, if $\mathcal{F}$ satisfied some slightly weaker condition, the game was in fact a draw.

In the case of a local condition like above, the Lovász Local Lemma gives us a very natural condition of $\mathcal{F}$ that ensures the existence of such a 2-colouring.

Lemma 13.10. *Suppose $H = (X, \mathcal{F})$ is an n-regular hypergraph in which every vertex is in at most $2^{n-2}/n$ edges, then there exits a 2-colouring of X with no monochromatic edge.*

Proof. This is a simple application of the Local Lemma, let us consider a 2-colouring of X picked uniformly at random. For each edge $F \in \mathcal{F}$ we let X_F be the event that F is monochromatic.

We see that $\mathbb{P}(X_F) = 2^{-n}$ for each F, and also that each event is mutually independent of all but the events $X_{F'}$ where F and F' intersect. Since each vertex is in $< 2^{n-2}/n$ edges, each edge intersects $< 2^{n-2}$ others, and there exists a dependency digraph of outdegree $< 2^{n-2}$.

Therefore, since

$$e 2^{n-2} 2^{-n} = \frac{1}{4} e < 1$$

we have by Corollary 9.3 that there exists a 2-colouring satisfying the conditions of the lemma. □

So, perhaps we might expect that a condition closer to Lemma 13.10 should imply that a game is a draw, perhaps by 'de-randomizing' in some sense the proof of the Local Lemma. That is it would be useful to have an algorithm that, under the conditions where the Local Lemma guarantees the existence of a point in a probability space where certain events holds, actually finds such a point.

Recently some major breakthroughs have been made in this area, in particular Moser and Tardos (2009) have developed an algorithm that works in almost all known applications of the Local Lemma, which we shall look at in the next section. Unfortunately, it still does not seem clear how one can use this to find a drawing strategy, and the following is still an open conjecture, of Beck.

Suppose we define $f(n)$ to be the smallest number such that the following is true: Every positional game $(X, \mathcal{F})$ in which every winning set has size n and each point in is at most $f(n)$ winning sets is a draw. Corollary 13.9 tell us that $f(n) \geq n/2$.

Conjecture 13.11 (The Neighbourhood Conjecture)**.**

$$f(n) = \frac{2^{n-2}}{n}.$$

Rather amazingly the best current bounds are that $n/2 \leq f(n) \leq 2^{n-1}/n$, that is, we don't even know if this function grows exponentially fast.

14 The Algorithmic Local Lemma

Suppose we have a probability space Ω which comes with an underlying set of mutually independent random variable $Z_1, \ldots, Z_n$. We are considering some set of events $(A_i : i \in I)$, as in the local lemma, where each A_i is determined by some subset $\text{vbl}(A_i) \subset \{Z_1, \ldots, Z_n\}$ of the Z_j.

In this case, a very natural dependency graph to take is that graph where $i \sim j$ if and only if $\text{vbl}(A_i) \cap \text{vbl}(A_j) \neq \emptyset$, when the events A_i and A_j depend on a common variable. Let use denote by D_i the neighbourhood of i in this graph.

Given an *assignment* of values to the variable Z_j, that is, some $\gamma = (\gamma_1, \ldots, \gamma_n)$ where $\gamma_j \in \text{range}(Z_j)$ for each j, we say that A_i is violated by this assignment if A_i is true under this assignment.

Our aim will be, given that the events $(A_i,\ i \in I)$ satisfy the conditions of the local lemma, to algorithmically find an assignment of values such that no event A_i is violated. The algorithm can be stated extremely simply:

- Pick a random assignment for each Z_j independently
- while there exists a violated A_i
  - Pick a violated A_i (according to some deterministic rule)
  - Re-sample the assignments for each $Z_j \in \text{vbl}(A_i)$
- return the values of the Z_j

Where we're sampling the Z_js according to their distribution. It is clear that if this algorithm terminates, then we have found our desired requirement. We will show that the expected time until it terminates is 'small'.

Theorem 14.1. *Let $Z_1, \ldots, Z_n$ and $(A_i : i \in I)$ be as above. If there exists real numbers $0 < x_i < 1$ such that*

$$\mathbb{P}(A_i) \leq x_i \prod_{j \in D_i} (1 - x_j)$$

for each i then the algorithm finds an assignment of values to the Z_j such that no event A_i is violated in expected time at most

$$\sum_{i \in I} \frac{x_i}{1 - x_i}$$

Note that, as long as the numbers x_i can be bounded away from 1, the runtime is essentially linear in $|I|$, the number of events. The bound will come from showing that we expect to re-sample each $\text{vbl}(A_i)$ at most $\frac{x_i}{1-x_i}$ many times.

Definition. The *log* of the algorithm is a sequence $L = (L(1), L(2), L(3) \ldots)$ where $L(t)$ is the event A_i such that $\text{vbl}(A_i)$ was re-sample in the tth step of the algorithm. We note that the log may be an infinite sequence.

A *witness tree* is a rooted tree T whose vertices are labelled with events A_i, where for a vertex $v \in V(T)$ we will denote by $[v]$ the i such that v is labelled with A_i, such that if u is a child of v, then $[u] \in D_{[v]} \cup \{[v]\}$. We call T *proper* if at each vertex the set of labels on its children is distinct.

Given a log L we can define a witness tree $T(t)$ for each step of the algorithm recursively. We first label the root with the event $L(t)$. Then, for each $i = t-1, t-2, \ldots$ we consider the event $L(i) = A_k$. If there exists a vertex v in the tree such that $[v] \in D_k \cup \{k\}$ then we pick one furthest from the root (breaking ties arbitrarily) and we add a leaf to the tree behind v labelled A_k. If no such vertex exists then we go on to the next $L(i-1)$.

The idea behind this process is to build a possible sequence of re-samplings that could have led to the log file $(L(1), \ldots, L(t))$. We say a witness tree T *occurs* in L if there is some t such that $T = T(t)$.

Lemma 14.2. *Let T be a witness tree and L the log file of a random execution of the algorithm.*

- *If T occurs in L then T is proper;*
- $\mathbb{P}(T \text{ occurs in } L) \leq \prod_{v \in T} \mathbb{P}(A_{[v]})$.

Proof. For the first part, let us denote for $v \in V(T)$ the *depth* of v in the tree by $d(v)$ and let us write $t(v)$ for the step in the log L at which v was attached to the tree.

We note that if $t(u) < t(v)$ and $\mathrm{vbl}(A_{[u]}) \cap \mathrm{vbl}(A_{[u]}) \neq \emptyset$, then $d(u) > d(v)$, since we must have attached u to either the vertex v, or one with strictly smaller depth. Hence, not only is T proper, but the set of vertices at a specific depth do not share variables.

For the second, we argue via a coupling argument. Given T we can define the *evaluation* of T to be the following process. Starting at the leaves we consider the vertices of T in reverse breadth-first search order and we re-sample in turn the variables in $\mathrm{vbl}(A_{[v]})$ for each vertex. We say the evaluation *succeeds* if at each step the event $A_{[v]}$ was violated by the re-sampling of $\mathrm{vbl}(A_{[v]})$. It is clear that

$$\mathbb{P}(\text{The evaluation succeeds}) = \prod_{v \in V(T)} \mathbb{P}(A_{[v]}).$$

So, we would like to show that it is less likely that T occurs in L then that the evaluation succeeds. To do so, let us imagine that we specified in advance, for each variable Z_j, an infinite sequence of independent random variables distributed as Z_j and, for both the evaluation and the algorithm, whenever we want to re-sample Z_j we do so by using the next random variable in the sequence.

This guarantees that, the algorithm and the evaluation will both use the same assignment for Z_j if it has been re-sampled the same number of times in both processes. We will show that if T occurs in the log, then the evaluation of T will succeed.

So suppose that T occurs in the log as $L(t)$. For every $i \leq t$ such that $L(i)$ labelled a vertex of the tree v we re-sampled $A_{[v]}$ because at this stage in the algorithm it was violated. Given

a $Z_j \in \text{vbl}(A_{[v]})$ how many times has it been sampled prior to this step? Well, it was sampled initially at the start of the algorimth, and then once for every $i' < i$ such that $Z_j \in \text{vbl}(L(i'))$. However, for each such i' there is a vertex of the tree labelled by $L(i')$ (since in particular it shares a variable with $A_{[v]}$), and by our previous observation these all appear at depth $> d(v)$ in the tree. Hence, the number of times Z_j was sampled prior to this was

$$1 + |\{u : d(u) > d(v) \text{ and } Z_j \in \text{vbl}(A_{[u]})\}| := 1 + n_{j,v}.$$

Furthermore, since $A_{[v]}$ was violated at this point in the algorithm (since we re-sampled $L(i) = A_{[v]}$) it follows that if we take the $(1 + n_{j,v})$th assignments for each $Z_j \in \text{vbl}(A_{[v]})$ then $A_{[v]}$ is violated.

Conversely, consider the point in the evaluation at which we are looking at a vertex v. For any $Z_j \in \text{vbl}(A_{[v]})$, we know from before there are no vertices at the same heigh as v which sample Z_j. Hence, since we're following the breadth-first search order, the number of times that Z_j has been sampled prior to this point in the evaluation is equal to the number of vertices u such that $d(u) > d(v)$ and $Z_j \in \text{vbl}(A_{[u]})$, that is, $n_{j,v}$.

Hence, when we re-sample each the $Z_j \in \text{vbl}(A_{[v]})$ the assignment we use is the $(1 + n_{j,v})$th assignment for each Z_j. However, we know that this assignment violates $A_{[v]}$.

Hence, if T occurs in the log, then the evaluation will succeed. It follows that

$$\mathbb{P}(T \text{ occurs in } L) \leq \mathbb{P}(\text{The evaluation succeeds}) = \prod_{v \in V(T)} \mathbb{P}(A_{[v]}).$$

□

For an event A_i let us define N_i to be the number of times that $\text{vbl}(A_i)$ is re-sampled in the algorithm. Note, given the log L, N_i is precisely the number of times that the witness tree $T(t)$ is labelled A_i. Hence, if we count over *all* proper witness trees whose root is labelled A_i we see that

$$\begin{aligned}
\mathbb{E}(N_i) &= \sum_{T \,:\, \text{root}=A_i} \mathbb{P}(T \text{ occurs in L}) \\
&\leq \sum_{T \,:\, \text{root}=A_i} \prod_{v \in V(T)} \mathbb{P}(A_{[v]}) \\
&\leq \sum_{T \,:\, \text{root}=A_i} \prod_{v \in V(T)} x_{[v]} \prod_{j \in D_{[v]}} x_j
\end{aligned}$$

by the previous Lemma, and then the local lemma.

We'd like to bound this sum, and we do so in quite an ingenious way. We will define a random process, called a *Galton-Watson branching process*, which produces a proper witness tree in some random manner and show that the terms

$$\prod_{v \in V(T)} x_{[v]} \prod_{j \in D_{[v]}} x_j$$

are (up to a fixed constant factor) the probability that we produce the tree T in this fashion. Hence we can bound the sum of these terms over all such trees by 1, since they represent probabilities of disjoint events.

So, let us define the Galton-Watson branching process. In this process we build a labelled tree by first picking a root with a label A_i and then, for each $j \in D_i \cup \{i\}$ we add with probability x_j a child of the root with label A_j. We then do the same for each child of the root according to the same rule, and so on. This process may die out eventually, or may produce an infinite tree.

Lemma 14.3. *Let T be a proper witness tree with root A_i. The probability that T is given by the above Galton-Watson branching process is*

$$p_T = \frac{1-x_i}{x_i} \prod_{v \in V(T)} x'_{[v]}$$

where $x'_i = x_i \prod_{j \in D_i}(1-x_j)$.

Proof. For a given $v \in V(T)$ labelled with A_i let us denote by W_v the subset of $D_i \cup \{i\}$ which *do not* occur as children of v in T. Then we have that

$$p_T = \frac{1}{x_i} \prod_{v \in V(T)} \left(x_{[v]} \prod_{w \in W_v} (1 - x_{[w]}) \right)$$

since p_T is the product over all vertices of the probability that that vertex appeared, and the set of non-children of that vertex didn't appear, with an extra factor of $1/x_i$ to account for the fact that the root always appears. To get rid of the dependence on W_v, for every v we add in a factor of $(1-x_{[u]})$ for each $u \in D_{[v]} \cup \{i\} \setminus W_v$, which is equivalent to adding a factor of $(1-x_{[v]})$ for every vertex v and hence

$$\frac{1}{x_i} \prod_{v \in V(T)} \left(x_{[v]} \prod_{w \in W_v} (1 - x_{[w]}) \right) = \frac{1-x_i}{x_i} \prod_{v \in V(T)} \left(\frac{x_{[v]}}{1-x_{[v]}} \prod_{w \in D_{[v]} \cup \{[v]\}} (1 - x_{[w]}) \right)$$

where again we had to deal with the root separately. However, we can then cancel some terms to see

$$\frac{1-x_i}{x_i} \prod_{v \in V(T)} \left(\frac{x_{[v]}}{1-x_{[v]}} \prod_{w \in D_{[v]} \cup \{[v]\}} (1 - x_{[w]}) \right) = \frac{1-x_i}{x_i} \prod_{v \in V(T)} \left(x_{[v]} \prod_{w \in D_{[v]}} (1 - x_{[w]}) \right)$$

which by our earlier definition is equal to

$$\frac{1-x_i}{x_i} \prod_{v \in V(T)} x'_{[v]}$$

□

So putting this all together we see that

$$\mathbb{E}(N_i) \leq \sum_{T \,:\, \text{root}=A_i} \prod_{v \in V(T)} x_{[v]} \prod_{j \in D_{[v]}} x_j = \sum_{T \,:\, \text{root}=A_i} \prod_{v \in V(T)} x'_{[v]}$$

which we can now bound by the previous lemmas as

$$\mathbb{E}(N_i) \leq \frac{x_i}{1-x_i} \sum_{T \,:\, \text{root}=A_i} p_T \leq \frac{x_i}{1-x_i}$$

where in the final line we've used the the trees T are distinct and so $\sum_T p_T \leq 1$.

This gives us a randomised algorithm that has a low expected running time, but it can be changed into a deterministic algorithm using the method of conditional expectations as in the previous section. Essentially the observation is that we bounded the run time by bounding the expected number of witness trees occurring in the log. If the probabilities x_i are bounded away from 0, this decreases exponentially quickly as the size of the tree increases, and so there is some N such that the expected number of witness trees of size $> N$ is less than $1/2$.

With a little work one can show that if there is a witness tree of size $> N$ there must be one whose size in in some finite interval $[N, (\Delta + 1)N]$ where Δ is the maximum degree of the dependency graph. At this point we can list all the possible witness trees in this range, and choose our values for the re-sampling of the variables Z_j so as the minimise the expected number of these trees which will appear in the log. Since by assumption this is $< 1/2$ at the start, we can choose assignments for all possible samplings so that this holds (noting that we never have to re-sample any variable more than N times). This gives us a list of assignment such that no witness trees of size $> N$ appear in the log, and hence the algorithm must terminate after $\leq N$ steps.