Random Graphs

Joshua Erde

Department of Mathematics,
TU Graz.

Contents

Preface

These notes were used to lecture a course at TU Graz for masters level students. These course notes were heavily based on a course given by Wojciech Samotij, which are themselves based on the books "Random graphs" by Svante Janson, Tomasz Łuczak and Andrzej Rucinski and "Introduction to random graphs" by Alan Frieze and Michał Karoński, as well as a lecture course of Michael Krivelevich. However any mistakes are my own.

1 Preliminaries

1.1 Graph Theory

We begin with a very quick recap of various graph theoretical notions and notations that should hopefully be familiar if you have attended an introductory graph theory course, mostly intended as a reference for when terms are introduced later in the course.

A *graph* is a pair (V, E) where V is a set of *vertices* and $E \subseteq V^{(2)}$ is a set of *edges*. Here and later we will write $X^{(k)}$ for the set of k-element subsets of a set X. For a graph G we will write $V(G)$ and $E(G)$ for its vertex set and edge set respectively and define $e(G) := |E(G)|$ and $v(G) := |V(G)|$. For ease of notation we will simply write xy for an edge $\{x, y\} \in E$.

Given a vertex $x \in V$ we will write $N_G(x)$ for the *neighbourhood* of x in G, that is $N_G(x) := \{y \in V : xy \in E\}$, and the *degree* of x is $d_G(x) := |N_G(x)|$. Given a pair of subsets $X, Y \subseteq V$ we will write $E_G(X, Y)$ for the set of edges with one endpoint in X and the other in Y and when $X = Y$ we will simplify this to $E_G(X) := E_G(X, X)$. We will also write $e_G(X, Y) := |E_G(X, Y)|$. Whenever the graph G is clear from the context we may omit the subscript G in the above notation.

Two graphs G and H are *isomorphic* if there exists a bijection $f : V(G) \to V(H)$ such that $xy \in E(G)$ if and only if $f(x)f(y) \in E(H)$. We say H is a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$, and write $H \subseteq G$. A subgraph is *spanning* if $V(H) = V(G)$. A subgraph is *induced* if $E(H) = E(G) \cap V(H)^{(2)}$. Given a subset $X \subseteq V(G)$ we write $G[X]$ for the induced subgraph of G with vertex set X.

Two vertices $x, y \in V(G)$ are *connected* if there exists a sequence of vertices

$$x = v_1, v_2, v_3, \ldots, v_k = y$$

such that $x_i x_{i+1} \in E(G)$ for each $i \in [k-1]$ (here and later we write $[k] := \{1, 2, \ldots, k\}$). Such a sequence is called a *walk*, or sometimes an *$x-y$-walk*, of *length* $k-1$. Being connected is an equivalence relation and the *components* of G are the subgraphs $G[X]$ where X is an equivalence class under the relation of being connected. We say G is *connected* if every pair of vertices are connected or equivalently if G itself is a component.

We write E_n for the *empty graph* on n vertices and K_n for the *complete* graph. A set of vertices $X \subseteq V(G)$ is *independent* if $G[X]$ is empty and a *clique* if $G[X]$ is complete.

A *path* of length k, denoted by P_k, has vertex set $\{v_1, \ldots, v_{k+1}\}$ and edge set $\{v_i v_{i+1} : i \in [k]\}$. A *cycle* of length k has vertex set $\{v_1, \ldots, v_k\}$ and edge set $\{v_i v_{i+1} : i \in [k-1]\} \cup \{v_k v_1\}$. A graph that doesn't contain any cycles is called a *forest* and a connected forest is a *tree*.

A *graph parameter* is some function whose range is the class of graphs, normally with domain in the real numbers, which is constant on classes of isomorphic graphs. For example $e(G)$ is a graph parameter. Let us briefly recall some important graph parameters.

The *minimum degree* $\delta(G)$ and *maximum degree* $\Delta(G)$ of a graph are given by

$$\delta(G) := \min_{v \in V} d_G(v) \text{ and } \Delta(G) := \max_{v \in V} d_G(v).$$

The *independence number* $\alpha(G)$ is the size of the largest independent set of vertices in G and the *clique number* $\omega(G)$ if the size of the largest set of vertices which is a clique in G.

We say a function $\chi : V(G) \to [k]$ is a *k-colouring* of G and a colouring is *proper* if $\chi(x) \neq \chi(y)$ for every $xy \in E(G)$. The *chromatic number* $\chi(G)$ is then the smallest k such that there exists a proper k-colouring of G. Finally the *girth* $g(G)$ is the length of a shortest cycle in G, if one exists, and we let $g(G) := \infty$ if G is a forest.

1.2 Probability Theory

This section is intended as a short introduction to the very basics of probability theory, covering only the basic facts about (discrete) 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 i.e

- $\emptyset \in \Sigma$;
- If $A \in \Sigma$ then $A^c \in \Sigma$;
- For all countable families of disjoint sets $(A_i : i \in \mathbb{N})$ in Σ, $\bigcup_{i\in\mathbb{N}} A_i \in \Sigma$,

and $\mathbb{P}$ is a measure on Σ with $\mathbb{P}(\Omega) = 1$ i.e

- $\mathbb{P}$ is non-negative;
- $\mathbb{P}(\emptyset) = 0$;
- For all countable families of disjoint sets $(A_i : i \in \mathbb{N})$ in Σ,

$$\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i).$$

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*.

The simplest example of a probability space is a *discrete probability spaces*, those where Ω is countable 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. For the most part in the course we will only have to consider discrete probability spaces, and in fact usually *finite* probability spaces, ones where Ω is finite.

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

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

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 countable family of events* $(A_i \colon i \in \mathbb{N})$ *in* Σ*,*

$$\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) \leq \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$$

Proof. For each $i \in \mathbb{N}$ let us define

$$B_i = A_i \setminus (\bigcup_{j=1}^{i-1} A_j).$$

Then $B_i \subseteq A_i$, and so $\mathbb{P}(B_i) \leq \mathbb{P}(A_i)$, and also $\bigcup_{i\in\mathbb{N}} B_i = \bigcup_{i\in\mathbb{N}} A_i$. Therefore, since the events $B_1, B_2, \ldots, B_n$ are disjoint, by the countable additivity of $\mathbb{P}$

$$\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \mathbb{P}\left(\bigcup_{i\in\mathbb{N}} B_i\right) = \sum_{i\in\mathbb{N}} \mathbb{P}(B_i) \leq \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$$

□

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

If we have a collection of events $(B_i \colon i \in \mathbb{N})$ which partition Ω then the *law of total probability* tells us that for any event $A \in \Sigma$

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Definition. A *random variable* on a probability space $(\Omega, \Sigma, \mathbb{P})$ is a $\mathbb{P}$-measurable function $X : \Omega \to E$ to some measurable space E. That is, E is a set together with a σ-algebra Σ_E on E such that for any measurable $A \in \Sigma_E$

$$\{\omega \in \Omega \colon X(\omega) \in A\} \in \Sigma.$$

For the most part we will be considering *discrete random variables*, that is random variables where the range of X is countable. Note that, in particular, this will be true whenever $(\Omega, \Sigma, \mathbb{P})$ is a discrete probability space.

Given a measurable set $A \subseteq E$ 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 $\mathbb{P}(X \in A)$. Normally we will want to think about random variables not as functions from some probability space to a measurable space, but just in terms of the *distributions* on the measurable space they determine.

What do we mean by a distribution? Well, for every measurable set $A \in \Sigma_E$ we can assign it a measure $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$. It is not hard to check that the triple $(E, \Sigma_E, \mathbb{P}_X)$ is then a probability space. So, in fact, this is just another word for a notion we already have, that of a probability measure on Σ_E, and indeed for every probability space $(\Omega, \Sigma, \mathbb{P})$ the function $\mathrm{id} : \Omega \to \Omega$ is a random variable whose distribution agrees with the measure $\mathbb{P}$.

During the course we will normally just introduce random variables by specifying their distributions, rather than making reference to any specific probability space. Given two random variables X and Y with the same range we write $X \sim Y$ if X and Y have the same distribution, that is, if $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ for every $A \in \Sigma_E$. Generally we will treat two random variables with the same distribution as the same random variable.

Often we will have two (or more) random variables X and Y defined on the same probability space, in which case we can specify their relationship to each other by giving their *joint distribution*, the function $(\mathbb{P}_X, \mathbb{P}_Y) : \Sigma_E \to [0,1]^2$ which maps a set A to $\big(\mathbb{P}(X \in A), \mathbb{P}(Y \in A)\big)$. On the other hand, given two arbitrary random variables X and Y, a useful tool for comparing their distributions can be to find or construct a *coupling*, a pair of random variables (Z_1, Z_2) defined on the same probability space whose *marginal distributions* Z_1 and Z_2 are distributed as X and Y, respectively.

A particularly common case is a *real random variable* when $E = \mathbb{R}$ and Σ_E is the borel σ-algebra on $\mathbb{R}$. In this case we can use the algebraic structure on $\mathbb{R}$ to define operations on random variables.

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

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

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

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

The set of real 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 real 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. To make the proof more instructive, let us consider the case where X and Y are random variables on a discrete probability space. In the general case the proof is essentially the same, with sums replaced by integration.

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

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

□

Let's briefly discuss some important random variables and their properties. Perhaps the simplest random variable is a *Bernoulli random variable* $\mathrm{Ber}(p)$ which takes the value 1 with probability p and 0 with probability $1 - p$. The expectation of $\mathrm{Ber}(p)$ is clearly p. A *binomial random variable* $\mathrm{Bin}(n, p)$ is the sum of n mutually independent $\mathrm{Ber}(p)$ random variables. One

normally thinks of this as measuring the number of success in n independent 'trials' with success probability p. For for any integer $0 \leq k \leq n$ we can calculate that

$$\mathbb{P}(\mathrm{Bin}(n,p) = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

By the linearity of expectation its expectation is np, and in fact it can be shown that np is also the *mode* and *median*.

The *Poisson distribution* $\mathrm{Po}(\lambda)$ has a distribution given by

$$\mathbb{P}(\mathrm{Po}(\lambda) = k) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ for all integers } k \geq 0.$$

One can think of the Poisson distribution as counting the number of occurences in a fixed time interval of some given event, if these events occur with a known constant rate and independently of the time since the last event. It can be shown that the expectation of $\mathrm{Po}(\lambda) = \lambda$.

1.3 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. Finally Stirlings formula, $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, gives a more precise asymptotic formula.

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

$$1 - p \geq e^{-\frac{p}{1-p}}.$$

Noting in particular that the latter is $\geq e^{-2p}$ if $0 \leq p \leq \frac{1}{2}$ and also $\geq e^{-p+O(p^2)}$ if $p \to 0$.

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 $C > 0$;
- $f = \Theta(g)$ if $f = O(g)$ and $f = \Omega(g)$;
- $f \approx g$ if $f = (1 + o(1))g$.

1.4 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 satisfying certain conditions 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.

So, for example, in order to show that there exists a graph satisfying a property $\mathcal{P}$, one strategy would be to find a graph-valued random variable G such that we can show that $\mathcal{P}(G \text{ satisfies } \mathcal{P}) > 0$. Initially these random variables G were considered as tools to prove the existence of certain graphs, but eventually it became clear, perhaps because these tools were so useful, that for certain natural distributions these 'random graphs' were interesting objects to study in their own right.

For example one of the simplest graph valued random variables would be to take G_U to be uniformly distributed over all graphs with vertex set $[n]$. Then, or any property $\mathcal{P}$ we have that

$$\mathbb{P}(G_U \text{ satisfies } \mathcal{P}) = \frac{|\{H \colon V(H) = [n], H \text{ satisfies } \mathcal{P}\}|}{2^{\binom{n}{2}}}$$

and so calculating $\mathbb{P}(G_U \text{ satisfies } \mathcal{P})$ is equivalent to counting the number of graphs which satisfy $\mathcal{P}$. So, it seems that if anything I've made my job harder. Before I needed to find or construct a graph with property $\mathcal{P}$, but now I need to be even cleverer and **count** graphs with certain properties.

However, there is another way of thinking about G_U. Suppose we take a family of $\binom{n}{2}$ many mutually independent $\mathrm{Ber}(1/2)$ random variables, $\left(B_e \colon e \in [n]^{(2)}\right)$ and let us define a random variable $G = (V, E)$ where $V = [n]$ and $E = \left\{e \in [n]^{(2)} \colon B_e = 1\right\}$. Clearly G is graph-valued, and perhaps surprisingly, as we will show later, $G \sim G_U$.

In this way switching to this probabilistic viewpoint, rather than an enumerative viewpoint, can be beneficial, since the product structure of this probability space allows you to argue at a local, rather than global level. Furthermore, once we start working with probability we can apply a whole host of tools from probability theory, many of which don't have obvious combinatorial counterparts.

Let us demonstrate this with an example. There is a famous theorem of Ramsey which asserts some relation between the independence and clique number of a graph, namely that for large enough graphs, these parameters cannot be simultaneously small.

Theorem 1.4 (Ramsey's Theorem)**.** *For every integer $k \geq 3$ there is some integer $n = n(k)$ such that if $v(G) \geq n$, then $\max\{\alpha(G), \omega(G)\} \geq k$.*

Since such an integer $n(k)$ exists, there is some smallest such integer

$$R(k) = \min\{n \colon \max\{\alpha(G), \omega(G)\} \geq k \text{ for every } n\text{-vertex graph } G\}.$$

Determining the value of $R(k)$ exactly seems to be a hopelessly difficult task, for example the value of $R(5)$ is not even known. Finding good bounds for the asymptotic behaviour of $R(k)$ is an important problem in graph theory. In terms of upper bounds Erdős and Szekeres showed that $R(k) \leq 4^k$, and it was a major open problem, which was only recently solved, as to whether this can be impoved to $(4-\varepsilon)^k$ for some $\varepsilon > 0$, and the exact value of $\lim_{k\to\infty} \frac{\log R(k)}{k}$ is still unknown. Whilst is not too hard to construct a graph giving a polynomial lower bound for $R(k)$, there is still no constructive exponential lower bound. However, as an early use of the probabilistic method Erdős gave an exponential lower bound.

Theorem 1.5 (Erdős)**.** *For every integer $k \geq 3$ we have $R(k) \geq 2^{\frac{k}{2}-1}$.*

Proof. Let $n = 2^{\frac{k}{2}-1}$ and let G be the random variable we defined earlier in the section. We will consider the property $\mathcal{P}$ that $\alpha(G) < k$ and $\omega(G) < k$. If we can show that $\mathbb{P}(G \text{ satisfies } \mathcal{P}) > 0$, then we can conclude that there must be at least one graph H on n vertices such that $\alpha(H) < k$ and $\omega(H) < k$, and hence $R(k) > n$.

So, let us estimate the probability that G satisfies $\mathcal{P}$. For any set $X \in [n]^{(k)}$ let A_X be the event that $G[X]$ is empty and let B_X be the event that $G[X]$ is complete. We note that G does not satisfy $\mathcal{P}$ if and only if some event A_X or B_X occurs.

Now, for any fixed X we have that

$$\mathbb{P}(A_X) = \left(\frac{1}{2}\right)^{\binom{k}{2}}$$

since each edge $e \in X^{(2)}$ is in G independently with probability $1/2$. Similarly

$$\mathbb{P}(B_X) = \left(\frac{1}{2}\right)^{\binom{k}{2}}.$$

Hence by the union bound

$$\mathbb{P}(G \text{ doesn't satisfy } \mathcal{P}) = \mathbb{P}\left(\bigcup_{X\in[n]^{(k)}} (A_X \cup B_X)\right) \leq 2 \sum_{X\in[n]^{(k)}} \left(\frac{1}{2}\right)^{\binom{k}{2}} = 2\binom{n}{k}2^{-\binom{k}{2}}.$$

It follows that

$$\mathbb{P}(G \text{ satisfies } \mathcal{P}) = 1 - \mathbb{P}(G \text{ doesn't satisfy } \mathcal{P}) \geq 1 - 2\binom{n}{k}2^{-\binom{k}{2}}.$$

So, it remains to find the largest n such that $2\binom{n}{k}2^{-\binom{k}{2}} < 1$. Let us show that this is satisfied with $n = 2^{\frac{k}{2}-1}$.

Using the weak estimate $\binom{n}{k} \leq n^k$ we see that

$$2\binom{n}{k}2^{-\binom{k}{2}} \leq 2n^k 2^{-\frac{k(k-1)}{2}} = 2^{1+k\left(\frac{k}{2}-1\right)-\frac{k(k-1)}{2}} = 2^{1-\frac{k}{2}} < 1.$$

□

2 Random Graph Models

Broadly, by a random graph, we will mean some random variable G which is graph-valued. In particular, if write $\mathcal{G}_n$ for the set of graphs whose vertex set is $[n]$, for the most part we will consider random variables taking values in $\mathcal{G}_n$ for some n. Note that, since we can associate each graph $H \in \mathcal{G}_n$ with the subset $E(H) \subseteq [n]^{(2)}$ we have that

$$|\mathcal{G}_n| = 2^{\binom{n}{2}}.$$

The simplest model of a random graph we could take would then be a uniform random graph, that is G is a random variable which takes the value H with probability $\frac{1}{2^{\binom{n}{2}}}$ for each $H \in \mathcal{G}_n$.

There is another useful way to view this random variable, we start with an empty graph with vertex set $[n]$ and for each edge $e \in [n]^{(2)}$ we add it to the graph with probability $1/2$, independently of all other edges. Let X be the random variable defined as above. For any particular graph $H \in \mathcal{G}_n$ it is easy to calculate

$$\mathbb{P}(X = H) = \prod_{e \in E(H)} \mathbb{P}(e \in E(X)) \cdot \prod_{e \notin E(H)} \mathbb{P}(e \notin E(X)) = \frac{1}{2^{\binom{n}{2}}}.$$

Hence X has the same distribution as G above.

More generally, given some $p \in [0,1]$ we can consider the *binomial random graph* $G(n,p)$, which is the random graph obtained by adding each edge to the empty n-vertex graph with probability p, independently of each other edge (and so each edge is not added with probability $1-p$). As above, it is easy to calculate the distribution of this random variable: For any graph $H \in \mathcal{G}_n$

$$\mathbb{P}(G(n,p) = H) = \prod_{e \in E(H)} \mathbb{P}(e \in E(G(n,p))) \cdot \prod_{e \notin E(H)} \mathbb{P}(e \notin E(G(n,p))) = p^{e(H)}(1-p)^{\binom{n}{2}-e(H)}.$$

When $p > 1/2$ this weights our choice of random graph towards graphs with more edges, and when $p < 1/2$ this weights our choice of random graph towards those with fewer edges.

We will also consider the random graph $G(n,m)$, which is a uniformly chosen graph on $[n]$ with exactly m edges. That is, it is the uniform random variable taking values in $\mathcal{G}_{n,m} = \{G \in \mathcal{G}_n \colon e(G) = m\}$. There is of course potential for confusion in this notation, however since $m \in \mathbb{N}$ and $p \in (0,1)$, it should be clear from the context which model we are referring to.

Often we are not interested in fixed p or m, but rather allowing p and m to vary as a function of n, the size of the vertex set. More precisely, often we are interested in a property $\mathcal{P}$ of graphs, for example being connectedness, or two colourability, which can be thought of as subsets of $\mathcal{P}_n \subseteq \mathcal{G}_n$ (those graphs having this property). In this case, given a function $p(n)$ or $m(n)$, we are interested in the asymptotic behaviour of $\mathbb{P}(G(n,p) \in \mathcal{P}_n)$ or $\mathbb{P}(G(n,m) \in \mathcal{P}_n)$. In particular, does this sequence have a limit as $n \to \infty$ and what is this limit. One special case is when this limit exists and is equal one, in which case we say that $G(n,p)/G(n,m)$ satisfies $\mathcal{P}$ *almost surely* or *with high probability* (whp).

A natural question to ask is how are these models of random graphs related to each other, both between the two different models, but also within each model as we vary p and m. Let us

consider the latter question first, since it introduces both an interesting new model, and a new technique.

The model is that of a *random graph process*, these were first considered by Erdős and Renyi. The idea is to start with an empty graph with vertex set $[n]$ and then add new edges to it randomly, one by one, until we get to the complete graph.

More formally, let's consider a random variable σ which is uniformly distributed on the set of permutations of $[n]^{(2)}$, i.e., we may think of σ as a uniformly chosen random ordering of the edges of K_n. For each $m \in \left\{0, 1, \ldots, \binom{n}{2}\right\}$ let

$$G_m = \left([n], \{\sigma(i) \colon i \leq m\}\right),$$

that is, the graph whose edge set is the first m edges in the order σ. This defines a sequence of random variables $\left(G_1, \ldots, G_{\binom{n}{2}}\right)$, and in fact it's not hard to show that $G_m \sim G(n, m)$.

This is useful, as it allows us to compare $G(n, m)$ and $G(n, m')$ for different $m \neq m'$. The fancy name for this idea is called *coupling*, and it will be useful later. Given two random variables X and Y, perhaps just in terms of their distributions, we might want to compare various probabilities (for example $\mathbb{P}(X \geq a)$ with $\mathbb{P}(Y \leq b)$). A useful tool to do so is to find a third random variable $Z = (X', Y')$ such that $X' \sim X$ and $Y' \sim Y$, and we call Z a *coupling* of X and Y. The useful thing here is that X' and Y', unlike perhaps X and Y, have a joint distribution. One obvious example of a coupling is just to take the *independent coupling* $Z = (X', Y')$ where X' and Y' are independently distributed as X and Y respectively. However, by finding a better coupling we can sometimes infer things about the relationship between X and Y.

For example, if we have $X = G(n, m)$ and $Y = G(n, m')$, say with $m < m'$, then we see that $Z = (G_m, G_{m'})$ is a coupling for X and Y. Consider then some property of graphs $\mathcal{P}$ which is preserved under taking supergraphs, which we call an *increasing* property. Since $G_m \subseteq G_{m'}$, it follows that whenever G_m satisfies $\mathcal{P}$ then so does $G_{m'}$. Hence

$$\mathbb{P}(G(n, m) \text{ satisfies } \mathcal{P}) = \mathbb{P}(G_m \text{ satisfies } \mathcal{P}) \leq \mathbb{P}(G_{m'} \text{ satisfies } \mathcal{P}) = \mathbb{P}(G(n, m') \text{ satisfies } \mathcal{P}).$$

Similarly there is a natural random graph process for the binomial model. Suppose we have a family of mutually independent random variables $\left(U_e \colon e \in \binom{n}{2}\right)$ which are all uniformly distributed on $[0, 1]$. For each $p \in [0, 1]$ let

$$G_p = \left([n], \left\{e \in \binom{n}{2} \colon U_e \leq p\right\}\right),$$

Again, this defines a family of random variables $(G_p \colon p \in [0, 1])$, and again it's not hard to show that $G_p \sim G(n, p)$. As this before this gives us a natural simultaneous coupling of all $G(n, p)$, which can be used to show that

$$\mathbb{P}(G(n, p) \text{ satisfies } \mathcal{P}) \leq \mathbb{P}(G(n, p') \text{ satisfies } \mathcal{P}).$$

whenever $p \leq p'$ and $\mathcal{P}$ is an increasing property, and analogous statements hold, in the obvious way, for *decreasing* properties.

In this way, we can think of the structure of the random graphs $G(n, p)$ and $G(n, m)$ as *evolving* as the parameters p and m increase, starting with an empty graph and gradually growing to a complete graph. This is an idea we will come back to in the next chapter.

What can we say about how $G(n,p)$ is related to $G(n,m)$? In fact, as it will turn out, when the parameters p and m are chosen appropriately, the two models are closely enough related that for most applications you can work in whichever model is most convenient, and use some standard results to transfer results between them.

Indeed, one simple thing to note is that, if we condition on having exactly m edges, then $G(n,p)$ is equally likely to be any of the graphs in $\mathcal{G}(n,m)$ and hence

$$\big(G(n,p)|e(G(n,p))=m\big) \sim G(n,m).$$

(Briefly, given a random variable X on a probability space $(\Omega,\Sigma,\mathbb{P})$ and an event B in Σ then $\mathbb{P}(\cdot|B)$ defines a new probability measure on Σ and the function $(X|B):\Omega\to E$ on the probability space $(\Omega,\Sigma,\mathbb{P}(\cdot|B))$ given by $(X|B)(\omega)=X(w)$ is the random variable X *conditioned on* B.)

More generally, we should expect that $G(n,p)$ and $G(n,m)$ should behave similarly when we expect $G(n,p)$ to have approximately m many edges. More precisely, we know that $\mathbb{E}(e(G(n,p)))=p\binom{n}{2}$ and hence we should expect the two models to behave similarly when $p\approx\frac{m}{\binom{n}{2}}$.

We don't quite have the tools yet to show the full extent of this behaviour, but for now we will be able to show a particular case - if the probability that $G(n,p)$ satisfies some property $\mathcal{P}$ is small, then so is the property that $G(n,m)$ satisfies this property.

Theorem 2.1. *Let $\mathcal{P}$ be a graph property, $m(n)\to\infty$, $\binom{n}{2}-m\to\infty$ and $p=\frac{m}{\binom{n}{2}}$. Then for sufficiently large n*

$$\mathbb{P}(G(n,m)\in\mathcal{P}) \leq 3\sqrt{m}\cdot\mathbb{P}(G(n,p)\in\mathcal{P})$$

Proof. By conditioning on the number of edges in $G(n,p)$ and using the law of total probability, we see that

$$\begin{aligned}
\mathbb{P}(G(n,p)\in\mathcal{P}) &= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}(e(G(n,p))=k)\mathbb{P}(G(n,p)\in\mathcal{P}|e(G(n,p))=k) \\
&= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}(e(G(n,p))=k)\mathbb{P}(G(n,k)\in\mathcal{P}) \\
&\geq \mathbb{P}(e(G(n,p))=m)\mathbb{P}(G(n,m)\in\mathcal{P})
\end{aligned}$$

Since the number of edges in a random graph $G(n,p)$ is distributed as a binomial distribution with $\binom{n}{2}$ many trials and success probability p we can directly calculate

$$\mathbb{P}(e(G(n,p))=m) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}.$$

This is the probability that a binomial random variable takes it's mean value. It can be shown that this is in fact also the *mode* of the binomial random variable, the most likely value, and so immediately we can say that it is at least $\frac{1}{\binom{n}{2}}\approx\frac{2}{n^2}$. However, we can be a bit more precise and estimate how this expression behaves asymptotically using our estimates from Section 1.3.

Let use write $N = \binom{n}{2}$, so that $p = \frac{m}{N}$, and recall that by Stirling's formula $k! = (1 + o(1))\sqrt{2\pi k}\left(\frac{k}{e}\right)^k$ and hence

$$\begin{aligned}\binom{N}{m} &= \frac{N!}{m!(N-m)!} \\ &= (1+o(1))\frac{\sqrt{2\pi N}N^N e^m e^{N-m}}{\sqrt{2\pi m}\sqrt{2\pi(N-m)}m^m(N-m)^{N-m}e^N} \\ &= (1+o(1))\sqrt{\frac{N}{2\pi m(N-m)}}\frac{N^N}{m^m(N-m)^{N-m}}.\end{aligned}$$

Since $p = \frac{m}{N}$, we can also evaluate

$$p^m(1-p)^{N-m} = p^m(1-p)^{N-m} = \frac{m^m}{N^m}\left(1-\frac{m}{N}\right)^{N-m} = \frac{m^m}{N^m}\left(\frac{N-m}{N}\right)^{N-m} = \frac{m^m(N-m)^{N-m}}{N^N}$$

Hence, it follows that

$$\mathbb{P}(e(G(n,p)) = m) = \binom{N}{m}p^m(1-p)^{N-m} = \big(1+o(1)\big)\sqrt{\frac{N}{2\pi m(N-m)}} \geq \big(1+o(1)\big)\frac{1}{\sqrt{2\pi m}} \geq \frac{1}{3\sqrt{m}}.$$

Hence

$$\mathbb{P}(G(n,m) \in \mathcal{P}) \leq \frac{\mathbb{P}(G(n,p) \in \mathcal{P})}{\mathbb{P}(e(G(n,p)) = m)} \leq 3\sqrt{m} \cdot \mathbb{P}(G(n,p) \in \mathcal{P}).$$

□

In fact, when the graph properties we consider are reasonably well behaved, we can do much better than Theorem 2.1. Recall that an increasing property of graphs is a property which is closed under taking supergraphs. Similarly we can define a *decreasing* property of graphs as one which is closed under taking subgraphs. A graph property which is either increasing or decreasing is called *monotone*.

Theorem 2.2. *Let $\mathcal{P}$ be a montone graph property and let m, n be integers. If we let $p = \frac{m}{\binom{n}{2}}$, then*

$$\mathbb{P}(G(n,m) \in \mathcal{P}) \leq 2\mathbb{P}(G(n,p) \in \mathcal{P}).$$

Proof. Let us suppose that $\mathcal{P}$ is increasing, the case that $\mathcal{P}$ is decreasing is similar. Then as before,

$$\begin{aligned}\mathbb{P}(G(n,p) \in \mathcal{P}) &= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}(e(G(n,p)) = k)\mathbb{P}(G(n,k) \in \mathcal{P}) \\ &\geq \sum_{k=m}^{\binom{n}{2}} \mathbb{P}(e(G(n,p)) = k)\mathbb{P}(G(n,k) \in \mathcal{P})\end{aligned}$$

However, by the coupling we demonstrated earlier we know that for every $k \geq m$,

$$\mathbb{P}(G(n,k) \in \mathcal{P}) \geq \mathbb{P}(G(n,m) \in \mathcal{P})$$

and hence

$$\begin{aligned}\mathbb{P}(G(n,p) \in \mathcal{P}) &\geq \mathbb{P}(G(n,m) \in \mathcal{P}) \sum_{k=m}^{\binom{n}{2}} \mathbb{P}(e(G(n,p)) = k) \\ &= \mathbb{P}(G(n,m) \in \mathcal{P})\mathbb{P}(e(G(n,p)) \geq m).\end{aligned}$$

However, $e(G(n,p))$ is distributed as a binomial distribution with $\binom{n}{2}$ trials and success probability p, and it can be shown that the *median* of the binomial distribution is also equal to its expectation, and hence $\mathbb{P}(e(G(n,p)) \geq m) \geq \frac{1}{2}$, and the result follows. □

The above proofs are then useful if we know that $\mathbb{P}(G(n,p) \in \mathcal{P}) \to 0$, since then we can conclude that $\mathbb{P}(G(n,m) \in \mathcal{P}) \to 0$ (although in the case of Theorem 2.1 we would need to know that $\mathbb{P}(G(n,p) \in \mathcal{P}) \to 0$ sufficiently quickly).

However, in general we can't hope to have a bound in the other direction, that is to bound $\mathbb{P}(G(n,p) \in \mathcal{P})$ from above in terms of $\mathbb{P}(G(n,m) \in \mathcal{P})$ since there are graph properties that are very sensitive to the number of edges. For example, if we let $\mathcal{P}$ be the event that G doesn't have *exactly* m edges then clearly $\mathbb{P}(G(n,m) \in \mathcal{P}) = 0$, however for any $p \neq 0, 1$ the probability that $G(n,p)$ doesn't have exactly m edges is bounded away from zero.

A little bit later in the course, when we have some more probabilistic tools available to us, we will be able to show that, in most cases, when $p \approx \frac{m}{\binom{n}{2}}$, if $\lim_{m\to\infty} \mathbb{P}(G(n,m) \in \mathcal{P})$ exists, then $\lim_{p\to\infty} \mathbb{P}(G(n,p) \in \mathcal{P})$ will also exist, and they will coincide. Furthermore, when $\mathcal{P}$ is monotone, the converse will also hold.

Such results allow us to easily transfer results between the two models, and allows us to choose to work in whichever model suits us better. As we will see, in many cases it will be significantly easier to work in one model rather than the other.

3 Thresholds

3.1 Thresholds

Through the standard coupling described in the previous section, we can view the structure of the random graph $G(n,p)$ as *evolving* as p increases from 0 to 1. Given a property $\mathcal{P}$ it is interesting to consider how the probability $\Phi(p) := \mathbb{P}(G(n,p) \in \mathbb{P})$ evolves we vary p.

Suppose for example we're interested in the probability $\Phi(p)$ that a random graph $G(n,p)$ is connected. Clearly $G(n,0)$ is connected with probability 0 and $G(n,1)$ is connected with probability 1. For $p \in (0,1)$, we know that $\Phi(p)$ is non-zero, but how does it behave? Well, in this case, for fixed n, we can see that $\Phi(p)$ is a continuous function of p, indeed it is in fact a polynomial in p that we can write down 'explicitly'

$$\mathbb{P}(G(n,p) \text{ is connected}) = \sum_{H \in \mathcal{G}_n \,:\, H \text{ connected}} p^{e(H)}(1-p)^{\binom{n}{2}-e(H)}.$$

Furthermore, whilst this is very much not clear from the expression above, we know that $\Phi(p)$ is increasing. Indeed, as we showed in the previous section, for any increasing property $\mathcal{P}$ the function $\Phi(p)$ is an increasing function of p, which for non-trivial p will satisfies $\Phi(0) = 0$ and $\Phi(1) = 1$.

So, we know that as we increas p from 0 to 1, $\Phi(p)$ also increases continuously from 0 to 1. We might expect this transition happens somehow 'smoothly', but a rather surprising thing we will discover is that, in the limit as $n \to \infty$, this transition from being disconnected to being connected is in fact incredibly abrupt, and that this isn't just true of connectedness, but of all monotone properties.

To make this precise let us introduce the concept of a *threshold*.

Definition. Let $\mathcal{P}$ be an increasing property.

1. A sequence $p(n)$ is a *threshold* for $\mathcal{P}$ if

$$\lim_{n\to\infty} \mathbb{P}(G(n,p') \in \mathcal{P}) = \begin{cases} 0 \text{ if } p' = o(p), \\ 1 \text{ if } p' = \omega(p). \end{cases} \tag{3.1}$$

2. A sequence $m(n)$ is a *threshold* for $\mathcal{P}$ if

$$\lim_{n\to\infty} \mathbb{P}(G(n,m') \in \mathcal{P}) = \begin{cases} 0 \text{ if } m' = o(m), \\ 1 \text{ if } m' = \omega(m). \end{cases} \tag{3.2}$$

Thresholds for decreasing properties are defined as thresholds for their complements. Note that threshold are not unique. For example if $p(n) = \frac{1}{n}$ is a threshold for $\mathcal{P}$ then so is $p(n) = \frac{10}{n}$ for example.

Note that by Theorem 2.2 it follows that if p is a threshold for $\mathcal{P}$ in the binomial model then $m = \binom{n}{2}p$ is a threshold in the other model.

We will show that every monotone property has a threshold, but to do so we will first need to introduce another useful probabilistic tool, which is sometimes known as *sprinkling*, or *multi-round exposure*.

The idea is to generate $G(n,p)$ in two independent steps as follows. Suppose we take some $p_1 < p$ and let p_2 be such that $p_1 + p_2 - p_1p_2 = p$ (which re-arranges to $p_2 = \frac{p-p_1}{1-p_1}$). Let $G_1 \sim G(n,p_1)$ and $G_2 \sim G(n,p_2)$ be independent random variables and define a random variable $G = G_1 \cup G_2$. We claim that $G \sim G(n,p)$.

Indeed, it is clear that the probability that any edge appears in G is independent of the probability that any other edge appears (since this is true individually in G_1 and G_2, and G_1 and G_2 are independent), and it's easy to calculate

$$\begin{aligned}\mathbb{P}(e \in E(G)) &= \mathbb{P}(e \in E(G_1) \text{ or } e \in E(G_2)) \\ &= \mathbb{P}(e \in E(G_1)) + \mathbb{P}(e \notin E(G_1) \text{ and } e \in E(G_2)) \\ &= p_1 + (1-p_1)p_2 \\ &= p_1 + p_2 - p_1p_2 = p.\end{aligned}$$

Hence $G \sim G(n,p)$.

Theorem 3.1 (Bollobás and Thomason)**.** *Every non-trivial monotone graph property has a threshold.*

Proof. Let $\mathcal{P}$ be a non-trivial monotone graph property, where without loss of generality we may assume that $\mathcal{P}$ is increasing.

Given $\varepsilon \in (0,1)$ let $p(\varepsilon)$ be such that $\mathbb{P}(G(n,p(\varepsilon)) \in \mathcal{P}) = \varepsilon$. Note that, since $\mathbb{P}(G(n,p) \in \mathcal{P})$ is a continuous function of p such a $p(\varepsilon)$ exists, and in fact since the function is strictly increasing it is in fact unique. We will show that $\hat{p} = p(1/2)$ is a threshold for $\mathcal{P}$.

Let k be an integer and let $G_1, G_2, \ldots, G_k$ be k independent copies of $G(n,q)$, where we will choose $q \in (0,1)$ later.

We consider $G_1, G_2, \ldots, G_k$ as a multi-round exposure of $G(n,q')$, which is to say that, by similar arguments as before

$$G_1 \cup G_2 \cup \ldots \cup G_k \sim G(n,q')$$

where $q' = 1-(1-q)^k \leq kq$ (This is an application of Bernoulli's inequality, or is easy to show inductively). Hence, since $\mathcal{P}$ is increasing

$$\mathbb{P}(G(n,kq) \notin \mathcal{P}) \leq \mathbb{P}(G(n,q') \notin \mathcal{P}) = \mathbb{P}(G_1 \cup G_2 \cup \ldots \cup G_k \notin \mathcal{P}).$$

Furthermore, again since $\mathcal{P}$ is increasing, if $G_i \in \mathcal{P}$ for any i then $G_1 \cup G_2 \cup \ldots \cup G_k \in \mathcal{P}$. Hence

$$\begin{aligned}\mathbb{P}(G(n,kq) \notin \mathcal{P}) &\leq \mathbb{P}(G_1 \cup G_2 \cup \ldots \cup G_k \notin \mathcal{P}) \\ &\leq \mathbb{P}(G_i \notin \mathcal{P} \text{ for all } i) \\ &= \mathbb{P}(G_1 \notin \mathcal{P})^k\end{aligned}$$

Hence, if $p \geq k\hat{p}$, then taking $q = \hat{p}$, we see that

$$\mathbb{P}(G(n,p) \notin \mathcal{P}) \leq \mathbb{P}(G(n,k\hat{p}) \notin \mathcal{P}) \leq \mathbb{P}(G(n,\hat{p}) \notin \mathcal{P})^k = 2^{-k}.$$

It follows that if $p = \omega(\hat{p})$ then $\mathbb{P}(G(n,p) \in \mathcal{P}) \to 1$.

Conversely, if $p \leq \frac{1}{k}\hat{p}$, then taking $q = p$ we see that

$$1/2 = \mathbb{P}(G(n,\hat{p}) \notin \mathcal{P}) \leq \mathbb{P}(G(n,p) \notin \mathcal{P})^k,$$

or in other words, $\mathbb{P}(G(n,p)) \notin \mathcal{P}) \geq 2^{-\frac{1}{k}} = 1 - o_k(1)$. Again, it follows that if $p = o(\hat{p})$ then $\mathbb{P}(G(n,p) \in \mathcal{P}) \to 0$.

□

3.2 Coarse and Sharp Thresholds

To recall, p is a threshold for $\mathcal{P}$ if whenever $p' = \omega(p)$ we have that with high probability $G(n,p')$ has property $\mathcal{P}$ and conversely whenever $p' = o(p)$ we have that with high probability $G(n,p')$ has property $\mathcal{P}$. This transition from almost surely not having $\mathcal{P}$ to almost surely having $\mathcal{P}$ happens in a range of probability of size $\Theta(p)$, however for some properties $\mathcal{P}$ (and the correct threshold p) this transition could happen over a much narrower range.

Definition. Let $\mathcal{P}$ be an increasing property.

1. A sequence $p(n)$ is a *sharp threshold* for $\mathcal{P}$ if for every $\gamma > 0$

$$\lim_{n\to\infty} \mathbb{P}(G(n,p') \in \mathcal{P}) = \begin{cases} 0 \text{ if } p' \leq (1-\gamma)p, \\ 1 \text{ if } p' \geq (1+\gamma)p. \end{cases} \tag{3.3}$$

2. A sequence $m(n)$ is a *sharp threshold* for $\mathcal{P}$ if for every $\gamma > 0$

$$\lim_{n\to\infty} \mathbb{P}(G(n,m') \in \mathcal{P}) = \begin{cases} 0 \text{ if } m' \leq (1-\gamma)m, , \\ 1 \text{ if } m' \geq (1+\gamma)m. \end{cases} \tag{3.4}$$

As before, sharp thresholds are not unique functions, but only determined up to a $(1+o(1))$-multiplicative factor. Note that, perhaps confusingly, if p is a sharp threshold for $\mathcal{P}$ then $10p$ is still a threshold for $\mathcal{P}$, despite not being a sharp threshold. If p is a threshold for a property $\mathcal{P}$, but not a sharp threshold, then p is called a *coarse threshold*.

3.3 Hitting Times

Finally let us introduce another way of thinking about thresholds, in terms of the random graph process $G(m)$. Given a non-trivial monotone property $\mathcal{P}$ let us define $\tilde{m} = \tilde{m}(\mathcal{P})$ as

$$\tilde{m} = \min\{m \colon G(m) \in \mathcal{P}\}.$$

We call $\tilde{m}$ the *hitting time* of $\mathcal{P}$. Note that

$$\mathbb{P}(\tilde{m} \leq m) = \mathbb{P}(G(m) \in \mathcal{P}) = \mathbb{P}(G(n, m) \in \mathcal{P})$$

and so the location and width of a threshold for $\mathcal{P}$ have equivalent formulations in terms of the concentration of $\tilde{m}$. For example, $\hat{m}$ is a threshold for $\mathcal{P}$ if and only if $\mathbb{P}(\tilde{m} = \Theta(\hat{m})) \to 1$.

The study of hitting times can give us a better insight into some to threshold phenomena that we see. For example, we shall see that the property that $\delta(G) \geq 1$ and the property that G is connected have the same threshold function, but in fact one can even show that with high probability the hitting times for the two properties are the same. That is, in almost all random graph processes, $G(m)$ becomes connected as soon as the last isolated vertex dissapears.

4 Small Subgraphs

4.1 The First and Second Moment Methods

So, given a monotone property $\mathcal{P}$, how might we actually go about finding a threshold for $\mathcal{P}$? Let's take as an example the property $\mathcal{P}$ that $G(n,p)$ contains a triangle. A sensible random variable to consider is $T = T(n,p) =$ Number of triangles in $G(n,p)$. Then, to show that a function $\hat{p}$ is a threshold for $\mathcal{P}$, we need to prove two things:

1. $\mathbb{P}(T > 0) \to 1$ if $p = \omega(\hat{p})$;
2. $\mathbb{P}(T = 0) \to 1$ if $p = o(\hat{p})$.

Now, T is a particular well behaved random variable, since it can be written as the sum of *indicator random variables*: Given an event A the indicator random variable $\mathbb{1}_A$ takes the value 1 if A occurs and 0 otherwise. Equivalenty, $\mathbb{1}_A \sim \text{Ber}(\mathbb{P}(A))$. If, for any $U \in [n]^{(3)}$ we let A_U be the event that $G(n,p)[U]$ is a triangle, then clearly

$$T = \sum_{U \in [n]^{(3)}} \mathbb{1}_{A_U}.$$

It is easy to see that the expectation of an indicator random variable $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$ and hence, by the linearity of expectation if we have a random variable Y which can be written as a sum of indicator random variables $Y = \sum_{A \in \mathcal{A}} \mathbb{1}_A$ then $\mathbb{E}(Y) = \sum_{A \in \mathcal{A}} \mathbb{P}(A)$. It follows that

$$\mathbb{E}(T) = \sum_{U \in [n]^{(3)}} \mathbb{P}(A_U).$$

However $\mathbb{P}(A_U)$ is easy to calculate, there are 3 edges in U and each edge appears independently with probability p, and hence $\mathbb{P}(A_U) = p^3$ (independently of U). Thus

$$\mathbb{E}(T) = \binom{n}{3} p^3 = (1 + o(1)) \frac{1}{6} n^3 p^3.$$

Hence we can see that if $p = o(n^{-1})$ then $\mathbb{E}(T) \to 0$ and if $p = \omega(n^{-1})$ then $\mathbb{E}(T) \to \infty$. So, we should expect $p = n^{-1}$ to be a threshold for $\mathcal{P}$.

Now all we need is a way to go from the statements we have about the expectation of T to statements about the probability that T takes certain values. Two elementary, but useful, results that allow us to do this are Markov and Chebyshev's inequality.

Firstly, Markov's inequality tells us that it's unlikely that a non-negative random variable exceeds it's expectation significantly.

Lemma 4.1. *[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

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

□

So, Markov's inequality says that if a non-negative random variable has a small expectation then it's very likely that the random variable 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 value N^2 with probability $1/N$ for some large N. The expectation of such a random variable can be 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. One practical reason for taking the square here is that we are interested only in the 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 4.2. *Given a sequence of random variables $X_1, X_2, \ldots, X_n$, let $X = \sum_i X_i$. Then*

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

Proof.

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

□

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 4.3 (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 4.3 is best possible.

It follows from Chebyshev's inequality that if we have a sequence of random variables with $\mathbb{E}(X_n)\to\infty$ and $\mathrm{Var}(X_n) = o(\mathbb{E}(X_n)^2)$ then

$$\mathbb{P}\left(X_n \geq \frac{\mathbb{E}(X_n)}{2}\right) \geq 1 - \mathbb{P}\left(|X_n - \mathbb{E}(X_n)| \geq \frac{\mathbb{E}(X_n)}{2}\right) \geq 1 - \frac{\mathrm{Var}(X_n)}{\left(\frac{\mathbb{E}(X_n)}{2}\right)^2} = 1 - o(1),$$

and so with high probability $X_n = \Omega(\mathbb{E}(X_n))$.

We'd like to apply this now to our problem about the threshold for the appearance of a triangle, but in order to do so we need to be able to calculate the variance of the random variable T counting the number of triangles. Luckily, it is relatively simple to calculate the variance of a sum of indicator random variables. Indeed suppose that $Y = \sum_{A\in\mathcal{A}} \mathbb{1}_A$. By Lemma 4.2 we have that

$$\mathrm{Var}(Y) = \sum_{A\in\mathcal{A}} \mathrm{Var}(\mathbb{1}_A) + \sum_{A\neq B\in\mathcal{A}} \mathrm{Cov}(\mathbb{1}_A,\mathbb{1}_B).$$

Note that for any $A\in\mathcal{A}$

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

and for any $A \neq B \in \mathcal{A}$

$$\mathrm{Cov}(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{E}(\mathbb{1}_A \mathbb{1}_B) - \mathbb{E}(\mathbb{1}_A)\mathbb{E}(\mathbb{1}_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(\mathbb{P}(B|A) - \mathbb{P}(B)).$$

Hence we can express the variance of Y as follows:

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

This reduces a lot of variance calculations to a combinatorial problem.

For example, if we let T as before be the number of triangles in $G(n,p)$ then we have that $T = \sum_{U \in [n]^{(3)}} \mathbb{P}(A_U)$ and hence,

$$\mathrm{Var}(T) = \sum_{U \in [n]^{(3)}} \mathbb{P}(A_U) \left(\sum_{W \in [n]^{(3)}} \mathbb{P}(A_W|A_U) - \mathbb{P}(A_W) \right).$$

In order to calculate the inner term we split into cases depending on $|U \cap W|$. In the first case, when $|U \cap W| \leq 1$, the events A_U and A_W are independent, and so $\mathbb{P}(A_W|A_U) = \mathbb{P}(A_W)$. So, there is no contribution to the sum from these pairs.

In the second case, when $|U \cap W| = 2$, the two triangles share an edge, and so $\mathbb{P}(A_W|A_U) - \mathbb{P}(A_W) = p^2 - p^3$. For a fixed U, there are $3(n-3)$ such W.

Finally, when $|U \cap W| = 3$, i.e. $U = W$, we have that $\mathbb{P}(A_W|A_U) - \mathbb{P}(A_W) = 1 - p^3$, and for a fixed U there is one such W. Hence

$$\begin{aligned}
\mathrm{Var}(T) &= \sum_{U \in [n]^{(3)}} \mathbb{P}(A_U) \left(\sum_{W \in [n]^{(3)}} \mathbb{P}(A_W|A_U) - \mathbb{P}(A_W) \right) \\
&= \sum_{U \in [n]^{(3)}} \mathbb{P}(A_U) \left(3(n-3)(p^2 - p^3) + (1 - p^3) \right) \\
&= \binom{n}{3} p^3 \left(3(n-3)(p^2 - p^3) + (1 - p^3) \right). \qquad (4.1)
\end{aligned}$$

Theorem 4.4. *$\hat{p} = n^{-1}$ is a threshold function for the existence of a triangle in $G(n,p)$.*

Proof. Let T be the number of triangles in $G(n,p)$. Suppose first that $p = o(n^{-1})$. Then, since $\mathbb{E}(T) = \binom{n}{3} p^3 = \Theta(n^3 p^3) = o(1)$, Markov's inequality implies that

$$\mathbb{P}(T \geq 1) \leq \mathbb{E}(T) = o(1).$$

Hence, with high probability $T = 0$, that is, $G(n,p)$ contains no triangles.

Suppose then that $p = \omega(n^{-1})$. By (4.1)

$$\mathrm{Var}(T) = \binom{n}{3} p^3 \left(3(n-3)(p^2 - p^3) + (1 - p^3) \right) = O\left(n^3 p^3 \left(np^2 + 1 \right) \right).$$

Hence

$$\frac{\mathrm{Var}(T)}{\mathbb{E}(T)^2} = O\left(\frac{np^2 + 1}{n^3 p^3} \right) = O\left(\frac{1}{n^2 p^2} + \frac{1}{n^3 p^3} \right) = o(1).$$

Therefore, by Chebyshev's inequality

$$\mathbb{P}(T=0) \leq \mathbb{P}\left(|T-\mathbb{E}(T)| \geq \mathbb{E}(T)\right) \leq \frac{\mathrm{Var}(T)}{\mathbb{E}(T)^2} = o(1).$$

Hence, with high probability $T>0$, that is, $G(n,p)$ contains a triangle. □

Note that, Theorem 2.2 allows us to deduce that $\hat{m} = n^{-1}\binom{n}{2} = \frac{n-1}{2}$ is a threshold for the existence of a triangle in $G(n,m)$. Indeed, if $m = o(\hat{m})$ then $\frac{m}{\binom{n}{2}} = o(n^{-1})$ and we can conclude that

$$\mathbb{P}(G(n,m) \text{ doesn't contain a triangle}) \leq 2\mathbb{P}(G(n,p) \text{ doesn't contain a triangle}) = o(1).$$

Similarly if $m = \omega(\hat{m})$ then $\frac{m}{\binom{n}{2}} = \omega(n^{-1})$ and we can conclude that

$$\mathbb{P}(G(n,m) \text{ contains a triangle}) \leq 2\mathbb{P}(G(n,p) \text{ contains a triangle}) = o(1).$$

4.2 Equivalance Between the Two Models

We'll take a brief detour from small subgraphs, now that we've introduced Chebyshev's inequality, to quickly prove a useful theorem about the equivalence of the models $G(n,p)$ and $G(n,m)$.

Theorem 4.5. *1. Let $\mathcal{P}$ be a property, $p = p(n) \in [0,1]$ and $a \in [0,1]$. If for every sequence $m(n)$ such that $m = \binom{n}{2}p + O\big(n\sqrt{p(1-p)}\big)$ it holds that $\mathbb{P}(G(n,m) \in \mathcal{P}) \to a$ as $n \to \infty$, then also $\mathbb{P}(G(n,p) \in \mathcal{P}) \to a$.*

2. Let $\mathcal{P}$ be a monotone property, $m = m(n) \in [\binom{n}{2}]$ and $a \in [0,1]$. If for every $p = p(n) \in (0,1)$ such that $p = \frac{m}{\binom{n}{2}} + O\left(\sqrt{\frac{m\left(\binom{n}{2}-m\right)}{n^6}}\right)$ it holds that $\mathbb{P}(G(n,p) \in \mathcal{P}) \to a$ as $n \to \infty$, then also $\mathbb{P}(G(n,m) \in \mathcal{P}) \to a$.

Proof of 1. Let C be a large constant and define, for each n,

$$\mathcal{M}(C) = \left\{ m : \left| m - \binom{n}{2}p \right| \leq C\sqrt{\binom{n}{2}p(1-p)} \right\}.$$

Note that, if $m(n)$ is a sequence with $m(n) \in \mathcal{M}(C)$ for each m then, by assumption $\mathbb{P}(G(n,m) \in \mathcal{P}) \to a$. Let $m^- = m^-(n,C)$ be the element of $\mathcal{M}(C)$ which minimises $\mathbb{P}(G(n,m^-)) \in \mathcal{P}$. Then

$$\begin{aligned}
\mathbb{P}(G(n,p) \in \mathcal{P}) &= \sum_{m=0}^{\binom{n}{2}} \mathbb{P}(G(n,m) \in \mathcal{P})\mathbb{P}(e(G(n,p)) = m) \\
&\geq \sum_{m \in \mathcal{M}(C)} \mathbb{P}(G(n,m) \in \mathcal{P})\mathbb{P}(e(G(n,p)) = m) \\
&\geq \mathbb{P}(G(n,m^-) \in \mathcal{P})\mathbb{P}(e(G(n,p)) \in \mathcal{M}(C)).
\end{aligned}$$

Now, since we know $e(G(n,p))$ is distributed as a binomial random variable with $N := \binom{n}{2}$ trials and success probability p, we have that $\mathbb{E}\big(e(G(n,p))\big) = Np$ and furthermore that $\mathrm{Var}\big(e(G(n,p))\big) = Np(1-p)$.

Hence, by Chebyshev's inequality, Lemma 4.3,

$$\mathbb{P}(e(G(n,p)) \notin \mathcal{M}(C)) \leq \frac{\mathrm{Var}\big(e(G(n,p))\big)}{C^2 Np(1-p)} = \frac{1}{C^2}.$$

Hence,

$$\mathbb{P}(G(n,p) \in \mathcal{P}) \geq \left(1 - \frac{1}{C^2}\right) \mathbb{P}(G(n,m^-) \in \mathcal{P})$$

and by taking limits as $n \to \infty$ we see that $\liminf \mathbb{P}(G(n,p) \in \mathcal{P}) \geq \big(1 - \frac{1}{C^2}\big)\, a$.

Conversely, if $m^+ \in \mathcal{M}(C)$ maximises $\mathbb{P}(G(n,m^+) \in \mathcal{P})$ then

$$\begin{aligned}
\mathbb{P}(G(n,p) \in \mathcal{P}) &= \sum_{m=0}^{\binom{n}{2}} \mathbb{P}(G(n,m) \in \mathcal{P}) \mathbb{P}(e(G(n,p)) = m) \\
&\leq \sum_{m \in \mathcal{M}(C)} \mathbb{P}(G(n,m) \in \mathcal{P}) \mathbb{P}(e(G(n,p)) = m) + \sum_{m \notin \mathcal{M}(C)} \mathbb{P}(G(n,m) \in \mathcal{P}) \mathbb{P}(e(G(n,p)) = m) \\
&\leq \mathbb{P}(G(n,m^+) \in \mathcal{P}) + \mathbb{P}(e(G(n,p) \notin \mathcal{M}(C)) \\
&\leq \mathbb{P}(G(n,m^+) \in \mathcal{P}) + \frac{1}{C^2}.
\end{aligned}$$

and so as before, $\limsup \mathbb{P}(G(n,p) \in \mathcal{P}) \leq a + \frac{1}{C^2}$. Letting $C \to \infty$ we see that $\lim \mathbb{P}(G(n,p) \in \mathcal{P})$ exists and is equal to a. □

Proof of 2. As with Theorem 2.2 we'll just consider the case where $\mathcal{P}$ is increasing. If m is 0 or $N := \binom{n}{2}$ infinitely often, then it is simple to check that the statement holds, so we may assume that $m \neq 0, N$ Let C be a large constant, $p_0 = \frac{m}{N}$, and define

$$p^+ = \min\left\{1, p_0 + C\sqrt{\frac{p_0(1-p_0)}{N}}\right\} \text{ and } p^- = \max\left\{0, p_0 - C\sqrt{\frac{p_0(1-p_0)}{N}}\right\}.$$

Note that, by assumption $\lim_{n\to\infty} \mathbb{P}(G(n,p^+) \in \mathcal{P}) = a = \lim_{n\to\infty} \mathbb{P}(G(n,p^-) \in \mathcal{P})$.

We have

$$\begin{aligned}
\mathbb{P}(G(n,p^+) \in \mathcal{P}) &= \sum_{k=0}^{N} \mathbb{P}(G(n,k) \in \mathcal{P}) \mathbb{P}(e(G(n,p^+)) = k) \\
&\geq \mathbb{P}(G(n,m) \in \mathcal{P}) \mathbb{P}(e(G(n,p^+)) \geq m) \\
&\geq \mathbb{P}(G(n,m) \in \mathcal{P}) - \mathbb{P}(e(G(n,p^+)) < m).
\end{aligned}$$

and similarly

$$\begin{aligned}
\mathbb{P}(G(n,p^-) \in \mathcal{P}) &= \sum_{k=0}^{m} \mathbb{P}(G(n,k) \in \mathcal{P}) \mathbb{P}(e(G(n,p^+)) = k) + \sum_{k=m+1}^{N} \mathbb{P}(G(n,k) \in \mathcal{P}) \mathbb{P}(e(G(n,p^+)) = k) \\
&\leq \mathbb{P}(G(n,m) \in \mathcal{P}) + \mathbb{P}(e(G(n,p^-)) > m).
\end{aligned}$$

As before $\mathbb{E}\big(e(G(n,p^+))\big) = Np^+$ and

$$\mathrm{Var}\big(e(G(n,p^+))\big) = Np^+(1-p^+) \leq N\left(p_0 + C\sqrt{\frac{p_0(1-p_0)}{N}}\right)(1-p_0) \leq Np_0(1-p_0) + C\sqrt{Np_0(1-p_0)}$$

If $p^+ = 1$, then $\mathbb{P}(e(G(n,p^+)) < m) = 0$. Otherwise, by Chebyshev's inequality

$$\begin{aligned}\mathbb{P}(e(G(n,p^+)) < m) &\leq \frac{\mathrm{Var}\big(e(G(n,p^+))\big)}{\big(\mathbb{E}\big(e(G(n,p^+))\big) - m\big)^2} \\ &\leq \frac{Np_0(1-p_0) + C\sqrt{Np_0(1-p_0)}}{(Np^+ - Np_0)^2} \\ &= \frac{Np_0(1-p_0) + C\sqrt{Np_0(1-p_0)}}{C^2 Np_0(1-p_0)} \\ &= \frac{1}{C^2} + \frac{1}{C\sqrt{Np_0(1-p_0)}} \\ &\leq \frac{1}{C^2} + \frac{\sqrt{2}}{C}.\end{aligned}$$

Where the last line follows since, writing $m := \alpha N$, we have that

$$Np_0(1-p_0) = N\frac{m}{N}(1-\frac{m}{N}) = \frac{m(N-m)}{N} = \alpha(1-\alpha)N.$$

It is a simple exercise to show that this function is convex on $[0,1]$ and hence, since $m \neq 0, N$, it takes its minimum at $m = 1$ or $N-1$, and in particular is always at least $\frac{1}{2}$. A similar computation shows that $\mathbb{P}(e(G(n,p^-)) > m) \leq \frac{1}{C^2} + \frac{\sqrt{2}}{C}$.

Hence

$$a - \frac{1}{C^2} - \frac{\sqrt{2}}{C} \leq \liminf \mathbb{P}(G(n,m) \in \mathcal{P}) \leq \limsup \mathbb{P}(G(n,m) \in \mathcal{P}) \leq a + \frac{1}{C^2} + \frac{\sqrt{2}}{C}.$$

Letting $C \to \infty$ as before gives us the claimed result. □

4.3 Small Subgraphs

As we saw in the previous section, we can calculate explicitly the threshold for the existence of a triangle in $G(n,p)$. Can we do a similar thing for an arbitary fixed graph H? Well, we can calculate as before the expected number of copies of G in $G(n,p)$.

Indeed, let us denote by $\mathcal{C}(n,H)$ the set of copies of H in K_n then we have that

$$|\mathcal{C}(n,H)| = \binom{n}{v(H)} \frac{v(H)!}{|\mathrm{Aut}(H)|} = \Theta(n^{v(H)}).$$

Therefore, if we let $X_H = X_H(n,p)$ be the random variable which counts the number of copies of H in $G(n,p)$, then, since X_H can be decomposed as the sum of indicator random variables of the existence of each copy of H in $\mathcal{C}(n,H)$,

$$\mathbb{E}(X_H) = |\mathcal{C}(n,H)|p^{e(H)} = \Theta(n^{v(H)}p^{e(H)}).$$

Hence,

$$\lim_{n\to\infty} \mathbb{E}(X_H) = \begin{cases} 0 \text{ if } p = o\left(n^{-\frac{v(H)}{e(H)}}\right), \\ \infty \text{ if } p = \omega\left(n^{-\frac{v(H)}{e(H)}}\right). \end{cases}$$

Let us denote by $\rho(H) = \frac{e(H)}{v(H)}$, which we call the *density* of H. So, we might expect that $\hat{p} = n^{-\frac{1}{\rho(H)}}$ is a threshhold for containing a copy of H. However, if H contains a subgraph H' that is much denser that H itself, then even when $p \gg n^{-\frac{1}{\rho(H)}}$ we shouldn't expect to contain any copies of H'. However, we can't contain a copy of H without containing a copy of H'!

So, at the very least, we need to make sure that we contain every subgraph of H, so perhaps the proper quantity to be considering will be the *maximum subgraph density*

$$m(H) = \max\{\rho(H') \colon H' \subseteq H, e(H') \geq 1\}.$$

Let us also define the following quanitity, which will be useful

$$\Phi(H) = \Phi(H, n, p) = \min\{\mathbb{E}(X_{H'}) \colon H' \subseteq H, e(H') \geq 1\}.$$

Note that $\Phi(H) = o(1)$ if $p = o\left(n^{-\frac{1}{m(H)}}\right)$ and $\Phi(H) = \omega(1)$ if $p = \omega\left(n^{-\frac{1}{m(H)}}\right)$.

Theorem 4.6. *For every graph H with $e(H) \geq 1$, $\hat{p} = n^{-\frac{1}{m(H)}}$ is a threshold for the property that $G(n,p)$ contains a copy of H.*

Proof. Suppose first that $p = o\left(n^{-\frac{1}{m(H)}}\right)$. There is some subgraph H' of H with $\rho(H') = m(H)$ and hence

$$\mathbb{E}(X_{H'}) = \Theta\left(n^{v(H')}p^{e(H')}\right) = o(1).$$

Hence by Markov's inequality

$$\mathbb{P}(H \subseteq G(n,p)) \leq \mathbb{P}(H' \subseteq G(n,p)) \leq \mathbb{E}(X_{H'}) = o(1).$$

So, let us suppose instead that $p = \omega\left(n^{-\frac{1}{m(H)}}\right)$. We wish to show that $\mathbb{P}(H \subseteq G(n,p)) \to 1$. Since $m(H) \geq \rho(H)$, we know from the previous calculation that $\mathbb{E}(X_H) = \omega(1)$, so let us calculate the variance of X_H.

Given a copy $C \in \mathcal{C}(n,H)$ let us denote by A_C the event that $C \subseteq G(n,p)$. Note that, if $C, D \in \mathcal{C}(n,H)$, then A_C and A_D are independent if $E(C)$ and $E(D)$ don't intersect, and if they do intersect, then their intersection is a subgraph of H. Furthermore, for a fixed copy C of H and a fixed subgraph $H' \subseteq H$ there are $\Theta(n^{v(H)-v(H')})$ many $D \in \mathcal{C}$ such that $C \cap D = H'$.

Hence, by our standard expression

$$\begin{aligned}
\mathrm{Var}(X_H) &= \sum_{C\in\mathcal{C}} \mathbb{P}(A_C) \sum_{D\in\mathcal{C}} (\mathbb{P}(A_D|A_C) - \mathbb{P}(A_D)) \\
&= \Theta\left(n^{v(H)}p^{e(H)}\right) \sum_{\substack{H'\subseteq H \\ e(H')\geq 1}} \Theta\left(n^{v(H)-v(H')}\right)\left(p^{e(H)-e(H')} - p^{e(H)}\right) \\
&= \Theta\left(n^{2v(H)}p^{2e(H)} \sum_{\substack{H'\subseteq H \\ e(H')\geq 1}} \frac{1-p^{e(H')}}{n^{v(H')}p^{e(H')}}\right) \\
&= \Theta\left(\mathbb{E}(X_H)^2 \sum_{\substack{H'\subseteq H \\ e(H')\geq 1}} \frac{1-p^{e(H')}}{\mathbb{E}(X_{H'})}\right) \\
&= \Theta\left((1-p)\frac{\mathbb{E}(X_H)^2}{\Phi(H)}\right),
\end{aligned}$$

since there are only finitely many $H' \subseteq H$.

Hence by Chebyshev's inequality

$$\mathbb{P}(X_H = 0) \leq \frac{\mathrm{Var}(X_H)}{\mathbb{E}(X_H)^2} \leq \Theta\left(\frac{1}{\Phi(H)}\right).$$

□

Firstly note that, when $p = \omega(n^{-\frac{1}{m(H)}})$ we can apply Chebyshev's inequality to see that for any $\varepsilon > 0$

$$\mathbb{P}(|X_H - \mathbb{E}(X_H)| \geq \varepsilon\mathbb{E}(X_H)) \leq \Theta\left(\frac{\varepsilon^2}{\Phi(H)}\right) = o(1)$$

and so with high probability $X_H = (1+o(1))\mathbb{E}(X_H)$.

Also, we note that we have actually showed that, for any p

$$1 - \Phi(H) \leq \mathbb{P}(H \not\subseteq G(n,p)) \leq \Theta\left(\frac{1}{\Phi(H)}\right).$$

Can we say more about this probability? In order to do so we will need to introduce some new and useful probabilistic tools.

4.4 Harris' Inequality and Janson's Inequality

Suppose we have some finite set A, and some function $p : A \to [0,1]$. We let A_p be the random variable given by choosing a random subset of A, including each $a \in A$ independently with probability $p(a)$. Alternatively the distribution of A_p is given by

$$\mathbb{P}(A_p = B) = \prod_{a\in B} p(a) \prod_{a\notin B} (1-p(a)) \text{ for every } B \subseteq A.$$

Note that, when $A = E(G)$ and p is constant, this is precisely how we build the random graph $G(n,p)$.

We want to consider the following problem: Given a collection of subsets $\mathcal{B} \subseteq 2^A$, what can we say about the probability that no $B \in \mathcal{B}$ is contained in A_p? Well, if the subsets B were disjoint, then the probability that $B \subseteq A_p$ would be independent for different B and hence

$$\mathbb{P}(B \not\subseteq A_p \text{ for all } B \in \mathcal{B}) = \prod_{B\in\mathcal{B}} \mathbb{P}(B \not\subseteq A_p) = \prod_{B\in\mathcal{B}} (1 - \mathbb{P}(B \subseteq A_p)) = \Theta(e^{-\sum_{B\in\mathcal{B}} \mathbb{P}(B\subseteq A_p)}),$$

as long as no $\mathbb{P}(B \subseteq A_p)$ is too large. Note that $\mu = \sum_{B\in\mathcal{B}} \mathbb{P}(B \subseteq A_p)$ is the expected number of $B \in \mathcal{B}$ contained in A_p. The aim of this next section will be to show that, as long as the sets $B \in \mathcal{B}$ are 'not too dependent', the probability that no $B \in \mathcal{B}$ is contained in A_p does behave approximately like $e^{-\mu}$.

The first inequality we will need is Harris' inequality, which basically tells us that increasing or decreasing events are positively correlated. For example, suppose we're looking at a random graph and we're considering the events that a pair of vertices u and v are connected, call it $E_{u,v}$ and also that a pair of vertices x and y are connected, say $E_{x,y}$. Since adding extra edges only helps vertices to be connected, we should expect that knowing that $E_{u,v}$ happens should make it more likely that $E_{x,y}$ happens, that is

$$\mathbb{P}(E_{x,y}|E_{u,v}) \geq \mathbb{P}(E_{x,y}),$$

or equivalently

$$\mathbb{P}(E_{x,y} \text{ and } E_{u,v}) \geq \mathbb{P}(E_{x,y})\mathbb{P}(E_{u,v}).$$

Since the event that each $B \not\subseteq A_p$ is a decreasing event, this will allow us to give a lower bound on the probability that all the events happen, in terms of the probabilities that each individual event happens.

A function $f : 2^A \to \mathbb{R}$ is *increasing* if for every $C \subseteq B \subseteq A$ we have $f(C) \leq f(B)$, and decreasing if for every $C \subseteq B \subseteq A$ we have $f(B) \leq f(C)$. For example, the characteristic function of an increasing/decreasing property is increasing/decreasing.

Lemma 4.7. *[Harris' inequality] Let $f_1, f_2 \colon 2^A \to \mathbb{R}$ be increasing/decreasing functions and let $Y \sim A_p$ be as above. Then*

$$\mathbb{E}(f_1(Y)f_2(Y)) \geq \mathbb{E}(f_1(Y))\mathbb{E}(f_2(Y)).$$

Remark 4.8. *If we take f_i to be the indicator function for an increasing/decreasing event E_i for each i, we can conclude that*

$$\mathbb{P}(E_1 \text{ and } E_2) \geq \mathbb{P}(E_1)\mathbb{P}(E_2).$$

Proof. Let us write $X_i = f_i(Y)$ for $i = 1, 2$ and assume without loss of generality that f_i is increasing. We will induct on $|A|$. The base case, $A = \emptyset$ is clear, since Y takes only one value.

Suppose then that $|A| \geq 1$ and let us write $A = \{a_1, \ldots, a_n\}$. Let $A' = A \setminus \{a_n\}$ and let $Y' = Y \cap A'$. Note that $Y' \sim A'_{p'}$ where p' is the restriction of p to A'.

Let us consider the functions $f_{0,i} : 2^{A'} \to \mathbb{R}$ given by $f_{0,i}(B) = f_i(B)$ for $i = 1, 2$, and the random variables $Z_{0,i} = f_{0,i}(Y')$. Note that $f_{0,i}$ is increasing for each i and

$$Z_{0,i} \sim (X_i | a_n \notin Y)$$

Note that, for fixed a and b, by linearity of expectation

$$\mathbb{E}((X_1 - a)(X_2 - b)) - \mathbb{E}(X_1 - a)\mathbb{E}(X_2 - b) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$$

and hence we may assume that $\mathbb{E}(Z_{0,i}) = 0$ for $i = 1, 2$. Similarly we can define $f_{1,i}$ and $Z_{1,i}$ by taking $f_{1,i}(B) = f_i(B \cup \{a_n\})$ and $Z_{1,i} = f_{1,i}(Y')$. Note that, since the f_i are increasing we have that

$$\mathbb{E}(Z_{1,i}) \geq \mathbb{E}(Z_{0,i}) \geq 0 \text{ for } i = 1, 2.$$

However we can imply the induction hypothesis to both pairs of random variables to conclude that

$$\begin{aligned}\mathbb{E}(Z_{0,1}Z_{0,2}) &\geq \mathbb{E}(Z_{0,1})\mathbb{E}(Z_{0,2}) = 0,\\ \mathbb{E}(Z_{1,1}Z_{1,2}) &\geq \mathbb{E}(Z_{1,1})\mathbb{E}(Z_{1,2}) \geq 0,\end{aligned}$$

Also we can expand out by the definition of conditional probability

$$\begin{aligned}\mathbb{E}(X_1X_2) &= \mathbb{E}(X_1X_2|a_n \in Y)\mathbb{P}(a_n \in Y) + \mathbb{E}(X_1X_2|a_n \notin Y)\mathbb{P}(a_n \notin Y)\\ &= \mathbb{E}(Z_{1,1}Z_{1,2})p(a_n) + \mathbb{E}(Z_{0,1}Z_{0,2})(1 - p(a_n))\\ &\geq \mathbb{E}(Z_{1,1}Z_{1,2})p(a_n).\end{aligned}$$

Furthermore,

$$\begin{aligned}\mathbb{E}(X_1)\mathbb{E}(X_2) &= \left(\mathbb{E}(Z_{0,1})(1 - p(a_n)) + \mathbb{E}(Z_{1,1}p(a_n))\right)\left(\mathbb{E}(Z_{0,2})(1 - p(a_n)) + \mathbb{E}(Z_{1,2}p(a_n))\right)\\ &= \mathbb{E}(Z_{1,1})\mathbb{E}(Z_{1,2})p(a_n)^2\end{aligned}$$

and hence, since $p(a_n) \leq 1$

$$\mathbb{E}(X_1X_2) \geq \mathbb{E}(Z_{1,1}Z_{1,2})p(a_n) \geq \mathbb{E}(Z_{1,1})\mathbb{E}(Z_{1,2})p(a_n)^2 = \mathbb{E}(X_1)\mathbb{E}(X_2).$$

□

We note the following corollary that will be useful. We will use the fact that $1 - x \geq e^{-\frac{x}{1-x}}$.

Corollary 4.9. *Let $\mathcal{B}$ be a collection of subsets of a set A, let $p \in (0, 1)$ and let $Y \sim A_p$. If $\mu = \mathbb{E}(|\{B \colon B \in \mathcal{B}, B \subseteq Y\}|)$, then*

$$\mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B}) \geq \exp\left(-\frac{\mu}{1 - \max_{B\in\mathcal{B}} \mathbb{P}(B \subseteq Y)}\right).$$

Proof. Since the events $B \not\subseteq Y$ are decreasing, by Lemma 4.7

$$\begin{aligned}\mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B}) &\geq \prod_{B\in\mathcal{B}} \big(1 - \mathbb{P}(B \subseteq Y)\big)\\ &\geq \exp\left(-\sum_{B\in\mathcal{B}} \frac{\mathbb{P}(B \subseteq Y)}{1 - \mathbb{P}(B \subseteq Y)}\right)\\ &\geq \exp\left(-\frac{\mu}{1 - \max_{B\in\mathcal{B}} \mathbb{P}(B \subseteq Y)}\right).\end{aligned}$$

□

So, when none of the individual events $B \subseteq Y$ is very likely, we have a good lower bound on $\mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B})$, can we also get a good upper bound? Given $Y \sim A_p$ as before, let us define μ as before and

$$\Delta = \sum_{B_1, B_2 \in \mathcal{B}:\ B_1 \neq B_2, B_1 \cap B_2 \neq \emptyset} \mathbb{P}(B_1 \cup B_2 \subseteq Y)$$

where in the latter we note the sum is over unordered pairs.

Lemma 4.10. *[Janson's inequality] Let $\mathcal{B}$ be a collection of subsets of a set A and let $Y \sim A_p$, and let μ, Δ be as above. Then*

$$\mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B}) \leq e^{-\mu + \Delta}$$

Proof. Let us write $\mathcal{B} = \{B_1, \ldots, B_m\}$ and denote by E_i the event that $B_i \subseteq Y$. Our aim is to estimate $\mathbb{P}(\bigcap_i E_i^c)$, and to do so we'll bound, for each i, the probability that E_i doesn't happen conditioned on the event that none of the preceding E_j happen. More precisely, we're going to show that for every $i \in [m]$

$$\mathbb{P}(E_i^c | \bigcap_{j=1}^{i-1} E_j^c) \leq e^{-\mathbb{P}(E_i) + \sum_{j<i:\ B_j \cap B_i \neq \emptyset} \mathbb{P}(E_i \cap E_j)}.$$

The result will then follow by taking the product of both sides over $i \in [m]$. For example, when $i = 1$ we have

$$\mathbb{P}(E_1^c) = 1 - \mathbb{P}(E_1) \leq e^{-\mathbb{P}(E_1)}.$$

For an arbitrary i we're interested in the set of indices

$$D = \{j < i \colon B_i \cap B_j \neq \emptyset\},$$

$$I = \{j < i \colon B_i \cap B_j = \emptyset\}.$$

Let us define $E_D = \bigcup_{j \in D} E_j$ and $E_I = \bigcup_{j \in I} E_j$. We have that

$$\begin{aligned}
\mathbb{P}\left(E_i | \bigcap_{j=1}^{i-1} E_j^c\right) &= \mathbb{P}(E_i | E_D^c \cap E_I^c) \\
&= \frac{\mathbb{P}(E_i \cap E_D^c \cap E_I^c)}{\mathbb{P}(E_D^c \cap E_I^c)} \\
&\geq \frac{\mathbb{P}(E_i \cap E_D^c \cap E_I^c)}{\mathbb{P}(E_I^c)} \\
&= \frac{\mathbb{P}(E_i \cap E_D^c \cap E_I^c)}{\mathbb{P}(E_i \cap E_I^c)} \frac{\mathbb{P}(E_i \cap E_I^c)}{\mathbb{P}(E_I^c)}.
\end{aligned}$$

However, since E_i is indepedent of E_I^c we have that $\frac{\mathbb{P}(E_i \cap E_I^c)}{\mathbb{P}(E_I^c)} = \mathbb{P}(E_i)$. Hence

$$\mathbb{P}\left(E_i | \bigcap_{j=1}^{i-1} E_j^c\right) \geq \mathbb{P}(E_i) \mathbb{P}(E_D^c | E_i \cap E_I^c).$$

However, we note that we can think of $\mathbb{P}(E_D^c | E_i \cap E_I^c)$ in another way. Let's consider a random variable $Y' \sim (Y | B_i \subseteq Y)$. Note that Y' is just distributed as $A_{\hat{p}}$ where $\hat{p} = p$ on $A \setminus B_i$

and $\hat{p} = 1$ on B_i. If we define E'_j to be the event that $B_j \subseteq Y'$, and go on to define E'_D and E'_I as before, then we have that

$$\mathbb{P}(E_D^c | E_i \cap E_I^c) = \mathbb{P}(E_D'^c | E_I'^c)$$

However, $E_D'^c$ and $E_I'^c$ are decreasing events, and hence by Harris' inequality

$$\mathbb{P}(E_D'^c | E_I'^c) \geq \mathbb{P}(E_D'^c) = \mathbb{P}(E_D^c | E_i).$$

Then, by the union bound

$$\mathbb{P}(E_D^c | E_i) = 1 - \mathbb{P}(E_D | E_i) \geq 1 - \sum_{j \in D} \mathbb{P}(E_j | E_i).$$

Therefore we can conclude that

$$\mathbb{P}(E_D^c | E_i \cap E_I^c) \geq 1 - \sum_{j \in D} \mathbb{P}(E_j | E_i).$$

Hence,

$$\begin{aligned}
\mathbb{P}\left(E_i^c | \bigcap_{j=1}^{i-1} E_j^c\right) &= 1 - \mathbb{P}\left(E_i | \bigcap_{j=1}^{i-1} E_j^c\right) \\
&\leq 1 - \mathbb{P}(E_i)\mathbb{P}(E_D^c | E_i \cap E_I^c) \\
&\leq 1 - \mathbb{P}(E_i)\left(1 - \sum_{j \in D} \mathbb{P}(E_j | E_i)\right) \\
&= 1 - \mathbb{P}(E_i) + \sum_{j \in D} \mathbb{P}(E_j \cap E_i) \\
&\leq e^{-\mathbb{P}(E_i) + \sum_{j \in D} \mathbb{P}(E_i \cap E_j)}.
\end{aligned}$$

□

Hence if Δ is small compared to μ then we get an upper bound that is asymptotically very similar to the lower bound. If $\Delta \geq \mu$ then it seems that Janson's inequality doesn't tell us much, but actually we can still get an exponential upper bound on this probability as long as $\Delta = o(\mu^2)$.

Lemma 4.11. *[Generalised Janson's Inequality] Let $\mathcal{B}$, A, $Y \sim A_p$, μ and Δ be as above. If $2\Delta \geq \mu$, then*

$$\mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B}) \leq e^{-\frac{\mu^2}{4\Delta}}.$$

Proof. Clearly for any subset $\mathcal{B}' \subseteq \mathcal{B}$ we have

$$\mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B}) \leq \mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B}').$$

Let μ' and Δ' be the same quantities computed for $\mathcal{B}'$. The idea will be to form a random subset $\mathcal{B}'$ of $\mathcal{B}$ by keeping each $B \in \mathcal{B}$ independently with probability q, where we will choose q in an optimal fashion later. How will this affect μ' and Δ'?

Well, $\mathbb{E}(\mu') = q\mu$, since we keep each $B \in \mathcal{B}$ with probability q. Similarly $\mathbb{E}(\Delta') = q^2\Delta$. Hence by linearity of expectation

$$\mathbb{E}(\mu' - \Delta') = q\mu - q^2\Delta.$$

Choosing q to maximise this quantity we see that it is best to take $q = \frac{\mu}{2\Delta}$ (which is indeed ≤ 1 and so a valid choice of q), giving

$$\mathbb{E}(\mu' - \Delta') = \frac{\mu^2}{4\Delta}.$$

Since the expected value of $\mu' - \Delta'$ is large, there must be at least one choice of $\mathcal{B}'$ for which $(\mu' - \Delta') \geq \frac{\mu^2}{4\Delta}$. Hence, using Janson's inequality,

$$\mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B}) \leq \mathbb{P}(B \not\subseteq Y \text{ for all } B \in \mathcal{B}') \leq e^{-\mu' + \Delta'} \leq e^{-\frac{\mu^2}{4\Delta}}.$$

□

4.5 Small Subgraphs Revisited

Given a fixed graph H, recall that, letting $X_{H'}$ be the number of copies of H' in $G(n,p)$ for any H', we defined

$$\Phi(H) = \Phi(H, n, p) = \min\{\mathbb{E}(X_{H'}) \colon H' \subseteq H, e(H') \geq 1\}.$$

Theorem 4.12. *Let H be a graph with at least one edge. Then for any $p = p(n)$*

$$e^{-\frac{\Phi(H)}{1-p}} \leq \mathbb{P}(H \not\subseteq G(n,p)) \leq e^{-\Theta(\Phi(H))}.$$

Proof. We want to use the inequalities from the previous section, and to do so we think about $G(n,p)$ as coming from $E(G)_p$ in the obvious way. We will take our set $\mathcal{B}$ to be the set $\mathcal{C}(n,H)$ of all copies of H in K_n (or at least, the edge sets).

Then, by applying Corollary 4.9 to the set $\mathcal{B}'$ of copies of the densest subgraph $H' \subseteq H$, we see that

$$\mathbb{P}(H \not\subseteq G(n,p)) \geq \mathbb{P}(H' \not\subseteq G(n,p)) \geq \exp\left(-\frac{\mathbb{E}(X_{H'})}{1 - p^{e(H')}}\right) \geq e^{-\frac{\Phi(H)}{1-p}}.$$

For the upper bound we will use Janson's inequalities. So we need to calculate

$$\mu = \sum_{C \in \mathcal{C}} \mathbb{P}(C \subseteq G(n,p)) = \mathbb{E}(X_H)$$

and

$$\Delta = \sum_{\substack{C_1, C_2 \in \mathcal{C}(n,H) \colon C_1 \neq C_2 \\ C_1 \cap C_2 \neq \emptyset}} \mathbb{P}(C_1 \cup C_2 \subseteq G(n,p)).$$

Note that, by our standard arguments

$$\Delta \leq \sum_{\substack{H' \subseteq H \\ e(H') \geq 1}} n^{2v(H) - v(H')} p^{2e(H) - e(H')}$$

To estimate the latter we split into cases depending on the subgraph $H' \subseteq H$ given by $C_1 \cap C_2$. In this case $C_1 \cup C_2$ spans $2v(H) - v(H')$ many vertices, and so there at most $n^{2v(H)-v(H')}$ such pairs to consider. The probability that $C_1 \cup C_2 \subseteq G(n,p)$ is then given by $p^{2e(H)-e(H')}$ and hence

$$\begin{aligned}
\Delta &\leq \sum_{\substack{H' \subseteq H \\ e(H') \geq 1}} n^{2v(H)-v(H')} p^{2e(H)-e(H')} \\
&= n^{2v(H)} p^{2e(H)} \sum_{\substack{H' \subseteq H \\ e(H') \geq 1}} n^{-v(H')} p^{-e(H')} \\
&= \mathbb{E}(X_H)^2 \sum_{\substack{H' \subseteq H \\ e(H') \geq 1}} \frac{1}{\mathbb{E}(X_{H'})} \\
&\leq O\left(\frac{\mu^2}{\Phi(H)}\right)
\end{aligned}$$

However, we can't just immediately apply Jansons' inequality, since it might be that $\Delta \geq \mu$, so we split into two cases. If $\Delta \leq \frac{\mu}{2}$ then, by Janson's inequality (Lemma 4.10) we see that

$$\mathbb{P}(H \not\subseteq G(n,p)) \leq e^{-\mu+\Delta} \leq e^{-\frac{\mu}{2}} = e^{-\frac{\mathbb{E}(X_H)}{2}} \leq e^{-\frac{\Phi(H)}{2}}.$$

Otherwise, if $\Delta \geq \frac{\mu}{2}$ we use the generalised Janson's inequality (Lemma 4.11), and see that

$$\mathbb{P}(H \not\subseteq G(n,p)) \leq e^{-\frac{\mu^2}{4\Delta}} = e^{-\Omega(\Phi(H))}.$$

□

Note that in particular, we can conclude from the above that the existence of a copy of H in $G(n,p)$ does not have a sharp threshold. Indeed, since $\mathbb{E}(X_{H'}) = \Theta(n^{v(H')}p^{e(H')})$ we have that for any $\delta > 0$ there is some $\gamma > 0$ such that if $\delta n^{-\frac{1}{m(H)}} \leq p(n) \leq \delta^{-1} n^{-\frac{1}{m(H)}}$ then

$$\gamma \leq \Phi(H) \leq \gamma^{-1}$$

and hence there is some ε such that

$$\varepsilon \leq \mathbb{P}(H \not\subseteq G(n,p)) \leq 1 - \varepsilon.$$

Similarly, using our equivalence between the two models from Theorem 4.5, we can see that the same holds in $G(n,m)$.

Using some more probabilistic tools we can in fact say quite precisely how X_H is distributed within this range of p, at least for well behaved H. Here the notion of well behaved we will need is *strictly balanced*. This means, that for every proper subgraph $H' \subsetneq H$ we have that $\rho(H') < \rho(H) = m(H)$, or in other words, every subgraph of H has is strictly less dense than H itself.

Now, X_H is the sum of indicator random variables, one for each $C \in \mathcal{C}(n,H)$, however these random variables are not independent. However, we might still expect X_H to behave a little bit like $\mathrm{Bin}(N,q)$ where $N = |\mathcal{C}(n,H)| \approx \frac{n^{v(H)}}{|\mathrm{Aut}(H)|}$ is the number of potential copies of H and

$q = p^{e(H)}$ is the probability any one is contained in $G(n,p)$. In the range we're looking at, $p = \alpha n^{-\frac{1}{m(H)}} = \alpha n^{-\frac{v(H)}{e(H)}}$ and so

$$Nq = \alpha^{e(H)} n^{-v(H)} \frac{n^{v(H)}}{|\mathrm{Aut}(H)|} = \frac{\alpha^{e(H)}}{|\mathrm{Aut}(H)|} = \lambda(\alpha, H) := \lambda.$$

Hence $\mathrm{Bin}(N,q)$ tends in distribution to $\mathrm{Po}(\lambda)$, that is, for every $k \in \mathbb{N}$

$$\mathbb{P}(\mathrm{Bin}(N,q) = k) \to \mathbb{P}(\mathrm{Po}(\lambda) = k)$$

as $n \to \infty$, which we write as $\mathrm{Bin}(N,q) \xrightarrow{d} \mathrm{Po}(\lambda)$. So we might expect X_H to behave 'like' $\mathrm{Po}(\lambda)$. We shall see that this is the case, but to do so we will need to use some theorems that essentially tell us that the Poisson distribution is determined by its moments.

Definition. We say a real random variable X is *determined by its moments* if $\mathbb{E}(X^k) < \infty$ for every $k \in \mathbb{N}$ and every real random variable Y such that $\mathbb{E}(Y^k) = \mathbb{E}(X^k)$ for every $k \in \mathbb{N}$ satisfies $X \sim Y$.

Roughly, a random variable will be determined by its moments as long as the sequence of its moments doesn't grow too quickly. We will use, without proof, the fact that the Poisson distribution is determined by its moments. Furthermore, this property behaves nicely under limits.

Theorem 4.13. *Let Z be a real random variable which is determined by its moments. If $(X_i : i \in \mathbb{N})$ is a sequence of real random variables such that $\mathbb{E}\left(X_n^k\right) \to \mathbb{E}\left(Z^k\right)$ for every $k \in \mathbb{N}$ then $X_n \xrightarrow{d} Z$.*

In fact, rather than the moments of Z, it will be slightly easier to work with a similar concept called the *factorial moments*, which are

$$\mathbb{E}\left((Z)_k\right) = \mathbb{E}\big(Z(Z-1)(Z-2)\dots(Z-k+1)\big).$$

It is a simple exercise to show that both the monomials $\{1, x, x^2, \dots, x^k\}$ and the descending factorials $\{1, x, x(x-1), \dots, x(x-1)(x-2)\dots(x-k+1)\}$ form bases of the space of polynomials of degree $\leq k$ and hence, by the linearity of expectation

$$\mathbb{E}(X^k) \to \mathbb{E}(Z^k) \text{ for all } k \in \mathbb{N} \text{ if and only if } \mathbb{E}\left((X)_k\right) \to \mathbb{E}\left((Z)_k\right) \text{ for all } k \in \mathbb{N}.$$

For example the factorial moments of $\mathrm{Po}(\lambda)$ can be calculated to be

$$\mathbb{E}\left((\mathrm{Po}(\lambda))_k\right) = \sum_{\ell=0}^{\infty} e^{-\lambda} \frac{\lambda^\ell}{\ell!} (\ell)_k = \lambda^k e^{-\lambda} \sum_{\ell=k}^{\infty} \frac{\lambda^{\ell-k}}{(\ell-k)!} = \lambda^k.$$

Another useful fact is as follows - if $X = \sum_{A \in \mathcal{A}} \mathbb{1}_A$ then

$$(X)_k = \sum_{A_1,\dots,A_k \in \mathcal{A} \text{ distinct}} \mathbb{1}_{\bigcap_{i=1}^k A_k}$$

and so

$$\mathbb{E}\left((X)_k\right) = \sum_{A_1,\dots,A_k \in \mathcal{A} \text{ distinct}} \mathbb{P}\left(\bigcap_{i=1}^k A_i\right).$$

Theorem 4.14 (Bollobás / Karoński-Ruciński)**.** *If H is a strictly balanced graph, $C > 0$ and $pn^{\frac{1}{m(H)}} \to \alpha$ as $n \to \infty$, then $X_H \xrightarrow{d} Po(\lambda)$, where $\lambda = \frac{\alpha^{e(H)}}{|Aut(H)|}$.*

Proof. By the preceding discussion, it will be sufficient to show that for every integer $k \in \mathbb{N}$ we have that $\mathbb{E}\left((X_H)_k\right) \to \lambda^k$ as $n \to \infty$. Since X_H is a sum of indicator random variables we can write

$$\mathbb{E}\left((X_H)_k\right) = \sum_{C_1,\ldots,C_k \in \mathcal{C}(n,H) \text{ distinct}} \mathbb{P}\left(\bigcup_{i=1}^{k} C_i \subseteq G(n,p)\right).$$

Let us split this sum into two terms Σ_1 and Σ_2 where Σ_1 is the sum over the tuples $(C_1, \ldots, C_k)$ which are mutually vertex-disjoint, and Σ_2 is the rest. We expect that Σ_1 should be the main contriubtion to this sum, and Σ_2 should be negligible. Indeed

$$\begin{aligned}
\Sigma_1 &= \sum_{C_1,\ldots,C_k \in \mathcal{C}(n,H) \text{ mutually vertex-disjoint}} \mathbb{P}\left(\bigcup_{i=1}^{k} C_i \subseteq G(n,p)\right) \\
&= \prod_{i=0}^{k-1} \left(\frac{\prod_{j=0}^{v(H)-1}(n - v(H)i - j)}{|\mathrm{Aut}(H)|} \right) \cdot p^{e(H)} \\
&= (1+o(1)) \left(\frac{n^{v(H)} p^{e(H)}}{|\mathrm{Aut}(H)|} \right)^k \\
&= (1+o(1)) \left(\frac{\left(pn^{\frac{1}{m(H)}}\right)^{e(H)}}{|\mathrm{Aut}(H)|} \right)^k \\
&\to \left(\frac{\alpha^{e(H)}}{|\mathrm{Aut}(H)|} \right)^k = \lambda^k
\end{aligned}$$

So, let us show that Σ_2 is $o(1)$. Let e_t be the minimum number of edges contained in a union of k copies of H on t vertices. Then

$$\Sigma_2 \leq \sum_{t=v(H)}^{kv(H)-1} n^t p^{e_t} = \sum_{t=v(H)}^{kv(H)-1} \left(np^{\frac{e_t}{t}}\right)^t.$$

Hence, it would be enough to show that $\frac{e_t}{t} > m(H)$ when $t \leq kv(H)-1$, since then $np^{\frac{e_t}{t}} = o(1)$. In other words, if $C_1, \ldots, C_k$ are copies of H and $v\left(C_1 \cup \ldots, \cup C_k\right) < kv(H)$, then $\rho(C_1 \cup \ldots, \cup C_k) > m(H) = \rho(H)$.

Suppose first that $k = 2$. So, we have two copies C_1 and C_2 of H such that $v(H) \leq v(C_1 \cup C_2) \leq 2v(H) - 1$. We note that $v(C_1 \cup C_2) + v(C_1 \cap C_2) = v(C_1) + v(C_2)$ and similarly $e(C_1 \cup C_2) + e(C_1 \cap C_2) = e(C_1) + e(C_2)$. Hence

$$\frac{e(C_1 \cup C_2)}{v(C_1 \cup C_2)} = \frac{e(C_1) + e(C_2) - e(C_1 \cap C_2)}{v(C_1) + v(C_2) - v(C_1 \cap C_2)}$$

However, $\frac{e(C_1)}{v(C_1)} = \rho(H) = m(H) = \frac{e(C_2)}{v(C_2)}$ and if $\emptyset \neq C_1 \cap C_2 \neq H$ we have that $\frac{e(C_1 \cap C_2)}{v(C_1 \cap C_2)} < m(H)$ and so

$$\frac{e(C_1) + e(C_2) - e(C_1 \cap C_2)}{v(C_1) + v(C_2) - v(C_1 \cap C_2)} > m(H).$$

Otherwise, $v(C_1 \cap C_2) = 1$ and $e(C_1 \cap C_2) = 0$ and the above still holds.

Suppose then that $k \geq 3$. In this case there is some index j such that $(C_i \colon i \neq j)$ are still not mutually vertex disjoint, without loss of generality $j = k$. Let $F_1 = \bigcup_{i=1}^{k-1} C_i$ and $F_2 = C_i$. By the induction hypothesis $\rho(F_1) > m(H)$, $\rho(F_2) = m(H)$ and since $F_1 \cap F_2 \subseteq H$ if $F_1 \cap F_2 \neq \emptyset$ then, $\rho(F_1 \cap F_2) \leq m(H)$, and so

$$\frac{e(F_1 \cup F_2)}{v(F_1 \cup F_2)} = \frac{e(F_1) + e(F_2) - e(F_1 \cap F_2)}{v(F_1) + v(F_2) - v(F_1 \cap F_2)} > m(H).$$

Otherwise, $e(F_1 \cap F_2) = v(F_1 \cap F_2) = 0$, and the above still holds.

□

5 Evolution of the random graph

5.1 The Sub-Critical Phase

Let us return to the questions of the how the random graph process $G(1), G(2), \ldots, G(\binom{n}{2})$ develops over time, in particular in terms of its component structure. We will see that we can split this development into clearly distinguishable phases. For this purpose, we will be interested in making statements about $G(n,m)$ that hold with high probability, but in order to do so it will be easier to talk about the model $G(n,p)$ with $p = m/\binom{n}{2}$. Using Theorems 2.2 and 4.5 we can translate between statements in both models.

So, from the results of last section we know that, if $p = o(n^{-2})$ then we don't expect there to be any edges at all in the random graph (although in the uniform model this doesn't translate to an interesting statement).

As p gets larger, edges will start to appear, but at first they will be disjoint, and so $G(n,p)$ will likely be a matching. At certain points, larger subgraphs will start to appear. For example, by Theorem 4.6 the path of length two has a threshold at $\hat{p} = n^{-\frac{3}{2}}$ and so for $p = o\left(n^{-\frac{3}{2}}\right)$ whp $G(n,p)$ will contain no paths of length two, and so will be a matching, whereas when $p = \omega\left(n^{-\frac{3}{2}}\right)$ whp $G(n,p)$ will contain a path of length two. Translated to the uniform model, this happens around the threshold $\hat{m} = \hat{p}\binom{n}{2} = \Theta(\sqrt{n})$.

As p increases, larger and larger subgraphs will start to appear, and in particular similar statements can be made for any fixed tree of size t. Indeed, it's an easy calculation that $m(T) = \rho(T) = \frac{t-1}{t}$ and hence we see that, when

$$n^{-\frac{t}{t-1}} \ll p \ll n^{-\frac{t+1}{t}}$$

then with high probability each component of $G(n,p)$ will be a tree on at most t vertices. A similar statement holds in $G(n,m)$ when

$$n^{2-\frac{t}{t-1}} = n^{\frac{t-2}{t-1}} \ll m \ll n^{\frac{t-1}{t}} = n^{2-\frac{t+1}{t}}.$$

Furthermore, it is simple to check that any graph which contains a cycle has $m(H) \geq 1$. So, if m grows like n^{ε} and we let ϵ increase from 0 to 1, then in this range $G(n,m)$ will be a forest, with the largest component growing as a function of ϵ. In fact, $G(n,m)$ continues to be acyclic all the way up until m is almost of order n.

Theorem 5.1. *If $p = o\left(\frac{1}{n}\right)$, and so $p\binom{n}{2} = o(n)$, then with high probability $G(n,p)$ is a forest.*

Proof. Let us denote by X the number of cycles in $G(n,p)$. We can calculate that

$$\begin{aligned}\mathbb{E}(X) &\leq \sum_{k=3}^{n} \binom{n}{k} \frac{(k-1)!}{2} p^k \\ &\leq \sum_{k=3}^{n} (np)^k\end{aligned}$$

However since $np \to 0$, for large enough n we can bound this using the formula for the sum of a geometric series as

$$\mathbb{E}(X) \leq \frac{np}{1-np} = o(1).$$

Hence $\mathbb{E}(X) = o(1)$ and so, by Markov's inequality (Lemma 4.1),

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}(X) \to 0.$$

It follows that with high probability $G(n,p)$ contains no cycles, and hence is a forest. □

So, what happens when $m = \Theta(n)$? It turns out a really startling transition happens at this stage. When m is just a little smaller than $n/2$, $G(n,m)$ still only has very small components, and is quite close to a forest. On the other hand, when m is just a little large than $n/2$, suddenly the behaviour of $G(n,m)$ changes, and a single 'giant' component emerges, of linear order in n.

In order to analyse the component structure of $G(n,p)$ where $p = \frac{c}{n}$ we are going to use an algorithmic procedure called *breadth first search* (BFS). The idea is to recursively expose the component containing a vertex v using a 'First-in-first' out procedure. Given a graph G, and vertex v in $V(G)$ we maintain an ordered *list* of vertices A, the *active* vertices, and two sets of vertices W, the *visited* vertices, and U, the *unvisited* vertices, as well as a subgraph T of $G[W \cup A]$. Initially we set $A(0) = (v)$, $W(0) = \{v\}$, $T(0) = (\{v\}, \emptyset)$ and $S(0) = \emptyset$.

At each stage t in the algorithm we look at the first vertex in the list $A(t) = (a_1, \ldots, a_s)$, and we look at the neighbourhood of a_1 amongst the unvisited vertices, that is, $N_G(a_1) \cap U(t) = \{w_1, \ldots, w_r\}$. We mark $w_1, \ldots, w_r$ as active by appending them to the list $A(t+1)$ and deleting them from $U(t+1)$, and then mark a_1 as visited by adding it to $W(t+1)$ and deleting it from $A(t+1)$. Also we add the edges (a_1, w_i) to T for each $1 \leq i \leq r$ That is:

- $A(t+1) = (a_2, \ldots, a_s, w_1, \ldots, w_r)$;
- $W(t+1) = W(t) \cup \{a_1\}$;
- $U(t+1) = U(t) \setminus \{w_1, \ldots, w_r\}$;
- $T(t+1) = T(t) \cup \{(a_1, w_i) \colon i \in [r]\}$.

The algorithm terminates when A is empty.

Let us note a few things about this algorithm. Firstly, at each stage in the algorithm the triple (W, A, U) partitions the vertex set $V(G)$. Secondly, at each stage in the algorithm one vertex moves into W, and so after at most n steps the list A is empty and the algorithm terminates. Also, we note that the subgraph T is always a tree, with vertex set $W \cup A$. Indeed, whenever we add a vertex to A we add an edge to that vertex to T from a vertex of W, and at each stage of the algorithm we only add edges from $W \cup A$ to U, hence we can never form a cycle. Finally, we note that if G is connected then eventually every vertex of G will appear in A, which can be seen for example by inducting on the distance to the root v, and hence eventually every vertex will appear in T. In this case, it follows that the final tree $T(n)$ is a spanning tree of G.

If we run this algorithm on a random graph $G(n,p)$, starting at a fixed vertex v (and with some fixed underlying order of the vertices), then we will uncover a spanning tree T over the

component C_v containing v in $G(n,p)$. Each of the sets considered in the algorithm is now a random variable, and by analysing these random variables we can hope to discover some properties of $G(n,p)$.

One useful way to think about the analysis of this algorithm is via the *principal of deferred decisions*. Rather than choosing at the beginning of time a random graph $G(n,p)$ we can, in the process of analysing the algorithm, only choose whether or not an edge is present in the graph at the point at which that edge is relevant to the algorithm. In our case, this is when we look at the neighbourhood $N_G(a_1) \cap U(t)$. At the point in which we consider where a vertex a_1 is adjacent to a vertex $u \in U(t)$, we say that we *query* the edge (a_1, u) and the query is *succesful* if the edge lies in the graph.

We can formalise this by considering a sequence of independently identically distributed random variables $(X_i \colon i \in \mathbb{N})$ which are each $\mathrm{Ber}(p)$ random variables. We run the same algorithm as before, but without reference to any graph, and at stage t rather than adding $N_G(a_1) \cap U(t)$ to the active list we instead take the 'next' $|U(t)|$ many X_i in the list, let us call them $(X(w) \colon w \in U(t))$, and we add w to $A(t+1)$ if and only if $X(w) = 1$. It is relatively clear that the distribution of A, W, S and T are the same whether we run the algorithm on $G(n,p)$ or according to this list of random variables, but this change in perspective can help to analyse the process.

Suppose that run this procedure starting at a vertex v, how large do we expect the component C_v to be? The first thing we'll see is that, if $c < 1$ then we expect the component to die out before it gets too large. In order to see this we will need some good bounds on the tails of the binomial distribution.

Lemma 5.2 (The Chernoff bounds)**.** *Let $X \sim Bin(n,p)$, $\mu = \mathbb{E}(X) = np$ and let $t \geq 0$. Then,*

- $\mathbb{P}\Big(X \geq \mathbb{E}(X) + t\Big) \leq \exp\left(-\frac{t^2}{2(\mu + \frac{t}{3})}\right)$;
- $\mathbb{P}\Big(X \leq \mathbb{E}(X) - t\Big) \leq \exp\left(-\frac{t^2}{2\mu}\right)$.

Theorem 5.3. *If $c < 1$ and $p = \frac{c}{n}$, and so $p\binom{n}{2} \approx c\frac{n}{2}$, then with high probability the largest component of $G(n,p)$ has at most $\frac{3}{(1-c)^2} \log n$ vertices.*

Proof. Given an arbitrary vertex v, we will estimate the probability that v belongs to a component of size $\geq k$ by analysing the BFS procedure described above starting at v.

At each stage of the algorithm we look at the neighbourhood of some vertex $a_1(t)$ from within $U(t)$. Let us define X_t to be the number of neighbours we find in stage t. We note that $X_t \sim \mathrm{Bin}(m,p)$ where $m = |U(t)| \leq n$ and so X_t is *stochastically dominated* by $\mathrm{Bin}(n,p)$. By this we mean we can couple X_t with a random variable $X'_t \sim \mathrm{Bin}(n,p)$ such that $X_t \leq X'_t$ as functions on the same probability space. Note further that we can can take the X'_t to be independent for each t.

Now let us note a few things about the algorithm. Firstly, we have that

$$|W(t) \cup A(t)| = 1 + \sum_{i=1}^{t} X_i$$

and in each round where A is non-empty, a vertex is added to W. Hence, if v belongs to a component of size $> k$, then after the first k rounds of the algorithm

$$1 + \sum_{i=1}^{k} X_i = |W(k) \cup A(k)| \geq |W(k)| = k.$$

Hence, the probability that v is contained in a component of size $> k$ is at most

$$\mathbb{P}\left(\sum_{i=1}^{k} X_i \geq k\right) \leq \mathbb{P}\left(\sum_{i=1}^{k} X_i' \geq k\right)$$

However, $\sum_{i=1}^{k} X_i' \sim \mathrm{Bin}(kn, p)$, which has expectation $knp = ck < k$ and so by the Chernoff bounds

$$\begin{aligned}\mathbb{P}\Big(\sum_{i=1}^{k} X_i' \geq k\Big) &= \mathbb{P}\Big(\mathrm{Bin}(kn,p) \geq ck + (1-c)k\Big) \\ &\leq \exp\left(-\frac{(1-c)^2k^2}{2\left(ck + \frac{(1-c)k}{3}\right)}\right) \\ &\leq \exp\left(-\frac{(1-c)^2k}{2}\right).\end{aligned}$$

Hence if $k \geq \frac{3}{(1-c)^2}\log n$ then

$$\mathbb{P}(v \text{ is in a component of size } > k) \leq n^{-\frac{3}{2}}.$$

Hence by the union bound,

$$\mathbb{P}(G(n,p) \text{ has a component of size } > k) \leq nn^{-\frac{3}{2}} = n^{-\frac{1}{2}} = o(1).$$

□

So, for $m = \frac{cn}{2}$ with $c < 1$ all the components are still quite small. Let us show further that they are all simple in structure, in that they are close to being trees.

Definition. The *excess* of a connected component C of a graph is $e(C) - v(C)$. Note that the excess of a component if -1 if and only if C is a tree and 0 if and only if C contains a unique cycle. A component if *complex* if it's excess is positive, that is, it contains at least two cycles.

Theorem 5.4. *If $p = \frac{1}{n} - \frac{s(n)}{\binom{n}{2}}$, where $s(n) > 0$, and so $p\binom{n}{2} \approx \frac{n}{2} - s(n)$, then the probability that $G(n,p)$ contains a complex component is at most $\frac{n^2}{4s^3}$. In particular if $s = \omega(n^{\frac{2}{3}})$ then with high probability $G(n,p)$ contains no complex components.*

Proof. If a component of $G(n,p)$ contains at least two cycles, then in particular it has to contain a subgraph that consist of two cycles joined by by a path (possibly a degenerate path), or a theta graph (A cycle together with a path joining two points on the cycle).

We note that all of these graphs can be constructed by taking a path $P = v_1 v_2 \dots v_k$, and adding two edges to it, one incident to each endpoint $\{v_1, v_k\}$ of the graph.

In particular, if we fix a given set of k vertices, then there are at most $k^2 k!$ such subgraphs on this set of vertices. Let X denote the number of subgraphs of this type in $G(n,p)$ then we can calculate

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=4}^{n} \binom{n}{k} k^2 k! p^{k+1} \\ &\leq \sum_{k=4}^{n} \frac{n^k}{k!} k^2 k! \frac{1}{n^{k+1}} \left(1 - \frac{2s}{n-1}\right)^k \\ &\leq \sum_{k=4}^{n} \frac{k^2}{n} \exp\left(-\frac{2ks}{n}\right) \\ &\leq \int_0^{\infty} \frac{x^2}{n} \exp\left(-\frac{2xs}{n}\right) dx,\end{aligned}$$

where we can replace the sum by the integral since the function is decreasing on $[2, \infty]$ and non-negative on $[0, \infty]$. Taking a well-chosen change of variables, we set $y = \frac{2xs}{n}$ so that $dy = \frac{2s}{n} dx$ and $\frac{x^2}{n} = y^2 \frac{n}{4s^2}$

$$\mathbb{E}(X) \leq \int_0^{\infty} y^2 \frac{n}{4s^2} \frac{n}{2s} e^{-y} \, dy = \frac{n^2}{4s^3} \int_0^{\infty} \frac{y^2}{2} e^{-y} \, dy = \frac{n^2}{4s^3}.$$

Hence

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}(X) \leq \frac{n^2}{4s^3}.$$

□

So, if we want to know the order of the largest component of G_m whilst $m = \frac{cn}{2}$ with $c < 1$, then we only need to know the order of the largest component which contains either one (unicyclic) or zero (trees) excess edges. We state without proof the following lemma (although its proof is just a standard first moment calculation)

Lemma 5.5. *Let $p = \frac{c}{n}$, and so $p\binom{n}{2} \approx \frac{cn}{2}$, where $c < 1$ and let $f(n)$ be a function such that $f(n) \to \infty$. Then with high probability the number of vertices in unicyclic components of $G(n,p)$ is $O(f)$.*

Together with Theorems 5.3 and 5.4 it follows that when $m \approx \frac{cn}{2}$, with $c < 1$, $G(n,m)$ is still almost a forest with all components quite small. More precisely, each component of $G(n,m)$ is either a tree or unicyclic and of size at most $\frac{3}{(1-c)^2} \log n$, and the number of vertices in unicylic components is $O(f)$ for any slowly growing function f of n, and so most of the vertices are in tree components. As we shall see on the example sheet, there is in fact some component of size $\Omega(\log n)$, and so the largest component of $G(n,m)$ will be a tree of size $\Omega(\log n)$.

5.2 The Galton-Watson Process

Let us think more generally about the growth of the tree $T(t)$ we build in the BFS process. Whilst $|T(t)|$ is small, the size of $U(t)$ is approximately n, and so we expect the size of $N_G(a_1) \cap U(t)$

to be approximately distributed as $\mathrm{Bin}(n, \frac{c}{n})$. Hence, in the early stages of the BFS process, T will grow very much like a tree grown via the following *idealised* process - We start with a single vertex v and then recursively and independently for each leaf we add $\mathrm{Bin}(n, \frac{c}{n})$ many children. If we expect this ideal process to die out quickly, before our approximation for T starts to fail, then we should expect the tree we build to be small. On the other hand, whilst this approximation only holds whilst T is small, if we expect this ideal process to grow very large, then we should expect T to grow at least to the point that the approximation fails to hold.

So, let us analyse this idealised process, which are called *Galton-Watson trees*, in a more general setting. Let $(X_{i,j} \colon i, j \in \mathbb{N})$ be a family of independent, identically distributed random variables such that $X_{i,j} \sim X$ for every $i, j \in \mathbb{N}$. We will define a sequence of random variables $(Z_i \colon i \in \mathbb{N})$ by letting

- $Z_0 = 1$;
- $Z_{i+1} = \sum_{j=1}^{Z_i} X_{i,j}$.

If we think about this in terms of the tree-branching process defined above, we have that Z_i represents the number of vertices in the ith level of the tree. Let us define the *extinction probability*

$$\rho_X = \lim_{n \to \infty} \mathbb{P}(Z_n = 0)$$

that is, the probability that at some point the branching process stops. Note that, this sequence is monotone increasing, since if $Z_n = 0$ then $Z_{n+1} = 0$, and so the limit exists.

Lemma 5.6. *If $\mathbb{E}(X) < 1$, then $\rho_X = 1$.*

Proof. Note that $\mathbb{E}(Z_0) = 1$ and if $i \geq 0$ then

$$\begin{aligned}
\mathbb{E}(Z_{i+1}) &= \sum_{k=0}^{\infty} \mathbb{E}(Z_{i+1} | Z_i = k) \mathbb{P}(Z_i = k) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(Z_i = k) \mathbb{E}\left(\sum_{j=1}^{k} X_{i,j} \right) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(Z_i = k) k \mathbb{E}(X) \\
&= \mathbb{E}(Z_i) \mathbb{E}(X).
\end{aligned}$$

Hence, since $\mathbb{E}(Z_i) = \mathbb{E}(X)^i$ and so if $\mathbb{E}(X) < 1$ it follows by Markov's inequality that

$$1 - \rho_X = \lim_{n \to \infty} \mathbb{P}(Z_i \geq 1) \leq \lim_{n \to \infty} \mathbb{E}(Z_i) = \lim_{n \to \infty} \mathbb{E}(X)^n = 0.$$

□

So, perhaps not surprisingly, if the average number of children at each vertex is less than one, then the average number of vertices at each level is getting exponentially small.

Note that, if the branching process doesn't die out on the first step, then in order for it to die out each of the Z_1 many branching processes starting at the first level must die out, and these

are all just independent copies of the same process. Hence ρ_X satisfies the following recurrence relation

$$\rho_X = \mathbb{P}(X = 0) + \sum_{k=1}^{\infty} \mathbb{P}(X = k)\rho_X^k := f_X(\rho_X).$$

That is, ρ is a fixed point of the function $f_X(x) := \sum_{k=0}^{\infty} \mathbb{P}(X = k)x^k$. This is in fact a very well-known function, known as the *probability generating function* of X, and has a lot of nice properties.

For example, it is easy to see that $f_X(0) = \mathbb{P}(X = 0)$ and $f_X(1) = 1$ and, as long as X is not equal to 0 almost surely, it is clear that f_X is strictly increasing. Furthermore, as long as X is not equal to 0 or 1 almost surely, then

$$f_X'(x) = \sum_{k=1}^{\infty} k\mathbb{P}(X = k)x^{k-1}$$

is also strictly increasing. It follows that $f_X(x) = x$ has at most one solution in $[0, 1)$. Let us denote this solution by x_0, and let $x_0 = 1$ if there is not solution in $[0, 1)$.

Theorem 5.7. *The probability of extinction satisfies $\rho_X = x_0$.*

Proof. By the discussion above, ρ_X is a fixed point of f_X, so it will be sufficient to show that ρ_X is the smallest non-negative fixed point. To this end let us consider

$$F_i(x) = \mathbb{E}\left(x^{Z_i}\right) = \sum_{k=0}^{\infty} \mathbb{P}(Z_i = k)x^k.$$

That is, the probability generating function of Z_i. Note that $F_1 = f_X$ and $F_i(0) = \mathbb{P}(Z_i = 0)$. Morever

$$\begin{aligned}
F_{i+1}(x) &= \mathbb{E}\left(x^{Z_{i+1}}\right) \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left(x^{Z_{i+1}} | Z_i = k\right) \mathbb{P}(Z_i = k) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(Z_i = k)\mathbb{E}\left(x^{X_{i,1}+X_{i,2}+\ldots+X_{i,k}}\right) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(Z_i = k) \prod_{j=1}^{k} \mathbb{E}\left(x^{X_{i,j}}\right) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(Z_i = k) \left(f_X(x)\right)^k \\
&= F_i(f_X(x)).
\end{aligned}$$

Hence, $F_i(x) = f_X^{(i)}(x)$, that is, f_X applied i times to x. In particular, $F_i(x_0) = x_0$ for all i.

Now,

$$\rho_X = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_{n\to\infty} F_n(0)$$

However, since F_i is monotone, $F_i(0) \leq F_i(x_0) = x_0$ and so, taking the limit as $i \to \infty$ we can conclude that

$$\rho_X = \lim_{n\to\infty} F_n(0) \leq \lim_{n\to\infty} F_n(x_0) = x_0.$$

However, by definition x_0 was the smallest non-negative fixed point of f_X, and $f_X(\rho_x) = \rho_X$, and hence $x_0 = \rho_X$. □

Corollary 5.8. *If $\mathbb{E}(X) > 1$, then $\rho_X < 1$.*

Proof. We have that

$$f'_X(1) = \sum_{k=0}^{\infty} \mathbb{P}(X = k)k = \mathbb{E}(X) > 1$$

and hence $f_X(x) < x$ for some $x < 1$. On the other hand $f_X(0) = \mathbb{P}(X = 0) \geq 0$ and so either $f_X(0) = 0$ or $f_X(0) > 0$, in which case $f_X(x) = x$ must have a solution in $(0, 1)$ by the mean value theorem. □

We note that it is possible to show that $\rho_X = 1$ in the *critical* case when $\mathbb{E}(X) = 1$.

Let us look then at the concrete example we are interested in, where $X_n \sim \text{Bin}(n, p)$ with $p = \frac{c}{n}$. In this case

$$f_{X_n}(x) = \sum_{k=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} x^k = (1 - p + xp)^n.$$

So, ρ_{X_n} will be the smallest fixed point of this equation.

This is not a particularly pleasant expression to work with, however we're really only interested in the behaviour as $n \to \infty$, in which case we already know that $X_n \xrightarrow{d} \text{Po(c)}$.

So, let us look at the limiting distribution $Y = \text{Po(c)}$. In this case

$$f_Y(x) = \sum_{k=0}^{\infty} e^{-c} \frac{c^k}{k!} x^k = e^{-c} \sum_{k=0}^{\infty} \frac{(cx)^k}{k!} = e^{c(x-1)}.$$

Hence ρ_Y satisfies the equation $\rho_Y = e^{c(\rho_Y - 1)}$. It will often be convenient to phrase this in terms of the complementary probability $\beta_c = 1 - \rho_Y$, which satisfies the equation.

$$\beta_c + e^{-\beta_c c} = 1 - \rho_Y + e^{(\rho_Y - 1)c} = 1.$$

However, since

$$\lim_{n\to\infty} (1 - p + xp)^n = \lim_{n\to\infty} (1 + \frac{c(x-1)}{n})^n = e^{c(x-1)}.$$

it is an exercise in analysis to show that $\rho_{X_n} \to \rho_Y = 1 - \beta_c$.

5.3 The Emergence of the Giant Component

Let us suppose then that $p = \frac{c}{n}$ with $c > 1$. By analysing the breadth first search process described at the start of this section, with the help of our results on the branching processes, we will show that with high probability there will be unique linear sized component in $G(n, p)$, and that the remaining components are small.

Theorem 5.9. *Let $c > 1$ and let $p = \frac{c}{n}$, and so $p\binom{n}{2} \approx c\frac{n}{2}$. Then with high probably the largest component of $G(n,p)$ has size $(1+o(1))\beta_c n$. Furthermore, with high probability the second largest component has size at most $\frac{16c}{(c-1)^2}\log n$.*

Proof. We will start by showing there is a *gap* in the order of components. That is, if we define

$$k^- = \frac{16c}{(c-1)^2}\log n \quad \text{and} \quad k^+ = n^{\frac{2}{3}},$$

then whp no component of $G(n,p)$ will have order between k^- and k^+.

In order to do so, we will analyse the breadth first search algorithm starting at some fixed vertex $v \in [n]$.

Claim. With probability $1 - o\left(\frac{1}{n}\right)$ either v belongs to a component of size $\leq k^-$, or for every $k^- \leq k \leq k^+$ the number of active vertices satisfies

$$|A(k)| \geq \frac{(c-1)k}{2}.$$

Proof of Claim. Let B_v be the event that the claim does not hold. If B_v occurs, then there must be some $k^- \leq k \leq k^+$ such that the event

$$D_k = \text{“}|A(k)| < \frac{(c-1)k}{2}\text{”}$$

occurs, and so $\mathbb{P}(B_v) \leq \sum_{k=k^-}^{k^+} \mathbb{P}(D_k)$.

Recall that $|W(k) \cup A(k)| = 1 + \sum_{i=1}^{k} X_i$ and that $|W(k)| \leq k$ and hence if $A(k)$ holds then

$$1 + \sum_{i=1}^{k} X_i < k + \frac{(c-1)k}{2} = \frac{(c+1)k}{2}.$$

However, since $U(t)$ is decreasing, we have that $|U(t)| \geq |U(k)| \geq n - \frac{(c+1)k}{2} \geq n - \frac{(c+1)k^+}{2}$ for all $t \leq k$. In particular, the number of neighbours found at each step $X_i \geq X_i^- \sim \mathrm{Bin}(n', p)$, where $n' = n - \frac{(c+1)}{2}k^+$.

Writing $m = kn' = (1+o(1)kn$ and $\mu = mp = (1+o(1))ck$, we see from the Chernoff bounds

(Lemma 5.2)

$$\begin{aligned}
\mathbb{P}(B_v) &\leq \sum_{k=k^-}^{k^+} \mathbb{P}\left(D_k\right) \leq \sum_{k=k^-}^{k^+} \mathbb{P}\left(\sum_{i=1}^{k} X_i \leq \frac{(c+1)k}{2}\right) \\
&\leq \sum_{k=k^-}^{k^+} \mathbb{P}\left(\mathrm{Bin}(m,p) \leq \mu - (1+o(1))\frac{(c-1)k}{2}\right) \\
&\leq \sum_{k=k^-}^{k^+} \exp\left(-\frac{((1+o(1)\frac{(c-1)k}{2})^2}{2(1+o(1))ck}\right) \\
&\leq \sum_{k=k^-}^{k^+} \exp\left(-\frac{(c-1)^2 k}{9c}\right) \\
&\leq k^+ \exp\left(-\frac{(c-1)^2 k^-}{9c}\right) \\
&\leq n^{\frac{2}{3}} \exp\left(-\frac{16}{9}\log n\right) \\
&= n^{\frac{2}{3}-\frac{16}{9}} = n^{-\frac{10}{9}} = o\left(\frac{1}{n}\right).
\end{aligned}$$

□

Note that, if $|A(k)| > 0$ then the size of the component containing v has size at least k. An immediate consequence of the the claim, by the union bound, is that with high probability every vertex $v \in [n]$ either belongs to a component of order $\leq k^-$, or at least k^+.

So, what remains is to show that there is a unique component L of order at least k^+, and to estimate the size of L. For the first, it will be sufficient to show that for any two vertices v_1 and v_2, the probability that they belong to different components of size $\geq k^+$ is small.

Suppose that we run the breadth first search algorithm starting at v_1 for k^+ steps, and then run it again starting from v_2 for k^+ steps. If v_1 and v_2 belong to different components of size $\geq k^+$. then neither process dies out before k^+ steps, and the trees $T_1(k^+) \subseteq C_{v_1}$ and $T_2(k^+) \subseteq C_{v_2}$ that we discover do not share any vertices.

However then the sets $A_1(k^+)$ and $A_2(k^+)$ of active vertices are disjoint, and by the claim, we may assume that both have size at least $\frac{(c-1)k^+}{2}$. Furthermore, $A_1(k^+) \subseteq U_2(k^+)$ and $A_2(k^+) \subseteq U_1(k^+)$, and in the first part of the process we did not query any edges between $A_1(k^+)$ and $U_1(k^+)$, and in the second part of the process we did not query any edges between $A_2(k^+)$ and $U_2(k^+)$. Hence, in the whole process we did not query any of the edges between $A_1(k^+)$ and $A_2(k^+)$.

Therefore, by the principle of deferred decisions, the probability that v_1 and v_2 belong to different components of size $\geq k^+$ is at most the probability that none of these edges are in $G(n,p)$. There are $|A_1(k^+)||A_2(k^+)| \geq \left(\frac{(c-1)k^+}{2}\right)^2$ many potential edges, and so the probability that none of these are in $G(n,p)$ is

$$(1-p)^{\left(\frac{(c-1)k^+}{2}\right)^2} \leq \exp\left(\frac{(c-1)^2 c n^{\frac{1}{3}}}{4}\right) = o\left(n^{-2}\right).$$

Hence the probability that there is some pair of vertices v_1 and v_2 which belong to different components of size $\geq k^+$ is $o(1)$.

So, we have shown that with high probability every vertex of $G(n,p)$ is either in a component of size $\leq k^-$ or lies in a unique component of size $\geq k^+$. It remains then to show that this unique component is actually linear in size, and to estimate it's density. We will do this by estimating the number of vertices lying in small components.

Let us write $V(G(n,p)) = S \cup L$ where S are the vertices in small components and L the vertices in the large component. Let $q(n,p)$ be the probability that a given vertex v is in S. We claim that

$$\rho_{Y_1} - o(1) \leq q \leq \rho_{Y_2},$$

where $Y_1 \sim \mathrm{Bin}(n,p)$ and $Y_2 \sim \mathrm{Bin}(n-k^-,p)$, and as before ρ_Y is the extinction probability of the Galton-Watson process with offspring distribution Y.

For the upper bound we note that if $v \in S$ then the breadth first search process must stop before discovering k^- many vertices. In this case there are at most k^- many discovered vertices at any step of the process, and so the breadth first search process dominates the Galton-Watson process with offspring distribution $Y_2 \sim \mathrm{Bin}(n-k^-,p)$. Then the probability that the breadth first search process stops before discovering k^- many vertices is at most the probability that the Galton-Watson process with offspring distribution Y_2 dies before the tree grows to size k^-, which is definitely at most the probability that the Galton-Watson process dies at all, which is ρ_{Y_2}.

On the other hand the Galton-Watson process with offspring distribution $Y_1 \sim \mathrm{Bin}(n,p)$ dominates the breadth first search process and so the probability that the breadth first search process stops before discovering k^- many vertices is at least the probability that the Galton-Watson process dies before the tree grows to size k^-. This is clearly smaller than ρ_{Y_1}, but we will see in fact that the probability that the Galton-Watson process dies after the tree grows to size k^- is $o(1)$, and hence this probability is $\rho_{Y_1} - o(1)$.

Indeed, suppose the tree grows to size exactly $k \geq k^-$ before dying, and let $X_1, \ldots, X_k$ be the number of children of the k vertices in the tree. Then we have that $k = |T(k)| = |W(k) \cup A(k)| = 1 + \sum_{i=1}^k X_i$. However by the Chernoff bounds

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^k X_i \leq k\right) &= \mathbb{P}\Big(\mathrm{Bin}(kn,p) \leq k\Big) \\
&= \mathbb{P}\left(\mathrm{Bin}\Big(kn, \frac{c}{n}\Big) \leq ck - (c-1)k\right) \\
&\leq \exp\left(-\frac{(c-1)^2k^2}{2ck}\right).
\end{aligned}$$

However $k^- = \frac{16c}{(c-1)^2}\log n$ and so by summing the geometric series we have that

$$\begin{aligned}\mathbb{P}(\text{ Tree dies after growing to size } k^-) &\leq \sum_{k\geq k^-} \exp\left(-\frac{(c-1)^2 k}{2c}\right) \\ &\leq \frac{\exp\left(-\frac{(c-1)^2 k^-}{2c}\right)}{1-\exp\left(-\frac{(c-1)^2}{2c}\right)} \\ &= o(1).\end{aligned}$$

However since both Y_1 and Y_2 tend to $\mathrm{Po}(c)$ in distribution, it follows that

$$q = 1 - \beta_c + o(1).$$

In particular,

$$\mathbb{E}(|S|) = (1 - \beta_c + o(1))n.$$

It remains to show that $|S|$ is well-concentrated about its mean. Later in the lecture we will develop stronger tools which will make this calculation much simpler, but for now we will do a simple second moment argument. Let us suppose we wish to calculate

$$\mathbb{E}(|S|(|S|-1)) = \sum_u \mathbb{P}(u \in S) \sum_{v \neq u} \mathbb{P}(v \in S | u \in S)$$

If we condition not just on the fact that $u \in S$, but also on the component C_u of u, which by assumption has size $\leq k^-$, then

$$\sum_{v\neq u} \mathbb{P}(v \in S | u \in S, C_u) \leq k^- + \sum_{v \notin C_u} \mathbb{P}(v \in S | u \in S, C_u).$$

However, conditioned on the value of C_u, the rest of the graph is distributed as $G(n - |C_u|, p)$ and hence by the same arguments as before

$$\mathbb{P}(v \in S | u \in S, C_u) \leq \rho_{Y_2},$$

where as before $Y_2 \sim \mathrm{Bin}(n - k^-, p)$, since by assumption $|C_u| \leq k^-$.

However, as before $\rho_{Y_2} = 1 - \beta_c + o(1)$, and since this holds uniformly over all choices of C_u, by the law of total probability

$$\mathbb{E}(|S|(|S|-1)) \leq \mathbb{E}(|S|)\big(k^- + n\rho_{Y_2}\big) = (1+o(1))\mathbb{E}(|S|)^2.$$

Since $\mathbb{E}(|S|) = \Omega(n) = \omega(1)$, it follows that $\mathrm{Var}(S) = o\left(\mathbb{E}(|S|)^2\right)$ and so a standard application of Chebyshev's inequality gives that with high probability $|S| = (1+o(1)\mathbb{E}(|S|) = (1-\beta_c+o(1))n$. Hence with high probability $|L| = (1+o(1))\beta_c n$. □

5.4 Long Paths in the Super-Critical Phase

In the subcritical regime we saw that with high probability all the components were small, of logarithmic order, and moreover had a simple structure, almost all the vertices lie in tree components.

On the other hand, in the supercritical regime, we have seen that with high probability there is now a 'large' component L, covering a positive proportion of the vertices, and in fact this component can be seen to have a rich and complex internal structure.

One initial question that was considered in this context was how long is the longest path in L. Since we expect the vertices in $G(n,p)$ in this range to have constant degree, we should expect the diameter of the graph to be at least logarithmic, and in fact it can be shown that this is the case, so there will at least be a path of logarithmic length. However, since the component itself has linear size, there could potentially exist a path of much larger length.

In order to investigate this question we're going to consider a different graph exploration process to the one we considered in the previous section, this time looking at *depth first search*.

Given a graph G and an order on the vertices $V(G)$ we maintain an ordered list of vertices A the *stack* of *active* vertices and two sets of vertices W the *visited* vertices and U the *unvisited* vertices. Initially we set $A(0) = W(0) = \emptyset$ and $U(0) = V(G)$.

At each stage t in the algorithm we first look to see if $A(t)$ is empty. If it is then we move the smallest unvisited vertex u in U to A, that is

- $A(t+1) = (u)$;
- $U(t+1) = U(t) \setminus \{u\}$;
- $W(t+1) = W(t)$.

Otherwise, if $A(t)$ is non-empty we look at the first vertex in the stack $A(t) = (a_1, \dots, a_s)$, and we look through $U(t)$, in order according to our order on the vertices of G, and for each $u \in U(t)$ we check if $(a_1, u) \in E(G)$. The first time this happens, say for $u_1 \in U(t)$ we add u_1 to the top of the stack of active vertices, removing it from U. That is

- $A(t+1) = (u_1, a_1, \dots, a_s)$;
- $U(t+1) = U(t) \setminus \{u_1\}$;
- $W(t+1) = W(t)$.

Otherwise, if a_1 has no neighbours in $U(t)$, then we remove a_1 from A and add it to W, that is

- $A(t+1) = (a_2, \dots, a_s)$;
- $W(t+1) = W(t) \cup \{a_1\}$;
- $U(t+1) = U(t)$.

The algorithm terminates when A and U are both empty.

Let us note a few things about this algorithm. Firstly, at every stage A, U and W are disjoint, and form a partition of $V(G)$. Secondly, at each step of the algorithm one vertex moves, either from U to A or from A to W, hence the algorithm terminates in at most $2n$ steps. Thirdly, at any stage of the algorithm it had been revealed that there are no edges in G between U and W. Indeed, a vertex is only added to $W(t)$ when it has no neighbours in $U(t)$, and since $U(t)$ is non-increasing with t, this the remains true at later stages. Finally we note that the stack A always forms a path. Indeed, we only ever add a vertex a to the front of A if there is an edge between a and the first vertex of A and hence a augments the path spanned by A.

As before, if we run this algorithm on a random graph $G(n,p)$ we can consider these paramters A, U, W as random variables. Furthermore, again using the principal of deferred decisions, it will be useful to think of these random variables as coming from running the above algorithm using a sequence $(X_i \colon i \in \mathbb{N})$ of independent identically distributed $\mathrm{Ber}(p)$ random variables, where whenever we 'check' if an edge is in $E(G)$ we instead look at the 'next' X_i in our sequence and 'check' if $X_i = 1$. We will refer to each of these actions as a *query*.

Note that, every positive answer to a query results in a vertex being moved from U to A, and so after t queries we have that $|A \cup W| \geq \sum_{i=1}^{t} X_i$ (in fact, we get a strict inequality here, since each time A becomes empty we add a vertex to A without using a query). Conversely, since the addition of each vertex to A (after the first) is caused by a positive response to a query we have that after t queries $|A| \leq 1 + \sum_{i=1}^{t} X_i$.

We will need the following simple fact.

Lemma 5.10. *Let $\varepsilon > 0$, $(X_i \colon i \in \mathbb{N})$ be independent identically distributed Ber(p) random variables with $p = \frac{1+\varepsilon}{n}$ and let $N = \frac{\varepsilon n^2}{2}$. Then with high probability*

$$\left| \sum_{i=1}^{N} X_i - \frac{\varepsilon(1+\varepsilon)n}{2} \right| \leq n^{\frac{2}{3}}.$$

Proof. Since each $X_i \sim \mathrm{Ber}(p)$ it follows that $X := \sum_{i=1}^{N} X_i \sim \mathrm{Bin}(N,p)$, and so $\mathbb{E}(X) = Np = \frac{\varepsilon(1+\varepsilon)n}{2}$. Furthermore $\mathrm{Var}(X) = np(1-p) \leq n$ and so by Chebyshev's inequality

$$\mathbb{P}\left(\left| \sum_{i=1}^{N} X_i - \mathbb{E}(X) \right| \geq n^{\frac{2}{3}} \right) \leq \frac{\mathrm{Var}(X)}{n^{\frac{4}{3}}} = o(1).$$

□

The following theorem, asserting the existence of a linear path in the super-critical regime was initially a result of Atjai, Komlós and Szemerédi, although the proof we're giving based on the DFS process is due to Krivelevich and Sudakov.

Theorem 5.11. *Let $\varepsilon > 0$ be sufficiently small and let $p = \frac{1+\varepsilon}{n}$. Then with high probability $G(n,p)$ contains a path on at least $\frac{\varepsilon^2 n}{5}$ vertices.*

Proof. We will run the DFS algorithm described above on the graph $G(n,p)$, using the principle of deferred decisions to relate this to a sequence $(X_i \colon i \in \mathbb{N})$ of i.i.d $\mathrm{Ber}(p)$ random variables representing the answer to each query. Our aim is to show that, after the first $N := \frac{\varepsilon n^2}{2}$ queries the stack is large, of size at least $\frac{\varepsilon^2 n}{5}$.

By Lemma 5.10 we know that with high probability the sequence $(X_i \colon i \in \mathbb{N})$ is such that

$$\left| \sum_{i=1}^{N} X_i - \frac{\varepsilon(1+\varepsilon)n}{2} \right| \leq n^{\frac{2}{3}}.$$

and so we know approximately how many positive answers we get in the first N queries. Let us assume that this holds deterministically, and show that it follows that at some point during the process $|A|$ was large.

We first claim that after N queries there are not too many visited vertices. Indeed suppose that at this point $|W| \geq \frac{4\varepsilon}{7}n$, then there was some number of queries $t \leq N$ when $|W| = \frac{4\varepsilon}{7}n$. If at this point $|A| > \frac{\varepsilon^2}{5}n$ then we are done, and hence we may assume that $|A| < \frac{\varepsilon^2}{5}n$.

Since $V = A \cup U \cup W$ it follows that at this point

$$|U| > n(1 - \frac{4\varepsilon}{7} - \frac{\varepsilon^2}{5}).$$

However then the number of queries which the algorithm has made at this point is at least

$$t \geq |U||W| > n^2 \frac{4\varepsilon}{7}(1 - \frac{4\varepsilon}{7} - \frac{\varepsilon^2}{5}) > \frac{\varepsilon}{2}n^2 = N,$$

a contradiction.

Hence we may assume that after N queries $|W| < \frac{4\varepsilon}{7}n$. Suppose for a contradiction that at this point $|A| \leq \frac{\varepsilon^2 n}{5}$ and so in particular, since $V = A \cup U \cup W$, $U \neq \emptyset$ and so the algorithm has not yet terminated.

Since each positive query results in a vertex moving from U to A (which may later move to W) we have that

$$|W \cup A| \geq \sum_{i=1}^{N} X_i \geq \frac{\varepsilon(1+\varepsilon)n}{2} - n^{\frac{2}{3}}.$$

Hence, if $|A| \leq \frac{\varepsilon^2 n}{5}$ then

$$|W| \geq |W \cup A| - |A| \geq \frac{\varepsilon(1+\varepsilon)n}{2} - \frac{\varepsilon^2 n}{5} - n^{\frac{2}{3}} = \frac{\varepsilon n}{2} + \frac{3\varepsilon^2 n}{10} - n^{\frac{2}{3}}.$$

Furthermore

$$|U| = n - |W \cup A| \geq n - \frac{4\varepsilon}{7}n - \frac{\varepsilon^2 n}{5}.$$

However, since every edge between U and W has been queried at this point it follows that

$$\begin{aligned} N &\geq |U||W| \\ &\geq \big(n - \frac{4\varepsilon}{7}n - \frac{\varepsilon^2 n}{5}\big)\big(\frac{\varepsilon n}{2} + \frac{3\varepsilon^2 n}{10} - n^{\frac{2}{3}}\big) \\ &\geq \frac{\varepsilon n^2}{2} - \frac{2\varepsilon^2 n^2}{7} + \frac{3\varepsilon^2 n^2}{10} + O(\varepsilon^3 n^2) \\ &\geq \frac{\varepsilon n^2}{2} + \frac{\varepsilon^2 n^2}{70} + O(\varepsilon^3 n^2), \end{aligned}$$

contradicting our choice of $N = \frac{\epsilon n^2}{2}$.

Hence, at some point before the first N queries, $|A| \geq \frac{\epsilon^2 n}{5}$ and hence, since A spans a path, $G(n,p)$ must contain a path of at least this length. □

We note that it's not too hard to go from a long path to a long cycle.

Corollary 5.12. *Let $\varepsilon > 0$ be sufficiently small and let $p = \frac{1+\varepsilon}{n}$. Then with high probability $G(n,p)$ contains a cycle on at least $\frac{\varepsilon^2 n}{40}$ vertices.*

Proof. We will use the sprinkling method. Let us take $p_1 = \frac{1+\frac{\varepsilon}{2}}{n}$ and p_2 such that $p_1 + p_2 - p_1 p_2 = p$. Note in particular that $p_2 \geq \frac{\varepsilon}{2n}$.

By Theorem 5.11 with high probability there is a path P in $G(n,p_1)$ with at least $\frac{\varepsilon^2 n}{20}$ many vertices. Let F be the first $n^{\frac{2}{3}}$ many vertices on P and let L be the last $n^{\frac{2}{3}}$ many vertices on P.

Clearly, if there is any edge between F and L in $G(n,p_2)$, then $G(n,p)$ will contain a cycle on at least $|P| - 2n^{\frac{2}{3}} \geq \frac{\varepsilon^2 n}{40}$ many vertices.

However, there are $n^{\frac{4}{3}}$ many such edges and so the probability that there are no such edges in $G(n,p_2)$ is

$$\mathbb{P}\left(\mathrm{Bin}\left(n^{\frac{4}{3}}, p_2\right) = 0\right) = (1-p_2)^{n^{\frac{4}{3}}} \leq e^{-\frac{\varepsilon}{2n} n^{\frac{4}{3}}} \leq e^{-\varepsilon n^{\frac{1}{3}}} = o(1).$$

□

Note, if $c = 1 + \epsilon$, then the size of the giant component $|L| = (1 + o(1))\beta_c$ can be calculated to be $2\epsilon n + O(\epsilon^2 n)$, and so for sufficiently small ϵ, the path guaranteed by Theorem 5.11, whilst linear in a length, will still cover only a very small fraction of the giant component, and in fact it can be shown that the dependence on ϵ in Theorem 5.11 is optimal.

On the other hand, if $c \to \infty$, then again we can calculate that $\beta_c \to 1$, and so the giant component covers more and more of the vertex set of $G(n,p)$. Perhaps surprisingly, the same is true of the longest path - as c increases the longest path in L covers a larger and large proportion of L, and hence a larger and larger proportion of $G(n,p)$. We'll show the following in the exercises.

Theorem 5.13. *For every $\varepsilon > 0$ there is a $c > 0$ such that if $p = \frac{c}{n}$ then with high probability $G(n,p)$ contains a path or cycle of length $(1-\varepsilon)n$.*

So, as $c \to \infty$, if we take $p = \frac{c}{n}$ then the giant component starts to cover almost all of the vertices of $G(n,p)$. One might naively think that once $c = \omega(1)$ we will have covered everything, the largest component will be a spanning component, and $G(n,p)$ will be connected. However, as we will see, small local obstructions will force us to wait a bit longer until we start to see spanning subgraphs in $G(n,p)$. .

6 Spanning Subgraphs

6.1 Connectivity Threshold

There is a natural 'local' obstruction to a graph being connected, having a vertex of degree 0, an isolated vertex. On the example sheet we saw that the threshold for this property appears at $p = \frac{\log n}{n}$.

Lemma 6.1. *Let $p = \frac{\log n + C(n)}{n}$.*

1. *If $C(n) \to \infty$ then with high probability $\delta(G(n,p)) \geq 1$;*
2. *If $C(n) \to -\infty$ then with high probability $\delta(G(n,p)) = 0$;*

So, at the very least, if $p = o\left(\frac{\log n}{n}\right)$ then with high probability $G(n,p)$ will be disconnected, since it will contain an isolated vertex. However, it turns out that in fact this is also the threshold for connectivity. To see this we will need the following lemma.

Lemma 6.2. *Let $p = \frac{c \log n}{n}$ for some $c > \frac{1}{2}$. Then with high probability $G(n,p)$ contains no connected component of size between 2 and $\frac{n}{2}$.*

Proof. This will follow by a careful estimate of the expected number of such components. Let X_k be the numbrer of components of size k. Note that, if $A \subseteq [n]$ spans a connected component of $G(n,p)$ then two events must hold:

- $G(n,p)[A]$ contains a spanning tree;
- There are no edges between A and A^c in $G(n,p)$.

Furthermore, for a fixed A these two events are independent. We can estimate the probability of the former using Cayley's formula, that says the number of spanning trees of a complete graph on k vertices is k^{k-2}. Hence, if A has size k we see that

$$\mathbb{P}(A \text{ is a component of } G(n,p)) \leq k^{k-2} p^{k-1} (1-p)^{k(n-k)},$$

and so

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

We will show that $\sum_{k=2}^{\frac{n}{2}} \mu_k = o(1)$, and then Markov's lemma will imply the result claimed.

Using our standard inequalities that $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ and $(1-p) \leq e^{-p}$ we see that

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

Now, when k is small, say $2 \leq k \leq 5$ we have that

$$\begin{aligned}\mu_k &\leq \exp\Big(\log n + (k-1)\log\log n - pkn + O(1)\Big)\\ &\leq \exp\Big(\log n + (k-1)\log\log n - ck\log n + O(1))\Big)\\ &= O\left(\frac{(\log n)^{k-1}}{n^{ck-1}}\right) = o(1),\end{aligned}$$

since $ck - 1 > 0$.

On the other hand, since $k(n-k) \geq \frac{kn}{2}$ for $5 \leq k \leq \frac{n}{2}$ we have that

$$\begin{aligned}\mu_k &\leq \exp\Big(\log n + k + (k-1)\log c + (k-1)\log\log n - pk(n-k))\Big)\\ &\leq \exp\Big(\log n + 2k\log\log n - p\frac{kn}{2}\Big)\\ &= \exp\Big(\log n + 2k\log\log n - \frac{ck\log n}{2}\Big)\\ &= n\left(\frac{(\log n)^2}{n^{\frac{c}{2}}}\right)^k\\ &= o\left(\frac{1}{n}\right),\end{aligned}$$

since $\frac{ck}{2} > 2$.

So we can conclude that

$$\begin{aligned}\sum_{k=2}^{\frac{n}{2}} \mu_k &= \sum_{k=2}^{5} \mu_k + \sum_{k=5}^{\frac{n}{2}} \mu_k\\ &\leq \sum_{k=2}^{5} o(1) + \sum_{k=5}^{\frac{n}{2}} o\left(\frac{1}{n}\right)\\ &= o(1)\end{aligned}$$

□

Theorem 6.3. *Let* $p = \frac{\log n + C(n)}{n}$.

1. *If* $C(n) \to -\infty$ *then with high probability* $G(n,p)$ *is not connected;*
2. *If* $C(n) \to \infty$ *then with high probability* $G(n,p)$ *is connected.*

In particular, $\hat{p} = \frac{\log n}{n}$ *is a sharp threshold for* $G(n,p)$ *being connected.*

Proof. If $C(n) \to -\infty$ then by Lemma 6.1 it follows that with high probability $G(n,p)$ contains an isolated vertex, and hence is not connected.

Conversely, suppose that $C(n) \to \infty$. Since being connected is a monotone property, we may assume that $C(n) = o(\log n)$ and hence $p = \frac{c\log n}{n}$ where $c := c(n) = 1 + o(1)$. Hence, by Lemma

6.2 we have that

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

where X_k is the number of components of size k. Furthermore, by Lemma 6.1 we also have that $\mathbb{P}(X_1 > 0) = o(1)$. Also we note that $G(n,p)$ is disconected if and only if it has some component of size at most $\frac{n}{2}$. Hence by the union bound

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

It follows that with high probability $G(n,p)$ is connected. □

Hence, the threshold for connectivity really closely coincides with the threshold for having no isolated vertex. It turns out that this is not just a coincidence, in fact the two thresholds are very delicately related. Indeed, we will show that, in the random graph process, with high probability the moment that the graph becomes connected is precisely the moment when the last isolated vertex disappears.

To make this precise let us consider the following two hitting times:

- $\tilde{m}_1 = \min\{m \colon \delta(G(m)) \geq 1\}$;
- $\tilde{m}_c = \min\{m \colon G(m) \text{ is connected } \}$.

Since, as noted, whenever G is connected we have that $\delta(G) \geq 1$, it follows that $\tilde{m}_c \geq \tilde{m}_1$. We shall show that with high probability the converse also holds.

Theorem 6.4. *In the random graph process* $\left(G(m) \colon m \in \left[\binom{n}{2}\right]\right)$ *with high probability* $\tilde{m}_c = \tilde{m}_1$.

Proof. Let us consider the following two times in the random graph process:

- $m_1 = \frac{(n-1)}{2}(\log n - \log\log n)$;
- $m_2 = \frac{(n-1)}{2}(\log n + \log\log n)$.

During the proof we will show that a (finite) number of events hold whp, and we will show that if all of the events occur, then $\tilde{m}_c = \tilde{m}_1$.

By Lemma 6.1 and Theorem 4.5 with high probability $m_1 < \tilde{m}_1 < m_2$, and we call this event A_1. In other words, whp $G(m_1)$ has at least one isolated vertex and $G(m_2)$ has no isolted vertices. Let us write I for the set of isolated vertices in $G(m_1)$.

Furthermore, since $\frac{m_1}{\binom{n}{2}} \approx \frac{\log n}{n}$, it follows from Lemma 6.2 and Theorem 4.5 that with high probability $G(m_1)$ has no connected components of size between 2 and $\frac{n}{2}$, and we call this event A_2. In particular, if A_2 holds, then there is a unique component of $G(m_1)$ which is not an isolated vertex, which we denote by L.

We can estimate the number $|I|$ of isolated vertices in $G(m_1)$. This will be a little easier to calculate in $G(n,p_1)$ where $p_1 = \frac{m_1}{\binom{n}{2}} = \frac{\log n - \log\log n}{n}$, and the result will carry over to $G(m_1)$ via Theorem 4.5.

In this case, for an arbitrary $v \in [n]$

$$\begin{aligned}\mathbb{P}(v \text{ is isolated in } G(n,p_1)) &= (1-p_1)^{n-1}\\ &\leq \frac{e^{-p_1 n}}{1-p_1}\\ &= \frac{e^{\log n - \log\log n}}{1-p_1}\\ &\leq 2\frac{\log n}{n}.\end{aligned}$$

Hence the expected number of isolated vertices in $G(n,p_1)$ is at most $2\log n$. Then, by Markov's inequality, with high probability the number of isolated vertices in $G(n,p_1)$ is at most $(\log n)^2$ and it follows from Theorem 4.5 that the same holds in $G(m_1)$. That is, whp $|I| \leq (\log n)^2$, and we call this event A_3.

Now, if A_1 holds, then whp as the random graph process evolved from $G(m_1)$ to $G(m_2)$, for each of the vertices $i \in I$, we added an edge e_i adjacent to that vertex. If A_2 holds, then $|I|$ is small, and we will show that it is unlikely that any of the $m_2 - m_1$ edges that were added in this period go between vertices of I. In this case, if A_3 holds, then each edge e_i must be incident with a vertex in L (since L covers the vertices in $[n] \setminus I$. It will follow that whp at time $\tilde{m}_1$, each of the vertices in I has a neighbour in L, and hence $G(\tilde{m}_1)$ is connected.

So, there are precisely $m_2 m_1 = (n-1)\log\log n$ many edges in $G(m_2) \setminus G(m_1)$ and if we condition on the event A_3 then $|I| \leq (\log n)^2$ and by the union bound the probability that one of these edges lie completely in I is at most

$$(n-1)\log\log n \cdot \frac{(\log n)^4}{\binom{n}{2} - m_2} \leq \frac{4n\log\log n(\log n)^4}{n^2} = o(1).$$

So whp, conditioned on A_3, none of the edges in $G(m_2) \setminus G(m_1)$ lie in I. Let us call this A_4.

We have seen that if A_1, A_2, A_3 and A_4 occur then $\tilde{m}_1 = \tilde{m}_c$ and hence

$$\begin{aligned}\mathbb{P}(\tilde{m}_1 = \tilde{m}_c) &\geq \mathbb{P}\left(\bigcap_{i=1}^{4} A_i\right)\\ &\geq 1 - \mathbb{P}\left(\bigcup A_i^c\right)\\ &\geq 1 - \mathbb{P}(A_1^c) + \mathbb{P}(A_2^c) + \mathbb{P}(A_3^c) + \mathbb{P}(A_4^c \cap A_3)\\ &\geq 1 - \mathbb{P}(A_1^c) + \mathbb{P}(A_2^c) + \mathbb{P}(A_3^c) + \mathbb{P}(A_4^c | A_3)\\ &\geq 1 - o(1).\end{aligned}$$

□

6.2 Matching thresholds

Another simple spanning structure, whose appearance in $G(n,p)$ we can consider is that of a *perfect matching*. Recall, a *matching* in a graph G is an independent set of edges, one meeting

each vertex at most once. A matching is *perfect* if it meets each vertex if $|V(G)|$ is even, and all but one vertex if $|V(G)|$ is odd.

Since it is 'easier' to find matchings in bipartite graphs (c.f Hall's theorem), we will start by considering the bipartite case. That is, we will denote by $G(n,n,p)$ the random bipartite graph on two partition clases V_1 and V_2 of size n, where each edge is included independently with probability p from the set $V_1 \times V_2$.

We will need to use the following well-known theorem that gives a necessary and sufficient condition for the existence of a matching.

Theorem 6.5 (Hall's Marriage Theorem)**.** *Let G be a bipartite graph on vertex classes V_1 and V_2. Then G has a matching meeting every vertex in V_1 if and only if for every subset $S \subseteq V_1$ we have*

$$|S| \leq |N_G(S)|.$$

Theorem 6.6. *Let $p = \frac{\log n + C(n)}{n}$. Then*

1. *If $C(n) \to \infty$ then with high probability $G(n,n,p)$ has a perfect matching;*
2. *If $C(n) \to -\infty$ then with high probability $G(n,n,p)$ doesn't have a perfect matching.*

Proof. If $C(n) \to -\infty$ then, by similar arguments as before, $G(n,n,p)$ will have an isolated vertex with high probability. Indeed, for $v, v' \in V_1$ the events that $d(v) = 0$ and $d(v') = 0$ are independent and so that number of vertices of degree 0 in V_1 is distributed as $\mathrm{Bin}(n,q)$ where $q = \mathbb{P}(d(v) = 0) = (1-p)^n$. Therefore since

$$\mathbb{E}\left(\mathrm{Bin}(n,q)\right) = n(1-p)^n \geq ne^{-\frac{pn}{1-p}} = ne^{-np+O(np^2)} = \omega(1).$$

it follows from Lemma 5.2 that with high probability there is at least one isolated vertex in V_1.

So, suppose that $C(n) \to \infty$. Since containing a perfect matching is an increasing event, we may assume that $C(n) \leq \log n$, and so $p \leq \frac{2\log n}{n}$. If $G(n,n,p)$ does not contain a perfect matching then by Hall's Theorem there is some subset $S \subseteq V_i$ for $i \in \{1,2\}$ such that $|N_G(S)| < |S|$. Let us take such a subset with $|S|$ as small as possible. We note that

1. $|S| = |N(S)| + 1$;
2. $|S| \leq \lceil \frac{n}{2} \rceil$;
3. Every $w \in N(S)$ has at least two neighbours in S.

Indeed, the first is clear, since otherwise we could remove any vertex from S and find a smaller subset violating the condition. For the second we note that if this is not the case, then we can replace S with $V_{3-i} \setminus N(S)$ and we would get a smaller set than S which still violates the condition. Finally, if $w \in N(S)$ has only one neighbour x in S then $S \setminus \{x\}$ also violates the condition.

Now, if $|S| = 1$ then $S = \{x\}$ is an isolated vertex. However, since $C(n) \to \infty$ we may assume, again via similar calculations as before, that with high probability there are no isolated vertices.

If $|S| = 2$ then $S = \{x, y\}$ and x and y are joined only to some common neighbour z. By the union the bound the probability that such a pair exists is at most

$$2\binom{n}{2}np^2(1-p)^{2(n-1)} = O\left(n^3p^2e^{-2pn}\right) = n^3\left(\frac{2\log n}{n}\right)^2\frac{1}{n^2} = o(1).$$

Let us denote by $\mathcal{E}$ the event that there is a minimal such set S of size s at least 3. Since S is minimal it satisfies conditions 1., 2. and 3. and hence by the union bound

$$\begin{aligned}
\mathbb{P}(\mathcal{E}) &\leq \sum_{s=3}^{\lceil \frac{n}{2} \rceil} 2\binom{n}{s}\binom{n}{s-1}\binom{s}{2}^{s-1}(p^2)^{s-1}(1-p)^{s(n-s+1)} \\
&\leq \sum_{s=3}^{\lceil \frac{n}{2} \rceil} 2\left(\frac{en}{s}\right)^s\left(\frac{en}{s-1}\right)^{s-1}s^{2(s-1)}\left(\frac{2\log n}{n}\right)^{2(s-1)}e^{-\frac{spn}{2}} \\
&\leq 2n\sum_{s=3}^{\lceil \frac{n}{2} \rceil}(2e\log n)^{2(s-1)}e^{-\frac{s\log n}{2}} \\
&\leq 2n\sum_{s=3}^{\lceil \frac{n}{2} \rceil}\left(\frac{(\log n)^2}{\sqrt{n}}\right)^s \\
&\leq O\left(\frac{(\log n)^6}{\sqrt{n}}\right) = o(1).
\end{aligned}$$

The first line in the above follows as we have $2\binom{n}{s}$ potential choices for the set S, $\binom{n}{s-1}$ choices for $N_G(S)$ and then each vertex in $N(S)$ has at least 2 neighbours inside S (by property 3.) and each vertex in S has no neighbours outside of $N(S)$. The rest is just standard approximations.

Hence with high probability there is no minimal such set S and so with high probability $G(n, n, p)$ has a perfect matching. □

So, what about in the non-bipartite case? One could try and adapt the proof above using Tutte's Theorem rather than Hall's Theorem as a criterion for the existence of a matching, and a proof along these lines was given by Erdős and Rényi. Instead we will follow a proof of Łuczak and Ruciński.

The rough idea is as follows - if $p \approx \frac{\log n}{n}$ then p is large enough that whp any two *large* sets in $G(n, p)$ are connected by an edge, but on the other hand p is small enough that we don't expect *small* sets to be very dense.

Let us fix some bipartiton $A \cup B$ of $[n]$. On average, each vertex in A will have many neighbours in B, and vice versa. If this were to be true for all vertices, then we could find a perfect matching using Hall's condition.

Indeed, if a *small* subset $S \subseteq A$ violated Hall's condition, then the set $S \cup N(S)$ would be small, but contain many edges since all vertices in S have many neighbours in B, contradicting our assumptions.

However, if $S \subseteq A$ is *large* then, since there are no edges between S and $B \setminus N(S)$, $B \setminus N(S)$ must be small, and a similar argument shows that this cannot happen.

However, for any bipartiton $A \cup B$ of $[n]$ there will likely be some *bad* vertices, with few neighbours in the other partition class, however these should be relatively rare. Since they are relatively rare, we will be able to show that whp there is only a small number of such vertices, and moreover that they are not very close to each other. It will follow that we can greedily match these vertices to neighbours, and in the remaining graph each vertex will still have many unmatched neighbours in the opposite partition class, and we can finish the argument as above. It will turn out that in order to make the above precise, we will have to deal quite carefully with vertices of very low degree, which we call *small*.

Let us start then by showing the first property.

Lemma 6.7. *Let $p = \Theta\left(\frac{\log n}{n}\right)$ and let $u = n\frac{(\log\log n)^2}{\log n}$. Then with high probability in $G(n,p)$*

1. *There is an edge between every pair of disjoint subsets of vertices of size at least u; and*
2. *Every set S of vertices of size at most $2u$ is such that $G[S]$ contains fewer than $(\log\log n)^3|S|$ many edges.*

Proof. Let us suppose that that $c_1\frac{\log n}{n} \leq p \leq c_2\frac{\log n}{n}$.

For the first property we note that if there are two sets of vertices of size $\geq u$ with no edges between them, then there are two sets of size exactly u with no edges between them. Hence if we let $\mathcal{E}$ be the event that the first condition does not hold, a first moment calculation shows that

$$\begin{aligned}
\mathbb{P}(\mathcal{E}) &\leq \binom{n}{u}^2 (1-p)^{u^2} \\
&\leq \left(\frac{en}{u}\right)^{2u} e^{-pu^2} \\
&= (\log n)^{2n\frac{(\log\log n)^2}{\log n}}\, e^{-c_1 n\frac{(\log\log n)^4}{\log n}} \\
&= \exp\left(2n\frac{(\log\log n)^3}{\log n} - c_1 n\frac{(\log\log n)^4}{\log n}\right) \\
&\leq \exp\left(-\Omega\left(n\frac{(\log\log n)^4}{\log n}\right)\right) \\
&= o(1)
\end{aligned}$$

Similarly, if we let $\mathcal{E}'$ be the event that the second condition doesn't hold, then

$$\begin{aligned}
\mathbb{P}(\mathcal{E}') &\leq \sum_{s=1}^{2u} \binom{n}{s} \binom{\binom{s}{2}}{s(\log\log n)^3} p^{(\log\log n)^3 s} \\
&\leq \sum_{s\leq 2u} \left(\frac{en}{s}\right)^s \left(\frac{es^2}{s(\log\log n)^3}\right)^{s(\log\log n)^3} p^{(\log\log n)^3 s} \\
&\leq \sum_{s\leq 2u} \left(\frac{c_2 e^2 \log n}{(\log\log n)^3}\right)^s \left(\frac{2eup}{(\log\log n)^3}\right)^{s((\log\log n)^3-1)} \\
&\leq \sum_{s\leq 2u} (\log n)^s \left(\frac{2c_2 e}{\log\log n}\right)^{s((\log\log n)^3-1)} \\
&\leq \sum_{s\leq 2u} e^{s\log\log n} \left(\frac{2c_2 e}{\log\log n}\right)^{s\log\log n} \\
&\leq \sum_{s\leq 2u} \left(\left(\frac{2c_2 e^2}{\log\log n}\right)^{\log\log n}\right)^s \\
&= o(1)
\end{aligned}$$

□

We will see then, if the conclusion of Lemma 6.7 holds and we have two sets A and B of equal size in $G(n,p)$ where each vertex has many neighbours in the other set, then we can find a matching covering $A \cup B$.

Lemma 6.8. *Let $C > 0$ and let H be a bipartite graph on vertex sets $A, B \subseteq [n]$ with $|A| = |B|$ which satisfies the conclusions of Lemma 6.7 and with minimum degree $\delta(H) \geq c\log n$. Then H contains a perfect matching.*

Proof. Let $u = n\frac{(\log\log n)^2}{\log n}$ be as in Lemma 6.7. Let us suppose for a contradiction that H has no perfect matching, then by Hall's Theorem there is some subset $S \subseteq A$ such that $|N_H(S)| < |S|$. We will split into a few cases.

Firstly, if $|S| \leq u$, then $|S \cup N_H(S)| < 2u$, but the number of edges of $G(n,p)$ in $S \cup N_H(S)$ is at least $|S|c\log n$, since H has minimum degree at least $c\log n$. However, $c\log n \gg (\log\log n)^3$, and so this contradicts Property 2. of Lemma 6.7.

Secondly, if $|N_H(S)| \geq a-u$, then consider $B\setminus N_H(S) := U$. Then $|U| \leq u$ and $N_H(U) \subseteq A\setminus S$ and so $|N_H(U)| \leq |A \setminus S| = a - |S| < a - |N_H(S)| = |U|$. However then as in the first case we have that $|U \cup N_H(U)| < 2u$ and the number of edges of $G(n,p)$ in $U \cup N_H(U)$ is at least $|U|c\log n$, a contradiction.

Hence, we may assume that $|S| > u$ and $|N_H(S)| < a - u$. However then $|B \setminus N_H(S)| \geq u$, and there are no edges between S and $B \setminus N_H(S)$, contradicting Property 1. of Lemma 6.7. □

Using these two lemmas we can complete the proof that the perfect matching threshold coincides with the connectivity threshold.

Theorem 6.9. *Suppose that n is even and let $p = \frac{\log n + C(n)}{n}$. Then*

1. *If $C(n) \to -\infty$ then with high probability $G(n,p)$ doesn't have a perfect matching;*
2. *If $C(n) \to \infty$ then with high probability $G(n,p)$ has a perfect matching.*

Proof. The first statement follows from Theorem 6.1, since if $C(n) \to -\infty$ then with high probability there is some isolated vertex, and hence no perfect matching.

So, let us assume that $C(n) \to \infty$ and note that, since containing a matching is an increasing property, we may assume that $C(n) = o(\log n)$. Let us fix some arbitrary bipartition $[n] = A \cup B$ where $|A| = |B| = \frac{n}{2}$.

We call a vertex v *bad* if it has fewer than $\frac{\log n}{200}$ neighbours in either A or B. If we let X denote the number of bad vertices then we see by Lemma 5.2

$$\begin{aligned}
\mathbb{E}(X) &\leq 2\mathbb{P}\left(\mathrm{Bin}\left(\frac{n}{2}, p\right) < \frac{\log n}{200}\right) \\
&\leq 2n \exp\left(-\frac{\left(\frac{pn}{2} - \frac{\log n}{200}\right)^2}{2pn}\right) \\
&\leq 2n \exp\left(-\frac{1}{2}\left(\frac{1}{2} - \frac{1}{200} + o(1)\right)^2 \log n\right) \\
&\leq 2n^{1-0.11} \leq n^{0.89}
\end{aligned}$$

and hence by Markov's inequality

$$\mathbb{P}\left(X \geq n^{\frac{9}{10}}\right) = o(1).$$

Also let us call a vertex *small* if it has degree at most 20. We will need the following two claims. The first shows that with high probability the bad and small vertices are far apart in $G(n,p)$.

Claim 6.10. For every fixed k, with high probability $G(n,p)$ does not contain a tree T with k vertices such that either

- at least two vertices of T are small; or
- at least ten vertices of T are bad.

Proof of claim. If we let $Z \sim \mathrm{Bin}(n-k, p)$, then for any fixed integer z

$$\begin{aligned}
\mathbb{P}(Z \leq z) &\leq z \cdot \mathbb{P}(Z = z) = O\left(\binom{n}{z} p^z (1-p)^{n-z}\right) \\
&= O\left((np)^z e^{-np}\right) = O\left(\frac{(\log n)^z}{n}\right) = o\left(n^{-\frac{2}{3}}\right).
\end{aligned}$$

Hence, by a union bound, the expected number of tree T with k vertices and at least two small vertices, which we denote by Y_k, satisfies

$$\mathbb{E}(Y_k) \leq \binom{n}{k} k^{k-2} \binom{k}{2} p^{k-1} \mathbb{P}(Z \leq 19)^2 = O\left(n^k p^{k-1} n^{-\frac{4}{3}}\right) = o(1).$$

Indeed, if there is such a tree T with two small vertices x and y, all the edges of T must appear in $G(n,p)$ and at most 19 of the edges from x and y to $[n] \setminus V(T)$ can appear in $G(n,p)$.

Simlarly the expected number of trees with at least ten bad vertices, which we denote by Y'_k, can be seen to satisfy

$$\mathbb{E}(Y'_k) \leq \binom{n}{k} k^{k-2} \binom{k}{10} p^{k-1} \mathbb{P}\left(\mathrm{Bin}\left(\frac{n}{2} - k + 1, p\right) \leq \frac{\log n}{200}\right)^{10} = O\left(n^k p^{k-1} n^{-(0.11\times 10)}\right) = o(1).$$

The claim then follows as a simple consquence of Markov's inequality. □

The second claim bounds the likely maximum degree of $G(n,p)$.

Claim 6.11. With high probability $\Delta(G(n,p)) \leq 8 \log n$.

Proof of claim. This is a simple application of the Chernoff bounds.

$$\begin{aligned}
\mathbb{P}(\Delta(G(n,p)) \geq 8\log n) &\leq n\mathbb{P}(\mathrm{Bin}(n,p) > 8\log n) \\
&\leq n \exp\left(-\frac{(8\log n - np)^2}{2(np + \frac{8\log n}{3})}\right) \\
&\leq n \exp\left(-\frac{(6\log n)^2}{8\log n}\right) \\
&\leq n^{-2} = o(1).
\end{aligned}$$

□

So, we have a whole host of properties that hold with high probability, and so with high probability the following is true in $G(n,p)$:

- The conclusion of Lemma 6.8 holds with $c = \frac{1}{300}$;
- Claim 6.10 holds for all $k \leq 21$;
- $\Delta(G(n,p)) \leq 8\log n$;
- $\delta(G(n,p)) \geq 1$; and
- The number of bad vertices is at most $n^{\frac{9}{10}}$.

Let us order the bad vertices $v_1, \ldots, v_X$ such that $d(v_i) \leq d(v_{i+1})$. We will match these vertices greedily to some $u_1, \ldots, u_x$. Suppose that $v_1, \ldots, v_{i-1}$ have already been matched to $u_1, \ldots, u_{i-1}$. and let $V_{i-1} = \{v_1, u_1, v_2, u_2, \ldots, v_{i-1}, u_{i-1}\}$.

If $v_i \in V_{i-1}$, then we've already matched v_i to some v_j with $j < i$. Otherwise we split into two cases.

Firstly, if $d(v_i) \leq 20$. We note that v_i cannot have any neighbours in V_{i-1}, since if it did it would lie at distance ≤ 2 from some v_j with $j < i$. However then $d(v_j) \leq 20$ as well, and hence

there is some tree of size at most 3 with two small vertices, a contradiction. So, in this case, since $d(v_i) \geq \delta(G(n,p)) \geq 1$, we can match v_i arbitrarily to any one of its neighbours u_i.

Otherwise, $d(v_i) \geq 20$. Now, v_i cannot have ≥ 18 neighbours in V_{i-1}, since otherwise it would be at distance ≤ 2 from at least nine different bad vertices. It is an easy check that there is then a tree of size at most 19 containing ten bad vertices, a contradiction. Hence v_i has at least one neighbour not in V_{i-1} and we match it to this neighbour.

Furthermore, removing V_X from the graph can change the degree of each other vertex by at most 20. Indeed, if v is adjacent to more than 20 vertices in V_x, then it is at distance two from at least 10 different bad vertices, and again we can form a tree of size at most 21 containing 10 bad vertices.

Since every bad vertex is in V_x, every vertex in $G \setminus V_x$ has at least $\frac{\log n}{200} - 20$ neighbours in both $A' = A \setminus V_x$ and $B' = B \setminus V_x$. We would like to apply Lemma 6.8 to tthe bipartite subgraph of $G(n,p)$ between A' and B', however it is no longer necessarily balanced. Nevertheless, since all the vertices have large minimum degree into both sides of the bipartition, we might hope to 'swap' some vertices from A' to B'.

Let us suppose, without loss of generality, that $|A'| \geq |B'|$. Since $|A| = |B|$, in order to balance the bipartition we might need to move as many as $|V_X| = X \leq n^{\frac{9}{10}}$ many vertices from A' to B', which could ruin our minimum degree condition. However, suppose we can choose a set $S \subseteq A'$ of size $|S| = \frac{|A'|-|B'|}{2}$ such that for every $v_1, v_2 \in S$, $d(v_1, v_2) > 2$.

Then, for any $v \in [n] \setminus V_X$, v is adjacent to at most one vertex in S. Hence, if we consider the bipartite subgraph of $G \setminus V_X$ with partition sets $A' \setminus S$ and $B' \cup S$ we have that $|A' \setminus S| = |B' \cup S|$ and the degree of each vertex to the other partition class is at least

$$\frac{\log n}{200} - 20 - 1 \geq \frac{\log n}{300},$$

and so by our assumptions and Lemma 6.8, $G(n,p)$ contains a matching between $A' \setminus S$ and $B' \cup S$. Together with the matching on V_X, this gives us a complete matching on $G(n,p)$.

It remains then to find such a set S. To do so let us consider an auxilliary graph H where $V(H) = A'$ and $E(H) = \{(u,v) \colon d(u,v) \leq 2\}$. We note that, since $\Delta(G(n,p)) \leq 8 \log n$, it follows that $\Delta(H) \leq 64(\log n)^2$. However it then follows that

$$\alpha(H) \geq \frac{v(H)}{\Delta(H)+1} \geq \frac{\frac{n}{2} - X}{64(\log n)^2 + 1} = \omega\left(n^{\frac{9}{10}}\right).$$

Hence H contains an independent set S of size $\frac{|A'|-|B'|}{2} \leq X \leq n^{\frac{9}{10}}$, which is as required. □

6.3 Hamiltonicity threshold

So, what might we expect the threshold for containing a Hamiltonian path or cycle to be? Well, in the case of connectivity an obvious necessary condition was to have no isolated vertices. Similarly, if $G(n,p)$ has a Hamiltonian cycle then clearly each vertex has degree at least two, and so the threshold for Hamiltonicity must be at least the threshold for having minimum degree at least two. Recall that we proved on the example sheet the following:

Lemma 6.12. *Let* $p = \frac{\log n + \log\log n + C(n)}{n}$.

1. *If* $C(n) \to \infty$ *then with high probability* $\delta(G(n,p)) \geq 2$*;*
2. *If* $C(n) \to -\infty$ *then with high probability* $\delta(G(n,p)) \leq 1$.

In particular, if $C(n) \to -\infty$ then with high probability $G(n,p)$ will not contain a Hamiltonian cycle. In fact, moreover, the proof shows that if $C(n) \to -\infty$ then with high probability the number of vertices of degree at most one will tend to infinity, and so there cannot even be a Hamiltonian path. However, again perhaps surprisingly, once we are guaranteed to have degree at least two everywhere, the graph will with high probability be Hamiltonian.

Theorem 6.13. *Let* $p = \frac{\log n + \log\log n + C(n)}{n}$.

1. *If* $C(n) \to \infty$ *then with high probability* $G(n,p)$ *contains a Hamiltonian cycle;*
2. *If* $C(n) \to -\infty$ *then with high probability* $G(n,p)$ *does not contain a Hamiltonian path.*

In fact, a similar result as Theorem 6.4 holds; the hitting time for having minimum degree ≥ 2 is exactly the same as the hitting time for containing a Hamiltonian cycle. A key idea in the proof of Theorem 6.13 will be that of a booster.

We say that a non-edge $(u,v) \notin E(G)$ is a *booster* if its addition increases the length of the longest path in G (or makes G Hamiltonian). That is, (u,v) is a booster if either $G \cup \{(u,v)\}$ contains a strictly longer path than G, or $G \cup \{(u,v)\}$ is Hamiltonian. In particular, if G is already Hamiltonian, then every non-edge $(u,v) \notin E(G)$ is a booster.

Crucially, given any graph G, if we add a sequence of n boosters to G one-by-one (note that the set of boosters at each step might be different), then the resulting graph must be Hamiltonian.

Finding boosters will be key to our strategy for building a Hamiltonian cycle in $G(n,p)$, but how do we find boosters?

Pósa's rotation-extension technique

Our main tool will comes from an idea of Pósa, which he initially used to show that $G\left(n, \frac{c\log n}{n}\right)$ is Hamiltonian for c sufficiently large, and which has been incredibly useful for studying questions about Hamiltonicity.

We start with the following simple observation: If $P = (v_0, \dots, v_\ell)$ is a longest path in a connected graph G which is not Hamiltonian, then $e = (v_0, v_\ell)$ is a booster. Indeed, if P is a Hamiltonian path, then $C = P + e$ is a Hamiltonian cycle, and hence e is a booster. Otherwise, there is at least one vertex $w \in V(G) \setminus V(C)$. However, since G is connected there is a path from w to C in G, let P' be a shortest such path. Then, if $v_i = V(P') \cap V(C)$, the path

$$wP'v_iv_{i+1}\dots v_\ell v_0 v_1 \dots v_{i-1}$$

in $G+e$ has length $\geq \ell+1$. Hence, the longest path in $G+e$ is longer than the longest path in G, and so e is a booster.

So, each longest path in a graph will give us a booster (although many different paths might have the same endpoints and lead to the same booster). Pósa's key idea was to find a way to find many boosters from a single path by *rotating* it.

More precisely, suppose that $P=(v_0,\dots,v_\ell)$ is a longest path in a connected graph G which is not Hamiltonian. Then, by the above $(v_\ell,v_0)\notin E(G)$ is a booster. However, each edge $e\in E(G)$ which is adjacent to v_ℓ must have its other endpoint in P, since otherwise $P+e$ would be a longer path. Then, if say $e=(v_\ell,v_i)$ with $1<i<\ell$, we can *rotate* P at v_i by adding e and deleting (v_i,v_{i+1}) to obtain a new longest path P' (which has its own booster, (v_0,v_{i+1})). We call P' a *rotation* of P around v_i.

Not only do we find one such path P' for each edge adjacent to v_ℓ, we can also apply the same reasoning to P' and to find more longest paths and so more boosters.

For the next few lemmas let us fix a longest path P in G and let $\mathcal{Q}$ be the set of paths we can obtain from P by performing an arbitrary sequence of rotations, with v_0 fixed. Let us denote by R the set of endpoints of paths in $\mathcal{Q}$. Furthermore, let us write

$$R^+=\{v_i\colon v_{i-1}\in R\}\quad\text{and}\quad R^-=\{v_i\colon v_{i+1}\in R\}$$

and for any subset $U\subseteq V(G)$ let us write $\bar{N}(U)$ for the *external neighbourhood* of U, that is,

$$\bar{N}(U)=\{w\in V(G)\setminus U\colon (u,w)\in E(G)\text{ for some }u\in U\}.$$

Lemma 6.14. *[Pósa] Let G be a connected graph and P, $\mathcal{Q}$, R be as above. Then $\bar{N}(R)\subseteq R^+\cup R^-$.*

Proof. We first not that any vertex in $\bar{N}(R)$ must lie in $V(P)$. Indeed, if $v\in R$ then there is some v_0-v-path $Q\in\mathcal{Q}$ such that $V(P)=V(Q)$. However, since P is a longest path in G, so is Q, and hence each neighbour of v lies in $V(Q)=V(P)$.

Suppose then that some $v\in R$ has a neighbour $u\in V(P)$ and that $u\notin(R\cup R^+\cup R^-)$. Then, $u\in V(P)\setminus R$, and so u has two neighbours in each $Q\in\mathcal{Q}$. We claim that u in fact has the same neighbours in each Q. Indeed, suppose that Q is obtained from P via a sequence of rotations and Q does not contain both edges adjacent to u in P. Then one of the rotations caused one of these edges to be deleted. However, if an edge (u,x) is deleted in an extension then one of u or x is the endpoint of the new path, and hence is in R. It follows that either u is in R, u is in $R^+\cup R^-$, contradicting our assumption.

Consider then the path $Q\in\mathcal{Q}$ which has v as an endpoint. If $u=v_i$, then by the above, u is adjacent to v_{i-1} and v_{i+1} in Q. Then, since $(u,v)\in E(G)$ we can rotate Q around u to get a new path $Q'\in\mathcal{Q}$ in which either v_{i+1} or v_{i-1} is an endpoint (note, whilst u is adjacent to v_{i-1} and v_{i+1} on Q, it is not clear what order they appear in, as rotations will reverse segments of the path). However then either v_{i+1} or $v_{i-1}\in R$ and so $u\in R^-\cup R^+$, contradicting our assumption. □

Corollary 6.15. *Let G be a connected graph and P, $\mathcal{Q}$, R be as above. Then $|\bar{N}(R)|\leq 2|R|-1$.*

Proof. Clearly $|R^-| \leq |R|$ and also since $v_\ell \in R$ it follows that $|R^+| \leq |R| - 1$. □

Hence this set R has bad 'expansion' properties, its neighbourhood is not too much bigger than itself. However, as we will show, with high probability all small subsets of $G(n,p)$ will have good expansion properties, and hence R cannot be too small, and so in particular in $G(n,p)$ there will be many boosters. Formally the way we will phrase these 'expansion' properties will be to talk about expander graphs.

Given $k \in \mathbb{N}$ and $t > 0$ we say a graph G is a (k,t)-*expander* if $|\bar{N}(U)| \geq t|U|$ for every set $U \subseteq |V(G)|$ with $|U| \leq k$.

Lemma 6.16. *Suppose G is a $(k,2)$-expander which is connected and non-Hamiltonian. Then G has at least $\frac{(k+1)^2}{2}$ boosters.*

Proof. Let $P = (v_0, \dots, v_\ell)$ be a longest path in G and let $\mathcal{Q}$, R be as above. Since G is connected and non-hamiltonian, by our observation at the start of this section, (v_0, v) is a booster for every $v \in R$. However, by Corollary 6.15 $|\bar{N}(R)| \leq 2|R| - 1$ and hence, since G is $(k,2)$-expander, $|R| \geq k + 1$.

Therefore, there is a set B_{v_0} of at least $|R| \geq k + 1$ many boosters, each of the form (v_0, v). However, the path P was an arbitrary longest path in G and hence the same argument applies to each path $Q \in \mathcal{Q}$. In particular, for each $v \in R$ there is a path $Q_v \in \mathcal{Q}$ with endpoints v and v_0, and by the same argument applied to Q_v, with the roles of v and v_0 switched, we see that there is a set B_v of at least $k + 1$ many boosters of the form (v, w).

Since each booster (v, w) is contained in at most two such sets, B_v and B_w, by double counting it follows that the total number of boosters in G is at least

$$\frac{1}{2} \sum_{v \in R} |B_v| \geq \frac{|R|^2}{2} \geq \frac{(k+1)^2}{2}.$$

□

We will start by showing that, since whp $\delta(G(n,p)) \geq 2$, it is likely that $G(n,p)$ will be a $(k,2)$-expander for some large k when p is in the range we're considering.

Why should this be true? Well, as in the proof of Theorem 6.9, in this range of p we should have that any two *large* sets in $G(n,p)$ are connected by an edge, and we don't expect any *small* sets to be dense. Furthermore, the minimum degree of $G(n,p)$ should be at least 2, and we shouldn't have any low degree vertices close to each other. Also, a similar argument will show that no vertices of low degree are contained in short cycles.

Then, if $U \subseteq [n]$ is large, then since U is not connected to $[n] \setminus (U \cup \bar{N}(U))$, it will follow that $\bar{N}(U)$ must also be large.

However, if U is small, then we can split U into two parts $U_S \cup U_L$ where U_S are the vertices of low degree. Since $\delta(G) \geq 2$ and the vertices in U_S are far apart, $\bar{N}(U_S) \geq 2|U_S|$. Furthermore, vertices in U_L cannot have many neighbours in $U_S \cup \bar{N}(U_S)$, since the vertices in U_S are far apart and not contained in any short cycles.

However, since U_L is small, but each vertex of U_L has large degree and few neighbours in $U_S \cup \bar{N}(U_S)$, then $\bar{N}(U_L) \setminus (U_S \cup \bar{N}(U_S))$ is too dense to be small, and so

$$\bar{N}(U) = \bar{N}(U_S) \cup \bar{N}(U_L) \setminus (U_S \cup \bar{N}(U_S))$$

must be large.

Lemma 6.17. *Let $p = \frac{\log n + \log\log n + C(n)}{n}$ where $C(n) \to \infty$ then with high probability $G(n,p)$ is an $(\frac{n}{4}, 2)$-expander.*

Proof. Let us write $G = G(n,p)$ for brevity in what follows, and since being an expander is an increasing property we may assume that $C(n) = o(\log\log n)$. We say a vertex is *small* if $d_G(v) \leq (\log n)^{\frac{7}{8}}$ and write

$$\text{SMALL} = \{v \in [n] \colon v \text{ small}\}.$$

We first note that for each v

$$\begin{aligned}
\mathbb{P}(v \text{ is small}) &= \mathbb{P}(\text{Bin}(n-1,p) \leq (\log n)^{\frac{7}{8}}) \\
&= \sum_{k \leq (\log n)^{\frac{7}{8}}} \mathbb{P}(\text{Bin}(n-1,p) \leq k) \\
&= \sum_{k \leq (\log n)^{\frac{7}{8}}} \binom{n}{k} p^k (1-p)^{n-1-k} \\
&\leq \sum_{k \leq (\log n)^{\frac{7}{8}}} (np)^k e^{-pn} e^{p(k+1)} \\
&\leq (\log n)^{\frac{7}{8}} (2\log n)^{(\log n)^{\frac{7}{8}}} \frac{2}{n} \\
&\leq n^{-0.9}.
\end{aligned}$$

A similar calculation will show that for $u, v \in [n]$, $\mathbb{P}(u, v \in \text{SMALL}) \leq n^{-1.8}$.

Given the similarities between the heuristic for this proof and the proof of Theorem 6.9, it is reasonable to ask why we cannot use Lemma 6.7 again (and so define our small vertices to have degree $\alpha \log n$ for some small α). However, if we take our threshold for a vertex having small degree to be of order $\Theta(\log n)$, we will find that the probability of a vertex being small is always $\omega\left(n^{-\frac{1}{2}}\right)$. However, in the next claim it will be necessary that this probability is smaller than $n^{-\frac{1}{2}-\epsilon}$.

Claim 6.18. With high probability there is no pair $u, v \in \text{SMALL}$ with $d(u,v) \leq 4$ and no 4-cycle containing a vertex $u \in \text{SMALL}$.

Proof of claim. For any vertices $u, v \in [n]$

$$\mathbb{P}(d(u,v) \leq 4) \leq \mathbb{E}(\text{number of } u-v\text{-paths of length} \leq 4) \leq \sum_{i=1}^{4} n^{i-1} p^i = O\left(n^3 p^4\right) \leq n^{-0.9}.$$

Furthermore, since the property that $u, v \in \text{SMALL}$ is a decreasing property and the property that $d(u,v) \leq 4$ is an increasing property, it follows from Harris' inequality that

$$\begin{aligned}
\mathbb{P}(u, v \in \text{ SMALL and } d(u,v) \leq 4) &\leq \mathbb{P}(u, v \in \text{ SMALL}) \mathbb{P}(d(u,v) \leq 4) \\
&\leq n^{-1.8} n^{-0.9} \\
&\leq n^{-2.7}.
\end{aligned}$$

Hence by taking a union bound over all pairs $u, v \in [n]$

$$\mathbb{P}(\text{there exists } u, v \in \text{ SMALL with } d(u,v) \leq 4) \leq n^{-0.7}.$$

Similarly, by a union bound the probability that there is a 4-cycle containing a small vertex is at most

$$\binom{n}{4} p^4 \sum_{k=0}^{(\log n)^{\frac{7}{8}}} p^k (1-p)^{n-k-4} \leq (np)^4 (\log n)^{\frac{7}{8}} e^{-pn} \leq \frac{(\log n)^5}{n} = o(1).$$

□

Claim 6.19. *G contains an edge between every pair of disjoint vertex sets $A, B \subseteq [n]$ with $|A|, |B| \geq \frac{n}{\sqrt{\log n}}$.*

Proof of Claim. Note that if such a pair exists, then one exists with $|A| = |B| = \frac{n}{\sqrt{\log n}} := a$ and we can estimate the expected number of 'bad' pairs A, B as

$$\begin{aligned}\left(\binom{n}{a}\right)^2 (1-p)^{-a^2} &\leq \left(e\sqrt{\log n}\right)^{\frac{2n}{\sqrt{\log n}}} e^{-n} \\ &\leq e^{-n+o(n)} = o(1),\end{aligned}$$

and so the claim follows from Markov's inequality. □

Claim 6.20. *Every subset $U \subseteq [n]$ with $|U| \leq \frac{2n}{(\log n)^{\frac{3}{8}}}$ is such that*

$$e(G[U]) \leq 6|U|(\log n)^{\frac{5}{8}}.$$

Proof of Claim. Again this follows from a union bound. The probability that such a set U with $|U| \leq \frac{2n}{(\log n)^{\frac{3}{8}}} := u$ and $e(G[U]) \geq 3|U|(\log n)^{\frac{5}{8}} := r|U|$ exists is at most

$$\begin{aligned}&\sum_{k \leq u} \binom{n}{k} \binom{\binom{k}{2}}{rk} p^{rk} \\ &\leq \sum_{k \leq u} \left(\frac{en}{k}\right)^k \left(\frac{ek^2}{rk}\right)^{rk} \left(\frac{\log n}{n}\right)^{rk} \\ &= \sum_{k \leq u} \left(\frac{en}{k} \cdot \frac{ek}{r} \cdot \frac{\log n}{n}\right)^k \left(\frac{ek}{r} \cdot \frac{\log n}{n}\right)^{k(r-1)} \\ &= \sum_{k \leq u} \left(e^2 (\log n)^{\frac{3}{8}}\right)^k \left(\frac{e}{3}\right)^{k(r-1)}.\end{aligned}$$

Using the fact that $k \log n \leq u \log n \leq \frac{1}{3} rn$. Hence, the probability is at most

$$\leq \sum_{k \leq u} \exp\left(k\left(2 + \frac{3}{8}\log\log n - \Theta(r)\right)\right) = o(1).$$

□

Hence, by Lemma 6.12 and Claims 6.18-6.20, with high probability G has the following properties:

(a) $\delta(G) \geq 2$;

(b) There is no pair $u, v \in \text{SMALL}$ with $d(u,v) \leq 4$ and no 4-cycle containing a small vertex;

(c) G contains an edge between every pair of disjoint vertex sets $A, B \subseteq [n]$ with $|A|, |B| \geq \frac{n}{(\log n)^{\frac{1}{2}}}$;

(d) Every subset $U \subseteq [n]$ with $|U| \leq \frac{2n}{(\log n)^{\frac{3}{8}}}$ is such that $e(G[U]) \leq 6|U|(\log n)^{\frac{5}{8}}$.

We show that from these conditions it follows deterministically that G is an $(\frac{n}{4}, 2)$-expander.

Supposing then that G satisfies (a)-(d) and let $U \subseteq [n]$ be of size $|U| \leq \frac{n}{4}$. We wish to show that $|\bar{N}(U)| \geq 2|U|$. We split into two cases. Firstly, if $|U| \geq \frac{n}{(\log n)^{\frac{1}{2}}}$ then by (c)

$$|\bar{N}(U)| \geq n - |U| - \frac{n}{(\log n)^{\frac{1}{2}}} \geq \frac{n}{2} \geq 2|U|.$$

Hence we may assume that $U \leq \frac{n}{(\log n)^{\frac{1}{2}}}$. Let us split U into two parts:

$$U_S = U \cap \text{SMALL} \quad \text{and} \quad U_L = U \setminus U_S.$$

Note that

$$\bar{N}(U) = \left(\bar{N}(U_S) \setminus U_L\right) \cup \left(\bar{N}(U_L) \setminus (U_S \cup \bar{N}(U_S)\right).$$

So, in particular

$$|\bar{N}(U)| \geq |\bar{N}(U_S)| - |U_L| + |\bar{N}(U_L) \setminus (U_S \cup \bar{N}(U_S)|. \tag{6.1}$$

Since $\delta(G) \geq 2$ by (a), and by (b) no two vertices of U_S have common neighbours,

$$|\bar{N}(U_S)| \geq 2|U_S|. \tag{6.2}$$

Hence, U_S expands by a factor of at least two. We will see that U_L expands even by a factor of $\omega(1)$, which will be enough to finish the proof. More precisely, we claim that $|\bar{N}(U_L)| \geq |U_L|(\log n)^{\frac{1}{8}}$.

Indeed, suppose for a contradiction that $|\bar{N}(U_L)| \leq |U_L|(\log n)^{\frac{1}{8}}$. Then

$$|U_L \cup \bar{N}(U_L)| \leq (1 + (\log n)^{\frac{1}{8}})|U| \leq \frac{n}{(\log n)^{\frac{3}{8}}}$$

However, since no vertex in U_L is small,

$$e(G[U_L \cup \bar{N}(U_L)]) \geq \frac{1}{2}|U_L|(\log n)^{\frac{7}{8}} \geq \frac{1}{3}|U_L \cup N(U_L)|(\log n)^{\frac{6}{8}},$$

contradicting (d).

So, $|\bar{N}(U_L)| \geq |U_L|(\log n)^{\frac{1}{8}}$ is large, and furthermore each $u \in U_L$ has at most one neighbour in $U_S \cup \bar{N}(U_S)$ by (b). It follows that

$$|\bar{N}(U_L) \setminus (U_S \cup \bar{N}(U_S)| \geq |\bar{N}(U_L)| - |U_L| \geq |U_L| \left((\log n)^{\frac{1}{8}} - 1\right). \tag{6.3}$$

Hence, plugging (6.2) and (6.3) into (6.1),

$$\begin{aligned}|\bar{N}(U)| &\geq |\bar{N}(U_S)| - |U_L| + |\bar{N}(U_L) \setminus (U_S \cup \bar{N}(U_S)| \\ &\geq 2|U_S| + |U_L| \left((\log n)^{\frac{1}{8}} - 2\right) \\ &\geq 2(|U_S| + |U_L|) \\ &= 2|U|.\end{aligned}$$

□

Using this we can now prove Theorem 6.13.

Proof of Theorem 6.13. As mentioned before, Lemma 6.12 deals with the statement for $C(n) \to -\infty$ so let us assume that $C(n) \to \infty$. Furthermore, since containing a Hamiltonian cycle is an increasing property, we may assume that $C(n) = o(\log \log n)$.

Our plan will be to expose the edges of G in multiple rounds. In the first round we will take $p_1 = p - \frac{C'}{n}$ for some large, but finite, C', so that $p_2 \geq \frac{C'}{n}$. We will then split the second round into $2n$ subrounds each with equal probability q, so that $(1-q)^{2n} = 1 - p_2$. It follows that $q \geq \frac{p_2}{2n} \geq \frac{C'}{2n^2}$. We will show that in the first round $G(n, p_1)$ will be an $(\frac{n}{4}, 2)$-expander, and will also be connected. Since $G(n, p_1)$ is an expander, it will contain many boosters, and so with positive probability one such boosters will be contained in a random graph with edge probability q. However, since being an expander is an increasing property, after adding this edge this will remain true, and so if we have sufficiently many random graphs with edge probability q, we expect to be able to add many boosters in a row, and so eventually form a Hamiltonian cycle.

So, let us make the above sketch precise. Let p_1, p_2 and q be as above, $G_1 \sim G(n, p_1)$ and let $\{G_{2,i} \colon i \leq 2n\}$ be a sequence of mutually independent random graphs such that each $G_{2,i} \sim G(n, q)$. By standard coupling arguments we may assume that $G(n, p_1) \cup \bigcup_i G_{2,i} = G(n, p)$.

Since $p_1 = p - \frac{C'}{n} = \frac{\log n + \log \log n + C(n) - C'}{n}$, by Lemma 6.17, with high probability $G(n, p_1)$ is an $(\frac{n}{4}, 2)$-expander. In particular, we note that this implies that $G(n, p_1)$ is connected.

Indeed, if C is a connected component of G then $\bar{N}(C) = \emptyset$ and so $|C| > \frac{n}{4}$. However, if we take some subset $C_0 \subseteq C$ of size $|C_0| = \frac{n}{4}$ then the expansion of $G(n, p_1)$ implies that $|\bar{N}(C_0)| \geq 2|C_0| \geq \frac{n}{2}$ and hence $|C| \geq |C_0 \cup \bar{N}(C_0)| \geq \frac{3n}{4}$. Hence, every component of $G(n, p_1)$ has size at least $\frac{n}{2}$, and so there is only one component.

We will expose the edges in each $G_{2,i}$ one-by-one, referring to the exposure of $G_{2,i}$ as *round* i. Let us write $H_j = G_1 \cup_{i \leq j} G_{2,i}$.

We say that round i is *successful* if either

- H_{i-1} is Hamiltonian; or
- $G_{2,i}$ contains a booster of H_{i-1}.

Note that, if n rounds are succesful then $H_{2n} = G(n,p)$ contains a Hamiltonian cycle.

Let us consider round i. If H_{i-1} is Hamiltonian then we are already successful. Otherwise, since H_0 is a connected, $(\frac{n}{4}, 2)$-expander, so is H_{i-1}. Hence, by Lemma 6.16, H_{i-1} has at least $\frac{n^2}{32}$ many boosters. It follows that

$$\mathbb{P}(\text{round } j \text{ is not succesful}) \leq (1-q)^{\frac{n^2}{32}} \leq e^{-\frac{qn^2}{32}} \leq e^{-\frac{C'}{64}} \leq \frac{1}{3},$$

for a large enough choice of C'. Whilst the probability that future rounds are successful are not necessarily independent, it is clear that the number of successful rounds is at least stochastically dominated by a $\mathrm{Bin}\big(2n, \frac{2}{3}\big)$ random variable. Hence, if we left X denote the number of successful rounds then by Lemma 5.2

$$\mathbb{P}(X < n) \leq \mathbb{P}\left(\mathrm{Bin}\left(2n, \frac{2}{3}\right) < n\right) = e^{-\Omega(n)} = o(1).$$

□

7 Chromatic Number

7.1 Martingales and the Azuma-Hoeffding inequality

Given a real random variable X and a discrete random variable Z on the same probability space Ω we write $\mathbb{E}(X|\sigma(Z))$ for the random variable such that

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

We can think about this as restricting ourselves to the σ-algebra generated by the sets on which Z is constant, and replacing the random variable X with it's average on each of these sets.

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 for each i, X_i is determined by $\{Z_1, Z_2, \ldots, Z_i\}$ and

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

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

We note that this is a rather simplified definition of a martingale, but it will serve for our purposes. Roughly we can think of a martingale as being a gradual exposure of a random variable X_n, by revealing more and more information in terms of the random variables Z_i. Indeed, it is easy to check inductively that X_i is $\mathbb{E}(X_n|\sigma(Z_1, \ldots, Z_i))$, and so each X_i is a finer and fine approximation of X_n. The converse is also true, which gives us a rich source of martingales.

Lemma 7.1. *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 each 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 you can consider, the *edge-exposure martingale* and the *vertex-exposure martingale.*

For the first we order the set $[n]^{(2)}$ arbitrarily as $e_1, e_2, \ldots, e_m$ and define a sequence (Z_i) where Z_i is $\mathbb{1}_{e_i \in G(n,p)}$, the indicator function of the event that $e_i \in G(n, p)$. Then for any graph theoretic function f the random variable $f := f(G(n, p))$ is a function of the sequence (Z_i), and so as in Lemma 7.1 we can consider the martingale sequence $X_i = \mathbb{E}(f \,|\, \sigma(Z_1, Z_2, \ldots, Z_i))$. We can think of this martingale as revealing the edges of $G(n, p)$ one by one, and the random variables X_i is the expected value of that f on a random graph $G(n, p)$, which agrees with the first $i - 1$ revealed edges. In particular, $X_0 = \mathbb{E}(f)$ and $X_n = f$.

For the vertex exposure martingale we take $Z_i' \in \{0,1\}^{i-1}$ to be the vector of indicator random variables of the event that the edge (i,j) with $j < i$ 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, then X_i is the expected value of f on a random graph $G(n,p)$, which 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 7.2 (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 $\sigma^2 = \sum_{i=1}^n c_i^2$.

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 $f(Z_1, Z_2, \ldots, Z_n)$, then if changing any single coordinate 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 *c-edge-Lipschitz* if whenever H and H' differ in only one edge then $|f(H) - f(H')| \leq c$. Equivalently, if we consider f as a function of the variables $f(Z_1, Z_2, \ldots, Z_m)$ then we require that changing one coordinate does not change f by more than c. Similarly it is *c-vertex-Lipschitz* if whenever H and H' differ at only one vertex $|f(H) - f(H')| \leq c$.

Lemma 7.3. *For any graph theoretic function f, if f is c-edge-Lipschitz then the corresponding edge exposure martingale satisfies $|X_i - X_{i-1}| \leq c$ and similarly if f is c-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 mutually independent with $Z_1, \ldots, Z_n$. Then,

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

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

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

Where the last inequality holds by the Lipschitz condition for any particular event in $\sigma(Z_1, Z_2, \ldots, Z_i)$, and so holds in the average case. □

Combining Lemma 7.3 and Theorem 7.2 for a Lipschitz graph function f (and its 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).

7.2 The Chromatic Number of a Dense Random Graph

We note that the chromatic number of a graph $\chi(G)$ is an 1-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 7.2 to the edge exposure martingale will not tell us much. Indeed the Azuma-Hoeffding inequality, applied with $c_i = 1$, tells us that

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

where $m = \binom{n}{2} = \Omega\left(n^2\right)$ is the number of possible edges in $G(n,p)$. So, to have the right hand side tend to 0 we need $t = \omega(n)$. However χ only takes values in $[n]$, so in this case saying that with high probability χ is within $\omega(n)$ of its expectation is not very useful.

However we can instead consider the vertex exposure martingale. Again it is a simple check that χ is also a 1-vertex-Lipschitz function. In this case combining Lemma 7.3 and Theorem 7.2 gives us a better bound, first due to Shamir and Spencer

Theorem 7.4. *For any n and p and for all $t \geq 0$*

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

Proof. As noted, χ is a vertex Lipschitz function. So by Lemma 7.3 we can apply Theorem 7.2 to the associated vertex exposure martingale with $c_i = 1$ to conclude that, for all $t \geq 0$

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

□

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. This will be a useful bound as long

at this expectation is larger than $\sqrt{n}$. But what is the expected chromatic number of a random graph?

One simpler question to consider is the independence number $\alpha(G(n,p))$ of a random graph. Then, since any colour class must be an independent set,

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

Hence, a good upper bound on the independence number would give us a good lower bound on the chromatic number. It is relatively easy, for fixed p, to see what the independence number of $G(n,p)$ should be. Note that, since an independent set is a clique in the complement, and $G(n,p)^c \sim G(n,1-p)$ this also tells us what the clique number of a dense random graph is.

Theorem 7.5. *Let $p \in (0,1)$ be constant and let $b = \frac{1}{1-p}$. Then with high probability*

$$\alpha(G(n,p)) = (2+o(1))\log_b n.$$

Proof. Let X_k be the number of independent sets of size k, and suppose that $k = \lceil 2\log_b n \rceil$, so that $(1-p)^k \leq n^{-2}$. Then

$$\begin{aligned}
\mathbb{E}(X_k) &= \binom{n}{k}(1-p)^{\binom{k}{2}} \\
&\leq \left(\frac{en}{k}\right)^k (1-p)^{-\frac{k}{2}}(1-p)^{\frac{k^2}{2}} \\
&= \left(\frac{en}{k}(1-p)^{-\frac{1}{2}}(1-p)^{\frac{k}{2}}\right)^k \\
&\leq \left(\frac{e}{k\sqrt{1-p}}\right)^k = o(1).
\end{aligned}$$

Hence by Markov's inequality, with high probability there are no independent sets of size k, and hence no independent sets of size $\geq k$.

Suppose then that $k = \lfloor (2-\varepsilon)\log_b n \rfloor$ for some fixed $\varepsilon > 0$, so that $(1-p)^k \geq n^{\epsilon-2}$. In this case we have that

$$\begin{aligned}
\mathbb{E}(X_k) &= \binom{n}{k}(1-p)^{\binom{k}{2}} \\
&\geq \left(\frac{n}{k}\right)^k (1-p)^{\frac{k^2}{2}} \\
&= \left(\frac{n}{k}(1-p)^{\frac{k}{2}}\right)^k \\
&\geq \left(\frac{n^{\frac{\varepsilon}{2}}}{k}\right)^k = \omega(1)
\end{aligned}$$

So, we would like to establish the concentration of X_k. The original proof of this fact by Bollobás used the Azuma-Hoeffding inequality. However, since X_k is not very edge or vertex Lipschitz, he had to use a little trick and consider instead the size of a maximal family of 'disjoint' independent k-sets. This is then a 1-edge-Lipschitz random variable, and it is possible to show using an alteration trick that its expectation is not much smaller than that of X_k, and one can then apply the Azuma-Hoeffding inequality that it is non-zero with high probability.

However, we can save a bit of time by using Janson's inequality. Technically, in order to apply Janson's inequality we should work in the complement and consider the number of k-cliques, but for ease or presentation we won't go into quite so much detail.

We want then to calculate the quantity Δ from Janson's inequality, which is the sum over all pairs Y_i, Y_j of k-sets in $[n]$ which share at least two vertices of the probability that Y_i and Y_j are independent sets. Hence

$$\begin{aligned}\Delta &= \sum_{j=2}^{k-1} \binom{n}{k}\binom{k}{j}\binom{n-k}{k-j}(1-p)^{\binom{k}{2}}(1-p)^{\binom{k}{2}-\binom{j}{2}} \\ &= \binom{n}{k}(1-p)^{\binom{k}{2}} \sum_{j=2}^{k-1} \binom{k}{j}\binom{n-k}{k-j}(1-p)^{\binom{k}{2}-\binom{j}{2}}\end{aligned}$$

and so

$$\frac{\Delta}{\mathbb{E}(X_k)^2} = \sum_{j=2}^{k-1} \binom{n}{k}^{-1}\binom{k}{j}\binom{n-k}{k-j}(1-p)^{-\binom{j}{2}} := \sum_{j=1}^{k-1} \mu_j$$

We'd like to show that this is small. Firstly let us show that the first term dominates this expression. Indeed if $j \geq 2$ then

$$\begin{aligned}\frac{\mu_{j+1}}{\mu_j} &= \frac{\binom{k}{j+1}\binom{n-k}{k-j-1}(1-p)^{-\binom{j+1}{2}}}{\binom{k}{j}\binom{n-k}{k-j}(1-p)^{-\binom{j}{2}}} \\ &= \frac{k-j}{j+1}\frac{k-j}{n-2k+j+1}(1-p)^{-j} \\ &\leq \frac{k^2}{n}(1-p)^{-j}\end{aligned}$$

and so for $j \geq 3$

$$\begin{aligned}\frac{\mu_j}{\mu_2} &\leq \left(\frac{k^2}{n}\right)^{j-2}(1-p)^{-\binom{j}{2}+1} \\ &\leq \left(\frac{k^2}{n}(1-p)^{-\frac{j(j-1)+2}{2(j-2)}}\right)^{j-2} \\ &\leq \left(\frac{k^2}{n}(1-p)^{-\frac{j}{2}-\frac{1}{2}}\right)^{j-2} \\ &\leq \left(\frac{k^2}{n\sqrt{1-p}}(1-p)^{-\frac{k}{2}}\right)^{j-2} \\ &\leq \left(\frac{k^2}{n}\frac{n^{1-\frac{\varepsilon}{2}}}{\sqrt{1-p}}\right)^{j-2} \\ &\leq \left(\frac{1}{2}\right)^{j-2}.\end{aligned}$$

In particular, $\frac{\Delta}{\mathbb{E}(X_k)^2} = \sum_{j=1}^{k-1} \mu_j \leq 2\mu_2$, and so

$$\begin{aligned}\frac{\mathbb{E}(X_k)^2}{\Delta} &\geq \frac{1}{2\mu_2} \\ &= \frac{1-p}{2}\binom{n}{k}\binom{k}{2}^{-1}\binom{n-k}{k-2}^{-1} \\ &\geq \frac{1-p}{2}\binom{n}{k}\binom{n-2}{k-2}^{-1}k^{-2} \\ &= \frac{1-p}{2}\frac{n(n-1)}{k^3(k-1)} \\ &\geq \frac{(1-p)n^2}{2k^4}\end{aligned}$$

Hence by Lemma 4.11 we have that

$$\mathbb{P}(X_k = 0) \leq \exp\left(-\frac{\mathbb{E}(X_k)^2}{\Delta}\right) \leq \exp\left(-\frac{(1-p)n^2}{2k^4}\right) = o(1).$$

□

Let us note a few things about this proof. Firstly it follows quite easily that with high probability

$$\chi(G(n,p)) \geq \frac{n}{\alpha(G(n,p))} \geq (1+o(1))\frac{n}{2\log_b n}.$$

In particular, the concentration result we get from Theorem 7.4 is effective - we can conclude that $\chi(G(n,p))$ is whp $(1+o(1))\mathbb{E}(\chi(G(n,p)))$, even if we don't know what this expectation is.

However, as we say see, the bound above is in fact the right order of magnitude. Moreover, we note that the proof actually gives quite a strong bound on the probability that $\alpha(G(n,p)) \leq (2-\varepsilon)\log_b n$, and since we will need it in the following proof we make it explicit now

$$\mathbb{P}\left(\alpha(G(n,p)) \leq (2-\varepsilon)\log_b n\right) \leq \exp\left(-\Omega\left(\frac{n^2}{(\log n)^4}\right)\right) = o(2^{-n}). \tag{7.1}$$

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

Theorem 7.6. *Let $p \in (0,1)$ be constant and let $b = \frac{1}{1-p}$. Then with high probability*

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

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

We can use Theorem 7.5 to say that almost surely this subgraph will contain an independent set of size $(2+o(1))\log_b k$. since this probability is sufficiently large 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 $(2+o(1))\log_b k$, which we can do until there are less than k vertices left.

If k can be chosen to be large enough that $(2+o(1))\log_b k \sim (2+o(1))\log_b n$, then we will use about the right amount of colours in this process. Similarly, if k can be chosen to be small enough that $k = o\left(\frac{n}{\log_b n}\right)$, then if we colour the remaining vertices each by a different colour, it won't have a large effect on the total number of colours we've used.

Let $\varepsilon > 0$ be abitrary and let $k = n/(\log_b(n))^2$. Note that $\log_b k \approx \log_b n$ and $k = o\left(\frac{n}{\log_b k}\right)$. For any subset $S \subseteq [n]$ of size k, the graph $G(n,p)[S] \sim G(k,p)$. hence by Theorem 7.5

$$\mathbb{P}\left(\alpha\left(G(n,p)[S]\right) < (2-\varepsilon)\log_b k\right) = o(2^{-n}).$$

Since there are at most 2^n subsets of $G(n,p)$ of size k, by the union bound

$$\mathbb{P}\left(\alpha\left(G(n,p)[S]\right) < (2-\varepsilon)\log_b k \text{ for some } k\text{-set } S\right) = o(1).$$

Hence, with high probability, every set of k vertices contains an independent set of size $(2-\varepsilon)\log_b k$.

We greedily colour $G(n,p)$ by choosing, as long as the set of uncoloured vertices is of size $\geq k$, some independent set of size $(2-\varepsilon)\log_b k$ and colouring it with a new colour. Once there are less than k uncoloured vertices we colour each with a different new colour. Clearly this produces a proper colouring of $G(n,p)$ and so with high probability

$$\begin{aligned}\chi(G(n,p)) &\leq \frac{n}{(2-\varepsilon)\log_b k} + k \\ &\leq \frac{n}{(2-2\varepsilon)\log_b n} + \frac{n}{(\log(n))^2} \\ &\leq (1+2\varepsilon)\frac{n}{2\log_b n}\end{aligned}$$

for small enough ε. Together with the comment after Theorem 7.5 the result follows. □

7.3 The Chromatic Number of Sparse Random Graphs

What about for smaller values of p? Since Theorem 7.4 holds for for any p, as long as $\mathbb{E}(\chi(G(n,p)) = \omega(\sqrt{n})$ it will follow that $\chi(G(n,p))$ is with high probability $(1+o(1))\mathbb{E}(\chi(G(n,p)))$, although we might not be able to calculate the precise value of $\mathbb{E}(\chi(G(n,p))$. Since we can bound the chromatic number from below in terms of the independence number, a simple first moment calculation shows that this will happen as long as p is significantly larger than $n^{-\frac{1}{2}}$.

Lemma 7.7. *For any $p \in (0,1)$,*

$$\mathbb{P}\left(\alpha(G(n,p)) \geq \frac{4\log n}{p}\right) = o\left(n^{-2}\right).$$

Proof. Let $k = \lceil \frac{4\log n}{p} \rceil$ and let X_k be the number of independent sets of size k in $G(n,p)$. Then

$$\begin{aligned}
\mathbb{P}\left(\alpha(G(n,p)) \geq \frac{4\log n}{p}\right) &\leq \mathbb{E}(X_k) = \binom{n}{k}(1-p)^{\binom{k}{2}} \\
&\leq \left(\frac{en}{k}\right)^k e^{-p\frac{k(k-1)}{2}} \\
&\leq \left(ene^{-\frac{pk}{3}}\right)^k \\
&\leq \left(ene^{-\frac{4}{3}\log n}\right)^k \\
&\leq \left(en^{-\frac{1}{3}}\right)^k = o(n^{-2})
\end{aligned}$$

□

Hence, if $p \geq \frac{(\log n)^2}{\sqrt{n}}$ then

$$\begin{aligned}
\mathbb{E}(\chi(G(n,p))) &\geq \mathbb{P}\left(\alpha(G(n,p)) \leq 4\frac{\sqrt{n}}{\log n}\right)\mathbb{E}\left(\chi(G(n,p)) \mid \alpha(G(n,p)) \leq 4\frac{\sqrt{n}}{\log n}\right) \\
&\geq \mathbb{P}\left(\alpha(G(n,p)) \leq 4\frac{\log n}{p}\right)\mathbb{E}\left(\chi(G(n,p)) \mid \alpha(G(n,p)) \leq 4\frac{\sqrt{n}}{\log n}\right) \\
&\geq (1+o(1))\frac{\sqrt{n}\log n}{4} = \omega\left(\sqrt{(n-1)\log n}\right)
\end{aligned}$$

and by Theorem 7.4

$$\mathbb{P}\left(|\chi(G(n,p) - \mathbb{E}(\chi(G(n,p))| \geq \sqrt{(n-1)\log n}\right) \leq n^{-\frac{1}{2}}.$$

Hence, for this range of p as well $\chi(G(n,p) = (1+o(1))\mathbb{E}(\chi(G(n,p))$. How about for smaller p? It turns out that this sort of concentration will hold as long as p is sufficiently larger than $\frac{1}{n}$.

Theorem 7.8. *Suppose that $p \geq n^{-\alpha}$ for some $\alpha < 1$. Then there exists some function $h(n)$ such that with high probability $\chi(G(n,p)) = (1+o(1))h(n)$.*

Proof. By Lemma 7.7 we may assume that $p \leq \frac{(\log n)^2}{\sqrt{n}}$. Let $h(n)$ be the smallest integer such that

$$\mathbb{P}(\chi(G(n,p)) \leq h) \geq \frac{1}{\log n},$$

so that $\mathbb{P}(\chi(G(n,p)) < h) < \frac{1}{\log n}$. Note that, by Lemma 7.7 we know that

$$\mathbb{P}\left(\alpha(G(n,p)) \geq 4\frac{\log n}{p}\right) \leq o(n^{-2}) = o\left(\frac{1}{\log n}\right)$$

and so

$$\mathbb{P}\left(\chi(G(n,p)) \leq \frac{pn}{4\log n}\right) \leq \mathbb{P}\left(\alpha(G(n,p)) \geq 4\frac{\log n}{p}\right) = o\left(\frac{1}{\log n}\right)$$

and so $h(n) \geq \frac{pn}{4\log n} \geq \frac{n^{1-\alpha}}{4\log n}$.

Let us also define a graph function f where $f(G)$ is the smallest integer k such that there exists a subset $S \subseteq V(G)$ of size k with $\chi(G \setminus S) \leq h$.

We note that f is a 1-vertex-Lipschitz function. So let us consider the corresponding vertex exposure martingale $\mathbb{E}(f) = X_1, X_2, \dots, X_n = f(G(n,p))$. The Azuma-Hoeffding inequality tells us that

$$\mathbb{P}(\chi(G(n,p)) \leq h) = \mathbb{P}(X_n = 0) \leq \mathbb{P}(|X_n - X_1| \leq \mathbb{E}(X_n)) \leq 2e^{-\frac{(\mathbb{E}(X_n))^2}{2(n-1)}}.$$

However, since $\mathbb{P}(\chi(G(n,p)) \leq h) \geq \frac{1}{\log n}$ by assumption, it follows that $\mathbb{E}(X_n) \leq \sqrt{n \log n}$. Then, applying Azuma-Hoeffding again we see that

$$\mathbb{P}(f(G(n,p)) > 2\sqrt{n \log n}) \leq \mathbb{P}(|X_n - X_1| \geq \sqrt{n \log n}) \leq n^{-\frac{1}{2}}.$$

To put this into words, with high probability $G(n,p)$ will contain a subset S of size at most $2\sqrt{n \log n}$ such that $G(n,p) \setminus S$ can be h-coloured.

Claim 7.9. If $p \leq \frac{(\log n)^2}{\sqrt{n}}$ then with high probability every subset $A \subseteq [n]$ of size $|A| \leq 2\sqrt{n \log n}$ has $e(G(n,p)[A]) \leq |A|(\log n)^3$.

Proof of claim. If we let X be the number of subsets A which don't satisfy the conclusion of the claim then

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k \leq 2\sqrt{n \log n}} \binom{n}{k} \binom{\binom{k}{2}}{k(\log n)^3} p^{k(\log n)^3} \\
&\leq \sum_{k \leq 2\sqrt{n \log n}} \left(n^{\frac{1}{(\log n)^3}} \frac{ek^2}{k(\log n)^3} \frac{(\log n)^2}{\sqrt{n}} \right)^{k(\log n)^3} \\
&\leq \sum_{k \leq 2\sqrt{n \log n}} \left(e^{\frac{1}{(\log n)^2}} \frac{ek}{\sqrt{n} \log n} \right)^{k(\log n)^3} \\
&\leq \sum_{k \leq 2\sqrt{n \log n}} \left(\frac{2e^2 \sqrt{n \log n}}{\sqrt{n} \log n} \right)^{k(\log n)^3} \\
&\leq \sum_{k \leq 2\sqrt{n \log n}} \left(\frac{2e^2}{\sqrt{\log n}} \right)^{k(\log n)^3} \\
&= o(1)
\end{aligned}$$

□

We say a graph H is *r-degenerate* if there is an ordering of the vertices $V(H) = \{v_1, \dots, v_m\}$ such that $|\{j : j < i \text{ and } (v_i, v_j) \in E(H)\}| \leq r$ for all $i \leq m$. Note that an r-degenerate graph can easily be $r+1$ coloured by greedily colouring the graph starting from v_1 to v_m.

We note that Claim 7.9 implies that with high probability every set A of size $2\sqrt{n \log n}$ is such that $G(n,p)[A]$ is $2(\log n)^3$-degenerate. Indeed, if such an A were not, then there must exist a subset $A' \subseteq A$ with $\delta(G(n,p)[A']) > 2(\log n)^3$, as otherwise we could pick our ordering greedily. However then $|A'| \leq |A|$ and

$$e(G(n,p)[A']) \geq |A'| \frac{2(\log n)^3}{2} \geq |A'|(\log n)^3$$

contradicting Claim 7.9. In particular, the induced subgraph of $G(n,p)$ on S has chromatic number at most $2(\log n)^3 + 1$.

So, to recap, with high probability there is a subset $S \subseteq G(n,p)$ such that $\chi(G(n,p) \setminus S) \leq h$ and $\chi(G(n,p)[S]) \leq 2(\log n)^3 + 1$. Hence with high probability, by colouring these two sets with disjoint sets of colours

$$\chi(G(n,p)) \leq h(n) + 2(\log n)^3 + 1 = (1 + o(1))h(n)$$

since $h(n) \geq \frac{n^{1-\alpha}}{4 \log n} = \omega\left((\log n)^3\right)$. However, by assumption

$$\mathbb{P}(\chi(G(n,p)) < h) < \frac{1}{\log n} = o(1)$$

and so with high probability $\chi(G(n,p)) = (1 + o(1))h(n)$. □

In fact, rather remarkably, for small enough values of p with high probability $\chi(G(n,p))$ takes at most two values!

Theorem 7.10. *Let $\alpha > \frac{5}{6}$. If $p(n) \leq n^{-\alpha}$, then there exists a function $h(n)$ such that with high probability $h(n) \leq \chi(G(n,p)) \leq h(n) + 1$.*

Proof. The idea of the proof is relatively simple. In the previous proof we showed that, for an appropriate choice of $h := h(n)$, with high probability there is a set S of at most $2\sqrt{n \log n}$ vertices such that $G(n,p) \setminus S$ can be h-coloured, and with high probability $\chi(G(n,p)) \geq h$. Furthermore we showed that this set S was $2(\log n)^3$-degenerate, and hence we could colour this remainder with at most $2(\log n)^3$ more colours.

In fact, for small enough p, the conclusion of Claim 7.9 will hold for even a constant degeneracy, in fact degeneracy two. This will follow from the following fact, that no small subsets have average degree more than $3 - 2\delta$, where $\delta = \alpha - \frac{5}{6} > 0$.

Claim 7.11. If $p \leq n^{-\alpha}$ then with high probability every subset $A \subseteq [n]$ of size $|A| \leq \frac{4}{\delta}\sqrt{n \log n}$ has $e(G(n,p)[A]) \leq (\frac{3}{2} - \delta)|A|$.

Proof of claim. Let us assume first that $\alpha \leq 1$. If we let X be the number of subsets A which

don't satisfy the conclusion of the claim then

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k\le\frac{4}{\delta}\sqrt{n\log n}} \binom{n}{k}\binom{\binom{k}{2}}{(\frac{3}{2}-\delta)k} p^{(\frac{3}{2}-\delta)} \\
&\le \sum_{k\le\frac{4}{\delta}\sqrt{n\log n}} \left(\frac{en}{k}\left(\frac{ek^2}{(\frac{3}{2}-\delta)k}\right)^{(\frac{3}{2}-\delta)} p^{(\frac{3}{2}-\delta)}\right)^k \\
&\le O\left(\sum_{k\le\frac{4}{\delta}\sqrt{n\log n}} \left(k^{\frac{1}{2}-\delta} n^{1-(\frac{3}{2}-\delta)\alpha}\right)^k\right) \\
&\le O\left(\sum_{k\le\frac{4}{\delta}\sqrt{n\log n}} \left(n^{\frac{5}{4}-\frac{3}{2}\alpha-\frac{1}{2}\delta+\delta\alpha}(\log n)^{\frac{1}{4}}\right)^k\right) \\
&\le O\left(\sum_{k\le\frac{4}{\delta}\sqrt{n\log n}} \left(n^{\frac{3}{2}(\frac{5}{6}-\alpha)-\frac{1}{2}\delta+\delta\alpha}(\log n)^{\frac{1}{4}}\right)^k\right) \\
&\le O\left(\sum_{k\le\frac{4}{\delta}\sqrt{n\log n}} \left(n^{-\frac{3}{2}\delta-\frac{1}{2}\delta+\delta}(\log n)^{\frac{1}{4}}\right)^k\right) \\
&\le O\left(\sum_{k\le\frac{4}{\delta}\sqrt{n\log n}} \left(n^{-\delta}(\log n)^{\frac{1}{4}}\right)^k\right) \\
&= o(1)
\end{aligned}$$

□

So, already we can use the same strategy as before to show that with high proability

$$h(n) \le \chi(G(n,p)) \le h(n)+3.$$

However, by being a bit more careful we can bring this 3 down even further.

With high probability there exists an S as in Theorem 7.8, that is, $|S| \le 2\sqrt{n\log n}$ and $\chi(G(n,p)\setminus S) \le h$, where $h := h(n)$ be the smallest integer such that

$$\mathbb{P}(\chi(G(n,p)) \le h) \ge \frac{1}{\log n}.$$

If $N(S)$, the exclusive neighbourhood, were independent then we could use it as a 'buffer' of sorts: We colour $G(n,p)\setminus(S\cup N(S)$ using the colours $\{1,2,\dots,h\}$, and then use a new colour $h+1$ to colour $N(S)$. Then, S is 2-degenerate, and so can be three coloured, however since all of S's neighbours are coloured $h+1$, we can just use the colours $\{1,2,3\}$ to colour S as before, and not cause any problems with the previous colouring.

So what if $N(S)$ is not independent? Well, then we can find an edge (u,v) with $u,v\in N(S)$. But then $S' = S\cup\{u,v\}$ has two more vertices, but three more edges than S. Since by assumption

$\frac{2e(S)}{|S|} < 3$ we have that

$$\frac{2e(S')}{|S'|} = \frac{2e(S)+6}{|S|+2} > \frac{2e(S)}{|S|},$$

and so the average degree of S' is larger than that of S. So, let us define a sequence of subsets $S = S_0, S_1, \ldots, S_t$ where, if S_i has already been defined then either $N(S_i)$ is not independent, in which case we find $u, v \in N(S_i)$ with $(u, v) \in E(G(n,p))$ and set $S_{i+1} = S_i \cup \{u, v\}$, or $N(S_i)$ is independent and we let $S_{i+1} = S_i$. Suppose we run this process for $t = (\frac{2}{\delta} - 1)\sqrt{n \log n}$ steps, then if $N(S_t)$ is not independent we have that

$$|S_t| = |S| + 2t \quad \text{and} \quad e(S_t) \geq e(S) + 3t \geq 3t$$

Hence $|S_t| \leq 2\sqrt{n \log n} + (\frac{4}{\delta} - 2)\sqrt{n \log n} \leq \frac{4}{\delta}\sqrt{n \log n}$ and so

$$\frac{2e(S_t)}{|S_t|} \geq \frac{6t}{\frac{4}{\delta}\sqrt{n \log n}} \geq \frac{\frac{12}{\delta} - 6}{\frac{4}{\delta}} = \frac{12 - 6\delta}{4} = 3 - \frac{3}{2}\delta > 3 - 2\delta,$$

contradicting Claim 7.11. It follows that $N(S_t)$ is independent and hence, with high probability $G(n,p)$ contains some set S_t such that $|S_t| \leq \frac{4}{\delta}\sqrt{n \log n}$, $N(S_t)$ is independent and $\chi(G(n,p) \setminus S_t) \leq h$.

It follows that there is a h-colouring of $G(n,p) \setminus (S_t \cup N(S_t)$. Furthmore, by Claim 7.11 whp S_t is 2-degenerate, and so can be three coloured using the colours $\{1, 2, 3\}$. Finally we can give the set $N(S_t)$ the colour $h + 1$. As long as $h \geq 3$, this determines a proper colouring of $G(n,p)$ and so with high probability $h(n) \leq \chi(G(n,p)) \leq h(n) + 1$.

But what if $h \leq 3$? This is sort of a silly case, and note that even if $h \leq 3$ we can still use the above ideas to show that with high probability $h \leq \chi(G(n,p)) \leq h + 3$. However, with a bit of care we can deal with this case as well.

Firstly, when $p = o(n^{-1})$, by Theorem 5.1 with high probability $G(n,p)$ is a forest and so $\chi(G(n,p)) \leq 2$.

Also if $p = \frac{c}{n}$ for $c = 1 + \epsilon$ sufficiently close to one, then it's not too hard to show, via similar calculations as Claim 7.11, that with high probability every set of size at most $\approx 2\epsilon n$ has minimum degree at most two. However, since any subgraph with minimum degree at least three is a subgraph of the giant component, and the giant component has at most $\approx 2\epsilon n$ vertices, it follows that $G(n,p)$ has no subgraph of minimum degree at least three, and hence can be three coloured greedily.

So, all we really need to show if that, if $p \geq (1.001)/n$ say then $\chi(G(n,p)) \geq 3$, or in other words, G contains an odd cycle, with probability at least $1 - \frac{1}{\log n}$. We leave this as an exercise. □

8 Random Regular Graphs

Suppose we want to consider the properties of a random r-regular graph $G(n, \overline{r})$, for some fixed r, that is, a graph picked unfiformly from the set of all r-regular graphs on n vertices. For very small values of r, $0, 1$ or 2, we can say quite explicitly what such graphs look like, and so reasonably easily talk about the distribution of $G(n, \overline{r})$. For example, a random 1-regular graph is just a random matching, which will only exist when n is even, and can be generated by sequentially matching pairs of unmatched vertices uniformly at random.

However for larger values of r we have no easy description of the class of r-regular graphs, and so we don't know how to efficiently sample or generate a random r-regular graphs.

Of course, we could look at $G(n, p)$ for an appropriate choice of p and condition on the event that G is r-regular, but even for the 'best' choice of $p = \frac{r}{n-1}$ the probability that $G(n, p)$ is r-regular will be exponentially small, and furthermore, we have little control over how the edge probabilities interact.

In this section will describe a very useful model for generating random graphs with a fixed degree sequence, that in particular will allow us to talk about properties of $G(n, \overline{r})$.

8.1 The Configuration Model

Suppose we have a sequence $\overline{d} = (d_1, \dots, d_n) \in \mathbb{N}^n$ such that $\sum_i d_i$ is even. Let us denote by $G(n, \overline{d})$ the random variable which is uniformly distributed on the set $\mathcal{G}_{n,\overline{d}}$ of graphs with degree sequence $\overline{d}$. We will assume that $d_i \geq 1$ for each i.

If we have a graph with degree sequence $\overline{d}$ then we could imagine drawing d_i unfinished 'half-edges' from each vertex $i \in [n]$. Locally around each vertex we know that this is what the graph looks like, but we don't know where we will end up if we follow a particular edge from i. One way to build a graph with degree sequence $\overline{d}$ would be to match up the ends of these half edges in a uniformly random manner (you could think of this as happening sequentially, one by one, or done all at once by picking a random matching of the half-edges). Due to the inherent symmetry of this process, we might hope that this process shouldn't be biased towards any particular graph $\mathcal{G}_{n,\overline{d}}$, or in other words, that this produces a uniformly distributed graph in $\mathcal{G}_{n,\overline{d}}$. As we shall see, this will almost be the case.

So, let us make the process we descrived about more explicit. We take a collection of disjoint sets $\{W_i \colon i \in [n]\}$, which we call cells, such that $|W_i| = d_i$ for each i and let $W = \bigcup_i W_i$. Let $\phi : W \to [n]$ be the function which maps a point $x \in W_i$ to $\phi(x) = i$.

Now, $\sum_i d_i = 2m$ is even, and so there are many ways to partition W into a set F of m pairs, which we call a configuration. Given any configuration F we can build a multi-graph $\gamma(F)$ on $[n]$ by taking edges $(\phi(x), \phi(y))$ for each pair $(x, y) \in F$. Note that $\gamma(F)$ may have loops or parallel edges. In this way we get a map $\gamma : \Omega \to \mathcal{G}^*_{n,\overline{d}}$ to the set of multi-graphs with degree sequence $\overline{d}$.

If we let Ω be the set of all configurations of W then we can count

$$|\Omega| = \frac{(2m)!}{m!2^m}.$$

Indeed, we can build a configuration from a permutation of W by just taking the consecutive pairs, and if we fix a configuration F we can see that the number of permutations which give rise to F is exactly $2^m m!$.

Lemma 8.1. *If $G \in \mathcal{G}_{n,\overline{d}}$, then $|\gamma^{-1}(G)| = \prod_{i=1}^n d_i!$.*

Proof. Let us arrange the edges of G in some fixed order $E(G) = \{(i_1, i_2), (i_3, i_4), \ldots, (i_{2m-1}, i_{2m})\}$. If F is a configuration such that $\gamma(F) = G$, each edge (i_j, i_{j+1}) corresponds to a pair $(x, y) \in F$ where $\phi(x) = i_j$ and $\phi(y) = i_{j+1}$. Hence, we can generate each such configuration F by going through the i_j in order and choosing the corresponding element $x \in W_{i_j}$. Since for each W_i there are d_i many $i_j = i$, there are $d_i!$ many different ways to assign the elements of W_i and hence $\prod_{i=1}^n d_i!$ many configurations altogether. □

As a simple corollary, if F is uniformly chosen from Ω then $\gamma(F)$ is equally likely to be any particular graph in $\mathcal{G}_{n,\overline{d}}$.

Corollary 8.2. *Let F be uniformly distributed on Ω and let $G_1, G_2 \in \mathcal{G}_{n,\overline{d}}$, then*

$$\mathbb{P}(\gamma(F) = G_1) = \mathbb{P}(\gamma(F) = G_2).$$

However, if G is not simple, then the procedure in Lemma 8.1 will overcount the number of configurations leading to G, and we can see that the configuration model is less likely to produce multi-graphs with many loops or parallel edges. In particular, one can show that $\gamma(F)$ is not uniformly distributed on $\mathcal{G}^*_{n,\overline{d}}$. However, if we want to sample uniformly from $G(n, \overline{d})$, we can sample from $\gamma(F)$ and accept it only if the graph we get is simple. By Corollary 8.2, conditioned on $\gamma(F)$ being simple, it is equally likely to be any graph in $\mathcal{G}_{n,\overline{d}}$. However, this is only really useful if $\gamma(F)$ is quite likely to be simple, but we will see that this is in fact the case for 'simple' $\overline{d}$.

It will be useful to imagine that we can generate F in the following algorithmic manner: We choose our partition by sequentially choosing an arbitrary x from the unpaired points in W and then choosing y uniformly at random from the remaining unpaired points. It is clear that if we generate a configuration in this manner then it is uniformly distributed. Note that it doesn't matter if the xs are chosen randomly or according to some pre-determined procedure (which might depend on that choices of y). This will be useful, allowing us some control over the order in which we choose the elements x.

From the above it is clear that if we have a set $X = \{f_1, \ldots f_k\}$ of k disjoint pairs of points in W then

$$\mathbb{P}(f_i \in F | f_1 \ldots, f_{i-1} \in F) = \frac{1}{2m - 2i + 1}$$

and so

$$\mathbb{P}(X \subseteq F) = \frac{1}{(2m-1)(2m-3)\ldots(2m-2k+1)} \leq \frac{1}{(2m-2k)^k}. \tag{8.1}$$

Let us define $\lambda = \frac{\sum_i d_i(d_i-1)}{2\sum_i d_i}$ we will show the following:

Theorem 8.3. *Suppose* $\Delta = \max_i\{d_i\} \le n^\alpha$ *where* $\alpha < \frac{1}{6}$*, then*

$$\mathbb{P}(\gamma(F) \text{ is simple}) = (1 + o(1))e^{-\lambda(\lambda+1)}$$

In order to prove the theorem we will need to first show a number of simple lemmas about $\gamma(F)$.

Lemma 8.4. *Suppose* $\Delta = \max_i\{d_i\} \le n^\alpha$ *where* $\alpha < \frac{1}{6}$*. Then with high probability* $\gamma(F)$ *has*

(a) No double loops;

(b) At most $\Delta \log n$ *loops;*

(c) No triple edges;

(d) No adjacent double edges;

(e) At most $\Delta^2 \log n$ *double edges;*

Proof. For (a) we have by (8.1) that

$$\begin{aligned}\mathbb{P}(F \text{ contains a double loop}) &\le \sum_{i=1}^{n} 3\binom{d_i}{4}\left(\frac{1}{2m-4}\right)^2 \\ &\le \frac{\Delta^3}{(2m-4)^2}\sum_{i=1}^{n} d_i \\ &\le \frac{\Delta^3 m}{(2m-4)^2} = o(1)\end{aligned}$$

For (b) we have, with $k = \Delta \log n$,

$$\begin{aligned}\mathbb{P}(F \text{ contains} \ge k \text{ loops}) &\le \mathbb{P}(F \text{ contains a double loop}) + \sum_{\substack{x_1+x_2+\ldots+x_n=k \\ x_i=0,1}} \prod_{i=1}^{n}\left(\binom{d_i}{2}\frac{1}{2m-2k}\right)^{x_i} \\ &\le \left(\frac{\Delta}{2m}\right)^k \sum_{\substack{x_1+x_2+\ldots+x_n=k \\ x_i=0,1}} \prod_{i=1}^{n} d_i^{x_i} + o(1) \\ &\le \left(\frac{\Delta}{2m}\right)^k \frac{(\sum_i d_i)^k}{k!} + o(1) \\ &\le \left(\frac{\Delta e}{k}\right)^k + o(1) = o(1)\end{aligned}$$

where the x_i counts whether or not there is a loop at the vertex i.

For (c) we have

$$\begin{aligned}\mathbb{P}(F \text{ contains a triple edge}) &\le \sum_{i,j\le n} 6\binom{d_i}{3}\binom{d_j}{3}\left(\frac{1}{2m-6}\right)^3 \\ &\le \frac{\Delta^5 m^2}{(2m-6)^3} = o(1)\end{aligned}$$

For (d) we have

$$\begin{aligned}\mathbb{P}(F \text{ contains two adjacent double edges}) &\leq \sum_{i,j,k\leq n} 24\binom{d_i}{4}\binom{d_j}{2}\binom{d_k}{2}\left(\frac{1}{2m-8}\right)^4 \\ &\leq \frac{\Delta^5 m^3}{(2m-8)^4} = o(1)\end{aligned}$$

For (e) we have with $k = \Delta^2 \log n$

$$\begin{aligned}\mathbb{P}(F \text{ has} \geq k \text{ double edges}) &\leq \mathbb{P}(F \text{ contains a triple edge}) + \mathbb{P}(F \text{ contains two adjacent double edges}) \\ &\quad + \sum_{\substack{x_1+x_2+\ldots+x_n=k \\ x_i=0,1}} \prod_{i=1}^{n} \left(\binom{d_i}{2}\frac{\Delta}{2m-2k_1}\right)^{x_i} \\ &\leq \left(\frac{\Delta^2}{m}\right)^k \sum_{\substack{x_1+x_2+\ldots+x_n=k \\ x_i=0,1}} \prod_{i=1}^{n} d_i^{x_i} + o(1) \\ &\leq \left(\frac{2\Delta^2 e}{k}\right)^k + o(1) = o(1).\end{aligned}$$

Indeed, if F has no triple edges or adjacent parallel edges, then we can bound the probability that F has at least k double edges by the probability that at least k of the vertices i are incident to a pair of double edges. For a fixed i, this will happen only if one of the $\binom{d_i}{2}$ pairs x_1, x_2 in W_i are paired to points in the same W_j.

□

8.2 The Switching Lemma

The idea behind the proof of Theorem 8.3 will be to try and move from multigraphs with loops and parallel edges towards simple graphs via reversible local changes, reducing the number of loops or parallel edges. Our hope will be that by counting how many ways we can move from a 'less simple' graph to a 'more simple' graph and vice versa, we can estimate quite precisely how the number of possible multigraphs with a fixed degree sequence changes when we reduce the number of parallel edges or loops, and so estimate the proportion which have none of either.

So, let us define $\Omega_{i,j}$ to be the set of all configurations $F \in \Omega$ such that $\gamma(F)$ has i loops and j pairs of parallel edges and furthermore satisfy (a),(c) and (d) from Lemma 8.4.

Lemma 8.5. *[Switching Lemma] Suppose that $\Delta \leq n^\alpha$ where $\alpha < \frac{1}{6}$. Let $\lambda = \frac{\sum_i d_i(d_i-1)}{2\sum_i d_i}$. Then for any $i \leq \Delta \log n$ and $j \leq \Delta^2 \log n$,*

$$\frac{|\Omega_{i-1,j}|}{|\Omega_{i,j}|} = \left(1 + \tilde{O}\left(\frac{\Delta^3}{n}\right)\right)\frac{i}{\lambda},$$

and

$$\frac{|\Omega_{0,j-1}|}{|\Omega_{0,j}|} = \left(1 + \tilde{O}\left(\frac{\Delta^4}{n}\right)\right)\frac{j}{\lambda^2},$$

where $\tilde{O}$ hides a polylogarithmic factor in n.

Before we prove the lemma, let us show that Theorem 8.3 follows quickly from this.

Proof of Theorem 8.3. By Lemma 8.4 we have that

$$|\Omega| = (1 + o(1)) \sum_{i \leq \Delta \log n} \sum_{j \leq \Delta^2 \log n} |\Omega_{i,j}|.$$

Then, by repeated application of Lemma 8.5, for any $i \leq \Delta \log n$ and $j \leq \Delta^2 \log n$,

$$\begin{aligned} |\Omega_{i,j}| &= \left(1 + \tilde{O}\left(\frac{\Delta^3}{n}\right)\right)^i |\Omega_{0,j}| \frac{\lambda^i}{i!} \\ &= \left(1 + \tilde{O}\left(\frac{\Delta^4}{n}\right)\right)\left(1 + \tilde{O}\left(\frac{\Delta^4}{n}\right)\right)^{2j} |\Omega_{0,0}| \frac{\lambda^i}{i!}\frac{\lambda^{2j}}{j!} \\ &= \left(1 + \tilde{O}\left(\frac{\Delta^6}{n}\right)\right) |\Omega_{0,0}| \frac{\lambda^{i+2j}}{i!j!} \end{aligned}$$

Hence

$$\begin{aligned} |\Omega| &= (1 + o(1)) |\Omega_{0,0}| \sum_{i \leq \Delta \log n} \sum_{j \leq \Delta^2 \log n} \frac{\lambda^{i+2j}}{i!j!} \\ &= (1 + o(1)) |\Omega_{0,0}| e^{\lambda(\lambda+1)} \end{aligned}$$

□

In order to prove the switching lemma we will introduce two operations on configurations, one called an *ℓ-switch*, which will remove loops, and one called a *d-switch*, which will remove double edges.

For the first type of switch we take six points $x_1, x_2, x_3, x_4, x_5, x_6$ in a configuration F such that

- $\{x_1, x_6\}, \{x_2, x_3\}$ and $\{x_4, x_5\} \in F$;
- $\{x_2, x_3\}$ is a loop.

To perform an ℓ-switch we replace F by

$$F' = F \setminus \{\{x_1, x_6\}, \{x_2, x_3\}, \{x_4, x_5\}\} \cup \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}\}.$$

Let us consider a directed graph H on Ω whose arcs are pairs (F, F') such that F' can be obatined from F by performing an ℓ-switch. We would like to consider the bipartite subgraph H' of this consisting of the edges directed from $\Omega_{i,j}$ to $\Omega_{i-1,j}$.

Note that, for any $F \in \Omega_{i,j}$ there are at most $2i|W|^2 = 8im^2$ ℓ-switches that transform F to some F'. Indeed, in order to determine an ℓ switch we have to choose one of the i loops in F (and a labelling of the loop), and then we have at most $|W|^2$ choices for x_1 and x_4 (which determine x_6 and x_5 respectively).

On the other hand, letting $M := \sum_i d_i(d_i - 1)$, for an arbitrary $F' \in \Omega$ there are at most $M|W| = 2mM$ configurations F such that some ℓ-switch on F results in F'. Indeed, the pair $\{x_2, x_3\}$ has to lie in the same W_i, and so there are at most M choices for this pair, and then at most $|W|$ further choices for x_5. However, once we have fixed x_2, x_3 and x_5 the other points are determined.

Hence we know that in H the outdegree of any $F \in \Omega_{i,j}$ is at most $8im^2$ and the indegree of any $F' \in \Omega_{i-1,j}$ is at most $2mM$, and hence the same bounds hold in H'. We would like to show that these are not too far from the actual in/out degrees.

Lemma 8.6. *Let H' be a directed bipartite graph on vertex set $(\Omega_{i,j}, \Omega_{i-1,j})$ with an edge from $F \in \Omega_{i,j}$ to $F' \in \Omega_{i-1,j}$ if there is an ℓ-switch transforming F to F'. Then, for $i \leq \Delta \log n$ and $j \leq \Delta^2 \log n$,*

(I) For all $F \in \Omega_{i,j}$, $d^+_{H'}(F) = \left(1 + \tilde{O}\left(\frac{\Delta^2}{m}\right)\right) 8im^2$;

(II) For all $F' \in \Omega_{i-1,j}$, $d^-_{H'}(F') = \left(1 + \tilde{O}\left(\frac{\Delta^3}{M}\right)\right) 2mM$.

Proof. By the remark before the lemma, it suffices to give lower bounds. Let us first consider (I).

We are interested in when an ℓ-switch applied to $F \in \Omega_{i,j}$ produces an $F' \in \Omega_{i-1,j}$. For such a thing to happen we need that F' has $i - 1$ loops, j parallel edges, and satisfies (a),(c) and (d) from Lemma 8.4.

We note that this will definitely hold as long as

- $\{x_1, x_6\}$ and $\{x_4, x_5\}$ are neither a loop nor a parallel edge in F;
- None of $\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}$ are a loop or a parallel edge in F'.

So, let us count the sextuples $\{x_1, \ldots, x_6\}$ for which at least one of these conditions fails to hold.

We first note that there are at most $O(im\Delta)$ tuples such that one of $\{x_1, x_4, x_5, x_6\}$ lies in the same cell as $\{x_2, x_3\}$. Indeed, after choosing the loop $\{x_2, x_3\}$ in $\leq 2i$ ways, there are at most $2\Delta m$ choices for $\{x_1, x_4\}$ such that one of $\{x_1, x_4, x_5, x_6\}$ lies in the same cell W_i as $\{x_2, x_3\}$. In what follows we will assume that we exclude these tuples.

Since there are i loops and j parallel edges in F, the number such tuples where $\{x_1, x_6\}$ or $\{x_4, x_5\}$ forms a loop or a parallel edge in F is at most $16i(\Delta + \Delta^2)(\log n)m$, since we have $2i$ choices for the loop $\{x_2, x_3\}$ and then at most $2i + 2j \leq 2(\Delta + \Delta^2)\log n$ choices for the pair forming a loop or a parallel edge and at most $2m$ choices for the other pair. Hence there are at most $\tilde{O}(im\Delta^2)$ many such bad tuples.

The number of such tuples where $\{x_1, x_2\}$ or $\{x_3, x_4\}$ form a loop in F' is at most $8i\Delta m$. Indeed, after choosing the loop $\{x_2, x_3\}$ there are at most Δ many choices for x_1 which forms a loop with x_2 and then at most $|W| = 2m$ choices for x_4. A similar bound holds if $\{x_3, x_4\}$ forms a loop. Hence there are at most $O(im\Delta)$ many such bad tuples.

The number of such tuples where $\{x_1, x_2\}$ or $\{x_3, x_4\}$ form a parallel edge in F' is at most $8i\Delta^2 m$. Indeed, given $\{x_2, x_3\}$, since none of $\{x_1, x_4, x_5, x_6\}$ lies in the same cell W_i, there are at most Δ^2 choices for x_1 such that $\{x_1, x_2\}$ forms a parallel edge. Indeed, since $\{x_1, x_4, x_5, x_6\}$ do not lie in W_i, as long as x_1 is not chosen from the same cell as the neighbour (in F, and so also in F') of one of the $\leq \Delta$ many xs in W_i, then $\{x_1, x_2\}$ will not form a parallel edge. After choosing x_1 we then have at most $|W|$ many choices for x_4. A similar argument holds for x_4. Hence there are at most $O(im\Delta^2)$ many such bad tuples.

The number of such tuples where $\{x_5, x_6\}$ forms a loop in F' is at most $4i\Delta m$. Indeed, after choosing the loop $\{x_2, x_3\}$ and the edge $\{x_1, x_6\}$ in F there are at most Δ many choices for x_5 which lies in the same W_i as x_6. Hence there are at most $O(im\Delta)$ many such bad tuples.

The number of such tuples where $\{x_5, x_6\}$ forms a parallel edge in F' is at most $4i\Delta^2 m$. Indeed, given $\{x_2, x_3\}$ and $\{x_1, x_6\}$, where we can assume that none of $\{x_1, x_2, x_3\}$ lie in the same cell as x_6, there are at most Δ many choices for x_5 so that $\{x_5, x_6\}$ forms a parallel edge in F'. Indeed, there are at most Δ many xs in the same W_i as x_6, and if we choose an x_5 which is not in the same cell as the neighbour of any of these points in F (and so also in F'), then $\{x_5, x_6\}$ will not form a parallel edge in F'. Hence there are at most $O(im\Delta^2)$ many such bad tuples.

Hence the total number of bad tuples is $\tilde{O}(im\Delta^2) = \tilde{O}\left(\frac{\Delta^2}{m}\right) 8im^2$.

So, let us move onto (II). For every $F' \in \Omega_{i-1,j}$ and all $x_1, \ldots, x_6 \in W$ such that $\phi(x_2) = \phi(x_3)$ and $\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\} \in F'$ there is a unique $F \in \Omega$ such that an ℓ-switch applied to $x_1, \ldots, x_6$ transforms F to F'. This F will have i loops, j parallel edges and satisfy (a),(c) and (d) from Lemma 8.4 as long as

- None of $\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}$ are loops or parallel edges in F';
- Neither $\{x_1, x_6\}$ nor $\{x_4, x_5\}$ becomes a loop or parallel edge in F;
- $W_{\phi(x_2)}$ doesn't contain a loop in F'.

So, let us count the sextuples $\{x_1, \ldots, x_6\}$ for which at least one of these conditions fails to hold.

Since there are $i-1$ loops and j parallel edges in F', the number such tuples where $\{x_1, x_2\}$ or $\{x_3, x_4\}$ forms a loop or a parallel edge in F' is at most $8(i+j-1)\Delta m$, since we have $2(i+j-1)$ choices for the one forming a loop or parallel edge, then at most Δ choices for the element of $\{x_2, x_3\}$ not yet chosen, since $\phi(x_2) = \phi(x_2)$, which determines its partner in F', and finally at most $2m$ choices for the remaining pair. Hence there are at most $\tilde{O}(\Delta^3 m)$ many such bad tuples.

The number of such tuples where $\{x_5, x_6\}$ is a loop or parallel edge in F' is at most $4(i+j-1)\Delta m$. Indeed, There are at most $2(i+j-1)$ choices for x_5 and x_6, and then at most $2m\Delta$ choices for x_2 and x_3, which then determines x_1 and x_4. Hence there are at most $\tilde{O}(\Delta^3 m)$ many such bad tuples.

The number of such tuples where $\{x_1, x_6\}$ becomes a loop in F is at most $2m\Delta^2$. Indeed, there are at most $2m$ choices for x_5, x_6, and then at most Δ choices for x_1, which determines x_2,

and hence at most Δ choices for x_3, which determines x_4. A similar calculation holds if $\{x_4, x_5\}$ becomes a loop in F. Hence there are at most $O(\Delta^2 m)$ many such bad tuples.

The number of such tuples where $\{x_1, x_6\}$ becomes a parallel edge in F is at most $3m\Delta^3$. Indeed, we first note that there are at most $O(m\Delta^2)$ tuples such that one of x_1, x_4 lies in the same cell as x_5 or x_6. Assuming that this doesn't hold, once we've chosen x_5 and x_6, as long as we don't choose x_1 from the same cell as the neighbour of some vertex in the same cell as x_6, then $\{x_1, x_6\}$ will not form a parallel edge in F' (using the fact that $\{x_2, x_3\}$ is a loop, and so cannot be a parallel edge). Hence, there are at most Δ^2 bad choices for x_1. After we've chosen x_1, this fixes x_2 and there is at most Δ choices for x_3, which determines x_4. A similar calculation holds if $\{x_4, x_5\}$ becomes a parallel edge in F. Hence there are at most $O(\Delta^3 m)$ many such bad tuples.

Finally, the number of such tuples where $W_{\phi(x_2)}$ contains a loop in F' is at most $2i\Delta^2 m$. Indeed, there are at most i choices for the cell $\phi(x_2)$, and then at most Δ^2 choices for x_2 and x_3, which then determines x_1 and x_4. Finally there are at most $2m$ choices for x_5 and x_6. Hence there are at most $\tilde{O}(\Delta^3 m)$ many such bad tuples.

Hence the total number of bad tuples is $\tilde{O}(\Delta^3 m) = \tilde{O}\left(\frac{\Delta^3}{M}\right) 2mM$. □

The first part of Lemma 8.5 now follows by double counting the edges of H'. Indeed, on the one hand

$$e(H') = \sum_{F \in \Omega_{i,j}} d^+_{H'}(F) = |\Omega_{i,j}| 8im^2 \left(1 + \tilde{O}\left(\frac{\Delta^2}{m}\right)\right),$$

and on the other hand

$$e(H') = \sum_{F' \in \Omega_{i-1,j}} d^-_{H'}(F') = |\Omega_{i-1,j}| 2mM \left(1 + \tilde{O}\left(\frac{\Delta^3}{M}\right)\right).$$

Hence, since $m, M = \Omega(n)$,

$$\frac{|\Omega_{i-1,j}|}{|\Omega_{i,j}|} = \left(1 + \tilde{O}\left(\frac{\Delta^3}{n}\right)\right) \frac{4mi}{M}.$$

In order to prove the second half of Lemma 8.5 we will need to introduce a second type of switch, one that removes double edges, called a *d*-switch.

For this switch we take eight points $x_1, \ldots, x_8$ in a configuration F such that

- $\{x_1, x_5\}, \{x_2, x_6\}, \{x_3, x_7\}, \{x_4, x_8\} \in F$;
- $\phi(x_2) = \phi(x_3)$ and $\phi(x_6) = \phi(x_7)$;

To perform a *d*-switch we replace F by

$$F' = F \setminus \{\{x_1, x_5\}, \{x_2, x_6\}, \{x_3, x_7\}, \{x_4, x_8\}\} \cup \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}, \{x_7, x_8\}\}.$$

As before, for most tuples this will reduce the number of double edges in F by one. More precisely, if we consider the bipartite graph H from Ω to Ω whose arcs are pairs (F, F') such

that F' can be obtained from F by performing a d-switch, then we are interested in the subgraph H' of H between $\Omega_{0,j}$ and $\Omega_{0,j-1}$.

Again, for any $F \in \Omega_{0,j}$ there are at most $16jm^2$ many d-switches that transform F to some F'. Indeed, in order to determine a d-switch we have to first choose on of the j many double edges in F (and choose from one of the 4 possible labellings thereof) and then we have at most $|W|^2$ many choices for x_1 and x_4 (which determine x_5 and x_8 respectively).

On the other hand, for an arbitrary $F' \in \Omega$ there are at most M^2 many configurations F such that some d-switch on F results in F'. Indeed, the both $\{x_2, x_3\}$ and $\{x_6, x_7\}$ have to lie in the same cell, and so there are at most M^2 choices for these two pairs, which then determine the other points.

As with the previous case we will see that these estimates are actually not far from the true in/out-degrees in H'.

Lemma 8.7. *Let H' be a directed bipartite graph on vertex set $(\Omega_{0,j}, \Omega_{0,j-1})$ with an edge from F to F' if there is an d-switch transforming F to F'. Then, for $j \leq \Delta^2 \log n$,*

(I) For all $F \in \Omega_{0,j}$, $d^+_{H'}(F) = \left(1 + \tilde{O}\left(\frac{\Delta^2}{m}\right)\right) 16jm^2$;

(II) For all $F' \in \Omega_{0,j-1}$, $d^-_{H'}(F') = \left(1 + \tilde{O}\left(\frac{\Delta^4}{M}\right)\right) M^2$.

Proof. The proof follows very similar lines to Lemma 8.6, we will just outline the key claims, and leave their proofs as an exercise.

We are interested in when an d-switch applied to $F \in \Omega_{0,j}$ produces an $F' \in \Omega_{0,j-1}$. For such a thing to happen we need that F' has no loops, $j-1$ paralell edges, and satisfies (a),(c) and (d) from Lemma 8.4. Note that this will definitely hold as long as

- Neither $\{x_1, x_5\}$ or $\{x_4, x_8\}$ are a parallel edge in F;
- None of $\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}$ or $\{x_7, x_8\}$ are a loop or a parallel edge in F'.

Exercise : Show that there are at most $\tilde{O}(j\Delta^2 m)$ many bad tuples.

Similarly, given an $F' \in \Omega_{0.j-1}$ and a tuple $x_1, \ldots, x_8$ where $\phi(x_2) = \phi(x_3), \phi(x_6) = \phi(x_7)$ and $\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}, \{x_7, x_8\} \in F'$ there is a unique $F \in \Omega$ such that performing a d-switch to the tuple $x_1, \ldots, x_8$ in F transforms it to F', and we are interested in when $F \in \Omega_{0,j}$. This will definitely hold as long as

- None of $\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}$ or $\{x_7, x_8\}$ is a parallel edge in F';
- Neither $\{x_1, x_5\}$ nor $\{x_4, x_8\}$ form a loop or a double edge in F;
- Neither $\{x_2, x_6\}$ nor $\{x_3, x_7\}$ form a triple edge in F or are incident with another double edge in F or F'.

Exercise : Show that there are at most $\tilde{O}(\Delta^4 M)$ many bad tuples. □

Finally, given Lemma 8.7, the second part of Lemma 8.5 follows as before from double counting

$$|\Omega_{0,j-1}|M^2\left((1+\tilde{O}(\frac{\Delta^4}{M})\right)) = \sum_{F'\in\Omega_{0,j-1}} d^-_{H'}(F') = \sum_{F\in\Omega_{0,j}} d^+_{H'}(F) = |\Omega_{0,j}|16jm^2\left(1+\tilde{O}(\frac{\Delta^2}{m})\right)$$

and hence, since $m, M = \Omega(n)$,

$$\frac{|\Omega_{0,j-1}|}{|\Omega_{0,j}|} = \left(1+\tilde{O}\left(\frac{\Delta^4}{n}\right)\right)\frac{16m^2j}{M^2}.$$

As a simple application of Theorem 8.3 we can estimate the asymptotic number of r-regular graphs on n vertices.

Theorem 8.8. *Let $r \geq 3$ then the number of r-regular graphs on n vertices is*

$$|\mathcal{G}_{n,\overline{r}}| = (1+o(1))\sqrt{2}e^{-\frac{(r^2-1)}{4}}\left(\frac{r^{\frac{r}{2}}}{e^{\frac{r}{2}}r!}\right)^n n^{\frac{rn}{2}}.$$

Proof. Since $\mathbb{P}(\gamma(F) \text{ is simple}) = (1+o(1))e^{-\lambda(\lambda+1)}$, and $|\gamma^{-1}(G)| = (r!)^n$,it follows that

$$|\Omega|(1+o(1))e^{-\lambda(\lambda+1)} = (r!)^n|\mathcal{G}_{n,\overline{r}}|.$$

Note that, since $d_i = r$ for each i, we have that $\lambda = \frac{r-1}{2}$ and hence $\lambda(\lambda+1) = \frac{r^2-1}{4}$. However, we can count the number of configurations on W, as it's simply the number of perfect matchings on rn many points. We can calculate this number as

$$(rn-1)(rn-3)\dots(1) = \frac{(rn)!}{(rn)(rn-2)(rn-4)\dots 2} = \frac{(rn)!}{2^{\frac{rn}{2}}\left(\frac{rn}{2}\right)!}$$

Hence

$$\begin{aligned}
|\mathcal{G}_{n,\overline{r}}| &= (1+o(1))e^{-\frac{(r^2-1)}{4}}\frac{(rn)!}{2^{\frac{rn}{2}}\left(\frac{rn}{2}\right)!(r!)^n} \\
&= (1+o(1))e^{-\frac{(r^2-1)}{4}}\frac{\sqrt{2\pi rn}\left(\frac{rn}{e}\right)^{rn}}{\sqrt{\pi rn}\left(\frac{rn}{2e}\right)^{\frac{rn}{2}}2^{\frac{rn}{2}}(r!)^n} \\
&= (1+o(1))\left(\frac{r^{\frac{r}{2}}}{e^{\frac{r}{2}}r!}\right)^n n^{\frac{rn}{2}}.
\end{aligned}$$

□

8.3 Connectivity of Regular Graphs

Theorem 8.3 allows us to deduce that properties that hold in the multigraph $\gamma(F)$ with high probability, also hold in $G(n,\overline{d})$ with high probability (for suitably well behaved $\overline{d}$).

Indeed, given some property $\mathcal{P}$ of multi-graphs, then whenever $\mathbb{P}(\gamma(F)\in\mathcal{P})=o(1)$ then by Theorem 8.3

$$\mathbb{P}(G(n,\overline{d})\in\mathcal{F})=\mathbb{P}(\gamma(F)\in\mathcal{F}|\gamma(F)\text{ is simple})\leq\frac{\mathbb{P}(\gamma(F)\in\mathcal{F})}{\mathbb{P}(\gamma(F)\text{ is simple})}=\frac{\mathbb{P}(\gamma(F)\in\mathcal{F})}{(1+o(1))e^{-\lambda(\lambda+1)}}=o(1),$$

as long as the probability $\mathbb{P}(\gamma(F)\in\mathcal{F})$ is sufficiently small as a function of λ. Note, $\lambda=\frac{\sum_i d_i(d_i-1)}{2\sum_i d_i}\leq\frac{\Delta-1}{2}$ and so, if Δ is bounded, then *every* property which holds whp in the configuration model holds also whp in $G(n,\overline{d})$.

Using this we can talk about properties of $G(n,\overline{r})$.

Theorem 8.9. *Let $r\geq 3$, then with high probability $G(n,\overline{r})$ is r-connected.*

Proof. We might hope to show that in fact, with high probability $\gamma(F)$ is r-connected. However, there are ways in which $\gamma(F)$ can fail to be r-connected which can't happen in $G(n,\overline{r})$, and in fact this will happen with a non-zero probability. For example, if there is any loop or parallel edge, then there is a vertex v with $|N(v)|<r$, and so $\gamma(F)$ cannot be r-connected.

However, we still want to work within the configuration model. So, our plan will be to find structures which witness that a graph is not r-connected, and show that the expected number of such structures appearing in $\gamma(F)$ is $o(1)$, and hence by Markov's inequality with high probability there are no such structures. Whilst this doesn't imply that $\gamma(F)$ is r-connected, it does imply that with high probability $G(n,\overline{r})$ contains no such structures, and hence is r-connected.

So, how can a graph G fail to be r-connected? Precisely when there is a set B of size less than r whose deletion splits the graph into multiple components. By choosing the smallest of these components, and taking B to be its neighbourhood in the graph, it follows that we can partition the vertex set $V=[n]$ into three disjoint parts A,B,C such that $B=N(A)$ and $|B|<r$. In what follows we will write $a:=|A|,b:=|B|$ and $c:=|C|$. Furthermore, since A is the smallest component of $G\setminus B$, it holds that $a<n/2$.

We will show, splitting into a number of cases, that whp there are no such triples in $\gamma(F)$ with $a\geq 2$. This will be sufficient, since $G(n,\overline{r})$ deterministically cannot contain such a triple with $a=1$, since $|N(v)|=r$ for all v.

Suppose first that $a\in\{2,3\}$. Then, in $G(n,\overline{r})$ the set $S=A\cup B$, whose size $s:=|S|$ is at most $r+2$, must span at least $ar-\binom{a}{2}\geq s+1$ edges. However, as we will see, it is very unlikely that a set of size $s\leq r+2$ spans $s+1$ edges in $\gamma(F)$.

Indeed, given a fixed set $S\subseteq[n]$ of size s let us estimate the probability that S contains $s+1$ edges in $\gamma(F)$. The s cells $\{W_{i_1},\dots,W_{i_s}\}$ in W corresponding to S contain rs points in total. Let us call the union of these cells Z. If S spans at least $s+1$ edges in $\gamma(F)$, then there must be some subset $T=\{t_1,\dots,t_{s+1}\}\subseteq Z$ of size $s+1$ such that each point in T is paired to another point in Z.

For each choice of T, this probability can be bounded by

$$\left(\frac{rs-(s+1)}{rn-1}\right)\left(\frac{rs-(s+2)}{rn-3}\right)\cdots\left(\frac{rs-(2s+1)}{rn-(2s-1)}\right)\leq\left(\frac{2rs}{rn}\right)^{s+1},$$

for large enough n, using the fact that $r,s\in\mathbb{N}$. Hence,

$$\begin{aligned}
&\mathbb{P}(\text{There exists } S \text{ with } s\leq r+2 \text{ vertices containing } s+1 \text{ edges})\\
&\leq\sum_{s=2}^{r+2}\binom{n}{s}\binom{rs}{s+1}\left(\frac{2rs}{rn}\right)^{s+1}\\
&\leq\sum_{s=2}^{r+2}n^s 2^{rs}(2s)^{s+1}n^{-(s+1)}\\
&\leq r2^{rs}(2s)^{s+1}\frac{1}{n}\\
&=o(1).
\end{aligned}$$

The second case we will consider is $4\leq a\leq ne^{-100}$. In this case we note that there are at least $\frac{ra+b}{2}$ many edges incident with the set A in $G(n,\overline{r})$. Indeed, each vertex in A is incident to r edges, and each vertex in $B=N(A)$ is incident to at least one edge which is incident to A.

Given a fixed set A of size $4\leq a\leq ne^{-100}$ and a fixed set B of size $b\leq r-1$, by a similar calculation as before, the probability that there are $\frac{ra+b}{2}$ edges in $\gamma(F)$, with one endpoint in A and the other endpoint in $A\cup B$ is at most

$$\binom{ra}{\frac{ra+b}{2}}\prod_{i=1}^{\frac{ra+b}{2}}\frac{r(a+b)-i}{rn-2i}\leq\binom{ra}{\frac{ra+b}{2}}\left(\frac{2(a+b)}{n}\right)^{\frac{ra+b}{2}}$$

Hence,

$$\begin{aligned}
\mathbb{P}(\text{There exists such a pair } A,B)&\leq\sum_{a=4}^{ne^{-100}}\sum_{b=0}^{r-1}\binom{n}{a}\binom{n}{b}\binom{ra}{\frac{ra+b}{2}}\left(\frac{2(a+b)}{n}\right)^{\frac{ra+b}{2}}\\
&\leq\sum_{a=4}^{ne^{-100}}\sum_{b=0}^{r-1}n^{-\frac{ra}{2}+a+\frac{b}{2}}\frac{e^{a+b}}{a^a b^b}2^{ra+\frac{ra+b}{2}}(a+b)^{\frac{ra+b}{2}}\\
&\leq C_r\sum_{a=4}^{ne^{-100}}n^{-\frac{ra}{2}+a+\frac{r}{2}}\frac{e^a}{a^a}2^{\frac{3ra}{2}}(a+r)^{\frac{ra+r}{2}},
\end{aligned}$$

where C_r is some constant which depends only on r. Now

$$\left(\frac{a+r}{r}\right)^{\frac{r}{2}}=\left(1+\frac{a}{r}\right)^{\frac{r}{2}}\leq e^{\frac{a}{2}},$$

and similarly

$$\left(\frac{a+r}{a}\right)^{\frac{ra}{2}}=\left(1+\frac{r}{a}\right)^{\frac{ra}{2}}\leq e^{\frac{r^2}{2}}.$$

Hence,

$$(a+r)^{\frac{ra+r}{2}}\leq C_r a^{\frac{ra}{2}}e^{\frac{a}{2}},$$

where again C_r only depends on r (but might be a different constant).

It follows that, with C_r being some constant (perhaps changing line to line)

$$\begin{aligned}\mathbb{P}(\text{There exists such a pair } A,B) &\leq C_r \sum_{a=4}^{ne^{-100}} n^{-\frac{ra}{2}+a+\frac{r}{2}} e^{\frac{3a}{2}} 2^{\frac{3ra}{2}} a^{a\left(\frac{r}{2}-1\right)} \\ &\leq C_r \sum_{a=4}^{ne^{-100}} \left(e^{\frac{3}{2}} 2^{\frac{3r}{2}} \left(\frac{n}{a}\right)^{1-\frac{r}{2}} n^{\frac{r}{2a}}\right)^a.\end{aligned}$$

For small a, say $a \leq \log^2 n$, we note that since $r \geq 3$ and $a \geq 4$, $\frac{r}{2a} < \frac{r}{2} - 1$ and hence

$$n^{1-\frac{r}{2}+\frac{r}{2a}} a^{\frac{r}{2}-1} = o(1).$$

On the other hand, for $\log^2 n \leq a \leq ne^{-100}$, $n^{\frac{r}{2a}} = 1 + o(1)$ and $\left(\frac{n}{a}\right)^{1-\frac{r}{2}} \leq \left(\frac{n}{a}\right)^{-\frac{r}{8}} \leq e^{-10r}$. It follows that

$$\mathbb{P}(\text{There exists such a pair } A,B) = o(1).$$

The final case we will consider is $ne^{-10} \leq a \leq \frac{n}{2}$. Suppose there are exactly s edges between B and C, where $s \leq br \leq r^2$. Let us estimate the probability that this occurs in $\gamma(F)$.

Suppose we have chosen the vertex sets A and B, which determines the set C. Let $\hat{A}, \hat{B}$ and $\hat{C}$ be the ra, rb and rc many points in W corresponding to vertices in A,B and C respectively. Having chosen A and B, we have to choose s points in $\hat{B}$ which will be paired to points in $\hat{C}$, pair the remaining $r(a+b) - s$ points in $\hat{A} \cup \hat{B}$ to each other and then pair the remaining $r(n-a-b)+s$ points in $\hat{B} \cup \hat{C}$ together. If we let

$$\psi(2m) = \frac{(2m)!}{m!2^m} \approx \sqrt{2}\left(\frac{2m}{e}\right)^m$$

be the number of matchings on $2m$ points then we see that the probability of getting such a pairing for the chosen A, B, C and s points in $\hat{B}$ is at most

$$\frac{\psi(r(a+b)-s)\psi(r(n-a-b)+s)}{\psi(rn)} \approx 2\frac{(r(a+b)-s)^{\frac{r(a+b)-s}{2}}(r(n-a-b)+s)^{\frac{r(n-a-b)+s}{2}}}{(rn)^{\frac{rn}{2}}}$$

Hence we can estimate the probability that we contain such a triple A, B, C as follows, where again C_r is some constant (perhaps changing line to line)

$$\begin{aligned}&\sum_{a,b,s} \binom{n}{a}\binom{n}{b}\binom{rb}{s} 2\frac{(r(a+b)-s)^{\frac{r(a+b)-s}{2}}(r(n-a-b)+s)^{\frac{r(n-a-b)+s}{2}}}{(rn)^{\frac{rn}{2}}} \\ &\leq C_r \sum_{a,b,s} \left(\frac{en}{a}\right)^a n^b \frac{(ra)^{\frac{r(a+b)-s}{2}}(r(n-a))^{\frac{r(n-a-b)+s}{2}}}{(rn)^{\frac{rn}{2}}},\end{aligned}$$

where we used the fact that

$$\left(\frac{r(a+b)-s}{ra}\right)^{\frac{r(a+b)-s}{2}} = \left(1+\frac{rb-s}{ra}\right)^{\frac{ra}{2}+\frac{rb-s}{2}} = (1+o(1))\exp\left(-\frac{rb-s}{2}\right) \leq C_r,$$

and a similar fact for the term $r(n-a-b)+s)^{\frac{r(n-a-b)+s}{2}}$. Continuining, we can bound the probability above by

$$\leq C_r \sum_{a,b,s} \left(\frac{en}{a}\right)^a n^b \frac{a^{\frac{r(a+b)-s}{2}} (n-a)^{\frac{r(n-a-b)+s}{2}}}{(n)^{\frac{rn}{2}}}$$
$$\leq C_r \sum_{a,b,s} \left(\frac{en}{a}\right)^a n^b \left(\frac{a}{n}\right)^{\frac{ra}{2}} \left(1-\frac{a}{n}\right)^{\frac{r(n-a)}{2}},$$

where we used the fact that

$$\left(\frac{a}{n-a}\right)^{\frac{rb-s}{2}} \leq \left(2e^{-10}\right)^{\frac{rb-s}{2}} \leq C_r.$$

Continuing, we can bound the probability above by

$$\leq C_r \sum_{a,b,s} \left(\frac{en}{a} n^{\frac{b}{a}} \left(\frac{a}{n}\right)^{\frac{r}{2}} e^{-\frac{r}{2}(1-\frac{a}{n})}\right)^a$$
$$\leq C_r \sum_{a} \left(2e \left(\frac{a}{n}\right)^{\frac{r}{2}-1} e^{-\frac{r}{2}(1-\frac{a}{n})}\right)^a$$
$$\leq C_r \sum_{a} \left(2e^{-5}\right)^a$$
$$= o(1),$$

since $\frac{a}{n} \leq e^{-10}$, $\frac{r}{2}-1 \geq \frac{1}{2}$ and $\frac{r}{2}\left(1-\frac{a}{n}\right) \geq 1$. □

As a simple corollary we can conclude that with high probability $G(n,\overline{r})$ contains a perfect matching (if n is even). Indeed, it is a relatively simple consequence of Tutte's theorem that every r-regular, r-connected graph on an even number of vertices contains a perfect matching.

Theorem 8.10 (Tutte's Theorem)**.** *$G=(V,E)$ has a perfect matching if and only if for every subset $U \subseteq V$, $G[V \setminus U]$ has at most $|U|$ many connected components with an odd number of vertices.*

Corollary 8.11. *Let $r \geq 3$ and n be even, then with high probability $G(n,\overline{r})$ contains a perfect matching.*

Proof. By Theorem 8.9 with high probability $G := G(n,\overline{r})$ is r-connected. Furthermore, since vertex-connectivity is always larger than edge-connectivity, it follows that G is also r-edge-connected.

Suppose then that there is some subset $U \subseteq V$ such that $G[V \setminus U]$ has more than $|U|$ many connected components with an odd number of vertices. Since G is r-connected, it follows that $|U| > r$, else $G[V \setminus U]$ has exactly one component.

Suppose then that $C_1, \ldots, C_m$ are the components of $G[V \setminus U]$, and so $m > |U| > r$. However, since G is r-edge-connected, there must be at least r edges between U and C_i for each i. It follows that at least rm many edges meet U, however, since G is r-regular, at most $r|U|$ edges meet U, but $r|U| < rm$, a contradiction. □

So, in contrast to the uniform models, even though $G(n,\overline{r})$ is quite sparse, with high probability it will be connected, and contain a perfect matching. In fact even moreso it can be shown that it will contain a Hamiltonian cycle.

Indeed, we can calculate at least the expected number of Hamiltonian cycles in $G^*(n,\overline{r})$. Let us be a bit more general, and calculate the expected number of cycles of length k for any $k \leq n$. Each k-cycle in in $G^*_{n,\overline{r}}$ comes from a set of k pairs in a configuration whose endpoints match up in the correct way. In a slight abuse of notation we will call such a set of pairs a *k-cycle* on W. By symmetry, the probability that any particular k-cycle is included in an configuration is some fixed p_k and so the expected number of k-cycle in a configuration $\mathbb{E}(Z_k) = a_k p_k$, where a_k is the number of possible k-cycles on W.

In order to calculate a_k we can choose a k-cycle by first choosing an (ordered) sequence of k cells $W_{i_1},\ldots,W_{i_k}$, and in each cell choosing an (ordered) pair (x_{i_j},y_{i_j}) of points. This gives rise to a k-cycle given by $\{(y_{i_j},x_{i_{j+1}})\colon j=1\ldots k\}$ where $x_{i_{k+1}}=x_1$. This gives rise to a rooted, oriented cycle, and so each k-cycle is counted precisely $2k$ many times. It follows that

$$2ka_k = (n)_k(r(r-1))^k.$$

For small k we have that $p_k \approx \left(\frac{1}{|W|}\right)^k \approx (rn)^{-k}$ and hence

$$\mathbb{E}(Z_k) \approx \frac{1}{2k}(r-1)^k.$$

In order to calculate $\mathbb{E}(Z_n)$ we have to approximate p_n more precisely. We have that

$$p_n = \frac{1}{rn-1}\frac{1}{rn-1}\cdots,\frac{1}{rn-2n+1} = \frac{\psi(rn-2n)}{\psi(rn)}$$

where $\psi(2m) = \frac{(2m)!}{m!2^m} \approx \sqrt{2}\left(\frac{2m}{e}\right)^m$ as in the proof of Theorem 8.9. It follows that

$$p_n \approx e^n n^{-n}(r-2)^{\frac{rn}{2}-n}r^{-\frac{rn}{2}}$$

and so

$$\begin{aligned}
\mathbb{E}(Z_n) &\approx \frac{1}{2n}n!(r(r-1))^n e^n n^{-n}(r-2)^{\frac{rn}{2}-n}r^{-\frac{rn}{2}}\\
&\approx \frac{1}{2n}\sqrt{2\pi n}\left(\frac{n}{e}\right)^n (r(r-1))^n n^{-n}e^n(r-2)^{\frac{rn}{2}-n}r^{-\frac{rn}{2}}\\
&\approx \sqrt{\frac{\pi}{2n}}\left(r^{1-\frac{r}{2}}(r-1)(r-2)^{\frac{r}{2}-1}\right)^n.
\end{aligned}$$

The term in the brackets can be seen to be larger than one if $r \geq 3$. Indeed when $r=3$ it can be calculated to be $\frac{2}{\sqrt{3}} > 1$, and the ratio of successive terms can be approximated to be larger than one.

Hence, $\mathbb{E}(Z_n) \to \infty$ as $n \to \infty$ and so we expect $G^*_{n,\overline{r}}$ to have many Hamiltonian cycles when $r \geq 3$. With some quite careful work the variance of Z_n can also be computed, however the ratio

$$\frac{\mathrm{Var}(Z_n)}{\mathbb{E}(Z_n)^2} \to \frac{2}{r-2}$$

does not tend to 0, so we can't quite use Chebyshev's inequality to show that with high probability there is a Hamiltonian cycle. However, using a more refined version of the second moment method, allowing them to condition on the non-existence of loops and parallel edges, Robinson and Womald were able to calculate precisely the limiting distribution of the number of Hamiltonian cycles in $G(n, \overline{r})$ and hence show that with high probability $G(n, \overline{r})$ contains a Hamiltonian cycle.

Theorem 8.12. *Let $r \geq 3$ then with high probability $G(n, \overline{r})$ contains a Hamiltonian cycle.*

8.4 Contiguity

As we've seen, with high probability $G^*(n, \overline{3})$ contains a Hamiltonian cycle H, and so in fact $G^*(n, \overline{3}) = H \cup M$ where M is a 1-regular subgraph, that is, a matching. Perhaps a natural question to ask is, if we choose a Hamiltonian cycle H and a matching M uniformly at random, then how does the distribution of $M \cup H$ compare to that of $G^*(n, \overline{3})$?

We note that they cannot be identical in distribution, since $M \cup H$ surely contains a matching, but there is a (vanishingly small, but non-zero) probability that $G^*(n, \overline{3})$ doesn't. However, suppose we allow ourselves a slighty broader definition of 'the same', and say two distributions G_1 and G_2 on (multi)-graphs are *contiguous* is a property hold with high probability for G_1 if and only if it holds for G_2. Will $M \cup H$ and $G^*(n, \overline{3})$ be continguous?

In fact, there is a slightly problem here, in that $M \cup H$ surely has no loops, but $G^*(n, \overline{3})$ contains a loop with positive probability. However, for $r > 3$ this is no longer a problem, and the expected statement does hold, and when $r = 3$ this is the only problem.

Theorem 8.13. *Let $r \geq 4$ then the random multigraph $H \cup G^*(n, \overline{r-2})$ where H is a uniformly chosen Hamiltonian cycle is contiguous with $G^*(n, \overline{r})$.*

Furthermore if we let $G'(n, \overline{3})$ be $(G^*(n, \overline{3}) | G^*$ has no loops$)$ then $H \cup M$ is contiguous with $G'(n, \overline{3})$, where M is a uniformly chosen perfect matching.

Using some standard probabilistic tools one can lift these results from multigraphs to simple graphs as follows, letting $\oplus$ be the operation of unioning two graphs and then simplifying the resulting multigraph.

Theorem 8.14. *Let $r \geq 3$ then the random multigraph $H \oplus G(n, \overline{r-2})$ where H is a uniformly chosen Hamiltonian cycle is contiguous with $G(n, \overline{r})$.*