

# Random Graphs (2017)

Yuval Filmus

February 21, 2019

Most of these lecture notes are based on the textbook of Frieze and Karoński [FK16]. There is an essential difference between the approach we take and theirs: they work in  $G(n, m)$ , and we work in  $G(n, p)$ . However, even they work mostly in  $G(n, p)$  and then deduce results in  $G(n, m)$ .

I thank the students in my class and my TA, Itay Hazan, for help with these notes.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Ramsey graphs . . . . .	3
1.2	The countable random graph and zero-one laws . . . . .	3
<b>2</b>	<b>Sparse random graphs</b>	<b>4</b>
2.1	Graph couplings . . . . .	5
2.2	Evolution of random graphs . . . . .	5
2.3	Forest regime . . . . .	6
2.4	Unicyclic regime . . . . .	6
<b>3</b>	<b>Subcritical regime</b>	<b>7</b>
3.1	Bonus: Proof of Stirling’s formula . . . . .	10
3.1.1	Warm-up: harmonic numbers . . . . .	10
3.1.2	Stirling’s formula . . . . .	11
3.2	Bonus: Proof of Cayley’s formula . . . . .	11
<b>4</b>	<b>Supercritical regime</b>	<b>12</b>
<b>5</b>	<b>Critical regime</b>	<b>15</b>
<b>6</b>	<b>Connectivity</b>	<b>19</b>
<b>7</b>	<b>Subgraphs</b>	<b>22</b>
7.1	Subgraph thresholds . . . . .	22
7.2	Subgraph counts . . . . .	23
<b>8</b>	<b>Thresholds</b>	<b>24</b>
8.1	Sharp and coarse thresholds . . . . .	24
8.2	Friedgut–Kalai threshold theorem . . . . .	25
<b>9</b>	<b>Cliques and coloring</b>	<b>26</b>
9.1	Cliques . . . . .	26
9.2	Chromatic number . . . . .	28
9.3	Finding cliques in random graphs . . . . .	30

<b>10 Planted clique</b>	<b>31</b>
10.1 Degree-based algorithm . . . . .	31
10.2 More on the maximal degree of a graph* . . . . .	32
10.3 Spectral norm of random graph . . . . .	34
10.4 Spectral algorithm* . . . . .	35
10.4.1 Idea . . . . .	35
10.4.2 Proof sketch . . . . .	36
10.5 SDP-based algorithm . . . . .	38
10.6 Combinatorial algorithms* . . . . .	39
10.7 Lower bounds . . . . .	40
<b>11 Random regular graphs</b>	<b>40</b>
11.1 Connectedness . . . . .	42
11.2 Contiguity . . . . .	43
<b>12 Expanders</b>	<b>44</b>
12.1 Spectral expanders . . . . .	45
12.2 Constructions . . . . .	46
12.3 Properties . . . . .	46
12.4 Applications . . . . .	47
<b>13 Quasirandomness</b>	<b>48</b>
13.1 Graphons . . . . .	48
13.2 Quasirandom graphs . . . . .	48
13.3 Quasirandom and non-quasirandom sequences . . . . .	49
13.4 Flag algebras and extremal graph theory . . . . .	50
13.5 Graphings . . . . .	50
13.6 Permutons . . . . .	51
<b>References</b>	<b>51</b>

## 1 Introduction

**What are graphs?** We will mostly consider *finite undirected graphs*. During this lecture we will also be interested in *countable undirected graphs*.

**What are random graphs?** The two most popular models (both known as Erdős–Rényi random graphs) are:

- $G(n, m)$ : a random graph on  $n$  vertices with  $m$  random edges.
- $G(n, p)$ : a random graph on  $n$  vertices in which each edge appears with probability  $p$ , independently.

Roughly speaking,  $G(n, m) \approx G(n, p)$  for  $p = m/\binom{n}{2}$ . Often results can be translated from one model to the other, see [FK16, §1.1]. We will be exclusively interested in the  $G(n, p)$  model.

**Why random graphs?** There are many possible answers:

- Random graphs are an excellent example of the probabilistic method. The first two examples were Shannon’s theorem and Erdős’ existence proof for graphs with arbitrarily large chromatic number and girth (size of the shortest cycle). Often random graphs are the only way to construct graphs with a certain property. A case in point is Ramsey numbers (see below).
- Random graphs are similar to other models of interest to physicists and probabilists: for example, Ising model, percolation, random CSPs.
- Random graphs appear in theoretical computer science in many situations, and are used to define certain problems such as random CSPs and planted clique.

- A different model of random graphs (preferential attachment) is used to study social networks (e.g. Facebook).

**What about random graphs?** We will be interested in the *typical* properties of random graphs. We say that a property occurs *with high probability* (whp) if the probability that it happens tends to 1. For example,  $G(n, 1/n^2)$  is a forest with high probability.

### Other popular models of random graphs

- Random regular graphs: for  $dn$  even, a random  $d$ -regular graph on  $n$  vertices (chosen uniformly among all such graphs).
- Geometric graphs: put  $n$  points on  $[0, 1]^2$ , and connect two if at distance at most  $r$ .
- Stochastic block model: divide  $\{1, \dots, n\}$  into  $k$  classes according to some distribution, and then connect vertices in class  $i$  and  $j$  with probability  $p_{ij}$ .
- Graphons: a vast generalization of the stochastic block model (in some sense).
- Preferential attachment graphs: various models in which vertices appear one by one, and are attached to vertices with probability increasing with the degree. Exhibits several phenomena encountered in social networks (e.g. the Facebook graph) such as a heavy-tail degree distribution.

Quasirandom graphs are graphs that share some of the properties of random graphs. For example, if a graph contains  $\approx pn$  edges and  $\approx p^4 n^4 / 8$  squares, then it behaves like  $G(n, p)$  (in certain senses). This is very related to graphons.

## 1.1 Ramsey graphs

Ramsey showed that every infinite graph contains either an infinite clique or an infinite independent set. The finitary version of this theorem states that for all  $a, b$  there exists  $n$  such that every graph on  $n$  vertices contains either a clique on  $a$  vertices or an independent set on  $b$  vertices. The smallest such  $n$  is denoted  $R(a, b)$ . A simple induction shows that

$$R(a, b) \leq \binom{a+b-2}{a-1}.$$

This implies, in particular, that

$$R(k, k) < 4^k.$$

This upper bound is proved by taking an arbitrary graph on  $\binom{a+b-2}{a-1}$  vertices and extracting from it either a clique of size  $a$  or an independent set of size  $b$ .

The best lower bound is

$$R(k, k) = \Omega(2^{k/2}).$$

To prove this, we have to show that there exists a graph on  $C \cdot 2^{k/2}$  vertices which contains no  $k$ -clique and no  $k$ -independent set; such a graph is known as a Ramsey graph. The only way we know how to construct such a graph explicitly is by taking a random  $G(n, 1/2)$  graph with  $n = C \cdot 2^{k/2}$ ; it will be a Ramsey graph with high probability.

In terms on  $n$ , with high probability  $G(n, 1/2)$  contains no clique and no independent set of size roughly  $2 \log_2 n$ . The best deterministic construction (due to Gil Cohen and Xin Li, 2016) gives  $(\log n)^{O(\log \log n)}$ . It's very hard to construct such graphs explicitly!

## 1.2 The countable random graph and zero-one laws

The following can be found, for example, in Horowitz [Hor08] and (probably) in Spencer [Spe01].

What does  $G(\aleph_0, p)$  look like? We will show that for every constant  $p$ , almost surely we get the same graph (which doesn't depend on  $p$ ). This surprising result follows from combining the following two lemmas.

**Definition 1.1.** A (countable) graph is *saturated* if for any two disjoint sets of vertices  $A, B$  there exists a vertex  $v$  such that  $(v, a)$  is an edge for all  $a \in A$ , and  $(v, b)$  is *not* an edge for all  $b \in B$ .

**Lemma 1.2.** For all  $p \in (0, 1)$ , almost surely  $G(\aleph_0, p)$  is saturated.

**Lemma 1.3.** Every two saturated graphs are isomorphic.

The two lemmas imply the following theorem.

**Theorem 1.4.** Suppose  $G_1 \sim G(\aleph_0, p_1)$  and  $G_2 \sim G(\aleph_0, p_2)$ , where  $0 < p_1, p_2 < 1$ . Almost surely,  $G_1 \approx G_2$ .

Lemma 1.3 shows that there is a unique saturated countable graph (up to isomorphism), and the theorem justifies its name, *the countable random graph*. Saturated countable graphs can be constructed explicitly (see for example Horowitz [Hor08]), but this is best left as an exercise for the reader.

The proof of the first lemma is very easy.

*Proof of Lemma 1.2.* For every  $A, B$  and every  $v \notin A \cup B$ , the probability that  $v$  satisfies the condition is  $p^{|A|}(1-p)^{|B|} > 0$ . Since there are infinitely many potential vertices, almost surely one of them will satisfy the condition. (There are various ways to make this argument rigorous.) Since there are countably many choices for  $A, B$ , the condition holds for all of them almost surely (since if countable many events happen almost surely, then all of them happen almost surely).  $\square$

The proof of the second lemma uses the so-called *back and forth argument*.

*Proof of Lemma 1.3.* The idea is to construct an isomorphism from  $V(G_1)$  to  $V(G_2)$ , vertex by vertex. Suppose that we already constructed a bijection  $\pi$  from  $V_1 \subset V(G_1)$  to  $V_2 \subset V(G_2)$  which is an isomorphism between  $G_1|_{V_1}$  and  $G_2|_{V_2}$ , and consider a new vertex  $v \in V(G_1) \setminus V_1$ . Let  $A = \{a \in V_1 : (v, a) \in G_1\}$  and  $B = \{b \in V_1 : (v, a) \notin G_1\}$ . Since  $G_2$  is saturated, there is a vertex  $w \in V(G_2) \setminus V_2$  which is connected to all vertices in  $\pi(A)$  and to none in  $\pi(B)$ . We extend  $\pi$  by mapping  $v$  to  $w$ .

We perform infinitely many steps of this form. How can we ensure that in the end we get a bijection between all of  $V(G_1)$  and all of  $V(G_2)$ ? Order  $V(G_1), V(G_2)$  arbitrarily. We alternate between  $G_1$ -steps and  $G_2$ -steps. In a  $G_1$  step, we map the smallest vertex of  $G_1$  not already mapped. In a  $G_2$  step, we map the smallest vertex of  $G_2$  not already mapped. This easily implies that the  $t$ th vertex of  $G_1$  is mapped in the first  $2t$  steps, and the same holds for the  $t$ th vertex of  $G_2$ . So eventually all vertices are mapped.  $\square$

The same argument shows that every two countable dense linear orderings with no extreme elements are isomorphic (one example is the rationals), see Rosenstein [Ros82].

**Zero-one law** Theorem 1.4 essentially implies the following zero-one law for random graphs. Suppose that  $P$  is a first-order property (in the language including only the edge relation symbol and equality). Then for all constant  $p \in (0, 1)$ , either  $\Pr[P(G(n, p))] \rightarrow 0$  or  $\Pr[P(G(n, p))] \rightarrow 1$  as  $n \rightarrow \infty$ . A first-order property is one of the form

$$\exists x \exists y \exists z (x \sim y) \wedge (x \sim z) \wedge (y \sim z),$$

which states that the graph contains a triangle (we are also allowed to use  $\forall$ ); in this case, a  $G(n, p)$  random graph contains a triangle with high probability.

A difficult theorem of Shelah and Spencer states the same with  $p = n^{-\alpha}$  for any *irrational*  $\alpha$ . For rational  $\alpha$  this is not true, and we will see examples when we discuss appearance of subgraphs.

## 2 Sparse random graphs

Our first major topic will be *the evolution of random graphs*. Much of this comes from the seminal paper of Erdős and Rényi [ER60], which inaugurated the subject.

Erdős and Rényi considered the dynamical process in which we start with the empty graph on  $n$  vertices, and at each step add a random edge not already in the graph. After  $\binom{n}{2}$  steps, we reach the complete graph. We can look at *random times* at which some events happen. For example, at what step does the graph become connected?

More formally, let  $(G_0, \dots, G_{\binom{n}{2}})$  be a random variable such that  $G_0 = \emptyset$  and  $G_{m+1}$  is obtained from  $G_m$  by adding a random edge.

**Definition 2.1.** Let  $m_{\text{Con}}$  be the minimal  $m$  such that  $G_m$  is connected.

Let  $m_{\text{Iso}}$  be the minimal  $m$  such that  $G_m$  doesn't contain isolated vertices.

Clearly  $m_{\text{Iso}} \leq m_{\text{Con}}$ . Erdős and Rényi showed that whp,  $m_{\text{Con}} = m_{\text{Iso}}$ . It is not immediately clear how to formulate this property in the language of  $G(n, p)$ .

What we will do next is (i) sort out the correct *coupling* of  $G(n, p)$  graphs, (ii) make a list of some highlights in the evolution of random graphs, and (iii) start proving some of these properties.

The discussion of couplings doesn't appear in our textbook, and is somewhat non-orthodox.

## 2.1 Graph couplings

What are the salient properties of the process  $G_0, \dots, G_{\binom{n}{2}}$  discussed above?

1.  $G_m \sim G(n, m)$ .
2. If  $m_1 \leq m_2$  then  $G_{m_1} \subseteq G_{m_2}$ .

This is a *coupling* of the models  $G(n, 0), \dots, G(n, \binom{n}{2})$ . A coupling of two random variables  $X, Y$  is a joint distribution on pairs whose marginals are  $X, Y$ , and this can be extended to many random variables. The coupling shows that if  $P$  is any monotone property (if  $P(G)$  and  $H \supseteq G$  then  $P(H)$ ) then  $\Pr[P(G(n, m_2))] \geq \Pr[P(G(n, m_1))]$  whenever  $m_2 \geq m_1$ : this is since  $\Pr[P(G_{m_2})] \geq \Pr[P(G_{m_1})]$ , due to  $G_{m_2} \supseteq G_{m_1}$ .

We can think of these two properties as “equations” between the random variables  $G_0, \dots, G_{\binom{n}{2}}$ . These equations have a unique “solution”: any two random tuples satisfying these constraints are “isomorphic”.

The corresponding properties for  $G(n, p)$  are as follows. We are looking for a process  $G_p$  (for  $p \in [0, 1]$ ) such that

1.  $G_p \sim G(n, p)$ .
2. If  $p_1 \leq p_2$  then  $G_{p_1} \subseteq G_{p_2}$ .

The unique (up to measure zero) solution looks as follows. Associate with every edge  $e$  a random variable  $x_e \sim U([0, 1])$ , and define  $G_p = \{e : x_e \leq p\}$ .

Given this coupling, we can define  $p_{\text{Con}}, p_{\text{Iso}}$  in analogy to  $m_{\text{Con}}, m_{\text{Iso}}$ . It turns out that  $p_{\text{Con}} = p_{\text{Iso}}$  whp as well (this result is stated in the textbook only as  $m_{\text{Con}} = m_{\text{Iso}}$ , but the proof is very similar).

A much more sophisticated example of this kind of construction is Brownian motion, in which case constructing the set of random variables is somewhat non-trivial (see for example Mörters and Peres [MP10, Chapter 1]).

## 2.2 Evolution of random graphs

Here are some highlights of the evolution of random graphs:

1. If  $p = o(1/n)$  then whp  $G(n, p)$  is a forest. (Theorem 2.2)
2. If  $p = 1/n - \omega(1/n^{4/3})$  then whp all connected components of  $G(n, p)$  are trees or unicyclic (consist of a tree plus an edge). (Theorem 2.3)
3. If  $p = c/n$  for constant  $c < 1$  then whp the largest connected component of  $G(n, p)$  is a tree of size  $\Theta_c(\log n)$ . (Theorem 3.2)
4. If  $p = 1/n$  then whp the largest connected component of  $G(n, p)$  has size roughly  $n^{2/3}$ . (Theorem 5.3)
5. If  $p = c/n$  for constant  $c > 1$  then whp  $G(n, p)$  contains a giant component of size  $\Theta_c(n)$ , and all other components are trees of size  $O_c(\log n)$ . (Theorem 4.5)
6. If  $p = (\log n + c)/n$  then the probability that  $G(n, p)$  is connected tends to  $e^{-e^{-c}}$ . Furthermore, whp  $p_{\text{Con}} = p_{\text{Iso}}$ . (Theorem 6.3)

7. If  $p = (\log n + c)/n$  then the probability that  $G(n, p)$  has a matching of size  $\lfloor n/2 \rfloor$  tends to  $e^{-e^{-c}}$ .
8. If  $p = (\log n + \log \log n + c)/n$  then the probability that  $G(n, p)$  is hamiltonian tends to  $e^{-e^{-c}}$ .

We will prove most of these properties.

### 2.3 Forest regime

**Theorem 2.2** ([FK16, Theorem 2.1]). *If  $p = o(1/n)$  then whp  $G(n, p)$  is a forest.*

Before proving the theorem, let us explain what it means. Suppose that we are given a function  $p: \mathbb{N} \rightarrow [0, 1]$  that satisfies  $p(n) = o(1/n)$ , that is,  $np(n) \rightarrow 0$ . For any such function  $p$ , the probability that  $G(n, p(n))$  is a forest is  $1 - o(1)$ . In other words,

$$np(n) \rightarrow 0 \implies \Pr[G(n, p(n)) \text{ is a forest}] \rightarrow 1.$$

*Proof.* We will show that whp  $G(n, p)$  contains no cycle. This, in turn, we will show using the *first moment method*: let  $X$  be the number of cycles in  $G(n, p)$ . We will show that  $\mathbb{E}[X] = o(1)$ , and so Markov's inequality implies that  $\Pr[X > 0] = \Pr[X \geq 1] \leq \mathbb{E}[X] = o(1)$ .

We calculate the expected number of cycles by estimating the expected value of  $X_k$ , the number of cycles of length  $k$ . We use the following formula: the expected value of  $X_k$  is the number of potential cycles times the probability that each potential cycle is in  $G(n, p)$ . The formula follows from linearity of expectation by using indicator variables: if  $X_C$  is the event that  $G \sim G(n, p)$  contains the cycle  $C$ , then

$$\mathbb{E}[X_k] = \sum_{\substack{C \text{ a cycle} \\ |C|=k}} \mathbb{E}[X_C] = \sum_{\substack{C \text{ a cycle} \\ |C|=k}} \Pr[X_C = 1] = |\{C \text{ a cycle} : |C| = k\}| \cdot \Pr[X_{C_k} = 1],$$

where  $C_k$  is a particular cycle of size  $k$ .

We can specify a cycle of length  $k$  by specifying  $k$  vertices. Each cycle is counted  $2k$  times this way, so the number of cycles of size  $k$  is  $n^k/2k$ , where  $n^{\underline{k}} = n(n-1)\cdots(n-k+1) = n!/(n-k)!$  is a *falling power* (there are also rising powers). The probability that  $G(n, p)$  contains any fixed cycle of size  $k$  is  $p^k$  (since there are  $k$  edges in such a cycle), and so

$$\mathbb{E}[X_k] = \frac{n^{\underline{k}}}{2k} p^k.$$

Therefore

$$\mathbb{E}[X] = \sum_{k=3}^n \frac{n^{\underline{k}}}{2k} p^k \leq \sum_{k=3}^{\infty} (pn)^k = \frac{(pn)^3}{1-pn} = O((pn)^3) = o(1).$$

Here we used that if  $pn \rightarrow 0$  then from some point on  $pn \leq 1/2$ , and so  $1/(1-pn) \leq 2$ . This completes the proof.  $\square$

### 2.4 Unicyclic regime

**Theorem 2.3** ([FK16, Lemma 2.10]). *If  $p = 1/n - \omega(1/n^{4/3})$  then whp all connected components of  $G(n, p)$  are trees or unicyclic (contain at most one cycle).*

We will see from the proof where  $1/n^{4/3}$  comes from. This threshold is (probably) tight: if  $p = 1/n - c/n^{4/3}$  for constant  $c$  then (it seems) there is constant probability that the graph contain a component with at least two cycles.

The forbidden structure in the preceding case was a cycle. Here it will be a certain type of extended path.

**Lemma 2.4.** *If  $G$  has a connected component which is neither a tree nor unicyclic, then  $G$  contains a path  $P$  (of length at least 3 vertices) and two (distinct) edges connecting the two endpoints of  $P$  to other vertices in  $P$ .*

We leave the proof to the reader.

*Proof of Theorem 2.3.* Let  $p = 1/n - \alpha/n^{4/3}$ , where  $\alpha \rightarrow \infty$ . The number of potential arrangements as the extended path described in the lemma, when the path has length  $k$ , is at most  $n^k k^2$ : there are  $n^{\underline{k}}/2 \leq n^k$  choices for the path, and  $(k-2)^2 - 1 < k^2$  choices for the two extra edges. Each such structure occurs in  $G(n, p)$  with probability  $p^{k+1}$ , and so the expected number of structures appearing in  $G(n, p)$  is at most

$$\sum_{k=3}^n n^k k^2 p^{k+1} \leq \sum_{k=3}^n \frac{k^2}{n} (np)^k = \sum_{k=3}^n \frac{k^2}{n} \left(1 - \frac{\alpha}{n^{1/3}}\right)^k \leq \sum_{k=3}^n \frac{k^2}{n} e^{-\alpha k/n^{1/3}}.$$

We used  $p \leq 1/n$  in the first inequality and  $1 - x \leq e^{-x}$  (for  $x \geq 0$ ) in the second inequality.

There are many ways to estimate this sum. One of them approximates the sum by an integral. We will use the substitution  $x_k = (\alpha/n^{1/3})k$ .

$$\sum_{k=3}^n \frac{k^2}{n} e^{-\alpha k/n^{1/3}} = \sum_{k=3}^n \frac{(n^{2/3}/\alpha^2)x_k^2}{n} e^{-x_k} = \frac{1}{\alpha^3} \sum_{k=3}^n \frac{\alpha}{n^{1/3}} x_k^2 e^{-x_k}.$$

Since  $|x_{k+1} - x_k| = \alpha/n^{1/3}$ , we recognize this as a Riemann sum, and so

$$\sum_{k=3}^n \frac{k^2}{n} e^{-\alpha k/n^{1/3}} \rightarrow \frac{1}{\alpha^3} \int_0^\infty x^2 e^{-x} dx.$$

(Skipping some technical details.) The integral clearly converges (to 2, though the exact value isn't important for us), and so the expected number of bad structures is  $O(1/\alpha^3) \rightarrow 0$ .  $\square$

### 3 Subcritical regime

So far we have seen that when  $p = o(1/n)$ , whp  $G(n, p)$  is a forest, and when  $p = 1/n - \omega(1/n^{4/3})$ , whp all connected components contain at most one cycle. Our goal now is to describe the state of a  $G(n, c/n)$  random graph, for  $c < 1$ . We already know that all connected components contain at most one cycle. But do unicyclic components actually occur?

**Lemma 3.1** ([FK16, Lemma 2.11]). *If  $p = c/n$  for  $c < 1$  and  $f(n) = \omega(1)$  then whp  $G(n, p)$  contains no unicyclic components of size  $f(n)$  or larger. Moreover, whp the total number of vertices in unicyclic components is at most  $f(n)$ .*

*Proof.* Let  $U_k$  denote the number of unicyclic components of size  $k$ . Given a set of  $k$  vertices, Cayley's formula states that there are  $k^{k-2}$  possible trees on this vertex set, and so the number of potential unicyclic components of size  $k$  is at most  $\binom{n}{k} k^{k-2} k^2 = \binom{n}{k} k^k$ , since there are at most  $k^2$  ways to choose the extra edge. Such a structure appears in the graph as a *connected component* if the  $k$  implied edges are in the graph, the other  $\binom{k}{2} - k$  potential edges in the component are not in the graph, and the  $k(n-k)$  edges connecting the  $k$  vertices to the rest of the graph are also not in the graph. Thus

$$\mathbb{E}[U_k] \leq \binom{n}{k} k^k p^k (1-p)^{k(n-k) + \binom{k}{2} - k} = n^k \frac{k^k}{k!} p^k (1-p)^{k[n-k/2-3/2]}.$$

We now estimate these terms one by one. The first step is estimating  $1-p \leq e^{-p} = e^{-c/n}$ . The second step is using Stirling's approximation

$$k! \sim \sqrt{2\pi k} (k/e)^k \implies k! = \Theta\left(\sqrt{k} (k/e)^k\right).$$

These estimates show that

$$\mathbb{E}[U_k] \leq \frac{1}{\Omega(\sqrt{k})} \frac{n^k}{n^k} \left( e c e^{-[n-k/2-3/2]c/n} \right)^k.$$

When  $k = o(n)$ , the last factor becomes roughly  $ce^{1-c}$ , which is good for us: this function attains its maximum 1 at  $c = 1$  (since its derivative is  $c(1-c)e^{1-c}$ ), and the assumption  $c < 1$  implies that

$ce^{1-c} < 1$ ,<sup>1</sup> and so  $\mathbb{E}[U_k]$  decays exponentially fast. However, for large  $k$  we only get  $ce^{1-c/2}$ , which could exceed 1.

The solution to this conundrum is taking advantage of the difference between  $n^{\underline{k}}$  and  $n^k$ :

$$\frac{n^{\underline{k}}}{n^k} = \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} = \left(1 - \frac{0}{n}\right) \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \leq \exp - \left[ \frac{0}{n} + \frac{1}{n} + \cdots + \frac{k-1}{n} \right] = e^{-\binom{k}{2}/n}.$$

This factor becomes significant when  $k = \Omega(\sqrt{n})$ , and helps us obtain a better estimate:

$$\mathbb{E}[U_k] \leq \frac{1}{\Omega(\sqrt{k})} \left( ece^{-[n-2]c/n} \right)^k \leq O(1) \cdot (ce^{1-c})^k e^{2ck/n} \leq O_c(1)(ce^{1-c})^k.$$

Summing over all  $k \geq f(n)$ , we get

$$\mathbb{E}[U_{\geq f(n)}] \leq O_c(1) \sum_{k=f(n)}^n (ce^{1-c})^k = O_c(1)(ce^{1-c})^{f(n)}.$$

If  $f(n) \rightarrow \infty$  then  $\mathbb{E}[U_{\geq f(n)}] \rightarrow 0$ .

We leave the proof of the *moreover* part to the reader.  $\square$

Together with Theorem 2.3, this suggests that the largest connected component in  $G(n, c/n)$  is a tree when  $c < 1$ . This allows us to determine almost exactly the size of the largest connected component in  $G(n, c/n)$ .

**Theorem 3.2** ([FK16, Lemma 2.12]). *If  $p = c/n$  for  $c < 1$  and  $f(n) = \omega(1)$  then whp the size  $S$  of the largest component in  $G(n, p)$  satisfies*

$$\left| S - \frac{\log n - \frac{5}{2} \log \log n}{c - 1 - \log c} \right| < f(n).$$

Note that  $c - 1 - \log c = -\log[ce^{1-c}]$ . The origin of the mysterious threshold will become apparent during the proof.

*Proof.* Let  $\alpha = c - 1 - \log c$  and  $k_0 = (\log n - \frac{5}{2} \log \log n)/\alpha$ , and consider  $k = k_0 + \delta$ . We will estimate the number  $T_k$  of tree components of  $G(n, p)$  of size  $k$ . The number of potential components is  $\binom{n}{k} k^{k-2}$  by Cayley's formula, and the probability that  $G(n, p)$  contains one of them is  $p^{k-1} (1-p)^{k(n-k) + \binom{k}{2} - (k-1)}$ . Hence

$$\mathbb{E}[T_k] = n^{\underline{k}} \cdot \frac{1}{k^2} \cdot \frac{k^k}{k!} \cdot p^{k-1} \cdot (1-p)^{k(n-k) + \binom{k}{2} - (k-1)}.$$

Comparing this expression to the expression we used to upper bound  $\mathbb{E}[U_k]$  in Theorem 2.3, we see that the present expression is smaller by a factor of  $\frac{p}{1-p} k^2$ . The proof of Theorem 2.3 uses the upper bound  $O_c(1/\sqrt{k})(ce^{1-c})^k$  (bringing back the factor  $1/\sqrt{k}$  that we dropped in that theorem), and so

$$\mathbb{E}[T_k] \leq \frac{1-p}{pk^2} O_c \left( \frac{1}{\sqrt{k}} \right) (ce^{1-c})^k = O_c \left( \frac{n}{k^{5/2}} \right) (ce^{1-c})^k.$$

Let us now assume that  $\delta \geq 0$ . Then  $n/k^{5/2} \leq O_c(n/\log^{5/2} n)$ , and so the definition of  $k_0$  shows that  $\mathbb{E}[T_k] \leq O_c((ce^{1-c})^\delta)$ . Therefore

$$\mathbb{E} \left[ \sum_{k=k_0+f(n)}^n T_k \right] \leq O_c \left( \sum_{\delta=f(n)}^{\infty} (ce^{1-c})^\delta \right) = O_c((ce^{1-c})^{f(n)}) = o(1).$$

---

<sup>1</sup>Another argument uses the Taylor expansion of  $e^x$ :  $e^{c-1} = 1 + (c-1) + \frac{(c-1)^2}{2} e^\theta > c$ , for some  $c-1 \leq \theta \leq 0$ .



This show that the largest tree component in  $G(n, p)$  has size smaller than  $k_0 + f(n)$  with high probability. The same is true for unicyclic components by Theorem 2.3, and Lemma 3.1 shows that with high probability there are no other components. We conclude that  $S < k_0 + f(n)$  with high probability.

For the other direction, we first need to estimate  $\mathbb{E}[T_k]$  from below. Let us assume, therefore, that  $\delta \leq 0$ . Since  $k = O_c(\log n)$ ,  $n^k$  is a good approximation for  $n^{\underline{k}}$ :

$$n^{\underline{k}} = n^k \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \geq n^k \left(1 - \frac{\binom{k}{2}}{n}\right) = (1 - o(1))n^k,$$

where we used  $(1 - \alpha)(1 - \beta) = 1 - \alpha - \beta + \alpha\beta \geq 1 - \alpha - \beta$ . Stirling's formula gives the estimate  $k! = (1 \pm o(1))\sqrt{2\pi k}(k/e)^k$ . We estimate the final factor using the lower bound  $1 - x \geq e^{-x-x^2}$ , which holds for small enough  $x$  (for example,  $x \leq 1/2$  suffices). This estimate implies that

$$(1 - p)^{k(n-k) + \binom{k}{2} - (k-1)} \geq (1 - p)^{kn} \geq e^{-ck - c^2k/n} = (1 - o(1))e^{-ck}.$$

Putting all estimates together, we obtain

$$\begin{aligned} \mathbb{E}[T_k] &\geq (1 \pm o(1))n^k \cdot \frac{1}{k^2} \cdot \frac{e^k}{\sqrt{2\pi k}} \cdot \frac{n}{c} p^k \cdot e^{-ck} = \\ &(1 \pm o(1)) \frac{1}{c\sqrt{2\pi}} \cdot \frac{n}{k^{5/2}} (ce^{1-c})^k = (1 \pm o(1)) \frac{\alpha^{5/2}}{c\sqrt{2\pi}} \cdot \frac{n}{\log^{5/2} n} (ce^{1-c})^k. \end{aligned}$$

We chose  $k_0$  so that  $(ce^{1-c})^k$  cancels  $n/\log^{5/2} n$  exactly, and so

$$\mathbb{E}[T_k] = \Omega_c((ce^{1-c})^\delta).$$

If  $\delta = -f(n)$  then  $\mathbb{E}[T_k] = \omega(1)$ . However, this is not enough to conclude that  $S \geq \frac{\log n - \frac{5}{2} \log \log n}{\alpha} - f(n)$  with high probability; in principle, it could be that there is a low probability event that creates many tree components of that size, and this is the reason that  $\mathbb{E}[T_k] = \omega(1)$ . We will actually see such an example when we talk about subgraph counts. To complete the proof, we will use the *second moment method*.

The idea is to use Chebyshev's inequality:

$$\Pr[T_k = 0] \leq \Pr[|T_k - \mathbb{E}[T_k]| \geq \mathbb{E}[T_k]] \leq \frac{\mathbb{V}[T_k]}{\mathbb{E}[T_k]^2}.$$

We will estimate  $\mathbb{V}[T_k]$  using the formula  $\mathbb{V}[T_k] = \mathbb{E}[T_k^2] - \mathbb{E}[T_k]^2$ . For a particular tree  $t$  of size  $k$ , let  $X_t$  be the indicator variable for the random graph  $G \sim G(n, p)$  containing  $t$ . Then

$$\mathbb{E}[T_k^2] = \sum_{s, t \text{ } k\text{-trees}} \mathbb{E}[X_s X_t] = \sum_{s, t \text{ } k\text{-trees}} \Pr[s, t \in G].$$

What is  $\Pr[s, t \in G]$ ? There are three cases. If  $s = t$  then  $\Pr[s, t \in G] = \Pr[t \in G]$ . If  $s, t$  share vertices then  $\Pr[s, t \in G] = 0$ . If  $s, t$  are disjoint then they involve together  $2(k-1)$  edges,  $2\binom{k}{2} - (k-1)$  non-edges inside the components, and  $2k(n-k) - k^2$  non-edges separating the components from the rest of the graphs;  $k^2$  non-edges are counted twice. Therefore  $\Pr[s, t \in G] = (1-p)^{-k^2} \Pr[s \in G] \Pr[t \in G]$ . In total,

$$\mathbb{E}[T_k^2] \leq \sum_t \Pr[t \in G] + \sum_{s \neq t} (1-p)^{-k^2} \Pr[s \in G] \Pr[t \in G] = \mathbb{E}[T_k] + (1-p)^{-k^2} \mathbb{E}[T_k]^2.$$

Therefore

$$\frac{\mathbb{V}[T_k]}{\mathbb{E}[T_k]^2} = \frac{1}{\mathbb{E}[T_k]} + \left[(1-p)^{-k^2} - 1\right] \leq o(1) + e^{O(k^2/n)} - 1 = o(1) + O(k^2/n) = o(1).$$

We conclude that  $\Pr[T_k = 0] = o(1)$ , and so with high probability a tree component of size at least  $\frac{\log n - \frac{5}{2} \log \log n}{\alpha} - f(n)$  exists.  $\square$

### 3.1 Bonus: Proof of Stirling's formula

#### 3.1.1 Warm-up: harmonic numbers

Let us start with a warm-up: estimating the sum of the harmonic series. The  $n$ th harmonic number is defined by

$$H_n := \sum_{i=1}^n \frac{1}{i}.$$

Since  $1/x$  is increasing, we see that

$$\log(n+1) = \int_1^{n+1} \frac{dx}{x} \leq \sum_{i=1}^n \frac{1}{i} \leq 1 + \int_1^n \frac{dx}{x} = \log n + 1.$$

This shows that  $H_n = \log n + \Theta(1)$ . In fact, we can get a better estimate by being more careful:

$$H_n = \sum_{i=1}^n \frac{1}{i} = \sum_{i=1}^n \left[ \int_{i-1/2}^{i+1/2} \frac{dx}{x} + \int_{i-1/2}^{i+1/2} \left( \frac{1}{i} - \frac{1}{x} \right) dx \right].$$

Substituting  $x = i + \delta$  and using  $1/i - 1/(i + \delta) = \delta/i(i + \delta)$ , we get

$$H_n = \int_{1/2}^{n+1/2} \frac{dx}{x} + \sum_{i=1}^n \int_{-1/2}^{1/2} \frac{\delta}{i(i + \delta)} d\delta.$$

The first term is equal to

$$\int_{1/2}^{n+1/2} \frac{dx}{x} = \log(n+1/2) - \log(1/2) = \log n + \log 2 + \log \left( 1 + \frac{1}{2n} \right) = \log n + \log 2 + O\left(\frac{1}{n}\right).$$

Using the power series expansion

$$\frac{1}{i + \delta} = \frac{1/i}{1 + \delta/i} = \frac{1}{i} \sum_{m=0}^{\infty} (-1)^m \frac{\delta^m}{i^m},$$

we can calculate each summand in the second term:

$$\int_{-1/2}^{1/2} \sum_{m=0}^{\infty} (-1)^m \frac{\delta^{m+1}}{i^{m+2}} d\delta = \sum_{m=0}^{\infty} (-1)^m \frac{(1/2)^{m+2} - (-1/2)^{m+2}}{(m+2)i^{m+2}} = - \sum_{\ell=0}^{\infty} \frac{2 \cdot 2^{-2\ell-3}}{(2\ell+3)i^{2\ell+3}},$$

where  $m = 2\ell + 1$ . Altogether, we get

$$H_n = \log n + \log 2 + O\left(\frac{1}{n}\right) - \sum_{\ell=0}^{\infty} \frac{4^{-\ell-1}}{2\ell+3} \sum_{i=1}^n \frac{1}{i^{2\ell+3}}.$$

Estimating the sum by an integral, we see that  $\sum_{i=n+1}^{\infty} (1/i^{2\ell+3}) \leq 1/n^{2\ell+2}$ , and so using  $\sum_{\ell=0}^{\infty} (1/n^{2\ell}) = O(1)$ , we obtain

$$H_n = \log n + \log 2 - \sum_{\ell=0}^{\infty} \frac{4^{-\ell-1}}{2\ell+3} \sum_{i=1}^{\infty} \frac{1}{i^{2\ell+3}} + O\left(\frac{1}{n}\right).$$

In other words, there is a constant  $\gamma$  (known as Euler's constant) such that  $H_n = \log n + \gamma + O(1/n)$ . The constant  $\gamma$  can be estimated from the expression given above to be roughly  $\gamma \approx 0.5772$ . It is conjectured that  $\gamma$  is irrational, and probably also transcendental.

### 3.1.2 Stirling's formula

We can use exactly the same method to estimate  $n!$ . It is easier to estimate  $\log n! = \sum_{i=2}^n \log i$  instead. A rough estimate is given by

$$\int_2^{n+1} \log x \, dx \leq \sum_{i=2}^n \log i \leq \int_1^n \log x \, dx.$$

An antiderivative of  $\log x$  is  $x \log x - x$ , and this shows  $\log n! \approx n \log n - n = n \log(n/e)$ . To get a more precise estimate, we proceed as before, with slight differences to make the calculation easier:

$$\sum_{i=2}^n \log i = \sum_{i=2}^n \int_{i-1}^i [\log i + (\log x - \log i)] \, dx = \int_1^n \log x \, dx - \sum_{i=2}^n \int_0^1 \log \frac{i-\delta}{i} \, d\delta.$$

The first summand is equal to

$$\int_1^n \log x \, dx = x \log \frac{x}{e} \Big|_1^n = n \log \frac{n}{e} + 1.$$

For the second summand, we need to use the Taylor expansion of  $\log(1+x)$ :

$$\log \frac{i-\delta}{i} = \log \left(1 - \frac{\delta}{i}\right) = - \sum_{m=1}^{\infty} \frac{\delta^m}{mi^m}.$$

Integrating, this shows that the second summand equals

$$\sum_{i=2}^n \sum_{m=1}^{\infty} \frac{1}{(m+1)mi^m} = \frac{H_n - 1}{2} + \sum_{m=2}^{\infty} \frac{1}{(m+1)m} \sum_{i=2}^n \frac{1}{i^m}.$$

As before, we can estimate the finite sums over  $i$  by infinite sums to get

$$\log n! = n \log \frac{n}{e} + 1 + \frac{\log n + \gamma - 1}{2} + \sum_{m=2}^{\infty} \frac{1}{(m+1)m} \sum_{i=2}^{\infty} \frac{1}{i^m} + O\left(\frac{1}{n}\right).$$

This shows that there exists a constant  $C$  such that

$$\log n! = n \log \frac{n}{e} + \frac{\log n}{2} + C + O\left(\frac{1}{n}\right).$$

Since  $e^{O(1/n)} = 1 + O(1/n)$ , this implies that

$$n! \sim e^C \sqrt{n} (n/e)^n.$$

Surprisingly, it turns out that  $e^C = \sqrt{2\pi}$ .

## 3.2 Bonus: Proof of Cayley's formula

Here is a simple proof of Cayley's formula, which states that the number of trees on  $n$  vertices is  $n^{n-2}$ . This proof paraphrases the article *Cayley's Formula: A Page From The Book* by Arnon Avron and Nachum Dershowitz.

Let  $T$  be a tree on  $n$  vertices, and choose an order on its vertices,  $v_1, \dots, v_n$ . Initially, we think of  $T$  as a tree rooted at  $v_1$ . We then disconnect  $v_2$  from its parent, so that now we have two trees, one rooted at  $v_1$ , and the other rooted at  $v_2$ . We then disconnect  $v_3$  from its parent, so that we have three trees rooted at  $v_1, v_2, v_3$ . Continuing this way, after  $n-1$  steps we reach  $n$  disconnected vertices.

What information do we need to reverse this process? Suppose that we want to reverse the  $k$ th step. We are given  $k+1$  rooted trees, and we need to figure out how to turn them into  $k$  rooted trees. The information we need is the parent  $u_{k+1}$  of  $v_{k+1}$ , of which there are  $n$  options, and  $v_{k+1}$  itself, of which there are  $k$  options: it has to be the root of some tree, but cannot be the root of the tree containing  $u_{k+1}$ .

In total, we obtain that the number  $T_n$  of trees on  $n$  vertices satisfies

$$n!T_n = \prod_{k=1}^{n-1} kn = (n-1)!n^{n-1}.$$

Dividing by  $n!$ , this gives  $T_n = n^{n-2}$ .

## 4 Supercritical regime

We now switch to the supercritical regime,  $p = c/n$  for  $c > 1$ . In this regime, whp  $G(n, p)$  contains a unique *giant component* of linear size, all other connected components being logarithmic in size. The proof will involve several steps:

- First we will show a dichotomy for the size of the connected components: all of them are of size  $O(\log n)$  or  $\Omega(n)$ . (Lemma 4.1)
- Then we will count how many vertices in total participate in small components. (Lemma 4.3)
- Finally, we will show that there is only one linear-sized component. (Lemma 4.4)

We start with the size dichotomy.

**Lemma 4.1.** *If  $p = c/n$  for  $c > 1$  then whp every connected component of  $G(n, p)$  has size either  $O_c(\log n)$  (“small”) or  $\Omega_c(n)$  (“large”).*

*Proof.* Let  $C_k$  denote the number of spanning trees of connected components of size  $k$  of  $G(n, p)$ . There are  $\binom{n}{k} k^{k-2}$  potential spanning trees, and each of them belongs to the graph with probability  $p^{k-1}(1-p)^{k(n-k)}$  (the numbers in the exponents don’t sum up to  $\binom{n}{2}$  since we allow other edges inside the component). Thus

$$\begin{aligned} \mathbb{E}[C_k] &= \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \leq n^k \frac{k^k}{k!} k^{-2} p^{-1} p^k e^{-pk(n-k)} \leq \\ &O\left(\frac{n}{ck^{5/2}}\right) (ce^{1-c} \frac{n-k}{n})^k \leq O_c(n) (ce^{1-c})^k e^{ck^2/n}. \end{aligned}$$

If  $k \leq \sqrt{n}$  then  $ck^2/n \leq c$ , and so

$$\mathbb{E}[C_k] \leq O_c(n) (ce^{1-c})^k.$$

Since  $ce^{1-c} < 1$ , there is a constant  $A_c$  such that  $(ce^{1-c})^{A_c \log n} \leq 1/n^3$ . Whenever  $A_c \log n \leq k \leq \sqrt{n}$ , we thus have

$$\mathbb{E}[C_k] = O_c(1/n^2).$$

Taking a union bound, we see that whp there are no connected components of these sizes.

To see what happens when  $k \geq \sqrt{n}$ , consider the substitution  $k = Bn$ :

$$\mathbb{E}[C_k] \leq O_c(n) [(ce^{1-c})^B e^{cB^2}]^n = O_c(n) (ce^{1-(1-B)c})^{nB}.$$

Since  $c > 1$ , we can find a constant  $B_c$  such that  $ce^{1-(1-B_c)c} < 1$ . Whenever  $k \leq B_c n$ , the expectation  $\mathbb{E}[C_k]$  will be exponentially small, and so a union bound shows that whp there are no connected components of these sizes either.  $\square$

The next step is to consider small components. It turns out that all of them are trees, whp.

**Lemma 4.2.** *If  $p = c/n$  for  $c > 1$  then whp the total number of vertices in small non-tree components of  $G(n, p)$  is  $o(n)$ .*

Here *small* is as given by Lemma 4.1.

*Proof.* Let  $C_k$  denote the number of spanning trees of connected components of size  $k$  of  $G(n, p)$ , together with an extra edge. There are at most  $\binom{n}{k} k^{k-2} \cdot k^2$  of these, and each of them appears in the graph with probability  $p^k (1-p)^{k(n-k)}$ . Recycling our computations from the proof of Lemma 4.1, we see that when  $k \leq \sqrt{n}$ ,

$$\mathbb{E}[C_k] \leq O_c(1) \frac{n}{k^{5/2}} (ce^{1-c})^k k^2 p = O_c(1) \frac{(ce^{1-c})^k}{\sqrt{k}}.$$

Therefore the expected total number of vertices in small non-tree components is

$$\sum_{k=3}^{A_c \log n} \mathbb{E}[kC_k] = O_c(1) \sum_{k=3}^{A_c \log n} \sqrt{k} (ce^{1-c})^k = O_c(1).$$

Markov's inequality shows that whp, the total number of vertices is  $o(n)$ . (In fact, whp the total number of vertices is  $o(f(n))$  for any  $f(n) = \omega(1)$ .)  $\square$

It remains to count the number of trees.

**Lemma 4.3.** *If  $p = c/n$  for  $c > 1$  then whp the number of vertices in small tree components is*

$$(1 \pm o(1)) \frac{x}{c} n,$$

where  $x < 1$  is the unique solution to  $xe^{-x} = ce^{-c}$ .

Note that the function  $f(x) = xe^{1-x}$ , whose derivative is  $(1-x)e^{1-x}$ , increases from  $f(0) = 0$  to  $f(1) = 1$  and then decreases to  $\lim_{x \rightarrow \infty} f(x) = 0$ .

*Proof.* Let  $T_k$  be the number of tree components of size  $k$ . When  $k = O_c(\log n)$ ,

$$\mathbb{E}[T_k] = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k) + \binom{k}{2} - (k-1)} = (1 - o(1)) \frac{n}{c} \frac{k^{k-2}}{k!} c^k (1-p)^{kn} (1-p)^{-O(k^2)}.$$

Now  $(1-p)^{-O(k^2)} = 1 + o(1)$  and  $(1-p)^{kn} = (1 - o(1))e^{-ck}$  (using estimates we have seen above). Therefore

$$\mathbb{E}[T_k] = (1 \pm o(1)) \frac{n}{c} \frac{k^{k-2}}{k!} (ce^{-c})^k.$$

The expected number of vertices in small components is thus

$$\mathbb{E}[S] := \sum_{k=1}^{A_c \log n} k \mathbb{E}[T_k] = (1 \pm o(1)) \frac{n}{c} \sum_{k=1}^{A_c \log n} \frac{k^{k-1}}{k!} (ce^{-c})^k.$$

The general term of the series satisfies

$$\frac{k^{k-1}}{k!} (ce^{-c})^k = O(k^{-3/2}) (ce^{1-c})^k.$$

In particular, for some constant  $K_c$  we have

$$\sum_{k=A_c \log n + 1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k = O_c((ce^{1-c})^k) = O_c(n^{-K_c}) = o(1).$$

Therefore we can estimate the finite sum with an infinite one:

$$\mathbb{E}[S] := (1 \pm o(1)) \frac{n}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k.$$

Suppose now that  $xe^{-x} = ce^{-c}$  for  $x < 1$ , and consider  $G(n, x/n)$ . In that regime, all but  $o(n)$  vertices belong to small tree components (with a different constant  $A_x$ ), and so, repeating essentially the same calculations (with a bit more work) we obtain

$$n = (1 \pm o(1)) \frac{n}{x} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (xe^{-x})^k + o(n),$$

which in the limit  $n \rightarrow \infty$  implies that the infinite series equals  $x$ . Since  $xe^{-x} = ce^{-c}$ , it follows that

$$\mathbb{E}[S] = (1 \pm o(1)) \frac{x}{c} n.$$

However, we are not done yet: it could be that the number of vertices in small components tends to deviate a lot from its expectation. To show that this isn't the case, we will use *concentration of measure*, in the form of Chebyshev's inequality.

We will have two different arguments, depending on  $\mathbb{E}[T_k]$ . Let  $K_1 = \{k \leq A_c \log n : \mathbb{E}[T_k] \leq \log^2 n\}$  and  $K_2 = \{k \leq A_c \log n : \mathbb{E}[T_k] > \log^2 n\}$ . For every  $k \in K_1$ , Markov's inequality shows that  $T_k \leq \log^4 n$  with probability  $1 - 1/\log^2 n$ , and so whp  $T_k \leq \log^4 n$  for all  $k \in K_1$ . Since  $\mathbb{E}[T_k] \leq \log^2 n$ , this shows that whp  $|T_k - \mathbb{E}[T_k]| \leq O(\log^4 n)$  for all  $k \in K_1$ .

For  $k \in K_2$ , a calculation along the lines of Theorem 3.2 shows that

$$\mathbb{V}[T_k] \leq \mathbb{E}[T_k] + [(1-p)^{-k^2} - 1] \mathbb{E}[T_k]^2 = \left( \frac{1}{\mathbb{E}[T_k]} + [(1-p)^{-k^2} - 1] \right) \mathbb{E}[T_k]^2.$$

The second term is at most

$$(1-p)^{-k^2} - 1 \leq e^{O(pk^2)} - 1 = O(pk^2) = O\left(\frac{\log^2 n}{n}\right),$$

since for  $x \leq 1$  we have  $e^x \leq 1 + O(x)$ . In total, we get  $\mathbb{V}[T_k] = O(\mathbb{E}[T_k]^2 / \log^2 n)$ , and so Chebyshev's inequality shows that

$$\Pr[|T_k - \mathbb{E}[T_k]| \geq \mathbb{E}[T_k]/s] \leq \frac{\mathbb{V}[T_k]}{\mathbb{E}[T_k]^2} s^2 = O(s^2 / \log^2 n).$$

Choosing  $s = \log^{1/3} n$ , we get that  $|T_k - \mathbb{E}[T_k]| < \mathbb{E}[T_k] / \log^{1/3} n$  with probability  $O(1/\log^{4/3} n)$ . Hence whp,  $|T_k - \mathbb{E}[T_k]| < \mathbb{E}[T_k] / \log^{1/3} n$  for all  $k \in K_2$ .

Combining the contributions of both  $K_1$  and  $K_2$ , we get that whp

$$|S - \mathbb{E}[S]| \leq \sum_{k=1}^{A_c \log n} k |T_k - \mathbb{E}[T_k]| \leq \sum_{k \in K_1} A_c \log n \cdot O(\log^4 n) + \sum_{k \in K_2} \frac{k T_k}{\log^{1/3} n} \leq O(\log^6 n) + \frac{\mathbb{E}[S]}{\log^{1/3} n}.$$

Since  $\mathbb{E}[S] = \Theta(n)$  we have  $\log^6 n = o(\mathbb{E}[S])$ , and so in total, we get that whp  $|S - \mathbb{E}[S]| = o(\mathbb{E}[S])$ .  $\square$

This leaves roughly  $(1 - x/c)n$  vertices in large components. A short argument shows that all of these vertices must belong to the *same* connected component.

**Lemma 4.4.** *If  $p = c/n$  for  $c > 1$  then whp  $G(n, p)$  contains a unique large component.*

*Proof.* Let  $\delta = o(1/n)$  be a parameter to be chosen, and let  $p' = \frac{p-\delta}{1-\delta}$ . If  $G' \sim G(n, p')$  and  $G$  is obtained from  $G'$  by adding each edge not already in  $G'$  with probability  $\delta$ , then  $G \sim G(n, p)$ ; this is because the probability that an edge appears in the graph is  $p' + (1-p')\delta = (1-\delta)p' + \delta = p$ . Lemma 4.1, while nominally proved only for  $p$ , holds for  $p'$  as well, showing that whp each large component has size at least  $B'_c n$ . There are at most  $1/B'_c$  of these. Consider any two such components,  $C_1, C_2$ . The probability that none of the edges between  $C_1$  and  $C_2$  is added when moving to  $G$  is

$$(1-\delta)^{|C_1| \cdot |C_2|} \leq (1-\delta)^{B_c'^2 n^2} \leq e^{-\delta B_c'^2 n^2}.$$

If  $\delta = \omega(1/n^2)$  then this expression becomes  $e^{-\omega(1)} = o(1)$ , and so whp any two components will be connected in  $G$ .

Conversely, the expected number of edges added to  $G'$  is at most  $\binom{n}{2} \delta \leq n^2 \delta$ , and so if  $f(n) = \omega(1)$  then whp at most  $n^2 f(n) \delta$  edges are added. This means that any large component in  $G$  must have arisen by connecting at most  $n^2 f(n) \delta$  components of  $G'$ , one of which must have consisted of at least  $B_c n / (n^2 f(n) \delta) = B_c / (n f(n) \delta)$  vertices. If  $\delta = o(1/n \log n)$  then we can choose  $f(n)$  so that  $B_c / (n f(n) \delta) > A_c \log n$ , and so every large component in  $G$  contains a large component of  $G'$ .

Summarizing, if  $\delta = \omega(1/n^2)$  and  $\delta = o(1/n \log n)$ , then whp every large component of  $G$  contains a large component of  $G'$ , and any two large components of  $G'$  are connected in  $G$ . Choosing  $\delta = 1/n^{1.5}$  (say), this means that  $G$  contains a unique large component whp.  $\square$

Our work in this section is summarized in the following theorem.

**Theorem 4.5** ([FK16, Theorem 2.14]). *If  $p = c/n$  for  $c > 1$  then whp  $G(n, p)$  consists of a giant component containing  $(1 + o(1))(1 - x/c)n$  vertices, and tree components of size  $O(\log n)$ .*

## 5 Critical regime

So far we have addressed  $G(n, c/n)$  for  $c < 1$  (the subcritical regime) and for  $c > 1$  (the supercritical regime). Now we will analyze what happens at the threshold  $c = 1$ . This doesn't quite cover all the cases, since the behavior at  $c = 1 + \delta(n)$  for  $|\delta(n)| = o(1)$  doesn't quite match the behavior at  $c = 1$ , but it already reveals a curious phenomenon: whereas in the subcritical regime all connected components had logarithmic size and in the supercritical regime there was a giant component of linear size, here there will be several components of size roughly  $n^{2/3}$ .

The argument is divided into two parts. First we show that whp there exists a connected component of size roughly  $n^{2/3}$ . Then we show that whp there are no larger components. Surprisingly, the second part is much harder.

**Lemma 5.1.** *For every constant  $\epsilon > 0$ , with probability  $1 - \epsilon$ ,  $G(n, 1/n)$  has a tree component whose size is  $\Theta_\epsilon(n^{2/3})$ .*

*Proof.* Let  $k = Cn^{2/3}$ . The expected number of tree components of this size is

$$\mathbb{E}[T_k] := \binom{n}{k} k^{k-2} \left(\frac{1}{n}\right)^{k-1} \left(1 - \frac{1}{n}\right)^{k(n-k) + \binom{k}{2} - (k-1)} = \frac{n^k}{n^k} \frac{k^k}{k!} \frac{n}{k^2} \left(1 - \frac{1}{n}\right)^{k(n-k) + \binom{k}{2} - (k-1)}.$$

Since we are aiming at showing that this number is large, we will carefully estimate all factors in the expression. For the first factor, notice that  $\log(1 - x) = -x - x^2/2 - \Theta(x^3)$  implies that  $1 - x = e^{-x - x^2/2 - \Theta(x^3)}$ , and so

$$\begin{aligned} \frac{n^k}{n^k} &= \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \\ &= \exp \left[ -\frac{1 + \cdots + (k-1)}{n} - \frac{1^2 + \cdots + (k-1)^2}{2n^2} - \Theta \left( \frac{1^3 + \cdots + (k-1)^3}{n^3} \right) \right] \\ &= \exp \left[ -\frac{k(k-1)}{2n} - \frac{k(k-1)(2k-1)}{6n^2} - \Theta \left( \frac{k^4}{n^3} \right) \right] \\ &= \exp \left[ -\frac{C^2 n^{1/3}}{2} - \frac{C^3}{6} \pm \Theta \left( \frac{1}{n^{1/3}} \right) \right]. \end{aligned}$$

The factor  $k^k/k!$  equals  $(1 \pm o(1))e^k/\sqrt{2\pi k}$ . The remaining factor equals

$$\exp - \left[ \left( \frac{1}{n} + \frac{1}{2n^2} \right) \left( Cn^{5/3} - \frac{C^2 n^{4/3}}{2} \pm O(n^{2/3}) \right) \right] = \exp - \left[ Cn^{2/3} - \frac{C^2 n^{1/3}}{2} \pm O \left( \frac{1}{n^{1/3}} \right) \right].$$

In total, since  $n/k^{5/2} = n^{-2/3}/C^{5/2}$  we obtain

$$\begin{aligned} \mathbb{E}[T_k] &:= (1 \pm o(1)) \frac{n^{-2/3}}{C^{5/2} \sqrt{2\pi}} \exp - \left[ \left( \frac{C^2 n^{1/3}}{2} + \frac{C^3}{6} \right) - (Cn^{2/3}) + \left( Cn^{2/3} - \frac{C^2 n^{1/3}}{2} \right) \right] = \\ &= (1 \pm o(1)) \frac{n^{-2/3}}{C^{5/2} \sqrt{2\pi} e^{C^3/6}}. \end{aligned}$$

Let  $A \geq 1$  be some number. Let  $\mathcal{T}$  be the collection of all tree components whose size is between  $A^{-1}n^{2/3}$  and  $An^{2/3}$ , and for  $T \in \mathcal{T}$  let  $X_T$  denote the event that  $T$  is a component of  $G(n, p)$ . We define  $S = \sum_{T \in \mathcal{T}} X_T$ , that is,  $S$  is the number of tree components of  $G(n, p)$  whose size is between  $A^{-1}n^{2/3}$  and  $An^{2/3}$ . The preceding calculation shows that

$$\mathbb{E}[S] = \sum_{k=A^{-1}n^{2/3}}^{An^{2/3}} \mathbb{E}[T_k] = (1 \pm o(1)) \int_{1/A}^A \frac{1}{C^{5/2} \sqrt{2\pi} e^{C^3/6}} dC,$$

where the factor  $n^{-2/3}$  disappeared due to the substitution. Using the Taylor expansion of  $e^x$  it is not difficult to estimate this sum as

$$(1 \pm o(1)) \int_{1/A}^A \frac{1}{C^{5/2} \sqrt{2\pi} e^{C^3/6}} dC \approx (1 \pm o(1)) \int_{1/A}^A \frac{1}{C^{5/2} \sqrt{2\pi} \cdot 1} dC \approx (1 \pm o(1)) \frac{-2}{3\sqrt{2\pi}} C^{-3/2} \Big|_{A^{-1}}^A \approx \frac{\sqrt{2}}{3\sqrt{\pi}} A^{3/2} + O(\sqrt{A}).$$

Thus we expect there to be some tree component whose size is between  $A^{-1}n^{2/3}$  to  $An^{2/3}$ .

We now want to apply the second moment method:

$$\mathbb{E}[S^2] = \sum_{T_1, T_2 \in \mathcal{T}} \Pr[X_{T_1} \text{ and } X_{T_2}] = \mathbb{E}[S] + \sum_{\substack{T_1 \neq T_2 \\ T_1, T_2 \text{ compatible}}} (1-p)^{-|T_1| \cdot |T_2|} \Pr[X_{T_1}] \Pr[X_{T_2}].$$

In earlier applications of the method, it had been the case that the  $(1-p)^{-|T_1| \cdot |T_2|}$  factor were close to 1, but in our case it is roughly  $e^{|T_1| \cdot |T_2|/n} = e^{\Theta_A(n^{1/3})}$ . On the other hand, the sum is only over *compatible*  $T_1, T_2$ . The probability that random  $T_1, T_2$  are compatible is at most the probability that random sets of size  $|T_1|, |T_2| \approx n^{2/3}$  are disjoint, which is

$$\left(1 - \frac{|T_1|}{n}\right) \cdots \left(1 - \frac{|T_1|}{n - |T_2| + 1}\right) \approx \exp - \left[ \frac{|T_1|}{n} + \cdots + \frac{|T_1|}{n - |T_2| + 1} + \frac{|T_1|^2}{2n^2} + \cdots + \frac{|T_1|^2}{2(n - |T_2| + 1)^2} \right] \approx \exp - \left[ \frac{|T_1||T_2|}{n} + \frac{(|T_1| + |T_2|)|T_1||T_2|}{2n^2} \right].$$

Similarly,

$$(1-p)^{-|T_1||T_2|} \approx \exp \frac{|T_1||T_2|}{n}.$$

In both cases the approximation hides  $1 + o(1)$  factors. Since  $e^{-(|T_1|+|T_2|)|T_1||T_2|/2n^2} \leq 1$ , we deduce that

$$\mathbb{E}[S^2] \leq \mathbb{E}[S] + (1 + o(1)) \mathbb{E}[S]^2.$$

Chebyshev's inequality thus shows that

$$\Pr[S = 0] \leq \frac{\mathbb{V}[S]}{\mathbb{E}[S]^2} \leq \frac{1}{\mathbb{E}[S]} + o(1). \quad \square$$

If we choose  $\epsilon$  to be subconstant then we get a result holding with high probability, but we won't be able to pin down the order of magnitude of the tree component.

**Lemma 5.2.** *For every constant  $\epsilon > 0$ , with probability  $1 - \epsilon$ , all components of  $G(n, 1/n)$  have size  $O_\epsilon(n^{2/3})$ .*

*Proof.* Let  $V$  be the set of vertices of  $G(n, 1/n)$ . Fix some vertex  $x \in V$ . Let  $X_d$  be the number of vertices at distance  $d$  from  $x$ . We think of the sequence  $X_0, X_1, X_2, \dots$  as being determined by running BFS from  $x$ . In particular, at step  $d$ , after having discovered all vertices at distance at most  $d$ , we are unaware of the status of any edge other than those touching a vertex at distance less than  $d$  from  $x$ .

If  $X_d = \ell$  then a vertex is connected to one of these  $\ell$  vertices with probability at most  $\ell/n$ , and so the expected number of vertices at distance  $d+1$  is at most  $(n-\ell)\ell/n < \ell$  (since there are at most  $n-\ell$  potential vertices). Thus  $\mathbb{E}[X_{d+1}|X_d] < \mathbb{E}[X_d]$ , showing that the process  $X_d$  is a *submartingale*. In particular,  $\mathbb{E}[X_d] \leq 1$  for all  $d$ .

Let  $\pi_d = \Pr[X_d > 0]$ . We can estimate  $\pi_{d+1}$  as follows. First we find the  $X_1 \sim \text{Bin}(n-1, 1/n)$  vertices at distance one from  $x$ . Then we run a BFS process from each of them. Intuitively, for  $X_{d+1}$  to be positive, one of these processes must reach level  $d$ , and so  $\pi_{d+1} \leq 1 - (1 - \pi_d)^{X_1}$ .

Here is a more formal justification, using the technique of *coupling*. Let  $X_1 = m$  and let  $y_1, \dots, y_m$  be the neighbors of  $x$ . We will simulate  $m$  independent BFS processes, starting from  $y_1, \dots, y_m$ , generating



a coupled BFS process from  $x$  where in its first step the vertices  $y_1, \dots, y_m$  are discovered. All processes are run on the set of vertices  $V$ . We call them the  $x, y_1, \dots, y_m$  processes.

Each BFS process contains three types of vertices: active vertices (vertices discovered in the preceding step), old vertices (vertices discovered in previous steps), and undiscovered vertices. When running the BFS process from a vertex  $v$ , initially  $v$  is active, and all other vertices are undiscovered. At each step, we *expose* all edges from active vertices to undiscovered vertices: each such edge belongs to the graph with probability  $p = 1/n$ . All neighbors of active vertices are marked active, and then the previously active vertices are marked old. We will couple the  $x, y_1, \dots, y_m$  processes so that the  $x$  process individually behaves like a BFS from  $x$  conditioned on the neighbors of  $x$  being  $y_1, \dots, y_m$ , and the  $y_1, \dots, y_m$  behave like independent BFS processes from  $y_1, \dots, y_m$ . The coupling will have the property that if the  $x$  process survives for  $d + 1$  steps, then at least one of the  $y_1, \dots, y_m$  processes survives for  $d$  steps. The inequality  $\pi_{d+1} \leq 1 - (1 - \pi_d)^m$  will immediately follow.

For the purpose of the coupling, the active vertices of the  $y_1, \dots, y_m$  will be colored using the colors green and red. Red vertices are ones which are ignored in the  $x$  process. We initialize each  $y_i$  process with  $y_i$  as an active green vertex, and all other vertices are undiscovered. The  $x$  process is initialized with  $x$  as old and  $y_1, \dots, y_m$  as active. We then execute the following procedure  $d$  times. Run one step of the BFS process for the  $y_1, \dots, y_m$  processes. An active vertex is colored green if one of its previously active neighbors is green, and otherwise it is colored red. We then consider the set  $S$  of all green active vertices in the  $y_1, \dots, y_m$  processes, partitioning it into two parts:  $S_1$  consists of those already discovered by the  $x$  process, and  $S_2$  consists of the rest. We advance the  $x$  process by marking the vertices in  $S_2$  as active, and the previously active vertices as old. Then we adjust the colors: all active copies of vertices in  $S_1$  in the  $y_1, \dots, y_m$  processes are colored red. If a vertex in  $S_2$  is green active in several of the  $y_1, \dots, y_m$  processes, we choose one of them arbitrarily, and color all other copies red.

Our construction guarantees that at each step, there is a one to one correspondence between active vertices in the  $x$  process and green active vertices in the  $y_1, \dots, y_m$  processes. This ensures that the resulting  $x$  process behaves exactly like a BFS process (work it out!). The  $y_1, \dots, y_m$  processes are also independent BFS processes, as promised. Finally, by construction, if the  $x$  process lasts for  $d + 1$  steps then one of the  $y_1, \dots, y_m$  processes has lasted for  $d$  steps. This implies that  $\pi_{d+1} \leq 1 - (1 - \pi_d)^m$ . Considering all possible values of  $m$ , we conclude that

$$\pi_{d+1} \leq \mathbb{E}_{X \sim \text{Bin}(n-1, 1/n)} [1 - (1 - \pi_d)^X].$$

When  $n$  is large, the distribution  $\text{Bin}(n-1, 1/n)$  approaches a Poisson distribution with expectation 1, which we denote by  $\text{Po}(1)$ . Indeed, for each particular  $k$ ,

$$\Pr[\text{Bin}(n-1, 1/n) = k] = \binom{n-1}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-1-k} = \frac{(1-1/n)^n (n-1)^k}{k! n^k} (1-1/n)^{-1-k} \rightarrow \frac{e^{-1}}{k!}.$$

In this case we can say more<sup>2</sup>:  $\text{Po}(1)$  *stochastically dominates*  $\text{Bin}(n-1, 1/n)$ . That is, there is a coupling  $(X, Y)$  such that  $X \sim \text{Bin}(n-1, 1/n)$ ,  $Y \sim \text{Po}(1)$ , and  $X \leq Y$  always. Indeed, let  $\lambda = -\log(1-1/n)$ , and note that  $\tilde{\lambda} := 1 - (n-1)\lambda > 0$ , since  $(1-1/n)^{n-1} > 1/e$  classically. We define the coupling as follows:

- Let  $Y_1, \dots, Y_{n-1} \sim \text{Po}(\lambda)$  and  $\tilde{Y} \sim \text{Po}(\tilde{\lambda})$ .
- Let  $X_i$  indicate the event  $Y_i \geq 1$ .
- Define  $X = X_1 + \dots + X_{n-1}$  and  $Y = Y_1 + \dots + Y_{n-1} + \tilde{Y}$ .

Since  $\text{Po}(\alpha) + \text{Po}(\beta) \sim \text{Po}(\alpha + \beta)$ , it follows that  $Y \sim \text{Po}(1)$ . Furthermore, since

$$\Pr[Y_i \geq 1] = 1 - \Pr[Y_i = 0] = 1 - e^{-\lambda} = 1 - \left(1 - \frac{1}{n}\right) = \frac{1}{n},$$

we see that  $X_i \sim \text{Ber}(1/n)$  and so  $X \sim \text{Bin}(n-1, 1/n)$ . Finally, by construction  $X_i \leq Y_i$  and so  $X \leq Y$ .

<sup>2</sup>This folklore argument is adapted from [KM10]. They also prove the stronger result that  $\text{Bin}(n, 1/(n+1))$  stochastically dominates  $\text{Bin}(n-1, 1/n)$ .

This coupling easily implies that

$$\pi_{d+1} \leq \mathbb{E}_X[1 - (1 - \pi_d)^X] \leq \mathbb{E}_Y[1 - (1 - \pi_d)^Y] = 1 - e^{-1} \sum_{n=0}^{\infty} \frac{(1 - \pi_d)^n}{n!} = 1 - e^{-\pi_d}.$$

The right-hand side of the inequality is increasing in  $\pi_d$ , and so  $\pi_d \leq x_d$ , where  $x_d$  is given by the recurrence

$$x_{d+1} = 1 - e^{-x_d}, \quad x_0 = 1.$$

A Taylor expansion shows that  $x_{d+1} \leq x_d$ , and conversely  $x_d \geq 0$  can be proved by induction. Thus  $x_d$  approaches some non-negative limit  $x$ , which satisfies  $x = 1 - e^{-x}$ ; the only solution is  $x = 0$ . In other words,  $x_d \rightarrow 0$ .

For small  $x_d$ , we can approximate the right hand side by the first few terms of the Taylor series:

$$x_{d+1} \approx 1 - (1 - x_d + x_d^2/2 - \dots) = x_d - x_d^2/2 + \dots.$$

This suggests that the sequence  $x_d$  is comparable to the sequence  $y_d$  given by

$$y_{d+1} = y_d - y_d^2/2$$

and a suitable initial condition. This sequence, in turn, is comparable to the solution of the differential equation

$$Y' = -Y^2/2,$$

which is  $Y(d) = 2/(d + C)$  for some constant  $C$ . This leads us to conjecture that  $x_d = (1 + o(1))2/d$ .

Indeed, we now prove by induction that  $x_d \leq 2/d$ . This is clear for  $d \leq 2$ . For  $d > 2$  we have

$$x_{d+1} \leq 1 - e^{-x_d} \leq 1 - e^{-2/d} \leq 1 - \left(1 - \frac{2}{d} + \frac{4}{2d^2} - \frac{8}{6d^3}\right).$$

Thus

$$\frac{2}{d+1} - x_{d+1} \geq \frac{2}{d+1} - \frac{2}{d} + \frac{2}{d^2} - \frac{4}{3d^3} = \frac{2(d-2)}{3(d+1)d^3} \geq 0.$$

We can now estimate the probability that the connected component of  $x$  contains at least  $Cn^{2/3}$  vertices. Suppose that the BFS process terminates after  $d$  steps. With probability at most  $2/\sqrt{C}n^{1/3}$  it happens that  $d \geq 2\sqrt{C}n^{1/3}$ , and otherwise Markov's inequality shows that

$$\Pr[X_0 + \dots + X_d \geq Cn^{2/3}] \leq \frac{\mathbb{E}[X_0 + \dots + X_d]}{Cn^{2/3}} \leq \frac{2}{\sqrt{C}n^{1/3}}.$$

Thus the probability that the connected component of  $x$  contains at least  $Cn^{2/3}$  vertices is at most

$$\frac{4}{\sqrt{C}n^{1/3}}.$$

The expected number of vertices participating in such components is  $4n^{2/3}/\sqrt{C}$ . Markov's inequality shows that the probability that at least  $Cn^{2/3}$  vertices participate in such components is at most  $4/C^{3/2}$ , and this is an upper bound on the probability that such a component exists.  $\square$

Combining both results together, we obtain the following theorem.

**Theorem 5.3** ([FK16, Theorem 2.21]). *For every constant  $\epsilon > 0$ , with probability  $1 - \epsilon$ , the largest component of  $G(n, 1/n)$  has size  $\Theta_\epsilon(n^{2/3})$ .*

## 6 Connectivity

We have seen that at  $p = 1/n$ , the random graph  $G(n, p)$  undergoes a phase transition: a giant component emerges. Aside from the giant component, the graph also contains logarithmic-sized components. When are these swallowed completely by the giant component? It turns out that the toughest cookies to crack are isolated vertices, as the following lemma implies.

**Lemma 6.1.** *If  $p = (\log n + c)/n$ , for some constant  $c$ , then whp  $G(n, p)$  has no connected component whose size is in the range  $[2, \dots, n/2]$ .*

*Proof.* Given  $k$ , we will estimate the expectation of the number of spanning trees of connected components of size  $k \leq n/2$ :

$$\begin{aligned} \mathbb{E}[T_k] &:= \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \\ &\leq (1+o(1)) n^k \frac{k^k}{k! \log n} p^k (1-p)^{k(n-k)} \\ &\leq O\left(\frac{n}{\log n}\right) \left(\frac{e(\log n + c)}{(e^n)^{1-k/n}}\right)^k \end{aligned}$$

since  $1-p \leq (e^{-pn})^{1/n} = (e^n)^{-1/n}$ . Since  $k \leq n/2$ , we can bound this by

$$\mathbb{E}[T_k] \leq O\left(\frac{n}{\log n}\right) \left(\frac{e(\log n + c)}{\sqrt{e^n}}\right)^k.$$

We can assume that  $n$  is large enough so that  $e(\log n + c) \leq \sqrt{e^n}$ . This shows that for  $k \geq 5$ ,

$$\mathbb{E}[T_k] \leq O\left(\frac{n}{\log n} \cdot \frac{\log^5 n}{n^{5/2}}\right) = O\left(\frac{\log^4 n}{n^{3/2}}\right) = o\left(\frac{1}{n}\right),$$

and so whp there are no connected components of size  $5 \leq k \leq n/2$ .

To handle small  $k$ , we will need to be a bit more careful: for each constant  $k$ ,

$$\mathbb{E}[T_k] \leq O\left(\frac{n}{\log n}\right) \left(\frac{e(\log n + c)}{e^n}\right)^k (e^n)^{k^2/n} \leq O\left(\frac{\log^{k-1} n}{n^{k-1}}\right).$$

This is  $o(1)$  for  $k \geq 2$ , and so whp there are also no connected components of size  $2 \leq k \leq 4$ .  $\square$

The startling implication is that the only obstruction for connectedness is isolated vertices! (We'll work this out in detail later.) So we turn our focus to studying the number of isolated vertices. When  $p = (\log n + c)/n$ , the probability that a given vertex is isolated is  $(1-p)^{n-1} \approx e^{-c}/n$ , and so the expected number of isolated vertices is roughly  $e^{-c}$ . Intuitively, there is only slight dependence between different vertices, and so we expect the distribution of the number of isolated vertices to be roughly Poisson, and this explains the mysterious probability  $e^{-e^{-c}}$ .

**Theorem 6.2** ([FK16, Theorem 3.1(ii)]). *If  $p = \frac{\log n + c}{n}$  then for every  $k$ , the probability that  $G(n, p)$  has exactly  $k$  isolated vertices tends to*

$$e^{-e^{-c}} \frac{e^{-ck}}{k!},$$

*which matches the distribution of a  $\text{Po}(e^{-c})$  random variable.*

*Proof.* Let  $X$  denote the number of isolated vertices. We start with the case  $k = 0$ . The idea is to use inclusion-exclusion. For a set  $S$  of vertices, let  $I_S$  denote the event that all vertices in  $S$  are isolated. This event has probability

$$\Pr[I_S] = (1-p)^{\binom{|S|}{2} + |S|(n-|S|)}.$$

The inclusion-exclusion principle (known in this context as the Bonferroni inequalities) shows that for each fixed  $\ell$ ,

$$1 - \sum_{|S|=1} \Pr[I_S] + \cdots - \sum_{|S|=2\ell+1} \Pr[I_S] \leq \Pr[X=0] \leq 1 - \sum_{|S|=1} \Pr[I_S] + \cdots + \sum_{|S|=2\ell} \Pr[I_S].$$

There are  $\binom{n}{\ell}$  sets of size  $\ell$ , and so for any fixed  $\ell$ ,

$$\sum_{|S|=\ell} \Pr[I_S] = \binom{n}{\ell} (1-p)^{n\ell - O(\ell^2)} \rightarrow \frac{e^{-c\ell}}{\ell!},$$

since  $\binom{n}{\ell} \approx n^\ell/\ell!$ ,  $(1-p)^{n\ell} \approx e^{-(\log n+c)\ell} = e^{-c\ell}/n^\ell$ , and  $(1-p)^{O(\ell^2)} = 1 - o(1)$ . Therefore for any fixed  $\ell$ ,

$$\sum_{|S|=\ell} \Pr[I_S] \rightarrow \frac{e^{-c\ell}}{\ell!}.$$

Thus for every  $\ell$  and  $\epsilon > 0$ , for large enough  $n$

$$1 - \frac{e^{-c}}{1!} + \cdots - \frac{e^{-(2\ell+1)c}}{(2\ell+1)!} - \epsilon \leq \Pr[X=0] \leq 1 - \frac{e^{-c}}{1!} + \cdots + \frac{e^{-2\ell c}}{(2\ell)!} + \epsilon.$$

The series  $1 - e^{-c}/1! + e^{-2c}/2! - \cdots$  converges to  $e^{-e^{-c}}$ . Thus for any  $\epsilon > 0$  there exists  $\ell$  such that

$$1 - \frac{e^{-c}}{1!} + \cdots - \frac{e^{-(2\ell+1)c}}{(2\ell+1)!} \geq e^{-e^{-c}} - \epsilon, \quad 1 - \frac{e^{-c}}{1!} + \cdots + \frac{e^{-2\ell c}}{(2\ell)!} \leq e^{-e^{-c}} + \epsilon.$$

Altogether, we deduce that for any  $\epsilon > 0$ , for large enough  $n$ ,

$$e^{-e^{-c}} - 2\epsilon \leq \Pr[X=0] \leq e^{-e^{-c}} + 2\epsilon.$$

This implies that

$$\Pr[X=0] \rightarrow e^{-e^{-c}}.$$

The case of general  $k$  is similar, and left to the reader.  $\square$

We can conclude our main result, stating the threshold of connectedness.

**Theorem 6.3** ([FK16, Theorem 4.1, Theorem 4.2]). *If  $p = \frac{\log n + c}{n}$  then the probability that  $G(n, p)$  is connected tends to  $e^{-e^{-c}}$ .*

*Furthermore, whp  $p_{\text{Con}} = p_{\text{Iso}}$ .*

*Proof.* The graph  $G(n, p)$  is connected if there is no connected component whose size is at most  $n/2$ , since there can only be one connected component of larger size. This, in turn, happens with probability  $o(1) + (1 \pm o(1))e^{-e^{-c}}$ , and so  $G(n, p)$  is connected with probability tending to  $e^{-e^{-c}}$ .

It remains to prove that  $p_{\text{Con}} = p_{\text{Iso}}$ . The idea is to show that whp, at time  $p_- = \frac{\log n - \log \log n}{n}$  the graph isn't connected, while at time  $p_+ = \frac{\log n + \log \log n}{n}$  the graph is connected. The small gap makes it unlikely that isolated vertices will be connected. We will consider graphs  $G_- \sim G(n, p_-)$  and  $G_+ \sim G(n, p_+)$  generated using the coupling process; thus  $G_+ \setminus G_-$  consists of those edges whose timestamp is between  $p_-$  and  $p_+$ .

The graph coupling easily implies that when  $p = (\log n - \omega(1))/n$ , whp  $G(n, p)$  isn't connected, and when  $p = (\log n + \omega(1))/n$ , whp  $G(n, p)$  is connected. In particular, whp  $G_-$  is not connected whereas  $G_+$  is connected. Moreover, Lemma 6.1 holds for  $p_-$  as well; the only difference in the calculations is a few more logarithmic terms. Thus whp,  $p_- \leq p_{\text{Iso}} \leq p_{\text{Con}} \leq p_+$ .

The expected number of isolated vertices at time  $p_-$  is

$$\begin{aligned} n(1-p_-)^{n-1} &= n \exp[(n-1) \log(1-p_-)] = n \exp[(n-1)(-p_- - \Theta(p_-^2))] = \\ &= n \exp \left[ -\log n + \log \log n \pm \Theta \left( \frac{\log^2 n}{n} \right) \right] = (1 + o(1)) \log n. \end{aligned}$$

Markov's inequality shows that whp, the number of isolated vertices at time  $p_-$  is at most  $\log^2 n$ .

Conditioned on  $G_-$ , the probability that an edge gets added between time  $p_-$  and time  $p_+$  is  $\frac{2 \log \log n/n}{1-p_-} = O(\frac{\log \log n}{n})$ . There are at most  $\log^4 n$  edges which connect two isolated vertices, and the probability that any of them gets added between time  $p_-$  and time  $p_+$  is  $O(\frac{\log^4 n \log \log n}{n}) = o(1)$ . Thus, whp all edges in  $G_+ \setminus G_-$  touch the giant component. In particular, whp the edge that makes the graph connected connects an isolated vertex to the giant component, showing that whp  $p_{\text{Con}} = p_{\text{Iso}}$ .  $\square$

**Connection to the coupon collector problem** In the coupon collector problem, we are given an infinite sequence of coupons, each of which is a uniformly random number drawn from  $[n]$ . The question is how long we have to wait until we have collected all different coupons. If we denote by  $Y_i$  the first time we have  $i$  different coupons and  $X_i = Y_i - Y_{i-1}$ , then  $X_i$  has geometric distribution  $G(1 - (i-1)/n)$ , and so

$$\mathbb{E}[Y_n] = \sum_{i=1}^n \mathbb{E}[X_i] = n \left( \frac{1}{n} + \frac{1}{n-1} + \dots + 1 \right) = nH_n = n \log n + \gamma n + O(1),$$

where  $H_n$  is the  $n$ th harmonic number. We thus expect to collect about  $n \log n$  coupons until we see all different kinds. This suggests calculating the probability that after collection  $m = n \log n + cn$  coupons, we have collected all of them. For  $S \subseteq [n]$ , let  $E_S$  be the event that we haven't seen the coupons in  $S$  after collecting  $m$  coupons. The probability that we have seen all coupons is

$$1 - \sum_i E_{\{i\}} + \sum_{i \neq j} E_{\{i,j\}} - \dots = 1 - n \left(1 - \frac{1}{n}\right)^m + \binom{n}{2} \left(1 - \frac{2}{n}\right)^m - \dots.$$

Mimicking the proof of Theorem 6.2, this probability tends to

$$1 - ne^{-m/n} + \frac{n^2}{2} e^{-2m/n} - \dots = 1 - e^{-c} + \frac{e^{-2c}}{2!} - \dots = e^{-e^{-c}}.$$

What is the connection to connectivity? Consider the  $G(n, m)$  model. When the number of edges is relatively small, say  $o\left(\sqrt{\binom{n}{2}}\right)$ , there isn't a big difference between sampling edges with and without replacement. When we sample edges with replacement, this is like the coupon collectors problem, only we get two coupons for each edge (which also have to be different). This naturally leads to the conjecture that the critical  $m$  is about  $n \log n/2$  (although this is not  $o\left(\sqrt{\binom{n}{2}}\right)$ , which corresponds to  $p = \log n/n$ ). Moreover, the critical window  $n \log n + cn$  in the coupon collectors problem translates to  $m = (n \log n + cn)/2$  and to  $p = (\log n + c)/n$ . Our calculations above show that the various dependencies don't effect the process by much.

**More on Poisson distributions** There are two different models that give rise to Poisson distributions. The first one is  $\text{Bin}(n, \lambda/n)$ . As  $n$  tends to infinity, the binomial distribution converges in probability to  $\text{Po}(\lambda)$ . Indeed,

$$\Pr[\text{Bin}(n, \frac{\lambda}{n}) = k] = \binom{n}{k} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

since  $\binom{n}{k} \sim n^k/k!$ ,  $(1 - \lambda/n)^n \rightarrow e^{-\lambda}$ , and  $(1 - \lambda/n)^{-k} = 1 + o(1)$ .

The second model is the one with exponential clocks. Consider an infinite sequence  $T_i$  of variables with standard exponential distribution  $E(1)$  (so  $\Pr[T_i > t] = e^{-t}$ ), and the corresponding partial sums sequence  $T_1, T_1 + T_2, \dots$ . We can think of the partial sums sequence as describing the following process: at time zero, we start an exponential clock, and when it "arrives", we mark this, and start another one, and so on. The partial sums sequence marks the arrival times. The number of arrivals until time  $\lambda$  has distribution  $\text{Po}(\lambda)$ . We can see this by dividing the interval  $[0, \lambda]$  into  $n$  parts, and using the alternative definition of the exponential distribution as a memoryless distribution which on an interval of infinitesimal length  $\epsilon$  has a probability of  $\epsilon$  to "buzz" (given that it hasn't buzzed so far). In some sense, this exponential process is the limit of  $\text{Bin}(n, \lambda/n)$  at  $n = \infty$ .

## 7 Subgraphs

### 7.1 Subgraph thresholds

Fix a graph  $H$ . How large should  $p$  be to ensure that whp  $G(n, p)$  contains a (non-induced) copy of  $H$ ? It is easy to get a lower bound on  $p$ , by calculating the expected number of copies of  $H$ . The exact answer will depend on the number of automorphisms of  $H$ . We will use the following notation:

- $V(H)$  is the set of vertices of  $H$ , and  $v(H)$  is their number.
- $E(H)$  is the set of edges of  $H$ , and  $e(H)$  is their number.
- $A(H)$  is the automorphism group of  $H$ , and  $a(H)$  is its size.

**Lemma 7.1** ([FK16, Theorem 5.2]). *Given a graph  $H$ , the expected number of copies of  $H$  in  $G(n, p)$  is*

$$\frac{1 - o(1)}{a(H)} n^{v(H)} p^{e(H)}.$$

*Consequently, if  $p = o(n^{-v(H)/e(H)})$  then whp  $G(n, p)$  contains no copy of  $H$ , and if  $p = \omega(n^{-v(H)/e(H)})$  then the expected number of copies of  $H$  in  $G(n, p)$  is  $\omega(1)$ .*

*Proof.* The idea is to consider the expected number of mappings  $\{1, \dots, n\} \rightarrow V(H)$  which define a copy of  $H$ . This counts each copy of  $H$  in  $G$  exactly  $a(H)$  times, and so we will divide by  $a(H)$  to get expected number of copies. Each “ordered” copy of  $H$  appears with probability  $p^{e(H)}$ , and so the expected number of copies is

$$\frac{1}{a(H)} n^{v(H)} p^{e(H)} = \frac{1 - o(1)}{a(H)} n^{v(H)} p^{e(H)}. \quad \square$$

If  $p = \omega(n^{-v(H)/e(H)})$  then the expected number of copies of  $H$  is  $\omega(1)$ . Does that guarantee that  $G$  contains a copy of  $H$  whp? Consider the following example:  $K_4$  with an extra edge attached. Lemma 7.1 predicts a threshold of  $n^{-5/7}$ . However, the same lemma shows that  $K_4$  itself only appears at  $\Omega(n^{-2/3})$ , and  $n^{-5/7} = o(n^{-2/3})$ ! What is happening here? The reason that there are  $\omega(1)$  copies of  $H$  at  $\omega(n^{-5/7})$  is that each copy of  $K_4$  (if any) creates roughly  $\omega(n^{2/7})$  copies of  $H$ . So even though it is unlikely that  $H$  will appear at all, if it does appear (due to chance occurrence of  $K_4$ ) then many copies of  $H$  are likely to appear.

It is easy at this point to formulate another guess at the correct threshold.

**Definition 7.2.** The *density* of a graph  $H$  is  $d(H) = e(H)/v(H)$ , which is half the average degree of  $H$ . The *maximum subgraph density* (MSD) of a graph  $H$ , denoted  $m(H)$ , is the maximum density of a subgraph of  $H$ .

A graph  $H$  is *balanced* if  $m(H) = d(H)$ . It is *strictly balanced* if  $d(K) < d(H)$  for all proper subgraphs  $K$  of  $H$ .

Lemma 7.1 shows that if  $p = o(n^{-1/d(H)})$  then whp  $G(n, p)$  contains no copy of  $H$ . The following theorem finds the correct threshold.

**Theorem 7.3** ([FK16, Theorem 5.3]). *Let  $H$  be a graph. If  $p = o(n^{-1/m(H)})$  then whp  $G(n, p)$  contains no copy of  $H$ , and if  $p = \omega(n^{-1/m(H)})$  then whp  $G(n, p)$  does contain a copy of  $H$ .*

*Proof.* The first claim follows directly from Lemma 7.1. We will prove the second claim using the second moment method, showing that  $\mathbb{V}[N_H] = o(\mathbb{E}[N_H]^2)$  when  $p = \omega(n^{-1/m(H)})$ , where  $N_H$  is the number of copies of  $H$ . Since  $\mathbb{E}[N_H] = \omega(1)$  by Lemma 7.1, this will complete the proof.

Let  $H_i$  be an enumeration of all possible copies of  $H$ , each appearing exactly once. We have

$$\begin{aligned} \mathbb{V}[N_H] &= \mathbb{E}[N_H^2] - \mathbb{E}[N_H]^2 \\ &= \sum_{i,j} \Pr[H_i \cup H_j \in G(n, p)] - \sum_{i,j} \Pr[H_i \in G(n, p)] \Pr[H_j \in G(n, p)] \\ &= \sum_{i,j} (\Pr[H_i \cup H_j \in G(n, p)] - p^{2e(H)}) \\ &= \sum_{K \neq \emptyset} \sum_{\substack{i,j: \\ H_i \cap H_j \approx K}} p^{2e(H)} (p^{-e(K)} - 1). \end{aligned}$$

In the last step, we go over all possible isomorphism types of intersections of two copies of  $H$ ; if  $H_i \cap H_j = \emptyset$  then the corresponding sum vanishes, so we can ignore such terms. Given  $K$ , a pair of copies of  $H$  with intersection  $K$  is a structure consisting of  $2v(H) - v(K)$  vertices, and so there are  $O(n^{2v(H)-v(K)})$  of these. This shows that

$$\begin{aligned} \mathbb{V}[N_H] &= \sum_{K \neq \emptyset} O(n^{2v(H)-v(K)} p^{2e(H)-e(K)})(1 - p^{e(K)}) \\ &\leq O\left(n^{2v(H)} p^{2e(H)} \sum_{K \neq \emptyset} n^{-v(K)} p^{-e(K)}\right). \end{aligned}$$

Since  $d(K) \leq m(H)$ , for each  $K \neq \emptyset$  we have

$$n^{-v(K)} p^{-e(K)} = (n^{-1} p^{-d(K)})^{v(K)} = (n^{-1} o(n^{d(K)/m(H)}))^{v(K)} = o(1).$$

Since  $\mathbb{E}[N_H] = \Theta(n^{v(H)} p^{e(H)})$ , we conclude that  $\mathbb{V}[N_H] = o(\mathbb{E}[N_H]^2)$ , completing the proof.  $\square$

## 7.2 Subgraph counts

Suppose that  $p = c/n^{1/m(H)}$ . Lemma 7.1 shows that the expected number of copies of  $H$  in  $G(n, p)$  tends to

$$\frac{c^{e(H)}}{a(H)}.$$

As in the case of isolated vertices, if  $H$  is balanced, there is only slight dependence between the appearance of different copies of  $H$ , and so it is natural to expect that the distribution is roughly Poisson. This turns out not to be the case in general, but a Poisson law does hold when  $H$  is *strictly* balanced.

**Theorem 7.4** ([FK16, Theorem 5.4]). *Let  $H$  be a strictly balanced graph. If  $p = cn^{-1/d(H)}$  then for every  $k$ , the probability that  $G(n, p)$  has exactly  $k$  copies of  $H$  tends to*

$$e^{-\lambda} \frac{\lambda^k}{k!}, \text{ where } \lambda = \frac{c^{e(H)}}{a(H)},$$

which matches the distribution of a  $\text{Po}(\lambda)$  random variable.

*Proof.* As in the proof of Theorem 6.2, we will only prove the case  $k = 0$ , the general case being very similar. Let  $H_i$  be a list of all possible copies of  $H$ , as in the proof of Theorem 7.3. We will show that for every  $\ell \geq 1$ ,

$$\sum_{i_1 < \dots < i_\ell} \Pr[H_{i_1}, \dots, H_{i_\ell} \in G(n, p)] \rightarrow \frac{\lambda^\ell}{\ell!}.$$

Then we can conclude the proof as in Theorem 6.2.

We can decompose the sum into two parts: pairwise vertex-disjoint  $H_{i_1}, \dots, H_{i_\ell}$ , and all other cases. The number of disjoint  $\ell$ -tuples is  $(1 - o(1))n^{v(H)}/a(H)^\ell/\ell!$  (the  $\ell!$  factor comes from the condition  $i_1 < \dots < i_\ell$ ), and each of them appears in the graph with probability  $p^{e(H)\cdot\ell}$ . The total contribution of the disjoint  $\ell$ -tuples is thus

$$\frac{1 - o(1)}{\ell!} \left(\frac{n^{v(H)}}{a(H)}\right)^\ell \left(\frac{c^{e(H)}}{n^{e(H)/d(H)}}\right)^\ell = \frac{1 - o(1)}{\ell!} \left(\frac{c^{e(H)}}{a(H)}\right)^\ell \rightarrow \frac{\lambda^\ell}{\ell!}.$$

Our goal is, therefore, to show that all other cases contribute  $o(1)$  to the sum.

Consider a structure composed of  $\ell$  copies of  $H$ , not all of them disjoint. We will show that the density of each such structure is strictly larger than  $d(H)$ . As a result, the total contribution of all such structures  $S$  is

$$O(n^{v(S)} p^{e(S)}) = O((n^{1-d(S)/d(H)})^{v(S)}) = o(1).$$

It remains to estimate the density of graphs formed by putting together  $\ell$  copies of  $H$ . Let  $H_1, \dots, H_\ell$  be  $\ell$  graphs which are isomorphic to  $H$ , where  $H_i \neq H_j$  for all  $i \neq j$ , and let  $T_i = H_1 \cup \dots \cup H_i$ . We will

prove by induction that  $d(T_i) \geq d(H_i)$ , and furthermore  $d(T_i) > d(H_i)$  if  $H_1, \dots, H_i$  are not pairwise vertex-disjoint. This will complete the proof.

The claim is clear for  $i = 1$ . Suppose now that it holds for  $T_i$ . To prove it for  $T_{i+1}$ , consider the graph  $K = (V(T_i) \cap V(H_{i+1}), E(T_i) \cap E(H_{i+1}))$ . Since  $H_1, \dots, H_i$  are all different,  $K$  is isomorphic to a strict subset of  $H$ , and so either  $K = \emptyset$  or  $d(K) < d(H)$ . We have

$$d(T_{i+1}) = \frac{e(T_i) + e(H) - e(K)}{v(T_i) + v(H) - v(K)} \geq \frac{d(T_i)v(T_i) + d(H)v(H) - d(K)v(K)}{v(T_i) + v(H) - v(K)}.$$

Since  $d(T_i) \geq d(H)$  and either  $v(K) = 0$  or  $d(K) > d(H)$ , this is always at least  $d(H)$ . Moreover, if either  $d(T_i) > d(H)$  (which happens when  $H_1, \dots, H_i$  are not vertex-disjoint) or  $v(K) \neq 0$  (which happens when  $H_{i+1}$  vertex-intersects  $H_1, \dots, H_i$ ), this is strictly larger than  $d(H)$ . This completes the inductive proof.  $\square$

## 8 Thresholds

### 8.1 Sharp and coarse thresholds

We have seen several thresholds so far. Two of the most prominent of them are the connectivity threshold and the subgraph appearance thresholds. Let us compare the connectivity threshold and the threshold for the appearance of a triangle:

1. If  $p = \frac{\log n}{n} + \frac{c}{n}$  then the probability that  $G(n, p)$  is connected tends to  $e^{-e^{-c}}$ .
2. If  $p = \frac{c}{n}$  then the probability that  $G(n, p)$  contains a triangle tends to  $1 - e^{-c^3/6}$ .

The *critical scale* for connectivity has order of magnitude  $1/n$ , which is asymptotically smaller than the threshold itself  $\log n/n$ . In contrast, the critical scale for triangles has order of magnitude  $1/n$ , matching the threshold. We say that the former threshold is *sharp*, and the latter *coarse*. More formally:

**Definition 8.1.** A *graph property* is a property of graphs which is invariant under permutation of the vertices. A property is *non-trivial* if for all  $n$  there is some graph satisfying the property and some graph not satisfying the property.

Let  $P$  be a non-trivial monotone graph property, and let  $p^*(n)$  be the probability at which  $\Pr[G(n, p^*) \in P] = 1/2$ . The property  $P$  has a *sharp threshold* if:

1. For all  $\epsilon > 0$ ,  $\Pr[G(n, (1 - \epsilon)p^*) \in P] \rightarrow 0$ .
2. For all  $\epsilon > 0$ ,  $\Pr[G(n, (1 + \epsilon)p^*) \in P] \rightarrow 1$ .

Otherwise  $P$  has a *coarse threshold*.

Which properties have sharp thresholds? Friedgut's celebrated sharp threshold theorem [Fri99] roughly says that a monotone graph property has a sharp threshold unless it is "correlated" with the appearance of small graphs. In other words, the property of containing a subgraph is *more or less* the only kind of property having a coarse threshold.

On the other, every monotone graph property does have a threshold, in several senses. The first is very simple.

**Theorem 8.2.** Let  $P$  be a non-trivial monotone graph property, and let  $p^*(n)$  be the probability such that  $\Pr[G(n, p^*) \in P] = 1/2$ . Then:

1. If  $p(n) = o(p^*(n))$  and  $p^*(n) = o(1)$  then  $\Pr[G(n, p) \in P] \rightarrow 0$ .
2. If  $p(n) = \omega(p^*(n))$  then  $\Pr[G(n, p) \in P] \rightarrow 1$ .

*Proof.* We start with the second result. Let  $C(n) = \lfloor p(n)/p^*(n) \rfloor \rightarrow \infty$ . If  $G_1, \dots, G_C \sim G(n, p^*)$  then the probability that at least one of  $G_1, \dots, G_C$  has property  $P$  is  $1 - 2^{-C}$ , and so the union  $G_1 \cup \dots \cup G_C$  satisfies property  $P$  with probability at least  $1 - 2^{-C}$ . On the other hand, the union  $G_1 \cup \dots \cup G_C$



contains each edge with probability  $1 - (1 - p^*)^C \leq Cp^* \leq p$ . The graph coupling thus implies that  $G(n, p)$  satisfies property  $P$  with probability  $1 - 2^{-C} = 1 - o(1)$ .

The first result is very similar. Let  $C(n)$  be the largest integer satisfying  $1 - (1 - p^*)^C \leq p$ . Since

$$1 - (1 - p^*)^C \geq Cp^* - \binom{C}{2} p^{*2} = Cp^* (1 - \frac{1}{2}(C-1)p^*),$$

the assumption  $p^* = o(1)$  implies that  $C \rightarrow \infty$ . We obtain the required result by taking the *intersection* of  $C$  graphs  $G_1, \dots, G_C$ . Details left to the reader.  $\square$

## 8.2 Friedgut–Kalai threshold theorem

Another important result is due to Friedgut and Kalai [FK96].

**Theorem 8.3.** *Let  $P$  be a non-trivial monotone graph property, and let  $p^*(n)$  be the probability such that  $\Pr[G(n, p^*) \in P] = 1/2$ . If  $p^* \leq 1/2$  then:*

1. *For small enough  $C > 0$ ,  $\Pr[G(n, p^* - Cp^* \log(1/p^*)/\log n) \in P] \leq O(1/C)$ .*
2. *For small enough  $C > 0$ ,  $\Pr[G(n, p^* + Cp^* \log(1/p^*)/\log n) \in P] \geq 1 - O(1/C)$ .*

The theorem states that the critical window always has width  $O(p^* \log(1/p^*)/\log n)$ , and so there is a sharp threshold if  $p^* = 1/n^{o(1)}$ . Unfortunately we won't be able to provide a complete proof of the theorem; this would require us to delve into analysis of Boolean functions. However, we will indicate the proof of the theorem up to a basic result in that field, the KKL theorem [KKL88].

*Proof.* We start with a formula for the derivate of  $r(p) = \Pr[G(n, p) \in P]$ , the Russo–Margulis formula.

For a graph  $G$  and an edge  $e \notin G$ , we say that  $e$  is *influential* for  $G$  if  $G - e \notin P$  but  $G + e \in P$ . Define  $\iota(G)$  to be the number of edges influential in  $G$ . (This is known as *sensitivity* of  $P$  at  $G$ ; the *influence* of  $P$  at  $G$  is  $p(1-p)\iota(G)$ .)

Consider a new model for random graphs, in which each edge  $e$  is put in with probability  $p_e$ ; we denote this model by  $G(n, \vec{p})$ . The expression  $\Pr[G(n, \vec{p}) \in P]$ , as a function of the  $p_e$ , is multilinear, hence its derivative at  $p_e$  equals

$$\Pr_{G \sim G(n, \vec{p})} [G + e \in P] - \Pr_{G \sim G(n, \vec{p})} [G - e \in P],$$

since the derivative of  $ax + b$  is  $a = (a \cdot 1 + b) - (a \cdot 0 + b)$ . If we define indicator variables  $I_+, I_-$  for the events that  $G + e \in P$  and  $G - e \in P$ , then the expression above is just  $\mathbb{E}[I_+ - I_-]$ . This is non-zero precisely when  $G + e \in P$  but  $G - e \notin P$ . That is,

$$\frac{\partial}{\partial p_e} \Pr[G(n, \vec{p}) \in P] = \Pr[e \text{ is influential for } G(n, \vec{p})].$$

Since  $G(n, p)$  is the same as  $G(n, \vec{p})$  for the constant vector  $p_e = p$ , the chain rule shows that

$$\frac{d}{dp} \Pr[G(n, p) \in P] = \sum_e \frac{\partial}{\partial p_e} \Pr[G(n, \vec{p}) \in P] = \mathbb{E}[\iota(G(n, p))].$$

The KKL theorem states that there is an edge  $e$  such that the probability that  $e$  is influential for  $G \sim G(n, p)$  is

$$\Omega\left(\frac{r(p)(1-r(p)) \log n}{p \log(1/p) n^2}\right).$$

Since  $P$  is a graph property, the probabilities are the same for all edges, and so

$$r'(p) = \mathbb{E}[\iota(G(n, p))] = \Omega\left(\frac{r(p)(1-r(p)) \log n}{p \log(1/p)}\right).$$

When  $C$  is small enough,  $p \log(1/p) = \Theta(p^* \log(1/p^*))$  whenever  $|p - p^*| \leq Cp^* \log(1/p^*)/\log n$ .

Suppose now that  $r(p^* + Cp^* \log(1/p^*)/\log n) = 1 - \delta$ . Then

$$r(p^* + Cp^* \log(1/p^*)/\log n) = r(p^*) + \int_0^{Cp^* \log(1/p^*)/\log n} r'(p^* + x) dx \geq \frac{1}{2} + \frac{Cp^* \log(1/p^*)}{\log n} \cdot \frac{\frac{1}{2}\delta \log n}{\Theta(p^* \log(1/p^*))} = \frac{1}{2} + \Omega(C).$$

This directly implies that  $C\delta = O(1)$ , and so  $\delta = O(1/C)$ .

Similarly, if  $r(p^* - Cp^* \log(1/p^*)/\log n) = \delta$  then  $\delta = O(1/C)$ .  $\square$

It is a nice exercise to improve the dependence on  $C$  by shortening the interval from  $\frac{Cp^* \log(1/p^*)}{\log n}$  to  $\frac{\log(1/C)p^* \log(1/p^*)}{\log n}$ . It also suffices to consider  $C < 1/4$  (say) rather than just small enough  $C > 0$ .

Theorem 8.3 holds in greater generality. The property  $P$  need not be a graph property. Instead, we just require the property to have enough symmetries. Technically, we require it to be invariant under some transitive permutation group (a permutation group is *transitive* if for all  $i, j$ , it contains some permutation mapping  $i$  to  $j$ ). As an example, consider the *tribes* function:

$$f(x_1, \dots, x_n) = \bigvee_{i=1}^{n/m} \bigwedge_{j=1}^m x_{(i-1)m+j},$$

where  $m = \log_2(n/\log_2 n)$ . For this function,

$$r(p) = 1 - (1 - p^m)^{n/m} \approx 1 - e^{-(n/m)p^m}.$$

When  $p = 1/2$ ,  $p^m = \log_2 n/n \approx m/n$ , and so  $p^*$  is very close to  $1/2$ . When  $p = \frac{1}{2}(1 + c/\log_2 n)$ , we have  $p^m \approx 2^{-m}(1 + cm/\log_2 n) \approx 2^{-m}(1 + c)$ , and so  $r(p) \approx 1 - e^{-(1+c)}$ . This shows that the threshold interval for the tribes function is  $1/\log n$ , and so Theorem 8.3 is tight in its more general formulation.

Bourgain and Kalai [BK97] strengthened Theorem 8.3, showing that the threshold interval for monotone graph properties is always at most  $1/\log^2 n$ . This is tight for the property of containing a clique  $K_k$  for an appropriate value of  $k$ . Recently their result has been used to analyze the capacity of Reed–Muller codes on erasure channels [KKM<sup>+</sup>16].

## 9 Cliques and coloring

### 9.1 Cliques

Earlier we commented that a random  $G(n, 1/2)$  graph has good Ramsey-theoretic properties: its maximal clique and independent set are both at most roughly  $2\log_2 n$ . It turns out that with more effort, one can show that both of these are concentrated quite strongly, around one or two values.

We can obtain the figure  $2\log_2 n$  heuristically from the expected number of cliques of given size.

**Lemma 9.1.** *The expected number  $N_k$  of  $k$ -cliques in  $G(n, p)$  is*

$$N_k = \binom{n}{k} p^{\binom{k}{2}}.$$

The numbers  $N_k$  satisfy

$$\frac{N_{k+1}}{N_k} = \frac{n-k}{k+1} p^k.$$

Let  $k_0$  be the maximal  $k$  such that  $N_{k_0} \geq 1$ . Then

$$k_0 = 2\log_{1/p} n - 2\log_{1/p} \log_{1/p} n \pm O(1).$$

For  $k = k_0 \pm O(1)$ ,

$$\frac{N_{k+1}}{N_k} = \Theta\left(\frac{\log n}{n}\right).$$

*Proof.* The first claim is a simple calculation. The second claim follows from

$$\frac{N_{k+1}}{N_k} = \frac{n!}{(k+1)!(n-k-1)!} \cdot \frac{k!(n-k)!}{n!} \cdot \frac{p^{(k+1)k/2}}{p^{(k-1)k/2}} = \frac{n-k}{k+1} p^k.$$

For the third claim, note first that by Stirling's approximation,

$$N_k \lesssim \frac{n^k}{\sqrt{2\pi k} (k/e)^k} (p^{(k-1)/2})^k = \frac{1}{\sqrt{2\pi k}} \left( \frac{ep^{(k-1)/2} n}{k} \right)^k.$$

To obtain this approximation, we used the upper bound  $n^{\underline{k}} \leq n^k$ . When  $k \geq 3 \log_{1/p} n$ , say, this expectation is very small. Therefore  $k_0 = O(\log n)$ , and so the estimate above is a good lower bound as well, since in that case  $n^{\underline{k}} \sim n^k$ .

Let now  $k = 2 \log_{1/p} n - 2 \log_{1/p} \log_{1/p} n + 2\gamma + 1$ . Then

$$N_k = \frac{1}{\Theta(\sqrt{\log_{1/p} n})} \left( \frac{p^\gamma e \log_{1/p} n}{k} \right)^k = \frac{\Theta(p^\gamma)^{\Theta(\log_{1/p} n)}}{\Theta(\sqrt{\log_{1/p} n})}.$$

For large constant  $\gamma$ , the numerator is at most  $p^{\log_{1/p} n} = 1/n$  and so  $N_k = o(1)$ . Similarly, for large constant  $-\gamma$ , the numerator is at least  $p^{-\log_{1/p} n} = n$ , and so  $N_k = \omega(1)$ . Therefore  $k_0$  is obtained when  $|\gamma| = O(1)$ .

If  $k = k_0 \pm O(1)$  then

$$\frac{N_{k+1}}{N_k} = \frac{n-k}{k+1} p^k = (1 - o(1)) \cdot \frac{n}{2 \log_{1/p} n} \cdot \Theta\left(\frac{\log_{1/p}^2 n}{n^2}\right) = \Theta\left(\frac{\log_{1/p} n}{n}\right). \quad \square$$

The last part of the lemma already shows that with high probability,  $G(n, p)$  contains no clique of size  $k_0 + 2$ . On the other hand, it contains  $\omega(n/\log n)$  cliques of size  $k_0 - 1$ , in expectation. Using the second moment method we will show that with high probability,  $G(n, p)$  contains a clique of size  $k_0 - 1$ .

**Theorem 9.2** ([FK16, Theorem 7.3]). *Let  $p \in (0, 1)$  be a constant, and define  $k_0$  as in Lemma 9.1. With high probability, the clique number of  $G(n, p)$  is concentrated on two values.*

*Proof.* We first show that with high probability, the clique number is one of  $k_0 - 1, k_0, k_0 + 1$ . Then we slightly refine the argument to obtain the theorem.

The last part of Lemma 9.1 shows that  $N_{k_0+2} = O(\log n/n)$ , and so with high probability there are no cliques of size  $k_0 + 2$ . On the other hand,  $N_{k_0-1} = \Omega(n/\log n)$ . We proceed to estimate the second moment of the number of cliques of size  $k = k_0 \pm O(1)$ , which we denote by  $X_k$ . We will consider pairs of potential cliques of size  $k$  with intersection  $r$ . The probability that two such cliques appear together is  $p^{\binom{k}{2} - \binom{r}{2}}$ , and so

$$\mathbb{E}[X_k^2] = \sum_{r=0}^k \binom{n}{k-r, r, k-r} p^{\binom{k}{2} - \binom{r}{2}} = \mathbb{E}[X_k]^2 \sum_{r=0}^k \frac{\binom{n}{k-r, r, k-r}}{\binom{n}{k}^2} p^{-\binom{r}{2}}.$$

The ratio of binomial coefficients can be estimated as

$$\frac{\binom{n}{k-r, r, k-r}}{\binom{n}{k}^2} = \frac{n^{2k-r}}{(\underline{n^k})^2} \frac{k!^2}{(k-r)!^2 r!} \approx \frac{k!^2}{(k-r)!^2 r!} n^{-r}.$$

Therefore

$$\mathbb{E}[X_k^2] \approx \mathbb{E}[X_k]^2 \sum_{r=0}^k \frac{k!^2}{(k-r)!^2 r!} \frac{1}{n^r p^{\binom{r}{2}}} \approx \mathbb{E}[X_k]^2 \sum_{r=0}^k \frac{k!^2}{(k-r)!^2 r!^2} \frac{1}{\binom{n}{r} p^{\binom{r}{2}}} = \mathbb{E}[X_k]^2 \sum_{r=0}^k \frac{\binom{k}{r}^2}{N_r}.$$

Consider now the summands  $J_r = \binom{k}{r}^2 / N_r$ . We have  $J_0 = 1$ ,  $J_1 = O(\log^2 n/n)$ ,  $J_2 = O(\log^4 n/n^2)$ , and then the summands continue decreasing rapidly, until they increase back again. Eventually the sequence reaches  $J_k = 1/N_k$ . When  $k = k_0 - 1$  we have  $J_k = O(\log n/n)$ .

This picture can be seen by considering first the ratio

$$\frac{J_{r+1}}{J_r} = \frac{\binom{k}{r+1}^2 N_r}{\binom{k}{r}^2 N_{r+1}} = \frac{(k-r)^2 r+1}{(r+1)^2 n-r} p^{-r} = \frac{(k-r)^2}{(r+1)(n-r)} p^{-r}.$$

and then the ratio

$$\frac{J_{r+1}^2}{J_r J_{r+2}} = \frac{(k-r)^2}{(r+1)(n-r)} p^{-r} \frac{(r+2)(n-r-1)}{(k-r-1)^2} p^{r+1} = \frac{r+2}{r+1} \left( \frac{k-r}{k-r-1} \right)^2 \frac{n-r-1}{n-r} p,$$

which is less than 1 unless  $r$  is very close to 0 or  $k$ . Thus the ratio  $J_{r+1}/J_r$  is decreasing, which means that the sequence  $J_r$  is unimodal (decreases then increases) for  $r$  not very close to 0 or  $k$ . For  $r$  very close to 0 or  $k$ , this can be seen directly.

It follows that

$$\mathbb{E}[X_k^2] = \left( 1 + O\left(\frac{\log^2 n}{n}\right) \right) \mathbb{E}[X_k]^2.$$

The second moment method then shows that

$$\Pr[X_{k_0-1} = 0] \leq \frac{\mathbb{E}[X_k^2]}{\mathbb{E}[X_k]^2} - 1 = O\left(\frac{\log^2 n}{n}\right).$$

Up to now we have shown that the clique number is concentrated on the three values  $k_0 - 1, k_0, k_0 + 1$ . To improve on this, we have to consider the value of  $k_0$ . If  $k_0 \leq \sqrt{n/\log n}$  then with high probability there are no cliques of size  $k_0 + 1$  since  $N_{k_0+1} = O(\sqrt{\log n/n})$ . Otherwise, when  $k = k_0$  we have  $J_k = O(\sqrt{\log n/n})$  and so  $\mathbb{E}[X_k^2]/\mathbb{E}[X_k]^2 - 1 = O(\sqrt{\log n/n})$ , and so there is a clique of size  $k_0$  with high probability.  $\square$

## 9.2 Chromatic number

In the previous section we were interested in the size of the largest clique. If we switch  $p$  by  $1-p$  throughout, we obtain formulas for the independent set instead. Since a graph with chromatic number  $c$  must have an independent set of size  $n/c$ , namely, the largest color class, our work shows that with high probability, the chromatic number of  $G(n, p)$  is  $\Omega(n/\log n)$ . In the reverse direction, it is natural to try out the following strategy: repeatedly take out the largest independent set until  $O(n/\log n)$  vertices are left, and color the rest with fresh colors. We expect the largest independent sets throughout this process to have size  $\Omega(\log n)$ . However, the bounds that we got in Theorem 9.2 are not good enough to obtain this result: we need a bound for all potential graphs up to size  $n/\log n$ , whereas the error bound there is only  $\tilde{O}(1/n)$ . To rectify this, we will use a better concentration result, Janson's bound:

**Lemma 9.3** (Janson's inequality [FK16, Theorem 21.12]). *Let  $U$  be a universe and let  $\mathbf{R}$  be a random subset of  $U$  obtained by choosing each element  $x$  with some probability  $p_x \in (0, 1)$  independently. Let  $\{D_i\}$  be a collection of subsets of  $U$ . Let the random variable  $\mathbf{S}$  count the number of subsets  $D_i$  which  $\mathbf{R}$  contains, and let  $\mu = \mathbb{E}[\mathbf{S}]$ . Define*

$$\bar{\Delta} = \sum_{i,j: D_i \cap D_j \neq \emptyset} \Pr[D_i, D_j \subseteq \mathbf{R}].$$

Then for each  $0 \leq t \leq \mu$ ,

$$\Pr[\mathbf{S} \leq \mu - t] \leq e^{-t^2/2\bar{\Delta}}.$$

In particular,

$$\Pr[\mathbf{S} \leq 0] \leq e^{-\mu^2/2\bar{\Delta}}.$$

The lemma is proved using the exponential moment method, just like Chernoff's bound, though it employs the FKG inequality, and for this reason we skip it.

Let us compare this lemma to Chebyshev's inequality. Notice that

$$\mathbb{E}[\mathbf{S}^2] = \bar{\Delta} + \sum_{i,j: D_i \cap D_j = \emptyset} \Pr[D_i \subseteq \mathbf{R}] \Pr[D_j \subseteq \mathbf{R}] \leq \bar{\Delta} + \mathbb{E}[\mathbf{S}]^2.$$

In many cases this inequality is close to optimal. Chebyshev's inequality then gives

$$\Pr[\mathbf{S} \leq \mu - t] \leq \Pr[|\mathbf{S} - \mu| \geq t] \leq \frac{\overline{\Delta}}{t^2}.$$

Janson's inequality thus gives an exponential improvement.

We can now bound the chromatic number.

**Theorem 9.4** ([FK16, Theorem 7.7]). *Let  $p \in (0, 1)$  be a constant. With high probability, the chromatic number of  $G(n, p)$  satisfies*

$$\chi(G(n, p)) \sim \frac{n}{2 \log_{1/(1-p)} n}.$$

*Proof.* Theorem 9.2 shows that with high probability, the maximal independent set in  $G(n, p)$  has size  $\alpha(G(n, p)) \leq (1 - o(1))2 \log_{1/(1-p)} n$ . In that case

$$\chi(G(n, p)) \geq \frac{n}{\alpha(G(n, p))} = (1 + o(1)) \frac{n}{2 \log_{1/(1-p)} n}.$$

We are now going to replace Chebyshev's inequality with Janson's inequality in the proof of Theorem 9.2. The setup is as follows:

- $U$  consists of all non-edges in the empty graph on  $n$  vertices, and  $\mathbf{R}$  is chosen by putting each non-edge with probability  $1 - p$ .
- The  $D_i$  are all sets of size  $k$ .
- $\mathbf{S}$  is the number of independent sets of size  $k$ .

In the course of the proof of Theorem 9.2, we saw that for  $k = O(\log n)$  it holds that

$$\overline{\Delta} \approx \mu^2 \sum_{r=2}^k \frac{\binom{k}{r}^2}{N_r}.$$

We showed there that the largest two summands correspond to  $r = 2$  and  $r = k$ , and that all other values are smaller by a factor of at least  $n/\log^2 n$  from one of them. We also calculated that the  $r = 2$  summand is  $O(\log^4/n^2)$  and that the  $r = k_0 - 2$  summand is  $O(\log^5/n^2)$ . Thus, when  $k = k_0 - 2$ ,

$$\overline{\Delta} = O\left(\frac{\log^5 n}{n^2}\right) \mu^2.$$

Janson's inequality thus shows that

$$\Pr[X_{k_0-2} = 0] \leq e^{-O(n^2/\log^5 n)}.$$

In particular, with high probability, all graphs induced by at least  $n/\log^2 n$  vertices (of which there are fewer than  $2^n$ ) have independent sets of size  $(1 - o(1))2 \log_{1-p} n$ .

We now repeatedly remove independent sets of size  $(1 - o(1))2 \log_{1/(1-p)} n$  from the graph until  $n/\log^2 n$  or fewer vertices are left, and then color the rest with individual colors. In total, we have used  $(1 + o(1))n/2 \log_{1/(1-p)} n$  colors.  $\square$

With more effort, we could obtain bounds on the deviation of  $\chi(G(n, p))$  from its expectation. However, it is simpler to use a concentration bound directly. To this end, we will use Azuma's inequality, in the form of McDiarmid's inequality:

**Lemma 9.5** (McDiarmid's inequality, [FK16, Theorem 21.16]). *Let  $f$  be an  $n$ -variate function on  $U_1 \times \dots \times U_n$ , and suppose that there exist  $c_1, \dots, c_n$  such that  $|f(\vec{x}) - f(\vec{y})| \leq c_i$  whenever  $\vec{x}, \vec{y}$  differ only in the  $i$ th coordinate.*

*If  $X_1, \dots, X_n$  are independent random variables on  $U_1, \dots, U_n$  and  $Y = f(X_1, \dots, X_n)$  then for all  $t > 0$ ,*

$$\Pr[|Y - \mathbb{E}[Y]| \geq t] \leq 2 \exp - \frac{t^2}{2 \sum_{i=1}^n c_i^2}.$$

This inequality follows by applying Azuma's inequality, which is a Chernoff inequality for martingales, to the Doob martingale. Applications of this lemma to random graphs usually use either the *vertex exposure martingale* or the *edge exposure martingale*. In the latter, the  $x_i$  are the edges of the graph. In the former, we "expose" the graph vertex by vertex, and  $x_i$  is the list of edges from vertex  $i$  to vertices  $1, \dots, i-1$ . In this case we use the vertex exposure martingale to obtain the following result.

**Theorem 9.6** ([FK16, Theorem 7.8]). *Suppose that  $g(n) = \omega(\sqrt{n})$ . Then with high probability,*

$$|\chi(G(n, p)) - \mathbb{E}[\chi(G(n, p))]| < g(n).$$

*Proof.* Let  $f(x_1, x_2, \dots, x_n)$  be the function that accepts the list of all edges from vertex  $i$  to vertices  $1, \dots, i-1$  (so  $x_i \in \{0, 1\}^{i-1}$ ) and returns the chromatic number. Changing the edges adjacent to a single vertex can affect the chromatic number by at most 1 (since the chromatic number is at least the chromatic number of the rest of the graph, and at most that number plus one), and so McDiarmid's inequality shows that

$$\Pr[|\chi(G(n, p)) - \mathbb{E}[\chi(G(n, p))]| \geq g(n)] \leq 2e^{g(n)^2/2(n-1)} = o(1). \quad \square$$

Notice that this argument doesn't require knowledge of the expectation. This is a common feature of concentration of measure arguments. Moreover, the obtained concentration is better than what the argument of Theorem 9.4 gives. The theorem shows that with high probability,

$$\begin{aligned} \frac{n}{2 \log_{1/(1-p)} n - 2 \log_{1/(1-p)} \log_{1/(1-p)} n} &\leq \chi(G(n, p)) \leq \\ &\frac{n}{2 \log_{1/(1-p)}(n/\log^2 n) - 2 \log_{1/(1-p)} \log_{1/(1-p)}(n/\log^2 n)} + O\left(\frac{n}{\log^2 n}\right) \approx \\ &\frac{n}{2 \log_{1/(1-p)} n - 4 \log_{1/(1-p)} \log_{1/(1-p)} n} + O\left(\frac{n}{\log^2 n}\right) \approx \\ &\frac{n}{2 \log_{1/(1-p)} n - 2 \log_{1/(1-p)} \log_{1/(1-p)} n} + O\left(\frac{n \log \log n}{\log^2 n}\right). \end{aligned}$$

While the bounds can be slightly improved (since the size of the graph changes smoothly from  $n$  to  $n/\log^2 n$ ), this will only affect the hidden constant in the error term.

### 9.3 Finding cliques in random graphs

Theorem 9.2 shows that with high probability,  $G(n, p)$  contains a clique of size roughly  $2 \log_p n$ . Can we find this clique efficiently? No such algorithm is known. However, the trivial greedy algorithm finds a clique of size roughly  $\log_p n$  with high probability (the probability being with respect to the *graph* rather than the *algorithm*).

**Theorem 9.7** ([FK16, Theorem 7.9]). *Consider the following greedy algorithm: Start with the empty set, and repeatedly choose a vertex connected to all previously chosen vertices.*

*For fixed  $p \in (0, 1)$ , with high probability (over the choice of the graph), the greedy algorithm produces a clique of size  $(1 - o(1)) \log_{1/p} n$ .*

*Proof.* The algorithm terminates with a set  $T$  which is a maximal clique: no vertex can be added to it. The expected number of maximal cliques of size  $k$  is

$$E_k = \binom{n}{k} p^{\binom{k}{2}} (1 - p^k)^{n-k} \leq n^k (e/k)^k p^{k(k-1)/2} e^{-(n-k)p^k} \leq \left(p^{-1/2} n e^{1+p^k}\right)^k e^{-np^k}.$$

When  $k = \log_{1/p} n - C \log_{1/p} \log_{1/p} n$ , we get

$$E_k \leq (p^{-1/2} e^2 n)^{\log_{1/p} n} e^{-\log_{1/p}^C n}.$$

When  $C > 2$ , this is very small, and in fact  $\sum_{k \leq \log_{1/p} n - C \log_{1/p} \log_{1/p} n} E_k = o(1)$ . This shows that with high probability all maximal cliques have size at least  $\log_{1/p} n - C \log_{1/p} \log_{1/p} n$  (for any  $C > 2$ ).  $\square$

Intuitively, each vertex is connected to a  $p$ -fraction of the remaining vertices, and so it takes  $\log_{1/p} n$  steps to “kill” all vertices. The proof is some formal version of this argument, which also shows that the algorithm does terminate within  $\log_{1/p} n$  steps.

Repeated use of Theorem 9.7 in the case of independent sets yields an efficient coloring algorithm which uses  $(1 + o(1)) \frac{n}{\log_{1/1-p} n}$  colors, with high probability.

## 10 Planted clique

The maximum clique problem is one of the original NP-complete problems. It is known to be NP-hard to approximate to within  $n^{1-\epsilon}$  for every  $\epsilon > 0$ , and the best known algorithms only give an  $\tilde{O}(n/\log^3 n)$  approximation. As we have seen, in  $G(n, p)$  random graphs we can efficiently find cliques of size roughly  $\log_{1/p} n$ , although the maximum clique has size roughly  $2 \log_{1/p} n$ . Does the problem become easier if we “plant” a large clique?

The *planted clique* random graph  $G(n, p, k)$  is formed by taking a  $G(n, p)$  random graph and adding a clique on  $k$  random vertices. The central question in this area is:

For which values of  $k$  is there an efficient algorithm that finds a clique of size  $(1 - o(1))k$  in  $G(n, p, k)$  with high probability?

Importantly, the probability here is with respect to both  $G(n, p, k)$  and the algorithm (if it is randomized). In particular, we do not ask for the algorithm to succeed on every realization of  $G(n, p, k)$ .

We will only be interested in the case  $p = 1/2$ , the other cases being similar. We will see several different algorithms: a degree-based algorithm which works for  $k \approx \sqrt{n \log n}$ , and several algorithms which work for  $k \approx \sqrt{n}$ .

### 10.1 Degree-based algorithm

The first algorithm, due to Kučera [Kuč95], is based on the observation that the degree of the planted vertices is higher in expectation than the degree of the other vertices.

**Lemma 10.1.** *Let  $G \sim G(n, 1/2, k)$ . With high probability, the degree of every non-planted vertex is at most  $n/2 + \sqrt{2n \log n}$ , and the degree of every planted vertex is at least  $(n + k)/2 - \sqrt{2n \log k}$ .*

*Proof.* The degree of every non-planted vertex has distribution  $\text{Bin}(n - 1, 1/2)$ . Hoeffding’s inequality states that for such a vertex  $x$ ,

$$\Pr[|\deg(x) - (n - 1)/2| \geq t\sqrt{n - 1}] \leq 2e^{-t^2}.$$

If  $t = \sqrt{2 \log n}$  then this probability is at most  $2/n^2$ , and so with high probability, the degree of every non-planted vertex is at most  $n/2 + \sqrt{2n \log n}$ .

The degree of every planted vertex has distribution  $k - 1 + \text{Bin}(n - k, 1/2)$ . Hoeffding’s inequality implies, in the same way, that with high probability, the degree of each such vertex is at least  $(k - 1) + (n - k)/2 - \sqrt{1.9n \log k} \geq (n + k)/2 - \sqrt{2n \log k}$ .  $\square$

**Corollary 10.2.** *If  $k \geq \sqrt{8n \log n}$  then with high probability, the  $k$  vertices of  $G(n, 1/2, k)$  with largest degree are the planted vertices.*

*Proof.* The lemma shows that the degrees are separated given that  $k/2 - \sqrt{2n \log k} < \sqrt{2n \log n}$ . For this it suffices that  $k \leq 2\sqrt{2n \log n}$ .  $\square$

This implies a very quick algorithm that finds the hidden clique when  $k \geq \sqrt{8n \log n}$ .

We can obtain a polynomial time algorithm for  $k \geq c\sqrt{n \log n}$  for every  $c > 0$  using a trick from [AKS98]. The idea is to “guess”  $m = O(1)$  vertices from the clique. All vertices in the planted clique are neighbors of these  $m$  vertices, but there are only roughly  $n/2^m$  of these, while the induced clique has size  $k - m = k - O(1)$ .

**Theorem 10.3** (Kučera [Kuč95]). *For every  $c > 0$  there is a polynomial time algorithm that with high probability finds a clique of size  $k$  in  $G(n, 1/2, k)$  whenever  $k = c\sqrt{n \log n}$ .*

*Proof.* Let  $m$  be a parameter depending only on  $c$ , whose value will be decided later on. For every set of  $m$  vertices in  $G(n, 1/2)$ , the number of other vertices connected to all  $m$  vertices has binomial distribution  $\text{Bin}(n - m, 1/2^m)$ . Hoeffding's bound shows that with extremely high probability, there are at most  $1.9n/2^m$  such common neighbors, and a union bound shows that this holds for every set of  $m$  vertices with high probability. In the planted model there are  $k$  extra vertices, and so with high probability every set of  $m$  vertices has at most  $1.9n/2^m + k \leq 2n/2^m$  common neighbors. From now on we assume that this event happens.

Suppose that we somehow “guessed”  $m$  vertices of the planted clique. They have  $n' \leq 2n/2^m$  common neighbors by assumption. The graph induced by these common neighbors has distribution  $G(n', 1/2, k - m)$ . Since  $k = c\sqrt{n \log n}$ , in terms of  $n'$  we have  $k \geq c\sqrt{2^{m-1}n' \log n'}$ , and for large enough  $n$  we have  $k - m \geq c\sqrt{2^{m-2}n' \log n'}$ . We can choose  $m$  so that for large enough  $n$ ,  $k - m \geq \sqrt{8n' \log n'}$ , and so the algorithm of Corollary 10.2 will find the rest of the planted clique with high probability.

We can implement this idea as follows: We go over all  $m$ -cliques in the graph, compute the graph induced by the common neighbors of these vertices, and run the algorithm of Corollary 10.2 (with parameter  $k - m$ ). If at any point the algorithm of Corollary 10.2 succeeds, we output the corresponding clique of size  $m + (k - m) = k$ . When the  $m$ -clique is part of the planted clique, the algorithm of Corollary 10.2 will succeed (with high probability).  $\square$

It is not too hard to show that with high probability there is a unique  $k$ -clique in  $G(n, 1/2, k)$  (whenever  $k = \omega(\log n)$ ), and so Kučera's algorithm in fact returns the planted clique.

**Lemma 10.4.** *Suppose that  $k = \omega(\log n)$ . With high probability, there is a unique  $k$ -clique in  $G(n, 1/2, k)$ .*

*Proof.* Let  $T_\ell$  be the number of  $k$ -cliques whose intersection with the planted clique is  $\ell$ . We have

$$M_\ell := \mathbb{E}[T_\ell] = \binom{n-k}{k-\ell} \frac{1}{2^{\binom{k}{2} - \binom{\ell}{2}}}.$$

We have  $M_k = 1$  and

$$\frac{M_{\ell-1}}{M_\ell} = \frac{n-2k+\ell}{k-\ell+1} 2^{-(\ell-1)} \leq 2n2^{-\ell}.$$

This shows that as long as  $\ell \geq \log_2 n$  (say),  $M_\ell$  is very small. When  $\ell = O(\log n)$ , we can estimate directly (using  $\binom{k}{2} - \binom{\ell}{2} \geq \frac{1}{2}\binom{k}{2}$ )

$$M_\ell \leq \frac{n^k}{2^{k^2/4}} = \left(\frac{n}{2^{k/4}}\right)^k,$$

which is also very small. In total, we conclude that  $\sum_{\ell < k} M_\ell = o(1)$ , and so with high probability the unique  $k$ -clique is the planted one.  $\square$

## 10.2 More on the maximal degree of a graph\*

Kučera's algorithm only works for  $k = \Omega(\sqrt{n \log n})$ , and this is because we need the maximal degree in  $G(n, 1/2)$  to exceed its expectation by at most roughly  $k/2$ . Using Hoeffding's inequality, we showed that the maximal degree is at most roughly  $n/2 + \sqrt{n \log n}$ , with high probability. Is this tight? Using a second-moment calculation we can show that the maximal degree is indeed  $n/2 + \Theta(\sqrt{n \log n})$ .

**Theorem 10.5** ([FK16, Theorem 3.5]). *For every  $\epsilon > 0$ , with high probability the maximal degree  $\Delta$  in  $G(n, p)$  satisfies*

$$|\Delta(G(n, p)) - [(n-1)p + \sqrt{2(n-1)p(1-p) \log n}]| \leq \epsilon \sqrt{2(n-1)p(1-p) \log n}.$$

Morally speaking, the reason that this theorem holds is that the individual degrees have distribution  $\text{Bin}(n-1, p)$  which is very close to the normal distribution  $N(\mu, \sigma^2)$  with  $\mu = (n-1)p$  and  $\sigma = (n-1)p(1-p)$ . Thus roughly speaking, each degree has distribution  $\mu + \sigma \cdot N(0, 1) = \mu + \sigma F^{-1}(U(0, 1))$ , where  $F$  is the CDF of  $N(0, 1)$ , and  $U(0, 1)$  is distribution uniformly on  $[0, 1]$ .



If the degrees were completely independent, then the maximal degree would have distribution  $\mu + \sigma F^{-1}(X_n)$ , where  $X_n$  is the maximum of  $n$  copies of  $U(0, 1)$ . The expected value of this maximum is  $n/(n+1) \approx 1 - 1/n$ , although this maximum is not too concentrated. Now for large  $x$ , it is known that

$$1 - F(x) \approx \frac{e^{-x^2/2}}{\sqrt{2\pi x}}.$$

This implies that  $1 - F^{-1}(\sqrt{2\log n}) \approx 1/n$  (up to logarithmic factors), and so we expect  $F^{-1}(X_n)$  to be close to  $\sqrt{2\log n}$ . This implies precisely the formula in the theorem, though the formal proof is quite a bit different.

*Proof sketch.* Let  $X \sim \text{Bin}(n-1, p)$  be the distribution of the degree of a particular vertex. The local limit theorem, a version of the central limit theorem for discrete random variables, states that  $\Pr[X = d]$  is roughly equal to the density of the Gaussian approximation to  $\text{Bin}(n-1, p)$  at the point  $d$ . In our case, we can prove such a result using Stirling's approximation: if  $x \leq n^{1/3} \log^2 n$  then

$$\Pr[X = (n-1)p + x\sqrt{(n-1)p(1-p)}] = (1 \pm o(1)) \frac{1}{\sqrt{2\pi np(1-p)}} e^{-x^2/2}.$$

This is worked out diligently in [FK16, Lemma 3.6].

We will be interested in three specific values of  $d$ :

- $d_+ = (n-1)p + (1+\epsilon)\sqrt{2\log n} \cdot \sqrt{(n-1)p(1-p)}$  is the upper bound in the theorem.
- $d_- = (n-1)p + (1-\epsilon)\sqrt{2\log n} \cdot \sqrt{(n-1)p(1-p)}$  is the lower bound in the theorem.
- $d_L = (n-1)p + \log^2 n \cdot \sqrt{(n-1)p(1-p)}$  is a degree which is small enough for the local limit estimate to apply, and large enough so that in calculation we can replace it by  $\infty$ .

The binomial distribution is unimodal with mode near  $(n-1)p$ , and in particular  $\Pr[X = d]$  decreases beyond  $d_L$ , as can also be seen by direct calculation. Since  $\Pr[X = d_L]$  is readily seen to be very small, a simple union bound shows that  $\Delta(G(n, p)) \leq d_L$  with high probability.

Let  $A_d$  be the expected number of vertices whose degrees are between  $d$  and  $d_L$ , where  $d \geq d_-$ . Then

$$\mathbb{E}[A_d] = (1 + o(1)) \sqrt{\frac{n}{2\pi p(1-p)}} \sum_{\delta=d}^{d_L} \exp -\frac{1}{2} \left( \frac{\delta - (n-1)p}{\sqrt{(n-1)p(1-p)}} \right)^2.$$

The value of  $d_L$  is so large that extending the range of the summation to infinity doesn't affect the sum asymptotically. Estimating the sum by an integral and computing the integral, we finally obtain

$$\mathbb{E}[A_d] = (1 + o(1)) \frac{n}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{x},$$

where  $d = (n-1)p + x\sqrt{(n-1)p(1-p)}$ . When  $d = d_+$ , this gives  $\mathbb{E}[A_{d_+}] = O(n^{1-(1+\epsilon)^2}) = O(n^{-2\epsilon-\epsilon^2})$ , and so with high probability  $\Delta(G(n, p)) \leq d_+$ . Conversely, when  $d = d_-$ , we get  $\mathbb{E}[A_{d_-}] = \Omega(n^{1-(1-\epsilon)^2}) = \Omega(n^{2\epsilon-\epsilon^2}) = \Omega(n^\epsilon)$ .

In order to complete the proof, we use the second moment method. We have

$$\begin{aligned} \mathbb{E}[A_d^2] &= \sum_{d_1, d_2=d}^{d_L} \sum_{x_1, x_2} \Pr[\deg(x_1) = d_1 \text{ and } \deg(x_2) = d_2] \\ &= \mathbb{E}[A_d] + n(n-1) \sum_{d_1, d_2=d}^{d_L} \Pr[\deg(x_1) = d_1 \text{ and } \deg(x_2) = d_2]. \end{aligned}$$

Considering the two cases corresponding to whether the edge  $(x_1, x_2)$  is in the graph or not, we get

$$\begin{aligned} \mathbb{E}[A_d^2] &= \mathbb{E}[A_d] + n(n-1) \sum_{d_1, d_2=d}^{d_L} ((1-p) \Pr[\text{Bin}(n-2, p) = d_1] \Pr[\text{Bin}(n-2, p) = d_2] \\ &\quad + p \Pr[\text{Bin}(n-2, p) = d_1 - 1] \Pr[\text{Bin}(n-2, p) = d_2 - 1]). \end{aligned}$$

A short calculation shows that the summand can be estimated as

$$(1 + o(1)) \Pr[\text{Bin}(n-1, p) = d_1] \Pr[\text{Bin}(n-1, p) = d_2],$$

and so

$$\mathbb{E}[A_d^2] = \mathbb{E}[A_d] + (1 + o(1)) \mathbb{E}[A_d]^2.$$

Chebyshev's inequality thus shows that

$$\Pr[A_d = 0] \leq \frac{\mathbb{E}[A_d^2]}{\mathbb{E}[A_d]^2} - 1 = \frac{1}{\mathbb{E}[A_d]} + o(1).$$

When  $d = d_-$ , we obtain  $\Pr[A_d = 0] = o(1)$ . □

### 10.3 Spectral norm of random graph

Consider a  $G(n, 1/2)$  random graph. Each individual degree has distribution roughly  $N(n/2, n/4) = n/2 + \sqrt{n/2}N(0, 1)$ , but the maximal degree deviates  $\sqrt{2 \log n}$  standard deviations from the mean, and so is roughly  $n/2 + \sqrt{n \log n/2}$ . Surprisingly, when we consider the spectral norm of the adjacency matrix of the random graph, the noise goes down from order  $\sqrt{n \log n}$  to order  $\sqrt{n}$ , as shown by Füredi and Komlós [FK81].

Let  $A$  be the adjacency matrix of a  $G(n, 1/2)$  random graph. We can write  $A = (B - I + J)/2$ , where  $B$  is a symmetric matrix with zeroes on the diagonal in which every pair of off-diagonal entries is a uniform  $\pm 1$  random variable, and  $J$  is the all-ones matrix. We can analyze the eigenvalues of  $B$  using the following formula:

$$\sum_{i=1}^n \lambda_i(B)^{2k} = \text{Tr}(B^{2k}) = \sum_{i_0, \dots, i_{2k-1}} B_{i_0 i_1} \cdots B_{i_{2k-1} i_0}.$$

Let us now estimate the expectation of the right-hand side, assuming that  $k = o(n)$ . We can divide the sum into several different sums, according to the number of distinct indices appearing on the right-hand side:

$$S_m := \sum_{\substack{i_0, \dots, i_{2k-1} \\ |\{i_0, \dots, i_{2k-1}\}| = m}} \mathbb{E}[B_{i_0 i_1} \cdots B_{i_{2k-1} i_0}].$$

Since all moments of  $B_{ii}, B_{ij}$  are zero or one, the expectation of each term is either zero or one. Since  $\mathbb{E}[B_{ii}] = 0$  and  $\mathbb{E}[B_{ij}] = 0$  whenever  $i \neq j$ , each appearance of  $B_{ij}$  must be matched with another appearance of  $B_{ij}$  or  $B_{ji}$ . We can associate each term with a closed walk of length  $2k$  which goes through  $m$  vertices. The underlying undirected graph must contain a spanning tree consisting of  $m-1$  edges, each of which is traced at least twice, for a total of at least  $2(m-1)$  edges. Thus  $m \leq k+1$ .

Let us consider first the case  $m = k+1$ . Suppose we have chosen the vertices in the order in which they appear, in  $n^m$  ways. The walk itself encodes a *rooted tree*: a tree rooted at  $v_0$  in which the children of each node are ordered. Indeed, the walk corresponds to a breadth-first search of the tree using the given order. It is well-known that the number of ordered trees on  $k+1$  vertices is given by the Catalan number  $C_k = \binom{2k}{k}/(k+1) \sim 4^k/(\sqrt{\pi}k^{2/3})$  which satisfies  $C_k \leq 4^k$ , and so  $S_{k+1} = C_k n^{k+1} \leq (4n)^{k+1}$ .

Füredi and Komlós use an elaborate encoding to show that as long as  $k = o(n^{1/6})$ ,  $\text{Tr}(B^{2k}) = (1 + o(1))S_{k+1}$ . In total, denoting  $\lambda_{\max}(B) := \max_i |\lambda_i(B)|$ , we have

$$\mathbb{E}[\lambda_{\max}(B)^{2k}] \leq \mathbb{E} \left[ \sum_{i=1}^n \lambda_i(B)^{2k} \right] \leq (1 + o(1))(4n)^{k+1}.$$

For small  $\epsilon > 0$  we have

$$\Pr[\lambda_{\max}(B) \geq (1 + \epsilon)2\sqrt{n}] = \Pr[\lambda_{\max}(B)^{2k} \geq (1 + \epsilon)^{2k}(4n)^k] \leq (1 + \epsilon)^{-2k}(4n) \leq e^{-\Omega(\epsilon k)}(4n).$$

As a consequence, if  $\epsilon = \omega(\log n/k)$  then with high probability,  $\lambda_{\max}(B) \leq (1 + \epsilon)2\sqrt{n}$ . In particular, we get such a guarantee with high probability whenever  $\epsilon = \omega(\log n/n^{1/6})$ .

The same technique of moment calculation shows that the distribution of a random eigenvalue of  $B$  divided by  $\sqrt{n}$  converges (in an appropriate sense) to the semicircle distribution on  $[-2, 2]$ , which is the distribution whose density is given by the semicircle  $\sqrt{4-x^2}/2\pi$ .

For the next step, we need to use the inequality  $\lambda_2(M+L) \leq \lambda_1(M) + \lambda_2(L)$ , where  $\lambda_k(M)$  is the  $k$ th largest eigenvalue of  $M$ . This inequality follows from the variational formula for  $\lambda_k(M)$ , which is valid for Hermitian matrices:<sup>3</sup>

$$\lambda_k(M) = \min_{U: \dim U=k} \max_{\substack{x \perp U \\ \|x\|=1}} x^T M x = \max_{U: \dim U=n-k+1} \min_{\substack{x \perp U \\ \|x\|=1}} x^T M x.$$

Indeed, if  $M, L$  are Hermitian then this formula shows that

$$\lambda_k(M+L) = \min_{U: \dim U=k} \max_{\|x\|=1} x^T (M+L)x \leq \max_{\|x\|=1} x^T M x + \min_{U: \dim U=k} \max_{\|x\|=1} x^T L x = \lambda_1(M) + \lambda_k(L).$$

An analogous argument shows that  $\lambda_k(M+L) \geq \lambda_k(M) + \lambda_n(L)$ .

With high probability, all eigenvalues of  $B-I$  differ from 1 by at most  $(1+\epsilon)2\sqrt{n}$ . Since  $\lambda_2(J) = \lambda_n(J) = 0$ , the inequality above shows that  $\lambda_2(A) \leq -\frac{1}{2} + (1+\epsilon)\sqrt{n}$  and  $\lambda_n(A) \geq -\frac{1}{2} - (1-\epsilon)\sqrt{n}$  with high probability. (In fact, a glance at the proof shows that the same argument would work even if we analyzed  $B-I$  directly, thus allowing us to improve the bound on the second eigenvalue to  $(1+\epsilon)\sqrt{n}$ .)

What about the first eigenvalue? On the one hand,  $\lambda_1(\frac{(B-I)+J}{2}) \leq \frac{1}{2}\lambda_1(B-I) + \frac{1}{2}\lambda_1(J) \leq \frac{n-1}{2} + O(\sqrt{n})$ , with high probability. On the other hand, if  $j$  is the all-ones vector then  $j^T A j = 2 \sum_{i < j} A_{ij}$ , which has distribution roughly  $N(n(n-1)/2, n(n-1)/2)$ . Therefore  $\lambda_1(A) \geq j^T A j / j^T j \gtrsim \frac{n-1}{2} - \tilde{O}(\sqrt{n})$  with high probability. Füredi and Komlós showed that the distribution of  $\lambda_1(A)$  is  $1/\sqrt{n}$ -close (in CDF distance) to  $N(n/2, 1/2)$ .

## 10.4 Spectral algorithm\*

### 10.4.1 Idea

We can summarize the idea behind Kučera's algorithm using the following points:

- Degrees in a  $G(n, 1/2)$  random graph are in the range  $n/2 \pm \sqrt{n \log n/2}$ .
- Adding a  $k$ -clique boosts the degree of all vertices in the clique by roughly  $k/2$ , so now they are  $n/2 + k/2 \pm \sqrt{n \log n/2}$ .
- If  $k/2 \geq 2\sqrt{n \log n/2}$  then the  $k$  vertices of maximal degree are the planted clique.

The size of the clique that the algorithm is able to handle thus stems from fluctuations in the degree that can be as large as  $\sqrt{n \log n}$ .

Alon, Krivelevich and Sudan [AKS98] developed a different algorithm which is based (in some sense) on Wigner's *semicircle law*. Take any symmetric matrix whose entries are chosen iid from a "reasonable" distribution with zero mean and unit variance. The law states that if the matrix is  $n \times n$ , then the number of eigenvalues in the range  $[\alpha\sqrt{n}, \beta\sqrt{n}]$  is

$$\int_{\alpha}^{\beta} \frac{\sqrt{4-x^2}}{2\pi} dx \cdot n + o(n).$$

The density function  $\sqrt{4-x^2}/2\pi$  is a semicircle supported on the interval  $[-2, 2]$ . In particular, the number of eigenvalues outside the range  $[-2\sqrt{n}, 2\sqrt{n}]$  is  $o(n)$ . Füredi and Komlós [FK81] showed that

<sup>3</sup>A complex matrix  $A$  is *Hermitian* if  $A = A^*$ , where  $(A^*)_{ij} = \overline{A_{ji}}$ ; in other words  $A^*$  is obtained from the transpose  $A^T$  by complex-conjugating all entries. A real matrix is Hermitian iff it is symmetric. The adjacency matrix of an undirected graph is symmetric.

Let us now show how to prove that  $\lambda_k = \min_{\dim U=k} \max_{0 \neq x \in U} x^T M x / \|x\|^2$ . Since the matrix is Hermitian, there is an orthonormal basis of eigenvectors  $v_1, \dots, v_n$  corresponding to the eigenvalues  $\lambda_1, \dots, \lambda_n$ . In one direction, we take  $U = \text{span}(v_1, \dots, v_k)$ . Every vector  $x \in U$  decomposes as  $x = \sum_{i=1}^k c_i v_i$ , where  $\sum_{i=1}^k c_i^2 = \|x\|^2$  by orthonormality. Therefore  $x^T M x = \sum_{i=1}^k \lambda_i c_i^2 \leq \lambda_k \|x\|^2$ . In the other direction, let  $U$  be a  $k$ -dimensional subspace, and let  $V = \text{span}(v_k, \dots, v_n)$ . Since  $\dim U + \dim V > n$ , the two subspaces must intersect non-trivially, and so there exists some non-zero vector  $0 \neq x \in U \cap V$ . Since  $x \in V$ , we can write  $x = \sum_{i=k}^n c_i v_i$ , and so  $x^T M x = \sum_{i=k}^n \lambda_i c_i^2 \geq \lambda_k \|x\|^2$ .

with high probability, *all* eigenvalues lie in a range only slightly larger than  $[-2\sqrt{n}, 2\sqrt{n}]$ , and furthermore this holds even if the diagonal entries are constant.

A random graph almost conforms to this setting. Apart from the diagonal elements, the adjacency matrix of a graph can be formed by drawing a symmetric matrix  $B$  whose entries are uniform signs ( $\pm 1$ ), and taking  $A := (B + J)/2$ ,  $J$  being the all-ones matrix. The matrix  $J$  has the eigenvector  $\vec{1}$  with eigenvalue  $n$ . All vectors orthogonal to  $\vec{1}$  belong to the eigenspace of 0.

The spectral norm of  $B$  is roughly  $2\sqrt{n}$  with high probability (since the random perturbations have magnitude  $o(\sqrt{n})$ ). Adding  $B$  to  $J$  has the effect of slightly perturbing the main eigenvector  $\vec{1}$  (and its eigenvalue); the spectral norm of the rest is roughly  $2\sqrt{n}$ . Thus  $A$  has a main eigenvector, almost constant, corresponding to an eigenvalue close to  $n/2$ , and all other eigenvalues are at most roughly  $\sqrt{n}$ .

Let now  $K$  be a random  $k$ -clique (from now on we identify graphs with their adjacency matrix). Planting the clique  $K$  in the random graph  $A$  corresponds to taking the pointwise maximum  $A' = A \vee K$ . If we remove the clique edges which are already in  $A$  we get a new graph  $C$  such that  $A' = A + C$ . The graph  $C$  behaves like  $G(k, 1/2)$  on the clique vertices, and its main eigenvector is roughly constant on the clique (and zero everywhere else) and corresponds to an eigenvalue close to  $k/2$ . Denote the main eigenvector of  $A, C$  by  $v_A, v_C$ , normalized so that its entries are close to 1. Roughly speaking,

$$(A + C)v_A \approx (n/2)v_A, \quad (A + C)v_C \approx (k/2)v_A + (k/2)v_C.$$

Thus  $v_A$  is an approximate eigenvector of  $A + C$ , but  $v_C$  isn't. Solving the system of equations (or running one step of Gram–Schmidt), we get that

$$(A + C)((n - k)v_C - kv_A) \approx (k/2)((n - k)v_C - kv_A),$$

and so we expect  $A + C$  to have an eigenvector close to  $(n - k)1_K - k1_{\bar{K}}$  whose eigenvalue is close to  $k/2$ . All other eigenvalues are still roughly  $2\sqrt{n}$ . Summarizing, with high probability:

- $G(n, 1/2)$  has one strong eigenvalue, roughly  $n/2$ , and all other eigenvalues are  $O(\sqrt{n})$ .
- Adding a  $k$ -clique adds another strong eigenvalue, roughly  $k/2$ , which roughly encodes the vertices in the clique.

When  $k \geq C\sqrt{n}$  for some large enough constant  $C$ , the eigenvector corresponding to the second largest eigenvalue will be the one encoding the clique. It is tempting to partition the coordinates of this eigenvector according to their sign or magnitude (since  $n - k \gg k$ ), and then read off the planted clique. However, we are only promised that the eigenvector is *close* to  $(n - k)1_K - k1_{\bar{K}}$ . This is enough to conclude that a large fraction of the  $k$  vertices of largest magnitude in the eigenvector (forming the set  $S$ ) indeed belong to the planted clique. We can now identify the planted clique by taking all vertices which are connected to a significant fraction of vertices in  $S$ .

### 10.4.2 Proof sketch

We now repeat the exposition a bit more formally.

**Lemma 10.6.** *Let  $k = C\sqrt{n}$  for a large enough constant  $C$ . With high probability, the spectrum  $\lambda_1 \geq \dots \geq \lambda_n$  of  $G(n, 1/2, k)$  satisfies  $\lambda_1 \approx n/2$ ,  $\lambda_2 = k/2 \pm O(\sqrt{n})$ , and  $\lambda_3 \lesssim \sqrt{n}$  (the approximations hide  $1 \pm o(1)$  factors).*

*Furthermore, the eigenvector corresponding to  $\lambda_2$  is close to  $z = (n - k)1_K - k1_{\bar{K}}$ , where  $K$  is the hidden clique, in the sense that  $z - \delta$  is an eigenvector for some vector  $\delta$  satisfying  $\|\delta\|^2 \leq \|z\|^2/60$ .*

*Proof sketch.* Let  $M$  be the adjacency matrix of the graph. It is well-known that  $\lambda_1 = \max_x \frac{x^T M x}{\|x\|^2}$ . Choosing  $x = \vec{1}$ , this shows that  $\lambda_1 \geq 2|E|/n$  is at least the average degree in  $G(n, 1/2, k)$ , which with high probability is roughly  $n/2$ .

We can write  $M = M_1 + M_2$ , where  $M_1$  is the adjacency matrix of  $G(n, 1/2)$ , and  $M_2$  has the marginal distribution of  $G(k, 1/2)$  on a random subset of  $k$  vertices. Füredi and Komlós [FK81] showed that with high probability,

$$\max_{i \geq 2} |\lambda_i(M_1)| \leq \sqrt{n} + O(n^{1/3} \log n).$$

A similar result holds for the eigenvalue of  $M_2$  (with  $n$  replaced by  $k$ ). This implies that all vectors orthogonal to  $v_1(M_1), v_1(M_2)$  (the eigenvectors corresponding to  $\lambda_1(M_1), \lambda_1(M_2)$ ) satisfy  $\frac{x^T M_1 x}{\|x\|^2} \lesssim \sqrt{n}$  and  $\frac{x^T M_2 x}{\|x\|^2} \lesssim \sqrt{k}$ , and so  $\frac{x^T M x}{\|x\|^2} \lesssim \sqrt{n}$ . There is thus a subspace of codimension 2, restricted to which the spectral norm of  $M$  is at most roughly  $\sqrt{n}$ . This implies that  $\lambda_3 \lesssim \sqrt{n}$ .

It remains to estimate  $\lambda_2$  and its corresponding eigenvector. We start by analyzing  $t = (A - (k/2)I)z$ . The distribution of the individual entries of  $t$  is:

- If  $i$  is in the planted clique:  $(k-1)(n-k) - k \text{Bin}(n-k, 1/2) - k/2(n-k) = (k/2-1)(n-k) - k \text{Bin}(n-k, 1/2)$ .
- If  $i$  is not in the planted clique:  $(n-k) \text{Bin}(k, 1/2) - k \text{Bin}(n-k-1, 1/2) + k^2/2$ .

Use  $X_i \sim \text{Bin}(n-k, 1/2)$ ,  $Y_i \sim \text{Bin}(k, 1/2)$  and  $Z_i \sim \text{Bin}(n-k-1, 1/2)$  to stand for the implied random variables. Thus for vertices in the clique,  $t_i = (k/2-1)(n-k) - kX_i$ , and for vertices not in the clique,  $t_i = (n-k)Y_i - kZ_i + k^2/2$ . Since  $k^2/2 = -(n-k)(k/2) + k(n/2)$ , we can rewrite that as  $t_i = (n-k)(Y_i - k/2) - k(Z_i - n/2)$ . Therefore

$$\|t\|^2 = \sum_{i \in K} ((k/2-1)(n-k) - kX_i)^2 + \sum_{i \notin K} (n-k)^2 (Y_i - k/2)^2 + \sum_{i \notin K} k^2 (Z_i - n/2)^2 - \sum_{i \notin K} 2k(n-k)(Y_i - k/2)(Z_i - n/2).$$

Roughly speaking,  $(k/2-1)(n-k) - kX_i \sim N(-(n-k), k^2/4)$ , and so with high probability all terms in the first sum are at most  $O(k^2 \log k)$ , and in total amount to  $O(k^3 \log k)$ . Similarly, roughly  $Z_i - n/2 \sim N(-(k-1)/2, (n-k-1)/4)$ , and so all terms in the third sum are at most  $O(k^2 n \log n)$  with high probability, and in total amount to  $O(k^2 n^2 \log n)$ . In the same vein, roughly  $Y_i - k/2 \sim N(0, k/4)$ , and so all terms in the fourth sum are at most  $O(kn \cdot \sqrt{kn} \log n)$  with high probability, and in total amount to  $O(k^{1.5} n^{2.5} \log n)$ .

We could bound each term in the second sum by  $O(n^2 k \log n)$  with high probability, for a total of  $O(n^3 k \log n)$ , but we would like to improve on this bound. Notice that the variables  $Y_i$  count different edges, and so are independent. The summands  $(n-k)^2 (Y_i - k/2)^2$  are independent and have expectation  $(n-k)^2 k/4$ , and in total  $(n-k)^3 k/4$ . The variance of each of the summands is of order  $O(n^4 k^2)$ , for a total variance of  $O(n^5 k^2)$ . Chebyshev's inequality thus shows that the deviation of the second sum from its expectation is at most, say  $O(n^{2.5} k \log n)$  with high probability, and so the sum itself is at most  $O(n^3 k)$  with high probability.

Concluding, we have shown that  $\|z\|^2 = O(n^3 k)$  with high probability. Write now  $z = w + \delta$ , where  $w$  is in the eigenspace of  $\lambda_2$  and  $\delta$  is orthogonal to that eigenspace. We have

$$\|t\|^2 \gtrsim \left( \sqrt{n} - \frac{k}{2} \right)^2 \|\delta\|^2 = (C/2 - 1)^2 n \|\delta\|^2,$$

since  $\lambda_1(A - (k/2)I) \approx n/2 - k/2$  and  $\lambda_3(A - (k/2)I) \lesssim \sqrt{n} - k/2$ . We expect the component in the direction of the large eigenvalue to be small, and so it is reasonable to approximate  $(n/2 - k/2)^2$  by  $(\sqrt{n} - k/2)^2$  in this context.

Comparing our two estimates on  $\|t\|^2$ , we see that  $C^2 n \|\delta\|^2 = O(n^3 k)$ , and so  $\|\delta\|^2 = O(n^2 k)/C^2$ . Now  $\|z\|^2 = k(n-k)^2 + (n-k)k^2 = k(n-k)n \approx n^2 k$ , and so for an appropriate choice of  $C$ ,  $\|\delta\|^2 \leq \|z\|^2/60$ . Moreover,

$$O(n^3 k) \geq \|(A - (k/2)I)z\|^2 \geq \|(A - (k/2)I)w\|^2 \geq (\lambda_2 - k/2)^2 \|w\|^2 \geq (\lambda_2 - k/2)^2 \Omega(n^2 k),$$

using  $\|w\|^2 = \|z\|^2 - \|\delta\|^2 \geq (59/60)k(n-k)n$ . It follows that  $(\lambda_2 - k/2)^2 = O(n)$ , and so  $\lambda_2 \geq k/2 - O(\sqrt{n})$ .  $\square$

The next step is showing that the noise  $\delta$  still allows us to decode the planted clique from the second eigenvalue  $v_2$ , in two steps.

**Lemma 10.7.** *Let  $k = C\sqrt{n}$  for a large enough constant  $C$ . Let  $v_2$  be an eigenvector corresponding to the second eigenvalue of the adjacency matrix of  $G(n, 1/2, k)$ , and let  $S$  be its top  $k$  elements (in magnitude). With high probability, at least a 5/6 fraction of  $S$  belongs to the planted clique.*

*Proof.* Recall that  $v_2 = z - \delta$ , where  $\|\delta\|^2 \leq n^2 k/60$ , with high probability. In particular, at most  $k/6$  of the coordinates of  $\delta$  are at least  $n/3$  in magnitude (since  $k/6(n/3)^2 > n^2 k/60$ ). Thus, apart from at most  $k/6$  coordinates, all the clique coordinates have value at least  $(2/3)n - k$ , and all the non-clique coordinates have value at most  $n/3 - k$ , which is smaller.  $\square$

**Lemma 10.8.** *Let  $k = C\sqrt{n}$  for a large enough constant  $C$ . Let  $v_2$  be an eigenvector corresponding to the second eigenvalue of the adjacency matrix of  $G(n, 1/2, k)$ , and let  $S$  be its top  $k$  elements (in magnitude). Let  $T$  consist of all vertices neighboring at least a  $3/4$  fraction of  $S$ . With high probability,  $T$  is the planted clique.*

*Proof.* We have seen that with high probability,  $S$  contains at least  $(5/6)k$  vertices from the clique. Hence a clique vertex has at least  $(5/6)k$  neighbors in  $S$ . A non-clique vertex is adjacent to roughly half of the vertices in the clique, and so to only at most roughly  $k/2 + k/6 = (2/3)k$  vertices in  $S$  (we use the trivial bound for the at most  $k/6$  vertices not in the clique).  $\square$

The last lemma can be implemented efficiently. This completes our description of the spectral algorithm for planted clique. Using the technique of Theorem 10.3 (which actually originates in Alon et al.), we can replace  $C$  with an arbitrary constant  $c > 0$  at the cost of increasing the running time by a polynomial factor.

## 10.5 SDP-based algorithm

Another algorithm, due to Feige and Krauthgamer [FK00], uses the Lovász theta function, which we can define as follows:  $\theta(G)$  is the minimum  $\lambda_1(M)$  for a real symmetric matrix  $M$  indexed by vertices of the graph such that  $M_{ij} = 1$  whenever  $i, j$  are adjacent or identical. If  $v$  is the characteristic function of a  $k$ -clique in the graph then  $v^T M v = k^2 = k\|v\|^2$ , and so  $\lambda_1(M) \geq k$ . This shows that  $\theta(G) \geq \omega(G)$ . Surprisingly, with high probability there is a matching upper bound for  $G(n, 1/2, k)$  for  $k = \Omega(\sqrt{n})$ . In other words, for graphs with a large planted clique, we can compute the clique number using the Lovász theta function! Since the Lovász theta function can be computed using semidefinite programming, this provides a simple algorithm for recovering the clique, by repeatedly removing vertices (recall that by Lemma 10.4, there is a unique  $k$ -clique in  $G(n, 1/2, k)$  whenever  $k = \omega(\log n)$ ).

**Theorem 10.9.** *Let  $k = C\sqrt{n}$ , for a large enough  $C > 0$ . With very high probability  $1 - o(1/n)$ ,  $\theta(G(n, 1/2, k)) = k$ .*

*Proof.* We have seen above that  $\theta(G(n, 1/2, k)) \geq k$ , and it remains to prove the upper bound. To this end, denote by  $K$  the planted clique, and consider the following matrix  $M$ . If  $i = j$  or  $(i, j)$  is an edge, then we put  $M_{i,j} = 1$ . If  $(i, j)$  is not an edge and  $i, j$  both don't belong to the clique, we put  $M_{i,j} = -1$ . If  $(i, j)$  is not an edge,  $i$  is in the clique, and  $j$  isn't, then we put  $M_{i,j} = -1 + x_j$ , where  $x_j$  is chosen so that  $\sum_{i \in K} M_{i,j} = 0$ . When  $j$  is in the clique and  $i$  isn't, we take  $M_{i,j} = M_{j,i}$ .

Stated differently, we start with which contains  $+1$  for each edge,  $-1$  for each non-edge, and  $+1$  on the diagonal. We can think of this matrix schematically as follows:

$$\left( \begin{array}{c|c} J_{k \times k} & B_{k \times n-k} \\ \hline B_{n-k \times k}^T & C_{n-k \times n-k} \end{array} \right)$$

Here the first  $k$  rows and columns correspond to the variables of the planted clique. We add a constant  $x_j$  to all  $-1$  entries of row  $j$  of  $B^T$  to make the row sum to zero, and we update  $B$  accordingly.

By construction,  $M$  is a symmetric matrix which satisfies the constraints of the theta function. Moreover, by construction  $M1_K = k1_K$ . To complete the proof, we will show that all other eigenvalues of  $M$  are at most  $k$ .

We can write  $M = U + V + W$ , where  $U$  is a random symmetric sign matrix with 1's on the diagonal,  $V$  is a  $\{0, 2\}$ -valued matrix that corrects the clique edges to 1, and  $W$  correspond to the corrections  $x_j$ . Füredi and Komlós [FK81] showed that with high probability, the spectral norm of  $U$  is  $O(\sqrt{n})$ . The matrix  $V$  is obtained by choosing a random  $k \times k$  sign matrix  $V'$ , and padding  $V' + J$  (where  $J$  is the all-ones matrix). Since  $J$  has rank one, the result of Füredi and Komlós also shows that all but the largest eigenvalue of  $V$  is  $O(\sqrt{k})$ . This follows from the inequality  $\lambda_2(V' + J) \leq \lambda_1(V') + \lambda_2(J) = \lambda_1(V')$ .

We bound the spectral norm of  $W$  using its Frobenius norm  $\text{Tr } W^2$  and the inequality  $\lambda_1^2(W) \leq \text{Tr } W^2$ . The Frobenius norm is just the sum of squares of entries. If there are  $S_j$  non-edges connecting  $j \notin K$  to

the clique vertices then  $\text{Tr } W^2 = 2 \sum_{j \notin K} S_j x_j^2$ . We chose  $x_j$  so that  $(k - S_j) + S_j(x_j - 1) = 0$ , and so  $x_j = (2S_j - k)/S_j$ . Thus

$$\text{Tr } W^2 = 2 \sum_{j \notin K} S_j \left( \frac{2S_j - k}{S_j} \right)^2 = 2 \sum_{j \notin K} \frac{(2S_j - k)^2}{S_j}.$$

Now  $S_j \sim \text{Bin}(k, 1/2)$ , and so a Chernoff bound shows that with high probability  $S_j \geq k/3$  for all  $j$ . Roughly speaking,  $2S_j - k \sim N(0, k)$ . In particular,  $\mathbb{E}[(2S_j - k)^2] \approx k$  and  $\mathbb{E}[(2S_j - k)^4] \approx 3k$  (since  $\mathbb{E}[N(0, 1)^4] = 3$ ). Another Chernoff bound shows that with high probability,  $\sum_j (2S_j - k)^2$  doesn't deviate much from its expectation  $(n - k)k$ , say it is at most  $2nk$ . In total, with high probability

$$\lambda_1(W)^2 \leq \text{Tr } W^2 \leq \frac{2 \cdot 2nk}{k/3} = O(n).$$

It follows that with high probability,

$$\lambda_2(M) \leq \lambda_1(U) + \lambda_2(V) + \lambda_1(W) = O(\sqrt{n}) + O(\sqrt{k}) + O(\sqrt{n}) = O(\sqrt{n}).$$

When  $C$  is large enough, this is less than  $k$ , and so  $k$  is indeed the largest eigenvalue of  $M$ .  $\square$

This implies the following algorithm:

**Corollary 10.10.** *Let  $k = C\sqrt{n}$ , for a large enough  $C > 0$ . With high probability, a vertex  $i$  belongs to the planted clique if and only if  $\theta(G \setminus i) = k - 1$ .*

*Proof.* If  $v$  is in the planted clique then  $G \setminus i \sim G(n-1, 1/2, k-1)$ , and otherwise  $G \setminus i \sim G(n-1, 1/2, k)$ . In both cases, the theorem shows that  $\theta(G \setminus i)$  recovers the size of the remaining planted clique.  $\square$

Using the idea of Theorem 10.3, we can use this algorithm for a clique of size  $c\sqrt{n}$  for any  $c > 0$ . Feige and Krauthgamer also show how to decode the clique from the matrix  $M$  witnessing  $\theta(G) = k$ .

## 10.6 Combinatorial algorithms\*

Feige and Ron [FR10] analyzed the following simple two-phase algorithm:

1. Repeatedly remove a vertex of minimum degree, until the remaining graph is a clique.
2. Go over the vertices in reverse order of removal, adding back a vertex whenever the resulting graph is a clique.

They show that this algorithm finds the planted clique in  $G(n, 1/2, k)$  with constant probability when  $k = C\sqrt{n}$  for large enough  $C > 0$ .

Dekel, Gurel-Gurevich and Peres [DGGP14] gave a different combinatorial algorithm which succeeds with high probability in the same setting. Their algorithm consists of three phases:

1. **Enrichment:** Sample  $S$ , an  $\alpha$ -fraction of vertices, and remove  $S$  together with all vertices which are connected to less than  $|S|/2 + \beta\sqrt{|S|}/2$  of them. Repeat  $t = O(\log n)$  times.
2. **Seed:** Choose the  $\alpha^t k$  vertices of largest degree, forming a set  $T$ . With high probability, all vertices in  $T$  belong to the planted clique.
3. **Completion:** Consider the graph induced by  $T$  and its common neighbors, and choose the  $k$  vertices of largest degree.

Consider the first iteration of enrichment. A non-clique vertex has  $\text{Bin}(|S|, 1/2) \approx N(|S|, |S|/4)$  neighbors in  $S$ , and so the probability that it survives is roughly  $x = \Pr[N(0, 1) \geq \beta]$ . A clique vertex is connected to all the roughly  $\alpha k$  vertices in  $S$ , and so it has roughly  $\alpha k + \text{Bin}(\alpha(n - k), 1/2) \sim N(\alpha n/2 + \alpha k/2, \alpha n/4)$  neighbors in  $S$ , and so the probability that it survives is roughly  $y = \Pr[N(0, 1) \geq (\beta\sqrt{n} - \alpha k)/\sqrt{\alpha n}] = \Pr[N(0, 1) \geq (\beta - C\alpha)/\sqrt{\alpha}]$ . If we choose parameters so that  $y > x$ , then we have enriched the fraction of clique vertices.

After a logarithmic fraction of iterations, we will have enriched the clique so much that we can find it using Kučera's algorithm. The analysis is completed along the lines of Theorem 10.3, though there are a few subtleties resulting from the fact that  $T$  is not a random sample of the planted clique.

## 10.7 Lower bounds

We have explained several algorithms that reveal hidden cliques of size  $\Omega(\sqrt{n})$ . Other algorithms exist: Ames and Vavasis [AV11] give yet another spectral algorithm, based on nuclear norm minimization, and Deshpande and Montanari [DM15] given an algorithm based on belief propagation. All of these algorithm require the hidden clique to be of size  $\Omega(\sqrt{n})$ . Can we do better?

Feldman et al. [FGR<sup>+</sup>13] (see also subsequent work mentioned there) show that efficient algorithms that get samples of vertices along with their neighbors cannot identify the planted clique if it has size  $n^{1/2-\epsilon}$  for  $\epsilon > 0$ . Barak et al. [BHK<sup>+</sup>16] consider strengthenings of the Lovász theta function, corresponding to the sum-of-squares hierarchy (which is known to be “universal” for the class of SDP relaxations, in some sense), and show that they also cannot go beyond  $n^{1/2-o(1)}$  efficiently. It is conjectured that  $\Omega(\sqrt{n})$  is indeed the correct threshold; see Deshpande and Montanari for an even more refined conjecture for which the conjectured threshold is  $\sqrt{n/e}$ .

## 11 Random regular graphs

So far we have seen two models for random graphs:  $G(n, m)$  and  $G(n, p)$ . Today we will consider a third basic model, random regular graphs. This model is especially useful in theoretical computer science, since in many cases we are interested in sparse graphs. While  $G(n, d/n)$  is also a model of sparse random graphs, the strict bound on the degrees afforded by random regular graphs is important in some applications. One of the most important properties of random regular graphs are that they are expanders (in various senses) with high probability.

Our exposition is based on the textbook [FK16, Chapter 10] and on lecture notes of Ellis [Ell], as well as on [JLR00, Chapter 9].

Let  $n, d$  be such that  $nd$  is even (otherwise no  $d$ -regular graph on  $n$  vertices exists). A *random  $d$ -regular graph* is a  $d$ -regular graph on  $n$  vertices which is uniformly distributed over all such graphs. While this is a very simple and natural definition, it is not at all clear how to study it. It is not even clear how to generate a random  $d$ -regular graph. The trick is to use Bollobás’ *configuration model*. We think of every edge as a combination of two half-edge, attached to the two vertices it connects. In a  $d$ -regular graph, every vertex is adjacent to exactly  $d$  half-edges. It is thus natural to consider the following process:

1. Construct a random perfect matching on  $[n] \times [d]$ . Here  $\{x\} \times [d]$  are the  $d$  half-edges connected to  $x$ .
2. Add an edge  $(x, y)$  for each edge  $(x, i), (y, j)$  in the matching.

This process is easy to implement, but doesn’t quite satisfy our requirements, since it doesn’t always produce a simple graph. There are two types of problems: there could be self-loops, and there could be parallel edges. However, this only happens with constant probability, and moreover, conditioned on the result being simple, it is a completely uniform random regular graph. This is the contents of the following result, for which we introduce the following notations:

1.  $\mathbb{G}^*(n, d)$  is a random multigraph generated according to the process outlined above.
2.  $\mathbb{G}(n, d)$  is a random  $d$ -regular graph.

**Theorem 11.1** ([FK16, Corollary 10.2, Corollary 10.7]). *Fix  $d \geq 3$ . Let  $G \sim \mathbb{G}^*(n, d)$ , and let  $E$  be the event that  $G$  is simple. Then  $G|E \sim \mathbb{G}(n, d)$ , and*

$$\Pr[E] \rightarrow e^{-(d^2-1)/4}.$$

We require  $d \geq 3$ , since the cases  $d = 1$  and  $d = 2$  behave differently in a qualitative sense.

The fact that  $\Pr[E]$  is bounded below has the following important consequence: if a property occurs with high probability in  $\mathbb{G}^*(n, d)$ , then it also occurs with high probability in  $\mathbb{G}(n, d)$ . This will allow us to analyse random  $d$ -regular graphs, and even to count them.

The difficult part of Theorem 11.1 is estimating the probability that  $\mathbb{G}^*(n, d)$  is simple. That  $\mathbb{G}^*(n, d)|E \sim \mathbb{G}(n, d)$  is an elementary statement, whose proof allows us to estimate the number of random  $d$ -regular graphs, assuming the estimate  $\Pr[E] \rightarrow e^{-(d^2-1)/4}$ .



**Lemma 11.2** ([FK16, Theorem 10.4]). *Fix  $d \geq 3$ . The number of  $d$ -regular graphs on  $n$  vertices is asymptotic to*

$$\sqrt{2}e^{-(d^2-1)/4} \left( \frac{d^{d/2}}{e^{d/2}d!} \right)^n n^{nd/2}.$$

*Proof.* Take any  $d$ -regular graph, and a particular representation of it as a perfect matching on  $[n] \times [d]$ . All other representations are obtained by permuting the labels of the half-edges, and thus this graph has  $(d!)^n$  different representations. Since this number is the same for all graphs, it follows that  $\mathbb{G}^*(n, d)|E \sim \mathbb{G}(n, d)$ .

On the other hand, the overall number of perfect matchings on  $[n] \times [d]$  is

$$(nd-1)(nd-3)\cdots(1) = \frac{(nd)!}{(nd)(nd-2)(nd-4)\cdots 2} = \frac{(nd)!}{2^{nd/2}(nd/2)!}.$$

Since  $\Pr[E] \rightarrow e^{-(d^2-1)/4}$ , roughly a  $e^{-(d^2-1)/4}$  fraction of them correspond to simple graphs, and each such graph is represented  $(d!)^n$  many times. This implies that the number of  $d$ -regular graphs on  $n$  vertices is asymptotic to

$$\begin{aligned} e^{-(d^2-1)/4} \frac{(nd)!}{2^{nd/2}(nd/2)!(d!)^n} &\sim e^{-(d^2-1)/4} \frac{\sqrt{2\pi nd}(nd/e)^{nd}}{\sqrt{\pi nd}(nd/2e)^{nd/2}2^{nd/2}(d!)^n} \\ &\sim \sqrt{2}e^{-(d^2-1)/4} \frac{(nd)^{nd/2}}{e^{nd/2}(d!)^n}. \quad \square \end{aligned}$$

Let  $X_r$  denote the number of cycles of length  $r$  in  $\mathbb{G}^*(n, d)$ . The idea behind the proof of Theorem 11.1 is showing that  $X_r$  has roughly Poisson distribution, with expectation roughly  $\frac{(d-1)^r}{2r}$ . Given that, the graph is simple when  $X_1 = X_2 = 0$ , and so with probability roughly  $e^{-(d-1)/2} \cdot e^{-(d-1)^2/4} = e^{-(d^2-1)/4}$ .

We start by estimating  $\mathbb{E}[X_r]$ .

**Lemma 11.3.** *Fix  $d \geq 3$  and  $r$ . As  $n \rightarrow \infty$ , the expected number of  $r$ -cycles in  $\mathbb{G}^*(n, d)$  tends to*

$$\frac{(d-1)^r}{2r}.$$

*Proof.* We will count the expected number of  $r$ -tuples of vertices  $v_1, \dots, v_r$  such that there are edges  $(v_1, v_2), \dots, (v_{r-1}, v_r), (v_r, v_1)$ , showing that it is approximately  $(d-1)^r$ . This counts each cycle  $2r$  times, hence the result.

A cycle  $(v_1, v_2), \dots, (v_r, v_1)$  involves two half-edges for each vertex, and so the number of possible matchings of half-edges corresponding to this cycle is  $(d(d-1))^r$ . A random perfect matching contains  $r$  specific edges with probability  $1/(nd-1)(nd-3)\cdots(nd-(2r-1)) \sim 1/(nd)^r$ , and so the probability that this specific cycle appears is asymptotic to  $(d(d-1))^r/(nd)^r = (d-1)^r/n^r$ . Since there are  $n^r \sim n^r$  choices for the vertices, the expected number of cycles is asymptotic to  $(d-1)^r$ .  $\square$

In order to show that the distribution of  $X_r$  is roughly Poisson, we estimate higher moments of  $X_r$ . We will content ourselves in carrying out the calculation for  $X_r^2$ , the general calculation being very similar.

**Lemma 11.4.** *Fix  $d \geq 3$  and  $r$ . As  $n \rightarrow \infty$ , the expected number of ordered pairs of distinct  $r$ -cycles in  $\mathbb{G}^*(n, d)$  tends to*

$$\left( \frac{(d-1)^r}{2r} \right)^2.$$

*Proof.* As before, we will count the expected number of ordered pairs of distinct  $r$ -tuples corresponding to cycles, showing it to be asymptotic to  $(d-1)^{2r}$ . This counts each pair of distinct  $r$ -cycles exactly  $(2r)^2$  times, hence the formula.

Let  $C_1, C_2$  be two distinct  $r$ -tuples. If  $C_1, C_2$  are vertex-disjoint then the probability that both appear as cycles in  $\mathbb{G}^*(n, d)$  is

$$\frac{(d(d-1))^{2r}}{(nd-1)(nd-3)\cdots(nd-(4r-1))} \sim \frac{(d-1)^{2r}}{n^{2r}}.$$

There are  $n^{2r} \sim n^{2r}$  such pairs, and they contribute  $(d-1)^{2r}$  to the sum  $\mathbb{E}[X_r(X_r-1)] = \sum_{C_1 \neq C_2} \Pr[C_1, C_2 \subseteq \mathbb{G}^*(n, d)]$ . It remains to show that the contribution of non-vertex-disjoint tuples is small. Indeed, there are  $O(n^{2r-1})$  choice of realizations of such tuples, and each one occurs with probability  $1/n^{2r}$ , for a total contribution of  $O(1/n)$ .  $\square$

In exactly the same way, one can show the following more general result.

**Lemma 11.5.** *Fix  $d \geq 3$ ,  $r$  and  $t_1, \dots, t_r$ . Then*

$$\mathbb{E}[X_1^{t_1} \dots X_r^{t_r}] \rightarrow \prod_{i=1}^r \left( \frac{(d-1)^i}{2i} \right)^{t_i}.$$

A multidimensional version of the argument in Theorem 6.2 then implies the following corollary.

**Corollary 11.6.** *Fix  $d \geq 3$  and  $r$ . The joint distribution of  $(X_1, \dots, X_r)$  tends to independent Poisson distributions  $(\text{Po}(\lambda_1), \dots, \text{Po}(\lambda_r))$ , where  $\lambda_i = (d-1)^i/2i$ .*

As noted above, Theorem 11.1 immediately follows.

## 11.1 Connectedness

We demonstrate the configuration model by showing that with high probability,  $\mathbb{G}(n, d)$  is connected. In view of Theorem 11.1, it suffices to show that  $\mathbb{G}^*(n, d)$  is connected with high probability. We will show this using the first moment method.

**Theorem 11.7.** *Fix  $d \geq 3$ . With high probability,  $\mathbb{G}(n, d)$  is connected.*

*Proof.* A non-empty set  $A$  is a *separator* if  $|A| \leq n/2$  and there are no edges between  $A$  and its complement. We will show that the expected number of separators in  $\mathbb{G}^*(n, d)$  is  $o(1)$ , and so with high probability  $\mathbb{G}^*(n, d)$  is connected. Theorem 11.1 implies that with high probability,  $\mathbb{G}(n, d)$  is connected, since

$$\begin{aligned} \Pr[\mathbb{G}(n, d) \text{ not connected}] &= \frac{\Pr_{G \sim \mathbb{G}^*(n, d)}[G \text{ not connected} | G \text{ simple}]}{\Pr[\mathbb{G}^*(n, d) \text{ simple}]} \\ &= \frac{\Pr[\mathbb{G}^*(n, d) \text{ simple and not connected}]}{\Pr[\mathbb{G}^*(n, d) \text{ simple}]} \\ &\leq \frac{\Pr[\mathbb{G}^*(n, d) \text{ not connected}]}{\Pr[\mathbb{G}^*(n, d) \text{ simple}]} \rightarrow 0. \end{aligned}$$

Let  $|A| = a$ . If  $A$  is a separator then in the configuration model, the  $da$  vertices corresponding to vertices in  $A$  all get matched within themselves. This happens with probability

$$\frac{da-1}{dn-1} \cdot \frac{da-3}{dn-3} \cdots \frac{da-(da-1)}{dn-(da-1)} \leq \left( \frac{a}{n} \right)^{da/2}.$$

There are  $\binom{n}{a} \leq \left( \frac{en}{a} \right)^a$  choices for  $A$  of this size, and so the expected number of separators of size  $a$  is at most

$$\left( \frac{en}{a} \right)^a \cdot \left( \frac{a}{n} \right)^{da/2} = \left( \frac{ea^{d/2-1}}{n^{d/2-1}} \right)^a.$$

This is small when  $(a/n)^{d/2-1} < 1/e$ . To be concrete, pick an arbitrary constant  $c \in (0, 1/2)$  such that  $c^{d/2-1} < 1/e$  (this is possible since  $d/2-1 > 0$ ). The expected number of separators of size at most  $cn$  is at most

$$\begin{aligned} \sum_{a=1}^{n^{1/4}} \left( \frac{en^{(d/2-1)/4}}{n^{d/2-1}} \right)^a + \sum_{a=n^{1/4}+1}^{cn} (ec^{d/2-1})^a \\ \leq n^{1/4} \frac{e}{n^{(3/4)(d/2-1)}} + O((ec^{d/2-1})^{n^{1/4}}) \leq O(n^{-1/8} + (ec^{d/2-1})^{n^{1/4}}) = o(1), \end{aligned}$$

for  $n$  large enough so that  $\frac{en^{(d/2-1)/4}}{n^{d/2-1}} < 1$ .

When  $a$  is large, we use a more accurate asymptotic estimate. For even  $x$ , define

$$x!! = (x-1)(x-3)(x-5)\cdots 1 = \frac{x(x-1)(x-2)(x-3)\cdots}{x(x-2)\cdots} = \frac{x!}{2^{x/2}(x/2)!}.$$

The probability that  $A$  is a separator is thus

$$\frac{(da)!!(d(n-a))!!}{(dn)!!} = \frac{\binom{dn/2}{da/2}}{\binom{dn}{da}}.$$

(Another way to see this is to note that if  $A$  is a separator then the perfect matching decomposes into a perfect matching on  $A$  and a perfect matching on its complement.)

When  $a = xn$ , Stirling's approximation shows that

$$\binom{n}{a} \sim \frac{1}{2\pi x(1-x)} 2^{nh(x)},$$

where  $h(a) = -x \log_2 x - (1-x) \log_2(1-x)$  is the binary entropy function. Therefore the expected number of separators of size  $a = xn$  is at most

$$\binom{n}{a} \frac{\binom{dn/2}{da/2}}{\binom{dn}{da}} \sim \frac{1}{2\pi x(1-x)} 2^{nh(x)[1+d/2-d]} \leq \frac{1}{2\pi x(1-x)} 2^{-nh(x)/2}.$$

In view of the preceding calculation, we can assume that  $x \geq c$ , and so the expected number of separators of size  $a = xn$  is at most  $\Omega(2^{-\Omega(n)}) = o(1/n)$ , showing that the expected number of separators overall is  $o(1)$ .  $\square$

This result can be strengthened in several ways, where we always assume that  $d \geq 3$ . First, we can show that not only is  $\mathbb{G}(n, d)$  connected with high probability, but also that it is an expander; see Ellis [Ell, Theorem 5]. Second, we can show that  $\mathbb{G}(n, d)$  is not only connected, but also  $d$ -connected (remains connected after removing at most  $d-1$  vertices) with high probability; see Ellis [Ell, Theorem 6] or the textbook [FK16, Theorem 10.8]. Via Tutte's criterion, a generalization of Hall's criterion for non-bipartite graphs, a  $d$ -connected  $d$ -regular graph having an even number of vertices contains a perfect matching. We conclude that with high probability,  $\mathbb{G}(n, d)$  contains a perfect matching. In fact, more is true: with high probability,  $\mathbb{G}(n, d)$  is *Hamiltonian*. The proof, which uses an extended version of the second moment method, is unfortunately quite complicated.

## 11.2 Contiguity

The random  $d$ -regular graph model is hard to analyze directly, and so we have come up with the model  $\mathbb{G}^*(n, d)$  which is easy to analyze and has a very strong relation to  $\mathbb{G}(n, d)$ : an event holds with high probability on  $\mathbb{G}(n, d)$  if it holds with high probability on  $\mathbb{G}^*(n, d)$ . When two models are such that this relation holds both ways, we say that they are *contiguous*. Difficult results in the theory of random regular graphs show the following contiguity results, where  $\mathbb{G}'(n, d)$  is  $\mathbb{G}^*(n, d)$  conditioned on having no loops. (See [JLR00, Section 9.5] for proofs.)

1. When  $d \geq 4$  is even,  $\mathbb{G}'(n, d)$  is contiguous to a union of  $d/2$  random Hamiltonian cycles.
2. When  $d \geq 3$  is odd,  $\mathbb{G}'(n, d)$  is contiguous to a union of  $(d-1)/2$  random Hamiltonian cycles and a random perfect matching.
3. When  $d \geq 4$  is even,  $\mathbb{G}(n, d)$  is contiguous to a union of  $d/2$  random Hamiltonian cycles, conditioned on it being simple.
4. When  $d \geq 3$  is odd,  $\mathbb{G}(n, d)$  is contiguous to a union of  $(d-1)/2$  random Hamiltonian cycles and a random perfect matching, conditioned on it being simple.

A result along the lines of Theorem 11.1 shows that  $\mathbb{G}'(n, d)$  is simple with probability tending to  $e^{-(d-1)^2/4}$ , and so an event holds with high probability on  $\mathbb{G}(n, d)$  if it holds with high probability on  $\mathbb{G}'(n, d)$ . This allows us to prove results like the following.

**Theorem 11.8.** *A random 4-regular graphs can be partitioned, with high probability, to cycles of length at most  $4\sqrt{n \log n}$ .*

*Proof.* We will show that the union of two random Hamiltonian cycles can be so partitioned, with high probability. We can fix one of the Hamiltonian cycles to be  $1, 2, \dots, n$ , and let the other one be  $\pi(1), \pi(2), \dots, \pi(n)$ . Partition  $\{1, \dots, n\}$  into  $n/m$  intervals  $I_1, \dots, I_{n/m}$  of length  $m$ , where  $m = \sqrt{n \log n}$ . The probability that  $\pi(i) \notin I_j$  for all  $i \in I_j$  is

$$\frac{(n-m)(n-m-1)\cdots(n-2m+1)}{n(n-1)\cdots(n-m+1)} \leq \left(1 - \frac{m}{n}\right)^m \leq e^{-m^2/n} = \frac{1}{n}.$$

In this case, we say that  $I_j$  is *bad*. The probability that some  $I_j$  is bad is at most  $\frac{n/m}{n} = \frac{1}{m} = o(1)$ , and so with high probability all  $I_j$  are good. Thus for each interval  $I_j$  there exists a point  $x_j \in I_j$  such that  $\pi(x_j) \in I_j$ . The required partition into cycles is  $\pi(x_j), \pi(x_j+1), \dots, \pi(x_{j+1}), \pi(x_{j+1})-1, \dots, \pi(x_j)$ , the first part in the Hamiltonian cycle given by  $\pi$ , and the second part in  $n, \dots, 2, 1$ .  $\square$

We don't know whether  $\sqrt{n \log n}$  is the optimal order of magnitude.

## 12 Expanders

Random graphs have many desirable properties. Sometimes we would like to construct a graph with similar properties, but deterministically. Quasirandom graphs have the correct subgraph densities as well as other properties, but they are lacking in other regards; most importantly, they are dense graphs, whereas in many applications we are interested in bounded-degree graphs. Expanders are another type of quasirandom graphs which have many uses in theoretical computer science. In many cases it is important that they can be constructed deterministically, but in some cases their existence suffices.

Expanders come in many types and can be described in several different ways. There is a major distinction between *bipartite expanders* and *non-bipartite expanders*. In both cases we are most often interested in  $d$ -regular graphs for constant  $d$ .

**Non-bipartite expanders** General expanders are sometimes described as graphs which behave (in certain ways) like the complete graph, while being sparse (having  $O(n)$  edges). Expanders are, informally, graphs in which any set has many neighbors. This informal definition concerns *vertex expansion*, but it turns out that a more useful definition is about *edge expansion*. The *(edge) expansion* of a graph  $G$  on  $n$  vertices is given by

$$h(G) = \min_{|S| \leq n/2} \frac{|E(S, \bar{S})|}{|S|},$$

where  $E(S, T)$  is the set of edges between  $S$  and  $T$ . In words,  $h(G)$  is the best parameter such that every set  $S$  of size at most  $n/2$  is connected to its complement by at least  $h(G)|S|$  edges.

Sometimes we are interested instead in sizes of neighborhoods. Let  $N^*(S)$  denote the set of nodes in  $\bar{S}$  connected to vertices in  $S$ . We can define the vertex expansion of  $G$  as follows:

$$h_V(G) = \min_{|S| \leq n/2} \frac{|N^*(S)|}{|S|},$$

Clearly the two parameters differ by a constant (for constant  $d$ ):  $h(G)/d \leq h_V(G) \leq h(G)$ .

A sequence  $G_1, G_2, \dots$  of  $d$ -regular graphs whose number of vertices tends to infinity is an *expander* if  $h(G_n) \geq h$  for some positive constant  $h > 0$ . Usually we abuse the definition and call a *particular* graph an expander.<sup>4</sup>

---

<sup>4</sup>In non-standard arithmetic, the existence of a sequence of expanders is equivalent to the existence of a single expander with a *non-standard* number of vertices.

**Bipartite expanders** In many applications slightly different properties are needed, and often the natural graph to be considered is bipartite (for example, a random CNF can be described as a bipartite graph connecting variables or literals to clauses). A bipartite graph with bipartition  $L, R$  (where  $|R| \leq |L|$ ) which is  $d$ -regular on the left is called a  $(\gamma, \alpha)$ -expander if for all  $S \subseteq L$  of size  $|S| \leq \gamma|L|$ , we have  $|N(S)| \geq \alpha|S|$ . A sequence of  $d$ -regular bipartite graphs form an expander sequence if there are  $(\gamma, \alpha)$ -expanders for some positive constants  $\gamma, \alpha > 0$ .

Sometimes we want a stronger property, about *unique neighbors* rather than just neighbors. The *unique neighborhood* of  $S$  consists of all vertices in  $\bar{S}$  which have a *unique* neighbor in  $S$ . A bipartite graph is a  $(\gamma, \alpha)$ -unique expander if every subset  $S \subseteq L$  of size at most  $\gamma|L|$  has at least  $\alpha|S|$  unique neighbors.

## 12.1 Spectral expanders

Non-bipartite expanders can also be defined by examining their *spectrum*. The adjacency matrix  $A$  of every  $d$ -regular graph on  $n$  vertices has  $d$  as an eigenvalue, corresponding to the constant eigenspace. This is also the maximum magnitude of an eigenvalue of  $A$ , as follows from the Perron–Frobenius theorem. It can also be seen directly: if  $v$  is an eigenvector corresponding to the eigenvalue  $\lambda$  and  $v_i$  is an entry of maximum magnitude then

$$|\lambda v_i| = |(Av)_i| = \left| \sum_{j \in N(i)} v_j \right| \leq d|v_i|.$$

It is an easy exercise in spectral graph theory that the dimension of the eigenspace of  $d$  equals the number of connected components of the graph. Since  $A$  is symmetric, the only other possible eigenvalue with magnitude  $d$  is  $-d$ . Another easy exercise shows that this is an eigenvalue if and only if the graph is bipartite.

A graph can only be an expander if it is connected (since a connected component doesn't expand at all, and some connected component has size at most  $n/2$ ). The *spectral gap* of a graph is defined as  $d - \lambda_2$ , where  $\lambda_2$  is the second largest eigenvalue of  $A$  (we implicitly assume that the eigenspace of  $d$  has dimension 1). We can bound the edge expansion of a graph in terms of its spectral gap. Suppose that  $S$  is a set of at most  $n/2$  vertices, and let  $1_S$  be its characteristic vector. Note that  $1_{\bar{S}} = \mathbf{1} - 1_S$ , where  $\mathbf{1}$  is the constant 1 vector. We have

$$|E(S, \bar{S})| = 1_S^T A 1_{\bar{S}} = 1_S^T A (\mathbf{1} - 1_S) = d 1_S^T \mathbf{1} - 1_S^T A 1_S = d|S| - 1_S^T A 1_S.$$

We can write  $1_S = \frac{|S|}{n} \mathbf{1} + v$ , where  $v$  is orthogonal to  $\mathbf{1}$ . Since this is an orthogonal decomposition, we have  $|S| = \|1_S\|^2 = \left\| \frac{|S|}{n} \mathbf{1} \right\|^2 + \|v\|^2 = \frac{|S|^2}{n} + \|v\|^2$ , so that  $\|v\|^2 = |S| - \frac{|S|^2}{n}$ . On the other hand, since  $v$  is orthogonal to  $\mathbf{1}$ , we have  $v^T A v \leq \lambda_2 \|v\|^2$ . In total,

$$1_S^T A 1_S = \frac{|S|^2}{n^2} \mathbf{1}^T A \mathbf{1} + v^T A v \leq \frac{d|S|^2}{n} + \lambda_2 \left( |S| - \frac{|S|^2}{n} \right) = (d - \lambda_2) \frac{|S|^2}{n} + \lambda_2 |S|.$$

We conclude that

$$\frac{|E(S, \bar{S})|}{|S|} \geq d - (d - \lambda_2) \frac{|S|}{n} - \lambda_2 = (d - \lambda_2) \left( 1 - \frac{|S|}{n} \right) \geq \frac{d - \lambda_2}{2}.$$

In other words,  $h(G) \geq \frac{d - \lambda_2}{2}$  for any  $d$ -regular graph. In other words, a sequence of  $d$ -regular graphs whose spectral gap is bounded from below is an expander sequence. A more difficult argument shows that the converse holds as well:

$$\frac{d - \lambda_2}{2} \leq h(G) \leq \sqrt{2d(d - \lambda_2)}.$$

This is known as Cheeger's inequality.

Alon and Boppana showed that  $\lambda_2 \geq 2\sqrt{d-1} - o_n(1)$  for every  $d$ -regular graph, where the error term vanishes as  $n \rightarrow \infty$ . The quantity  $2\sqrt{d-1}$  is the second eigenvalue of the infinite  $d$ -regular tree, which is thus, in a sense, the best possible  $d$ -regular expander. Expander sequences which consist of  $d$ -regular graphs satisfying  $\lambda_2 \geq 2\sqrt{d-1}$  are known as *Ramanujan graphs*.

## 12.2 Constructions

It turns out that a random  $d$ -regular graph is an expander with high probability, and this can be shown using the first moment method (see Ellis [Ell]). Friedman [Fri08] and Puder [Pud15] showed that it has an almost optimal spectral gap. We show below a simpler result for bipartite expanders.

There are several explicit constructions of expanders:

1. Margulis/Gabber–Galil/Lubotzky–Phillips–Sarnak: A family of 8-regular expanders on  $m^2$  vertices for every  $m$ . The vertex set is  $\mathbb{Z}_m^2$ . A vertex  $(x, y)$  is connected to  $(x \pm y, y), (x \pm y + 1, y), (x, y \pm x), (x, y \pm x + 1)$ .
2. A family of 3-regular expanders on  $p$  vertices for every prime  $p$ . The vertex set is  $\mathbb{Z}_p$ . A vertex  $x$  is connected to  $x \pm 1$  and to  $x^{-1}$  (where  $0^{-1} = 0$ ).

Other constructions use lifts [Coh16] or the zig-zag product, used by Reingold [Rei08] to give a log-space algorithm for undirected connectivity.

There are also explicit constructions of bipartite expanders of various types. We will be content here to show that when  $d \geq 2$ , a random bipartite left- $d$ -regular graph is a  $(\gamma, \alpha)$ -expander with high probability, for appropriate values of  $\gamma, \alpha$ . In fact, for each  $\gamma < 1$  some value  $\alpha > 0$  will work. A *random bipartite left- $d$ -regular graph* on  $2n$  vertices is one chosen at random among all bipartite graphs with two bipartitions of size  $n$  in which all vertices on the left have degree  $d$ . Such a graph can be generated by choosing for each vertex on the left  $d$  (not necessarily distinct) random neighbors on the right.

The expected number of 2-cycles is  $n \binom{d}{2} \cdot \frac{1}{n} = \binom{d}{2}$ , and the number of 2-cycles has roughly Poisson distribution; thus the probability that the resulting graph is simple tends to  $e^{-\binom{d}{2}}$ . This shows that if we choose a random *simple* bipartite left- $d$ -regular graph, it will also be a  $(\gamma, \alpha)$ -expander with high probability: indeed, the probability that a random simple bipartite left- $d$ -regular graph is not a  $(\gamma, \alpha)$ -expander is large by a factor of at most roughly  $e^{\binom{d}{2}}$  than the probability that a random bipartite left- $d$ -regular graph is not a  $(\gamma, \alpha)$ -expander.

Let  $S \subseteq L$  be a set of size  $s \leq \gamma n$ . If  $T \subseteq R$  is a particular set of size  $\alpha s$ , then the probability that  $N(S) \subseteq T$  is at most  $(\alpha s/n)^{ds}$ . Thus the probability that some set of size  $s$  has at most  $\alpha s$  neighbors is at most

$$\binom{n}{s} \binom{n}{\alpha s} \left(\frac{\alpha s}{n}\right)^{ds} \leq \left(\frac{en}{s}\right)^s \left(\frac{en}{\alpha s}\right)^{\alpha s} \left(\frac{\alpha s}{n}\right)^{ds} = \left(\frac{en}{s} \cdot \frac{(en)^\alpha}{(\alpha s)^\alpha} \cdot \frac{(\alpha s)^d}{n^d}\right)^s = \left(e^{1+\alpha} \alpha^{d-\alpha} \cdot \frac{s^{d-1-\alpha}}{n^{d-1-\alpha}}\right)^s.$$

Let  $\rho = e^{1+\alpha} \alpha^{d-\alpha} \gamma^{d-1-\alpha}$ , so that  $\rho^s$  is an upper bound on the probability that some set of size  $s$  has at most  $\alpha s$  neighbors. Since  $\rho$  vanishes as  $\alpha \rightarrow 0$ ,  $\rho < 1$  for small enough  $\alpha > 0$ . The probability that some set of size  $s \leq \gamma n$  has at most  $\alpha s$  neighbors is thus at most

$$\sum_{s=1}^{\gamma n} \left(e^{1+\alpha} \alpha^{d-\alpha} \cdot \frac{s^{d-1-\alpha}}{n^{d-1-\alpha}}\right)^s \leq n^\epsilon e^{1+\alpha} \alpha^{d-\alpha} \cdot \frac{1}{n^{(1-\epsilon)(d-1-\alpha)}} + \sum_{s=n^\epsilon+1}^n \rho^s \leq O(n^{\epsilon-(1-\epsilon)(d-1-\alpha)}) + O(x^{n^\epsilon}).$$

For small enough  $\epsilon > 0$  we have  $\epsilon < (1-\epsilon)(d-1-\alpha)$  (since  $d \geq 2$ ), and we deduce that the probability that the graph is not a  $(\gamma, \alpha)$ -expander is  $o(1)$ , assuming that  $\alpha$  is small enough so that  $\rho < 1$ .

## 12.3 Properties

Expander graphs (non-bipartite ones!) satisfy many useful properties. Here we prove just two: they have diameter  $O(\log n)$  (where  $n$  is the number of vertices), and they are uniform in terms of the number of edges connecting two sets of vertices, a property known as the *expander mixing lemma*.

We start with the diameter.

**Lemma 12.1.** *Suppose that  $G$  is a  $d$ -regular graph on  $n$  vertices that has expansion  $h(G) \geq h$ . Then  $G$  has diameter at most  $O((d/h) \log n)$ .*

*Proof.* Suppose that  $x, y$  are any two vertices. Denote by  $N^i(x)$  the set of vertices at distance at most  $i$  from  $x$ . If  $|N^i(x)| \leq n/2$  then

$$|N^{i+1}(x)| \geq |N^i(x)| + \frac{|E(N^i(x), \overline{N^i(x)})|}{d} \geq \left(1 + \frac{h}{d}\right) |N^i(x)|.$$

We conclude that if  $|N^i(x)| \leq n/2$  then  $|N^i(x)| \geq (1 + h/d)^i$ , which leads to a contradiction if  $i = c(d/h) \log n$  for an appropriate constant  $c > 0$ . We conclude that  $|N^i(x)| > n/2$  for some  $i = O((d/h) \log n)$ . The same holds for  $y$ , and since  $|N^i(x)|, |N^i(y)| > n/2$ , the two sets must intersect, showing that the distance between  $x$  and  $y$  is at most  $2i = O((d/h) \log n)$ .  $\square$

Another important result is the expander mixing lemma, which requires a slightly stronger notion of spectral expansion: instead of just requiring  $\lambda_2$ , the second largest eigenvalue, to be bounded away from  $d$ , we require  $\lambda = \max(\lambda_2, |\lambda_{\min}|)$  to be bounded away from  $d$ , where  $\lambda_{\min}$  is the minimal eigenvalue. In particular, the graph should not be bipartite. Families of non-bipartite expanders usually satisfy this additional property, and the Alon–Boppana bound  $2\sqrt{d-1} - o(1)$  is actually a bound on  $\lambda$ .

**Lemma 12.2.** *Suppose that  $G$  is a  $d$ -regular graph on  $n$  vertices. For all sets of vertices  $S, T$ ,*

$$\left| |E(S, T)| - \frac{d|S||T|}{n} \right| \leq \lambda \sqrt{|S||T|}.$$

*Proof.* The proof is very similar to the case  $T = \overline{S}$  considered above. Let  $\mathbf{1}_S, \mathbf{1}_T$  be the characteristic vectors of  $S, T$ , and decompose them orthogonally as  $\mathbf{1}_S = \frac{|S|}{n} \mathbf{1} + s$ ,  $\mathbf{1}_T = \frac{|T|}{n} \mathbf{1} + t$ , where  $\|s\|^2 = |S| - \frac{|S|^2}{n}$  and  $\|t\|^2 = |T| - \frac{|T|^2}{n}$  (as above). If  $A$  is the adjacency matrix of  $G$  then

$$|E(S, T)| = \mathbf{1}_S^T A \mathbf{1}_T = \frac{|S||T|}{n^2} \mathbf{1}^T A \mathbf{1} + s^T A t = \frac{d|S||T|}{n} + s^T A t.$$

The Cauchy–Schwartz inequality implies (after decomposing  $s, t$  into the eigenspaces of  $A$ ) that

$$|s^T A t| \leq \lambda \|s\| \|t\| \leq \lambda \sqrt{|S||T|}. \quad \square$$

Every  $d$ -regular graph has  $\lambda \leq d$ , so perhaps the error bound doesn't look impressive. To put it into perspective, suppose that  $|S| = \alpha n$  and  $|T| = \beta n$ . In that case, the bound states that

$$\left| |E(S, T)| - d\alpha\beta n \right| \leq \lambda \sqrt{\alpha\beta} n.$$

This is useful mostly when  $\lambda \sqrt{\alpha\beta} < d\alpha\beta$ , that is when  $\lambda/d \leq \sqrt{\alpha\beta}$ .

The expander mixing lemma can be used to bound the size of an independent set in  $G$ : if  $S$  is an independent set of size  $\alpha n$  then  $E(S, S) = 0$  and so  $d\alpha^2 n \leq \lambda \alpha n$ , and so  $\alpha \leq \lambda/d$ . This implies that  $\chi(G) \geq d/\lambda$ .

## 12.4 Applications

Expanders have many applications. We briefly mention a few of them:

**AKS sorting network** Expanders are used to construct a sorting network of asymptotically optimal depth  $O(\log n)$ .

**Approximate majority** Ajtai used expanders to construct an  $AC^0$  circuit which can tell apart inputs of weight  $(1/2 - \epsilon)n$  from inputs of weight  $(1/2 + \epsilon)n$ , where  $\epsilon = \frac{1}{\log^d n}$  for arbitrary  $d$ . (Note that such circuits famously *cannot* compute majority.)

**Embedding into Euclidean space** Bourgain showed that any  $n$ -point metric can be embedded into Euclidean space with distortion  $O(\log n)$ . Linial, London and Rabinovich [LLR95] showed that this is tight for expanders.

**Proof complexity** Expanders can be used (via Tseitin formulas) to prove strong lower bounds on Resolution. Other expansion properties pop up in lower bounds for random  $k$ -CNFs.

**Derandomization** Reingold [Rei08] used expanders to derandomize a random walk algorithm, thus proving that undirected reachability can be decided in logspace. (Directed reachability is complete for NL.)

## 13 Quasirandomness

Our exposition is based mostly on Lovász's monograph [Lov12]. The original work on quasirandom graphs is [CGW89], and quasirandom graphs are also discussed in [AS16, Section 9.3].

### 13.1 Graphons

Graphons are a generalization of the  $G(n, p)$  model which is complete in some sense that we sketch below. A *graphon* is a measurable function  $W: [0, 1]^2 \rightarrow [0, 1]$  which is symmetric:  $W(x, y) = W(y, x)$ . A random  $G(n, W)$  is sampled as follows:

1. Sample  $n$  numbers  $x_1, \dots, x_n \in [0, 1]$  uniformly and independently.
2. For  $i \neq j$ , put an edge between  $i$  and  $j$  with probability  $W(x_i, x_j)$ , independently for all edges.

When  $W \equiv p$  is constant, a  $G(n, W)$  random graph is the same as a  $G(n, p)$  random graph.

Earlier we have counted the expected number of subgraphs of a particular type in  $G(n, p)$ , and we can do the same in  $G(n, W)$ . Instead of counting subgraphs, we will count the number of *homomorphisms* from a specific small graph  $H = (V(H), E(H))$ . Given two graphs  $G, H$ , a homomorphism from  $H$  to  $G$  is a function  $h: V(H) \rightarrow V(G)$  such that  $(h(i), h(j)) \in E(G)$  whenever  $(i, j) \in E(H)$ . We denote by  $t(H, G)$  the probability that a random function from  $V(H)$  to  $V(G)$  is a homomorphism; this is the normalized number of copies of  $H$  inside  $G$ . A simple calculation shows that

$$\mathbb{E}[t(H, G(n, W))] \rightarrow t(H, W) := \int_{x_1, \dots, x_{|V(H)|}} \prod_{(i, j) \in E(H)} W(x_i, x_j) \, d\vec{x},$$

where  $x_1, \dots, x_n$  are integrated against the Lebesgue measure on  $[0, 1]$ . Moreover, if  $G_n \sim G(n, W)$  for all  $n$ , then it is not too hard to show that almost surely, for every  $H$  it holds that  $t(H, G_n) \rightarrow t(H, W)$ .

A *graph sequence*  $G_n$  is a sequence of graphs in which  $|V(G_n)| \rightarrow \infty$ . A graph sequence  $G_n$  *converges* to the graphon  $W$  if the statement above holds for all graphs  $H$ . A deep result shows that *any graph sequence has a convergent subsequence*. In this sense the graphon model is complete.

Every finite graph  $G = (V, E)$  has a graphon counterpart  $W$  formed by dividing  $[0, 1]$  into  $|V|$  intervals of equal length, and letting each square be the constant 1 if the corresponding edge is in the graph, and the constant 0 otherwise. This graphon satisfies  $t(H, G) = t(H, W)$  for every  $H$ . In this sense, graphons are generalizations of finite graphs.

Another example is the *stochastic block model*. There are  $k$  types of vertices, a random vertex belongs to type  $i$  with probability  $p_i$ , and vertices of type  $i, j$  are connected with probability  $w_{ij}$ . This is captured by a piecewise constant graphon which the reader can surely construct by herself. One basic question about the stochastic block model, which has recently been answered, is whether the types of the vertices can be recovered given the graph, with high probability. This is known as the problem of *community detection*. See Abbé's recent survey [Abb16].

### 13.2 Quasirandom graphs

We say that a graph sequence  $G_n$  is  $p$ -quasirandom if  $t(H, G_n) \rightarrow p^{|E(H)|}$ , that is, if it behaves like  $G(n, p)$  in terms of graph densities. Chung, Graham and Wilson [CGW89] proved (for  $p = 1/2$ ) the surprising result that a graph sequence  $G_n$  is  $p$ -quasirandom if and only if  $t(-, G_n) \rightarrow p$  and  $t(\square, G_n) \rightarrow p^4$ , that is, if the condition above holds for the two particular graphs  $H = -, \square$ . In such a case we say that the corresponding graphon (in this case, the constant  $p$  graphon) is *finitely forcible*. All piecewise constant graphons are finitely forcible.

**Theorem 13.1.** *A graph sequence  $G_n$  is  $p$ -quasirandom if and only if  $t(-, G_n) \rightarrow p$  and  $t(\square, G_n) \rightarrow p^4$ .*

*Proof.* The only if part is obvious, so it suffices to prove that  $G_n$  is  $p$ -quasirandom if  $t(-, G_n) \rightarrow p$  and  $t(\square, G_n) \rightarrow p^4$ . Suppose to the contrary that for some graph  $H$ ,  $t(H, G_n) \not\rightarrow p^{|E(H)|}$ . We can find a subsequence  $G'_n$  and  $c \neq p^{|E(H)|}$  such that  $t(H, G'_n) \rightarrow c$  (since the interval  $[0, 1]$  is compact). The compactness result quoted above implies that there is a subsequence  $G''_n$  of  $G'_n$  converging to a graphon  $W$ . The graphon  $W$  satisfies  $t(-, W) = p$  and  $t(\square, W) = p^4$ , but  $t(H, W) \neq p^{|E(H)|}$ .



As the following calculation, which uses the inequality  $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$  twice, shows, it is always the case that  $t(\square, W) \geq t(-, W)^4$ :

$$\begin{aligned} t(\square, W) &= \mathbb{E}_{a,b,c,d} [W(a,b)W(b,c)W(c,d)W(d,a)] \\ &= \mathbb{E}_{a,b} \left[ \mathbb{E}_c [W(a,b)W(b,c)]^2 \right] \\ &\geq \mathbb{E}_{a,b,c} [W(a,b)W(b,c)]^2 \\ &= \mathbb{E}_b \left[ \mathbb{E}_a [W(a,b)]^2 \right]^2 \\ &\geq \mathbb{E}_{a,b} [W(a,b)]^4 \\ &= t(-, W)^4. \end{aligned}$$

In our case we have equality, and so  $\int_c W(a,b)W(b,c)$  must be constant almost surely; in fact, equal to  $p^2$  almost surely. If we treat  $W$  as an operator on  $L^2([0,1])$ , this shows that  $W^2 = p^2$ , where  $p^2$  is the operator that maps  $f \in L^2([0,1])$  to the constant function  $p^2 \mathbb{E}[f]$  (note that operators, and graphons, are defined up to measure zero). This implies that  $W = \pm p$  (since  $W^2$  has a unique eigenspace, consisting of all constant functions, which doesn't belong to its kernel). Since  $W \geq 0$  (as a graphon), it follows that  $W = p$ . This contradicts the assumption that  $t(H, W) \neq p^{|E(H)|}$ .  $\square$

Quasirandom graphs enjoy many other properties:

- All eigenvalues other than the maximal one are  $o(n)$ .
- Every set  $S$  contains  $(p/2)|S|^2 \pm o(n^2)$  edges.
- Every two disjoint sets  $S, T$  are connected by  $p|S||T| \pm o(n^2)$  edges.
- The average deviation of  $|\{z : (x, z), (y, z) \in E\}|$  from  $p^2n$  is  $o(n)$ .

### 13.3 Quasirandom and non-quasirandom sequences

The material in this part is taken from Chung, Graham, and Wilson [CGW89].

**Paley graphs** As an application of the theory, we will show that the Paley graphs form a quasirandom sequence for  $p = 1/2$ . For a prime  $q = 4m + 1$ , the Paley graph  $P_q$  is a graph on  $\mathbb{Z}_q$  in which  $i, j$  are connected whenever  $i - j$  is a quadratic residue modulo  $q$  (that is,  $i - j \equiv k^2 \pmod{q}$  for some  $k \in \mathbb{Z}_q$ ). The condition  $q = 4m + 1$  guarantees that  $-1$  is a quadratic residue, and so this defines an undirected graph. It is known that exactly  $\frac{q-1}{2}$  non-zero elements of  $\mathbb{Z}_q$  are quadratic residues (essentially since  $x^2 = (-x)^2$ ), and so the graph is  $\frac{q-1}{2}$ -regular. This easily implies that  $t(-, P_q) \rightarrow 1/2$ .

To estimate the number of squares, consider a pair of vertices  $x \neq y$ . A third vertex  $z$  is adjacent to both or neither  $x, y$  iff  $\frac{z-x}{z-y}$  is a quadratic residue (since quadratic residuity is multiplicative). If  $a \neq 1$  is a quadratic residue then  $\frac{z-x}{z-y} = a$  implies that  $a - 1 = \frac{y-x}{z-y}$ , and so  $z = y + \frac{y-x}{a-1}$ . It follows that there are exactly  $\frac{q+1}{2} - 2 = \frac{q-3}{2}$  vertices  $z$  which are adjacent to both or neither  $x, y$ . Denote by  $N(x)$  the set of neighbors of  $x$ . Then

$$2|N(x) \cap N(y)| = |N(x)| + |N(y)| + (|N(x) \cap N(y)| + |\overline{N(x)} \cap \overline{N(y)}|) - q = (q-1) + \frac{q-3}{2} - q = \frac{q-5}{2},$$

and so  $|N(x) \cap N(y)| = \frac{q-5}{4}$ . This easily implies that  $t(\square, P_q) \rightarrow (1/4)^2 = 1/16$ . We conclude that  $P_q$  is a pseudorandom sequence.

**Intersection parity graphs** Here is another example for  $p = 1/2$ . For every  $n$ , consider the graph whose vertex set is  $\binom{[2n]}{n}$ , connecting two vertices  $A, B$  if their intersection contains an even number of points. Roughly speaking the intersection of two random vertices is distributed like  $\text{Bin}(2n, 1/4)$ , and so it is even roughly half the time. We leave it to the reader to show that, for reasons similar to those in the previous examples, the relative number of squares is roughly  $1/16$ .

**Triangle density is not enough** We have seen that if  $t(-, G_n) \rightarrow p$  and  $t(\square, G_n) \rightarrow p^4$  then the graph sequence is  $p$ -quasirandom. It is natural to ask whether  $\square$  can be replaced by  $\triangle$ . The following counterexample shows it not to be the case. For  $p \leq 1/2$ , consider the following graphon  $W_p$ :

$$\begin{array}{cccc} 2p & 0 & p & p \\ 0 & 2p & p & p \\ p & p & 0 & 2p \\ p & p & 2p & 0 \end{array}$$

This corresponds to a graphon by replacing each cell by a constant  $1/4 \times 1/4$  square. One checks that  $t(-, W_p) = p$ ,  $t(\wedge, W_p) = p^2$ , and  $t(\triangle, W_p) = p^3$ , but  $t(\square, W_p) = \frac{9}{8}p^4$ . When  $p \geq 1/2$ , we similarly take

$$\begin{array}{cccc} 2p-1 & 1 & p & p \\ 1 & 2p-1 & p & p \\ p & p & 1 & 2p-1 \\ p & p & 2p-1 & 1 \end{array}$$

One checks that  $t(-, W_p) = p$ ,  $t(\wedge, W_p) = p^2$ , and  $t(\triangle, W_p) = p^3$ , but  $t(\square, W_p) = p^4 + \frac{1}{8}(1-p)^4$ .

### 13.4 Flag algebras and extremal graph theory

Graphons are also useful in extremal graph theory. Consider any inequality involving graph densities. If this inequality holds for all graphons, then it holds for all graphs, since each graph is equivalent to some graphon (from this point of view). On the other hand, if an inequality holds for all graphs then it also holds for all graphons: if it didn't hold for some graphon  $W$ , then with high probability it wouldn't hold for  $G(n, W)$  for large enough  $n$ . What we gain by moving from graphs to graphons is that the space of possible graph densities for graphons can be described using an infinite semidefinite program. If we aim at proving an inequality of the form  $I \geq 0$  (where  $I$  is a linear combination of graph densities), then we can try to compute the minimum of  $I$  given a finite subset of the infinite SDP. In many cases, some finite subset is enough to prove that  $I \geq 0$ . In general, it is known that if indeed  $I \geq 0$  then for every  $\epsilon > 0$ , some finite subset proves  $I > -\epsilon$ ; and that some  $I$  exist for which no finite subset of the constraints implies that  $I \geq 0$ .

When the semidefinite program proves that  $I \geq 0$ , we can translate the proof into a sequence of applications of the Cauchy–Schwartz inequality. We have already seen such an example above, where we proved that  $t(\square, W) \geq t(-, W)^4$ . Although this is not a linear inequality, we can turn it into a linear one by noticing that  $t(-, W)^4 = t(\lll, W)$ . The linear inequality  $t(\square, W) - t(\lll, W) \geq 0$  can be proved automatically by choosing an appropriate subset of the infinite SDP.

The most spectacular demonstration of these techniques is Razborov's theorem [Raz08] on the edges vs. triangles problem. Razborov determined, for each edge density  $p$ , what is the minimal possible triangle density. He also finds the extremal graphons, thus showing how to compute graphs with almost optimal edge and triangle densities.

### 13.5 Graphings

The theory of graphons is appropriate for dense graphs. If, however, we consider a sequence of  $\mathbb{G}(n, d)$  graphs, then it converges to the zero graphon. The correct notion of convergence for bounded-degree graphs, known as *Benjamini–Schramm convergence*, is convergence of the distribution of the  $\ell$ th neighborhood of a random vertex, for all  $\ell$ . As we have shown above, with high probability a random  $d$ -regular graph contains few short cycles (although it is Hamiltonian!), and so locally the graph looks like a tree. In other words, in this case the  $\ell$ th neighborhood is just a  $d$ -regular tree of depth  $\ell$ . Other graph sequences could converge to different distributions. The appropriate limiting object here is called a *graphing*.

The definition of graphing is slightly technical, and so we skip it. Let us just mention a few other examples of random (or non-random) bounded-degree graphs:

- Let  $\pi$  be a probability distribution on  $\{0, \dots, d\}$ . A random  $\pi$ -graph on  $n$  vertices is a graph that looks locally like a tree in which the degree of the root has distribution  $\pi$ .
- The sequence of  $n \times n$  grids converges to the infinite grid.
- The sequence of  $d \times n$  grids (for constant  $d$ ) converges to an infinite grid of height  $d$ .

## 13.6 Permutons

Counterparts of graphons have been studied in other settings. Perhaps the most natural one is that of *permutations*. The counterpart of homomorphism density is the relative order of  $k$  random points. For every (large) permutation  $\tau$  and (small) permutation  $\pi$ , the density  $t(\pi, \tau)$  is the probability that if we sample  $|\pi|$  inputs at random (where  $|\pi|$  is the size of the domain of  $\pi$ ), then their relative order in  $\tau$  is  $\pi$ .

A *permuton* is a probability distribution  $\mu$  on  $[0, 1]^2$  such that if  $(X, Y) \sim \mu$  then  $X \sim U([0, 1])$  and  $Y \sim U([0, 1])$ . To draw a permutation on  $n$  elements, draw  $n$  elements  $(X_1, Y_1), \dots, (X_n, Y_n)$  from  $\mu$ , order them according to the  $X_i$ , and output the relative order of the  $Y_i$ . This is a limit object in the sense that every sequence of permutations has a subsequence converging to a permuton, where convergence is defined via the permutation densities defined above.

We say that a permutation  $\tau$  is  $k$ -uniform if  $t(\pi, \tau) = 1/k!$  for every  $\pi \in S_k$ ; this concept generalizes to permutons. A permuton is *uniform* if it is  $k$ -uniform for all  $k$ . For example,  $U([0, 1]^2)$  is a uniform permuton. If a convergent sequence of permutations  $\tau_n$  satisfies  $t(\pi, \tau_n) \rightarrow 1/k!$  for every  $\pi \in S_k$ , then their limit is a  $k$ -uniform permuton.

In the context of graphs, for a sequence of graphs to be  $p$ -quasirandom it sufficed for it to be  $p$ -quasirandom with respect to two graphs:  $-$  and  $\square$ . Is the same true for permutons?

A 3-uniform permuton need not be uniform, as shown by Copper and Petrarca [CP08]. For example, the permutation 349852167 is 3-uniform but not 4-uniform. Kral' and Pikhurko [KP13] showed that a 4-uniform permuton is uniform. Their proof resembles the argument for graphs.

## References

- [Abb16] Emmanuel Abbé. Community detection and stochastic block models: recent developments. [http://princeton.edu/~eabbe/publications/sbm\\_jmlr\\_2.pdf](http://princeton.edu/~eabbe/publications/sbm_jmlr_2.pdf), 2016.
- [AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3–4):457–466, 1998.
- [AS16] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc, 4th edition, 2016.
- [AV11] Brendan P. W. Ames and Stepehn A. Vavasis. Nuclear norm minimization for the planted clique and biclique problems. *Mathematical Programming, Series B*, 129:68–89, 2011.
- [BHK<sup>+</sup>16] Boaz Barak, Samuel B. Hopkins, Jonathan Kelner, Pravesh K. Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *Proceedings of the fifty-seventh annual IEEE symposium on Foundations of Computer Science (FOCS '16)*, pages 428–437, 2016.
- [BK97] Jean Bourgain and Gil Kalai. Influences of variables and threshold intervals under group symmetries. *Geometric and Functional Analysis*, 7(3):438–461, 1997.
- [CGW89] Fan R. K. Chung, Ronald L. Graham, and Richard M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.
- [Coh16] Michael B. Cohen. Ramanujan graphs in polynomial time. In *FOCS 2016*, pages 276–280, 2016.
- [CP08] Joshua N. Cooper and Andrew Petrarca. Symmetric and asymptotically symmetric permutations. arXiv:0801.4181, 2008.
- [DGGP14] Yael Dekel, Ori Gurel-Gurevich, and Yuval Peres. Finding hidden cliques in linear time with high probability. *Combinatorics, Probability and Computing*, 23(1):2949, 2014.
- [DM15] Yash Deshpande and Andrea Montanari. Finding hidden cliques of size  $\sqrt{N/e}$  in nearly linear time. *Foundations of Computational Mathematics*, 15(4):1069–1128, 2015.
- [Ell] David Ellis. The expansion of random regular graphs. <https://www.dpmms.cam.ac.uk/~dce27/randomreggraphs3.pdf>.

- [ER60] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61, 1960.
- [FGR<sup>+</sup>13] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing (STOC '13)*, pages 655–664, 2013.
- [FK81] Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.
- [FK96] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124:2993–3002, 1996.
- [FK00] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures & Algorithms*, 16(2):195–208, 2000.
- [FK16] Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.
- [FR10] Uriel Feige and Dana Ron. Finding hidden cliques in linear time. In *AOFA*, 2010.
- [Fri99] Ehud Friedgut. Sharp thresholds of graph properties, and the  $k$ -SAT problem. *Journal of the American Mathematical Society*, 12(4):1017–1054, 1999.
- [Fri08] Joel Friedman. *A proof of Alon’s second eigenvalue conjecture and related problems*, volume 910 of *Memoirs of the American Mathematical Society*. American Mathematical Society, 2008.
- [Hor08] Joshua Horowitz. Zero-one laws, random graphs, and Fraïssé limits. <http://web.mit.edu/joshuah/www/projects/zeroone.pdf>, 2008.
- [JLR00] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., 2000.
- [KKL88] Jeff Kahn, Gil Kalai, and Nati Linial. The influence of variables on Boolean functions. In *Proceedings of the 29th annual Symposium on Foundations of Computer Science (SFCS '88)*, pages 68–80, 1988.
- [KKM<sup>+</sup>16] Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D. Pfister, Eren Şaşoğlu, and Rüdiger Urbanke. Reed–Muller codes achieve capacity on erasure channels. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing (STOC '16)*, pages 658–669, 2016.
- [KM10] Achim Klenke and Lutz Mattner. Stochastic ordering of classical discrete distributions. *Advances in Applied Probability*, 42:392–410, 2010.
- [KP13] Daniel Král’ and Oleg Pikhurko. Quasirandom permutations are characterized by 4-point densities. *Geometric and Functional Analysis*, 23(2):570–579, 2013.
- [Kuĉ95] Luďek Kuĉera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2–3):193–212, 1995.
- [LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
- [Lov12] László Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012.
- [MP10] Peter Mörters and Yuval Peres. *Brownian motion*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010.
- [Pud15] Doron Puder. Expansion of random graphs: New proofs, new results. *Inventiones Mathematicae*, 201(3):845–908, 2015.

- [Raz08] Alexander A. Razborov. On the minimal density of triangles in graphs. *Combinatorics, Probability and Computing*, 17(4):603–618, 2008.
- [Rei08] Omer Reingold. Undirected connectivity in log-space. *Journal of the ACM*, 55(4):1–24, 2008.
- [Ros82] Joseph G. Rosenstein. *Linear orderings*. Academic Press, 1982.
- [Spe01] Joel Spencer. *The strange logic of random graphs*. Springer Verlag, 2001.