# LECTURE NOTES: RANDOM GRAPHS AND PERCOLATION THEORY

MATTHEW KAHLE

# INTRODUCTION

These notes are based on a graduate course "Random graphs and percolation theory" at Ohio State University in Fall 2012. The notes were typed by several participants in the course. In particular, many thanks go to: Huseyin Acan, Charles Baker, Chris Kennedy, Jung Eun Kim, Yoora Kim, Greg Malen, Fatih Olmez, Kyle Parsons, Dan Poole, and Younghwan Son.

Disclaimer: we have not had a chance to carefully edit all these notes yet, so please proceed with caution!

Date: January 1, 2013.

### 1. (WEDNESDAY, AUGUST 22)

This course will be about various types of random graphs. In today's class, we partly motivate this study with a discussion of "the probabilistic method". This method is essential in many areas of mathematics now, but Ramsey theory is an important classical one.

The function R(s,t) is defined to be the smallest number N such that for every coloring of the edges of the complete graph  $K_n$  with red and blue, there is either a red  $K_s$  or a blue  $K_t$  subgraph. For example R(3,3) = 6. For large s and t there is little hope for an exact formula for R(s,t), but we might at least hope to understand the asymptotics.

An easy induction argument shows that  $R(s,t) \leq R(s-1,t) + R(s,t-1)$ . Together with the fact that R(s,1) = R(1,t) = 1, this gives by induction that

$$R(s,t) \le \binom{s+t}{s}.$$

**Exercise 1.1** (easy). Show that this implies that  $R(s,s) \leq 4^s$ . Better yet, show that  $R(s,s) = O(4^s/\sqrt{s})$ .

**Exercise 1.2** (open). Improve the base of the exponent 4, or else show that it can not be improved. E.g. prove or disprove that  $R(s,s) = O(3.999^s)$ .

The best upper bound I know of is due to David Conlon [4].

Theorem 1.3 (Conlon, 2009). There exists a contain c such that

$$R(s+1,s+1) \le s^{-c\log s/\log\log s} \binom{2s}{s}.$$

For lower bounds we apply the probabilistic method. Taking a random coloring of the edges gives the following bound.

Theorem 1.4 (Erdős). If

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

then R(k,k) > n.

The proof is a simple union bound.

**Exercise 1.5.** Show as a corollary that

$$R(s,s) \ge \sqrt{2}^{s}$$

The proof is easy but the technique is still powerful. In fact the following is wide open.

**Exercise 1.6** (open). Give an explicit description of any sequence of graphs which gives an exponential lower bound on R(s, s).

This can be rephrased as follows.

**Exercise 1.7** (open). Give an explicit infinite sequence of graphs  $\{G_n\}_{n=1}^{\infty}$ , so that  $G_n$  has n vertices and such that the largest clique and the largest independent set in  $G_n$  are both of order  $O(\log n)$ .

It seems strange that although asymptotically almost every graph according to the measure G(n, 1/2) has this property, <sup>1</sup> it is hard to exhibit any explicit examples. This is sometimes referred to as the problem of finding hay in a haystack.

**Exercise 1.8** (open). Improve the base of the exponent  $\sqrt{2}$ , or else show that it can not be improved. E.g. prove or disprove that

$$R(s,s) = \Omega\left(1.415^s\right).$$

The main source for this lecture was Chapter 1 of Alon and Spencer's book [1].

<sup>&</sup>lt;sup>1</sup>In the next lecture we will define the binomial random graph G(n, p).

### 2. (FRIDAY, AUGUST 24)

**Definition 2.1.** For a natural number n and  $0 \le p \le 1$ , an Erdős-Rényi graph G(n,p) is defined to be a graph on n vertices where each pair of vertices is joined by an edge with probability p, with the edges chosen jointly and independently.

Note that if G is any particular graph on n vertices, then the probability of obtaining G is  $p^{e_G}(1-p)^{\binom{n}{2}-e_G}$ .

**Definition 2.2.** For  $0 \le m \le {n \choose 2}$ , the Erdős-Rényi graph G(n,m) is a graph on n vertices with m edges, chosen among all  $\binom{\binom{n}{2}}{m}$  such graphs uniformly.

Each definition has its advantages and disadvantages; one particular advantage of G(n, p) is that all edges are chosen independently, so local phenomena are modeled more easily than in G(n, m).

Very roughly, G(n,m) resembles G(n,p) with  $p = m/\binom{n}{2}$ . Conversely, G(n,p)resembles G(n,m) with  $m = p \cdot \binom{n}{2}$ .

We will use the following notation extensively. Let f(n), g(n) be sequences of positive real numbers.

- f = O(g) iff there exists a constant M > 0 such that  $f(n) \leq Mg(n)$  for all sufficiently large n; we also write  $g = \Omega(f)$ .
- f = o(g) iff  $\lim_{n\to\infty} f(n)/g(n) = 0$ ; we also write  $f \ll g, g \gg f$ , or  $q = \omega(f).$
- $f = \Theta(g)$  iff f = O(g) and g = O(f); we also write  $f \asymp g$ .
- $f \sim g$  iff  $\lim_{n \to \infty} f(n)/g(n) = 1$ .

If  $\mathcal{Q}$  is any graph property (e.g. connected, contains a  $K_3$ , etc.), we say G(n,p)has property Q a.a.s. (asymptotically almost surely) or w.h.p. (with high probability) if  $\Pr[G(n,p) \in \mathcal{Q}] \to 1$  as  $n \to \infty$ . For example, we can say the following about connectedness:

- (1) If  $p \gg \frac{\log n}{n}$ , then G(n,p) is connected a.a.s. (2) If  $p \ll \frac{\log n}{n}$ , then G(n,p) is disconnected a.a.s. (3) If  $p \ll n^{-2}$ , then G(n,p) has no edges a.a.s.

In the above statements, we have suppressed the implicit variation of p with n; that is, we might more properly write G(n, p(n)). As another example, we will prove the following proposition as a special case of a more general theorem on subgraph containment later:

**Proposition 2.3.** Let Q be the set of graphs containing  $K_4$  as a subgraph. Then

$$Pr[G(n,p) \in \mathcal{Q}] \to \begin{cases} 0 & \text{if } p \ll n^{-2/3} \\ f(c) & \text{if } p = cn^{-2/3} \\ 1 & \text{if } p \gg n^{-2/3}, \end{cases}$$

#### where f(c) is a function independent of p.

We are often interested in graph properties  $\mathcal{Q}$  that are monotone in the following sense: let G and H be graphs on n vertices such that  $G \subseteq H$ . Then  $\mathcal{Q}$  is said to be *monotone increasing* if  $G \in \mathcal{Q}$  implies  $H \in \mathcal{Q}$  (e.g. connected, contains  $K_m$ , contains a cycle). Similarly,  $\mathcal{Q}$  is *monotone decreasing* if  $H \in \mathcal{Q}$  implies  $G \in \mathcal{Q}$ (e.g. k-colorable, maximum degree at most 5).

**Definition 2.4.** Let Q be a monotone increasing graph property and  $\hat{p}$  a sequence of probabilities. Then  $\hat{p}$  is a threshold for Q if

$$Pr[G(n,p) \in \mathcal{Q}] \to \begin{cases} 0 & \text{if } p \ll \hat{p} \\ 1 & \text{if } p \gg \hat{p}, \end{cases}$$

**Theorem 2.5.** Every nontrivial monotone graph property has a threshold.

In the context of the above statement, a property is "trivial" if it holds for every graph or no graph, and nontrivial otherwise. Now, fix a nontrivial monotone increasing property Q. To prove the theorem, the following lemma is useful:

**Lemma 2.6.** If  $p_1 \leq p_2$ , then  $Pr[G(n, p_1) \in \mathcal{Q}] \leq Pr[G(n, p_2) \in \mathcal{Q}]$ .

*Proof.* The proof uses the useful notion of "2-round sprinkling": if G(n,p) and G(n,q) are chosen independently on [n], then their union is G(n, p + q - pq) by inclusion-exclusion. Setting  $p = p_1$  and  $q = \frac{p_2 - p_1}{1 - p_1}$ , view  $G(n, p_2)$  as the union of G(n,p) and G(n,q). Then  $G(n,p_1) \subseteq G(n,p_2)$ , and the conclusion follows since Q is increasing.

**Definition 2.7.** Let  $a \in (0, 1)$ . Then  $p(a) \in (0, 1)$  is defined to be the number such that  $Pr[G(n, p(a)) \in \mathcal{Q}] = a$ .

Note that since  $\Pr[G(n, p) \in \mathcal{Q}]$  is simply a (potentially very complicated) polynomial in p, p(a) exists and is unique, and indeed is continuous and increasing as a function of a. Using this definition, we have the following fact, to be proved next time:

**Proposition 2.8.** Let  $\hat{p}$  be a sequence of probabilities. Then  $\hat{p}$  is a threshold for Q if and only if  $p(a,n) \simeq \hat{p}(n)$  for all  $a \in (0,1)$ .

#### 3. (Monday, August 27)

We began by rehashing some definitions from last time.

**Definition 3.1.** We recall the symbols defining how sequences p(n), q(n) of strictly positive real numbers relate.

- f = O(g) if and only if  $\exists c > 0, N \in \mathbb{N}$  such that n > N implies  $f(n) \leq cg(n)$ .
- $p(n) \ll q(n)$  if and only if p(n) = o(q(n)) if and only if  $\lim_{n \to \infty} \frac{p(n)}{q(n)} = 0$ .
- $p(n) \gg q(n)$  if and only if  $p(n) = \omega(q(n))$  if and only if  $\lim_{n \to \infty} \frac{p(n)}{q(n)} = \infty$ .
- $p(n) \approx q(n)$  if and only if  $p(n) = \Theta(q(n))$  if and only if p(n) = O(q(n))and q(n) = O(p(n)); i.e., there are constants 0 < c < C,  $N \in \mathbb{N}$  such that n > N implies  $cq(n) \leq p(n) \leq Cq(n)$ .

**Definition 3.2.** If Q is a non-trivial, monotone, increasing graph property, then  $\hat{p} = \hat{p}(n)$  is said to be a threshold for Q if, as  $n \to \infty$ ,

$$\Pr[G(n, p(n)) \in Q] \to \begin{cases} 1, & p \gg \hat{p} \\ 0, & p \ll \hat{p} \end{cases}$$

**Definition 3.3.** p(a) = p(a; n) is defined to be the unique number q in (0, 1) such that  $Pr[G(n, q) \in Q] = a$ . (The well-defined-ness of this number more or less follows from the intermediate value theorem, and the probability a polynomial in p: For fixed  $n, p \mapsto Pr[G(n, p) \in Q]$  is continuous and strictly increasing.)

3.1. Every monotone graph property has a threshold. We now wish to show that every increasing (nontrivial) graph property has a threshold. To start, we now clarify the proof of the fact given at the end of class on Friday, Aug. 24th, 2012.<sup>2</sup>

**Lemma 3.4.**  $\hat{p}$  is a threshold if and only if  $\hat{p} \simeq p(a)$  for any fixed a, 0 < a < 1.

*Proof.* ("only if"): Assume  $\hat{p}$  is a threshold. If 0 < a < 1 but  $\hat{p} \not\geq p(a)$ , then there exists a subsequence  $n_1, n_2, \ldots$  along which  $\frac{p(a)}{\hat{p}} \to 0$  or  $\frac{p(a)}{\hat{p}} \to \infty$ .

In the first case, as  $k \to \infty$ ,  $\frac{p(a; n_k)}{\hat{p}(n_k)} \to 0$ . Extend the subsequence  $p(a; n_k)$  to any full sequence q(n) such that  $q(n_k) = p(a; n_k)$  and that  $\frac{q(n)}{\hat{p}(n)} \to 0$  as  $n \to \infty$ .<sup>3</sup> Therefore, by the definition of a threshold,

$$\Pr[G(n,q(n)) \in Q] \to 0 \text{ as } n \to \infty,$$

 $<sup>^{2}</sup>$ There was no real error in the proof from last time; only one observation fixes the proof.

<sup>&</sup>lt;sup>3</sup>For example, for any  $n \neq n_k$  for some k, just define  $q(n) = 2^{-n} \hat{p}(n)$ .

and hence, since subsequential limits equal the full limits,

$$\Pr[G(n_k, p(a; n_k) \in Q] \to 0 \text{ as } n \to \infty.$$

Yet by definition of p(a),

$$\Pr[G(n_k, p(a; n_k) \in Q] = a \forall k \in \mathbb{N}.$$

Contradiction. The second possibility yields a similar contradiction. Therefore, we must have  $\hat{p} \simeq p(a)$  for all a, 0 < a < 1.

("if"): Assume  $\hat{p}$  is not a threshold. Then there exists a sequence p = p(n) such that  $\frac{p}{\hat{p}} \to 0$  and  $\liminf \Pr[G(n,p) \in Q] \geqq 0$ , or p exists such that  $\frac{p}{\hat{p}} \to \infty$  and  $\limsup \Pr[G(n,p) \in Q] \leqq 0$ .

In the first case, there exists a > 0 and a subsequence  $n_1 < n_2 < \cdots$  along which  $\Pr[G(n,p) \in Q] \ge a$ , so that by definition of p(a),  $p(a;n_k) \le p(n_k)$ .<sup>4</sup> Thus, since  $p \ll \hat{p}$ , for all c > 0, there exists  $N \in \mathbb{N}$  such that n > N implies  $p(n) < c\hat{p}$ . Hence, for k > N,  $n_k > N$ , so that  $p(a;n_k) < c\hat{p}(n_k)$ . Therefore,  $\hat{p}$  is not O(p(a)), so  $\hat{p} \neq p(a)$ .

The second case proceeds similarly.

#### **Theorem 3.5.** Every nontrivial monotone increasing property has a threshold.

The proof method relies on an extension of the 2-fold sprinkling process began last time to an m-fold sprinking process.

**Proof.** Fix  $\epsilon$ ,  $0 < \epsilon < \frac{1}{2}$ . Choose  $m = m(\epsilon)$  large enough that  $(1 - \epsilon)^m < \epsilon$ . Fix n temporarily, and take m independent copies of  $G(n, p(\epsilon))$ , named  $G_1, \ldots, G_m$ . By m-fold sprinkling and the inclusion-exclusion principle, their union (literally superimposing the graphs, and identifying common edges) is a G(n, p'), where  $p' = 1 - (1 - p(\epsilon))^m \le mp(\epsilon)$ . Therefore, we have that

$$\Pr[G_1 \cup \ldots \cup G_m \in Q] = \Pr[G(n, p') \in Q] \le \Pr[G(n, mp(\epsilon)) \in Q].$$

Yet by the independence of the choices, and since  $\Pr[G_i \in Q] = \Pr[G(n, p(\epsilon)) \in Q] = \epsilon$ , we have that

$$\Pr[G_1 \cup \ldots \cup G_m \notin Q] = \prod_{i=1}^m \Pr[G_i \notin Q] = (1-\epsilon)^m < \epsilon,$$

and hence

$$\Pr[G(n, mp(\epsilon)) \in Q] \ge 1 - \Pr[G_1 \cup \ldots \cup G_m \notin Q] \ge 1 - \epsilon.$$

Therefore, by definition of p(a),  $p(1-\epsilon) \leq mp(\epsilon)$ , since  $p \to \Pr[G(n,p) \in Q]$  is increasing in p by Q an increasing graph property. By the monotonicity of p(a),

<sup>&</sup>lt;sup>4</sup>This depends in the monotonicity of  $\Pr[G(n,p) \in Q]$  on p, proved last time.

however, we have

$$p(\epsilon) \le p\left(\frac{1}{2}\right) \le p(1-\epsilon) \le mp(\epsilon).$$

Note also that m depends on  $\epsilon$ , and not on n. Therefore, the same inequalities hold for every m, and hence

$$p(\epsilon) \asymp p\left(\frac{1}{2}\right) \asymp p(1-\epsilon),$$

for every  $0 < \epsilon < \frac{1}{2}$ . Therefore, by the Lemma,  $\hat{p} = p\left(\frac{1}{2}\right)$  is a threshold.

3.2. Sharp Thresholds. We next discussed various definitions of a stricter kind of threshold.

**Definition 3.6.**  $\hat{p}$  is said to be a sharp threshold for Q if

$$Pr[G(n,p) \in Q] \to \begin{cases} 1, & p \ge (1+\eta)\hat{p} \text{ for some } \eta > 0\\ 0, & p \le (1-\eta)\hat{p} \text{ for some } \eta > 0 \end{cases}$$

A (non-sharp) threshold, by contrast, requires that p is eventually greater than any constant multiple of  $\hat{p}$  before requiring that the limit goes to 1. Therefore, sharp thresholds are indeed thresholds.

Another characterization of sharp thresholds comes from the concept of the "widths of the windows of change."

**Definition 3.7.** Fix  $0 < \epsilon < \frac{1}{2}$ . Then define the  $2(1 - \epsilon)$ -width  $\delta = \delta(\epsilon) = p(1 - \epsilon) - p(\epsilon)$ .

Small widths imply that a relatively small change in the input edge probability controls whether or not the output probability of a graph having property Q is likely or unlikely. In fact, we have the following fact.

**Lemma 3.8.**  $\hat{p}$  is a sharp threshold if and only if for all  $\epsilon$ ,  $0 < \epsilon < \frac{1}{2}$ ,  $\delta = \delta(\epsilon) = o(\hat{p})$ .

*Proof.* ("only if") Suppose  $\hat{p}$  is a threshold. Then we claim that for all a, 0 < a < 1, that for all  $\eta, 0 < \eta < 1$ , there exists  $N = N(a, \eta)$  such that n > N implies  $(1 - \eta)\hat{p} < p(a) < (1 + \eta)\hat{p}$ .

Suppose, by way of contradiction, that the above does not hold. Then for some a, for some  $\eta > 0$ , either  $p(a) \leq (1 - \eta)\hat{p}$  infinitely often or  $(1 + \eta)\hat{p} \leq p(a)$  infinitely often. In the first case, we have an infinite subsequence  $n_1 < n_2 < \cdots$  such that  $p(a; n_k) \leq (1 - \eta)\hat{p}(n_k)$ . Re-extend  $p(a; n_k)$  to a sequence q(n) such that  $q(n_k) = p(a; n_k)$  and that  $q(n) \leq (1 - \eta)\hat{p}(n)$  for all n.<sup>5</sup> Then  $q(n) \leq (1 - \eta)\hat{p}(n)$ ,

<sup>&</sup>lt;sup>5</sup>For example, for any  $n \neq n_k$  for some k, just define  $q(n) = (1 - \eta)\hat{p}(n)$ .

so by the fact that  $\hat{p}$  is a threshold, we have that  $\Pr(G(n,q) \in Q) \to 0$  as  $n \to \infty$ ; in particular,  $\Pr[G(n_k, p(a; n_k) \in Q] = \Pr[G(n_k, q(n_k)) \in Q] \to 0$  as  $k \to \infty$ . Yet  $\Pr[G(n_k, p(a; n_k) \in Q] = a$  by definition, for all k. Hence,  $a \to 0$  as  $k \to \infty$ . Contradiction. The second possibility  $((1 + \eta)\hat{p} \leq p(a)$  infinitely often) also causes a contradiction in the same way. Therefore, the claim holds.

From the claim, and from monotonicity of p(a), it follows that for any a, b, 0 < a < b < 1, we have that for  $n > \max\{N(a, \eta), N(b, \eta)\}$ ,

$$0 < p(b) - p(a) < [(1+\eta) - (1-\eta)]\hat{p} = 2\eta\hat{p}$$

Since  $\eta$  is arbitrary, it is clear that  $p(b) - p(a) = o(\hat{p})$ . Fixing  $0 < \epsilon < \frac{1}{2}$ , and setting  $a = \epsilon$  and  $b = 1 - \epsilon$ , we have that  $\delta(\epsilon) = p(1 - \epsilon) - p(\epsilon) = o(\hat{p})$ . This works for all such  $\epsilon$ .

("if") Exercise.

Another basic fact in this setting shows that if sharp thresholds exist, then  $p\left(\frac{1}{2}\right)$  reprises its role as the universal threshold.

**Lemma 3.9.** (1) If a sharp threshold for a nontrivial, increasing property Q exists, then for all  $\epsilon$ ,  $0 < \epsilon < 1$ ,

(1) 
$$\frac{p(\epsilon)}{p(\frac{1}{2})} \to 1 \text{ as } n \to \infty$$

(2) If for a given nontrivial, increasing property Q the limit in 1 holds for all  $\epsilon, 0 < \epsilon < 1$ , then  $p\left(\frac{1}{2}\right)$  is a sharp threshold.

In particular, if any sharp threshold exists,  $p\left(\frac{1}{2}\right)$  is also a sharp threshold.

*Proof.* Suppose  $\hat{p}$  is a sharp threshold for a nontrivial, increasing property Q. First, take  $\epsilon$  with  $0 < \epsilon < \frac{1}{2}$ . Then by the previous lemma, for any c > 0, for  $n > \max\{N(\epsilon, c), N(1 - \epsilon, c\},$ 

$$p\left(\frac{1}{2}\right) - p(\epsilon) \le p(1-\epsilon) - p(\epsilon) \le c\hat{p}.$$

Dividing both sides by  $p\left(\frac{1}{2}\right)$  and moving some terms around, we get that

$$1 - c\frac{\hat{p}(n)}{p\left(\frac{1}{2}\right)} \le \frac{p(\epsilon)}{p\left(\frac{1}{2}\right)}.$$

Yet since  $\hat{p}$  is a sharp threshold, it is a regular threshold, and hence  $\hat{p} \approx p\left(\frac{1}{2}\right)$ . In particular, then, there exists a constant k such that  $\frac{\hat{p}(n)}{p\left(\frac{1}{2}\right)} \leq k$ . Therefore, for large

enough n,

$$1 - ck \le 1 - c\frac{\hat{p}(n)}{p\left(\frac{1}{2}\right)} \le \frac{p(\epsilon)}{p\left(\frac{1}{2}\right)}.$$

Since k is fixed and c is arbitrary, this shows that  $\liminf \frac{p(\epsilon)}{p(\frac{1}{2})} \ge 1$ . Yet  $\epsilon < \frac{1}{2}$ , so  $p(\epsilon) \le p(\frac{1}{2})$  and hence  $\limsup \frac{p(\epsilon)}{p(\frac{1}{2})} \le 1$ .

The case  $\frac{1}{2} < \epsilon < 1$  is similar, and of course the case  $\epsilon = \frac{1}{2}$  is trivial.

Now, suppose that Q is a nontrivial, increasing graph property such that for all  $\epsilon$ ,  $0 < \epsilon < 1$ , 1 holds. Therefore, for all C > 1, there is  $M = M(\epsilon, C)$  such that n > M implies

$$\frac{1}{C}p\left(\frac{1}{2}\right) < p(\epsilon) < Cp\left(\frac{1}{2}\right).$$

Suppose that for some  $\eta > 0$ ,  $q(n) \ge (1 + \eta)p(\frac{1}{2})$  for all n. Then for  $C := 1 + \eta$ , and fixing  $\epsilon \in (0, 1)$ ,  $n > M = M(\epsilon, C)$ , the above gives

$$q(n) \ge (1+\eta)p\left(\frac{1}{2}\right) = Cp\left(\frac{1}{2}\right) > p(\epsilon),$$

and hence, by the monotonicity of  $\Pr[G(n, p) \in Q]$  in p proved last time, for n > M,

 $\Pr[G(n,q)\in Q]\geq \Pr[G(n,p(\epsilon)\in Q]=\epsilon.$ 

Since this works for all  $\epsilon \in (0, 1)$ , we have that  $\Pr[G(n, q) \in Q] \to 1$  as  $n \to \infty$ .

Similarly, if for some  $\eta > 0$ ,  $q(n) \le (1 - \eta)p(\frac{1}{2})$ , then  $\Pr[G(n, q) \in Q] \to 0$  as  $n \to \infty$ .

3.3. An example. We now apply our work on thresholds to a common choice of graph property. We begin today and continue on Wednesday, Aug. 29th.

**Exercise 3.10.** One nontrivial, increasing graph property is the existence of subgraphs isomorphic to complete graphs of a specified size. To set notation, let a graph G be in  $Q_m$  if and only if G contains a subgraph isomorphic to  $K_m$ , the complete graph on m vertices. Let  $N_m$  be the random variable counting the number of  $K_m$ 's in a given graph; then  $G \in Q_m$  if and only if  $N_m(G) \ge 0$ .

Our first claim is that for fixed n, and taking expectation over the space of graphs

$$G(n,p), \mathbb{E}[N_m] = \binom{n}{m} p^{\binom{m}{2}}$$
. Note that for fixed n, there are only finitely many

distinct m-tuples of vertices with which to create a  $K_m$ . Therefore, if  $i \in {\binom{[n]}{m}}$  is an m-tuple of vertices of G(n), then let the set  $A_i$  be the subset of graphs in G(n, p)in which the m-tuple of vertices i creates a  $K_m$ . Since individual edges exist with

probability p, and since  $\binom{m}{2}$  edges are needed to form  $K_m$ , we have for each i that  $Pr(A_i) = p \binom{m}{2}$ . For notational convenience, let  $Y_i = \mathbb{K}_{A_i}$  be the indicator function of  $A_i$ ; then  $\mathbb{E}Y_i = p \binom{m}{2}$ . Further, since  $N_m$  is the total number of  $K_m$ 's, and each possible  $K_m$  is indexed by some  $i \in \binom{m}{2}$ , we have that  $N_m$  is merely the sum of the indicator variables  $Y_i: N_m = \sum_{i \in \binom{[n]}{m}} Y_i$ . Since expectation is linear, the expectation operator passes  $i \in \binom{[n]}{m}$ 

through the finite sum indicated, so we have that

$$\mathbb{E}(N_m) = \mathbb{E}\left[\sum_{\substack{i \in \binom{[n]}{m}}{Y_i}} Y_i\right]$$
$$= \sum_{\substack{i \in \binom{[n]}{m}}{p}} \mathbb{E}(Y_i)$$
$$= \sum_{\substack{i \in \binom{[n]}{m}}{p}} p\binom{m}{2} = \binom{n}{m} p\binom{m}{2}$$

Thus, the claim is proved.

**Remark 3.11.** The above work demonstrated that the summability of expectations happens despite the lack of independence of some pairs of subsets; for example, if  $m = 4 \text{ and } n \ge 5$ , the existence of  $K_4$ 's on the vertex sets  $\{1, 2, 3, 4\}$  and  $\{1, 2, 3, 5\}$ are certainly not independent. We resolve this issue in two ways. First, we note that independence of random variables X, Y implies that the expectation of their product is 0:  $\mathbb{E}(XY) = 0$ . Thus, independence of dependence of random variables does not matter until we start studying variances and the second-moment method. Second, we note that if m is fixed and n is large, then there are many pairwise independent events: e.g., in the m = 4 example above, the existence of  $K_4$ 's on the vertex sets  $\{1, 2, 3, 4\}$ ,  $\{5, 6, 7, 8\}$ ,  $\{9, 10, 11, 12\}$ , are independent. The heuristic is that if n is large relative to m, then "most" pairs are independent.

Exercise 3.12. Continuing the example above, we also note that by Stirling's

formula, if m is fixed, then  $\mathbb{E}[N_m] = \binom{n}{m} p^{\binom{m}{2}} \approx n^m p^{\binom{m}{2}}$ . Therefore, if

 $p(n) = n^{\frac{-2}{(m-1)}}$ , then by the fact that  $\binom{m}{2} = \frac{m(m-1)}{2}$ , we have that  $n^m p^{\binom{m}{2}} = \frac{m(m-1)}{2}$ 

 $n^{m-m} = 1$ . Therefore, it is simple to check that if  $p \gg n^{\frac{-2}{m-1}}$ , then  $\mathbb{E}[N_m] \to \infty$ , and if  $p \ll n^{\frac{-2}{(m-1)}}$ , then  $\mathbb{E}[N_m] \to 0$ . From this, we can make the conjecture that a threshold of the graph property  $Q_m$  is in fact  $n^{\frac{-2}{(m-1)}}$ . In fact, we can show the following.

(1) If  $p \ll n^{\frac{-2}{(m-1)}}$ , then  $Pr[G(n,p) \supset K_m] \to 0$  as  $n \to \infty$ . Theorem 3.13.

- (2) If  $p \gg n^{\frac{-2}{(m-1)}}$ , then  $Pr[G(n,p) \supset K_m] \to 1$  as  $n \to \infty$ .
- (3) If  $p = cn^{\frac{-2}{(m-1)}}$ , then  $Pr[G(n,p) \supseteq K_m] \rightarrow f(c)$  as  $n \rightarrow \infty$ , where  $0 < \infty$ f(c) < 1 is a constant depending on c alone, not n or m.

This result demonstrates that  $n^{\frac{-2}{(m-1)}}$  is in fact a threshold, but not a sharp threshold (by specifics of the constant f(c), c can be greater than 1 but  $Pr[G(n,p) \supseteq$  $K_m \not \to 1$ ).

For an example of a sharp threshold, we consider the property of connectedness.

(1) If  $p \ge \frac{\log(n) + \omega(1)}{n}$ , then  $\Pr[G(n, p) \text{ is connected}] \to 1$ . Theorem 3.14. (2) If  $p \leq \frac{\log(n) - \omega(1)}{n}$ , then  $\Pr[G(n, p) \text{ is connected}] \to 0$ . (3) If  $p = \frac{\log(n) + c}{n}$ , then  $\Pr[G(n, p) \text{ is connected}] \to e^{-e^{-c}}$ .

Recall that  $q = \omega(1)$  means that some function is diverging, however slowly, to infinity; in particular,  $\eta \log(n)$  is a reasonable choice (showing that we have a sharp threshold), but  $\log(\log(\log(\cdots \log(n) \cdots)))$  also works. The width of this window is slightly larger than  $\frac{2}{n}$ , since you have to allow slow-growing functions.

(Our discussion of thresholds followed Section 1.5 of [7].)

# 4. Wednesday, August 29

We will use the second moment method to prove that the threshold function for G(n, p) includes  $K_4$  is  $p = n^{-2/3}$ .

# Definition 4.1.

$$Var[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Notation  $\sigma^2 = Var[X]$ , where  $\sigma$  is the standard variation.

**Theorem 4.2** (Chebyshev's Inequality). Let  $\mu := \mathbb{E}[X]$ . Then,

$$\Pr[|X - \mu| \ge \lambda\sigma] \le \frac{1}{\lambda^2}$$

Proof. We have

$$\sigma^2 = Var[X] = \mathbb{E}[(X - \mu)^2] \ge \lambda^2 \sigma^2 \Pr[|X - \mu| \ge \lambda \sigma].$$

Dividing by  $\sigma^2 \lambda^2$  we get

$$\frac{1}{\lambda^2} \ge \Pr[|X - \mu| \ge \lambda\sigma]$$

Suppose  $X = X_1 + \cdots + X_m$ . Then,

$$Var[X] = \sum_{i=1}^{m} Var[X_i] + \sum_{i \neq j} Cov[X_i, X_j]$$

where

$$Cov[X_i, X_j] := \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]$$

If  $X_i$  and  $X_j$  are independent r.v.'s, then  $Cov[X_i, X_j] = 0$ . The converse of this statement is not true. For a counterexample, let Y be the uniform random variable on the interval [-1, 1] and let  $Z = Y^2$ . Clearly Y and Z are not independent but Cov[Y, Z] = 0.

If  $X = X_1 + \cdots + X_m$  where each  $X_i$  is a Bernoulli r.v. with

$$\Pr[X_i = 1] = p_i, \quad \Pr[X_i = 0] = 1 - p_i,$$

then  $Var[X] = p_i(1 - p_i) \le p_i$ . SO,

$$Var[X] \le \mathbb{E}[X] + \sum_{i \ne j} Cov[X_i, X_j].$$

In general, if X is a non-negative integer r.v., then

$$\Pr[X > 0] = \sum_{i \ge 1} \Pr[X = i] \le \sum_{i \ge 1} i \Pr[X + 1] = \mathbb{E}[X].$$

This is a special case of the following theorem.

Theorem 4.3 (Markov's Inequality).

$$Pr[X \ge a] \le \frac{\mathbb{E}[X]}{a}$$

for any non-negative r.v. X.

In particular, if  $X_n$  is a sequence of non-negative integer r.v.'s such that  $\mathbb{E}[X_n] \to 0$ , then  $X_n = 0$  a.a.s. In other words,

$$\Pr[X_n = 0] \to 1 \text{ as } n \to \infty.$$

What if  $\mathbb{E}[X_n] \to \infty$ ? Is it true that  $X_n > 0$  a.a.s.? **Example** Let  $X_n$  be a sequence of r.v.'s such that

$$X_n = \begin{cases} e^n & \text{with probability } \frac{1}{n} \\ 0 & \text{with probability } 1 - \frac{1}{n} \end{cases}$$

Then, the expected value  $\mathbb{E}[X_n] = e^n/n \to \infty$  but  $\Pr[X_n > 0] = 1/n \to 0$ .

**Theorem 4.4.** Let X be a non-negative integer valued r.v. Then,

$$Pr[X=0] \le \frac{Var[X]}{\mathbb{E}[X]^2}.$$

*Proof.* Set  $\lambda = \mu/\sigma$ . Then,

$$\Pr[X=0] \le \Pr[|X-\mu| \ge \mu] = \Pr[|X-\mu| \ge \lambda\sigma] \le \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2}.$$

**Corollary 4.5.** If  $\mathbb{E}[X] \to \infty$  and  $Var[X] = o(\mathbb{E}[X]^2)$ , then X > 0 a.a.s.

the same proof shows that

$$\Pr[|X - \mu| \ge \epsilon \mu] \le \frac{Var[X]}{\epsilon^2 \mathbb{E}[X]^2}.$$

So if  $\mathbb{E}[X] \to \infty$  and  $Var[X] = o(\mathbb{E}[X]^2)$ , then  $X \sim \mathbb{E}[X]$  a.a.s. In other words, for any fixed  $\epsilon > 0$ ,

$$(1-\epsilon)\mathbb{E}[X] \le X \le (1+\epsilon)\mathbb{E}[X]$$

with probability approaching 1 as  $n \to \infty$ .

Now set  $X = X_1 + \cdots + X_m$  where  $X_i$  is the indicator random variable for the event  $A_i$ . Set

$$\Delta = \sum_{i \sim j} \Pr[A_i \text{ and } A_j]$$

where  $i \sim j$  means that  $i \neq j$  and  $A_i, A_j$  are not independent.

In particular we have

$$Cov[X_i, X_j] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i][X_j] \le \mathbb{E}[X_i X_j] = \Pr[A_i \text{ and } A_j],$$

so  $Var[X] \leq \mathbb{E}[X] + \Delta$ .

**Corollary 4.6.** If  $\mathbb{E}[X] \to \infty$  and  $\Delta = o(\mathbb{E}[X]^2)$ , then X > 0 a.a.s. (Actually  $X \sim \mathbb{E}[X]$  a.a.s).

**Theorem 4.7.**  $n^{-2/3}$  is a threshold for  $G(n,p) \supset K_4$ .

*Proof.* Let  $X = X_n$  be the number of  $K_4$ 's in G(n, p). Previously we have seen that  $\mathbb{E}[X_n] = \binom{n}{4}p^6 \approx n^4p^6$ .

(1) If  $p \ll n^{-2/3}$ , then  $\mathbb{E}[X] \to 0$ , so X = 0 a.a.s.

(2) If  $p \gg n^{-2/3}$  then  $\mathbb{E}[X] \to \infty$ . To apply second moment method we need to compute  $\Delta$ . For  $i \in {\binom{[n]}{4}}$ , let  $A_i$  be the event that "i spans a clique". Then,

$$\Pr[A_i] = p^6$$

The events  $A_i$  and  $A_j$  are not independent iff  $|i \cap j| = 2$  or 3. Let  $\Delta = \Delta_2 + \Delta_3$ where  $\Delta_2$  is the contribution of the pairs (i, j) with  $|i \cap j| = 2$  and  $\Delta_3$  is the contribution of the pairs (i, j) with  $|i \cap j| = 3$ . Then

$$\Delta_2 \asymp n^6 p^{11}, \quad \Delta_3 \asymp n^5 p^9.$$

On the other hand,

$$\mathbb{E}[X]^2 \asymp n^8 p^{12}.$$

We conclude the proof by noting that,

$$\frac{\Delta_2}{\mathbb{E}[X]^2} \asymp \frac{n^6 p^{11}}{n^8 p^{12}} = \frac{1}{n^2 p} \to 0,$$

and

$$\frac{\Delta_3}{\mathbb{E}[X]^2} \asymp \frac{n^5 p^9}{n^8 p^{12}} = \frac{1}{n^3 p^3} \to 0$$

as  $n \to \infty$ .

Exercise 4.8.	What is a	threshold	for $G(n, p)$	contains $K_m$	for fixed $m$ ?
<b>L</b> MOI 0100 1101	<i>iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii</i>	010100100000	101 0110, 01		101 100000 110.

**Exercise 4.9.** Is it always true that threshold for "contains H" is  $n^{-v_H/e_H}$  for any fixed graph H?

#### 5. FRIDAY, AUGUST 31

**Question 5.1.** Is  $n^{-v_H/e_H}$  always a threshold for the property containment of H in G(n, p) as a subgraph, i.e.  $G(n, p) \supseteq H$ ?

The answer is no. Note that when  $H = K_4$ , it is shown in previous lectures that  $n^{-4/6} = n^{-2/3}$  is a threshold.

Consider  $H = K_4 \cup \{\text{extra edge}\}$ . Is  $n^{-5/7}$  a threshold? If  $p \ll n^{-5/7}$ , then  $\mathbb{E}[X_H] \to 0$  and hence  $X_H = 0$  a.a.s. If  $p \gg n^{-5/7}$ , p can still be much less than the threshold for containment of a  $K_4$ , since  $n^{-2/3} \gg n^{-5/7}$ . So if  $p \gg n^{-5/7}$ , it may even be that  $X_{K_4} = 0$  a.a.s. (eg.  $p = n^{-20/29}$ .) Therefore  $n^{-5/7}$  is not the threshold for containment of H.

**Definition 5.2.** p(H) is defined to be  $e_H/v_H$  and called the density of H. H is said to be balanced if  $p(H') \leq p(H)$  for every subgraph of  $H' \subseteq H$ .

**Proposition 5.3.** Threshold of containment of H is  $n^{-v_H/e_H} \iff H$  is balanced.

*Proof.* [Following chapter 4 of [1].]

 $\implies$ : Suppose  $n^{-v_H/e_H}$  is a threshold for containment of H and assume to the contrary that H is not balanced. Then there exists a subgraph H' with  $p(H') = \frac{e_{H'}}{v_{H'}} > \frac{e_H}{v_H} = p(H)$ . Choose  $\alpha$  such that  $p(H') < \alpha < p(H)$  and set  $p = n^{-\alpha}$ . Then  $p \ll n^{-p(H')}$  which implies  $\mathbb{E}(\# \text{ of } H' \text{ subgraphs}) \to 0$ . So a.a.s. there are no H' subgraphs in G(n, p) and hence no H subgraphs.

 $\Rightarrow$ : We need a couple of facts for this direction.

Take  $X = \sum_{i=1}^{m} X_i = \sum_{i=1}^{m} 1_{A_i}$ , that is each  $X_i$  is 1 on an event  $A_i$  and 0 otherwise. Define  $\Delta = \sum_{i \sim j} \Pr(A_i \cap A_j)$ , where  $i \sim j$  is the same as  $i \neq j$  and the events  $A_i$  and  $A_j$  are not independent. It was showed in previous lectures that

 $\mathbb{E}(X) \to \infty \text{ and } \Delta = o(\mathbb{E}(X)^2) \implies X > 0 \text{ and } X \sim \mathbb{E}(X) \text{ a.a.s.}$ 

Define  $X_1, X_2, \ldots, X_m$  to be symmetric random variables when there exists a measure preserving automorphism of underlying probability space taking  $X_i$  to  $X_j$  for every i, j.

$$\Delta = \sum_{i \sim j} \Pr(A_i \cap A_j) = \sum_i \Pr(A_i) \sum_{j \sim i} \Pr(A_j | A_i).$$

Notice that when  $X_1, X_2, \ldots, X_m$  are symmetric,  $\sum_{j \sim i} \Pr(A_j | A_i)$  doesn't depend on *i*, so call it  $\Delta^*$ . Then  $\Delta = \Delta^* \cdot \mathbb{E}(X)$ , which implies the following corollary:

**Corollary 5.4.** Let  $X_1, X_2, \ldots, X_m$  be symmetric indicator random variables of  $A'_i$ s and  $X = \sum_{i=1}^m X_i$ . Then

$$\mathbb{E}(X) \to \infty \ and \ \Delta^* = o(\mathbb{E}(X)) \implies X > 0 \ and \ X \sim \mathbb{E}(X) \ a.a.s.$$

Now we are ready to prove the other direction.

For each V-set  $S \in {\binom{[n]}{V}}$ , let  $A_S$  be the event that  $G|_S$  contains H-subgraph and  $X_S = 1_{A_S}$ . Note that  $p^v \leq \Pr(A_S) \leq v! p^v$ .

Now set  $X = \sum_{S \in \binom{[n]}{V}} X_S$ . Then X > 0 if and only if there exists at least one *H*-subgraphs. Note that *X* doesn't give the exact count of *H*-subgraphs.

$$\mathbb{E}(X) = \sum \mathbb{E}(X_S) = \binom{n}{v} \Pr(A_S) = n^v p^e.$$

So if  $p \ll n^{-v/e}$ , then X = 0 a.a.s. Suppose  $p \gg n^{-v/e}$ . Let's compute  $\Delta^* = \sum_{T \sim S} \Pr(A_T | A_S)$ . Notice that here  $T \sim S$  also means that  $T \neq S$  and S, T share a common edge. Fix S.

$$\Delta^* = \sum_{i=2}^{v-1} \sum_{|S \cap T|=i} \Pr(A_T | A_S).$$

For each *i*, there are  $O(n^{v-i})$  choices for *T*. Fix S, T with  $|T \cap S| = i$ . Let's compute  $\Pr(A_T|A_S)$ . There are O(1) copies of *H* on *T*. Each has at least  $i\frac{e}{v}$  edges on vertices of *S*, since *H* is balanced. This leaves at least  $e - i\frac{e}{v}$  edges outside of *S*. Then

$$\Delta^* = \sum_{i=2}^{v-1} O(n^{v-1} p^{e-i\frac{e}{v}}) = \sum_{i=2}^{v-1} O((n^v p^e)^{1-\frac{i}{v}}) = \sum_{i=2}^{v-1} o(n^v p^e).$$

Corollary applies. X > 0 a.a.s.

**Theorem 5.5.** *H* is said to be strictly balanced if p(H') < p(H) for every subgraph *H'*. Let *H* have *v* vertices, *e* edges and a automorphisms. Let *X<sub>H</sub>* be the number of copies of *H* in *G*(*n*,*p*). Then,

$$p \gg n^{-v/e} \implies X_H \approx \frac{n^v p^e}{a}$$

**Theorem 5.6.** If H is any graph and p such that  $\mathbb{E}(X_H) \to \infty$  for every subgraph  $H' \subseteq H$ , then  $X_H \approx \mathbb{E}(X_H)$  a.a.s. and  $X_H > 0$  a.a.s.

6. (Wednesday, September 5)

Let  $G = K_3$  and  $p = \frac{c}{n}$  for  $c \in (0, \infty)$  a constant. We seek the probability that there exists any G subgraph of G(n, p). If we let  $X_G$  be the random variable counting the number of G subgraphs of G(n, p) we'll see that  $\Pr[X_G = m] \to f(m)$ .

**Definition 6.1.**  $z \in P_0(\mu)$  means z is chosen according to a Poisson distribution with mean  $\mu$ . That is  $Pr[z = t] = \frac{\mu^t e^{-\mu}}{t!}$ 

Note that

$$\sum_{t=0}^{\infty} \frac{\mu^t}{t!} e^{-\mu} = e^{-\mu} \sum_{t=0}^{\infty} \frac{\mu^t}{t!} = e^{-\mu} e^{\mu} = 1$$

so  $P_0(\mu)$  is in fact a probability distribution on the nonnegative integers. Also note that

$$\sum_{t=0}^{\infty} t \Pr[z=t] = \sum_{t=1}^{\infty} \frac{\mu^t}{(t-1)!} e^{-\mu} = \mu e^{-\mu} \sum_{t=0}^{\infty} \frac{\mu^t}{t!} = \mu e^{-\mu} e^{\mu} = \mu$$

so the mean of  $P_0(\mu)$  is  $\mu$  as desired.

Now again letting  $G = K_3$ ,  $p = \frac{c}{n}$  and  $X_G$  be the number of  $K_3$  subgraphs of G(n,p) we have that  $\mathbb{E}[X_G] = \binom{n}{3}p^3$  which tends to  $\frac{c^3}{6}$  as  $n \to \infty$ . In fact  $X_G \xrightarrow{D} P_0\left(\frac{c^3}{6}\right)$  (approaches in distribution). That is  $\Pr[X_G = t] \to \frac{\mu^t e^{-\mu}}{t!}$  as  $n \to \infty$  where  $\mu = \frac{c^3}{6}$ . In particular the probability that there are no  $K_3$  subgraphs is approaching  $e^{-c^3/6}$ .

Note that if all  $\binom{n}{3}$  events were independent, then we would have  $\Pr[X_G = 0] = (1 - p^3)^{\binom{n}{3}} \approx e^{-p^3 n^3/6} \rightarrow e^{-c^3/6}$ . We will next apply the method of (factorial) moments also know as Brun's sieve.

**Theorem 6.2.** Suppose  $X = X_n$  is a distribution on nonnegative integers such that  $E[X] \to \mu$  as  $n \to \infty$  and for every fixed r,  $\mathbb{E}\left[\binom{X}{r}\right] \to \frac{\mu^r}{r!}$ . Then  $X \xrightarrow{D} P_0(\mu)$ .

Now note that If  $X = X_1 + X_2 + \dots + X_m$  is a sum of indicator random variables corresponding to events  $A_1, A_2, \dots, A_m$  then  $\mathbb{E}\left[\binom{X}{r}\right] = \sum_{i_1, i_2, \dots, i_r \in \binom{[m]}{r}} \Pr[A_{i_1} \wedge A_{i_2} \wedge \dots \wedge A_{i_r}].$ 

Returning to our example where  $G = K_3$  we know that  $\mathbb{E}[X_G] \to \frac{c^3}{6}$ . Now when considering  $\mathbb{E}\left[\binom{X_G}{2}\right]$  we realize that two triangles will either be disjoint, intersect in a vertex, or intersect in an edge. The expected number of two disjoint triangles is  $p^6\binom{n}{6}\binom{6}{3}\frac{1}{2} \approx \frac{1}{2}\left(\frac{n^3}{6}\right)^2 p^6 \approx \frac{\mu}{2}$ . While on the other hand the expected number of two triangles intersecting in a vertex is  $p^6\binom{n}{3}\binom{n-3}{2}\frac{3}{2} = O(n^5p^6) = o(1)$ . Similarly the expected number of two triangles sharing an edge is  $p^5\binom{n}{4}\binom{4}{2} = O(n^4p^5) = o(1)$ . So  $\mathbb{E}\left[\binom{X_G}{2}\right] \to \frac{\mu^2}{2}$  as  $n \to \infty$  as desired. Now for  $\mathbb{E}\left[\binom{X_G}{3}\right]$  consider that the expected number of a disjoint triangles is  $\frac{1}{3!}\binom{n}{3}\binom{n-3}{3}\binom{n-6}{3}p^9 \approx \frac{\mu^3}{3!}$  and the contribution to

 $\mathbb{E}\left[\binom{X_G}{3}\right] \text{ from 3 intersecting triangles will be small. Thus again we will see that} \\ \mathbb{E}\left[\binom{X_G}{3}\right] \to \frac{\mu^3}{3!}. \text{ Continuing in this way and applying the previous theorem we see that } X_G \xrightarrow{D} P_0\left(\frac{c^3}{6}\right).$ 

**Theorem 6.3.** If H is strictly balanced  $\left(\frac{e_{H'}}{v_{H'}} < \frac{e_H}{v_H}\right)$  for every proper subgraph H' and if  $np^{e_H/v_H} \to c > 0$  as  $n \to \infty$  then  $X_H \xrightarrow{D} P_0(\mu)$  where  $\mu = \frac{c^{v_H}}{aut(H)}$ . (aut(H) is the number of automorphisms of H.)

**Lemma 6.4.** Let  $e_t$  be the minimum number of edges in a t vertex union of k not mutually disjoint copies of a strictly balanced graph G, and suppose  $2 \le k \le t < kv_G$ . Then for  $m(G) = \frac{e_G}{v_G}$  we have  $e_t > tm(G)$ .

7. (FRIDAY, SEPTEMBER 7)

Reminded:

- $m_G = \max \frac{e_{G'}}{v_{G'}}$ , where maximum is taken over all subgraphs G'
- G is strictly balanced if  $m_{G'} < m_G$  for every proper subgraph G'

We want to show:

#### Theorem 7.1. If

(1) G is strictly balanced, and

(2) p is such that  $\mathbb{E}[X_G] \to \mu \in (0, \infty)$ ,

then  $X_G \to Po(\mu)$ .

Idea: method of moments

- If  $\mathbb{E} \to \mu \in (0, \infty)$  and  $\mathbb{E}[\binom{X}{k}] \to \frac{\mu^k}{k!}$ , then  $X \xrightarrow{D} \operatorname{Po}(\mu)$ .
- If  $X = X_1 + \ldots + X_m$  and  $X_i$  is indicator r.v. for event  $B_i$ , then  $\mathbb{E}\left[\begin{pmatrix} X \\ \end{pmatrix}\right] = \sum_{i=1}^{n} \Pr[B_i \cap \dots \cap B_i]$

$$\mathbb{E}\left[\binom{A}{k}\right] = \sum_{\text{all } k \text{ subsets of } \{1, 2, \dots, m\}} \Pr[B_{i_1} \cap \dots \cap B_{i_k}].$$

**Proof of Theorem 7.1:** Write  $\mathbb{E}\left[\binom{X_G}{k}\right] = X'_k + X''_k$ . Here,  $X_G$  is sum of indicator r.v.s. How many?  $\binom{n}{v_G}(v_G)!/a_G$  where  $a_G := \operatorname{aut}(G)$ .  $X'_k$  denotes contribution to sum from mutually vertex disjoint copies of G, and  $X''_k$  denotes everything else. Note:

$$X'_{k} = \frac{1}{k!} \frac{\binom{n}{v_{G}}(v_{G})! p^{e_{G}}}{a_{G}} \frac{\binom{n-v_{G}}{v_{G}}(v_{G})! p^{e_{G}}}{a_{G}} \cdots \frac{\binom{n-(k-1)v_{G}}{v_{G}}(v_{G})! p^{e_{G}}}{a_{G}} \approx \frac{1}{k!} \mu^{k}$$

since  $\mu = \mathbb{E}[X_G] = p^{e_G} \binom{n}{v_G} \frac{(v_G)!}{a_G} \approx \frac{n^{v_G} p^{e_G}}{a_G}, \binom{n}{v_G} \approx \frac{n^{v_G}}{(v_G)!}$  and  $\binom{n-(k-1)v_G}{v_G} \approx \frac{n^{v_G}}{(v_G)!}$ . Let  $e_t$  be the minimum number of edges in t vertex union of k not mutually

Let  $e_t$  be the minimum number of edges in t vertex union of k not mutually disjoint copies of G.

**Lemma 7.2.** For  $k \ge 2$  and  $k \le t < kv_G$ , we have  $e_t > tm_G$ , i.e.,  $\frac{e_t}{t} > m_G$  (density of union > density of G).

**Proof of Lemma 7.1:** For arbitrary graph F, define  $f_F = m_G v_F - e_F$ . Note:

- (1)  $f_G = 0$
- (2)  $f_H > 0$  for any proper subgraph H of G because G is strictly balanced.
- (3)  $f_{F_1 \cup F_2} = f_{F_1} + f_{F_2} f_{F_1 \cap F_2}$  for arbitrary graphs  $F_1, F_2$

Let  $F = \bigcup_{i=1}^{k} G_i$ . Assume without loss of generality  $G_1 \cap G_2 \neq \emptyset$ . Induction on k: we want to show  $f_F < 0$ 

$$\begin{split} &k = 2: \, f_{G_1 \cup G_2} = f_{G_1} + f_{G_2} - f_{G_1 \cap G_2} < 0 \, (\because f_{G_1} = f_{G_2} = 0, \, f_{G_1 \cap G_2} > 0) \\ &k \geq 3: \, \text{Let} \, \, F' = \cup_{i=1}^{k-1} G_i \, \, \text{and} \, \, H = F' \cap G_k. \ \text{Then,} \, \, f_F = f_{F'} + f_{G_k} - f_H < 0 \, (\because f_{G_k} - f_{G_k}) \\ &k \geq 0 \, \text{Let} \, F' = 0 \, \text{Let}$$

 $f_{F'} < 0$  by induction,  $f_{G_k} = 0$ , and  $f_H \ge 0$  since H can be any subgraph of G including null graph or G itself)

To finish proof of Theorem 7.1, we want  $X_k'' = o(1)$  for every k.

$$X_{k}'' = \sum_{t=k}^{kv_{G}-1} O(n^{t} p^{e_{t}})$$

There are only finite many possibilities for F, for any fixed k, t. Note that  $n^t p^{e_t} \approx n^t n^{-e_t/m_G}$  for  $p \approx n^{-1/m_G}$ . But  $t - e_t/m_G < 0$  by Lemma 7.2. So,  $X''_k = o(1)$ .  $\Box$ 

A few balanced but not strictly balanced examples:

(1)  $G = \triangle \triangle, T = \triangle, p = \frac{c}{n}$ We have a.a.s.  $X_G = \binom{X_T}{2} = \frac{1}{2}X_T(X_T - 1)$  (because a.a.s. no.  $\triangle \triangle < \triangle$ ) We know  $X_T \to z \in \operatorname{Po}(\frac{c^3}{6})$ . Continuous functions preserve converges in distribution. So,  $X_G \stackrel{\mathrm{D}}{\to} \frac{1}{2}z(z-1), z$  as above

In particular, 
$$\Pr[X_G = 0] \to (1 + \frac{c}{6})e^{-\frac{c}{6}}$$
  
 $(\because \Pr[Z = 0] = e^{-\frac{c^3}{6}}, \Pr[Z = 1] = \frac{c^3}{6}e^{-\frac{c^3}{6}})$ 

(2)  $G = \triangle \Box, T = \triangle, S = \Box, p = \frac{c}{n}$ a.a.s.  $X_G = X_T X_S$  $X_T \to z_1 \in \operatorname{Po}(\frac{c^3}{6})$  $X_S \to z_2 \in \operatorname{Po}(\frac{c^4}{8})$ 

It turns out that  $(X_T, X_S) \xrightarrow{D} (\operatorname{Po}(\frac{c^3}{6}), \operatorname{Po}(\frac{c^4}{8}))$  and limit variables are independent (see chapter 6 of JLR).

$$X_G \to z_1 z_2$$
  
 $\Pr[X_G = 0] = 1 - (1 - e^{-\frac{c^3}{6}})(1 - e^{-\frac{c^4}{8}})$ 

(3)  $G = \triangle_{-}, p = \frac{c}{n}$ It can be shown that  $X_G \to \sum_{i=1}^{Z_T} W_i$  $Z_T \in \operatorname{Po}(\frac{c^3}{6})$ Each  $W_i \in \operatorname{Po}(3c)$  all independent

Idea:  $Z_t$  triangles, Each  $W_i$  perpends edges

**Exercise 7.3.** What is  $Pr[X_G = 0]$  in Example (3)?

#### 8. (Monday, September 10)

### Connection of Three Distributions; Binomial, Poisson and Normal

8.1. **Binomial distribution.** Bin(n, p), a distribution on  $\{0, 1, \dots, n\}$  means when flipping a coin with the weighted head with probability p and the tail with probability 1 - p n times, counting the number of heads. (or it can be considered as a sum of i.i.d Bernolli random variables.)

**Example** The number of edges in G(n, p) is Bin(N, p) for  $N = \binom{n}{2}$ , and p = p. Note If  $Y \in Bin(n, p)$ , then  $\mathbb{E}[Y] = np$ .

To see connections of Bin(n, p) and  $Po(\mu)$ , set  $\mu = pn \in (0, \infty)$  and let  $n \to \infty$ ( or  $p \to 0$ ). Let  $Y \in Bin(n, p)$ , then for a fixed  $k \ge 0$ ,

$$Pr[Y = k] = \binom{n}{k} p^k (1-p)^{n-k}$$
$$\sim \frac{n^k}{k!} \left(\frac{\mu}{n}\right)^k e^{-\frac{\mu}{n}(n-k)} \quad (\because (1-p) \sim e^{-p})$$
$$\rightarrow \frac{\mu^k}{k!} e^{-\mu} \text{ as } n \to \infty$$

In other words, if  $pn = \mu$ , and  $n \to \infty$  then

$$\operatorname{Bin}(n,p) \xrightarrow{\mathcal{D}} \operatorname{Po}(\mu).$$

Why is Poisson so ubiquitous? "law of rare events"

**Example** Let  $X_T$  be the number of triangles in G(n, p) for  $p = \frac{c}{n}$ . So  $\mathbb{E}[X_T] \to \frac{c^3}{6}$  as  $n \to \infty$ . We showed by method of moments that  $X_T \xrightarrow{\mathcal{D}} \operatorname{Po}(\frac{c^3}{6})$ . Is  $X_T$  is a binomial random variable? No, but seems to get like one. Model  $X_T$  as  $\operatorname{Bin}(N, P)$ , where  $N = \binom{n}{3}$ ,  $P = p^3$ .

**Binomial Revisited** Let  $Y \in Bin(n, p)$ . If set  $p \in (0, 1)$  and let  $n \to \infty$ , what can we say about the limiting behavior?

(2) 
$$\frac{Y - \mathbb{E}[Y]}{\sqrt{\operatorname{Var}(Y)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where  $\mathcal{N}(0,1)$  a normal distribution with mean 0, variance 1 and probability density function  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ . So if  $X \in \mathcal{N}(0,1)$  then

$$P(X \le a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx,$$

and (2) is equivalent that

$$P(Y \le a) \to \sqrt{2\pi} \int_{-\infty}^{a} e^{-x^2/2} dx$$

for every fixed  $a \in \mathbb{R}$  and  $n \to \infty$ . This holds in greate generality, even when variance does not exist.

## 8.2. Central limit theorem. See Feller Vol. 1 and Vol. 2 for more general statement.

**Theorem 8.1.** Let  $S_n = X_1 + \cdots + X_n$  where  $X_i$  are *i.i.d* random variables such that  $\mathbb{E}[X_i] = M$  and  $Var[X_i] = \sigma^2$  exist. Then

$$Pr\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} < \beta\right] \to \int_{-\infty}^{\beta} e^{-x^2/2} dx.$$

In other word,

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{Var[S_n]}} \to \mathcal{N}(0, 1).$$

**Corollary 8.2.** If  $Y \in Bin(n,p)$  for a fixed  $p \in (0,\infty)$ , then

$$\frac{Y - \mathbb{E}[Y]}{\sqrt{Var[Y]}} \to \mathcal{N}(0, 1).$$

**Theorem 8.3.** Let  $X = X_n$  be Poisson distribution with mean $\mu = \mu_n$  and suppose  $\mu \to \infty, n \to \infty$ . Then

$$\frac{X - \mathbb{E}[X]}{\sqrt{Var[X]}} \to \mathcal{N}(0, 1)$$



# 8.3. Random Graphs. Recall that

$$m_G = \max_{\substack{H \le G, \\ \text{subgraphs}}} \frac{e_H}{v_H}.$$

Recently, we have proved

(1)  $p = n^{-\frac{1}{m_G}}$  is the threshold for appearance of G subgraphs.

(2) If G is strictly balanced and  $\mathbb{E}[X_G] \to \mu$ , then  $X_G \xrightarrow{\mathcal{D}} \operatorname{Po}(\mu)$ .

What if  $p >> n^{-\frac{1}{m_G}}$ ? (See Ch.6 and Thm 6.5 on JTR.)

**Theorem 8.4.** Let G be a fixed graph, with  $e_G > 0$ . Suppose that

(1)  $np^{m(G)} \to \infty$ , (2)  $n^2(1-p) \to \infty$ , (i.e. p is not too large) then

$$\frac{X_G - \mathbb{E}[X_G]}{\sqrt{Var[X_G]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Note Suppose  $n^2(1-p) \to c \in (0,\infty)$ , then  $\mathbb{E}[\# \text{ non-edges }] = \binom{n}{2}(1-p) \to c/2$ . In fact,  $X_n = \#$  non-edges is binomial with finite mean, so it is Poisson in limit with  $\mu = c/2$ , and  $\Pr[\text{no non-edges}] \to e^{-c/2} > 0$ . That is,  $\Pr[G(n,p)$  is a complete graph] is bounded away from zero.

#### 9. (Wednesday, September 12)

Let  $\omega(G)$  denote the clique number of a graph, G, which is defined as the size of the largest complete subgraph of G.

# **Exercise 9.1.** Suppose $p = n^{-0.99}$ . How does $\omega(G(n, p))$ behave?

Since we know that the threshold for having a  $K_3$  subgraph is  $n^{-1}$  and the threshold for having a  $K_4$  subgraph is  $n^{-\frac{2}{3}}$ , we have *a.a.s.* at least one  $K_3$  subgraph and zero  $K_4$  subgraphs. So  $\Pr[\omega(G(n, p)) = 3] \to 1$ .

More generally, since a complete graph is strictly balanced, we know the threshold for having a  $K_k$  subgraph (k fixed) is at  $p = n^{-\frac{k}{\binom{k}{2}}} = n^{\frac{-2}{k-1}}$ 

**Exercise 9.2.** Fix k and let  $p = cn^{\frac{-2}{k-1}}$ . What can we say about  $\omega(G(n,p))$ ?

First, we know that *a.a.s* we have  $K_{k-1}$  subgraphs and zero  $K_{k+1}$  subgraphs. Moreover sometimes, we have  $K_k$  and sometimes we don't. Hence *a.a.s.*,  $\omega(G(n,p))$  is either k-1 or k.

If X = the number of  $K_k$  subgraphs, then X converges in distribution to a poisson random variable with mean  $\mu$ , where  $\mu = \lim \mathbb{E}[X]$ .

$$\mathbb{E}[X] = \binom{n}{k} p^{\binom{k}{2}} \sim \frac{n^k}{k!} \frac{c^{\binom{k}{2}}}{n^k} \sim \frac{c^{\binom{k}{2}}}{k!}$$

So we have that

$$\Pr[\omega \ge k] = \Pr[\exists K_k \subset G] \to \Pr[Po(\frac{c^{\binom{k}{2}}}{k!}) \ge 1] = 1 - \exp\left(-\frac{c^{\binom{k}{2}}}{k!}\right)$$
  
So  $\Pr[\omega = k - 1] \to \exp\left(-\frac{c^{\binom{k}{2}}}{k!}\right)$  and  $\Pr[\omega = k] \to 1 - \exp\left(-\frac{c^{\binom{k}{2}}}{k!}\right)$ .

**Remark 9.3.** For a fixed  $\alpha > 0$ , if  $p = O(n^{-\alpha})$ , then the clique number should be concentrated on at most 2 values.

**Exercise 9.4.** What can we say about larger p? For example, suppose  $p = \frac{1}{2}$ . What is  $\omega(G(n, p))$ ?

Heuristically if we solve  $n^{\frac{-2}{k-1}} = p$  for k, we find  $k \sim \frac{2\log(n)}{-\log(p)}$ . So we might suspect that  $k \sim 2\log_2 n$ .

Define  $f(k) = \mathbb{E}[\text{number of } k - cliques] = {n \choose k} \left(\frac{1}{2}\right)^{\binom{k}{2}}$ .

**Exercise 9.5.** Let  $\delta > 0$ . Show that if  $k \ge (2+\delta)\log_2 n$ , then  $f(k) \to 0$  and if  $k \le (2-\delta)\log_2 n$ , then  $f(k) \to \infty$ , as  $n \to \infty$ .

**Remark 9.6.** If  $N \to \infty$  and  $M = o(\sqrt{N})$ , then  $\binom{N}{M} \sim \frac{N^M}{M!}$  as  $N \to \infty$ 

Suppose that  $k = O(\log(n))$ , then

$$\binom{n}{k} \left(\frac{1}{2}\right)^{\binom{k}{2}} \sim \frac{n^k}{k!} \left(\frac{1}{2}\right)^{\binom{k}{2}} \sim \frac{1}{\sqrt{2\pi k}} \frac{e^k n^k}{k^k} \left(\frac{1}{2}\right)^{\binom{k}{2}} = \frac{1}{\sqrt{2\pi}} \exp\left(k + k \log(n) - (k + \frac{1}{2})\log(k) - \binom{k}{2}\log(2)\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(k \log(n) - \frac{k^2}{2}\log(2) + O(k \log(k))\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(k \log(n) - \frac{k^2}{2}\log(2) + O(\log(n)\log(\log n))\right)$$
If  $k = (2 + \epsilon) \log_2(n)$ , then

$$f(k) \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\epsilon \left(1 + \frac{\epsilon}{2}\right) \frac{1}{\log(2)} \log^2(n) + o(\log^2(n))\right)$$

Depending on the sign of  $\epsilon$ , we get the desired result. Following theorem is in Chap 4 of Alon and Spencer.

**Theorem 9.7.** Let k = k(n) satisfying  $k \sim 2 \log_2 n$  and  $f(k) \to \infty$  Then a.a.s.  $\omega(G(n,p)) \ge k$ .

Proof using Second Moment method. (Also see Chap 10 of Alon and Spencer for more precise results using Janson's Inequality).

Let  $A_1, A_2, \ldots, A_{\binom{n}{k}}$  be the events that correspond to one of the  $\binom{n}{k}$  possible k-cliques is present. Let  $X_1, \ldots$  be the corresponding indicator random variables and  $X = \sum_l X_l$ .

Fix l, and let  $\Delta^* = \sum_{j \sim l} \Pr[A_j | A_l].$ 

If we can show that  $\mathbb{E}[X] \to \infty$  and  $\Delta^* = o(\mathbb{E}[X])$ , then we have that a.a.s.X > 0.

By hypothesis,  $\mathbb{E}[X] = f(k) \to \infty$ .

If i corresponds to the number of vertices in the intersection with the fixed clique l of size k, then

$$\Delta^* = \sum_{i=2}^{k-1} \binom{k}{i} \binom{n-k}{k-i} \left(\frac{1}{2}\right)^{\binom{k}{2} - \binom{i}{2}}$$
$$= \mathbb{E}[X] \sum_{i=2}^{k-1} g(i)$$

where

$$g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}$$

**Exercise 9.8.** Show that  $\sum_i g(i) = o(1)$ 

We will show that g(2) and g(k-1) tend to zero.

$$g(2) = \frac{\binom{k}{2}\binom{n-k}{k-2}}{\binom{n}{k}} 2^{\binom{2}{2}} \sim \frac{k^2}{2} \frac{n^{k-2}}{(k-2)!} \frac{k!}{n^k} 2 \sim \frac{k^4}{n^2} = o(1)$$

 $g(k-1) = \frac{\binom{k}{k-1}\binom{n-k}{1}}{\binom{n}{k}} 2^{\binom{k-1}{2}} = \frac{k(n-k)2^{-(k-1)}}{\binom{n}{k}2^{\binom{k}{2}}} \sim \frac{2k \cdot n \cdot 2^{-k}}{f(k)} \sim \frac{2}{f(k)} \exp\left(\log(k) + \log(n) - k\log(2)\right)$ 

Since  $k \sim 2\frac{\log(n)}{\log(2)}$ , this exponent tends to  $-\infty$  giving the result we want. Hence X > 0 a.a.s., which implies that  $\omega(G(n, \frac{1}{2})) \geq k$ .

**Remark 9.9.** Since  $\frac{f(k+1)}{f(k)} = \frac{n-k}{k+1}2^{-k}$  and  $k \sim 2\log_2 n$ , we have that  $\frac{f(k+1)}{f(k)} = o(n^{-1})$ .

We can let  $k_0 = k_0(n)$  be such that  $f(k_0) \ge 1 > f(k_0 + 1)$ . For "most" n, f(k) jumps from large  $f(k_0)$  to small  $f(k_0 + 1)$ .

**Corollary 9.10** (Bollobas, Erdos 76; Matule 76). There exists a sequence k = k(n) such that  $Pr[\omega(G(n, \frac{1}{2})) = k \text{ or } k + 1] \to 1$ 

Moreover in Bollobas' book (on the chapter on cliques), it is shown that  $k_0 = 2\log_2 n - 2\log_2 \log_2 n + \log_2 \frac{e}{2} + o(1)$ 

**Question 9.11.** What is the threshold for G(n, p) to be connected?

One possible approach is to count the number of spanning trees in G(n, p)From Cayley's Theorem, we know that the number of trees on n vertices is  $n^{n-2}$ .

**Exercise 9.12.** Consider  $\mathbb{E}[number \ of \ spanning \ trees]$  and the threshold of connectivity. How many spanning trees should we expect to see?

10. (FRIDAY, SEPTEMBER 14)

Today: Threshold for Connectivity of G(n, p).

A naive approach is to count spanning trees. Recall that the graph  $K_n$  has  $n^{n-2}$  spanning trees. The probability that a particular spanning tree occurs in G(n,p) is  $p^{n-1}$ , so

$$\mathbb{E}[\# \text{ of spanning trees}] = n^{n-2}p^{n-1} = n(np)^{p-1}$$

**Remark 10.1.** If  $p \leq \frac{c}{n}$  for c < 1, then  $\mathbb{E}[\# \text{ of spanning trees}] \to 0 \Rightarrow a.a.s \nexists$ spanning trees  $\Rightarrow a.a.s G(n,p)$  is not connected. And if  $p \geq \frac{c}{n}$  for c > 1, then  $\mathbb{E}[\# \text{ of spanning trees}] \to \infty$ .

In the case that  $p \geq \frac{c}{n}$ , even though  $\mathbb{E} \to \infty$ , this does not necessarily imply that G(n,p) is connected a.a.s. Instead we conclude that  $Var[X] \neq 0$ .

**Theorem 10.2.** Threshold for Connectivity of G(n, p)

(1) If  $p \ge \frac{\log n + \omega(1)}{n}$ , then G(n, p) is connected a.a.s. (2) If  $p \le \frac{\log n - \omega(1)}{n}$ , then G(n, p) is disconnected a.a.s. (3) If  $p = \frac{\log n + c}{n}$ , c > 1 constant, then Pr[G(n, p) is connected]  $\rightarrow e^{-e^{-c}}$ 

Erdős-Rényi's idea was to look for components of order  $i = 1, 2, \ldots, \lfloor \frac{n}{2} \rfloor$ 

<u>i=1</u>: Isolated vertices. Let  $X_i = \#$  components of order *i*.  $\mathbb{E}[X_1] = n(1-p)^{n-1}$ . So if  $p = \frac{\log n+c}{n}$ , then

$$\mathbb{E}[X_1] = n\left(1 - \frac{\log n + c}{n}\right)^{n-1} \approx ne^{-\frac{\log n + c}{n}(n-1)} \to ne^{-\log n}e^{-c} = e^{-c} \text{ as } n \to \infty$$

**Exercise 10.3.** Show that  $X_1 \xrightarrow{\mathcal{D}} Po(e^{-c})$ , though none of the events are pairwise disjoint for distinct vertices.

 $\underline{i=2}$ : Isolated edges.

$$\mathbb{E}[X_2] = \binom{n}{2} p(1-p)^{2(n-2)}. \text{ If } p \sim \frac{\log n}{n},$$
$$\mathbb{E}[X_2] \approx \frac{n^2}{2} \frac{\log n}{n} e^{-\frac{\log n}{n} 2(n-2)} \to \frac{n \log n}{2} e^{-2\log n} = O\left(\frac{\log n}{n}\right) \to 0$$

 $\underline{i=3}$ : Triangles and paths of length two on three vertices.

$$\mathbb{E}[X_3] \le {\binom{n}{3}} 3p^2 (1-p)^{3(n-3)}. \text{ Again, for } p \sim \frac{\log n}{n},$$
$$\mathbb{E}[X_3] \approx \frac{n^3}{6} \frac{\log^2 n}{n^2} e^{-\frac{\log n}{n} 3(n-3)} \to \frac{n \log^2 n}{6} e^{-3 \log n} = O\left(\frac{\log^2 n}{n^2}\right) \to 0$$

For an upper bound that a set of k vertices is connected, recall

$$\mathbb{E}[\# \text{ spanning trees}] = k^{k-2}p^{k-1}$$

Now for  $4 \le k \le \lceil \frac{n}{2} \rceil$ 

$$\begin{split} \mathbb{E}[X_k] &\leq \sum_{i=4}^{\left\lceil \frac{n}{2} \right\rceil} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \leq \sum_{i=4}^{\left\lceil \frac{n}{2} \right\rceil} \frac{n^k}{k!} k^{k-2} p^{k-1} e^{-pk(n-k)} \leq \sum_{i=4}^{\left\lceil \frac{n}{2} \right\rceil} \frac{n^k}{\sqrt{2\pi k} \left(\frac{k}{e}\right)^k} k^{k-2} p^{k-1} e^{-pk(n-k)} \\ &= \frac{1}{\sqrt{2\pi}} \sum_{i=4}^{\left\lceil \frac{n}{2} \right\rceil} \frac{1}{k^{5/2}} (ne)^k p^{k-1} e^{-pk(n-k)} = \frac{1}{p\sqrt{2\pi}} \sum_{i=4}^{\left\lceil \frac{n}{2} \right\rceil} \frac{1}{k^{5/2}} \left(\frac{enp}{e^{p(n-k)}}\right)^k \leq \frac{1}{p\sqrt{2\pi}} \sum_{i=4}^{\left\lceil \frac{n}{2} \right\rceil} \left(\frac{enp}{e^{p(n-k)}}\right)^k \end{split}$$

A standard trick is to bound by a geometric series with  $a \to 0$  and  $r \to 0$ 

So 
$$k \leq \lceil \frac{n}{2} \rceil \Rightarrow e^{p(n-k)} \geq e^{\frac{\log n}{n} \frac{n}{2}} = e^{\frac{\log n}{2}} = \sqrt{n}$$
, and  $enp \sim e \log n \Rightarrow \frac{enp}{e^{p(n-k)}} \to 0$ 

Look at a, the k = 4 term:

$$\frac{1}{p\sqrt{2\pi}} \left(\frac{enp}{e^{p(n-4)}}\right)^4 = \frac{1}{\sqrt{2\pi}} \frac{e^4 n^4 p^3}{e^{4p(n-4)}} \sim \frac{e^4 n^4 \frac{\log^3 n}{n^3}}{e^{4\frac{\log^3 n}{n}(n-4)}} \to \frac{e^4 n \log^3 n}{n^4} = \frac{e^4 \log^3 n}{n^3} \to 0$$
$$\implies \frac{1}{p\sqrt{2\pi}} \sum_{i=4}^{\left\lceil \frac{n}{2} \right\rceil} \left(\frac{enp}{e^{p(n-k)}}\right)^k \to 0$$

Hence  $\mathbb{E}[\# \text{ components of order } 2 \leq k \leq \left\lceil \frac{n}{2} \right\rceil] \to 0$ 

Thus if  $p = \frac{\log n + c}{n}$ , then a.a.s. G(n, p) consists of a (unique) giant component of order (1 - o(1))n isolated vertices. And  $\mathbb{E}[X_1] = e^{-c}$ ,  $X_1 \in \text{Po}(e^{-c})$ . So

 $\Pr[G(n,p) \text{ connected}] \sim \Pr[G(n,p) \text{ has no isolated vertices}] \rightarrow e^{-e^{-c}}$ 

In fact, for  $p = \frac{\log n + c}{n}$ ,  $X_1 = \#$  components -1, so we have  $\tilde{\beta_0} \xrightarrow{\mathcal{D}} \operatorname{Po}(e^{-c})$ .

**Exercise 10.4.** What is the threshold for the appearance of cycles? Is it sharp?

11. (Monday, September 17)

Today : Expander-like qualities of G(n, p), Stronger notions of connectivity.

**Definition 11.1.** Cheeger number of a graph G

$$h(G) = \min \frac{\#e(s,S)}{|S|},$$

where minimum is taken over all  $S \leq V(G)$  and  $1 \leq |S| \leq \frac{|V(G)|}{2}$ .

Remark: This measures how hard it is to disconnect the graph. There's other ways to measure this: Vertex connectivity, edge connectivity.

Exercise: Show that  $p = \frac{\log n + (k-1) \log \log n}{n}$  is sharp threshold for k-connectivity. Clearly,  $h(G) \leq \min$  degree (G).

Example.  $G = K_n$ : complete graph.

Let k = |S|. Then  $e(S, \overline{S}) = k(n-k)$ .

$$h(K_n) = \min_{1 \le k \le \lfloor \frac{n}{2} \rfloor} \frac{k(n-k)}{k} \sim \frac{n}{2}$$

Our goal is to understand h(G(n, p)) to see that once G(n, p) is connected, it is very connected.

Often we want "concentration of measure" results showing  $\Pr[|X - E[X]| > t]$  is small. In many cases, one can do much better than Chebyshev's inequality / 2nd moment.

# Theorem 11.2. Chernoff-Hoeffding bounds.

Let  $X = \sum_{i \in [m]} X_i$ , where  $X_i$  are independent distributed in [0,1]. (For example, *i.i.d.* indicator random variables.)

$$\Pr[X > E[X] + t] \le e^{-2t^2/m}$$
$$\Pr[X < E[X] - t] \le e^{-2t^2/m}$$
$$\Pr[X > (1 + \epsilon)E[X]] \le \exp(-\frac{\epsilon^3}{3}E[X])$$
$$\Pr[X < (1 - \epsilon)E[X]] \le \exp(-\frac{\epsilon^3}{2}E[X])$$

**Theorem 11.3.** If  $p = w\left(\frac{\log n}{n}\right)$  and  $\epsilon > 0$  fixed, then a.a.s

$$\frac{(1-\epsilon)np}{2} \leq h(G(n,p)) \leq \frac{(1+\epsilon)np)}{2}$$

*Proof.* Upper bound is clear. (Use Chernoff-Hoeffding boudns.  $E[\#\text{edges}] = \frac{n}{4}p$  and  $\frac{n^2/4p}{n/2} = np/2$ .)

For any set S of s vertices, E[#edges] = ps(n-s). Since  $s \le n/2$ ,

$$\Pr[\# \text{edges} \le (1-\epsilon)ps(n-2)] \le \exp(-\frac{\epsilon^2}{2}ps(n-2)) \le \exp(-\frac{\epsilon^2}{2}ps\frac{n}{2}).$$

Set  $p \ge \frac{C \log n}{n}$  and C > 1 to be determined.

$$\Pr[\#\text{edges} \le (1-\epsilon)ps(n-2)] \le \exp(-\frac{\epsilon^2}{4}Cs\log n).$$

Then, for  $c > \frac{4}{\epsilon^2}$ ,

$$\begin{aligned} \Pr(h(G(n,p))) &\leq \sum_{s=1}^{\lfloor \frac{n}{2} \rfloor} \exp(-\frac{\epsilon^2}{4} Cs \log n) \\ &\ll \sum_{s=1}^{l} \left(\frac{ne}{s}\right)^s \exp(-\frac{\epsilon^2}{4} Cs \log n) \\ &= \sum_{s=1}^{l} \left(\frac{ne}{sn^{\epsilon^2/4C}}\right)^s = o(1). \end{aligned}$$

**Definition 11.4.** A family of bounded degree graphs  $G_1, G_2, \cdots$  with number of vertexes  $\rightarrow \infty$  is called expander family if

$$\liminf_{n \to \infty} h(G_n) > 0.$$

We will define Normalized Laplacian of a graph, which is a linear operator on functions on vertices

$$C^0(G) = \{ f : V(G) \to \mathbb{R} \}$$

Assume that the minimum degree of G satisfies  $\delta(G) \ge 1$ .

Define the averaging operator by

$$A(f)(v) = \frac{1}{\deg v} \sum_{u \sim v} f(u).$$

Define the identity operator by

$$I(f)(v) = f(v).$$

Then the Laplacian is defined by

$$\triangle = I - A.$$

Remark: Eigenvalues of  $\triangle$  are  $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$ . Multiplicity of 0-eignevalues is number of connected components of G. If G is connected,  $\lambda_2$  is spectral gap or algebraic connectivity.

**Theorem 11.5** (Hoffman, Kahle, Paquette). There exists C > 0 such that if

$$p \le \frac{\log n + C\sqrt{\log n} \log \log n}{n}$$

then  $\Pr(\lambda_2 < 1 - \epsilon) \to 0$  as  $n \to \infty$  for any fixed  $\epsilon > 0$ .

12. (Wednesday, September 19)

12.1. **Simplicial Complexes.** Today's lecture discussed higher-dimensional analogues of the Erdös-Rényi Theorem. To do so, we must discuss the appropriate higher-dimensional extension of a (finite) graph. In this case, the appropriate extension is an (abstract, finite) simplicial complex.

**Definition 12.1.** • Again, let  $[n] = \{1, 2, ..., n\}$ .

- An abstract finite simplicial complex, denoted S, is a collection of nonempty subsets of [n] with the following two properties.
  (1) if σ ∈ S and Ø ≠ τ ⊂ σ, then τ ∈ S.
  - (2) for all  $i \in \mathbb{N}$ ,  $1 \leq i \leq n$ ,  $\{i\} \in S$ .
- Elements of S are termed faces or simplices of S; similarly, if Ø ≠ τ ⊂ σ, then τ is termed a face or facet of the face σ. The dimension of a face is one less than its cardinality: if σ ∈ S, dim σ = |σ| − 1. For example, if {a, b, c} ∈ S, the dimension of that face is 2.
- Define dim S to be  $\max_{\sigma \in S} \dim \sigma$ .

The above definition is designed to run parallel to the topological definition of a simplex. For example, if  $\{a, b, c\} \in S$ , a, b, and c are thought of as the vertices of a 2-simplex. In fact, to every face of S we explicitly associate a simplex of the appropriate dimension: singleton sets are vertices,  $\{a, b\}$  is a line segment, etc. In particular if dim S = 1, then S can be associated with the graph on the vertex set [n] with edges corresponding to the 1-cells.

It is important to note (though we do not discuss the details here) that S can be considered as a topological simplex (a topological space), where requirement 1) of the definition ensures that the (topological) boundary of any simplex is included in the set of faces, and the second requirement just ensures that we have all vertices. To create a "geometric realization" of the abstract simplex, one starts with  $\mathbb{R}^n$ , with the standard basis vectors  $\{e_1, \ldots, e_n\}$ . Then the simplex  $\{i_1, \ldots, i_k\}$ is realized as the convex hull of the vectors  $\{e_{i_1}, \ldots, e_{i_k}\}$ . For a picture, see Figure 1. For more details, including the technicality of distinguishing between an abstract simplicial complex and its geometric realization, see [2]. We will intentionally blur the distinction between abstract complexes and their corresponding topological complexes.

12.2. (Simplicial,  $\mathbb{Z}/2\mathbb{Z}$ -)Cohomology. Recall that the Erdös-Rényi Theorem gave a sharp threshold for the connectivity of a graph. To generalize connectivity



FIGURE 1. Geometric Realization of some Two-Dimensional Simplices

to higher-dimensional analogues, we first realize connectivity of a graph as a statement about the graph's (reduced) 0-homology, or equivalently, 0-cohomology.<sup>6</sup> For higher-dimensional variants, we will define a "nice" (simplicial,  $\mathbb{Z}/2\mathbb{Z}$ ) cohomology, and start looking at the higher levels of cohomology.

# **Definition 12.2.** • Let $F^i(S)$ , $i \ge 1$ , denote the collection of *i*-dimensional faces of *S*.

- Let C<sup>i</sup>(S) denote the vector space of maps F<sup>i</sup>(S) → Z/2Z. This Z/2Z vector space is named the set of i-(Z/2Z-)cochains of S.
- Define the  $\mathbb{Z}/2\mathbb{Z}$ -linear map, the coboundary operator, as  $d^i : C^i(S) \to C^{i+1}(S)$ , as follows: if  $f \in C^i(S)$ , and  $\sigma \in F^{i+1}(S)$ , define  $d^i f(\sigma) = \sum_{\substack{\tau \text{ a face of } \sigma \\ \sigma = \tau \cup \{j\} \text{ for some } j \in \sigma}} f(\tau).^7$  In other words, we just add up f's values on all

codimension-1 faces of  $\sigma$ . In practice, the reference to a specific level is suppressed, and we just write d, not  $d^i$ , when the level of the cochain vector space is understood.

• Z<sup>i</sup> is defined to be the kernel of d<sup>i</sup>. d<sup>i-1</sup> maps into C<sup>i</sup>, and the image of d<sup>i-1</sup> is termed B<sup>i</sup>.

<sup>&</sup>lt;sup>6</sup>The equivalence is governed by the "universal coefficient theorem for homology," described in [6, p.195].

<sup>&</sup>lt;sup>7</sup>In general, powers of -1 according to orientations of the various faces enter into this formula. One advantage of working in  $\mathbb{Z}/2\mathbb{Z}$  is that (-1) = 1, so no minus signs are necessary.



FIGURE 2. Example of the coboundary map

An example is in order. See Figure 2. Looking at the upper-right-hand triangle above, note that there are two ways to have a function f on 1-cells in the boundary of a 2-cell such that df maps the 2-cell to 0: to have all the edges go to 0, or to have two of them going to 0 (since 1 + 1 = 2 = 0 in  $\mathbb{Z}/2\mathbb{Z}$ . In particular, then, it is likely that the kernel of  $d^i$  is a significant subspace of  $C^i$ . The key observation that enables our proceeding is the following exercise.

# **Lemma 12.3.** $B^i \subset Z^i$ ; that is, $d \circ d = 0$ .

*Proof.* Fix  $f \in C^i(S)$ ,  $i \ge 1$ ; we wish to show that  $d^{i+1} \circ d^i f = 0$ . By definition, the first coboundary gives  $d^i f(\sigma) = \sum_{\substack{\tau \text{ a face of } \sigma \\ \sigma = \tau \cup \{j\} \text{ for some } j \in \sigma}} f(\tau)$ . Then applying  $d^{i+1}$ 

gives, for any i + 2-dimensional face  $\rho$ ,

$$d^{i+1}(d^{i}(f))(\rho) = \sum_{\substack{\sigma \text{ a face of } \rho \\ \rho = \sigma \cup \{k\} \text{ for some } k \in \rho}} d^{i}f(\sigma)$$
$$= \sum_{\substack{\sigma \text{ a face of } \rho \\ \rho = \sigma \cup \{k\} \text{ for some } k \in \rho}} \sum_{\substack{\tau \text{ a face of } \sigma \\ \sigma = \tau \cup \{j\} \text{ for some } j \in \sigma}} f(\tau)$$

Yet note that since  $\sigma = \rho \setminus \{k\}$ , we can rewrite " $\sigma = \tau \cup \{j\}$  for some  $j \in \sigma$ " as " $\sigma = \tau \cup \{j\}$  for some  $j \in \rho \setminus \{k\}$ ." The point is not only that  $\tau$  is a codimension-2

face of  $\rho$ , but that for any codimension-two face  $\tau$  of  $\rho$ , say  $\tau = \rho \setminus \{p, q\}$ , we can find  $\tau$  in the final double-summation in two ways: by eliminating p and then q, or by eliminating q and then p. This happens for every such  $\tau$ , so in the end

$$d^{i+1} \circ d^{i}(f)(\rho) = \sum_{\substack{\tau \text{ a face of } \rho \\ \rho = \sigma \cup \{p,q\} \text{ for some } p,q \in \rho}} 2f(\tau).$$

Yet 2 = 0 in  $\mathbb{Z}/2\mathbb{Z}$ , so this becomes  $0.^8$  This works for all  $\rho \in F^{i+2}(S)$ , and for all  $f \in C^i(S)$ , so  $d^{i+1} \circ d^i = 0$ .

Since  $B^i$  is a sub-vector space of  $Z^i$ , we can define the *i*th cohomology group with  $\mathbb{Z}/2\mathbb{Z}$  coefficients, denoted  $H^i(S, \mathbb{Z}/2\mathbb{Z})$  or simply  $H^i$ , as the quotient vector space

$$\frac{Z^i}{B^i} = \frac{\ker d^i}{\operatorname{Im} d^{i-1}}.$$

We leave it as a fact that the cohomology groups are *topological invariants*; that is, homeomorphic spaces have the same cohomology groups (in fact, homotopy equivalent spaces have the same cohomology groups, but this will only concern us with concrete examples). For details, see Chapter 3 of [6].

The other major property that will concern us is that the (singular) cohomology map is a contravariant functor; that is, a continuous map  $X \xrightarrow{f} Y$  induces a map  $H^i(Y) \xrightarrow{f^*} H^i(X)$  by the definition  $f^* : \phi \to \phi \circ f$ . Further, with  $(g \circ f)^* = f^* \circ g^*$ and  $(Id_X)^* = Id_{H^i(X)}$ , where  $Id_X$  is the identity on X.

For our purposes, we need only concern ourselves with the dimension of the homology group dim  $H^i(S, \mathbb{Z}/2\mathbb{Z})$ .<sup>9</sup> Let  $\beta^i(S) = \beta^i := \dim H^i(S, \mathbb{Z}/2\mathbb{Z})$  denote the *ith*  $(\mathbb{Z}/2\mathbb{Z})$  *Betti number* of the simplicial complex S.

**Remark 12.4.** A circle,  $S^1$ , has  $\beta^0 = 1$  and  $\beta^1 = 1$ , and all other Betti numbers are 0. A sphere,  $S^2$ , has  $\beta^0 = 1$ ,  $\beta^1 = 0$ , and  $\beta^2 = 1$ , and all other Betti numbers are 0. More generally, the d-dimensional sphere  $S^d \subset \mathbb{R}^{d+1}$  has  $\beta^0 = 1$  and  $\beta^d = 1$ as the only nonzero Betti numbers.

The sphere is a good intuitive picture of what objects correspond to the simplest nontrivial d-dimensional homology groups; very roughtly speaking,  $\beta^d$  counts the number of d-dimensional holes in the topological space modeled by a simplicial complex.

<sup>&</sup>lt;sup>8</sup>In the non- $\mathbb{Z}/2\mathbb{Z}$  case, the trick is that the different orders of removing the vertices also shifts the sign of the orientation by one, so that the two copies of  $f(\tau)$  have competing signs and cancel in that way.

<sup>&</sup>lt;sup>9</sup>This is emphatically not the case in most algebraic topology, where various long exact sequences of (co)homology groups become very important.

We now show that remarks about the connectedness of a graph (e.g., the conclusions of the Erdös-Rényi theorem) are equivalent to statements about cohomology groups.

**Lemma 12.5.**  $\beta^0$  counts the number of connected components. Hence, S is connected if and only if  $\beta^0(S) = 1$ .<sup>10</sup>

*Proof.* There is no (-1)-st level in convential cohomology; or, to be more precise, we define it to be the single-point vector space  $\{0\}$ . Therefore,  $B^0 = \{0\}$  (since the boundary map is a vector space homomorphism, hence sends  $0 \in C^{-1}$  to  $0 \in C^0$ ). Therefore,  $H^0$  is equated with  $Z^0$ , the kernel of the map  $d^0$ . Yet a map f on vertices, mapping each vertex to 0 or 1, is transformed by the map  $df \in C^1$  such that for all edges  $e = \{i, j\}$  in the set,

$$df(e) = f(i) + f(j) = \begin{cases} 0, & \text{if } f(i) \cong f(j) \mod 2\\ 1, & \text{if } f(i) \not\cong f(j) \mod 2 \end{cases}$$

Therefore, any element f of the kernel  $Z^0$  must have identical inputs on the vertices of every edge. If f is a nonzero element of the kernel, it must send at least one vertex  $\{i\}$  to 1. Yet for any j adjacent to i,  $df(\{i, j) = 0$  (by  $f \in Z^0$ ), but  $df(\{i, j) = f(i) + f(j) = 1 + f(j)$ , so  $1 + f(j) = 0 \mod 2$  and hence f(j) = 1. This happens similarly for all k adjacent to j, and so forth, so f must send the entire connected component of i to 1. It can, however, send all other vertices to 0. Therefore, for each connected component  $\alpha$  of S, there is a map  $f_{\alpha} \in Z^0$ sending all vertices in  $\alpha$  to 1 and all vertices in other components to 0. These are clearly linearly independent functions, having disjoint support, and by the fact that elements of  $Z^0$  are constant on connected components, they span  $Z^0$ . Therefore,  $\dim(Z^0) = \dim(H^0) = \beta^0$  is equal to the number of connected components.

12.3. Reduced Homology Groups. The above work means that we can rephrase the topological property of connectivity in terms of a specific topological invariant, simplicial cohomology, which is "multilayered," hence admits generalization. It is not intuitively clear, however, where to go from here. In particular, our example of spheres demonstrates that "most" homology groups of a simple space are 0dimensional, not 1-dimensional. Also, if we are thinking algebraically, the simplest results should be the vanishing of such-and-such an invariant. We would like a situation where connectivity should correspond to  $\beta^0$  being a trivial element of  $\mathbb{Z}$ , namely 0. We do not wish, however, to change any of the other Betti numbers.<sup>11</sup>

<sup>&</sup>lt;sup>10</sup>The proof is optional.

<sup>&</sup>lt;sup>11</sup>The uninterested reader may now skip to the next subsection.
Therefore, we would like to define an altered set of Betti numbers  $\tilde{\beta}^i$ , where

$$\widetilde{\beta}^i = \begin{cases} \beta^i, & \text{if } i \ge 1\\ \beta^0 - 1 & \text{if } i = 0 \end{cases}$$

To accomplish this, in the abstract simplicial complex, we now allow the empty set to be in our collection of subsets of [n], and modify the closure-under-subsets rule to eliminate the  $\emptyset \neq \tau$  criterion. Since we have already included all vertices, and the  $\emptyset \subset \{i\}$  for any *i*, the empty set is now a guaranteed member of our revised (which, for reasons that will become clear later, we call "reduced") simplicial complex. We declare dim $(\emptyset) = 0$ , so that the empty set is a new codimension-1 face of every vertex.

Pushing this forward to the level of cochains, we may now define  $F^{-1} = \{\emptyset\}$ (the set whose sole element is the empty set); thus,  $C^{-1}$  is the set of maps  $\{\emptyset\} \rightarrow \mathbb{Z}/2\mathbb{Z}$ , namely  $f_1 : \emptyset \mapsto 1$  and  $f_0 : \emptyset \mapsto 0$ . Thus, this is a 1-dimensional  $\mathbb{Z}/2\mathbb{Z}$ -vector space. Therefore,  $d^{-1} : C^{-1} \to C^0$  maps  $f_0$  to the 0-map, but the map  $f_1$  has, for any  $i \in [n], d^{-1}f_1(\{i\}) = f_1(\emptyset) = 1$ , and hence is the constant-1 map.<sup>12</sup> Hence,  $B^{-1}$  is now 1-dimensional, so the homology group  $\widehat{H}^0 = \frac{Z^0}{\widehat{B}^{-1}}$  is reduced in dimension by 1, since we lose one degree of freedom by identifying maps that differ by a constant. Hence, we have defined the *reduced* homology groups, where  $\widehat{H}^0$  is adjusted as above, and nothing happens to the other parts of the cochain complexes, etc. Therefore, we have successfully defined a reduction in homology groups for which connectivity corresponds to the vanishing of some topological invariant.

**Remark 12.6.**  $\tilde{\beta}^0(S) = 0$  if and only if S is connected, for S any simplicial complex.

Therefore, we are now in a natural enough setting that we are able to fruitfully generalize the Erdös-Rényi Theorem by focusing our attention on  $\beta^1$ , and asking when it vanishes.

12.4. Random 2-Dimensional Simplicial Complexes. Following Linial and Meshulam [8], we now create a model of random 2-complexes. In order to avoid undue complexity, the simplest models should have a connected graph as their 1-skeleton (the collection of vertices and edges), since otherwise we have to start worrying about the interactions between the 0th and 1st cohomology groups.

**Definition 12.7.** Define Y(n, p) to be the 2-dimensional simplicial complex on the vertex set [n], where all vertices and edges (all sets of the form  $\{i\}$  or  $\{i, j\}$ ) are

<sup>&</sup>lt;sup>12</sup>More generally, for any coefficient group G,  $d^{-1}(C^{-1}(X,G))$  consists of all the constant maps from vertices to G; see, e.g., [6, p. 199].

included. For the  $\binom{[n]}{3}$  faces  $\{i, j, k\}$ , we add them to the simplex with probability p, jointly independently.

The paper [8] studies  $H^1(Y(n, p); \mathbb{Z}/2\mathbb{Z})$ , and in so doing gives guidance on the behavior of  $\beta^1$ . The extreme case p = 0 corresponds to the case of an very large  $\beta^1$ , since the lack of two-cells means that *every* edge can create its own cocycle (since  $C^2$  is empty and hence *every* function in  $C^1$  has coboundary map 0), whereas the set of coboundaries is fixed to those functions that give either 0 or 2 1's to the edges of any given triangle (as can be seen by case-by-case enumeration). At the other extreme, p = 1 causes  $H^1(Y(n, 1), \mathbb{Z}/2\mathbb{Z}) = 0$ , because f is a cocycle now if and only if the number of 1's f assigns to each edge of a triangle must be even (since the triangle's value in df equals the mod-2 sum of the values of the edges), hence 0 or 2, and hence a coboundary.

**Lemma 12.8.** The property of having a trivial  $H^1$ , i.e.,  $H^1(Y(n,p), \mathbb{Z}/2\mathbb{Z}) = 0$ , is a monotone-increasing property in p = p(n).

Proof. Exercise.

We are now in a position to state a major theorem of Linial and Meshulam.

**Theorem 12.9** ([8]).

$$Pr[H^{1}(Y(n,p);\mathbb{Z}/2\mathbb{Z})=0] = \begin{cases} 1, & p \ge \frac{2\log(n)+\omega(1)}{n} \\ 0, & p \le \frac{2\log(n)-\omega(1)}{n} \end{cases}$$

Compare this to the Erdös-Rényi Theorem:

$$\Pr[\widetilde{H}^0(G(n,p);\mathbb{Z}/2\mathbb{Z})=0] = \begin{cases} 1, & p \ge \frac{1\log(n)+\omega(1)}{n} \\ 0, & p \le \frac{1\log(n)-\omega(1)}{n} \end{cases}$$

There are similarities with parts of the proof, not just the statement. Recall that the final obstruction to the connectivity of G(n, p) was isolated vertices (because asymptotically almost surely, in the first case, the components of size less than or equal to  $\lfloor \frac{n}{2} \rfloor$  disappeared, so that we had a large component and isolated vertices). Similarly, we want to show in the present case that if there are any isolated edges (i.e., an edge with *no* incident triangles), then  $H^1 \neq 0$  and  $\beta^1 > 0$ .

We note that any given edge uses two vertices, and hence there are n-2 other vertices that combine with the given edge to give a triangle of edges; the face for that triangle is *not* added with probability (1 - p). There are  $\binom{n}{2}$  such edges; hence, by linearity of expectation,  $\mathbb{E}(\# \text{ isolated edges}) = \binom{n}{2}(1-p)^{n-2}$ . Therefore,

if 
$$p \approx \frac{2\log(n) + c}{n}$$
, then we have that

$$\begin{split} E(\#isolatededges) &\asymp \quad \frac{n^2}{2}e^{-\frac{2\log(n)+c}{n}(n-2)} \\ &\asymp \quad \frac{n^2}{2}e^{-(2\log(n)+c)} \\ &\asymp \quad \frac{n^2}{n^2}*\frac{1}{2}e^{-c} \\ &= \quad \frac{1}{2}e^{-c}. \end{split}$$

Note that each isolated edge indeed generates some 1-cochains that are not 1coboundaries. More specifically, taking the map f that takes the value 1 on an isolated edge  $e = \{i, j\}$  and 0 on all other edges gives a cocycle (since there are no adjacent triangles, so the coboundary of this map is necessarily 0). It is not, however, a 1-coboundary (since dg = f for  $g \in C^0(Y(n, p), \mathbb{Z}/2\mathbb{Z})$  would require that g assign a value of 1 to one of the endpoints of e (say i), and 0 to the other endpoint j. Any other vertex k must be assigned a 1 so that  $\{i, k\}$  gets mapped to 0, but simultaneously must be assigned a 0 so that  $\{j, k\}$  gets mapped to 0. Therefore, f is not a coboundary. Therefore,  $H^1(Y(n, p), \mathbb{Z}/2\mathbb{Z}) \neq 0$  in such a case.

**Conjecture 12.10.** If  $p = \frac{2 \log n + c}{n}$ , then

$$\beta^1(Y(n,p),\mathbb{Z}/2\mathbb{Z}) \xrightarrow{D} Po\left(\frac{1}{2}e^{-c}\right),$$

where  $\xrightarrow{D}$  denotes convergence in distribution.

The above conjecture is in fact true, but is stated as a conjecture because for random *d*-dimensional simplicial complexes analogous to the above two-dimensional case, the analogous statement for  $\beta^d$  is only known for large *d*. Note, however, that this is an extension of a preliminary result used in proving the Erdös-Rényi Theorem:

$$\widetilde{\beta}^0(G(n,p)) \xrightarrow{D} \operatorname{Po}\left(e^{-c}\right),$$

when  $p = \frac{\log n + c}{n}$ .

## 13. (FRIDAY, SEPTEMBER 21)

First, some thoughts about homology versus cohomology. Sometimes homology is easier to think about.

# 13.1. $H_0(X, \mathbb{Z}/2)$ measures the number of connected components of a graph.

# Definition 13.1.

- The zeroth homology group with Z/2 coefficients H<sub>0</sub>(X, Z/2) is a (Z/2)vector space whose dimension is the number of connected components of X.
- (2) The zeroth reduced homology group  $\widetilde{H}_0$  is a  $(\mathbb{Z}/2)$ -vector space whose dimension is the number of connected components of X minus one.
- (3) A 0-chain is a function  $\phi: V \to \mathbb{Z}/2$ .
- (4) A 0-cycle φ is a function in the kernel the boundary operator, i.e. dφ =
  0. In reduced homology, this is equivalent to a function supported on an even number of vertices. Then notice that the vector space of all cycles is generated by functions supported on pairs of vertices.
- (5) By definition,  $H_0 = cycles/boundaries$ .

So we say  $\tilde{H}_0 = 0$  if and only if every cycle is a boundary i.e., every pair of vertices is connected by a path. It turns out that  $\tilde{H}_0 = 0 \Leftrightarrow \tilde{H}^0 = 0$ . But when we proved Erdős–Rényi theorem, we actually proved the cohomological version  $\tilde{H}^0 = 0$ , i.e. every cocycle is a coboundary.

The only coboundary is the coboundary of the empty set, defined to be the constant function  $\phi : V(H) \to 1$ . A cocycle is a collection of vertices so that every edge meets an even number of them, or equivalently, a union of connected components generated by connected components. So we see that "every cocycle is a coboundary" really is equivalent to "H is connected."

To prove the Erdős–Rényi theorem, we show that for  $\phi =$  collection of vertices,  $\Pr(\phi \text{ is nontrivial cocycle})$  is small.

$$\Pr(\phi \text{ is nontrivial cocycle}) \le k^{k-2} p^{k-1} (1-p)^{k(n-k)},$$

k(n-k) is the size of  $d\phi$  in complex graph.

# 13.2. $H_1(G)$ , G a connected graph.

## Definition 13.2.

- (1) Cycles = collection of edges meeting every vertex in an even number of edges. generated by primitive cycles.
- (2) 1-coboundaries = cut spanning complete bipartite graph.

(3) 1-cocyles=collection of edges meeting every triangle in an even number.

 $H_1 = 0$  means every cycle is boundary means no cycles. In a 2-dimensional complex  $H_1 = 0$  means every cycle is a boundary. It is known that for 2-dimensional simplicial complex Y,  $H_1(Y, \mathbb{Z}/2) \equiv H^1(Y, \mathbb{Z}/2)$ .

Linial–Meshulam considers the case for 2-dim  $\mathbb{Z}/2$  coefficients.

Meshulam–Wallach: d-dim,  $\mathbb{Z}/m$  coefficients.

Let  $\triangle_n^{(2)}$  be a 2-skeleton of simplex on *n*-vertices. i.e., *n* vertices,  $\binom{n}{2}$  edges,  $\binom{n}{3}$  faces.

Let 
$$\phi \in C^1(\triangle_n^{(2)})$$
,

 $b(\phi) = |\operatorname{supp} d\phi| = \operatorname{number} of triangles with an odd number of edges from <math>\phi$ .  $w(\phi) = \min(\operatorname{supp} \phi + d\tau), \ \tau \in C^0.$ 

**Theorem 13.3** (Meshulam-Wallach).  $b(\phi) \geq \frac{n}{3}w(\phi)$ . (co-isoperimetric inequality).

$$\Pr[\exists any non-trivial cocyles] \le \sum_{k=1}^{\binom{n}{2}/2} \binom{\binom{n}{2}}{k} (1-p)^{\frac{n}{3}k}.$$

Exercise 13.4. Show that if

$$p \ge \frac{6\log n + \omega(1)}{n}$$

then

$$\Pr[\exists any non-trivial cocyles] \rightarrow 0.$$

To get all the way down to the true threshold of

$$p = \frac{2\log n}{n}$$

requires careful cocyle counting.

See also Hoffman, Kahle, Paquette's recent preprint "A sharp threshold for Kazhdan's property (T)" (arXiv:1201.0425), where spectral analogues of the Linial– Meshulam theorem are discussed, with applications to geometric group theory.

14. (Monday, September 24)

This lecture discusses the evolution of G(n, p) as a process with varying p, and in particular the phase transition that occurs around  $p = \frac{1}{n}$ . We begin by considering  $p = O(n^{-1-\epsilon})$  for fixed  $\epsilon > 0$ :

**Proposition 14.1.** Let k > 0 be fixed and  $p = o(n^{-(k+1)/k})$ ; then w.h.p. all connected components are of order at most k.

*Proof.* For k = 1:  $p = o(n^{-2})$  and  $\mathbb{E}(e) = p\binom{n}{2} = O(pn^2) = o(1)$ , so w.h.p. there are no edges.

For k = 2: Any component of order 3 contains a path of length 2, and

$$\mathbb{E}(\#2\text{-paths}) = \Theta(n^3 p^2) = o(1),$$

so w.h.p. any component has order at most 2.

For k arbitrary, there are only finitely many connected graphs of order k + 1, each of which has at least k edges. Thus the expected number of copies of any particular connected graph with k + 1 vertices is  $O(n^{k+1}p^k) = o(1)$  by assumption, and summing only finitely many (as k is fixed) is once again o(1).

Thus if p is bounded by any power of n less than  $n^{-1}$ , G(n, p) has only very small components (i.e. of bounded size). We can also establish how many cycles G(n, p) is likely to have. For fixed k, the threshold for finding a cycle  $C_k$  is  $p = n^{-v/e} = n^{-1}$ . By the same logic as the proposition above, if  $p = o(n^{-1-\epsilon})$  then w.h.p. G(n, p) is a forest (a collection of trees). We can say more:

# **Proposition 14.2.** If $p = o(n^{-1})$ , then G(n, p) is a forest w.h.p.

*Proof.* Let  $k \geq 3$  be a fixed integer. Then the expected number of  $C_k$  subgraphs in G(n, p) is  $n^k p^k / 2k$ , since the automorphism group of  $C_k$  is the dihedral group  $D_{2k}$ . Now, write  $p = cn^{-1}$ ; then

$$\mathbb{E}(\# \text{ cycles}) = \sum_{k=3}^{\infty} \frac{n^k p^k}{2k} \le \sum_{k=3}^{\infty} \frac{c^k}{2k} \le c^3 + c^4 + c^5 + \cdots$$

Since  $p = o(n^{-1})$ , we can take c < 1 and let  $c \to 0$ . Then the geometric series above converges to  $\frac{c^3}{1-c} \to 0$ . So w.h.p. the expected number of cycles is 0, meaning G(n, p) is a forest.

In fact, if p = c/n with c > 1, then w.h.p. G(n, p) contains a cycle. Moreover, it can be proved that the probability of finding a cycle approaches a constant for c = 1, and the probability that the first cycle is  $C_k$  approaches a limiting distribution which is bounded away from zero for all k.

Next is the case where  $p \gg n^{-1}$ ; in particular, we examine  $p \ge n^{-1/2+\epsilon}$  for positive  $\epsilon$ :

**Proposition 14.3.** If  $p \ge n^{-1/2+\epsilon}$ , then w.h.p. G(n,p) has diameter at most 2.

*Proof.* In fact it is only necessary to assume that  $p \ge \left(\frac{C\log n + \omega(1)}{n}\right)^{1/2}$  for some fixed C > 0 (to be determined). With that assumption, for any two vertices  $x, y \in [n]$ ,

$$\mathbb{E}(\# \text{ paths}\{x, v, y\}) = (n-2)p^2 = (n-2)\left(\frac{C\log n + \omega(1)}{n}\right) \ge C\log n$$

Since there are  $O(n^2)$  pairs of points x, y, we need  $\Pr[\text{no path}\{x, v, y\}] = o(n^{-2})$ . The number of paths of length 2 from x to y is a binomial random variable  $Bin(n - 2, p^2)$ . The relevant Chernoff bound is  $\Pr[X \leq (1 - \epsilon)\mathbb{E}[X]] \leq \exp(\mathbb{E}[X]\epsilon^2/2)$ . Fix  $\epsilon = 1/2$ ; then

$$\Pr[\text{no path}\{x, v, y\}] \le \exp(-\frac{1}{8}C\log n) = n^{-C/8}$$

so if C > 16, this probability is  $o(n^{-2})$ , as desired.

**Exercise 14.4.** Show that if  $p \ge \left(\frac{2\log n + \omega(1)}{n}\right)^{1/2}$ , then the diameter of G(n, p) is  $\le 2$  w.h.p.

To determine whether G(n, p) has diameter at most k, we apply the same principle as above: fix vertices x and y. Then  $\mathbb{E}[\# \text{ paths of length}k]$  is of order  $n^{k-1}p^k$ . We would expect that if  $n^{k-1}p^k \to \infty$  at least as fast as  $\log n$ , then w.h.p. every pair of vertices in G(n, p) is joined by a path of length at most k:

**Proposition 14.5.** If 
$$p \ge \left(\frac{C_k \log n}{n}\right)^{k/(k+1)}$$
, then w.h.p.  $diam(G(n,p)) \le k+1$ 

Note, however, that for k > 1, the paths between x and y are no longer described by a binomial distribution. Instead, the appropriate tools are Jansen inequalities, which stand in for Chernoff bounds.

**Proposition 14.6.** Let  $p = \frac{c \log n}{n}$ . If  $\frac{1}{2} < c \le 1$ , then w.h.p. G(n,p) consists of a unique giant component of order (1 - o(1))n and isolated vertices.

**Exercise 14.7.** Show that if  $p = \frac{c \log n}{n}$ , then w.h.p. G(n,p) consists of a giant component of order (1 - o(1))n and components of order  $\leq k$ , where k depends on c.

15. (Wednesday, September 26)

Cycles in G(n, p)

We have showed that if p = o(1/n), then whp there is no cycle in G(n, p). What about for p = c/n,  $c \in (0, \infty)$ ?

For  $k \geq 3$  fixed,

$$\mathbb{E}[\text{number of cycles in } G(n,p)] \sim \frac{n^k p^k}{2k} = \frac{c^k}{2k}$$

We have showed that  $X_k \to \operatorname{Po}\left(\frac{c^k}{2k}\right)$  where  $X_k$  denotes the number of k-cycles in G(n, p). It can be shown for any fixed k that

$$(X_3, X_4, \dots, X_k) \to \left( \operatorname{Po}\left(\frac{c^3}{6}\right), \operatorname{Po}\left(\frac{c^4}{8}\right), \dots, \operatorname{Po}\left(\frac{c^k}{2k}\right) \right)$$

where the Poisson random variables are independent (See Chapter 6, JLR). Then,

$$\Pr[\text{there is no cycle of length } \leq M] \to \exp\left(-\frac{c^3}{6} - \frac{c^4}{8} - \dots - \frac{c^M}{2M}\right).$$

If c < 1, then the sum  $\frac{c^3}{6} + \frac{c^4}{8} + \cdots$  converges and we would guess that

$$\Pr[\text{no cycles in } G(n,p)] \to \exp\left(-\frac{c^3}{6} - \frac{c^4}{8} - \dots\right)$$

For c = 1, the sum  $1/6 + 1/8 + \cdots$  diverges, and the probability that there is no cycle approaches 0 in this case.

**Question**: If p = 1/n, how fast does  $\Pr[\text{no cycle in } G(n, p)] \to 0$ ? **Answer**:  $\Pr[\text{ no cycle in } G(n, 1/n)] \sim Cn^{-1/6}$  where

$$C = \frac{1}{\sqrt{2\pi}} e^{3/4} \int_0^\infty e^{-4t^{3/2}} \cos\left(4t^{3/2}\right).$$

Let  $L_n^+$  be the length of the first cycle in  $\{G(n,m)\}_{m=0}^{\binom{n}{2}}$ .

Theorem 15.1.

$$\Pr[L_n^+ = \ell] \to \frac{1}{2} \int_0^1 x^{\ell-1} (1-x)^{1/2} e^{x/2 + x^2/4} dx$$

A consequence of this theorem is that, we have  $\mathbb{E}[L_n^+] \asymp n^{1/6}$  even though  $\Pr[L_n^+ \ge \omega] \to 0$  as  $\omega \to \infty$  however slowly.

**Diameter of** G(n, p)

**Janson Inequalities** [Reference: Alon & Spencer, Chapter 8 ] *Notation:* 

•  $\Omega$  is a finite universal set,

- R is a random subset of  $\Omega$  given by  $\Pr[r \in R] = p_r$ . These events are mutually independent over  $r \in \Omega$ ,
- $\{A_i\}_{i \in I}$  is a collection of subsets where I is a finite index set,
- $B_i$  is the event that  $A_i \subset R$ , i.e., when we flip a coin for each  $r \in \Omega$ ,  $B_i$  is the event that all the coins for  $r \in A_i$  comes up heads.
- $X_i$  is the indicator random variable for  $B_i$  and  $X = \sum_{i \in I} X_i$ ,
- For  $i, j \in I$ , we write  $i \sim j$  if  $i \neq j$  and  $A_i \cap A_j = \emptyset$  (hence  $B_i$  and  $B_j$  are independent).

*Note:*  $X = 0 \iff$  none of the events  $B_i$  occur

**Theorem 15.2** (Janson Inequality). Let  $\mu = \mathbb{E}[X]$  and  $\Delta = \sum_{i \sim j} Pr[B_i \text{ and } B_j]$ . Then,

$$Pr[X=0] \le e^{-\mu + \Delta/2}.$$

**Example**  $I = {\binom{[n]}{3}}$  and  $A_i$  is the triangle for  $i \in I$ . For p = c/n, we have  $\mu \to c^3/6$ . Check that  $\Delta = o(1)$ . Then,

Pr[no triangles in G(n, p)]  $\leq e^{-c^3/6 + o(1)}$ .

This theorem often gives some bound but if  $\mu \simeq \Delta$ , maybe not the best one. If  $\Delta \ge 2\mu$ , obviously it is useless.

**Theorem 15.3** (Extended Janson Inequality). Under additional assumption that  $\Delta \ge \mu$ , we have

$$Pr[X=0] \le e^{-\mu^2/2\Delta}.$$

**Claim**  $n^{-2/3}$  is (roughly) the threshold for the event {diameter of  $G(n, p) \leq 3$ }. Let  $p^3n^2 \geq C \log(n)$ , (C is a constant to be determined).

Given vertices x and y, the expected number of paths of length 3 with end vertices x and y is  $(n-2)(n-3)p^3 \sim n^2p^3$ . Hence  $\mu \sim n^2p^3$ . We proved the following lemma previously.

**Lemma 15.4.** Let  $e_t$  be the minimum number of edges in a union of (not all vertex disjoint) k copies of strictly balanced graph G on t vertices. Then  $e_t > m_G \cdot t$ .

Let *i* be the number of vertices in intersection of two such paths. Then either i = 1 or i = 2.

**Exercise 15.5.** Show that  $n^{-k/(k+1)}$  is (roughly) threshold for diameter  $\leq k+1$ .

## 16. (FRIDAY, SEPTEMBER 28)

**Example** Let X be the number of triangles in G(n, p) for  $p = \frac{c}{n}$  then  $\mu \to \frac{c^3}{6}$ and  $\Delta = o(1)$  So  $\Pr[X = 0] \le e^{-c^3/6 + o(1)}$  by Janson. In general  $\mu = \frac{n^3 p^3}{6}$  and  $\Delta = \frac{n^4 p^5}{4}$ . Once  $p \gg n^{-1/2}$  we have  $\Delta > \mu$  so for example for  $p = \frac{1}{2}$  we have  $\Pr[X = 0] \le e^{-cn^2}$  by extended Janson.

**Example** Consider paths of length k + 1 from  $x \leftrightarrow y$ . Set  $\mu = n^k p^{k+1} = 2\log(n) + \omega(1)$ . In this case  $\Delta = o(1)$ . Observe that two distinct paths  $x \leftrightarrow y$  that intersect in *i* vertices intersect in at most *i* edges. Then the largest contribution to  $\Delta$  is  $n^{2k-1}p^{2k+1}$  when i = 1. Then we conclude if  $n^k p^{k+1} \ge (2 + \varepsilon)\log(n)$  for  $\varepsilon > 0$  fixed then with high probability the diameter of G(n, p) is less than or equal to k + 1.

# **Random Regular Graphs and Expanders**

Note once  $p \gg \frac{\log(n)}{n}$  then with high probability  $deg(v) \approx (n-1)p$  for every vertex v in G(n,p) by Chernoff bounds. However, with high probability G(n,p) is not normal and the average degree  $(n-1)p \to \infty$ .

We want to have a model to generate regular random graphs. Our idea here is to construct them out of permutations. Begin with a permutation  $\sigma \in S_n$  chosen uniformly. Then  $\sigma : [n] \to [n]$  bijectively. We can then think of  $\sigma$  as a random 2 regular graph. The two problems we encounter are fixed points of  $\sigma$  and transpositions. Now  $\mathbb{E}[\#\text{of fixed points}] = 1$  and  $\Pr[\text{no fixed points}] \to \frac{1}{e}$ . So it's true that the number of fixed points approaches in distribution to  $P_0(1)$ . Now  $\mathbb{E}[\#\text{of transpositions}] = {n \choose 2} \frac{1}{n(n-1)} = \frac{1}{2}$ . So we have a small number of bad events that we will be able to control for.

Take  $\sigma_1, \sigma_2, \ldots, \sigma_l$  to give a 2l-regular graph on [n]. There are two types of bad events we may encounter in this construction. The first happens when  $\sigma_i(x) = \sigma_j(x)$ for  $i \neq j$  and the second happens when  $\sigma_i(x) = \sigma_i^{-1}(x)$  for  $i \neq j$ . Note by the previous discussion the expected number of bad events tends to a constant as  $n \to \infty$ . Using Poisson approximation one can show that there is a positive probability of having no bad events. Thus to create our 2l-regular graph we just pick our permutations until we have no bad events.

Now for some fixed l large enough we want show that with high probability our random regular graph is an expander. That is we wish to show that the Cheeger number h(G) > c. Recall  $h(G) = \min_{\substack{|S| \le n/2}} \frac{\#e(S,\bar{S})}{|S|}$ . It will suffice to show for all  $F \subseteq [n]$  with  $|F| \le \frac{n}{2}$  that  $|N(F)| \ge (1+c')|F|$  for some c' > 0. This will exclude the possibility that there exists  $F \subseteq F' \subseteq [n]$  with  $|F'| \le (1+c')|F|$  and with all edges in F' ending up in F. Let r = |F|, r + r' = |F'|, and  $r' = \lfloor c'r \rfloor + 1$ . Then for each r there are  $\frac{n!}{r!r'!(n-r-r')!} \le \frac{n^{r+r'}}{(r+r')!}c^r$  choices for F and F'. Now  $\begin{aligned} &\Pr[\text{all edges of } F' \text{ end in } F] = \frac{\binom{r+r'}{r}}{\binom{n}{r}} \leq \left(\frac{r+r'}{n}\right)^r. \text{ Then we know the total probabil-}\\ &\text{ity of failure for some choice of } F \text{ and } F' \text{ is at most } \left(\frac{r+r'}{n}\right)^{rl}. \text{ So by Sterling's approximation the total failure probability is bounded by } \sum_{r=1}^{\lfloor n/2 \rfloor} o(1)^r \left(\frac{r+r'}{n}\right)^{lr-r-r'}.\\ &\text{So for small enough } c \text{ we have that } \frac{r+r'}{n} \leq 0.6 \text{ and so for large enough } l \text{ the whole } \end{aligned}$ 

sum goes to o(1).

**Definition 16.1.** Given a graph G(V, E),  $S \subseteq V$  is a <u>vertex cut set</u> if  $G(V \setminus S, E)$  is not connected. A graph G is <u>k-connected</u> if there is no cut set of size k - 1. The connectivity number of G denoted  $K(G) = max\{k : G \text{ is } k\text{-connected}\}$ .

Note that we know the threshold for 1-connectedness in G(n,p) is  $p = \frac{\log(n)}{n}$ . We will see that the threshold for k-connectedness is  $p = \frac{\log(n) + (k-1)\log(\log(n))}{n}$ .

**Theorem 16.2.** Let  $p = \frac{\log(n) + (k-1) \log(\log(n))}{n}$ . Then

$$\Pr[G(n,p) \text{ is } k\text{-connected}] \to \begin{cases} 0 & \text{if } x_n \to -\infty \\ 1 & \text{if } x_n \to \infty \\ e^{e^c/(k+1)!} & \text{if } x_n \to c \end{cases}$$

Note that the probability of G(n, p) being k-connected is approximately the probability that G(n, p) has no vertex of degree k - 1.

17. (Monday, October 8)

17.1. The phase transition in G(n, p). Very roughly, set  $p = \frac{c}{n}$ , where  $c \in (0, \infty)$  is constant.

- If c < 1, then w.h.p. largest component has order  $O(\log n)$ .
- If c = 1, then w.h.p. largest component has order  $O(n^{2/3})$ .
- If c > 1, then w.h.p. largest component has order  $\geq \lambda n$ , where  $\lambda = \lambda(c)$ . (aka giant component)

Erdos and Renyi called this a *double jump*.

Now let's state the theorem in detail following Alon–Spencer. First some notation:

- (1) A connected component is said to have *complexity* e v + 1. (*eg:* a tree has complexity 0, a unicycle has complexity 1)
- (2)  $C_i = \#i$ -th largest component,  $L_i = \#$  vertices of  $C_i$ . ( $L_1$  =size of largest component)

Now the theorem in five parts:

- (1) Very subcritical  $(p = \frac{c}{n}, c < 1)$ 
  - All components are trees or unicyclic
  - $L_1 = \Theta(\log n)$
  - $L_k \sim L_1$  for every fixed k
- (2) Barely subcritical  $(p = \frac{1-\epsilon}{n}, \epsilon = \lambda n^{-1/3}$  and assume  $\epsilon \to 0, \lambda \to \infty)$ 
  - All components are trees or unicyclic
  - $L_1 = \Theta(\epsilon^{-2} \log \lambda)$
  - $L_k \sim L_1$  for every fixed k

(3) Critical  $(p = \frac{1-\epsilon}{n}, \epsilon = \lambda n^{-1/3}$  where  $\lambda \in (-\infty, \infty)$  is constant)

- Largest k components (k fixed), all have size  $L_k = \Theta(n^{2/3})$
- Parametrizing,  $L_k = c_k n^{2/3}$  and  $d_k = complexity(c_k)$ ; then there is a nontrivial limiting point distribution for  $(c_1, c_2, \ldots, c_k, d_1, \ldots, d_k)$
- (4) Barely supercritical  $(p = \frac{1+\epsilon}{n}, \epsilon = \lambda n^{-1/3} \text{ and assume } \epsilon \to 0, \lambda \to \infty)$ 
  - $L_1 \sim 2\epsilon n$
  - $complexity(C_1) \to \infty$
  - All other components are trees or unicyclic
  - $L_2 = \Theta(\epsilon^{-2} \log \lambda) = \Theta(n^{2/3} \lambda^{-2} \log \lambda) \text{ (Note } \frac{L_1}{L_2} \to \infty)$

(5) Very supercritical  $(p = \frac{c}{n}, c > 1)$ 

- $L_1 \sim yn$  where y = y(c)
- $complexity(C_1) \to \infty$
- All other components are trees or unicyclic
- $L_2 = \Theta(\log n)$

17.2. GALTON-WATSON process. Let X be a distribution on  $\mathbb{Z}_{\geq 0}$ . Set  $Z_0 = 1$ . Recursively define  $Z_n = \text{sum of } Z_{n-1}$  i.i.d. copies of X, for  $n \geq 1$ . (in other words, X-offspring distribution,  $Z_0$ -root of a tree,  $Z_n$ -size of n-th generation) Examples:

- (1)  $\mathbb{P}(X = c) = 1$ . Then,  $Z_n = 0$  if c = 0 and  $Z_n > 0$  if c > 0.
- (2)  $\mathbb{P}(X=0) = q$ ,  $\mathbb{P}(X=m) = p$ ; p+q=1.

Question: Does this process continue forever?

Answer:  $p_X$ -extinction probability =  $\lim_{n\to\infty} \mathbb{P}(Z_n = 0)$ , and if  $p_X < 1$ , the process continues forever with positive probability.

Probability generating function of X:  $f(x) = f_X(x) := \sum_{i=0}^{\infty} \mathbb{P}(X = i) x^i$ . (eg:  $f_X(x) = q + px$ ).

**Theorem 17.1.** (1) If  $\mathbb{E}X \leq 1$ , then we have  $p_X = 1$  (except the degenerate case  $\mathbb{P}(X = 1) = 1$ )

(2) If  $\mathbb{E}X > 1$  and  $\mathbb{P}(X = 0) > 0$ , then  $p_X = x_0$  where  $x_0$  is the unique solution of f(X) = X in (0, 1).

Continuing on eg:

 $f_X(x) = q + px^m$ ,  $\mathbb{E}X = mp$ ; then  $p_X = 1$  if  $mp \le 1$  and  $p_X < 1$  if mp > 1

Two *ancient* references: "Probability problems in nuclear chemistry", Schrodinger 1945, "Survival of family names", Galton, Watson (1874)

18. (Wednesday, October 10)

Let  $v_0 \in [n]$  in G(n, p). Explore as follows:

Let  $v_1, \ldots, v_m$  be neighbors of  $v_0$  in G(n, p). Mark  $v_0$  as saturated.

Let  $v_{11}, v_{12}, \ldots$  be neighbors of  $v_1$  in  $[n] \setminus \{v_0\}$ . Mark  $v_1$  as saturated.

Let  $S_i$  be the set of vertices that are labelled, and i + 1 is the smallest non-saturated vertex. Look for neighbors of  $v_{i+1}$  to  $[n] \setminus S_i$ .

Equivalent Galton-Watson process:  $Z_0 = 1$ ,  $Z_{n+1} = Z_n + X - 1$ , where  $Z_n \sim \{seen\} \setminus \{saturated\}$  and X is a RV on  $\mathbb{Z}_{\geq 0}$ . Exploring G(n, p),  $X_i = \#$ new neighbors at step  $i = Bin(n - |S_{i-1}|, p)$ .

Note that if  $S_i$  is small compared to  $n, X_i \sim Bin(n, p)$ .

- (\*) Also note that, if  $p = \frac{c}{n}$ , c < 1, then in GW branching process with offspring distribution Bin(n, p),  $p_X = 1$  (extinction probability) and hence we expect that all components are small.
- (\*) If  $p = \frac{c}{n}, c > 1$ , then  $p_X < 1$  and we might expect large components to appear.

**Theorem 18.1.** Let  $p = \frac{c}{n}, c < 1$ . Then whp,

- (1)  $L_1 \leq \frac{3}{(1-c)^2} \log n$
- (2) All components are trees or unicyclic.

We need one more tool: Chernoff bound revisited; Let X = Bin(N, p) and  $\mu = Np$ . Then  $\mathbb{P}(X \ge \mu + t) \le \exp\left(\frac{-t^2}{2(\mu + t/3)}\right)$  and  $\mathbb{P}(X \le \mu - t) \le \exp\left(\frac{-t^2}{2\mu}\right)$ .

Proof of (1). Let  $v \in [n]$ . Explore G(n, p). Let  $X_i = \#$ vertices discovered at step *i*.

 $\mathbb{P}(v \text{ belongs to a component of size} \ge k = k(n)) \le \mathbb{P}(\sum_{i=1}^{k-1} X_i \ge k-1)$ 

then bound with  $X_i \leq X_i^+ \sim Bin(n, p)$ ,

$$\leq \mathbb{P}(\sum_{i=1}^{k-1} X_i^+ \geq -1)$$
$$\leq \mathbb{P}(Bin(kn, p) \geq k - 1)$$

Now set  $p = \frac{c}{n}, c < 1$  and  $k \ge \frac{3}{(1-c)^2} \log n$ . Take N = kn.

$$\mathbb{E}[Bin(kn,p)] = knp \ge \frac{3\log n}{(1-c)^2}c$$

$$\begin{split} \mathbb{P}(Bin(kn,p) \ge k-1) \le \exp\left(\frac{-t^2}{2(k\mu+t/3)}\right) & \text{where } \mu = ck, t = (1-c)k-1\\ \le \exp\left(\frac{-((1-c)k-1)^2}{2(ck+\frac{(1-c)k}{3})}\right)\\ \le \exp\left(\frac{-(1-c)^2}{2}k\right) = O(n^{-3/2}) = o(n^{-1}) \end{split}$$

Summing over all possible choices for  $\boldsymbol{v}$ 

$$\mathbb{P}\left(\exists \text{ component of size} \ge \frac{3\log n}{(1-c)^2}\right) \le O(n^{-0.5}) \to 0$$

Proof of (2). By (1), WLOG we may assume that  $L_1 = O(\log n)$ .

**Exercise 18.2.** Any connected graph with  $e - v + 1 \ge 2$  contains a subgraph (i) two cycles connected by a path, (ii) shares a vertex, (iii) shares common edge(s).

For such a connected graph,

 $\mathbb{E}(\#\text{copies in } G(n,p)) \leq n^{k+l+m} p^{k+l+m+1} \leq (np)^{k+l+m} p \leq c^{k+l+m} p = O(n^{-1}).$ We may assume  $k, l, m \leq O(\log n)$ , so  $\#\text{choices} = O((\log n)^3)$ Whp all components have e - v + 1 < 2, hence either is a tree or unicyclic.

**Exercise 18.3.** If X = Po(c), c > 1, what is  $p_X$ ?

г		
L		
L		
L		

## 19. (FRIDAY, OCTOBER 12)

Consider  $X \in P_0(c)$  for c > 1. We seek to calculate  $\rho_X$ , the extinction probability for the branching process determined by X. We have  $f_X(x) = \sum_{i=0}^{\infty} \frac{c^i x^i}{i!} e^{-c} = e^{-x} e^{cx} = e^{c(1-x)}$ . So  $\rho_X$  is the unique solution to  $x = e^{c(x-1)}$  in the interval (0, 1). Let y = 1 - x, then our equation becomes  $1 - y = e^{cy}$ . Then  $\rho_X = 1 - \beta$  where  $\beta$  is the unique solution in (0, 1) to  $\beta + e^{-c\beta} = 1$ . Note that  $\beta$  is the survival probability of our branching process. Also note that as  $c \to \infty$  we have  $\beta \to 1$ .

Now let  $Y_n \in Bin(n,p)$  for  $p = \frac{c}{n}$  and c > 1. Consider a sequence of branching processes determined by  $Y_n$ . We seek  $\rho_{Y_n}$ . Then we have  $f_{Y_n}(x) = \sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} x^i = \sum_{i=0}^n \binom{n}{i} (px)^i (1-p)^{n-i} = (1-p+px)^n$ . Remember  $p = \frac{c}{n}$  and let  $n \to \infty$ , then  $f_{Y_n}(x) \to e^{c(x-1)}$  for every fixed x so  $\rho_{Y_n} \to 1-\beta$  where  $\beta = \beta(c)$ is the unique solution of  $\beta + e^{-c\beta} = 1$  in the interval (0,1). In fact the preceding

**Theorem 19.1.** If  $p = \frac{c}{n}$  for c > 1 constant, then with high probability  $L_1 \approx \beta n$ where  $\beta \in (0,1)$  is the unique solution to  $\beta + e^{-c\beta} = 1$  and  $L_2 \leq \frac{16c}{(1-c)^2} \log(n)$ .

argument holds if we merely have  $pn \to c$  as  $n \to \infty$  for c > 1 fixed.

To prove this we'll let  $k_{-} = \frac{16c}{(1-c)^2} \log(n)$  and  $k_{+} = n^{2/3}$ . We want to show that G(n, p) has no components of order k for  $k_{-} \leq k \leq k_{+}$ . To do this we consider a searching process beginning at a single vertex. Either this process ends in fewer than  $k_{-}$  steps or at the kth step there are at least  $\frac{(c-1)k}{2}$  unsaturated vertices. Let  $X_i$  be the number of vertices found at step i, and let  $X_i^- \in Bin(n - \frac{c+1}{2}k_{+}, p)$  each  $X_i^-$  i.i.d.. Note in order to check if the process starting at v produces after step k at least  $\frac{(c-1)k}{2}$  unsaturated vertices of this component. Now Pr  $\left[ \text{after } k \text{ steps we have fewer than } \frac{(c-1)k}{2} = \frac{(c+1)k}{2} \text{ vertices of this component. Now Pr } \left[ \text{after } k \text{ steps we have fewer than } \frac{(c-1)k}{2} \text{ unsaturated vertices} \right] \leq \Pr\left[ \sum_{i=1}^{k} X_i^- \leq k - 1 + \frac{(c-1)k}{2} \right] \approx Bin(kn, \frac{c}{n})$  where  $\sum_{i=1}^{k} X_i^- = Bin\left(k(n - \frac{c+1}{2}k^+), \frac{c}{n}\right)$ . Recall now for  $X \in Bin(N, p)$  we have  $\mu = Np$  and  $\Pr(X < \mu - t) \leq e^{-t^2/2\mu}$ . In our case  $\mu = kn\frac{c}{n} = kc$  and  $t = kc - \frac{(c+1)k}{2} = \frac{k(c-1)}{2}$ . Then  $\Pr[\text{claim fails}] \leq \sum_{k=k_-}^{k_+} e^{-((c-1)k)^2/8kc} = \sum_{k=k_-}^{k_+} e^{-(c-1)^2k_-/9c}$ . This simplifies to  $n^{2/3}n^{-16/9}$ .

However this is only for v a vertex in an intermediate size component. Summing over all vertices we have an upper bound on the number of intermediate size components  $nn^{2/3}n^{-16/9} = n^{-1/9} \to 0$  as  $n \to \infty$ . Next we will show that there is at most one component of size at least  $k_+$ . Suppose the exploring processes starting with v' and v'' both result in a component of size at least  $k_+$ . By the previous argument we will have at least  $\frac{(c-1)k}{2}$  unsaturated vertices so  $\Pr[\text{no edges between } V' \text{ and } V''] \leq (1-p)^{(c-1)^2k^2/4}$  where V' is the unsaturated vertices for the process starting at v' and the same for V'' and v''. However this probability is  $(1-\frac{c}{n}) < e^{-c/n}$  so this probability is at most  $e^{-c(c-1)^2n^{1/3}/4} = o(n^{-2}) \to 0$  so there is at most one large component.

Now let  $\rho = \rho(n, p)$  be the probability that a vertex v is in a small component. We have  $\rho(n, p) \leq \rho_{X'}$  where  $X' \in Bin(n - k_-, p)$ . On the other hand  $\rho_{X''} + o(1) \leq \rho(n, p)$  where  $X'' \in Bin(n, p)$ . Both  $\rho_{X'}$  and  $\rho_{X''}$  tend to  $1 - \beta$  where  $\beta + e^{-c\beta} = 1$ . So we also have  $\rho(n, p) \to 1 - \beta$ . Now let Y be the number of vertices in small components. Then  $\mathbb{E}[Y] \approx (1 - \beta)n$ . In fact it can be show that  $\mathbb{E}[Y^2] = (1 + o(1))\mathbb{E}[Y]^2$  and  $\mathbb{E}[Y^2] = n^2\rho(n, p)\rho(n - O(k), p)$ . So by Chebyshev's inequality, with high probability  $Y \approx \mathbb{E}[Y]$ . Then by the second moment method the theorem holds.

## 20. Monday, October 22

There are many other kinds of random graphs.

- Bernoulli random subgraphs of non-complete graphs
- Sequence of finite graphs  $G_i$  with  $V(G_i) \to \infty$
- Infinite graphs, eg. lattice, branching tree, Cayley graphs of an infinite group
- Random regular graphs
- Uniform spanning trees

We are especially interested in random geometric graphs.

20.1. Random geometric graphs. (Ref: M. Penrose: Random Geometric Graphs)

**Definition 20.1.** Fix  $d \ge 2$ , take *n* points iid in  $\mathbb{R}^d$  according to your favorite distribution, eg. Gaussian, or uniform on a convex body. These *n* points are your vertices. There is an edge between two vertices *x* and *y* if the distance between them is less than r = r(n), that is d(x, y) < r. If there is an edge between *x* and *y* we write  $x \sim y$ . This random graph is sometimes denoted by  $G(X_n; r)$ .

Remark 20.2. In the graph above, edge events are not independent.

$$Pr[y \sim z | x \sim y, x \sim z] \ge Pr[y \sim z].$$

**Exercise 20.3.** These edges are pairwise independent.

**Definition 20.4.** Instead of  $X_n = \{n \text{ points in } \mathbb{R}^d\}$ , take Poisson point process for vertices.

**Poisson Process:** Continuous time counting process, a stochastic process  $\{N(t); t \ge 0\}$  such that

- $N(t) \ge 0$
- N(t) is an integer
- If  $s \le t$ , then  $N(s) \le N(t)$ ; N(t) N(s) is the number of events in (s, t].
- (1) Homogeneous Poisson Process

$$\Pr[N(t+\tau) - N(t) = k] = \frac{e^{-\lambda \tau} (\lambda \tau)^k}{k!}, (k = 0, 1, \dots)$$

where  $\lambda$  is the intensity constant.

- If intervals  $I_1, \ldots, I_m$  are disjoint then  $N(I_j)$  are independent
- N(I) is a Poisson r.v. only dependent on length of I.
- (2) Non-homogeneous Poisson Point Process

Let  $\lambda : [0, \infty) \to [0, \infty)$ . For every  $a \leq b$  set

$$\mu_{a,b} = \int_{a}^{b} \lambda(t) dt$$

N(b) - N(a) is Poisson $(\mu_{a,b})$ .

- Still have disjoint intervals are independent
- These are continuous Markov processes ("memoryless").

More generally, consider a Poisson point process on  $\mathbb{R}^d$ . Let  $f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$  be a measurable, bounded density function  $(\int_{\mathbb{R}^d} f dx = 1)$ . Let  $\lambda \in (o, \infty)$  be the intensity. Then there is a point process so that

(1) for every open set  $U \subset \mathbb{R}^d$ , the number of points in U is Poisson with mean

$$\mu_U = \lambda \int_U f dx$$

(2) Disjoint regimes  $\implies$  independent r.v.'s.  $\mathbb{E}[$ number of points in  $\mathbb{R}^d ] = \lambda.$ 

**Note:** Uniform Poisson point processes are totally fine but they give you infinitely many points ("continuum percolation").

Letting  $\lambda \to \infty$  is a possibility.  $\tilde{G}(\lambda; r)$  is a geometric random graph on Poisson process of intensity  $\lambda$ .

Idea:  $\tilde{G}(\lambda; r)$ ,  $(\lambda = n)$  looks a lot like  $G(X_n; r)$ . In  $G(X_n; r)$  choice of points is independent. In  $\tilde{G}(\lambda; r)$  number of points in disjoint regimes are independent.  $\tilde{G}(\lambda; r)$  and  $G(X_Z; r)$  look similar where  $Z \in Po(n)$ .

**Exercise 20.5.** Consider  $G(X_n; r)$  on  $[0, 1]^d$  where the distribution is uniform. Write down approximate formula for  $\mathbb{E}[number of triangles]$  as a function of n and r = r(n) up to a constant factor. Assume  $r \to 0$ .

21. October 29

## Random geometric graphs

For  $n \ge 2$ , let  $G(X_n; r)$  be a random geometric graph where  $X_n = \{x_1, x_2, \dots, x_n\}$  is a set of n i.i.d. points in  $\mathbb{R}^d$  underlying density f which is bounded and measurable.

Induced subgraph  $\Gamma$  is a connected and feasible graph on k vertices. (i.e. possible to appear as an induced subgraph)

**Exercise 21.1.**  $K_{1,m}$  is not feasible in  $\mathbb{R}^2$  for  $m \geq 7$ .

**Remark 21.2.** By Penrose,  $\mathbb{E}[\# \text{ of induced copies of } \Gamma] \sim c n^k r^{d(k-1)}$  where the constant c only depends on  $d, f, \Gamma$ .

**Exercise 21.3.** Same in  $\mathbb{R}^d$  for  $K_{1,m_d}$ . How does  $m_d$  grow with d?

**Remark 21.4.** 1. If  $n^k r^{a(k-1)} \to \mu \in (0,\infty)$  as  $n \to \infty$ , then the number of induced copies of  $\Gamma$  is Poisson distributed as  $n \to \infty$ .

2. If  $n^k r^{d(k-1)} \to \infty$  then CLT for the number of induced  $\Gamma$ .

3. Threshold for  $\Gamma$  :  $r \sim n^{-k/d(k-1)}$ .

**Exercise 21.5.** From induced subgraph counts, one may deduce subgraph counts as a linear combination. For example,



**Question** 1. Critical exponent  $r = n^{-1/d}$ ?? True, but what does this mean? Getting things up to constant average degree makes for c qualitative / global changes on graphs, a.k.a giant component average, etc.

2. Threshold for connectivity for  $G(X_n, ; r)$ ? This is tricky; in particular, answer depends a lot on underlying distribution function f.

Case 1. Uniform distribution on convex body (i.e. convex, compact set with nonempty interior). Look at cube  $[0,1]^d$ . Try to guess threshold for connectivity. Certainly we need minimum degree  $\delta(G) \geq 1$  in order to ensure connectivity. Fix  $i \in [n]$ .

> $Pr[x_i \text{ is near the boundary within } r] \to 0$  $\mathbb{E}[\# \text{ points in a ball } B(x_i; r)] = ncr^d$

56

In  $G(X_n; r)$ ,

$$\begin{split} \mathbb{E}[\deg(v)] &\sim np \\ &\sim c \log n, \quad \text{ if } p = \frac{c \log n}{n}. \end{split}$$

deg(v) is distributed like  $Po(\mu)$  with mean  $\mu$ .

$$Pr[\deg(v) = 0] \sim e^{-\mu} \sim e^{-c \log n} \sim n^{-c}$$
$$\mathbb{E}[\#\text{isolated vertices}] \sim n^{1-c}$$

More precisely, set  $p = \frac{\log n + c}{n}$  for  $c \in (-\infty, \infty)$ .

 $\mathbb{E}[\#$ isolated vertices]  $\rightarrow e^{-c}$ 

For 
$$G(X_n, r)$$
, makes  $\mathbb{E}[\deg v] \sim \log n$ . Is  $\deg(v)$  Poisson distributed??  
By proposition,  $X_n$  'is close to' Poisson point process of intensity  $nf$ .  $cnr^d = \log n$   
 $Guess = \theta \left(\frac{\log n}{n}\right)^{1/d}$  (threshold of conectivity)  
Penrose: For nice bounded region with  $f$  bounded below in this region

$$\lim_{n \to \infty} \frac{nr^d}{\log n} = \frac{1}{\delta_o},$$

where  $\delta_0$  is the infimum of f on support if uniformly distributed.

Let R = threshold for connectivity = min r such that  $G(X_n; r)$  is connected, and let B = volume of a ball of radius r in  $\mathbb{R}^d$ .

Standard multivariable Gaussian:  $\mu = 0$  mean,  $\Sigma =$  identity matrix. If

$$R\sqrt{2\log n} - (d-1)\log\log n + \frac{d-1}{2}\log\log\log n + \log k_d \to \infty,$$

then w.h.p connected.

If it goes to  $-\infty$ , w.h.p disconnected.

W.h.p., largest point  $||x|| \sim x \sim \sqrt{2 \log n}$ Threshold:  $R \sim \frac{(d-1) \log \log n}{\sqrt{2 \log n}}$ 

Lesson: Must r much bigger to assume connectivity when f is Gaussian.

22. Wednesday, October 24

22.1. More on Random Geometric Graphs and Poisson point processes. First recall the following definition concerning Poisson point processes:

**Definition 22.1.** For a Poisson point process with intensity  $\lambda$ , density function f, and a Borel set  $A \subset \mathbb{R}^d$ ,  $P_{\lambda}(A) := \mathbb{E}[\# pts \text{ in } A]$  is a Poisson distributed random variable with mean  $\mu = \lambda \int_{A} f(x) dx$ .

**Lemma 22.2.** Given a fixed  $\lambda > 0$ , let  $N_{\lambda} \in Po(\lambda)$ , and  $P_{\lambda} = \{x_1, \ldots, x_{N_{\lambda}}\}$  be *i.i.d.* random points. Then  $P_{\lambda}$  is a Poisson point process with intensity  $\lambda$ .

**Proof of Lemma.** Let  $A_1, \ldots, A_k$  be a Borel-set partition of  $\mathbb{R}^d$ . Let  $n_1, \ldots, n_k$  be positive integers such that  $n_1 + \ldots + n_k = n$ . Then

$$\Pr[P_{\lambda}(A_i) = n_i \forall i] = \Pr[N_{\lambda} = n] \cdot \Pr[P_{\lambda}(A_i) = n_i \forall i \mid N_{\lambda} = n]$$
$$= \frac{e^{-\lambda}\lambda^n}{n!} \cdot \frac{n!}{n_1!n_2!\dots n_k!} \cdot \prod_{i=1}^k \left(\int_{A_i} f(x) dx\right)^{n_i}$$
$$= \prod_{i=1}^k \frac{\exp\left(-\lambda \int_{A_i} f(x) dx\right) \left(\lambda \int_{A_i} f(x) dx\right)^{n_i}}{n_i!}$$

The first part of the second line is because  $N_{\lambda} \in \text{Po}(\lambda)$ , and then the *n* points must be partitioned into the  $A_i$ . The extra integral in the last line comes from  $1 = \int_{\mathbb{R}^d} f(x) dx = \sum_{i=1}^k \int_{A_i} f(x) dx.$ 

So the  $P_{\lambda}(A_i)$  are independent random variables in  $Po(\mu_i)$  with  $\mu_i = \lambda \int_{A_i} f(x) dx$ . Hence  $P_{\lambda}$  is a Poisson point process with intensity  $= \lambda$ .

**Upshot**: One can prove something about  $X_n = \{x_1, \ldots, x_n\}$  as  $n \to \infty$  by proving something about  $P_{\lambda} = \{x_1, \ldots, x_{N_{\lambda}}\}, N_{\lambda} \in \text{Po}(\lambda)$  and letting  $\lambda \to \infty$ . ["Poissonization."]

**Exercise 22.3.** Let  $X_n$  be *n* points *i.i.d.* uniformly in  $[0,1]^d$ ,  $d \ge 2$ . Assume  $r \to 0$ . What is  $\mathbb{E}[\# \text{ of triangle subgraphs}]$  in  $G(X_n, r)$ ?

Let  $p = \Pr[x_i, x_j, x_k \text{ form a triangle}]$ , then  $\mathbb{E}[\# \text{ of triangles}] = \binom{n}{3}p$  by linearity of expectation.

$$p = \Pr[x_i \sim x_j, x_j \sim x_k, x_k \sim x_i]$$
  
=  $\Pr[x_i \sim x_j] \cdot \Pr[x_j \sim x_k] \cdot \Pr[x_k \sim x_i \mid x_i \sim x_j \sim x_k]$ 

This last term is the probability that two points in a unit ball around a point are also at most unit distance apart, and it depends only on the dimension d. Call this probability  $c'_d$ . The first two terms are independent, and hence equal, and will be some fixed constant depending on d multiplied by the volume of the ball around a point. So this is  $c_d r^d$  for some constant  $c_d$ , which depends only on d and the distribution function f.

$$p \sim (c_d r^d)^2 \cdot c'_d \Longrightarrow \mathbb{E}[\# \text{ of triangles}] \sim cn^3 r^{2d}, \text{ with } c = c(d, f)$$

Suppose  $n^3 r^{2d} = 1$ ,  $r = n^{\frac{-3}{2d}}$ :

If  $r = o(n^{\frac{-3}{2d}})$ , whp there are no triangles. If  $r = \omega(n^{\frac{-3}{2d}})$ , whp there are lots of triangles.

In fact, if  $n^3 r^{2d} \to c$ , Penrose shows that the number of triangles is Poisson distributed in the limit. Similarly,  $\mathbb{E}[\# \text{ of } C_4\text{'s}] \sim cn^4 r^{3d}$ . More generally, in Ch. 3 of Penrose, he shows the following:

**Lemma 22.4.** Let  $\Gamma$  be a connected graph on  $k \geq 2$  vertices. Then  $n^{-k}r^{-d(k-1)}\mathbb{E}[\# \Gamma \text{ subgraphs}] \to \mu$ , which only depends on  $\Gamma, d$ , and f.

## 23. Wednesday, October 31

23.1. Cleanup discussion. This lecture, we continue our discussion on the random geometric graphs  $G(X_n; r)$ .<sup>13</sup> We discussed that the threshold for connectivity of  $G(X_n, r)$  depends on the underlying density. For uniform densities on a convex body, the threshold is  $R \sim C\left(\frac{\log(n)}{n}\right)^{1/d}$  (somewhat akin to our results for G(n, p)), whereas for the standard multivariate normal distribution, the threshold is  $R \sim \frac{(d-1)\log(\log(n))}{\sqrt{2\log(n)}}$  (a much larger threshold radius than the G(n, p) case).

These cases, however distinct their bounds, have in common with the standard G(n, p) case that the threshold for connectivity is the *same* as that for the necessary condition of  $\delta(G) \geq 1$  whp as  $n \to \infty$ , where  $\delta(G)$  stands for minimum degree of G. We cannot surmise, however, that all the properties of G(n, p) carry over to  $G(X_n; r)$ ; the next exercise is a case in point.

**Exercise 23.1.** Let f be a uniform distribution on a cube,  $[0,1]^d$ , and assume we have  $G(X_n;r)$  with underlying distribution f for  $X_n$  and  $r \to 0$  as  $n \to \infty$ . Show that  $G(X_n,r)$  is not an expander graph: as  $n \to \infty$ , when  $h(G(X - n;r)) \to 0$ .<sup>14</sup> This is in contrast to the case of G(n,p), since if p is large enough so that when G(n,p) is connected, then h(G(n,p)) is bounded away from 0 whp.

**Exercise 23.2.** For comparison, try finding  $h(G(X_n; r))$  on the torus  $\mathbb{T}^d$ , again with uniform distribution, and r the constant so that the probability of any two vertices connecting is  $\frac{1}{2}$ .

**Question 23.3.** Compare the fixed-r case of  $G(X_n; r)$  hinted at above with  $G\left(n, \frac{1}{2}\right)$ . Look at clique numbers and independence numbers. Weak bounds are possible even with simple estimates; how well can you do?

23.2. Giant components on Infinite Random Geometric Graphs. Again rehashing the same sorts of questions that we did considering characterizations of G(n, p), the next problem is to characterize when a giant component appears in random geometric graphs. Rather than looking at sequences of finite graphs with n vertices as  $n \to \infty$ , we choose to look at an infinite graph, instead. To do so, we recall the process cousin to  $G(X_n; r)$ , the  $G(P_\lambda; r)$  process. We still connect vertices with mutual distance less than r, but now the points are chosen by a Poisson point process with intensity  $\lambda, \lambda \in (0, \infty)$ . Recall the salient properties of such a process.

<sup>14</sup>Recall that here, we define  $h(G) = \min_{|S| \le \lfloor \frac{n}{2} \rfloor} \left\{ \frac{\#E[S, S^{\complement}]}{|S|} \right\}.$ 

<sup>&</sup>lt;sup>13</sup>As a reminder, these graphs are defined by choosing n points according to i.i.d. random variables with a common density function f, then connecting vertices within r = r(n) of each other; r is usually a function of n.

- (1) For any Borel set  $A \subset \mathbb{R}^d$ ,  $P_{\lambda}(A) := \mathbb{E} \{ \# \text{ points in } A \}$  is a Poissondistributed random variable with mean  $\lambda \int_A f(x) dx$  for some measurable, bounded, *density* function  $f : \mathbb{R}^d \to \mathbb{R}$ .
- (2) For disjoint Borel Sets A, B,  $P_{\lambda}(A)$  and  $P_{\lambda}(B)$  are disjoint.

For today's purposes, we drop the requirement that f is density, so that we can take  $f \equiv 1$ ; in other words, replace requirement 1 with

(1') For any Borel set  $A \subset \mathbb{R}^d$ ,  $P_{\lambda}(A) := \mathbb{E} \{ \# \text{ points in } A \}$  is a Poissondistributed random variable with mean  $\lambda \cdot \operatorname{vol}(A)$ , where  $\operatorname{vol}(A)$  is the volume (Lebesgue measure) of A.

Therefore, our process will be called *uniform* on all of  $\mathbb{R}^d$ . We will give this process the special name  $H_{\lambda}$ . Note that this process  $H_{\lambda}$  will have infinitely many vertices whp.

To discuss giant components, it is easier to discuss a particular giant component's emergence than to discuss its existence in general; therefore, we add  $\vec{0} \in \mathbb{R}^d$  into our vertex set so that we can talk about the size of the component containing  $\vec{0}$ .

Therefore, we consider the process  $G \sim G(H_{\lambda} \cup \{\vec{0}\}; 1)$ . That is, we connect vertices separated by a distance strictly less than  $1.^{15}$ 

**Definition 23.4.** For  $k \ge 1$ , let  $p_k(\lambda) := Pr(\text{the component of } G \text{ containing } \vec{0} \text{ is of order } k)$ . Also define

$$p_{\infty} := 1 - \sum_{k=1}^{\infty} p_k(\lambda)$$
  
=  $Pr(\text{the component of } G \text{ containing } \vec{0} \text{ is of infinite order.})$ 

Note that 0's component contains 0, so there is no need for a  $p_0$ . Let  $\lambda_C := \inf \{\lambda > 0 : p_{\infty}(\lambda) > 0\}.$ 

This setup is akin to a model we will discuss later in the course: an infinite lattice, including vertices with uniform probability, and then connecting adjacent lattice points according to the lattice structure. Such a model is called *site percolation*, whose only real difference from our model is that the vertices can only appear at "fixed" positions. Our model, allowing vertices everywhere, is therefore called *continuum percolation*. Our goal today will be to use the following fact.

**Theorem 23.5** (Fundamental Result of (Continuum) Percolation).  $0 \nleq \lambda_C \gneqq \infty$ .

This proof can be found in the aptly named reference, *Continuum Percolation* by Meester and Roy ([9]).

<sup>&</sup>lt;sup>15</sup>Note that the Poisson point processes have no  $n \to \infty$  mechanism, so we cannot let  $r \to 0$ .

23.3. Giant components of Random Geometric Subgraphs on Sub-Cubes. M. Penrose, in turn, addressed the question of how this infinite-graph model relates to the limit of finite random geometric graphs with  $|V| \rightarrow \infty$ . He writes up this work in Chapters 9 and 10 of [12]. We now overview this discussion, and attempt to emphasize in what ways we fulfill our strategy of an analogue of the Erdös-Rényi phase transition at the critical value for the existence of a giant component.

**Definition 23.6.** Let  $H_{\lambda,s}$  be the restriction of  $H_{\lambda}$  to the cube  $\left[-\frac{s}{2}, \frac{s}{2}\right]^d$ . (The parameterization is chosen such that the side-length is s when the parameter is s.) Recall that  $L_1(G)$  is the size of the largest component of G a graph.

**Theorem 23.7.** If  $\lambda \neq \lambda_C$ , then as  $s \to \infty$ ,

$$s^{-d}L_1(G(H_{\lambda,s};1)) \to \lambda p_\infty(\lambda).$$

We interpret this by noting that by our parameterization,  $s^d$  is the volume of the cube of the support of the distribution for  $H_{\lambda,s}$ . Therefore, we may rewrite the above as saying that

$$\frac{L_1\left(G(H_{\lambda,s};1)\right)}{\operatorname{Vol}\left(\left[-\frac{s}{2},\frac{s}{2}\right]^d\right)} \to \lambda p_{\infty}(\lambda).$$

Again, we have a sub-critical and super-critical regime, depending on the value of  $\lambda$ .

(1) In the supercritical case,  $\lambda > \lambda_C = \inf \{\lambda > 0 : p_{\infty}(\lambda) > 0\}$ , then  $p_{\infty}(\lambda) > 0$  (since the set whose infimum is taken is clearly upward closed; we will elaborate on this point later) and hence

$$\frac{L_1\left(G(H_{\lambda,s};1)\right)}{\operatorname{Vol}\left(\left[-\frac{s}{2},\frac{s}{2}\right]^d\right)} \to \lambda p_{\infty}(\lambda) > 0.$$

Therefore, the largest component has a positive fraction of *all* points in the given cube. This parallels the very supercritical regime on Erdös-Rényi graphs: if  $p = \frac{c}{n}$  for c > 1, then  $L_1 \sim yn$  whp, where y is some constant depending on c, so whp  $\frac{L_1}{n}$  is a positive constant; i.e., the number of vertices in the given component over the total number of vertices is (whp) at least some positive constant.

(2) In the subcritical case,  $\lambda < \lambda_C$ , then  $p_{\infty}(\lambda) = 0$ , so whp

$$\frac{L_1\left(G(H_{\lambda,s};1)\right)}{\operatorname{Vol}\left(\left[-\frac{s}{2},\frac{s}{2}\right]^d\right)} \to 0.$$

This again agrees with the Erdös-Rényi case: in the very subcritical case, where  $p = \frac{c}{n}$  for c < 1, then  $L_1$  is  $\Theta(\log(n))$ , and hence the proportion

of vertices in the largest component to all vertices is whp bounded by a constant times  $\frac{\log(n)}{n}$ , which tends to 0 as n tends to  $\infty$ .

Things are subtler, however, for  $\lambda = \lambda_C$ . In particular,  $p_{\infty}(\lambda)$  is not known, and our current state of knowledge suggests the following conjecture.

# Conjecture 23.8. $p_{\infty}(\lambda_C) = 0.$

This is known for continuum percolation on  $\mathbb{R}^d$  when either d = 2 or  $d \ge 19$ ; the inbetween dimensions are unknown.<sup>16</sup> This work (of S. K. Smirnov and others) relies heavily on the rich geometry of certain lattices.

In addition, we have even better bounds on the exact size of the largest component in the sub-critical case, and the second-largest component in the supercritical case (but not so large as to be connected). As an aside, since results on  $L_1$  in the Erdös-Rényi case sometimes involved expression in n and  $\log(n)$ , the "size" of the graph, here we would expect expressions in the volume of the cube, namely  $s^d$  and  $\log(s^d)$ . Yet  $\log(s^d) = d \log(s)$ , so up to a constant, we can just write  $\log(s)$ .

(1) In the supercritical case,  $\lambda > \lambda_C$ , (but below the threshold for connectivity),

$$c_1 \le \frac{L_2(G)}{\log(s)^{d(d-1)}} \le c_2.$$

(2) In the subcritical case,  $\lambda < \lambda_C$ ,

 $L_1 \sim c \log(s).$ 

**Question 23.9.** Can we determine the size of  $L_1(G(H_{\lambda_C,s};1))$ ? A possible approach might be getting some better estimates by considering  $L_1(G(H_{\lambda(s),s};1))$ , where  $\lambda(s) \to \lambda_C$  as  $s \to \infty$ . Presumably, we would take the cases  $\lambda(s) \nearrow \lambda_C$  and  $\lambda(s) \searrow \lambda_C$  to get computable and tractable bounds.

**Question 23.10.** In the supercritical case,  $\lambda < \lambda_C$ , does  $\lim_{s \to \infty} \frac{L_2(G)}{\log(s)^{d(d-1)}}$  exist whp?

23.4. **Overview of Methods.** We do not choose to try to copy the proofs of the above statements here (they are in [12]), but we will discuss some tools used in the proofs. The results are mostly about Poisson point processes.<sup>17</sup>

The first exercise and theorem allow you to regard a lower-probability situation as a subset of a higher-probability situation, akin to the situation for the Erdös-Rényi random graphs, where if p < p', we could regard G(n, p) as a subgraph of G(n, p').

**Exercise.** If  $\lambda \in Po(\mu)$  and  $\lambda' \in Po(\mu')$  are independent, then  $\lambda + \lambda' \in po(\mu + \mu')$ .

 $<sup>^{16}\</sup>mathrm{Prof.}$  Kahle suggested that the d=3 case would be considerable work, possibly even to the level of a Fields Medal.

<sup>&</sup>lt;sup>17</sup>A common abbreviation for "Poisson point Process" is p.p.p.

**Theorem 23.11** (Superposition). Let P and P' be independent Poisson point processes on  $\mathbb{R}^d$ , with intensities  $g(\cdot), g'(\cdot)$ . Then  $P \cup P'$  is a Poisson point process with intensity  $(g + g')(\cdot)$ .

The next theorem allows you to reduce to any "sub-intensity" of a given intensity while maintaining the Poisson-ness of the point process.

**Theorem 23.12** (Thinning). Let P be a Poisson point process with intensity  $g(\cdot)$ . Let  $p : \mathbb{R}^d \to [0, 1]$  be measurable. For each point x in P, declare x to be **accepted** with probability p(x). The new point process, restricting the old process to the set of accepted points, is again a Poisson point process with intensity  $p * g(\cdot)$ .

The next theorem allows you to move homogenous point processes radially, if you scale by the appropriate constant. If you have a *homogenous* Poisson point process H on a region  $A \subset \mathbb{R}^d$  of intensity  $\lambda$ , and a > 0 is some positive constant, consider the point process aH on  $aA = \{a \cdot x : x \in A\}$  defined by scaling the points in H out by the constant a.

**Theorem 23.13** (Scaling). If H is a homogeneous, Poisson point process H on a region  $A \subset \mathbb{R}^d$  of intensity  $\lambda$ , and a > 0 is some positive constant, then aH is also homogenous and Poisson.

Finally, for lack of a better position, we mention this result, emphasizing that the giant component should be much larger than all other components.

**Theorem 23.14.** If  $\lambda > \lambda_C$ , then almost surely,  $G(H_{\lambda}, 1)$  has a unique infinite component.

## 24. (November 2):

This lecture covers some basic results of percolation theory, following Bollobás & Riordan. The idea is to consider an infinite connected graph  $\Lambda$ , and look at properties of random subgraphs. Often,  $\Lambda$  will have some symmetry, sometimes including a vertex-transitive automorphism group, as is the case for the lattices  $\mathbb{Z}^d$  for  $d \geq 2$  and the *d*-regular Cayley tree.

There are two main perspectives from which to view percolation on a graph:

- Bond percolation: Each edge (bond) is included (open) with probability p, where individual bonds are independent random variables; all vertices are included. The probability measure on the space of subgraphs is denoted  $\mathbb{P}^{b}_{\Lambda,p}$ , or simply  $\mathbb{P}_{p}$  if  $\Lambda$  and b are clear from context.
- Site percolation: Each site is included with probability p; the open bonds are those from the induced subgraph on the included sites. The probability measure is denoted  $\mathbb{P}^{s}_{\Lambda,p}$ , or simply  $\mathbb{P}_{p}$ .

The resulting subgraph is denoted  $\Lambda_p^b$  (or  $\Lambda_p^s$ ), or usually  $\Lambda_p$  if the type of percolation is clear from context.

Note that site percolation is in some sense "more general" than bond percolation since bond percolation on  $\Lambda$  is equivalent to site percolation on  $L(\Lambda)$ , the line graph of  $\Lambda$  (ignoring sites in  $\Lambda_p^b$  which have no open bonds). Here  $L(\Lambda)$  is the graph with a vertex  $v_e$  associated to each edge e of  $\Lambda$ , and  $v_e \tilde{v}_{e'}$  if and only if e, e' meet at a vertex in  $\Lambda$ .

It is often convenient to use the natural "coupling" of measures  $\mathbb{P}^b_{\Lambda,p}$  with  $0 \leq p \leq 1$ , in which  $\Lambda^b_{p_1}$  is viewed as a subgraph of  $\Lambda^b_{p_2}$  if  $p_1 \leq p_2$ . In other words, we can think of choosing a subgraph of  $\Lambda$  by labeling each edge e with a random value  $\alpha(e) \in [0, 1]$ , and include e in  $\Lambda_p$  if  $p \geq \alpha(e)$ . A similar idea holds for site percolation.

We will use the following notation:

- For sites x and  $y, x \to y$  means there exists an open path from x to y;
- For a site  $x, x \to \infty$  means there exists an infinite open path starting at x;
- Denote the connected component including x by  $C_x$ ;
- Write  $\Theta_x(p)$  for  $\mathbb{P}_p(x \to \infty)$ , the probability that  $x \to \infty$  in  $\Lambda_p$ .

Note that for locally finite graphs  $\Lambda$ , König's infinity lemma implies that  $|C_x| = \infty$  if and only if  $x \to \infty$ ; for this reason, we will assume from now on that  $\Lambda$  is locally finite and has countably many vertices.

Let x and y be sites in  $\Lambda$  at distance d from one another. Then  $\Theta_x(p) \ge p^d \Theta_y(p)$ , since the right-hand side is the probability that  $y \to \infty$  and the entire length-d path  $y \to x$  is open. This has the following very strong implication: either  $\Theta_x(p) = 0$ for all  $x \in \Lambda$ , or  $\Theta_x(p) > 0$  for all  $x \in \Lambda$ . This is true even if  $\Lambda$  does not have a vertex-transitive automorphism group, or indeed any symmetry at all; it relies only on the connectedness of  $\Lambda$ .

Now, for a given  $x \in \Lambda$ , clearly  $\Theta_x(p)$  is an increasing function of p. Hence for any  $\Lambda$  there exists a critical probability  $p_H \in [0, 1]$  such that  $\Theta_x(p) = 0$  when  $p < p_H$ , and  $\Theta_x(p) > 0$  when  $p > p_H$  (independent of x). As with  $p_\infty(\lambda)$  previously, this says nothing about the behavior of  $\Theta_x(p)$  at the critical probability  $p = p_H$ ; in general, it is an open problem whether or not  $\Theta_x(p) = 0$  when  $p = p_H$ , as one might expect. Note that it is possible that  $p_H = 0$  or  $p_H = 1$  (as an example of the latter case, take  $\Lambda = \mathbb{Z}$ ). However, for many interesting choices of  $\Lambda$ , including for example  $\mathbb{Z}^d$  for  $d \geq 2, 0 < p_H < 1$ .

Before proving the next proposition, we state the following theorem without proof:

**Theorem 24.1.** (Kolmogorov zero-one law) Let  $X = (X_1, X_2, ...)$  be an infinite sequence of independent random variables, and A an event in the  $\sigma$ -field generated by X such that A is independent of  $(X_1, ..., X_n)$  for each n (i.e. A is a tail event). Then either P(A) = 0 or P(A) = 1.

**Proposition 24.2.** For fixed  $\Lambda$ , let E be the event "there is an infinite open cluster." If  $p < p_H$ , then  $\mathbb{P}_p(E) = 0$ , and if  $p > p_H$ , then  $\mathbb{P}_p(E) = 1$ .

Proof. First, assume  $p < p_H$ . Then  $\mathbb{P}_p(E) \leq \sum_{x \in \Lambda} \Theta_x(p)$ ; this is a sum of infinitely many numbers, all of which are exactly 0 (not merely tending to 0) so  $\mathbb{P}_p(E) = 0$ . On the other hand, assume  $p > p_H$ . Order the edges of  $\Lambda$  so that the *j*th edge corresponds to the random variable  $X_j$ . Then E is independent of  $(X_1, \ldots, X_n)$ for any n, since existence of an infinite component is unaffected by the inclusion or exclusion of any finite set of edges. Hence E is a tail event, so by Kolmogorov's zero-one law,  $\mathbb{P}_p(E) = 0$  or  $\mathbb{P}_p(E) = 1$ . But for any x,  $\mathbb{P}_p(E) \geq \Theta_x(p) > 0$ , so in fact  $\mathbb{P}_p(E) = 1$ .

One says that percolation occurs if  $\Theta_x(p) > 0$ ; that is, if  $\mathbb{P}_p(E) = 1$ .

Here is one example where  $p_H$  is easy to calculate exactly: let  $\Lambda$  be the kbranching tree  $T_k$ , consisting of a root vertex x with k descendants  $x_1, x_2, \ldots x_k$ , each of which has k descendants, and so on (so  $T_k$  is an infinite tree where one vertex has degree k and all other vertices have degree k + 1). Then in site percolation,  $\Theta_x(p) = 1 - p_X$ , the complement of the extinction probability  $p_X$  for a Galton-Watson process with X = Bin(k, p), where the *n*th step of the Galton-Watson process corresponds to the *n*th-level descendants of x (those at distance n from x). We know that the critical value of p for that Galton-Watson process is at p = 1/k, so  $p_H = 1/k$  for  $\Lambda$ . We can say more, namely that when  $p = p_H$ ,  $\Theta_x(p) = 0$ , since extinction occurs with probability 1 at p = 1/k. 25. (Nov 5, 2012)

**Question 25.1.** What is the critical probability for bond percolation on  $\mathbb{Z}^2$ ?

Another interesting critical probability is

$$p_T = \inf\{p : \mathbb{E}_p[|C_x|] = \infty\} ( \text{ or } \sup\{p : \mathbb{E}_p[|C_x|] < \infty\})$$

Compare this definition with the previous critical probability we defined

$$p_H = \inf\{p : \Theta_x(p) > 0\} ( \text{ or } \sup\{p : \Theta_x(p) = 0\})$$

By their definitions, we have  $p_T \leq p_H$ .

# Proposition 25.2.

$$\frac{1}{3} \le p_T \le p_H \le \frac{2}{3}$$

Let  $\Lambda$  be the  $\mathbb{Z}^2$  lattice. A useful notion is the dual of  $\Lambda$ , denoted  $\Lambda^*$ , is also a  $\mathbb{Z}^2$  lattice and that each edge in  $\Lambda$  corresponds to a unique edge in  $\Lambda^*$ .

**Remark 25.3.** Let H be a finite connected subgraph of  $\Lambda$  with vertex set C. Then there is a unique infinite component,  $C_{\infty}$ , of  $\Lambda - C$ .

## Definition 25.4.

 $\delta^{\infty}C = \text{ set of bonds of } \Lambda^* \text{ dual to bonds joining } C \text{ to } C_{\infty}$ 

**Proposition 25.5.**  $\delta^{\infty}C$  is a cycle with C in its interior

Proof. Let  $\vec{F}$  be set of  $C - C_{\infty}$  bonds oriented from C to  $C_{\infty}$ . For  $\vec{f} \in \vec{F}$ , let  $f^*$  be dual edge (oriented  $\frac{\pi}{2}$  rotated counterclockwise).

Claim: If  $f^* = uv$ , then there is a unique bond of  $\delta^{\infty}C$  leaving v.

Set  $R = \mathbb{Z}^2 - C - C_{\infty}$ . Note: There are no  $C_{\infty} - R$  bonds.

Suppose *abcd* is a  $1 \times 1$  square with v in the middle and u below. So  $a \in C$  and  $b \in C_{\infty}$ .

Case 1.  $d \in C$ . Then necessarily  $c \notin R$ . So  $c \in C$  or  $c \in C_{\infty}$ . If  $c \in C_{\infty}$ , then  $\vec{dc} \in \vec{F}$  and you leave through the top. If  $c \in C$ , then you leave through the right.

Case 2.  $c \in C_{\infty}$ . Then  $d \in R$ . If  $d \in C$ , then you leave through the top. If  $d \in C_{\infty}$ , then you leave through the left.

Case 3.  $c \notin C_{\infty}$ . Since  $c \notin R$ , we have  $c \in C$ . If  $d \in R$ , then you leave through the right. We get a contradiction with  $d \in C_{\infty}$ , since we can not get disjoint ac-bd paths on exterior of cycle *abcd*, by Kuratowski's theorem. This would result in a topological embedding of  $K_5$  into plane.

As a consequence of the proposition, either the origin is in an infinite component or some cycle separates 0 from  $\infty$  in  $\Lambda^*$ . Let  $\mu_n$  be the number of self-avoiding walks in  $\Lambda$  starting at 0 of length n. We get the following inequality

$$2 \cdot 2^n \le \mu_n \le 4 \cdot 3^{n-1}$$

Now we can prove the proposition.

Suppose  $p < \frac{1}{3}$ . Let  $C_0$  be the open cluster containing 0. For  $x \in C_0$ , there exists an open path from 0 to x. So  $|C_0| \leq$  the number of open paths starting from 0 =: X.

$$\mathbb{E}_p[|C_0|] \le \mathbb{E}_p[X] = \sum_{n=0} \mu_n p^n \le 1 + \frac{4}{3} \sum_{n=1}^{\infty} (3p)^{n-1} < \infty$$

Hence  $p_T \ge p$ . Since p was arbitrarily less than  $\frac{1}{3}$ , we have  $p_T \ge \frac{1}{3}$ .

Suppose  $p > \frac{2}{3}$ . Let  $L_k$  be the path of length k from (0,0) to (k,0) and S be a dual cycle surrounding  $L_k$  of length 2l. This forces S to cross the x-axis somewhere between  $(k + \frac{1}{2}, 0)$  and  $(l - \frac{3}{2}, 0)$ . Hence there are fewer than l choices for crossing edge  $e^*$ . So there are at most  $l\mu_{2l-1}$  choices for S.

Let  $Y_k$  be the number of open dual cycles surrounding  $L_k$ .

Then  $\mathbb{E}_p[Y_k] \le \sum l \ge k + 2l\mu_{2l-1}(1-p)^{2l}$ .

Since 3(1-p) < 1, this sum converges.

So  $\mathbb{E}_p[Y_k] \to 0$  as  $k \to \infty$ . There exists some k such that  $\mathbb{E}_p[Y_k] < 1$ .

Let  $A_k$  be event that  $Y_k = 0$ . Then  $\Pr_p[A_k] > 0$ . Let  $B_k$  be event that all k bonds on  $L_k$  are open.  $\Pr_p[B_k] = p^k$ .

Since  $A_k$  and  $B_k$  look at different bonds, the events are independent.

Hence  $\Pr_p[A_k \cap B_k] = p^k \Pr_p[A_k] > 0$ . There is an infinite open path on the event  $A_k \cap B_k$ . So we have positive probability of an infinite open path from 0. Hence  $p_H \leq p$ .

**Remark 25.6.** Actually it can be shown that  $\lim_{n\to\infty} \mu_n^{\frac{1}{n}} = \lambda$ . This  $\lambda$  is called the connective constant of  $\mathbb{Z}^2$ . We showed that  $2 \leq \lambda \leq 3$ . This method that we did can show that

$$\frac{1}{\lambda} \le p_T \le p_H \le 1 - \frac{1}{\lambda}$$

It has been shown that  $2.62 \leq \lambda \leq 2.68$ .

**Exercise 25.7.** Show that  $\lambda$  exists. Hint: Show that  $\mu_{m+n} \leq \mu_m \cdot \mu_n$ .

68

## 26. (NOVEMBER 7)

Our bounds for the previous critical probabilities can be improved upon.

# Theorem 26.1. $p_H = \frac{1}{2}$ .

In 1960's, Harris showed that if  $p < \frac{1}{2}$ , then  $\Pr_p(\exists \text{ infinite component}) = 0$ . In 1982, Kesten showed that if  $p > \frac{1}{2}$ ,  $\Pr_p(\exists \text{ infinite component}) = 1$ . In fact, it can be showed that at critically, there are no infinite components.

Given a  $k \times (l-1)$  rectangle, R, in  $\Lambda = \mathbb{Z}^2$ , there exists a  $(k-1) \times l$  dual rectangle  $R^h$  in  $\Lambda^*$ . The *h* corresponds to the horizontal dual rectangle.

Let H(R) be event of horizontal crossing in R (i.e. open path from left to right) and let  $V(R^h)$  be the event of vertical crossing in  $R^h$ .

**Lemma 26.2.** Whatever the states of bonds in R, exactly one of H(R) and  $V(R^h)$  occurs

Proof. Consider  $R \cup R^h$  as a part of Archimedian tiling of  $\mathbb{R}^2$  by regular octagons and squares. The octagons are around the vertices of the lattice and dual lattice and the squares correspond to the bonds. Color the lattice octagons blue and the dual lattice vertices orange. We color the squares depending on which bond is open. Draw boundary graph between orange and blue regions. This graph has vertices of degree 1 or 2 and the degree 1 vertices occur on corners.

Let x, y be vertices along the top and w, z be the vertices along the bottom. So connected component containing x is a path that ends at one of w, y, z. The path from x can not end at z, since path from x always has orange on its left.

If x is connected to w, then vertical crossing occurs to the left of x-w path. This is called the left-most crossing. If x is connected to y, then a horizontal crossing occurs to the right of the path. This is called the top-most crossing.

## **Corollary 26.3.** 1)

$$Pr_p[H(R)] + Pr_{1-p}[V(R^h)] = 1$$

for any  $p \in [0, 1]$  and any rectange R.

2) If R is  $(n+1) \times n$ , then

$$Pr_{\frac{1}{2}}(H(R)) = \frac{1}{2}$$

for any  $n \in \mathbb{N}$ .

3) If S is  $n \times n$  square, then

$$Pr_{\frac{1}{2}}(V(S)) = Pr_{\frac{1}{2}}(H(S)) \ge \frac{1}{2}$$

We will be using the Russo-Seymour-Welsh method to prove Harris' result. The following proof is by Bollobas and Riordan 2006.

Let  $R = [m] \times [2n], m \ge n$  and  $S = [n] \times [n]$ . Let X(R) be the event that there are open paths  $P_1$  and  $P_2$  such that  $P_1$  is a vertical crossing of S and  $P_2$  lies within R and connectes  $P_1$  to right side of R.

# Lemma 26.4.

$$\Pr_p[X(R)] \ge \Pr_p[H(R)] \cdot \Pr_p[V(S)] \cdot \frac{1}{2}$$

Proof. Suppose V(S) holds. Let LV(S) be the left-most open vertical crossing. For every possible path  $P_1$ , the event  $LV(S) = P_1$  does not depend on state of bonds of S to right of  $P_1$ .

Claim: For every possible  $P_1$ , we have  $\Pr_p[X(R)|LV(S) = P_1] \ge \frac{1}{2}\Pr_p[H(R)]$ . We get the proof of the claim by symmetry and reflecting  $P_1$  to above square. Let  $Y(P_1)$  be the event that  $P_1$  is joined to right side of R.

 $\Pr_p[Y(P_1)|LV(S) = P_1] = \Pr_p[Y(P_1)] \ge \frac{1}{2} \cdot \Pr_p[H(R)]$ 

If  $Y(P_1)$  holds and  $LV(S) = P_1$ , then X(R) holds. This implies that  $\Pr_p[X(R)|LV(S) = P_1] \ge \frac{1}{2} \cdot \Pr_p[H(R)]$ 

 $\operatorname{So}$ 

$$\Pr_p[X(R)] \ge \sum_{P_1} \Pr_p[X(R) \cap LV(S) = P_1] = \sum_{P_1} \Pr_p[X(R)|LV(S) = P_1]\Pr[LV(S) = P_1]$$

Hence we have

$$\Pr_p[X(R)] \ge \sum_{P_1} \frac{1}{2} \cdot \Pr_p[H(R)] \Pr[LV(S) = P_1] = \frac{1}{2} \cdot \Pr_p[H(R)] \Pr_p[V(S)]$$

Let  $h_p(m,n) = \Pr_p[H(R)]$ , where R is an  $m \times n$  rectangle. Let  $h(m,n) = h_{\frac{1}{2}}(m,n)$ .

**Corollary 26.5.** For  $n \in \mathbb{N}$ , we have

$$h(3n, 2n) \ge 2^{-7}$$

We would guess that  $\lim_{n\to\infty} h(3n, 2n)$  exists since for all n, we have

$$\frac{1}{2^7} \leq h(3n,2n) \leq \frac{1}{2}$$

**Conjecture 26.6.** Does  $h(an, bn) \to f(\frac{a}{b})$  as  $n \to \infty$ ?

70

## 27. November 9

# Continuing Harris-Kesten: $p_H \geq \frac{1}{2}$

Last time: Let  $R = m \times n$  rectangle,  $h_p(m,n) := \mathbb{P}(H(R))$ , and  $h(m,n) := h_{\frac{1}{2}}(m,n)$ . It was shown via duality that  $h(n+1,n) = \frac{1}{2} \forall n \geq 1$ . Then by monotonicity,  $h(n,n) \geq \frac{1}{2}$  for  $n \geq 1$ . The next step is to use the Russo-Seymour-Welsh method to go from squares to rectangles. Recall the following lemma and its corollary:

**Lemma 27.1.** For R an  $m \times 2n$  rectangle,  $m \ge n$ , and S an  $n \times n$  square, let X(R) be the event that a vertical crossing of S intersects a horizontal crossing of R. Then, by symmetry,  $\mathbb{P}(X(R)) \ge \mathbb{P}(H(R)) \cdot \mathbb{P}(V(S)) \cdot \frac{1}{2} \ge \mathbb{P}(H(R)) \cdot \frac{1}{4}$ .

Corollary 27.2.  $h(3n, 2n) \ge 2^{-7}$ .

**Proof of Corollary.** Let the bottom middle  $n \times n$  square be S, the leftmost  $2n \times 2n$  square be  $R_1$ , and the rightmost  $2n \times 2n$  square be  $R_2$ . Then let  $X_1$  be the event  $X(R_1)$ ,  $X_2$  be the event  $X(R_2)$ , and  $X_3$  be the event H(S). Then by the lemma, since  $R_1, R_2$  are in fact squares and thus  $\mathbb{P}(H(R_i) \geq \frac{1}{2})$ , we have  $\mathbb{P}(X_1), \mathbb{P}(X_2) \geq \frac{1}{8}$ . And  $\mathbb{P}(X_3) \geq \frac{1}{2}$ . Then  $\mathbb{P}(X_1, X_2, X_3) \geq \mathbb{P}(X_1) \cdot \mathbb{P}(X_2) \cdot \mathbb{P}(X_3) \geq \frac{1}{8} \cdot \frac{1}{8} \cdot \frac{1}{2} = 2^{-7}$ .

**Proposition 27.3.** For  $k \geq 3$  and all  $n \geq 1$ ,  $h(kn, 2n) \geq 2^{17-8k}$ 

**Proof of Prop.** Start with a  $2n \times M$  rectangle, R, with  $M \ge 2n$ . Write M as  $M = m_1 + m_2 - 2n$ , where  $m_1$  is measured from the left of the rectangle, and  $m_2$  from the right, so that  $R_1 := 2n \times m_1$  and  $R_2 := 2n \times m_2$  intersect on a  $2n \times 2n$  square, S. Let  $X_1 = H(R_1), X_2 = H(R_2)$ , and  $X_3 = V(S)$ , so that  $\mathbb{P}(H(R)) \ge \mathbb{P}(X_1, X_2, X_3) \ge \mathbb{P}(X_1) \cdot \mathbb{P}(X_2) \cdot \mathbb{P}(X_3)$ . So  $h(m_1 + m_2 - 2n) \ge h(m_1, 2n) \cdot h(m_2, 2n) \cdot \frac{1}{2}$ . In particular, for  $m = m_1 \ge 2n, 3n = m_2, h(m + n, 2n) \ge h(m, 2n) \cdot h(3n, 2n) \cdot \frac{1}{2} \ge 2^{-8} \cdot h(m, 2n)$ . Hence for  $m + n = kn, k \ge 3$ ,

$$h(kn, 2n) \ge 2^{-8} \cdot h((k-1)n, 2n) \ge \ldots \ge 2^{-7-8(k-3)} = 2^{17-8k}$$

In particular,  $h(6n, 2n) \ge 2^{-31}$ .

**Exercise 27.4.** Show that for constants a and b, there exists  $h_{a,b}$ , a constant depending only on a, b, such that  $h(an, bn) \ge h_{a,b} > 0$ .

**Remark 27.5.** The conjecture here is that  $\lim_{n \to \infty} h(an, bn)$  exists and is  $> 0 \forall a, b$ .

Another approach to proving the proposition is to use the construction X from the lemma. Here, let  $M = m_1 + m_2 - n$  with  $m_1, m_2$  intersecting on the middle n coordinates, and so S is the bottom  $n \times n$  square in this intersection. Let  $X_i =$ 

 $X(R_i)$  for i = 1, 2, and let  $X_3 = H(S)$ . Then  $\mathbb{P}(X_i) \ge h(m_i, 2n) \cdot \frac{1}{4}, \mathbb{P}(X_3) \ge \frac{1}{2}$ . So  $h(m_1 + m_2 - n, 2n) \ge h(m_1, 2n) \cdot h(m_2, 2n) \cdot 2^{-5}$ . Hence

$$h(5n,2n) \ge h(3n,2n)^2 \cdot 2^{-5} \ge 2^{-19}$$
  
$$h(6n,2n) \ge h(5n,2n) \cdot h(2n,2n) \cdot 2^{-5} \ge 2^{-19} \cdot \frac{1}{2} \cdot 2^{-5} = 2^{-25} > 2^{-31}$$

**Definition 27.6.** Recall that  $C_0$  is the open cluster containing the origin. Let  $r(C_0) = \max\{d(x,0) : x \in C_0\}$ , where d(x,0) is the graph distance, and not the shortest distance through the open cluster.

**Theorem 27.7.** (Harris)  $\Theta(\frac{1}{2}) = 0$ . In fact,  $\mathbb{P}_{\frac{1}{2}}(r(C_0) \ge n) \le n^{-c}$  for some absolute constant c > 0.

**Proof of Theorem.** When looking at the dual lattice  $\Lambda^*$ , the edge  $e^*$  is open iff e is closed. Let  $A_k$  be the square  $6N \times 6N$  annulus, with  $N = 4^k$ , centered on the origin. Let  $E_k$  be the event that  $A_k$  contains an open cycle in  $\Lambda^*$ , so that  $E_k \Rightarrow r(C_0) \leq 4^{k+1}$ . Then  $A_k$  is covered by the four  $6N \times 2N$  rectangles. Hence  $\mathbb{P}(E_k) \geq \mathbb{P}(\text{all four } 6N \times 2N \text{ rect's have longways crossing}) \geq h(6N, 2N)^4 \geq 2^{-100}$ . Let  $\epsilon = 2^{-100} > 0$ . So  $\mathbb{P}(r(C_0) \geq 4^{l+1}) \leq (1 - \epsilon)^l$ . To relate this back to n, let  $n = 4^{l+1} \Rightarrow l = \frac{\log(n) - \log(4)}{\log(4)} \geq \frac{\log(n)}{2}$ , for large n. Then

$$1 - \epsilon < 1 \Rightarrow (1 - \epsilon)^l \le (1 - \epsilon)^{\frac{\log(n)}{2}} = n^{\frac{\log(1 - \epsilon)}{2}}, \text{ with } \frac{\log(1 - \epsilon)}{2} = -\epsilon$$

For small  $\epsilon$ ,  $\log(1-\epsilon) \sim -\epsilon \Rightarrow c \sim 2^{-101}$ .

**Remark 27.8.**  $\mathbb{P}_{\frac{1}{2}}(H(S)) \leq n \cdot \mathbb{P}_{\frac{1}{2}}(r(C_0) \geq n)$  by union bound, since H(S) implies that some point on the left of S is connected to some point on the right, creating a path of length at least n. Hence  $\mathbb{P}_{\frac{1}{2}}(r(C_0) \geq n) \geq \frac{1}{2n}$ .

Then 
$$\frac{1}{2n} \le \mathbb{P}_{\frac{1}{2}}(r(C_0) \ge n) \le n^{-c} \Longrightarrow 2^{-101} < \frac{\log[\mathbb{P}_{\frac{1}{2}}(r(C_0) \ge n)]}{\log(n)} \le 1.$$

Which leads to the following question:

Question 27.9. Does 
$$\lim_{n \to \infty} \frac{\log[\mathbb{P}_{\frac{1}{2}}(r(C_0) \ge n)]}{\log(n)}$$
 exist?

72
## 28. (Wednesday, November 14)

Goal Prove Harris-Kesten Theorem

$$p_H(\mathbb{Z}^2) = \frac{1}{2}$$

(Harris showed that  $p_H(\mathbb{Z}^2) \geq \frac{1}{2}$ , and Kesten proved that  $p_H(\mathbb{Z}^2) \leq \frac{1}{2}$ , see Chapter 2 of BR book.)

## **Probability Preliminaries**

- Margulis(1974)-Russo(1981) Formula
- Friegat Kata Theorem 1998)
- 1. MR formula

Let  $Q^n = n$  dimensional cube=boolean lattice  $(w_1, \dots, w_n)$  where each  $w_i = 0$  or 1, and let  $Q_p^n =$  probability measure on  $Q^n$ , where  $p = (p_1, \dots, p_n)$  and each coordinate is 1 with probability  $p_i$  and 0 with probability  $1 - p_i$  (i.i.d). Let A = increasing event (i.e.  $x \in A, y \ge x$ , then  $y \in A) \subseteq Q^n$  (up-set) Let  $w \in Q^n$ , then  $w_i$ , ith coordinate variable of w is pivotal for A iff precisely one of  $w = (w_1, \dots, w_n)$  and  $r_i(w) := (w_1, \dots, w_{i-1}, 1 - w_i, w_{i+1}, \dots, w_n)$  is in A.

## Influence of *i*th variable on A

Let  $\beta_i(A) := \beta_{p,i}(A) = \mathbb{P}_p(w_i \text{ is pivotal for } A)$ , where  $\mathbb{P}_p(A)$  is a function of p. (In fact, it is a polynomial, hence smooth.)

Lemma 28.1. (MR)

$$\frac{\partial}{\partial p_i} \mathbb{P}_p(A) = \beta_i(A)$$

Lemma 28.2. (FK 1996, Proc. AMS) "Every monotone graph property has a sharp threshold"

But, containing  $K_4$  subgroup property is 'not' sharp.

Let A be an increasing subset of  $Q_p^n$  with  $\mathbb{P}_p(A) = t$ . If  $\beta_i(A) \leq \delta$  for every *i*, then

$$\sum_{i=1}^{n} \beta_i(A) \ge Ct(1-t)\log\frac{1}{\delta},$$

where c > 0 is an absolute constant.

Let *E* be an event (eg.  $H(r) = \exists$  horizontal open coring of  $m \times n$  rectangle *R*.) *e* is a pivotal for *E* in configuration *w* iff  $w^+$  is in *E* and  $w^-$  is not in *E*, where  $w^+$  agrees with everywhere except possible at  $e_j$  included and  $w^-$  agrees with everywhere except possible at  $e_j$  excluded.

Let  $I_p(e, E) = \mathbb{P}_p(e \text{ is pivotal for } E)$ . If E is increasing,  $I_p(e, E) = \mathbb{P}_p(w^+ \in E \text{ and } w^- \in E)$ .

**Lemma 28.3.** Let  $R = m \times n$  rectangle in  $\mathbb{Z}^2$  and e be a bond in R. Then,

$$I_p(e, H(R)) \le 2\mathbb{P}_{\frac{1}{2}}(r(C_0) \ge \min(m/2 - 1, \frac{n-1}{2}))$$

for all 0 .

(Proof) Suppose e is a pivotal for H(R), and  $w^+ \in H(R)$ ,  $w^+ \notin H(R)$ , and every horizontal crossing in H(R) must use e. So one end point of e joined to left of R, and another end point joined to right side of R. So at least one end point is start at open path of at least m/2 - 1. Hence,

(3) 
$$I_p(e, H(R)) \le 2\mathbb{P}_p(r(C_0) \ge m/2 - 1)$$

Similarly, we also have  $w^- \in V(\mathbb{R}^h)$ . So in  $w^-$ , there is open dual path crossing  $\mathbb{R}^h$  vertically using  $e^*$ .

(4) 
$$I_p(e, H(R)) \le 2\mathbb{P}_{1-p}(r(C_0) \ge \frac{n-1}{2}$$

For each  $a, r(C_0) \ge a$  is increasing event. so  $\mathbb{P}_p(r(C_0) \ge a)$  is increasing function of p. So, claim follows from (1) for  $p \le \frac{1}{2}$  and (2) for  $p \ge \frac{1}{2}$ .

**Lemma 28.4.** Let  $p > \frac{1}{2}$ , and let  $k \ge 1$  be fixed. Then,  $\exists \gamma = r(p) > 0$  and  $n_0 = n_0(p,k)$  such that  $h_p(kn,n) \ge 1 - n^{-r}$  for all  $n \ge n_0$ .

### 29. (FRIDAY, NOVEMBER 16)

29.1. Kesten's theorem. If  $p > \frac{1}{2}$ , then  $\mathbb{P}_{\vec{p}}(E_{\infty}) = 1$ .

**Lemma 29.1.** Let  $p > \frac{1}{2}$  and  $k \ge 1$  be fixed. Then  $\exists \gamma = \gamma(p) > 0$ , and  $n_0 = n_0(p,k)$  such that  $h_p(kn,n) \ge 1 - n^{\gamma}$ .

(Recall that  $h_p(kn, n) \ge h_k > 0$  for every  $n \ge 1$ .)

*Proof.* By previous results, we have  $\mathbb{P}_{1/2}(r(C_0) \ge n) \le n^{-c}$ . By lemma,

$$I_{p'}(e, H(R)) \le n^{-a} = \gamma$$

for some absolute constant a, for every bond e of R and  $p' \in [1/2, p]$ .

Write  $f(p') := \mathbb{P}_{p'}(H(R))$ . By Friedgut-Kalai, we have

$$\sum_{e \in R} I_{p'}(e, H(R)) \ge cf(p')(1 - f(p')) \log(1/\delta),$$

for some absolute constant c > 0, for all  $p' \in [1/2, p]$ .

By MR, this sum is the derivative of f(p') with respect to p'. Write  $g(p') := \log\left(\frac{f(p')}{1-f(p')}\right)$ . Then

$$\frac{d}{dp'}g(p') = \frac{1}{f(p')(1 - f(p'))} \frac{d}{dp'}(f(p')) \ge c \log(1/\delta) = ac \log n.$$

Taking *n* large enough, we have  $g(p) \ge ac(p - \frac{1}{2})\log n + g(\frac{1}{2})$ , where the second term is bounded below by some constant. Hence  $g(p) \ge \frac{ac(p-\frac{1}{2})\log n}{2}$ .

Now

$$g(p) = \log\left(\frac{f(p)}{1 - f(p)}\right) \ge \frac{ac(p - \frac{1}{2})\log n}{2}$$

 $\mathbf{SO}$ 

$$\frac{f(p)}{1-f(p)} \ge e^{\frac{ac(p-\frac{1}{2})\log n}{2}} \Rightarrow f(p) \ge \frac{n^{ac(p-\frac{1}{2})/2}}{1+n^{ac(p-\frac{1}{2})/2}} \ge 1-n^{ac(p-\frac{1}{2})/2}.$$

**Theorem 29.2** (Kesten). If p > 1/2, then  $\mathbb{P}_p(E_{\infty}) = 1$ . ( $E_{\infty}$ : event that there exists an infinite open cluster)

*Proof.* Fix p > 1/2. Let  $\gamma = \gamma(p) = ac(p - \frac{1}{2})\frac{1}{2}$  and  $n_0 = n_0(p, k)$  be as before. Assume  $n \ge n_0$ .

Let k = 0, 1, 2, ... and  $R_k$  be the rectangle where bottom-left corner at 0, and  $2^k n \times 2^{k+1}n$  is k is even, and  $2^{k+1}n \times 2^k n$  if k is odd. Let  $E_k$  be the event that  $R_k$  is crossed the long way by an open path. Any two  $E_k$ ,  $E_{k+1}$  meet. So if all  $E_k$ 's hold, so does  $E_{\infty}$ .

Apply union bound:

$$\sum_{k \ge 0} \mathbb{P}_p(E_k \text{ fails}) \le \sum_{k \ge 0} (2^k n)^{-\gamma} = \frac{n^{-\gamma}}{1 - 2^{-\gamma}} < 1,$$

for large enough n. So  $\mathbb{P}_p(E_{\infty}) > 0$ . By Kolmogorov 0-1 law,  $\mathbb{P}_p(E_{\infty}) = 1$ .

**Exercise 29.3.** Show that  $0 < \mathbb{P}_H(\mathbb{Z}^d) < 1$  for every  $d \ge 3$ .

## 30. (NOVEMBER 19)

## Question 30.1. What is the notion of higher dimensional percolation

This lecture is based on the paper "On a sharp threshold transition from area law to perimeter law" by Aizenman, Chayes, Chayes, Fröhloh, Russo 1983.

Take the complete lattice  $\mathbb{Z}^3$  sites and bonds alike. Square plaquettes appear *i.i.d.* with probability *p*. Consider  $(M \times N)$  rectangular loops  $\gamma$  in lattice plane.

Let  $W_{\gamma}$  be the event that  $\gamma$  is a boundary of some subset of plaquettes. If X is a collection of 2-dimensional plaquettes, then  $\delta X$  is collection of all bonds in an odd number of squares in X.

So  $W_{\gamma}$  is an monotone increasing event as p increases. We get the following bounds on the probability of  $W_{\gamma}$ .

$$p^{M \cdot N} \le \Pr_p[W_{\gamma}] \le \left(1 - (1 - p)^4\right)^{2(M-1) + 2(N-1)} \le (4p)^{2(M-1) + 2(N-1)}$$

We get the lower bound since if all plaquettes are present within the loop, then  $W_{\gamma}$  holds. We get the upper bound from needing at least one of the four possible plaquettes needs to be turned on for each bond in  $\gamma$ . We disregard the corners because we want independent events.

Another way to write the previous inequality is

$$p^{\operatorname{Area}(\gamma)} \leq \Pr_p[W_{\gamma}] \leq (4p)^{\operatorname{Perimeter}(\gamma)-4}$$

This implies that there are some absolute constants  $c_1, c_2 > 0$  such that for any  $\gamma$ , we have

$$\exp\left(-c_1\operatorname{Area}\right) \leq \Pr_p[W_{\gamma}] \leq \exp\left(-c_2\operatorname{Perimeter}\right)$$

This plaquette model is dual to bond percolation on the  $\mathbb{Z}^3$  lattice. If  $p_c$  is the critical probability for bond percolation in  $\mathbb{Z}^3$ , then we would expect some sort of transition to occur at  $1 - p_c$  in the plaquette model.

## Theorem 30.2.

$$Pr_p[W_{\gamma}] \sim \begin{cases} \exp\left(-\alpha(p)A(\gamma)\right) & : p > 1 - p_c \\ \exp\left(-c(p)P(\gamma)\right) & : p < 1 - p_c \end{cases}$$

where  $\alpha, c$  are positive constants depending only on p

The previous ~ means that  $\frac{\log(\Pr[W_{\gamma}])}{A(\gamma)} \to -\alpha$  as  $A(\gamma) \to \infty$ .

Question 30.3.

$$\lim_{p \to 1-p_c^+} \alpha(p) = 0?$$

This is known as the surface tension at critically.

**Question 30.4.** What is the right notion of plaquette percolation? What corresponds to an infinite object when  $p > 1 - p_c$  but not when  $p < 1 - p_c$ ?

**Question 30.5.** Also what happends when  $p = \frac{1}{2}$  in  $\mathbb{Z}^4$  for 2-dimensional paquettes?

This system is self-dual, so we would expect transition occurs at  $p = \frac{1}{2}$ 

Let D(0, y) be the chemical distance in the random graph in lattice and let |y| denote the graph distance from 0 to y.

Antal, Piszhora (1996) in On chemical distances in supercritical Bernoulli percolation show that if 0 and y are in the same component, then

$$\frac{D(0,y)}{|y|} \le \rho(p,d)$$

with probability tending to 1 as  $|y| \to \infty$ 

## 31. Monday, November 26

**Goal:** Proving the uniqueness of infinite open cluster. In Aizenmann-Kesten-Newman Theorem this holds under mild conditions.

**Lemma 31.1.** Let  $\Lambda$  be a connected, locally finite, finite type ( $V(\Lambda)$  has finitely many orbits under  $aut(\Lambda)$ ), infinite graph. Let  $E \subset \{0,1\}^{V(\Lambda)}$  be an automorphism invariant event. Then  $Pr^s_{\Lambda,p}[E] = Pr[E] \in \{0,1\}$ .

*Proof.* Fix  $\epsilon > 0$ . Since E is measurable, there exists an event  $E_F$  which only depends on the states of finitely many sites F, and such that  $\Pr(E\Delta E_F) \leq \epsilon$ . (Exercise!). Let  $x_0 \in V(\Lambda)$ , and let

$$M = \max\{d(x_0, y) : y \in F\}.$$

Since  $\Lambda$  is locally finite

$$B_{2M}(x_0) = \{ z : d(x_0, x) \le 2M \}$$

is finite. Let x be a site equivalent to  $x_0$  (x and  $x_0$  are the same type) such that  $d(x, x_0) > 2M$ . Let  $\varphi(x_0) = x$ . For  $y \in F$ , we have

$$d(x_0, \varphi(y)) \ge d(x_0, \varphi(x_0)) - d(\varphi(x_0), \varphi(y))$$
$$= d(x_0, x) - d(x_0, y)$$
$$> 2M - M = M,$$

so  $\varphi(y) \notin F$ . Since  $F \cap \varphi(F) = \emptyset$ , the events  $E_F$  and  $E_{\varphi(E_F)}$  are independent. Then,

$$\Pr[E_F \cap \varphi(E_F)] = \Pr[E_F] \Pr[\varphi(e_f)] = \Pr[E_F]^2.$$

Since  $\Pr[A] - \Pr[B] \leq \Pr[A\Delta B]$  we have

$$\begin{aligned} \left| \Pr[E] - \Pr[E_F]^2 \right| &= \left| \Pr[E \cap E] - \Pr[E_F \cap \varphi(E_F)] \right| \\ &\leq \Pr[(E \cap E) \Delta(E_F \cap \varphi(E_F))] \end{aligned}$$

For any sets A, B, C, D

$$(A \cap B)\Delta(C \cap D) \subset (A\Delta C) \cup (B\Delta D).$$

So,

$$|\Pr[E] - \Pr[E_F]^2| \le \Pr[E\Delta E_F] + \Pr[E\Delta\varphi(E_F)]$$
  
=  $\Pr[E\Delta E_F] + \Pr[\varphi(E)\Delta\varphi(E_F)]$  (E is automorphism invariant)  
=  $2\Pr[E\Delta E_F] \le 2\epsilon$ 

Since 
$$|\Pr[E_F] - \Pr[E]| \le \Pr[E\Delta E_F] \le \epsilon$$
 we have  
 $\Pr[E] - \Pr[E]^2 \le |\Pr[E] - \Pr[E]^2| \le |\Pr[E] - \Pr[E_F]^2| + |\Pr[E_F]^2 - \Pr[E]^2| \le 2\epsilon + 2\epsilon = 4\epsilon$ 

Since  $\epsilon$  is arbitrary, we have  $\Pr[E] - \Pr[E]^2 = 0$ , so  $\Pr[E] \in \{0, 1\}$ .

**Lemma 31.2.** Let  $\Lambda$  be a connected, infinite, locally finite, finite type graph. Let  $p \in (0, 1)$ . Then either

- Pr[I<sub>0</sub>] = 1, or
  Pr[I<sub>1</sub>] = 1, or
- $Pr[I_{\infty}] = 1$ ,

where  $I_k$  is the event that there are exactly k infinite ope clusters.

*Proof.* Fix  $x_0 \in V(\Lambda)$ . Let  $2 \leq k < \infty$ . Assume  $\Pr[I_k] > 0$ . Let

$$T_{n,k} := I_k \cap \{ \text{ every infinite cluster intersects the ball } B_n(x_0) \}.$$

Note that  $I_k = \bigcup_{n \ge 1} T_{n,k}$  since  $\bigcup_{n \ge 1} B_n(x_0) = V(\Lambda)$ . If we assume  $\Pr[I_k] > 0$ , then  $\Pr[T_{n,k}] > 0$  for some *n*. Also  $T_{n,k}$  is the union of disjoint events

$$T_{n,k,\vec{s}} = T_{n,k} \cap \{S = \vec{s}\}$$

where S denotes the state of sites in  $B_n(x_0)$ . So, there is an  $\vec{s}$  such that  $\Pr[T_{n,k,\vec{s}}]$  is positive. If  $w \in T_{n,k,\vec{s}}$ , let w' be the configuration obtained by changing the closed sites in  $\vec{s}$  to open. Then  $w' \in I_1$ , so

$$\Pr[I_1] \ge \Pr[\{w' : w \in T_{n,k,\vec{s}}\}]$$
$$\ge \left(\frac{p}{1-p}\right)^c \Pr[T_{n,k,\vec{s}}] > 0$$

where c denotes the number of closed sites in  $\vec{s}$ . By previous lemma, if  $\Pr[I_k] > 0$  for some  $2 \le k < \infty$  then

$$\Pr[I_1] = \Pr[I_k] = 1,$$

so  $I_1 \cap I_k = \emptyset$ .

**Lemma 31.3.** let G be a finite graph with k components. Let

$$L \subset V(G), \ C = \{c_1, \ldots, c_s\} \subset V(G), \ and \ L \cap C = \emptyset$$

such that at least one  $c_i$  is in each component of G and deleting  $C_i$  disconnects its component into smaller components, at least  $m_i \geq 2$  of which contain vertices of L. Then,

$$|L| \ge 2k + \sum_{i=1}^{s} (m_i - 2).$$

*Proof.* Suffices to assume G is connected, i.e., k = 1. Removing an edge only increases the number of components of  $G - c_i$ , so assume G is minimal wrt G being connected and containing  $C \cup L$ . G is a tree, all leaves are in L. But for a tree

number of leaves 
$$= 2 + \sum_{v} d(v) - 2$$

where the sum is taken over all internal vertices.

## 32. (Wednesday, Nov. 28)

32.1. **Amenability.** Today, we will discuss the Aizenman-Kesten-Newman Theorem. We begin by making a variation on a very standard definition.

**Definition 32.1.** (1) As usual, if X is some metric space, and  $x \in X$ , let  $B_n(x)$  denote the open balls of radius n centered at x.

- (2) Similarly, let  $S_n(x)$  denote the sphere of radius n; i.e., those points exactly n units from x.
- (3) Say that an infinite, locally finite graph is amenable if for all  $x \in X$ ,

$$\frac{|S_n(x)|}{|B_n(x)|} \to 0 \quad \text{ as } n \to \infty$$

Compare this definition to the definition of an amenable group. Recall (see, e.g., [5], Section 11.1) that every locally compact (Hausdorff) group G has a left-invariant measure, called a Haar measure, that is unique up to constant multiplication. We will use  $\lambda$  to refer to one such measure. Recall a piece of notation: for S a subset of a group G and  $x \in G$ ,  $xS := \{x \cdot s : s \in S\}$ .

**Definition 32.2.** A locally compact group G is amenable provided that for every compact  $K \subset G$  and for all  $\epsilon > 0$  one can find a set  $U \subset G$  with  $0 < \lambda(U) < \infty$  with the property that for all  $x \in K$ ,<sup>18</sup>

$$\frac{\lambda(U\Delta xU)}{\lambda(U)} < \epsilon.$$

In other words, no matter how large a compact set of shifts is, there is a finitemeasure set that is moved very little by it. For example, we can take  $G = (\mathbb{Z}, +)$ with the usual discrete topology and counting measure, and for any  $K \subset [-n, n]$ and for any  $\epsilon > 0$ , let  $N = \left\lceil \frac{n}{\epsilon} \right\rceil + 1$  and U = [-N, N].<sup>19</sup> Then for all  $x \in K$ , |x| < n, so

$$\frac{\lambda(U\Delta(x+U))}{\lambda(U)} \le \frac{2n}{2N+1} < \frac{n}{N} \lneq \epsilon.$$

In short, the set does not get moved because it does not have a large boundary. In fact, this correspondence with the "isoperimetric inequality" formulation of amenability of graphs is not accidental.

**Theorem 32.3.** A locally compact (Hausdorff) group G is amenable if and only if every Cayley graph is amenable as a graph.

Sadly, the only place online where I can find a reference is the statement on a conference website, [10] (with a slightly more restrictive definition of amenable graph).

 $<sup>^{18}\</sup>text{Recall}$  that  $\Delta$  denotes the symmetric difference of sets.

<sup>&</sup>lt;sup>19</sup>Recall that  $\lceil x \rceil$  is the ceiling function, the least integer greater than or equal of x.

32.2. **Theorem.** Recall that  $I_k$  denotes the event that there are exactly k infinite clusters in a random graph, where  $0 \le k \le \infty$ . For convenience, we declare  $I_{\ge k}$  to denote the event that there are at least k infinite clusters. Also recall that for  $\Lambda$  a (countably?) infinite graph,  $\Pr_{\Lambda,p}^s$  denotes the site-percolation on  $\Lambda$  where vertices are included with probability p.

We now come to the main theorem. Rather than the original proof, we follow the proof of Burton and Kane (1989), as discussed in [3].

**Theorem 32.4** (AKN, 1987). Let  $\Lambda$  be a connected, locally finite, finite type, infinite, amenable graph. Let 0 . Then

(5) either 
$$Pr^s_{\Lambda,p}(I_0) = 1$$
 or  $Pr^s_{\Lambda,p}(I_1) = 1$ 

Recall from last time that using all the assumptions except amenability, we get the weaker statement

(6) 
$$\operatorname{Pr}_{\Lambda,p}^{s}(I_{0}) = 1, \operatorname{Pr}_{\Lambda,p}^{s}(I_{1}) = 1, \text{ or } \operatorname{Pr}_{\Lambda,p}^{s}(I_{\infty}) = 1$$

Therefore, we must rule out the possibility of infinitely many infinite-size components.<sup>20</sup> To do so, we must use a technical lemma on cut-sets of vertices and the sets they separate.

**Lemma 32.5.** Let G be a finite graph with k components. Suppose that there exists  $L \subseteq V(G)$ , and  $C = \{c_1, c_2, \ldots, c_s\} \subseteq V(G)$ , with the following properties:

(1)  $L \cap C = \emptyset$ 

- (2) at least one  $c_i$  is in each component of G,
- (3) for each  $i, 1 \le i \le k, c_i$  is a cut-vertex, and the graph induced from G with vertex-set  $V(G) \setminus \{c_i\}$  disconnects its component into smaller components,  $m_i \ge 2$  of which contain vertices of L.

Then 
$$|L| \ge 2k + \sum_{i=1}^{3} (m_i - 2).$$

32.3. **Proof of the AKN Theorem.** We will demonstrate that  $\operatorname{Pr}_{\Lambda,p}^{s}(I_{\infty}) = 0$  in the amenable case by showing that  $\operatorname{Pr}_{\Lambda,p}^{s}(\bigcup_{k\geq 3}I_{k}) = 0$ . By the above,  $\operatorname{Pr}_{\Lambda,p}^{s}(I_{2}) = 0$  already, so this suffices. Our proof is by contradiction; suppose that  $\operatorname{Pr}_{\Lambda,p}^{s}(\bigcup_{k\geq 3}I_{k}) \geqq 0$ .

Fix  $x_0 \in V(\Lambda)$ . Let  $X_0 \subset V(\Lambda)$  be the set of vertices equivalent to  $x_0$  under automorphisms.<sup>21</sup> Recall by the locally-finite assumption that there are only finitely many such equivalence classes.

Since  $\Lambda$  is connected, we have that  $V(\Lambda) = \bigcup_{r \geq 1} B_r(x_0)$ ; that is, the vertex-set is covered by various metric balls centered at  $x_0$ . Since  $\Pr_{\Lambda,p}^s(\bigcup_{k \geq 3} I_k) \geq 0$  by

<sup>&</sup>lt;sup>20</sup>This can happen, for example, when  $\Lambda = F_2$ , the free group/graph on two generators.

<sup>&</sup>lt;sup>21</sup>i.e.,  $y \in X_0$  if and only if there exists  $\phi \in Aut(\Lambda)$  with  $\phi(x_0) = y$ .



FIGURE 3. Illustration of  $T_r(x)$ 

assumption, there is a large enough r such that with positive probability,  $B_r(x_0)$  contains sites from at least 3 infinite open clusters (since if this did not hold, with high probability no ball would contain vertices of three open clusters, and by the covering property, the graph would not have as many as three infinite open clusters). Fix one such r, and let  $T_r(x)$  be the event that both of the following occur:

- (1) Every site in  $B_r(x)$  is open.
- (2) There exists an infinite cluster, O, such that when all the sites in  $B_r(x)$  are clused, O is disconnected into at least three infinite open clusters.

An illustration is given in Figure 3.

**Proposition 32.6.**  $Pr(T_r(x_0)) =: a > 0$ . In fact, for all sites  $x \in X_0$ , we have  $Pr(T_r(x)) = a$  for some a > 0.

*Proof.* First, note that since every  $x \in X_0$  is moved by graph automorphisms to  $x_0$ , we have  $\Pr(T_r(x)) = \Pr(T_r(x_0))$  for all  $x \in X_0$ . Thus, we immediately reduce to the situation at  $x_0$ .

 $B_r(x_0)$  contains sites from at least 3 infinite open clusters with nonzero probability by assumption. Further, every configuration in  $\bigcup_{k\geq 3}I_k$  creates a new configuration in  $T_r(x_0)$  by changing all the sites in  $B_r(x_0)$  to open.<sup>22</sup> Unfortunately, the map is not injective, and also not necessarily measure-preserving. We must, therefore, determine how the map works with respect to the measure.

To be precise, let f be the map of configurations of  $\Lambda$ , that switches every site in  $B_r(x_0)$  to open. This map is clearly  $2^{|B_r(x_0)|} : 1$ . By the above, f maps  $\bigcup_{k\geq 3}I_k$  inside  $T_r(x_0)$ , so we wish to show that f maps positive-measure sets to positive-measure sets. Further, for all of the configurations  $2^{|B_r(x_0)|}$  of  $B_r(x_0)$ , switching all closed vertices to open changes the probability by a multiplicative factor of  $\left(\frac{p}{1-p}\right)^m$ , where  $0 \leq m \leq |B_r(x_0)|$  denotes the number of closed vertices. Therefore, if  $p \geq \frac{1}{2}$ , then for all subsets E of the configurations of  $\Lambda$ ,  $\Pr(f(E)) \geq \Pr(E)$ . If  $p < \frac{1}{2}$ , then  $\Pr(f(E)) \geq \left(\frac{p}{1-p}\right)^{|B_r(x_0)|}$   $\Pr(E)$ , since at worst  $|B_r(x_0)|$  many switches from closed to open occur for any element of E.  $\Box$ 

To continue the main proof, for any  $n \ge r$ , let W = W(n) be a subset of  $X_0 \cap B_{n-r}(x_0)$  maximal with respect to the property that the balls  $\{B_{2r}(w) : w \in W\}$  are disjoint. Therefore, by maximality, if  $w' \in X_0 \cap B_{n-r}(x_0)$ , and  $w' \notin W$ , there exists  $w_1 \in W$  with d(w, w') < 4r.<sup>23</sup>

Now, as  $\Lambda$  is connected and of finite type, there exists a constant  $\ell$  such that every site is within a distance  $\ell$  of a site in  $X_0$ ; else, the sequence of points farther and farther from sites in  $X_0$  would of necessity be in different equivalence classes under automorphism, violating the finite-type assumption. Therefore, if we now set  $n \ge r + \ell$ , every  $y \in B_{n-r-\ell}(x_0)$  is within  $\ell$  units of some  $w' \in X_0 \cap B_{n-r}(x_0)$ .<sup>24</sup> Therefore, since w' is either a member of W or within 4r of a member of W, y is strictly less than  $4r + \ell$  units away from some  $w \in W$ . Therefore, for  $n \ge r + \ell$ , the balls  $\bigcup_{w \in W} \{B_{4r+\ell}(w)\}$  cover  $B_{n-r-\ell}(x_0)$ . By a simple counting argument, then (and noting that all elements in W are elements of  $X_0$ , hence have the same type as  $x_0$ , so that the cardinality of given-radius balls around any  $w \in W \subset X_0$  have

 $<sup>^{22}</sup>$ Since every site is forced open, we create a single open cluster O from the infinite open clusters we had before. If then every site in the ball were converted to closed, the remaining parts of the three (or more) infinite open clusters would remain open and infinite; thus, O would be disconnected into three or more pieces.

<sup>&</sup>lt;sup>23</sup>Otherwise, we could adjoin w' to W, and the open, radius-2r balls around each point would be disjoint, since the minimum distance between w' and any other w is 4r; this is just a standard triangle-inequality argument. Yet this would violate maximality of W.

<sup>&</sup>lt;sup>24</sup>The membership in  $X_0$  follows from the previous sentence; the membership in  $B_{n-r}(x_0)$  follows from the triangle inequality.

the same cardinality as a same-sized ball around  $x_0$ ),

$$|W| \ge \frac{|B_{n-r-\ell}(x_0)|}{|B_{4r+\ell}(x_0)|}.$$

Yet we also have, by  $\Lambda$  locally finite and of finite type, that  $\Delta = \max_{v \in V(\Lambda)} \{ \deg(v) \} < \infty$ . Therefore, we also have by a naive counting argument that

$$|B_{n+1}(x_0)| \le |B_{n-r-\ell} \cdot \left(1 + \Delta + \Delta^2 + \dots + \Delta^{r+\ell+1}\right)$$

Putting the inequalities together, we have that there exists some  $c \ge 0$  such that for all  $n \ge r + \ell$ ,  $|W| \ge c \cdot |B_{n+1}(x_0)|$ . For such n, then,

$$|W| \ge c \cdot \frac{|B_n(x_0)|}{|S_n(x_0)|} |S_{n+1}(x_0)|$$

Finally using the amenability hypothesis, we know that as  $n \to \infty$ ,  $\frac{|S_n(x_0)|}{|B_n(x_0)|} \to 0$ , and hence  $\frac{|B_n(x_0)|}{|S_n(x_0)|} \to \infty$ . Therefore, there exists  $N_0 \ge r + \ell$  such that for all  $n \ge N_0$ ,  $\frac{|B_n(x_0)|}{|S_n(x_0)|} \ge (ac)^{-1}$ , where c is as above, and recalling that  $a = \Pr(T_r(x_0)) > 0$ . Therefore,

$$|W| \ge a^{-1} |S_{n+1}(x_0)|, \text{ if } n \ge N_0$$

To continue, declare, for any  $x \in V(\lambda)$  to be a *cut-3-ball*, or more succintly a *cut-ball*, if  $T_r(x)$  holds.<sup>25</sup> Recall that for all  $w \in W \subset X_0$ ,  $T_r(w) = T_r(x_0) = a > 0$  by assumption and the common type of vertices in  $x_0$ . Therefore, by the linearity of expectation, for  $n \geq N_0$ ,

$$\mathbb{E}(\# \text{ cut-balls in } W) = \sum_{w \in W} \Pr(T_r(w)) = a|W| \ge |S_{n+1}(x_0)|$$

Note that this reduces, for any fixed  $n \ge N_0$ , to the configurations on the finite set W: for some configuration  $\omega$  of W,<sup>26</sup> for any fixed  $n \ge N_0$ , with positive probability, we have that

(7) 
$$s(\omega) := (\# \text{ cut-balls in } W) \ge |S_{n+1}(x_0)|$$

This is one of the inequalties to get a contradiction.

To get an inequality going the other direction, note that the turn towards the emphasis on cut-balls makes it possible to consider reducing the graph to a finite variant to which we can apply the lemma we mentioned at the beginning.

For any configuration  $\chi$  of  $\Lambda$  agreeing with  $\omega$  on W (for  $n = N_0$ , say), let  $\mathcal{O}(\chi)$  denote the union of all open clusters intersecting  $B_n(x_0)$ . Define  $\chi'$  to agree with  $\chi$ , but changing all sites in the cut-balls centered at points in W from open to closed.

<sup>&</sup>lt;sup>25</sup>That is,  $B_r(x)$  is open and part of an infinite open component, but when the vertices in it are set to closed, it splits into at least 3 pieces.

 $<sup>^{26}</sup>$ and since W is a finite set, all configurations of it occur with positive probability

Then the infinite open clusters in  $\mathcal{O}(\chi)$  are disconnected into several open clusters. Let  $L_1, \ldots, L_t$  denote the infinite clusters in  $\chi'$  (at most 3 per cut-ball, at least 3 total with positive probability by assumption), and let  $F_1, \ldots, F_u$  denote the finite clusters in  $\chi'$ . Each  $L^i$  contains a site in  $S_{n+1}(x_0)$  (by connectivity arguments, since all the cut-balls of radius r centered at elements of  $W \subset B_{n-r}(x_0)$  are contained in  $B_n(x)$  by the triangle inequality), and distinct clusters must pass through distinct points of the boundary  $S_{n+1}(x_0)$ , so

(8) 
$$t \le |S_{n+1}(x_0)|$$

Name the cut-balls  $C_1, \ldots, C_s$ -note that they are disjoint, since W was defined so that the 2r-balls around each  $w \in W$  were disjoint. Define a finite graph Hdepending on  $\chi$ 's and  $\chi$ ''s configurations on  $B_n(x_0)$  as follows: contract each  $C_i$  to a single vertex  $c_i$ , each  $F_j$  to a single vertex  $f_j$ , and each  $L_k$  to a single vertex  $\ell_k$ . Connect each  $c_i$  to the  $f_j$ 's and/or  $\ell_k$ 's depending on whether or not  $C_i$ ,  $F_j$ , and  $L_k$  were adjacent in the configuration  $\chi$  (note that there are no edges between the  $f_j$ 's and  $\ell_k$ 's, since they are separated components in  $\chi'$  by hypothesis). Certainly,  $L = \bigcup_{k=1}^t \{l_k\}$  and  $C = \bigcup_{i=1}^s \{c_i\}$  are disjoint by definition. The infinite components of  $\mathcal{O}_{\chi}$  correspond to components of H containing at least one vertex in L. By each  $C_i$  a cut-ball, deleting  $C_i$  from the configuration  $\chi$  of  $\Lambda$  disconnected its component into at least 3 infinite open clusters, so deleting  $c_i$  from H disconnects its component into at least  $m_i \geq 3$  components containing vertices of L. Therefore, the hypotheses of the lemma are satisfied, so we may apply the lemma, and conclude that

(9) 
$$t = |L| \ge 2 + \sum_{i=1}^{s} (m_i - 2) \ge 2 + \sum_{i=1}^{s} 1 = 2 + s$$

Note that although t and s depend on  $\chi$ , by hypothesis, with positive probability, s is at least 1 and t is at least 3, since  $T_r(x_0) = a > 0$  by assumption, so that any vertex in W is a cut-ball with positive probability. Combining Equations 7, 8, and 9, we have

$$|S_{n+1}(x_0)| \ge t \ge s + 2 \ge |S_{n+1}(x_0)| + 2$$
  
 $0 \ge 2$ 

This is a contradiction. Therefore,  $\Pr^{s}_{\Lambda,p}(\bigcup_{k\geq 3}I_k)=0.$ 

Note that the above techniques can be used to show that  $\theta_x(p)$ , the probability that x is in an infinite open cluster under (site-)percolation with probability p, is continuous in p, except possibly at the critical probability  $p = p_c$ .

32.4. For More Information. For more information about the very, very many equivalent formulations of amenability, a taste is given by [13].

The conference hinted at in [10] has a fuller writeup in [11]. In particular, the work of Mark Sapir, and his student Iva Kozáková, is mentioned with regard to percolation theory.

33. (MONDAY, DECEMBER 3)

## Future directions for random graphs

## 33.1. More flexible model of random graphs - especially for application.

- theoretical cognitive neurosciences
  - (see Kologrov, Barzdin, 1967)
- epidemiology
- computer science
  - (model www, social network)
- Topological data analysis

(see Carlson's survey article "Topology and data") in Bulletin of AMS. persistent homology)

## 33.2. Probability Method. Randomness is a nice way to prove existence.

- Higher-dimensional expanders
- Buser-Cheeger inequalities
  - $\lambda_2 =$ smallest positive eigenvalue of  $\mathcal{L}(G)$

h = Cheeger numer, expansion constant

- Lplacian of k-forms,  $\delta d + d\delta$ , for analysis of h (see Dotterrer, Kahle)
- Application of higher dimensional expanders
- Random examples in DK have, for example,

$$f_1 = \binom{n}{2}, \ f_2 = n^2 \log n.$$

. .

- Can we find random expanders with  $f_2 = O(f_2)$ ?
- Topological Turan theory (extream in graph theory)
- Extreme in graph theory  $ex(n; H) := \max \#$  of edges with n vertices and no H-subgroups.
- Turan theorem characterizes for  $H = K_m$ .
- Erdös -Stone for non bipartite group.
- Easier questions ex(n; cycles) = n 1
- 2-dimensional simplicial complxes
- Let S = 2-complex on n-vertices. How may two dimensional faces f(n) can you put in with no embedded copies of  $S^2$ .
- Sos, Erdös, Brown, "On the existence of triangulated spheres in 3 groups..."  $f(n) = n^{5/2}$
- (Linial) What about torus?

You can get  $\sim n^{5/2}$  faces with no torus with modified random constructions. Is that optimal exponent?

• Random groups

LECTURE NOTES: RANDOM GRAPHS AND PERCOLATION THEORY

$$- < \underbrace{g_1, g_2, \cdots, g_n}_{\text{generators}} | \underbrace{r_1, r_2, \cdots, r_m}_{\text{random relations}} >$$

- Triangular model  $r_i$  are all of length 3.
- Total number of possible words of length  $3 \sim n^3$ .
- Let  $m = (n^3)^{\lambda}$ , where  $\lambda \in (0, 1)$  is the density.
- Known :
- If  $\lambda > \frac{1}{2}$ , then w.h.p group is trivial. (exercise)
- (Žuk) If  $\lambda < \frac{1}{2}$ , then w.h.p (word) hyperbolic group  $\exists c > o$  such that  $A(r) \leq CL(\gamma)$  for all contractible  $\gamma$
- If  $\frac{1}{3} < \lambda < \frac{1}{2}$ , the group has Kazhdan's property w.h.p.
- Is every hyperbolic group residually finite?
- Intersections of all finite index subgroup is trivial.
- Conjectured answer is NO. What about random groups?

## References

- Noga Alon and Joel H. Spencer. The probabilistic method. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.
- [2] Anders Björner. Topological methods. In R. Graham, M. Grötschel, and L. Lovász, editors, Handbook of Combinatorics, pages 1819–1872. Elsevier, Amsterdam, 1995.
- [3] Béla Bollabás and Oliver Riordan. Percolation. Wiley-Interscience Series in Discrete Mathematics and Optimization. Cambridge U. Press, New York, 2008.
- [4] David Conlon. A new upper bound for diagonal Ramsey numbers. Ann. of Math. (2), 170(2):941–960, 2009.
- [5] Gerald Folland. Real Analysis: Modern Techniques and their Applications. John Wiley & Sons, New York, second edition, 1999.
- [6] Allen Hatcher. Algebraic Topology. Cambridge U. Press, New York, 2010.
- [7] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
- [8] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. Combinatorica, 26(4):475–487, 2006.
- [9] Ronald Meester and Rahul Roy. Continuum Percolation. Cambridge Tracts in Mathematics (No. 119). Cambridge U. Press, New York, first edition, 2008 (1996).
- [10] Tatiana Nagnibeda and Mark Sapir. Dichotomy amenable/nonamenable in combinatorial group theory: October 8, 2007 - october 12, 2007. http://www.aimath.org/~skrantz/ Blurbs/amenable.html, October 2007. Accessed: 4 Dec. 2012.
- [11] Tatiana Nagnibeda and Mark Sapir. Dichotomy amenable/nonamenable in combinatorial group theory: Workshop summary. http://www.aimath.org/pastworkshops/ nonamenablerep.pdf, October 2007. Accessed: 4 Dec. 2012.
- [12] Mathew Penrose. Random Geometric Graphs. Oxford Studies in Probability (No. 5). Oxford U. Press, New York, 2003.
- [13] Terry Tao. Some notes on amenability. http://terrytao.wordpress.com/2009/04/14/ some-notes-on-amenability/, April 2009. Accessed: 4 Dec. 2012.

89

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY

90