MATH595
Stochastic Processes on Graphs

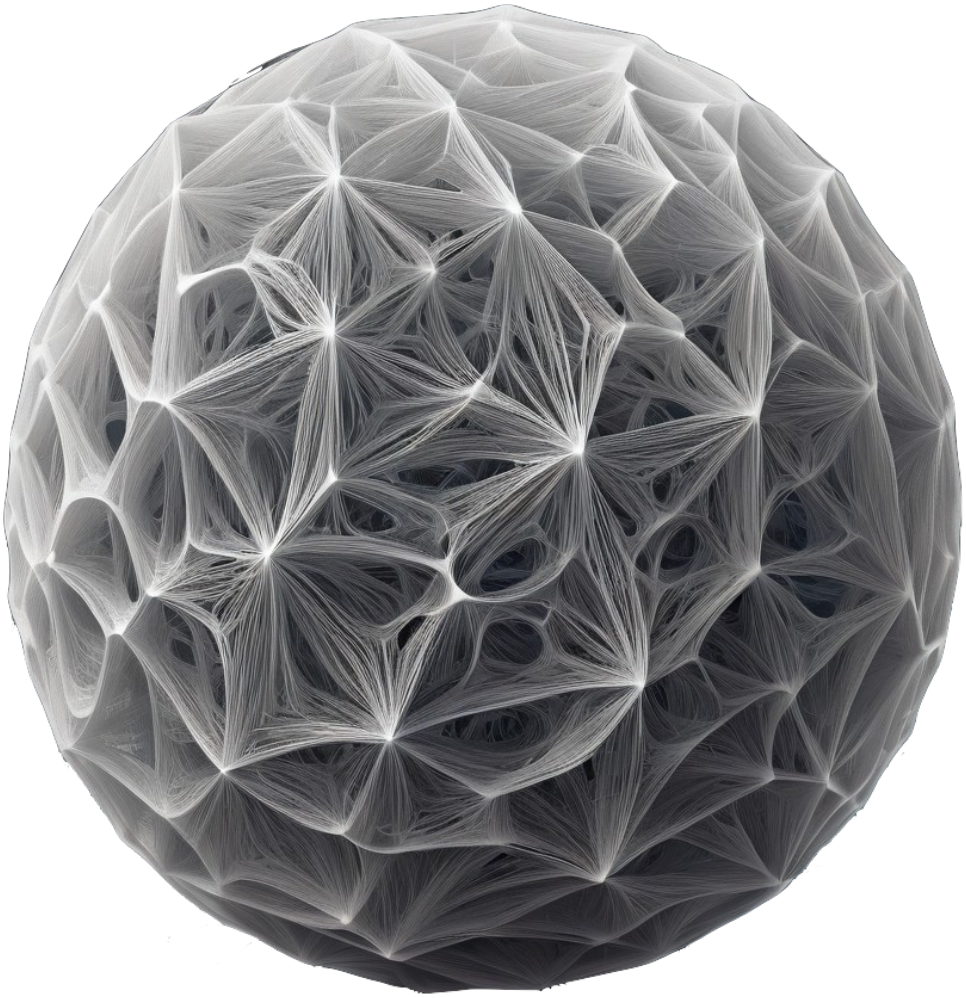
Pingbang Hu

February 10, 2025

Abstract

This is an advanced graduate-level math course taught by Partha Dey at University of Illinois Urbana-Champaign.

We list some references of this course, although we will not follow any particular book page by page: *Random Graph Dynamics* [Dur10], *Random Graphs* [JLR11], *Random Graphs and Complex Networks* [Van24].



This course is taken in Spring 2025, and the date on the cover page is the last updated time.

Contents

1 Introduction **2**
1.1 Structure 2
1.2 Random Graph and Random Graph Process 3

2 Erdős-Rényi Random Graph **4**
2.1 Density and Phase Transition 4
2.2 Degree in Sparse Regime 7
2.3 Size of Connected Component in Sparse Regime 8
2.4 Structure Counting in Various Regime 16

Chapter 1

Introduction

Lecture 1: Overview

In this course, we will consider undirected, unweighted, and finite graph $G = (V, E)$. Given a graph $G = (V, E)$, for any $x, y \in V$, we define $\omega_{xy} \coloneqq \mathbb{1}_{(x,y)\in E}$ as the indicator of (x, y) in E. 21 Jan. 9:30

1.1 Structure[1]

One of the fundamental structures in a graph is the connected component, where we now define.

Definition 1.1.1 (Connected)**.** Given a graph $G = (V, E)$, we say $x, y \in V$ is *connected*, denoted as $x \leftrightarrow y$, if there exists a path $x = v_1, \ldots, v_k = y$ such that $\omega_{v_i v_{i+1}} = 1$ for all $1 \le i \le k - 1$.

It's easy to see that $\leftrightarrow$ is an equivalent relation, hence, one can define the so-called connected component, which is an equivalent class of G with $\leftrightarrow$.

Definition 1.1.2 (Connected component)**.** Given a graph G, a *connected component* $\mathcal{C} \subseteq V$ is a maximal[a] size subset of V such that for all $x, y \in \mathcal{C}$, $x \leftrightarrow y$.

[a] Note the wording: it's not equivalent to maximum.

Notation. For a particular vertex $v \in V$, we define $\mathcal{C}(v, G) \coloneqq \{u \mid u \leftrightarrow v \text{ in } G\}$ as the connected component containing v. If G is realized, we simply write $\mathcal{C}(v)$.

Connected component is an example of *structure*. We list some common structures below:

Definition 1.1.3 (Triangle)**.** A *triangle* (v_1, v_2, v_3) in a graph $G = (V, E)$ is such that $(v_1, v_2), (v_2, v_3)$, and (v_3, v_1) are in E.

Definition 1.1.4 (Cycle)**.** A *n-cycle* $(v_1, \ldots, v_n)$ in a graph $G = (V, E)$ is such that (v_i, v_{i+1}) and (v_k, v_0) are in E.

Definition 1.1.5 (Clique)**.** A *n-clique* $K_n \subseteq V$ in a graph $G = (V, E)$ is such that for every $v_i, v_j \in K_n$, $(v_i, v_j) \in E$.

Example. It's clear that a triangle is just a 3-cycle while also a 3-clique.

A central problem we will be asking is the following:

[1] Later (after this section), we will not reference back to definitions defined here due to their elementary nature.

Problem (Subgraph count)**.** Whether a graph contains a certain structure; if yes, how many?

1.2 Random Graph and Random Graph Process

We are interested in certain graph models where when the number of vertices grows, some structures emerge. The most famous (and simple) random graph model is the Erdős-Rényi random graph model.

Definition 1.2.1 (Erdős-Rényi random graph)**.** The *Erdős-Rényi random graph* model, denoted as $G(n,p)$ or $\mathrm{ER}(n,p)$, is a random graph generated on n vertices such that any two vertices are connected with probability $p \in (0,1)$ independently.

Note. There are lots of independence and symmetry, leading to closed forms for many calculations.

To get a less restrictive model, one can also consider inhomogeneous model, where we let p_{xy} differ for different pairs of $(x,y) \in V \times V$. On the other hand, to relax edge independence, the so-called *exponential random graph model* exists.

Remark. These model all have light-tail. There are also models with heavy tail behavior, e.g., random graph with specified degree distribution, and preferential attachment model.

It's natural to view these random graph model by a random sequence of graphs, which we call graph process. People are interested in several optimization problems of such a graph process.

Example (Optimization on graph process)**.** Given a graph process, what's the (expected) number of the largest cycle, or what's the minimum spanning tree, or some maximum weight problem.

On the other hand, we can also consider another layer of randomness, where we are given a fixed graph, and consider stochastic processes on this graph.

Example. Infection model on a social network, or a growth process.

Some other more advanced topics include Gibbs measures, spin model (Ising model and its generalization Potts model), and spin glass model.

Chapter 2

Erdős-Rényi Random Graph

In this chapter, we first look at the simplest random graph model, the Erdős-Rényi random graph.

> **As previously seen** (Erdős-Rényi random graph). Let $V = [n] \coloneqq \{1, \dots, n\}$ and $p \in [0, 1]$. For every $1 \le i < j \le n$, we let $\omega_{ij} \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(p)$, which induces $E \coloneqq \{(i, j) \mid \omega_{ij} = 1, 1 \le i < j \le n\}$.

Due to the independence and the simplicity, we get several immediate results.

> **Claim.** The number of edges converges in distribution to a standard normal, in particular,
>
> $$\frac{|E| - \binom{n}{2}p}{\sqrt{\binom{n}{2}p(1-p)}} \xrightarrow{D} \mathcal{N}(0, 1),$$
>
> if and only if $\binom{n}{2}p(1-p) \to \infty$. As a corollary, we have $|E|/\binom{n}{2}p \approx 1$.

> **Proof.** We see that $|E| = \sum_{1 \le i < j \le n} \omega_{ij} \sim \mathrm{Bin}(\binom{n}{2}, p)$, hence, $\mathbb{E}[|E|] = \binom{n}{2}p = n(n-1)p/2$. Then, the result follows directly from the central limit theorem. ⊛

Now it's a good time to bring up another random graph model, $\overline{\mathrm{ER}}(n, m)$, where we sample a graph with n vertices and m edges uniformly. This is actually the original Erdős-Rényi random graph model.

> **Remark.** If $m \approx \binom{n}{2}p$, the results often transfer between $\mathrm{ER}(n, p)$ and $\overline{\mathrm{ER}}(n, m)$.

2.1 Density and Phase Transition

2.1.1 Dense and Sparse Graph[1]

We now introduce the concept of *dense* and *sparse* graph, which is decided by the parameter $|E|/\binom{n}{2}$.

> **Definition.** Consider a graph $G = (V, E)$ with $|V| = n$ and $|E| = m$.
>
> > **Definition 2.1.1** (Dense graph). G is *dense* if there exists a constant $\epsilon > 0$ such that $m/\binom{n}{2} > \epsilon$.
>
> > **Definition 2.1.2** (Sparse graph). G is *sparse* if the average degree is constant, i.e., $m = O(n)$.

Let's first observe an interesting property for the Erdős-Rényi random graph model. Note that the typical degree of the Erdős-Rényi random graph is some constant since for $\mathrm{ER}(n, p)$,

$$\frac{1}{|V|}\sum_{v \in V} \deg(v) = \frac{2|E|}{|V|} \approx \frac{2n(n-1)}{2}\frac{p}{n} = (n-1)p.$$

[1] Again, since sparse/dense are so elementary, we will not reference back to definitions defined here.

Note. Regime hence depends on $\lambda := np$ for some λ. When $\lambda \in (0, \infty)$, we are in the sparse regime.

In particular, when $\lambda \in (0, \infty)$, the degree of a particular vertex follows $\text{Bin}(n-1, p) = \text{Bin}(n-1, \lambda/n)$.

Claim. If $\lambda \in (0, \infty)$, $\text{Bin}(n-1, \lambda/n) \xrightarrow{D} \text{Pois}(\lambda)$ as $n \to \infty$.

Proof. We see this in a straightforward way: for any k, $X \sim \text{Bin}(n-1, \lambda/n)$ has a pmf

$$\Pr(X = k) = \binom{n-1}{k} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-1-k} \to \frac{\lambda^k}{k!} e^{-\lambda},$$

which is the pmf of $\text{Pois}(\lambda)$. Hence, by definition, $\text{Bin}(n-1, \lambda/n) \xrightarrow{D} \text{Pois}(\lambda)$. Another proof is based on the total variational distance d_{TV}.

> **As previously seen** (Total variational distance). For the discrete case, given two discrete probability distributions p, r with a finite support Ω,
>
> $$d_{\text{TV}}((p_k)_{k\in\Omega}, (r_k)_{k\in\Omega}) := \frac{1}{2} \sum_{k\in\Omega} |p_k - r_k|.$$

Now, consider the empirical degree distribution defined as $d^{(n)} := \frac{1}{n} \sum_v \delta_{\deg(v)}$. We see that

$$d_{\text{TV}}\left(d^{(n)}, \text{Pois}(\lambda)\right) = \frac{1}{2} \sum_{k=0}^{n} \left| \frac{|\{v \mid \deg(v) = k\}|}{n} - \frac{e^{-\lambda}\lambda^k}{k!} \right|,$$

and by Jensen's inequality,

$$\mathbb{E}\left[d_{\text{TV}}\left(p^{(n)}, \text{Pois}(\lambda)\right)\right] \le \frac{1}{2} \sum_{k=0}^{n} \sqrt{\mathbb{E}\left[\left(\frac{|\{v \mid \deg(v) = k\}|}{n} - \frac{e^{-\lambda}\lambda^k}{k!}\right)^2\right]} \approx \sqrt{\frac{p_k}{n}} = O\left(\frac{1}{\sqrt{n}}\right),$$

where $p_k = e^{-\lambda}\lambda^k/k!$. ⊛

The above gives a distance-one neighborhood characterization of $\text{ER}(n, p)$. However, this actually gives a higher-level picture on larger neighborhoods, in particular, the connected component.

Notation. Given a graph G, let $\mathcal{C}_{\max_i}$ denotes the i^{th} largest connected component in G. For convenient, we use $\mathcal{C}_{\max}$ to denote $\mathcal{C}_{\max_1}$ when it's clear from the context.

2.1.2 Phase Transition of Component Size

Our goal in this section is to prove the following theorem about the components size in $\text{ER}(n, \lambda/n)$:

Theorem 2.1.1. Consider the Erdős-Rényi random graph model $\text{ER}(n, \lambda/n)$ for some $\lambda > 0$.

(a) If $\lambda < 1$, the graph is disconnected with high probability such that $|\mathcal{C}_{\max_1}| = \Theta_p(\log n)$. In particular, if $a(\lambda - 1 - \log \lambda) > 1$, as $n \to \infty$, we have $\Pr\left(|\mathcal{C}_{\max_1}| \ge a \log n\right) \to 0$.

(b) If $\lambda > 1$, $\frac{1}{n}|\mathcal{C}_{\max_1}|$ converges to a constant, i.e., there exists a giant component. Moreover, $\mathcal{C}_{\max_2}$ has size of $O(\log n)$.

(c) At $\lambda = 1$, the random vector $\frac{1}{n^{2/3}}(|\mathcal{C}_{\max_1}|, |\mathcal{C}_{\max_2}|, \dots)$ converges in distribution to a non-trivial limit.

Theorem 2.1.1 says that in the sparse regime, there is a phase transition at $\lambda = 1$. When $\lambda < 1$, there will not exist large component; if $\lambda > 1$, the largest component is of constant fractional of the entire graph, and at $\lambda = 1$, it's something in between.

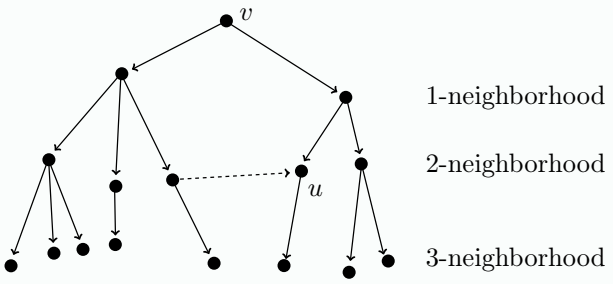
Intuition. Consider the extremely sparse regime where $\lambda < 1$. We give a heuristic argument of why there can't exist a large component. The neighborhood structure of some vertex v, which should be tree-like, at least locally. This is because, for any $k \geq 2$, the expected number of cycles of length k in this structure is

$$\binom{n}{k} \cdot k! \cdot \left(\frac{\lambda}{n}\right)^k \approx n^k \cdot \frac{\lambda^k}{n^k} = \lambda^k,$$

which implies that when $\lambda < 1$,

$$\sum_{k=2}^{n} \mathbb{E}[\#\text{cycle of length } k] \leq \frac{1}{1-\lambda}. \tag{2.1}$$

Hence, in this regime, for a random vertex v, up to any finite distance k, we will only see few cycles.



Formally, by viewing the neighborhood structure as a branching process, one can bound its size.

Lecture 2: Erdős-Rényi Random Graph Model

23 Jan. 9:30

As previously seen. We mainly focus on the following three types of questions for both the degree and components:

1. Typical (local) behavior: single/multiple points view.
2. Global behavior: empirical behavior.
3. Extremal behavior: maxima or minima of various objects.

Toward proving Theorem 2.1.1 (a), we will need the following idea:

Definition 2.1.3 (Stochastic domination)**.** Let X and Y be two real-valued random variables. We say that X is *stochastically dominated* by Y, denoted as $X \preceq Y$, if there exists a coupling of X, Y such that $X \leq Y$.

The reason why stochastic domination is useful is because of the following:

Exercise. $X \preceq Y$ if and only if $\Pr(X > t) \leq \Pr(Y > t)$ for all $t \in \mathbb{R}$.

Here we give some elementary examples of stochastic domination.

Example. $\mathrm{Bin}(n,p) \preceq \mathrm{Bin}(m,p)$ for $m \geq n$.

Proof. Since we have $\mathrm{Bin}(m,p) \overset{D}{=} \mathrm{Bin}(n,p) + \mathrm{Bin}(m-n,p)$. ⊛

Example. $\mathrm{Ber}(p) \preceq \mathrm{Ber}(r)$ if $p \leq r$.

Example. $\mathrm{Ber}(p) \preceq \mathrm{Pois}(\theta)$ by letting $\theta e^{-\theta} = p$. More generally, we just need $1 - p \geq e^{-\theta}$.

Proof. This follows from an application of the above exercise. ⊛

As we will soon see, by using stochastic domination, one can provide a nice bound for proving Theorem 2.1.1 easily.

Intuition. We will often construct objects that are stochastically dominating the object of interest, such that bounds on the dominating quantity imply bounds on the desired quantities in the graph.

2.2 Degree in Sparse Regime

As a warm-up toward proving Theorem 2.1.1 (a), let's first look at an easier problem: the degree.

2.2.1 Single Point Viewpoint

When $p = \lambda/n$, recall what we have proven.

As previously seen. The expected degree of any vertex v is approximately $\lambda \in (0, \infty)$. We also have $\deg_{G_n}(v) \xrightarrow{D} \text{Pois}(\lambda)$ as $n \to \infty$ where $G_n \sim \text{ER}(n, \lambda/n)$.

2.2.2 Joint Distribution

This is for a single point, what about their joint behaviors?

Claim. For any finite k, $(\deg(1), \deg(2), \ldots, \deg(k)) \xrightarrow{D} (\text{Pois}(\lambda), \text{Pois}(\lambda), \ldots, \text{Pois}(\lambda))$.

Proof. Consider any two vertices i, j, we see that

$$\deg(i) = \mathbb{1}_{(i,j)\in E} + \sum_{v\neq j} \mathbb{1}_{(i,v)\in E} \text{ and } \deg(j) = \mathbb{1}_{(i,j)\in E} + \sum_{v\neq i} \mathbb{1}_{(j,v)\in E}.$$

Note that the remaining parts, $\sum_{v\neq j} \mathbb{1}_{(i,v)\in E}$ and $\sum_{v\neq i} \mathbb{1}_{(j,v)\in E}$, are independent. The same argument generalizes to any fixed k vertices.

Moreover, for any fixed k, the number of edges among these k vertices follows $\text{Bin}(\binom{k}{2}, \lambda/n)$, which goes to 0 as $n \to \infty$. Hence, only the remaining parts in the above degree expression survive, which are independent. As k is finite, the remaining parts again follow $\text{Pois}(\lambda)$. ⊛

Intuition. Since the graph is sparse, for any fixed, finite k, $n \to \infty$, independence emerges.

The above is for finite k, serving as the multiple points view.

2.2.3 Extremal Viewpoint

For a global view of the degree distribution, recall the following:

As previously seen. Consider the empirical distribution of degree, defined as $\frac{1}{n}\sum_{v=1}^{n} \delta_{\deg(v)}$, converges to $\text{Pois}(\lambda)$ in the total variation distance.

The last question is the extremal behavior, where we are interested in either bounding or approximating the maximum degree $\deg_{\max,n} := \max_{v\in V} \deg(v)$ for $G \sim \text{ER}(n, p)$.

Proposition 2.2.1. Consider the Erdős-Rényi random graph model $\text{ER}(n, \lambda/n)$ for $\lambda \in (0, \infty)$. Then for all $\epsilon > 0$, as $n \to \infty$, we have

$$\Pr\left(\deg_{\max,n} \geq (1+\epsilon)\frac{\log n}{\log\log n}\right) \to 0.$$

Proof. By a simple union bound, for any $x \in \mathbb{R}$, we have

$$\Pr\left(\max_{v\in[n]} \deg(v) \geq x\right) = \Pr\left(\bigcup_{v=1}^{n} \{\deg(v) \geq x\}\right) \leq n \Pr(\deg(1) \geq x).$$

Now we focus on $\Pr(\deg(1) \geq x)$. With the Chernoff-Cramér method, for any $\theta > 0$, we have

$$\begin{aligned}
\Pr(\deg(1) \geq x) &\leq e^{-\theta x}\mathbb{E}[e^{\theta \deg(1)}] \\
&= e^{-\theta x} \cdot \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^{\theta}\right)^{n-1} \\
&\leq \exp\left(-\theta x + (n-1)\frac{\lambda}{n}(e^{\theta}-1)\right) \leq \exp\big(-\theta x + \lambda(e^{\theta}-1)\big). \qquad (1+t \leq e^{t})
\end{aligned}$$

Optimizing θ, we see that $\theta_0 = \ln(x/\lambda)$ minimizes the above, and it's positive if $x > \lambda$. In the end, we have an upper bound $\exp(-x\ln(x/\lambda) + x - \lambda)$. Plugging it back, we have

$$\Pr\left(\max_{v\in[n]} \deg(v) \geq x\right) \leq n \exp\Big(-x \ln \frac{x}{\lambda} + x - \lambda\Big).$$

By choosing $x = (1+\epsilon)\log n / \log\log n$, the upper bound goes to 0 as $n \to \infty$. ■

Remark. The proof technique of Proposition 2.2.1 will be used extensively in this course.

Let's summarize all results we have for degree so far in the following:

Theorem 2.2.1 (Degree of sparse Erdős-Rényi graph)**.** Let $G \sim \mathrm{ER}(n, \lambda/n)$ for some $\lambda \in (0, \infty)$.

(a) $\deg(1) \xrightarrow{D} \mathrm{Pois}(\lambda)$ as $n \to \infty$.

(b) For any finite k, $(\deg(1), \ldots, \deg(k)) \xrightarrow{D} \mathrm{Pois}(\lambda) \otimes \cdots \otimes \mathrm{Pois}(\lambda)$ as $n \to \infty$.[a]

(c) The empirical degree distribution $\frac{1}{n}\sum_{v=1}^{n} \delta_{\deg(v)} \xrightarrow{D} \mathrm{Pois}(\lambda)$ as $n \to \infty$.

(d) For any $\epsilon > 0$, as $n \to \infty$, we have

$$\Pr\left(\deg_{\max,n} \geq (1+\epsilon)\frac{\log n}{\log\log n}\right) \to 0.$$

[a] I.e., the joint distribution of k many i.i.d. $\mathrm{Pois}(\lambda)$.

2.3 Size of Connected Component in Sparse Regime

Getting back to Theorem 2.1.1, we start by consider the *subcritical regime*, i.e., when $\lambda < 1$.

2.3.1 Subcritical Regime $\lambda < 1$

We start from a similar technique and argument from Theorem 2.2.1, without loss of generality we consider $|\mathcal{C}(1)|$. To see how to compute the size of a connected component, consider the breadth-first search algorithm starting from vertex 1.

Intuition. We see that the induced distance tree $\mathcal{T}$ is in some sense *dominated* by the tree $\mathcal{T}$ where we do not mark the already explored vertices.

The latter is considered as a Galton-Watson branching process with progeny $\mathrm{Bin}(n-1, p)$, denoted as $\mathrm{GWBP}(\mathrm{Bin}(n-1, p))$. The crucial observation is that, the size of this branching process stochastically dominates the size of $\mathcal{C}(1)$. We can now see some intuition of how to prove Theorem 2.1.1 (a), where we aim to show that $\Pr(|\mathcal{C}_{\max_1}| \geq a\log n) \to 0$ as $n \to \infty$ if $a(\lambda - 1 - \log\lambda) > 1$.

Intuition (Proof intuition of Theorem 2.1.1 (a)). For any t, as we discussed above, we will have $\Pr(|\mathcal{C}(1)| \geq t) \leq \Pr(|\mathrm{GWBP}(\mathrm{Bin}(n-1, \lambda/n))| \geq t)$. Next, we observe that we can maintain the number of vertices in the queue when we do the breadth-first search, we see that the tree $\mathcal{T}$ can be (uniquely) embedded in a sequence.

Formally, let (u_i) be the sequence of vertices ordered in terms of the order of exploration. Then, consider the size of the tree $\mathcal{T}$, which is the length of the sequence (s_n) that records the number of vertices in the queue,[a] where $s_0 = 1$, $s_k = s_{k-1} + (x_k - 1)$ such that x_k is the number of children of u_k in $\mathcal{T}$. It is easy to verify that this embedding is indeed a bijection. Finally, we see that the size of the tree is the hitting time to 0, i.e., $|\mathcal{T}| = \inf\{n \geq 1 \mid s_n = 0\}$.

With the above two ingredients, consider the branching process. In this case, the embedded sequence has i.i.d. increments, and is therefore a random walk given by $s_0 = 1$, $s_k = s_{k-1} + X_k - 1$ with $X_k \sim \mathrm{Bin}(n-1, p)$ for all k. The final observation is that when $\lambda < 1$, the above process has a negative drift, hence the hitting time is almost surely finite and can be bounded.

[a] Note that the sequence stops whenever the exploration stops, i.e., an entire connected component is explored.

Lecture 3: Component Size in Subcritical Regime

28 Jan. 9:30

As previously seen. Consider $\mathrm{ER}(n, \lambda/n)$ for some $\lambda > 0$. As in Theorem 2.2.1, we have proved: (a) $\deg(1) \xrightarrow{D} \mathrm{Pois}(\lambda)$ and (b) $(\deg(1), \dots, \deg(k)) \xrightarrow{D} \mathrm{Pois}(\lambda) \otimes \cdots \otimes \mathrm{Pois}(\lambda)$ as $n \to \infty$. (c) Also, the empirical distribution $\frac{1}{n}\sum_{v=1}^{n} \delta_{\deg(v)} \xrightarrow{D} \mathrm{Pois}(\lambda)$. (d) Finally, we have a maximum degree bound such that for any $\epsilon > 0$, $\Pr(\deg_{\max,n}) \geq (1+\epsilon)\log n / \log\log n$.

With the build-up from the previous lecture, we are almost ready to prove Theorem 2.1.1 (a). However, as noted above, it's expected that the result will depend on $\mathrm{Pois}(\lambda)$ in various ways. Hence, we note the following results from standard probability analysis.

Exercise. Let $X_1, \dots, X_r \overset{\text{i.i.d.}}{\sim} \mathrm{Pois}(\lambda)$. Prove the following.

(a) As $n \to \infty$, we have $\max_{i=1,\dots,n} X_i \log\log n / \log n \xrightarrow{p} 1$.

(b) Moreover, we can show that as $n \to \infty$, $\max_{i=1,\dots,n} X_i \in \{m_n, m_n + 1\}$ with probability converging to 1 for some integer m_n satisfying $m_n \cdot \log\log n / \log n \to 1$.

(c) Similarly, we can prove the above two by replacing $\max_{i=1,\dots,n} X_i$ with $\deg_{\max,n}$.

Answer. We quickly sketch the proof for $\deg_{\max,n}$, specifically for the lower bound. Consider a bipartition (V_1, V_2) of V, each with $n/2$ vertices. Then, for any $v \in V_1$, $\deg_{V_2}(v)$ lower bounds $\deg_{\max,n}$. Analyzing $\deg_{V_2}(v)$ turns out to be manageable. ⊛

Now, we're ready to prove Theorem 2.1.1 (a).

Lemma 2.3.1 (Component of subcritical Erdős-Rényi graph). Let $G \sim \mathrm{ER}(n, \lambda/n)$ with $\lambda < 1$.

(a) As $n \to \infty$, $\mathcal{C}(1) \xrightarrow{D} \mathrm{BP}(\mathrm{Pois}(\lambda))$. In particular, $|\mathcal{C}(1)| \xrightarrow{D} |\mathcal{T}_\lambda|$ where $\mathcal{T}_\lambda \sim \mathrm{BP}(\mathrm{Pois}(\lambda))$.

(b) For any finite k, $(\mathcal{C}(1), \dots, \mathcal{C}(k)) \xrightarrow{D} \mathcal{T}_\lambda \otimes \cdots \otimes \mathcal{T}_\lambda$ as $n \to \infty$, where $\mathcal{T}_\lambda \sim \mathrm{BP}(\mathrm{Pois}(\lambda))$.

(c) The empirical distribution of components converges weakly to $\mathrm{BP}(\mathrm{Pois}(\lambda))$ as $n \to \infty$.

(d) $|\mathcal{C}_{\max,n}| \leq (1/I_\lambda + \epsilon) \cdot \log n$ as $n \to \infty$ with high probability where $I_\lambda = \lambda - 1 - \log\lambda > 0$.[a]

[a] Note that I_λ equals to 0 at 1, and diverges to ∞ at both $+\infty$ and $-\infty$.

Proof. Let's prove (a) first. Last time, we have shown that $|\mathcal{C}_n(1)| \preceq |\mathrm{BP}(\mathrm{Bin}(n-1, p))|$ for $p = \lambda/n$. To show that in general, $\mathcal{C}(1) \xrightarrow{D} \mathrm{BP}(\mathrm{Pois}(\lambda))$, we need to compute the pmf of $\mathcal{C}(1)$. Obviously, the support of the distribution of $\mathcal{C}(1)$ is on the set of connected rooted graphs $G' = (V', E')$. Suppose

G' is not a tree such that $|V'| = k$ and $|E'| = d \geq k$. By a simple counting argument, we have

$$\Pr(C_n(1) = G') = \left(\frac{\lambda}{n}\right)^d \left(1 - \frac{\lambda}{n}\right)^{k(n-k)+\binom{k}{2}-d} \cdot \binom{n-1}{k-1} f(G') \to 0,$$

where $f(G)$ is the number of automorphisms of G, which is finite for any fixed G'. Hence, we see that only when $k = d - 1$, this probability is not 0. That is to say, in the limit, the component will be a tree. In fact, with the same calculation, we have the following.

Claim (Borel-Tenner distribution)**.** For $\lambda \leq 1$, and for $k \geq 1$, we have

$$\Pr(|\mathcal{T}_\lambda| = k) = e^{-\lambda k} \frac{(\lambda k)^{k-1}}{k!}.$$

Moreover, $\Pr(\mathcal{C}_n(1) = \mathcal{T}) \to \Pr(\mathrm{BP}(\mathrm{Pois}(\lambda)) = \mathcal{T})$ for all rooted finite tree $\mathcal{T}$.

Proof. For a leveled tree $\mathcal{T}$ of k vertices with $k - 1$ edges,[a]

$$\Pr(\mathcal{C}(1) = \mathcal{T}) = \left(\frac{\lambda}{n}\right)^{k-1} \cdot \left(1 - \frac{\lambda}{n}\right)^{k(n-1)+\binom{k}{2}-(k-1)} \cdot \binom{n-1}{k-1} \to \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda k}.$$

From Cayley's formula, the number of leveled trees on k nodes is k^{k-2}, proving the claim. ⊛

[a] Note that we don't need $f(\mathcal{T})$ since we're considering leveled tree, which is already labeled, making it unique (in terms of automorphisms). In some sense $f(\mathcal{T})$ is handled by the Cayley's formula below.

We omit (b) and (c) since they can be easily shown. To prove (d), we have

$$\Pr(|\mathcal{C}_{\max,n}| \geq t) \leq n \cdot \Pr(|\mathcal{C}_n(1)| \geq t) \leq n \cdot \Pr(|\mathrm{BP}(\mathrm{Bin}(n, \lambda/n))| \geq t).$$

Recall our algorithmic notation:

As previously seen. We denote the set of active vertices as $\mathcal{A}_t$ at time t, and $A_t := |\mathcal{A}_t|$.

Specifically, we have $A_0 = |\mathcal{A}_0| = 1$, $A_1 = |\mathcal{A}_1| = X_1 + 1 - 1 = A_0 + (X_1 - 1)$, $A_2 = A_1 + (X_2 - 1)$, etc., where $X_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}(n, \lambda/n)$. Then, $|\mathrm{BP}(\mathrm{Bin}(n, \lambda/n))|$ is the hitting time at 0, $H^{\{0\}} = \inf\{t \geq 1 \mid A_t = 0\}$. Hence, we have $\Pr(|\mathcal{C}_n(1)| > t) \leq \Pr(A_t \geq 1)$, where $A_t = 1 + (X_1 - 1) + \cdots + (X_t - 1)$. Combining the above, for all $\theta > 0$,

$$\begin{aligned}\Pr(|\mathcal{C}_{\max,n}| \geq t) &\leq n \cdot \Pr\left(\sum_{i=1}^{t} (X_i - 1) \geq 0\right) \\ &\leq n \left(\mathbb{E}[e^{\theta(X_1-1)}]\right)^t = n \left(e^{-\theta}\left(1 - \frac{\lambda}{n} + \frac{\lambda}{n} e^\theta\right)^n\right)^t \leq n \exp\big(t(-\theta + \lambda(e^\theta - 1))\big)\end{aligned}$$

Minimizing over θ, we have $\lambda e^\theta = 1$, hence $\theta = \log 1/\lambda > 0$, which gives

$$n \exp(-t(-\log \lambda - 1 + \lambda)) =: n \exp(-t I_\lambda) = \exp(\log n - t I_\lambda)$$

By taking $t = (1/I_\lambda + \epsilon) \log n$, the probability goes to 0, proving the result. ■

Remark. From (b), for $\lambda < 1$ and any $k \geq 1$, $|C_{\max,k}|/\log n \xrightarrow{D} 1/I_\lambda$ as $n \to \infty$.

Note. We note that for (a), we can also prove it by observing that in the exploration tree, each vertex has $\mathrm{Bin}(n - 1 - c, p)$ children where c is some constant depending on the same level. Hence, as long as we're considering a fixed level neighborhood, everything converges to $\mathrm{Pois}(\lambda)$. In all, for any finite connected rooted tree $\mathcal{T}$, we have $\Pr(\mathcal{C}(1) = \mathcal{T}) \to \Pr(\mathrm{BP}(\mathrm{Pois}(\lambda)) = \mathcal{T})$.

2.3.2 Supercritical Regime $\lambda > 1$

Next, we consider the *supercritical regime* when $\lambda > 1$. Specifically, we want to show Theorem 2.1.1 (b). This is proved in Lemma 2.3.2 below.

Lemma 2.3.2 (Component of supercritical Erdős-Rényi graph)**.** Let $G \sim \mathrm{ER}(n, \lambda/n)$ with $\lambda > 1$.

(a) As $n \to \infty$, $|C_{\max_1,n}|/n \xrightarrow{p} \zeta_\lambda$ where $\zeta_\lambda = \Pr(\mathrm{BP}(\mathrm{Pois}(\lambda))$ survives forever).

(b) As $n \to \infty$, $|C_{\max_2,n}|/\log n \xrightarrow{p} 1/I_\lambda$ where $I_\lambda = \lambda - 1 - \log\lambda$.

(c) Outside $\mathcal{C}_{\max_1,n}$, the graph looks like $\mathrm{ER}(m, \mu/m)$ for some $m \approx n(1-\zeta_\lambda)$ with $\mu < 1$.

To prove Lemma 2.3.2, we divide it into three steps. Fix $k = k_n \approx A\log n$ for some large A. Define $Z_{\ge k_n} := \sum_{v=1}^n \mathbb{1}_{|\mathcal{C}_n(v)|\ge k_n}$ as the number of vertices that has a component size greater than k_n. Then:

(i) $\mathbb{E}[Z_{\ge k_n}] \approx n\cdot\zeta_\lambda + o(n^{1-\epsilon})$ and $\mathrm{Var}[Z_{\ge k_n}] \ll (\mathbb{E}[Z_{\ge k_n}])^2$. By using the second-moment method to control $\Pr(Z_{\ge k_n} = 0)$, e.g., Chebyshev's inequality, we have a concentration bound.

(ii) $\Pr(B\log n \le |\mathcal{C}(1)| \le an) \to 0$ for some $B > 0$ and for any $\zeta_\lambda > a$, i.e., either the component is small or large.

(iii) $Z_{\ge k_n} \approx |\mathcal{C}_{\max,n}|$. Since $Z_{\ge k_n} = \sum_{v=1}^n \mathbb{1}_{|\mathcal{C}(v)|\ge k_n} = \sum_{v:\ |\mathcal{C}(v)|\ge k_n} |\mathcal{C}(v)|$.

Now, to analyze $Z_{\ge k_n}$, we need to consider the exploration algorithm again. However, for convenience, we will now maintain three sets $(\mathcal{A},\mathcal{U},\mathcal{R})$, corresponding to *active*, *unexplored*, and *already explored* set.

Intuition. We see that:

- At time 0, $\mathcal{A}_0 = \{1\}$, $\mathcal{U}_0 = \{2,\dots,n\}$, $\mathcal{R}_0 = \varnothing$ with $A_0 = 1$, $U_0 = n-1$, $R_0 = 0$.
- At time 1, $A_1 = A_0 + \mathrm{Bin}(U_0,p) - 1$, $U_1 = U_0 - \mathrm{Bin}(U_0,p) - 1$, $R_1 = 1$.
- At time 2, $A_2 = A_1 + \mathrm{Bin}(U_1,p) - 1$, $U_2 = U_1 - \mathrm{Bin}(U_1,p)$, and $R_2 = 2$.
- In general, $A_{t+1} - A_t \overset{D}{\sim} \mathrm{Bin}(U_t,p) - 1$, $U_{t+1} - U_t = \mathrm{Bin}(U_t,p)$, and $R_{t+1} = t+1$ for all $t \ge 1$.

Eventually, the above (A,U,R) structure embeds the graph. Then, we can simply look at R_t, and look at the parts whenever it hits 0 to determine the components.

Lecture 4: Component Size in Supercritical Regime

Let's first simplify the notations. At time t, we choose $i_t \in \mathcal{A}_t$, and let $\mathcal{C}_{t+1} = \mathrm{Children}(i_t,\mathcal{U}_t)$. Then, 30 Jan. 9:30

$$\mathcal{A}_{t+1} = (\mathcal{A}_t \setminus \{i_t\}) \cup \mathcal{C}_{t+1}, \quad \mathcal{U}_{t+1} = \mathcal{U}_t \setminus \mathcal{C}_{t+1}, \quad \mathcal{R}_{t+1} = \mathcal{R}_t \cup \{i_t\},$$

where $\mathcal{A}_0 = \{1\}$, $\mathcal{U}_0 = [n] \setminus \mathcal{A}_0$, and $\mathcal{R}_0 = \varnothing$. Let $\xi_{t+1} := |\mathcal{C}_{t+1}| \sim \mathrm{Bin}(U_t,p)$ where $p = \lambda/n$, then

$$A_{t+1} = A_t - 1 + \xi_{t+1}, \quad U_{t+1} = U_t - \xi_{t+1} \sim \mathrm{Bin}(U_t, 1-p), \quad R_{t+1} = R_t + 1 = t+1.$$

Claim. For all $t \ge 0$, $U_t \sim \mathrm{Bin}(U_0,(1-p)^t) = \mathrm{Bin}(n-1,(1-p)^t)$. In particular, $A_t = n - U_t - R_t = n - t - U_t$ with

$$\mathbb{E}[A_t] = n - t - (n-1)\cdot(1-p)^t = (n-1)\left[1-(1-\lambda/n)^t\right] - t + 1 \cong n\left(1 - e^{-\lambda\cdot\frac{t}{n}} - t/n\right)$$

and $\mathrm{Var}[A_t] = (n-1)(1-p)^t\cdot[1-(1-p)^t] \approx \lambda t e^{-\lambda t/n}$.

Let $f_\lambda(x) = 1 - e^{-\lambda x} - x$, hence, $\mathbb{E}[A_t] \cong n f_\lambda(t/n)$. Also, let $\zeta_\lambda > 0$ be the unique positive solution of $f_\lambda(x) = 0$, where one can easily check that when $\lambda > 1$, $f(x) = 0$ for $x > 0$ has a unique solution.

Intuition. Since $\mathrm{Var}[A_t] < (\mathbb{E}[A_t])^2$, so $A_t \neq 0$ with high probability away from $t = 0$ and $t = \zeta_\lambda$.

We prove that this is indeed the case, i.e., there are no intermediate clusters.

Lemma 2.3.3. When $\lambda > 1$ and $0 < t < n\zeta_\lambda$, $\Pr(A_t = 0) \le \exp\big(-t \cdot I_{g(t/n)}\big)$ [a] where $g(x) = \frac{1-e^{-\lambda x}}{x}$.

[a] Recall that we define $I_x = x - 1 - \log x$.

Proof. We see that

$$\begin{aligned}
\Pr(A_t = 0) &= \Pr(n - 1 - U_t = t - 1) \\
&\le \Pr(\mathrm{Bin}(n-1, 1-(1-p)^t) \le t-1) && (n-1-U_t \sim \mathrm{Bin}(n-1, 1-(1-p)^t)) \\
&\le \Pr(\mathrm{Bin}(n-1, 1-e^{-pt}) \le t-1) && (1-p \le e^{-p}) \\
&\le \Pr(\mathrm{Bin}(n, 1-e^{-pt}) \le t) \\
&\le \inf_{\theta>0} e^{\theta t} \cdot \Big[\mathbb{E}[e^{-\theta \cdot \mathrm{Ber}(1-e^{-pt})}]\Big]^n \\
&= \inf_{\theta>0} e^{\theta t} \cdot \big[1 - (1-e^{-\theta})(1-e^{-pt})\big]^n \\
&\le \inf_{\theta>0} \exp\left(n \cdot \frac{\theta t}{n} - n(1-e^{-\theta})(1-e^{-\frac{\lambda t}{n}})\right) = \inf_{\theta>0} \exp\big(n \cdot (\theta a - (1-e^{-\theta})(1-e^{-\lambda a}))\big),
\end{aligned}$$

where we let $a \coloneqq t/n$. Setting $a \coloneqq e^{-\theta} \cdot (1 - e^{-\lambda a})$ will minimize the above, i.e., $e^\theta = \frac{1-e^{-\lambda a}}{a}$. When $a \in [0, c)$ for some c, we can find $\theta > 0$ since $f(x) = 1 - e^{-\lambda x} - x = x(\frac{1-e^{\lambda x}}{x} - 1)$. Specifically, we want $0 < a = t/n < \zeta_\lambda$, where $\zeta_\lambda > 0$ is the solution of $1 - e^{-\lambda \zeta_\lambda} = \zeta_\lambda$. With such a, we have

$$\begin{aligned}
\Pr(A_t = 0) &\le \exp\big(n\theta a - n(1-e^{-\lambda a}) + na\big) \\
&= \exp(-t(g(a) - 1 - \ln g(a))) = \exp(-t(g(t/n) - 1 - \ln g(t/n))) = \exp\big(-tI_{g(t/n)}\big),
\end{aligned}$$

where $\ln g(a) = \theta$ since $e^\theta = g(a) = \frac{1-e^{-\lambda a}}{a}$. ∎

From Lemma 2.3.3, the following is immediate.

Corollary 2.3.1. When $\lambda > 1$ and $t/n \le \alpha < \zeta_\lambda$, $\Pr(A_t = 0) \le e^{-tc(\alpha)}$ where $c(\alpha) \coloneqq I_{g(\alpha)}$.

Proof. Since $g'(x) = \frac{e^{-\lambda x}}{x^2}(1 + \lambda x - e^{\lambda x}) < 0$ when $x > 0$, we have $1 < g(\alpha) \le g(t/n)$, which implies $I_{g(\alpha)} \le I_{g(t/n)}$. The first result then follows from Lemma 2.3.3. ∎

From Corollary 2.3.1, the following is immediate.

Corollary 2.3.2. When $\lambda > 1$ and $\alpha < \zeta_\lambda$, there is some $k > 0$ such that

$$\Pr(\exists t\colon k \log n \le t \le \alpha n \text{ such that } A_t = 0) \le \frac{n^{-kc(\alpha)}}{1 - e^{-c(\alpha)}}.$$

By using a union bound, Corollary 2.3.2 leads to the following key lemma.

Lemma 2.3.4. When $\lambda > 1$ and $\alpha < \zeta_\lambda$, as $k > 1/c(\alpha)$, as $n \to \infty$,

$$\Pr(\exists v \in [n]\colon k \log n \le |\mathcal{C}(v)| \le \alpha n) \le n \Pr(k \log n \le |\mathcal{C}(1)| \le \alpha n) \lesssim n^{1-kc(\alpha)} \to 0.$$

Remark. Lemma 2.3.4 basically proves that when $\lambda > 1$, if a cluster survives after some initial size $(k \log n)$, it'll stay alive until it reaches a size of a constant fraction of n with high probability.

Now, we just need to worry about the size of $\mathcal{C}_{\max}$. To do this, define a random variable that counts

the number of vertices having a small component:

$$Z_{\le k_n} \coloneqq \sum_{v\in[n]} \mathbb{1}_{|\mathcal{C}(v)|\le k_n},$$

where we let $k_n \coloneqq k\log n$.

Lemma 2.3.5. When $\lambda > 1$, $|Z_{\le k_n} - n(1-\zeta_\lambda)| \le n^{1/2+\epsilon}$ with high probability for all $\epsilon > 0$.

Proof. Consider using mean control as our primary tool. We see that

$$\begin{aligned}
\mathbb{E}[Z_{\le k_n}] &= n\Pr(|\mathcal{C}(1)| \le k_n) \\
&= n\Pr(|\mathrm{BP}(\mathrm{Bin}(n-1,p))| \le k_n) + O(k_n\cdot p) \\
&= n\Pr(|\mathrm{BP}(\mathrm{Pois}(\lambda))| \le k_n) + O(k_n\cdot p) \\
&= n(1-\underbrace{\Pr(|\mathcal{T}_\lambda| = \infty)}_{\zeta_\lambda} - \underbrace{\Pr(k_n < |\mathcal{T}_\lambda| < \infty)}_{e^{-k_n\cdot I_\lambda}}) + O\left(\frac{\lambda\log n}{n}\right) = n(1-\zeta_\lambda) + O\left(\frac{\log n}{n}\right).
\end{aligned}$$

For the variance, we have

$$\begin{aligned}
\mathrm{Var}[Z_{\le k_n}] &= \sum_{u,v=1}^{n} \mathrm{Cov}[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}, \mathbb{1}_{|\mathcal{C}(u)|\le k_n}] \\
&= n\mathbb{E}\left[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}\sum_{u=1}^{n}\left(\mathbb{1}_{|\mathcal{C}(u)|\le k_n} - \Pr(|\mathcal{C}(1)|\le k_n)\right)\right] \\
&= n\mathbb{E}\left[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}\sum_{u=1}^{n}\mathbb{1}_{1\leftrightarrow u}\left(\mathbb{1}_{|\mathcal{C}(u)|\le k_n} - \Pr(|\mathcal{C}(1)|\le k_n)\right)\right] \\
&\quad + n\mathbb{E}\left[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}\sum_{u=1}^{n}\mathbb{1}_{1\not\leftrightarrow u}\left(\mathbb{1}_{|\mathcal{C}(u)|\le k_n} - \Pr(|\mathcal{C}(1)|\le k_n)\right)\right].
\end{aligned}$$

Let's first look at the first term, where we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}\sum_{u=1}^{n}\mathbb{1}_{1\leftrightarrow u}\left(\mathbb{1}_{|\mathcal{C}(u)|\le k_n} - \Pr(|\mathcal{C}(1)|\le k_n)\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}\sum_{u=1}^{n}\mathbb{1}_{1\leftrightarrow u}(1-\Pr(|\mathcal{C}(1)|\le k_n))\right] = \mathbb{E}\left[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}|\mathcal{C}(1)|\right](1-\Pr(|\mathcal{C}(1)|\le k_n)) \le k_n.
\end{aligned}$$

As for the second term, we see that

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}\sum_{u=1}^{n}\mathbb{1}_{1\not\leftrightarrow u}\left(\mathbb{1}_{|\mathcal{C}(u)|\le k_n} - \Pr(|\mathcal{C}(1)|\le k_n)\right)\right] \\
&= \sum_{u=2}^{n}\sum_{\ell=1}^{k_n}\Pr(|\mathcal{C}(1)| = \ell)\cdot\Pr(1\not\leftrightarrow u \mid |\mathcal{C}(1)| = \ell) \\
&\qquad\cdot\left(\Pr(|\mathcal{C}(u)|\le k_n \mid 1\not\leftrightarrow u, |\mathcal{C}(1)| = \ell) - \Pr(|\mathcal{C}(1)|\le k_n)\right) \\
&\le \sum_{u=2}^{n}\sum_{\ell=1}^{k_n}\Pr(|\mathcal{C}(1)| = \ell)\cdot 1\cdot \ell k_n\frac{\lambda}{n} \\
&= \frac{(n-1)k_n\lambda}{n}\mathbb{E}[\mathbb{1}_{|\mathcal{C}(1)|\le k_n}|\mathcal{C}(1)|] \le \lambda k_n^2,
\end{aligned}$$

where the first inequality comes from the fact that when $|\mathcal{C}(1)| = \ell$ and $1\not\leftrightarrow u$, $|\mathcal{C}(u)|$ follows the law of $|\mathcal{C}(1)|$ in $\mathrm{ER}(n-\ell,p)$. Then, we couple $\mathrm{ER}(n-\ell,p)$ and $\mathrm{ER}(n,p)$ by adding vertices

$\{n-\ell+1,\dots,n\}$ of $\mathrm{ER}(n-\ell,p)$ and missing edges (sampled i.i.d. from $\mathrm{Ber}(p)$). Hence,

$$\begin{aligned}&\Pr(|\mathcal{C}(u)|\le k_n \mid 1\not\leftrightarrow u, |\mathcal{C}(1)|=\ell)-\Pr(|\mathcal{C}(1)|\le k_n)\\&=\Pr(|\mathcal{C}(1)|\le k_n \text{ in } \mathrm{ER}(n-\ell,p))-\Pr(|\mathcal{C}(1)|\le k_n \text{ in } \mathrm{ER}(n,p))\\&=\Pr(|\mathcal{C}(1)|\le k_n \text{ in } \mathrm{ER}(n-\ell,p) \text{ and } |\mathcal{C}(1)|> k_n \text{ in } \mathrm{ER}(n,p))\\&\le \Pr(\exists u\in\{n-\ell+1,\dots,n\}\colon 1\leftrightarrow u)\le \ell k_n p,\end{aligned}$$

since there are at most ℓk_n many edges between $\mathcal{C}(1)$ in $\mathrm{ER}(n-\ell,p)$ and $\{n-\ell+1,\dots,n\}$ in $\mathrm{ER}(n,p)$. Putting everything together, we have, $\mathrm{Var}[Z_{\le k_n}]\le n(k\log n+\lambda k^2\log^2 n)$, which gives

$$\Pr(|Z_{\le k_n}-\mathbb{E}[Z_{\le k_n}]|\ge n^{1/2+\epsilon})\le \frac{n\log^2 n}{n^{1+2\epsilon}}\to 0$$

as $n\to\infty$ from the Chebyshev's Inequality. ■

Combining Lemma 2.3.4 and Lemma 2.3.5, we have the following.

Corollary 2.3.3. For $\lambda>1$, for all α such that $0<\alpha<\zeta_\lambda$, $|Z_{\ge\alpha n}-n\zeta_\lambda|\le n^{1/2+\epsilon}$. Moreover, $|\mathcal{C}_{\max}|=Z_{\ge\alpha n}$ with high probability for all $\alpha\in(\zeta_\lambda/2,\zeta_\lambda)$.

Putting all results we have, we see that:

(i) No middle ground: no clusters between $[k\log n,\alpha n]$ for $\alpha<\zeta_\lambda$ (Lemma 2.3.4).

(ii) The number of vertices with $|\mathcal{C}(v)|\le k\log n$ is concentrated at $n(1-\zeta_\lambda)$ (Lemma 2.3.5).

(iii) Everything else is in a single component.

Putting everything together, Lemma 2.3.2 (a) and (b) are proved.

Lecture 5: Component Size in Critical Regime

2.3.3 Critical Regime $\lambda=1$

4 Feb. 9:30

What is left is the *critical regime*, where we want to prove Theorem 2.1.1 (c): the random vector $\frac{1}{n^{2/3}}(|\mathcal{C}_{\max_1}|,|\mathcal{C}_{\max_2}|,\dots)$ converges in distribution to a non-trivial limit. To analyze the component size when $\lambda=1$, as what we have done previously, we have $|\mathcal{C}(1)|\preceq\mathrm{Bin}(n-1,1/n)$. Moreover,

- $\mathrm{Ber}(p)\preceq\mathrm{Pois}(\theta)$ for $\theta=-\log(1-p)$;
- $\mathrm{Bin}(n-1,p)\preceq\mathrm{Pois}(-(n-1)\log(1-p))\preceq\mathrm{Pois}(1)$ with $p=1/n$ since

$$-(n-1)\log\left(1-\frac{1}{n}\right)=(n-1)\left(\frac{1}{n}+\frac{1}{2n^2}+\frac{1}{3n^3}+\dots\right)\le(n-1)\cdot\frac{1}{n}\cdot\frac{1}{1-\frac{1}{n}}=1.$$

Hence, for $\lambda=1$, we have $|\mathcal{C}(1)|\preceq\mathrm{Pois}(1)$. This gives the following.

Claim. For any $k>0$, $\Pr(|\mathcal{C}(1)|\ge k)\le 1/\sqrt{k}$.

Proof. Let $\mathcal{T}_1\sim\mathrm{BP}(\mathrm{Pois}(1))$. We see that

$$\Pr(|\mathcal{C}(1)|\ge k)\le\Pr(|\mathcal{T}_1|\ge k)=\sum_{i=k}^{\infty}e^{-i}\frac{(1\cdot i)^{i-1}}{i!}\le\sum_{i=k}^{\infty}\frac{e^{-i}\cdot i^{i-1}}{\sqrt{2\pi}\cdot i^{1/2+i}\cdot e^{-i}}=\frac{1}{\sqrt{2\pi}}\sum_{i=k}^{\infty}\frac{1}{i^{3/2}}\le\frac{1}{\sqrt{k}},$$

where we use the Stirling approximation with $i!\ge\sqrt{2\pi i}\cdot e^{-i}\cdot i^i$. ⊛

Given the above bound, if we want to use the usual union bound to bound the maximum component size, the bound is too weak. However, we can improve upon the union bound in this case as

$$\Pr(|\mathcal{C}_{\max}|\ge k)=\Pr(Z_{\ge k}\ge k)\le\frac{1}{k}\mathbb{E}[Z_{\ge k}]=\frac{n}{k}\Pr(|\mathcal{C}(1)|\ge k)\le\frac{n}{k^{3/2}},$$

hence $k=a\cdot n^{2/3}$ for some $a>0$ suffices. We now restate and prove Theorem 2.1.1 (c) in Lemma 2.3.6:

Lemma 2.3.6 (Component of critical Erdős-Rényi graph)**.** Let $G \sim \mathrm{ER}(n, \lambda/n)$ with $\lambda = 1$.

(a) For any $\epsilon > 0$, for some large $a = a(\epsilon)$, $\liminf_{n\to\infty} \Pr\left(n^{2/3}/a \le |\mathcal{C}_{\max}| \le a \cdot n^{2/3}\right) \ge 1 - \epsilon$.

(b) For any $k > 0$, $\frac{1}{n^{2/3}}(|\mathcal{C}_{\max_1}|, |\mathcal{C}_{\max_2}|, \ldots, |\mathcal{C}_{\max_k}|)$ converges in distribution to some non-degenerated random vectors as $n \to \infty$.

Proof. We already proved the upper bound part of (a). For the lower bound, consider $Z_{\ge n^{2/3}/a}$. We can show that it is concentrated at the mean tightly, and as $a \to \infty$, the mean is small.

For (b), recall the exploration algorithm, where we maintain $(\mathcal{A}_t, \mathcal{U}_t, \mathcal{R}_t)$. We know that $U_t \sim \mathrm{Bin}(n-1, (1-p)^t)$ and $A_t = n - 1 - U_t$ with $A_0 = 1$. We want to study when $A_t = 0$ for some t since this indicates the completion of the exploration of the component.

However, since we want to control k components at once, after a component is fully explored, we continue the exploration by adding a new random vertex as the seed into the set of active vertices. Hence, the corresponding process is defined as

$$\hat{A}_t \coloneqq A_t + \#0 \text{ hitting in } [0, t-1] \text{ in } \hat{A}_t = A_t - \min_{s<t} A_s + 1,$$

where we *add one* to A_t after the current component is fully explored ($\hat{A}_t = 0$).

Intuition. We see that $\hat{A}_t$ again encodes the entire graph into a single path.

To make sense of $\hat{A}_t \coloneqq A_t - \min_{s<t} A_s + 1$, it's worth recalling that $A_t \to -\infty$ as $t \to \infty$:

As previously seen. We have $A_t \overset{D}{=} n - t - \mathrm{Bin}(n-1, (1-p)^t)$ with

$$\mathbb{E}[A_t] = n - t - (n-1)\left(1 - \frac{1}{n}\right)^t \approx 1 - \frac{t}{n} + \frac{t^2}{2n} + \ldots.$$

When $t \ge \sqrt{n}$, the quadratic term dominates, contributing a negative drift. Moreover,

$$\mathrm{Var}[A_t] = (n-1)\left(1 - \frac{1}{n}\right)^t \left(1 - \left(1 - \frac{1}{n}\right)^t\right) \approx te^{-t/n}.$$

One can check that when $1 \ll t \ll n$, by CLT, the standard Binomial converges in distribution to $\mathcal{N}(0,1)$ if and only if $np(1-p) \to \infty$. This implies $A_t \approx -t^2/2n + \sqrt{t} \cdot \mathcal{N}(0,1)$.

Intuition. The timescale where $t^2/n \approx \sqrt{t}$, i.e., $t \approx n^{2/3}$, is necessary to maintain the balance between the subcritical and supercritical behavior.

From the recursive definition of U_t, we can make a martingale. Let B_s denotes the standard Brownian motion, then with martingale CLT, one can prove that

$$\left(\frac{1}{n^{1/3}} A_{\lfloor s \cdot n^{2/3} \rfloor}\right)_{s \ge 0} \xrightarrow{D} \left(-\frac{s^2}{2} + B_s\right)_{s \ge 0}.$$

Hence, for $\hat{A}_t = A_t - \min_{s<t} A_s + 1$, scaling by $n^{2/3}$ now, we have

$$\left(\frac{1}{n^{2/3}} \hat{A}_{s \cdot n^{2/3}}\right)_{s \ge 0} \xrightarrow{D} \left(\left(B_s - \frac{s^2}{2}\right) - \inf_{t \le s}\left(B_t - \frac{t^2}{2}\right)\right)_{s \ge 0},$$

which is precisely the reflected (and thus non-negative) version of the process $(B_s - s^2/2)_{s \ge 0}$. With this, we can use a martingale CLT argument to show that the component sizes converge to the excursion lengths of the $\hat{A}_t$ process. ■

In particular, from the proof of Lemma 2.3.6, we know that $|\mathcal{C}_{\max_1}| = \Theta_p(n^{2/3})$. Furthermore, we can zoom in at the critical region and study what will happen when λ is very close to 1.

Remark (Critical window). When $\lambda = 1 + \theta/n^{1/3}$ for some fixed $\theta \in \mathbb{R}$, the above becomes

$$\left(\frac{1}{n^{1/3}} A_{\lfloor s \cdot n^{2/3} \rfloor}\right)_{s \geq 0} \xrightarrow{D} \left(-\frac{s^2}{2} + B_s + \theta s\right)_{s \geq 0},$$

and

$$\left(\frac{1}{n^{2/3}} \hat{A}_{s \cdot n^{2/3}}\right)_{s \geq 0} \xrightarrow{D} \left(\left(B_S - \frac{s^2}{2} + \theta s\right) - \inf_{t \leq s} \left(B_t - \frac{t^2}{2} + \theta t\right)\right)_{s \geq 0}.$$

Hence, when λ is in a small window $[1 - \theta/n^{1/3}, 1 + \theta/n^{1/3}]$ around 1, we're effectively in the critical regime where the phase transition happens.

This concludes the discussion for the component sizes on the sparse regime where $\lambda = \Theta(1)$.

2.4 Structure Counting in Various Regime

Next, we're interested in understanding the structural emergence behavior as λ varies.

Example (Disconnected edge). Again consider $\mathrm{ER}(n, \lambda/n)$ for some $\lambda \in (0, \infty)$. Then

$$\mathbb{E}[\#\text{disconnected edge}] = \frac{n(n-1)}{2} \cdot \frac{\lambda}{n} \left(1 - \frac{\lambda}{n}\right)^{2(n-2)}.$$

We have done such a counting several times. For instance, one can consider other structures such as 3-chains, cycles, etc. In general, we have the following:

Example. Let k and ℓ be the number of vertices and edges of a specific disconnected structure S, respectively. Then we see that

$$\mathbb{E}[\#S \text{ in } \mathrm{ER}(n, \lambda/n)] = \binom{n}{k} \left(\frac{\lambda}{n}\right)^{\ell} \left(1 - \frac{\lambda}{n}\right)^{k(n-k)} \approx \frac{n^k}{k!} \frac{\lambda^\ell}{n^\ell} e^{-\lambda k}.$$

Intuition. We see that it gets increasingly difficult (with k grows) for k-components to remain isolated. This hints that the bottleneck to connectivity of a graph are isolated vertices.

It turns out that we can characterize this. In particular, we will see that the connectivity threshold is $\log n$: as $\lambda < \log n$, there are single vertices, while after $\lambda > \log n$, the whole graph is connected. We will also study the behavior when $\lambda = \Theta(n)$ later.

Lecture 6: Cycle Counting, Star Counting, and Tree Counting

2.4.1 Stein-Chen Method for Poisson Approximation

6 Feb. 9:30

Let's first summarize some common proof techniques we have seen so far:

As previously seen. For some counting random variable Z (i.e., non-negative integer-valued):

- $\Pr(Z > 0) = \Pr(Z \geq 1) \leq \mathbb{E}[Z]$. For example, $\Pr(|\mathcal{C}_{\max}| \geq k) = \Pr(Z_{\geq k} \geq k) \leq \mathbb{E}[Z_{\geq k}]/k$.
- $\Pr(Z = 0) = \Pr(Z - \mathbb{E}[Z] = -\mathbb{E}[Z]) \leq \Pr(|Z - \mathbb{E}[Z]| \geq \mathbb{E}[Z]) \leq \mathrm{Var}[Z]/(\mathbb{E}[Z])^2$.

To proceed, we will need some tools on $\mathrm{Pois}(\lambda)$. The following notation will be heavily used.

Notation. For $X, k \in \mathbb{N}$, we let $(X)_k \coloneqq X!/k! = X(X-1)\cdots(X-k+1)$.

By some calculation, the following can be shown.

Lemma 2.4.1. For $X \sim \text{Pois}(\lambda)$, $\mathbb{E}[(X)_k] \coloneqq \mathbb{E}[X(X-1)\dots(X-k+1)] = \lambda^k$ for all $k = 1, 2, \dots$.

Surprisingly, if all moments of a random variable converges to what is stated in Lemma 2.4.1, then it indeed will converge in distribution to a Poisson random variable.

Lemma 2.4.2. For a non-negative integer random variable X_n, if $\mathbb{E}[(X_n)_k] \to \lambda^k$ as $n \to \infty$ for all $k = 1, 2, \dots$, then $X_n \xrightarrow{D} \text{Pois}(\lambda)$.

With Lemma 2.4.2, the main tool we will be utilized can be proven (omit due to its length):

Theorem 2.4.1 (Stein-Chen method)**.** Let $(A_i)_{i\geq 1}^n$ be a sequence of events with $p_i = \Pr(A_i)$ for all $i \in [n]$, and let $X = \sum_{i=1}^n \mathbb{1}_{A_i}$ with $\lambda = \mathbb{E}[X] = \sum_{i=1}^n p_i$. If $(A_i)_{i\geq 1}^n$'s are positively associated, i.e., $(A_i)_{i\neq j} \mid A_j \geq (A_i)_{i\neq j}$ for all j, then,

$$d_{\text{TV}}(X, \text{Pois}(\lambda)) = \frac{1}{2}\sum_{k\geq 0}\left|\Pr(X=k) - e^{-\lambda}\frac{\lambda^k}{k!}\right| \leq \min(1, 1/\lambda)\left(\text{Var}[X] - \lambda + 2\sum_{i=1}^n p_i^2\right).$$

On the other hand, if $(A_i)_{i\geq 1}^n$ are negatively associated, i.e., $(A_i)_{i\neq j} \mid A_j \leq (A_i)_{i\neq j}$ for all j, then

$$d_{\text{TV}}(X, \text{Pois}(\lambda)) = \frac{1}{2}\sum_{k\geq 0}\left|\Pr(X=k) - e^{-\lambda}\frac{\lambda^k}{k!}\right| \leq \min(1, 1/\lambda)\left(\lambda - \text{Var}[X]\right).$$

2.4.2 Cycle Counting

Consider the cycle counting problem for $\text{ER}(n, \lambda/n)$ for some $\lambda > 0$:

Problem 2.4.1 (Cycle counting)**.** For some fixed $k \geq 3$, we're interested in controlling

$$X_k \coloneqq \sum_{\substack{(v_1,\dots,v_k), v_i\in[n] \\ v_i\neq v_{i'} \text{ for } i\neq i'}} \Big/ \substack{\text{starting point} \\ \text{orientation}} \mathbb{1}_{(v_1,\dots,v_k) \text{ is a } k\text{-cycle}},$$

where the summation is over all k distinct vertices modulo the starting one and the orientation.

Note. For the cycle counting problem, it's okay that the k-cycle has additional edges, i.e., we care about induced subgraph rather than the exact structure.

The following is easy to see.

Lemma 2.4.3. Let $G \sim \text{ER}(n, \lambda/n)$. When $\lambda < 1$, the expected number of cycles is less than $\sum_{k=3}^\infty \lambda^k/2k < \infty$. Moreover, the expected number of vertices in a cycle is less than $\sum_{k=3}^\infty \mathbb{E}[k\cdot X_k] \leq \sum_{k=3}^\infty \lambda^k/2 < \infty$.

Proof. By a simple counting argument, we see that

$$\mathbb{E}[X_k] = \left(\frac{\lambda}{n}\right)^k \cdot \binom{n}{k}\frac{k!}{2\cdot k} = \frac{\lambda^k}{n^k}\cdot\frac{n(n-1)\dots(n-k+1)}{2k} = \frac{\lambda^k}{2k}\prod_{i=1}^k\left(1-\frac{i}{n}\right) \approx \frac{\lambda^k}{2k}e^{-\frac{k(k-1)}{2n}}.$$

Hence, when $k \ll \sqrt{n}$, we have $\mathbb{E}[X_k] \lessapprox \lambda^k/2k$, which becomes more vacuous as k increases. ∎

We can also calculate the variance of X_k. In particular, we see that

$$\text{Var}[X_k] = \binom{n}{k}\frac{k!}{2k}\cdot\left(\frac{\lambda}{n}\right)^k\left(1-\left(\frac{\lambda}{n}\right)^k\right) + O\left(\sum_{s=1}^{k-2}\binom{n}{k}\cdot\frac{k!}{2k}\cdot n^{k-s-1}\cdot\left(\frac{\lambda}{n}\right)^{k+k-s}\right),$$

where the big-O (second) term is the covariance: If two cycles don't share edges, then the covariance is 0. Otherwise, it is strictly greater than 0, with s being the number of shared edges between these two

cycles. In particular, we can show that as $n \to \infty$, we have

$$\sum_{s=1}^{k-2} \binom{n}{k} \cdot \frac{k!}{2k} \cdot n^{k-s-1} \left(\frac{\lambda}{n}\right)^{k+k-s} \leq \sum_{s=1}^{k-2} \frac{\lambda^{2k-s}}{2k} \cdot \frac{1}{n} \to 0.$$

From the Stein-Chen method, we can show the following.

Theorem 2.4.2. Let $G \sim \mathrm{ER}(n, \lambda/n)$. For a fixed $\lambda > 0$ and $k \geq 3$, we have

- $X_k \xrightarrow{D} \mathrm{Pois}(\lambda^k/2k)$ as $n \to \infty$.
- For any fixed d, $(X_k)_{k=3}^d \xrightarrow{D} \bigotimes_{k=3}^d \mathrm{Pois}(\lambda^k/2k)$ as $n \to \infty$.
- For any fixed d, $\sum_{k=3}^d X_k \xrightarrow{D} \mathrm{Pois}(\sum_{k=3}^d \lambda^k/2k)$ for all $\lambda > 0$.
- If $\lambda < 1$, the above converges, i.e., $\sum_{k=3}^\infty X_k \xrightarrow{D} \mathrm{Pois}(\sum_{k=3}^\infty \lambda^k/2k)$.

2.4.3 Tree Counting

The next elementary object after cycles might be trees. Let $G \sim \mathrm{ER}(n, \lambda/n)$, consider the problem of degree counting, which simply corresponds to the star graph. Fix $k \geq 0$, then the number of vertices with degree k is defined as

$$N_k = \sum_{v=1}^n \mathbb{1}_{\deg(v)=k}.$$

We see that

$$\mathbb{E}[N_k] = n \cdot \binom{n-1}{k} \cdot \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-1-k} \approx n \cdot \frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k} e^{-\lambda} = n \cdot \frac{\lambda^k}{k!} e^{-\lambda}.$$

Hence, for all $k \geq 0$, as $n \to \infty$,

$$\frac{1}{n}\mathbb{E}[N_k] \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

We can also calculate the variance of N_k, which is

$$\mathrm{Var}[N_k] = n\left(\frac{\lambda^k e^{-\lambda}}{k!}\left(1 - \frac{\lambda^k e^{-\lambda}}{k!}\right) + o(1)\right) + n(n-1)\left(\Pr(\deg(1) = k, \deg(2) = k) - \Pr(\deg(v) = k)^2\right)$$

We see that $\Pr(\deg(v) = k)^2 = \Pr(\mathrm{Bin}(n-1, p) = k)^2$ and

$$\Pr(\deg(1) = k, \deg(2) = k) = \frac{\lambda}{n} \Pr(\mathrm{Bin}(n-2, p) = k-1)^2 + \left(1 - \frac{\lambda}{n}\right) \Pr(\mathrm{Bin}(n-2, p) = k)^2.$$

Overall, we have $\Pr(\deg(1) = k, \deg(2) = k) - \Pr(\deg(v) = k)^2 \approx c_\lambda/n + \ldots$.

Theorem 2.4.3. Let $G \sim \mathrm{ER}(n, \lambda/n)$. For any fixed $k \geq 0$, $\lambda > 0$, and $\ell \geq 1$, as $n \to \infty$, we have

$$\left(\frac{N_k - \mathbb{E}[N_k]}{\sqrt{n}}\right)_{k=1}^\ell \xrightarrow{D} \mathcal{N}_\ell(0, D)$$

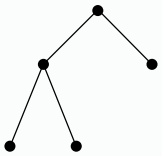
where $D \in \mathbb{R}^{\ell \times \ell}$ is a positive definite covariance matrix.

Note. In the above calculation, we assume that there can be edges presented between the non-center vertices of the star.

Remark. For the counting problem, the Poisson approximation holds if and only if $\mathrm{Var}[Z]/\mathbb{E}[Z] \to 1$ as $n \to \infty$.

Using the same idea, we can consider any given tree structure.

Intuition. Consider the following tree with 5 vertices and 4 edges:



We see that the number of components with this tree structure has mean $\approx n^5 \cdot \lambda^4/n^4 = n\lambda^4$.

One can actually consider a more restrictive version of the structure counting, where we require the structure to be presented *exactly*, i.e., in the induced subgraph-sense. In general, one can show that for any tree $\mathcal{T}$,

$$\left(\frac{N_{\mathcal{T}} - \mathbb{E}[N_{\mathcal{T}}]}{\sqrt{n}}\right)_{\mathcal{T}} \xrightarrow{D} \mathcal{N}_\ell(0, D)$$

for some non-degenerate D, where ℓ is the number of trees $\mathcal{T}$'s we considered jointly.

Example. Consider a tree $\mathcal{T}$ that is a 3-chain. Then, $\mathbb{E}[N_{\mathcal{T}}] \approx n\lambda^2 e^{-3\lambda}$, where $N_{\mathcal{T}}$ counts the number of induced subgraph of 3 vertices being $\mathcal{T}$.

Proof. We see that by a similar calculation,

$$\mathbb{E}[N_{\mathcal{T}}] = \binom{n}{3} \cdot 3! \cdot \left(\frac{\lambda}{n}\right)^2 \left(1 - \frac{\lambda}{n}\right)^{3(n-3)+1} \approx n\lambda^2 e^{-3\lambda},$$

with the difference being not allowing extra edges present between the two vertices at the end. ⊛

Example. More generally, given a cluster $\mathcal{T}_k$ of k nodes and $k-1$ edges, for some small constant c_k,

$$\mathbb{E}[N_{\mathcal{T}_k}] = n\lambda^{k-1}e^{-k\lambda} \cdot \Theta_k(1) = \exp(\log n - k\lambda + (k-1)\log\lambda + c_k).$$

The above calculation has some interesting implications. If $\lambda > (1+\epsilon)\log n/k$: $\mathbb{E}[N_{\mathcal{T}_k}] \to 0$ as $n \to \infty$, i.e., no $\mathcal{T}_k$ will appear. In particular, when $k = 1$, we see that as $\lambda > \log n$, the graph becomes connected since no isolating vertices (i.e., the $\mathcal{T}_1$ structure) will appear.

Theorem 2.4.4. Let $\lambda = \log n + c$ for some constant $c > 0$. Then, as $n \to \infty$

$$\Pr\left(G \sim \mathrm{ER}(n, \lambda/n) \text{ is connected}\right) \to e^{-e^{-c}}.$$

Moreover, let $N_1 \coloneqq N_{\mathcal{T}_1}$ be the number of isolating vertices. Then as $n \to \infty$,

$$N_1 \xrightarrow{D} \mathrm{Pois}(e^{-c}).$$

Appendix

Bibliography

[Dur10] Rick Durrett. *Random graph dynamics*. Vol. 20. Cambridge university press, 2010.

[JLR11] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*. John Wiley & Sons, 2011.

[Van24] Remco Van Der Hofstad. *Random graphs and complex networks*. Vol. 54. Cambridge university press, 2024.