MATH595
Stochastic Processes on Graphs

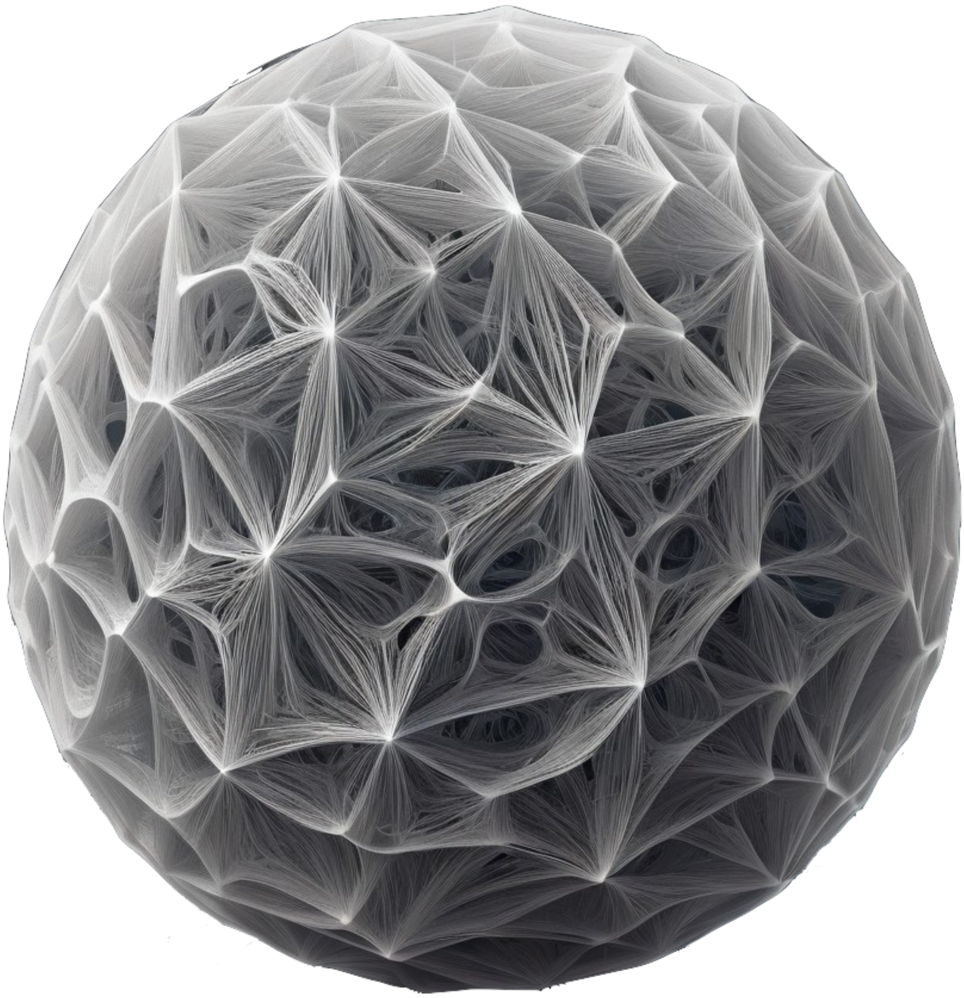
Pingbang Hu

March 13, 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

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 $1 \leq i \leq k - 1$.

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?

For instance, the following is famous.

Theorem 1.1.1 (Cayley's formula)**.** It states that for every positive integer n, the number of trees on n labeled vertices is n^{n-2}.

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 $V = [n] \coloneqq \{1, \ldots, n\}$ and $p \in [0, 1]$. For every $1 \leq i < j \leq n$, we let $\omega_{ij} \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(p)$, which induces $E \coloneqq \{(i, j) \mid \omega_{ij} = 1, 1 \leq i < j \leq n\}$.

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,

$$\frac{|E| - \binom{n}{2}p}{\sqrt{\binom{n}{2}p(1-p)}} \overset{D}{\to} \mathcal{N}(0, 1),$$

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 $|E| = \sum_{1 \leq i < j \leq n} \omega_{ij} \sim \mathrm{Bin}(\binom{n}{2}, p)$, 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 $\mathrm{Bin}(n-1, p) = \mathrm{Bin}(n-1, \lambda/n)$.

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

Proof. We see this in a straightforward way: for any k, $X \sim \mathrm{Bin}(n-1, \lambda/n)$ 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 $\mathrm{Pois}(\lambda)$. Hence, by definition, $\mathrm{Bin}(n-1, \lambda/n) \xrightarrow{D} \mathrm{Pois}(\lambda)$. 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,

$$d_{\mathrm{TV}}((p_k)_{k\in\Omega}, (r_k)_{k\in\Omega}) := \frac{1}{2} \sum_{k\in\Omega} |p_k - r_k|.$$

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

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

and by Jensen's inequality,

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

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

The above gives a distance-one neighborhood characterization of $\mathrm{ER}(n, p)$. 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 $\mathrm{ER}(n, \lambda/n)$:

Theorem 2.1.1. Consider the Erdős-Rényi random graph model $\mathrm{ER}(n, \lambda/n)$ 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 $\Pr(|\mathcal{C}_{\max_1}| \geq a \log n) \to 0$.

(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.

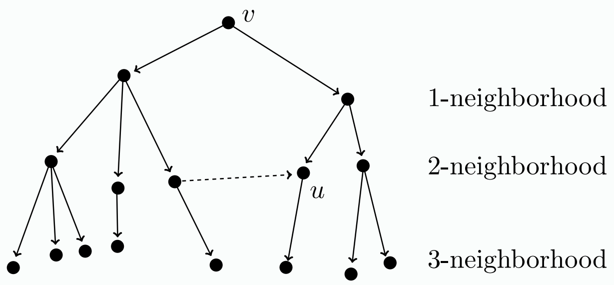
Intuition. Consider the extremely sparse regime where $\lambda < 1$. We give a heuristic argument of why there can't exist a large component. The neighborhood structure of some vertex v, which should be tree-like, at least locally. This is because, for any $k \geq 2$, the expected number of cycles of 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 $\deg_{G_n}(v) \xrightarrow{D} \mathrm{Pois}(\lambda)$ as $n \to \infty$ where $G_n \sim \mathrm{ER}(n, \lambda/n)$.

2.2.2 Joint Distribution

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

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

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 $\mathrm{Bin}(\binom{k}{2}, \lambda/n)$, 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 $\mathrm{Pois}(\lambda)$. ⊛

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 $\mathrm{Pois}(\lambda)$ in the total variation distance.

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

Proposition 2.2.1. Consider the Erdős-Rényi random graph model $\mathrm{ER}(n, \lambda/n)$ 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

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

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

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

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 $\theta > 0$,

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

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 $Z_{\geq k_n} \coloneqq \sum_{v=1}^{n} \mathbb{1}_{|\mathcal{C}_n(v)| \geq k_n}$ as the number of vertices that has a component size greater than k_n. Then:

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

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

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

Now, to analyze $Z_{\geq k_n}$, 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 $t \geq 1$.

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 $\xi_{t+1} \coloneqq |\mathcal{C}_{t+1}| \sim \mathrm{Bin}(U_t, p)$ 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 $t \geq 0$, $U_t \sim \mathrm{Bin}(U_0, (1-p)^t) = \mathrm{Bin}(n-1, (1-p)^t)$. 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 $0 < t < n\zeta_\lambda$, $\Pr(A_t = 0) \leq \exp\left(-t \cdot I_{g(t/n)}\right)$[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

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

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

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

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 $t/n \leq \alpha < \zeta_\lambda$, $\Pr(A_t = 0) \leq e^{-tc(\alpha)}$ 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 $1 < g(\alpha) \leq g(t/n)$, which implies $I_{g(\alpha)} \leq I_{g(t/n)}$. 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

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

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 $n \to \infty$,

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

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:

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

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

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

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

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

For the variance, we have

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

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

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

As for the second term, we see that

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

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,

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

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, $\mathrm{Var}[Z_{\leq k_n}] \leq n(k \log n + \lambda k^2 \log^2 n)$, which gives

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

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 $0 < \alpha < \zeta_\lambda$, $|Z_{\geq \alpha n} - n\zeta_\lambda| \leq n^{1/2+\epsilon}$. Moreover, $|\mathcal{C}_{\max}| = Z_{\geq \alpha n}$ 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 $|\mathcal{C}(v)| \leq k \log n$ 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

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

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

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

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

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

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

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

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

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 $a = a(\epsilon)$, $\liminf_{n\to\infty} \Pr\left(n^{2/3}/a \leq |\mathcal{C}_{\max}| \leq a \cdot n^{2/3}\right) \geq 1 - \epsilon$.

(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 $Z_{\geq n^{2/3}/a}$. 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 $t \geq \sqrt{n}$, 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

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

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

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

which is precisely the reflected (and thus non-negative) version of the process $(B_s - s^2/2)_{s \geq 0}$. 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

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

and

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

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 Connectivity Threshold via Structure Counting

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 v_S and e_S be the number of vertices and edges of a specific structure S, respectively. Then we see that with c_S being the number of S structure on v_S many labeled vertices,

$$\mathbb{E}[\#S \text{ in } \mathrm{ER}(n, \lambda/n)] = \binom{n}{v_S} \cdot c_S \cdot \left(\frac{\lambda}{n}\right)^{e_S}\left(1 - \frac{\lambda}{n}\right)^{v_S(n-v_S) + (\binom{v_S}{2} - e_S)} \approx \frac{n^{v_S}}{v_S!}\frac{\lambda^{e_S}}{n^{e_S}} e^{-\lambda v_S}.$$

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, where we can count the frequency of a particular cluster and (induced/injective) structure appears in $\mathrm{ER}(n,p)$. In particular, this gives the connectivity threshold to be $\log n$: as $\lambda < \log n$, there are single vertices, while after $\lambda > \log n$, the whole graph is connected since all finite cluster disappears. We will also study the behavior when $\lambda = \Theta(n)$ later.

Lecture 6: Subgraph Counting in Sparse Regime

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 $(X)_k := X!/k! = X(X-1)\cdots(X-k+1)$.

By some calculation, the following can be shown.

Lemma 2.4.1. For $X \sim \mathrm{Pois}(\lambda)$, $\mathbb{E}[(X)_k] \coloneqq \mathbb{E}[X(X-1)\dots(X-k+1)] = \lambda^k$ 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 $X_n \overset{D}{\to} \mathrm{Pois}(\lambda)$.

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,

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

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

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

2.4.2 Injective Cycle Counting

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

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

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

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 $G \sim \mathrm{ER}(n, \lambda/n)$. 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

$$\mathrm{Var}[X_k] = \binom{n}{k}\frac{k!}{2k}\cdot\left(\frac{\lambda}{n}\right)^k\left(1 - \left(\frac{\lambda}{n}\right)^k\right) + O\left(\sum_{s=1}^{k-2}\binom{n}{k}\cdot\frac{k!}{2k}\cdot n^{k-s-1}\cdot\left(\frac{\lambda}{n}\right)^{k+k-s}\right), \tag{2.2}$$

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 Injective 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

$$\mathbb{E}[N_k] = n \cdot \binom{n-1}{k} \cdot \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-1-k} \approx n \cdot \frac{n^k}{k!} \cdot \frac{\lambda^k}{n^k} e^{-\lambda} = n \cdot \frac{\lambda^k}{k!} e^{-\lambda}. \tag{2.3}$$

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

$$\begin{aligned} \mathrm{Var}[N_k] = n &\left(\frac{\lambda^k e^{-\lambda}}{k!}\left(1 - \frac{\lambda^k e^{-\lambda}}{k!}\right) + o(1)\right) \\ &+ n(n-1)\left(\Pr(\deg(1) = k, \deg(2) = k) - \Pr(\deg(v) = k)^2\right). \end{aligned} \tag{2.4}$$

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

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

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

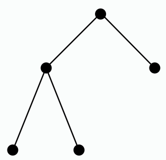
where $D \in \mathbb{R}^{\ell \times \ell}$ is a positive definite covariance matrix.

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 $\operatorname{Var}[Z]/\mathbb{E}[Z] \to 1$ 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 $\approx n^5 \cdot \lambda^4/n^4 = n\lambda^4$.

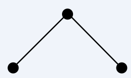
2.4.4 Connected Component Counting and Connectivity Threshold

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 exact connected component sense. In general, one can show that for any tree $\mathcal{T}$, let $N_{\mathcal{T}}$ to be the number of clusters that look like $\mathcal{T}$ in $\mathrm{ER}(n, p)$. Then,

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

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 since there are 3 possible edge configurations for $\mathcal{T}$ among three vertices,

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

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 $\lambda > (1+\epsilon)\log n/k$, then $\mathbb{E}[N_{\mathcal{T}_k}] \to 0$ as $n \to \infty$, i.e., no $\mathcal{T}_k$ will appear. In particular, we will show that as $\lambda > \log n$, i.e., when isolating vertices stop appearing, G becomes connected.

Intuition. Consider $\lambda > (1+\epsilon)\log n/k$ and k decreases from a large number until $k = 1$. We see that $\mathcal{T}_k$ stop showing in order, and in the end, the graph becomes connected.

Lecture 7: Connectivity Threshold from Counting

As previously seen. When $\lambda = pn > 1$, a giant component exists.

11 Feb. 9:30

Now, let's formalize the cluster counting calculation: when $\lambda > \log n/k$ for any fixed $k \geq 1$, with

Cayley's formula,

$$\begin{aligned}\mathbb{E}[\#\text{clusters of size } k] &\leq \mathbb{E}[\#\text{clusters having a spanning tree of size } k] \\ &\leq k^{k-2}\cdot\binom{n}{k}\cdot\left(\frac{\lambda}{n}\right)^{k-1}\left(1-\frac{\lambda}{n}\right)^{k(n-k)} \\ &\leq k^{k-2}\cdot\frac{n^k}{e^{-k}k^{k+1/2}}\cdot\frac{\lambda^{k-1}}{n^{k-1}}\cdot e^{-\lambda k(n-k)/n} = \frac{n}{\lambda k^{5/2}}\left(e\lambda e^{-\frac{\lambda(n-k)}{n}}\right)^k.\end{aligned} \tag{2.5}$$

When $k\geq 1$ is fixed and $\lambda k^2 \ll n$, the above is bounded by $\frac{n}{\lambda}(e\lambda e^{-\lambda})^k$. Hence, when $\lambda=(1+\epsilon)\log n/k$ for some $\epsilon>0$, this bound goes to 0 as $n\to\infty$.

Remark. When $\lambda > \log n/k$, we start to see clusters of size k vanish as $n\to\infty$.

Theorem 2.4.4. Let $G\sim \mathrm{ER}(n,\lambda/n)$ with $\lambda=\log n + c$ for some constant $c>0$. Let Z be the number of isolating vertices in G. Then, as $n\to\infty$, $Z \overset{D}{\to} \mathrm{Pois}(e^{-c})$, and in particular, $\Pr(\text{no isolated vertices}) \to e^{-e^{-c}}$.

Proof. One can use Stein-Chen method or moment method. However, here we consider computing the falling factorial moment directly. Firstly,

$$\mathbb{E}[Z] = n\Pr(\text{vertex 1 is isolated}) = n(1-p)^{n-1} = n\left(1-\frac{\log n + c}{n}\right)^{n-1} \approx ne^{-\log n - c + o(1)} \to e^{-c}.$$

In general, for a fixed $k\geq 2$, we have

$$\begin{aligned}\mathbb{E}[(Z)_k] &= \mathbb{E}\left[\sum_{i_1,\ldots,i_k\in[n]} \mathbb{1}_{i_1,\ldots,i_k \text{ are isolated}}\right] \\ &= (n)_k\cdot\Pr(\text{vertices } 1,2,\ldots,k \text{ are isolated}) \\ &\approx n^k(1-p)^{k(n-k)}(1+o(1)) \approx (n(1-p)^n)^k \to (e^{-c})^k.\end{aligned}$$

Since this is true for all k, from Lemma 2.4.2, we're done. ■

It turns out that our intuition is correct: i.e., when isolating vertices stop showing, the whole graph becomes connected.

Theorem 2.4.5. Let $G\sim\mathrm{ER}(n,\lambda/n)$ with $\lambda=\log n + c$ for some constant $c>0$. Then, as $n\to\infty$,

$$\Pr(G \text{ is connected}) \to e^{-e^{-c}}.$$

Proof. It's clear that $\{G \text{ is connected}\}\subseteq\{G \text{ has no isolated vertices}\}$,

$$\begin{aligned}0 &\leq \Pr(G \text{ has no isolated vertices}) - \Pr(G \text{ is connected}) \\ &= \Pr(\exists \text{clusters of size } k \text{ for } k\in\{2,3,\ldots,\lceil n/2\rceil\}) \\ &\leq \sum_{k=2}^{\lceil n/2\rceil}\mathbb{E}[\#\text{clusters of size } k],\end{aligned}$$

then from Equation 2.5, we have

$$\leq \sum_{k=2}^{\lceil n/2\rceil}\frac{n}{\lambda k^{5/2}}\left(e\lambda e^{-\frac{\lambda(n-k)}{n}}\right)^k = \sum_{k=2}^{c}\frac{n}{\lambda k^{5/2}}\left(e\lambda e^{-\frac{\lambda(n-k)}{n}}\right)^k + \sum_{k=c+1}^{\lceil n/2\rceil}\frac{n}{\lambda k^{5/2}}\left(e\lambda e^{-\frac{\lambda(n-k)}{n}}\right)^k,$$

where we split the sum into two at $k=c$ for some constant c. It's easy to see that the first term goes to 0 as $n\to\infty$, while the second term is upper bounded by $n/\lambda(e\lambda e^{-\lambda/2})^c$. By selecting $c\geq 3$, this also goes to 0 as $n \to \infty$. Hence, we conclude that

$$\Pr(G \text{ is connected}) = \Pr(G \text{ has no isolated vertices}) \to e^{-e^{-c}}$$

from Theorem 2.4.4, proving the result. ■

In fact, not only the number of isolating vertices follows $\mathrm{Pois}(e^{-c})$ (assuming $\lambda = \log n + c$). Under suitable regime of λ, the number of degree k vertices also follows $\mathrm{Pois}(e^{-c})$:

Theorem 2.4.6. For $G \sim \mathrm{ER}(n, \lambda/n)$ with $\lambda = \log n + k \log\log n + c$ for some fixed $c \in \mathbb{R}$. Then, $N_k \xrightarrow{D} \mathrm{Pois}(e^{-c})$ for any fixed $k \geq 0$, where $N_k \coloneqq \sum_{v=1}^{n} \mathbb{1}_{\deg(v)=k}$.

Proof. From Equation 2.3, with a more careful calculation,

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

We want $n\lambda^k e^{-\lambda} = \Theta(1)$, hence $\lambda = \log n + \log\log^k n + c$. When we choose this λ, $\mathbb{E}[N_k] \to e^{-c}$. ■

Remark. In general, *rare* events are modeled by Poisson.

Here, we mention one last example about Hamiltonian cycle without proving it.

Definition 2.4.1 (Hamiltonian cycle). A *Hamiltonian cycle* in a graph $G = (V, E)$ is an n-cycle with $n = |V|$, i.e., a cycle passing through all vertices.

Theorem 2.4.7. For $G \sim \mathrm{ER}(n, \lambda/n)$ with $\lambda = \log n + \log\log n + c$ for some fixed $c \in \mathbb{R}$,

$$\Pr(G \text{ contains a Hamiltonian cycle}) \to e^{-e^{-c}}.$$

Proof idea. It's obvious that

$$\{G \text{ contains a Hamiltonian cycle}\} \subseteq \{\deg(v) \geq 2 \text{ for all } v \in [n]\} = \{N_0 = N_1 = 0\}.$$

It turns out that $\{N_1 = 0\}$ dominates, hence the probability converges to $e^{-e^{-c}}$ as desired. ■

Note. When p increases, small clusters and small degree vertices vanish, while "more structures" appear.

2.5 Existence Threshold and Subgraph Density

We have looked at the counting problem of a given injective structure (e.g., cycles, trees), which leads to the characterization of connectivity threshold of $G \sim \mathrm{ER}(n, p)$ (Theorem 2.4.5).

As previously seen (Cycle). For the cycle counting with random variable X_k, Lemma 2.4.3 and Equation 2.2 gives $\mathbb{E}[X_k]$ and $\mathrm{Var}[X_k]$.

As previously seen (Degree). For the degree counting with random variable N_k, Equation 2.3 and Equation 2.4 gives $\mathbb{E}[N_k]$ and $\mathrm{Var}[X_k]$.

However, for general structure of size k, computing the variance of the counting random variable becomes intractable due to the correlations, unlike the case for cycles and trees. To get a finer control on the counting random variable, e.g., concentration, we introduce the so-called *chaos decomposition*.

2.5.1 Chaos Decomposition

We consider the triangle graph as our running example to illustrate the idea of *chaos decomposition*.

Example (Running example). Fix a triangle graph $H = \triangle$. Then, we're interested in $X_n(H)$, the number of copies (in the injective-sense) of H in the graph $G \sim \mathrm{ER}(n, p)$. We see that

$$X_n(\triangle) = \sum_{1 \leq i < j < k \leq n} \mathbb{1}_{(i,j),(j,k),(k,i) \in E}.$$

It's easy to see that $\mathbb{E}[X_n(\triangle)] = \binom{n}{3} p^3 \approx \frac{(np)^3}{6}$. But what about its variance?

As previously seen. Recall that $\omega_{ij} = \mathbb{1}_{(i,j) \in E} \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}(p)$ for all $i < j$.

The chaos decomposition starts by **centering** ω_{ij}:

$$\begin{aligned} X_n(\triangle) = \sum_{i<j<k} \omega_{ij}\omega_{jk}\omega_{ki} &= \sum_{i<j<k} (\overline{\omega}_{ij} + p)(\overline{\omega}_{jk} + p)(\overline{\omega}_{ki} + p) \\ &= \sum_{i<j<k} \overline{\omega}_{ij}\overline{\omega}_{jk}\overline{\omega}_{ki} + p(\overline{\omega}_{ij}\overline{\omega}_{jk} + \overline{\omega}_{ij}\overline{\omega}_{ki} + \overline{\omega}_{jk}\overline{\omega}_{ki}) + p^2(\overline{\omega}_{ij} + \overline{\omega}_{jk} + \overline{\omega}_{ki}) + p^3, \end{aligned}$$

where $\overline{\omega} \coloneqq \omega - p$. By regrouping, we then have the following decomposition

$$= \underbrace{\sum_{i<j<k} \overline{\omega}_{ij}\overline{\omega}_{jk}\overline{\omega}_{ki}}_{A_3} + \underbrace{p \sum_{i,j<k} \overline{\omega}_{ij}\overline{\omega}_{ik}}_{A_2} + \underbrace{(n-2)p^2 \sum_{i<j} \overline{\omega}_{ij}}_{A_1} + \binom{n}{3} p^3,$$

which is the so-called *chaos decomposition*. This is useful since $\overline{\omega}_{ij}$ is a mean zero, independent random variables. Hence, the correlation between two sums of *different* orders (of $\overline{\omega}_{ij}$'s) will be zero, since one of the $\overline{\omega}_{ij}$'s will be of odd order, resulting in 0. We hence have

$$\begin{cases} \mathbb{E}[A_3] = 0, \quad \mathrm{Var}[A_3] = \binom{n}{3} p^3 (1-p)^3 \approx n^3 p^3 = (np)^3; \\ \mathbb{E}[A_2] = 0, \quad \mathrm{Var}[A_2] = p^2 \cdot n \binom{n-1}{2} \cdot p^2 (1-p)^2 \approx n^3 p^4 = (np)^3 \cdot p; \\ \mathbb{E}[A_1] = 0, \quad \mathrm{Var}[A_1] = (n-2)^2 p^4 \cdot \binom{n}{2} p(1-p) \approx n^4 p^5 = (np)^3 \cdot np^2. \end{cases}$$

Claim. Let $np^2 \to \infty$, and $n^2 p(1-p) \to \infty$, then

$$\frac{X_n(\triangle) - \binom{n}{3} p^3}{\sqrt{\mathrm{Var}[A_1]}} \overset{D}{\to} \mathcal{N}(0, 1).$$

Proof. Since we have

$$\frac{\overline{X}_n(\triangle)}{\sqrt{\mathrm{Var}[A_1]}} = \frac{A_3}{\sqrt{\mathrm{Var}[A_1]}} + \frac{A_2}{\sqrt{\mathrm{Var}[A_1]}} + \frac{A_1}{\sqrt{\mathrm{Var}[A_1]}}.$$

For the first term, we see that it is $\mathbb{E}[A_3]^2 = \mathrm{Var}[A_3] / \mathrm{Var}[A_1] \to 0$, same for the second term. However, for the last term, we have

$$\frac{A_1}{\sqrt{\mathrm{Var}[A_1]}} = \frac{\sum_{i<j} \overline{\omega}_{ij}}{\sqrt{\binom{n}{2} p(1-p)}},$$

where $\sum_{i<j} \overline{\omega}_{ij} \sim \mathrm{Bin}(\binom{n}{2}, p) - \binom{n}{2} p$, and CLT applies. ⊛

Exercise. Find the CLT threshold for N_H for a fixed connected graph H.

Theorem 2.5.1. For $\mathrm{ER}(n, \lambda/n)$ with $\lambda > 1$. $\mathrm{diam}(\mathcal{C}_{\max_1}) \approx c_\lambda \log n$.

Lecture 8: Existence Threshold using Chaos Decomposition

2.5.2 Existence Threshold for General Structure

13 Feb. 9:30

We now consider using chaos decomposition to compute the subgraph counts of any finite connected graph $F = (V_F, E_F)$ with $v_F = |V_F|$ and $e_F = |E_F|$.

Note. We're interested in the *injective* subgraph rather than the induced subgraph.

Let $X_n(F)$ be the number of copies of F in a graph $G = (V, E)$ with $n = |V|$, which is

$$X_n(F) = \sum_{i_1, \ldots, i_{v_F} \text{ distinct } / \ \mathrm{Aut}(F)} \mathbb{1}_{(i_1, \ldots, i_{v_F}) \text{ contains edges in } F}.$$

Remark (F-density)**.** Clearly, the maximum value of $X_n(F)$ is $(n)_{v_F}/|\mathrm{Aut}(F)|$ when G is a complete graph. Hence, we can define the density of F (called *F-density*) in G as

$$t(F, G) := \frac{X_n(F)}{(n)_{v_F}/|\mathrm{Aut}(F)|}.$$

If $G \sim \mathrm{ER}(n, p)$, then the expectation of $X_n(F)$ is

$$\mathbb{E}[X_n(F)] = \frac{(n)_{v_F}}{|\mathrm{Aut}(F)|} p^{e_F} \approx n^{v_F} p^{e_F}.$$

We see that to decide the existence threshold for the first moment, we want to see $n^{v_F} p^{e_F} \to 0$, which happens if and only if $p \ll 1/n^{v_F/e_F}$. Hence, the existence threshold of F is $p = 1/n^{v_F/e_F}$.

Example. If F is a tree, $v_F/e_F = v_F/(v_F - 1) = 1 + 1/(v_F - 1) > 1$, hence $1/n^{v_F/e_F} \ll 1/n$. This implies that trees always exist in all regimes when $p \cdot n$ is bounded away from 0.

Example. If F is connected but not a tree, $v_F/e_F \leq v_F/v_F = 1$, hence $1/n^{v_F/e_F} \geq 1/n$. This means that the regime must be super-critical for non-tree components to exist with positive probability.

To get concentration of X_n, we need to compute the variance. First, we have

$$X_n(F) - \mathbb{E}[X_n(F)] = \sum_{\varnothing \neq H \subseteq F} n^{v_F - v_H} \cdot p^{e_F - e_H} \cdot c_n(H, F) \cdot \hat{X}_n(H),$$

where $c_n(H, F)$ is the number of copies of H in F, and $\hat{X}_n(\cdot)$ is the version of $X_n(\cdot)$ by replacing all edges variables ω with the centered version $\omega - p$.

Note. We can interpret that $\hat{X}_n(\cdot)$ being the version of $X_n(\cdot)$ where all edge variables ω is replaced by its centered version $\omega - p$.

From the chaos decomposition, we have independence, and hence

$$\mathrm{Var}[X_n(F)] = \sum_{\varnothing \neq H \subseteq F} c_n(H, F)^2 n^{2(v_F - v_H)} p^{2(e_F - e_H)} n^{v_H} p^{e_H} (1-p)^{e_H} \approx \sum_{\varnothing \neq H \subseteq F} n^{2v_F - v_H} p^{2e_F - e_H} (1-p)^{e_H}.$$

For p bounded away from 1, we have $(1-p)^{e_H} \approx 1$. Hence, the ratio of the variance to the square of the mean is

$$\frac{\mathrm{Var}[X_n(F)]}{(\mathbb{E}[X_n(F)])^2} \approx \sum_{\varnothing \neq H \subseteq F} \frac{1}{n^{v_H} p^{e_H}}.$$

From Chebyshev's inequality, for any $\epsilon > 0$, we have

$$\Pr\left(\left|\frac{X_n(F)}{\mathbb{E}[X_n(F)]} - 1\right| > \epsilon\right) = \Pr\left(|X_n(F) - \mathbb{E}[X_n(F)]| > \epsilon\mathbb{E}[X_n(F)]\right) \leq \frac{\mathrm{Var}[X_n(F)]}{\epsilon^2(\mathbb{E}[X_n(F)])^2}.$$

Combining the above, if we want concentration for $X_n(F)$ to $\mathbb{E}[X_n(F)]$, we need $n^{v_H}p_{e_H} \to \infty$ for all $\varnothing \neq H \subseteq F$. Equivalently,

$$n \cdot \min_{\varnothing \neq H \subseteq F} p^{e_H/v_H} = n \cdot p^{\max_{\varnothing \neq H \subseteq F} e_H/v_H} \to \infty.$$

Notation. Define the *concentration parameter* $\theta(F) \coloneqq \max_{\varnothing \neq H \subseteq F} e_H/v_H \geq e_F/v_F$.

We can then rewrite the condition for concentration as $np^{\theta(F)} \to \infty$, or equivalently, $p \gg 1/n^{1/\theta(F)}$. Hence, we conclude that the concentration threshold of $X_n(F)$ is $p = 1/n^{1/\theta(F)}$.

Remark (Balanced subgraph)**.** If the above maximum is taken when $H = F$ with $\theta_F = e_F/v_F$, then $1/n^{v_F/e_F} = 1/n^{1/\theta(F)}$, i.e., the concentration threshold and existence threshold are the same. In this case, we say F is balanced. However, for an unbalanced F, there is a gap between the existence threshold and concentration threshold. We need to look at the maximum $H \subsetneq F$ in this case.

If we now consider the contribution to the variance from edges (i.e., H with $v_H = 2$ and $e_H = 1$), that is, $n^{2v_F-2}p^{2e_F-1}(1-p) = O(n^{2v_F-2}p^{2e_F-1})$, we see that its contribution to $\mathrm{Var}[X_n(F)]$ is of order

$$\frac{\mathrm{Var}[X_n(F)]}{n^{2v_F-2}p^{2e_F-1}} \approx \Theta(1) + \sum_{H \subseteq F\colon e_H > 1} n^{2-v_H}p^{1-e_H}.$$

We see that the term from edges will dominate $\mathrm{Var}[X_n(F)]$ when $n^{v_H-2}p^{e_H-1} \to \infty$ for all $H \subseteq F$ with $e_H > 1$, or equivalently, as $n \to \infty$,

$$n \cdot \min_{H \subseteq F\colon e_H > 1} p^{\frac{e_H-1}{v_H-2}} \to \infty.$$

Notation. Define the *edge-dominant parameter* $\theta_1(F) \coloneqq \max_{H \subseteq F\colon e_H > 1} \frac{e_H-1}{v_H-2} \geq \theta(F)$.

Now, recall that $\mathbb{E}[X_n(F)] \approx n^{v_F}p^{e_F}$, hence as the edge term is dominating in $\mathrm{Var}[X_n(F)]$, we have

$$\mathrm{Var}\left[\frac{X_n(F)}{\mathbb{E}[X_n(F)]}\right] \leq \frac{c_F \cdot n^{2v_F-2}p^{2e_F-1}}{n^{2v_F}p^{2e_F}} \leq \frac{c_F}{n^2p} \to 0$$

for some constant c_F as $n \to \infty$. Now, observe that for a fixed $p \in (0,1)$, $\mathbb{E}[X_n(F)] \approx n^{v_F}$. Hence, we conclude that the density of F converges in probability to p^{e_F} as $n \to \infty$ for all F, i.e., as $n \to \infty$,

$$\frac{X_n(F)}{\mathbb{E}[X_n(F)]} \xrightarrow{p} p^{e_F}.$$

2.5.3 Subgraph Density from Dense Graph Limit Point of View

We digress a bit to talk about a limit of the dense graph sequence. The motivation is that consider a sequence of dense graphs G_n with increasing number of vertices, and we want to have a unified way to define the limiting object for such a sequence. In particular, we can study various properties of this limiting object, e.g., the subgraph density $t(F, G)$. Graphon normalized vertex set to $[0,1]$ and define a function $W(x,y)$ on $[0,1]^2$ that generalizes the adjacency matrix structure.

Definition 2.5.1 (Graphon)**.** A *graphon* is a symmetric measurable function $W\colon [0,1]^2 \to [0,1]$, i.e., $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$.

We see that we can embed any fixed, simple graph F as a graphon W_F.

Example. For a finite simple graph $G = (V, E)$ with $V = [n]$, we can associate a graphon

$$W_G(x, y) \coloneqq \begin{cases} 0, & \text{if } x, y \text{ belongs to a region representing an edge;} \\ 1, & \text{otherwise.} \end{cases}$$

That is, we partition $[0, 1]$ into n equal-length intervals $I_i = [(i-1)/n, i/n]$ for $i \in [n]$, and define $W_G(x, y) = 1$ if and only if $(i, j) \in E$ for $x \in I_i$ and $y \in I_j$.

Example (Constant graphon)**.** An important graphon is the constant graphon $W(x, y) = p \cdot \mathbb{1}_{[0,1]^2}$ with some $p \in (0, 1)$. Intuitively, a constant graphon should correspond to a uniform random graph where each edge is present independently with probability p, i.e., $\mathrm{ER}(n, p)$. We will see this later.

Next, since the described embedding is not unique due to various vertex relabeling, we define the space of graphons by identifying two graphons when one is a relabeling version of another:

Definition 2.5.2 (Graphon space)**.** The *graphon space*, denoted as $\widetilde{\mathcal{W}}$, is defined as

$$\widetilde{\mathcal{W}} \coloneqq \{W \colon [0, 1]^2 \to [0, 1] \mid W \text{ is a graphon}\} \big/ {\sim},$$

where the equivalence reaction $W \sim W'$ holds if there exists a measure-preserving bijection $\varphi\colon [0, 1] \to [0, 1]$ such that $W'(x, y) = W(\varphi(x), \varphi(y))$ for almost all $(x, y) \in [0, 1]^2$.

Now, to introduce the distance over the graphon space, we introduce the following first.

Definition 2.5.3 (Cut norm)**.** The *cut norm* of a graphon W is defined as

$$\lVert W \rVert_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x, y) \, \mathrm{d}x \, \mathrm{d}y \right| .$$

It's probably trivial to see that in Definition 2.5.3, the definition is a bit redundant since we only care about graphons, which is non-negative, i.e., we always have

$$\lVert W \rVert_\square = \int_{[0,1]^2} W(x, y) \, \mathrm{d}x \, \mathrm{d}y.$$

However, what we really care is the "metric" induced by this norm:

Definition 2.5.4 (Cut metric)**.** The *cut metric* between two graphons $\widetilde{W}, \widetilde{Y} \in \widetilde{\mathcal{W}}$ is defined as

$$d_\square(\widetilde{W}, \widetilde{Y}) = \inf_{\varphi\colon [0,1] \to [0,1]} \lVert W - Y \circ \varphi \rVert_\square,$$

where the infimum is taken over all measure-preserving bijections φ and W and Y are any representation graphon of $\widetilde{W}$ and $\widetilde{Y}$, respectively.

Remark. It's also possible to consider defining graphon space by first defining the cut metric and identify two graphons W_1, W_2 to be the same when $d_\square(W_1, W_2) = 0$.

Now, given a graphon W, we can define the F*-density* (or *homomorphism density*) in W similarly as in the usual graph G, which is the continuous version of the F-density in a finite graph:

$$t(F, W) \coloneqq \int_{[0,1]^{v_F}} \prod_{(i,j) \in E(F)} W(x_i, x_j) \, \mathrm{d}x_i \cdots \mathrm{d}x_{v_F}.$$

This again measures how frequently a given finite simple graph F appears as a subgraph within a graphon W. It turns out that the F-density characterizes a graphon exactly:

Theorem 2.5.2 (Graphon convergence). Given a graphon sequence $(W_n) \in \widetilde{\mathcal{W}}$, $d_\square(W_n, W) \to 0$ as $n \to \infty$ if and only if $t(F, W_n) \to t(F, W)$ for all finite connected graph F.

This leads to the following.

Corollary 2.5.1 (Erdős-Rényi random graphs converge to constant graphon). Let $G_n \sim \mathrm{ER}(n, p)$ for $p \in (0, 1)$ be a sequence of Erdős-Rényi random graphs. Then, $d_\square(G_n, p \cdot \mathbb{1}_{[0,1]^2}) \xrightarrow{p} 0$ as $n \to \infty$.

Proof. For a fixed $p \in (0, 1)$, we have proven that $X_n(F)/\mathbb{E}[X_n(F)] \xrightarrow{p} p^{e_F}$ in $G_n \sim \mathrm{ER}(n, p)$. This is equivalent to $t(F, G_n) \xrightarrow{p} p^{e_F}$. On the other hand, let $W \equiv p$, a constant function, then $t(F, W) = p^{e_F}$ as well. From Theorem 2.5.2, the result follows. ■

Lecture 9: Minimum Spanning Tree in Erdős-Rényi Graph

2.6 Minimum Spanning Tree

18 Feb. 9:30

Let $\mathcal{T}_n$ be the set of all spanning trees on $[n]$, with $|\mathcal{T}_n| = n^{n-2}$ from Cayley's formula.

Note. All spanning trees $\tau \in \mathcal{T}_n$ has $(n-1)$ many edges.

We are now interested in the minimum spanning tree problem:

Problem 2.6.1 (Minimum spanning tree). Given a connected graph $G = (V, E)$ with edge capacity $\omega \colon E \to \mathbb{R}_+$, the *minimum spanning tree* (MST) problem aims to find the min-cost spanning tree $\tau^* \in \mathcal{T}_n$ with cost $W_n := \min_{\tau \in \mathcal{T}_n} W(\tau) = W(\tau^*)$, where $W(\tau) := \sum_{e \in \tau} \omega_e$.

MST is a well-studied problem in combinatorial optimization. We will focus on one famous algorithm called the Kruskal's algorithm, which works by sequentially (w.r.t. the edge weight) add edges with the largest weight as long as no cycle is created.

2.6.1 Thresholding View of Erdős-Rényi Graph

In particular, consider the complete graph $G_n = K_n$ on $[n]$ nodes. Given a cdf F, let $\omega_e = \omega_{ij} = \omega_{ji} \overset{\text{i.i.d.}}{\sim} F$ on $[0, \infty)$ with $e = (i, j)$ for all $1 \leq i < j \leq n$. Without loss of generality, let $0 \in \operatorname{supp}(F)$, and we may assume that F is continuous and $F(x)/x^d \to 1$ as $x \searrow 0^+$ for some $d > 0$.

Intuition. This means that the cdf generating the edge weight follows a polynomial decay.

Example. For $\mathcal{U}(0, 1)$ and $\mathrm{Exp}(1)$, the assumption holds for $d = 1$.

Theorem 2.6.1. Let G_n be the complete graph with n vertices and edge weight generated from $F(x)/x^d \to 1$ as $x \searrow 0^+$ for some $d > 0$. The following convergences hold as $n \to \infty$.

- When $d = 1$, $\mathbb{E}[W_n] \to \zeta(3) = \sum_{k=1}^{\infty} 1/k^3$ and $\sqrt{n}(W_n - \zeta(3)) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$ for some $\sigma^2 > 0$.
- For general $d > 0$, we have

$$n^{1/d-1}\mathbb{E}[W_n] \to \sum_{k=1}^{\infty} \frac{1}{dk^3} \frac{\Gamma(k + 1/d - 1)}{\Gamma() k^{1/d-}} =: c_d,$$

  and $\frac{1}{\sqrt{n}}(n^{1/d} W_n - n c_d) \xrightarrow{D} \mathcal{N}(0, \sigma_d^2)$ for some $\sigma_d^2 > 0$.

Proof. Consider the MST τ_n^* found by the Kruskal's algorithm. Focus on the case of $d = 1$ such that $F(x) = x$ for all $0 \leq x \leq 1$. Then, we see that

$$W_n = \sum_{e \in \tau_n^*} \omega_e = \sum_{e \in \tau_n^*} \int_0^1 \mathbb{1}_{x \leq \omega_e} \,\mathrm{d}x = \int_0^1 \sum_{e \in \tau_n^*} (1 - \mathbb{1}_{x > \omega_e}) \,\mathrm{d}x = \int_0^1 \left(n - 1 - \sum_{e \in \tau_n^*} \mathbb{1}_{\omega_e < x} \right) \mathrm{d}x.$$

We see that:

- all edges e with weight $\omega_e \leq x$ gives the edge set of $\mathrm{ER}(n, x)$;
- $n - \sum_{\mathcal{C} \colon \tau \text{ spans } \mathcal{C}} |E(\tau)|$ is the total number of components.

Hence, by letting $C_n(x)$ denotes the number of components in $\mathrm{ER}(n, x)$, we further have

$$W_n = \int_0^1 (\#\text{components in } \mathrm{ER}(n, x) - 1) \,\mathrm{d}x = \int_0^1 (C_n(x) - 1) \,\mathrm{d}x.$$

Note. For general F, we have $W_n = \int_0^1 (C_n(x) - 1) \,\mathrm{d}F^{-1}(x)$.

For some $A > 1$ chosen later, consider decomposing the integral into

$$W_n = \int_0^{A \log n / n} C_n(x) \,\mathrm{d}x - \underbrace{\int_0^{A \log n / n} 1 \,\mathrm{d}x}_{A \log n / n} + \int_{A \log n / n}^1 (C_n(x) - 1) \,\mathrm{d}x.$$

The last integral can be bounded with Theorem 2.4.5 as

$$\begin{aligned} \int_{A \log n / n}^1 (C_n(x) - 1) \,\mathrm{d}x &= \int_{A \log n / n}^1 \mathbb{E}[|C_n(x) - 1|] \,\mathrm{d}x \\ &\leq \int_{A \log n / n}^1 n \Pr(\mathrm{ER}(n, x) \text{ is not connected}) \,\mathrm{d}x \leq n^{2-A}. \end{aligned}$$

Take $A = 3$, and let $x = \lambda / n$, the first integral becomes

$$\int_0^{3 \log n / n} C_n(x) \,\mathrm{d}x = \int_0^{3 \log n} \frac{1}{n} C_n(\lambda / n) \,\mathrm{d}\lambda.$$

Since we care about the number of components, in this regime (i.e., $\lambda \in (0, 3 \log n)$), we know that only one giant component can exist, which doesn't matter since it contributes only $1/n$, we're really interested in the number of small components.

Next, consider decomposing $C_n(\lambda / n)$ into the number of components that are trees plus the number of components that are non-trees. The second quantity, in expectation, is bounded by $O(1 + 1/n)$. While for the first quantity, we see that

$$\sum_{k=1}^n \underbrace{(\#\text{tree-component of size } k)}_{N_n(k, \lambda)} + O_p(1),$$

where $N_n(k, \lambda)$ is the number of tree-components of size k in $\mathrm{ER}(n, \lambda / n)$. Overall, we have

$$W_n = \Theta\left(\frac{\log n}{n} \right) + \sum_{k=1}^n \int_0^{3 \log n} \frac{1}{n} N_n(k, \lambda) \,\mathrm{d}\lambda.$$

From Cayley's formula,

$$\mathbb{E}[N_n(k, \lambda)] = k^{k-2} \binom{n}{k} \left(\frac{\lambda}{n} \right)^{k-1} \left(1 - \frac{\lambda}{n} \right)^{k(n-k) + \binom{k}{2} - k + 1},$$

as $n \to \infty$, we have

$$\mathbb{E}[W_n] \to \sum_{k=1}^{\infty} \frac{k^{k-2}}{k!} \int_0^{\infty} \lambda^{k-1} e^{-\lambda k} \, dx = \sum_{k=1}^{\infty} \frac{k^{k-2}}{k!} k^{-k} (k-1)! = \sum_{k=1}^{\infty} \frac{1}{k^3} = \zeta(3).$$

Now, for k and λ fixed, we know that as $n \to \infty$,

$$\frac{1}{n} \mathbb{E}[N_n(k, \lambda)] \to \frac{k^{k-2}}{k!} \lambda^{k-1} e^{-\lambda k} =: m(k, \lambda)$$

and

$$\frac{1}{n} \operatorname{Var}[N_n(k, \lambda)] \to \sigma^2_{k,\lambda} > 0.$$

Then, we have

$$\frac{N_n(k, \lambda) - n \cdot m(k, \lambda)}{\sqrt{n}} \overset{D}{\to} \mathcal{N}(0, \sigma^2_{k,\lambda}).$$

Moreover, we can let λ varies and make the above a process of λ, and one can show that this actually converges:

$$\left(\frac{N_n(k, \cdot) - n \cdot m(k, \cdot)}{\sqrt{n}} \right)_{\lambda \in (0, \infty)} \overset{D}{\to} \mathrm{GP}(0, \Sigma),$$

i.e., some mean zero Gaussian process with certain covariance structure. One can further consider all k, and the above further becomes

$$\left(\left(\frac{N_n(\cdot, \cdot) - n \cdot m(\cdot, \cdot)}{\sqrt{n}} \right)_{\lambda \in (0, \infty)} \right)_{k=1,2,\dots} \overset{D}{\to} (\mathrm{GP}(0, \Sigma))_{k=1,2,\dots}.$$

■

Exercise. Give a simple proof with a rate of convergence.

2.6.2 Extension

One can also consider the Gibbs measure on $\mathcal{T}_n$: let $0 < \beta < \infty$, then for $\tau \in \mathcal{T}_n$, define

$$\mathbb{P}_{\beta,n}(\tau) = \frac{\exp(-\beta \cdot n \cdot W(\tau))}{Z_n(\beta)}$$

with the partition function $Z_n(\beta) = \sum_{\tau \in \mathcal{T}_n} e^{-\beta n W(\tau)}$.

Intuition. This measure interpolates the uniform spanning tree and the minimum spanning tree:

- when $\beta \to 0$, all $\tau \in \mathcal{T}_n$ receives the same probability mass, hence sampling from $\mathbb{P}_{\beta,n}$ gives a uniformly random spanning tree;
- when $\beta \to \infty$, the difference between $W(\theta)$ is amplified, hence the MST τ^* with W_n receives the largest probability mass.

More questions can be asked in this context:

Problem. If we sample τ_n^* from $\mathbb{P}_{\beta,n}$,

(a) What is $\lim_{n \to \infty} \mathbb{E}[W(\tau_n^*)]$?

(b) What is $\lim_{n \to \infty} \frac{1}{n} \log Z_n(\beta)$ and the σ^2 in $\frac{1}{\sqrt{n}}(\log Z_n(\beta) - \mathbb{E}[\log Z_n(\beta)]) \overset{D}{\to} \mathcal{N}(0, \sigma^2)$?

(c) What is the structure of τ_n^* (local sense/metric sense)?

Chapter 3

Other Random Graph Models

We have a thorough discussion on Erdős-Rényi random graph model. However, as we have seen, the degree distribution is thin-tailed: due to the independent structure of various random variables involved. Real-life networks oftentimes are heavy-tailed, especially the degree distribution. In this chapter, we aim to explore other random graph models that exhibit heavy-tailed behaviors.

3.1 Some Proposed Models

Let's begin by first looking at a closely related model called exponential random graph model, which, unfortunately, is still thin-tailed. In some sense, it tries to specific the subgraph densities for various subgraphs. Next, we look at the inhomogeneous random graph model, which tries to directly specify the degrees. However, both of the proposals fail to generate heavy-tailed behaviors.

3.1.1 Exponential Random Graph

Consider $G \sim \mathrm{ER}(n, p)$ for some $p \in (0, 1)$. We see that

$$\Pr(G) = p^m (1-p)^{\binom{n}{2}-m} = (1-p)^{\binom{n}{2}} e^{\log \frac{p}{1-p} \cdot m}.$$

Notation. Let $\beta(p) := \log(p/(1-p))$ where $\beta \colon [0, 1] \to (-\infty, \infty)$.

Hence, given β, $\mathrm{ER}(n, p)$ with $p = e^\beta/(1 + e^\beta)$, which is the same model $\Pr(G) \propto e^{\beta \cdot m}$. This motivates us to define the following random graph model:

Definition 3.1.1 (Exponential random graph). Fix some finite subgraphs $H_1, \dots, H_k$ with $\beta_1, \dots, \beta_k \in (-\infty, \infty)$. Then, the *exponential random graph*, denoted as $G \sim \mathrm{ERGM}(\beta_1, \dots, \beta_k)$, satisfies

$$\Pr(G) \propto \exp\left(n^2 \sum_{i=1}^{k} \beta_i \frac{\#H_i \text{ in } G}{\#H_i \text{ in } K_n} \right) = \exp\left(n^2 \sum_{i=1}^{k} \beta_k t(H_k, G) \right).$$

Intuition. By specifying the probability $p(\beta_i)$, $\mathrm{ER}(n, p(\beta_i))$ gives a graph with probability $e^{\beta m}$. Here, we do the similar construction such that by specifying all the desired subgraphs H_i and β_i, $\mathrm{ERGM}(\beta_1, \dots, \beta_k)$ encourages each H_i appear with the desired subgraph density according to β_i.

Problem (Open problem). CLT for subgraph count in $\mathrm{ERGM}(\beta_1, \dots, \beta_k)$.

Lecture 10: Failure Attempt to Enforcing Degree Distribution

20 Feb. 9:30

3.1.2 Inhomogeneous Random graph

The next model we will be looking at is the so-called inhomogeneous random graph. We first state its simplest form and abstract from it along the way.

Definition 3.1.2 (Inhomogeneous random graph). The *inhomogeneous random graph* $\mathrm{IRG}(n, (p_{ij})_{i,j\in V})$ is defined with $V = [n]$ and $\omega_{ij} = \mathbb{1}_{(i,j)\in E} \overset{\text{ind.}}{\sim} \mathrm{Ber}(p_{ij})$ for some $p_{ij} > 0$ for all $i < j$.

We see that it is just like $\mathrm{ER}(n, p)$, but now we have different $p = p_{ij}$ for every pair $(i, j) \in V \times V$.

As previously seen. We know that $\deg(i) = \sum_{j\in[n]\setminus\{i\}} \omega_{ij}$. From the Stein-Chen method,

$$d_{\mathrm{TV}}\Big(\deg(i), \mathrm{Pois}\,\big(\underbrace{\textstyle\sum_{j\neq i} p_{ij}}_{\lambda_i}\big)\Big) \leq \min(1, 1/\lambda_i) \sum_{j\neq i} p_{ij}^2 \leq \min(1, \lambda_i) \cdot \max_{j\neq i} p_{ij}.$$

By some calculation, we have $d_{\mathrm{TV}}(\mathrm{Pois}(\lambda), \mathrm{Pois}(\lambda')) \leq |\lambda - \lambda'|$. Hence, we want $\max_{i,j} p_{ij} \ll 1$. To have bounded degree, or at least have finite mean of every $\mathrm{Pois}(\lambda_i)$, we also want $\sum_{i,j} p_{ij} < \infty$.

To choose p_{ij}, consider using a *kernel* $\kappa(\cdot, \cdot)$, which gives the probability between two nodes of specified *types*: specifically, let $i \in [n] \mapsto x_i \in S$, where x_i is called a type.

Example. $S = [0, 1]$ or $\mathbb{R}^+$, i.e., types can be 0.02, 784329, etc.

Now, let $p_{ij} \approx \kappa_n(x_i, x_j)/n$, where $\kappa_n \colon S \times S \to [0, \infty)$ is a symmetric kernel. Define the empirical distribution of types as $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, then, we assume that (S, μ_n, κ_n) satisfies the following.

- The discrete measure on the type converges, i.e., $\mu_n \overset{D}{\to} \mu$ on S as $n \to \infty$. Furthermore, $\frac{1}{n}\sum_{i=1}^n x_i \to \int x\mu(\mathrm{d}x)$.
- κ_n is *graphical* on S with a limiting kernel κ, i.e.,
  (a) κ is a continuous in both arguments.
  (b) $\kappa_n(a_n, b_n) \to \kappa(a, b)$ $\mu \times \mu$-almost surely as $n \to \infty$ when $a_n \to a$ and $b_n \to b$.
  (c) $\frac{1}{n}\sum_j \kappa_n(x, x_j) \to \int_y \kappa(x, y)\mu(\mathrm{d}y) =: \kappa(x, \cdot)$ is well-defined μ-almost everywhere. Moreover, $\frac{1}{n}\sum_{i<j} \kappa_n(x_i, x_j) \to \frac{1}{2}\iint \kappa(x, y)\mu(\mathrm{d}x)\mu(\mathrm{d}y) < \infty$.

Example (Single type). Consider only one type exists, say $S = \{1\}$. Then $\mu_n = \delta_1$ and $\kappa_n(1, 1) = \lambda$. This gives back $\mathrm{ER}(n, \lambda/n)$.

To make sure that $p_{ij} < 1$, we consider the following different (basically equivalent) models:

(a) Chang-Lu '02: $p_{ij} = \min(\kappa_n(x_i, x_j)/n, 1)$.

(b) Norros-Reittu '06 (Poisson Graph Process): $p_{ij} = 1 - e^{-\kappa(x_i,x_j)/n}$.

(c) Boritton-Martin-Lof (Generalized Random Graph) $p_{ij}/(1 - p_{ij}) = \kappa_n(x_i, x_j)/n$, or equivalently, $p_{ij} = \frac{\kappa_n(x_i,x_j)}{n+\kappa_n(x_i,x_j)}$.

Example (Finite type). Consider $S = [r]$, $\mu = (\lambda_1, \dots, \lambda_r)$ such that $\sum_{i=1}^r \lambda_i = 1$ with $\lambda_i > 0$. In this case, $\big(\kappa(i, j)\big)_{i,j\in[r]}$ is an $r \times r$ symmetric matrix.

Basically, we have r blocks B_i's, with each of the size n_i such that $n_i/n \approx \lambda_i$. Between blocks i, j, the edge probability follows $\kappa(i, j)/n$, while within a block i, edges form with probability $\kappa(i, i)/n$.

Rank-1 IRG model: $\kappa(x, y) \approx \varphi(x) \cdot \varphi(y)$ for $x, y \in S$ with $\varphi \colon S \to [0, \infty)$, and $\hat{S} = \varphi(S)$. By reparametrization, we can simply consider $\kappa(x, y) := xy$ for $x, y \in \varphi(S) = [0, \infty)$.

Now, consider the simplification: let every vertex i having a type $w_i := \varphi(x_i)$. Then $\frac{1}{n}\sum_{i=1}^n \delta_{w_i} \overset{D}{\to} F$, where F is just some cdf of μ on $[0, \infty)$, i.e., $\frac{1}{n}\#\{w_i \leq x\} \to F(x)$ for all continuity point x of F. Also, $\frac{1}{n}\sum_{i=1}^n w_i \to \mathbb{E}[W]$ where $W \sim F$.

Consider $p_{ij} = \frac{w_i w_j}{n + w_i w_j}$. Then,

$$\sum_{j \neq i} p_{ij} = w_i \cdot \left(\frac{1}{n} \sum_{j=1}^{n} w_j \right).$$

We see that we can choose $(w_i)_{i \in [n]}$ satisfying the above assumptions, such that

$$p_{ij} = \frac{w_i w_j}{\sum_{k=1}^{n} w_k + w_i w_j}$$

for $i < j$. Then, w_i gives the degree sequence. One can show that $\deg(i) \approx \mathrm{Pois}(w_i)$, hence $\mathbb{E}[\deg(i)] = \mathbb{E}[w_i]$.

What about the general degree distribution? Choose $v_n \sim \mathcal{U}([n])$. Then for any fixed $k = 0, 1, 2, \ldots$,

$$\Pr\left(\deg(v_n) = k\right) = \frac{1}{n} \sum_{i=1}^{n} \Pr\left(\deg(i) = k\right) \cong \frac{1}{n} \sum_{i=1}^{n} e^{-w_i} \cdot \frac{w_i^k}{k!} \to \mathbb{E}\left[e^{-W} \frac{W^k}{k!} \right],$$

which is called a mixed Poisson distribution, which is not the original degree distribution, i.e., W, we want.

In general, we have the following.

Theorem 3.1.1. Let $v_1^{(n)}, \ldots, v_\ell^{(n)}$ be ℓ many samples from $[n]$ chosen uniformly at random without replacement. Then, under this rank-1 model, $(\deg(v_1^{(n)}), \ldots, \deg(v_\ell^{(n)})) \xrightarrow{D} \bigotimes_{i=1}^{\ell} \mathrm{MixedPois}(W)$.

In a general (S, μ, κ),

$$\Pr\left(\deg(v^{(n)}) = k\right) \to \int_S e^{-\kappa(x,\cdot)} \frac{\kappa(x,\cdot)^k}{k!} \mu(\mathrm{d}x).$$

Remark. This is still a thin-tailed degree distribution. For real-life network, degree distribution is often heavy-tailed.

We can also look at the cluster structure. In this case, we will be looking into the so-called *multi-type branching process*. Consider (S, μ, κ). Given any vertex x, we know that the number of children follows $\mathrm{Pois}(\kappa(x, \cdot)) = \int \kappa(x, y) \mu(\mathrm{d}y)$. Then, for each child y_i, the probability follows $\kappa(x, y_i)/\kappa(x, \cdot)$.

Intuition.

$$\tau_\kappa f(x) = \int \kappa(x, y) f(y) \mu(\mathrm{d}y)$$

is a positive integral operator on $L^1(S, \mu)$, with the operator norm being

$$\|\tau_\kappa\|_{\mathrm{op}} = \sup_{f > 0} \frac{\|\tau_\kappa f\|_2}{\|f\|_2}.$$

The MTBP survives with positive probability if and only if $\|\tau_\kappa\|_{\mathrm{op}} > 1$.

Similar to the previous case, when $\|\tau_\kappa\|_{\mathrm{op}} > 1$, then what's the survival probability of this MTBP? Let $\Phi f := 1 - e^{-\tau_\kappa f}$. One can check that there exists a unique non-negative maximum solution ρ of $f = \Phi f$, such that the survival probability starting at type x is $\rho(x)$.

Theorem 3.1.2. Under the (S, μ, k) model and all the assumptions.

(a) If $\|\tau_\kappa\|_{\mathrm{op}} < 1$, $|\mathcal{C}_{\max}|/n \xrightarrow{p} 0$.

(b) If $\|\tau_\kappa\|_{\mathrm{op}} > 1$,

$$\frac{|\mathcal{C}_{\max_1}|}{n} \to \int_S \rho(x)\mu(\mathrm{d}x) \in (0,1),$$

and

$$\frac{|\mathcal{C}_{\max_2}|}{n} \xrightarrow{p} 0.$$

Example. For the rank-1 IRG, let $\kappa(x, y) = xy/\mathbb{E}[W]$ where $x_i \sim \mu_n \approx W$. The only unit eigenvector is $y/\sqrt{\mathbb{E}[W^2]}$, with $\|\tau_\kappa\|_{\mathrm{op}} = \sqrt{\mathbb{E}[W^2]/\mathbb{E}[W]}$. We claim that $\mathbb{E}[W^2] < \mathbb{E}[W]$, then it's sub-critical with all components of size $o(n)$. If $\mathbb{E}[W^2] > \mathbb{E}[W]$, then there exists one giant component with all other components of size $o(n)$. For the critical regime, it depends on $\mathbb{E}[W^3]$. If it's finite, then we have the previous behavior (all components are of order $\Theta(n^{2/3})$); otherwise, the exponent varies.

Remark. When the edge probability follows Bernoulli, the degree is the sum over these independent Bernoulli random variables, which makes it highly concentrated. To escape this, we need to avoid this mean-field structure.

Example. Consider $\kappa(x, y) = c/\max(x, y)$, $S = [0, 1]$, and $\mu \sim \mathcal{U}(0, 1)$. Intuitively, the j^{th} vertex will connect to $i < j$ with probability $p_{ij} = c/i \vee j$.

Lecture 11: Configuration Model

3.2 Configuration Model

25 Feb. 9:30

Now, we want to sample a random graph on n vertices with degree sequence $d_1, d_2, \ldots, d_n$. We first look at the existence theorem.

Theorem 3.2.1 (Erdős-Gallai theorem)**.** Let $\ell_n := \sum_{i=1}^n d_i$ is even and for all $k = 1, 2, \ldots, n$, $\sum_{i=1}^n d_{(i)} \leq k(k-1) + \sum_{i=k+1}^n k \wedge d_{(i)}$, where $d_{(i)}$'s are the order statistics of d_i's. Then, a graph with degree sequence $(d_1, \ldots, d_n)$ exists if and only if the above holds.

Problem. How to create a random d-regular graph?

Answer. Simply conditioning doesn't work since the probability is exponentially small. On the other hand, random adjacency matrix $A = (\mathbb{1}_{(i,j)\in E})_{i,j\in[n]}$ that is symmetric, 0-1 matrices with all diagonals 0. This is actually possible by permutation matrix $P_\pi = (\mathbb{1}_{j=\pi_i})_{i,j}$ with a permutation $\pi \colon [n] \to [n]$ with $i \mapsto \pi_i$. Then, sum of d many i.i.d. permutation matrices will be a random adjacency matrix for d-regular graph with positive probability that is bounded below. Hence, conditioning works here. ⊛

Finally, the configuration method. Given a degree sequence $(d_1, \ldots, d_n)$ satisfying the requirements. Next, consider attaching d_i "half-edges" of the form (i, a) for vertex i with $1 \leq a \leq d_i$. By choosing a random matching between all the half edges, we're done. More specifically, if (i, a) and (j, b) is matched, then we connect i and j by an edge.

Note. It's easy to see that the number of possible matchings are $(\ell_n - 1)!!$.

However, since this will potentially give us a non-simple graph. One can again condition given that the probability is strictly positive. Another way is to fix these self-loops or multi-edges.

Problem. The latter will causes the resulting graph doesn't satisfy the desired degree sequence.

Consider a multi-graph (i.e., self-loops and multi-edges), let x_{ii} be the number of self-loops and x_{ij} be the number of edges between vertices i and j.

Example. For a simple graph, $x_{ii} = 0$ and $x_{ij} \in \{0, 1\}$.

We see that $\deg(i) = 2x_{ii} + \sum_{j \neq i} x_{ij} = x_{ii} + \sum_{j=1}^{n} x_{ij}$. Let $S_n := \sum_{i=1}^{n} x_{ii}$ is the total number of self-loops, and $M_n := \sum_{i<j} (x_{ij} - 1) \mathbb{1}_{x_{ij} \geq 2}$ is the total number of multi-edges.

How to choose the degree sequence? Consider there exists a D such that

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{d_i} \xrightarrow{\text{w}} D \Leftrightarrow \sum_{k=0}^{\infty} \left| \frac{1}{n} |\{i \mid d_i = k\}| - \Pr(D = k) \right|,$$

where the equivalence is due to the fact that we're considering countable elements.

A deterministic way is that given D put n_k many k's for $k \geq 0$ where $n_k \approx n \cdot \Pr(D = k)$.

Example. $\lceil n \cdot \Pr(D \leq k) \rceil - \lceil n \cdot \Pr(D \leq k - 1) \rceil$.

A random way is that given $d_i, \ldots, d_n \overset{\text{i.i.d.}}{\sim} D$.

We also assume $\mathbb{E}[D] < \infty$ and $\frac{1}{n} \sum_{i=1}^{n} d_i \to \mathbb{E}[D]$ as $n \to \infty$.

Then, $\ell_n = \sum_{i=1}^{n} d_i \approx n \cdot \mathbb{E}[D]$.

We now compute the probability of getting a specific multi-graph.:

$$\Pr\left(\mathrm{CM}((d_i)_{i \in [n]}) \text{ gives } (x_{ij})_{i \leq j}\right) = \frac{1}{(\ell_n - 1)!!} \frac{\prod_{i=1}^{n} d_i!}{\prod_{i=1}^{n} 2^{x_{ii}} \prod_{i \leq j} x_{ij}!},$$

where we note that $d_i = x_{ii} + \sum_j x_{ij}$.

Intuition. If we condition on the fact that the graph $(x_{ij})_{ij}$ needs to be simple, it's clear that

$$\Pr\left(\mathrm{CM}((d_i)_{i \in [n]}) \text{ is simple}\right) = \frac{\prod_{i=1}^{n} d_i!}{(\ell_n - 1)!!},$$

which is truly uniform.

Next, we want to see what's the probability exactly. First, $\mathbb{E}[S_n] = \sum_{i=1}^{n} \mathbb{E}[X_{ii}]$, where

$$X_{ii} = \sum_{1 \leq a < b \leq d_i} \mathbb{1}_{(i,a) \text{ matched to } (i,b)}.$$

We see that $\mathbb{E}[X_{ii}] = \frac{d_i(d_i - 1)}{2(\ell_n - 1)}$, and hence

$$\mathbb{E}[S_n] = \frac{1}{2(\ell_n - 1)} \sum_{i=1}^{n} d_i(d_i - 1).$$

If we have second moment convergence as well, then

$$\mathbb{E}[S_n] \approx \frac{\mathbb{E}[D(D-1)]}{2\mathbb{E}[D]}.$$

Next, consider $M_n = \sum_{i<j} (x_{ij} - 1) \mathbb{1}_{x_{ij} \geq 2}$.

Note. $(x_{ij} - 1) \mathbb{1}_{x_{ij} \geq 2} \leq (x_{ij} - 1) \frac{x_{ij}}{2}$ with the equality holds if $x_{ij} \in \{0, 1, 2\}$.

Hence,

$$M_n \leq \frac{1}{2} \sum_{i<j} x_{ij}(x_{ij} - 1)$$

with

$$\mathbb{E}[M_n] = \frac{1}{2} \sum_{i<j} \sum_{1 \leq a < b \leq d_i} \sum_{1 \leq c \neq f \leq d_j} \frac{1}{(\ell_n - 1)(\ell_n - 3)} = \frac{1}{4} \sum_{i \neq j} \frac{d_j(d_j - 1) d_i(d_i - 1)}{(\ell_n - 1)(\ell_n - 3)} \approx (\mathbb{E}[S_n])^2.$$

Lemma 3.2.1. If $\mathbb{E}[D^2] < \infty$ and all the assumptions hold., then with $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]$, as $n \to \infty$,

$$(S_n, M_n) \xrightarrow{D} \mathrm{Pois}(\nu/2) \otimes \mathrm{Pois}(\nu^2/4).$$

In particular, $\Pr\left(\mathrm{CM}((d_i)_{i\in[n]}) \text{ is simple}\right) \to e^{-\nu/2-\nu^2/4}$ as $n \to \infty$.

Hence, we see that

$$\frac{\prod_{i=1}^n d_i}{(\ell_n-1)!!} \cdot |\{\text{simple graphs with degree sequence } (d_1, \dots, d_n)\}| = \Pr\left(\mathrm{CM}((d_i)_{i\in[n]}) \text{ is simple}\right) \approx e^{-\nu/2-\nu^2/4},$$

which implies the number of simple graphs with degree sequence $(d_i)_{i\in[n]}$ is

$$e^{-\nu/2-\nu^2/4} \cdot \frac{(\ell_n-1)!!}{\prod_{i=1}^n d_i}(1+o(1)).$$

Note. If $\Pr(B) \to 1$ (or equivalently, $\Pr(B^c) \to 0$), then

$$\Pr(B \mid A) = 1 - \Pr(B^c \mid A) = 1 - \frac{\Pr(B^c \cap A)}{\Pr(A)} = 1 - O(\Pr(B^c)) \to 1.$$

Hence, for A being hte even of simple graphs, it's okay to prove what we have shown.

3.2.1 Erased Configuration Model

Remove self-loops and keep only 1 of multiple edges. That is, $x_{ii} \mapsto 0 =: \widetilde{x}_{ii}$, and $x_{ij} \mapsto x_{ij} \wedge 1 =: \widetilde{x}_{ij}$. Then,

$$d_i = x_{ii} + \sum_{j=1}^n x_{ij} \mapsto \widetilde{d}_i = \sum_{j=1}^n x_{ij} \wedge 1$$

Problem. Without the second moment convergence assumption, can we show

$$\sum_{k=0}^\infty \left| \frac{1}{n} |\{i \mid \widetilde{d}_i = k\}| - \Pr(D = k) \right| \to 0?$$

Lemma 3.2.2. The above is actually true!

Proof. $0 \le d_i - \widetilde{d}_i \le 2x_{ii} + \sum_{j\neq i}(x_{ij}-1)\mathbb{1}_{x_{ij}\ge 2}$, and hence

$$\sum_{i=1}^n |d_i - \widetilde{d}_i| \le \sum_{i=1}^n 2x_{ii} + \sum_{i=1}^n \sum_{j\neq i}(x_{ij}-1)\mathbb{1}_{x_{ij}\ge 2} = 2(S_n + M_n).$$

One can also check that

$$\sum_{k=0}^\infty \left| \frac{1}{n} |\{i \mid \widetilde{d}_i = k\}| - \Pr(D = k) \right| \le \frac{1}{n}\sum_{i=1}^n (d_i - \widetilde{d}_i) \le \frac{2}{n}(S_n + M_n).$$

We have

$$\frac{1}{n}\mathbb{E}[S_n] = \frac{\sum_{i=1}^n d_i(d_i-1)}{2n(\ell_n-1)} \le \frac{d_{\max}\ell_n}{n\ell_n} \le \frac{d_{\max}}{n} \to 0$$

under the assumptions.

Exercise. If $X_i, \dots, X_n \overset{\text{i.i.d.}}{\sim} X$ and $\mathbb{E}[X] < \infty$, then $\max|X|/n \xrightarrow{p} 0$ as $n \to \infty$.

One can also control $\mathbb{E}[M_n]/n$ by a careful cut-off argument and show that it also converges to 0 as $n \to \infty$, hence $\mathbb{E}[S_n + M_n]/n \to 0$. ■

Lecture 12

27 Feb. 9:30

As previously seen. (a) $D_n \sim \frac{1}{n}\sum_{i=1}^n \delta_{d_i} \xrightarrow{D} \alpha(D)$

(b) $\frac{1}{n}\sum_{i=1}^n d_i = \ell_n/n = \mathbb{E}[D_n] \to \mathbb{E}[D]$.

(c) $\frac{1}{n}\sum_{i=1}^n d_i^2 = \mathbb{E}[D_n^2] \to \mathbb{E}[D^2]$.

Under all assumptions,

$$\Pr\left(\mathrm{CM}(d) \text{ is simple}\right) = \Pr\left(\text{no self-loops \& no multiple edges}\right) \to \exp\left(-\nu/2 - \nu^2/4\right)$$

where $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D]$.

Intuition. Say $D \in \{0, 1, 2, \dots\}$ is a random variable. The *size-biased* transform $\hat{D}$ of D is defined as

$$\Pr\left(\hat{D} = k\right) = \frac{k \cdot \Pr\left(D = k\right)}{\mathbb{E}[D]}, \quad k = 0, 1, \dots.$$

We see that

$$\mathbb{E}[\hat{D} - 1] = \sum_{k=0}^{\infty} (k-1) \cdot k \cdot \frac{\Pr\left(D = k\right)}{\mathbb{E}[D]} = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}.$$

Remark. If $D \sim \mathrm{Pois}(\lambda)$, then $\hat{D} - 1 \sim \mathrm{Pois}(\lambda)$.

Proof. We see that for all $k \geq 0$, we have

$$\Pr\left(\hat{D} - 1 = k\right) = \Pr\left(\hat{D} = k + 1\right) = \frac{(k+1)}{\lambda} \frac{e^{-\lambda}\lambda^{k+1}}{(k+1)!} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

⊛

Exercise. If $D \in \{0, 1, \dots\}$ and $D \overset{D}{=} \hat{D} - 1$, then D must be $\mathrm{Pois}(\lambda)$ for some $\lambda > 0$.

Let $D_1, \dots, D_n$ all i.i.d. from some distribution be the degree sequence.

Exercise. We see that:

- Choose a random node and observe its degree. We will see the D_i distribution.
- Choose a half-edge uniformly at random and observe the degree of the corresponding node. We will see $\hat{D}$ distribution instead.

Answer. The number of degree k nodes is roughly $n \cdot \Pr\left(D = k\right)$. Hence, $\Pr\left(\text{observe degree } k\right) = \frac{k \cdot n \cdot \Pr(D=k)}{\sum_{i=1}^n D_i} \approx \frac{k \cdot \Pr(D=k)}{\mathbb{E}[D]}$. ⊛

Remark (Branching process). Say we choose one of the random node. Then it will have D children. Now, choose one of its random child, its new degree will be $\hat{D} - 1$. Hence, we get $\mathrm{BP}(D)$ at the root and $\mathrm{BP}(\hat{D} - 1)$ for everything else.

One can check that this exploration tree will extinct with probability 1 if and only if $\mathbb{E}[\hat{D} - 1] \leq 1$.

Theorem 3.2.2. For the configuration model under all assumptions, we have

(a) $\frac{1}{n}|\mathcal{C}_{\max}| \to 0$ almost surely if $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \leq 1$.

(b) If $\nu > 1$, then $\frac{1}{n}|\mathcal{C}_{\max}| \to \Pr$ (root survive in the 2-step branching process).

Exercise. If $\mathbb{E}[\hat{D}^2] < \infty$ or $\mathbb{E}[D^3] < \infty$, then the universality holds at $\nu = 1$ (all clusters have size $n^{2/3}$).

Example. We can also count the number of triangles. Let

$$N_{\triangle} = \sum_{1 \leq i < j < k \leq n} \mathbb{1}_{(i,j),(j,k),(k,i) \in E},$$

with

$$\mathbb{E}[N_{\triangle}] \approx \frac{n^3}{6} \cdot \mathbb{E}\left[\frac{D_1 D_2}{n\mathbb{E}[D]} \cdot \frac{(D_1-1)D_3}{n\mathbb{E}[D]} \cdot \frac{(D_2-1)(D_3-1)}{n\mathbb{E}[D]}\right] \approx \frac{1}{6}\left(\frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}\right)^3 = \frac{\nu^3}{6}.$$

Exercise. $N_{\triangle} \xrightarrow{D} \operatorname{Pois}(\nu^3/6)$ as $n \to \infty$.

Under (a) + (b), erase the self-loops and keep only one edge (out of many multiple edges).

Lemma 3.2.3. The empirical degree distribution $(\hat{d}_i)_{i=1}^n \xrightarrow{D} D$.

Fix D, and take $d_1, \ldots, d_n$. Let $\Pr(D \geq k) = k^{-(\tau-1)}L(k)$ as $k \to \infty$ where L is a slowly varying function such that $L(ak)/L(k) \to 1$ as $k \to \infty$ for all $a > 0$. Then,

$$\Pr(D = k) \approx \frac{1}{k^\tau}$$

as $k \to \infty$.

- If $\tau \in (2, \infty)$, then $\mathbb{E}[D] = \sum_{i=1}^\infty k 1/k^\tau < \infty$.
- If $\tau \in (3, \infty)$, then $\mathbb{E}[D^2] < \infty$.
- If $\tau \in (2, 3)$, then $\mathbb{E}[D] < \infty$ but $\mathbb{E}[D^2]$ diverges to ∞.

Let $d_{\max}^{(n)} := \max_{i \in [n]} d_i \approx \overline{F}^{-1}(1/n)$ if $\overline{F}(x) = \Pr(D \geq x)$.
A heuristic argument is that

$$\Pr\left(d_{\max}^{(n)} \leq \overline{F}^{-1}(1/n)\right) = \Pr\left(d_1 \leq \overline{F}^{-1}(1/n)\right)^n = \left(1 - \overline{F}(\overline{F}^{-1}(1/n))\right)^n = (1-1/n)^n \approx e^{-1}.$$

When $\tau \in (2,3)$, $d_{\max}^{(n)} = n^{\frac{1}{\tau-1}} \gg \sqrt{n}$.

Note. $|\{i \mid d_i \geq \sqrt{n}\}| \sim \operatorname{Bin}(n, \overline{F}(\sqrt{n})) = \Theta(n^{1-\frac{\tau-1}{2}}) = o(n)$ where $\overline{f}(\sqrt{n}) = 1/n^{\frac{\tau-1}{2}}$.

Intuition. Since we're assuming $\mathbb{E}[D] < \infty$, the behavior is not dominated by $d_{\max}^{(n)}$. Hence, the above argument works, and the behavior is similar to $\operatorname{ER}(n,p)$.

However, what if $\mathbb{E}[D]$ also diverges? That is, $\mathbb{E}[D] = \infty$ with $\tau \in (1,2)$, where $\Pr(D = k) \approx 1/k^\tau$ as $k \to \infty$.

- $\ell_n = \sum_{i=1}^n (d_i \wedge n)$, and hence $d_{\max}^{(n)} \approx n^{\frac{1}{\tau-1}} \geq n$. With some heuristic argument, since $\ell_n = \sum_{i=1}^n d_i \approx n^{\frac{1}{\tau-1}}$, from extremal value theory,

$$\frac{1}{n^{\frac{1}{\tau-1}}} \cdot d_{(1)} \geq d_{(1)} \geq d_{(2)} \geq \ldots \xrightarrow{D} (\xi_1 > \xi_2 > \ldots) \sim \operatorname{PPP}()$$

  and $\frac{\ell_n}{n^{\frac{1}{\tau-1}}} \xrightarrow{D} \sum_{i=1}^\infty \xi_i$.

- Hub dominate (no bulk influence).
- $$\left(\frac{d_{(1)}}{\ell_n}, \dots, \frac{d_{(n)}}{\ell_n}\right) \overset{D}{\to} \left(\frac{\xi_1}{\sum_{i=1}^n \xi_i}, \dots, \frac{\xi_n}{\sum_{i=1}^n \xi_i}\right)$$

Theorem 3.2.3. The degree of a typical node has a limit given by the number of non-zero values in Multinomial$(D, (\xi_1/\sum_{i=1}^n \xi_i, \xi_n/\sum_{i=1}^n \xi_i))$.

3.2.2 Preferential Attachment Model

At each time t, for $v = v_1, \dots, v_{t-1}$, $\Pr_t(v_t \to v) = \deg_{t-1}(v)/\sum_{i=1}^{t-1}\deg_{t-1}(v_i) = \deg_{t-1}(v)/2(t-1)$ as $t \to \infty$.

Note. We can also consider $\Pr_t(v_t \to v) = f(\deg_{t-1}(v))/\sum_{i=1}^{t-1} f(\deg_{t-1}(v_i))$.

Theorem 3.2.4. As $n \to \infty$, with $q_k \approx 1/k^3$ as $k \to \infty$,

$$\left(\frac{1}{t}\lvert\{v \mid \deg_t(v) = k\}\rvert\right)_{k\geq 0} \overset{D \text{ in } p}{\to} (q_k)_{k\geq 0}.$$

Lecture 13: Preferential Attachment Model

4 Mar. 9:30

Lecture 14: Degree Distribution in Preferential Attachment Model

6 Mar. 9:30

As previously seen. PA$(m = 1, \delta)$ model. At time 1, there's a self loop at vertex 1, and for $i \in [t]$,

$$\Pr(t+1 \to i \mid \mathcal{F}_t) = \frac{D_i(t) + \delta}{(2+\delta)t}.$$

We have shown that

- $\frac{1}{t^{1/2+\delta}}(D_1(t), D_2(t), \dots, D_k(t)) \overset{\text{a.s.}}{\to} (\hat{\xi}_1, \dots, \hat{\xi}_k)$ for all k as $t \to \infty$.
- $\Pr\left(\max_{k\geq 1}\lvert N_k(t) - \mathbb{E}[N_k(t)]\rvert \geq c\sqrt{t\log t}\right) \leq t^{1-c^2/8}$.
- $P_k(t) = N_k(t)/t \approx \mathbb{E}[P_k(t)] + O_p(\sqrt{\log t/t})$ where the last term is uniform in k.

Theorem 3.2.5. For all $k \geq 1$, $\mathbb{E}[P_k(t)] \to P_k$ as $t \to \infty$ where $P_k = \cdot P_{k-1}$ for $k \geq 2$ and $P_1 =$, and $(P_1, P_2, \dots)$ is a pmf.

3.2.3 Heuristic Approach

Let

$$P_k(t) = \frac{1}{t}\sum_{i=1}^{t} \mathbb{1}_{D_i(t)=k},$$

hence $\mathbb{E}[P_k(t)]$ is the probability that a uniform at random chosen node at time t has degree k. Fix some t, choose a uniform at random node $\approx t_0 = \lceil t \cdot u\rceil$ for $u \sim \mathcal{U}(0,1)$. Let t_i be the node that is i^{th} connected to t_0, and let T_i's be the corresponding random variables. We see that for $x > 1$,

$$\Pr\left(\frac{T}{t_0} \geq x\right) = \Pr(T_1 \geq x \cdot t_0) = \prod_{t=t_0+1}^{\lceil x\cdot t_0\rceil}\left(1 - \frac{1+\delta}{(2+\delta)t}\right) \approx \exp\left(-\frac{1+\delta}{2+\delta}\sum_{t=t_0+1}^{\lceil x\cdot t_0\rceil}\frac{1}{t}\right) \approx \exp\left(-\frac{1+\delta}{2+\delta}\log x\right).$$

Equivalently, we may write $\Pr\left(T_1/t_0 \geq e^y\right) \approx e^{-\frac{1+\delta}{2+\delta}y}$ for $y > 0$ for large t_0. Hence,

$$\log T_1 - \log t_0 \approx \exp\left(\frac{1+\delta}{2+\delta}\right).$$

Lemma 3.2.4. Consider a node at time $t_0 =: T_{-1\to 0}$. Let $T_{k-1\to k}$ be the time when the degree of the node changes from $k-1$ to k. Then, as $t_0 \to \infty$,

$$(\log T_{k-1\to k} - \log T_{k-2\to k-1})_{k\geq 2} \xrightarrow{D} \left(\operatorname{Exp}\left(\frac{k-1+\delta}{2+\delta}\right)\right)_{k\geq 2}.$$

Denote $X_k := \operatorname{Exp}(\frac{k+\delta}{2+\delta})$, then starting from t_0, at time $t_0 \cdot \exp(X_1)$, a new vertex is connected, and this goes on with $t_0 \exp(X_1 + X_2 + \ldots)$.

Remark. X_i's are independent as long as $t_0 \to \infty$ as each X_i converges to the same limit regardless of the starting time.

We see that as $t \approx t_0/u \approx t_0 e^{-\log u} = t_0 e^X$ for $u \sim \mathcal{U}(0,1)$ and $X \sim \operatorname{Exp}(1)$. Hence,

$$\begin{aligned} P_k &\approx \Pr\left(\text{a uniform at random chosen node has degree } k \text{ at time } t\right) \\ &\approx \Pr\left(X_1 + X_2 + \cdots + X_{k-1} \leq X < X_1 + \cdots + X_k\right). \end{aligned}$$

Let $P_{\geq k} := \Pr\left(X \geq X_1 + \cdots + X_{k-1}\right)$, then

$$P_{\geq k} = \mathbb{E}[e^{-(X_1+\cdots+X_{k-1})}] = \prod_{i=1}^{k-1}\mathbb{E}[e^{-X_i}] = \prod_{i=1}^{k-1}\frac{\frac{i+\delta}{2+\delta}}{1+\frac{i+\delta}{2+\delta}}. = \prod_{i=1}^{k-1}\frac{i+\delta}{2+i+2\delta}.$$

Example. $P_1 = \Pr\left(X < X_1\right) = \frac{1}{1+\frac{1+\delta}{2+\delta}} = \frac{2+\delta}{3+2\delta}$.

3.2.4 Martingale Approach

Consider $(N_1(t))_{t\geq 0}$ with $N_1(1) = 0$, and $N_1(2) = 1$. Then,

$$\mathbb{E}[N_1(t+1) - N_1(t) \mid \mathcal{F}_t] = 1 - \frac{1+\delta}{(2+\delta)t}N_1(t) \Rightarrow \mathbb{E}_t[N_1(t+1)] = 1 + \left(1 - \frac{1+\delta}{2+\delta}\frac{1}{t}\right)\cdot N_1(t)$$

for all $t \geq 0$. Hence,

$$\begin{aligned} \mathbb{E}[N_1(t+1)] &= 1 + \left(1 - \frac{1+\delta}{2+\delta}\frac{1}{t}\right) + \left(1 - \frac{1+\delta}{2+\delta}\frac{1}{t}\right)\left(1 - \frac{1+\delta}{2+\delta}\frac{1}{t-1}\right) + \ldots \\ &= \sum_{i=0}^{t}\prod_{j=t+1}^{t}\left(1 - \frac{1+\delta}{2+\delta}\frac{1}{j}\right). \end{aligned}$$

We then get

$$\mathbb{E}[P_1(t+1)] = \frac{1}{t+1}\sum_{i=0}^{t}\prod_{j=t+1}^{t}\left(1 - \frac{1+\delta}{2+\delta}\frac{1}{j}\right) \approx \frac{1}{t+1}\sum_{i=1}^{t}\left(\frac{i}{t}\right)^{\frac{1+\delta}{2+\delta}} \to \int_0^1 x^{\frac{1+\delta}{2+\delta}}\,\mathrm{d}x = \frac{1}{1+\frac{1+\delta}{2+\delta}} = \frac{2+\delta}{3+2\delta}$$

as $t \to \infty$. Fix $k \geq 2$, and consider induction on k, we have

$$\mathbb{E}_t[N_k(t+1) - N_k(t)] = 1\frac{k-1+\delta}{(2+\delta)t}N_{k-1}(t) - \frac{k+\delta}{(2+\delta)t}N_k(t),$$

hence

$$\begin{aligned}
&\mathbb{E}[N_k(t+1)] \\
&= \left(1 - \frac{k+\delta}{(2+\delta)t}\right)\mathbb{E}[N_k(t)] + \frac{k-1+\delta}{2+\delta}\mathbb{E}\left[\frac{N_{k-1}(t)}{t}\right] \\
&= \frac{k-1+\delta}{2+\delta}\left[\mathbb{E}[P_{k-1}(t)] + \left(1 - \frac{k+\delta}{(2+\delta)t}\mathbb{E}[P_{k-1}(t-1)] + \cdots + \left(1-\frac{1}{t}\right)\cdots\left(1-\frac{1}{i}\right)\mathbb{E}[P_{k-1}(i+1)]\right)\right] \\
&= \frac{k-1+\delta}{2+\delta}\sum_{i=0}^{t}\prod_{j=i+1}^{t}\left(1 - \frac{k+\delta}{2+\delta}\frac{1}{j}\mathbb{E}[P_{k-1}(i)]\right),
\end{aligned}$$

which gives

$$\mathbb{E}[P_k(t+1)] = \frac{k-1+\delta}{2+\delta}\frac{1}{t+1}\sum_{i=0}^{t}\left(\frac{i}{t}\right)^{\frac{k+\delta}{2+\delta}}\mathbb{E}[P_{k-1}(i)] \to \frac{k-1+\delta}{2e\delta}\int_0^1 x^{\frac{k+\delta}{2+\delta}}\,\mathrm{d}x P_{k-1} = \frac{k-1+\delta}{k+2+2\delta}P_{k-1}$$

as $t \to \infty$.

Remark. $P_k = \frac{k-1+\delta}{k+2+2\delta}P_{k-1}$ and $P_1 = \frac{2+\delta}{2+2\delta}$.

Lemma 3.2.5. $\log P_{\geq k}/\log k \to -(2+\delta)$ as $k \to \infty$. In particular, $(P_k)_{k\geq 1}$ has $(2+\delta-\epsilon)^{\text{th}}$ moment finite, but $(2+\delta)^{\text{th}}$ moment infinity.

Proof. $P_{\geq k} = \prod_{i=1}^{k-1}\frac{i+\delta}{i+2+2\delta} = \prod_{i=1}^{k-1}\left(1+\frac{2+\delta}{i+\delta}\right)^{-1} \approx \exp\left(-(2+\delta)\sum_{i=1}^{k-1}\frac{1}{i+\delta}\right) \approx k^{-(2+\delta)}$. ■

3.2.5 General

the general $\mathrm{PA}(m,\delta)$ can be connected out of $\mathrm{PA}(1,\delta/m)$ model:

$$\underbrace{1,2,\ldots,m}_{[1]},\underbrace{m+1,\ldots,2m}_{[2]},\underbrace{2m+1,\ldots,3m}_{[3]},\ldots$$

where we contract every block of size m into one super node.

Also, for $\mathrm{PA}(1,\delta)$, as $\delta \to \infty$, we have the so-called uniform attachment tree, where $\Pr_t(t+1 \to i) = 1/t$. This is also called a random recursive tree.

Lecture 15: Mean Field First Passage Percolation

Given a connected finite graph $G = (V, E)$ with $\omega_e \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(1)$ for $e \in E$. 11 Mar. 9:30

As previously seen. For $\omega \sim \mathrm{Exp}(\lambda)$, then $(\omega \mid \omega > x) \overset{D}{=} x + \mathrm{Exp}(\lambda)$.

Let $X_{ij} \coloneqq \min_{\Pi\colon i\to j}\sum_{e\in\Pi}\omega_e$ be the "minimum" distance between i and j in the randomly weighted graph. Also, let $X_{i*} = \max_j X_{ij}$ and $X_{**} = \max_{i,j} X_{ij}$.

Theorem 3.2.6. Focus on $\mathrm{ER}(n, p_n)$ with $np_n \gg \log^A n$ for some large A, then

- $np_n X_{ij}/\log n \overset{p}{\to} 1$ and $np_n \cdot X_{ij} - \log n \overset{D}{\to} ?$.
- $np_n X_{i*}/\log n \overset{p}{\to} 2$, and $np_n X_{i*} - 2\log n \overset{D}{\to} ?$.
- $np_n X_{**}/\log n \overset{p}{\to} 3$, and $np_n X_{**} - 3\log n \overset{D}{\to} ?$.

Proof. Consider X_{1*} in $K_n = \mathrm{ER}(n,1)$. We see that since 1 has $n-1$ many neighbors, $\min_j \omega_{1j} \sim \frac{1}{n-1}\mathrm{Exp}(1)$. Consider the exploration process from 1, and let $\hat{X}_k$ be the size of the exploration tree at time k. We see that at time 0, $\hat{X}_0 = 1$. Assuming time t flows in a uniform rate, at i, the vertex j is revealed when the time further exceeds ω_{ij}. It's easy to see that the waiting time for $\hat{X}_1 = 2$ is $\mathrm{Exp}(n-1)$, and for $\hat{X}_2 = 3$ is $\mathrm{Exp}(2(n-2))$. In general, the waiting time for $\hat{X}_{k-1} = k$ to $\hat{X}_k = k+1$ is $\mathrm{Exp}(k(n-k))$.

Let X_{1*} be the sum of independent $\mathrm{Exp}(k(n-k))$ random variables for $k = 1, 2, \dots, n-1$. We see that with $\xi_k \overset{\text{i.i.d.}}{\sim} \mathrm{Exp}(1)$

$$nX_{1*} \overset{D}{=} n\sum_{k=1}^{n-1} \frac{1}{k(n-k)} \cdot \xi_k = n\sum_{k=1}^{n-1} \frac{1}{k(n-k)}(\xi_k - 1) + \sum_{k=1}^{n-1} \underbrace{\frac{n-k+k}{k(n-k)}}_{\frac{1}{k}+\frac{1}{n-k}},$$

and hence $n\mathbb{E}[X_{1*}] = 2\sum_{k=1}^{n-1} 1/k \sim 2\log n$. Moreover,

$$\mathrm{Var}[nX_{1*}] = \sum_{k=1}^{n-1} \frac{n^2}{k^2(n-k)^2} = 2\sum_{k=1}^{\lfloor n/2 \rfloor} \frac{n^2}{k^2(n-k)^2} + o(1) \leq 8\sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{k^2} \leq \pi^2.$$

Hence, from Chebyshev's inequality, $nX_{1*}/\log n \xrightarrow{p} 2$ as $n \to \infty$. Moreover,

$$\begin{aligned}
&nX_{1*} - 2\log n \\
&= \sum_{k=1}^{n-1} \frac{n-k+k}{n(n-k)}(\xi_k - 1) + 2H_{n-1} - 2\log n \\
&= \sum_{k=1}^{n-1} \frac{1}{k}(\xi_k - 1) + \sum_{k=1}^{n-1} \frac{1}{n-k}(\xi_k - 1) + 2(H_{n-1-\log n}) \\
&= \sum_{k=1}^{n-1} \frac{1}{k}(\xi_k - 1) + \sum_{k=1}^{n-1} \frac{1}{k}(\xi_{n-k} - 1) + 2(H_{n-1-\log n}) \\
&= \sum_{k=1}^{n/2} \frac{1}{k}(\xi_k - 1) + \underbrace{\sum_{k=n/2+1}^{n-1} \frac{1}{k}(\xi_k - 1)}_{\to 0} + \sum_{k=1}^{n/2} \frac{1}{k}(\xi_{n-k} - 1) + \underbrace{\sum_{k=n/2+1}^{n-1} \frac{1}{k}(\xi_{n-k} - 1)}_{\to 0} + 2(H_{n-1-\log n}).
\end{aligned}$$

Observe that the remaining terms are independent since they are the first and the second half of ξ_i's.

Claim. $\sum_{k=1}^{n} \xi_k/k \overset{D}{=} \max_{i\in[n]} \xi_i$.

Proof. Starting from 0, the first gap is given by $\mathrm{Exp}(1)/n$, and the second will be $\mathrm{Exp}(1)/(n-1)$, and so on. Hence, by summing over all gaps starting from 0, it's exactly $\sum_{k=1}^{n} \xi_k/k$, which is the same as the maximum over all ξ_i's. ⊛

Hence,

$$\Pr\left(\max_{i\in[n]} \xi_i - \log n \leq x\right) = \Big(\Pr\left(\xi \leq \log n + x\right)\Big)^n = \left(1 - \frac{1}{n}e^{-x}\right)^n \to e^{-e^{-x}}.$$

We conclude that $nX_{1*} - 2\log n \xrightarrow{D}$ sum of two i.i.d. Gumbel distributions.

Exercise. $\Pr\left(nX_{1*} \geq (2+\epsilon)\log n \leq A_\epsilon \log^2 n / n^\epsilon\right)$ for all $\epsilon > 0$ and some constant A_ϵ. As a corollary,

$$\Pr\left(X_{**} = \max_{i\in[n]} nX_{i*} \geq (2+1+\epsilon)\log n\right) \leq n\frac{A_\epsilon \log^2 n}{n^{1+\epsilon}} \to 0$$

as $n \to \infty$, hence $nX_{**}/\log n \leq 3 + \epsilon$ with high probability.

Now, for X_{12}, we have

$$X_{12} \overset{D}{\to} \sum_{k=1}^{U_n} \frac{1}{k(n-k)} \xi_k = \sum_{k=1}^{n-1} \frac{1}{k(n-k)} \xi_k \mathbb{1}_{k \leq U_n}$$

for $U_n \sim \mathcal{U}([n-1])$. Hence,

$$\mathbb{E}[nX_{12}] = \sum_{k=1}^{n-1} \frac{n}{k(n-k)} \cdot 1 \cdot \frac{n-k}{n-1} = \frac{n}{n-1} H_{n-1} \approx \log n.$$

To compute the variance, we can first center X_{12} by

$$X_{12} = \sum_{k=1}^{n-1} \frac{1}{k(n-k)} (\xi_k - 1) \mathbb{1}_{k \leq U_n} + \sum_{k=1}^{n-1} \frac{1}{k(n-k)} \mathbb{1}_{k \leq U_n},$$

and we have $\mathrm{Var}[nX_{12}]$ is bounded by some constant.

Note. The first term will disappear since ξ_k is independent of U_n, and since $\xi_k - 1$ has mean 0, any cross-products will disappear.

One can then check that

$$nX_{12} - \log n \cong \underbrace{\sum_{k=1}^{n-1} \frac{1}{k} \xi_k}_{\max_{i \in [n-1]} \xi_i - \log n} + \underbrace{(H_{U_n - 1} - H_{n - U_n - 1})}_{\log \frac{U_n - 1}{n - U_n - 1} \to \log \frac{U}{1-U}, U \sim \mathcal{U}(0,1)} \overset{D}{\to} \mathrm{Gumbel}_1 + \log \frac{U}{1-U}$$

Intuition.

■

Remark. We can also consider $\mathrm{ER}(n,p)$, then the number of edges of the exploration process is $\mathrm{Bin}(k(n-k), p)$, all correlated between each time step k. Hence,

$$npX_{1*} = \sum_{k=1}^{n-1} \underbrace{\frac{pk(n-k)}{\mathrm{Bin}(k(n-k), p)}}_{\frac{p}{p + \frac{\mathcal{N}(0,1)}{\sqrt{k(n-k)p \ldots}}}} \cdot \xi_k \cdot \frac{n}{k(n-k)},$$

which is again the sum of two Gumbels.

Problem (Open problem). Suppose $n_1, \ldots, n_d$ such that $n_i / n \approx \lambda_i$ for $i \in [d]$ with $\sum_{i=1}^d \lambda_i = 1$ and $P = (P_{ij})_{i,j \in [d]}$ be the edge probability matrix. Then, we don't know whether $n \cdot X_{IJ} / \log n \overset{p}{\to} C(\lambda, P)$ and $nX_{IJ} - C(\lambda, P) \log n \overset{D}{\to} ?$.

Lecture 16: Random Recursive Tree/Uniform Attachment Model

Definition 3.2.1 (Random recursive tree). The *random recursive tree* (RRT) is a model such that given the tree $\mathcal{T}_{n-1}$ at time $n-1$, choose $v_n \sim \mathcal{U}(V_{n-1})$ and connect a new node $n \to v_n$. 13 Mar. 9:30

Note. We see that for random recursive tree, we have $V_1 = [1], \ldots, V_n = [n]$.

Definition 3.2.2 (Uniform attachment model). The *uniform attachment model*, denoted as $\mathrm{UA}_n(m)$ is a model such that at time n, a new vertex n has $m \geq 2$ many out-going edges such that the end points are chosen from $\mathcal{U}([n-1])$ independently.

Equivalently, create random recursive tree up to time $m \cdot n$ such that nodes $\{m(k-1)+1, m(k-1)+2, \ldots, mk\}$ are merged to create a vertex $k' := (k)$ for $k = 1, \ldots, n$.

This is the limit of the preferential attachment model, $\mathrm{PA}(m, \delta)$, as $\delta \to \infty$ since

$$\lim_{\delta \to \infty} \frac{D_t(v) + \delta}{(2+\delta)t} = \frac{1}{t}.$$

As previously seen. Recall that $D_t(1) \approx t^{\frac{1}{2+\delta}}$.

Consider $\mathrm{RRT} = \mathrm{UA}(1)$, and we have

$$D_n(1) = \sum_{v=2}^{n} \mathbb{1}_{v \to 1},$$

where $\mathbb{1}_{v \to 1} \sim \mathrm{Ber}(1/(v-1))$. Hence,

$$\mathbb{E}[D_n(1)] = \sum_{v=2}^{n} \frac{1}{v-1} \approx \log n,$$

also, from the Stein-Chen method,

$$d_{\mathrm{TV}}\left(D_n(1), \mathrm{Pois}\left(\sum_{k=1}^{n-1} \frac{1}{k}\right)\right) \leq \min(1, 1/\lambda_n) \sum_{k=1}^{n-1} \left(\frac{1}{k}\right)^2 \leq \frac{c}{\log n} \to 0$$

Lemma 3.2.6. Given a fixed k, as $n \to \infty$, we have

$$d_{\mathrm{TV}}\left((D_n(1), \ldots, D_n(k)), \bigotimes_{i=1}^{k} \mathrm{Pois}(\log n)\right) \to 0.$$

In particular, as $n \to \infty$,

$$\left(\frac{D_n(i) - \log n}{\sqrt{\log n}}\right)_{i \in [n]} \overset{D}{\to} \mathcal{N}_k(0, I_k)$$

Remark. For any constant m, the same calculation goes through for $\mathrm{UA}_n(m)$. But since m is a constant, as $n \to \infty$, the above still holds.

Now, we want to understand the typical degree distribution for a uniformly chosen random node $I_n \sim \mathcal{U}([n])$ at time n. Consider $m = 1$ again, then

$$D_n(I_n) \overset{D}{=} \sum_{v=I_n+1}^{n} \mathbb{1}_{v \to I_n} + 1.$$

From the above lemma, let $u \sim \mathcal{U}(0,1)$, we have $I_n \overset{D}{\approx} \lceil nu \rceil$, hence

$$\sum_{v=I_n+1}^{n} \mathbb{1}_{v \to I_n} \mathbb{1}_{v \to I_n} \sim \mathrm{Pois}(\log n / I_n) \overset{D}{\approx} \mathrm{Pois}(\log 1/u).$$

Hence, for $k \geq 0$,

$$\Pr\left(D_n(I_n) = k+1\right) = \mathbb{E}_{y \sim \mathrm{Exp}(1)} \underbrace{[\Pr(\mathrm{Pois}(y) = k)]}_{e^{-y} \cdot y^k / k!} = \frac{1}{k!} \int_0^\infty e^{-y} y^k e^{-y} \, \mathrm{d}y = \frac{1}{2^k}.$$

Exercise. For $\mathrm{UA}_n(m)$ for a general $m > 1$, show that as $n \to \infty$,

$$(\#\text{self-loops}, \#\text{extra multi-edges}) \xrightarrow{D} \mathrm{Pois}(\lambda_1) \otimes \mathrm{Pois}(\lambda_2).$$

We can also view the model from an embedding viewpoint. Consider embed in a continuous time branching process (CTBP). We know that previously, our time goes like $0, 1, 2, \ldots$. However, since we now know that the corresponding RRT will be a Yule tree.

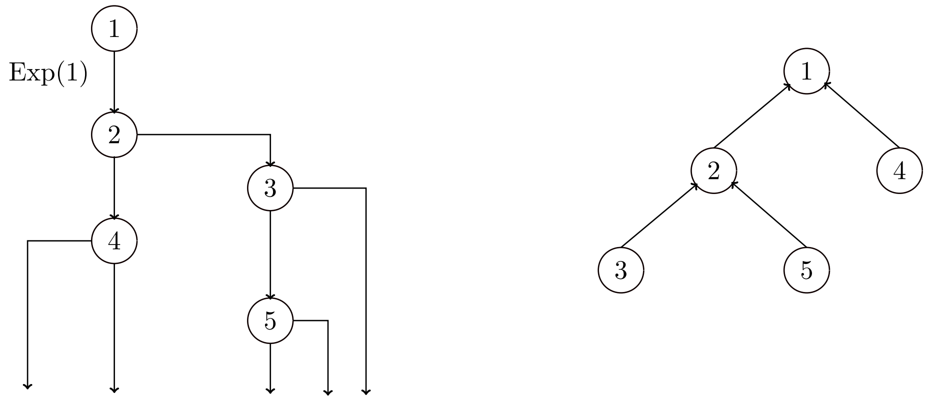


Figure 3.1: title

This is a RRT with $\mathrm{Exp}(k)$ intended waiting time for $(k+1)^{\text{th}}$ node. Let N_t be the number of nodes at time t, then $(e^{-t}N_t)_{t\geq 0}$ is a continuous time martingale.

Note. One can show that $(e^{-t}N_t)_{t\geq 0}$ is a positive martingale and $e^{-t}N_t \to Z_\infty \sim \mathrm{Exp}(1)$ almost surely (or in L_1) as $t \to \infty$.

Lemma 3.2.7. $\mathrm{YT}(\gamma_n) \overset{D}{=} \mathrm{RRT}_n$ where $\gamma_n \coloneqq \min\{t \geq 0 \mid N_t = n\}$.

For large t, heuristically, we have $N_t \approx e^t \cdot Z_\infty$ where $Z_\infty \sim \mathrm{Exp}(1)$. Hence, at time $t_n = \log n - \log Z_\infty$, $N_{t_n} \approx n$. As for a typical node which has an index $\lceil n \cdot u \rceil \approx nu$ with $u \sim \mathcal{U}(0,1)$, in this case, for

$$t_n' \approx \log \lceil nu \rceil - \log Z_\infty = \log n - \log Z_\infty - \log \frac{1}{u} =: \log n - \log Z_\infty - \eta$$

where $\eta \sim \mathrm{Exp}(1)$, we have $N_{t_n'} \approx \lceil nu \rceil$.

Intuition. The intuition is that the size of the tree grows exponentially fast, and if we want to find a uniformly random node, we simply need to look at the time just above t_n by η, i.e., most of the nodes are concentrated at leaves.

The picture of the tree viewed from a typical vertex (as $n \to \infty$) is exactly the same as the root vertex 1, i.e.,

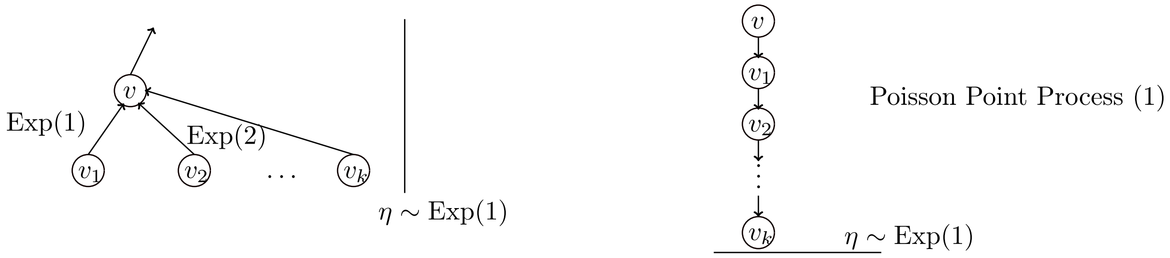


Figure 3.2: title

Since we want to look at time t_n, and we're starting at time $t_n - \eta$, we look at the cut-off at time η. We have

$$\Pr\left(\deg(v) \geq 1 + k\right) = \Pr\left(\sum_{i=1}^{k} \eta_i \leq \eta\right) = \mathbb{E}[e^{-\eta_1 - \eta_2 - \cdots - \eta_k}] = \left(\mathbb{E}[e^{ac;e_1}]\right)^k = \frac{1}{2^k},$$

where $\eta_i \stackrel{\text{i.i.d.}}{\sim} \mathrm{Exp}(1)$.

Theorem 3.2.7. The tree rooted at $v_n \sim \mathcal{U}([n])$ converges in distribution as the Yule tree stopped at $\eta \sim \mathrm{Exp}(1)$.

Figure 3.3

- Ancestor line, starting from a uniform at random node to the root
- Tree hanging below any node on the line (remove descendant line).

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.