MATH595
Stochastic Processes on Graphs

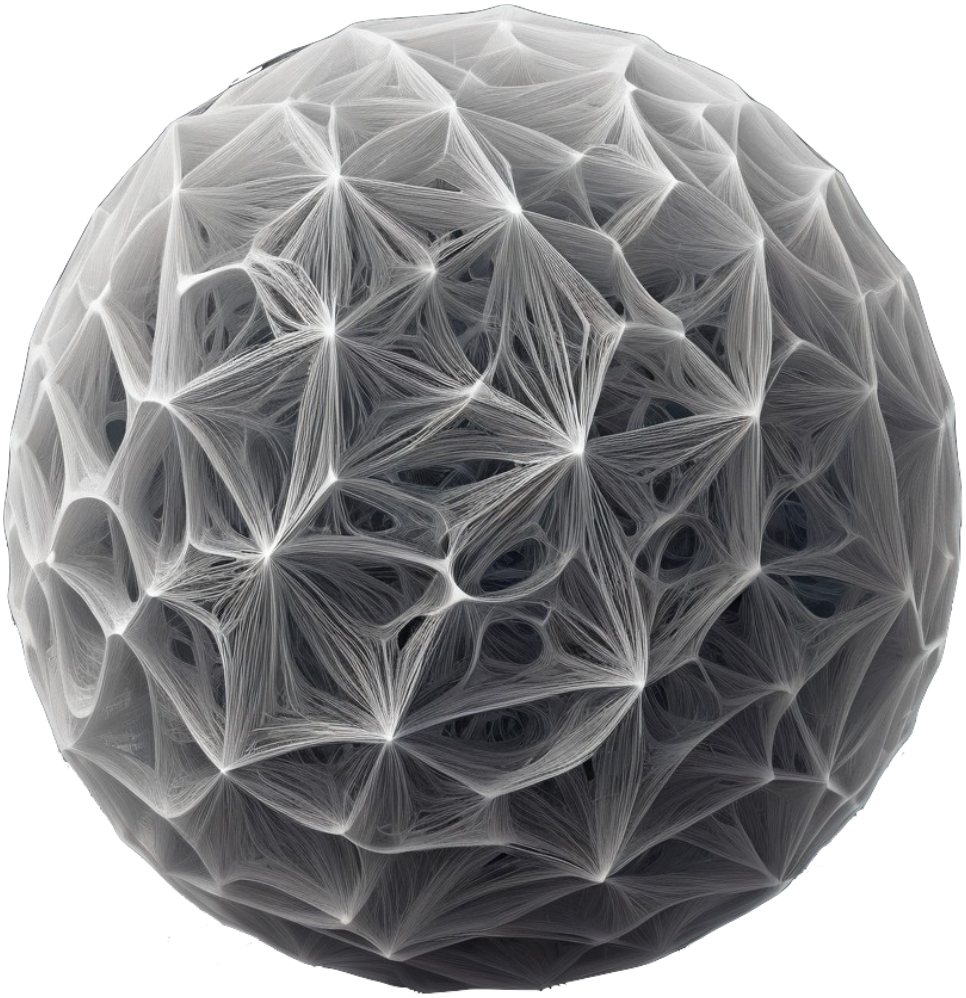
Pingbang Hu

January 23, 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

21 Jan. 9:30

1.1 Emergence of Graph Structure

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.

1.1.1 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* if there exists a path $x = v_1, \dots, v_k = y$ such that $\omega_{v_i v_{i+1}} = 1$ for all $1 \le i \le k - 1$.

Notation. We will write $x \to y$ if x is connected to y.

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

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

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

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

Connected component is an example of *structure*. Other common structures include triangle, cycle, clique, etc. We also define them formally for concreteness.

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 *k-cycle* $(v_1, \dots, v_k)$ 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 *k-clique* $K_k \subseteq V$ in a graph $G = (V, E)$ is such that for every $v_i, v_j \in K_k$, $(v_i, v_j) \in E$.

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

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:

Problem (Property). Given a graph G, whether it contains a certain structure; if yes, how many?

1.1.2 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 1.1.6 (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.

1.2 Erdős-Rényi Random Graph

Recall the Erdős-Rényi random graph model.

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

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

Claim. The number of edges converges in distribution to a standard normal, in particular,

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

if and only if $\binom{n}{2}p(1-p) \to \infty$. As a corollary, we have $|E|/\binom{n}{2}p \approx 1$.

Proof. We see that $|E| = \sum_{1 \le i < j \le n} \omega_{ij} \sim \mathrm{Bin}(\binom{n}{2}, p)$, hence, $\mathbb{E}[|E|] = \binom{n}{2}p = n(n-1)p/2$. Then,

the result follows directly from the central limit theorem. ⊛

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

Remark. If we choose p such that $m \approx \binom{n}{2}p$, the results often transfer between $G(n, p)$ and $G(n, m)$.

1.2.1 Density and Phase Transition

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

Definition 1.2.1 (Dense graph)**.** A graph $G = (V, E)$ is *dense* if there exists a constant $\epsilon > 0$ such that $|E|/\binom{n}{2} > \epsilon$.

Definition 1.2.2 (Sparse graph)**.** A graph $G = (V, E)$ is *sparse* if it is not dense.

Putting the motivation of defining Definition 1.2.1 and Definition 1.2.2 in this way aside, let's first observe an interesting property for the Erdős-Rényi random graph model. Firstly, note that the typical degree of the Erdős-Rényi random graph is some constant since for $G(n, p)$,

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

Note. Regime we consider hence depend on $\lambda \coloneqq np$ for some $\lambda \in (0, \infty)$.

Then, given a particular vertex, its degree follows $\mathrm{Bin}(n-1, p) = \mathrm{Bin}(n-1, \lambda/n)$.

Claim. $\mathrm{Bin}(n-1, \lambda/n) \xrightarrow{D} \mathrm{Pois}(\lambda)$ 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,
>
> $$d_{\mathrm{TV}}((p_k)_{k\in\Omega}, (r_k)_{k\in\Omega}) \coloneqq \frac{1}{2}\sum_{k\in\Omega}|p_k - r_k|.$$

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

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

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

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

Notation. Given a graph G, let $C_{\max_i}$ denotes the i^{th} largest connected component in G.

Theorem 1.2.1. Consider the Erdős-Rényi random graph model $G(n,\lambda/n)$ for some $\lambda > 0$.

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

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

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

Theorem 1.2.1 says that there is a phase transition right 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.

We first focus on the case when $\lambda < 1$. Consider the following heuristic argument of why there can't exist a large component.

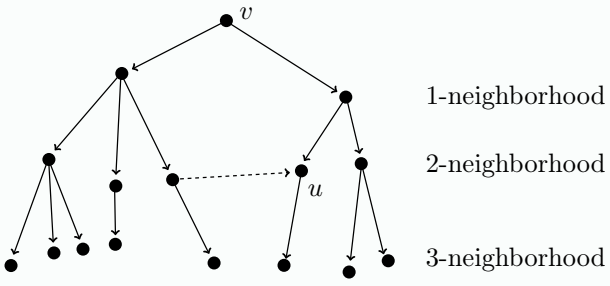
Intuition. The neighborhood structure of some vertex v, which should be tree-like, at least locally. This is because, for any $k \geq 2$, the expected number of cycles of length k in this structure is

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

which implies that when $\lambda < 1$,

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

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 (Types of questions). We mainly focus on the following three types of questions:

1. Typical behavior: single/multiple points view.
2. Global behavior: empirical behavior.
3. Extremal behavior: maxima or minima.

The degree distribution ($\text{Bin}(n-1,\lambda/n) \xrightarrow{D} \text{Pois}(\lambda)$) for a single vertex is a single point view.

Toward proving Theorem 1.2.1 (a), we will need the following idea:

Definition 1.2.3 (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. $\mathrm{Bin}(n,p) \preceq \mathrm{Bin}(m,p)$ 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 $1 - p \geq e^{-\theta}$.

As we will soon see, by using stochastic domination, one can provide a nice bound for proving Theorem 1.2.1 easily.

1.2.2 Small Digression to Degree Distribution

As a warm-up toward proving Theorem 1.2.1, let's first look at an easier problem: the degree. 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 $\deg_{G_n}(v) \overset{D}{\to} \mathrm{Pois}(\lambda)$ as $n \to \infty$ where $G_n \sim \mathrm{ER}(n, \lambda/n)$.

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

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

Proof. Consider any two vertices i, j, we see that $\deg(i) = \mathbb{1}_{(i,j)\in E} + \sum_{v \neq j} \mathbb{1}_{(i,v)\in E}$ 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. 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 $\deg_{\max,n} := \max_{v\in V} \deg(v)$ for $G \sim G(n,p)$.

Proposition 1.2.1. Consider the Erdős-Rényi random graph model $G(n, \lambda/n)$ for $\lambda < 1$. Then,

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

as $n \to \infty$ for all $\epsilon > 0$.

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

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

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

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

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

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

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 1.2.1 will be used extensively in this course.

1.2.3 Cluster Distribution

Getting back to Theorem 1.2.1, with a similar technique and argument, without loss of generality we consider $|C(1)|$. 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 is in some sense *dominated* by the tree 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 $C(1)$. We can now prove Theorem 1.2.1 (a), where we aim to show that $\Pr(|C_{\max_1}| \geq a\log n) \to 0$ as $n \to \infty$ if $a(\lambda - 1 - \log\lambda) > 1$.

Proof of Theorem 1.2.1 (a). For any x, $\Pr(|C(1)| \geq x) \leq \Pr(|\mathrm{GWBP}(\mathrm{Bin}(n-1,\lambda/n))| \geq x)$. The intuition is the following.

Intuition. If we maintain the number of vertices in the queue when we do the breadth-first search, we see that the tree can be (uniquely) embedded in a sequence. It's clear that the size of the tree is the length of this sequence (s_n) such that $s_0 = 1$, $s_1 = s_0 + (x_1 - 1)$, and so on.

The above sequence can be viewed as a random work, we see that $|T| = \inf\{n \geq 1 \mid s_n = 0\}$, where T is the tree corresponding to $\mathrm{GWBP}(\mathrm{Bin}(n-1,\lambda/n))$. At the end, we have

$$\Pr(|C(1)| \geq x) \leq e^{-\theta x}\mathbb{E}[e^{\theta \cdot H^{\{0\}}}],$$

where $H^{\{0\}}$ is the hitting at 0 for the random work $s_0 = 1$, $s_k = s_{k-1} + (\mathrm{Bin}(n-1,p) - 1)$. ■ Finish

It's tempting to say that since we're considering the sparse regime, as we have seen, very few cycles exist. Hence, the branching process should be pretty close to the actual breadth-first search tree. In fact, this is the case.

Theorem 1.2.2. For $G \sim \mathrm{ER}(n,\lambda/n)$ with $\lambda < 1$, $|C(1)| \xrightarrow{D} |T_\lambda|$ where $T_\lambda \sim \mathrm{GWBP}(\mathrm{Pois}(\lambda))$.

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.