CO749 - Random Graph Theory

(Lecture Summaries)

University of Waterloo
Nicholas Pun

Winter 2020

Contents

Lecture 1: Introduction

Definition 1.1. The probability space we'll work in is denoted with the triple $(G, \mathbb{P}, F)$, where G is a class of graphs, $\mathbb{P}$ a probability measure and F a sigma algebra.

Normally, G is a finite set, $\mathbb{P}$ is a discrete probability measure and $F = 2^G$

Definition 1.2 (Erdős-Rényi Random Graph Model)**.**

- The $\mathcal{G}(n,p)$ model: A graph with vertex set $[n]$ is constructed randomly by including each edge in $k_{[n]}$ with probability p
- The $\mathcal{G}(n,m)$ model: A graph is chosen uniformly at random from all graphs with vertex set $[n]$ and has m edges.

(Aside: We can think of $\mathcal{G}(n,m)$ as labelling the edges)

Other models:

- $\mathcal{G}(n,d)$ is the model of random d-regular graphs
- $\mathcal{G}(n,\tilde{d})$ where $\tilde{d} = (d_1, \ldots, d_n)$ is a vector representing the degrees of vertices. (This is a generalization of $G(n,d)$)
- $\mathcal{G}(n,r)$ is the model of random geometric graphs. The construction is as follows: Pick n points uniformly in the unit square, then, add an edge if and only if the distance between two points is $\leq r$
- Random trees. A tree is chosen uniformly at random from the n^{n-2} trees on n vertices.

In this class, we will primarily focus on the Erdős-Rényi Model.

1.1 Probability Primer

Definition 1.3. A discrete probability space consists of a countable set Ω and a probability function $\mathbb{P} : \Omega \to [0,1]$ such that $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$

A subset of Ω is called an event. The probability of $A \subseteq \Omega$ is $\sum_{\omega \in A} \mathbb{P}(\omega)$, denoted $\mathbb{P}(A)$.

Proposition 1.1 (Inclusion-Exclusion)**.** For events A, B:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

and, in general:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i_1 < i_2} \mathbb{P}(A_{i_1} \cap A_{i_2}) + \sum_{i_1 < i_2 < i_3} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) + \ldots + (-1)^{n-1}\, \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

Corollary 1.1. $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$

Definition 1.4. Two events are <u>independent</u> if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$

Definition 1.5. A <u>random variable</u> (r.v) X is a function $X : \Omega \to \mathbb{R}$. In a discrete probability space, the <u>expectation</u> of X is defined by: $\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega)$

Proposition 1.2 (Linearity of Expectation)**.** $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

Proof. $\mathbb{E}(X+Y) = \sum_{\omega\in\Omega}(X+Y)(\omega)\,\mathbb{P}(\omega) = \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega) + \sum_{\omega\in\Omega} Y(\omega)\,\mathbb{P}(\omega) = \mathbb{E}(X) + \mathbb{E}(Y)$ $\square$

Lemma 1.1.

- For any $n \geq k \geq 1$

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \frac{n^k}{k!} \leq \left(\frac{en}{k}\right)^k$$

- (Stirling's Formula)

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \mathcal{O}(n^{-2})\right)$$

- For every $t \in \mathbb{R}, e^t \geq 1 + t$

Lemma 1.2. Assume $k = o(\sqrt{n})$ Then, $\binom{n}{k} \sim \frac{n^k}{k!}$

Proof.

$$\begin{aligned}
\binom{n}{k} &= \frac{1}{k!}\prod_{i=0}^{k-1}(n-i) \\
&= \frac{n^k}{k!}\prod_{i=0}^{k-1}\left(1 - \frac{i}{n}\right) \\
&= \frac{n^k}{k!}\prod_{i=0}^{k-1} e^{\mathcal{O}(i/n)} \qquad (\log(1-x) = \mathcal{O}(x)) \\
&= \frac{n^k}{k!}\exp\left(\mathcal{O}\left(\frac{1}{n}\sum_{i=0}^{k-1} i\right)\right) \\
&= \frac{n^k}{k!}\exp\left(\mathcal{O}\left(\frac{k^2}{n}\right)\right) \\
&= (1 + o(1))\frac{n^k}{k!} \qquad \left(\text{as } k = o(\sqrt{n})\right)
\end{aligned}$$

$\square$

Remark 1.1. $k = o\left(n^{\frac{2}{3}}\right)$, then $\binom{n}{k} \sim e^{-\frac{k^2}{n}} \cdot \frac{n^k}{k!}$

Lecture 2: Concentration Inequalities, Coupling, Connection Theorem

Definition 2.1. Given a sequence of probability spaces $(\Omega_n, P_n)_{n\geq 1}$. We say that A_n holds asympototically almost surely (a.a.s) if $P_n(A_n) \to 1$ as $n \to \infty$

2.1 Concentration Inequalities

Theorem 2.1 (Markov's Inequality)**.** Let X be a nonnegative random variable. Then, for any real $t > 0$, $\Pr(X \geq t) \leq \frac{\mathbb{E}\,X}{t}$

Proof. Let I_t be the indicator r.v. that $X \geq t$. Then, $X \geq t \cdot I_t$, so:

$$\mathbb{E}\,x \geq t \cdot \mathbb{E}\,I_t = t \cdot \mathbb{P}(X \geq t)$$

□

Theorem 2.2 (Chebyshev's Inequality)**.** For any $t \geq 0$

$$\mathbb{P}(|X - \mathbb{E}\,X| \geq t) \leq \frac{\operatorname{Var} X}{t^2}$$

Example 2.1. Let X be the number of edges in $\mathcal{G}(n,p), N = \binom{n}{2}$. $X \sim \operatorname{Bin}(N,p)$ so $\mathbb{E}\,X = Np$ and $\operatorname{Var} X = p(1-p)N$.

Further, by Chevyshev's Inequality, for all $t > 0$:

$$\mathbb{P}(|X - \mathbb{E}\,X| \geq t) \leq \frac{p(1-p)N}{t^2}$$

△

This leads us to the following proposition:

Proposition 2.1. Let $f_n \to \infty$ as $n \to \infty$. Then,

$$\mathbb{P}(|X - Np| \geq f_n\sqrt{p(1-p)N}) \leq \frac{1}{f_n^2} = o(1)$$

So a.a.s:

$$pN - f_n\sqrt{p(1-p)N} \leq X \leq pN + f_n\sqrt{p(1-p)N}$$

2.2 Coupling

Definition 2.2. Given 2 r.v.s X, Y, a coupling of X and Y is a construction of a join distribution of $(\hat{X}, \hat{Y})$ into the probability space such that marginally $\hat{X} \sim X$ and $\hat{Y} \sim Y$

Lemma 2.1.

(a) Let $0 \leq m_1 < m_2 \leq N$ and $0 \leq p_1 < p_2 \leq 1$. There exist couplings such that:

$$\mathcal{G}(n, m_1) \subseteq \mathcal{G}(n, m_2) \quad \text{and} \quad \mathcal{G}(n, p_1) \subseteq \mathcal{G}(n, p_2)$$

where by $\mathcal{G}(n, p_1) \subseteq \mathcal{G}(n, p_2)$ (and respectively, m_1 and m_2), we mean that there exists a coupling (G_1, G_2) such that:

- Marginally, $G_1 \sim \mathcal{G}(n, p_1), G_2 \sim \mathcal{G}(n, p_2)$, and
- jointly, $G_1 \subseteq G_2$ always

(b) Let $m_1 = pN - f\sqrt{p(1-p)N}, m_2 = pN + f\sqrt{p(1-p)N}$ ($f = f(n)$ as before). Then, there exists a coupling (G_1, H, G_2) such that:

- $G_1 \sim \mathcal{G}(n, m_1), G_2 \sim \mathcal{G}(n, m_2), H \sim \mathcal{G}(n, p)$
- $\mathbb{P}(G_1 \subseteq H \subseteq G_2) = 1 - o(1)$

Proof.

(a) Let $G_1 \sim \mathcal{G}(n, p_1)$. For G_2, include every non-edge in G_1, include it independently with probability $q = 1 - \frac{1-p_2}{1-p_1}$. Clearly, $G_1 \subseteq G_2$. Then, check the probability that an edge is not included in G_2:

$$(1-p_1)(1-q) = (1-p_1)\left(1 - \left(1 - \frac{1-p_2}{1-p_1}\right)\right) = 1 - p_2$$

For $G_1 \sim \mathcal{G}(n, m_1), G_2 \sim \mathcal{G}(n, m_2)$, we choose permutation and the first m_1, m_2 edges

□

2.3 Connection Theorem

Definition 2.3. Let Ω be the set of graphs on $[n]$. $Q \subseteq \Omega$ is a graph property if it is invariant under graph isomorphism. We say Q is monotone increasing if:

$$G \in Q \Rightarrow H \in Q \quad \forall H \supseteq G$$

Further, we say Q is convex if:

$$G_1, G_2 \in Q, G_1 \subseteq G_2 \Rightarrow H \in Q \quad \forall G_1 \subseteq H \subseteq G_2$$

Theorem 2.3. Suppose Q is monotone. Then, for any $0 \leq m_1 \leq m_2 \leq N$, $0 \leq p_1 \leq p_2 \leq 1$:

$$\mathbb{P}(\mathcal{G}(n, m_1) \in Q) \leq \mathbb{P}(\mathcal{G}(n, m_2) \in Q)$$
$$\mathbb{P}(\mathcal{G}(n, p_1) \in Q) \leq \mathbb{P}(\mathcal{G}(n, p_2) \in Q)$$

Theorem 2.4 (Connection Theorem)**.** Let Q be a graph property:

(i) Given $p = p(n)$. Suppose for all $m = pN + \mathcal{O}(\sqrt{p(1-p)N})$ we have $\mathcal{G}(n, m) \in Q$ a.a.s. Then, a.a.s $\mathcal{G}(n, p) \in Q$

(ii) Suppose Q is convex. Given $m = m(n)$ and suppose $\mathcal{G}(n, m/N) \in Q$ a.a.s. Then, a.a.s $\mathcal{G}(n, m) \in Q$

Proof Sketch.

(i) Write $\mathbb{P}(\mathcal{G}(n, p) \in G)$ in terms of the number of edges in the graph, i.e. $\mathbb{P}(\mathcal{G}(n, p) \in G) = \sum_{m=0}^{N} \mathbb{P}(X = m, \mathcal{G}(n, p) \in Q)$ (Law of total probability). Then, use Proposition 2.1.

(ii) Condition on the number of edges, and analyze the probabilities of having a graph with: less edges than 1 standard deviation (m_1), more edges than 1 standard deviation (m_2), and a number of edges within 1 standard deviation (m). Then construct graphs from $\mathcal{G}(n, m_1)$ and $\mathcal{G}(n, m_2)$ and use convexity to show $\mathbb{P}(\mathcal{G}(n, m) \in Q) = 1 - o(1)$

$\square$

Lecture 3: Threshold, First Order Logic of Graphs

3.1 Threshold

Definition 3.1. We say a property Q has a threshold p_0 if:

$$\mathbb{P}(\mathcal{G}(n,p) \in Q) \to \begin{cases} 0 & \text{if } p \ll p_0 \\ 1 & \text{if } p \gg p_0 \end{cases}$$

Theorem 3.1 (Bollobás & Thomason, 1987)**.** Every non-trivial monotone property has a threshold

Definition 3.2. We say a property Q has a sharp threshold p_0 if $\forall \epsilon > 0$:

$$\mathbb{P}(\mathcal{G}(n,p) \in Q) \to \begin{cases} 0 & \text{if } p \leq (1-\epsilon)p_0 \\ 1 & \text{if } p \geq (1+\epsilon)p_0 \end{cases}$$

Definition 3.3. The window of a threshold is $\delta(\epsilon) = p_{1-\epsilon} - p_\epsilon$

3.2 First Order Logic of Graphs

Example 3.1.

$$\forall x \forall y \exists z \, (x = y \vee x \sim y \vee (x \sim z \wedge y \sim z))$$

is the statement characterizing the graphs of diameter ≤ 2 △

Fix $k > 0$. Let P_k be the property that for any disjoint sets W and V of order at most k, there exists a vertex $x \in V(G) \setminus (W \cup V)$ such that x is adjacent to all vertices in W and is adjacent to none of V

Lemma 3.1. Suppose $m(n), p(n)$ satisfy the following:

For every fixed $\epsilon > 0$

$$\begin{aligned} &mn^{-2+\epsilon} \to \infty, \qquad (N-m)n^{-2+\epsilon} \to \infty \\ &pn^{\epsilon} \to \infty, \qquad (1-p)n^{\epsilon} \to \infty \end{aligned}$$

For every fixed $k > 0$, a.a.s $\mathcal{G}(n,p) \in P_k$ and $\mathcal{G}(n,m) \in P_k$

Theorem 3.2 (0-1 law of the 1st order logic of random graphs)**.** Suppose $m(n), p(n)$ satisfy the conditions of the lemma. Suppose Q is a graph property given by a 1st order sentence. Then, either Q holds a.a.s or does not hold a.a.s.

Proof Sketch. We play a k-round Ehrenfeucht-Fraïssé Game. Player 1 chooses vertices from either graph and Player 2 must choose vertices from the other graph.

After k rounds, this produces two sequences $v_1, v_2, \dots, v_k$ in G_1 and $v_1', v_2', \dots, v_k'$ in G_2. Player 2 wins if $v_i \mapsto v_i'$ $\forall 1 \leq i \leq k$ is an isomorphism between $G_1[v_1, v_2, \dots, v_k]$ and $G_2[v_1', v_2', \dots, v_k']$, and Player 1 wins otherwise.

The idea is that if G_1, G_2 are similar, then player 2 will win, but if they are not similar, then player 1 can exploit the dissimilarity.

<u>Claim</u>: Let $Th_k(G)$ be the set of graph properties of G expressible by 1st-order logic sentences with quantifier depth at most k. Player 2 has a winning strategy if and only if $Th_k(G_1) = Th_k(G_2)$ (i.e. They share the same set of properties) $\square$

Lecture 4: Evolution of Graphs and Theorem E

Theorem 4.1 (Theorem E)**.**

(a) Fix $k \geq 2$ integer. If $n^{\frac{k-2}{k-1}-2} \ll p \ll n^{\frac{k-1}{k}-2}$, then a.a.s $\mathcal{G}(n,p)$ is a forest and the largest component is of order k.

(b) If $p \ll \frac{1}{n}$, then a.a.s $\mathcal{G}(n,p)$ is a forest and the largest component is of order $o(\log n)$

(c) If $p = \frac{c}{n}$, $0 < c < 1$, then a.a.s every component of $\mathcal{G}(n,p)$ is a tree or unicyclic and the largest component has order $\Theta(\log n)$

(d) If $p = \frac{c}{n}$, $c > 1$, then a.a.s $\mathcal{G}(n,p)$ contains a unique component of linear order and all other components of order $\mathcal{O}(\log n)$

(e) When $p \geq \frac{\log n + \log \log n \omega(1)}{2n}$, a.a.s $\mathcal{G}(n,p)$ has a giant component and a few isolated vertices

(f) When $p \geq \frac{\log n + \omega(1)}{n}$, a.a.s $\mathcal{G}(n,p)$ connected and has a perfect matching if n even, or a matching of size $\frac{n-1}{2}$ if n is odd.

(g) When $p \geq \frac{\log n + \log \log n + \omega(1)}{n}$, a.a.s $\mathcal{G}(n,p)$ is Hamiltonian

4.1 Small Subgraphs

Lemma 4.1. If $p = o(1/n)$, then a.a.s $\mathcal{G}(n,p)$ has no cycles

Proof. Use Markov's Inequality □

Proof. (of Theorem 4.1 (a))

That $\mathcal{G}(n,p)$ is a forest is directly implied by above. It remains to show that every tree has order $\leq k$ and there is one tree of order k.

Let X_t be the number of trees of order t in $\mathcal{G}(n,p)$. First, we show that that $\mathbb{E}\left(\sum_{t \geq k+1} X_t\right) = o(1)$, so a.a.s $\sum_{t \geq k+1} X_t = 0$ by Markov's inequality. This tells us that we don't have trees of order $> k$

Then, we show the existence of a tree of order k by using the 2nd moment method. □

Lecture 5:

References

[1] Bollobás Béla. *Random graphs.* Academic Press, 1985.