CO749 - Random Graph Theory

(Lecture Summaries ... sort of)

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 independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$

Definition 1.5. A random variable (r.v) X is a function $X : \Omega \to \mathbb{R}$. In a discrete probability space, the expectation 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. $\mathbb{E}(X+Y) = \sum_{\omega\in\Omega}(X+Y)(\omega)\,\mathbb{P}(\omega) = \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega) + \sum_{\omega\in\Omega} Y(\omega)\,\mathbb{P}(\omega) = \mathbb{E}(X) + \mathbb{E}(Y)$ □

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

□

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 $v_1, v_2, \ldots, v_k$ in G_1 and $v_1', v_2', \ldots, v_k'$ in G_2. Player 2 wins if $v_i \mapsto v_i' \ \forall 1 \le i \le k$ is an isomorphism between $G_1[v_1, v_2, \ldots, v_k]$ and $G_2[v_1', v_2', \ldots, v_k']$, 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.

Claim: 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 $\square$

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. $\square$

Lecture 5: Cycles, Degrees of Vertices, Critical Window Analysis

Let $X \sim \text{Po}(\lambda)$ and recall that $\mathbb{P}(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}\ \forall i \geq 0$ and $\mathbb{E}(X)_r = \lambda^r\ \forall r \geq 0$.

Theorem 5.1. Given a sequence of random variables $(X_n)_{n\geq 1}$ and $\lambda > 0$, suppose $\mathbb{E}(X_n)_r \to \lambda^r\ \forall r \geq 0$. Then, $X_n \to \text{Po}(\lambda)$ as $n \to \infty$

Let Y_k denote the number of k-cycles in $\mathcal{G}(n,p)$.

Theorem 5.2. Let $p = \frac{c}{n}$, where c fixed, then for every $k \geq 3$, $Y_k \to \text{Po}\left(\frac{c^k}{2k}\right)$

Proof. $\mathbb{E}[Y_k] = \binom{n}{k}\frac{(k-1)!}{2}p^k \sim \frac{n^k}{k!}\frac{(k-1)!}{2}p^k = \frac{n^k p^k}{2k} = \frac{c^k}{2k}$

Now, fix $r \geq 2$ and consider:

$$(Y_k)_r = |\{C_1, \ldots, C_r \ :\ C_i \text{ is a } k\text{-cycle}\}|$$

So, $(Y_k)_r$ can be treated like the number of ordered lists of k-cycles. First, consider r k-cycles that are all vertex-disjoint, then consider the case where there are vertex intersections. Apply Theorem 5.1 to finish the proof. □

5.1 Degrees of Vertices

Theorem 5.3. Let $p = \frac{c\log n}{n}$ ($c > 0$ fixed).

(a) If $c < 1$, a.a.s $\exists$ isolated vertices

(b) If $c > 1$, a.a.s there are no isolated vertices

(Note: This is a sharp threshold)

Proof. Let X be the number of isolated vertices.

$$\mathbb{E}[X] = n \cdot (1-p)^{n-1} \sim n^{1-c}$$

Then, if $c > 1$, then use Markov's inequality to find (b). If $c < 1$, then use Chebyshev's inequality to find (a). □

5.2 Critical Window Analysis

Theorem 5.4. Let $p = \frac{\log n + c}{n}$, c fixed. Then

$$\mathbb{P}(X = 0) \sim e^{-e^{-c}} \tag{5.1}$$

Corollary 5.1. Let $p = \frac{\log n + c(n)}{n}$. If $c(n) \to -\infty$, then a.a.s $\mathcal{G}(n,p)$ has isolated vertices. And, if $c(n) \to +\infty$, then a.a.s $\mathcal{G}(n,p)$ has no isolated vertices

Let X_k be the number of vertices with degree k

Theorem 5.5. Let $\epsilon > 0$ be fixed, $\epsilon n^{-\frac{3}{2}} \leq p \leq 1 - \epsilon n^{-\frac{3}{2}}$. Let $k = k(n)$ be a nonnegative integer and $\lambda_k(n) = n \cdot \mathbb{P}(\mathrm{Bin}(n-1, p) = k)$ Then,

(i) If $\lambda_k(n) = o(1)$, then a.a.s $X_k = 0$

(ii) If $\lambda_k(n) \to o(1)$, then a.a.s $X_k \geq t$ for any fixed t.

(iii) If $0 < \lambda_k := \lim_{n\to\infty} \lambda_k(n) < \infty$, then $\mathbb{P}(X_k = t) \sim e^{\lambda_k} \cdot \frac{\lambda_k^t}{t!}$

Theorem 5.6. Let $p = \frac{\log n + c}{n}$, c fixed. Then, $\mathbb{P}(\mathcal{G}(n,p) \text{ is connected}) \to e^{-e^{-c}}$

In a feat of ultimate laziness (as the true computer scientist I am), the rest of the notes are going to be scans.

7CQ - ~~Colouring~~ Random. Jan 20th

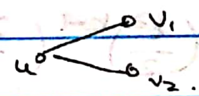
Lecture 6:

$\mathcal{G}(n,n,p)$ denotes the random bipartite graph with $V = V_1 \cup V_2$, $|V_1| = |V_2| = n$ and uv is an edge with prob p for $u \in V_1$, $v \in V_2$

Exercise: let $p = \frac{\log n + c}{n}$. Let X denote # of isolated vertices in $\mathcal{G}(n,n,p)$. Then,

- $P(X=0) \sim e^{-2e^{-c}}$
- $P(\mathcal{G}(n,n,p)$ is connected$) \sim e^{-2e^{-c}}$

A triple of vertices $\{u, v_1, v_2\}$ is called a cherry if $d(v_1) = d(v_2) = 1$ and $v_1 \sim u$ and $v_2 \sim u$.



Exercise: Let $p \sim c \log n / n$ where $c > 3/5$. Then a.a.s there are no cherries in $\mathcal{G}(n,n,p)$

Theorem (Hall's theorem)
A bipartite graph G with vertex set $V_1 \cup V_2$ has a PM if $|V_1| = |V_2|$ and $|N(S)| \geq |S|$ $\forall S \subseteq V_1$

Exercise: Suppose G fails Hall's condition. Then, there exists a set ~~[illegible]~~ $S \subseteq V_1$ or $S \subseteq V_2$.

(i) $N(S) = |S| - 1$
(ii) $|S| \leq \lceil \frac{n}{2} \rceil$
(iii) Every vertex in $N(S) \geq 2$ neighbours in S

Let X_k denote a k-set S satisfying (i), (ii), (iii)

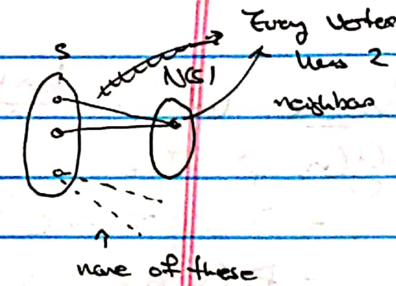
X_1 : # isolated vertices

X_2 : # cherries

Lemma:

Let $(3/4)\frac{\log n}{n} < p < \frac{2\log n}{n}$. Then, $\sum_{k=3}^{\lceil \frac{n}{2} \rceil} \mathbb{E}X_k = o(1)$

Proof:



For each $3 \le k \le \lceil \frac{n}{2} \rceil$

$$\mathbb{E}X_k \le \underbrace{2\cdot\binom{n}{k}}_{\text{Choose from either } V_1 \text{ or } V_2}\cdot\underbrace{\binom{n}{k-1}}_{\text{neighbors}}\underbrace{\left(\binom{k}{2}p^2\right)^{k-1}}_{\text{Each vertex in } N(S) \text{ have 2 neighbors}}\underbrace{(1-p)^{k(n-k+1)}}_{\text{No edges going from } S \text{ to not } N(S)}$$

$$\le 2\left(\frac{en}{k}\right)^k\left(\frac{en}{k-1}\right)^{k-1}\left(\frac{k^2}{2}\right)^{k-1}\left(\frac{2\log n}{n}\right)^{2k-2}\exp\left(-\frac{3\log n}{4n}\cdot k\cdot \underbrace{\frac{n}{2}}_{\text{since } n-k+1 \le \frac{n}{2}}\right)$$

$$\le 2n\left(\frac{k}{k-1}\right)^{k-1}\left(\frac{2e^2\log^2 n}{n^{3/8}}\right)^k$$

$$= O(n\alpha^k), \text{ where } \alpha = \frac{3e^2\log^2 n}{n^{3/8}}.$$

So: $\sum_{k=3}^{\lceil \frac{n}{2} \rceil} \mathbb{E}X_k = o(1)$ □

Theorem

Let $p = \frac{\log n + c}{n}$, c fixed $\mathbb{R}$. Then $\Pr(\mathcal{G}(n,n,p) \text{ has a PM}) \sim e^{-2^{-c}}$

Proof:

Let ~~E_1~~ C_1 denote the event that $X_1 = 0$, i.e. no isolated vertices,

C_2 ——— " ——— $X_k = 0$ $\forall\, 2 \le k \le \lceil \frac{n}{2} \rceil$

Hall's theorem and Exercise $\Rightarrow C_1 \cap C_2 \Rightarrow$ $\mathcal{G}(n,n,p)$ has a PM

So, $\Pr(\mathcal{G}(n,n,p) \text{ has PM}) \ge \Pr(C_1 \cap C_2)$

$= \Pr(C_1) - \Pr(C_1 \cap \overline{C_2})$

$\ge \Pr(C_1) - \Pr(\overline{C_2}) = P(C_1) - o(1)$ ← By lemma + exercise

Proof (con't)

So, $G(n,n,p)$ has a PM $\Rightarrow C_1$ ①

And,

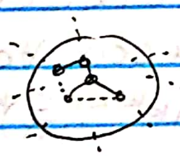
$$\Pr(G(n,n,p) \text{ has a PM}) \le \Pr(C_1) \quad ②$$

So:

$$\Pr(G(n,n,p) \text{ has a PM}) = \Pr(C_1) + o(1) \sim e^{-2e^{-c}} \qquad \square$$

Let τ_k denote the number of components that are trees of order k.

Let $\tau = \sum_{k=1}^{n} k\tau_k$.



$$E\tau_k = \underbrace{\binom{n}{k}}_{k \text{ vertices}} \underbrace{k^{k-2}}_{\# \text{ of trees}} \overbrace{p^{k-1}}^{\text{Edges in tree}} (1-p)^{\underbrace{\binom{k}{2}-(k-1)}_{\text{No other edges except for tree edges}} + \underbrace{k(n-k)}_{\text{edges coming out of component}}}$$

$$E\tau = \sum_{k=1}^{n} k\, E\tau_k$$

Theorem Let $p = \frac{c}{n}$, $c > 0$ fixed. τ is a function of the random graph

(a) If $0 < c < 1$, then $E[\tau(G(n,p))] = n + o(1)$

(b) If $c > 1$, then $E[\tau(G(n,p))] = t(c)n + o(1)$,

where $t(c) = \frac{1}{c} \sum_{k \ge 1} \frac{k^{k-1}}{k!} (ce^{-c})^k$

Note: $t(c) \le \sum_{k \ge 1} \frac{1}{k} \left(\frac{ek}{k}\right)^k (ce^{-c})^k = \sum_{k \ge 1} \frac{1}{k} (\underbrace{ce^{1-c}}_{})^k$

Can show this is < 1 for $c \ne 1$.

So, $t(c)$ is well-defined.

Cont

Exercise: (a) $s(c) = c\,t(c)$ is the unique solution of $se^{-s} = ce^{-c}$ in the range $0 < s \le 1$

(b) $t(c) = 1 \ \forall\, 0 < c \le 1$.

Note: $\mathbb{E}[T(G(n,p))]$ is cts. in p and $t(c) \to 1$ as $c \to 1$.
So, $\mathbb{E}[T(G(n,\tfrac{1}{n}))] = n + o(n)$. (Since $t(c)n = n +$ "some error").

Proof:
Since $t(c) = 1 \ \forall\, 0 < c < 1$, it's sufficient to show that $\mathbb{E}T = t(c)n + O(1) \ \forall c \ne 1$.

$$\mathbb{E}T_k = \binom{n}{k} k^{k-2} \left(\frac{c}{n}\right)^{k-1} \left(1 - \frac{c}{n}\right)^{kn - \frac{k(k+3)}{2} + 1}.$$

Consider $1 \le k \le \sqrt{n}$. (In this range $\binom{n}{k} \sim e^{-\frac{k^2}{n}} \cdot \frac{n^k}{k!}$)
...uh... actually we'll use:
$$\binom{n}{k} = \frac{n^k}{k!} \exp\left(-\frac{k^2}{n} + O\left(\frac{k^3}{n^2}\right)\right).$$

Then,
$$\mathbb{E}T_k = n \cdot \frac{k^{k-2}}{n!} c^{k-1} e^{-ck} \exp\left(-\frac{k^2}{n} + \frac{ck^2}{2n} + O\left(\frac{k}{n}\right)\right).$$

Hence,
$$\left|\sum_{k=1}^{\sqrt{n}} k\,\mathbb{E}T_k - \sum_{k=1}^{\sqrt{n}} n \cdot \frac{k^{k-1}}{k!} c^{k-1} e^{-ck}\right|$$
$$= \sum_{k=1}^{\sqrt{n}} O\left(\frac{k^2}{n}\right) n \cdot \frac{k^{k-1}}{k!} c^{k-1} e^{-ck}$$
$$= \sum_{k=1}^{\sqrt{n}} O(k \cdot e^{k - ck} c^k)$$
$$= \sum_{k=1}^{\sqrt{n}} O(k(\underbrace{ce^{1-c}}_{<1})^k) = O(1).$$

$$\frac{\mathbb{E}\,(k+1)T_{k+1}}{\mathbb{E}\,kT_k} = (n-k)\underbrace{\left(1+\frac{1}{k}\right)^{k-2}}_{\le e} \cdot \frac{c}{n}\left(1-\frac{c}{n}\right)^{n-k-2}$$

$$\le \left(1-\frac{k}{n}\right) \cdot e c e^{-\frac{c}{n}(n-k)} \cdot \left(1-\frac{c}{n}\right)^{-2}$$

$$= \left(1-\frac{k}{n}\right) e c e^{-c\left(1-\frac{k}{n}\right)} \left(1-\frac{c}{n}\right)^{-2}$$

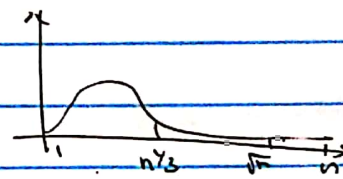
Let $\eta = \left(1-\frac{k}{n}\right)$.

$$= \underbrace{\eta e c e^{-c\eta}}_{c\eta e^{1-c\eta} \le 1} \left(1-\frac{c}{n}\right)^{-2}$$

$$\le \left(1-\frac{c}{n}\right)^{-2} := \lambda.$$

It is easy to verify (using formulae we've shown for $\mathbb{E}T_k$), that $\mathbb{E}T_{k_1} = o(n^{-M})$ for any fixed $M>0$, where $k_1 = n^{1/3}$

$$\sum_{k=k_1}^{n} k\,\mathbb{E}[T_k] \le k_1 \mathbb{E}[T_{k_1}] \sum_{i=0}^{n} \lambda^i$$



$$\le O(n) \cdot k_1 \mathbb{E}[T_{k_1}]$$

Since $\lambda^i \le \lambda^n = O(1)$

$$= o(n^{-3})$$

It is easy to check that $\sum_{k=\sqrt{n}}^{\infty} \frac{k^{k-1}}{k!} c^{k-1} e^{-ck} = \cancel{o(1)}\ o(n^{-1})$

$$|\mathbb{E}T - t(c)n| \le \left|\sum_{k=1}^{\sqrt{n}} k\,\mathbb{E}T_k - n \cdot \frac{k^{k-1}}{k!} c^{k-1} e^{-ck}\right|$$

$$+ \sum_{k=\sqrt{n}}^{n} k\,\mathbb{E}T_k + \sum_{k=\sqrt{n}}^{\infty} n \cdot \frac{k^{k-1}}{k!} c^{k-1} e^{-ck}$$

$$= O(1) + o(1) + o(1) = O(1)$$

□

749 - Random

Feb 3rd

Lecture 7:

Clarification on thresholds:

A property Q has a threshold $p_0 = p_0(n)$ if:

$$\Pr(\mathcal{G}(n,p) \in Q) \to \begin{cases} 0 & \text{if } p = o(p_0) \iff p \ll p_0 \\ 1 & \text{if } p = \omega(p_0) \iff p \gg p_0 \end{cases}$$

$\iff$

Q has a threshold $p_0 = p_0(n)$ if

$\forall \varepsilon > 0, \exists c > 0, n_0 > 0$ s.t. for all $n > n_0$

$\Pr(\mathcal{G}(n,p) \in Q) < \varepsilon$ if $p < \frac{1}{c} p_0$

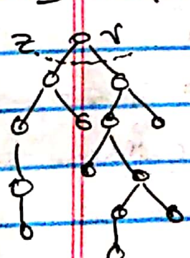
& $\Pr(\mathcal{G}(n,p) \in Q) > 1 - \varepsilon$ if $p > c p_0$

Recall: (Theorem 2(d))

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

Intuition for proof:

Branching / Exploration Process



$\deg(v) \sim \text{Bin}(n-1, p) \approx Po(c)$ ← approximately (same poisson theorem)

↑ Same parameter

$Z \sim Po(c)$

↑ The out has poisson distribution

- Start at a vertex v, and explore the graph by doing a BFS/DFS on its children.
- Degrees (# of children) is roughly the binomial ~~distrib~~ distribution.

①

Formalizing this idea:

A Galton-Watson branching process.

Let Z be a random variable on nonnegative integers. At time $t=0$, a single particle comes into existence. At each $t \geq 1$, each existing particle gives birth to a random number of new particles and then it dies. These random numbers are i.i.d copies of Z.

Let X_t denote the number of particles born during step $t \geq 0$. Thus, $X_0 = 1$, and X_1 has the same distribution as Z.

Let $\mathcal{E}$ denote the event "that $X_t = 0$ for some $t \geq 0$", i.e. the process terminates after a finite # of steps.

Let $p_Z = \Pr(\mathcal{E})$

Exercise: For every $t \geq 1$, $\mathbb{E} X_t = (\mathbb{E} Z)^t$

Theorem

(i) If $\mathbb{E} Z < 1$, then $p_Z = 1$.

(ii) If $\mathbb{E} Z = 1$, and $\Pr(Z=0) > 0$, then $p_Z = 1$.

(iii) If $\mathbb{E} Z > 1$ and $\underbrace{\Pr(Z=0) > 0}$, then $0 < p_Z < 1$

to avoid triviality (if $\Pr(Z=0) = 0$, then we always give birth to new children, so $p_Z = 0$)

$\rightarrow$ Proof

Proof:

(i) $\lim_{t\to\infty} \Pr(X_t > 0) \underset{\text{Exercise + Markov}}{\leq} \lim_{t\to\infty} (\mathbb{E} Z)^t \overset{\text{Since } \mathbb{E}Z<1}{=} 0$

(ii) and (iii):

Let ρ_j be the probability that the process becomes extinct after j steps. This blah blah occurs. Let $v_1, \dots, v_{X_1}$ denote the set of children born at step 1. The process terminates after j steps iff the X_1 processes starting from $v_1, \dots, v_{X_1}$ all terminate after $j-1$ steps. These X_1 processes are independent.

Then:

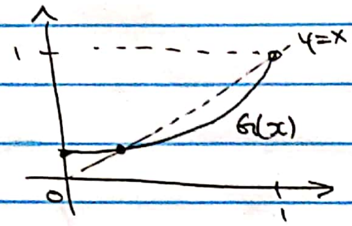
$$\rho_j = \sum_{i=0}^{\infty} \underbrace{\Pr(Z=i)}_{i \text{ children}} \underbrace{\rho_{j-1}^{i}}_{\text{each terminate in } j-1 \text{ steps}}$$

Let $G(x) = \mathbb{E}\, x^Z = \sum_{i=0}^{\infty} \Pr(Z=i) x^i$, then $\rho_j = G(\rho_{j-1})$

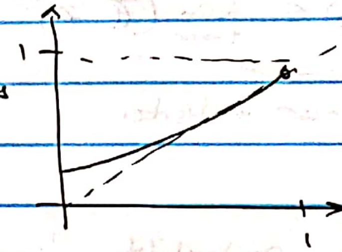
Note:

- $G(1) = 1$
- $G(0) > 0$ since $G(0) = \Pr(Z=0) > 0$
- G convex (since G is a probability generating function and they are always convex)
- $G'(1) = \sum_{i=1}^{\infty} i x^{i-1} \Pr(Z=i) \Big|_{x=1} = \mathbb{E} Z \underset{\text{By assumption.}}{\geq} 1.$

These facts can help us visualize G.



*Shape comes from that G convex

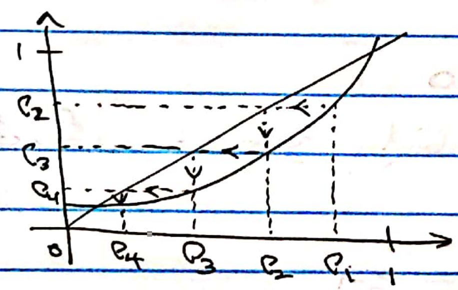


If $EZ > 1$, then $G(x) = x$ has 2 roots, 1 root at $x=1$ and the other $0 < P_Z < 1$

If $EZ = 1$, then $G(x) = x$ has a unique root at $x=1$ (the line $y=x$ is tangent to G at $x=1$)

For (iii) (and (ii) is similar), it is sufficient to show that $P_t \to P_Z$ for $t \to \infty$, $0 < P_1 < 1$, and the conclusion follows from the properties of G. $\square$

Proof by picture!



$P_2 = G(P_1)$, then find P_2 on the x-axis by following $y=x$ etc. Then $P_3 = G(G(P_2))$. And continue until we converge at the fixed point $G(P_Z) = P_Z$.

Theorem: (Chernoff Bound)

Let $X_1, \dots, X_n$ be independent $\{0,1\}$-value random variables. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}X$. Then the following probability bounds hold:

(a) For any $\delta > 0$:

$$\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

(a) ⇒ (b)

(b) For any $0 < \delta \leq 1$

$$\Pr(X \geq (1+\delta)\mu) \leq \exp\left(-\frac{\mu\delta^2}{3}\right)$$

(c) For any $0 < \delta \leq 1$

$$\Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

$$\Rightarrow \Pr(X \leq (1-\delta)\mu) \leq \exp\left(-\frac{\mu\delta^2}{2}\right)$$

Proof:

Let $X_i \sim \text{Bernoulli}(p_i)$

$$\Pr(X \geq (1+\delta)\mu) = \Pr(e^{tX} \geq e^{t(1+\delta)\mu}) \quad \text{holds for any } t>0,\ t \in \mathbb{R}$$

$$\leq \frac{\mathbb{E}e^{tX}}{e^{t(1+\delta)\mu}} \quad \text{Markov}$$

$$= \frac{\prod_{i=1}^{n} \mathbb{E}e^{tX_i}}{e^{t(1+\delta)\mu}} \quad \text{as } X_1, \dots, X_n \text{ are independent}$$

$$= \frac{\prod_{i=1}^{n} (1 - p_i + p_i e^t)}{e^{t(1+\delta)\mu}}.$$

$$\leq \frac{\prod_{i=1}^{n} e^{p_i(e^t - 1)}}{e^{t(1+\delta)\mu}} \xrightarrow{\sum p_i = \mu\ (\ast)} = \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}}.$$

For part (a), take $t = \log(1+\delta) > 0$

Proof: (con't)

Part (b) follows by noting that $\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \le e^{-\delta^2/3}$ for $0<\delta\le 1$.

For part (c), for ~~some~~ $t<0$

$$\Pr(X \le (1-\delta)\mu) = \Pr(e^{tX} \ge e^{t(1-\delta)\mu})$$
$$\le \frac{E(e^{tX})}{e^{t(1-\delta)\mu}}$$
$$\le \frac{e^{\mu(e^t-1)}}{e^{t(1-\delta)\mu}}$$

Proceed as before

And part (c) follows by taking $t=\log(1-\delta)<0$.
The 2nd inequality follows by $\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \le e^{-\delta^2/2}$ for $0<\delta<1$ $\square$.

Remark:
The Chernoff bounds hold for ~~any~~ r.v. $X_1,\dots,X_n$ that are independent $[0,1]$-valued as well.
Modifications for the proof:
At the step

$$\frac{\prod_{i=1}^{n} E\, e^{tX_i}}{e^{t(1+\delta)\mu}}$$ we cannot split this to $$\frac{\prod_{i=1}^{n}(1-p_i+p_ie^t)}{e^{t(1+\delta)\mu}}.$$

But, we will use that e^x is a convex function and $e^x \le \alpha x+\beta$ for all $x\in[0,1]$ where $\alpha=e^t-1$ and $\beta=1$.
Then,

$$E e^{tX_i} \le E(\alpha(tX_i)+\beta) = p_ie^t+(1-p_i)$$

and the remaining steps are the same.

Scanned with CamScanner

Lecture 8:

Thm E(d): Let $c > 1$. a.a.s $\mathcal{G}(n, c/n)$ has a unique giant component and all other components are of order $O(\log n)$

Still need a bit more probability theory before we can prove this theorem!

Stochastic Dominance:
Given 2 real r.v.s X and Y. We say that X stochastically dominates Y if:

$$P(X \geq x) \geq P(Y \geq x) \quad \forall x$$

We can couple X and Y such that $X \geq Y$ always. Then X stochastically dominates Y always. (Proof below)

Compare: If we have 2 ~~ran~~ random graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ such that $\mathcal{G}_1$ and $\mathcal{G}_2$ can be coupled so that $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then for any monotone increasing property Q:

$$Pr(\mathcal{G}_2 \in Q) \geq Pr(\mathcal{G}_1 \in Q)$$

Proof:
$\forall x$. Let $\Omega^X(x) = \{\omega \in \Omega \text{ s.t. } X(\omega) \geq x\}$

$\Omega^Y(x) = \{\omega \in \Omega \text{ s.t. } Y(\omega) \geq x\}$

Since in Ω we have $X(\omega) \geq Y(\omega)$ $\forall \omega \in \Omega$

$\Rightarrow \Omega^X(x) \supseteq \Omega^Y(x)$ $\forall x$

So:

$$P(X \geq x) = \int_{\omega \in \Omega^X(x)} dP(\omega) \geq \int_{\omega \in \Omega^Y(x)} dP(\omega) = P(Y \geq x)$$

$\square$

* The p's are p's (rho's)

Proof (of Thm (iii))

Let $k_0 = C \log n$ and $k_1 = n^{2/3}$ where $C > 0$ is a sufficiently large constant. We will prove a.a.s.:

(i) There are no components of $G(n, c/n)$ with order between k_0 and k_1

(ii) There is only ≤ 1 component of order greater than k_1

(iii) The total # of vertices in small components (order $\leq k_0$) is asymptotic to ρn, where $\rho e^{c} = e^{\rho c}$, $0 < \rho < 1$

(Compare ρ with $t(c)$ in the theorem of tree components $\Rightarrow$ "$\rho = t(c)$" (Check yourself!)).

Intuition about ρ:

Take a uniformly random vertex u and explore its neighbours, its 2^{nd} neighbours, etc. This mimics the branching process with $Z \sim Po(c)$. Recall that P_Z is the probability of extinction. Then, P_Z is the root of $P = G(P)$, $0 < P < 1$. (where $G(x)$ is the PGF for this distribution), and $G(x) = e^{-c+cx}$ (this is the PGF for Poisson

(Aside:

$$G(x) = \sum_{j=0}^{\infty} \Pr(Z=j) x^j = \sum_{j=0}^{\infty} e^{-c} \frac{c^j}{j!} x^j$$

$$= e^{-c} \sum_{j=0}^{\infty} \frac{c^j}{j!} x^j$$

$$= e^{-c+cx}$$

).

And P_Z is the fixed point for this PGF.

Proof: (con't)

Parts (i) - (iii) says that $G(n,p)$ contains a unique component of linear order and all other components are small. Intuitively, the prob. that v lies in a small component coincides with P_2.

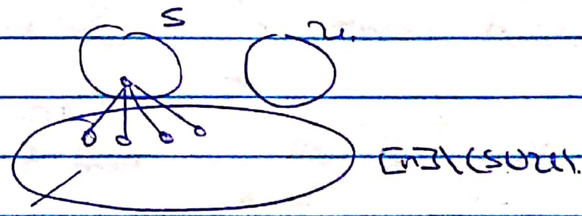
Consider the following graph exploration process. It starts with $t=0$ and $S=\{v\}$, $\mathcal{U}=\emptyset$. In every step $t\geq 1$, the process takes an arbitrary vertex $u\in S$, adds all neighbors of u in $G\setminus(S\cup \mathcal{U})$ into S, and then moves u from S to $\mathcal{U}$. The process terminates when $S=\emptyset$, and $\mathcal{U}$ contains all vertices in the same component as v and the size of $\mathcal{U}$ equals the # of steps the graph exploration process takes.

Let X_t be the size of S after step t. $X_0=1$

By the construction of the process, $|\mathcal{U}|=t$ and $|\mathcal{U}\cup S|=X_t+t$ after t steps of the process.

Let $Z_t = X_t - X_{t-1} + 1$ (# of new vertices added)

Then, $Z_{t+1} \sim Bin(n-(X_t+t), p)$



Z_t counts all the new vertices. This is a binomial distribution since an edge appears in $[n]\setminus(S\cup \mathcal{U})$ with probability p.

Proof:

To prove part (i), it is sufficient to prove that with probability $o(1/n)$ X_t reaches 0 for some $k_0 < t < k_1$, as by union bound, with probability $o(1)$, there is a component of order between k_0 and k_1.

Let $k_0 < \tau < k_1$ and E_τ be the event that $(X_t)_{t \geq 0}$ becomes 0 after step τ.

- Given E_τ, define $W_\tau := \sum_{t=1}^{\tau} Z_t = X_\tau - X_0 + \tau = \tau - 1$
- $Z_t \geq 0 \Rightarrow X_t \geq X_{t-1} - 1$ (In the worst case, we add no new ~~nodes~~ vertex and only remove ~~one~~ ourself).

  If for some $t' < \tau$ we have $X_{t'} > k_1$, then consequently $X_t > 0$ for all $t < k_1$, contradicting that $X_\tau = 0$.

So $E_\tau \Rightarrow X_{t'} \leq k_1$ for all $t' < \tau$.

In Summary:

$$E_\tau \Rightarrow \begin{cases} W_\tau = \tau - 1 & (1) \\ X_{t'} \leq k_1 \text{ for all } t' < \tau & (2) \end{cases}$$

(2) implies that $W_\tau := \sum_{t=1}^{\tau} Z_t$ stochastically dominates $\hat{W}_\tau := \sum_{t=1}^{\tau} \hat{Z}_t$ where $\hat{Z}_t$ are i.i.d copies of $Z \sim Bin(n - 2k_1, p)$

since both $X_t \leq k_1$ and $t \leq k_1$

This r.v. will be easier to analyze since Z ~~always~~ changes with each step.

Exercise ??
Aside: $\mathrm{Bin}(n_1, p)$ dominates $\mathrm{Bin}(n_2, p)$ if $n_1 \geq n_2$)

Proof (con't)

since each Z_t binomial.
$\hat{W}_\tau \sim \mathrm{Bin}(\tau(n-2k_1), p)$
$\mathbb{E}\hat{W}_\tau = \tau(n-2k_1)\cdot\frac{c}{n} = c\tau$ (where $\tau > 1$)
~~Let Event ...~~

Let $\sigma = c-1 > 0$ and let $\varepsilon = \sigma/2c$. Then $(1-\varepsilon)c = 1+\frac{\sigma}{2}$
By Chernoff Bound:

$Pr(E_\tau) \leq Pr(① \text{ and } ②)$

$\leq Pr(\hat{W}_\tau \leq 1)$ since $\tau \leq (1-\varepsilon)\tau c(1-\frac{2k_1}{n}) = (1+\frac{\sigma}{2})\tau(1-\frac{2k_1}{n})$

$= Pr(\hat{W}_\tau < (1-\varepsilon)\tau c(1-\frac{2k_1}{n}))$

$\leq \exp(-k\varepsilon^2\tau)$ (for some constant $k>0$)

$= o(\frac{1}{n})$ since $\tau \geq k_0 = C\log n$ and C is sufficiently large

And part (i) follows by taking union bound over $v \in [n]$.

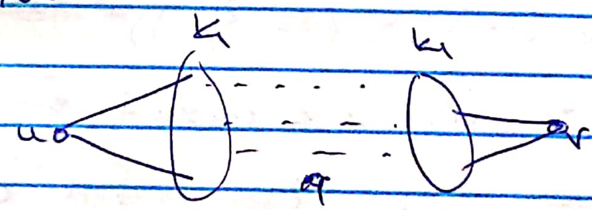
Proof (con't)

Part (ii). For any 2 vertices u and v, we prove that with prob $o(n^{-2})$, they lie in 2 different components, each of order greater than k_1.

Let $X_t(u)$ and $X_t(v)$ denote the size of $S(u)$ and $S(v)$ with respect to the graph exploration processes starting from u and v respectively.

We may assume that $X_t(u)$ and $X_t(v)$ are positive for all $t<k_1$ (since o/w ~~they~~ one of them lie in order k_1, i.e. the component has order $< k_1$)

Take $\varepsilon = \delta/2e$ and let $W_{k_1}(i) = \sum_{t=1}^{k_1} Z_t(i) = X_{k_1}(i) - X_0(i) + k_1$ for $i \in \{u, v\}$



We want to show w.h.p that if after k_1 steps, if we still have a large # of vertices to explore there must be an edge b/w the component containing u and the component containing v.

Scanned with CamScanner

Proof (cont.)

By Chernoff Bound, (and similar argument as (i)).

$$\Pr(W_{k_1}(i) < (1-\varepsilon)k_1(1-\frac{2k_1}{n})) = o(n^{-2})$$

Hence with prob $\geq 1-o(n^{-2})$, $X_{k_1}(u), X_{k_1}(v) \leq (\frac{\sigma}{4})k_1$

If $S(u) \cap S(v) \neq \emptyset$ after k_1 steps, then we're done. Else we may assume $|S(u)|, |S(v)| \geq (\frac{\sigma}{4})k_1$ and

$$\Pr(\text{no edge } b(w\ S(u) \text{ and } \{S(v)\})$$
$$\leq (1-p)^{(\frac{\sigma}{4}k_1)^2}$$
$$\leq o(n^{-2})$$

$\Rightarrow$ Part (ii) (By union Bound)

Lecture 9:

* All p's are thds.

Recall:

$k_0 = C\log n$, $k_1 = n^{2/3}$. Prove a.a.s

(i) No components of order (k_0, k_1)

(ii) Only one component of order $\geq k_1$

(iii) # vertices in small component $\sim \rho n$, where $\rho e^c = e^{\rho c}$ $(0<\rho<1)$

We will prove (iii) this class:

Let Y denote the # of vertices lying in small components (i.e. of order $< k_0$). We will prove that the probability for a ~~random~~ uniformly random vertex to be in a small component is $\rho + o(1)$

Lemma: Fix $c>1$ and let $Z_n \sim Bin(n, p)$ where $p \sim c/n$

Then $\rho_{Z_n} \to \rho$ as $n\to\infty$ where $\rho e^c = e^{\rho c}$, $(0<\rho<1)$

Proof:

Let $G(x) = E(x^{Z_n}) = (1-p+px)^n = e^{\underbrace{-c+cx+O(\frac{x^2}{n})+o(1)}_{\text{why?}}}$,

which converges to e^{-c+cx} pointwise as $n\to\infty$.

By the proof of the theorem for the branching process, part (iii):

ρ_{Z_n} is the unique solution of $G(x) = x$, where $0<x<1$.

As $G(x) \to \underbrace{e^{-c+cx}}_{F(x)}$ pointwise and ρ is the unique solution of

$x = F(x)$, $0<x<1$, it follows immediately $\rho_{Z_n} \to \rho$ as $n\to\infty$.

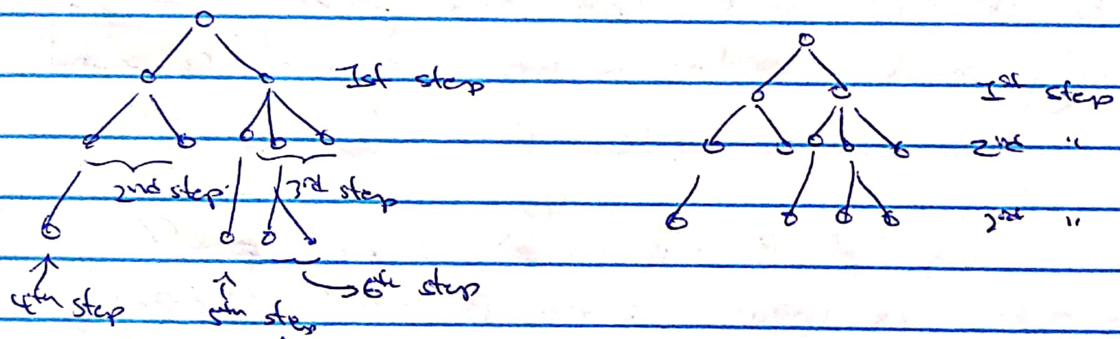
□

Lemma $\mathbb{E}Y = (1+o(1))\rho n$.

Proof:

Consider the graph exploration process starting from v, $\hat{\varphi}_v$.
Let E_v denote the event that $\hat{\varphi}_v$ lasts $\le k_0$ steps.
Consider the slowed down version of the branching process where $Z_n \sim Bin(n, (1+\xi)\frac{c}{n})$, where $0 < \xi = o(1)$.

(Slowed-down Branching process:



i.e. reveal grandchildren one child at a time)

$Bin(n, (1+\xi)\frac{c}{n})$ stochastically dominates $Bin(n-(\sum X_t + t), \frac{c}{n})$, so we can couple $(\hat{\varphi}_v, \varphi)$ such that $\hat{\varphi}_v$ terminates no later than φ. (Exercise!)

Let ρ_t denote the probability that φ terminates during the first t steps. We know $\rho_t \to \rho_{Z_n}$.
That is $\forall \varepsilon > 0, \exists n_0 > n \ \forall n > n_0$: $\rho_{k_0} \ge \rho_{Z_n} - \varepsilon$
And also, $\rho_{Z_n} \to \rho$ as $n \to \infty \Rightarrow \rho_{Z_n} \ge \rho - \varepsilon$
$\Rightarrow \rho_{k_0} \ge \rho - 2\varepsilon$

Feb 10th

By the coupling $(\hat{\varphi}_v, \varphi)$, we have $P(E_v) \geq P(\varphi$ terminates before k_0 steps$) = P_{k_0} > \rho - \varepsilon \Rightarrow P(E_v) \geq \rho - o(1)$

For the upper bound, consider the stoned-down branching process with $Z_n \sim Bin(n-k_1, (1-\varepsilon)\frac{c}{n})$. Z_n is stochastically dominated by $Bin(n-(Z_t+t), \frac{c}{n})$ for $t < k_0$

Again, we can couple these two processes such that during the first k_1 steps the graph exploration process will terminate no earlier than the branching process

$\Rightarrow P(E_v) \leq P($Branching process lasts $\leq k_0$ steps$)$

$= \rho + o(1)$., using similar argument as above

$\Rightarrow P(E_v) = \rho + o(1)$

$\Rightarrow EY = (1+o(1))\rho n$ $\square$

$EY(Y-1)$. Fix a pair of vertices u and v.

$Pr(u$ is in a small component$) = \rho + o(1)$

Condition on that. The graph exploration process only "exposes" $\leq k_0$ vertices.

(Exercise: Easy to show that a.a.s. v is not exposed) → i.e. In the same component that u is in.

Run the graph exploration process starting from v, condition on C_u, the component containing u. This is equivalent to running the graph exploration process in $G(n',p)$ where $n' = n - O(k)$.

Repeat the proof as for EY. We have ...

We have $\Pr(v \text{ is in a small component} \mid u \text{ is in a small component})$
$= \rho + o(1)$.

$\mathbb{E}Y(Y-1) = n(n-1)(\rho^2 + o(1)) \sim (\mathbb{E}Y)^2$.
By Chebyshev's Inequality: $\forall \varepsilon > 0$
$$\Pr(|Y - \mathbb{E}Y| > \varepsilon \mathbb{E}Y) \le \frac{\operatorname{Var} Y}{\varepsilon^2 (\mathbb{E}Y)^2} = \frac{o(1) \cdot (\mathbb{E}Y)^2}{\varepsilon^2 (\mathbb{E}Y)^2} = o(1)$$
So a.a.s. $Y \sim \mathbb{E}Y = (\rho + o(1))n$. □

Hamiltonian Cycles in $\mathcal{G}(n,p)$: ρ's are now p's

Theorem:
Let $p = \frac{\log n + \log\log n + x(n)}{n}$.

(a) If $x(n) \to \infty$, then a.a.s. $\mathcal{G}(n,p)$ is Hamiltonian
(b) If $x(n) \to -\infty$, then " Not Hamiltonian

(Note: When degree 1 vertices disappear, then our graph becomes Hamiltonian (w.h.p)).

Claims:
Let D_2 denote the property that min. degree is ≥ 2
HAM " graph is Hamiltonian
$G \in HAM \Rightarrow G \in D_2$

Lemma (Exercise)

(a) If $pn-(\log n+\log\log n)\to-\infty$ then a.a.s $G(n,p)\notin D2$

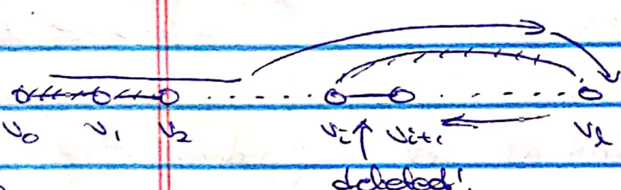
(b) If $pn-(\log n+\log\log n)\to\infty$ then a.a.s $G(n,p)\in D2$

Lemma $\Rightarrow$ Part (b) of the theorem.

Idea for proof:

We try to build a longer and longer path until we build a Hamiltonian cycle.
What is the probability of us creating a longer path? $\to$ We can only build from the 2 ends

Posá Rotation:



$v_0v_1v_2\dots v_iv_kv_{k-1}\dots v_{i+1}$ is a Posá Rotation

Def'n:

Let $P=v_0v_1\dots v_k$ be a longest path in G. Assume v_iv_k is an edge. Then, the operation of deleting edge v_iv_{i+1} from P and adding edge v_iv_k to P is called a Posá Rotation.

The Posá Rotation converts P to another longest path

Given a longest path $P = v_0 v_1 \dots v_\ell$. Let $\mathcal{P}$ be the set of longest paths obtained by fixing end v_0 and repeatedly performing Pósa Rotations.

Let $End(v_0)$ be the set of ends of paths in $\mathcal{P}$ other than v_0.

Given a subset S of vertices, let

$$N(S) = \cup\{u \notin S : \exists v \in S \text{ s.t. } v \sim u\}$$

Lemma: $|N(End(v_0))| < 2|End(v_0)|$

Proof:

It is sufficient to show that if $v_i \in N(End(v_0))$, then one of v_{i-1}, v_{i+1} must be in $End(v_0)$

(Extremal Case:

)

Consider a sequence of Pósa rotations which produces $\mathcal{P}'$ with $v_i \sim x$, where $x \neq v_0$ is an end of $\mathcal{P}'$. Let y and z be the left and right neighbors of v_i on $\mathcal{P}'$. If $\{y, z\} = \{v_{i-1}, v_{i+1}\}$, the one of them can be added to $End(v_0)$ by performing a Pósa Rotation on $\mathcal{P}'$. O/w, one of the edges $v_i v_{i-1}$, $v_i v_{i+1}$ have been deleted in a previous Pósa Rotation $\Rightarrow$ one of v_{i-1}, v_{i+1} have been added to $End(v_0)$.

$\square$

Lecture 10:

Recall:

① Want to prove if $pn - (\log n + \log\log n) \to \infty$, then a.a.s $\mathcal{G}(n,p) \in$ HAM

② Posa Rotation:

$|N(End(v_0))| < 2|End(v_0)|$.

Lemma: Assume $pn - (\log n + \log\log n) \to \infty$ and $pn \le 2\log n$. Then there exists a constant $\varepsilon_0 > 0$ s.t. a.a.s there is no $S \subseteq [n]$, $|S| \le \varepsilon_0 n$ and $|N(S)| < 2|S|$.

Proof: Later! (We'll need this for the upcoming proof)

Back to the proof of hamiltonicity of $\mathcal{G}(n,p)$:

Let $\varepsilon_0 > 0$ be the constant in the lemma.

Let Exp denote the property that $\forall S \subseteq [n]$ where $|S| \le \varepsilon_0 n$ $|N(S)| \ge 2|S|$

CNT denote the property that G is connected

Let $f = pn - (\log n + \log\log n)$

By our assumption, $f \to \infty$ as $n \to \infty$.

Let $p_1 = \frac{\log n + \log\log n + f/2}{n}$. Let $G' \sim \mathcal{G}(n, p_1)$. Let G be the random graph by adding $fn/8$ uniformly random edges to G'. We can show (exercise!) that G and $\mathcal{G}(n,p)$ can be coupled so that a.a.s $G \subseteq \mathcal{G}(n,p)$. So, it is sufficient to show that a.a.s G is Hamiltonian

Proof (cont)

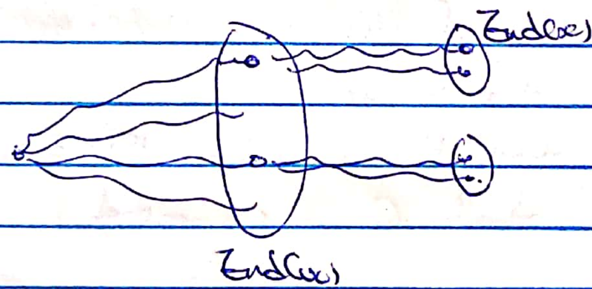
We know that a.a.s $G' \in Exp \cap CN7$.

Take a longest path $P = v_0 v_1 \ldots v_\ell$ in G'. Let $\mathcal{P}$ be the set of paths obtained from Posá rotations with v_0 fixed.

Let $End(v_0)$ denote the other ends of paths in $\mathcal{P}$.

Since $G' \in Exp$ and $|N(End(v_0))| < 2|End(v_0)|$. $\Rightarrow$ $|End(v_0)| \geq \varepsilon_0 n$.

Next, for each $x \in End(v_0)$ let $\mathcal{P}_x$ denote the set of paths obtained via Posá rotations with x fixed and let $End(x)$ be the set of the other ends of $p \in \mathcal{P}_x$.



Let $\xi = \{\{x,y\}, x \in End(v_0), y \in End(x)\}$. ξ is called the set of boosters.

As before, we have $|End(x)| \geq \varepsilon_0 n$ for every $x \in End(v_0)$.
It follows now that $|\xi| \geq \varepsilon_0^2 n^2$. Consider G. We will sprinkle the $\varepsilon n/8$ edges in sequence. We call each sprinkled edge a "trial". A trial is successful if it is in ξ and fails otherwise. If a trial is successful, then since G' is connected, it will result in finding a path longer than P or finding a Hamiltonian cycle if P is a Hamiltonian path

Proof (cont)

Each trial is successful with probability $\ge c\varepsilon_0^2$ (for some constant $c>0$). Only at most n successful trials are needed to guarantee a Hamiltonian cycle in G.

$$\Pr(G \notin HAM) \le \Pr(\mathrm{Bin}(tn/8, c\varepsilon_0^2) \le n) = e^{-\Omega(n)} = o(1)$$

□

~~[illegible]~~: $\Rightarrow \log n \le pn \le 2\log n$.

Lemma: Assume ~~$pn \ge 2\log n$~~. A.a.s.:

(a) Each $S \subseteq [n]$ with $|S| \le \frac{n}{\log n}$ induces at most $3|S|$ edges (Sublinear-sized subgraphs are sparse)

(b) No two vertices with degree <100 are within distance 5

(c) No vertex with degree ≤ 100 lies in a cycle of length ≤ 5

Proof: Exercise!

Back to 2-expander lemma:

Proof:

Let $n^{3/4} \le s \le \varepsilon_0 n$ where ε_0 is a sufficiently small constant

Let $E(S,W)$ be the expected number of points of disjoint sets (S,W) s.t. $|S|=s$, $|W|=w$ and $W=N(S)$

$$E(S,W) = \underbrace{\binom{n}{s}}_{\text{Choose } S}\underbrace{\binom{n-s}{w}}_{\text{Choose } W}\left(1-\underbrace{(1-p)^s}_{\text{Don't see any vertex in } S}\right)^{w}\underbrace{(1-p)^{s(n-s-w)}}_{\text{S cannot see any other vertices}}$$

every vertex in W sees a vertex in S

Complement: See a vertex in S

$$\xrightarrow{\quad} \text{Since } 1-(1-p)^s \le ps$$

$$E(s,w) \le \left(\frac{en}{s}\right)^s \left(\frac{en}{w}\right)^w (ps)^w \exp((1-ps(n-s-w)))$$

$$= \left(\frac{en}{s}\right)^s \left(\frac{en}{w}\right)^w (ps)^w \left(\exp((1-pn(1-\frac{s+w}{n})))\right)^s.$$

(Since $pn \ge \log n$, $ps \le p\varepsilon_0 n < \log n$, if $\varepsilon_0 < \frac{1}{2}$)

$$\le (\log n)^w \left(\frac{e}{s}\right)^s \left(\frac{en}{w}\right)^w \exp(\underbrace{ps}_{<\log n}(s+w))$$

$$\le (\log n)^w \left(\frac{e}{s}\right)^s \left(\frac{en}{w}\right)^w \exp(\underbrace{pn}_{\le 2\log n}.s.\frac{s+w}{n})$$

$$\le (\log n)^w \left(\frac{e}{s}\right)^s \left(\frac{en}{w}\right)^w n^{2s(s+w)/n}$$

Then

$$\sum_{s=n^{3/4}}^{\varepsilon_0 n} \sum_{w=1}^{2s} E(s,w) \le \sum_{s=n^{3/4}}^{\varepsilon_0 n} \sum_{w=1}^{2s} (\log n)^w \left(\frac{e}{s}\right)^s \left(\frac{en}{w}\right)^w n^{2s(s+w)/n}$$

$$\le \sum_{s=n^{3/4}}^{\varepsilon_0 n} (\log n)^{2s} \left(\frac{e}{s}\right)^s \left(\frac{en}{2s}\right)^{2s} n^{2s.3s/n}$$

$$= \sum_{s=n^{3/4}}^{\varepsilon_0 n} \left(\log^2 n \cdot \frac{e}{s} \cdot \frac{e^2n^2}{4s^2} \cdot n^{6s/n}\right)^s$$

$$\le C\sum_{s=n^{3/4}}^{\varepsilon_0 n} \left(\frac{c\log^2 n}{s^3} \cdot n^{2+\frac{6s}{n}}\right)^s$$

$$\le C\sum_{s=n^{3/4}}^{\varepsilon_0 n} \left(\frac{c\log^2 n}{s^3} \cdot n^{2+6\varepsilon_0}\right)^s.$$

Choose ε_0 small so that $2+6\varepsilon_0 < 9/4$. Therefore,

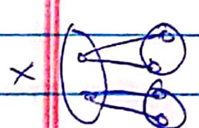
$$\sum_{s=n^{3/4}}^{\varepsilon_0 n} \sum_{w=1}^{2s} E(s,w) \le \sum_{s=n^{3/4}}^{\varepsilon_0 n} n^{-\alpha s} = o(1) \text{ for some } \alpha > 0$$

Now for $S < n^{3/4}$. Assume G is a graph satisfying properties (a)-(c) from the lemma. and $G \in D2$ (min. degree of G is ≥ 2)

We know that a.a.s $G(n,p)$ has these properties. It is sufficient to show that any graph G satisfying these above properties satisfies
$\forall S \subseteq [n]$ where $|S| \leq n^{3/4}$, $|N(S)| \geq 2|S|$

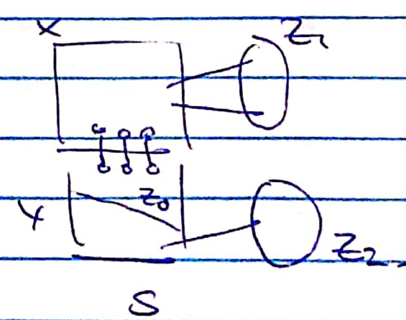
We call a vertex "light" if its degree is at most 100, otherwise call it "heavy". For any $S \subseteq [n]$, let $X \subseteq S$ be the set of light vertices in S and let
$Y = S \setminus X$.

Case I: $Y = \phi$ (i.e. All vertices in S are light)
Then $|N(S)| = |N(X)| \geq 2|X| = 2|S|$. by $G \in D2$ and property (b).



{ All vertex disjoint

Case 2: $Y \neq \emptyset$.
Assume $|N(S)| < 2|S|$
$Z_0 = N(X) \cap Y$
$Z_1 = N(X) \setminus Y$
$Z_2 = N(Y) \setminus (X \cup Z_1)$



Proof (cont)

$\underbrace{}_{\text{By assumption}}$

We have $|N(S)| = |Z_1| + |Z_2| + |Z_3| < 2|S|$
$= 2(|X| + |Y|)$

Since $G \in D2$ and by properties (b) and (c), we will have.

$|Z_0| + |Z_1| \geq 2|X|$

$|Z_0| = e(X, Y) \leq |Y|$

$e(Y, Z_1) \leq |Y|$.

(We will continue after reading week!)

References

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