PACS Part 2, Lecture 4

Connectivity Threshold in ER Graphs

- $p = \frac{\log n}{n}$ is a threshold function for connectivity in Erdős–Rényi graphs
- If $p \le \lambda \frac{\log n}{n}$ with $\lambda < 1$, then there is an isolated vertex a.a.s.
- If $p \ge \lambda \frac{\log n}{n}$ with $\lambda > 1$, then the graph is connected a.a.s.

Size of Connected Components

- we will study the maximum size of a connected component in the disconnected case
- the connected component $C(u)$ of vertex u can be constructed in the following manner
  - We keep a set L of live vertices, a set N of neutral vertices, and a set D of dead vertices.
  - Initially, at time $t = 0$, we have $L(0) = \{u\}$, $N(0) = V \setminus \{u\}$, and $D(0) = 0$.
  - In every step $t \ge 1$, we choose a live vertex $w \in L(t-1)$, move it from L to D, and move all neutral neighbors of w from N to L.
  - The process stops at the earliest time T with $L(T) = \emptyset$.
  - We then have $C(u) = D(T)$ and $|D(T)| = T$.
- Setting $Z(t) = |N(t-1)| - |N(t)|$, we have the recurrence formulas

$$\begin{aligned} |L(t)| &= |L(t-1)| - 1 + Z(t) \\ |N(t)| &= |N(t-1)| - Z(t) \\ |D(t)| &= t \end{aligned}$$

  with $|L(0)| = 1$, $|N(0)| = n - 1$, and $|D(0)| = 0$.
- In particular, $|N(t)| = n - t - |L(t)|$ and

$$|L(t)| = 1 + \sum_{s=1}^{t} (Z(s) - 1) = 1 - t + \sum_{s=1}^{t} Z(s)$$

- Thus: $Z(t) \sim \mathrm{Bin}\big(|N(t-1)|\,,\, p\big) = \mathrm{Bin}\big(n - t + 1 - |L(t-1)|\,,\, p\big)$

Regime with Only Small Components

- set $p = c/n$
- since we always have $|N(t-1)| \le n$, we can upper bound the size T of the connected component by the length of a process $Y(t)$ that satisfies the recurrence for $|L(t)|$ and in which $Z(t) \sim \mathrm{Bin}(n, p)$

- that is, $Y(t) = Y(t-1) - 1 + Z(t)$ and $Y(0) = 1$, with the process stopping when $Y(T) = 0$
- denote by $T_{n,p}$ the length of the original graph process and by $\tilde{T}_{n,p}$ the length of the process $Y(t)$
- if $c < 1$, applying Chernoff's bound, we have:

$$\begin{aligned}
\mathbb{P}(T_{n,p} > t) &\leq \mathbb{P}(\tilde{T}_{n,p} > t) \\
&\leq \mathbb{P}(\text{Bin}(nt, p) \geq t) \\
&= \mathbb{P}(\text{Bin}(nt, p) \geq ct(1 + (1-c)/c)) \\
&\leq \exp\left(-\frac{ct}{3}\frac{(1-c)^2}{c^2}\right)
\end{aligned}$$

- choosing $t = a \log n$ with an appropriate constant a, this is $\leq 1/n^2$
- the union bound then implies that all connected components have size $\leq a \log n$ with high probability

Birth of the Giant Component

- let now $p = c/n$ with $c > 1$
- setting $t^- = b \log n$ and $t^+ = n^{2/3}$, we define for a vertex v:
  - v is *small* if $|C(v)| \leq t^-$
  - v is *big* if $|L_v(t)| \geq \frac{c-1}{2}t$ for all $t^- \leq t \leq t^+$
  - v is *bad* if it is neither big nor small
- If there are no bad vertices, then there is at most one big component (of super-logarithmic size).

  Proof: For any pair (u, v) of big vertices, we have:

$$\begin{aligned}
\mathbb{P}\big(C(u) \neq C(v)\big) &\leq \mathbb{P}\big(\text{there are no edges between } L_u(t^+) \text{ and } L_v(t^+)\big) \\
&\leq (1-p)^{\left(\frac{c-1}{2}t^+\right)^2} \leq \exp\left(-\frac{c}{n}\frac{(c-1)^2}{4}n^{4/3}\right) \\
&= \exp\left(-\frac{c(c-1)^2}{4}n^{1/3}\right) = O(1/n^3)
\end{aligned}$$

  The union bound then shows that there is no such pair with high probability.
- If there are no bad vertices, then there is a giant component (of linear size).

  Proof: Let X be the number of small vertices. We show that

  We have:

$$\mathbb{P}\left(\tilde{T}_{n,p} \leq t^-\right) \leq \mathbb{P}\left(T_{n,p} \leq t^-\right) \leq \mathbb{P}\left(\tilde{T}_{n-t^-,p} \leq t^-\right)$$

We will later show that, for $n \geq \infty$, the two outer terms converge to the same quantity p_e, which then shows $\mathbb{E}X = (p_e + o(1))n$. But first we will study the variance of N_s and apply Chebyshev's inequality.

Define the indicator variable $X_u = 1$ iff u is a small vertex. Then $X = \sum_{u \in V} X_u$. We have

$$\mathrm{Var}(X) \leq \mathbb{E}X^2 = \mathbb{E}X + \sum_{v} \mathbb{P}(X_v = 1) \sum_{u \neq v} \mathbb{P}(X_u = 1 \mid X_v = 1)$$

and

$$\sum_{u \neq v} \mathbb{P}(X_u = 1 \mid X_v = 1) = \sum_{\substack{u \neq v \\ u \in C(v)}} \mathbb{P}(X_u = 1 \mid X_v = 1) + \sum_{\substack{u \neq v \\ u \notin C(v)}} \mathbb{P}(X_u = 1 \mid X_v = 1)$$
$$\leq t^- + (p_e + o(1))n$$

which gives

$$\mathrm{Var}(X) \leq \mathbb{E}X + (p_e + o(1))^2 n^2 = \mathbb{E}X + o\left((\mathbb{E}X)^2\right)$$

Applying Chebyshev's inequality with $a = \delta \mathbb{E}X$ gives

$$\mathbb{P}\left(X/n \geq p_e\left(1+\delta\right)\right) \leq \frac{1}{\delta}\left(\frac{1}{\mathbb{E}X} + o(1)\right) = o(1)$$

We can even let $\delta \to 0$ very slowly.

- We left two tasks open: show that there are no bad vertices and the convergence to p_e
- We first show that there are no bad vertices with high probability:

  Let v be a bad vertex. Then there is some $t^- < t \leq t^+$ with $L_v(t) < \frac{c-1}{2}t$. We have

$$\mathbb{P}\left(L_v(t) \leq \frac{c-1}{2}t\right) \leq \mathbb{P}\left(\mathrm{Bin}\left(t\left(n - t - \frac{c-1}{2}t\right), \frac{c}{n}\right) \leq \frac{c-1}{2}t\right)$$
$$\leq \mathbb{P}\left(\mathrm{Bin}\left(t\left(n - \frac{c+1}{2}t^+\right), \frac{c}{n}\right) \leq \frac{c-1}{2}t\right)$$

  We have $\mu = ct\left(1 - \frac{c+1}{2n}t^+\right)$. Choosing δ such that $(1-\delta)\mu = \frac{c+1}{2}t$, we get $\delta = \frac{c-1}{2c} + o(1)$ as $n \to \infty$ and, by Chernoff's bound,

$$\mathbb{P}\left(L_u(t) \leq \frac{c-1}{2}t\right) \leq \exp\left(-\left(\frac{(c-1)^2}{8c} + O(n^{-1/3})\right)t\right)$$

  and thus

$$\mathbb{P}\left(u \text{ is bad}\right) \leq n^{2/3} \exp\left(-\left(\frac{(c-1)^2}{8c} + O(n^{-1/3})\right) b \log n\right)$$

  which is $O(1/n)$ for an appropriate choice of b.

Galton–Watson Process

- To show convergence of $\mathbb{P}(\tilde{T}_{n,p} \le t^-)$ and $\mathbb{P}(\tilde{T}_{n-t^-,p} \le t^-)$ to p_e, we introduce a parallel version of the process: the Galton–Watson branching process.
- There, we set $Y_0 = 1$, and $Y_{n+1} = \sum_{i=1}^{Y_n} Z_i^{(n)}$ where the $Z_i^{(n)}$ are i.i.d.
- Using the probability-generating function $g_X(s) = \mathbb{E}s^X$ for integer-valued random variables X, we get $g_{X_{n+1}}(s) = g_Z(g_{X_n}(s))$.
- We have $\mathbb{P}(X_n = 0) = g_Z^n(0)$ and thus $p_e = g_Z(p_e)$ for the extinction probability $p_e = \mathbb{P}(\exists n\colon X_n = 0)$.
- For a binomial random variable Z, we have $g_Z(s) = \big((1-p) + ps\big)^n$.
- Both probability-generating functions $\left(1 + \frac{c(s-1)}{n}\right)^n$ and $\left(1 + \frac{c(s-1)}{n}\right)^{n-t^-}$ converge to $e^{c(1-s)}$.
- Let p_e be the nontrivial ($s \neq 1$) solution of the equation $s = e^{c(s-1)}$. This solution is unique and asmptotically equal to the extinction probability.
- Using the approximation $|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \le np^2$ for all sets A of nonnegative integers whenever $X \sim \mathrm{Bin}(n,p)$ and $X \sim \mathrm{Poi}(np)$, we can show convergence of the finite-cutoff probabilities to p_e.