

5.1 Thresholds for Erdős-Rényi random graphs (cont'd)

Recall that in the last lecture, we used average analysis to give a derivation of threshold functions for two properties of Erdős-Rényi random graphs $G(n, p)$. We now present arguments that illustrate how this heuristic, along with the so-called *second moment method*, can be used to obtain rigorous proofs of these transitions.

Theorem 5.1. *The threshold for the existence of an edge (i.e., a tree with 2 nodes) is $t(n) = n^{-2}$.*

Proof: Let N denote the number of edges. There are two claims to prove:

(a) $\mathbb{P}(N \geq 1) \rightarrow 0$ as $n \rightarrow \infty$ if $p(n) = o(n^{-2})$

(b) $\mathbb{P}(N \geq 1) \rightarrow 1$ as $n \rightarrow \infty$ if $p(n) = \omega(n^{-2})$.

By Markov's inequality, we have $\mathbb{P}(N \geq 1) \leq \mathbb{E}[N]$. Therefore, if $p(n) = o(n^{-2})$ then $\mathbb{E}[N] = \binom{n}{2}p(n) = \Theta(n^2p(n)) \rightarrow 0$ and we have shown (a) holds. To prove (b), notice that

$$\begin{aligned} \mathbb{P}(N \geq 1) &= 1 - \mathbb{P}(N = 0) \\ &= 1 - (1 - p(n))^{\binom{n}{2}} \\ &= 1 - \exp\left[\binom{n}{2} \log(1 - p(n))\right]. \end{aligned}$$

From the Taylor series representation of $f(x) = \log(1 - x)$ about $x = 0$, we observe that

$$\log(1 - p(n)) = -\sum_{k=1}^{\infty} \frac{p(n)^k}{k}.$$

Now suppose $p(n) = \omega(n^{-2})$. Since $\binom{n}{2} = \Theta(n^2)$, we find that

$$-\binom{n}{2} \left(p(n) + \frac{p(n)^2}{2} + \dots \right) \rightarrow -\infty$$

so $\exp\left[\binom{n}{2} \log(1 - p(n))\right] \rightarrow 0$ as $n \rightarrow \infty$. Therefore, (b) holds as well. \square

5.1.1 Second-moment method

Proving a transition for the appearance of a tree with k nodes will require bounds on the variance of a sum of random variables. While we only illustrate this method in the case $k = 3$, a similar argument can be used to prove that $t(n) = n^{-k/(k-1)}$ is a threshold for any given k . We leave this to the interested reader.

Theorem 5.2. *The threshold for the existence of a tree with 3 nodes is $t(n) = n^{-3/2}$.*

To show this first requires the following two lemmas:

Lemma 5.3. (Chebyshev's inequality) *For any random variable X ,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq k) \leq \frac{\text{Var}(X)}{k^2}.$$

Proof: Apply Markov's inequality to the non-negative random variable $|X - \mathbb{E}[X]|$. \square

Lemma 5.4. *If $X = \sum_{i=1}^n Y_i$ where Y_1, \dots, Y_n are random variables such that $0 \leq Y_i \leq 1$ for every i , then*

$$\text{Var}(X) \leq \mathbb{E}[X] + \sum_{i \neq j} \text{Cov}(Y_i, Y_j).$$

Proof: To begin,

$$\begin{aligned} \text{Var}(X) &= \text{Var}\left(\sum_i Y_i\right) \\ &= \sum_i \text{Var}(Y_i) + \sum_{i \neq j} \text{Cov}(Y_i, Y_j). \end{aligned}$$

Now observe that

$$\text{Var}(Y_i) = \mathbb{E}[Y_i^2] - (\mathbb{E}[Y_i])^2 \leq \mathbb{E}[Y_i^2] \leq \mathbb{E}[Y_i]$$

since $0 \leq Y_i \leq 1$ gives $Y_i^2 \leq Y_i$. Substituting $\sum_i \mathbb{E}[Y_i] = \mathbb{E}[X]$, we conclude the proof. \square

We will now be able to prove Theorem 5.2:

Proof: Let N be the number of trees with 3 nodes. Again we must show

- (a) $\mathbb{P}(N \geq 1) \rightarrow 0$ as $n \rightarrow \infty$ if $p(n) = o(n^{-3/2})$
- (b) $\mathbb{P}(N \geq 1) \rightarrow 1$ as $n \rightarrow \infty$ if $p(n) = \omega(n^{-3/2})$.

As before, the proof of (a) follows by applying Markov's inequality and using the average analysis from the previous lecture. To prove (b), note that we cannot simply use Markov's

inequality again since the bound would be in the wrong direction. Instead, we begin by noting the following containment of events:

$$\{N = 0\} \subseteq \{|N - \mathbb{E}[N]| \geq \mathbb{E}[N]\}.$$

Combining with Chebyshev's inequality yields

$$\mathbb{P}(N = 0) \leq \frac{\text{Var}(N)}{\mathbb{E}[N]^2}.$$

Our plan is to bound the right hand side of this inequality by applying the lemma above. To do so, we must express N as a sum of $[0, 1]$ -valued random variables. For an unordered pair of nodes $e = (i, j)$, let $I_e = I_{(i,j)}$ denote the indicator function of the event that an edge exists between (i, j) . Then we have

$$N = \sum_{(i,j),(j,k)} I_{(i,j)} I_{(j,k)} = \sum_{\{e,e'\} \text{ adj.}} I_e I_{e'}.$$

where the last sum is taken over pairs that share a node (i.e., over all adjacent edges). By the previous lemma,

$$\text{Var}(N) \leq \mathbb{E}[N] + \sum_{\{e,e'\} \neq \{f,f'\}} \text{Cov}(I_e I_{e'}, I_f I_{f'}).$$

Here, $\{e, e'\}$ and $\{f, f'\}$ are unordered pairs of adjacent edges. Now if $\{e, e'\} \cap \{f, f'\} = \emptyset$ then $\text{Cov}(I_e I_{e'}, I_f I_{f'}) = 0$. How many non-zero terms remain? To begin with, there are $\binom{n}{3} \binom{3}{2}$ pairs of adjacent edges $\{e, e'\}$. Given $\{e, e'\}$ there are two possibilities for an edge f such that either $f = e$ or $f = e'$. Finally, there are $n - 2$ remaining choices for an additional edge f' adjacent to f . Altogether, this gives $\binom{n}{3} \binom{3}{2} 2(n - 2) = \Theta(n^4)$ non-zero terms of the form $\text{Cov}(I_e I_{e'}, I_f I_{f'})$. For these terms, we have

$$\text{Cov}(I_e I_{e'}, I_f I_{f'}) \leq \mathbb{E}[I_e I_{e'} I_f I_{f'}] = p(n)^3.$$

To see that the last equality holds, suppose that $f = e'$. Then $I_e I_{f'} = I_e^2 = I_e$, so

$$\mathbb{E}[I_e I_{e'} I_f I_{f'}] = \mathbb{E}[I_e I_{e'} I_{f'}] = \mathbb{E}[I_e] \mathbb{E}[I_{e'}] \mathbb{E}[I_{f'}] = p(n)^3.$$

Combining these computations with the lemma above and the fact that $\mathbb{E}[N] = \Theta((n^{3/2}p(n))^2)$ (shown in the last lecture), we obtain

$$\begin{aligned} \mathbb{P}(N = 0) &\leq \frac{\text{Var}(N)}{\mathbb{E}[N]^2} \\ &\leq \frac{\Theta((n^{3/2}p(n))^2) + \Theta(n^4 p(n)^3)}{\Theta((n^{3/2}p(n))^4)} \\ &\leq C \left(\frac{1}{(n^{3/2}p(n))^2} + \frac{1}{n^2 p(n)} \right) \end{aligned}$$

for some constant $C > 0$. The right hand side converges to 0 as $n \rightarrow \infty$ if $p(n) = \omega(n^{-3/2})$. \square

5.1.2 Giant component, connectivity, and diameter

There are several significant phase transitions in the Erdős-Rényi model, which we now summarize.

Let C_i denote the i^{th} largest connected component of the graph $G(n, p)$. A dramatic change in the structure of the graph occurs at the threshold $t(n) = n^{-1}$. Below the threshold, all components of $G(n, p)$ are disconnected trees. Above it, we observe the formation of a single *giant component* which contains a positive fraction of the nodes:

- $|C_1| = O(\log n)$, if $p(n) = o(n^{-1})$
- $|C_1| = \Theta(n^{2/3})$ and $|C_2| = O(\log n)$, if $p(n) = \Theta(n^{-1})$
- $|C_1| = \Theta(n)$ and $|C_2| = O(\log n)$, if $p(n) = \omega(n^{-1})$.

If we increase $p(n)$ past the threshold $t(n) = n^{-1}$, we observe transitions in the connectedness and diameter of the random graph as well. Below the threshold $t(n) = n^{-1}(\log n + \omega(1))$, the graph is disconnected (i.e., it has isolated nodes), while above it the graph is connected. Furthermore, we should expect that by increasing $p(n)$ further the diameter of the connected graph will become small. Indeed, above the threshold $t(n) = n^{-1}\omega(\log(n))$ the diameter of the graph is $O(\log n)$ and $G(n, p)$ has the small-world property. We will discuss these transitions at length in the next lecture.

References

1. Acemoglu, D., & Ozdaglar A. (2009). 6.207/14.15, Networks, Fall 2009 [Lecture notes]. Retrieved from:
<http://economics.mit.edu/faculty/acemoglu/courses/>
2. Chaintreau, A. (2012). COMS 4995-1, Introduction to Social Networks, Fall 2012 [Lecture notes]. Retrieved from:
<http://www.cs.columbia.edu/~augustin/>
3. Jackson, M. O. (2010). Social and economic networks. Princeton University Press.

Last edited: March 21, 2013