

Before we start with the lecture, let us review two simple results in probability that will be very useful in this and future lectures.

The first one is Chernoff bounds for binomial random variables:

**Proposition 1.1** *If  $X \in \text{Binomial}(n, p)$  and  $0 \leq \epsilon \leq 1$ , then*

$$\Pr(x > np(1 + \epsilon)) \leq \exp\left(\frac{-\epsilon^2 np}{3}\right)$$

and

$$\Pr(x < np(1 - \epsilon)) \leq \exp\left(\frac{-\epsilon^2 np}{2}\right)$$

For proof, see [5] Chapter 4, Section 1.

The second result is on branching processes. Let  $X$  be a random variable that takes non-negative integer values. The *Galton-Watson branching process*  $Y$  determined by  $X$  is defined as follows: Set  $Y_0 = 1$ . At step  $i$ , set  $Y_i = Y_{i-1} + Z_i - 1$ , where  $Z_i \stackrel{D}{=} X$ . If  $Y_i > 0$ , increment  $i$  and repeat from step 2, otherwise set  $Y_j = 0$  for all  $j > i$ .

Denote the probability of extinction as  $\rho_X = \Pr(Z < \infty)$ , where  $Z = \sum_{i \geq 0} Z_i$  is the total number of offspring produced. Then the following property of the branching process defined by  $X$  holds:

**Proposition 1.2** *For  $EX \leq 1$ , we have  $\rho_X = 1$ , unless  $\Pr(X = 1) = 1$ . If  $EX > 1$  and  $\Pr(X = 0) > 0$ , then  $\rho_X = x_0$ , where  $x_0$  is the unique solution of the equation  $f_X(x) = x$  which belongs to the interval  $(0, 1)$ , where  $f_X : [0, 1] \rightarrow \mathbb{R}$  is the probability generating function of  $X$ .*

## 1. Evolution of Random Graphs

The focus of this lecture is on the emergence of the giant component in the Erdős-Rényi random graph model  $G(n, p)$ . In this model, a graph  $G \in G(n, p)$  on  $n$  vertices is chosen by placing an edge between each pair of vertices independently with probability  $p$ . A series of seminal paper by Erdős and Rényi from 1959-61 [1, 2, 3] helped to develop the theory behind this model.

One of the key observations about Erdos-Renyi graphs is that several important properties of the graph had “threshold values”, above which they are almost surely true and below which they are almost surely false. Properties with threshold values are said to exhibit a “phase transition”, due to the analogous principle observed in physical systems.

One such property is the presence or absence of a “giant component”, i.e. a component that has  $O(n)$  vertices. We will show that there exists such a threshold at  $np = 1$ .

**Theorem 1.1** *Let  $np = c$ , where  $c > 0$  is a constant.*

- *If  $c < 1$ , then with high probability all connected components of  $G \in G(n, p)$  are of size  $O(\log n)$ .*
- *If  $c > 1$ , with high probability,  $G \in G(n, p)$  has a unique giant connected component of size  $(\beta + o(n))n$  for constant  $\beta = \beta(c)$ . The sizes of the rest of the components will be of  $O(\log n)$ .*

**Proof:** We can construct each connected component of  $G$  as a variation of the Galton-Watson branching process as follows. Pick a vertex  $w \in G$ . Initialize the set of “live” vertices  $L_0 \leftarrow \{w\}$ , and the set of “saturated” or “dead” vertices  $D_0 \leftarrow \emptyset$ . At step  $i$ , if  $L_{i-1} \neq \emptyset$ , pick a vertex  $v_i \in L_{i-1}$ . Set  $L_i \leftarrow L_{i-1}/v_i$ . Set  $D_i \leftarrow D_{i-1} \cup \{v_i\}$ . Find all of the neighbors of  $v_i$  in  $G/\{D_{i-1} \cup L\}$  add them to  $L_i$ .

Let  $Y_i = |L_i|$ . Then  $Y_0 = 1$ , and  $Y_i = Y_{i-1} + Z_i - 1$ , where  $Z_i$  is the number of new neighbors added in step  $i$ . If  $Y_{i-1} > 0$ , then  $Z_i$  is distributed according to  $Bin(n - i + 1 - Y_i, p)$ .

Now, assume that  $c < 1$ . We can bound  $Z_i$  from above by  $Z^+ = Bin(n, p)$ . The probability that  $Y_k > 0$  is upper bounded by

$$\begin{aligned} \Pr\left(\sum_{i=1}^k Z_i \geq k - 1\right) &\leq \Pr\left(\sum_{i=1}^k Z_i^+ \geq k - 1\right) \\ &= \Pr\left(\sum_{i=1}^k Z_i^+ \geq ck + (1 - c)k - 1\right) \\ &\leq \exp\left(-\frac{((1 - c)k - 1)^2}{3ck}\right) \\ &\leq \exp\left(-\frac{(1 - c)^2}{4}k\right), \end{aligned} \tag{1}$$

where the second to last inequality is by Chernoff bound. Plugging in  $k = 5 \log n / (1 - c)^2$  gives the probability that any particular vertex is part of a component of size  $k$  or greater as  $\leq n^{-5/4}$ . Applying the union bound gives Theorem (1.1).

Now let  $c > 1$ . Let  $k^- = \frac{16c}{(c-1)^2} \log n$  and  $k^+ = n^{2/3}$ .

**Proposition 1.3** For  $k^- \leq k \leq k^+$ , w.h.p. the process either dies before time  $k^-$  or reaches time  $k$  with at least  $k(c-1)/2$  live nodes.

We bound the  $Z_i$ 's from below by  $Z_i^- \sim \text{Bin}(n - k^+, p)$ . So given the process reaches  $k^-$ , the probability of it having less than  $k(c-1)/2$  live nodes before step  $k^+$  is at most

$$\sum_{k=k^-}^{k^+} \Pr\left(\sum_{i=1}^k Z_i^- \leq k + \frac{c-1}{2}k\right) \leq \sum_{k=k^-}^{k^+} \exp\left(-\frac{(c-1)^2 k^2}{9ck}\right) \quad (2)$$

$$\leq k^+ \exp\left(-\frac{(c-1)^2 k^-}{9c}\right) = O(1/n) \quad (3)$$

Taking the union bound gives the proposition.

So far, we know that, w.h.p., every connected component is either of size bigger than or equal to  $n^{2/3}$  or is smaller than  $k^-$ . To show the uniqueness of the giant component, suppose that processes started at node  $u$  and  $w$  are both alive but have not intersected by step  $k^+$ . Note that, w.h.p., each process has generated at least  $k^+(c-1)/2$  live nodes. Therefore, the probability that in the next step the processes do not intersect is at most

$$(1-p)^{\left(k^+(c-1)/2\right)^2} = (1-c/n)^{\left((c-1)/2\right)^2 n^{4/3}} \leq \exp\left(-\frac{c(c-1)^2}{4} n^{1/3}\right) = o(1/n^2).$$

To establish the existence of the giant connected component, we estimate the number of vertices that are in small components. Note that the probability of extinction before  $k^-$ ,  $\rho(n, p)$ , is bounded from above by  $\rho_+(n, p)$ , where  $\rho_+(n, p)$  is governed by the branching process with new nodes added according to  $\text{Bin}(n - k^-, p)$ .  $\rho(n, p)$  is bounded from below by the process  $\rho_-(n, p) + o(1)$ , governed by  $\text{Bin}(n, p)$ , where the extra term comes from the probability the process has more than  $k^-$  vertices. It can be shown using Proposition 1.2 (see [4] example 5.3 p.108) that  $\rho_-$  and  $\rho_+$  converge to a constant  $\beta_c = \beta(c)$ , where  $\beta(c)$  is the unique solution to the equation

$$\beta + e^{-\beta c} = 1.$$

Therefore, the expectation of number of vertices  $Y$  in small components is  $(1 - \beta_c + o(1))n$ . We can bound the variance of this process by

$$\mathbb{E}(Y(Y-1)) \leq n\rho(n, p)(k^- + n\rho(n - O(k^-), p)) = (1 + o(1))(\mathbb{E}Y)^2.$$

From Chebyshev's inequality follows that the number of vertices in small components is  $(1 - \beta_c + o(1))n$ , proving the theorem. ■

The above theorem statement and proof are adopted from section 5.2 of [4].

**Remark** There is actually a double jump, or a double phase transition happening here. One can observe yet another phase transition by increasing  $p$  more slowly. In that case the maximum component size jumps from  $O(\log n)$  to  $n^{2/3}$ . For getting a giant component of size  $n^{2/3}$ , you can set  $np = 1 + \frac{\lambda}{n^{(1/3)}}$  for some  $\lambda > 0$ .

## References

- [1] P. Erdős and A. Rényi. On random graphs. *Publ. Math. Debrecen*, 6(290), 1959.
- [2] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61, 1960.
- [3] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12:261–267, 1961.
- [4] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. John Wiley New York, 2000.
- [5] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.