October 10, 2005
prepared by Paul Constantine

1. Evolution of Random Graphs

In this lecture, we will talk about random graph model known as Erdos-Renyi graph or $G(n,p)$. In this model, given a set of n vertices, we place an edge between pairs of vertices with probability p independently at random. In 1960 Erdos and Reyni observed that these graphs have very interesting properties. For example, when $np < 1$, with high probability G consists of a few small connected components of size around $\mathcal{O}(\log n)$. But at $np = 1 + \epsilon$, with high probability G contains a giant connected component of size linear in n. This phenomenon is called phase transition. As np increases, the giant connected component subsumes all the smaller surrounding components. As np gets close to $\log(n)$, G begins to look like a large connected component with a few isolated vertices. When np reaches $\log(n)/2$, the graph becomes connected almost surely.

In our analysis of these graphs, we will use the following branching process: Start at a vertex v. Find all neighbors of v, $v_1, v_2, \cdots v_n$ and add them to the list. Thus call v saturated. Now at every step take the next unsaturated node in the list and add its neighbors to the list that are not already in the list. The distribution of the number of nodes that an unsaturated node adds to a list (the number of children of that node) is

$$X_k \in \text{Binomial}(n-k, p)$$

where k is the number of vertices that are already in the list. For small k, we have an approximation

$$X_k \in \text{Binomial}(n, p).$$

Before continuing, we state the Chernoff bounds, which we will use later.

Proposition 1.1 *If $X \in Binomial(n,p)$, then*

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

and

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

For proof of this, see Motwani and Raghavan's *Randomized Algorithms*, Chapter 4, Section 1.

Proposition 1.2 *Let $c = np$. For $c < 1$, the size of the components in $G(n, p)$ is at most $\mathcal{O}(\log n)$.*

Proof: The probability that vertex v is in a component of size at most k is at least

$$\Pr(X_1 + X_2 + \cdots + X_k \geq k - 1)$$

where X_i can be bounded from above by $X_i^+ \in \text{Binomial}(n, p)$. Therefore

$$\begin{aligned}
\Pr(X_1 + X_2 + \cdots + X_k \geq k - 1) &\leq \Pr(\Sigma_i X_i^+ \geq k - 1) \\
&= \Pr(\Sigma_i X_i^+ \geq ck + (1 - c)k - 1) \\
&\leq \exp\left(\frac{-(1 - c)^2}{2k}\right)
\end{aligned}$$

Also the probability that there is no vertex in a component of size larger than k is

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

Let $k = 3/(c - 1)^2 \log n$. Then

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

In other words, the probability that there is no vertex in a component of size larger than k becomes very small. ∎

For $c > 1$, if c is the expected number of children of a vertex k in the branching process, then the probability of survival for that process is β such that $\exp(-\beta c) + \beta = 1$.

Theorem 1.1 *For $G(n, p)$, if $np = c > 1$, the graph contains a unique giant component of size βn, almost surely. The sizes of the rest of the components are $\mathcal{O}(\log n)$.*

Proof: Let $K^+ = n^{(2/3)}$ and $K^- = (16c \log n)/(c - 1)^2$. We will show that there are no components with size between K^+ and K^-.

1 Fix the root v. Observe that the probability that the branching process starting at v and reaching K^- vertices dies out before K^+ or ends up with fewer than $\frac{k(c-1)}{2}$ unsaturated vertices at that time is

$$\sum_{k=K^-}^{K^+} Pr(X_1 + X_2 + \cdots + X_k \leq k + \frac{(c - 1)k}{2})$$

but each X_i can be bounded from below by $X_i^- \in \text{Binomial}(n - K^+, p)$. So

$$\begin{aligned} & \textstyle\sum_{k=K^-}^{K^+} \Pr(X_1 + X_2 + \cdots + X_k \le \frac{k+(c-1)k}{2}) \\ \le \; & \textstyle\sum_{k=K^-}^{K^+} \Pr(\sum_{i=1}^{k} X_i \le ck - (ck/2 - k/2)) \\ \le \; & \textstyle\sum_{k=K^-}^{K^+} \exp\left(\frac{-(c-1)^2 k}{9c}\right) \\ \le \; & K^+ \exp\left(\frac{-(c-1)^2 K^-}{9c}\right) \quad \text{(by plugging in Chernoff bound)} \\ = \; & o(1) \end{aligned}$$

So either the branching process dies out before K^- or it has many unsaturated vertices at K^+.

2 Next we claim that the number of giant connected components is no more than 1. Suppose there are two giant connected components. It means that for two vertices v and u the branching process survived until time K^+. At that time both branchings have $\frac{(c-1)k}{2} \times \frac{(c-1)l}{2}$ vertices on the boundaries. But

$$(1-p)^{\frac{(c-1)k}{2} \times \frac{(c-1)l}{2}} \le \exp\left(\frac{-(c-1)cn^{(1/3)}}{4}\right) \le o(1/n^2).$$

In other words, the two giant components connect to each other with high probability.

3 Finally we claim that the size of the giant component is linear in n. To show this, we bound the number of vertices in smaller components. The probability of the branching process dying out is $1 - \beta < 1$. So the expected number of vertex branchings dying out is $(1-\beta)n$. The rest of the nodes create one giant connected component. The expected size of this giant component is $(1 - \beta + o(1))n$. Using Chebyshev's inequality one can show that the size of the giant component is of the same order with high probability.

■

Remarks:

- Actually, there are two phase transitions if you increase p much more slowly: (i) one from all components of size $\log n$ to the formation of a component of size $n^{\frac{2}{3}}$ and (ii) another one to a giant component of linear size. To see the giant component of size $n^{\frac{2}{3}}$ you can set $np = 1 + \frac{\lambda}{n^{(1/3)}}$.
- By using Chernoff bounds, one can show that when $np = (\log n)/2$ then graph becomes connected with diameter $\mathcal{O}(\log n)$. Moreover, the diameter of the large connected component goes is $O(\log n)$.