October 3, 2007

Prepared by Adam Guetz and Paul Constantine

1. Evolution of Random Graphs

In this lecture, we will talk about the emergence of the giant component in the Erdös-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 Erdös and Rényi from 1959-61 [1, 2, 3] helped to develop the theory behind this model. One of their key contributions was to show 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 principal 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 $np = c < 1$. For $G \in G(n, p)$, w.h.p. the size of the largest connected component is $O(\log n)$.*

Theorem 1.2 *Let $np = c > 1$. For $G \in G(n, p)$, w.h.p. G has a giant connected component of size $(\beta + o(n))n$ for constant $\beta = \beta_c$. w.h.p, the sizes of the rest of the components are $O(\log n)$.*

Before presenting the proof, we will use the following argument to give an intuition about the main reason behind the phase transition. Let X be a random variable take 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 $Z_i \stackrel{\mathcal{D}}{=} X$. 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:

Theorem 1.3 *For* $\mathrm{E}X \le 1$*, we have* $\rho_X = 1$*, unless* $\Pr(X = 1) = 1$*. If* $\mathrm{E}X > 1$ *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* $f_X : [0, 1] \to I\!R$ *is the probability generating function of* X*.*

This provides the main motivation for the proof of Theorems 1.1 and 1.2.

Proof: We can discover the component structure of G step by step using the following procedure. Pick a vertex $w \in G$. Initialize the set of "live" vertices $L_0 \leftarrow \{w\}$, and the set of "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 above by $Z^+ = Bin(n, p)$. The probability that $Y_k > 0$ is bounded above by

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

where the second to last inequality is due to Chernoff bound (see proposition (1.2) below). Plugging in $k = 3 \log n/(1 - c)^2$ gives the probability that any particular vertex is part of a component of size k or greater as $\le n^{-3/2}$. 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 1.1 *For* $k^- \le k \le k^+$*, with high probability 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 $Z_i^- \sim Bin(n - k^+, p)$. So given the process reaches k^-, the probability of it having less than $k(c - 1)/2$ live nodes before step k^+ is bounded above by

$$\begin{aligned}
\sum_{k=k^-}^{k^+} \Pr(\sum_{i=1}^{k} Z_i^- \le k + \frac{c - 1}{2} k) &\le \sum_{k=k^-}^{k^+} \exp\left(-(c - 1)^2 k^2/9ck\right) && (5)\\
&\le k^+ \exp\left(-(c - 1)^2 k^-/9c\right) = O(1/n) && (6)
\end{aligned}$$

Taking the union bound gives the proposition.

So far, we know that every connected component is either at least size $n^{2/3}$ or is smaller than k^- w.h.p. 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 at for each process, there are at least $k^+(c-1)/2$ live nodes w.h.p. The probability that in the next step the processes won't intersect is then

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

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 $Bin(n-k^-,p)$. $\rho(n,p)$ is bounded from below by the process $\rho_-(n,p)+o(1)$, governed by $Bin(n,p)$, where the extra term comes from the probability the process has more than k^- vertices. It can be shown (see [4] example 5.3 p.108) that ρ_- and ρ_+ 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

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

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

Remarks:

- The above theorem statement and proof are adapted from section 5.2 of [4].
- There is another phase transitions if you increase p more slowly, from maximum component size of $\log n$ to the formation of a component of size $n^{\frac{2}{3}}$. The threshold value to get the giant component of size $n^{\frac{2}{3}}$ is at $np = 1 + \frac{\lambda}{n^{(1/3)}}$.
- Recall the Chernoff bounds for binomial random variables:

  Proposition 1.2 *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, see [5] Chapter 4, Section 1.

References

[1] P. Erdös and A. Rényi. On random graphs. *Publ. Math. Debrecen*, 6(290), 1959.

[2] P. Erdös and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.

[3] P. Erdös 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.