Lecture 06 — 01/31/13

Lecturer: Ravi Srinivasan *Scribe: Tianran Geng*

6.1 Proofs for emergence of giant component

We now sketch the main ideas underlying the formation of a giant component in the Erdős-Rényi model, referring the interested reader to the references for additional details.

To begin, let $\lambda = np$ denote the average degree as $n \to \infty$ and $p = p(n) \to 0$. First we consider the subcritical regime $\lambda < 1$ and the supercritical regime $\lambda > 1$, for which we can use branching processes to estimate the size of the largest component. For the critical case $\lambda = 1$, we provide an elegant scaling argument that utilizes more advanced topics in probability theory: martingales and Brownian motion.

6.1.1 Subcritical regime ($\lambda < 1$)

As before, we let C_i be the i^{th} largest component of $G(n, p)$. In addition, we will say that an event A happens *almost surely* (denoted a.s.) if $\mathbb{P}(A) = 1$.

Theorem 6.1. *If $\lambda < 1$, a graph in $G(n, p)$ has no connected component of size larger than $O(\log n)$ a.s. That is, there exists a constant α such that $\mathbb{P}(|C_1| > \alpha \log(n)) \to 0$ as $n \to \infty$.*

Proof: (Sketch) Recall the Chernoff bound on the tail distributions in the Galton-Watson branching process:

$$\mathbb{P}(|\mathcal{T}| > k) \le e^{-kh(a)}, \qquad h(x) = \sup_{\theta \ge 0} \left\{ \theta x - \log \mathbb{E}\left[e^{\theta \xi}\right] \right\}. \tag{6.1}$$

We examine the size of the largest component as follows. Let v be some node in the graph $G(n, p)$. We will bound the size of the component to which v belongs by comparing its breadth-first spanning tree to a Galton-Watson branching process $\mathcal{T}$ with offspring distribution $\xi \sim \text{Binomial}(n-1, p) \approx \text{Poisson}(\lambda)$. This is seen as follows.

Let w be a neighbor of v and j be the number of neighbors of v. We see that v has $n-j-1$ neighbors other than w, and each has a probability p of being w's neighbor. We can view these nodes as possible children of w in the spanning tree. Note that w has $n-j-1$ possible children, less than the $n-1$ possible children we had from the branching process $\mathcal{T}$. Therefore, repeating this procedure for every node in the breadth-first spanning tree of

v, we obtain a modified Galton-Watson process in which the number of possible children decreases with each generation and thus dominated in size by $\mathcal{T}$.

Now denote by $C(v)$ be the connected component to which v belongs. According to the preceding argument,

$$\mathbb{P}\left(|C(v)| > k\right) \leq \mathbb{P}\left(|\mathcal{T}| > k\right) \leq e^{-kh(1)}$$

where h is the rate function corresponding to an approximately Poisson(λ) offspring distribution. But what is h? Assuming ξ is exactly Poissonian, we have

$$\mathbb{E}\left[e^{\theta\xi}\right] = \sum_{k=0}^{\infty} e^{\theta k}\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=0}^{\infty}\frac{(\lambda e^{\theta})^k e^{-\lambda}}{k!} = e^{-\lambda}e^{\lambda e^{\theta}} = e^{\lambda(e^{\theta}-1)}.$$

Now set $g(\theta) = \theta x - \log \mathbb{E}\left[e^{\theta\xi}\right] = \theta x - \lambda(e^{\theta} - 1)$. Note that $g'(\theta) = x - \lambda e^{\theta} = 0$ at $\hat{\theta} = \log(x/\lambda)$. Since $g''(\theta) = -\lambda e^{\theta} < 0$ for all $\theta \geq 0$, we know that $g(\theta)$ reaches its maximum at $\hat{\theta}$. Therefore,

$$h(x) = \sup_{\theta\geq 0} g(\theta) = g(\hat{\theta}) = x\log\left(\frac{x}{\lambda}\right) - \lambda\left(\frac{x}{\lambda} - 1\right)$$

and $h(1) = -\log(\lambda) + \lambda - 1 > 0$ for any $\lambda \neq 1$. Fix $\delta > 0$ and let $k = h(1)^{-1}(1+\delta)\log n$. Then,

$$\mathbb{P}\left(|C(v)| > k\right) \leq e^{-(1+\delta)\log n} = n^{-(1+\delta)}$$

and the probability that the size of the largest component exceeds k satisfies

$$\mathbb{P}\left(|C_1| > k\right) \leq \mathbb{P}\left(\max_{j=1..n}|C(v_j)| > k\right) \leq \sum_{j=1}^{n}\mathbb{P}\left(|C(v_j)| > k\right) = nn^{-(1+\delta)} = n^{-\delta}.$$

Since δ is arbitrary, we conclude $\mathbb{P}\left(|C_1| > \alpha\log(n)\right) \to 0$ as $n \to \infty$ for all α sufficiently large . □

6.1.2 Supercritical regime ($\lambda > 1$)

Theorem 6.2. *If $\lambda > 1$, a graph in $G(n,p)$ has a unique giant component containing a positive fraction of the vertices while no other component contains more than $O(\log n)$ vertices. Precisely, let $p_{\text{ext}}(\mu)$ be the extinction probability of a Galton-Watson branching process with parameter μ. Then for any $\varepsilon > 0$ and some constant α, we have*

$$\mathbb{P}\left(\left|\frac{|C_1|}{n} - (1 - p_{\text{ext}}(\lambda))\right| \leq \varepsilon \text{ and } |C_2| \leq \alpha\log n\right) \to 1$$

as $n \to \infty$.

Proof: (Sketch) For simplicity, we only consider the case $1 < \lambda < 2$. We again identify the connected component $C(v)$ of some vertex v by considering its breadth-first spanning tree. In this case, we will bound the size of $C(v)$ from below using another Galton-Watson branching process. The procedure is less straightforward than in the subcritical case since we would like to show existence of a giant component, but the breadth-first spanning tree of v has a *decreasing* number of possible offspring with each generation. To remedy this, we consider a branching process with a strictly smaller average number of offspring but which still remains supercritical.

Let $d = \sqrt{\lambda} > 1$, and define $\tilde{n} = (2-d)n$ and $\tilde{p} = d/n$. Note that as long as our breadth-first process has explored less than $(d-1)n$ nodes, there are at least $\tilde{n}$ unexplored nodes. In this case, each node in the spanning tree has at least $\tilde{n}$ possible children, each with probability $p > \tilde{p}$. This implies that this tree is larger in size than a branching process with offspring distribution $\xi \sim \text{Binomial}(\tilde{n}, \tilde{p})$. Ordering the vertices in $C(v)$, we now apply the following procedure starting with $i = 1$:

1) Consider node v_i.

2) The component to which v_i belongs to in the limiting graph as $n \to \infty$ has infinite size with constant probability at least $1 - p_{\text{ext}}(d)$. In this case, we have successfully identified a giant component of size $|C(v_i)| \geq (d-1)n$.

3) Otherwise, by the argument given in the proof of the subscritical case we have that $|C(v_i)| \geq \alpha \log n$ with probability at most $n^{-\delta}$ for some $\delta > 0$.

Now repeat the same process with the next unexplored vertex v_{i+1}. The probability we have not found a giant component after $k < n$ steps equals $(p_{\text{ext}}(d))^k$, which is less than $n^{-\gamma}$ for any given $\gamma > 0$ if $k > -\gamma \log n / \log p_{\text{ext}}(d)$. The number of nodes belonging to small components which have been explored after k steps is then $O\left((\log n)^2\right) \ll n$, so there remain enough unexplored nodes for this procedure to work.

Lastly, we would not only like the existence of a giant component with high probability, but that $|C_1| \approx (1 - p_{\text{ext}}(\lambda))\, n$. While we do not show this here, it is reasonable that this should be true by the law of large numbers (in the number of nodes). That is, $|C_1|/n$ should converge to the probability that the spanning tree started from a node does not become extinct, which is approximately $1 - p_{\text{ext}}(\lambda)$. □

6.1.3 Critical regime ($\lambda = 1$)

Theorem 6.3. *If $\lambda = 1$, then $G(n,p)$ has a largest component whose size is of order $n^{2/3}$ a.s. That is, there exist constants $a_1, a_2 > 0$ such that $\mathbb{P}\left(a_1 n^{2/3} \leq |C_1| \leq a_2 n^{2/3}\right) \to 1$ as $n \to \infty$.*

To show this, we will use a result from probability theory known as the martingale central limit theorem.

Definition. A stochastic process $\{X_k : k \geq 0\}$ is a *martingale* if $\mathbb{E}[|X_k|] < \infty$ and

$$\mathbb{E}[X_k - X_{k-1} | X_0, \ldots, X_{k-1}] = 0$$

for every $k \geq 1$. That is, a martingale is a process in which the average of any increment is zero given its past history.

Lemma 6.4. (Martingale central limit theorem) *Suppose $\{X_k : k \geq 0\}$ is a martingale such that $|X_k - X_{k-1}| < M$ and $|X_0| < M$ a.s. for some fixed M and for all $k \geq 1$. Assume further that $Var[X_k - X_{k-1}|X_0, \ldots, X_{k-1}] = 1$ as $k \to \infty$. Then for all $t \geq 0$,*

$$\frac{X_1 + \cdots + X_{\lfloor mt \rfloor}}{\sqrt{m}} \xrightarrow{d} W_t$$

as $m \to \infty$, where $(W_t)_{t \geq 0}$ is a standard Brownian motion.

Using this we obtain a proof for the critical regime:

Proof: (Sketch) Let $\{A_k : k \geq 0\}$ be the random walk associated to the one-by-one exploration of the breadth-first spanning forest of the graph (where the forest is comprised of the spanning trees of the various disconnected components). Recall that

$$A_k = A_{k-1} - 1 + \xi_k, \quad k \geq 1$$
$$A_0 = 1$$

where A_k is the size of the queue of explored nodes at time k, and ξ_k is the number of children of the k^{th} deactivated node. Given $A_0, \ldots, A_{k-1}$,

$$\xi_k \sim \text{Binomial}(n - k - 1 - A_{k-1}, p) \approx \text{Binomial}(n-k, p)$$

for $n \gg k$.

To examine the size of largest component at the critical value, we consider $\lambda = 1 + cn^{-1/3}$ $-\infty < c < \infty$. As c passes 0 we should observe the emergence of a giant component whose size scales like $n^{2/3}$. As the following argument shows, the $-1/3$ exponent of the perturbative term is the appropriate scaling in which to see this critical behavior.

To begin, since $p = \lambda/n = n^{-1} + cn^{-4/3}$,

$$\mathbb{E}[A_k - A_{k-1}|A_0, \ldots, A_{k-1}] = \mathbb{E}[\xi_k] - 1 \approx (n-k)p - 1 = \frac{c}{n^{1/3}} - \frac{k}{n} - \frac{ck}{n^{4/3}},$$

$$Var[A_k - A_{k-1}|A_0, \ldots, A_{k-1}] = Var[\xi_k] \approx (n-k)p(1-p) \approx 1.$$

Letting

$$X_k = A_k - A_{k-1} - \left(\frac{c}{n^{1/3}} - \frac{k}{n} - \frac{ck}{n^{4/3}}\right),$$

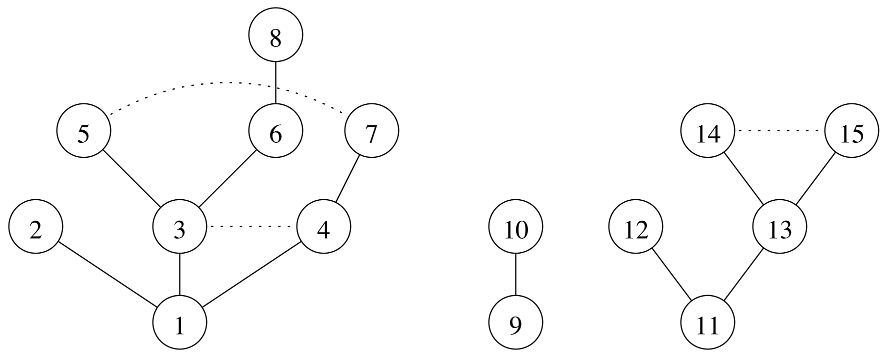

Figure 6.1. Example of a breadth-first spanning forest.

we see that $\{X_k : k \geq 0\}$ is a martingale. Now, for any $m \geq 1$ define the process

$$\bar{X}_t^{(m)} = \sum_{k=1}^{mt} X_k = A_{mt} - 1 - \left[\frac{cmt}{n^{1/3}} - \frac{mt\,(mt-1)}{2n} - \frac{cmt\,(mt-1)}{2n^{4/3}}\right].$$

for $t \in m^{-1}\mathbb{N}$ and by linear interpolation for all nonnegative $t \notin m^{-1}\mathbb{N}$. Taking $m = n^{2/3}$ yields

$$\bar{X}_t^{(m)} = A_{n^{2/3}t} - n^{1/3}\left(ct - \frac{1}{2}t^2\right) + O(1).$$

Therefore, by the martingale central limit theorem

$$\frac{A_{n^{2/3}t}}{n^{1/3}} - \left(ct - \frac{1}{2}t^2\right) = \frac{\bar{X}_t^{(n^{2/3})}}{n^{1/3}} + O\left(n^{-2/3}\right) \xrightarrow{d} W_t$$

for all $t \geq 0$ as $n \to \infty$, where $(W_t)_{t\geq 0}$ is Brownian motion.

To summarize, we have that $A_{n^{2/3}t} \approx n^{1/3} Z_t$ for large n, where

$$Z_t = W_t + ct - \frac{1}{2}t^2.$$

The size of the i^{th} component is the i^{th} time the queue is empty:

$$|C_i| = \min\{k \geq 0 : A_k = 1 - i\} \approx n^{2/3} \min\left\{t \geq 0 : Z_t = (1-i)\, n^{-1/3}\right\}.$$

We conclude that the size of the largest component can be estimated by $n^{2/3}$ times the length τ of the largest excursion of Z_t above zero. Since $\mathbb{E}[Z_t] = ct - \frac{1}{2}t^2 = 0$ when $t = 2c$, we expect the distribution of τ to be concentrated about $2c$ when $c > 0$ and about zero when $c \leq 0$. This yields the desired result. $\square$

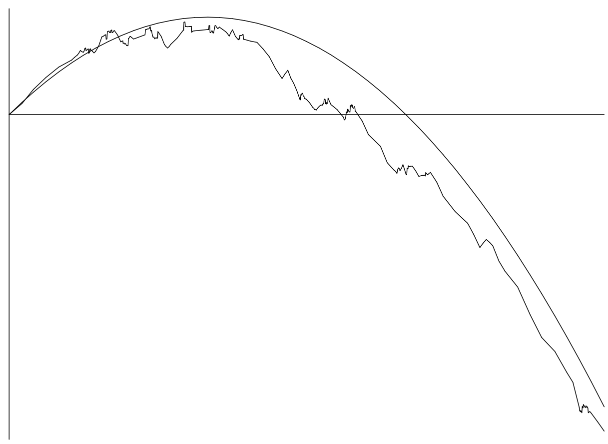

Figure 6.2. Plot of $Z_t = W_t + ct - \frac{1}{2}t^2$ for $c > 0$.

6.2 Connectivity and diameter

For completeness, let us state the existence of thresholds for the connectivity and diameter of Erdős-Rényi random graphs. Additional discussion along with proofs can be found in the references.

For connectedness we have:

Theorem 6.5. *A sharp threshold for the connectedness of $G(n,p)$ is $t(n) = n^{-1}\log n$. In other words, if $p(n) < (1-\varepsilon)n^{-1}\log n$ for some $\varepsilon > 0$, then $G(n,p)$ almost surely contains isolated nodes (so it is disconnected) in the limit $n \to \infty$. If $p(n) > (1+\varepsilon)n^{-1}\log n$ for some $\varepsilon > 0$, $G(n,p)$ is almost surely connected as $n \to \infty$.*

Moving past the threshold for connectivity, the diameter of the graph can be estimated with high probability:

Theorem 6.6. *Assume that for n large, $\log n \ll \lambda \ll n^{1/2}$. Then as $n \to \infty$,*

$$\mathbb{P}\left(\left\lceil \frac{\log n}{\log \lambda} \right\rceil - 4 \leq diam\left(G(n,p)\right) \leq \left\lceil \frac{\log n}{\log \lambda} \right\rceil + 1\right) \longrightarrow 1.$$

References

1. Aldous, D. (2007). STAT 260, Random Graphs and Complex Networks, Spring 2007 [Lecture notes]. Retrieved from:
`http://www.stat.berkeley.edu/~aldous/Networks/`

2. Draief, M., & Massoulié, L. (2010). Epidemics and rumours in complex networks. Cambridge University Press.

3. Durrett, R. (2006). Random graph dynamics (Vol. 20). Cambridge University Press.

4. Spielman, D. (2010). Applied Mathematics 462, Graphs and Networks, Fall 2010 [Lecture notes]. Retrieved from:
`http://www.cs.yale.edu/homes/spielman/462/`

Last edited: March 21, 2013