Lecture 17: The size of the largest component in Erdős-Rényi random graphs

We are now in a position to begin to understand the component structure of the Erdős-Rényi random graphs we are studying. For $\lambda \neq 1$, define

$$I_\lambda := \lambda - 1 - \log(\lambda) > 0,$$

and recall from Lemma 16.5 from Lecture 16 that if $Y_{j\lambda} \sim \mathrm{Po}(j\lambda)$, then there is a constant c depending only on λ such that

$$\mathbb{P}(Y_{j\lambda} = j - 1) \leq ce^{-jI_\lambda}.$$

For $\lambda > 1$, let ζ_λ be the probability that a branching process with Poisson(λ) offspring distribution does not go extinct. Recall from Example 12.3 of Lecture 12 that ζ_λ satisfies

$$\zeta_\lambda = 1 - e^{-\lambda\zeta_\lambda}.$$

Lemma 17.1. *Let C_v be the number of vertices in the component containing vertex v in $G = G(n, \lambda/n) \sim \mathrm{ER}(n, \lambda/n)$. If $\lambda < 1$, then there is a constant c depending only on λ such that for any $\kappa > 0$,*

$$\mathbb{P}\Big(C_v \geq (1+\kappa)I_\lambda^{-1}\log(n)\Big) \leq cn^{-1-\kappa}. \tag{17.1}$$

If $\lambda > 1$, then

$$\mathbb{P}\Big(C_v \geq (1+\kappa)I_\lambda^{-1}\log(n)\Big) = \zeta_\lambda + \mathrm{O}\left(\frac{\log(n)^3}{n}\right).$$

Proof. Write $k_n := \lceil (1+\kappa)I_\lambda^{-1}\log(n)\rceil$, and let W_λ be distributed as the total population size in a branching process with Poisson(λ) offspring distribution.

If $\lambda < 1$, then Lemmas 13.1, 15.4, and 16.5 imply there is a constant c' depending only on λ such that

$$\begin{aligned}
\mathbb{P}(C_v \geq (1+\kappa)I_\lambda^{-1}\log(n)) &\leq \mathbb{P}(W_\lambda = \infty) + \sum_{j \geq k_n} \mathbb{P}(W_\lambda = j) \\
&\leq \mathbb{P}(W_\lambda = \infty) + c' \sum_{j \geq k_n} e^{-jI_\lambda} \\
&= e^{-k_n I_\lambda} \frac{c'}{1 - e^{-I_\lambda}} \\
&\leq cn^{-(1+\kappa)},
\end{aligned} \tag{17.2}$$

where $c = c'/(1 - e^{-I_\lambda})$, and the equality is because Proposition 12.2 implies $\mathbb{P}(W_\lambda = \infty) = 0$, and by simplifying the geometric sum. Thus the first assertion follows.

If $\lambda > 1$, then Lemmas 13.1, 15.4, and 16.5 imply that for any $\kappa > 0$,

$$\begin{aligned}
\mathbb{P}(C_v \geq (1+\kappa)I_\lambda^{-1}\log(n)) &= \mathbb{P}(W_\lambda = \infty) + \sum_{j \geq k_n} \mathbb{P}(W_\lambda = j) + \mathrm{O}\left(\frac{\log(n)^3}{n}\right) \\
&= \zeta_\lambda + \mathrm{O}\big(n^{-(1+\kappa)}\big) + \mathrm{O}\left(\frac{\log(n)^3}{n}\right) \\
&= \zeta_\lambda + \mathrm{O}\left(\frac{\log(n)^3}{n}\right),
\end{aligned}$$

where the second equality is because by definition $\zeta_\lambda = \mathbb{P}(W_\lambda = \infty)$, and the sum is the bounded the same as when $\lambda < 1$. Thus the second assertion follows. □

From this lemma we can see the outline of what is going on. The exploration process for each vertex initially looks like a Poisson branching process. If $\lambda < 1$, the probability that any given exploration process captures at least $(1+\kappa)I_\lambda^{-1}\log(n)$ vertices is $\mathrm{O}(n^{-1-\kappa})$. Since there are only n vertices, the probability that any of the exploration processes captures at least this many vertices tends to zero (first moment method). With a little more work (second moment method), we can show that with probability tending to one, at least one captures at least $(1-\kappa)I_\lambda^{-1}\log(n)$ vertices, and so the size of the largest component is approximately $I_\lambda^{-1}\log(n)$.

If $\lambda > 1$, then each exploration process is "big" with probability ζ_λ, the probability that the associated branching process survives forever. We will show that in fact when an exploration process is big, it is actually of order n, and all the vertices with big processes end up in the same component. Thus, if the exploration processes are not too correlated, then the number of exploration processes that are big (which equals the size of the large component) is order $n\zeta_\lambda$. We state this formally in the following result which has been our goal for the last several lectures.

Theorem 17.2. *Let $C_{\max}$ be the size of the largest component in $G = G(n,\lambda/n) \sim \mathrm{ER}(n,\lambda/n)$. If $\lambda < 1$, then for any $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{C_{\max}}{\log(n)} - \frac{1}{I_\lambda}\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0. \tag{17.3}$$

If $\lambda > 1$, then for any $\alpha \in (0,1/2)$ and $\beta \in (0, 1-2\alpha)$

$$\mathbb{P}\left(\left|\frac{C_{\max}}{n} - \zeta_\lambda\right| > n^{-\alpha}\right) = \mathrm{O}(n^{-\beta}).$$

Moreover, in this case, with probability tending to one, there is a unique largest component, and the second largest component is $\Theta(\log(n))$.

We currently have the tools to prove the "subcritical" $\lambda < 1$ case of the theorem.

Proof of (17.3) *of Theorem 17.2.* We show that for any $0 < \kappa < 1$,

$$\mathbb{P}\big(C_{\max} \geq (1+\kappa)I_\lambda^{-1}\log(n)\big) \xrightarrow[n\to\infty]{} 0, \tag{17.4}$$

$$\mathbb{P}\big(C_{\max} < (1-\kappa)I_\lambda^{-1}\log(n)\big) \xrightarrow[n\to\infty]{} 0, \tag{17.5}$$

from which the result easily follows. For (17.4), first note that

$$\big\{C_{\max} \geq k\big\} = \cup_{v=1}^n \big\{C_v \geq k\big\},$$

and so a union bound and (17.1) implies

$$\mathbb{P}\big(C_{\max} \geq (1+\kappa)I_\lambda^{-1}\log(n)\big) \leq \sum_{v=1}^n \mathbb{P}\big(C_v \geq (1+\kappa)I_\lambda^{-1}\log(n)\big) = \mathrm{O}\big(n^{-\kappa}\big),$$

which implies (17.4). For (17.5), let $X_k = \sum_{v=1}^n \mathbb{I}[C_v \geq k]$ and note that

$$\big\{C_{\max} < k\big\} = \big\{X_k = 0\big\}, \tag{17.6}$$

and so we need to show that $\mathbb{P}(X_k = 0) \to 0$ as $n \to \infty$ for $k := \lceil (1-\kappa)I_\lambda^{-1}\log(n)\rceil$. We use the second moment method:

$$\mathbb{P}(X_k = 0) = 1 - \mathbb{P}(X_k \geq 1) \leq 1 - \frac{\mathbb{E}[X_k]^2}{\mathbb{E}[X_k^2]}, \tag{17.7}$$

and so we need a lower bound on $\mathbb{E}[X_k]$ and an upper bound on $\mathbb{E}[X_k^2]$. For the first moment, for any vertex v, W_λ distributed as the total population size of a branching process with Poisson(λ) offspring distribution, and $Y_{\lambda k} \sim \mathrm{Po}(\lambda k)$, we have

$$\begin{aligned}
\mathbb{E}[X_k] &= n\mathbb{P}(C_v \geq k) \\
&= n\sum_{j\geq k} \mathbb{P}(W_\lambda = j) + \mathrm{O}(k^3) \\
&\geq n\mathbb{P}(W_\lambda = k) + \mathrm{O}(k^3) \\
&= n\frac{\mathbb{P}(Y_{\lambda k} = k-1)}{k} + \mathrm{O}(k^3) \\
&= ne^{-\lambda k}\frac{(\lambda k)^{k-1}}{k!} + \mathrm{O}(k^3),
\end{aligned}$$

where we have used Lemma 13.1 in the second equality, and Lemma 15.4 in the second to last line. Using Stirling's formula (noting $k \to \infty$ with n),

$$k! = \big(1+\mathrm{o}(1)\big)\sqrt{2\pi k}e^{-k}k^k,$$

in the previous display, for k large enough

$$\begin{aligned}
\mathbb{E}[X_k] &\geq \frac{1+\mathrm{o}(1)}{\lambda k^{3/2}\sqrt{2\pi}} ne^{-k(\lambda-\log(\lambda)-1)} + \mathrm{O}(k^3) \\
&= \frac{1+\mathrm{o}(1)}{\lambda k^{3/2}\sqrt{2\pi}} ne^{-kI_\lambda} + \mathrm{O}(k^3) \\
&\geq \frac{\big(1+\mathrm{o}(1)\big)e^{-I_\lambda}}{\lambda k^{3/2}\sqrt{2\pi}} ne^{-(1-\kappa)\log(n)} + \mathrm{O}(k^3) \\
&\geq c'\frac{n^\kappa}{\log(n)^{3/2}} \xrightarrow[n\to\infty]{} \infty,
\end{aligned} \tag{17.8}$$

where the second inequality is because $k \leq (1-\kappa)I_\lambda^{-1}\log(n)+1$, and c' is a constant depending only on λ.

For the second moment, in next lecture we show there is a constant $c = c(\lambda)$ only depending on λ such that

$$\mathbb{E}[X_k^2] \leq \mathbb{E}[X_k]^2 + \mathbb{E}[X_k] + c\log(n)n^\kappa,$$

which, combined with (17.8), (17.7), and (17.6) proves (17.5). □