Lecture 18: Components sizes in subcritical Erdős-Rényi random graphs

Continuing from Lecture 17, we need to show the following lemma.

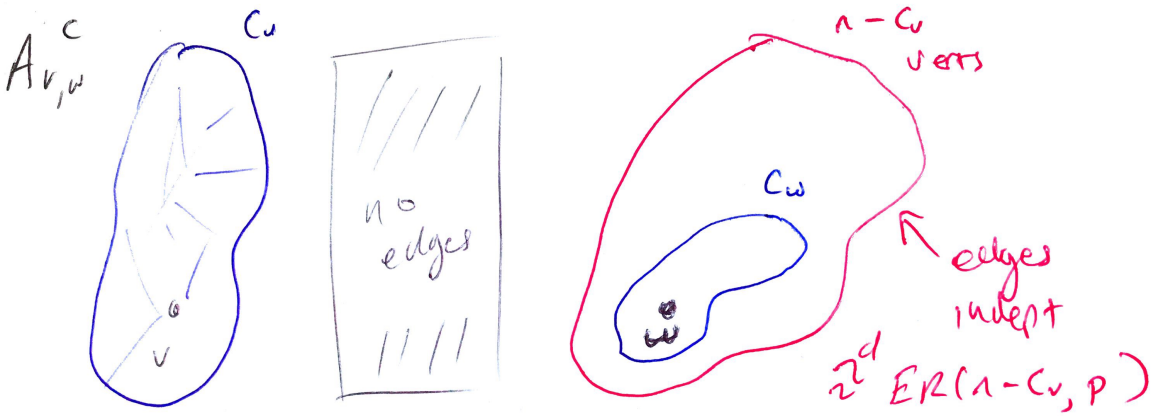
Lemma 18.1. *For $v \in \{1, \dots, n\}$, let C_v denote the size of the component containing vertex v in $G \sim \mathrm{ER}(n, \lambda/n)$ with $\lambda < 1$, and let $X_k =: \sum_{v=1}^n \mathbb{I}[C_v \geq k]$, where $k := \lceil (1-\kappa) I_\lambda^{-1} \log(n) \rceil$. Then there is a constant $c = c(\lambda)$ such that*

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

Proof. We begin by computing

$$\mathbb{E}[X_k^2] = \mathbb{E}[X_k] + \sum_{v=1}^n \sum_{w \neq v} \mathbb{P}(C_v \geq k, C_w \geq k). \tag{18.1}$$

To bound the sum, let $A_{v,w}$ denote the event that v and w are in the same component, and note that under $A_{v,w}$, $C_v = C_w$ and so $\{C_v \geq k\} = \{C_w \geq k\}$, and that under $A_{v,w}^c$, and conditional on C_v, C_w is distributed as the size of the component containing w in an Erdős-Rényi random graph on the remaining $n - C_v$ vertices. The following picture illustrates this case.



We first bound

$$\mathbb{P}(C_v \geq k, C_w \geq k, A_{v,w}^c) \leq \mathbb{P}(C_v \geq k) \mathbb{P}(C_w \geq k \,|\, C_v \geq k, A_{v,w}^c). \tag{18.2}$$

Now, as already remarked, $\mathscr{L}(C_w | \{C_v = \ell\} \cap A_{v,w}^c)$ is the same as the size of the component containing vertex w in an Erdős-Rényi random graph on $n - \ell$ vertices, which is stochastically dominated by $\mathscr{L}(C_w)$, since the component can only get bigger by adding vertices (think of first generating the component on $n - \ell$ vertices, and then adding the remaining ℓ vertices and possible edges). In particular,

$$\mathbb{P}\big(C_w \geq k | \{C_v = \ell\} \cap A_{v,w}^c\big) \leq \mathbb{P}(C_w \geq k),$$

which isn't exactly what we need for (18.2), since the conditioning is different, but clearly the same principle will hold, and we can make this rigorous by noting

$$\begin{aligned} \mathbb{P}(C_w \geq k \,|\, C_v \geq k, A_{v,w}^c) &= \sum_{\ell \geq k} \mathbb{P}\big(C_w \geq k \,|\, C_v = \ell, A_{v,w}^c\big) \frac{\mathbb{P}(C_v = \ell \,|\, A_{v,w}^c)}{\mathbb{P}(C_v \geq k \,|\, A_{v,w}^c)} \\ &\leq \mathbb{P}(C_w \geq k). \end{aligned}$$

Thus using this in (18.2), we find

$$\sum_{v \neq w} \mathbb{P}(C_v \geq k, C_w \geq k, A_{v,w}^c) \leq \sum_{v \neq w} \mathbb{P}(C_v \geq k) \mathbb{P}(C_w \geq k) \leq \mathbb{E}[X_k]^2. \tag{18.3}$$

For the other case, since under $A_{v,w}$, $\{C_v > k\} = \{C_w > k\}$, we have

$$\begin{aligned}\sum_{v=1}^{n}\sum_{w\neq v}\mathbb{P}(C_v \geq k, C_w \geq k, A_{v,w}) &= \sum_{v=1}^{n}\sum_{w\neq v}\mathbb{P}(C_v \geq k, A_{v,w}) \\ &= \sum_{v=1}^{n}\mathbb{E}\left\{\sum_{w\neq v}\mathbb{I}\big[\{C_v \geq k\}\cap A_{v,w}\big]\right\} \\ &\leq \sum_{v=1}^{n}\mathbb{E}\big[C_v\mathbb{I}[C_v \geq k]\big] \\ &= n\mathbb{E}\big[C_v\mathbb{I}[C_v \geq k]\big], \end{aligned} \tag{18.4}$$

where the inequality is because there are exactly $C_v - 1$ other vertices in the component containing vertex v. Now rewrite the expectation appearing in the last term in terms of tail probabilities in the usual way:

$$\begin{aligned}\mathbb{E}\big[C_v\mathbb{I}[C_v \geq k]\big] &= \sum_{\ell\geq k}\ell\,\mathbb{P}(C_v = \ell) \\ &= \sum_{\ell\geq k}\sum_{j=1}^{\ell}\mathbb{P}(C_v = \ell) \\ &= \sum_{j=1}^{k}\sum_{\ell\geq k}\mathbb{P}(C_v = \ell) + \sum_{j>k}\sum_{\ell\geq j}\mathbb{P}(C_v = \ell) \\ &= k\mathbb{P}(C_v \geq k) + \sum_{j>k}\mathbb{P}(C_v \geq j).\end{aligned}$$

To bound these tail probabilities, we use, as in (17.2) from Lecture 17, that Lemmas 13.1, 15.4, and 16.5 imply there is a constant c' depending only on λ such that

$$\mathbb{P}(C_v \geq j) \leq c'e^{-jI_\lambda};$$

here is where we are using that $\lambda < 1$ so that the associated branching process is finite with probability one. Plugging this into the previous display, we have

$$\begin{aligned}\mathbb{E}\big[C_v\mathbb{I}[C_v > k]\big] &\leq c'e^{-kI_\lambda}\big(k + (1-e^{-I_\lambda})^{-1}\big) \\ &\leq c\log(n)n^{\kappa-1},\end{aligned} \tag{18.5}$$

where the first inequality is by computing the geometric sum, and the second is because

$$(1-\kappa)I_\lambda^{-1}\log(n) \leq k \leq (1-\kappa)I_\lambda^{-1}\log(n) + 1.$$

Combining (18.5) and (18.4) with (18.3) in (18.1) yields the lemma. □

Before going on to proving the remaining part ($\lambda > 1$) of Theorem 17.2 from Lecture 17, we take a moment to summarise what has happened in the subcritical $\lambda < 1$ case. The key facts are as follows:

1. For $1 \ll k \ll n^{1/3}$, and the notation as in previous lectures,

$$\mathbb{P}(C_v \geq k) \approx \mathbb{P}(W_\lambda \geq k) = \sum_{j\geq k}\mathbb{P}(Y_{\lambda j} = j-1) \approx e^{-kI_\lambda}.$$

The first approximation uses the upper bound on k and we showed it by coupling the Erdős-Rényi exploration process to binomial and Poisson branching processes, leading

to Lemma 13.1; the equality follows by the random walk representation of the Poisson branching process and the hitting time result culminating in Lemma 15.4; and the second approximation uses the lower bound on k and it follows by applying Lemma 16.5 (or using Stirling's approximation) and summing the geometric series.

2. Since $\{C_{\max} \geq k\} = \cup_{v=1}^{n} \{C_v \geq k\}$, we have

$$\mathbb{P}(C_{\max} \geq k) \leq n\mathbb{P}(C_v \geq k) \approx ne^{-kI_\lambda}.$$

Thus, if $k = (1+\kappa)I_\lambda^{-1}\log(n)$, the right hand side goes to zero, and if $k = (1-\kappa)I_\lambda^{-1}\log(n)$, the right hand side goes to infinity. This suggests $C_{\max}/\log(n)$ may be close to I_λ^{-1}; only suggests because the right hand side going to infinity is not meaningful, but we can make a proof by applying the second moment method.