Lecture 21: Component sizes in supercritical Erdős-Rényi random graphs

In this lecture we prove Lemma 20.5 from Lecture 20. It is similar in structure to Lemma 18.1 from Lecture 18, but the detailed bounds are different.

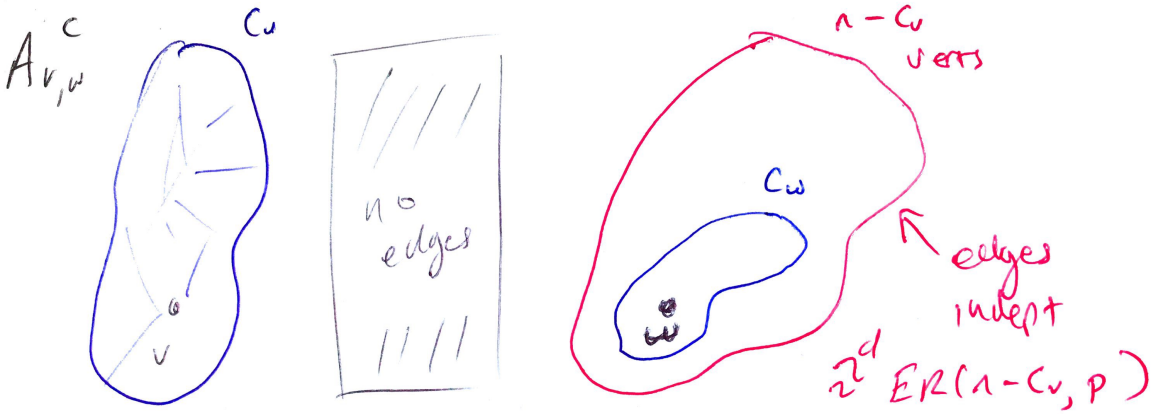
Lemma 20.5. *For $v \in \{1, \dots, n\}$, let C_v denote the size of the component containing vertex v in $G \sim \mathrm{ER}(n, \lambda/n)$ and $Y_k =: \sum_{v=1}^n \mathbb{I}[C_v \le k]$, where $k < n$. Then*

$$\mathbb{E}[Y_k^2] \le \mathbb{E}[Y_k]^2 + \mathbb{E}[Y_k](\lambda k^3 + k + 1)$$

Proof. We begin by computing

$$\mathbb{E}[Y_k^2] = \mathbb{E}[Y_k] + \sum_{v=1}^n \sum_{w \neq v} \mathbb{P}(C_v \le k, C_w \le k). \tag{21.1}$$

To bound the sum, let $A_{v,w}$ denote the event that v and w are in the same component, and note that under $A_{v,w}$, $\{C_v \le k\} = \{C_w \le k\}$, and that under $A_{v,w}^c$, 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; if in addition $C_v \le k$, then we can quantify how the distribution of the size of the component containing w changes compared to an Erdős-Rényi random graph on n vertices. The following picture (also used in Lecture 18) illustrates this case.



With this observation, the first case is easy to bound:

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

where the third equality is because there are exactly $C_v - 1$ other vertices in the component containing vertex v. For the second case, we have

$$\mathbb{P}(C_v \le k, C_w \le k, A_{v,w}^c) \le \mathbb{P}(C_v \le k)\mathbb{P}(C_w \le k \,|\, C_v \le k, A_{v,w}^c)$$

Thus, we can now bound (21.1) by, for $v \neq w$,

$$\begin{aligned}\mathbb{E}[Y_k^2] &\leq \mathbb{E}[Y_k](k+1) + n(n-1)\mathbb{P}(C_v \leq k)\mathbb{P}(C_w \leq k \,|\, C_v \leq k, A_{v,w}^c) \\ &\leq \mathbb{E}[Y_k](k+1) + \mathbb{E}[Y_k]^2 + n^2\mathbb{P}(C_v \leq k)\big(\mathbb{P}(C_w \leq k \,|\, C_v \leq k, A_{v,w}^c) - \mathbb{P}(C_v \leq k)\big),\end{aligned}$$

and the result will follow once we show, noting $C_v \stackrel{d}{=} C_w$,

$$\mathbb{P}(C_w \leq k \,|\, C_v \leq k, A_{v,w}^c) - \mathbb{P}(C_w \leq k) \leq \frac{\lambda k^2(k+1)}{2n}, \tag{21.2}$$

since the right hand side is no greater than $\lambda k^3/n$ for $k \geq 1$. For each $\ell = 1, \ldots, k$ and $j \geq 1$, we show that

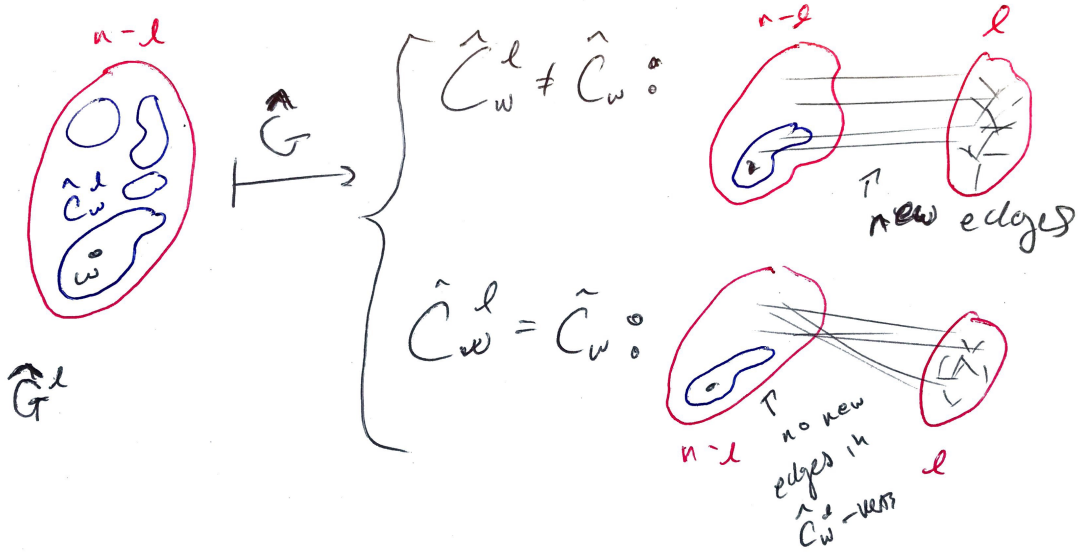
$$\mathbb{P}(C_w = j \,|\, C_v = \ell, A_{v,w}^c) - \mathbb{P}(C_w = j) \leq \frac{\lambda jk}{n}, \tag{21.3}$$

and then (21.2) follows by using that for $j = 1, \ldots, k$,

$$\begin{aligned}\mathbb{P}(C_w = j \,|\, C_v \leq k, A_{v,w}^c) &= \sum_{\ell=1}^{k} \mathbb{P}\big(C_w = j \,|\, C_v = \ell, A_{v,w}^c\big) \frac{\mathbb{P}(C_v = \ell \,|\, A_{v,w}^c)}{\mathbb{P}(C_v \leq k \,|\, A_{v,w}^c)} \\ &\leq \sum_{\ell=1}^{k} \left(\mathbb{P}(C_w = j) + \frac{\lambda jk}{n}\right) \frac{\mathbb{P}(C_v = \ell \,|\, A_{v,w}^c)}{\mathbb{P}(C_v \leq k \,|\, A_{v,w}^c)} \\ &= \mathbb{P}(C_w = j) + \frac{\lambda jk}{n};\end{aligned}$$

the inequality is by (21.3); and then summing over j.

To show (21.3), we want to do something like bounding the total variation distance between $\mathscr{L}(C_w \,|\, C_v = \ell, A_{v,w}^c)$ and $\mathscr{L}(C_w)$, but only over a range of the support. So we create a coupling $(\widehat{C}_w^\ell, \widehat{C}_w)$ of these two distributions by constructing a graph first on $n - \ell$ vertices, including w, such that the component containing w has distribution $\mathscr{L}(C_w \,|\, C_v = \ell, A_{v,w}^c)$, and then expanding the graph to n vertices so that the component containing w has distribution $\mathscr{L}(C_w)$. The key observation (already mentioned) is that conditional on $C_v = \ell$ and $A_{v,w}^c$, C_w is distributed as the size of the component containing w in an Erdős-Rényi random graph on the remaining $n - \ell$ vertices. Thus we let $\widehat{G}^\ell \sim \mathrm{ER}(n - \ell, \lambda/n)$ and $\widehat{C}_w^\ell$ be the size of the component containing w in $\widehat{G}^\ell$ (we assume w is one of the vertices of G^ℓ). Now, to generate $\widehat{C}_w$, we construct $\widehat{G} \sim \mathrm{ER}(n, \lambda/n)$ from $\widehat{G}^\ell$ by adding ℓ vertices to $\widehat{G}^\ell$ and independently putting edges, each with probability λ/n, between (i) pairs of new vertices and (ii) pairs of new and existing vertices. Then it's clear $\widehat{G}$ is distributed as desired, and thus we let $\widehat{C}_w$ be the size of the component containing w in $\widehat{G}$. The following picture illustrates.



Then

$$\mathbb{P}(\widehat{C}_w^\ell = j) = \mathbb{P}(\widehat{C}_w^\ell = j, \widehat{C}_w^\ell = \widehat{C}_w) + \mathbb{P}(\widehat{C}_w^\ell = j, \widehat{C}_w^\ell \neq \widehat{C}_w),$$
$$\mathbb{P}(\widehat{C}_w = j) = \mathbb{P}(\widehat{C}_w = j, \widehat{C}_w^\ell = \widehat{C}_w) + \mathbb{P}(\widehat{C}_w = j, \widehat{C}_w^\ell \neq \widehat{C}_w),$$

so that

$$\left|\mathbb{P}(\widehat{C}_w^\ell = j) - \mathbb{P}(\widehat{C}_w = j)\right| \leq \left|\mathbb{P}(\widehat{C}_w^\ell = j, \widehat{C}_w^\ell \neq \widehat{C}_w) - \mathbb{P}(\widehat{C}_w = j, \widehat{C}_w^\ell \neq \widehat{C}_w)\right|. \tag{21.4}$$

Looking again at the picture above, since the only way $\widehat{C}_w^\ell \neq \widehat{C}_w$ is if there are added edges of type (ii) between the $\widehat{C}_w^\ell$ vertices in the component containing w in G^ℓ and the ℓ added vertices; note there are $\ell \times \widehat{C}_w^\ell$ such pairs; we have

$$\mathbb{P}(\widehat{C}_w^\ell = j, \widehat{C}_w^\ell \neq \widehat{C}_w) \leq \mathbb{P}(\widehat{C}_w^\ell \neq \widehat{C}_w | \widehat{C}_w^\ell = j) \leq j\ell(\lambda/n) \leq \frac{\lambda jk}{n}. \tag{21.5}$$

Similarly, noting that $\widehat{C}_w^\ell \leq \widehat{C}_w$ (since we only add edges to obtain $\widehat{G}$ from $\widehat{G}^\ell$) and using the second inequality of (21.5), we have

$$\begin{aligned}
\mathbb{P}(\widehat{C}_w = j, \widehat{C}_w^\ell \neq \widehat{C}_w) &= \sum_{i=0}^{j-1} \mathbb{P}(\widehat{C}_w = j | \widehat{C}_w^\ell = i)\mathbb{P}(\widehat{C}_w^\ell = i) \\
&\leq \sum_{i=0}^{j-1} \mathbb{P}(\widehat{C}_w^\ell \neq \widehat{C}_w | \widehat{C}_w^\ell = i)\mathbb{P}(\widehat{C}_w^\ell = i) \\
&\leq \frac{\lambda jk}{n} \sum_{i=0}^{j-1} \mathbb{P}(\widehat{C}_w^\ell = i) \\
&\leq \frac{\lambda jk}{n}.
\end{aligned}$$

Thus, the left hand side of (21.4), which is equal to that of (21.3), is bounded by $\lambda jk/n$, as desired. □