Lecture 14: Branching process approximation

We prove the following lemma introduced in the last lecture.

Lemma 13.1. *Let C_v be the size of the component containing vertex v in $G \sim \mathrm{ER}(n, \lambda/n)$, where $\lambda > 0$ is fixed. Let W_λ be the total population of a branching process with* Poisson(λ) *offspring distribution. Then for any $k \geq 0$,*

$$\mathbb{P}(W_\lambda \geq k) - \frac{k^2\lambda(k+\lambda)}{n} \leq \mathbb{P}(C_v \geq k) \leq \mathbb{P}(W_\lambda \geq k).$$

Proof. For the upper bound, recall the exploration process recursions (11.1) of Lemma 11.2 from Lecture 11:

$$A_t = A_{t-1} + \mathbb{I}[A_{t-1} > 0](-1 + Z_t), \tag{14.1}$$

where $Z_t \sim \mathrm{Bi}\big(n - (t-1) - A_{t-1}, \lambda/n\big) = \mathrm{Bi}\big(N_{t-1}, \lambda/n\big)$ on the event that $A_{t-1} > 0$. From Example 13.4 and Proposition 13.7 from Lecture 13, we know that for any $m \leq n$

$$\mathrm{Bi}\big(m, \lambda/n\big) \leq_{\mathrm{st}} \mathrm{Bi}\big(n, \lambda/n\big) \leq_{\mathrm{st}} \mathrm{Po}(\lambda).$$

Thus, for the Poisson offspring BP recursion (12.1) from Lecture 12:

$$A'_t = A'_{t-1} + \mathbb{I}[A'_{t-1} > 0](-1 + Z'_t),$$

where $Z'_t \sim \mathrm{Po}(\lambda)$ are i.i.d., we have that

$$A'_t \geq_{\mathrm{st}} A_t,$$

for all t, and thus that

$$\tau'_0 := \inf\{t > 0 : A'_t = 0\} \geq_{\mathrm{st}} \inf\{t > 0 : A_t = 0\} =: \tau_0.$$

Note this means that $\mathbb{P}(\tau'_0 \geq k) \geq \mathbb{P}(\tau_0 \geq k)$ for all k. Using Lemmas 11.1 and 12.4 relating τ_0, τ'_0 to the component size C_v and total population size W_λ, we have

$$\mathbb{P}(W_\lambda \geq k) = \mathbb{P}(\tau'_0 \geq k) \geq \mathbb{P}(\tau_0 \geq k) = \mathbb{P}(C_v \geq k),$$

as desired.

For the lower bound, we first establish the intermediate bound

$$\mathbb{P}(C_v \geq k) \geq \mathbb{P}(V_{n-k,p} \geq k), \tag{14.2}$$

where $V_{n-k,p}$ is the total population size of a BP with $\mathrm{Bi}(n-k, p)$ offspring distribution. Note that this is not a stochastic domination result, since the parameter of the binomial depends on k, but it is in the same spirit and we will prove it with coupling. The idea is that the exploration process stochastically dominates the branching process until it has reached k vertices (since the first parameter of the binomial driving it is bigger). After this point the domination switches, but it doesn't ruin the result since we are only considering the event $\{\cdot \geq k\}$. The second step is to compare the binomial BP to the Poisson BP, which are close until they have k vertices since the total variation distance between the Poisson and Binomial offspring distributions is small.

To prove (14.2), let $T_k = \inf\{t : N_t \leq n - k\}$ be the first time the number of neutral vertices in the Erdős-Rényi exploration process is no greater than $n - k$, and note that

$$\{C_v \geq k\} = \{T_k < \infty\}.$$

For $t \leq T_k$, $N_{t-1} > n - k$ and so, using the binomial domination Example 13.4, we have

$$Z_t \sim \mathrm{Bi}(N_{t-1}, p) \geq_{\mathrm{st}} \mathrm{Bi}(n-k, p), \;\; t \leq T_k, \tag{14.3}$$

where Z_t is defined at (14.1). Letting $Z_t'' \sim \mathrm{Bi}(n-k,p)$, $t = 1, 2, \ldots$ be i.i.d., define the binomial BP recursion

$$A_t'' = A_{t-1}'' + \mathbb{I}[A_{t-1}'' > 0](-1 + Z_t'').$$

By (14.3), we have that $A_t'' \leq_{\mathrm{st}} A_t$ for $0 \leq t \leq T_k$. Saving notation, let $(A_t, A_t'')_{t\geq 0}$, be a coupling so that $A_t'' \leq A_t$ for $0 \leq t \leq T_k$. We claim that if $A_k'' > 0$, then $T_k \leq k$. This is true because if $A_k'' > 0$ and $T_k > k$, then the monotonicity of the coupling is in force and so $A_k \geq A_k'' > 0$, which implies that E_k, the number of explored vertices at step k, satisfies $E_k = k$. In particular,

$$N_k = n - k - A_k < n - k \implies T_k \leq k,$$

a contradiction. Thus the claim implies

$$\{V_{n-k,p} \geq k\} = \{A_k'' > 0\} \subseteq \{T_k < \infty\} = \{C_v \geq k\},$$

from which the inequality (14.2) follows.

To now compare the Poisson and binomial BPs, note that

$$\begin{aligned} \big|\mathbb{P}(V_{n-k,p} \geq k) - \mathbb{P}(W_\lambda \geq k)\big| &= \big|\mathbb{P}(V_{n-k,p} < k) - \mathbb{P}(W_\lambda < k)\big| \\ &\leq \sum_{i=0}^{k-1} \big|\mathbb{P}(V_{n-k,p} = i) - \mathbb{P}(W_\lambda = i)\big| \\ &\leq k \max_{i=0,\ldots,k-1} \big|\mathbb{P}(V_{n-k,p} = i) - \mathbb{P}(W_\lambda = i)\big|. \end{aligned} \tag{14.4}$$

Letting $U_{n,p}$ be the total population size in a BP with $\mathrm{Bi}(n,p)$ offspring distribution, we bound

$$\big|\mathbb{P}(V_{n-k,p} = i) - \mathbb{P}(W_\lambda = i)\big| \leq \big|\mathbb{P}(V_{n-k,p} = i) - \mathbb{P}(U_{n,p} = i)\big| + \big|\mathbb{P}(U_{n,p} = i) - \mathbb{P}(W_\lambda = i)\big|.$$

To bound each of these, we closely couple the offspring variables generating the pairs of processes. For the first term, let $(\widehat{Z}_t'', \widetilde{Z}_t)_{t\geq 1}$ be i.i.d. couplings of $\mathrm{Bi}(n-k,p)$ with $\mathrm{Bi}(n,p)$ such that (as in Lecture 6),

$$\mathbb{P}(\widehat{Z}_t'' \neq \widetilde{Z}_t) \leq kp = \frac{k\lambda}{n}. \tag{14.5}$$

Then

$$\begin{aligned} \mathbb{P}(V_{n-k,p} = i) &= \mathbb{P}\big(V_{n-k,p} = i, \cap_{t=1}^i \{\widehat{Z}_t'' = \tilde{Z}_t\}\big) + \mathbb{P}\big(V_{n-k,p} = i, \cup_{t=1}^i \{\widehat{Z}_t'' \neq \tilde{Z}_t\}\big) \\ &= \mathbb{P}\big(U_{n,p} = i, \cap_{t=1}^i \{\widehat{Z}_t'' = \tilde{Z}_t\}\big) + \mathbb{P}\big(V_{n-k,p} = i, \cup_{t=1}^i \{\widehat{Z}_t'' \neq \tilde{Z}_t\}\big), \end{aligned}$$

where the second equality is since if $V_{n-k,p} = i$, then BP only depends on the first i offspring variables, and so if these are the same for the BPs, then their total populations will also be the same. Similarly,

$$\mathbb{P}(U_{n,p} = i) = \mathbb{P}\big(U_{n,p} = i, \cap_{t=1}^i \{\widehat{Z}_t'' = \tilde{Z}_t\}\big) + \mathbb{P}\big(U_{n,p} = i, \cup_{t=1}^i \{\widehat{Z}_t'' \neq \tilde{Z}_t\}\big),$$

and so, using (14.5), for $i \leq k$,

$$\begin{aligned} \big|\mathbb{P}(V_{n-k,p} = i) - \mathbb{P}(U_{n,p} = i)\big| &\leq \mathbb{P}\big(\cup_{t=1}^i \{\widehat{Z}_t'' \neq \tilde{Z}_t\}\big) \\ &\leq \sum_{t=1}^i \mathbb{P}(\widehat{Z}_t'' \neq \tilde{Z}_t) \\ &\leq \frac{k^2\lambda}{n}. \end{aligned} \tag{14.6}$$

A similar argument for the second term, starting from taking $(\widetilde{Z}_t, \widehat{Z}_t')_{t\geq 1}$ to be i.i.d. couplings of $\mathrm{Bi}(n,p)$ and $\mathrm{Po}(\lambda)$ variables satisfying (as in Lecture 6)

$$\mathbb{P}(Z_t' \neq Z_t'') \leq np^2 = \frac{\lambda^2}{n},$$

implies that

$$\left|\mathbb{P}(U_{n,p} = i) - \mathbb{P}(W_\lambda = i)\right| \leq \frac{k\lambda^2}{n}. \tag{14.7}$$

Combining (14.2), (14.4), (14.6), and (14.7) implies the lower bound. □