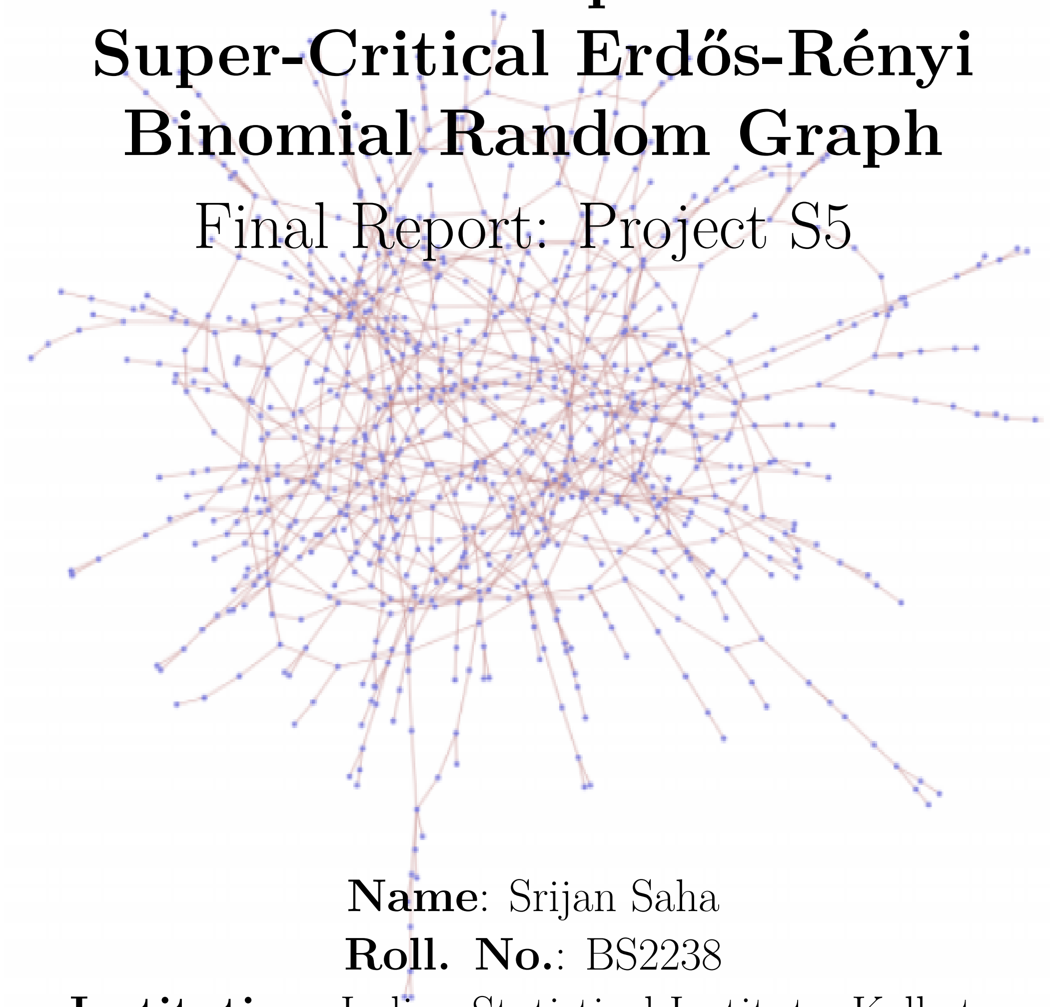
Random Graphs (Optional Course)
B. Stat Year III 2nd Semester, Summer 2024

Central Limit Theorem of the Size of the Giant Component for the Super-Critical Erdős-Rényi Binomial Random Graph

Final Report: Project S5



Name: Srijan Saha
Roll. No.: BS2238
Institution: Indian Statistical Institute, Kolkata

Course Instructor: Prof. Antar Bandyopadhyay
(Indian Statistical Institute, Delhi)

Abstract

The goal of this report is to study the asymptotic distribution of the "giant" connected component in supercritical Erdős-Rényi binomial random graph model i.e. in $\mathrm{ER}_n(p)$ model with $\lambda \coloneqq np > 1$. In the supercritical regime, the famous law of large numbers for giant component states that the largest connected component is linear in size with the total number of vertices, and engulfs roughly ζ_λ proportion of the vertex set $[n]$, where $\zeta_\lambda > 0$ is the survival probability of a Poisson(λ) Galton-Watson branching process. Further, the size of largest component is precisely $\zeta_\lambda n + o_p(n^{1/2+\epsilon})$, where $\epsilon > 0$ is arbitrarily small. Further, the second largest component is of size $O(\log n)$, i.e., much smaller compared to the total number of vertices (or the size of giant component).

In this report, we establish a central limit for the size of giant connected component, i.e., $|\mathscr{C}_{\max}|$, which claims the appropriate scaling for its non-degenerate asymptotic distribution is precisely $\sqrt{n}$. Needless to say, this is a much stronger result in the sense that it talks about the distributional convergence of $|\mathscr{C}_{\max}|$.

Acknowledgement
I would like to express my sincere gratitude to my course instructor, Prof. Antar Bandyopadhyay, Indian Statistical Institute, Delhi for their profound guidance during the course and for the novel initiative of introducing presentations as a part of the course, which enabled me to learn and appreciate the subject substantially. I also thank my peers for their invaluable feedback, unwavering support and enormous enthusiasm, and last but not the least, the zeal for actually understanding the subject.

Contents

Chapter 1

LLN for Giant Component in Supercritical Regime

1.1 Statement

The law of large numbers for "giant" Component for supercritical regime is the following,

Theorem 1.1 (Law of large numbers for giant component)**.** *Fix $\lambda > 1$. Then for every $\nu \in (1/2, 1)$, there exists $\delta \equiv \delta(a, \lambda) > 0$ such that,*

$$\mathbb{P}_\lambda[||\mathscr{C}_{\max}| - \zeta_\lambda n| \geq n^\nu] = O(n^{-\delta}). \tag{1.1}$$

$$\mathbb{P}\left[\textit{There exists a connected component with size in } (K \log n, \zeta_\lambda n)\right] = O(n^{-\delta}). \tag{1.2}$$

i.e., the largest connected component roughly occupies ζ_λ proportion of the graph (size $\approx \zeta_\lambda n$, more precisely $\zeta_\lambda n + o_p(n^{1/2+\epsilon})$ for any small $\epsilon > 0$) and is the unique $\theta(n)$ connected component. The second largest connected component is of size $O(logn)$, much smaller than the largest one. Hence the giant component is "unique", in the sense that it is linear in order with the size of the vertex set.

An immediate corollary for the earlier result is given below.

Theorem 1.2. *(Asymptotic size in probability) Let $\mathscr{C}_i$ denote the i-th largest connected component in $ER_n(p)$. Then*

$$\frac{|\mathscr{C}_1|}{n} \xrightarrow{\mathbb{P}} \zeta_\lambda, \tag{1.3}$$

$$\frac{|\mathscr{C}_i|}{n} \xrightarrow{\mathbb{P}} 0 \ \forall\ i \geq 2. \tag{1.4}$$

The proof of the main law of large numbers theorem is given in five main steps. We discuss each of them in separate sections.

1.2 STEP 1: Expected number of vertices in the giant component

Theorem 1.3 (Cluster tail probability is branching process survival probability)**.** *Fix $\lambda > 1$. Then for $K > a$ where $a > I_\lambda^{-1}$ and $I_\lambda = \lambda - 1 - \log\lambda$, and for sufficiently large n,*

$$\mathbb{P}_\lambda[|\mathscr{C}(v)| \geq K\log n] = \zeta_\lambda + O(K\log n/n). \tag{1.5}$$

i.e.,

$$\mathbb{E}_\lambda[Z_{\geq K\log n}] = n\mathbb{P}_\lambda[|\mathscr{C}(v)| \geq K\log n] = n\zeta_\lambda + O(K\log n). \tag{1.6}$$

Proof. Upper Bound:

$$\begin{aligned}
\mathbb{P}_\lambda[|\mathscr{C}(v)| \geq K\log n] &\leq \mathbb{P}_{n,\lambda/n}[T \geq K\log n] \quad \text{[By Theorem A.4]} \\
&\leq \mathbb{P}^*_\lambda[T^* \geq K\log n] + O(K\log n/n) \quad \text{[By Theorem A.3]} \\
&= \mathbb{P}^*_\lambda[T^* = \infty] + \mathbb{P}^*_\lambda[K\log n \leq T^* < \infty] + O(K\log n/n) \\
&= \zeta_\lambda + O(e^{-K\log n I_\lambda}) + O(K\log n/n) \\
&= \zeta_\lambda + O(n^{-aI_\lambda}) + O(K\log n/n) \quad [K \geq a > I_\lambda^{-1}] \\
&= \zeta_\lambda + O(K\log n/n).
\end{aligned} \tag{1.7}$$

Lower Bound:

$$\begin{aligned}
\mathbb{P}_\lambda[|\mathscr{C}(v)| \geq K\log n] &\geq \mathbb{P}_{n-K\log n,\lambda/n}[T \geq K\log n] \quad \text{[By Theorem A.5]} \\
&\geq \mathbb{P}^*_{\lambda(1-K\log n/n)}[T^* \geq K\log n] + O(K\log n/n) \quad \text{[By Theorem A.3]} \\
&= \zeta_{\lambda_n} + O(K\log n/n) \\
&= \zeta_\lambda + O(K\log n/n) \quad \text{[By Mean Value Theorem and Theorem A.2]}.
\end{aligned} \tag{1.8}$$

□

1.3 STEP 2: Concentration of number of vertices in large clusters

Theorem 1.4 (Concentration of number of Vertices in large clusters)**.** *Fix* $\nu \in (1/2, 1)$*. Then for sufficiently large* K *and every* $\delta_1 < 2\nu - 1$*, as* $n \to \infty$*,*

$$\mathbb{P}_\lambda[|Z_{\geq K \log n} - n\zeta_\lambda| > n^\nu] = O(n^{-\delta_1}). \tag{1.9}$$

Proof of STEP 2. We need estimate of variance of $Z_{\geq k}$.

Theorem 1.5 (Estimate of variance of $Z_{\geq k}$)**.** *For every* n *and* $k \in [n]$*,*

$$\mathrm{Var}_\lambda[Z_{\geq k}] \leq (\lambda k + 1) n \mathbb{E}[|\mathscr{C}(v)| \mathbb{1}_{|\mathscr{C}(v) < k|}] \leq nk(\lambda k + 1). \tag{1.10}$$

We omit the tedious calculation, which proves the variance estimate. It is similar with the proof of subcritical regime (but better estimate). For reference, see [1], Section 4.4.3, Proposition 4.10.

Now, by Theorem 1.3,

$$\mathbb{E}_\lambda[Z_{\geq K \log n}] = n\zeta_\lambda + O(K \log n). \tag{1.11}$$

and $K \log n = o(n^\nu)$, then for sufficiently large n,

$$\{|Z_{\geq K \log n} - n\zeta_\lambda| > n^\nu\} \subseteq \{|Z_{\geq K \log n} - \mathbb{E}_\lambda[Z_{\geq K \log n}]| > n^\nu/2\}. \tag{1.12}$$

By Chebyshev Inequality and Theorem 1.5,

$$\begin{aligned} \mathbb{P}_\lambda[|Z_{\geq K \log n} - n\zeta_\lambda| > n^\nu] &\leq \mathbb{P}_\lambda[|Z_{\geq K \log n} - \mathbb{E}_\lambda[Z_{\geq K \log n}]| > n^\nu/2] \\ &\leq 4n^{-2\nu} \mathrm{Var}[Z_{\geq K \log n}] \\ &\leq 4n^{-2\nu} \cdot n(K \log n)(\lambda K \log n + 1) \\ &\leq n^{-\delta_1} \end{aligned} \tag{1.13}$$

for any $\delta_1 < 2\nu - 1$ and sufficiently large n. □

1.4 STEP 3: No middle ground

Theorem 1.6 (No intermediate clusters)**.** *Fix* $\alpha < \zeta_\lambda$ *and let* $\delta_2 = K \cdot I_{(1-e^{-\lambda\alpha})\alpha} - 1$*, such that* $\delta_2 > 0$ *for sufficiently large* K*. Then*

$$\mathbb{P}[\textit{There exists a connected component with size} \in (K \log n, \alpha n)] = O(n^{-\delta_2}). \tag{1.14}$$

Proof of STEP 3. Let S_t be defined according to the exploration process:

$$S_t = \begin{cases} 1, & t = 0 \\ S_{t-1} + X_t - 1. & t \geq 1, \end{cases} \tag{1.15}$$
$$X_t \sim \text{Bin}(n - S_{t-1} - (t-1), p).$$

$$\begin{aligned} \mathbb{P}_\lambda[K \log n \leq |\mathscr{C}(v)| \leq \alpha n] &= \sum_{t=K\log n}^{\alpha n} \mathbb{P}_\lambda[|\mathscr{C}(v)| = t] \\ &\leq \sum_{t=K\log n}^{\alpha n} \mathbb{P}_\lambda[S_t = 0] \\ &\leq \sum_{t=K\log n}^{\alpha n} e^{-tJ(\alpha;\lambda)} \quad \text{[By Theorem A.6]} \\ &\leq \frac{e^{-(K\log n)J(\alpha;\lambda)}}{1 - e^{-J(\alpha;\lambda)}} \\ &= Cn^{-KJ(\alpha;\lambda)}. \end{aligned} \tag{1.16}$$

So,

$$\begin{aligned} \mathbb{P}_\lambda[\exists\, v : K \log n \leq |\mathscr{C}(v)| \leq \alpha n] &\leq n\mathbb{P}_\lambda[K \log n \leq |\mathscr{C}(v)| \leq \alpha n] \\ &\leq Cn^{1-KJ(\alpha;\lambda)} \\ &= O(n^{-\delta_2}) \quad \text{[Take } K \text{ sufficiently large]}. \end{aligned} \tag{1.17}$$

$\square$

1.5 STEP 4: Finding the high probability event

Theorem 1.7 ($|\mathscr{C}_{\max}|$ equals $Z_{\geq K\log n}$ whp)**.** *Fix $\alpha \in (\zeta_\lambda/2, \zeta_\lambda)$. Define an event*

$$\mathscr{E}_n := \left\{ |Z_{\geq K\log n} - n\zeta_\lambda| \leq n^\nu,\ \nexists\, v \in [n] \text{ such that } K \log n \leq |\mathscr{C}(v)| \leq \alpha n \right\}. \tag{1.18}$$

Then

$$\mathbb{P}_\lambda(\mathscr{E}_n^c) = O(n^{-\min(\delta_1,\delta_2)}), \tag{1.19}$$

$$|\mathscr{C}_{\max}| = Z_{\geq K\log n} \quad \text{on } \mathscr{E}_n. \tag{1.20}$$

Proof of STEP 4.

$$\begin{aligned}\mathbb{P}_\lambda(\mathscr{E}_n^c) &\le \mathbb{P}_\lambda[|Z_{\ge K\log n} - n\zeta_\lambda| > n^\nu] + \mathbb{P}_\lambda[\exists\ v \in [n] : K\log n \le |\mathscr{C}(v)| \le \alpha n] \\ &= O(n^{-\min(\delta_1,\delta_2)}) \quad \text{[By Theorem 1.4 and Theorem 1.6]}.\end{aligned} \tag{1.21}$$

On $\mathscr{E}_n$,

- $\{|Z_{\ge K\log n} - \zeta_\lambda n| \le n^\nu\} \subseteq \{Z_{\ge K\log n} \ge 1\} \Rightarrow |\mathscr{C}_{\max}| \le Z_{\ge K\log n}$.
- Assume, $|\mathscr{C}_{\max}| < Z_{\ge K\log n} \Rightarrow$. Then there are two connected components with size at least $K\log n$. Also on $\mathscr{E}_n$, there are no connected components of size $\in (K\log n, \alpha n)$. So the two components are of size $\ge \alpha n$. Which means $Z_{\ge K\log n} \ge 2\alpha n$. But since $2\alpha > \zeta_\lambda$ and for large n, $Z_{\ge K\log n} \le \zeta_\lambda n + n^\nu$ (Contradiction).

So,

$$|\mathscr{C}_{\max}| = Z_{\ge K\log n} \text{ on } \mathscr{E}_n.$$

□

1.6 STEP 5: Finalizing the proof of LLN

Proof for LLN.

$$\begin{aligned}\mathbb{P}_\lambda[||\mathscr{C}_{\max}| - \zeta_\lambda n| \le n^\nu] &\ge \mathbb{P}_\lambda[\{||\mathscr{C}_{\max}| - \zeta_\lambda n| \le n^\nu\} \cap \mathscr{E}_n] \\ &= \mathbb{P}_\lambda[\{|Z_{\ge K\log n} - \zeta_\lambda n| \le n^\nu\} \cap \mathscr{E}_n] \quad \text{[By Equation 1.20]} \\ &= \mathbb{P}_\lambda(\mathscr{E}_n) \quad \text{[By Definition of } \mathscr{E}_n] \\ &\ge 1 - O(n^{-\min(\delta_1,\delta_2)}) \quad \text{[By Equation 1.20]} \\ &= 1 - O(n^{-\delta}) \quad [\delta \coloneqq \min(\delta_1,\delta_2)].\end{aligned} \tag{1.22}$$

Hence proved. □

Chapter 2

CLT for the Giant Component

We now extend the law of large numbers for giant component in super-critical regime, to the Central Limit Theorem.

2.1 Statement

Theorem 2.1 (CLT for giant component in supercritical regime)**.** *Fix* $\lambda > 1$. *Then,*

$$\frac{|\mathscr{C}_{\max}| - \zeta_\lambda n}{\sqrt{n}} \xrightarrow{d} Z \sim N(0, \sigma_\lambda^2), \tag{2.1}$$

where,

$$\sigma_\lambda^2 \coloneqq \frac{\zeta_\lambda(1-\zeta_\lambda)}{(1-\lambda(1-\zeta_\lambda))^2}. \tag{2.2}$$

The main ideas of the proof are as following.

- Fix a $k \coloneqq k_n$ for now.
- Then we explore the union of the connected components of the vertices $[k] = \{1, \cdots, k\}$.
- Then we show that for an appropriate $k \longrightarrow \infty$ and with high probability, as $n \longrightarrow \infty$, this union contains the largest connected component $\mathscr{C}_{\max}$, and it cannot be larger than $|\mathscr{C}_{\max}| + (k-1)K \log n$.
- Taking $k = o(n^\nu)$ with $\nu < \frac{1}{2}$, the union of components is equal to $|\mathscr{C}_{\max}| + o(\sqrt{n})$.

- Then as a result, CLT for the union of connected components implies one for $|\mathscr{C}_{\max}|$ (by Slutsky's Theorem).

2.2 Exploration Process Revisited

Define S_t as following,

$$S_t = \begin{cases} k, & t = 0, \\ S_{t-1} + X_t - 1, & t \geq 1, \end{cases} \tag{2.3}$$

where

$$X_t \sim \mathrm{Bin}(n - S_{t-1} - (t-1), p). \tag{2.4}$$

Theorem 2.2 (The law of S_t)**.** *For all $t, k \in [n]$,*

$$S_t + (t - k) \sim \mathrm{Bin}((n-k), 1 - (1-p)^t). \tag{2.5}$$

Proof. Define $N_t \coloneqq n - t - S_t$. Enough to show,

$$N_t \sim \mathrm{Bin}((n-k), (1-p)^t). \tag{2.6}$$

This is immediate from the recursion

$$N_t = N_{t-1} - X_t \sim \mathrm{Bin}(N_{t-1}, (1-p)), \quad N_0 = (n-k). \tag{2.7}$$

□

Theorem 2.3 (The conditional law of S_t)**.** *For all $l, t \in [n]$ satisfying $l \geq t$, conditionally on S_t,*

$$S_l + (l - t) - S_t \sim \mathrm{Bin}((n - t - S_t), 1 - (1-p)^{l-t}). \tag{2.8}$$

Proof. Similarly,

$$N_l \sim \mathrm{Bin}(N_t, (1-p)^{l-t}). \tag{2.9}$$

The theorem immediately follows from the recursion. □

Now, we derive a central limit theorem for $S_{\lfloor nt \rfloor}$. We use the asymptotic mean μ_t and asymptotic variance ν_t to the state the theorem.

Define

$$\mu_t \coloneqq (1 - t - e^{-\lambda t}),\ t \in (0,1). \tag{2.10}$$

$$\nu_t \coloneqq e^{-\lambda t}(1 - e^{-\lambda t}),\ t \in (0,1). \tag{2.11}$$

Theorem 2.4 (Central limit theorem for number of active vertices)**.** *Fix* $k = k_n = o(\sqrt{n})$*. Let* $S_0 = k$*. Then for all* $t \in (0,1)$*,*

$$\frac{S_{\lfloor nt \rfloor} - n\mu_t}{\sqrt{n\nu_t}} \xrightarrow{d} Z \sim N(0,1). \tag{2.12}$$

Proof. By the central limit theorem for binomial distribution,

$$\frac{S_{\lfloor nt \rfloor} - \mathbb{E}[S_{\lfloor nt \rfloor}]}{\sqrt{\mathrm{Var}[S_{\lfloor nt \rfloor}]}} \xrightarrow{d} Z \sim N(0,1). \tag{2.13}$$

We have,

$$\begin{aligned}\mathbb{E}[S_{\lfloor nt \rfloor}] &= (n-k)\left(1 - (1-\lambda/n)^{\lfloor nt \rfloor}\right) - (\lfloor nt \rfloor - k) \\ &= n\mu_t + o(\sqrt{n}),\end{aligned} \tag{2.14}$$

$$\begin{aligned}\mathrm{Var}[S_{\lfloor nt \rfloor}] &= (n-k)(1-\lambda/n)^{\lfloor nt \rfloor}\left(1 - (1-\lambda/n)^{\lfloor nt \rfloor}\right) \\ &= n\nu_t + o(n).\end{aligned} \tag{2.15}$$

since $k = o(\sqrt{n})$. The proof now follows by simple application of Slutsky's Theorem. □

2.3 Why CLT for $|\mathscr{C}([k])|$ is sufficient for CLT for $|\mathscr{C}_{\max}|$?

Theorem 2.5. *whp,*

$$|\mathscr{C}_{\max}| \leq |\mathscr{C}([k])| \leq |\mathscr{C}_{\max}| + (k-1)K\log n, \tag{2.16}$$

if $k = o(\sqrt{n})$ *is appropriately chosen.*

Proof of Upper Bound. There are at maximum k disjoint connected components. One of them may be the maximum component and other components have size $\leq K\log n$ whp. Hence proved. □

Proof of Lower Bound. Consider $G' = G \setminus [k], \mathscr{C}'_{\max} = \mathscr{C}_{\max} \setminus [k]$. Fix any $\epsilon > 0$. So,

$$|\mathscr{C}'_{\max}| = n\zeta_\lambda + o_p(n^{1/2+\epsilon}) \quad \text{[By Law of Large Numbers and } k = o(\sqrt{n})]. \tag{2.17}$$

Now, for $i \in [k]$,

$$\begin{aligned}\mathbb{P}_\lambda\Big[\exists \text{ direct edge between } \mathscr{C}'_{\max} \text{ and } i\text{-th vertex}\Big|\mathscr{C}_{\max}\Big] &= 1-(1-\lambda/n)^{(n\zeta_\lambda+o_p(n^{1/2+\epsilon}))}] \\ &= 1-e^{-\lambda\zeta_\lambda}(1+o_p(1)) \\ &= \zeta_\lambda+o_p(1) \quad \text{[By Theorem 2.6]}.\end{aligned} \tag{2.18}$$

Hence,

$$\mathbb{P}_\lambda[\exists \text{ direct edge between } \mathscr{C}'_{\max} \text{ and } i\text{-th vertex}] = \zeta_\lambda + o_p(1). \tag{2.19}$$

So,

$$\begin{aligned}\mathbb{P}_\lambda[\{\mathscr{C}_{\max} \not\subseteq \mathscr{C}([k])\}] &= \mathbb{P}_\lambda[\nexists\, i \in [k] : i \in \mathscr{C}_{\max}] \\ &\le \mathbb{P}_\lambda[\nexists \text{ direct edge from } \mathscr{C}'_{\max} \text{ to } i\text{-th vertex}] \\ &= (1-\zeta_\lambda+o_p(1))^k \\ &\xrightarrow{k\to\infty} 0.\end{aligned} \tag{2.20}$$

i.e., whp,

$$\mathscr{C}_{\max} \subseteq \mathscr{C}([k]). \tag{2.21}$$

□

So, as a corollary of the previous theorem and specific choice of $k = o(\sqrt{n})$,

$$\frac{|\mathscr{C}([k])| - |\mathscr{C}_{\max}|}{\sqrt{n}} \xrightarrow{\mathbb{P}} 0. \tag{2.22}$$

Hence, for the CLT of giant component, it suffices to show that,

$$\frac{|\mathscr{C}([k])| - \zeta_\lambda n}{\sqrt{n}} \xrightarrow{d} Z \sim N(0, \sigma_\lambda^2). \tag{2.23}$$

2.4 Proof of the Giant Component CLT

Proof of the "giant" component CLT. By exploration process argument,

$$|\mathscr{C}([k])| = \min\{m : S_m = 0\}. \tag{2.24}$$

Let $k \coloneqq k_n = o(\sqrt{n}), k \longrightarrow \infty$.

We want to prove that

$$\frac{|\mathscr{C}([k])| - \zeta_\lambda n}{\sqrt{n}} \xrightarrow{d} Z \sim N(0, \sigma_\lambda^2), \tag{2.25}$$

where,

$$\sigma_\lambda^2 := \frac{\zeta_\lambda(1-\zeta_\lambda)}{(1-\lambda(1-\zeta_\lambda))^2} \tag{2.26}$$

In other words, we want to show,

$$\mathbb{P}\left[\frac{|\mathscr{C}([k])| - n\zeta_\lambda}{\sqrt{n}} > x\right] \longrightarrow 1 - \Phi\left(\frac{x}{\sigma_\lambda}\right)$$

$$\Rightarrow \quad \mathbb{P}\left[|\mathscr{C}([k])| > m_x\right] \longrightarrow 1 - \Phi\left(\frac{x}{\sigma_\lambda}\right), \quad m_x := \left\lfloor n\zeta_\lambda + x\sqrt{n}\right\rfloor. \tag{2.27}$$

To further simplify the expression of variance, as well as for further steps, we claim the following:

Theorem 2.6. *Fix $\lambda > 1$*

$$1 - \zeta_\lambda = e^{-\lambda\zeta_\lambda} \ \textit{i.e.,}\ \mu_{\zeta_\lambda} = 0. \tag{2.28}$$

Moreover ζ_λ (the survival probability of Poisson(λ) Galton-Watson Branching Process) is the unique solution of the equation

$$1 - x = e^{-\lambda x}, x \in (0,1) \tag{2.29}$$

Proof of Claim. The survival equation of a Galton-Watson branching process is $\rho = \psi(\rho)$, where ψ is the probability generating function of the progeny distribution and ρ is extinction probability. For Poisson(λ) Branching Process, $\psi_\lambda(t) = e^{-\lambda(1-t)}$. Since λ (progeny mean) > 1, there exists a unique solution of the equation $\rho = \psi(\rho)$ in (0,1), say ρ_λ. Also $\rho_\lambda = 1 - \zeta_\lambda$. So the equation is $(1 - \zeta_\lambda) = e^{-\lambda\zeta_\lambda}$. $\square$

So,

$$\sigma_\lambda^2 = \frac{\zeta_\lambda(1-\zeta_\lambda)}{(1-\lambda(1-\zeta_\lambda))^2} = \frac{e^{-\lambda\zeta_\lambda}(1-e^{-\lambda\zeta_\lambda})}{(1-\lambda e^{-\lambda\zeta_\lambda})^2} = \frac{\nu_{\zeta_\lambda}}{(1-\lambda e^{-\lambda\zeta_\lambda})^2} \tag{2.30}$$

Now we proceed in two steps, which suffices the proof,

Upper bound of probability:

$$\limsup_{n\to\infty} \mathbb{P}\Big[\frac{|\mathscr{C}([k])| - \zeta_\lambda n}{\sqrt{n}} > x\Big] \le 1 - \Phi(x/\sigma_\lambda). \tag{2.31}$$

Lower bound of probability:

$$\liminf_{n\to\infty} \mathbb{P}\Big[\frac{|\mathscr{C}([k])| - \zeta_\lambda n}{\sqrt{n}} > x\Big] \ge 1 - \Phi(x/\sigma_\lambda). \tag{2.32}$$

Proof of Upper Bound. We know,

$$\begin{aligned}
\mathbb{P}_\lambda\left[\frac{|\mathscr{C}([k])| - n\zeta_\lambda}{\sqrt{n}} > x\right] &= \mathbb{P}_\lambda[|\mathscr{C}([k])| > m_x] \quad [m_x \coloneqq \lfloor n\zeta_\lambda + x\sqrt{n}\rfloor] \\
&= \mathbb{P}_\lambda[S_m > 0 \ \forall\ m \le m_x] \\
&\le \mathbb{P}_\lambda[S_{m_x} > 0].
\end{aligned} \tag{2.33}$$

Now

$$\begin{aligned}
\mathbb{E}[S_{m_x}] &= \mathbb{E}[S_{\lfloor n\zeta_\lambda + x\sqrt{n}\rfloor}] \\
&= n\mu_{(\lfloor n\zeta_\lambda + x\sqrt{n}\rfloor)/n} + o(\sqrt{n}) \\
&= n\left(\mu_{\zeta_\lambda} + x\mu'_{\zeta_\lambda}/\sqrt{n} + o(1/\sqrt{n})\right) + o(\sqrt{n}) \quad [\because t \mapsto \mu_t \text{ is differentiable}] \\
&= \sqrt{n}x(\lambda e^{-\lambda\zeta_\lambda} - 1) + o(\sqrt{n}) \quad [\because \mu_{\zeta_\lambda} = 0].
\end{aligned} \tag{2.34}$$

Also,

$$\mathrm{Var}[S_{m_x}] = n\nu_{\zeta_\lambda} + o(n) \quad (\nu_{\zeta_\lambda} > 0), \tag{2.35}$$

So,

$$\begin{aligned}
\mathbb{P}_\lambda\Big[S_{m_x} > 0\Big] &= \mathbb{P}_\lambda\left[\frac{S_{m_x} - \mathbb{E}[S_{m_x}]}{\sqrt{\mathrm{Var}[S_{m_x}]}} > \frac{x(1 - \lambda e^{-\lambda\zeta_\lambda})}{\sqrt{\nu_{\zeta_\lambda}}}\right] \\
&\longrightarrow 1 - \Phi\left(\frac{x(1 - \lambda e^{-\lambda\zeta_\lambda})}{\sqrt{\nu_{\zeta_\lambda}}}\right) \quad [\text{By Theorem 2.4}] \\
&= 1 - \Phi\left(\frac{x}{\sigma_\lambda}\right).
\end{aligned} \tag{2.36}$$

□

Proof of Lower Bound. Fix $\epsilon > 0$.

$$\begin{aligned}\mathbb{P}_\lambda\left[\frac{|\mathscr{C}([k])| - n\zeta_\lambda}{\sqrt{n}} > x\right] &= \mathbb{P}_\lambda[S_m > 0 \; \forall \; m \leq m_x] \\ &\geq \mathbb{P}_\lambda[S_m > 0 \; \forall \; m \leq m_x, S_{m_x} > \epsilon\sqrt{n}] \\ &= \mathbb{P}_\lambda[S_{m_x} > \epsilon\sqrt{n}] - \mathbb{P}_\lambda[S_{m_x} > \epsilon\sqrt{n}, \; \exists m < m_x : Sm = 0].\end{aligned} \tag{2.37}$$

Using Theorem 2.4,

$$\mathbb{P}_\lambda[S_{m_x} > \epsilon\sqrt{n}] \longrightarrow \Phi\left(\frac{x(1-\lambda e^{-\lambda\zeta_\lambda}) + \epsilon}{\sqrt{\nu_{\zeta_\lambda}}}\right). \tag{2.38}$$

We claim, $\mathbb{P}_\lambda[S_{m_x} > \epsilon\sqrt{n}, \; \exists m < m_x : S_m = 0] \longrightarrow 0$ For the second term in lower bound,

$$\begin{aligned}\mathbb{P}_\lambda[S_{m_x} > \epsilon\sqrt{n}, \; \exists m < m_x : S_m = 0] &\leq \sum_{m=k}^{m_x-1} \mathbb{P}_\lambda[S_m = 0, S_{m_x} > \epsilon\sqrt{n}] \\ &\leq \sum_{m=k}^{\lfloor \alpha n \rfloor} \mathbb{P}_\lambda[S_m = 0] + \sum_{m=\lfloor \alpha n \rfloor}^{m_x-1} \mathbb{P}_\lambda[S_m = 0, S_{m_x} > \epsilon\sqrt{n}].\end{aligned} \tag{2.39}$$

Remark: For $m \leq \alpha n$, we use a Chernoff bound. For $m \geq \alpha n$, we use another Chernoff bound. Intuitively, the probability that $S_m = 0$ yet $S_{m_x} > \epsilon\sqrt{n}$ is small for such a large n, where it behaves like random walk with negative drift [$\mathbb{E}_\lambda[X_m] < 1$ for m close to $\zeta_\lambda n$].

For the first part,

$$\begin{aligned}\sum_{m=k}^{\lfloor \alpha n \rfloor} \mathbb{P}_\lambda[S_m = 0] &\leq \sum_{m=k}^{\lfloor \alpha n \rfloor} e^{-mJ(m/n,\lambda)} \quad \text{[By Theorem A.7]} \\ &\longrightarrow 0.\end{aligned} \tag{2.40}$$

Now observe, conditionally on $S_m = 0$,

$$S_{m_x} + (m_x - m) \sim \text{Bin}\left((n-m), 1 - \left(1 - \frac{\lambda}{n}\right)^{m_x - m}\right). \tag{2.41}$$

So,

$$
\begin{aligned}
\mathbb{P}_\lambda[S_m = 0, S_{m_x} > \epsilon\sqrt{n}] &= \mathbb{P}_\lambda[S_{m_x} > \epsilon\sqrt{n} | S_m = 0]\mathbb{P}_\lambda[S_m = 0] \\
&\leq \mathbb{P}_\lambda[S_{m_x} > (m_x - m) + \epsilon\sqrt{n} | S_m = 0] \\
&= \mathbb{P}_\lambda[X > (m_x - m) + \epsilon\sqrt{n}]. \qquad (2.42) \\
&\left[\text{Assume, } X \sim \text{Bin}\left((n-m), 1 - \left(1 - \frac{\lambda}{n}\right)^{m_x - m}\right)\right]
\end{aligned}
$$

We fix $\alpha = \zeta_\lambda - \epsilon$ for some ϵ sufficiently small. So, for $m \geq \alpha n$

$$
\begin{aligned}
\mathbb{E}_\lambda X &= (n-m)\left[1 - \left(1 - \frac{\lambda}{n}\right)^{m_x - m}\right] \\
&\leq (1-\alpha)n\left[1 - \left(1 - \frac{\lambda}{n}\right)^{m_x - m}\right] \\
&\leq (1 - \zeta_\lambda + \epsilon)n\frac{\lambda(m_x - m)}{n} \qquad (2.43) \\
&= \lambda(1 - \zeta_\lambda + \epsilon)(m_x - m) \\
&\leq (1 - \epsilon)(m_x - m) \\
&[\text{Using } \lambda(1 - \zeta_\lambda) = \lambda e^{-\lambda\zeta_\lambda} < 1 \text{ as } \mu'_{\zeta_\lambda} < 0 \text{ and } \epsilon \text{ sufficiently small}].
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}_\lambda[S_m = 0, S_{m_x} > \epsilon\sqrt{n}] &= \mathbb{P}_\lambda[X > (m_x - m) + \epsilon\sqrt{n}] \\
&\leq \mathbb{P}_\lambda[X - \mathbb{E}_\lambda X > \epsilon((m_x - m) + \sqrt{n})] \\
&= \mathbb{P}_\lambda[X - \mathbb{E}_\lambda X > t] \quad [\text{Let, } t = \epsilon((m_x - m) + \sqrt{n})] \\
&\leq \exp\left(-\frac{t^2}{2((1-\epsilon)(m_x - m) + \frac{t}{3})}\right) \quad [\text{By Theorem A.8}]. \qquad (2.44)
\end{aligned}
$$

We split, depending on whether $m_x - m \geq \epsilon\sqrt{n}$ or $m_x - m \leq \epsilon\sqrt{n}$.

For $m_x - m \geq \epsilon\sqrt{n}$, and since $t \geq \epsilon(m_x - m)$ and $(1 - \epsilon) + \epsilon/3 \leq 1$, we have,

$$
\begin{aligned}
\mathbb{P}_\lambda(S_m = 0,\, S_{mx} > \epsilon\sqrt{n}) &\le \exp\left(-\frac{t}{2\left(\frac{(1-\epsilon)(m_x-m)}{t} + \frac{1}{3}\right)}\right) \\
&\le \exp\left(-\frac{t}{2\left(\frac{(1-\epsilon)}{\epsilon} + \frac{1}{3}\right)}\right) \\
&\le e^{-\epsilon t/2} \\
&\le \exp\left(-\frac{\epsilon^2(m_x-m)}{2}\right) \\
&\le \exp\left(-\frac{\epsilon^2\sqrt{n}}{2}\right).
\end{aligned} \tag{2.45}
$$

For $m_x - m \le \epsilon\sqrt{n}$, since $t \ge \epsilon\sqrt{n}$, we have,

$$
\begin{aligned}
\mathbb{P}[S_m = 0,\, S_{mx} > \epsilon\sqrt{n}] &\le \exp\left(-\frac{t^2}{2\left((1-\epsilon)\sqrt{n} + \frac{t}{3}\right)}\right) \\
&\le \exp\left(-\frac{t^2}{2\left((1-\epsilon)t + \frac{t}{3}\right)}\right) \\
&\le \exp\left(-\frac{3t}{8}\right) \\
&\le \exp\left(-\frac{3\epsilon\sqrt{n}}{8}\right).
\end{aligned} \tag{2.46}
$$

So,

$$
\sum_{m=\lfloor \alpha n \rfloor}^{m_x-1} \mathbb{P}_\lambda[S_m = 0, S_{m_x} > \epsilon\sqrt{n}] \le n[\exp(-\epsilon^2\sqrt{n}/2) + \exp(-3\epsilon\sqrt{n}/4)] \longrightarrow 0. \tag{2.47}
$$

□

Now using equation 2.37, 2.38, 2.40, 2.47, and $\epsilon \downarrow 0$ as $\epsilon > 0$ is arbitrary, we obtain the result. □

Appendix A

Some additional results

Theorem A.1 (Tail bound on Poisson GWBP extinction)**.** *For a branching process with i.i.d. offspring X having mean $\mu = \mathbb{E}X > 1$,*

$$\mathbb{P}[k \le T < \infty] \le \frac{e^{-Ik}}{1 - e^{-I}}, \tag{A.1}$$

where the exponential rate I is given by

$$I = \sup_{t \le 0}(t - \log \mathbb{E}[e^{tX}]) > 0. \tag{A.2}$$

Theorem A.2 (Differentiability of survival function)**.** *Let ζ_λ denote the survival probability of Poisson(λ) Galton-Watson branching process. Then, for all $\lambda > 1$,*

$$\frac{d}{d\lambda}\zeta_\lambda = \frac{\zeta_\lambda(1 - \zeta_\lambda)}{1 - \mu_\lambda} \in (0, \infty), \tag{A.3}$$

where $\mu_\lambda < 1, \mu_\lambda e^{-\mu_\lambda} = \lambda e^{-\lambda}$ (uniquely). When $\lambda \downarrow 1$,

$$2(\lambda - 1)(1 + o(1)). \tag{A.4}$$

Theorem A.3 (Link between Poisson and binomial GWBP)**.** *For* $\mathrm{Bin}(n, p)$ *GWBP with $np = \lambda$ and Poisson(λ) GWBP, for each $k \ge 1$,*

$$\mathbb{P}_{n,p}[T \ge k] = \mathbb{P}^*_\lambda[T^* \ge k] + O(k\lambda^2/n), \tag{A.5}$$

where T and T^* are total progeny of the binomial and Poisson branching processes, respectively.

Theorem A.4 (Stochastic domination on $|\mathscr{C}(1)|$)**.** *Let $\lambda = np$. For each $k \geq 1$,*

$$\mathbb{P}_\lambda[|\mathscr{C}(1)| \geq k] \leq \mathbb{P}_{n,p}[T^{\geq} \geq k], \tag{A.6}$$

where $T^{\geq}$ is the total progeny of a binomial branching process with parameter n and p, and $\mathbb{P}_\lambda$ denotes the law of $ER_n(p)$.

Theorem A.5 (Lower bound on cluster tail)**.** *Let $\lambda = np$. For each $k \in [n]$,*

$$\mathbb{P}_\lambda[|\mathscr{C}(1)| \geq k] \geq \mathbb{P}_{n-k,p}[T^{\leq} \geq k], \tag{A.7}$$

where $T^{\leq}$ is the total progeny of a binomial branching process with parameter $n-k$ and p, and $\mathbb{P}_\lambda$ denotes the law of $ER_n(p)$.

Theorem A.6 (Chernoff Bound for Probability that $S_m = 0$ with $S_0 = 1$)**.** *Let S_t be defined as in 1.4. Then for $t < \zeta_\lambda n$,*

$$\mathbb{P}_\lambda[S_t = 0] \leq e^{-tJ(\alpha;\lambda)}, \tag{A.8}$$

where,

$$J(\alpha, \lambda) = I_{(1-e^{-\lambda\alpha})/\alpha}. \tag{A.9}$$

Theorem A.7 (Chernoff Bound for Probability that $S_m = 0$ with $S_0 = k$)**.** *Let S_t be defined as in 2.2. Then for $\alpha < \zeta_\lambda, m \leq \alpha n$,*

$$\mathbb{P}[S_m = 0] \leq e^{-mJ(m/n,\lambda)}, \tag{A.10}$$

where,

$$J(\alpha, \lambda) = I_{(1-e^{-\lambda\alpha})/\alpha} \tag{A.11}$$

Theorem A.8 (Upper Deviation Chernoff Bound for Binomial)**.** *Let $X \sim Bin(.)$. Then*

$$\mathbb{P}[X \geq \mathbb{E}X + t] \leq \exp\left(-\frac{t^2}{2(\mathbb{E}X + t/3)}\right), \quad t \geq 0. \tag{A.12}$$

Bibliography

[1] R. van der Hofstad. *Random Graphs and Complex Networks*. Cambridge Series in Statistical and Probabilistic Mathematics v. 1. Cambridge University Press, 2017. ISBN: 9781107172876. URL: https://books.google.co.in/books?id=1baSDQAAQBAJ.