Central Limit Theorem for the Largest Component of Random Intersection Graph

Liang Dong Zhishui Hu

International Institute of Finance, School of Management
University of Science and Technology of China
Heifei 230026, Anhui, China

dl040@mail.ustc.edu.cn huzs@ustc.edu.cn

Submitted: Sep 7, 2021; Accepted: Mar 15, 2022; Published: May 20, 2022

Abstract

Random intersection graphs are models of random graphs in which each vertex is assigned a subset of objects independently and two vertices are adjacent if their assigned subsets are adjacent. Let n and $m = [\beta n^{\alpha}]$ denote the number of vertices and objects respectively. We get a central limit theorem for the largest component of the random intersection graph $G(n, m, p)$ in the supercritical regime and show that it changes between $\alpha > 1$, $\alpha = 1$ and $\alpha < 1$.

Mathematics Subject Classifications: 05C80; 60F05

1 Introduction

Given positive integers n and m, let $\mathbf{V} = \{v_1, v_2, \ldots, v_n\}$ and $\mathbf{W} = \{w_1, w_2, \ldots, w_m\}$. For $p \in [0, 1]$, we construct a random bipartite graph $B(n, m, p)$ with bipartition $(\mathbf{V}, \mathbf{W})$ in which each one of the nm possible edges between vertices from $\mathbf{V}$ and vertices from $\mathbf{W}$ is occupied independently with probability p. The random intersection graph model $G(n, m, p)$ is a graph with vertex set $\mathbf{V}$ in which $v_i, v_j \in \mathbf{V}$ are adjacent if and only if there exists some $w \in \mathbf{W}$ so that both v_i and v_j are adjacent to w in $B(n, m, p)$.

The random intersection graph $G(n, m, p)$ was introduced by Singer [14] and Karoński, Scheinerman and Singer-Cohen [9] and has been further studied and generalized by Godehardt and Jaworski [6], Stark [15], Barbour and Reinert [2], Bloznelis [4]. Random intersection graph also has been used in various applications. These applications include, but are not restricted to, secure wireless sensor networks [12], social networks [1] and circuit design [14].

Erdős-Rényi random graph model $G(n, p)$ considers a fixed set of n vertices and edges that exist with a certain probability p independently of all other edges. Fill, Scheinerman and Singer-Cohen [5] showed that the total variation distance between $G(n, m, p)$ and

$G(n,\hat{p})$ tends to 0 for any $0 \leqslant p = p(n) \leqslant 1$ if $m = [n^\alpha]$ with $\alpha > 6$, where $\hat{p}$ is chosen so that the expected numbers of edges in the two graphs are the same, i.e., $\hat{p} = 1-(1-p^2)^m$. Kim, Lee and Na [10] proved that the total variation distance still tends to 0 for any $0 \leqslant p = p(n) \leqslant 1$ whenever $m \gg n^4$.

Let $\mathcal{N}(G(n,m,p))$ denote the number of the largest component of the random intersection graph $G(n,m,p)$. In this paper, we assume that $m = [\beta n^\alpha]$ and $nmp^2 = \lambda$, where α, β, λ are fixed positive constants.

Behrisch [3] studied $\mathcal{N}(G(n,m,p))$ for $\alpha \neq 1, \beta = 1$ and $\lambda \neq 1$. Lagerås and Lindholm [11] considered $\mathcal{N}(G(n,m,p))$ when $\alpha = 1$ and $\lambda \neq 1$. In the supercritical regime, i.e., $mnp^2 = \lambda > 1$, Behrisch [3] and Lagerås and Lindholm [11] derived the following weak law of large numbers:

$$\frac{\mathcal{N}(G(n,m,p))}{b_n} \xrightarrow{p} 1-b, \tag{1}$$

where

$$b_n = \begin{cases} n, & \alpha \geqslant 1, \\ nmp, & \alpha < 1, \end{cases} \qquad b = \begin{cases} \rho, & \alpha \neq 1, \\ \rho_\beta, & \alpha = 1, \end{cases} \tag{2}$$

$\rho \in (0,1)$ is the smallest nonnegative solution to

$$x = \exp(\lambda(x-1)), \tag{3}$$

and $\rho_\beta \in (0,1)$ is the smallest nonnegative solution to

$$x = \exp\Big\{\sqrt{\lambda\beta}\Big(e^{\sqrt{\lambda/\beta}(x-1)} - 1\Big)\Big\}. \tag{4}$$

The aim of this paper is to establish a central limit theorem for $\mathcal{N}(G(n,m,p))$ in the supercritical regime. Our main result is stated as follows.

Theorem 1. *Assume that $m = [\beta n^\alpha]$ and $nmp^2 = \lambda > 1$. Let $\zeta_{n,m,p} \in (0,1)$ be the unique positive solution to*

$$xb_n/n + \exp\Big\{-mp\Big(1 - e^{-xb_np}\Big)\Big\} = 1, \tag{5}$$

and let

$$\sigma^2 = \begin{cases} \big(\lambda c(1-c)\rho_\beta^2 + \rho_\beta(1-\rho_\beta)\big)(1-\lambda c\rho_\beta)^{-2}, & \alpha = 1, \\ \rho(1-\rho)(1-\lambda\rho)^{-2}, & \alpha > 1, \\ \lambda\rho(1-\rho)(1-\lambda\rho)^{-2}, & \alpha < 1, \end{cases}$$

where $c = e^{(\rho_\beta - 1)\sqrt{\lambda/\beta}}$. Then, for $\alpha > 1/2$, we have

$$\frac{\mathcal{N}(G(n,m,p)) - \zeta_{n,m,p}b_n}{\sqrt{n}} \xrightarrow{d} N(0,\sigma^2).$$

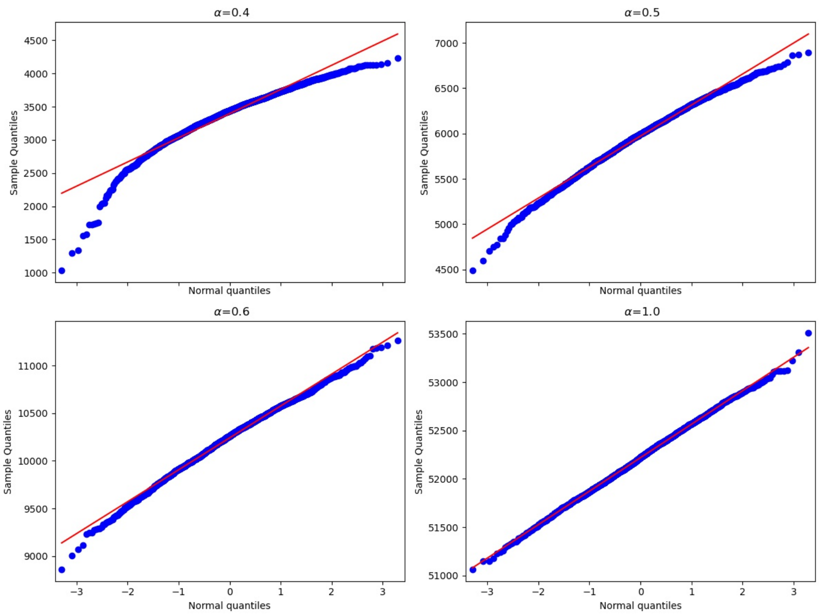

Figure 1: Q-Q plots of sample data versus normal distribution.

Remark 2. For $\alpha < 1$, the largest component of $G(n,m,p)$ is $O_P(b_n) = O_P(n^{(\alpha+1)/2})$ (see 1), and the second largest component of $G(n,m,p)$ has size $O_P(\sqrt{n/m}\log^2 n) = O_P(n^{(1-\alpha)/2}\log^2 n)$ (see Theorem 12 in [8]), the number of the largest component and the second largest component are both close to $\sqrt{n}$ for small α. Therefore, we guess that the central limit theorem for $\mathcal{N}(G(n,m,p)$ doesn't hold for small positive α.

Through numerical simulations, we obtain the number of the largest component in random intersection graph models, all on $n = 100000$ vertices, with $\beta = 1$, $\lambda = 2$ and different α. For each model, we take 2000 replications and our results are shown in Figure 1. The Q-Q plots show samples' quantiles compared to the normal distribution. When $\alpha < 0.5$, the Q-Q plot shows that the points do not align along a line. While when looking at the Q-Q plots for $\alpha > 0.5$, we see the points match up along a straight line which shows that the quantiles match. This leads us to reason that the largest component is most likely asymptoticallynormallydistributed when $\alpha > 0.5$. Q-Q plots show that the limit distributions of order of the largest components change greatly near $\alpha = 0.5$.

Remark 3. From Lemma 7 in Section 2, we always have $\sigma^2 > 0$ for $\lambda > 1$.

Remark 4. We write

$$f(x) = \exp\left\{mp\left(e^{-xnp} - 1\right)\right\} + x - 1,$$

then $f''(x) > 0$ for $x \in \mathbb{R}$. This implies that $f(x)$ is a convex function. Note that $f(0) = 0, f(1) > 0$ and $f'(0) = -\lambda + 1 < 0$. The equation $f(x) = 0$ has only one non-zero solution $x_{n,m,p} \in (0,1)$. Hence $\zeta_{n,m,p} := x_{n,m,p} n/b_n$ is the unique positive solution to (5) and $\zeta_{n,m,p} = x_{n,m,p} \in (0,1)$ for $\alpha \geqslant 1$. As for $\alpha < 1$, by noting that $mp \to 0$ and applying the inequality $1-e^{-x} \leqslant x$ for all $x \in \mathbb{R}$, we also have

$$0<\zeta_{n,m,p}=\frac{1-\exp\big\{-mp(1-e^{-\zeta_{n,m,p}\lambda})\big\}}{mp} \leqslant 1-e^{-\zeta_{n,m,p}\lambda}<1.$$

Furthermore, if $\alpha>1$, then $np \to 0$ and

$$\log(1-\zeta_{n,m,p})=-mp\Big(1-e^{-\zeta_{n,m,p}np}\Big) \sim -\lambda\zeta_{n,m,p}.$$

If $\alpha<1$, then $mp \to 0$ and

$$\zeta_{n,m,p}=\frac{1-\exp\big\{-mp(1-e^{-\zeta_{n,m,p}\lambda})\big\}}{mp} \sim 1-e^{-\zeta_{n,m,p}\lambda}.$$

Therefore by some basic calculations, we can get that $\zeta_{n,m,p} \to 1-\rho$ for $\alpha \neq 1$.

If $\alpha=1$, then we have $np \to \sqrt{\lambda/\beta},\ mp \to \sqrt{\lambda\beta}$ and

$$\begin{aligned}\log(1-\zeta_{n,m,p}) &= -mp(1-e^{-\zeta_{n,m,p}np}) \\ &= -mp(1-e^{-\zeta_{n,m,p}\sqrt{\lambda/\beta}})\Big(1+\frac{1-e^{\zeta_{n,m,p}(\sqrt{\lambda/\beta}-np)}}{e^{\zeta_{n,m,p}\sqrt{\lambda/\beta}}-1}\Big) \\ &\sim -\sqrt{\lambda\beta}(1-e^{-\zeta_{n,m,p}\sqrt{\lambda/\beta}}).\end{aligned}$$

We can also get that $\zeta_{n,m,p} \to 1-\rho_\beta$ for $\alpha=1$.

Combining the above facts, we always have

$$\zeta_{n,m,p} \to 1-b. \tag{6}$$

Hence the weak law of large numbers (1) is an immediate consequence of Theorem 1.

The basic idea of the proof of Theorem 1 follows from the proof of corresponding result for Erdős-Rényi random graph $G(n,p)$ (see, for instance, Chapter 4 in [13]). In section 2, we construct a related random variable S_t, get a central limit theorem for S_t and estimate the probability that $S_t=0$. The proof of Theorem 1 is given in Sections 3. Throughout this paper, all limits are taken as $n\to\infty$ and $m\to\infty$. We denote by $a\wedge b := \min\{a,b\}$ for any $a,b\in\mathbb{R}$.

2 Preliminaries

For two vertices $v, v' \in \mathbf{V}$, we write $v \longleftrightarrow v'$ when there exists a path of occupied edges connecting v and v' in $G(n,m,p)$. For $v \in \mathbf{V}$, we denote the connected component containing v by

$$\mathcal{C}(v)=\{v' \in \mathbf{V} : v \longleftrightarrow v'\}.$$

Assume that $\mathbf{V}=\{v_1,v_2,\dots,v_n\}$. Fix $k=k_{n,m,p} \leqslant n$, which will be chosen later on, and let

$$\mathcal{C}_k=\cup_{i=1}^k \mathcal{C}(v_i).$$

To study the growth of $\mathcal{C}_k$, we consider the random bipartite graph $B(n,m,p)$ with bipartition $(\mathbf{V},\mathbf{W})$ which is defined in Section 1. For any $v \in \mathbf{V}$ and $w \in \mathbf{W}$, if v and w are adjacent in $B(n,m,p)$, then we set $\eta_{v,w}=1$, otherwise $\eta_{v,w}=0$.

In the beginning, we construct $U_t, \bar{U}_t, V_t$ and W_t recursively with $U_t, V_t \subset \mathbf{V}$ and $\bar{U}_t, W_t \subset \mathbf{W}$ for $t = 0,1,2,\ldots$. U_t and $\bar{U}_t$ are the sets of active vertices which are investigated at t in $\mathbf{V}$ and $\mathbf{W}$ respectively, V_t and W_t are the unexplored vertices in $\mathbf{V}$ and $\mathbf{W}$ respectively. For $t=0$, we let $U_0=\{v_1,\cdots,v_k\}$, $\bar{U}_0=\varnothing$, $V_0=\mathbf{V}-U_0$ and $W_0=\mathbf{W}$. If $U_t \neq \varnothing$, then we pick v_{i_t} from U_t according to some rule that is measurable with respect to $\mathscr{F}_t=\sigma(U_0,\ldots,U_t)$ and let

$$\begin{aligned} U_{t+1} &= (U_t - \{v_{i_t}\}) \cup N_{t+1}, \\ V_{t+1} &= V_t - N_{t+1}, \\ W_{t+1} &= W_t - \bar{U}_{t+1}, \end{aligned}$$

where

$$\bar{U}_{t+1} = \left\{ w \in W_t : \eta_{v_{i_t},w}=1 \right\}, \qquad N_{t+1} = \left\{ v \in V_t : \eta_{v,w}=1 \text{ for some } \ w \in \bar{U}_{t+1} \right\}.$$

At time $\tau=\inf\{t : U_t=\varnothing\}$ the process stops. For $t \geqslant 0$, let $S_t=|U_t|$. This implies that

$$|\mathcal{C}_k| \stackrel{d}{=} \min\{t : S_t = 0\}. \tag{7}$$

Lemma 5. *For $t \geqslant 0$, let $H_t=\sum_{i=0}^{t}|\bar{U}_i|$. Then we have*

$$H_t \sim \mathrm{B}(m, 1-(1-p)^t), \tag{8}$$

and conditionally on H_t,

$$S_t + t - k \sim \mathrm{B}(n-k, 1-(1-p)^{H_t}). \tag{9}$$

Moreover, for $0 \leqslant l \leqslant t \leqslant n$,

$$H_t - H_l \sim \mathrm{B}(m, (1-p)^l - (1-p)^t), \tag{10}$$

and conditionally on H_l, H_t and S_l,

$$S_t - S_l + (t-l) \sim \mathrm{B}(n-l-S_l, 1-(1-p)^{H_t-H_l}). \tag{11}$$

Proof. Conditionally on H_t, we have

$$m - H_{t+1} = m - H_t - |\bar{U}_{t+1}| = m - H_t - \mathrm{B}(m-H_t, p) \sim \mathrm{B}(m-H_t, 1-p).$$

Note the fact that if $N \sim \mathrm{B}(n,p)$, and coditionally on N, $M \sim \mathrm{B}(N,q)$, then $M \sim \mathrm{B}(n,pq)$. For any $0 \leqslant l \leqslant t \leqslant n$, we have $m - H_l \sim \mathrm{B}(m,(1-p)^l)$ and conditionally on H_l,

$$\begin{aligned} m - H_t &\sim \mathrm{B}(m-H_l,\ (1-p)^{t-l}) \\ &\sim m - H_l - \mathrm{B}(m-H_l,\ 1-(1-p)^{t-l}). \end{aligned}$$

Hence, conditionally on H_l,

$$H_t - H_l \sim \mathrm{B}(m - H_l,\ 1 - (1-p)^{t-l}).$$

This implies (10).

Conditionally on S_t and $|\bar{U}_{t+1}|$, we have $|N_{t+1}| \sim B(n-t-S_t, 1-(1-p)^{|\bar{U}_{t+1}|})$ and then

$$n - t - S_{t+1} = n - t - S_t - |N_{t+1}| \sim B(n - t - S_t, (1-p)^{|\bar{U}_{t+1}|}).$$

Therefore for any $0 \leqslant l \leqslant t \leqslant n$, conditionally on $H_t,\ H_l$ and S_l,

$$\begin{aligned} n - t - S_t &\sim \mathrm{B}(n - l - S_l, (1-p)^{H_t - H_l}) \\ &\sim n - l - S_l - \mathrm{B}(n - l - S_l, 1 - (1-p)^{H_t - H_l}). \end{aligned}$$

This proves (11).

By taking $l = 0$, (8) and (9) follow from (10) and (11), respectively. The proof of Lemma 5 is completed. □

Let $\gamma \in \mathbb{R}$ and $\theta \in (0,1)$. Assume that $\{\theta_n, n \geqslant 1\}$ is a sequence of real numbers such that $\theta_n \in (0,1)$ and $\theta_n \to \theta$. Define

$$\mu_n := \Big(1 - \theta_n b_n/n - \exp\Big\{-mp\Big(1 - e^{-\theta_n b_n p}\Big)\Big\}\Big)n + \gamma\mu(\theta)\sqrt{n},$$

where

$$\mu(\theta) := \begin{cases} \lambda e^{-\theta\sqrt{\lambda/\beta} - c_0(\theta)} - 1, & \alpha = 1, \\ \lambda e^{-\theta\lambda} - 1, & \alpha \neq 1, \end{cases} \tag{12}$$

and

$$c_0(\theta) := \sqrt{\lambda\beta}(1 - e^{-\theta\sqrt{\lambda/\beta}}).$$

Let

$$\nu(\theta) := \begin{cases} \sqrt{\lambda/\beta}e^{-\theta\sqrt{\lambda/\beta}}c_0(\theta)e^{-2c_0(\theta)} + e^{-c_0(\theta)}(1 - e^{-c_0(\theta)}), & \alpha = 1, \\ e^{-\theta\lambda}(1 - e^{-\theta\lambda}), & \alpha > 1, \\ \lambda e^{-\theta\lambda}(1 - e^{-\theta\lambda}), & \alpha < 1. \end{cases} \tag{13}$$

Lemma 6. *Assume that $m = [\beta n^\alpha],\ nmp^2 = \lambda$ and $k = o(\sqrt{n})$. Then we have*

$$\frac{S_{[\theta_n b_n + \gamma\sqrt{n}]} - \mu_n}{\sqrt{n\nu(\theta)}} \xrightarrow{d} N(0,1), \tag{14}$$

where b_n is defined in (2)

Proof. Let $a_n := [\theta_n b_n + \gamma\sqrt{n}]$ and

$$\begin{aligned} A_n &:= -a_n + k + (n-k)(1-(1-p)^{\mathbb{E}(H_{a_n})}), \\ \widetilde{A}_n &:= -a_n + k + (n-k)(1-(1-p)^{H_{a_n}}), \\ B_n^2 &:= (1-(1-p)^{H_{a_n}})(1-p)^{H_{a_n}}. \end{aligned}$$

By (9) and applying the classical Berry-Esseen inequality for the binomial distribution, there exists some absolute constant $C > 0$ such that

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big(\frac{S_{a_n} - \widetilde{A}_n}{\sqrt{n-k}B_n} \leqslant x \Big| H_{a_n} \Big) - \Phi(x) \right| \leqslant \frac{C}{B_n\sqrt{n-k}}. \tag{15}$$

This implies that

$$\begin{aligned} & \sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big(\frac{S_{a_n} - A_n}{\sqrt{n-k}B_n} \leqslant x \Big| H_{a_n} \Big) - \Phi\Big(x + \frac{A_n - \widetilde{A}_n}{\sqrt{n-k}B_n} \Big) \right| \\ = & \sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big(\frac{S_{a_n} - \widetilde{A}_n}{\sqrt{n-k}B_n} \leqslant x + \frac{A_n - \widetilde{A}_n}{\sqrt{n-k}B_n} \Big| H_{a_n} \Big) - \Phi\Big(x + \frac{A_n - \widetilde{A}_n}{\sqrt{n-k}B_n} \Big) \right| \\ \leqslant & \frac{C}{B_n\sqrt{n-k}} \wedge 1. \end{aligned}$$

Hence

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big(\frac{S_{a_n} - A_n}{\sqrt{n-k}B_n} \leqslant x \Big) - \mathbb{E}\Big(\Phi\Big(x + \frac{A_n - \widetilde{A}_n}{\sqrt{n-k}B_n} \Big) \Big) \right| \leqslant \mathbb{E}\Big(\frac{C}{B_n\sqrt{n-k}} \wedge 1 \Big).$$

If $(\sqrt{n}B_n)^{-1} \xrightarrow{p} 0$, then we have

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big(\frac{S_{a_n} - A_n}{\sqrt{n-k}B_n} \leqslant x \Big) - \mathbb{E}\Big(\Phi\Big(x + \frac{A_n - \widetilde{A}_n}{\sqrt{n-k}B_n} \Big) \Big) \right| \to 0. \tag{16}$$

Note that

$$\begin{aligned} 1-(1-p)^{a_n} &= 1 - e^{a_n \ln(1-p)} = 1 - e^{-a_n p} + O(p) \\ &= 1 - e^{-\theta_n b_n p - \gamma\sqrt{\lambda}/\sqrt{m}} + O(p) \\ &= 1 - e^{-\theta_n b_n p}\Big(1 - \frac{\gamma\sqrt{\lambda}}{\sqrt{m}} \Big) + O(p + m^{-1}). \end{aligned}$$

By (8), we have

$$\mathbb{E}(H_{a_n}) = m(1-(1-p)^{a_n}) \sim \begin{cases} m(1-e^{-\theta\sqrt{\lambda/\beta}}), & \alpha = 1, \\ \theta nmp, & \alpha > 1, \\ m(1-e^{-\theta\lambda}), & \alpha < 1, \end{cases} \tag{17}$$

and

$$\mathrm{Var}(H_{a_n}) = m(1-p)^{a_n}(1-(1-p)^{a_n}) \sim \begin{cases} me^{-\theta\sqrt{\lambda/\beta}}(1-e^{-\theta\sqrt{\lambda/\beta}}), & \alpha = 1, \\ \theta nmp, & \alpha > 1, \\ me^{-\theta\lambda}(1-e^{-\theta\lambda}), & \alpha < 1. \end{cases} \tag{18}$$

Therefore,

$$(1-p)^{\mathbb{E}(H_{a_n})} \to e^{-c_1}, \qquad p\mathrm{Var}(H_{a_n}) \to c_2, \tag{19}$$

where

$$c_1 = c_1(\theta) := \begin{cases} \sqrt{\lambda\beta}(1-e^{-\theta\sqrt{\lambda/\beta}}), & \alpha = 1, \\ \theta\lambda, & \alpha > 1, \\ 0, & \alpha < 1, \end{cases}$$

and

$$c_2 = c_2(\theta) := \begin{cases} \sqrt{\lambda\beta}e^{-\theta\sqrt{\lambda/\beta}}(1-e^{-\theta\sqrt{\lambda/\beta}}), & \alpha = 1, \\ \theta\lambda, & \alpha > 1, \\ 0, & \alpha < 1. \end{cases}$$

Furthermore, we get that

$$\begin{aligned} (1-p)^{\mathbb{E}(H_{a_n})} &= \exp\{-mp(1-(1-p)^{a_n})\} + O(p) \\ &= \exp\Big\{-mp\Big(1-e^{-\theta_n b_n p}\Big(1-\frac{\gamma\sqrt{\lambda}}{\sqrt{m}}\Big)\Big)\Big\} + O(p+n^{-1}) \\ &= \exp\Big\{-mp\Big(1-e^{-\theta_n b_n p}\Big)\Big\}\Big(1-\frac{\gamma\lambda}{\sqrt{n}}e^{-\theta_n b_n p}\Big) + O(p+n^{-1}). \end{aligned}$$

Hence

$$\begin{aligned} A_n &= (n-k)\Big(1-\exp\Big\{-mp\Big(1-e^{-\theta_n b_n p}\Big)\Big\}\Big(1-\frac{\gamma\lambda}{\sqrt{n}}e^{-\theta_n b_n p}\Big) + O(p+n^{-1})\Big) - a_n + k \\ &= \Big(1-\theta_n b_n/n - \exp\Big\{-mp\Big(1-e^{-\theta_n b_n p}\Big)\Big\}\Big)n \\ &\qquad + \gamma(\lambda e^{-\theta_n b_n p - mp(1-e^{-\theta_n b_n p})} - 1)\sqrt{n} + o(\sqrt{n}) \\ &= \mu_n + o(\sqrt{n}), \end{aligned} \tag{20}$$

where we have used the fact that

$$\frac{np}{\sqrt{n}} = \sqrt{np^2} = \sqrt{\lambda/m} \to 0.$$

Since $\mathrm{Var}(H_{a_n}) \to \infty$ (by (19)), by using (8) and applying the central limit theorem for the binomial distribution, we have

$$\frac{H_{a_n} - \mathbb{E}(H_{a_n})}{\sqrt{\mathrm{Var}(H_{a_n})}} \xrightarrow{d} N(0,1). \tag{21}$$

It follows from (19) and (21) that $p(H_{a_n} - \mathbb{E}(H_{a_n})) \xrightarrow{p} 0$. Hence

$$\frac{(1-p)^{H_{a_n}}}{(1-p)^{\mathbb{E}(H_{a_n})}} = e^{(H_{a_n}-\mathbb{E}(H_{a_n}))\ln(1-p)} \xrightarrow{p} 1. \tag{22}$$

This together with (19) implies that, for $\alpha \geqslant 1$,

$$B_n^2 \xrightarrow{p} (1-e^{-c_1})e^{-c_1}. \tag{23}$$

By the mean value theorem, we have

$$\begin{aligned} & \sqrt{n}((1-p)^{H_{a_n}} - (1-p)^{\mathbb{E}(H_{a_n})}) \\ = \ & \sqrt{n}\ln(1-p)(1-p)^{\delta_n}(H_{a_n} - \mathbb{E}(H_{a_n})) \\ = \ & \sqrt{np}\sqrt{p\mathrm{Var}(H_{a_n})}(1-p)^{\delta_n}\frac{\ln(1-p)}{p}\frac{H_{a_n}-\mathbb{E}(H_{a_n})}{\sqrt{\mathrm{Var}(H_{a_n})}}, \end{aligned}$$

where δ_n lies between H_{a_n} and $\mathbb{E}(H_{a_n})$.

If $\alpha = 1$, then by (19), (21), (22) and Slutsky's theorem, we have

$$\sqrt{n}((1-p)^{H_{a_n}} - (1-p)^{\mathbb{E}(H_{a_n})}) \xrightarrow{d} N(0, \sqrt{\lambda/\beta}c_2e^{-2c_1}).$$

This implies that

$$\frac{A_n - \widetilde{A}_n}{\sqrt{n-k}B_n} = \frac{\sqrt{n-k}((1-p)^{H_{a_n}} - (1-p)^{\mathbb{E}(H_{a_n})})}{B_n} \xrightarrow{d} N\Big(0, \frac{\sqrt{\lambda/\beta}c_2e^{-c_1}}{1-e^{-c_1}}\Big).$$

Therefore, for any $x \in \mathbb{R}$,

$$\mathbb{E}\Big(\Phi\Big(x + \frac{A_n - \widetilde{A}_n}{\sqrt{n-k}B_n}\Big)\Big) \longrightarrow \mathbb{E}(\Phi(x+Y)) = \mathbb{P}(X \leqslant x+Y) = \mathbb{P}(Z \leqslant x),$$

where X, Y, Z are independent random variables such that

$$X \sim N(0,1), \quad Y \sim N\Big(0, \frac{\sqrt{\lambda/\beta}c_2e^{-c_1}}{1-e^{-c_1}}\Big), \quad Z \sim N\Big(0, \frac{\sqrt{\lambda/\beta}c_2e^{-c_1}}{1-e^{-c_1}} + 1\Big).$$

Summarizing the above facts, it follows from (16) that

$$\frac{S_{a_n} - A_n}{\sqrt{n-k}B_n} \xrightarrow{d} N\Big(0, \frac{\sqrt{\lambda/\beta}c_2e^{-c_1}}{1-e^{-c_1}} + 1\Big).$$

Hence, by applying (20), (23) and Slutsky's theorem, we have

$$\frac{S_{a_n} - \mu_n}{\sqrt{n}} \xrightarrow{d} N\Big(0, \sqrt{\lambda/\beta}c_2e^{-2c_1} + e^{-c_1}(1-e^{-c_1})\Big).$$

Similarly, if $\alpha > 1$, then we have

$$\frac{A_n - \widetilde{A}_n}{\sqrt{n-k}B_n} \xrightarrow{p} 0, \qquad \frac{S_{a_n} - A_n}{\sqrt{n-k}B_n} \xrightarrow{d} N(0,1),$$

and

$$\frac{S_{a_n} - \mu_n}{\sqrt{n}} \xrightarrow{d} N\Big(0, e^{-c_1}(1-e^{-c_1})\Big).$$

If $\alpha < 1$, then by applying (17), (18) and (21), we have $m^{-1}H_{a_n} \xrightarrow{p} 1 - e^{-\theta\lambda}$. This together with the facts that $mp \to 0$ and $H_{a_n} \leqslant m$ implies that

$$\frac{1-(1-p)^{H_{a_n}}}{mp} = \frac{-\ln(1-p)H_{a_n} + O(p^2H_{a_n}^2)}{mp} = \frac{pH_{a_n} + O(p^2H_{a_n}^2 + p^2H_{a_n})}{mp} \xrightarrow{p} 1-e^{-\theta\lambda}.$$

Therefore,

$$\frac{B_n^2}{mp} \xrightarrow{p} 1 - e^{-\theta\lambda}. \tag{24}$$

By the mean value theorem, we have

$$\begin{aligned} & \frac{1}{\sqrt{mp}}((1-p)^{H_{a_n}} - (1-p)^{\mathbb{E}(H_{a_n})}) \\ = \;& \frac{1}{\sqrt{mp}} \ln(1-p)(1-p)^{\delta_n'}(H_{a_n} - \mathbb{E}(H_{a_n})) \\ = \;& \sqrt{\mathrm{Var}(H_{a_n})/m}(1-p)^{\delta_n'}\frac{\ln(1-p)}{p}\frac{H_{a_n} - \mathbb{E}(H_{a_n})}{\sqrt{\mathrm{Var}(H_{a_n})}}, \end{aligned}$$

where δ_n' lies between H_{a_n} and $\mathbb{E}(H_{a_n})$. By (18), (19), (21), (22) and Slutsky's theorem, we have

$$\frac{1}{\sqrt{mp}}((1-p)^{H_{a_n}} - (1-p)^{\mathbb{E}(H_{a_n})}) \xrightarrow{d} N(0, e^{-\theta\lambda}(1-e^{-\theta\lambda})).$$

Hence,

$$\frac{A_n - \widetilde{A}_n}{\sqrt{n}} = \frac{\sqrt{\lambda}(n-k)((1-p)^{H_{a_n}} - (1-p)^{\mathbb{E}(H_{a_n})})}{\sqrt{m}np} \xrightarrow{d} N(0, \lambda e^{-\theta\lambda}(1-e^{-\theta\lambda})). \tag{25}$$

By (24) and the fact that $nmp \to \infty$, we have $(B_n\sqrt{n-k})^{-1} \xrightarrow{p} 0$. Therefore, it follows from (15) that

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\frac{S_{a_n} - \widetilde{A}_n}{\sqrt{n-k}B_n} \leqslant x\Big) - \Phi(x)\Big| \leqslant \mathbb{E}\Big(\frac{C}{B_n\sqrt{n-k}} \wedge 1\Big) \to 0.$$

This means that

$$\frac{S_{a_n} - \widetilde{A}_n}{\sqrt{n-k}B_n} \xrightarrow{d} N(0,1).$$

Therefore, by applying (24) and (25), we have

$$\frac{S_{a_n} - A_n}{\sqrt{n}} = \frac{\sqrt{(n-k)mp}}{\sqrt{n}} \frac{B_n}{\sqrt{mp}} \frac{S_{a_n} - \widetilde{A}_n}{\sqrt{n-k}B_n} - \frac{A_n - \widetilde{A}_n}{\sqrt{n}} \xrightarrow{d} N(0, \lambda e^{-\theta\lambda}(1 - e^{-\theta\lambda})).$$

Now (14) follows from (20) for $\alpha < 1$.

The proof of Lemma 6 is completed. $\square$

Lemma 7. *Assume that $\lambda > 1$ and $\beta > 0$, then we have*

$$\lambda \rho_\beta e^{\sqrt{\lambda/\beta}(\rho_\beta - 1)} < 1, \quad \lambda\rho < 1,$$

where ρ and ρ_β are defined in (3) and (4) respectively.

Proof. Since $\lambda\rho < 1$ is a well-known result for Branching processes, we only need to prove $\lambda \rho_\beta e^{\sqrt{\lambda/\beta}(\rho_\beta - 1)} < 1$. Let N be a Poisson random variable with mean $\sqrt{\lambda\beta}$ and we define

$$H = X'_1 + \cdots X'_N,$$

where $(X'_i)_{i \geqslant 1}$ are i.i.d. Poisson random variables with mean $\sqrt{\lambda/\beta}$ and are independent of N. Then the probability generating function of the distribution H is

$$\begin{aligned} \mathbb{E}(x^H) &= \mathbb{E}(\mathbb{E}(x^{X'_1 + \cdots X'_N} | N)) \\ &= \exp\Big\{\sqrt{\lambda\beta}\Big(e^{\sqrt{\lambda/\beta}(x-1)} - 1\Big)\Big\}. \end{aligned}$$

We define $f(x) = \mathbb{E}(x^H) - x$ for $x \geqslant 0$. By noting that $f(x)$ is strictly convex on $\mathbb{R}$ and $f(1) = f(\rho_\beta) = 0$, we have $f'(\rho_\beta) < 0$. Recalling the definition of ρ_β, we have

$$\begin{aligned} f'(\rho_\beta) &= \lambda \exp\Big\{\sqrt{\lambda\beta}(e^{\sqrt{\lambda/\beta}(\rho_\beta - 1)} - 1)\Big\} e^{\sqrt{\lambda/\beta}(\rho_\beta - 1)} - 1 \\ &= \lambda \rho_\beta e^{\sqrt{\lambda/\beta}(\rho_\beta - 1)} - 1. \end{aligned}$$

Therefore,

$$\lambda \rho_\beta e^{\sqrt{\lambda/\beta}(\rho_\beta - 1)} < 1.$$

The proof of Lemma 7 is completed. $\square$

Lemma 8. *Let $x \in \mathbb{R}$, $\alpha > 0$ and $l_x = [\zeta_{n,m,p} b_n + x\sqrt{n}]$, where $\zeta_{n,m,p}$ is defined in Theorem 1. There exists $\varepsilon_0 \in (0,1)$ such that for any $0 < \varepsilon < \varepsilon_0$ and $l_x > l > (1 - b - \varepsilon) b_n$,*

$$\mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}, S_l = 0) \leqslant \exp\{-\varepsilon^2 \sqrt{n}/4\} + \exp\{-\varepsilon (n \wedge m)^{1/3}/3\}$$

holds for sufficiently large n, where b_n and b are defined in (2).

Proof. Let ε_0 be a positive constant which will be chosen later on. For any $\varepsilon \in (0, \varepsilon_0)$, we denote by

$$h_1 := \varepsilon \mathbb{E}(H_{l_x} - H_l) + (n \wedge m)^{1/3}, \qquad h_2 := l_x - l + \varepsilon\sqrt{n} - (n-l)(1-(1-p)^{\mathbb{E}(H_{l_x}-H_l)+h_1}).$$

Let T be a random variable such that $T \sim \mathrm{B}(n-l, 1-(1-p)^{\mathbb{E}(H_{l_x}-H_l)+h_1})$.

By applying Lemma 5 and using the mean value theorem, there exists $\delta \in [l, l_x]$ such that

$$\begin{aligned}
\mathbb{E}(H_{l_x} - H_l) &= m((1-p)^l - (1-p)^{l_x}) \\
&= m(1-p)^{\delta} \ln((1-p)^{-1})(l_x - l) \\
&\leqslant mp(1+O(p))e^{-lp}(l_x - l),
\end{aligned}$$

where we have used the inequalities $1 - p \leqslant e^{-p}$ and

$$\ln((1-p)^{-1}) = \ln(1 + p(1-p)^{-1}) \leqslant p(1-p)^{-1} = p(1+O(p)).$$

Then by using the fact that $p(l_x - l) \leqslant pl_x \leqslant pb_n + xp\sqrt{n} = o(np(n \wedge m)^{1/3})$ and the inequality $1 - (1-y)^z \leqslant yz$ for $0 < y < 1$ and $z > 1$, we have

$$\begin{aligned}
\mathbb{E}(T) &= (n-l)(1-(1-p)^{\mathbb{E}(H_{l_x}-H_l)+h_1}) \\
&\leqslant (n-l)p((1+\varepsilon)\mathbb{E}(H_{l_x} - H_l) + (n \wedge m)^{1/3}) \\
&\leqslant (1+\varepsilon)(1-(1-b-\varepsilon)b_n/n)\lambda \exp\{-(1-b-\varepsilon)b_n p\}(l_x - l) + O(np(n \wedge m)^{1/3}).
\end{aligned}$$

By Lemma 7, we have $\lambda e^{\lambda(\rho-1)} = \lambda\rho < 1$ and

$$\lambda\rho_\beta \exp\{-\sqrt{\lambda/\beta}(1-\rho_\beta)\} < 1.$$

Hence, by some basic calculations, for any $\alpha > 0$, we can choose $\varepsilon_0 > 0$ so small that

$$\limsup_{n,m\to\infty} (1+\varepsilon_0)(1-(1-b-\varepsilon_0)b_n/n)\lambda \exp\{-(1-b-\varepsilon_0)b_n p\} \leqslant 1 - \varepsilon_0.$$

By using the fact that

$$np(n \wedge m)^{1/3} = O(m^{1/3}np) = O\left(\sqrt{mnp^2}m^{-1/6}\sqrt{n}\right) = o(\sqrt{n}),$$

then for any $0 < \varepsilon < \varepsilon_0$, we have $\mathbb{E}(T) \leqslant (1-\varepsilon)(l_x - l) + o(\sqrt{n})$ for sufficiently large n. Therefore, for sufficiently large n, we have

$$h_2 \geqslant (\varepsilon/2)\sqrt{n}, \qquad \mathbb{E}(T) \leqslant \frac{1-\varepsilon}{\varepsilon} h_2. \tag{26}$$

By applying Chernoff's bound for the binomial distribution (see, for instance, Theorem 2.21 in [13]) and the inequality $h_1 \geqslant (n \wedge m)^{1/3}$, we obtain

$$\begin{aligned}
\mathbb{P}(H_{l_x} - H_l - \mathbb{E}(H_{l_x} - H_l) > h_1) &\leqslant \exp\left\{-\frac{h_1^2}{2\mathbb{E}(H_{l_x} - H_l) + 2h_1/3}\right\} \\
&\leqslant \exp\left\{-\frac{3\varepsilon h_1}{2(3+\varepsilon)}\right\} \\
&\leqslant \exp\{-\varepsilon(n \wedge m)^{1/3}/3\}. \qquad (27)
\end{aligned}$$

By (26), for sufficiently large n, we have

$$\begin{aligned}\mathbb{P}(T > l_x - l + \varepsilon\sqrt{n}) &= \mathbb{P}(T - \mathbb{E}(T) > h_2)\\ &\leqslant \exp\Big\{-\frac{h_2^2}{2\mathbb{E}(T) + 2h_2/3}\Big\}\\ &\leqslant \exp\Big\{-\frac{h_2}{2(1-\varepsilon)/\varepsilon + 2/3}\Big\}\\ &\leqslant \exp\{-\varepsilon^2\sqrt{n}/4\}. \qquad (28)\end{aligned}$$

It follows from (11) that, conditionally on H_l, H_{l_x} and $S_l = 0$,

$$S_{l_x} + (l_x - l) \sim \mathrm{B}(n - l, 1 - (1-p)^{H_{l_x} - H_l}).$$

We can conclude from (27)-(28) that

$$\begin{aligned}\mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}, S_l = 0) &\leqslant \mathbb{P}(S_{l_x} + (l_x - l) > l_x - l + \varepsilon\sqrt{n},\ H_{l_x} - H_l \leqslant \mathbb{E}(H_{l_x} - H_l) + h_1)\\ &\quad + \mathbb{P}(H_{l_x} - H_l - \mathbb{E}(H_{l_x} - H_l) > h_1)\\ &\leqslant \mathbb{P}(T > l_x - l + \varepsilon\sqrt{n}) + \mathbb{P}(H_{l_x} - H_l - \mathbb{E}(H_{l_x} - H_l) > h_1)\\ &\leqslant \exp\{-\varepsilon^2\sqrt{n}/4\} + \exp\{-\varepsilon(n \wedge m)^{1/3}/3\}.\end{aligned}$$

The proof of Lemma 8 is completed. □

Lemma 9. *Let $0 < \eta < 1/2$ be a fixed constant and set $k = k_{n,m,p} = [(m \wedge n)^\eta np]$. Then for any fixed $r \in (0, 1-b)$, we have*

$$\sum_{t=k}^{[rb_n]} \mathbb{P}(S_t = 0) = o(1).$$

Proof. By (8), for $k \leqslant t \leqslant rb_n$ and sufficiently large n, we have

$$\mathbb{E}H_t = m(1 - (1-p)^t) \geqslant m(1 - (1-p)^k) \geqslant \frac{1}{2}mpk. \qquad (29)$$

By applying Chernoff's bound for binomial distribution (see Theorem 2.1 in [7]), we have

$$\begin{aligned}\mathbb{P}(H_t - \mathbb{E}H_t \leqslant -(\mathbb{E}H_t)^{2/3}) &\leqslant \exp\Big\{-\frac{(\mathbb{E}H_t)^{4/3}}{2\mathbb{E}H_t}\Big\}\\ &\leqslant \exp\Big\{-\frac{1}{4}(mpk)^{1/3}\Big\} \qquad \text{for large } n.\end{aligned}$$

Therefore, for large n, we apply (9) to obtain

$$\begin{aligned}\mathbb{P}(S_t = 0) &= \mathbb{P}(\mathrm{B}(n-k, 1-(1-p)^{H_t}) = t - k)\\ &\leqslant \mathbb{P}(\mathrm{B}(n, 1-(1-p)^{H_t}) \leqslant t)\\ &\leqslant \mathbb{P}(\mathrm{B}(n, 1-(1-p)^{H_t}) \leqslant t, H_t - \mathbb{E}H_t \geqslant -(\mathbb{E}H_t)^{2/3})\\ &\qquad + \mathbb{P}(H_t - \mathbb{E}H_t \leqslant -(\mathbb{E}H_t)^{2/3})\\ &\leqslant \exp\Big(-\frac{1}{4}(mpk)^{1/3}\Big) + \mathbb{P}(\mathrm{B}(n, 1-(1-p)^{\mathbb{E}H_t - (\mathbb{E}H_t)^{2/3}}) \leqslant t). \qquad (30)\end{aligned}$$

By Markov's inequality, we have that, for $s > 0$,

$$\begin{aligned}
\mathbb{P}(\mathrm{B}(n, 1-(1-p)^{\mathbb{E}H_t - (\mathbb{E}H_t)^{2/3}}) \leqslant t) &\leqslant e^{st}\mathbb{E}\left(e^{-s\mathrm{B}(n,1-(1-p)^{\mathbb{E}H_t-(\mathbb{E}H_t)^{2/3}})}\right) \\
&\leqslant e^{st}\Big(1-(1-e^{-p(\mathbb{E}H_t-(\mathbb{E}H_t)^{2/3})})(1-e^{-s})\Big)^n \\
&\leqslant \exp\Big\{st - n(1-e^{-p(\mathbb{E}H_t-(\mathbb{E}H_t)^{2/3})})(1-e^{-s})\Big\}.
\end{aligned}$$

Let $d > 0$ be a fixed constant and for $t > 0$, let

$$g(t) = \frac{1-e^{-pd\mathbb{E}H_t}}{t} = \frac{1-e^{-pdm(1-(1-p)^t)}}{t}.$$

Note that, for any fixed $d' > 0$, both $(1-(1-p)^t)/t$ and $(1-e^{-d't})/t$ are strictly decreasing in $(0, \infty)$ for t. Therefore, for $0 < t_1 < t_2$,

$$g(t_2) = \frac{1-e^{-pdmt_2\frac{1-(1-p)^{t_2}}{t_2}}}{t_2} < \frac{1-e^{-pdmt_2\frac{1-(1-p)^{t_1}}{t_1}}}{t_2} < \frac{1-e^{-pdmt_1\frac{1-(1-p)^{t_1}}{t_1}}}{t_1} = g(t_1).$$

Then $g(t)$ is strictly decreasing in $(0, \infty)$. Let $\varepsilon_n = \sup_{k\leqslant t\leqslant [rb_n]}(\mathbb{E}H_t)^{-1/3}$, then $\varepsilon_n \to 0$. Hence, we obtain

$$\begin{aligned}
\inf_{k\leqslant t\leqslant [rb_n]} \frac{n(1-e^{-p(\mathbb{E}H_t-(\mathbb{E}H_t)^{2/3})})}{t} &\geqslant \inf_{k\leqslant t\leqslant [rb_n]} \frac{n(1-e^{-p\mathbb{E}H_t(1-\varepsilon_n)})}{t} \\
&\geqslant \frac{n(1-e^{-p\mathbb{E}H_{[rb_n]}(1-\varepsilon_n)})}{[rb_n]} \\
&\to g(r,\lambda,\beta) := \begin{cases} \frac{1-e^{-\sqrt{\lambda\beta}(1-e^{-r\sqrt{\lambda/\beta}})}}{r}, & \alpha = 1, \\ \frac{1-e^{-r\lambda}}{r}, & \alpha \neq 1. \end{cases}
\end{aligned}$$

By some basic calculations, we obtain

$$\frac{\partial g(r,\lambda,\beta)}{\partial r} \leqslant 0.$$

Then, for $\lambda > 1$ and $r \in (0, 1-b)$, we have $g(r,\lambda,\beta) > g(1-b,\lambda,\beta) = 1$. Therefore, for $\lambda > 1$ and $r \in (0, 1-b)$, we have $g(r,\lambda,\beta) > 1$. Choose $s = \log g(r,\lambda,\beta)$, then

$$\begin{aligned}
\mathbb{P}(\mathrm{B}(n, 1-(1-p)^{\mathbb{E}H_t-(\mathbb{E}H_t)^{2/3}}) \leqslant t) &\leqslant \exp\Big\{st - n(1-e^{-p(\mathbb{E}H_t-(\mathbb{E}H_t)^{2/3})})(1-e^{-s})\Big\} \\
&\leqslant \exp\{-t(g(r,\lambda,\beta) - 1 - \log g(r,\lambda,\beta) + o(1))\}
\end{aligned}$$

holds uniformly for $k \leqslant t \leqslant [rb_n]$. Since $g(r,\lambda,\beta) - 1 - \log g(r,\lambda,\beta) > 0$, we can choose a constant $J(r,\lambda,\beta)$ such that $0 < J(r,\lambda,\beta) < g(r,\lambda,\beta) - 1 - \log g(r,\lambda,\beta)$ and, for sufficiently large n, $\mathbb{P}(\mathrm{B}(n, 1-(1-p)^{\mathbb{E}H_t-(\mathbb{E}H_t)^{2/3}}) \leqslant t) \leqslant e^{-tJ(r,\lambda,\beta)}$ holds uniformly for $k \leqslant t \leqslant [rb_n]$.

This together with (30) implies that, for any $0 < r < 1-b$,

$$\sum_{t=k}^{[rb_n]} \mathbb{P}(S_t = 0) \leqslant \sum_{t=k}^{[rb_n]} \Big(\exp\Big\{-\frac{1}{4}(mpk_n)^{1/3}\Big\} + e^{-tJ(r,\lambda,\beta)}\Big) = o(1).$$

This completes the proof of Lemma 9. □

3 Proof of Theorem 1

Proposition 10. *Let $\eta \in (0, 1/2)$ be a fixed constant. Assume that $m = [\beta n^{\alpha}]$, $nmp^2 = \lambda > 1$ and $k = k_{n,m,p} := [(m \wedge n)^{\eta} np]$. Then for $\alpha > 0$, we have*

$$\frac{|\mathcal{C}_k| - \zeta_{n,m,p} b_n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2),$$

where $\zeta_{n,m,p}$ and σ^2 are defined in Theorem 1.

Proof. At first, we have $k \leqslant m^{\eta} np = m^{\eta - 1/2}\sqrt{\lambda n} = o(\sqrt{n})$. Let $l_x = [\zeta_{n,m,p} b_n + x\sqrt{n}]$ for any $x \in \mathbb{R}$. Then by (7) we have

$$\mathbb{P}\Big(\frac{|\mathcal{C}_k| - \zeta_{n,m,p} n}{\sqrt{n}} > x\Big) = \mathbb{P}(|\mathcal{C}_k| > l_x) = \mathbb{P}(S_i > 0 \text{ for all } i \leqslant l_x) \leqslant \mathbb{P}(S_{l_x} > 0).$$

Note that $\zeta_{n,m,p} \to 1 - b$ (see (6)), where b is defined in (2). By applying Lemma 6 with $\theta_n = \zeta_{n,m,p}$ and $\gamma = x$, we have

$$\mathbb{P}(S_{l_x} > 0) = \mathbb{P}\Big(\frac{S_{l_x} - x\mu(1-b)\sqrt{n}}{\sqrt{n\nu(1-b)}} > \frac{-x\mu(1-b)}{\sqrt{\nu(1-b)}}\Big) \to \mathbb{P}\Big(Z > \frac{-x\mu(1-b)}{\sqrt{\nu(1-b)}}\Big), \tag{31}$$

where Z is a standard normal random variable, $\mu(\theta)$ and $\nu(\theta)$ are defined in (12) and (13). Therefore,

$$\limsup_{n\to\infty} \mathbb{P}\Big(\frac{|\mathcal{C}_k| - \zeta_{n,m,p} n}{\sqrt{n}} > x\Big) \leqslant \mathbb{P}\Big(Z > \frac{-x\mu(1-b)}{\sqrt{\nu(1-b)}}\Big). \tag{32}$$

For the lower bound, we have that for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\Big(\frac{|\mathcal{C}_k| - \zeta_{n,m,p} n}{\sqrt{n}} > x\Big) &= \mathbb{P}(S_i > 0 \text{ for all } i \leqslant l_x) \\
&\geqslant \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n},\ S_i > 0 \text{ for all } i \leqslant l_x) \\
&= \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}) - \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n},\ S_i = 0 \text{ for some } i < l_x) \\
&\geqslant \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}) - \sum_{l=k}^{l_x - 1} \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}, S_l = 0).
\end{aligned}$$

Similar arguments as in the proof of (31) show that

$$\lim_{\varepsilon\downarrow 0} \lim_{n\to\infty} \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}) = \lim_{\varepsilon\downarrow 0} \mathbb{P}\Big(Z > \frac{-x\mu(1-b) + \varepsilon}{\sqrt{\nu(1-b)}}\Big) = \mathbb{P}\Big(Z > \frac{-x\mu(1-b)}{\sqrt{\nu(1-b)}}\Big).$$

By applying Lemmas 8 and 9, we obtain that, for every $0 < \varepsilon < \varepsilon_0$, $r \in (1 - b - \varepsilon, 1 - b)$ and sufficiently large n,

$$\begin{aligned}
\sum_{l=k}^{l_x - 1} \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}, S_l = 0) &= \sum_{l=k}^{[rb_n]} \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}, S_l = 0) + \sum_{l=[rb_n]+1}^{l_x - 1} \mathbb{P}(S_{l_x} > \varepsilon\sqrt{n}, S_l = 0) \\
&\leqslant o(1) + b_n\Big(\exp\{-\varepsilon^2\sqrt{n}/4\} + \exp\{-\varepsilon(n\wedge m)^{1/3}/3\}\Big) = o(1),
\end{aligned}$$

We conclude that

$$\liminf_{n\to\infty} \mathbb{P}\Big(\frac{|\mathcal{C}_k| - \zeta_{n,m,p} n}{\sqrt{n}} > x\Big) \geqslant \mathbb{P}\Big(Z > \frac{-x\mu(1-b)}{\sqrt{\nu(1-b)}}\Big). \tag{33}$$

Noting that $e^{-c(1-\rho_\beta)} = \rho_\beta,\ e^{-(1-\rho)\lambda} = \rho$,

$$\mu(1-b) = \begin{cases} \lambda c\rho_\beta - 1, & \alpha = 1, \\ \lambda\rho - 1, & \alpha \neq 1, \end{cases}$$

and

$$\nu(1-b) := \begin{cases} \lambda c(1-c)\rho_\beta^2 + \rho_\beta(1-\rho_\beta), & \alpha = 1, \\ \rho(1-\rho), & \alpha > 1, \\ \lambda\rho(1-\rho), & \alpha < 1, \end{cases}$$

where $c = e^{(\rho_\beta - 1)\sqrt{\lambda/\beta}}$, we have

$$\sigma^2 = \frac{\nu(1-b)}{\mu^2(1-b)}.$$

The proof of Proposition (10) is completed by (32) and (33). $\square$

Proof of Theorem 1. For $m = [\beta n^\alpha]$, the second largest component of $G(n,m,p)$ has size less than $O_P(a_{n,m,p})$ (see Theorem 1 in [11] and Theorem 12 in [8]), where $a_{n,m,p} = \frac{\sqrt{mn}}{m\wedge n}\log^2 n$. Let $k = k_{n,m,p} = [(m\wedge n)^\eta np]$, then

$$ka_{n,m,p} \leqslant \sqrt{mn}(m\wedge n)^{\eta-1} np\log^2 n = O_P(n(m\wedge n)^{\eta-1}\log^2 n).$$

For any $\alpha > 1/2$, there exists $0 < \eta_\alpha < 1/2$ such that $ka_{n,m,p} = o_P(\sqrt{n})$. We can conclude that, for any $\alpha > 1/2$, with high probability,

$$\mathcal{N}(G(n,m,p)) \leqslant |\mathcal{C}_k|.$$

Otherwise, with high probability,

$$|\mathcal{C}_k| = O_P(ka_{n,m,p}) = o_P(\sqrt{n}).$$

This is a contradiction to Proposition 10. So, we get that with high probability,

$$\mathcal{N}(G(n,m,p)) \leqslant |\mathcal{C}_k| \leqslant \mathcal{N}(G(n,m,p)) + O_P(ka_{n,m,p}) = \mathcal{N}(G(n,m,p)) + o_P(\sqrt{n}).$$

Then Theorem 1 follows from Proposition 10. $\square$

Acknowledgements

We thank the anonymous referee for the useful suggestions that greatly improved the presentation of this work. This work is supported by NSFC (grant number 11671373).

References

[1] F. G. Ball, D. J. Sirl, and P. Trapman. Epidemics on random intersection graphs. *Ann. Appl. Probab.*, 24(3):1081-1128, 2014.

[2] A. D. Barbour and G. Reinert. The shortest distance in random multi-type intersection graphs. *Random Structures and Algorithms*, 39(2):179-209, 2011.

[3] M. Behrisch. Component evolution in random intersection graphs. *Electron. J. Combin.*, 14(1): #R17, 2007.

[4] M. Bloznelis. Degree and clustering coefficient in sparse random intersection graphs. *Ann. Appl. Probab.*, 23(3):1254-1289, 2013.

[5] J. A. Fill, E. R. Scheinerman, and K. B. Singer-Cohen. Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models. *Random Structures and Algorithms*, 16(2):156-176, 2000.

[6] E. Godehardt and J. Jaworski. Two models of random intersection graphs for classification. *Studies in Classification, Data Analysis and Knowledge Organization*, pages 67-81. Springer, 2003.

[7] S. Janson, T. Łuczak, and A. Ruciński. Random Graphs. John Wiley & Sons, New York, 2000.

[8] T. Johansson. The giant component of the random bipartite graph. Master thesis, Department of Mathematical Sciences, Chalmers University of Technology, 2012.

[9] M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen. On random intersection graphs: the subgraph problem. *Combinatorics, Probability and Computing*, 8(1-2):131-159, 1999.

[10] J. Kim, S. J. Lee, and J. Na. On the total variation distance between the binomial random graph and the random intersection graph. *Random Structures and Algorithms*, 52(4):662-679, 2018.

[11] A. N. Lagerås and M. Lindholm. A note on the component structure in random intersection graphs with tunable clustering. *Electron. J. Combin.*, 15(1): #N10, 2008.

[12] R. D. Pietro, L. V. Mancini, A. Mei, A. Panconesi, and J. Radhakrishnan. How to design connected sensor networks that are provably secure, in: Proceedings of the 2nd IEEE International Conference on Security and Privacy for Emerging Areas in Communication Networks, 2006.

[13] R. van der Hofstad. Random Graphs and Complex Networks. The Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2016.

[14] K. B. Singer-Cohen. Random intersection graphs. PhD thesis, Department of Mathematical Sciences, The John Hopkins University, 1995.

[15] D. Stark. The vertex degree distribution of random intersection graphs. *Random Structures and Algorithms.* 24(3):249-258, 2004.