Loglog distances in a power law random intersection graph

Mindaugas Bloznelis

Faculty of Mathematics and Informatics, Vilnius University, LT-03225 Vilnius, Lithuania
and Institute of Mathematics and Informatics, Akademijos 4, LT-08663 Vilnius, Lithuania
E-Mail: mindaugas.bloznelis@mif.vu.lt

Abstract

We consider the typical distance between vertices of the giant component of a random intersection graph having a power law (asymptotic) vertex degree distribution with infinite second moment. Given two vertices from the giant component we construct $O_P(\log \log n)$ upper bound for the length of the shortest path connecting them.

key words: intersection graph, random graph, power law, giant component.
AMS 2000 Subject Classification: Primary 05C80, Secondary 05C12

1 Introduction

Given a collection of subsets $S(1), \dots, S(n)$ of the set $W = \{w_1, \dots, w_m\}$ define the intersection graph on the vertex set $V = \{v_1, \dots, v_n\}$ such that v_i and v_j are joined by an edge (denoted $v_i \sim v_j$) whenever $S(i) \cap S(j) \neq \emptyset$, for $i \neq j$. Assuming that the sets $S(i)$, $i = 1, \dots, n$, are drawn at random we obtain a random intersection graph.
Random intersection graphs have applications in various fields: design and analysis of secure wireless sensor networks [7], [5], modelling of social networks [6], statistical clasification [8], see also [12], [13]. Usually, in applications the number of interacting nodes (vertices) is large and it is convenient to study the statistical properties of parameters of interest.
We consider a class of random intersection graphs, where m is much larger than n and where the random subsets $S(i)$, $i = 1, \dots, n$, are independent. Moreover, we assume that for every i, the distribution of $S(i)$ is a mixture of uniform distributions. That is, for every k, conditionally on the event $|S(i)| = k$ the random set $S(i)$ is uniformly distributed in the class of all subsets of W of size k. In particular, with P_{*i} denoting the distribution of $|S(i)|$ we have, for every $A \subset W$, $\mathbf{P}(S(i) = A) = \binom{m}{|A|}^{-1} P_{*i}(|A|)$. The random intersection graph corresponding to the sequence of distributions $\mathbf{P}_* = (P_{*1}, \dots, P_{*n})$ is denoted $G(n, m, \mathbf{P}_*)$.
Assuming that as $n, m \to \infty$ the asymptotic distributions of $\sqrt{n/m}\,|S(i)|$ have power tails and infinite second moment we obtain the random intersection graph $G(n, m, \mathbf{P}_*)$ with asymptotically heavy tailed vertex degree distribution without second moment, see [6] and [1]. In the present paper we show that the typical distance between vertices of the giant component of

such graph is of order $O_P(\log\log n)$. Similar results, but for different power law random graph models were obtained in [14], [4], [15], see also [10].
The paper is organized as follows: results are stated in Section 2. Proofs are given in Section 3.

2 Results

Given an integer sequence $\{m_1, m_2, \dots\}$, let $\{(Z_{n1}, \dots, Z_{nn}), n = 1, 2, \dots\}$ be a sequence of random vectors with independent coordinates such that for every n, Z_{ni} takes values in $\{0, 1, \dots, m_n\}$, $1 \le i \le n$. Let P_{ni} denote the distribution of Z_{ni}. Write $\mathbf{P}_n = (P_{n1}, \dots, P_{nn})$. Fix two countable sets $\{v_1, v_2, \dots\}$ and $\{w_1, w_2, \dots\}$ and define the sequence of random intersection graphs $\{G_n = G(n, m_n, \mathbf{P}_n), n = 1, 2, \dots\}$ as follows. Given n, let $S_n(v_1), \dots, S_n(v_n)$ be independent subsets of $W_n = \{w_1, \dots, w_{m_n}\}$ of sizes $Z_{ni} = Z_n(v_i) := |S_n(v_i)|$, $1 \le i \le n$, such that $\mathbf{P}(S_n(v_i) = A) = \binom{m_n}{|A|}^{-1} P_{ni}(|A|)$, for $A \subset W_n$. G_n is the graph on the vertex set $V_n = \{v_1, \dots, v_n\}$, where v_i and v_j are adjacent whenever $S_n(v_i) \cap S_n(v_j) \ne \emptyset$. Let $\tilde{P}_{ni}$ denote the distribution of the random variable $\tilde{Z}_{ni} = \tilde{Z}_n(v_i) := |S_n(v_i)| \sqrt{n/m_n}$.
Let $d_n(u, v)$ denote the distance between vertices $u, v \in V_n$ in G_n (=number of edges in the shortest path of G_n connecting u and v). Let $C_1 = C_1(G_n) \subset V_n$ denote the vertex set of the largest connected component of G_n. Therefore, the subgraph of G_n induced by C_1 is connected and the number of vertices of any other connected subgraph of G_n is not greater than $|C_1|$. A vertex $u \in V_n$ is called maximal in G_n if $Z_n(u) = \max_{v \in V_n} Z_n(v)$.

Theorem 1. *Let $0 < \alpha < 1$ and $c_0, c_1, c_2 > 0$. Let $\{\omega_1, \omega_2, \dots\}$ be a sequence of positive numbers satisfying $\lim_n \omega_n = +\infty$. Let $\{G(n, m_n, \mathbf{P}_n), n = 1, 2, \dots\}$ be a sequence of random intersection graphs such that*
(i) $n \ln^2 n = o(m_n)$ as $n \to \infty$;
(ii) $\exists\, n_0$ such that $\forall\, n > n_0$ we have

$$c_1 t^{-1-\alpha} \le \mathbf{P}(\tilde{Z}_{ni} > t) \le c_2 t^{-1-\alpha}, \qquad \forall t \in [c_0, n^{1/(1+\alpha)}\omega_n], \quad \forall i \in \{1, \dots, n\}. \tag{1}$$

Let $\{u_n\}$ be a sequence of maximal vertices, i.e., for every n, the vertex $u_n \in V_n$ is maximal in G_n. For every $\varepsilon > 0$ we have as $n \to \infty$

$$\mathbf{P}\Big(d(v_1, u_n) \le (1+\varepsilon) \ln^{-1}(1/\alpha)\, \ln(\ln(2+n)) \,\Big|\, d(v_1, u_n) < \infty\Big) \to 1, \tag{2}$$

$$\mathbf{P}\Big(d(v_1, v_2) \le (2+\varepsilon) \ln^{-1}(1/\alpha) \ln(\ln(2+n)) \,\Big|\, v_1, v_2 \in C_1\Big) \to 1. \tag{3}$$

Here $'\ln'$ denotes the natural logarithm.

It follows from (3), by the symmetry, that given two vertices v, v' drawn uniformly at random from the giant component C_1 we have $d(v, v') = O_P(\ln\ln n)$. Recall that such a distance is of much larger order $O_P(\ln n)$ in the corresponding Erdős-Rényi graph ($G(n, p)$ with $1 < c_1 \le np \le c_2$). This remarkable difference is explained by an effect of very large nodes whose degrees realize the extremes from a power law distribution, see [14], [15].
Note that with probability tending to 1 (with high probability) every maximal vertex belongs to the giant component C_1. In addition, as $n \to \infty$ we have $|C_1| > \rho n$, for some $\rho \in (0, 1)$. We collect these statements in Remark 1.
Remark 1. Assume that conditions of Theorem 1 are satisfied. Then

$$\exists\, \rho \in (0, 1) \qquad \text{such that} \qquad \mathbf{P}(|C_1| > \rho\, n) \to 1 \qquad \text{as} \qquad n \to \infty. \tag{4}$$

Let $\{u_n\}$ be a sequence of maximal vertices, i.e., for every n, the (random) vertex $u_n \in V_n$ is maximal in G_n. Then

$$\mathbf{P}\big(u_n \in C_1(G_n)\big) \to 1 \qquad as \qquad n \to \infty. \tag{5}$$

Acknowledgement. I thank Ilkka Norros for valuable discussion.

3 Proofs

We start with auxiliary Lemmas 1-4. Then we prove Remark 1, see Lemma 5 below, and Theorem 1.

In what follows we write $l_2(n) := \ln(\ln(n))$, where ln denotes the natural logarithm. $H_{j,k,m}$ denotes the hypergeometric random variable with parameters $j, k \le m$ and the distribution $\mathbf{P}(H_{j,k,m} = r) = \frac{\binom{k}{r}\binom{m-k}{j-r}}{\binom{m}{j}}$.

Lemma 1. *Let S_1, S_2 be independent random subsets of the set $W = \{1, \dots, m\}$ such that S_1 (respectively S_2) is uniformly distributed in the class of subsets of W of size j (respectively k). Then $H = |S_1 \cap S_2|$ is the hypergeometric random variable with parameters j, k, m and mean $\mathbf{E}H = jk/m$. The probability $p' := \mathbf{P}(H = 0) = (m-k)_j/(m)_j$ satisfies, for $j + k < m$,*

$$1 - \frac{jk/m}{1 - (j+k)/m} \le p' \le 1 - \frac{jk}{m} + (\frac{jk}{m})^2. \tag{6}$$

Here we denote $(m)_j = m(m-1)\cdots(m-j+1)$. For $0 < s < 1$ and $j + k \le s\,m$ we have

$$\frac{jk}{m} + \frac{2}{1-s}(\frac{jk}{m})^2 \ge \mathbf{P}(S_1 \cap S_2 \ne \emptyset) \ge \frac{jk}{m} - (\frac{jk}{m})^2. \tag{7}$$

For $\lambda = \mathbf{E}H$ and $t \ge 0$ we have

$$\mathbf{P}(H \ge \lambda + t) \le \exp\{-\frac{t^2}{2(\lambda + t/3)}\}, \qquad \mathbf{P}(H \le \lambda - t) \le \exp\{-\frac{t^2}{2\lambda}\}. \tag{8}$$

In particular, we have

$$\mathbf{P}(H = 0) \le e^{-jk/2m}. \tag{9}$$

Proof of Lemma 1. Inequalities (6) are shown in [12]. Inequalities (7) are simple consequences of (6). We only show the left-hand side inequality for $i, j \ge 1$. In this case $j + k \le 2jk$ and we have

$$a := \frac{1}{1 - (j+k)/m} = 1 + \frac{j+k}{m}a \le 1 + \frac{2jk}{m}\frac{1}{1-s}.$$

Now, desired inequality follows from the left-hand side inequality (6).

Exponential inequalities for hypergeometric probabilities (8) can be derived from the corresponding inequalities for binomial probabilities, see [9]. Their proof can be found in, e.g., [11]. The right-hand side inequality (8) applied to $t = \mathbf{E}H$ gives (9). □

Lemma 2. *Given integer m and constants $0<\gamma_1<\gamma_2<1$ let $z_1,z_2,\dots,z_r$ be integers such that $z=\sum_{h=1}^r z_h \le \ \gamma_1 m$ and $z_h \ge 6\gamma_2(\gamma_2-\gamma_1)^{-2}\ln n \ge 1$, for $1\le h\le r$. Let $S_1,S_2,\dots,S_r$ be independent random subsets of $W=\{1,\dots,m\}$ such that, for every h, S_h is uniformly distributed in the class of subsets of W of size z_h. Then*

$$\mathbf{P}\Big(\big|\cup_{i=1}^r S_i\big| \ge (1-\gamma_2)\sum_{i=1}^r |S_i|\Big) \ge 1-rn^{-3}. \tag{10}$$

Proof of Lemma 2. Write $D_{[0]}=\emptyset$ and, for $h\ge 1$, denote $D_{[h]}=\cup_{k\le h}S_k$ and $S_h'=S_h\setminus D_{[h-1]}$. Note that $|D_{[r]}|=\sum_{h=1}^r|S_h'|\le z$. In order to prove (10) we show that uniformly in h and $D_{[h-1]}$ (satisfying $|D_{[h-1]}|\le\gamma_1 m$) we have $p_h:=\mathbf{P}(|S_h'|\le(1-\gamma_2)z_h\,\big|\,D_{[h-1]})\le n^{-3}$. It is convenient to write this probability in the form $p_h=\mathbf{P}(H\ge\gamma_2 a)$, where H denotes the hypergeometric random variable with parameters $a=z_h$, $b=|D_{[h-1]}|$ and m. We have $\mathbf{E}H=ab/m\le\gamma_1 a$. An application of (8) shows $p_h\le\exp\{-a(\gamma_2-\gamma_1)^2/(2\gamma_2)\}$. For $a=z_h\ge(6\gamma_2/(\gamma_2-\gamma_1)^2)\ln n$ we obtain $p_h\le n^{-3}$, thus completing the proof. □

Lemma 3. *Given integers $1\le a,b,d\le m$, let $\mathcal{S}_a\subset\mathcal{S}_d$ be subsets of the set $W=\{1,2,\dots,m\}$ of sizes $|\mathcal{S}_a|=a$ and $|\mathcal{S}_d|=d$. Here $a\le d$. Let S_b be a random subset of W uniformly distributed over the subsets of W of size b. For integers $0<s\le r<t$ satisfying $s\le a\wedge b$, we have*

$$\mathbf{P}\Big(|S_b\cap\mathcal{S}_d|\ge t\,\Big|\,|S_b\cap\mathcal{S}_a|\ge s\Big)\le\max_{s\le i\le r}\mathbf{P}(H_{b-i,d-a,m-a}\ge t-i)+\frac{\mathbf{P}(H_{a,b,m}>r)}{\mathbf{P}(H_{a,b,m}\ge s)}. \tag{11}$$

Assume that $d\le m/100$. Then we have

$$\mathbf{P}\Big(|S_b\cap\mathcal{S}_d|\ge b/2\,\Big|\,S_b\cap\mathcal{S}_a\ne\emptyset\Big)\le e^{-b/8}\big(1+4\frac{m}{ab}\mathbb{I}\{a>b/4,ab\le m,b\ge 3\}\big). \tag{12}$$

Proof of Lemma 3. Let us prove (11). Introduce events $\mathbb{B}=\{|S_b\cap\mathcal{S}_d|\ge t\}$, $\mathbb{A}=\{|S_b\cap\mathcal{S}_a|\ge s\}$ and write $p:=\mathbf{P}(\mathbb{B}|\mathbb{A})$. Denote $p_i=\mathbf{P}(H_{b-i,d-a,m-a}\ge t-i)$. Let $\sum_j$ denote the sum over subsets $\mathcal{A}_j\subset\mathcal{S}_a$ of size $|\mathcal{A}_j|=j$. We have

$$\begin{aligned}\mathbf{P}(\mathbb{B}\cap\mathbb{A}) &= \sum_{s\le j\le a\wedge b}\sum_j\mathbf{P}\big(\mathbb{B}\cap\{S_b\cap\mathcal{S}_a=\mathcal{A}_j\}\big)\\ &= \sum_{s\le j\le a\wedge b}\sum_j\mathbf{P}\big(\mathbb{B}\big|S_b\cap\mathcal{S}_a=\mathcal{A}_j\big)\mathbf{P}(S_b\cap\mathcal{S}_a=\mathcal{A}_j)\\ &= \sum_{s\le j\le a\wedge b}p_j\sum_j\mathbf{P}(S_b\cap\mathcal{S}_a=\mathcal{A}_j)=\sum_{s\le j\le a\wedge b}p_j\mathbf{P}(H_{a,b,m}=j)\\ &\le \max_{s\le i\le r}p_i\,\mathbf{P}(s\le H_{a,b,m}\le r)+\mathbf{P}(H_{a,b,m}>r).\end{aligned} \tag{13}$$

(11) follows from (13) and the identity $p=\mathbf{P}(\mathbb{B}\cap\mathbb{A})/\mathbf{P}(\mathbb{A})$.
Let us prove (12). Put $t=\lceil b/2\rceil$, $s=1$ and $r=\lfloor b/4\rfloor$ and apply (11). We obtain

$$\mathbf{P}\Big(|S_b\cap\mathcal{S}_d|\ge b/2\,\Big|\,S_b\cap\mathcal{S}_a\ne\emptyset\Big)\le\max_{1\le i\le r}p_i+p_1^*/p_2^*. \tag{14}$$

Here we denote $p_1^*:=\mathbf{P}(H_{a,b,m}>r)$, $p_2^*=\mathbf{P}(H_{a,b,m}\ge 1)$. Let us show that

$$p_i\le e^{-b/8},\qquad 1\le i\le r. \tag{15}$$

For this purpose we apply the first inequality of (8). Denote $\lambda_i = \mathbf{E}H_{b-i,d-a,m-a}$ and $t_i = t - i - \lambda_i$. We have, for $1 \le i \le r$ and $d/m \le 100$,

$$\lambda_i = (b-i)(d-a)/(m-a) \le bd/m \le b/100,$$
$$t_i \ge \lceil b/2 \rceil - \lfloor b/4 \rfloor - (b/100) \ge (b/4) - (b/100),$$
$$t_i \le \lceil b/2 \rceil - i \le b/2.$$

These inequalities combined with the inequality, which follows from (8), $p_i \le e^{-t_i^2/(2(\lambda_i + t_i/3))}$ imply (15). Note that, for $a \le b/4$, we have $p_1^* = 0$ and, therefore, (12) follows from (15) and (14).
Now assume that $a > b/4$. Denote $\lambda_* = \mathbf{E}H_{a,b,m}$ and $t_* = r + 1 - \lambda_*$. We have

$$\lambda_* = ab/m \le b/100, \qquad 1 + b/4 > t_* > b/4 - b/100.$$

Note that $b \ge 3$ implies $t_* < (7/12)b$. These inequalities combined with the inequality, which follows from (8), $p_1^* \le e^{-t_*^2/(2(\lambda_* + t_*/3))}$ imply

$$p_1^* \le e^{-b/8}. \tag{16}$$

Finally, we apply (9) to get the lower bound

$$p_2^* \ge 1 - e^{-ab/2m} \ge ab/(4m), \tag{17}$$

for $ab < m$. Invoking (15, 16, 17) in (14) we obtain (12).

□

Lemma 4. *Let $0 < \alpha < 1$ and $c_0, c_1, c_2 > 0$. Let $\{\omega_n\}$ be a positive sequence satisfying $\omega_n \to +\infty$ as $n \to \infty$. Let $\{(\tilde{Z}_{n1}, \dots, \tilde{Z}_{nn})\}$ be a sequence of random vectors with independent non-negative coordinates satisfying condition (ii) of Theorem 1. We have as $n \to \infty$*

$$\mathbf{P}\big(n^{1/(1+\alpha)}/\omega_n < \max_{1 \le i \le n} \tilde{Z}_{ni} \le n^{1/(1+\alpha)}\, \omega_n\big) \to 1. \tag{18}$$

Let $L_n(t) = \sum_{i=1}^n \tilde{Z}_{ni} \mathbb{I}_{\{t < \tilde{Z}_{ni} \le n^{1/(1+\alpha)}\, \omega_n\}}$. There exists an integer $n_1 \ge n_0$ depending on α, c_1, c_2 and the sequence $\{\omega_n\}$ such that, for $n > n_1$ and $t \in (c_0, n^{1/(1+\alpha)})$, we have

$$c_1/2 \le \frac{\alpha}{1+\alpha} \frac{t^\alpha}{n} \mathbf{E}L_n(t) \le c_2, \tag{19}$$

For $1 < \tau < 1 + \alpha$ there exists an integer $n_2 \ge n_0$ and number $c^* > 0$ both depending on α, τ, c_1, c_2 and the sequence $\{\omega_n\}$ such that, for $n > n_2$ and $t \in (c_0, n^{1/(1+\alpha)})$, we have

$$\mathbf{P}\big(\big|L_n(t) - \mathbf{E}L_n(t)\big| > \gamma \mathbf{E}L_n(t)\big) \le c^* \gamma^{-\tau} n^{1-\tau} t^{(\tau-1)(\alpha+1)}. \tag{20}$$

Proof of Lemma 4. The proof is routine. We include it for the sake of completeness. Denote for short $t_* = n^{1/(1+\alpha)}/\omega_n$ and $T_* = n^{1/(1+\alpha)}\, \omega_n$.
Let us prove (18). Write $p_n^*(t) := \mathbf{P}(\max_{1 \le i \le n} \tilde{Z}_{ni} \le t)$. It follows from (1) as $n \to \infty$

$$p_n^*(t_*) = \prod_i \mathbf{P}(\tilde{Z}_{ni} \le t_*) \le \big(1 - c_1/t_*^{1+\alpha}\big)^n \le \exp\{-c_1 n/t_*^{1+\alpha}\} = o(1), \tag{21}$$

$$p_n^*(T_*) = \prod_i \mathbf{P}(\tilde{Z}_{ni} \le T_*) \ge \big(1 - c_2/T_*^{1+\alpha}\big)^n \ge \exp\{-c_2 n/(T_*^{1+\alpha} - c_2)\} = 1 - o(1). \tag{22}$$

In (21) we apply $1-x\le e^{-x}$ to $x=c_1/t_*^{1+\alpha}$. In (22) we apply $1-y\ge e^{-y/(1-y)}$ to $y=c_2/T_*^{1+\alpha}<1$.
Let us prove (19-20). Given $1\le\tau<1+\alpha$ and n, write $a_i^{(\tau)}(t)=\mathbf{E}\tilde Z_{ni}^{\tau}\mathbb{I}_{\{t<\tilde Z_{ni}\le T_*\}}$. It follows from (1) and the identity

$$a_i^{(\tau)}(t)=t^{\tau}\mathbf{P}(t<\tilde Z_{ni}\le T_*)+\tau\int_t^{T_*}x^{\tau-1}\mathbf{P}(x<\tilde Z_{ni}\le T_*)dx$$

that, for sufficiently large n and $t\in(c_0,n^{1/(1+\alpha)})$,

$$c_1/2\ \le\ a_i^{(\tau)}(t)\ t^{1+\alpha-\tau}\ \frac{1+\alpha-\tau}{1+\alpha}\ \le\ c_2. \tag{23}$$

Note that the right hand side inequality holds for $n>n_0$, while the left hand side inequality holds for $n>n_0'$, where $n_0'=n_0'(\alpha,\tau,c_1,c_2,\{\omega_n\})\ge n_0$.
An application of (23) to $\mathbf{E}L_n(t)=\sum_{i=1}^n a_i^{(1)}(t)$ shows (19).
Let us show (20). Denote $T_i=\tilde Z_{ni}\mathbb{I}_{\{t<\tilde Z_{ni}\le T^*\}}-\mathbf{E}\tilde Z_{ni}\mathbb{I}_{\{t<\tilde Z_{ni}\le T^*\}}$. Write, for short, $b:=\gamma\mathbf{E}L_n(t)$. By Chebyshev's inequality,

$$p_n(t):=\mathbf{P}\big(\big|\sum_{i=1}^n T_i\big|\ge b\big)\le b^{-\tau}\mathbf{E}\big|\sum_{i=1}^n T_i\big|^{\tau}. \tag{24}$$

Invoking the inequalities $\mathbf{E}|\sum T_i|^\tau\le\sum\mathbf{E}|T_i|^\tau$ and $\mathbf{E}|T_i|^\tau\le 2a_i^{(\tau)}(t)$, $1\le\tau\le 2$, we obtain

$$p_n(t)\le 2b^{-\tau}\sum_{i=1}^n a_i^{(\tau)}(t). \tag{25}$$

It follows from (23) that $\sum_{i=1}^n a_i^{(\tau)}(t)\le c_2\frac{1+\alpha}{1+\alpha-\tau}\frac{n}{t^{1+\alpha-\tau}}$. Substitution of this inequality and of (19) in (25) gives

$$p_n(t)\le\frac{8}{1+\alpha-\tau}\frac{c_2}{c_1^\tau}\frac{1}{\gamma^\tau}\frac{t^{(\tau-1)(\alpha+1)}}{n^{\tau-1}}$$

thus proving (20). □

Lemma 5. *Assume that conditions of Theorem 1 are satisfied. Then (4) holds. Let $V_n^0=\{v_i:\tilde Z_{ni}>n^{1/(1+\alpha)}/l_2^\alpha(n)\}\subset V_n$. We have as $n\to\infty$*

$$\mathbf{P}\big(V_n^0\subset C_1(G_n)\big)\to 1, \tag{26}$$
$$\mathbf{P}\big(|V_n^0|\ge 2c_2(l_2(n))^{\alpha(1+\alpha)}\big)\to 0. \tag{27}$$

Here $|V_n^0|$ denotes the number of elements of the set V_n^0 and $l_2(n)$ denotes $\ln(\ln(n))$.

Observe that (18) implies that every maximal vertex of G_n belongs whp to V_n^0. Therefore, (26) combined with (18) imply (5).

Proof of Lemma 5. Let us prove (27). Write $t_{0n}=n^{1/(1+\alpha)}/l_2^\alpha(n)$. We have

$$|V_n^0|=\sum_{i=1}^n\mathbb{I}_i^0,\qquad \mathbb{I}_i^0:=\mathbb{I}_{\{\tilde Z_{ni}>t_{0n}\}},\qquad 1\le i\le n. \tag{28}$$

For $i = 1, \ldots, n$, let $\mathbb{I}_i^+$ be independent Bernoulli random variables with success probability $\mathbf{P}(\mathbb{I}_i^+ = 1) = c_2 t_{0n}^{-1-\alpha}$. It follows from (1), (28) that the random variable $L^+ := \sum_{1 \le i \le n} \mathbb{I}_i^+$ is stochastically larger than $|V_n^0|$. Therefore, for every $a > 0$ we have

$$\mathbf{P}(|V_n^0| \ge a) \le \mathbf{P}(L^+ \ge a).$$

Recall that exponential inequalities (8) remain valid if we replace the hypergeometric random variable H by a Binomial random variable, see e.g., [11]. The first inequality of (8) applied to Binomial probability $\mathbf{P}(L_+ \ge a)$ with $a = 2\mathbf{E}L^+ = 2c_2(l_2(n))^{\alpha(1+\alpha)}$ shows (27).

Let us prove (4). Let G_n^0 be the subgraph of G_n obtained by deleting the edges incident to vertices from V_n^0. Note that G_n^0 is a random intersection graph defined by the random sets $S_n^0(v_i)$, $v_i \in V_n$, such that $S_n^0(v_i) = S_n(v_i)$ for $\tilde{Z}_{ni} \le t_{0n}$ and $S_n^0(v_i) = \emptyset$, for $\tilde{Z}_{ni} > t_{0n}$. Denote $\tilde{Z}_{ni}^0 := \tilde{Z}_{ni} \mathbb{I}_{\{\tilde{Z}_{ni} \le t_{0n}\}} = |S_n^0(v_i)| \sqrt{n/m}$. Write

$$\varkappa^{1+\alpha} = 2c_2/c_1, \qquad a_0 := c_0, \qquad a_{i+1} = a_i \varkappa, \qquad i = 0, 1, \ldots,$$

and note that $\varkappa^{1+\alpha} \ge 2$. Let $\tilde{Y}_n$ be a discrete random variable with values $0, a_0, a_1, \ldots, a_{j_n}$, where $j_n + 1 = \max\{i : a_i \le t_{0n}\}$, and with probabilities $\mathbf{P}(\tilde{Y}_n = a_j) = \tilde{p}_j$, defined by

$$\tilde{p}_j := c_1 a_j^{-1-\alpha} - c_2 a_{j+1}^{-1-\alpha} = c_1 a_j^{-1-\alpha}/2, \qquad j = 0, 1, \ldots, j_n.$$

Put $\mathbf{P}(\tilde{Y}_n = 0) = 1 - \tilde{p}_0 - \tilde{p}_1 - \cdots - \tilde{p}_{j_n}$. Note that $\tilde{Y}_n$ is stochastically smaller than $\tilde{Z}_{ni}^0$, for every $1 \le i \le n$. Indeed, (1) implies, for $j = 0, 1, \ldots, j_n$,

$$\begin{aligned} \mathbf{P}(a_j < \tilde{Z}_{ni}^0 \le a_{j+1}) &= \mathbf{P}(\tilde{Z}_{ni} > a_j) - \mathbf{P}(\tilde{Z}_{ni} > a_{j+1}) \\ &\ge c_1 a_j^{-1-\alpha} - c_2 a_{j+1}^{-1-\alpha} \\ &= \tilde{p}_j = \mathbf{P}(\tilde{Y}_n = a_j). \end{aligned}$$

Let $\tilde{Y}_{n1}, \ldots, \tilde{Y}_{nn}$ be independent copies of $\tilde{Y}_n$ defined on the same probability space as $\tilde{Z}_{ni}^0, 1 \le i \le n$ and such that almost surely $\tilde{Y}_{ni} \le \tilde{Z}_{ni}^0$, for every $1 \le i \le n$ (such coupling is possible because $\tilde{Y}_{ni}$ is stochastically smaller than $\tilde{Z}_{ni}^0$). For $1 \le i \le n$, let $\tilde{S}_n^0(v_i)$ be a random subset of $S_n^0(v_i)$ of size $|\tilde{S}_n^0(v_i)| = \lfloor \tilde{Y}_{ni} \sqrt{m/n} \rfloor$ (which is uniformly distributed over the class of subsets of $S_n^0(v_i)$ of size $\lfloor \tilde{Y}_{ni} \sqrt{m/n} \rfloor$). Random subsets $\tilde{S}_n^0(v_i)$, $v_i \in V$ are independent and identically distributed. They define random intersection graph (denoted) $\tilde{G}_n^0$ which is a subgraph of G_n^0. It is easy to see that $\mathbf{E}\tilde{Y}_n^2 \to \infty$. Therefore, using Theorem 1 and Remark 2 of [2], one can show that there exists $\rho \in (0, 1)$ such that the number of vertices $C_1(\tilde{G}_n^0)$ of the largest connected component of $\tilde{G}_n^0$ satisfies

$$\mathbf{P}\big(|C_1(\tilde{G}_n^0)| > \rho\, n\big) \to 1. \tag{29}$$

The inclusions $\tilde{G}_n^0 \subset G_n^0 \subset G_n$ imply $|C_1(\tilde{G}_n^0)| \le |C_1(G_n^0)| \le |C_1(G_n)|$ and, by (29), we obtain

$$\mathbf{P}\big(|C_1(G_n)| > \rho\, n\big) \ge \mathbf{P}\big(|C_1(G_n^0)| > \rho\, n\big) \ge \mathbf{P}\big(|C_1(\tilde{G}_n^0)| > \rho\, n\big) \to 1. \tag{30}$$

Note that (4) follows from (30).

Let us prove (26). Denote $\delta = c_1/(12(1 + c_0)^{1+\alpha}) > 0$. Write $t_* = (1 + c_0)(2c_2/c_1)^{1/(1+\alpha)}$ and note that for large n we have $t_* < t_{0n}$. (1) implies, for $1 \le i \le n$,

$$\mathbf{P}\big(1 + c_0 < \tilde{Z}_{ni} \le t_*\big) \ge \frac{c_1}{(1 + c_0)^{1+\alpha}} - \frac{c_2}{t_*^{1+\alpha}} = 6\delta. \tag{31}$$

We assume that n is large so that $\mathbf{P}(\tilde{Z}^0_{ni} > 1) \geq 6\delta$. Denote

$$D = \cup_{v \in C_1(G^0_n)} S^0_n(v), \qquad d^* = \lfloor 2\delta\rho\sqrt{m\,n} \rfloor, \qquad k^* = \lfloor 2c_2(l_2(n))^{\alpha(1+\alpha)} \rfloor.$$

Introduce the events

$$\mathbb{A} = \{\forall v \in V^0_n : \, S_n(v) \cap D \neq \emptyset\}, \qquad \mathbb{B} = \{|D| > d^*\}, \qquad \mathbb{D} = \{|V^0_n| \leq k^*\}.$$

Note that (26) follows from the limit $\mathbf{P}(\mathbb{A}) \to 1$, which itself follows from (27) and the limits

$$\mathbf{P}(\mathbb{B}) \to 1, \qquad \mathbf{P}(\mathbb{A} \cap \mathbb{B} \cap \mathbb{D}) \to 1. \tag{32}$$

Therefore, in order to prove (26) it suffices to show (32).
Let us show the first limit of (32). Denote

$$A = \sum_{v \in C_1(G^0_n)} |S_n(v)|, \qquad B = \sum_{\{v,u\} \subset C_1(G^0_n)} |S_n(v) \cap S_n(u)|.$$

The obvious inequality $|D| \geq A - B$ combined with the bounds

$$\mathbf{P}(B > \delta\rho\sqrt{mn}) = o(1), \qquad \mathbf{P}(A < 4\delta\rho\sqrt{mn}) = o(1) \tag{33}$$

implies $\mathbf{P}(\mathbb{B}) \to 0$. It remains to prove (33). It follows from (1) that there exists a number $C > 0$ (depending only on α, c_0, c_1, c_2) such that $\mathbf{E}\tilde{Z}_{ni} \leq C$ uniformly in $n > n_0$ and $1 \leq i \leq n$. We have

$$\mathbf{E}B \leq \sum_{1 \leq i < j \leq n} \mathbf{E}|S_n(v_i) \cap S_n(v_j)| = \sum_{1 \leq i < j \leq n} \frac{\mathbf{E}\tilde{Z}_{ni}\tilde{Z}_{nj}}{n} \leq \frac{C^2}{2} n.$$

The bound $\mathbf{E}B = O(n)$ in combination with condition (i) of Theorem 1 implies the first bound of (33). Let us prove the second bound of (33). Write $V^*_n = V_n \setminus V^0_n$. We call a vertex $v \in V^*_n$ large if $\tilde{Z}_n(v) > 1$. Other vertices of V^*_n are called small. Let N^* denote the number of large vertices in V^*_n. Note that large vertices have higher probabilities of belonging to $C_1(G^0_n)$ than small ones. Therefore, the number $\tilde{N}$ of large vertices in $C_1(G^0_n)$ is stochastically larger than the number N_0 of large vertices in the simple random sample of size $|C_1(G^0_n)|$ drawn without replacement and with equal probabilities from the set V^*_n. The obvious inequality $A \geq \tilde{N}\sqrt{m/n}$ implies, for $s \geq 0$,

$$\mathbf{P}(A > s) \geq \mathbf{P}(\tilde{N} > s\sqrt{n/m}) \geq \mathbf{P}(N_0 > s\sqrt{n/m}). \tag{34}$$

We shall show that, for $s_n = 4\delta\rho n$,

$$\mathbf{P}(N_0 > s_n) \to 1. \tag{35}$$

Introduce the events $\mathbb{H} = \{|C_1(G^0_n)| > \rho\, n\}$ and $\mathbb{B}^* = \{N^* \geq 5\delta n\}$ and denote

$$p(n) = \mathbf{P}(\{N_0 > s_n\} \cap \mathbb{D} \cap \mathbb{B}^* \cap \mathbb{H}). \tag{36}$$

By the total probability formula,

$$p(n) = \sum_{h > \rho n} \sum_{b > 5\delta n} \sum_{k \leq k^*} p_{h,b,k}(n)\, \mathbf{P}\big(|C_1(G^0_n)| = h, \, N^* = b, \, |V^*_n| = n - k\big). \tag{37}$$

Here $p_{h,b,k}(n)$ denotes the conditional probability of the event $\{N_0 > s_n\}$ given $|C_1(G_n^0)| = h$, $N^* = b$, $|V_n^*| = n-k$. (8) applies to the hypergeometric probability $p_{h,b,k}(n) = \mathbf{P}(H_{h,b,n-k} > s_n)$ and, for large n, shows $p_{h,b,k}(n) \ge 1 - n^{-10}$. From (37) we obtain

$$p(n) \ge \mathbf{P}(\mathbb{D} \cap \mathbb{B}^* \cap \mathbb{H})(1 - n^{-10}). \tag{38}$$

Note that the law of large numbers combined with (27) shows $\mathbf{P}(\mathbb{B}^*) \to 1$. This limit together with (30) and (27) implies $\mathbf{P}(\mathbb{D} \cap \mathbb{B}^* \cap \mathbb{H}) \to 1$. The latter limit, (36) and (38) shows (35). Finally, (35) combined with (34) implies the second bound of (33), thus completing the proof of the limit $\mathbf{P}(\mathbb{B}) \to 1$.

Let us show the second limit of (32). The total probability formula gives

$$\mathbf{P}(\mathbb{A} \cap \mathbb{B} \cap \mathbb{D}) = \sum_{k \le k^*} \sum_{d > d^*} \mathbf{P}_{k,d}(\mathbb{A}) \mathbf{P}(|D| = d,\, |V_n^0| = k). \tag{39}$$

Here $\mathbf{P}_{k,d}$ denotes the conditional probability given $|D| = d$, $|V_n^0| = k$. Let $S^* = \{|S_n(v)|, v \in V_n^0\}$ denote the collection of sizes of sets $S_n(v)$ of vertices $v \in V_n^0$. Note that for $|V_n^0| = k$, the collection $S^* = \{s_1, \dots, s_k\}$ is a multiset. We have

$$\mathbf{P}_{k,d}(\mathbb{A}) = \sum_{\{s_1,\dots,s_k\}} \mathbf{P}_{k,d}\bigl(\mathbb{A} \big| S^* = \{s_1, \dots, s_k\}\bigr)\, \mathbf{P}_{k,d}(S^* = \{s_1, \dots, s_k\}). \tag{40}$$

Here the sum is taken over all possible values $\{s_1, \dots, s_k\}$ of the multiset S^* of cardinality k. The identity

$$\mathbf{P}_{k,d}\bigl(\mathbb{A} \big| S^* = \{s_1, \dots, s_k\}\bigr) = \prod_{j=1}^{k} \mathbf{P}(H_{s_j,d,m} \ge 1)$$

combined with (9) implies, for large n, the inequality $\mathbf{P}_{k,d}\bigl(\mathbb{A} \big| S^* = \{s_1, \dots, s_k\}\bigr) \ge 1 - n^{-10}$ uniformly in d, k and $s_1, \dots, s_k$ satisfying the inequalities $s_j \ge t_{0n}\sqrt{m/n}$, $d > d^*$, $k \le k^*$. Now (40) implies the inequality $\mathbf{P}_{k,d}(\mathbb{A}) \ge 1 - n^{-10}$, for $d > d^*$ and $k \le k^*$. Invoking the latter inequality in (39) we obtain

$$\mathbf{P}(\mathbb{A} \cap \mathbb{B} \cap \mathbb{D}) \ge (1 - n^{-10}) \mathbf{P}(\mathbb{B} \cap \mathbb{D}) = 1 - o(1).$$

In the last step we used (27) and the first bound of (32). The proof of (32) is complete.

□

Proof of Theorem 1. Before the proof we introduce some notation. Denote

$$t_0 = n^{1/(1+\alpha)} l_2^{-\alpha}(n), \qquad t_k = n^{\alpha^k/(1+\alpha)} l_2(n), \quad k = 1, 2, \dots, \tag{41}$$
$$k_* = \max\{k : n^{\alpha^k/(1+\alpha)} \ge 100 + c_0\}.$$

We use the following simple properties of the sequence $\{t_k\}$. For $n > 9$ we have

$$t_0\, t_1/n = l_2^{1-\alpha}(n), \qquad t_k\, t_{k-1}^{-\alpha} = l_2^{1-\alpha}(n), \qquad k = 2, 3, \dots, \tag{42}$$
$$100\, l_2(n) < t_{k_*} < (100 + c_0)^{1/\alpha}\, l_2(n), \qquad k_* \le l_2(n)/\ln(1/\alpha). \tag{43}$$

Given $\mathcal{U} \subset V_n$ we denote $S(\mathcal{U}) = \cup_{v \in \mathcal{U}} S_n(s)$. Throughout the proof limits are taken as $n \to \infty$. Given n, write $m = m_n$ and $T = T_n = n^{1/(1+\alpha)}\omega_n$. Fix $1 < \tau < 1 + \alpha$. By $c_1^*, c_2^*, \dots$ we denote positive constants that may depend only on $\alpha, \tau, c_0, c_1, c_2$.

Let us prove (2). Fix a maximal vertex u_n of G_n. We have

$$\mathbf{P}\big(d(v_1,u_n) > k_* + \varepsilon l_2(n) \,\big|\, d(v_1,u_n) < \infty\big) = \frac{\mathbf{P}\big(d(v_1,u_n) > k_* + \varepsilon l_2(n),\ d(v_1,u_n) < \infty\big)}{\mathbf{P}\big(d(v_1,u_n) < \infty\big)}.$$

In order to prove (2) we shall show that

$$\mathbf{P}\big(d(v_1,u_n) > k_* + \varepsilon l_2(n),\ d(v_1,u_n) < \infty\big) = o(1), \tag{44}$$
$$\liminf_n \mathbf{P}\big(d(v_1,u_n) < \infty\big) > 0. \tag{45}$$

Let us prove (45). Write $C_1 = C_1(G_n)$. It follows from (4) that $\mathbf{P}(u_n \in C_1) = 1 - o(1)$. Therefore, we have

$$\mathbf{P}(d(v_1,u_n) < \infty) \geq \mathbf{P}(v_1, u_n \in C_1) = \mathbf{P}(v_1 \in C_1) - o(1). \tag{46}$$

Inequalities (30) imply $\mathbf{E}|C_1| \geq \rho n(1-o(1))$ and, by symmetry, we obtain

$$\mathbf{P}(v_1 \in C_1) = n^{-1}\sum_{v\in V}\mathbf{P}(v \in C_1) = n^{-1}\mathbf{E}|C_1| \geq \rho(1-o(1)).$$

This inequality combined with (46) implies (45).
Let us prove (44). Introduce the sets

$$\mathcal{U}_0 = \{u_n\}, \qquad \mathcal{U}_k = \{v_j :\ t_k \leq \tilde{Z}_{nj} \leq T\}, \qquad k = 1,2,\dots,k_*,$$
$$\mathcal{U}_* = \{v_j :\ 1 \leq \tilde{Z}_{nj} \leq t_*\}.$$

Denote $Q_k = \sum_{v\in\mathcal{U}_k}|S_n(v)|$ and $q_k = \mathbf{E}Q_k$. Introduce the events

$$\mathbb{A}_0 = \{t_0 \leq \tilde{Z}_n(u_n) \leq T\},$$
$$\mathbb{A}_k = \{q_k/2 \leq Q_k \leq (3/2)q_k\}, \qquad k = 1,2,\dots,k_*.$$
$$\mathbb{A}_{*1} = \big\{|\mathcal{U}_*| \geq 5n\delta\big\}, \qquad \mathbb{A}_{*2} = \big\{|\mathcal{U}_{k_*}| \leq n/l_2(n)\big\},$$

Here $\delta > 0$ is defined in (31) above. Denote $\tilde{\mathbb{A}} = \Big(\cap_{k=0}^{k_*}\mathbb{A}_k\Big) \cap \mathbb{A}_{*1} \cap \mathbb{A}_{*2}$. Let us show that

$$\mathbf{P}(\tilde{\mathbb{A}}) \to 1. \tag{47}$$

(47) follows from the limits

$$\mathbf{P}(\mathbb{A}_{*i}) \to 1, \quad i = 1,2, \qquad \mathbf{P}(\mathbb{A}_0) \to 1, \qquad \mathbf{P}\big(\cap_{1\leq k\leq k_*}\mathbb{A}_k\big) \to 1. \tag{48}$$

An application of Chebyshev's inequality to the binomial random variables $|\mathcal{U}_{k_*}|$ and $|\mathcal{U}_*|$ gives the first limit of (48). The second limit of (48) is shown in (18). To show the third limit of (48) we write

$$1 - \mathbf{P}(\cap_{1\leq k\leq k_*}\mathbb{A}_k) = \mathbf{P}(\cup_{1\leq k\leq k_*}\overline{\mathbb{A}}_k) \leq \sum_{1\leq k\leq k_*}\mathbf{P}(\overline{\mathbb{A}}_k).$$

Here $\overline{\mathbb{A}}_k$ denotes the event complement to $\mathbb{A}_k$. Combining the bound, which follows from (20),

$$\mathbf{P}(\overline{\mathbb{A}}_k) \leq c_1^* n^{(\alpha^k-1)(\tau-1)}\big(l_2(n)\big)^{(\alpha+1)(\tau-1)}$$

and the bound, see (43), $k_* = O(l_2(n))$ we obtain $\sum_{1\le k\le k_*} \mathbf{P}(\overline{\mathbb{A}}_k) = o(1)$, thus showing the third limit of (48). We arrive at (47).
In the remaining part of the proof we shall assume that the event $\tilde{\mathbb{A}}$ holds. Let $\bar{\mathbf{P}}$, $\bar{\mathbf{E}}$ and $\tilde{G}$ denote the conditional probability, the conditional expectation, and the conditional random graph G_n given $\tilde{Z}_{n1}, \dots, \tilde{Z}_{nn}$. Write $V^* = V \setminus \mathcal{U}_{k_*}$ and let G^* denote the subgraph of $\tilde{G}$ induced by V^*. Given $v \in V^*$ define $d_*(v) = \min\{d(w,v) : w \in \mathcal{U}_{k_*}\}$. We shall show that uniformly in $\tilde{Z}_{n1}, \dots, \tilde{Z}_{nn}$ satisfying $\tilde{\mathbb{A}}$ and uniformly in $v \in V^*$, $u \in \mathcal{U}_{k_*}$

$$\bar{\mathbf{P}}(d_*(v) > \varepsilon l_2(n), d_*(v) < \infty) = o(1), \quad (49)$$

$$\bar{\mathbf{P}}(d(u, u_n) \ge k_*) = o(1). \quad (50)$$

It follows from (49) that a vertex $v \in V$ satisfying $d(v, u_n) < \infty$ finds whp a path of length at most $l_2(n)$ to a vertex $u \in \mathcal{U}_{k_*}$. (50) then applies to u and together with (49) imply

$$\bar{\mathbf{P}}\bigl(d(v_1, u_n) > k_* + \varepsilon l_2(n),\, d(v_1, u_n) < \infty\bigr) = o(1). \quad (51)$$

The bound (51) combined with (47) shows (44). It remains to prove (49, 50).
Proof of (49). For simplicity of notation we put $\varepsilon = 1$. Given $v \in V^*$ denote $L_* = \{v' \in V^* : d^*(v, v') \le l_2(n)\}$. Here d^* denotes the distance between vertices of the graph G^*. Introduce the event $\mathbb{B}_* = \{|S(L_*)| \ge \delta\, l_2(n)\sqrt{m/n}\}$. The event

$$\begin{aligned} \{d_*(v) > l_2(n),\, d_*(v) < \infty\} \;&\subset\; \{S(L_*) \cap S(\mathcal{U}_{k_*}) = \emptyset,\, |L_*| > l_2(n)\} \\ &\subset\; \bigl(\{S(L_*) \cap S(\mathcal{U}_{k_*}) = \emptyset\} \cap \mathbb{B}_*\bigr) \cup \bigl(\overline{\mathbb{B}}_* \cap \{|L_*| > l_2(n)\}\bigr). \end{aligned}$$

Here $\overline{\mathbb{B}}_*$ denote the event complement to $\mathbb{B}_*$. We have

$$\bar{\mathbf{P}}\bigl(d_*(v) > l_2(n),\, d_*(v) < \infty\bigr) \le p' + p'', \quad (52)$$

where

$$p' = \bar{\mathbf{P}}\bigl(\{S(L_*) \cap S(\mathcal{U}_{k_*}) = \emptyset\} \cap \mathbb{B}_*\bigr), \qquad p'' = \bar{\mathbf{P}}\bigl(\overline{\mathbb{B}}_* \cap \{|L_*| > l_2(n)\}\bigr).$$

We shall show that

$$p' = o(1), \qquad p'' = o(1) \qquad \text{as} \qquad n \to \infty. \quad (53)$$

Let us prove the first bound. (19) and (43) imply $q_{k_*} > 4c_2^*\sqrt{mn}/(l_2(n))^\alpha$. Invoking the inequality $Q_{k_*} \ge q_{k_*}/2$ and the inequality, which follows from Lemma 2, $\bar{\mathbf{P}}(|S(\mathcal{U}_{k_*})| \ge Q_{k_*}/2) = 1 - o(1)$ we obtain the bound $1 - \bar{\mathbf{P}}(\mathbb{B}') = o(1)$ for the event $\mathbb{B}' = \{|S(\mathcal{U}_{k_*})| > c_2^*\sqrt{mn}/(l_2(n))^\alpha\}$. Therefore, we have

$$\begin{aligned} p' &= \bar{\mathbf{P}}(\{S(L_*) \cap S(\mathcal{U}_{k_*}) = \emptyset\} \cap \mathbb{B}' \cap \mathbb{B}_*) + o(1) \\ &\le \bar{\mathbf{P}}(S(L_*) \cap S(\mathcal{U}_{k_*}) = \emptyset \bigl| \mathbb{B}', \mathbb{B}_*) + o(1) \\ &= o(1). \end{aligned}$$

In the last step we applied (9) to the random variable $H = \bigl|S(L_*) \cap S(\mathcal{U}_{k_*})\bigr|$ conditionally, given $|S(L_*)|$ and $|S(\mathcal{U}_{k_*})|$.
Let us show the second bound of (53). Denote $k' = \lfloor l_2(n) \rfloor$. Let $\{v_1', v_2', \dots, v_{n'}'\}$ be an enumeration of elements of V^*. We call v_i' smaller than v_j' whenever $i < j$. We call $v' \in V^*$ large if $\tilde{Z}_n(v') \ge 1$. Paint elements of V^* white. Given $v \in V^*$ we construct the 'breath first search' tree T_v in G^* with the root v as follows. Paint vertex v black and write $\tau_0 = v$.

White vertices are checked in increasing order and those found adjacent to τ_0 are painted black. Denote them $\tau_1 < \tau_2 < \cdots < \tau_{j_1}$. After all neighbours of τ_0 have been found the vertex τ_0 is called saturated. Then proceed recursively: take the first available black unsaturated vertex, say τ_i (here $i = \min\{j : \tau_j$ is black and unsaturated $\}$), and find its neighbours among remaining white vertices. Do this by checking white vertices in increasing order. After all white neighbours of τ_i have been found the vertex τ_i is called saturated, the neighbours are denoted $\tau_{j_{i-1}+1} < \tau_{j_{i-1}+2} < \cdots < \tau_{j_i}$ and painted black. We call τ_i the parent vertex of its children $\tau_{j_{i-1}+1}, \dots, \tau_{j_i}$. In this way we obtain the list $L = \{\tau_0, \tau_1, \dots\}$ of vertices of the tree T_v. Denote $L_r = \{\tau_0, \dots, \tau_r\}$. Let $\tilde{N}$ denote the number of large vertices in the set $L_{k'}$. We say that (player) v receives a yellow card at step $r \geq 1$ if vertex τ_r is large and $|S(L_{r-1}) \cap S_n(\tau_r)| \geq 2^{-1}|S_n(\tau_r)|$. The event that v receives the first yellow card at step r is denoted B_r. On the event $\mathbb{H} := \left(\cap_{i=1}^{k'} \overline{B}_i\right) \cap \{\tilde{N} \geq 4\delta k'\}$ we have

$$|S(L_{k'})| \geq 2^{-1}\tilde{N}\sqrt{m/n} \geq \delta\, l_2(n)\sqrt{m/n}. \tag{54}$$

Note that the inequality $|L_*| > l_2(n)$ implies $|L| \geq k' + 1$. Therefore, we have

$$p'' \leq \bar{\mathbf{P}}(\overline{\mathbb{B}}_* \cap \{|L| > k' + 1\}).$$

Furthermore, for $|L| \geq k' + 1$, the inclusion $L_{k'} \subset L_*$ implies $|S(L_{k'})| \leq |S(L^*)|$ and in view of (54) we conclude that events $\mathbb{H}$ and $\overline{\mathbb{B}}_* \cap \{|L| > k' + 1\}$ do not intersect. We have

$$\bar{\mathbf{P}}(\overline{\mathbb{B}}_* \cap \{|L| > k' + 1\}) = \bar{\mathbf{P}}(\overline{\mathbb{B}}_* \cap \{|L| > k' + 1\} \cap \overline{\mathbb{H}}) \leq p_1^* + p_2^*,$$

where

$$p_1^* := \bar{\mathbf{P}}(\{|L| > k' + 1\} \cap \{\tilde{N} < 4\delta k'\}), \qquad p_2^* := \bar{\mathbf{P}}(\cup_{r=1}^{k'} \mathbb{B}_r).$$

In order to prove the bound $p'' = o(1)$ we shall show that $p_i^* = o(1)$, $i = 1, 2$.
Write

$$p_1^* = \bar{\mathbf{P}}(|L| > k' + 1)\tilde{p}, \qquad \tilde{p} := \bar{\mathbf{P}}(\tilde{N} < 4\delta k' \,\big|\, |L| > k' + 1). \tag{55}$$

Since large vertices have higher probabilities to join the list L than the other vertices we conclude that the random variable $\tilde{N}$ is stochastically larger than the number N_0 of large vertices in the simple random sample of size $k' + 1$ drawn without replacement and with equal probabilities from the set V^*. In particular, we have

$$\tilde{p} \leq \bar{\mathbf{P}}(N_0 < 4\delta k'). \tag{56}$$

Note that on the event $\mathbb{A}_{*1} \cap \mathbb{A}_{*2}$ we have $\mathbf{E}N_0 \geq 5\delta(k' + 1)$ and $\mathbf{Var}N_0 = O(k')$. Therefore, Chebyshev's inequality implies $\bar{\mathbf{P}}(N_0 < 4\delta k') = O(1/k') = o(1)$. This bound combined with (55) and (56) implies the bound $p_1^* = o(1)$.
In order to prove the bound $p_2^* = o(1)$ we write $p_2^* \leq \sum_{r=1}^{k'} \bar{\mathbf{P}}(\mathbb{B}_r)$ and show that

$$\bar{\mathbf{P}}(\mathbb{B}_r) \leq n^{-10}, \tag{57}$$

for every r and large n. Before the proof of (57) we introduce some notation. For $i \geq 1$ denote $W_i = W \setminus S(L_{i-1})$, $S'(\tau_i) = S_n(\tau_i) \setminus S(L_{i-1})$, $m_i = |W_i|$, $s_i = |S_n(\tau_i)|$, $s_i' = |S'(\tau_i)|$. Put $W_0 = W$. Fix $r \geq 1$. Let τ_{r^*} denote the parent vertex of τ_r. Denote $D_r = S(L_{r-1}) \cap W_{r^*}$ and $d_r = |D_r|$. We have $\bar{\mathbf{P}}(\mathbb{B}_r \big| W_{r^*}, D_r, S'(t_{r^*}), s_r) \leq p_*$, where

$$p_* := \bar{\mathbf{P}}_{r^*}\left(|D_r \cap S_n(\tau_r)| \geq 2^{-1}s_r \,\Big|\, S_n(\tau_r) \cap S'(\tau_{r^*}) \neq \emptyset\right). \tag{58}$$

Here $\bar{\mathbf{P}}_{r^*}$ denotes the conditional probability $\bar{\mathbf{P}}$ given $W_{r^*}, D_r, S'(t_{r^*}), s_r$. Note that in (58) values of all random variables are fixed (given), but $S_n(\tau_r)$ which is a random set uniformly distributed in the class of subsets of W_{r^*} of given size s_r satisfying $s_r \geq \sqrt{m/n}$ (because τ_r is a large vertex). It follows from (12) that for large n we have

$$p_* \leq e^{-s_r/8}(1 + 16m_{r^*}/s_r^2) \leq e^{-8^{-1}\sqrt{m/n}}(1 + 16n) < n^{-10}. \tag{59}$$

In the last step we applied condition (i) of Theorem 1. (59) implies (57) thus completing the proof of (53). We arrive to (49).

Proof of (50). Given $u'_0 \in \mathcal{U}_{k_*}$ finds a neighbour in $\mathcal{U}_{k_*-1}$, say, u'_1 with probability at least

$$\min_{u \in \mathcal{U}_{k_*}} \bar{\mathbf{P}}\big(S_n(u) \cap S(\mathcal{U}_{k_*-1}) \neq \emptyset\big) =: p_0^*.$$

Similarly, given $u'_j \in \mathcal{U}_{k_*-j}$ finds a neighbour in $\mathcal{U}_{k_*-j-1}$, say u'_{j+1}, with probability at least

$$\min_{u \in \mathcal{U}_{k_*-j}} \bar{\mathbf{P}}\big(S_n(u) \cap S(\mathcal{U}_{k_*-j-1}) \neq \emptyset\big) =: p_j^*,$$

and so on. In this way we may construct a path (namely, $u'_0, u'_1, u'_2, \ldots, u'_{k_*} = u_n$) of length at most k_* connecting u_n with an arbitrary vertex u'_0 from $\mathcal{U}_{k_*}$. The probability that such a construction fails is at most $\sum_{j=0}^{k_*-1}(1 - p_j^*)$. In particular, for any given $u \in \mathcal{U}_{k_*}$, we have

$$\bar{\mathbf{P}}(d(u, u_n) > k_*) \leq \sum_{j=0}^{k_*-1}(1 - p_j^*).$$

In order to prove (50) we shall show that, for some $c_4^* > 0$ and large n,

$$1 - p_j^* \leq e^{-c_4^*(l_2(n))^{1-\alpha}} + n^{-2}, \qquad 0 \leq j \leq k_* - 1. \tag{60}$$

Fix $1 \leq i \leq k_* - 1$ and $u \in \mathcal{U}_{i+1}$. On the event $\tilde{\mathbb{A}}$ we have, for large n,

$$Q_i \geq \frac{q_i}{2} \geq \frac{c_1}{4}\frac{1+\alpha}{\alpha}\frac{\sqrt{mn}}{t_i^\alpha}, \tag{61}$$

where the second inequality of (61) follows from (19). Denote $c_4^* = \frac{c_1}{16}\frac{1+\alpha}{\alpha}$ and introduce the event $\mathbb{B} = \{|S(\mathcal{U}_i)| \geq 2c_4^*\sqrt{m\,n}\,t_i^{-\alpha}\}$. It follows from Lemma 2 (applied to $\gamma_1 = 1/10$ and $\gamma_2 = 1/2$) that

$$1 - n^{-2} \leq \bar{\mathbf{P}}(|S(\mathcal{U}_i)| \geq Q_i/2) \leq \bar{\mathbf{P}}(\mathbb{B}). \tag{62}$$

Here in the last step we invoke (61). Next we apply (9) to the hypergeometric random variable $H = |S_n(u) \cap S(\mathcal{U}_i)|$, where $|S_n(u)|$ and $|S(\mathcal{U}_i)|$ are given and satisfy $|S_n(u)| \geq t_{i+1}\sqrt{m/n}$ and $|S(\mathcal{U}_i)| \geq 2c_4^*\sqrt{m\,n}\,t_i^{-\alpha}$. We obtain

$$\begin{aligned}\bar{\mathbf{P}}\big(S_n(u) \cap S(\mathcal{U}_i) = \emptyset\,\big|\mathbb{B}\big) &\leq \exp\{-c_4^*\frac{t_{i+1}}{t_i^\alpha}\} \\ &= \exp\{-c_4^*(l_2(n))^{1-\alpha}\}.\end{aligned} \tag{63}$$

Combining (62) and (63) we obtain (60), for $j = k_* - i - 1$ satisfying $0 \leq j \leq k_* - 2$. The proof of (60) for $j = k_* - 1$ is similar but simpler. We arrive to (50) thus completing the proof of (2).

Let us prove (3). Denote $a_n = (1+\varepsilon/2)(1/\alpha)\ln(\ln(2+n))$ and introduce the events

$$\mathbb{D} = \{v_1, v_2 \in C_1\}, \qquad \mathbb{G} = \{d(v_1,v_2) > 2a_n\}, \qquad \mathbb{G}_i = \{d(v_i,u_n) > a_n\}, \qquad i=1,2.$$

Note that (3) is equivalent to the limit $\mathbf{P}(\mathbb{G}|\mathbb{D}) = o(1)$. In order to prove (3) we shall show that that there exists $\rho > 0$ such that

$$\liminf_n \mathbf{P}(\mathbb{D}) > \rho^2, \tag{64}$$

$$\mathbf{P}(\mathbb{G}\cap\mathbb{D}) = o(1). \tag{65}$$

Let us prove (64). It follows from the identity $|C_1| = \sum_{v\in V}\mathbb{I}_{\{v\in C_1\}}$, by the symmetry, that

$$\begin{aligned}\mathbf{E}|C_1|^2 \quad &= \mathbf{E}\sum_{v\in V}\mathbb{I}^2_{\{v\in C_1\}} + \mathbf{E}\sum_{\{u,v\}\in V}\mathbb{I}_{\{u,v\in C_1\}} \\ &= \mathbf{E}|C_1| + n(n-1)\mathbf{P}(\mathbb{D}).\end{aligned}$$

This identity combined with $|C_1| \le n$ and the inequality, which follows from (4), $\mathbf{E}|C_1|^2 \ge n^2\rho^2(1-o(1))$ shows (64).

Let us prove (65). In view of (5) in suffices to show that $p := \mathbf{P}(\mathbb{G}\cap\mathbb{D}\cap\{u_n\in C_1\}) = o(1)$. We have $p \le p_1 + p_2$, where $p_i = \mathbf{P}(d(v_i,u_n) > a_n,\ d(v_i,u_n) < \infty)$, $i=1,2$. Finally, (44) implies $p_i = o(1)$ thus completing the proof of (65).

□

References

[1] M. Bloznelis, Degree distribution of a typical vertex in a general random intersection graph, Lithuanian Mathematical Journal, **48** (2008), 38–45

[2] M. Bloznelis, Component evolution in general random intersection graphs, Submitted to SIAM journal on Discrete Mathematics (2008)

[3] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogenious random graphs, Random Structures Algorithms **31** (2007), 3–122

[4] F. Chung and L. Lu, The average distance in a random graph with given expected degrees, Internet Mathematics, **1** (2003), 91–114

[5] R. Di Pietro, L.V. Mancini, A. Mei and 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 (SecureComm 2006)

[6] M. Deijfen and W. Kets, Random intersection graphs with tunable degree distribution and clustering, Discussion paper No. 2007-08, Tilburg University, 2007.

[7] L. Eschenauer and V. D. Gligor, A key-management scheme for distributed sensor networks, in: Proceedings of the 9th ACM conference on computer and communications security (2002), 41–47

[8] E. Godehardt and J. Jaworski, Two Models of Random Intersection Graphs for Classification, in: Studies in Classification, Data Analysis and Knowledge Organization, Springer, Berlin–Heidelberg– New York, 2003, 67–81

[9] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc., **58** (1963), 13–30

[10] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski, Distances in random graphs with finite mean and infinite variance degrees, Electron. J. Probab. **12** (2007), 703-766

[11] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, Wiley, New York, 2001

[12] J. Jaworski, M. Karoński and D. Stark, The degree of a typical vertex in generalized random intersection graph models, Discrete Mathematics **306** (2006), 2152–2165

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

[14] I. Norros and H. Reittu, On the effect of very large nodes in Internet graphs, In *Globecom'02*, Taipei, Taiwan, (2002)

[15] I. Norros and H. Reittu, On a conditionally Poissonian graph process, Adv. Applied Probability, 38 (2006), 59-75