Component evolution in general random intersection graphs

M. Bloznelis

Faculty of Mathematics and Informatics, Vilnius University, LT-03225 Vilnius, Lithuania
E-mail *mindaugas.bloznelis@mif.vu.lt*

Abstract

Given integers n, m and a probability distribution P_* on $[m] = \{1, \dots, m\}$, consider the random intersection graph on the vertex set $[n]$, where $i, j \in [n]$ are declared to be adjacent whenever $S(i) \cap S(j) \neq \emptyset$. Here $S(1), \dots, S(n)$ denote iid random subsets of $[m]$ with the distribution $\mathbf{P}(S(i) = A) = \binom{m}{|A|}^{-1} P_*(|A|)$, for $A \subset [m]$. Assuming that m is much larger than n, we show that the order of the largest connected component $N_1 = n\rho + o_P(n)$ as $n, m \to \infty$. Here ρ denotes the nonextinction probability of a related multi-type Poisson branching process.

key words: intersection graph, random graph, giant component, component evolution

1 Introduction

Given subsets $S(1), \dots, S(n)$ of a 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.

We consider a class of random intersection graphs where the random subsets $S(i) = S(v_i)$, $i = 1, \dots, n$, are independent and identically distributed. Moreover, we assume that 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_* denoting the distribution of $|S(i)|$, we have, for every $A \subset W$, $\mathbf{P}(S(i) = A) = \binom{m}{|A|}^{-1} P_*(|A|)$. The random intersection graph corresponding to P_* is denoted $G(n, m, P_*)$.

An attractive property of $G(n, m, P_*)$ is that by a proper choice of P_* one can obtain the vertex degree distribution with some desirable properties, e.g., heavy-tailed distribution, see [3], [8]. Note also that $G(n, m, P_*)$ has (statistically) *dependent* edges.

Random intersection graphs (RIG) with binomial distribution P_* were studied by Karoński, Scheinerman, and Singer-Cohen [20] and Singer-Cohen [21]. Up to our best knowledge, these are the first papers devoted to RIG. Random intersection graphs with general P_* were first

studied in Godehard and Jaworski [12]. They are applied in the analysis of secure wireless networks [9], [11], [14], social networks [8], and statistical classification [12].
We study the component evolution of random intersection graphs $G(n, m, P_*)$ in the case where m is much larger than n. Note that if we choose P_* such that $P_*(\sqrt{cm/n}) = 1$, where $c \geq 0$ is a constant, then the expected value of the degree of a typical vertex is approximately c, see [12], [17]. Therefore, one may expect to observe a fast growth of the largest connected component in $G(n, m, P_*)$ when $|S(i)|$ (the size of a typical random set) is of stochastic order $\Theta_P(\sqrt{m/n})$. Assuming that, as $n, m \to \infty$, the probability distribution of $\sqrt{n/m}\,|S(i)|$ converges to some limiting probability distribution on $[0, +\infty)$, say $\tilde{P}$, we describe the asymptotics of the size of the largest connected component in terms of $\tilde{P}$.
The paper is organized as follows: results are stated and discussed in Section 2, and proofs are given in Section 3.

2 Results

We start with mentioning related results for random graphs with *independent* edges. Erdős and Rényi [10] showed the fast growth of the largest connected component of the random graph $G_{n,p}$ (obtained from the complete graph $\mathcal{K}_n$ by removing edges independently and with probability $1-p$) in the case where the edge density p passes the threshold $p = 1/n$ as $n \to \infty$. Namely, for a constant $c > 0$ and $p = c/n$, the order of the largest connected component (denoted N_1) is $N_1 = O_P(\log n)$ if $c < 1$ and $N_1 = \rho_c n + o_P(n)$ if $c > 1$. Here, ρ_c denotes the survival probability of the Galton–Watson branching process with Poisson offspring distribution having mean value c. Recently, the component evolution of a general inhomogeneous random graph with independent edges was studied by Bollobás, Janson, and Riordan [5]. Let $V = \{v_1, \dots, v_n\}$ be the vertex set, and let $p_{ij}^{(n)} = P(v_i \sim v_j)$ denote the probability of the event that vertices $v_i, v_j \in V$ are adjacent. Assuming mild regularity conditions on the sequence of probabilities $\{p_{ij}^{(n)}, 1 \leq i < j \leq n\}_{n=1}^{\infty}$, Bollobás, Janson, and Riordan [5] showed that

$$N_1 = \rho_* n + o_P(n) \qquad \text{as} \qquad n \to \infty. \tag{1}$$

Here, $\rho_* > 0$ is the expected survival probability of the multi-type Galton–Watson branching process with Poisson offspring distribution that approximates the neighborhood discovery process performed by the breath first search procedure starting at a randomly chosen vertex from V.
Behrisch [2] extended the above-mentioned result of Erdős and Rényi [10] to the random intersection graph $G(n, m, P_*)$ with binomial distribution P_* and $m = n^\alpha$, $\alpha > 1$. For degenerate P_*, a similar result was shown in [4]. In the present paper, we extend (1) to random intersection graphs $G(n, m, P_*)$ with general P_* in the case where $n \ln^2 n = o(m)$.
In order to trace the connection between $G(n, m, P_*)$ and the random graph model studied in [5], it is convenient first to look at $G(n, m, P_*)$ conditioned on $\mathbb{S} = \{|S(v_1)|, \dots, |S(v_n)|\}$. In this way, we obtain an inhomogeneous random intersection graph, say $G_{\mathbb{S}}$, where the sizes of the sets are now nonrandom. For m much larger than n, edges of $G_{\mathbb{S}}$ are approximately independent as $m, n \to \infty$, and we have $\mathbf{P}\big(v_i \sim v_j \,\big|\, |S(v_i)|, |S(v_j)|\big) \approx |S(v_i)| \times |S(v_j)|/m$. Therefore, $G_{\mathbb{S}}$ resembles the inhomogeneous random graph with independent edges having probabilities of the form $p_{ij}^{(n)} = p_i^{(n)} p_j^{(n)}/m$. Here $p_i^{(n)} = |S(v_i)|$, $1 \leq i \leq n$, represents the idd sample from the probability distribution P_*. Note that the giant component of an inhomogeneous random graph with independent edges having probabilities of the form $P(v_i \sim v_j) = p_i p_j$ (with nonrandom weights p_i) has been studied in [7].

Before stating our result, we introduce some notation. Given a probability distribution $\tilde{P}$ on $[0,+\infty)$, let $\mathcal{X}_{\tilde{P}}$ denote the multi-type Galton–Watson branching process, where particles are of prescribed types from the set $[0,+\infty)$ and where a particle of type $y \in [0,+\infty)$ is replaced in the next generation by a set of particles distributed as a Poisson process on $[0,+\infty)$ with intensity $xy\tilde{P}(dx)$. In particular, the number of children of y with types $x \in [a,b]$ has Poisson distribution with mean $\int_a^b xy\tilde{P}(dx)$, see, e.g., [5]. By $\mathcal{X}_{\tilde{P}}(y)$ we denote the process $\mathcal{X}_{\tilde{P}}$ starting at a particle of type y. $\rho(\tilde{P},y)$ denotes the probability of nonextinction of $\mathcal{X}_{\tilde{P}}(y)$, and $\rho(\tilde{P}) := \int \tilde{\rho}(y)\tilde{P}(dy)$. We consider the sequence of random intersection graphs $G(n, m_n, P_n)$, $n = 1, 2, \ldots$. Define the sequence $\{\tilde{P}_n\}$ of probability distributions on $[0,+\infty)$ by putting $\tilde{P}_n(A) = P_n\big((n/m_n)^{1/2}A\big)$ for every Borel set $A \subset [0,+\infty)$. Let $\tilde{Z}_n$ be a random variable with the distribution $\tilde{P}_n$. Note that $(m_n/n)^{1/2}$ is the scale of the size of a typical random set, say $S_n(v)$, in the case where the degrees of $G(n, m_n, P_n)$ are stochastically bounded as $n \to \infty$. Therefore, it is convenient to state conditions in terms of distributions $\tilde{P}_n$ of rescaled random variables $(n/m_n)^{1/2}|S_n(v)|$. Let $N_1(G)$ denote the order of the the largest connected component of a graph G (i.e., $N_1(G)$ is the number of vertices of the connected component which has the largest number of vertices).

Theorem 1. *Let $\tilde{P}$ be a probability distribution on $[0,+\infty)$. Let $\tilde{Y}$ be a random variable with the distribution $\tilde{P}$. Assume that $\mathbf{E}\tilde{Y} < \infty$ and $\mathbf{E}\tilde{Z}_n < \infty$ for all n. Assume that, as $n \to \infty$,*
(i) $\{\tilde{P}_n\}$ converge weakly to $\tilde{P}$;
(ii) $\mathbf{E}\tilde{Z}_n$ converge to $\mathbf{E}\tilde{Y}$;
(iii) $n \ln^2 n = o(m_n)$.
Then $N_1(G(n, m_n, P_n)) = n\rho(\tilde{P}) + o_P(n)$ as $n \to \infty$.

Remark 1. Conditions (i) and (ii) together imply the uniform integrability of the sequence of the random variables $\{\tilde{Z}_n\}$, i.e.,

$$\forall\, \varepsilon > 0\ \exists\ \Delta > 0 \ \text{ such that } \forall\ n \geq 1 \ \text{ we have } \ \mathbf{E}\tilde{Z}_n \mathbb{I}_{\{\tilde{Z}_n > \Delta\}} < \varepsilon. \qquad (2)$$

For a sequence of numbers a_n, we write $a_n = \Theta(n)$ if there exist positive constants $c_1 < c_2$ and n_0 such that $c_1 n \leq a_n \leq c_2 n$ for $n > n_0$.

Remark 2. Assume that the conditions of Theorem 1 are satisfied and $\mathbf{E}\tilde{Y}^2 < \infty$. It this case, we have $\mathbf{E}\tilde{Y}^2 \leq 1$ whenever $\rho(\tilde{P}) = 0$, see Lemmas 5.11–16 in [5]. Therefore, from Theorem 1 it follows that, with probability tending to 1, we have $N_1 = o(n)$ for $\mathbf{E}\tilde{Y}^2 \leq 1$ and $N_1 = \Theta(n)$ for $\mathbf{E}\tilde{Y}^2 > 1$.

Remark 3. The order $N_2(G)$ of the second largest component satisfies $N_2(G) = o_P(n)$ as $n \to \infty$.

It would be interesting to learn more about the order N_1 of the largest connected component in the case where $\rho(\tilde{P}) = 0$, and about the order $N_2(G)$ of the second largest component in the case where $\rho(\tilde{P}) > 0$. It is likely that a more precise asymptotics is related to the order of decay of the tail $P(\tilde{Y} > t)$ as $t \to \infty$, c.f. [16].
The result of Theorem 1 should be considered as a first step in understanding the structure of general random intersection graphs with $n = o(m)$. The interesting and important class of random intersection graphs $G(n, m, P_*)$ with $n = O(m)$ exhibit stronger clustering and is beyond the scope of this paper.
Acknowledgment. I am grateful to Michał Karoński who introduced me to random intersection graphs and suggested the problem.

3 Proof

The section is organized as follows. In Section 3.1, we collect auxiliary inequalities for hypergeometric probabilities. In Section 3.2, we collect some facts about survival probabilities of Galton–Watson (G–W)processes. Theorem 1 and Remark 3 are proved in Section 3.3.

3.1. In what follows, $H_{j,k,m}$ 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_a, S_b be independent random subsets of the set $W' = \{1, \dots, m\}$ such that S_a (respectively S_b) is uniformly distributed in the class of subsets of W' of size a (respectively b). Then $H = |S_a \cap S_b|$ is a hypergeometric random variable with parameters $a, b \le m$. The probability $p' := \mathbf{P}(H = 0) = (m-b)_a/(m)_a$ satisfies, for $a + b < m$,*

$$1 - \frac{ab/m}{1-(a+b)/m} \le p' \le 1 - \frac{ab}{m} + \big(\frac{ab}{m}\big)^2. \qquad (3)$$

Here we denote $(m)_a = m(m-1)\cdots(m-a+1)$. For $0 < \alpha < 1$ and $a + b \le \alpha m$, we have

$$\frac{ab}{m} + \frac{2}{1-\alpha}\big(\frac{ab}{m}\big)^2 \ge \mathbf{P}(S_a \cap S_b \ne \emptyset) \ge \frac{ab}{m} - \big(\frac{ab}{m}\big)^2. \qquad (4)$$

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

$$\mathbf{P}(H \ge \lambda + t) \le \exp\big\{-\frac{t^2}{2(\lambda + t/3)}\big\}, \qquad \mathbf{P}(H \le \lambda - t) \le \exp\big\{-\frac{t^2}{2\lambda}\big\}. \qquad (5)$$

Note that inequalities (3, 4) remain valid if one of the sets S_a or S_b is nonrandom.

Proof of Lemma 1. Inequalities (3) are shown in [17]. Inequalities (4) are simple consequences of (3). Exponential inequalities for hypergeometric probabilities (5) can be derived from the corresponding inequalities for binomial probabilities, see [13]. Their proof can be found, e.g., in [15]. □

Lemma 2. *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 $0 < s \le r < t \le b$ satisfying $r \le a \wedge b$, we have*

$$\mathbf{P}\Big(|S_b \cap \mathcal{S}_d| \ge t \,\Big|\, |S_b \cap \mathcal{S}_a| \ge s\Big) \le \mathbf{P}(H_{b-r,d-a,m-a} \ge t - r) + \frac{\mathbf{P}(H_{a,b,m} > r)}{\mathbf{P}(H_{a,b,m} \ge s)}. \qquad (6)$$

Assume, in addition, that $bd \le m$ and $2ab < m$. Then, for $t \ge 10$, we have

$$\mathbf{P}\Big(|S_b \cap \mathcal{S}_d| \ge 2t \,\Big|\, S_b \cap \mathcal{S}_a \ne \emptyset\Big) \le e^{-t}(1 + 2m/(ab)). \qquad (7)$$

Proof of Lemma 2. Let us prove (6). Introduce the events $\mathbb{B} = \{|S_b \cap \mathcal{S}_d| \ge t\}$, $\mathbb{A} = \{|S_b \cap \mathcal{S}_a| \ge s\}$ and write the left-hand side of (6) in the form $\mathbf{P}(\mathbb{B} \cap \mathbb{A})/\mathbf{P}(\mathbb{A})$. Denote $p_j = \mathbf{P}(H_{b-j,d-a,m-a} \ge$

$t-j$). Let $\sum_j$ denote the sum over subsets $\mathcal{A}_j \subset \mathcal{S}_a$ of size $|\mathcal{A}_j| = j$. We have

$$\begin{aligned}
\mathbf{P}(\mathbb{B} \cap \mathbb{A}) &= \sum_{s \le j \le a \wedge b} \sum_j \mathbf{P}\big(\mathbb{B} \cap \{S_b \cap \mathcal{S}_a = \mathcal{A}_j\}\big) \\
&= \sum_{s \le j \le a \wedge b} \sum_j \mathbf{P}\big(\mathbb{B} | S_b \cap \mathcal{S}_a = \mathcal{A}_j\big) \mathbf{P}(S_b \cap \mathcal{S}_a = \mathcal{A}_j) \\
&= \sum_{s \le j \le a \wedge b} p_j \sum_j \mathbf{P}(S_b \cap \mathcal{S}_a = \mathcal{A}_j) \\
&= \sum_{s \le j \le a \wedge b} p_j \mathbf{P}(H_{a,b,m} = j) \\
&\le \mathbf{P}(s \le H_{a,b,m} \le r) p_r + \mathbf{P}(H_{a,b,m} > r). \qquad (8)
\end{aligned}$$

In the last step, we estimate $p_j \le 1$ for $j > r$ and invoke the simple inequalities $p_j \le p_r$ for $1 \le j \le r-1$. (8) implies (6).
Inequality (7) follows from (6) and the inequalities of Lemma 1. Let $r = \lfloor t \rfloor$ and write $H_1 = H_{b-r,d-a,m-a}$, $H_2 = H_{a,b,m}$. The simple inequalities

$$\lambda_1 := \mathbf{E}H_1 = (b-r)(d-a)/(m-a) \le bd/m \le 1 \qquad \text{and} \qquad \lambda_2 = \mathbf{E}H_2 = ab/m \le 1 \qquad (9)$$

imply (see (5))

$$\mathbf{P}(H_i \ge t) \le \exp\{-\frac{3}{2}\frac{(t-\lambda_i)^2}{2\lambda_i + t}\} \le \exp\{-\frac{3}{2}\frac{(t-1)^2}{2+t}\} \le e^{-t}. \qquad (10)$$

In the last step, we used the assumption $t \ge 10$. The inequality $2ab/m \le 1$, combined with (4), shows that $\mathbf{P}(H_2 \ge 1) \ge ab/(2m)$. Invoking this inequality and (10) in (6), we obtain (7). $\square$

3.2. Given a finite measure Q on $[0, +\infty)$ with $\int_{[0,+\infty)} tQ(dt) < \infty$, let $\mathcal{X}_Q$ and $\mathcal{X}_Q(y)$, $y \in [0, +\infty)$, be multi-type Poisson branching processes with type space $[0, +\infty)$, defined in the same way as $\mathcal{X}_{\tilde{P}}$ and $\mathcal{X}_{\tilde{P}}(y)$ in Section 2 above, but with respect to a more general measure Q. Let $|\mathcal{X}_Q(y)|$ denote the total progeny of the process starting at a particle of type $y \in [0, +\infty)$. Denote $\rho^{(k)}(Q, y) = \mathbf{P}(|\mathcal{X}_Q(y)| \ge k)$ and $\rho(Q, y) = \mathbf{P}(|\mathcal{X}_Q(y)| = \infty)$. Therefore, $\rho(Q, y)$ is the survival probability of the process $\mathcal{X}_Q(y)$. Denote $\rho(Q) = Q^{-1}([0, +\infty)) \times \int_{[0,+\infty)} \rho(Q, t) Q(dt)$ and $\rho^{(k)}(Q) = Q^{-1}([0, +\infty)) \times \int_{[0,+\infty)} \rho^{(k)}(Q, t) Q(dt)$. Given $\Delta > 0$, let $\tilde{P}_\Delta$ denote the truncation of $\tilde{P}$ at level Δ, that is, for every $0 < a < b$, we have that $\tilde{P}_\Delta([a, b]) = \tilde{P}([a, b] \cap (0, \Delta))$ and $\tilde{P}_\Delta$ assigns the mass $\tilde{P}(\{0\}) + \tilde{P}([\Delta, +\infty))$ to the point 0.
In Lemma 3 we collect some facts from [5] about survival probabilities of Poisson branching processes that are relevant to the random graph model considered in Theorem 1. For a general theory and proofs, we refer to Sect. 5 of [5].

Lemma 3. *(i) Given $\varepsilon \ge 0$, $\mathcal{B} = \{y_1, y_2, \dots, y_M\} \subset [0, +\infty)$, and a measure Q defined on $\mathcal{B}$, let $Q_{+\varepsilon}$ and $Q_{-\varepsilon}$ denote measures on $\mathcal{B}_{+\varepsilon} = \{y_1(1+\varepsilon), \dots, y_M(1+\varepsilon)\}$ and $\mathcal{B}_{-\varepsilon} = \{y_1(1-\varepsilon), \dots, y_M(1-\varepsilon)\}$, respectively, defined by $Q_{+\varepsilon}(y_i(1+\varepsilon)) = Q_{-\varepsilon}(y_i(1-\varepsilon)) = Q(y_i)$, $1 \le i \le M$. We have*

$$\forall \varepsilon \ge 0 \qquad \lim_{k \to \infty} \rho^{(k)}(Q_{+\varepsilon}) = \rho(Q_{+\varepsilon}), \qquad \lim_{k \to \infty} \rho^{(k)}(Q_{-\varepsilon}) = \rho(Q_{-\varepsilon}), \qquad (11)$$

$$\lim_{\varepsilon \downarrow 0} \rho(Q_{+\varepsilon}) = \lim_{\varepsilon \downarrow 0} \rho(Q_{-\varepsilon}) = \rho(Q). \qquad (12)$$

(ii) For all $\Delta > 0$ and $k \geq 1$, we have $\rho^{(k)}(\tilde{P}_\Delta) \leq \rho^{(k)}(\tilde{P})$. We also have

$$\lim_{\Delta \uparrow \infty} \rho(\tilde{P}_\Delta) = \rho(\tilde{P}), \qquad \lim_{k \to \infty} \rho^{(k)}(\tilde{P}) = \rho(\tilde{P}). \tag{13}$$

(iii) Fix $\Delta > 0$ such that $\tilde{P}(\{\Delta\}) = 0$. Given $\varepsilon \in (0, \Delta)$, let $0 = y_1 < y_2 < \cdots < y_M = \Delta$ be a partition of $[0, \Delta]$ such that $\mathbf{P}(\tilde{Y} = y_i) = 0$ and $y_i - y_{i-1} < \varepsilon$ for $i > 1$. Let Q_ε^- (respectively Q_ε^+) denote the probability distribution on the set $\mathcal{B}_M = \{y_1, \ldots, y_M\}$ that assigns the mass $q_i = \tilde{P}_\Delta([y_i, y_{i+1}))$ to y_i (respectively y_{i+1}), $1 \leq i < M$, and $Q_\varepsilon^-(y_M) = Q_\varepsilon^+(y_1) = 0$. Consider a sequence $\varepsilon_M \downarrow 0$ and the corresponding sequences of partitions $\{\mathcal{B}_M\}$ and measures $\{Q_{\varepsilon_M}^-\}$, $\{Q_{\varepsilon_M}^+\}$. We have

$$\lim_{\varepsilon_M \downarrow 0} \rho(Q_{\varepsilon_M}^-) = \lim_{\varepsilon_M \downarrow 0} \rho(Q_{\varepsilon_M}^+) = \rho(\tilde{P}_\Delta). \tag{14}$$

3.3. In the proof of Theorem 1, we use some ideas and techniques developed in [5] and [2]. Let us mention that the vertices of a random intersection graph resemble the particles of a related Poisson branching process and the sizes of their random sets resemble the types of the particles. We start with Lemma 4, where we establish the result in the case where the sizes of the sets $S(v_1), \ldots, S(v_n)$ (defining an intersection graph) are nonrandom and may attain a finite number of different values only. The proof of the general result (Theorem 1) is postponed up until the end of the section.

Given integers n, m and vector $\overline{s} = (s_1, \ldots, s_n)$ with integer coordinates, let $S(v_1), \ldots, S(v_n)$ be independent random subsets of $W_m = \{w_1, \ldots, w_m\}$ such that, for every $1 \leq i \leq n$, the subset $S(v_i)$ is uniformly distributed in the class of all subsets of W_m of size s_i (we assume that $0 \leq s_i \leq m$). Let $G_{\overline{s}}(n, m)$ denote the random intersection graph on the vertex set $V_n = \{v_1, \ldots, v_n\}$ defined by the random sets $S(v_1), \ldots, S(v_n)$.

Lemma 4. *Let Q be a finite measure defined on $\mathcal{B} = \{y_1, y_2, \ldots, y_M\} \subset [0, +\infty)$. Let $\{m_n\}$ be a sequence of integers, and $\{\overline{s}_n = (s_{n1}, \ldots, s_{nn})\}$ be a sequence of vectors with integer coordinates $s_{ni} \in \{d_{nt} = \lceil y_t \sqrt{m_n/n} \rceil, 1 \leq t \leq M\}$ for $1 \leq i \leq n$. Let n_t denote the number of coordinates of $\overline{s}_n$ attaining the value d_{nt}.*

Assume that, for some integer n_0 and positive sequences $\{\varepsilon_n\}$, $\{\varepsilon_n'\} \subset [0, 1]$ converging to zero, we have, for every $n > n_0$,

$$\begin{aligned} &(i) \qquad \max\{ |Q(y_t) - (n_t/n)| : 1 \leq t \leq M \} \leq \varepsilon_n, \\ &(ii) \qquad n \ln^2 n \leq \varepsilon_n' m_n. \end{aligned}$$

Then there exists a sequence $\{\varepsilon_n^*\}_{n \geq 1}$ converging to zero (depending only on $\mathcal{B}$, $\{m_n\}$, $\{\varepsilon_n\}$, and $\{\varepsilon_n'\}$) such that, for $n > n_0$, we have

$$\mathbf{P}\big(|N_1(G_{\overline{s}_n}(n, m_n) - n\rho(Q)| > \varepsilon_n^* n \big) < \varepsilon_n^*. \tag{15}$$

Throughout the proofs of Lemma 4 and Theorem 1, we call the elements of $W = W_m$ attributes. $S(v_i)$ is called the attribute set of v_i. Given a sequence of random variables $\{X_n\}$ and a number $a \in R$, we write $X_n \geq na - o_P(n)$ if there exists another sequence of random variables $\{\xi_n\}$ such that $X_n \geq na - \xi_n$ for every n and, for each $\varepsilon > 0$, we have $\mathbf{P}(\xi_n > \varepsilon n) \to 0$ as $n \to \infty$. The notation $X_n \leq na + o_P(n)$ is defined in a similar way.

Proof of Lemma 4. Before the proof, we introduce some more notation. By $c_{\mathcal{B}}, c_{\mathcal{B}}'$, etc. we denote positive constants which depend on $\mathcal{B}$ only. Similarly, $c_{\mathcal{B},\tau}$ depend on $\mathcal{B}$ and a parameter τ only. Denote $q_t = Q(y_t)$, $t \in [M]$. In what follows, we drop the subscript n and write $m = m_n$, $\varepsilon = \varepsilon_n$,

$d_j = d_{nj}$, $\overline{s} = \overline{s}_n$, $V = V_n$, $W = W_m$, $G = G_{\overline{s}} = G_{\overline{s}_n}(n, m)$. We say that a vertex $v_i \in V$ is of type t if the size s_i of its attribute set $S(v_i)$ equals d_t. Introduce independent random sets $S_1^*, \dots, S_M^*$ such that S_t^* is uniformly distributed in the class of subsets of W of size d_t, $t \in [M]$. We assume that these random sets are independent of $S(v_1), \dots, S(v_n)$.
Given $0 < \delta < 1$, let $\tilde{\mathcal{Y}}_{-\delta}$ and $\tilde{\mathcal{Y}}_{+\delta}$ denote multi-type G–W processes with the type space $[M]$ such that the number of u-type children of a particle of type t has Poisson distribution with the mean $(1-\delta)^2 q_u y_t y_u$ and $(1+\delta)^2 q_u y_t y_u$, respectively. Write

$$\rho_{-\delta}^{(k)}(t) = \mathbf{P}(|\tilde{\mathcal{Y}}_{-\delta}(t)| \geq k), \qquad \rho_{-\delta}(t) = \mathbf{P}(|\tilde{\mathcal{Y}}_{-\delta}(t)| = \infty),$$
$$\rho_{-\delta}^{(k)} = \sum_{t \in [M]} \rho_{-\delta}^{(k)}(t) Q(t), \qquad \rho_{-\delta} = \sum_{t \in [M]} \rho_{-\delta}(t) Q(t).$$

Here $\tilde{\mathcal{Y}}_{-\delta}(t)$ denotes the process $\tilde{\mathcal{Y}}$ starting at a particle of type t. For $\tilde{\mathcal{Y}}_{+\delta}$, the corresponding quantities $\rho_{+\delta}^{(k)}(t)$, $\rho_{+\delta}(t)$, $\rho_{+\delta}^{(k)}$, $\rho_{+\delta}$ are defined in the same way.
We assume without loss of generality that $0 < y_1 < y_2 < \cdots < y_M$. Indeed, if $y_1 = 0$, then all vertices of type $t = 1$ are isolated, and $N_1(G)$ belongs to the subgraph, say G', of G spanned by the vertices of types $t \in \{2, \dots, M\}$. If, in addition, $q_1 = 1$, then the order of G' is at most $n\varepsilon_n = o(1)$, and (15) becomes trivial. In the case where $y_1 = 0$ and $q_1 < 1$, it suffices to establish (15) for the random intersection graph G' and measure Q' defined on $\mathcal{B}' = \{y_2, \dots, y_M\}$ by $Q'(y_t) = q_t$, $2 \leq t \leq M$.
Let us prove (15). We shall show the following two bounds, which together imply (15):

$$N_1(G) \leq n\rho(Q) + o_P(n), \tag{16}$$
$$N_1(G) \geq n\rho(Q) - o_P(n). \tag{17}$$

First, we prove the upper bound (16). Given an integer function ω such that $\omega(n) \to \infty$ and $\omega(n) = o(n)$, we call a component of G big (ω-big) if it has at least $\omega(n)$ vertices. Let $B = B_\omega$ be the union of vertices of big components. We shall show that, for any such $\omega(\cdot)$,

$$|B| = \rho(Q)n + o_P(n) \tag{18}$$

as $n \to \infty$. This identity and the obvious inequality $N_1(G) \leq \max\{\omega(n), |B|\}$ imply (16).
The proof of (18) consists of several steps. We start with showing that

$$\mathbf{E}|B|/n \to \rho(Q). \tag{19}$$

Proof of (19). Given $v \in V$, we evaluate the probability $\mathbf{P}(v \in B)$. For this purpose, we explore the connected component containing v as follows. Color v white and add to the list L (now L consists of a single white vertex v). Then proceed recursively: pick up the oldest white vertex from the list L, color it black, reveal its neighbors that have not been discovered earlier (i.e., the neighbors from outside L), color them white and add these neighbors to the list. Stop when we have found at least $\omega(n)$ vertices (so $v \in B$) or when there are no white vertices left in L (so we have found the entire component, and $v \notin B$).
Let $L_v = \{x_1, \dots, x_k\}$ denote the complete list of the discovered vertices numbered according to the time of discovery. Here, $x_1 = v$ and $k \leq \omega(n)$. Let T_v be the tree of black vertices and their children. T_v has root v, contains all vertices of L_v and all children of the last black vertex (some of them might be not included in L_v because of the size limit $|L_v| \leq \omega(n)$). Introduce the sets $D_i = \cup_{1 \leq j \leq i} S(x_j)$, $D_0 = \emptyset$, and $S'(x_i) = S(x_i) \setminus D_{i-1}$. We label x_i if the event $\overline{\mathcal{A}}_i$, complement to $\mathcal{A}_i = \{|S'(x_i)| \geq |S(x_i)| - 8 \ln n\}$, occurs. Let T_v^* denote the tree obtained from T_v by deleting

all descendants of labeled vertices. Given i, let i_* ($i_* < i$) denote the index of vertex that has discovered x_i, and let $n_{i\,u}$ denote the number of the vertices of type u that remain outside L after $x_1, \dots, x_{i-1}$ have discovered their neighbors (and added them to the list L). Let $X_{i\,u}$ denote the number of the neighbors of type u discovered by the vertex x_i (the number of u-type children of x_i). The conditional distribution of $X_{i\,u}$, given $\mathcal{D}_i = \{S(x_1), \dots, S(x_i)\}$ and $n_{i\,u}$, is binomial $Bi(n_{i\,u}, p_{i\,u})$, where

$$p_{i\,u} = 1 - \frac{\binom{m-|D_{i-1}|-|S'(x_i)|}{d_u}}{\binom{m-|D_{i-1}|}{d_u}}. \tag{20}$$

When needed, we also indicate the type, say t, of x_i and write X_{itu}, p_{itu} instead of X_{iu}, p_{iu}. Note that, on the event $\mathcal{A}_i$, we have $p_{itu} = y_t y_u n^{-1} + o(n^{-1})$. To show this, we apply (4) to $a = |S'(x_i)|$, $b = d_u$, and $W' = W \setminus D_{i-1}$ and use the inequalities

$$d_t \geq |S'(x_i)| \geq d_t - 8\ln n = d_t(1 - o(1)),$$
$$|W'| = m - |D_{i-1}| \geq m - \max_t \{d_t\}\,\omega(n) = m(1 - o(1)).$$

We obtain

$$\frac{y_t y_u}{n}\big(1 + c_{\mathcal{B}}(n^{-1} + \sqrt{\varepsilon_n'})\big) \geq p_{itu} \geq \frac{y_t y_u}{n}\big(1 - c_{\mathcal{B}}(n^{-1} + \sqrt{\varepsilon_n'})\big). \tag{21}$$

In addition, it follows from condition (i) of the lemma that

$$(q_u - \varepsilon_n)n - \omega(n) \leq n_{i\,u} \leq (q_u + \varepsilon_n)n. \tag{22}$$

It follows from (21)–(22) that, given $\delta > 0$, there exists n_δ such that, for every $n \geq n_\delta$ and every unlabeled vertex $x_i \in L_v$, we have

$$(1-\delta)y_t y_u/n \leq p_{itu} \leq (1+\delta)y_t y_u/n, \tag{23}$$
$$(1-\delta)q_u n \leq n_{iu} \leq (1+\delta)q_u n. \tag{24}$$

In view of (23)–(24), for large n, we can couple offspring numbers X_{itu} of an unlabeled x_i with the binomial random variables

$$Y_{itu}^{-} \sim Bi\big((1-\delta)nq_u, (1-\delta)y_t y_u/n\big), \tag{25}$$
$$Y_{itu}^{+} \sim Bi\big((1+\delta)nq_u, (1+\delta)y_t y_u/n\big), \tag{26}$$

so that

$$Y_{itu}^{-} \leq X_{itu} \leq Y_{itu}^{+}. \tag{27}$$

In addition, we can couple the exploration process (until the first labeled vertex) with two Galton–Watson processes, say $\mathcal{Y}^-$ and $\mathcal{Y}^+$, that have binomial offspring distributions (25) and (26) as follows. In order to obtain the Galton–Watson tree of $\mathcal{Y}^-$, we perform the Breath First Search in T_v^* starting at the root v and delete discovered vertices (together with subtrees of their descendants) independently at random so that offspring numbers of vertices of the resulting pruned tree would follow binomial distributions (25). In addition, we replace white leaves and labeled leaves of the pruned three by independent Galton–Watson trees with offspring distributions (25). The process $\mathcal{Y}^+$ is defined similarly, but now we insert independent Galton–Watson trees (generated using offspring distributions (26)) instead of deleting subtrees of T_v^*. Since (21) fails on the event $\overline{\mathcal{A}}_i$ complement to $\mathcal{A}_i$, inequalities (27) are valid for $x_1, x_2, \dots, x_{i-1}$ until the first occurrence of $\overline{\mathcal{A}}_i$ (until the first labeled vertex x_i). Let $\mathcal{A}_{i*} = \{k \geq i\} \cap \big(\cap_{1 \leq j \leq i-1} \mathcal{A}_j\big) \cap \overline{\mathcal{A}}_i$ denote the event that x_i is the first vertex that receives the label. On the event $\overline{\mathcal{A}}_*$, complement to $\mathcal{A}_* = \cup_{1\le i<\omega(n)}\mathcal{A}_{i*}$, the coupling (27) is valid and implies the relations

$$\Big(\{|\mathcal{Y}^-| \ge \omega(n)\} \cap \overline{\mathcal{A}}_*\Big) \subset \Big(\{|L_v| = \omega(n)\} \cap \overline{\mathcal{A}}_*\Big) \subset \Big(\{|\mathcal{Y}^+| \ge \omega(n)\} \cap \overline{\mathcal{A}}_*\Big). \tag{28}$$

Note that $v \in B \Leftrightarrow |L_v| = \omega(n)$. Therefore, from (28) we obtain

$$\mathbf{P}\big(|\mathcal{Y}^-| \ge \omega(n)\big) - \mathbf{P}(\mathcal{A}_*) \le \mathbf{P}(v \in B) \le \mathbf{P}\big(|\mathcal{Y}^+| \ge \omega(n)|\big) + \mathbf{P}(\mathcal{A}_*). \tag{29}$$

Let us show that the probability of the event $\mathcal{A}_*$ is negligibly small,

$$\mathbf{P}(\mathcal{A}_*) = O(\omega(n)/n^3). \tag{30}$$

We write $\mathbf{P}(\mathcal{A}_*) \le \sum_{1\le i<\omega(n)}\mathbf{P}(\mathcal{A}_{i*})$ and prove that $\mathbf{P}(\mathcal{A}_{i*}) \le c'_{\mathcal{B}} n^{-3}$ for $i < \omega(n)$. In order to prove the latter inequality, we establish the same bound for the conditional probability

$$p^* = \mathbf{P}\big(\overline{\mathcal{A}}_i \big| \{k \ge i\}, \mathcal{A}_1, \dots, \mathcal{A}_{i-1}, \mathcal{D}_{i-1}\big).$$

Denote $S_a := S'(x_{i_*})$, $S_d := D_{i-1} \backslash D_{i^*-1}$, and $W' := W \backslash D_{i_*-1}$. On the event $\mathcal{A}_{i_*} \subset \cap_{1\le j\le i-1}\mathcal{A}_j$, we have $a = |S'(x_{i_*})| \ge |S(x_{i_*})| - 8\ln n = |S(i_*)|(1 - o(1))$. In addition, we always have $d \le \omega(n)\max_t d_t$, and, therefore, $|W'| \ge m - d = m(1 - o(1))$. Choosing $S_b = S(x_i)$, we obtain from (7) the bound $p^* \le c'_{\mathcal{B}} n^{-3}$, thus, completing the proof of (30).

Next, we replace the probabilities $\mathbf{P}(|\mathcal{Y}^-| \ge \omega(n))$ and $\mathbf{P}(|\mathcal{Y}^+| \ge \omega(n))$ in (29) by $\rho_{-\delta}^{(\omega(n))}(t_v)$ and $\rho_{+\delta}^{(\omega(n))}(t_v)$, where t_v denotes the type of vertex v. The total variation distance between the binomial distribution $Bi(r, p)$ and the Poisson distribution with the same mean is at most p; see, e.g., inequality (1.23) in [1]. Therefore, the error of this replacement is $O(\omega(n)/n)$. Now, from (29) and (30) we obtain (uniformly in $v \in V$) the inequalities

$$\rho_{-\delta}^{(\omega(n))}(t_v) - O(\omega(n)/n) \le \mathbf{P}(v \in B) \le \rho_{+\delta}^{(\omega(n))}(t_v) + O(\omega(n)/n). \tag{31}$$

Invoking (31) in the sum $\mathbf{E}|B| = \sum_{v\in V}\mathbf{P}(v \in B)$ and using condition (i) of the lemma, we get

$$\rho_{-\delta}^{(\omega(n))} - M\varepsilon_n - O(\omega(n)/n) \le \mathbf{E}|B|/n \le \rho_{+\delta}^{(\omega(n))} + M\varepsilon_n + O(\omega(n)/n). \tag{32}$$

Finally, letting $n \to \infty$, we conclude from (11) and (32) that

$$\rho_{-\delta} \le \liminf_n \mathbf{E}|B|/n \le \limsup_n \mathbf{E}|B|/n \le \rho_{+\delta}. \tag{33}$$

Since $\rho_{-\delta}, \rho_{+\delta} \to \rho(Q)$ as $\delta \downarrow 0$ (see (12)), the latter inequalities imply (19).

Let $\omega(n) = \lceil \ln n \rceil$. For this particular $\omega(\cdot)$, we prove that

$$\mathbf{Var}|B| = o(n^2). \tag{34}$$

Proof of (34). In the proof, we use the identity

$$|B|^2 - |B| = 2\sum_{\{v,w\}\subset V} \mathbb{I}_{\{v,w\in B\}} \tag{35}$$

and exploit the fact that events $\{v \in B\}$ and $\{w \in B\}$ are asymptotically independent as $n \to \infty$. This fact is not surprising as each of the events $\mathcal{L}_v := \{|L_v| = \omega(n)\}$ and $\mathcal{L}_w := \{|L_w| = \omega(n)\}$ (equivalent to $\{v \in B\}$ and $\{w \in B\}$) refers to at most $O(\ln n)$ vertices. We first generate the list

$L_v = \{x_1, \dots, x_k\}$. Then we generate the list $L_w = \{z_1, \dots, z_{k'}\}$ (here $z_1 = w$ and $k' \le \omega(n)$). Denote $D = \cup_{x \in L_v} S(x)$ and introduce the events $\mathcal{H}_i = \{S(z_i) \cap D = \emptyset\}$ and $\mathcal{H} = \cap_{1 \le i \le k'} \mathcal{H}_i$. We shall show below that

$$\mathbf{P}(\mathcal{H}) = 1 - O(\omega^2(n)/n). \tag{36}$$

On the event $\mathcal{H}$, the exploration process starting at w does not encounter any vertex of L_v. Conditionally, given D and $\mathcal{H}$, we can couple this (starting at w) exploration process with two G–W processes as in the proof of (19) above but with the attribute set $W \setminus D$ instead of W and the vertex set $V \setminus L_v$ instead of V. Since

$$|D| \le d_M \omega(n) = o(m), \tag{37}$$

the size $|W \setminus D| = m(1 - o(1))$, while the number of u-type vertices in $V \setminus L_v$ is at least $n_u - \omega(n) = n_u(1 - o(1))$ for every $u \in [M]$. Therefore, (31) extends to the conditional probability $p_w(D) := \mathbf{P}\big(\mathcal{L}_w \big| D, \mathcal{H}, \mathcal{L}_v\big)$. Uniformly in D, we have

$$\rho_{-\delta}^{(\omega(n))}(t_w) - O(\omega(n)/n) \le p_w(D) \le \rho_{+\delta}^{(\omega(n))}(t_w) + O(\omega(n)/n). \tag{38}$$

Here, t_w denotes the type of w. Combining (38) with the total probability formulae, where sums are taken over all possible values Δ of the random set D,

$$\begin{aligned} \mathbf{P}\big(\mathcal{L}_w \cap \mathcal{L}_v \cap \mathcal{H}\big) &= \sum_{\Delta} p_w(\Delta)\, \mathbf{P}\big(\{D = \Delta\} \cap \mathcal{H} \cap \mathcal{L}_v\big), \\ \mathbf{P}\big(\mathcal{H} \cap \mathcal{L}_v\big) &= \sum_{\Delta} \mathbf{P}\big(\{D = \Delta\} \cap \mathcal{H} \cap \mathcal{L}_v\big), \end{aligned}$$

we obtain

$$\mathbf{P}(\mathcal{L}_v \cap \mathcal{H})\, \rho_{-\delta}^{(\omega(n))}(t_w) - o(1) \le \mathbf{P}\big(\mathcal{L}_w \cap \mathcal{L}_v \cap \mathcal{H}\big) \le \mathbf{P}\big(\mathcal{L}_v \cap \mathcal{H}\big)\, \rho_{+\delta}^{(\omega(n))}(t_w) + o(1).$$

In view of (36), we can replace $\mathbf{P}(\mathcal{L}_v \cap \mathcal{H})$ by $\mathbf{P}(\mathcal{L}_v)$ and $\mathbf{P}\big(\mathcal{L}_w \cap \mathcal{L}_v \cap \mathcal{H}\big)$ by $\mathbf{P}(\mathcal{L}_w \cap \mathcal{L}_v)$. Then, invoking (31), we obtain (uniformly in v, w)

$$\rho_{-\delta}^{(\omega(n))}(t_v)\ \rho_{-\delta}^{(\omega(n))}(t_w) - o(1) \le \mathbf{P}\big(\mathcal{L}_w \cap \mathcal{L}_v\big) \le \rho_{+\delta}^{(\omega(n))}(t_v)\ \rho_{+\delta}^{(\omega(n))}(t_w) + o(1). \tag{39}$$

It follows from (35) and (39) and the obvious identity $\mathbf{E}\mathbb{I}_{\{v,w \in B\}} = \mathbf{P}(\mathcal{L}_v \cap \mathcal{L}_v)$ that

$$\Big(\sum_{v \in V} \rho_{-\delta}^{(\omega(n))}(t_v)\Big)^2 - o(n^2) \le \mathbf{E}(|B|^2 - |B|) \le \Big(\sum_{v \in V} \rho_{+\delta}^{(\omega(n))}(t_v)\Big)^2 + o(n^2). \tag{40}$$

Letting first $n \to \infty$ and then $\delta \downarrow 0$, we obtain, by the same argument as in (32–33) above, that $n^{-2}\mathbf{E}(|B|^2 - |B|) = (\rho(Q))^2 + o(1)$. The latter identity, combined with (19), shows (34).
It remains to prove (36). For this purpose, we write the complement event $\overline{\mathcal{H}}$ in the form $\overline{\mathcal{H}} = \cup_{1 \le i \le \omega(n)} \mathcal{H}_i^*$, where $\mathcal{H}_1^* = \overline{\mathcal{H}}_1$ and $\mathcal{H}_i = (\cap_{1 \le j < i} \mathcal{H}_j) \cap \overline{\mathcal{H}}_i \cap \{k' \ge i\}$, $i \ge 2$, and show that

$$\mathbf{P}(\overline{\mathcal{H}}_i^*) \le c''_{\mathcal{B}} \omega(n) n^{-1}, \qquad 1 \le i \le \omega(n). \tag{41}$$

Let $i = 1$. Given D, we apply (4) to $S_a = S(z_1)$, $S_b = D$, and $W' = W$. Using the inequalities $|S(z_1)| \le d_M$ and (37), we obtain (41). For $i = 2$, we expand, by the total probability formula,

$$\mathbf{P}(\mathcal{H}_2^*) = \sum_{h \ge 1} p_h\, \mathbf{P}(\mathcal{H}_1 \cap \{|S(z_2) \cap S(z_1)| = h\}), \tag{42}$$

where $p_h := \mathbf{P}\big(S(z_2) \cap D \neq \emptyset \big| \mathcal{H}_1 \cap \{|S(z_2) \cap S(z_1)| = h\}\big)$. Choosing $S_a = S(z_2) \setminus S(z_1)$, $S_b = D$, and $W' = W \setminus S(z_1)$, we obtain from (4) that $p_h \leq c''_B \omega(n) n^{-1}$ uniformly in h. This bound, combined with (42), shows (41) for $i = 2$. For $i = 3, 4, \ldots, \omega(n)$, the proof of (41) is much the same.

Proof of (18). For $\omega(n) = \lceil \ln n \rceil$, (18) follows from (19) and (34). We extend (18) to any B' defined using another function $\omega'(\cdot)$. To this aim, we apply (19) to $\max\{\omega(n), \omega'(n)\}$ and $\min\{\omega(n), \omega'(n)\}$ and obtain the bound for the expected size of the symmetric difference $\mathbf{E}|B \triangle B'| = o(n)$. This bound implies that both $|B|/n$ and $|B'|/n$ have the same limit in probability.

Let us prove the lower bound (17). To this aim, we show that, for every $0 < \tau < 1$,

$$N_1(G) \geq n\rho_{-\tau} - o_P(n). \tag{43}$$

Indeed, this inequality implies (17), because $\rho_{-\tau} \to \rho(Q)$ as $\tau \downarrow 0$, see (12).

We fix $\tau \in (0, 1)$ and prove (43). Let G_0 be the intersection graph on the vertex set V defined by the collection of sets $\mathbb{S}_0 := \{S_0(v), v \in V\}$, where each $S_0(v)$ is a subset of $S(v)$ of size $(1 - \tau)|S(v)|$ (round if necessary) drawn uniformly at random. Let G_* denote the intersection graph defined by the collection of sets $\{S_*(v) = S(v) \setminus S_0(v), v \in V\}$. Note that G_0 and G_* are spanning subgraphs of G. Fix the function $\omega(n) = \lceil n^{2/3} \rceil$ and let B_0 denote the union of ω-big vertices of G_0. It follows from (18) that

$$|B_0| = n\rho_{-\tau} + o_P(n). \tag{44}$$

We show that the event $\mathcal{G} = \{$all vertices of B_0 lie in at most one component of $G\}$ has the probability

$$\mathbf{P}(\mathcal{G}) = 1 - o(1) \tag{45}$$

as $n \to \infty$. From (44) and (45) we obtain (43). It remains to prove (45).

Proof of (45). Given $\mathbb{S}_0$, let $C_1, C_2, \ldots$ denote the connected components of G_0 of order at least $\omega(n)$, and let $V'_1, V'_2, \ldots$ denote their vertex sets (we assume that there are at least two such components since otherwise $\mathcal{G}$ is automatically satisfied). In the proof of (45), we use the fact that the random sets $W_*(V'_i) := \cup_{v \in V'_i} S_*(v)$, $i = 1, 2, \ldots$, are large enough so that they intersect with high probability. Such intersections produce links between V'_i and V'_j in G_* that imply the event $\mathcal{G}$. Introduce the events $\mathcal{A}_{ij} = \{$there is a link in G_* between V'_i and $V'_j\}$, $i \neq j$. Note that $\mathcal{A}_{ij} = \{W_*(V'_i) \cap W_*(V'_j) \neq \emptyset\}$ and

$$\begin{aligned} \mathbf{P}(\mathcal{G}|\mathbb{S}_0) \quad &\geq \mathbf{P}(\cap_{i \neq j} \mathcal{A}_{ij} \big| \mathbb{S}_0) \\ &\geq 1 - \sum_{i \neq j} \mathbf{P}(W_*(V'_i) \cap W_*(V'_j) = \emptyset \big| \mathbb{S}_0). \end{aligned} \tag{46}$$

We show below that, uniformly in $\mathbb{S}_0$, we have, for every $V'_i \neq V'_j$,

$$\mathbf{P}(W_*(V'_i) \cap W_*(V'_j) = \emptyset \big| \mathbb{S}_0) \leq c_{\mathcal{B},\tau} n^{-5}. \tag{47}$$

Invoking this bound in (46) and using the observation that there can be at most $n^{2/3}$ pairs $C_i \neq C_j$ of components of order at least $\lceil n^{2/3} \rceil$, we obtain (45).

Proof of (47). We first show that $W_*(V'_i)$ is large with high probability,

$$\mathbf{P}(|W_*(V'_i)| \geq 2^{-1} \tau d_1 \omega(n) \,\big|\, \mathbb{S}_0) \geq 1 - c_{\mathcal{B},\tau} n^{-5}. \tag{48}$$

Recall that $d_1 = \min_t d_t$ and $d_M = \max_t d_t$. Write $V'_i = \{w_1, \ldots, w_k\}$. (Note that, conditionally, given $\mathbb{S}_0$, the number $k \in [\omega(n), n]$ is nonrandom.) Introduce the sets $D_{[0]} = \emptyset$ and, for $h \geq 1$, put $D_{[h]} = \cup_{1\le j\le h} S_*(w_j)$ and $S'_*(w_h) = S_*(w_h) \setminus D_{[h-1]}$. (48) follows from the simple identity $|W_*(V'_i)| = \sum_{1\le h\le k} |S'_*(w_h)|$ and from the following inequalities which hold uniformly in $h \le k$ and $D_{[h-1]}$:

$$\mathbf{P}(|S'_*(w_h)| < 2^{-1}\tau d_1 \,\big|\, D_{[h-1]}, \mathbb{S}_0) \le c_{\mathcal{B}\tau} n^{-6}. \tag{49}$$

To show this bound, we write the probability in the form $\mathbf{P}(H_{a,b,m'} > y)$, where $y = a - 2^{-1}\tau d_1$ and $a = |S_*(w_h)|$, $b = |D_{[h-1]} \setminus S_0(w_h)|$, $m' = m - |S_0(w_h)|$ (note that, conditionally, given $\mathbb{S}_0$, the set $S_0(w_h)$ is nonrandom). The inequalities $\tau d_1 \le a \le \tau d_M$, $b \le (\omega(n) - 1)\tau d_M$ and $m' \ge m - d_M = m(1 - o(1))$ imply

$$\mathbf{E}H_{a,b,m'} \le \tau^2 y_M^2 O(\omega(n)/n) = o(1) \qquad \text{and} \qquad y \ge 2^{-1}\tau d_1 = 2^{-1}\tau y_1 (\varepsilon'_n)^{-1/2} \ln n.$$

Now (5) shows (49). We have arrived at (48).

Next, we show that $W_*(V'_i)$ and $W_*(V'_j)$ intersect with high probability. By the conditional independence, given $W_*(V'_i)$ and $\mathbb{S}_0$, of the random variables $H_v := |W_*(V_i) \cap S_*(v)|$, $v \in V'_j$, we have

$$\mathbf{P}\big(W_*(V'_i) \cap W_*(V'_j) = \emptyset \big| W_*(V'_i), \mathbb{S}_0\big) = \prod_{v\in V'_j} \mathbf{P}(H_v = 0 \big| W_*(V'_i), \mathbb{S}_0). \tag{50}$$

The second inequality of (5) implies $\mathbf{P}(H_v = 0 \,\big|\, W_*(V'_i), \mathbb{S}_0) \le \exp\{-\lambda_v/2\}$, where $\lambda_v = \mathbf{E}(H_v | W_*(V'_i), \mathbb{S}_0)$. In addition, invoking the inequalities

$$\lambda_v = |S_*(v)| \times |W_*(V'_i) \setminus S_0(v)|/(m - |S_0(v)|) \ge \tau d_1(|W_*(V'_i)| - d_M)/m$$

for $v \in V'_j$, we obtain from (50) that

$$\begin{aligned}\mathbf{P}\big(W_*(V'_i) \cap W_*(V'_j) = \emptyset \big| W_*(V'_i), \mathbb{S}_0\big) \quad &\le \exp\{-2^{-1} \sum_{v\in V'_j} \lambda_v\} \\ &\le \exp\{-2^{-1}\tau\omega(n) d_1(|W_*(V'_i)| - d_M)/m\}.\end{aligned}$$

Observe that, for $|W_*(V'_i)| \ge 2^{-1}\tau d_1\omega(n)$, the right-hand side is bounded from above by $\exp\{-c'_{\mathcal{B},\tau} n^{1/3}\}$. This bound, together with (48), implies (47). □

Proof of Theorem 1. In what follows, we write, for short, $m = m_n$ and $G_n = G(n, m, P_n)$ and denote $Z_{ni} := \sqrt{n/m}|S(v_i)|$, $1 \le i \le n$. Observe that the random variables $Z_{n1}, \dots, Z_{nn}$ are independent and have the common distribution $\tilde{P}_n$.

We start with a short outline of the proof. First, we truncate $\tilde{Z}_{n1}, \dots, \tilde{Z}_{nn}$ at level $\Delta > 0$ and approximate the truncated random variables by discrete ones. Then we establish the result for discrete random variables (see Lemma 4), extend it to bounded random variables $\tilde{Z}_{n1}\mathbb{I}_{\{\tilde{Z}_{n1}\le\Delta\}}, \dots, \tilde{Z}_{nn}\mathbb{I}_{\{\tilde{Z}_{nn}\le\Delta\}}$, and, finally, letting the truncation level $\Delta \to +\infty$, complete the proof.

Step 1 (result for truncated random variables). Given $\Delta > 0$ satisfying $\mathbf{P}(\tilde{Y} = \Delta) = 0$ and $\mathbf{P}(\tilde{Y} < \Delta) > 0$, let $G_n(\Delta)$ be the random intersection graph on the vertex set V defined by the collection $\mathbb{S}_\Delta = \{S_\Delta(v),\, v \in V\}$ of subsets of W. Here, $S_\Delta(v) := S(v)$ if $|S(v)| < \Delta\sqrt{m/n}$ and $S_\Delta(v) := \emptyset$ if $|S(v)| \ge \Delta\sqrt{m/n}$. In the first step of the proof, we show that

$$N_1(G_n(\Delta)) = n\rho(\tilde{P}_\Delta) + o_P(n). \tag{51}$$

Here, $\tilde{P}_\Delta$ denotes the distribution of $\tilde{Y}\mathbb{I}_{\{\tilde{Y}\le\Delta\}}$.

Fix $\varepsilon \in (0, \Delta)$ and an increasing sequence $0 = y_1 < y_2 < \cdots < y_M = \Delta$ such that $\mathbf{P}(\tilde{Y} = y_i) = 0$ and $y_i - y_{i-1} < \varepsilon$, $i > 1$. We say that v is of type $j \in \{1, \dots, M-1\}$ if $|S_\Delta(v)|\sqrt{n/m} \in [y_j, y_{j+1})$.

Every $v \in V$ is assigned random subsets $S^-(v)$ and $S^+(v)$ of sizes $d_j^- := \lfloor y_j \sqrt{m/n} \rfloor$ and $d_j^+ := \lceil y_{j+1}\sqrt{m/n}\rceil$ (here j is the type of v) such that

$$S^-(v) \subset S_\Delta(v) \subset S^+(v). \tag{52}$$

(We draw $S^-(v)$ uniformly at random from the class of subsets of $S_\Delta(v)$ of size d_j^-. Similarly, we draw $W \setminus S^+(v)$ uniformly at random from the class of subsets of $W \setminus S_\Delta(v)$ of size $m - d_j^+$.) Let G_n^- and G_n^+ be the intersection graphs defined by the collections of subsets $\{S^-(v),\, v \in V\}$ and $\{S^+(v),\, v \in V\}$, respectively. (52) implies $G_n^- \subset G_n(\Delta) \subset G_n^+$ (one is a subgraph of another), and, therefore, we have

$$N_1(G_n^-) \le N_1(G_n(\Delta)) \le N_1(G_n^+). \tag{53}$$

We are going to apply Lemma 4 to $N_1(G_n^-)$ and $N_1(G_n^+)$. Let Q_ε^- (respectively Q_ε^+) denote the probability distribution on the set $\mathcal{B} = \{y_1, \dots, y_M\}$ that assigns the mass $q_i = \tilde{P}_\Delta([y_i, y_{i+1}))$ to y_i (respectively y_{i+1}), $1 \le i < M$, and $Q_\varepsilon^-(y_M) = Q_\varepsilon^+(y_1) = 0$. We shall show that

$$N_1(G_n^-) = \rho(Q_\varepsilon^-) + o_P(n), \qquad N_1(G_n^+) = \rho(Q_\varepsilon^+) + o_P(n). \tag{54}$$

In view of the limits $\rho(Q_\varepsilon^-) \uparrow \rho(\tilde{P}_\Delta)$ and $\rho(Q_\varepsilon^+) \downarrow \rho(\tilde{P}_\Delta)$ as $\varepsilon \downarrow 0$, see (14), from (54) and (53) we obtain (51).

Let us prove (54). By assumption (i), the probability that given vertex is of type i,

$$q_{ni} := \mathbf{P}(|S_\Delta(v)|\sqrt{n/m} \in [y_i, y_{i+1})) = \mathbf{P}\big(\tilde{Z}_{n1}\mathbb{I}_{\{\tilde{Z}_{n1} \le \Delta\}} \in [y_i, y_{i+1})\big)$$

converges to q_i as $n \to \infty$. In particular, $\delta_n := \max_{1 \le i < M} |q_{ni} - q_i|$ converges to zero as $n \to \infty$. In addition, Chernoff's exponential bound applied to the number n_i of vertices $v \in V$ of type i shows

$$\mathbf{P}(|n_i - q_{ni}n| \ge n^{1/2}\ln^2 n) < \exp\{-3^{-1}\ln^2 n\} \le cn^{-10} \qquad \text{for} \qquad 1 \le i < M.$$

We conclude that the events $\mathcal{Z}_r := \{\max_{1 \le i < M} |q_i - n_i/n| < \delta_n + n^{-1/2}\ln^2 n \ \forall n \ge r\}$ have the probabilities $\mathbf{P}(\mathcal{Z}_r) = 1 - o(1)$ as $r \to \infty$. Finally, given $\delta > 0$ and r, Lemma 4 implies the bound $\mathbf{P}\big(|N_1(G_n^-) - \rho(Q_\varepsilon^-)n| > \delta n \big| \mathcal{Z}_r\big) = o(1)$, which yields the first part of (54). The proof of the second part is the same.

Step 2 (result for arbitrary random variables). We show two bounds

$$N_1(G_n) \ge n\rho(\tilde{P}) - o_P(n), \tag{55}$$
$$N_1(G_n) \le n\rho(\tilde{P}) + o_P(n), \tag{56}$$

which together imply the result $N_1(G_n) = n\rho(\tilde{P}) + o_P(n)$.

(55) is a simple consequence of (51). Indeed, the inequality $N_1(G_n) \ge N_1(G_n(\Delta))$ implies $N_1(G_n) \ge n\rho(\tilde{P}_\Delta) + o_P(n)$. Invoking the limit $\rho(\tilde{P}_\Delta) \uparrow \rho(\tilde{P})$ as $\Delta \uparrow +\infty$ (see (13)), we obtain (55).

Let us prove (56). Given k, G_n, and $G_n(\Delta)$, let B_k and $B_k(\Delta)$ denote the unions of vertices of components of order $\ge k$ in G_n and $G_n(\Delta)$, respectively. Let V_Δ denote the set of vertices $v \in V$ with $|S(v)| \ge \Delta\sqrt{m/n}$ (such vertices are isolated in $G_n(\Delta)$ but not necessarily in G_n). Let X_Δ denote the number of edges of G_n with one endpoint in V_Δ an another in $V \setminus V_\Delta$. The proof of (56) is based on the simple inequalities $|B_k| \le |B_k(\Delta)| + kX_\Delta + |V_\Delta|$ and $N_1(G_n) \le k + |B_k|$. These inequalities imply

$$N_1(G_n) \le k + |B_k(\Delta)| + kX_\Delta + |V_\Delta|. \tag{57}$$

In order to show (56), we choose $k, \Delta \to \infty$ such that $k = o(n)$, $|B_k(\Delta)| \le n\rho(\tilde{P}) + o_P(n)$, and $kX_\Delta, |V_\Delta| = o_P(n)$ as $n \to \infty$.
Let us construct an upper bound for X_Δ. Introduce the event $\mathcal{A}_\Delta = \{\tilde{Z}_{ni} < n/(2\Delta),\ 1 \le i \le n\}$. It follows from assumption (i) of the theorem and the inequalities

$$1 - \mathbf{P}(\mathcal{A}_\Delta) \le n\mathbf{P}(\tilde{Z}_{n1} \ge n/(2\Delta)) \le 2\Delta \mathbf{E}\tilde{Z}_{n1}\mathbb{I}_{\{\tilde{Z}_{n1} \ge n/(2\Delta)\}}$$

that, for each $\Delta > 0$, we have

$$\mathbf{P}(\mathcal{A}_\Delta) = 1 - o(1) \qquad \text{as} \qquad n \to \infty. \tag{58}$$

Given v, let Y_v denote the number of the neighbors of v in $V \setminus V_\Delta$. From the identity $X_\Delta = \sum_{v \in V} \mathbb{I}_{\{v \in V_\Delta\}} Y_v$ we obtain, by symmetry,

$$\begin{aligned} \mathbf{E}\, \mathbb{I}_{\mathcal{A}_\Delta} X_\Delta \quad &= n\mathbf{E}\, \mathbb{I}_{\mathcal{A}_\Delta} \mathbb{I}_{\{v_1 \in V_\Delta\}} Y_{v_1} \\ &= n\mathbf{E}\, \mathbb{I}_{\mathcal{A}_\Delta} \mathbb{I}_{\{v_1 \in V_\Delta\}} \sum_{v_i \in V \setminus v_1} \mathbb{I}_{\{v_i \notin V_\Delta\}} \mathbb{I}_{\{v_1 \sim v_i\}} \\ &= n(n-1)\mathbf{E}\, \mathbb{I}_{\mathcal{A}_\Delta} \mathbb{I}_{\{v_1 \in V_\Delta\}} \mathbb{I}_{\{v_2 \notin V_\Delta\}} \mathbb{I}_{\{v_1 \sim v_2\}}. \end{aligned} \tag{59}$$

Note that, on the event $\mathcal{A}_\Delta \cap \{v_2 \notin V_\Delta\}$, we have $|S(v_1)| \times |S(v_2)| \le m/2$. In this case, (4) implies

$$\mathbf{E}\big(\mathbb{I}_{\{v_1 \sim v_2\}} \big|\, |S(v_1)|, |S(v_2)|\big) = \mathbf{P}(v_1 \sim v_2 | \tilde{Z}_{n1}, \tilde{Z}_{n2}) \le 3n^{-1}\tilde{Z}_{n1}\tilde{Z}_{n2}.$$

Invoking this inequality in (59), we obtain

$$\mathbf{E}\, \mathbb{I}_{\mathcal{A}_\Delta} X_\Delta \le 3n\mathbf{E}\tilde{Z}_{n2}\, \mathbf{E}\tilde{Z}_{n1}\mathbb{I}_{\{\tilde{Z}_{n1} \ge \Delta\}} \le n\varphi^2(\Delta), \tag{60}$$

where we denote $\varphi^2(\Delta) = 3\sup_n\big\{\mathbf{E}\tilde{Z}_{n2}\, \mathbf{E}\tilde{Z}_{n1}\mathbb{I}_{\{\tilde{Z}_{n1} \ge \Delta\}}\big\}$. Observe that, by (2), $\varphi(\Delta) \to 0$ as $\Delta \to \infty$. Finally, invoking the inequality $\mathbf{P}(\mathbb{I}_{\mathcal{A}_\Delta} X_\Delta \ge n\varphi(\Delta)) \le \varphi(\Delta)$, which follows from (60) by Chebyshev's inequality, we obtain

$$\mathbf{P}\big(X_\Delta \ge n\varphi(\Delta)\big) \le \mathbf{P}(\overline{\mathcal{A}}_\Delta) + \mathbf{P}(\mathbb{I}_{\mathcal{A}_\Delta} X_\Delta \ge n\varphi(\Delta)) \le \mathbf{P}(\overline{\mathcal{A}}_\Delta) + \varphi(\Delta). \tag{61}$$

Next, we construct the upper bound for $|B_k(\Delta)|$,

$$|B_k(\Delta)| \le n\rho^{(k)}(\tilde{P}) + o_P(n). \tag{62}$$

Proceeding in the same way as in the proof of (18) and (52–54), we show that $|B_k(\Delta)| = n\rho^{(k)}(\tilde{P}_\Delta) + o_P(n)$. Then, invoking the inequality $\rho^{(k)}(\tilde{P}_\Delta) \le \rho^{(k)}(\tilde{P})$ of Lemma 3, we obtain (62).
Finally, we show that, for each $0 < \varepsilon < 1$, we have, as $n \to \infty$,

$$\mathbf{P}\big(N_1(G_n) \le n\rho(\tilde{P}) + 5n\varepsilon\big) \ge 1 - 2\varepsilon - o(1). \tag{63}$$

Fix ε. In view of (13), we can choose (sufficiently large) $k = k(\varepsilon)$ such that

$$\rho^{(k)}(\tilde{P}) \le \rho(\tilde{P}) + \varepsilon. \tag{64}$$

Given k, we choose (sufficiently large) Δ such that

$$k\varphi(\Delta) < \varepsilon \qquad \text{and} \qquad \forall n, \qquad \mathbf{P}(|V_\Delta| \ge n\varepsilon) \le \varepsilon. \tag{65}$$

Here, the last inequality, for large Δ, follows from the inequalities

$$\mathbf{P}(|V_\Delta| \ge n\varepsilon) \le (n\varepsilon)^{-1}\mathbf{E}|V_\Delta| = \varepsilon^{-1}\mathbf{P}(\tilde{Z}_{n1} \ge \Delta) \le (\varepsilon\Delta)^{-1}\max_n \mathbf{E}\tilde{Z}_{n1}.$$

Letting $n \to \infty$, we obtain from (58), (61), and (65) that $\mathbf{P}(kX_\Delta + |V_\Delta| \ge 2\varepsilon n) \le 2\varepsilon + o(1)$. From (62) and (64) we obtain $\mathbf{P}\big(|B_k(\Delta)| \le n\rho(\tilde{P}) + 2n\varepsilon\big) = 1 - o(1)$. Collecting these bounds and the simple inequality $k < \varepsilon n$ (which holds for all sufficiently large n) in (57), we obtain (63), thus completing the proof of (56). □

Proof of Remark 3. For $\rho(\tilde{P}) = 0$, the bound $N_2(G) = o_P(n)$ is an immediate consequence of Theorem 1. For $\rho(\tilde{P}) > 0$, this bound follows from the result of Theorem 1 combined with the bound $|B_k| \le \rho(\tilde{P}) + o_P(n)$ as $n, k \to \infty$. Here B_k and k are the same as in the proof of (56). □

Note added in proof. The result of Theorem 1 remains valid if we replace condition (iii) by the weaker condition $n = o(m_n)$ as $n \to \infty$. The proof remains almost the same with the only difference that instead of using Chernoff type bound (the first inequality of (5)) we now use the bound $\mathbf{P}(H_{i,j,m} \ge k) \le c(\Delta, k)(ij/m)^k$, for $k = 1, 2, \ldots$, and $j^2/m \le \Delta^2$, which is an immediate consequence of Sródka (1963) inequality, see, e.g., formula (6.70) of [18]. In particular, in the proof of Theorem 1, we consider functions $\omega(n) = O(n^\alpha)$, $0 < \alpha < 1$, and may safely assume that with a high probability intersections of random sets have at most a finite number $k = k(\alpha)$, elements (instead of $t = \Theta(\ln n)$).

References

[1] A.D. Barbour, L. Holst, and S. Janson, *Poisson approximation*, Oxford University Press, Oxford, 1992.

[2] M. Behrisch, Component evolution in random intersection graphs, The Electronic Journal of Combinatorics 14(1) (2007).

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

[4] M. Bloznelis, J. Jaworski, and K. Rybarczyk, Component evolution in a wireless sensor network, Networks 53 (2009), 19–26.

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

[6] F. Chung and L. Lu, Complex graphs and networks, CBMS regional conference series in mathematics, no 107, American Mathematical Society, Providence, Rhode Island, 2006.

[7] F. Chung and L. Lu, The volume of the giant component of a random graph with given expected degrees, SIAM J. Discrete Math. 20 (2006), 395–411.

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

[9] 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).

[10] P. Erdős and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hungar. Acad. Sci., 5 (1960), 17–61.

[11] 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.

[12] 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.

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

[14] J. Hwang and Y. Kim, Revisiting random key pre-distribution schemes for wireless sensor networks, in: Proceedings of the 2nd ACM Workshop on Security of ad hoc and Sensor Networks, SASN 2004, Washington, DC, USA, October 25 (2004), 43–52.

[15] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley, New York, 2001.

[16] S. Janson, The largest component in a subcritical random graph with a power law degree distribution. Ann. Appl. Probab. 18 (2008), 1651–1668.

[17] 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.

[18] N.L. Johnson, S. Kotz, and A. W. Kemp, *Univariate discrete distributions*, Wiley, New York, 1992.

[19] O. Kallenberg, *Foundations of modern probability*, 2nd ed., Springer, New York, 2002

[20] 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.

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

[22] B. Söderberg, General formalism for inhomogeneous random graphs, Phys. Rev. E 66 (2002), 066121.

[23] T. Sródka, On approximation of hypergeometric distribution, Zeszyty Naukowe Politechniki Lódzkiej, 10 (1963), 5–17.

[24] D. Stark, The vertex degree distribution of random intersection graphs, Random Structures Algorithms 24 (2004), 249–258.