

The largest component in an inhomogeneous random intersection graph with clustering

Mindaugas Bloznelis

Faculty of Mathematics and Informatics
Vilnius University, Vilnius, LT-03225, Lithuania

`mindaugas.bloznelis@mif.vu.lt`

Submitted: Feb 24, 2010; Accepted: Jul 18, 2010; Published: Aug 9, 2010

Mathematics Subject Classifications: 05C80, 05C82, 60J85

Abstract

Given integers n and $m = \lfloor \beta n \rfloor$ and a probability measure Q on $\{0, 1, \dots, m\}$, consider the random intersection graph on the vertex set $[n] = \{1, 2, \dots, n\}$ where $i, j \in [n]$ are declared adjacent whenever $S(i) \cap S(j) \neq \emptyset$. Here $S(1), \dots, S(n)$ denote the iid random subsets of $[m]$ with the distribution $\mathbf{P}(S(i) = A) = \binom{m}{|A|}^{-1} Q(|A|)$, $A \subset [m]$. For sparse random intersection graphs, we establish a first-order asymptotic as $n \rightarrow \infty$ for the order of the largest connected component $N_1 = n(1 - Q(0))\rho + o_P(n)$. Here ρ is the average of nonextinction probabilities of a related multitype Poisson branching process.

1 Introduction

Let Q be a probability measure on $\{0, 1, \dots, m\}$, and let $S_1, \dots, S_n$ be random subsets of a set $W = \{w_1, \dots, w_m\}$ drawn independently from the probability distribution $\mathbf{P}(S_i = A) = \binom{m}{|A|}^{-1} Q(|A|)$, $A \subset W$, for $i = 1, \dots, n$. A random intersection graph $G(n, m, Q)$ with vertex set $V = \{v_1, \dots, v_n\}$ is defined as follows. Every vertex v_i is prescribed the set $S(v_i) = S_i$, and two vertices v_i and v_j are declared adjacent (denoted $v_i \sim v_j$) whenever $S(v_i) \cap S(v_j) \neq \emptyset$. The elements of W are sometimes called attributes, and $S(v_i)$ is called the set of attributes of v_i .

Random intersection graphs $G(n, m, Q)$ with the binomial distribution $Q \sim Bi(m, p)$ were introduced in Singer-Cohen [15] and Karoński et al. [13], see also [10] and [16]. The emergence of a giant connected component in a sparse binomial random intersection graph was studied by Behrishi [2] for $m = \lfloor n^\alpha \rfloor$, $\alpha \neq 1$, and by Lagerås and Lindholm [14] for $m = \lfloor \beta n \rfloor$, where $\beta > 0$ is a constant. They have shown, in particular, that, for $\alpha \geq 1$, the largest connected component collects a fraction of all vertices whenever the average

vertex degree, say d , is larger than $1 + \varepsilon$. For $d < 1 - \varepsilon$, the order of the largest connected component is $O(\log n)$.

The graph $G(n, m, Q)$ defined by an arbitrary probability measure Q (we call such graphs inhomogeneous) was first considered in Godehardt and Jaworski [11], see also [12]. Deijfen and Kets [8] and Bloznelis [3] showed (in increasing generality) that the typical vertex degree of $G(n, m, Q)$ has the power law for a heavy-tailed distribution Q . Another result by Deijfen and Kets [8] says that, for $m \approx \beta n$, sparse intersection graphs $G(n, m, Q)$ possess the clustering property, that is, for any triple of vertices v_i, v_j, v_k , the conditional probability $\mathbf{P}(v_i \sim v_j \mid v_i \sim v_k, v_j \sim v_k)$ is bounded away from zero as $m, n \rightarrow \infty$.

The emergence of a giant connected component in a sparse inhomogeneous intersection graph with $n = o(m)$ (graph without clustering) was studied in [4]. The present paper addresses inhomogeneous intersection graphs with clustering, i.e., the case where $m \approx \beta n$.

2 Results

Given $\beta > 0$, let $\{G(n, m_n, Q_n)\}$ be a sequence of random intersection graphs such that

$$\lim_n m_n n^{-1} = \beta. \quad (1)$$

We shall assume that the sequence of probability distributions $\{Q_n\}$ converges to some probability distribution Q defined on $\{0, 1, 2, \dots\}$,

$$\lim_n Q_n(t) = Q(t) \quad \forall t = 0, 1, \dots, \quad (2)$$

and, in addition, the sequence of the first moments converges,

$$\lim_n \sum_{t \geq 1} t Q_n(t) = \sum_{t \geq 1} t Q(t) < \infty. \quad (3)$$

2.1. Degree distribution. Let $V_n = \{v_1, \dots, v_n\}$ denote the vertex set of $G_n = G(n, m_n, Q_n)$, and let $d_n(v_i)$ denote the degree of vertex v_i . Note that, by symmetry, the random variables $d_n(v_1), \dots, d_n(v_n)$ have the same probability distribution, denoted D_n . In the following proposition we recall a known fact about the asymptotic distribution of D_n .

Proposition 1. *Assume that (1), (2), and (3) hold. Then we have, as $n \rightarrow \infty$,*

$$\mathbf{P}(D_n = k) \rightarrow \sum_{t \geq 0} \frac{(at)^k}{k!} e^{-at} Q(t), \quad k = 0, 1, \dots, \quad (4)$$

where $a = \beta^{-1} \sum_{t \geq 0} t Q(t)$.

Roughly speaking, the limit distribution of D_n is the Poisson distribution $\mathcal{P}(\lambda)$ with random parameter $\lambda = aX$, where X is a random variable with distribution Q . In particular, for a heavy-tailed distribution Q , we obtain the heavy-tailed asymptotic distribution

for D_n . For $Q \sim Bi(m, p)$, (4) is shown in [16]. For arbitrary Q , (4) is shown (in increasing generality) in [8] and [5].

2.2. The largest component. Let $N_1(G)$ denote the order of the largest connected component of a graph G (i.e., $N_1(G)$ is the number of vertices of a connected component which has the largest number of vertices). We are interested in a first-order asymptotic of $N_1(G(n, m_n, Q_n))$ as $n \rightarrow \infty$.

The most commonly used approach in investigating the order of the largest component of a random graph is based on tree counting, see [9], [7]. For inhomogeneous random graphs, it is convenient to count trees with a help of branching processes, see [6]. Here large trees correspond to surviving branching processes, and the order of the largest connected component is described by means of the survival probabilities of a related branching process. In the present paper we use the approach developed in [6].

Before formulating our main result, Theorem 1, we will introduce some notation. Let $\mathcal{X} = \mathcal{X}_{Q,\beta}$ denote the multitype Galton–Watson branching process, where particles are of types $t \in \mathbb{T} = \{1, 2, \dots\}$, and where the number of children of type t of a particle of type s has the Poisson distribution with mean $(s-1)tq_t\beta^{-1}$. Here we write $q_t = Q(t)$, $t \in \mathbb{T}$. Let $\mathcal{X}(t)$ denote the process $\mathcal{X}$ starting at a particle of type t , and $|\mathcal{X}(t)|$ denote the total progeny of $\mathcal{X}(t)$. Let $\rho_{Q,\beta}(t) = \mathbf{P}(|\mathcal{X}(t)| = \infty)$ denote the survival probability of the process $\mathcal{X}(t)$. Write $\rho_{Q,\beta}^{(k)}(t) = \mathbf{P}(|\mathcal{X}(t)| \geq k)$,

$$\tilde{\rho}_{Q,\beta} = \sum_{t \in \mathbb{T}} \rho_{Q,\beta}(t+1)q_t, \quad \tilde{\rho}^{(k)}(Q) = \sum_{t \in \mathbb{T}} \rho_{Q,\beta}^{(k)}(t+1)q_t.$$

Note that for every $t \in \mathbb{T}$, we have $\rho_{Q,\beta}^{(k)}(t) \downarrow \rho_{Q,\beta}(t)$ as $k \uparrow \infty$ (by the continuity property of probabilities). Hence, $\tilde{\rho}^{(k)}(Q) \downarrow \tilde{\rho}(Q)$ as $k \uparrow \infty$.

Notation $o_P(n)$. We write $\eta_n = o_P(1)$ for a sequence of random variables $\{\eta_n\}$ that converges to 0 in probability. We write $\eta_n = o_P(n)$ if $\eta_n n^{-1} = o_P(1)$.

Theorem 1. *Let $\beta > 0$. Let $\{m_n\}$ be a sequence of integers satisfying (1). Let $Q, Q_1, Q_2, \dots$ be probability measures defined on $\{0, 1, 2, \dots\}$ such that $\sum_{t=0}^{m_n} Q_n(t) = 1$ for $n = 1, 2, \dots$. Assume that (2) and (3) hold. Then we have, as $n \rightarrow \infty$,*

$$N_1(G(n, m_n, Q_n)) = n\rho + o_P(n). \quad (5)$$

Here $\rho = \tilde{\rho}_{Q,\beta}$ for $Q(0) < 1$ and $\rho = 0$ otherwise.

We briefly explain the result. Following [6], we discover vertices of the giant component, by exploring the neighborhood of each vertex using the Breadth-First search (BFS): vertices producing large BFS trees (such vertices are called large) are likely to belong to the giant component. What we need is to evaluate the fraction of large vertices or, equivalently, to calculate the probability that the BFS tree rooted at a given vertex, say $v \in V_n$, is large. We denote this probability p_n (it does not depend on v) and expand it, by the total probability formula, $p_n = \sum_{t \geq 1} p_n(t)Q_n(t)$, where $p_n(t) = \mathbf{P}(v \text{ is large} \mid |S(v)| = t)$. Replacing $Q_n(t)$ and $p_n(t)$ by their asymptotic values $Q(t)$ and $\rho_{Q,\beta}(t+1)$, we obtain the

asymptotic value ρ of p_n given in (5). The approximation $p_n(t) \approx \rho_{Q,\beta}(t+1)$ is obtained by coupling the neighborhood exploration process with the branching process $\mathcal{X}(t+1)$. We explain this approximation in more detail. For notational convenience, we assign types to vertices: a vertex u is assigned type $t_u = |S(u)|$. We remark that, for large n , the number of vertices of type t is approximately nq_t , and the probability that vertices of types s and t establish a link is approximately stm^{-1} . In addition, with high probability, each pair of adjacent vertices shares only one common attribute. Now, consider the BFS tree rooted at v . In view of the remarks above, the number of children of type t of the root v is approximately binomially distributed with mean $t_v tm^{-1} \times nq_t \approx t_v tq_t \beta^{-1}$. We couple this number with the Poisson variable of mean $t_v tq_t \beta^{-1}$. Similarly, the number of children of type t of another vertex, say u , of the tree is coupled with the Poisson variable of mean $(t_u - 1)tm^{-1} \times nq_t \approx (t_u - 1)tq_t \beta^{-1}$. Here we use the observation that the attribute connecting u to its closest predecessor (in the BFS tree) attracts children to the predecessor, not to u , while u has remaining $t_u - 1$ attributes to attract its own children. In this way we couple the first $o(n)$ steps of the neighborhood exploration process with the Poisson branching process $\mathcal{X}$. As a result, we obtain the approximation of the probability $p_n(t)$ by the nonextinction probability $\rho_{Q,\beta}(t+1)$.

Remark 1. The correspondence $\rho > 0 \Leftrightarrow \mathbf{E}D_n > c > 1$ established for binomial random intersection graphs in [2], [14] cannot be extended to general inhomogeneous graphs $G(n, m_n, Q_n)$. To see this, consider the graph obtained from a binomial random intersection graph by replacing $S(v_i)$ by $\emptyset$ for a randomly chosen fraction of vertices. This way we can make the expected degree arbitrarily small and still have the giant connected component spanned by a fraction of unchanged vertices.

Remark 2. The kernel $(s, t) \rightarrow (s-1)t\beta^{-1}$ of the Poisson branching process which determines the fraction ρ in the case $m_n \approx \beta n$ differs from the kernel $(s, t) \rightarrow st$ which appears in the case $n = o(m_n)$, see [4]. This difference is explained as follows. For $n = o(m_n)$, the size of the attribute set of a typical vertex of a sparse random intersection graph increases to infinity at the rate $\sqrt{m_n/n}$ as $n \rightarrow \infty$, see [3]. Now the type t_u of a vertex u is set $t_u = |S(u)|\sqrt{n/m_n}$, and the fractions $|S(u)| \times |S(r)|/m$ and $(|S(u)| - 1) \times |S(r)|/m$ (describing the link probability between vertices u and r , and the probability that r is a child of u in a BFS tree) have the same asymptotic value $t_u t_r / n$. Therefore, the link probabilities and the growth of BFS trees are described by the same kernel $(s, t) \rightarrow st$.

3 Proof

The section is organized as follows. First, we collect some notation and formulate auxiliary results. We then prove Theorem 1. The proofs of auxiliary results are given at the end of the section.

Let W' be a finite set of size $|W'| = k$. Let B, H be subsets of W' of sizes $|B| = b$ and $|H| = h$ such that $B \cap H = \emptyset$. Let A be a random subset of W' uniformly distributed in

the class of subsets of W' of size a . Introduce the probabilities

$$\begin{aligned} p(a, b, k) &= \mathbf{P}(A \cap B \neq \emptyset), \\ p_1(a, b, k) &= \mathbf{P}(|A \cap B| = 1), \quad p_2(a, b, k) = \mathbf{P}(|A \cap B| \geq 2), \\ p(a, b, h, k) &= \mathbf{P}(|A \cap B| = 1, A \cap H = \emptyset), \\ p_1(a, b, h, k) &= \mathbf{P}(|A \cap B| = 1, A \cap H \neq \emptyset). \end{aligned}$$

Lemma 1. *Let $k \geq 4$. Denote $\varkappa = ab/k$ and $\varkappa' = ab/(k - a)$. For $a + b \leq k$, we have*

$$\varkappa(1 - \varkappa') \leq p_1(a, b, k) \leq p(a, b, k) \leq \varkappa, \quad (6)$$

$$p_2(a, b, k) \leq 2^{-1} \varkappa^2. \quad (7)$$

Denote $\varkappa'' = (a - 1)h/(k - b)$. For $a + b + h \leq k$, we have

$$\varkappa(1 - \varkappa' - \varkappa'') \leq p(a, b, h, k) \leq \varkappa, \quad (8)$$

$$p_1(a, b, h, k) \leq \varkappa'' \varkappa. \quad (9)$$

Given integers n, m and a vector $\bar{s} = (s_1, \dots, s_n)$ with coordinates from the set $\{0, 1, \dots, m\}$, let $S(v_1), \dots, S(v_n)$ be independent random subsets of $W_m = \{w_1, \dots, w_m\}$ 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 . Let $G_{\bar{s}}(n, m)$ denote the random intersection graph on the vertex set $V_n = \{v_1, \dots, v_n\}$ defined by the random sets $S(v_1), \dots, S(v_n)$. That is, we have $v_i \sim v_j$ whenever $S(v_i) \cap S(v_j) \neq \emptyset$.

Lemma 2. *Let $\beta > 0$. Let $M > 0$ be an integer, and let Q be a probability measure defined on $[M] = \{1, \dots, M\}$. Let $\{m_n\}$ be a sequence of integers, and $\{\bar{s}_n = (s_{n1}, \dots, s_{nn})\}$ be a sequence of vectors with integer coordinates $s_{ni} \in [M]$, $1 \leq i \leq n$. Let n_t denote the number of coordinates of $\bar{s}_n$ attaining the value t . Assume that, for some integer n' and a sequence $\{\varepsilon_n\} \subset (0, 1)$ converging to zero, we have, for every $n > n'$,*

$$\max_{1 \leq t \leq M} |(n_t/n) - Q(t)| \leq \varepsilon_n, \quad (10)$$

$$|m_n(\beta n)^{-1} - 1| \leq \varepsilon_n. \quad (11)$$

Then there exists a sequence $\{\varepsilon_n^\}_{n \geq 1}$ converging to zero such that, for $n > n'$, we have*

$$\mathbf{P}(|N_1(G_{\bar{s}_n}(n, m_n)) - n\tilde{\rho}_{Q, \beta}| > \varepsilon_n^* n) < \varepsilon_n^*. \quad (12)$$

Several technical steps of the proof of Lemma 2 are collected in the separate Lemma 3.

Lemma 3. *Assume that conditions of Lemma 2 are satisfied. For any function $\omega(\cdot)$ satisfying $\omega(n) \rightarrow +\infty$ as $n \rightarrow \infty$, bounds (24), (25), and (27) hold.*

Proof of Theorem 1. Write, for short, $G_n = G(n, m_n, Q_n)$ and $N_1 = N_1(G(n, m_n, Q_n))$. Given $t = 0, 1, \dots$, let n_t denote the number of vertices of G_n with the attribute sets

of size t . Let Q^* denote the probability measure on $\mathbb{T} = \{1, 2, \dots\}$ defined by $Q^*(t) = (1 - Q(0))^{-1}Q(t)$, $t \geq 1$. Write $q_{nt} = Q_n(t)$, $q_t = Q(t)$, and $q_t^* = Q^*(t)$.

Note that vertices with empty attribute sets are isolated in G_n . Hence, the connected components of order at least 2 of G_n belong to the subgraph $G_{[\infty]} \subset G_n$ induced by the vertices with nonempty attribute sets.

In the case where $q_0 = 1$, from (2) we obtain that the expected number of vertices in $G_{[\infty]}$ is $\mathbf{E}(n - n_0) = n(1 - q_{n0}) = o(n)$. This identity implies $N_1 = o_P(n)$. We obtain (5) for $q_0 = 1$.

Let us prove (5) for $q_0 < 1$. Let $G_{[M],n}$ denote the subgraph of G_n induced by the vertices with attribute sets of sizes from the set $[M] = \{1, \dots, M\}$. In the proof we approximate $N_1(G_n)$ by $N_1(G_{[M],n})$ and use the result for $N_1(G_{[M],n})$ shown in Lemma 2.

We need some notation related to $G_{[M],n}$. The inequality $q_0 < 1$ implies that, for large M , the sum $q_{[M]} := q_1 + \dots + q_M \approx 1 - q_0$ is positive. Given such M , let Q_M^* be the probability measure on $[M]$ which assigns the mass $q_{Mt}^* = q_t/q_{[M]}$ to $t \in [M]$. Denote $\tilde{\rho}_{[M]} = \tilde{\rho}_{Q_M^*, \beta_M}$, where $\beta_M = \beta/q_{[M]}$, and write $\beta^* = \beta(1 - q_0)^{-1}$. Clearly, β_M converges to β^* as $M \rightarrow \infty$, and we have

$$\forall t \geq 1, \quad \lim_M q_{Mt}^* = q_t^* \quad \text{and} \quad \lim_M \sum_{t \geq 1} t q_{Mt}^* = \sum_{t \geq 1} t q_t^* < \infty. \quad (13)$$

It follows from (13) that

$$\lim_M \tilde{\rho}_{[M]} = \tilde{\rho}_{Q^*, \beta^*}. \quad (14)$$

For the proof of (14), we refer to Chapter 6 of [6].

We are now ready to prove (5). For this purpose, we combine the upper and lower bounds

$$N_1 \geq n(1 - q_0)\tilde{\rho}_{Q^*, \beta^*} - o_P(n) \quad \text{and} \quad N_1 \leq n(1 - q_0)\tilde{\rho}_{Q^*, \beta^*} + o_P(n),$$

and use the simple identity $(1 - q_0)\tilde{\rho}_{Q^*, \beta^*} = \tilde{\rho}_{Q, \beta}$. We give the proof of the lower bound only. The proof of the upper bound is almost the same as that of the corresponding bound in [4], see formula (56) in [4].

In the proof we show that, for every $\varepsilon \in (0, 1)$,

$$\mathbf{P}(N_1 > n(1 - q_0)\tilde{\rho}_{Q^*, \beta^*} - 2\varepsilon n) = 1 - o(1) \quad \text{as} \quad n \rightarrow \infty. \quad (15)$$

Fix $\varepsilon \in (0, 1)$. In view of (14), we can choose M such that

$$\tilde{\rho}_{Q^*, \beta^*} - \varepsilon < \tilde{\rho}_{[M]} < \tilde{\rho}_{Q^*, \beta^*} + \varepsilon. \quad (16)$$

We apply Lemma 2 to $G_{[M],n}$ conditionally given the event

$$\mathcal{A}_n = \left\{ \max_{1 \leq t \leq M} |n_t - q_t n| < n\delta_n + n^{2/3} \right\}.$$

Here $\delta_n = \max_{1 \leq t \leq M} |q_{nt} - q_t|$ satisfies $\delta_n = o(1)$, see (2). In addition, we have

$$\begin{aligned} 1 - \mathbf{P}(\mathcal{A}_n) &\leq \mathbf{P}\left(\max_{1 \leq t \leq M} |n_t - q_{nt}n| \geq n^{2/3}\right) \\ &\leq \sum_{1 \leq t \leq M} \mathbf{P}(|n_t - q_{nt}n| \geq n^{2/3}) \\ &\leq M n^{-1/3} = o(1). \end{aligned}$$

In the last step we have invoked the bounds $\mathbf{P}(|n_t - q_{nt}n| \geq n^{2/3}) \leq n^{-1/3}$, which follow by Chebyshev's inequality applied to binomial random variables n_t , $t \in [M]$. Now, combining the bound

$$\mathbf{P}(|N_1(G_{[M],n}) - n\tilde{\rho}_{[M]}| > n\varepsilon | \mathcal{A}_n) = o(1) \quad (17)$$

(which follows from Lemma 2) with (16) and the bound $\mathbf{P}(\mathcal{A}_n) = 1 - o(1)$, we obtain

$$\mathbf{P}(|N_1(G_{[M],n}) - n\tilde{\rho}_{Q^*,\beta^*}| > 2n\varepsilon) = o(1).$$

Finally, (15) follows from the obvious inequality $N_1 \geq N_1(G_{[M],n})$. □

Proof of Lemma 2. The proof consists of two steps. First, we show that the components of order at least $n^{2/3}$ contain $n\tilde{\rho}_{Q,\beta} + o_P(n)$ vertices in total. This implies the upper bound for $N_1 = N_1(G_{\bar{s}_n}(n, m_n))$

$$N_1 \leq n\tilde{\rho}_{Q,\beta} + o_P(n). \quad (18)$$

Secondly, we prove that with probability tending to one, such vertices belong to a common connected component. This implies the lower bound

$$N_1 \geq n\tilde{\rho}_{Q,\beta} - o_P(n). \quad (19)$$

Clearly, (18) and (19) yield (12). Before the proof of (18) and (19), we introduce some notation.

Notation. Denote $\tilde{\rho} = \tilde{\rho}_{Q,\beta}$. In what follows, we drop the subscript n and write $m = m_n$, $V = V_n$, $W = W_m$, $G = G_{\bar{s}_n}(n, m)$. We say that a vertex $v \in V$ is of type t if the size $s_v = |S(v)|$ of its attribute set $S(v)$ is t . An edge $u' \sim u''$ of G is called regular if $|S(u') \cap S(u'')| = 1$. In this case, u' and u'' are called regular neighbors. The edge $u' \sim u''$ is called irregular otherwise. We say that v_i is smaller than v_j if $i < j$.

Given $v \in V$, let C_v denote the connected component of G containing vertex v . In order to count vertices of C_v , we explore this component using the BFS procedure. This procedure discovers vertices one by one and collects them in the list, denoted L_v . In what follows, we say that $u' \in L_v$ is older than $u'' \in L_v$ if u' has been added to the list before u'' .

Component exploration. In the beginning all vertices are uncolored. Color v white and add it to the list L_v (now L_v consists of a single white vertex v). Next, we proceed recursively. We choose the oldest white vertex in the list, say u , scan the current set of uncolored vertices (in increasing order) and look for neighbors of u . Each new discovered

neighbor immediately receives white color and is added to the list. In particular, neighbors with smaller indices are added to the list before ones with larger indices. Once all the uncolored vertices are scanned, color u black. Neighbors of u discovered in this step are called children of u . Exploration ends when there are no more white vertices in the list available.

By $L_v^* = \{v = u_1, u_2, u_3, \dots\}$ we denote the final state of the list after the exploration is complete. Here vertices are arranged according to the order of their inclusion in the list (e.g., u_2 was added to the list before u_3). Clearly, L_v^* is the vertex set of C_v . Denote $L_v(k) = \{u_i \in L_v^* : i \leq k\}$. Note that $|L_v(k)| = \min\{k, |L_v^*|\}$. By u_{j^*} we denote the vertex which has discovered u_j (u_j is a child of u_{j^*}). Introduce the sets

$$D_k = \cup_{1 \leq j \leq k} S(u_j), \quad S'(u_i) = S(u_i) \setminus D_{i-1}, \quad k \geq 1, \quad i \geq 2, \quad (20)$$

and put $D_0 = \emptyset$, $S'(u_1) = S(u_1)$.

Regular exploration is performed similarly to the “ordinary” exploration, but now only regular neighbors are added to the list. We call them regular children. A regular child u' of u is called simple if $S(u') \setminus S(u)$ does not intersect with $S(e)$ for any vertex e that has already been included in the list before u' . Otherwise, the regular child is called *complex*. *Simple* exploration is performed similarly to the *regular* exploration, but now simple children are added to the list only.

In the case of regular (respectively simple) exploration, we use the notation $L_v^r, L_v^{r*}, L_v^r(k), D_k^r, S'^r(u_i)$ (respectively $L_v^s, L_v^{s*}, L_v^s(k), D_k^s, S'^s(u_i)$) which is defined in much the same way as above. Similarly, i^* denotes the number in the list (L_v^r or L_v^s depending on the context) of the vertex that has discovered u_i (u_i is a child of u_{i^*}). For an element u_j of the list $L_v^{s*} = \{v = u_1, u_2, \dots\}$, we denote $H(u_j) = (\cup_{j^* < r < j} S(u_r)) \setminus D_{j^*}^s$. Consider the simple exploration at the moment where the current oldest white vertex, say u_i of evolving list $L_v^s = \{v = u_1, u_2, \dots\}$ starts the search of its simple children. Let $U_i = \{v_{j_1}, \dots, v_{j_r}, \dots, v_{j_k}\}$ denote the current set of uncolored vertices (the set of potential simple children). Here $j_1 < j_2 < \dots < j_k$. First, allow u_i to discover its simple children among $\{v_{j_1}, \dots, v_{j_{r-1}}\}$. Define the set $H_i(v_{j_r}) = (\cup_{u \in L} S(u)) \setminus D_i^s$, where L denotes the set of current white elements of the list that are younger than u_i . In particular, L includes the simple children of u_i discovered among $v_{j_1}, \dots, v_{j_{r-1}}$. Observe that any $u' \in U_i$ becomes a simple child of u_i if it is a regular neighbor of u_i and $H_i(u') \cap S(u') = \emptyset$, that is,

$$|S(u') \cap S(u_i)| = 1 \quad \text{and} \quad S(u') \cap H_i(u') = \emptyset. \quad (21)$$

Observe that for any element of the list $u_j \in L_v^{s*}$, we have $H(u_j) = H_{j^*}(u_j)$.

Note that irregular neighbors discovered during *regular* exploration receive white color but are not added to the list L_v^r . Similarly, irregular neighbors and complex children discovered during *simple* exploration receive white color but are not added to the list L_v^s . Note also that L_v^{s*} does not need to be a subset of L_v^{r*} .

Let $\omega(n)$ be an integer function such that $\omega(n) \rightarrow +\infty$ and $\omega(n) = o(n)$ as $n \rightarrow \infty$. A vertex $v \in V$ is called big (respectively, br-vertex and bs-vertex) if $|L_v^*| \geq \omega(n)$ (respectively, $|L_v^{r*}| \geq \omega(n)$ and $|L_v^{s*}| \geq \omega(n)$). Let B, B^r , and B^s denote the collections of

big vertices, br-vertices, and bs-vertices, respectively. Clearly, we have $B^s, B^r \subset B$. Note that in order to decide whether a vertex v is big, we do not need to explore the component C_v completely. Indeed, we may stop the exploration after the number of colored vertices reaches $\omega(n)$. In what follows, we assume that the exploration was stopped after the number of colored vertices had reached $\omega(n)$ (in this case $v \in B$) or ended even earlier because the last white vertex of the list failed to find an uncolored neighbor (in this case $v \notin B$).

The upper bound. Fix $\omega(\cdot)$. We show that

$$|B| - n\tilde{\rho} = o_P(n). \quad (22)$$

Note that (22), combined with the simple inequality $N_1 \leq \max\{\omega(n), |B|\}$, implies (18). We obtain (22) from the bounds

$$|B| - |B^s| = o_P(n), \quad (23)$$

$$|B^s| - n\tilde{\rho} = o_P(n). \quad (24)$$

(24) is shown in Lemma 3. (23) follows from the bound $\mathbf{E}(|B| - |B^s|) = o(n)$. In order to prove this bound, we show that

$$\mathbf{E}|B^s| - n\tilde{\rho} = o(n), \quad (25)$$

$$\mathbf{E}|B| \leq n\tilde{\rho} + o(n). \quad (26)$$

(25) is shown in **Lemma 3**. (26) follows from the bounds

$$\mathbf{E}|B^r| \leq n\tilde{\rho} + o(n), \quad (27)$$

$$\mathbf{E}|B \setminus B^r| = o(n). \quad (28)$$

(27) is shown in **Lemma 3**. In order to show (28), we write $\mathbf{E}|B \setminus B^r| = \sum_{v \in V} \mathbf{P}(v \in B \setminus B^r)$ and invoke the bounds that hold uniformly in $v \in V$,

$$\mathbf{P}(v \in B \setminus B^r) = O(\omega(n)n^{-2}). \quad (29)$$

In the proof of (29) we inspect the list $L_v(\omega(n))$ and look for an irregular child. The probability that given $u_i \in L_v(\omega(n))$ is an irregular child is $O(n^{-2})$, see (7). Now (29) follows from the fact that $L_v(\omega(n))$ has at most $\omega(n) = o(n)$ elements. The proof of (23) is complete.

The lower bound. We start with a simple observation that, with high probability, each attribute $w \in W$ is shared by at most $O(\ln n)$ vertices. Denote $f(w) = \sum_{v \in V} \mathbb{I}_{\{w \in S(v)\}}$, $w \in W$. We show that the inequality

$$\max_{w \in W} f(w) \leq 2M \ln n \quad (30)$$

holds with probability $1 - o(1)$. Since $f(w)$ is a sum of independent Bernoulli random variables with success probabilities at most M/m , Chernoff's inequality implies $\mathbf{P}(f(w) > 2M \ln n) \leq c_{M,\beta} n^{-2}$. Hence, the complementary event to (30) has the probability

$$\mathbf{P}(\max_{w \in W} f(w) > 2M \ln n) \leq \sum_{w \in W} \mathbf{P}(f(w) > 2M \ln n) = o(1).$$

Let us prove (19). Fix $\varepsilon \in (0, 1)$. For each $t \in [M]$, choose $\lceil n_t \varepsilon \rceil$ vertices of type t and color them red. Let G' denote the subgraph of G induced by uncoloured vertices, and let $C_1, C_2, \dots$ denote the (vertex sets of) connected components of G' of order at least $n^{2/3}$. Observe that the number, say k , of such components is at most $(1 - \varepsilon)n^{1/3}$. We apply (22) to the intersection graph G' and function $\omega(n) = \lceil n^{2/3} \rceil$ and obtain $|\cup_{i \geq 1} C_i| = (1 - \varepsilon)n\tilde{\rho}_{Q,\beta'} + o_P(n)$, where $\beta' = \beta(1 - \varepsilon)^{-1}$. We show below that, with high probability, all vertices of $\cup_{i \geq 1} C_i$ belong to a single connected component of the graph G . Hence, $N_1 \geq (1 - \varepsilon)\tilde{\rho}_{Q,\beta'} + o_P(n)$. Letting $\varepsilon \rightarrow 0$, we then immediately obtain lower bound (19).

We assume that G is obtained in two steps. First, the uncolored vertices generate G' , and, secondly, the red vertices add the remaining part of G . Let us consider the second step where the red vertices add their contribution. Write $\mathbb{I}_{ij} = 1$ if C_i and C_j are not connected by a path in G and $\mathbb{I}_{ij} = 0$ otherwise. Let $N = \sum_{1 \leq i < j \leq k} \mathbb{I}_{ij}$ denote the number of disconnected pairs. Clearly, the event $N = 0$ implies that all vertices from $\cup_{i \geq 1} C_i$ belong to the same connected component of G . Therefore, it suffices to show that $\mathbf{P}(N = 0) = 1 - o(1)$. For this purpose, we prove the bound $\mathbf{P}(N \geq 1 | G') = o(1)$ uniformly in G' satisfying (30), see (32) below.

In what follows, we assume that (30) holds. Let $\hat{f}(C_i) = \cup_{v \in C_i} S(v)$ denote the set of attributes occupied by vertices from C_i . Here $\hat{f}(C_i) \cap \hat{f}(C_j) = \emptyset$ for $i \neq j$. Note that if a red vertex finds neighbors in C_i and C_j simultaneously, then it builds a path in G that connects components C_i and C_j . Clearly, only vertices with attribute sets of size at least 2 (i.e., vertices of types 2, 3, ...) can build such a path. The probability of building such a path is minimized by vertices of type 2. This minimal probability is

$$p_{ij} = 2 \frac{|\hat{f}(C_i)| \times |\hat{f}(C_j)|}{m(m-1)}.$$

Note that (30), combined with $|C_i| \geq \lceil n^{2/3} \rceil$, implies that $|\hat{f}(C_i)| \geq n^{2/3}(2M \ln n)^{-1}$. Hence,

$$p_{ij} \geq \frac{1}{2M^2} \frac{n^{4/3}}{(m \ln n)^2} =: p_*.$$

Let $r := \lfloor n_2 \varepsilon \rfloor + \dots + \lfloor n_M \varepsilon \rfloor$ denote the number of red vertices of types 2, 3, Observe that, for large n , (10) implies $r \approx \varepsilon q' n$. Here $q' = q_2 + \dots + q_M$. In particular, we have

$$\mathbf{P}(I_{ij} = 1 | G') \leq (1 - p_{ij})^r \leq (1 - p_*)^r \leq e^{-p_* r}. \quad (31)$$

Here $p_* r \geq c' n^{7/3} (\ln n)^{-2}$, and the constant c' depends on β, M , and q' . Next, we apply Markov's inequality to the conditional probability

$$\mathbf{P}(N \geq 1 | G') \leq \mathbf{E}(N | G') = \sum_{1 \leq i < j \leq k} \mathbf{P}(I_{ij} = 1 | G').$$

Invoking (31) and the inequality $k \leq (1 - \varepsilon)n^{1/3}$, we obtain

$$\mathbf{P}(N \geq 1|G') \leq k^2 e^{-p_* r} \leq n^{2/3} e^{-c' n^{1/3} \ln^{-2} n}. \quad (32)$$

□

Proof of Lemma 3. Throughout the proof, we use the notation of Lemma 2.

Fix $\omega(\cdot)$. Given $0 < \varepsilon < 1$, let $\mathcal{Y}^{+\varepsilon}$ and $\mathcal{Y}^{-\varepsilon}$ be multitype Galton–Watson processes with type space $[M]$ where the numbers of children $Y_{st}^{+\varepsilon}$ and $Y_{st}^{-\varepsilon}$ of type t of a particle of type s have binomial distributions $Bi(\lfloor q_t n(1 + \varepsilon) \rfloor, p_{st}(1 + \varepsilon))$ and $Bi(\lfloor q_t n(1 - \varepsilon) \rfloor, p_{st}(1 - \varepsilon))$, respectively. Here $p_{st} := (s - 1)t(\beta n)^{-1}$.

Let $\mathcal{X}^{+\varepsilon}$ (and $\mathcal{X}^{-\varepsilon}$) be a multitype Galton–Watson process with type space $[M]$ where the number of children $X_{st}^{+\varepsilon}$ (and $X_{st}^{-\varepsilon}$) of type t of a particle of type s has the Poisson distribution with mean $\lambda_{st}(1 + \varepsilon)$ (and $\lambda_{st}(1 - \varepsilon)$). Here $\lambda_{st} := (s - 1)tq_t\beta^{-1}$.

Given a multitype G–W process $\mathcal{Z}$ with type space $[M]$, by $\mathcal{Z}(t)$ we denote the process starting at a particle of type t ; $|\mathcal{Z}(t)|$ denotes the total progeny of $\mathcal{Z}(t)$, $\rho(\mathcal{Z}, t) := \mathbf{P}(|\mathcal{Z}(t)| = \infty)$, and $\rho^{(k)}(\mathcal{Z}, t) := \mathbf{P}(|\mathcal{Z}(t)| \geq k)$.

It is known, see, e.g., inequality (1.23) in [1], that the total-variation distance between the binomial distribution $Bi(r, p)$ and the Poisson distribution with the same mean is at most p . Therefore, by a coupling of the offspring numbers of binomial and Poisson branching processes we obtain

$$\rho^{(\omega(n))}(\mathcal{Y}^{+\varepsilon}, t) = \rho^{(\omega(n))}(\mathcal{X}^{+\varepsilon'}, t) + o(\omega(n)/n), \quad (33)$$

$$\rho^{(\omega(n))}(\mathcal{Y}^{-\varepsilon}, t) = \rho^{(\omega(n))}(\mathcal{X}^{-\varepsilon''}, t) + o(\omega(n)/n). \quad (34)$$

Here $\varepsilon' = (1 + \varepsilon)^2 - 1$ and $\varepsilon'' = 1 - (1 - \varepsilon)^2$. Letting $n \rightarrow \infty$, we obtain

$$\rho^{(\omega(n))}(\mathcal{X}^{+\varepsilon'}, t) \rightarrow \rho(\mathcal{X}^{+\varepsilon'}, t), \quad \rho^{(\omega(n))}(\mathcal{X}^{-\varepsilon''}, t) \rightarrow \rho(\mathcal{X}^{-\varepsilon''}, t). \quad (35)$$

Furthermore, letting $\varepsilon \downarrow 0$, we obtain

$$\rho(\mathcal{X}^{+\varepsilon'}, t) \rightarrow \rho_{Q, \beta}(t), \quad \rho(\mathcal{X}^{-\varepsilon''}, t) \rightarrow \rho_{Q, \beta}(t). \quad (36)$$

Proof of (27). We shall show that

$$\mathbf{P}(v \in B^r) \leq \rho_{Q, \beta}(s_v + 1) + o(1) \quad (37)$$

uniformly in $v \in V$. Collecting these bounds in the identity $\mathbf{E}|B^r| = \sum_{v \in V} \mathbf{P}(v \in B^r)$ and using (10), we then obtain (27). Therefore, it suffices to prove (37). In the proof, we couple *regular* exploration starting at v with the process $\mathcal{Y}^{+\varepsilon}(s_v + 1)$. Let Y_{it}^r denote the number of regular children of type t discovered by $u_i \in L_v^r = \{v = u_1, u_2, \dots\}$. Let n_{it} denote the number of uncolored vertices of type t at the moment when u_i starts exploration of its neighborhood. Then Y_{it}^r has the binomial distribution $Bi(n_{it}, p'_{it})$ with success probability $p'_{it} = p_1(t, |S''(u_i)|, |W \setminus D_{i-1}|)$. Note that for large n , we have

$$n_{it} \leq \lfloor q_t n(1 + \varepsilon) \rfloor, \quad p'_{it} \leq |S''(u_i)| t(\beta n)^{-1}(1 + \varepsilon). \quad (38)$$

The first inequality follows from (10). The second inequality follows from (6) combined with the inequalities

$$m \geq |W \setminus D_{i-1}^r| = m - |D_{i-1}^r| > m - M\omega(n) = m - o(m). \quad (39)$$

In addition, in view of (11), we can replace m by βn in (38). (38) shows that the parameters of the binomial distribution of Y_{it}^r are smaller than the corresponding parameters of the offspring distribution of the branching process $\mathcal{Y}^{+\varepsilon}(s_v + 1)$. Therefore, particles of the branching process produce at least as many children of each type as the vertices u_i , $i < \omega(n)$. Note that $v = u_1$ corresponds to a particle of type $|S''(v)| = s_v + 1$ of the branching process, while the remaining vertices u_i , $i \geq 2$, correspond to particles of types $s_{u_i} = |S(u_i)|$, respectively. Hence, we have

$$\mathbf{P}(v \in B^r) \leq \mathbf{P}(|\mathcal{Y}^{+\varepsilon}(s_v + 1)| \geq \omega(n)). \quad (40)$$

Inequality (40), in combination with (33), (35), and (36), implies (37).

Proof of (25). Given $v \in V$, we start simple exploration at v . Let K_t (I_t) denote the number of complex (irregular) children of type t discovered by the exploration until the list $L_v^s(\omega(n))$ was completed. We put a label on v whenever $\max_t \{K_t, I_t\} \geq \omega(n)$.

Let A denote the set of labeled vertices, and $p'_v := \mathbf{P}(v \in B^s | v \notin A)$ be the probability that the simple exploration of unlabeled vertex v discovers at least $\omega(n)$ vertices. We show below that

$$\mathbf{P}(v \in A) = O(n^{-1}), \quad (41)$$

$$p'_v = \rho_{Q,\beta}(s_v + 1) - o(1). \quad (42)$$

It follows from (41) and (42) that

$$\mathbf{P}(v \in B^s) = p'_v + O(n^{-1}) = \rho_{Q,\beta}(s_v + 1) + o(1). \quad (43)$$

Invoking the latter identity in the expression $\mathbf{E}|B^s| = \sum_{v \in V} \mathbf{P}(v \in B^s)$, we obtain (25).

Proof of (42). Given $\varepsilon > 0$, we show that for large n ,

$$\mathbf{P}(|\mathcal{Y}^{+\varepsilon}(s_v + 1)| \geq \omega(n)) \geq p'_v \geq \mathbf{P}(|\mathcal{Y}^{-\varepsilon}(s_v + 1)| \geq \omega(n)). \quad (44)$$

These inequalities, in combination with (33–36), imply (42).

In order to generate events of probability p'_v , we use rejection sampling. In the course of exploration, we keep track of the number of colored vertices and interrupt the exploration at the moment when this number exceeds $3\omega(n)$. Exploration is rejected if it is interrupted before the list $L_v^s(\omega(n))$ is completed. Otherwise, it is accepted. Clearly, p'_v is the probability that the list $L_v^s(\omega(n))$ of an accepted exploration has collected all $\omega(n)$ elements.

In the proof of (44), we couple the simple exploration process with branching processes $\mathcal{Y}^{-\varepsilon}(s_v + 1)$ and $\mathcal{Y}^{+\varepsilon}(s_v + 1)$ so that the number of simple children of type t of the vertex v is at least (most) as large as the number of particles of type t in the first generation of

$\mathcal{Y}^{-\varepsilon}(s_v + 1)$ ($\mathcal{Y}^{+\varepsilon}(s_v + 1)$), $t \in [M]$. In the further steps of exploration, the number $Y_t(u)$ of simple children of type t discovered by a particle $u \in L_v^s(\omega(n)) \setminus \{v\}$ is at least (most) as large as the number of children of type t produced by the corresponding particle of type s_u of the process $\mathcal{Y}^{-\varepsilon}$ ($\mathcal{Y}^{+\varepsilon}$).

To make sure that such a coupling is possible, we fix $u = u_i \in L_v^s(\omega(n))$ and count its simple children. Recall that u_i selects simple children from the current set of uncolored vertices. These are checked one after another in increasing order, and each newly discovered simple child is added to the list L_v^s before the next uncolored vertex is checked. At the moment when a vertex g is checked, its probability to be a simple child of u is $p_i(g) = p(|S(g)|, |S'^s(u)|, |H_i(g)|, |W \setminus D_{i-1}|)$. It is a conditional probability given $\{S(u'), u' \in L_v^s\}$. Here L_v^s is the set of vertices that have been added to the list before g was checked. Note that, as far as the probability of the event $\{v \in B^s\} \equiv \{L_v^s(\omega(n)) = \omega(n)\}$ is considered, we may safely assume that $|D_{i-1}|, |H_i(g)| \leq M(\omega(n) - 1)$. It follows from these inequalities and (8) that for large n , we have

$$\frac{|S'^s(u)|s_g}{m}(1 - \varepsilon) \leq p_i(g) \leq \frac{|S'^s(u)|s_g}{m}(1 + \varepsilon). \quad (45)$$

In addition, in view of (11), we can replace m by βn in the denominator. Let n_{it}^* denote the number of uncolored vertices of type t at the moment when $u = u_i$ starts searching its simple children. Until the exploration is not interrupted, we have $n_{it}^* \geq n_t - 3\omega(n)$. For large n , this inequality implies $n_{it}^* \geq (1 - \varepsilon/2)n_t$. Invoking (10), we obtain

$$q_t n(1 - \varepsilon) \leq n_{it}^* \leq q_t n(1 + \varepsilon) \quad t \in [M]. \quad (46)$$

It follows from (45)–(46) that we can couple $Y_t(u)$ with binomial random variables

$$Y_t^\pm(u) \sim Bi\left(\lfloor q_t n(1 \pm \varepsilon) \rfloor, \frac{|S'^s(u)|s_g}{\beta n}(1 \pm \varepsilon)\right),$$

so that almost surely we have $Y_t^-(u) \leq Y_t(u) \leq Y_t^+(u)$. These inequalities imply (44).

Proof of (41). We write $\mathbf{P}(v \in A) \leq \sum_{t \in [M]} (\mathbf{P}(K_t \geq \omega(n)) + \mathbf{P}(I_t \geq \omega(n)))$ and show that

$$\mathbf{P}(K_t \geq \omega(n)) = o(n^{-1}), \quad \mathbf{P}(I_t \geq \omega(n)) = o(n^{-1}). \quad (47)$$

We prove the first bound only. The proof of the second bound is much the same. Given $i \leq \omega(n)$, the number of complex children of type t discovered by $u_i \in L_v^s$ is the sum of at most n_t independent Bernoulli random variables, each with success probability at most

$$p^* = p_1(M, M, M\omega(n), m - M\omega(n)) \leq cM^4m^{-2},$$

see (9). Therefore, K_t is at most the sum of $n_t\omega(n)$ independent Bernoulli random variables with success probability p^* . In particular, we have

$$\mathbf{P}(K_t \geq \omega(n)) \leq \mathbf{P}(\xi \geq \omega(n)), \quad (48)$$

where $\xi \sim \text{Bi}(n_t \omega(n), p^*)$. By Chebychev's inequality,

$$\mathbf{P}(\xi \geq \omega(n)) \leq (\omega(n) - \mathbf{E}\xi)^{-2} \mathbf{Var}\xi = O(n^{-1}). \quad (49)$$

In the last step, we invoke the simple bounds

$$\mathbf{Var}\xi \leq \mathbf{E}\xi = n_t \omega(n) p_t^* = O(\omega^2(n) n^{-1}) = o(\omega(n)).$$

(48) and (49) imply the first bound of (47).

Proof of (24). It suffices to establish (24) for one particular function ω , because for any other $\tilde{B}^s$ defined by another such function $\tilde{\omega}$, we have

$$|B^s| - |\tilde{B}^s| = o_P(n). \quad (50)$$

To see this, write $|B^s| - |\tilde{B}^s| \leq |B^s \cup \tilde{B}^s| - |B^s \cap \tilde{B}^s|$ and observe that $B^s \cup \tilde{B}^s$ and $B^s \cap \tilde{B}^s$ represent the sets of bs-vertices defined by the functions $\omega_1 = \min\{\omega, \tilde{\omega}\}$ and $\omega_2 = \max\{\omega, \tilde{\omega}\}$, respectively. An application of (25) to ω_1 and ω_2 yields the bound $\mathbf{E}(|B^s| - |\tilde{B}^s|) = o(n)$. This bound implies (50).

We show (24) for $\omega(n) = \lfloor \ln n \rfloor$. For this purpose, we prove the bound for the variance

$$\mathbf{E}|B^s|^2 - (\mathbf{E}|B^s|)^2 = o(n^2), \quad (51)$$

which tells us that $|B^s| - \mathbf{E}|B^s| = o_P(n)$. In particular, (51), combined with (25), shows (24).

In the proof of (51), we use the observation that the first $\omega(n)$ steps of any two explorations starting at distinct vertices are almost independent. More precisely, we show below that, uniformly in $\{u, v\} \subset V$,

$$\mathbf{P}(u, v \in B^s) = \rho_{Q, \beta}(s_u + 1) \rho_{Q, \beta}(s_v + 1) + o(1). \quad (52)$$

It follows from (52) that

$$\begin{aligned} 2 \sum_{\{u, v\} \subset V} \mathbf{P}(u, v \in B^s) &= \sum_{u, v \in V} \rho_{Q, \beta}(s_u + 1) \rho_{Q, \beta}(s_v + 1) + o(n^2). \\ &= n^2 \tilde{\rho}^2 + o(n^2). \end{aligned} \quad (53)$$

In the last step, we use (10). Observe, that the left-hand sum of (53) is the expected value of $2 \sum_{\{u, v\} \subset V} \mathbb{I}_{\{u, v \in B^s\}} = |B^s|^2 - |B^s|$. Therefore, from (53) we obtain

$$\mathbf{E}|B^s|^2 = n^2 \tilde{\rho}^2 + \mathbf{E}|B^s| + o(n^2).$$

This identity, combined with (25), implies (51).

Let us prove (52). We first explore u and then v . In each case, we stop simple exploration after the number of vertices in the corresponding list reaches $\omega(n)$. Note that, with high probability, these two explorations do not meet. Indeed, let T_u (T_v) denote the set of vertices colored by the first (second) exploration, and let $\mathcal{H}$ denote the event that

the second exploration does not encounter any vertex from T_u , i.e., $\mathcal{H} = \{D_u \cap S(v') = \emptyset \text{ for each } v' \in T_v\}$. Here we denote $D_u = \cup_{u' \in T_u} S(u')$ and $D_v = \cup_{v' \in T_v} S(v')$. Now assume that u, v are unlabeled vertices, i.e., $u, v \notin A$. Then

$$|T_u|, |T_v| \leq (2M + 1)\omega(n) =: \hat{T},$$

and $|D_u|, |D_v| < M\hat{T} \leq M(2M + 1)\omega(n) =: \hat{D}$. In this case, for each $v' \in T_v$, the probability that $S(v')$ does not hit D_u is at least $(\frac{m-2\hat{D}}{m})^M$. Here we use the fact that $S(v')$ has at most $|S(v')| \leq M$ elements (trials) to hit the set D_u which occupies $|D_u| \leq \hat{D}$ attributes among those (at least $m - \hat{D}$) that have not been used by the current collection of vertices of evolving list L_v^s . Since there are at most $\hat{T}$ vertices in T_v , we obtain

$$\mathbf{P}(\mathcal{H} | u, v \notin A) \geq \left(\frac{m-2\hat{D}}{m}\right)^{M\hat{T}} = 1 - O(\omega^2(n)n^{-1}).$$

For arbitrary u, v , from (41) we obtain

$$\mathbf{P}(\mathcal{H}) \geq \mathbf{P}(\mathcal{H} \cap \{u, v \notin A\}) = \mathbf{P}(\mathcal{H} | u, v \notin A) \mathbf{P}(u, v \notin A) = 1 - o(1). \quad (54)$$

Now assume that $\rho_{Q,\beta}(s_u + 1) > 0$ (otherwise, (52) trivially follows from (43)) and write

$$\mathbf{P}(u, v \in B^s) = \mathbf{P}(v \in B^s | u \in B^s) \mathbf{P}(u \in B^s). \quad (55)$$

We can replace $\mathbf{P}(v \in B^s | u \in B^s)$ by $p_{v,u} := \mathbf{P}(v \in B^s | \{u \in B^s\} \cap \{u, v \in A\} \cap \mathcal{H})$ and $\mathbf{P}(u \in B^s)$ by $\rho_{Q,\beta}(s_u + 1)$. It follows from (41), (54), and (43) that the error due to such replacement is of order $o(1)$. From (55) we obtain

$$\mathbf{P}(u, v \in B^s) = p_{v,u} \rho_{Q,\beta}(s_u + 1) + o(1). \quad (56)$$

Finally, (52) follows from (56) and the identity $p_{v,u} = \rho_{Q,\beta}(s_v + 1) + o(1)$, which is shown in much the same way as (42) above. $\square$

Proof of Lemma 1. Let $(x_1, \dots, x_k)$ be a random permutation of elements of the set W' . For $A = \{x_1, \dots, x_a\}$, we have, by symmetry,

$$p(a, b, k) \leq \sum_{1 \leq i \leq a} \mathbf{P}(x_i \in B) = a \mathbf{P}(x_1 \in B), \quad (57)$$

$$p_1(a, b, k) = \sum_{1 \leq i \leq a} \mathbf{P}(A \cap B = x_i) = a \mathbf{P}(A \cap B = x_1), \quad (58)$$

$$p_2(a, b, k) \leq \sum_{1 \leq i < j \leq a} \mathbf{P}(x_i, x_j \in B) = 2^{-1} a(a-1) \mathbf{P}(x_1, x_2 \in B), \quad (59)$$

$$\begin{aligned} p(a, b, h, k) &= \sum_{1 \leq i \leq a} \mathbf{P}(A \cap B = x_i) \mathbf{P}(H \cap A = \emptyset | A \cap B = x_i) \\ &= p_1(a, b, k) (1 - p(a-1, h, k-b)). \end{aligned} \quad (60)$$

The right-hand side inequality of (6) follows from (57) and the identity $\mathbf{P}(x_1 \in B) = b/k$. The left-hand side inequality follows from (58) combined with the identity $\mathbf{P}(A \cap B = x_1) = \frac{b(k-b)_{a-1}}{(k)_a}$ and inequalities

$$1 \geq \frac{(k-b)_{a-1}}{(k-1)_{a-1}} \geq \left(\frac{k-a-b}{k-a}\right)^{a-1} \geq 1 - \frac{ab}{k-a}.$$

(7) follows from (59) and the identity $\mathbf{P}(x_1, x_2 \in B) = \frac{(b)_2}{(k)_2}$. (8) follows from (60) combined with (6). (9) follows from the inequality $p_1(a, b, h, k) = p_1(a, b, k)p(a-1, h, k-b)$, which is shown in the same way as (60). $\square$

Acknowledgments. I would like to thank the anonymous referee for remarks and suggestions that have improved the presentation. The work was supported in part by the Research Council of Lithuania grant MIP-52/2010.

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, Component evolution in general random intersection graphs. *SIAM J. Discrete Math.* 24 (2010), 639–654.
- [5] M. Bloznelis, A random intersection digraph. Indegree and outdegree distributions, *Discrete Mathematics* 310 (2010), 2560–2566.
- [6] B. Bollobás, S. Janson, and O. Riordan, The phase transition in inhomogeneous random graphs, *Random Structures Algorithms* 31 (2007), 3–122.
- [7] F. Chung and L. Lu, Complex graphs and networks, CBMS regional conference series in mathematics, no 107, American Mathematical Society, Providence, Rhode Island, 2006.
- [8] M. Deijfen and W. Kets, Random intersection graphs with tunable degree distribution and clustering, *Probab. Eng. Inform. Sci.* 23 (2009), 661–674.
- [9] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.*, 5 (1960), 17–61.
- [10] J. A. Fill, E. R. Scheinerman, K. B. Singer-Cohen, Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models, *Random Structures and Algorithms*, 16 (2000), 156–176.

- [11] 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.
- [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] A. N. Lagerås and M. Lindholm, A note on the component structure in random intersection graphs with tunable clustering, *Electronic Journal of Combinatorics* 15(1) (2008).
- [15] K. B. Singer-Cohen, Random intersection graphs, PhD thesis, Department of Mathematical Sciences, The John Hopkins University, 1995.
- [16] D. Stark, The vertex degree distribution of random intersection graphs, *Random Structures Algorithms* 24 (2004), 249–258.