Large cliques in sparse random intersection graphs (extended version) *

Valentas Kurauskas and Mindaugas Bloznelis

Vilnius University

2013-09-21

Abstract

Given positive integers n and m, and a probability measure P on $\{0, 1, \ldots, m\}$, the random intersection graph $G(n, m, P)$ on vertex set $V = \{1, 2, \ldots, n\}$ and with attribute set $W = \{w_1, w_2, \ldots, w_m\}$ is defined as follows. Let $S_1, S_2, \ldots, S_n$ be independent random subsets of W such that for any $v \in V$ and any $S \subseteq W$ we have $\mathbb{P}(S_v = S) = P(|S|)/\binom{m}{|S|}$. The edge set of $G(n, m, P)$ consists of those pairs $\{u, v\} \subseteq V$ for which $S_u \cap S_v \neq \emptyset$.

We study the asymptotic order of the clique number $\omega(G(n, m, P))$ of sparse random intersection graphs. For instance, for $m = \Theta(n)$ we show that the maximum clique is of size

$$(1 - \alpha/2)^{-\alpha/2} n^{1-\alpha/2} (\ln n)^{-\alpha/2} (1 + o_P(1))$$

in the case where the vertex degree distribution is a power-law with exponent $\alpha \in (1; 2)$, and it is of size $\frac{\ln n}{\ln \ln n}(1 + o_P(1))$ in the case where the degree distribution has a finite variance. In each case there is a polynomial algorithm which finds a clique of size $\omega(G(n, m, P))(1 - o_P(1))$.

keywords: clique, random intersection graph, greedy algorithm, complex network, power-law, clustering

corresponding author:

*Supported by the Research Council of Lithuania (MIP-052/2010, MIP-067/2013).

Valentas Kurauskas, Faculty of Mathematics and Informatics, Didlaukio 47, LT-08303 Vilnius, Lithuania

email:

Mindaugas Bloznelis: mindaugas.bloznelis@mif.vu.lt

Valentas Kurauskas: valentas@gmail.com

1 Introduction

Bianconi and Marsili observed in 2006 [4] that “scale-free” real networks can have very large cliques; they gave an argument suggesting that the rate of divergence is polynomial if the degree variance is unbounded [4]. In a more precise analysis, Janson, Łuczak and Norros [11] showed exact asymptotics for the clique number in a power-law random graph model where edge probabilities are proportional to the product of weights of their endpoints.

Another feature of a real network that may affect formation of cliques is the clustering property: the probability of a link between two randomly chosen vertices increases dramatically after we learn about the presence of their common neighbour. An interesting question is whether and how the clustering property is related to the clique number.

With conditionally independent edges, the random graph of [11] does not have the clustering property and, therefore, can not explain such a relation.

In the present paper we address this question by showing precise asymptotics for the clique number of a related random intersection graph model that admits a tunable clustering coefficient and power-law degree distribution. We find that the effect of clustering on the clique number only shows up for the degree sequences having a finite variance. We note that the finite variance is a necessary, but not sufficient condition for the clustering coefficient to attain a non-trivial value, see [6] and (5) below.

In the language of hypergraphs, we ask what is the largest intersecting family in a random hypergraph on the vertex set $[m]$, where n identically distributed and independent hyperedges have random sizes distributed according to P. A related problem for uniform hypergraphs was considered by Balogh, Bohman and Mubayi [2]. Although the motivation and the approach of [2] are different from ours, the result of [2] yields the clique number, for a particular class of random intersectipon graphs based on the subsets having the same (deterministic) number of elements.

The random intersection graph model was introduced by Karoński, Scheinerman and Singer-Cohen in 1999 [13] and further generalised by Godehardt and Jaworski [10] and others. With appropriate parameters, it yields graphs that are sparse [8, 7], have a positive clustering coefficient [8, 6] and assortativity [5]. We will consider a sequence $\{G(n)\} = \{G(n), n = 1, 2, \dots\}$ of random intersection graphs $G(n) = G(n, m, P)$, where $P = P(n)$ and $m = m(n) \to +\infty$ as $n \to +\infty$. Let $X(n)$ denote a random variable distributed according to $P(n)$ and define $Y(n) := \sqrt{\frac{n}{m}} X(n)$. If not explicitly stated otherwise, the limits below will be taken as $n \to \infty$. In this paper we use the standard notation $o()$, $O()$, $\Omega()$, $\Theta()$, $o_P()$, $O_P()$, see, for example, [12]. For positive sequences (a_n), (b_n) we write $a_n \sim b_n$ if $a_n/b_n \to 1$ $a_n \ll b_n$ if $a_n/b_n \to 0$. For a sequence of events $\{\mathcal{A}_n\}$, we say that $\mathcal{A}_n$ occurs *whp*, if $\mathbb{P}(\mathcal{A}_n) \to 1$.

We will assume in what follows that

$$\mathbb{E}\, Y(n) = O(1). \tag{1}$$

This condition ensures that the expected number of edges in $G(n)$ is $O(n)$. Hence

$G(n)$ is sparse. We remark, that if, in addition, $Y(n)$ converges in distribution to an integrable random variable, say Z, and $\mathbb{E}\,Y(n) \to \mathbb{E}\,Z$, then $G(n)$ has asymptotic degree distribution $Poiss(\lambda)$, where $\lambda = Z\mathbb{E}\,Z$, see, e.g., [6]. In particular, if $Y(n)$ has asymptotic square integrable distribution, then $G(n)$ has asymptotic square integrable degree distribution too. Furthermore, if $Y(n)$ has a power-law asymptotic distribution, then $G(n)$ has asymptotic power-law degree distribution with the same exponent.

Our first result, Theorem 1.1, shows that in the latter case the clique number diverges polynomially. In fact, we do not require $Y(n)$ to have a limiting power-law distribution, but consider a condition that only involves the tail of $Y(n)$. Namely, we assume that for some $\alpha > 0$ and some slowly varying function L there is $0 < \epsilon_0 < 0.5$ such that for each sequence x_n with $n^{1/2-\epsilon_0} \le x_n \le n^{1/2+\epsilon_0}$ we have

$$\mathbb{P}\left(Y(n) \ge x_n\right) \sim L(x_n)x_n^{-\alpha}. \tag{2}$$

We recall that a function $L : \mathbb{R}_+ \to \mathbb{R}_+$ is called *slowly varying* if $\lim_{x\to\infty} L(tx)/L(x) = 1$ for any $t > 0$.

Theorem 1.1 *Let $1 < \alpha < 2$. Assume that $\{G(n)\}$ is a sequence of random intersection graphs satisfying (1), (2). Suppose that for some $\beta > \max\{2-\alpha, \alpha-1\}$ we have $m = m(n) = \Omega(n^\beta)$. Then the clique number of $G(n)$ is*

$$\omega(G(n)) = (1 + o_P(1))\,(1 - \alpha/2)^{-\alpha/2}\,K(n) \tag{3}$$

where

$$K(n) = L\left((n \ln n)^{1/2}\right) n^{1-\alpha/2}(\ln n)^{-\alpha/2}.$$

We remark that adjacency relations of neighbouring vertices of a random intersection graph are statistically dependent events and this dependence is not negligible for $m = O(n)$. Theorem 1.1 says that in the case where the asymptotic degree distribution has infinite second moment ($\alpha < 2$), the asymptotic order (3) of a power-law random intersection graph is the same as that of the related model of [11] which has conditionally independent edges. Let us mention that the lower bound for the clique number $\omega(G(n))$ is obtained using a simple and elegant argument of [11], which is not sensitive to the statistical dependence of edges of $G(n)$. To show the matching upper bound we developed another approach based on a result of Alon, Jiang, Miller and Pritkin [1] in Ramsey theory.

In the case where the (asymptotic) degree distribution has a finite second moment we not only find the asymptotic order of $\omega(G(n))$, but also describe the structure of a maximal clique. To this aim, it is convenient to interpret attributes $w \in W$ as colours. The set of vertices $T(w) = \{v \in V : w \in S_v\}$ induces a clique in $G(n)$ which we denote (with some ambiguity of notation) $T(w)$. We say that every edge of $T(w)$ receives colour w and call this clique *monochromatic*. Note that $G(n)$ is covered by the union of monochromatic cliques $T(w)$, $w \in W$. We denote the size of the largest monochromatic clique by $\omega'(G(n))$. Clearly, $\omega(G(n)) \ge \omega'(G(n))$.

Denote $x \vee y = \max(x, y)$. The next theorem shows that the largest clique is a monochromatic clique (plus possibly a few extra vertices).

Theorem 1.2 *Assume that $\{G(n)\}$ is a sequence of random intersection graphs satisfying (1). Suppose that $Var(Y(n)) = O(1)$. Then*

$$\omega(G(n)) = \omega'(G(n)) + O_P(1).$$

If, in addition, for some positive sequence $\{\epsilon_n\}$ converging to zero we have

$$n\mathbb{P}(Y(n) > \epsilon_n n^{1/2}) \to 0 \tag{4}$$

then, for an absolute constant C,

$$\mathbb{P}\left(\omega(G(n)) \le C \vee (\omega'(G(n)) + 3)\right) \to 1.$$

The condition (4) is not very restrictive. It is satisfied by uniformly square integrable sequences $\{Y(n)\}$. In particular, (4) holds if $\{Y(n)\}$ converges in distribution to a square integrable random variable, say Y_*, and $\mathbb{E}\,Y^2(n)$ converges to $\mathbb{E}\,Y_*^2$.

Next, we evaluate the size of the largest monochromatic clique. For this purpose we relate the random intersection graph to the balls into bins model. Let every vertex $v \in V$ throw $X_v := |S_v|$ balls into the bins $w_1, \dots, w_m$ uniformly at random, subject to the condition that every bin receives at most one ball from each vertex. Then $\omega'(G(n))$ counts the maximum number of balls contained in a bin. Let $M(N, m)$ denote the maximum number of balls contained in any of m bins after N balls were thrown into m bins uniformly and independently at random. Our next result says that the probability distribution of $\omega'(G(n))$ can be approximated by that of $M(N, m)$, with $N \approx n\mathbb{E}\,X(n) = \mathbb{E}\,(X_1 + \cdots + X_n)$. The asymptotics of $M(N, m)$ are well known, see, e.g., Section 6 of Kolchin et al [14].

Denote by $d_{TV}(\xi, \eta) = 2^{-1}\sum_{i\ge 0}|\mathbb{P}(\xi = i) - \mathbb{P}(\eta = i)|$ the total variation distance between probability distributions of non-negative integer valued random variables ξ and η.

Theorem 1.3 *Assume that $\{G(n)\}$ is a sequence of random intersection graphs satisfying $\mathbb{E}\,Y = \Theta(1)$ and $Var(Y) = O(1)$. Then*

$$d_{TV}(\omega'(G(n)), M(\lfloor (mn)^{1/2}\mathbb{E}\,Y(n)\rfloor, m)) \to 0.$$

Remark 1.4 *For $n, m \to +\infty$ the relations $\mathbb{E}\,Y = \Theta(1), VarY = O(1)$ imply $n = O(m)$. In particular, the conditions of Theorem 1.3 rule out the case $m = o(n)$.*

Let us summarize our results about the clique number of a sparse random intersection graph $G(n)$ with a square integrable (asymptotic) degree distribution. We note that the conditional probability (called the clustering coefficient of $G(n)$)

$$\mathbb{P}(v_1 \sim v_2 | v_1 \sim v_3, v_2 \sim v_3) \approx (n/m)^{1/2}\mathbb{E}\,Y(n)/\mathbb{E}\,Y^2(n) \tag{5}$$

only attains a non-trivial value for $m = \Theta(n)$ and $\mathbb{E}\,Y^2(n) = \Theta(1)$. (Here $u \sim v$ is the event that u and v are adjacent in $G(n)$, i.e., $uv \in E(G(n))$.) In the latter case

Theorems 1.2 and 1.3 together with the asymptotics for $M(N,m)$ (Theorem II.6.1 of [14]), imply that

$$\omega(G(n)) = \frac{\ln n}{\ln \ln n}\,(1 + o_P(1))\,.$$

In contrast, the clique number of a sparse Erdős-Rényi random graph $G(n, c/n)$ is at most 3, and in the model of [11], with square integrable asymptotic degree distribution, the largest clique has at most 4 vertices.

Each of our main results, Theorem 1.1 and Theorem 1.2, have corresponding simple polynomial algorithms that construct a clique of the optimal order whp. For a power-law graph with $\alpha \in (1; 2)$, it is the greedy algorithm of [11]: sort vertices in descending order according to their degree; traverse vertices in that order and 'grow' a clique, by adding a vertex if it is connected to each vertex in the current clique. For a graph with a finite degree variance we propose even simpler algorithm: for each pair of adjacent vertices, take any maximal clique formed by that pair and their common neighbours. Output the biggest maximal clique found in this way. More details and analysis of each of the algorithms are given in Section 4 below.

In practical situations a graph may be assumed to be distributed as a random intersection graph, but information about the subset size distribution may not be available. In such a case, instead of condition (2) for the tail of the normalised subset size $Y(n)$, we may consider a similar condition for the tail of the degree $D_1(n)$ of the vertex $1 \in V$ in $G(n)$: there are constants $\alpha' > 1, \epsilon' > 0$ and a slowly varying function $L'(x)$ such that for any sequence t_n with $n^{1/2-\epsilon'} \le t_n \le n^{1/2+\epsilon'}$

$$\mathbb{P}(D_1(n) \ge t_n) \sim L'(t_n) t_n^{-\alpha'}. \tag{6}$$

The following lemma shows that, subject to an additional assumption, there is equivalence between conditions (2) and (6) .

Lemma 1.5 *Assume that $\{G(n)\}$ is a sequence of random intersection graphs such that for some $\epsilon > 0$ we have*

$$\mathbb{E}\, Y(n) \mathbb{I}_{Y(n) \ge n^{1/2-\epsilon}} \to 0. \tag{7}$$

Suppose that either $(\mathbb{E}\, Y(n))^2$ or $\mathbb{E}\, D_1(n)$ converges to a positive number, say, d.

Then both limits exist and are equal, $\lim \mathbb{E}\, D_1(n) = \lim(\mathbb{E}\, Y(n))^2 = d$. Furthermore, the condition (6) holds if and only if (2) holds. In that case, $\alpha' = \alpha$ and $L'(t) = d^{\alpha/2} L(t)$.

Thus, under a mild additional assumption (7), condition (2) of Theorem 1.1 can be replaced by (6). Similarly, the condition $VarY(n) = O(1)$ of Theorem 1.2 can be replaced by the condition $VarD_1(n) = O(1)$.

Lemma 1.6 *Assume that $\{G(n)\}$ is a sequence of random intersection graphs and for some positive sequence $\{\epsilon_n\}$ converging to zero we have*

$$\mathbb{E}\, Y^2(n) \mathbb{I}_{Y(n) > \epsilon_n n^{1/2}} \to 0. \tag{8}$$

Suppose that either $\mathbb{E}\,Y(n) = \Theta(1)$ *or* $\mathbb{E}\,D_1(n) = \Theta(1)$. *Then*

$$\mathbb{E}\,D_1(n) = (\mathbb{E}\,Y(n))^2 + o(1) \tag{9}$$

$$VarD_1(n) = (\mathbb{E}\,Y(n))^2(VarY(n) + 1) + o(1). \tag{10}$$

Cliques of random intersection graphs have been studied in [13], where edge density thresholds for emergence of small (constant-sized) cliques were determined, and in [17], where the Poisson approximation to the distribution of the number of small cliques was established. The clique number was studied in [16], see also [3], in the case, where $m \approx n^{\beta}$, for some $0 < \beta < 1$. We note that in the papers [13], [17], [16] a particular random intersection graph with the binomial distribution $P \sim Bin(p, m)$ was considered.

The rest of the paper is organized as follows. In Section 2 we study sparse random power-law intersection graphs with index $\alpha \in (1; 2)$, introduce the result on "rainbow" cliques in extremal combinatorics (Lemma 2.8) and prove Theorem 1.1. In Section 3 we relate our model to the balls and bins model and prove Theorem 1.2. In Section 4 we present and analyse algorithms for finding large cliques in $G(n, m, P)$. In Section 5 we prove Lemmas 1.5 and 1.6.

2 Power-law intersection graphs

2.1 Proof of Theorem 1.1

We start with introducing some notation. Given a family of subsets $\{S_v, v \in V'\}$ of an attribute set W', we denote $G(V', W')$ the *intersection graph* on the vertex set V' defined by this family: $u, v \in V'$ are adjacent (denoted $u \sim v$) whenever $S_u \cap S_v \neq \emptyset$. We say that an attribute $w \in W'$ *covers* the edge $u \sim v$ of $G(V', W')$ whenever $w \in S_u \cap S_v$. In this case we also say that the edge $u \sim v$ receives colour w. In particular, an attribute w covers all edges of the (monochromatic) clique subgraph T_w of $G(V', W')$ induced by the vertex set $T_w = \{v \in V' : w \in S_v\}$. Given a graph H, we say that $G(V', W')$ *contains a rainbow* H if there is a subgraph $H' \subseteq G(V', W')$ isomorphic to H such that every edge of H' can be prescribed an attribute that covers this edge so that all prescribed attributes are different.

We denote by $e(G)$ the size of the set $E(G)$ of edges of a graph G. Given two graphs $G = (V(G), E(G))$ and $R = (V(R), E(R))$ we denote by $G \vee R$ the graph on vertices $V(G) \cup V(R)$ and with edges $E(G) \cup E(R)$. In what follows we assume that $V(G) = V(R)$ if not mentioned otherwise. Let t be a positive integer and let R be a non-random graph on the vertex set V'. Assuming that subsets S_v, $v \in V'$ are drawn at random, introduce the event $Rainbow(G(V', W'), R, t)$ that the graph $G(V', W') \vee R$ has a clique H of size $|V(H)| = t$ with the property that every edge of the set $E(H) \setminus E(R)$ can be prescribed an attribute that covers this edge so that all prescribed attributes are different.

In the case where every vertex v of the random intersection graph $G(n, m, P)$ includes attributes independently at random with probability $p = p(n)$, the size $X_v := |S_v|$ of the attribute set has binomial distribution $P \sim Binom(m, p)$. We denote such graph $G(n,m,p)$ and call it a *binomial* random intersection graph. We note that for $mp \to +\infty$ the sizes X_v of random sets are concentrated around their mean value $\mathbb{E}\, X_v = mp$. An application of Chernoff's bound (see, e.g., [15])

$$\mathbb{P}(|B - mp| > \epsilon mp) \le 2e^{-\frac{1}{3}\epsilon^2 mp}, \tag{11}$$

where B is a binomial random variable $B \sim Binom(m,p)$ and $0 < \epsilon < 3/2$, implies

$$\mathbb{P}(\exists v \in [n] : |X_v - mp| > y) \le n\mathbb{P}(|X_v - mp| > y) \to 0 \tag{12}$$

for any $y = y(n)$ such that $y/\sqrt{mp \ln n} \to \infty$ and $y/(mp) < 3/2$.

We write $a \wedge b = \min\{a,b\}$ and $a \vee b = \max\{a,b\}$.

Let us prove Theorem 1.1. For every member $G(n) = G(V,W)$ of a sequence $\{G(n)\}$ satisfying conditions of Theorem 1.1 and a number $\epsilon_1 \in (0, \epsilon_0)$ define the subgraphs $G_i \subseteq G(n)$, $i = 0,1,2$, induced by the vertex sets

$$\begin{aligned} V_0 = V_0(n) &= \{v \in V(G(n)) :\ X_v < \theta_1\}; \\ V_1 = V_1(n) &= \{v \in V(G(n)) :\ \theta_1 \le X_v \le \theta_2\}; \\ V_2 = V_2(n) &= \{v \in V(G(n)) :\ \theta_2 < X_v\}, \end{aligned}$$

respectively. Here $X_v = |S_v|$ denotes the size of the attribute set prescribed to a vertex v and the numbers

$$\theta_1 = \theta_1(n) = m^{1/2} n^{-\epsilon_1}; \qquad \theta_2 = \theta_2(n) = \left((1 - \alpha/2) m \ln n + m e_1\right)^{1/2},$$

with $e_1 = e_1(n) = \max(0, \ln L((n \ln n)^{1/2}))$. Note that $e_1 \equiv 0$ for $L(x) \equiv 1$. We have $V = V_0 \cup V_1 \cup V_2$ and $V_i \cap V_j = \emptyset$ for $i \ne j$. Theorem 1.1 follows from the three lemmas below. Let $K = K(n)$ be as in Theorem 1.1. The first lemma gives a lower bound for the clique number of $G(n)$.

Lemma 2.1 *For any $m = m(n)$*

$$\omega(G_2) = |V_2|(1 - o_P(1)) = (1 - o_P(1))\,(1 - \alpha/2)^{-\alpha/2} K.$$

The next two lemmas provide an upper bound.

Lemma 2.2 *Suppose there is $\beta > \alpha - 1$ such that $m = \Omega(n^\beta)$. If $\epsilon_1 < \frac{\beta}{6}$ then there is $\delta > 0$ such that*

$$\mathbb{P}\left(\omega(G_0) \ge n^{1-\alpha/2-\delta}\right) \to 0.$$

Lemma 2.3 *Suppose there is $\beta > 2 - \alpha$ such that $m = \Omega(n^\beta)$. If $\epsilon_1 < \frac{\beta-2+\alpha}{24}$ then*

$$\omega(G_1) = o_P(K).$$

Proof of Theorem 1.1 We choose $0 < \epsilon_1 < \min\{(\alpha-1)/6, (\beta - 2 + \alpha)/24, \epsilon_0\}$. The theorem follows from the inequalities $\omega(G_2) \le \omega(G) \le \omega(G_0) + \omega(G_1) + \omega(G_2)$ and Lemmas 2.1, 2.2 and 2.3. $\square$

2.2 Proof of Lemma 2.1

In this section we use ideas from [11] to give a lower bound on the clique number. We first note the following auxiliary facts.

Lemma 2.4 *Suppose $a = a_n, b = b_n$ are sequences of positive reals such that $0 < \ln 2b + 2a \to +\infty$. Let z_n be the positive root of*

$$a - \ln z - bz^2 = 0. \tag{13}$$

Then $z_n \sim \sqrt{\frac{2a+\ln(2b)}{2b}}$.

Proof Changing the variables $t = 2bz^2$ we get

$$t + \ln(t) = 2a + \ln(2b).$$

From the assumption it follows that $t + \ln t \sim t$ and therefore $z_n \sim \sqrt{\frac{2a+\ln(2b)}{2b}}$. □

Lemma 2.5 ([9]) *Let $x \to +\infty$. For any slowly varying function L and any $0 < t_1 < t_2 < +\infty$ the convergence $L(tx)/L(x) \to 1$ is uniform in $t \in [t_1, t_2]$. Furthermore, we have $\ln L(x) = o(\ln x)$.*

Proof of Lemma 2.1 Write $N = |V_2|$ and let

$$v^{(1)}, v^{(2)}, \dots, v^{(N)}$$

be the vertices of V_2 listed in an arbitrary order.

Consider a greedy algorithm for finding a clique in G proposed by Janson, Łuczak and Norros [11] (they use descending ordering by the set sizes, see also Section 4). Let $A^0 = \emptyset$. In the step $i = 1, 2, \dots, N$ let $A^i = A^{i-1} \cup \{v^{(i)}\}$ if $v^{(i)}$ is incident to each of the vertices $v^{(j)}$, $j = 1, \dots, i-1$. Otherwise, let $A^i = A^{i-1}$. This algorithm produces a clique H on the set of vertices A^N, and H demonstrates that $\omega(G_2) \geq |A^N|$.

Write $\theta = \theta_2$ and let $L_\theta = V_2 \setminus A^N$ be the set of vertices that failed to be added to A^N. We will show that

$$\frac{|L_\theta|}{N \vee 1} = o_P(1)$$

and

$$N = (1-\alpha/2)^{-\alpha/2}\, L\left((n \ln n)^{1/2}\right) (\ln n)^{-\alpha/2} n^{1-\alpha/2} (1 - o_P(1)).$$

From (2) we obtain for $N \sim Binom(n, q)$ with $q = \mathbb{P}(X_n > \theta)$

$$\begin{aligned} \mathbb{E}\, N = nq &= n\mathbb{P}\left((m/n)^{1/2} Y_n > \theta\right) \\ &\sim L\left((n/m)^{1/2}\theta\right) n^{1-\alpha/2} m^{\alpha/2} \theta^{-\alpha} \\ &\sim (1-\alpha/2)^{-\alpha/2}\, L(\sqrt{n \ln n}) (\ln n)^{-\alpha/2} n^{1-\alpha/2}. \end{aligned}$$

Here we used $L((n/m)^{1/2}\theta) \sim L(\sqrt{n \ln n})$ and $\ln L(\sqrt{n \ln n}) = o(\ln n)$, see Lemma 2.5. Furthermore, by the concentration property of the binomial distribution, see, e.g., (11), we have $N = (1 - o_P(1))\mathbb{E}\, N$.

The remaining bound $|L_\theta|/(N \vee 1) \le |L_\theta|/(N+1) = o_P(1)$ follows from the bound $\mathbb{E}\,(L_\theta/(N+1)) = o(1)$, which is shown below.

Let p_1 be the probability that two random independent subsets of $W = [m]$ of size $\lceil \theta \rceil$ do not intersect. The number of vertices in L_θ is at most the number of pairs in $x, y \in V_2$ where S_x and S_y do not intersect. Therefore by the first moment method

$$\mathbb{E}\, \frac{|L_\theta|}{N+1} = \mathbb{E}\,\mathbb{E}\left(\frac{|L_\theta|}{N+1}\Big| N\right) \le \mathbb{E}\,\mathbb{E}\left(\frac{\binom{N}{2} p_1}{N+1}\Big| N\right) \le \frac{p_1 \mathbb{E}\, N}{2},$$

where

$$p_1 = \frac{\binom{m-\theta}{\theta}}{\binom{m}{\theta}} \le \left(1 - \frac{\theta}{m}\right)^\theta \le e^{-\theta^2/m}.$$

Now it is straightforward to check that for some constant c we have $p_1 \mathbb{E}\, N \le c(\ln n)^{-\alpha/2} \to 0$. This completes the proof.

Let us briefly explain the intuition for the choice of θ. For simplicity assume $L(x) \equiv 1$ so that $e_1 = 0$. Could the same method yield a bigger clique if θ_2 is smaller? We remark that the product $p_1 \mathbb{E}\, N$ as well as its upper bound $n^{1-\alpha/2} m^{\alpha/2} \theta^{-\alpha} e^{-\theta^2/m}$ (which we used above) are decreasing functions of θ. Hence, if we wanted this upper bound to be $o(1)$ then θ should be at least as large as the solution to the equation

$$n^{1-\alpha/2} m^{\alpha/2} \theta^{-\alpha} e^{-\theta^2/m} = 1$$

or, equivalently, to the equation

$$\alpha^{-1} \ln n + \frac{1}{2} \ln(m/n) - \ln \theta - \frac{\theta^2}{\alpha m} = 0. \tag{14}$$

After we write the latter relation in the form (13) where $a = \alpha^{-1} \ln n + (1/2) \ln(m/n)$ and $b = (\alpha m)^{-1}$ satisfy $be^{2a} = \alpha^{-1} n^{\frac{2}{\alpha} - 1} \to +\infty$, we obtain from Lemma 2.4 that the solution θ of (14) satisfies

$$\theta \sim \sqrt{\frac{(2/\alpha) \ln n - \ln(n/m) + \ln(2/\alpha m)}{2/\alpha m}} \sim \sqrt{(1 - \alpha/2) m \ln n}.$$

□

2.3 Proof of Lemma 2.2

Before proving Lemma 2.2 we collect some preliminary results.

Lemma 2.6 *Let h be a positive integer. Let $\{G(n)\}$ be a sequence of binomial random intersection graphs $G(n) = G(n, m, p)$, were $m = m(n)$ and $p = p(n)$ satisfy $pn^{1/(h-1)}m^{1/2} \to a \in \{0, 1\}$. Then*

$$\mathbb{P}(G \text{ contains a rainbow } \mathcal{K}_h) \to a.$$

Proof The case $a = 1$ follows from Claim 2 of [13]. For the case $a = 0$ we have, by the first moment method,

$$\begin{aligned}\mathbb{P}(G \text{ contains a rainbow } \mathcal{K}_h) &\le \binom{n}{h}(m)_{\binom{h}{2}} p^{2\binom{h}{2}} \\ &\le \left(n^{1/(h-1)}m^{1/2}p\right)^{h(h-1)} \to 0.\end{aligned}$$

□

Next is an upper bound for the size $\omega'(G)$ of the largest monochromatic clique.

Lemma 2.7 *Let $1 < \alpha < 2$. Assume that $\{G(n)\}$ is a sequence of random intersection graphs satisfying (1), (2). Suppose that for some $\beta > \alpha - 1$ we have $m = \Omega(n^\beta)$. Then there is a constant $\delta > 0$ such that $\omega'(G(n)) \le n^{1-\alpha/2-\delta}$ whp.*

Proof Let $X = X(n)$ and $Y = Y(n)$ be defined as in (1). Since for any $w \in W$ and $v \in V$

$$\mathbb{P}(w \in S_v) = \sum_{k=0}^{\infty} \frac{k}{m}\mathbb{P}(|S_v| = k) = \frac{\mathbb{E}\,X}{m} = \frac{\mathbb{E}\,Y}{\sqrt{mn}},$$

and the number of elements of the set $T_v = \{v : w \in S_v\}$ is binomially distributed

$$|T_w| \sim Binom\left(n, \frac{\mathbb{E}\,Y}{\sqrt{mn}}\right), \tag{15}$$

we have, for any positive integer k

$$\mathbb{P}(|T_w| \ge k) \le \binom{n}{k}\left(\frac{\mathbb{E}\,Y}{\sqrt{mn}}\right)^k \le \left(\frac{en}{k}\right)^k\left(\frac{\mathbb{E}\,Y}{\sqrt{mn}}\right)^k \le \left(\frac{c_1}{k}\sqrt{\frac{n}{m}}\right)^k$$

for $c_1 = e\sup_n \mathbb{E}\,Y$. Therefore, by the union bound,

$$\mathbb{P}\left(\omega'(G(n)) \ge k\right) \le m\left(\frac{c_1}{k}\sqrt{\frac{n}{m}}\right)^k.$$

Fix δ with $0 < \delta < \min((\beta - \alpha + 1)/4, 1 - \alpha/2, \beta/2)$. We have

$$\begin{aligned}&\mathbb{P}\left(\omega'(G(n)) \ge n^{1-\alpha/2-\delta}\right) \le m\left(c_1 n^{\alpha/2-1/2+\delta}m^{-1/2}\right)^{\lceil n^{1-\alpha/2-\delta}\rceil} \\ &= m^{1-(\delta/\beta)\lceil n^{1-\alpha/2-\delta}\rceil}\left(c_1 n^{\alpha/2-1/2+\delta}m^{-1/2+\delta/\beta}\right)^{\lceil n^{1-\alpha/2-\delta}\rceil} \to 0\end{aligned}$$

since $m \to \infty$, $n^{1-\alpha/2-\delta} \to \infty$ and $m = \Omega(n^\beta)$ implies

$$n^{\alpha/2-1/2+\delta} m^{-1/2+\delta/\beta} \to 0.$$

□

The last and the most important fact we need relates the maximum clique size with the maximum rainbow clique size in an intersection graph. An edge-colouring of a graph is called t-good if each colour appears at most t times at each vertex. We say that an edge-coloured graph contains a rainbow copy of H if it contains a subgraph isomorphic to H with all edges receiving different colours.

Lemma 2.8 ([1]) *There is a constant c such that every t-good coloured complete graph on more than $\frac{cth^3}{\ln h}$ vertices contains a rainbow copy of $\mathcal{K}_h$.*

Proof of Lemma 2.2 Fix an integer $h > 1 + \frac{1}{\epsilon_1}$ and denote $t = n^{1-\alpha/2-\delta}$ and $k = \lceil \frac{cth^3}{\ln h} \rceil$, where positive constants δ and c are from Lemmas 2.7 and 2.8, respectively.

We first show that

$$\mathbb{P}(G_0 \text{ contains a rainbow } \mathcal{K}_h) = o(1). \tag{16}$$

We note that for the binomial intersection graph $\tilde{G} = G(n, m, p)$ with $p = p(n) = m^{-1/2} n^{-\epsilon_1} + m^{-2/3}$ Lemma 2.6 implies

$$\mathbb{P}(\tilde{G} \text{ contains a rainbow } \mathcal{K}_h) = o(1). \tag{17}$$

Let $\tilde{S}_v$ (respectively S_v), $v \in V$, denote the random subsets prescribed to vertices of $\tilde{G}$ (respectively $G(n)$). Given the set sizes $|S_v|, |\tilde{S}_v|$, $v \in V$, satisfying $|\tilde{S}_v| > \theta$, for each v, we couple the random sets of G_0 and $\tilde{G}$ so that $S_v \subseteq \tilde{S}_v$, for all $v \in V_0$. Now G_0 becomes a subgraph of $\tilde{G}$ and (16) follows from (17) and the fact that $\min_v |\tilde{S}_v| > \theta$ whp, see (12).

Next, we colour every edge $x \sim y$ of G_0 by an arbitrary element of $S_x \cap S_y$ and observe that the inequality $\omega'(G(n)) \le t$ (which holds with probability $1 - o(1)$, by Lemma 2.7) implies that the colouring obtained is t-good. Furthermore, by Lemma 2.8, every k-clique of G_0 contains a rainbow clique; however the probability of the latter event is negligibly small by (16). We conclude that $\mathbb{P}(\omega(G_0) \ge k) = o(1)$ thus proving the lemma. □

2.4 Proof of Lemma 2.3

We start with a combinatorial lemma which is of independent interest.

Lemma 2.9 *Given positive integers $a_1, \ldots, a_k$, let $\{A_1, \ldots, A_k\}$ be a family of subsets of $[m]$ of sizes $|A_i| = a_i$. Let $d \ge k$ and let S be a random subset of $[m]$ of size d. Suppose that $a_1 + \cdots + a_k \le m$. Then the probability*

$$\mathbb{P}\left(\{S \cap A_1, \ldots, S \cap A_k\} \text{ has a system of distinct representatives}\right) \tag{18}$$

is maximised when $\{A_i\}$ are mutually disjoint.

Proof Call any of $\binom{m}{d}$ possible outcomes c for S a configuration. Given $\mathcal{F} = \{A_1, \ldots, A_k\}$ let $\mathcal{C}_{DR}(\mathcal{F})$ be the set of configurations c such that $c \cap \mathcal{F} = \{c \cap A_1, \ldots, c \cap A_k\}$ has a system of distinct representatives. Write

$$p(\mathcal{F}) = \sum_{1 \le i < j \le k} |A_i \cap A_j|.$$

Suppose the claim is false. Out of all families that maximize (18) pick a family $\mathcal{F}$ with smallest $p(\mathcal{F})$. Then $p(\mathcal{F}) > 0$ and we can assume that there is an element $x \in [m]$ such that $x \in A_1 \cap A_2$. Since $\sum_{i=1}^{k} |A_i| \le m$, there is an element y in the complement of $\bigcup_{A \in \mathcal{F}} A$.

Define $A_1' = (A_1 \setminus \{x\}) \cup \{y\}$ and consider the family $\mathcal{F}' = \{A_1', A_2, \ldots, A_k\}$. Observe that the family of configurations $\mathcal{C} = \mathcal{C}_{DR}(\mathcal{F}) \setminus \mathcal{C}_{DR}(\mathcal{F}')$ has the following property: for each $c \in \mathcal{C}$ we have $x \in c$ and it is not possible to find a set of distinct representatives for $c \cap \mathcal{F}$ where A_1 is matched with an element other than x (indeed such a set of distinct representatives, if existed, would imply $c \in \mathcal{C}_{DR}(\mathcal{F}')$). Consequently, there is a set of distinct representatives for sets $c \cap A_2, \ldots, c \cap A_k$ which does not use x. Since the latter set of distinct representatives together with y is a set of distinct representatives for $c \cap \mathcal{F}'$, we conclude that $c \notin \mathcal{C}_{DR}(\mathcal{F}')$ implies $y \notin c$.

Now, for $c \in \mathcal{C}$, let $c_{xy} = (c \cup \{y\}) \setminus \{x\}$ be the configuration with x and y swapped. Then $c_{xy} \notin \mathcal{C}_{DR}(\mathcal{F})$ and $c_{xy} \in \mathcal{C}_{DR}(\mathcal{F}')$, because $y \in c_{xy}$ and can be matched with A_1. Thus each configuration $c \in \mathcal{C}$ is assigned a unique configuration $c_{xy} \in \mathcal{C}_{DR}(\mathcal{F}') \setminus \mathcal{C}_{DR}(\mathcal{F})$. This shows that $|\mathcal{C}_{DR}(\mathcal{F}')| \ge |\mathcal{C}_{DR}(\mathcal{F})|$. But $p(\mathcal{F}') \le p(\mathcal{F}) - 1$, which contradicts our assumption about the minimality of $p(\mathcal{F})$. $\square$

The next lemma is a version of a result of Erdős and Rényi about the maximum clique of the binomial random graph $G(n,p)$ (see, e.g., [12]).

Lemma 2.10 *Let $n \to +\infty$. Assume that probabilities $p_n \to 1$. Let $\{r_n\}$ be a positive sequence, satisfying $r_n = o(\tilde{K}^2)$, where $\tilde{K} = \frac{2 \ln n}{1 - p_n}$.*

There are positive sequences $\{\delta_n\}$ and $\{\epsilon_n\}$ converging to zero, such that $\delta_n \tilde{K} \to +\infty$ and for any sequence of non-random graphs $\{R_n\}$ with $V(R_n) = [n]$ and $e(R_n) \le r_n$ the number X_n of cliques of size $\lfloor \tilde{K}(1 + \delta_n) \rfloor$ in $G(n, p_n) \vee R_n$ satisfies

$$\mathbb{E}\, X_n \le \epsilon_n.$$

Proof Write $p = p_n, r = r_n$ and $h = 1 - p$. Pick a positive sequence $\delta = \delta_n$ so that $\delta_n \to 0$ and $\ln^{-1} n + h + \frac{r}{\tilde{K}^2} = o(\delta)$. Let $a = \left\lfloor \tilde{K}(1 + \delta) \right\rfloor$. We have

$$\mathbb{E}\, X_n \le \binom{n}{a} p^{\binom{a}{2} - r} \le \left(\frac{en}{a}\right)^a p^{\frac{a(a-1)}{2} - r} = e^{aB}, \tag{19}$$

where, by the inequality $\ln p \le -h$, for n large enough,

$$\begin{aligned} B &\le \ln(en/a) - \left(\frac{a-1}{2} - \frac{r}{a}\right) h \\ &\le \ln n - \frac{ah}{2} + \frac{rh}{a} \le (-1 + o(1)) \delta \ln n \to -\infty. \end{aligned}$$

□

Lemma 2.11 *Let $\{G(n)\}$ be a sequence of binomial random intersection graphs, where $m = m_n \to +\infty$ and $p = p_n \to 0$ as $n \to +\infty$. Let $\{r_n\}$ be a sequence of positive integers. Denote $\bar{K} = 2e^{mp^2} \ln n$. Assume that $r_n \ll \bar{K}^2$ and*

$$mp^2 \to +\infty, \qquad \ln n \ll mp, \qquad \bar{K}p \to 0, \qquad \bar{K} \le n/2. \tag{20}$$

There are positive sequences $\{\epsilon_n\}, \{\delta_n\}$ converging to zero such that $\delta_n \bar{K} \to +\infty$ and for any non-random graph sequence $\{R_n\}$ with $V(R_n) = V(G(n))$ and $e(R_n) \le r_n$

$$\mathbb{P}\left(Rainbow(G(n), R_n, \bar{K}(1+\delta_n))\right) \le \epsilon_n, \qquad n = 1, 2, \dots \tag{21}$$

Here we choose $\{\delta_n\}$ such that $\bar{K}(1+\delta_n)$ were an integer.

Proof Let $\{x_n\}$ be a positive sequence such that

$$px_n \to 0, \quad x_n \ll mp \quad \text{and} \quad \sqrt{mp \ln n} \ll x_n$$

(one can take, e.g., $x_n = \varphi_n \sqrt{mp \ln n}$, with $\varphi_n \uparrow +\infty$ satisfying $\varphi_n^2 \bar{K} p \to 0$).

Given n, we truncate the random sets S_v, prescribed to vertices $v \in V$ of the graph $G = G(n, m, p)$, to the size $M = \lfloor mp + x_n \rfloor$. Denote

$$\bar{S}(v) = \begin{cases} S_v, \text{ if } |S_v| \le M, \\ M \text{ element random subset of } S_v, \text{ otherwise.} \end{cases}$$

We remark that for the event $B = \{S_v = \bar{S}_v, \forall v \in V\}$ Chernoff's bound implies

$$\mathbb{P}(B) = 1 - o(1). \tag{22}$$

Now, let $t \in [K; 2K]$ and let $T = \{u_1, \dots, u_t\}$ be a subset of V of size t. By R_T we denote the subgraph of R_n induced by the vertex set T. Given $i \in \{1, \dots, t\}$, let $T_i \subseteq \{u_1, \dots u_{i-1}\}$ denote the subset of vertices which are not adjacent to v_i in R_n. Let $A_T(i)$ denote the event that sets $\{\bar{S}_u \cap S_{u_i},\ u \in T_i\}$ have distinct representatives (in particular, none of the sets is empty). Furthermore, let A_T denote the event that all $A_T(i)$, $1 \le i \le t$ hold simultaneously

$$A_T = \bigcap_{i=1}^{t} A_T(i).$$

We shall prove below that whenever n is large enough

$$\mathbb{P}(A_T) \le \left(1 - (1-p)^M\right)^{\binom{t}{2} - e(R_T)}. \tag{23}$$

Next, proceeding as in Lemma 2.10 we find positive sequences $\{\delta_n'\}$, $\{\epsilon_n'\}$ converging to zero such that the number X_n' of subsets $T \subseteq V$ of size

$$a' = \left\lfloor \frac{2 \ln n}{(1-p)^M} (1 + \delta_n') \right\rfloor$$

that satisfy the event A_T has expected value $\mathbb{E}\, X'_n \le \epsilon'_n$. For this purpose, we apply (19) to a' and $p' = 1-(1-p)^M$, and use (23). We remark that $a' = \bar{K}(1+\delta''_n)$, where $\{\delta''_n\}$ converges to zero and $\delta''\bar{K} \to +\infty$. Indeed, we have $\delta'_n \ln n/(1-p)^M \to +\infty$, by Lemma 2.10, and we have $(1-p)^M = e^{-mp^2 - O(px+mp^3)}$ with $px + mp^3 = o(1)$. In particular, for large n, we have $a' \in [\bar{K}, 2\bar{K}]$.

The key observation of the proof is that events B and $Rainbow(G, R_n, a')$ imply $X'_n > 0$. Hence,

$$\mathbb{P}(Rainbow(G, R_n, a') \cap B) \le \mathbb{P}(X'_n > 0) \le \mathbb{E}\, X'_n \le \epsilon'_n.$$

In the last step we used Markov's inequality. Finally, invoking (22) we obtain (21).

It remains to show (23). We write

$$\mathbb{P}(A_T) = \prod_{i=1}^{t} \mathbb{P}\left(A_T(i) | A_T(1), \ldots, A_T(i-1)\right)$$

and evaluate, for $1 \le i \le t$,

$$\mathbb{P}(A_T(i) | A_T(1), \ldots, A_T(i-1)) \le \left(1 - (1-p)^M\right)^{|T_i|}. \tag{24}$$

Now (23) follows from the simple identity $\sum_{1 \le i \le t} |T_i| = \binom{t}{2} - e(R_T)$. Let us prove (24). For this purpose we apply Lemma 2.9. We first condition on $\{\bar{S}_u,\ u \in T_i\}$ and the size $|S_{v_i}|$ of S_{v_i}. By Lemma 2.9 the conditional probability

$$\mathbb{P}(A_T(i) \,\big|\, \bar{S}_u,\ u \in T_i,\ |S_{v_i}|)$$

is maximized when the sets $\bar{S}_u$, $u \in T_i$ are mutually disjoint (at this step we check the condition of Lemma 2.9 that $\sum_{u \in T_i} |\bar{S}_u| \le tM < m$, for large n). Secondly, we drop the conditioning on $|S_{v_i}|$ and allow S_{v_i} to choose its element independently at random with probability p. In this way we obtain (24). $\square$

Lemma 2.12 *Let $\{G(n)\}$ be a sequence of random binomial intersection graphs, where $m = m(n) \to +\infty$ and $p = p(n) \to 0$ as $n \to +\infty$. Assume that*

$$np = O(1), \qquad m(np)^3 \ll \bar{K}^2,$$

where $\bar{K} = 2e^{mp^2} \ln n$. Assume, in addition, that (20) holds.

Then there is a sequence $\{\delta_n\}$ converging to zero such that $\delta_n \bar{K} \to +\infty$ and

$$\mathbb{P}\left(\omega(G(n)) > \bar{K}(1+\delta_n)\right) \to 0.$$

Proof Given n, let U be a random subset of $V = V(G(n))$ with binomial number of elements $|U| \sim Bin(n,p)$ and such that, for any $k = 0, 1, \ldots$, conditionally, given the event $|U| = k$, the subset U is uniformly distributed over the class of subsets of V of size k. Recall that $T_w \subseteq V$ denotes the set of vertices that have chosen an attribute $w \in W$. We remark that T_w, $w \in W$ are iid random subsets having the same probability distribution as U.

We call an attribute w *big* if $|T_w| \ge 3$, otherwise w is *small*. Let W_B and W_S denote the sets of big and small attributes. Denote by G_B (respectively, G_S) the subgraph of $G = G(n)$ consisting of edges covered by big (respectively, small) attributes. We observe that, given G_B, the random sets T_z, $z \in W_S$, defining the edges of G_S are (conditionally) independent. We are going to replace them by bigger sets, denoted T'_z, by adding some more elements as follows. Given T_z, we first generate independent random variables $\mathbb{I}_z$ and $|\Delta_z|$, where $\mathbb{I}_z$ has Bernoulli distribution with success probability $p' = \mathbb{P}(|U| \le 2)$ and where $\mathbb{P}(|\Delta_z| = k) = \mathbb{P}(|U| = k)/(1 - p')$, $k = 3, 4, \dots$. Secondly, for $\mathbb{I}_z = 1$ we put $T'_z = T_z$. Otherwise we put $T'_z = T_z \cup \Delta_z$, where Δ_z is a subset of $V \setminus T_z$ of size $|\Delta_z| - |T_z| \ge 1$ drawn uniformly at random. We note that given G_B, the random sets T'_z, $z \in W_S$ are (conditionally) independent and have the same probability distribution as U. Next we generate independent random subsets T'_w of V, for $w \in W_B$, so that they have the same distribution as U and were independent of G_S, G_B and T'_z, $z \in W_S$. Given G_B, the collection of random sets $\{T'_w, w \in W_B \cup W_S\}$ defines the binomial random intersection graph G' having the same distribution as $G(n, m, p)$.

We remark that $G_S \subseteq G'$ and every edge of G_S can be assigned a unique small attribute that covers this edge and the assigned attributes are all different. On the other hand, the graph G_B is relatively small. Indeed, since each w covers $\binom{|T_w|}{2}$ edges, the expected number of edges of G_B is at most

$$\mathbb{E} \sum_{w \in W} \binom{T_w}{2} \mathbb{I}_{\{|T_w| \ge 3\}} = m \mathbb{E} \binom{T_w}{2} \mathbb{I}_{\{|T_w| \ge 3\}} \le m \sum_{k \ge 3} \binom{k}{2} \binom{n}{k} p^k.$$

Invoking the simple bound

$$\sum_{k \ge 3} \binom{k}{2} \binom{n}{k} p^k \le (np)^2 (e^{np} - 1)/2 = O((np)^3)$$

we obtain $\mathbb{E}\, e(G_B) = O(m(np)^3)$.

Now we choose an integer sequence $\{r_n\}$ such that $m(np)^3 \ll r_n \ll \bar{K}^2$ and write, for an integer $K' > 0$,

$$\mathbb{P}\left(\omega(G) \ge K'\right) \le \mathbb{E}\, \mathbb{P}\left(\omega(G) \ge K' | G_B\right) \mathbb{I}_{\{e(G_B) \le r_n\}} + \mathbb{P}\left(e(G_B) \ge r_n\right). \qquad (25)$$

Here, by Markov's inequality, $\mathbb{P}(e(G_B) \ge r_n) \le r_n^{-1} \mathbb{E}\, e(G_B) = o(1)$. Furthermore, we observe that $\omega(G) \ge K'$ implies the event $Rainbow(G', G_B, K')$. Hence,

$$\mathbb{P}\left(\omega(G) \ge K' | G_B\right) \le \mathbb{P}\left(Rainbow(G', G_B, K') | G_B\right).$$

We choose $K' = \bar{K}(1 + \delta_n)$ and apply Lemma 2.11 to the conditional probability on the right. At this point we specify $\{\delta_n\}$ and find $\epsilon_n \downarrow 0$ such that $\mathbb{P}\left(Rainbow(G', G_B, K') | G_B\right) \le \epsilon_n$ uniformly in G_B satisfying $e(G_B) \le r_n$. Hence, (25) implies $\mathbb{P}\left(\omega(G) \ge \bar{K}(1 + \delta_n)\right) \le \epsilon_n + o(1) = o(1)$. □

Now we are ready to prove Lemma 2.3.

Proof of Lemma 2.3 Let

$$0 < \epsilon < 2^{-1} \min\{1,\, 1 - 2^{-1}\alpha,\, \beta - 2 + \alpha - 6\alpha\epsilon_1\} \tag{26}$$

and let $\bar{G}_1$ be the subgraph of G_1 induced by vertices $v \in V_1$ with $X_v \leq \theta$. Here $\theta^2 = (1 - \varepsilon - 2^{-1}\alpha)m \ln n$. Let $D = |V(G_1) \setminus V(\bar{G}_1)|$ denote the number of vertices of G_1 that do not belong to $\bar{G}_1$.

To prove the lemma we write $\omega(G_1) \leq D + \omega(\bar{G}_1)$ and show that each summand on the right is of order $o_P(K)$ for appropriately chosen $\epsilon = \epsilon(n) \to 0$.

Using (2) and Lemma 2.5 we estimate the expected value of D for $n \to +\infty$

$$\mathbb{E}\, D = n\left(\mathbb{P}(X_v \geq \theta) - \mathbb{P}(X_v \geq \theta_2)\right) \leq (h(\epsilon) + o(1))K. \tag{27}$$

Here $h(\epsilon) := (1 - \epsilon - 2^{-1}\alpha)^{-\alpha/2} - (1 - 2^{-1}\alpha)^{-\alpha/2} \to 0$ as $\varepsilon \to 0$. Letting $\epsilon \to 0$ we obtain from (27) that $D = o_P(K)$.

We complete the proof by showing that for any ε satisfying (26)

$$\mathbb{P}\left(\omega(\bar{G}_1) \geq 4n^{1-2^{-1}\epsilon-2^{-1}\alpha} \ln n\right) = o(1). \tag{28}$$

Note that $n^{1-2^{-1}\epsilon-2^{-1}\alpha} \ln n \ll K$.

Let $\bar{N}$ be a binomial random variable, $\bar{N} \sim Bin(n, \mathbb{P}(X_v > \theta_1))$, and let

$$\bar{n} = (1+\epsilon)n^{1-2^{-1}\alpha+\alpha\epsilon_1} L(n^{0.5-\epsilon_1}) \qquad \text{and} \qquad \bar{p}^2 = (1 - 2^{-1}\epsilon - 2^{-1}\alpha)m^{-1} \ln n.$$

We couple $\bar{G}_1$ with the binomial random intersection graph $G' = G(\bar{n}, m, \bar{p})$ so that the event that $\bar{G}_1$ is isomorphic to a subgraph of G', denoted $\bar{G}_1 \subseteq G'$, has probability

$$\mathbb{P}(\bar{G}_1 \subseteq G') = 1 - o(1). \tag{29}$$

We argue that such a coupling is possible because the events $A = \{$every vertex of G' is prescribed at least θ attributes$\}$ and $B = \{|V(\bar{G}_1)| \leq \bar{n}\}$ have very high probabilities. Indeed, the bound $\mathbb{P}(A) = 1 - o(1)$ follows from Chernoff's inequality (12). To get the bound $\mathbb{P}(B) = 1 - o(1)$ we first couple binomial random variables $|V(\bar{G}_1)| \sim Bin(n, \mathbb{P}(\theta_1 < X_v < \theta))$ and $\bar{N}$ so that $\mathbb{P}(|V(\bar{G}_1)| \leq \bar{N}) = 1$ and then invoke the bound $\mathbb{P}(\bar{N} \leq \bar{n}) = 1 - o(1)$, which follows from Chernoff's inequality.

Next we apply Lemma 2.12 to G' and obtain the bound

$$\mathbb{P}\left(\omega(G') > 4n^{1-2^{-1}\epsilon-2^{-1}\alpha} \ln \bar{n}\right) = o(1), \tag{30}$$

which together with (29) implies (28). □

3 Finite variance

In this section we prove Theorem 1.2. We note that the random power-law graph studied by Janson, Łuczak and Norros [11] whp does not contain $\mathcal{K}_4$ as a subgraph if the degree distribution has a finite second moment. In our case a similar result holds for the rainbow $\mathcal{K}_4$. Given a sequence of random intersection graphs $\{G(n)\}$, we show that the number of rainbow $\mathcal{K}_4$ subgraphs of $G(n)$ is stochastically bounded as $n \to +\infty$ provided that the sequence of the second moments of the degree distributions is bounded. If, in addition, the sequence of degree distributions is uniformly square integrable, then $G(n)$ has no rainbow $\mathcal{K}_4$ whp, see Lemma 3.3 below. We use these observations in the proof of Theorem 1.2.

3.1 Large cliques and rainbow $\mathcal{K}_4$

Let U be a finite set and let $\mathcal{C} = \{C_1, \dots, C_r\}$ be a collection of (not necessarily distinct) subsets of U. We consider the complete graph $\mathcal{K}_U$ on the vertex set U and interpret subsets C_i as colours: an edge $x \sim y$ receives colour C_i (or just i) whenever $\{x, y\} \subseteq C_i$. We call $\mathcal{C}$ a *clique cover* if every edge of the clique $\mathcal{K}_U$ receives at least one colour. The edges spanned by the vertex set C_i form a subclique, which we call the *monochromatic clique* of colour i. We say that a vertex set $S \subseteq U$ is a *witness of a rainbow clique* if every edge of the clique $\mathcal{K}_S$ induced by S receives a non-empty collection of colours and it is possible to assign each edge one of its colours so that all edges of $\mathcal{K}_S$ were assigned different colours. For example, the collection $\mathcal{C} = \{A, B, C\}$, where $A = \{1, 2, 3\}$, $B = \{1, 3, 4\}$ and $C = \{2, 4, 3\}$ is a clique cover of the set $\{1, 2, 3, 4\}$. It produces three monochromatic triangles and four rainbow triangles.

We start with a result that relates clique covers to rainbow clique subgraphs. For a clique cover $\mathcal{C} = \{C_1, \dots, C_r\}$ denote by $p(\mathcal{C}) = \max_{i \neq j} |C_i \cap C_j|$ the size of maximum pairwise intersection.

Lemma 3.1 *Let k and p be positive integers. Let $h = h(k) > 0$ denote the smallest integer such that $\binom{h}{4} \geq k$. Let $\mathcal{C}$ be a clique cover of a finite set U and assume that $\max_{C \in \mathcal{C}} |C| \geq |U| - h$ and $p(\mathcal{C}) \leq p$.*

If, in addition, $|U| \geq t(k, p)$, where $t(k, p) = c \frac{h^3}{\ln h} p \left(\sqrt{2k} + 5 + 2p \right)$, then $\mathcal{C}$ produces at least k witnesses of rainbow $\mathcal{K}_4$. Here c is the absolute constant of Lemma 2.8.

Proof Write $b = \max_i |C_i|$. We note that $\mathcal{C}$ has no rainbow $\mathcal{K}_h$ since otherwise there would be at least $\binom{h}{4} \geq k$ copies of rainbow $\mathcal{K}_4$. Observe, that every monochromatic subclique of $\mathcal{K}_U$ has at most b vertices. Hence, each colour appears at most $b - 1$ times at each vertex of $\mathcal{K}_U$. By Lemma 2.8, $\mathcal{K}_U$ has at most $c(b-1)h^3 / \ln h$ vertices. That is, $b > a|U|$, where $a = \frac{\ln h}{ch^3}$ and c is an absolute constant. Fix $B \in \mathcal{C}$ with $|B| = b$ and a subset $S \subseteq U \setminus B$ of size h, say $S = \{x_1, \dots, x_h\}$. Here we use the assumption $|U| \geq b + h$ telling that $U \setminus B$ has at least h elements, $|U \setminus B| = |U| - b \geq h$. We remark, that at least one pair of vertices of S, say $\{x_1, x_2\}$, receives at most 5 colours (it is covered by at most 5 sets from $\mathcal{C}$). Indeed, otherwise every edge of $\mathcal{K}_S$ received at least 6 distinct colours and, thus, each $S' \subseteq S$ of size $|S'| = 4$ induced a rainbow $\mathcal{K}_4$. This contradicts to our assumption that there are fewer than $k \leq \binom{h}{4}$ rainbow copies of $\mathcal{K}_4$.

We observe that the set of colours received by the pair $\{x_1, x_2\}$ is non-empty (since $\mathcal{C}$ is a clique cover) and fix one such colour, say $C_{x_1,x_2} \in \mathcal{C}$. Now, consider the set of pairs $\{\{x_1, y\}, y \in B\}$ and pick a smallest family of sets from $\mathcal{C}$ such that each pair were covered by a member of the family (the smallest family means that any other family with fewer members would leave at least one uncovered pair). Since each member of the family intersects with B in at most p vertices (condition of the lemma) we conclude that such a family contains at least $\lceil b/p \rceil$ members. Furthermore, since the family is minimal, every member covers a pair $\{x_1, y\}$ which is not covered by other members. Hence, we can pick a set $B_1 \subseteq B$ of size $\lceil b/p \rceil$ so that every $\{x_1, y\}$, $y \in B_1$ is covered by a unique member, say $C_{x_1,y}$, of the family.

Next, remove from B_1 the elements y such that $x_2 \in C_{x_1,y}$ (there are at most 5 of them). Then remove those elements y which belong to the set C_{x_1,x_2} (there are at most p of them, since $|C_{x_1,x_2} \cap B| \le p$). Call the newly formed set B'. Notice that

$$b' := |B'| \ge \frac{b}{p} - 5 - p > \frac{a|U|}{p} - 5 - p.$$

Let us consider the clique $\tilde{K}$ on the vertex set $B' \cup \{x_1, x_2\}$. For $y \in B'$, colour each edge $\{x_1, y\}$ of $\tilde{K}$ with the colour $C_{x_1,y}$. Colour the edge $\{x_1, x_2\}$ with C_{x_1,x_2} and for every edge $\{y_i, y_j\} \in B'$ use the colour B. Finally, for $y \in B'$, assign $\{x_2, y\}$ an arbitrary colour from the set of colours received by $\{x_2, y\}$ from the clique cover $\mathcal{C}$.

We claim that for any $y_1 \in B'$ and any $y_2 \in B' \setminus C_{x_2,y_1}$, the set $\{x_1, x_2, y_1, y_2\}$ witnesses a rainbow $\mathcal{K}_4$. Indeed, by the construction, the colour C_{x_1,x_2} of the edge $\{x_1, x_2\}$ occurs only once, because $B' \cap C_{x_1,x_2} = \emptyset$. Similarly, for $x_1, x_2 \notin B$, the colour B of $\{y_1, y_2\}$ occurs only once. The colours of the two other edges incident to x_1 occur only once, since we removed all candidates y such that $x_2 \in C_{i_{x_1,y}}$, while constructing the set B'. Finally, we have $C_{x_2,y_1} \ne C_{x_2,y_2}$ since we chose y_2 outside C_{x_2,y_1}.

How many such witnesses can we form? For any y_1 we choose $|B'| - |B' \cap C_{x_2,y_1}| \ge |B'| - p$ suitable y_2. Repeating this for each y_1 we will produce every 4-set at most twice. Therefore $\tilde{K}$ contains at least

$$\frac{b'(b'-p)}{2} \ge \frac{1}{2}\left(\frac{a|U|}{p} - 5 - 2p\right)^2 \tag{31}$$

witnesses of rainbow $\mathcal{K}_4$. But since the total number of witnesses of rainbow $\mathcal{K}_4$ produced by $\mathcal{C}$ is less that k, the right-hand side of (31) is less than k. We obtain the inequality

$$|U| < \frac{p}{a}\left(\sqrt{2k} + 5 + 2p\right) = t(k, p),$$

which contradicts to the condition $|U| \ge t(k, p)$. □

In the remaining part of the subsection 3.1 we interpret attributes $w \in W$ as colours assigned to edges of a random intersection graph.

Lemma 3.2 *Let $G = G(k, m, P)$ be a random intersection graph and let $X_1, \dots, X_k$ denote the sizes of random sets defining G. For any integers $x_1, \dots, x_k$ such that* *the event* $B = \{X_1 = x_1, \ldots, X_k = x_k\}$ *has positive probability, we have*

$$\mathbb{P}(G \text{ has a rainbow } \mathcal{K}_k | B) \le m^{-\frac{k(k-1)}{2}} (x_1 x_2 \ldots x_k)^{k-1}.$$

Proof Our intersection graph produces a rainbow clique on its k vertices whenever for some injective mapping, say f, from the set of pairs of vertices to the set of attributes, the event $A_f = \{\text{every pair } \{x,y\} \text{ is covered by } f(\{x,y\})\}$ occurs. By the independence, $\mathbb{P}(A_f|B) = \prod_i \frac{(x_i)_{k-1}}{(m)_{k-1}}$. Since there are $(m)_{\binom{k}{2}}$ possibilities to choose the map f, we obtain, by the union bound,

$$\mathbb{P}(G \text{ has a rainbow } \mathcal{K}_k | B) \le (m)_{\binom{k}{2}} \prod_i \frac{(x_i)_{k-1}}{(m)_{k-1}} \le \frac{(x_1 x_2 \ldots x_k)^{k-1}}{m^{k(k-1)/2}}.$$

□

Lemma 3.3 *Let* $\{G(n)\}$ *be a sequence of random intersection graphs such that* $\mathbb{E}\, Y(n)^2 = O(1)$*. Then the number* $R = R(n)$ *of 4-sets* $S \subseteq V(G(n))$ *that witness a rainbow* $\mathcal{K}_4$ *in* $G(n)$ *satisfies as* $n \to +\infty$

$$\mathbb{E}\, R \le \frac{(\mathbb{E}\, Y^2)^4}{4!} = O(1).$$

Furthermore, if for some positive sequence $\epsilon_n \to 0$ *we have* $n\mathbb{P}(Y(n) \ge \epsilon_n n^{1/2}) \to 0$ *then* $G(n)$ *does not contain a rainbow* $\mathcal{K}_4$ *whp.*

Proof of Lemma 3.3 Denote $X_v = |S_v(n)|$ and $Y = Y(n)$. We write, using symmetry and the bound of Lemma 3.2,

$$\mathbb{E}\, R = \sum_{S \subseteq V, |S|=4} \mathbb{P}(S \text{ witnesses a rainbow } \mathcal{K}_4) \le \binom{n}{4} \mathbb{E}\left(\frac{(X_1 X_2 X_3 X_4)^3}{m^6} \wedge 1 \right).$$

Next, we apply the simple inequality $a^6 \wedge 1 \le a^4$ and bound the right-hand side from above by $\frac{n^4}{4!} \frac{\mathbb{E}\,(X_1 X_2 X_3 X_4)^2}{m^4} = \frac{(\mathbb{E}\, Y^2)^4}{4!}$.

For the second part of the lemma, let $b = b(n) = \epsilon_n \sqrt{m}$ and let $A = A(n)$ be the event that $\max_{v \in V} X_v \le b$. Let $\bar{A}$ denote the complement event. We write

$$\mathbb{P}(R \ge 1) \le \mathbb{P}(R \ge 1, A) + \mathbb{P}(\bar{A}) \le \mathbb{E}\, R\mathbb{I}_A + \mathbb{P}(\bar{A}). \tag{32}$$

By the union bound the second term is at most

$$n\mathbb{P}(X > b) = n\mathbb{P}(Y > \epsilon_n n^{1/2}) \to 0.$$

The first term by Lemma 3.2 satisfies

$$\mathbb{E}\, R\mathbb{I}_A \le \binom{n}{4} m^{-6} \mathbb{E}\,(X_1 X_2 X_3 X_4)^3 \mathbb{I}_A \le \frac{(\mathbb{E}\, X^2)^4 n^4 b^4}{4! m^6} = (\epsilon_n \mathbb{E}\, Y^2)^4 = o(1).$$

□

The next result shows that the structure of random intersection graphs with $\mathbb{E}\, Y(n)^2 = O(1)$ is relatively simple.

Lemma 3.4 *Let $\{G(n)\}$ be a sequence of random intersection graphs. Assume that $\mathbb{E}\, Y(n)^2 = O(1)$ and $m(n) \to \infty$ as $n \to +\infty$. Then whp each pair $\{w', w''\}$ of attributes is shared by at most two vertices of $G(n)$.*

The lemma says that the intersection of any two monochromatic cliques of $G(n)$ consists of at most one edge whp.

Proof For any pair of attributes w', w'' and a vertex v of $G(n)$, we have

$$\begin{aligned}\mathbb{P}(w', w'' \in S_v) &= \sum_{k=0}^{m} \mathbb{P}(|S_v| = k) \frac{k(k-1)}{m(m-1)} = \frac{\mathbb{E}\, X^2 - \mathbb{E}\, X}{m(m-1)} \\ &\leq \frac{\mathbb{E}\, Y^2}{n(m-1)} \leq \frac{c}{nm}.\end{aligned}$$

Here $c > 0$ does not depend on m and n. By the union bound, the probability that there is a pair of attributes shared by k or more vertices is at most

$$\binom{m}{2}\binom{n}{k}\mathbb{P}(w', w'' \in S_v)^k \leq m^2 \left(\frac{en}{k}\right)^k \left(\frac{c}{nm}\right)^k \leq m^2 \left(\frac{ec}{km}\right)^k.$$

This probability tends to zero for any $k \geq 3$. □

Proof of Theorem 1.2 Let $R = R(n)$ denote the number of 4-sets $S \subseteq V(G(n))$ witnessing rainbow $\mathcal{K}_4$ in $G(n)$. By Lemma 3.4, the intersection of any two monochromatic cliques has at most 2 vertices whp. In that case, by Lemma 3.1 (applied to the set of vertices U of the largest clique) either $\omega(G(n)) < t(R+1, 2)$ or $\omega(G) \leq \omega'(G) + h(R+1)$. Thus, $\leq$

$$\omega(G(n)) \leq \omega'(G(n)) + Z(n)$$

where $Z(n) = t(R+1, 2) + h(R+1) = O_P(1)$, by Lemma 3.3.

If $n\mathbb{P}(Y(n) > \epsilon_n n^{1/2}) \to 0$ for some $\epsilon_n \to 0$ then by Lemma 3.3 $G(n)$ whp does not contain a rainbow $\mathcal{K}_4$, so whp $\omega(G) \leq t(1, 2) \vee (\omega'(G) + 3)$. □

3.2 Monochromatic cliques and balls and bins

Here we prove Theorem 1.3. In the proof we use the fact that the maximum bin load $M(N, m)$ is a "smooth" function of the first argument N, see lemma below.

Lemma 3.5 *Let $\{N_n\}$ and $\{m_n\}$ be sequences of positive integers such that $N = N_n \to \infty$ and $m = m_n \to \infty$. Let $\{\delta_n\}, \{\epsilon_n\}$ be positive sequences converging to zero such that $\epsilon_n = o(\delta_n)$. For every n there is a coupling between random variables $M' = M'_n = M(\lfloor N(1+\epsilon_n) \rfloor, m)$ and $M = M_n = M(N, m)$ such that $M \leq M'$ with probability one, and*

$$\mathbb{P}(M' - \delta_n \mathbb{E}\, M' \leq M) \to 1. \tag{33}$$

If, additionally, $\mathbb{P}(M' > \delta_n^{-1}) \to 0$, then $M = M'$ whp.

Proof Given n, we label m bins by numbers $1, \ldots, m$. Throw $\lfloor N(1+\epsilon_n)\rfloor$ balls into bins. This gives an instance of M'. Denote by L the label of the bin with the lowest index realising the maximum.

Now delete uniformly at random $\lfloor \epsilon_n N\rfloor$ balls. The configuration with the remaining N balls gives an instance of $M \leq M'$. We remark that conditionally, given M', the number Δ of balls deleted from the bin L has a hypergeometric distribution with the mean value

$$\frac{M' \times \lfloor \epsilon_n N\rfloor}{\lfloor N(1+\epsilon_n)\rfloor} \leq \epsilon_n M'.$$

Now the bin L contains $M' - \Delta \leq M$ balls and, by Markov's inequality,

$$\mathbb{P}(M' - M \geq t) \leq \mathbb{P}(\Delta \geq t) \leq t^{-1}\mathbb{E}\,\Delta \leq t^{-1}\epsilon_n \mathbb{E}\,M'.$$

Choosing $t = \delta_n \mathbb{E}\,M'$ yields (33). Similarly, if $\mathbb{P}(M' \geq \delta_n^{-1}) = o(1)$, then

$$\mathbb{P}(M' - M \geq 1) \leq \mathbb{E}\,\Delta \mathbb{I}_{M' \leq \delta_n^{-1}} + \mathbb{P}(M' > \delta_n^{-1}) \leq \epsilon_n \delta_n^{-1} + o(1) \to 0.$$

□

Proof of Remark 1.4 Suppose $m = o(n)$, $\mathbb{E}\,Y = \Theta(1)$ and $\mathbb{E}\,Y^2 = O(1)$. Since $X = X(n)$ is a non-negative integer, we have $\mathbb{E}\,X^2 \geq \mathbb{E}\,X$. But $\mathbb{E}\,X^2 = O(m/n)$ and $\mathbb{E}\,X = \Theta((m/n)^{1/2})$, so $\mathbb{E}\,X^2 = o(\mathbb{E}\,X)$, a contradiction. □

Proof of Theorem 1.3 In view of Remark 1.4 it suffices to consider the case $m = \Omega(n)$. Denote $\epsilon_n = (2 + \ln^2 n)^{-1}$ so that $\epsilon_n \ln n = o(1)$ and $n\epsilon_n^2 \to +\infty$. Given n, write $\epsilon = \epsilon_n$ and denote $\bar{N} = n\mathbb{E}\,X_1 = \sqrt{mn}\,\mathbb{E}\,Y$ and

$$\bar{N}^- = \lfloor \bar{N}(1-4\epsilon)\rfloor, \qquad \bar{N}^+ = \lceil \bar{N}(1+4\epsilon)\rceil.$$

In order to generate an instance of $G(n)$ we draw a random sample $X_1, \ldots, X_n$ from the distribution $P(n)$. Then choose random subsets $S_{v_i} \subseteq W$ of size X_i, $v_i \in V$, by throwing balls into m bins labelled $w_1, \ldots, w_m$ (the j-th bin has label w_j and index j) as follows. Keep throwing balls labelled $i = 1$ until there are exactly X_i different bins containing a ball labelled i. Do the same for $i = 2, \ldots, n$. Now, for each i, the bins containing balls labelled i make up the set S_{v_i}. In this way we obtain an instance of $G(n)$. Let X'_i denote the number of balls of label i thrown so far. Clearly, $X'_1, \ldots, X'_n$ is a sequence of independent random variables and $X'_i \geq X_i$, for each i. We stop throwing balls if the number of balls $N' = \sum_i X'_i$ at least as large as $\bar{N}^+$. Otherwise we throw additional $\bar{N}^+ - N'$ unlabelled balls into bins.

Let us inspect the bins after j balls have been thrown. Let $\mathcal{M}(j)$ denote the set of balls contained in the bin with the largest number of balls and the smallest index. We note that the number $M(j) = |\mathcal{M}(j)|$ of balls in that bin has the same distribution as $M(j, m)$ (random variable defined before Theorem 1.3).

Denote, for short, $\omega' = \omega'(G(n))$ and $\bar{M} = M(\lfloor \bar{N}\rfloor)$. We observe that the event $\mathcal{A}_1 = \{$all balls of $\mathcal{M}(N')$ have different labels$\}$ implies $\omega'(G(n)) = M(N')$.

Furthermore, if both events $\mathcal{A}_2 = \{M(\bar{N}^-) = M(\bar{N}^+)\}$ and $\mathcal{A}_3 = \{\bar{N}^- \le N' \le \bar{N}^+\}$ hold, then $\bar{M} = M(N')$. We shall show below that

$$\mathbb{P}(\mathcal{A}_r) = 1 - o(1), \qquad \text{for} \qquad r = 1, 2, 3. \tag{34}$$

Now, (34) implies $\mathbb{P}(\omega' = \bar{M}) = 1 - o(1)$ and, since the distributions of $M(\lfloor \bar{N} \rfloor, m)$ and $\bar{M}$ coincide, we obtain

$$d_{TV}\big(\omega', M(\lfloor \bar{N} \rfloor, m)\big) = d_{TV}\big(\omega', \bar{M}\big) \le \mathbb{P}\big(\omega' \ne \bar{M}\big) = o(1).$$

It remains to prove (34). Let us consider $\mathbb{P}(\mathcal{A}_3)$. We first replace X_i and X_i' by the truncated random variables

$$\tilde{X}_i = X_i \mathbb{I}_{\{X_i \le \epsilon\, m\}} \qquad \text{and} \qquad \tilde{X}_i' = X_i' \mathbb{I}_{\{X_i \le \epsilon\, m\}}, \quad 1 \le i \le n.$$

Denote $\tilde{N}' = \sum_i \tilde{X}_i'$ and introduce events $\tilde{\mathcal{A}}_3 = \{\bar{N}^- \le \tilde{N}' \le \bar{N}^+\}$ and $\mathcal{A}_4 = \{\max_{1 \le i \le n} X_i \le \epsilon\, m\}$. Let $\bar{\mathcal{A}}_4$ denote the complement of $\mathcal{A}_4$. From the relation $\mathcal{A}_3 \cap \mathcal{A}_4 = \tilde{\mathcal{A}}_3 \cap \mathcal{A}_4$ we obtain

$$\mathbb{P}(\mathcal{A}_3) \ge \mathbb{P}(\mathcal{A}_3 \cap \mathcal{A}_4) = \mathbb{P}(\tilde{\mathcal{A}}_3 \cap \mathcal{A}_4) \ge \mathbb{P}(\tilde{\mathcal{A}}_3) - \mathbb{P}(\bar{\mathcal{A}}_4).$$

Furthermore, by the union bound and Markov's inequality, we have

$$\mathbb{P}(\bar{\mathcal{A}}_4) \le n\mathbb{P}(X_1 > \epsilon\, m) \le n \frac{\mathbb{E}\, X_1^2}{\epsilon^2 m^2} = \frac{\mathbb{E}\, Y^2}{\epsilon^2 m} = o(1),$$

since $m = \Omega(n)$ and $\epsilon^2 n \to +\infty$. Hence, $\mathbb{P}(\mathcal{A}_3) \ge \mathbb{P}(\tilde{\mathcal{A}}_3) - o(1)$. Secondly, we prove that $\mathbb{P}(\tilde{\mathcal{A}}_3) = 1 - o(1)$. For this purpose we show that, for large n,

$$\bar{N}(1 - \epsilon) \le \mathbb{E}\, \tilde{N}' \le \bar{N}(1 + 2\epsilon) \quad \text{and} \quad \mathbb{P}\big(|\tilde{N}' - \mathbb{E}\, \tilde{N}'| \ge \epsilon\, \mathbb{E}\, \tilde{N}'\big) = o(1). \tag{35}$$

The proof of (35) is routine. Notice that conditionally, given $\tilde{X}_i = k$, we have $\tilde{X}_i' = \sum_{j=1}^k \xi_j$, where $\xi_1, \xi_2, \dots, \xi_k$ are independent geometric random variables with parameters

$$\frac{m}{m}, \ \frac{m-1}{m}, \ \dots, \ \frac{m-k+1}{m},$$

respectively. Since $\tilde{X}_i \le \epsilon\, m$, we only consider $k < \epsilon m$, so

$$\mathbb{E}\,(\tilde{X}_i' | \tilde{X}_i = k) = \frac{m}{m} + \frac{m}{m-1} + \dots + \frac{m}{m-k+1} \le \frac{k}{1-\epsilon} \le k(1 + 2\epsilon).$$

In the last step we used $\epsilon \le 1/2$. We conclude that

$$\tilde{X}_i \le \mathbb{E}\,(\tilde{X}_i' | \tilde{X}_i) \le \tilde{X}_i(1 + 2\epsilon). \tag{36}$$

From (36) we obtain

$$n\mathbb{E}\, \tilde{X}_1 \le \mathbb{E}\, \tilde{N}' \le (1 + 2\epsilon) n \mathbb{E}\, \tilde{X}_1. \tag{37}$$

Furthermore, invoking in (37) the inequalities $\mathbb{E}\,X_1 - s \le \mathbb{E}\,\tilde{X}_1 \le \mathbb{E}\,X_1$, where

$$s = \mathbb{E}\,X_1 \mathbb{I}_{\{X_1 > \epsilon\, m\}} \le (\epsilon\, m)^{-1} \mathbb{E}\,X_1^2 = (\epsilon\, n)^{-1} \mathbb{E}\,Y_1^2 = o(\epsilon),$$

we obtain the first part of (35). The second part of (35) follows from the inequalities $\tilde{N}' \ge N(1-\epsilon)$ and

$$Var \tilde{N}' \le 2n \mathbb{E}\,X_1^2 = 2m \mathbb{E}\,Y^2, \tag{38}$$

by Chebyshev's inequality. Let us show (38). Proceeding as in the proof of (36) we evaluate the conditional variance

$$Var(\tilde{X}_i' | \tilde{X}_i = k) = \sum_{j=1}^{k} Var(\xi_j) = \sum_{j=0}^{k-1} \frac{jm}{(m-j)^2} \le \frac{k^2}{2(1-\epsilon)^2 m} \le \frac{k^2}{m},$$

and obtain

$$\mathbb{E}\,Var(\tilde{X}_i' | \tilde{X}_i) \le \frac{\mathbb{E}\,\tilde{X}_i^2}{m}.$$

Furthermore, using (36) we write

$$Var(\mathbb{E}\,(\tilde{X}_i' | \tilde{X}_i)) \le \mathbb{E}\,(\mathbb{E}\,(\tilde{X}_i' | \tilde{X}_i))^2 \le \mathbb{E}\,\tilde{X}_i^2 (1+2\epsilon)^2 \le \mathbb{E}\,\tilde{X}_i^2 (1+8\epsilon).$$

Collecting these estimates we obtain an upper bound for the variance

$$Var(\tilde{X}_i') = \mathbb{E}\,Var(\tilde{X}_i' | \tilde{X}_i) + Var(\mathbb{E}\,(\tilde{X}_i' | \tilde{X}_i)) \le \mathbb{E}\,\tilde{X}_i^2 (1 + 8\epsilon + m^{-1}) \le 2\mathbb{E}\,X_i^2.$$

This bound implies (38). We have shown (34) for $r = 2$.

Let us prove (34) for $r = 1$. We start with an auxiliary inequality. Given integers $x_1, \dots, x_n \ge 0$ consider a collection of $k = x_1 + \cdots + x_n > 0$ labelled balls, containing x_i balls of label i, $1 \le i \le n$. The probability of the event that a random subset of r balls contains a pair of equally labelled balls is

$$\mathbb{P}(L \ge 1) \le \mathbb{E}\,L = \binom{r}{2} \binom{k}{2}^{-1} \sum_i \binom{x_i}{2} \le \left(\frac{r}{k}\right)^2 \sum_i x_i^2. \tag{39}$$

Here L counts pairs of equally labelled balls in the random subset.

We will show that $\mathbb{P}(\bar{\mathcal{A}}_1) = o(1)$. To this aim, we introduce events

$$\mathcal{A}_5 = \{M(\tilde{N}') \le \ln n\}, \qquad \mathcal{A}_6 = \{\sum_{1 \le i \le n} (\tilde{X}_i')^2 \le m \ln n\},$$

estimate

$$\mathbb{P}(\bar{\mathcal{A}}_1) \le \mathbb{P}(\bar{\mathcal{A}}_1 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6) + \mathbb{P}(\bar{\mathcal{A}}_3) + \mathbb{P}(\bar{\mathcal{A}}_4) + \mathbb{P}(\bar{\mathcal{A}}_5) + \mathbb{P}(\bar{\mathcal{A}}_6),$$

and show that each summand on the right is $o(1)$. For the first summand we estimate using (39)

$$\begin{aligned}
&\mathbb{P}(\bar{\mathcal{A}}_1 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6) = \mathbb{E}\,\mathbb{P}(\bar{\mathcal{A}}_1 | X_1, \dots, X_n) \mathbb{I}_{\mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6} \\
&\le \mathbb{E}\left(\frac{M(\tilde{N}')^2}{(\tilde{N}')^2} \sum_i (\tilde{X}_i')^2 \mathbb{I}_{\mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6} | X_1, \dots, X_n \right) \\
&\le \left(\frac{\ln n}{\bar{N}^+}\right)^2 m \ln n = O\left(\frac{\ln^3 n}{n}\right).
\end{aligned}$$

It remains to show $\mathbb{P}(\bar{\mathcal{A}}_r) = o(1)$, for $r = 5, 6$. We write $\mathbb{P}(\bar{\mathcal{A}}_5) = \mathbb{P}(\bar{\mathcal{A}}_5 \cap \mathcal{A}_3) + o(1)$ and estimate

$$\mathbb{P}(\bar{\mathcal{A}}_5 \cap \mathcal{A}_3) \le \mathbb{P}(M(\bar{N}^+) > \ln n) = \mathbb{P}(\max_{j\in[m]} Z_j > \ln n) \le m\mathbb{P}(Z_1 > \ln n) = o(1). \tag{40}$$

Here Z_j denotes the number of balls in the jth bin after $\bar{N}^+$ balls have been thrown. In the second inequality we applied the union bound and used the fact that $Z_1, \ldots, Z_m$ are identically distributed. To get the very last bound we write for binomially $Bin(\bar{N}^+, m^{-1})$ distributed Z_1 and $t = \lfloor \ln n \rfloor$,

$$\mathbb{P}(Z_1 \ge t) \le \binom{\bar{N}^+}{t} m^{-t} \le \left(\frac{e\bar{N}^+}{tm}\right)^t = o\left(m^{-1}\right).$$

To estimate $\mathbb{P}(\bar{\mathcal{A}}_6)$ we apply Markov's inequality,

$$\mathbb{P}(\bar{\mathcal{A}}_6) \le (m \ln n)^{-1} n \mathbb{E}\,(\tilde{X}_1')^2 = \ln^{-1} n (Var(\tilde{X}_1') + (\mathbb{E}\,\tilde{X}_1')^2) = O(\ln^{-1} n).$$

Finally, we prove (34) for $r = 2$. Notice that the coupling between $M(\bar{N}^+)$ and $M(\bar{N}^-)$ is equivalent to the coupling provided by Lemma 3.5. Choose ϵ' solving $N^+ = (1 + \epsilon')N^-$ and note that $\epsilon' \sim 8\epsilon = O(\ln^{-2} n)$. The bound $\mathbb{P}(\mathcal{A}_2) = 1 - o(1)$ follows by Lemma 3.5 and the bound $\mathbb{P}(M(\bar{N}^+) > \ln n) = o(1)$, shown above. □

4 Algorithms for finding the largest clique

Random intersection graphs provide theoretical models for real networks, such as the affilation (actor, scientific collaboration) networks. Although the model assumptions about the distribution of the family of random sets defining the intersection graph are rather stringent (independence and a particular form of the distribution), these models yield random graphs with clustering properties similar to those found in real networks, [6]. While observing a real network we may or may not have information about the sets of attributes prescribed to vertices. Therefore it is important to have algorithms suited to random intersection graphs that do not use any data related to attribute sets prescribed to vertices. In this section we consider two such algorithms that find cliques of order $(1 + o(1))\omega(G)$ in a random intersection graph G.

The Greedy-Clique algorithm of [11] finds a clique of the optimal order $(1 - o_P(1))\omega(G)$ in a random intersection graph, in the case where (asymptotic) degree distribution is a power-law with exponent $\alpha \in (1; 2)$.

```
GREEDY-CLIQUE(G):
    Let v^(1), ..., v^(n) be V(G) sorted by their degrees, descending
    M ← ∅
    for i = 1 to n
        if v^(i) is adjacent to each vertex in M then
            M ← M ∪ {v^(i)}
    return M
```

Here we assume that graphs are represented by the adjacency list data structure. The implicit computational model behind our running time estimates in this section is random-access machine (RAM).

Proposition 4.1 *Assume that conditions of Theorem 1.1 hold. Suppose that* $\mathbb{E}\, Y = \Theta(1)$ *and that (7) holds for some* $\epsilon > 0$*. Then on input* $G = G(n)$ GREEDY-CLIQUE *outputs a clique of size* $\omega(G(n))(1 - o_P(1))$ *in time* $O(n^2)$*.*

By Lemma 1.5, the above result remains true if the conditions (2) and $\mathbb{E}\, Y(n) = \Theta(1)$ are replaced by the conditions (6) and $\mathbb{E}\, D_1 = \Theta(1)$. Proposition 4.1 is proved in a similar way as Lemma 2.1, but it does not follow from Lemma 2.1, since GREEDY-CLIQUE is not allowed to know the attribute subset sizes.

Proof The running time bound is obvious. We have to check that the algorithm returns a clique of the correct size. Fix any $\delta \in (0; 1)$. Let $C = C(n)$ be the clique returned by the algorithm on input $G = G(n)$, and write $\omega = \omega(G(n))$. Let $A_\delta = A_\delta(n)$ be the event that $|C| < (1-\delta)\omega$. To prove the lemma, we have to show that $\mathbb{P}(A_\delta) \to 0$. Fix positive a, b, such that $a < 1/4$ and $b < (1 - 0.1\delta)^{-1/\alpha} - 1$. Let θ_2, e_1, K be as in Section 2.1. Set $\tilde{\theta} = \tilde{\theta}(n) = (1 + b)\theta_2$ and

$$\tau = \tau(n) = ((1 - \alpha/2)\mathbb{E}\, Y \ln n + e_1)n^{1/2}(1 + n^{-a}).$$

We will assume that n is large enough, so that $\lfloor \tilde{\theta} + 1 \rfloor \leq m$. Let $V_2 = \{v \in [n] : X_v > \theta_2\}$ as before, and define random sets $Q = Q(n)$ and $R = R(n)$

$$Q = \{v \in [n] : D_v > \tau\} \quad \text{and} \quad R = \{v \in [n] : X_v > \tilde{\theta}\}.$$

Here D_v is the degree of v in $G(n)$. Let $B = B(n)$ be the event that $R \subseteq Q \subseteq V_2$. Write $\tilde{K} = \tilde{K}(n) = (1 - \alpha/2)^{-\alpha} K$. Recall that in the proof of Lemma 2.1 it does not matter in which order the vertices of V_2 are considered when constructing the set A^N, in particular, the order may be random. We will assume that the vertices in $V_2 \cap Q$ are always considered first, in the order provided by the GREEDY-CLIQUE algorithm. Let L_θ be as in Lemma 2.1. We claim that

$$\begin{aligned} \mathbb{P}(A_\delta) \leq & \mathbb{P}\left(|R| < (1 - 0.1\delta)\tilde{K}\right) + \mathbb{P}(\bar{B}) + \mathbb{P}\left(|V_2| > (1 + 0.1\delta)\tilde{K}\right) \\ & + \mathbb{P}\left(L_\theta > 0.1\delta |V_2|\right) + \mathbb{P}\left(\omega > (1 + 0.5\delta)\tilde{K}\right). \end{aligned} \tag{41}$$

Indeed, suppose that $|R| \geq (1 - 0.1\delta)\tilde{K}$, $L_\theta \leq 0.1\delta|V_2|$, $|V_2| \leq (1 + 0.1\delta)\tilde{K}$, $\omega \leq (1 + 0.5\delta)\tilde{K}$ and B holds. Then A_δ does not hold, since

$$\begin{aligned} |C| &\geq |Q| - L_\theta \geq |R| - L_\theta \geq (1 - 0.1\delta)\tilde{K} - 0.1\delta(1 + 0.1\delta)\tilde{K} \\ &\geq (1 - 0.5\delta)\tilde{K} \geq (1 - \delta)\omega. \end{aligned}$$

Now the last three terms on the right side of (41) tend to zero by the proof of Lemma 2.1 and Theorem 1.1. Furthermore, since $\mathbb{E}\,|R| = n\mathbb{P}(X > \tilde{\theta}) \sim (1+b)^{-\alpha}\tilde{K}$ and $(1+b)^{-\alpha} > 1 - 0.1\delta$, we get that $\mathbb{P}(|R| < (1-0.1\delta)\tilde{K}) \to 0$ by the concentration of $|R|$ (using, for example, (12)).

It remains to prove that B holds whp. Let us first show that $Q \subseteq V_2$ whp. By the union bound

$$\mathbb{P}(|Q \setminus V_2| > 0) \leq n\mathbb{P}(X_1 \leq \theta_2, D_1 > \tau) \leq n\mathbb{P}(D_1 > \tau | X_3 = \lfloor \theta_2 \rfloor)$$
$$\leq n\mathbb{P}\left(\sum_{v=2}^{n} \mathbb{I}_{S_v \cap S_1 \neq \emptyset} > \tau | X_1 = \lfloor \theta_2 \rfloor \right) \leq n\mathbb{P}(Z_1 > \tau), \qquad (42)$$

where $Z_1 = Z_1(n)$ is a random variable with distribution $Binom(n, p_1)$ and $p_1 = p_1(n) = \frac{\theta_2 \mathbb{E} X}{m}$. The last inequality follows by monotonicity, since the probability that S_2 intersects an independent uniformly random subset of $[m]$ of size $\lfloor \theta_2 \rfloor$ is at most p_1. We have

$$\mathbb{E}\, Z_1 = np_1 = n^{1/2+o(1)} \quad \textit{and} \quad \tau = (1 + n^{-a})\mathbb{E}\, Z_1.$$

Now by the Chernoff bound (11)

$$\mathbb{P}(Z_1 > \tau) = \mathbb{P}(Z_1 > (1 + n^{-a})\mathbb{E}\, Z_1) \leq \exp\left(-n^{\frac{1}{2}-2a+o(1)}\right).$$

Putting this in (42) we get that $\mathbb{P}(|Q \setminus V_2| > 0) \to 0$.

Similarly we will show that $R \subseteq Q$ whp. We have

$$\mathbb{P}(|R \setminus Q| > 0) \leq n\mathbb{P}(D_1 \leq \tau, X_1 > \tilde{\theta}) \leq \mathbb{P}(D_1 \leq \tau | X_1 = \lfloor \tilde{\theta} + 1 \rfloor) = n\mathbb{P}(Z_2 \leq \tau),$$

where $Z_2 = Z_2(n) = Binom(n-1, p_2)$ and $p_2 = p_2(n)$ is the probability that S_2 intersects an independent uniformly random subset of $[m]$ of size $\lfloor \tilde{\theta} + 1 \rfloor$. Now (7), $\mathbb{E}\, Y = \Theta(1)$ and Lemma 2.5 imply that for $\tilde{a} = \tilde{a}(n) = n^{-\epsilon/2}$ we have

$$\mathbb{E}\, \tilde{\theta} X_2 \mathbb{I}_{\tilde{\theta} X_2 > \tilde{a} m} = \left(\frac{m}{n}\right)^{1/2} \tilde{\theta} \mathbb{E}\, Y \mathbb{I}_{Y > \tilde{a} n^{1/2}((1-\alpha/2)\ln n + e_1)^{-1}} = o(\tilde{\theta} \mathbb{E}\, X). \qquad (43)$$

Next, observe that the bounds (55), (56) and (57) apply also when X_1 and X_2 are independent but with different distributions, in particular when $X_1 = \lfloor \tilde{\theta} + 1 \rfloor$ and X_2 has distribution $P(n)$; these inequalities together with (43) yield that

$$p_2 \sim \frac{\tilde{\theta} \mathbb{E}\, X}{m} \quad \text{and} \quad \mathbb{E}\, Z_2 \sim np_2 \sim (1+b)\tau.$$

Applying (11) again, we get that $\mathbb{P}(Z_2 \leq \tau) = \exp\left(-n^{1/2+o(1)}\right)$ and $\mathbb{P}(|R \setminus Q| > 0) \to 0$. □

For random intersection graphs with square integrable degree distribution we suggest the following simple algorithm.

```
MONO-CLIQUE(G):
    for uv ∈ E(G)
        D(uv) ← |Γ(u) ∩ Γ(v)|
    for uv ∈ E(G) in the decreasing order of D(uv)
        S ← Γ(u) ∩ Γ(v)
        if S is a clique then
            return S ∪ {u, v}
    return {1} ∩ V(G)
```

Here $\Gamma(v)$ denotes the set of neighbours of v.

Theorem 4.2 *Assume that $\{G(n)\}$ is a sequence of random intersection graphs such that $n = O(m)$ and $\mathbb{E}\, Y^2(n) = O(1)$. Let $C = C(n)$ be the clique constructed by* Mono-Clique *on input $G(n)$. Then $\mathbb{E}\, (\omega(G(n)) - |C|)^2 = O(1)$. Furthermore, if there is a sequence $\{\omega_n\}$, such that $\omega_n \to \infty$ and $\omega(G(n)) \geq \omega_n$ whp, then $|C| = \omega(G(n))$ whp.*

Proof Given distinct vertices $v_1, v_2, v_3, v_4 \in [n]$, let $\mathcal{C}(v_1, v_2, v_3, v_4)$ be the event that $G(n)$ contains a cycle with edges $\{v_1v_2, v_2v_3, v_3v_4, v_1v_4\}$ and $S_{v_2} \cap S_{v_4} = \emptyset$. Let Z denote the number of tuples (v_1, v_2, v_3, v_4) of distinct vertices in $[n]$ such that $\mathcal{C}(v_1, v_2, v_3, v_4)$ hold. We will show below that

$$\mathbb{E}\, Z = O(1). \tag{44}$$

Let $S \subseteq [n]$ be the (lexicographically first) largest clique of $G(n)$. Denote $s = |S|$. If $s \leq 2$ or there is a pair $\{x, y\} \subseteq S$, $x \neq y$ such that $G(n)[\Gamma(x) \cap \Gamma(y)]$ is a clique, then the algorithm returns a clique of size s. Otherwise, for each such pair $\{x, y\}$ there are $x', y' \in \Gamma(x) \cap \Gamma(y)$, $x' \neq y'$ with $x'y' \notin E(G(n))$. That is, $\mathcal{C}(x, x', y, y')$ holds and $\binom{s}{2} \leq Z$. Thus, if $\binom{s}{2} > Z$, the algorithm returns a clique C of size s. Otherwise, the algorithm may fail and return a clique C of size 1. In any case we have that

$$s - |C| \leq \sqrt{2Z} + 1$$

and using (44)

$$\mathbb{E}\, (\omega(G(n)) - |C|)^2 \leq \mathbb{E}\, (\sqrt{2Z} + 1)^2 = O(1).$$

Also if $\omega(G(n)) \geq \omega_n$ whp, then by (44) and Markov's inequality

$$\mathbb{P}(|C| \neq \omega(G(n))) \leq \mathbb{P}(\omega(G(n)) < \omega_n) + \mathbb{P}\left(Z \geq \binom{\omega_n}{2}\right) \to 0.$$

It remains to show (44). What is the probability of the event $\mathcal{C}(1, 2, 3, 4)$? Clearly, $\mathcal{C}(1, 2, 3, 4)$ implies at least one of the following events:

- $\mathcal{A}_1$: there are distinct attributes $w_1, w_2, w_3, w_4 \in W$ such that $w_1 \in S_1 \cap S_2$, $w_2 \in S_2 \cap S_3$, $w_3 \in S_3 \cap S_4$ and $w_4 \in S_1 \cap S_4$;
- $\mathcal{A}_2$: there are distinct $w_1, w_2, w_3 \in W$, such that $w_1 \in S_1 \cap S_2 \cap S_3$, $w_2 \in S_3 \cap S_4$ and $w_3 \in S_1 \cap S_4$;
- $\mathcal{A}_3$: there are distinct $w_1, w_2, w_3 \in W$, such that $w_1 \in S_1 \cap S_2$, $w_2 \in S_2 \cap S_3$ and $w_3 \in S_1 \cap S_3 \cap S_4$;
- $\mathcal{A}_4$: there are distinct $w_1, w_2 \in W$, such that $w_1 \in S_1 \cap S_2 \cap S_3$ and $w_2 \in S_1 \cap S_3 \cap S_4$.

Conditioning on X_1, X_2, X_3, X_4 and using the union bound and independence we obtain, similarly as in Lemma 3.2

$$\mathbb{P}(\mathcal{A}_1) \le (m)_4 \mathbb{E}\, \frac{(X_1)_2 (X_2)_2 (X_3)_2 (X_4)_2}{(m)_2^4} \le \frac{(\mathbb{E}\, Y^2)^4}{n^4};$$
$$\mathbb{P}(\mathcal{A}_2) = \mathbb{P}(\mathcal{A}_3) \le (m)_3 \mathbb{E}\, \frac{(X_1)_2 X_2 (X_3)_2 (X_4)_2}{(m)_2^3 m} \le \frac{(\mathbb{E}\, Y^2)^3 (\mathbb{E}\, Y)}{m^{0.5} n^{3.5}};$$
$$\mathbb{P}(\mathcal{A}_4) \le (m)_2 \mathbb{E}\, \frac{(X_1)_2 X_2 (X_3)_2 X_4}{(m)_2^2 m^2} \le \frac{(\mathbb{E}\, Y^2)^2 (\mathbb{E}\, Y)^2}{m n^3}.$$

Furthermore, by symmetry,

$$\mathbb{E}\, X \le (n)_4 \left(\mathbb{P}(\mathcal{A}_1) + \mathbb{P}(\mathcal{A}_2) + \mathbb{P}(\mathcal{A}_3) + \mathbb{P}(\mathcal{A}_4)\right) = O(1).$$

□

Proposition 4.3 *Consider a sequence of random intersection graphs $\{G(n)\}$ as in Theorem 1.3.* Mono-Clique *can be implemented so that its expected running time on $G(n)$ is $O(n)$.*

Proof Let $\tilde{Z}$ denote the number of 4-cycles in $G(n)$, i.e., the number of tuples (v_1, v_2, v_3, v_4) of distinct vertices in $[n]$, such that $v_1v_2, v_2v_3, v_3v_4, v_1v_4 \in E(G(n))$. We will prove below that

$$\mathbb{E}\, \tilde{Z} = O(n). \tag{45}$$

Consider the running time of the first loop. We can assume that the elements in each list in the adjacency list structure are sorted in increasing order (recall that vertices are elements of $V = [n]$). Otherwise, given $G(n)$, they can be sorted using any standard sorting algorithm in time $O(n + \sum_{v \in [n]} D_v^2)$, where $D_v = d_{G(n)}(v)$ is the degree of v in $G(n)$. The intersection of two lists of lengths k_1 and k_2 can be found in $O(k_1 + k_2)$ time, so that expected total time for finding common neighbours is

$$O\left(n + \mathbb{E} \sum_{uv \in E(G(n))} (D_u + D_v)\right) = O\left(n + \mathbb{E} \sum_{v \in [n]} D_v^2\right) = O(n).$$

The last estimate follows by (66) in the proof of Lemma 1.6.

The second loop can be implemented so that the next edge uv with largest value of $D(uv)$ is found at each iteration (i.e., we do not sort the list of edges in advance). In this way picking the next edge requires at most $ce(G(n))$ steps c is a universal constant. We recall that the number of edges $uv \in E(G)$ with $\Gamma(u, v) := \Gamma(u) \cap \Gamma(v) \ne \emptyset$ that fail to induce a clique is at most the number Z of cycles considered in the proof of Theorem 4.2 above. Therefore, the total number of steps used in picking $D(uv)$ in decreasing order is at most

$$Z\, e(G(n)) = \sum_{(i,j,k,l)} \mathbb{I}_{\mathcal{C}(i,j,k,l)} e(G(n)).$$

Now
$$e(G(n) = \sum_{s<t\colon \{s,t\}\cap\{i,j,k,l\}=\emptyset} \mathbb{I}_{\{s\sim t\}} + \sum_{s<t\colon \{s,t\}\cap\{i,j,k,l\}\neq\emptyset} \mathbb{I}_{\{s\sim t\}}.$$
Note, that the second sum on the right is at most $4n$. Also, if $\{s,t\}\cap\{i,j,k,l\} = \emptyset$, the events $s \sim t$ and $\mathcal{C}(i,j,k,l)$ are independent, therefore

$$\begin{aligned}\mathbb{E}\left(\mathbb{I}_{\mathcal{C}(i,j,k,l)} \sum_{s<t\colon \{s,t\}\cap\{i,j,k,l\}=\emptyset} \mathbb{I}_{\{s\sim t\}}\right) &= \mathbb{P}(\mathcal{C}(i,j,k,l)) \sum_{s<t\colon \{s,t\}\cap\{i,j,k,l\}=\emptyset} \mathbb{P}(s\sim t)\\ &\le \mathbb{P}(\mathcal{C}(i,j,k,l))\mathbb{E}\, e(G(n)).\end{aligned}$$

Finally, invoking the simple bound $\mathbb{E}\, e(G(n)) = \binom{n}{2}\mathbb{P}(u\sim v) = O(n)$, and (44) we get

$$\mathbb{E}\, Z\, e(G(n)) \le (\mathbb{E}\, e(G(n)) + 4n) \sum_{(i,j,k,l)} \mathbb{P}(\mathcal{C}(i,j,k,l)) = (\mathbb{E}\, e(G(n)) + 4n)\mathbb{E}\, Z = O(n).$$

Now let us estimate the time of the rest of the iteration of the second loop. The total expected time to find common neighbours is again $O(n)$, so we only consider the time spent for checking if $\Gamma(u,v)$ is a clique. This requires $c\, s_{uv}^2$ steps, where we denote $s_{uv} = |\Gamma(u,v)|$. Observe that u, v and $\Gamma(u,v)$ yield at least $s_{uv}(s_{uv}-1)$ 4-cycles in $G(n)$ of the form (u,x,v,y), $x,y \in \Gamma(u,v)$. Summing over all edges uv and noticing that each 4-tuple corresponding to 4-cycle in $G(n)$ can be obtained at most once, we get

$$\tilde{Z} \ge \sum_{uv\in E(G(n))} s_{uv}(s_{uv}-1) \ge \sum_{uv\in E(G(n))} (s_{uv}^2 - 1)/2.$$

So using (45) and the fact that $\mathbb{E}\, e(G(n)) = O(n)$ we obtain

$$\mathbb{E} \sum_{uv\in E(G(n))} s_{uv}^2 \le 2\mathbb{E}\, \tilde{Z} + \mathbb{E}\, e(G(n)) = O(n).$$

Finally, let us bound $\mathbb{E}\, \tilde{Z}$. Let $\mathcal{A}_i$, $1 \le i \le 4$ be as in the proof of Theorem 4.2. Let $\mathcal{A}_5$ be the event that there is $w \in W$ such that $w \in S_1 \cap S_2 \cap S_3 \cap S_4$. Using the union bound
$$\mathbb{P}(\mathcal{A}_5) \le m\mathbb{E}\, \frac{X_1X_2X_3X_4}{m^4} = \frac{(\mathbb{E}\, Y)^4}{mn^2}.$$
Similarly as in the proof of Theorem 4.2 (we have to consider three other events similar to A_2 and A_4),

$$\mathbb{E}\, \tilde{Z} \le (n)_4\,(\mathbb{P}(A_1) + 4\mathbb{P}(A_2) + 2\mathbb{P}(A_4) + \mathbb{P}(A_5)) = O(n).$$

□

Combining the next lemma with Theorem 1.3 we can show that Mono-Clique whp finds a clique of size at least $\omega'(G(n))$.

Lemma 4.4 *Let $\{G(n)\}$ be as in Theorem 1.3 and let $M = M(G(n))$ be the monochromatic clique of size $\omega'(G(n))$ generated by the attribute with the smallest index. Then whp $G(n)$ has an edge uv such that $\{u, v\} \cup (\Gamma(u) \cap \Gamma(v)) = M$.*

Before we prove the lemma, we need several definitions. A *balls and bins configuration* $\mathcal{C}$ with parameters (N, m, n) is a pair $(bin_{\mathcal{C}}, label_{\mathcal{C}})$ of functions $bin_{\mathcal{C}} : [N] \to [m]$ and $label_{\mathcal{C}} : [N] \to [n]$. These functions determine the placement of N distinguishable balls with labels from $[n]$ into m bins. Recall that $f(S)$ and $f^{-1}(S)$ denote the image and preimage of a set S under a function f respectively. Given a balls and bins configuration $\mathcal{C}$, for each $i \in [n]$ we define $S_i(\mathcal{C}) = bin_{\mathcal{C}}(label_{\mathcal{C}}^{-1}(\{i\}))$, $X_i(\mathcal{C}) = |S_i(\mathcal{C})|$ and $X'_i(\mathcal{C}) = |label_{\mathcal{C}}^{-1}(\{i\})|$. Also define for $j \in [m]$ the set $T_j(\mathcal{C}) = bin_{\mathcal{C}}^{-1}(\{j\})$, write $Z_j(\mathcal{C}) = |T_j(\mathcal{C})|$ and interpret $T_j(\mathcal{C})$ as the set of balls in the bin j. We denote by $L(\mathcal{C})$ the (smallest) index $j \in [m]$ that maximizes $Z_j(\mathcal{C})$. Also, write $\omega'(\mathcal{C}) = Z_L(\mathcal{C})$, where $L = L(\mathcal{C})$.

The intersection graph $G(\mathcal{C})$ corresponding to $\mathcal{C}$ is the intersection graph on vertex set $[n]$ of the family of sets $\{S_i(\mathcal{C}) : i \in [n]\}$.

Given a function $f : [N] \to [m]$ and $x, y \in [N]$ let $exchange(f, x, y)$ be an operation, the result of which is again a function $g : [N] \to [m]$, defined as follows.

$$g(w) = \begin{cases} f(x), & \text{if } w = y \\ f(y), & \text{if } w = x; \\ f(w), & \text{otherwise.} \end{cases}$$

Given a balls and bins configuration $\mathcal{C}$ with parameters (N, m, n), define random variables b_1, b_2, b'_1, b'_2 taking values[1] in the set

$$\{undefined\} \cup [N]$$

as follows. If $\omega'(\mathcal{C}) \leq 1$ let $b_1 = b_2 = undefined$. Otherwise, let (b_1, b_2) be chosen uniformly at random from all ordered pairs of distinct balls in the bin $L(\mathcal{C})$. We let $b'_1 = b'_2 = undefined$ if $N \leq 1$. Otherwise, we define (b'_1, b'_2) as a uniformly random pair from all $N(N-1)$ ordered pairs of distinct balls in $[N]$.

Now if $\omega'(\mathcal{C}) \leq 1$, define $bin_{\mathcal{C}'} = bin_{\mathcal{C}}$, otherwise define

$$bin_{\mathcal{C}'} = exchange(exchange(bin, b_1, b'_1), b_2, b'_2),$$

that is, we exchange the bins of b_1 and b'_1, then exchange the bins of b_2 and b'_2. Finally, let $T(\mathcal{C}) = T(\mathcal{C}, b_1, b_2, b'_1, b'_2)$ be a balls and bins configuration with parameters (N, m, n) and functions $(bin_{\mathcal{C}'}, label_{\mathcal{C}})$.

Lemma 4.5 *Let n, m be positive integers, let $(x_1, \ldots, x_n)$ be a sequence of nonnegative integers and let $N = x_1 + \cdots + x_n$. Let f be an arbitrary function $[N] \to [n]$, such that $|f^{-1}(\{j\})| = x_j$ for each $j \in [n]$. Let $\mathcal{C}$ be a random balls and bins configuration with parameters (N, m, n) defined as follows:*

[1] We use a special value "undefined" to avoid the need of extra conditioning later.

- *$bin_{\mathcal{C}}$ is chosen uniformly at random from all m^N functions from $[N]$ to $[m]$ (i.e., each ball is thrown into a uniformly random bin);*
- *$label_{\mathcal{C}} = f$.*

Then $T(\mathcal{C})$ has the same distribution as $\mathcal{C}$.

Proof Let $z = (z_1, \dots, z_m)$ be a vector of non-negative integers, such that $\sum_{i=1}^m z_i = \sum_{j=1}^n x_j = N$ and write $z_{max} = \max_{i\in[n]} z_i$. For a balls and bins configuration $\mathcal{C}'$, write $Z(\mathcal{C}') = (Z_1(\mathcal{C}'), \dots, Z_m(\mathcal{C}'))$

Define a Markov chain M_z with state space $\mathcal{S}_z$ consisting of all balls and bins configurations $\mathcal{C}'$ with paramaters (N, m, n), such that $label_{\mathcal{C}'} = f$ and $Z(\mathcal{C}') = z$, and with transitions given by the operation T.

Suppose $z_{max} \geq 2$. We claim, that the transition probabilities p_{st} of M_z satisfy

$$\sum_{s\in\mathcal{S}_z} p_{ts} = \sum_{s\in\mathcal{S}_z} p_{st} \text{ for any state } t \in \mathcal{S}_z. \tag{46}$$

(Note that the above sum must be equal to 1). This follows by a standard argument: let $H = H(V_1, V_2)$ be an edge-weighted bipartite graph, where $V_1 = \mathcal{S}_z$, $V_2 = \{s' : s \in \mathcal{S}_z\}$ is a disjoint copy of V_1 and for each $st' \in V_1 \times V_2$, let the weight $w(st')$ be the number of quadruples (b_1, b_2, b_1', b_2') leading from s to t. By symmetry, for any $s \in \mathcal{S}$,

$$\sum_{t\in V_2} w(st') = \sum_{t\in V_1} w(ts') = \frac{1}{|\mathcal{S}_z|} \sum_{s,t} w(st').$$

For any state in $\mathcal{S}_z$, any valid quadruple has the same probability $p = (z_{max}(z_{max} - 1)N_0(N_0 - 1))^{-1}$ to be picked by the random transition T. Therefore (46) follows by multiplying both sides of the last equality by p. (By considering "inverse" transitions one may further show that the transition matrix of this Markov chain is actually symmetric.)

Now by (46), the uniform distribution over $\mathcal{S}_z$ is a stationary distribution for M_z. Clearly, conditionally on $Z(\mathcal{C}) = z$, $\mathcal{C}$ is uniformly distributed over $\mathcal{S}_z$, and so $T(\mathcal{C})$ is also distributed uniformly over $\mathcal{S}_z$. The last statement is also true in the case where $z_{max} < 2$, since in this case T is the identity operation. We conclude that $T(\mathcal{C})$ has the same distribution as $\mathcal{C}$. □

Proof of Lemma 4.4 Using the iid random variables $X_1, \dots, X_n$ (distributed according to $P(n)$), we will construct two random balls and bins configurations $\mathcal{C}'$ and $\mathcal{C}''$, such that

(i) $\mathcal{C}'$ and $\mathcal{C}''$ have the same distribution;

(ii) $G(\mathcal{C}')$ and $G(\mathcal{C}'')$ are distributed as $G(n)$.

At the same time we will construct a pair of random vertices (u^*, v^*), such that

- whp $u^*, v^* \in T_L(\mathcal{C}'')$, where $L = L(\mathcal{C}'')$;
- $u^*, v^* \in \{0\} \cup [n]$;

- Given $X_1, \dots, X_n$, such that $N = X_1 + \cdots + X_n$, u^*, v^* are iid, and for any $x \in [n]$ $\mathbb{P}(u^* = x) = \frac{X_i}{N+1}$;
- Given $X_1, \dots, X_n$, u^*, v^* are conditionally independent of $\mathcal{C}'$.

Finally we will show that whp $\mathcal{C}'$ and $\mathcal{C}''$ differ only by the positions of at most four balls, so we will be able to use the simple distribution of (u^*, v^*) in $G(\mathcal{C}')$ to finish the proof of the lemma.

Construction of $\mathcal{C}'$ and $\mathcal{C}''$. We proceed in several steps. We start by drawing the sequence $X_1, \dots, X_n$. Define a random balls and bins configuration $\mathcal{C}_0$ with parameters (N, m, n), $N = \sum_{i=1}^n X_i$ and X_i balls labelled i as in Lemma 4.5. For $b \in [N]$ let $label_{\mathcal{C}_0}(b) = \min\{i \in [n] : X_1 + \cdots + X_i \geq b\}$ and notice that $label_{\mathcal{C}_0}$ depends only on $X_1, \dots, X_n$. Now let $\mathcal{C}_0' = \mathcal{C}_0$ and let $\mathcal{C}_0'' = T(\mathcal{C}_0)$. Let $b_1, b_2, b_1', b_2' \in \{undefined\} \cup [N]$ be the corresponding random balls used in the operation T as in Lemma 4.5.

In the *second phase* we complete the construction of $\mathcal{C}'$ and $\mathcal{C}''$ by adding more balls to $\mathcal{C}_0'$ and $\mathcal{C}_0''$ respectively so that $X_i(\mathcal{C}') = X_i(\mathcal{C}'') = X_i$ for each $i \in [n]$. We use pairing as much as we can. Specifically, assume that $\mathcal{C}_t'$ and $\mathcal{C}_t''$ are already defined. If there is no index i, such that $\max(X_i(\mathcal{C}_t'), X_i(\mathcal{C}_t'')) < X_i$, we stop. Otherwise, let i be smallest such index. Pick b_{t+1}^{**} independently and uniformly at random from $[m]$ and place $N+t+1$ to the bin b_{t+1}^{**} in both $\mathcal{C}_t'$ and $\mathcal{C}_t''$, to obtain $\mathcal{C}_{t+1}'$ and $\mathcal{C}_{t+1}''$ respectively with parameters $(N+t+1, m, n)$ (formally, $bin_{\mathcal{C}_{t+1}'}(N+t+1) = b_{t+1}^{**}$, $label_{\mathcal{C}_{t+1}'}(N+t+1) = i$, and $bin_{\mathcal{C}_{t+1}'}, label_{\mathcal{C}_{t+1}'}$ agree on $[N+t]$; $\mathcal{C}_{t+1}''$ is obtained from $\mathcal{C}_t''$ similarly).

Let t be the largest integer, such that $\mathcal{C}_t'$ and $\mathcal{C}_t''$ are defined. For $s = t, t+1, \dots,$ while there is an index i such that $X_i(\mathcal{C}_s') < X_i$ add a ball $N+s+1$ labelled i to a uniformly random bin in $\mathcal{C}_s'$ to obtain $\mathcal{C}_{s+1}'$. Call the final configuration $\mathcal{C}'$. Similarly, but now independently, while there is an index i with $X_i(\mathcal{C}_s'') < X_i$, add a ball $N+s+1$ labelled i to a uniformly random bin from $\mathcal{C}_s''$ to obtain $\mathcal{C}_{s+1}''$, and call the final configuration $\mathcal{C}''$.

Proof of (i) and (ii). By Lemma 4.5, $\mathcal{C}_0'$ and $\mathcal{C}_0''$ are identically distributed. The procedure to obtain $\mathcal{C}'$ from $\mathcal{C}_0'$ is the same as the procedure to obtain $\mathcal{C}''$ from $\mathcal{C}_0''$: we may ignore the coupling of the second phase since the distribution for each of the final configuration does not depend on the sequence of labels used in this phase. Therefore $\mathcal{C}'$ and $\mathcal{C}''$ are also identically distributed. Using a similar argument as in the proof of Theorem 1.3, we see that $G(\mathcal{C}')$ has the same distribution as $G(n)$.

Properties of $\mathcal{C}'$ and $\mathcal{C}''$. Let N' and N'' be the number of balls in $\mathcal{C}'$ and $\mathcal{C}''$ respectively. For $i \in \{1, 2\}$ let $l_i' = label_{\mathcal{C}_0}(b_i')$ in the case where $b_i' \in [N]$, and let $l_i' = 0$ in the case where $b_i' = undefined$. Similarly let $l_i = label_{\mathcal{C}_0}(b_i)$ in the case where $b_i \in [N]$ and let $l_i = 0$ otherwise. Finally, define $w_i' = bin_{\mathcal{C}_0}(b_i')$ if $b_i' \neq undefined$ and $w_i' = 0$ otherwise. Also define $L := L(\mathcal{C}'')$.

We will prove that all of the following events hold whp

- A : $N' = N''$, $label_{\mathcal{C}'} = label_{\mathcal{C}''}$, the functions $bin_{\mathcal{C}'}$, $bin_{\mathcal{C}''}$ agree on $[N'] \setminus \{b_1, b_2, b'_1, b'_2\}$ and for $i \in \{1,2\}$ $bin_{\mathcal{C}'}(b'_i) = bin_{\mathcal{C}''}(b_i)$, $bin_{\mathcal{C}'}(b_i) = bin_{\mathcal{C}''}(b'_i)$;
- A' : for each $k \in [n] \setminus \{l'_1, l'_2\}$ there is $i \in \{1,2\}$ such that $S_k(\mathcal{C}') \cap S_{l'_i}(\mathcal{C}') = \emptyset$;
- B : for $i \in \{1,2\}$, $X_{l'_i}(\mathcal{C}_0) = X_{l'_i}$;
- C : b_1, b_2, b'_1, b'_2 are distinct and belong to $[N]$;
- D : $w'_1 \notin S_{\mathcal{C}_0}(l'_2)$ and $w'_2 \notin S_{\mathcal{C}_0}(l'_1)$;
- E : $L = L(\mathcal{C}_0)$, $T_L(\mathcal{C}''_0) = T_L(\mathcal{C}'')$ and all balls in $T_L(\mathcal{C}'')$ have distinct labels; also all balls in $T_L(\mathcal{C}')$ have distinct labels;
- F_1 : $L(\mathcal{C}_0) \notin S_{l'_1}(\mathcal{C}_0) \cup S_{l'_2}(\mathcal{C}_0)$;
- F_2 : for $j \in \{1,2\}$, b_j is the unique ball in $T_{w'_j}(\mathcal{C}'')$ with $label_{\mathcal{C}''}(b_j) = l_j$.

Assume that B, C, D, E, F_1 and F_2 occur, we will show that A occurs (this event is not essential but helps to make the proof clearer). Since B and F_1 hold, for $i \in \{1,2\}$, each ball in the set $label^{-1}_{\mathcal{C}_0}(\{l'_i\})$ is in a distinct bin of $\mathcal{C}'_0 = \mathcal{C}_0$, and not in the bin $L(\mathcal{C}_0)$. Using also C, each ball with label l'_i is in a distinct bin of $\mathcal{C}''_0$. Therefore no new balls with label l'_i are added in the random construction (of both $\mathcal{C}'$ and $\mathcal{C}''$), once $\mathcal{C}'_0$ and $\mathcal{C}''_0$ are defined and

$$S_{l'_i}(\mathcal{C}') \setminus \{w'_i\} = S_{l'_i}(\mathcal{C}'') \setminus \{L\}. \tag{47}$$

Now let $j \in \{1,2\}$. C, E, F_1 and F_2 imply that $L \in S_{l_j}(\mathcal{C}'_0)$ and $w'_j \notin S_{l_j}(\mathcal{C}'_0)$; meanwhile $L \notin S_{l_j}(\mathcal{C}''_0)$ and $w'_j \in S_{l_j}(\mathcal{C}''_0)$. By E, no new ball labelled l_j is added to the bin L in the second phase of the construction. By F_2, no new ball labelled l_j is added to the bin w'_j in the second phase. This means that

$$S_{l_j}(\mathcal{C}') \setminus \{L\} = S_{l_j}(\mathcal{C}'') \setminus \{w'_j\}. \tag{48}$$

Also, notice that the operation T does not modify the balls with labels $k \in [n] \setminus \{l_1, l_2, l'_1, l'_2\}$ therefore by the pairing $S_k(\mathcal{C}') = S_k(\mathcal{C}'')$ and A holds.

Now let A'' be the event that l'_1 and l'_2 belong to the set $M = label(T_L(\mathcal{C}''))$, $l'_1 \neq l'_2$, $|M| = \omega'(G(\mathcal{C}''))$ and $\Gamma_{G(\mathcal{C}'')}(l'_1) \cap \Gamma_{G(\mathcal{C}'')}(l'_2) = M \setminus \{l'_1, l'_2\}$. To prove the lemma, it suffices to show that A'' holds whp.

Suppose A', B, C, D, E, F_1, F_2 hold, but A'' does not. Then there is $k \in [n] \setminus \{l'_1, l'_2\}$, such that for both $i \in \{1,2\}$, $(S_{l'_i}(\mathcal{C}'') \setminus \{L\}) \cap S_k(\mathcal{C}'') \neq \emptyset$. Suppose $k \notin \{l_1, l_2\}$. We have $S_k(\mathcal{C}') = S_k(\mathcal{C}'')$, so if $S_k(\mathcal{C}'')$ intersects $S_{l'_i}(\mathcal{C}'') \setminus \{L\}$ then $S_k(\mathcal{C}')$ intersects $S_{l'_i}(\mathcal{C}')$ and A' does not hold, a contradiction. Therefore we can assume that $k \in \{l_1, l_2\}$.

Since $w'_1, w'_2 \notin S_{l'_i}(\mathcal{C}'')$ (by B, b'_i is the unique ball b in $T_{w'_i}(\mathcal{C}_0)$ with $label_{\mathcal{C}_0}(b) = l'_1$, and so by C, D and F_1, $w'_1, w'_2 \notin S_{l'_i}(\mathcal{C}''_0) = S_{l'_i}(\mathcal{C}'')$) we have that for $i \in \{1,2\}$,

$$(S_{l'_i}(\mathcal{C}'') \setminus \{L, w'_1, w'_2\}) \cap (S_k(\mathcal{C}'') \setminus \{L, w'_1, w'_2\}) \neq \emptyset.$$

But using (47) and (48), $(S_x(\mathcal{C}') \setminus \{L, w_1', w_2'\}) = (S_x(\mathcal{C}'') \setminus \{L, w_1', w_2'\})$ for $x \in \{l_1, l_2, l_1', l_2'\}$, therefore A' does not occur. This is a contradiction. We conclude that indeed the events $A', \dots, F_2$ imply A''.

To complete the proof of the lemma, it suffices to show that A', B, C, D, E, F_1, F_2 all occur whp.

The event B. For a random variable Y and an event A we will write $\mathbb{E}_* Y = \mathbb{E}(Y | X_1, \dots, X_n)$ and $\mathbb{P}_*(A) = \mathbb{E}_* \mathbb{I}_A$.

Let $i \in \{1, 2\}$. We have $X_i(\mathcal{C}_0) = X_i$ if and only if each ball with label l_i' falls into a distinct bin of $\mathcal{C}_e$. Given $X_1, \dots, X_n$ and l_i', the probability of this is

$$\frac{m-1}{m} \times \cdots \times \frac{m - X_{l_i'} + 1}{m} \geq 1 - \frac{1 + \cdots + (X_{l_i'} - 1)}{m} \geq 1 - \frac{X_{l_i'}^2}{2m}.$$

Therefore

$$\begin{aligned}
\mathbb{P}(\bar{B}) &= \mathbb{E}\,\mathbb{E}\,(\mathbb{I}_{\bar{B}} | X_1, \dots, X_n, l_i') \leq \mathbb{E}\,\mathbb{E}_* \left(1 \wedge \frac{X_{l_i'}^2}{2m}\right) \\
&\leq \mathbb{E}\left(\sum_{k=1}^{n} \left(1 \wedge \frac{X_k^2}{2m}\right) \frac{2X_k}{\mathbb{E}\,N}\right) + \mathbb{P}\left(N < \frac{\mathbb{E}\,N}{2}\right). \\
&= \frac{2n}{m^{1/2} n^{1/2} \mathbb{E}\,Y} \mathbb{E}\left(1 \wedge \frac{X_1^2}{2m}\right) X_1 + \mathbb{P}\left(N < \frac{\mathbb{E}\,N}{2}\right). \qquad (49)
\end{aligned}$$

We can bound the second term using Chebyshev's inequality:

$$\mathbb{P}\left(N < \frac{\mathbb{E}\,N}{2}\right) \leq \frac{4Var(N)}{(\mathbb{E}\,N)^2} \leq \frac{4\mathbb{E}\,Y^2}{n(\mathbb{E}\,Y)^2} = O(n^{-1}). \qquad (50)$$

We have

$$\begin{aligned}
&\mathbb{E}\left(1 \wedge \frac{X_1^2}{2m}\right) X_1 = \mathbb{E}\left(1 \wedge \frac{X_1^2}{2m}\right) X_1 \mathbb{I}_{X_1 \leq m^{1/2}} + \mathbb{E}\left(1 \wedge \frac{X_1^2}{2m}\right) X_1 \mathbb{I}_{X_1 > m^{1/2}} \\
&\leq \frac{m^{1/2} \mathbb{E}\,X_1^2}{2m} + \mathbb{E}\,X_1 \mathbb{I}_{X_1 > m^{1/2}} \leq \frac{m^{1/2} \mathbb{E}\,Y^2}{2n} + m^{-1/2} \mathbb{E}\,X_1^2.
\end{aligned}$$

Putting the last two bounds in (49) we get that

$$\mathbb{P}(\bar{B}) \leq \frac{\mathbb{E}\,Y^2}{n^{1/2} \mathbb{E}\,Y} + \frac{2\mathbb{E}\,Y^2}{n^{1/2} \mathbb{E}\,Y} + o(n^{-1}) = O(n^{-1/2})$$

The event C**.** First, using (50), we have that

$$\mathbb{P}(\bar{C}) = \mathbb{P}(\bar{C}, N \geq 0.5\mathbb{E}\,N) + O(n^{-1}).$$

The balls b_1, b_2 are "undefined" if and only if all N balls fall into different bins. For $N > m$ this cannot happen. For $N \le m$, the probability of this is

$$\left(1 - \frac{1}{m}\right) \dots \left(1 - \frac{N-1}{m}\right) \le e^{-\frac{N(N-1)}{m}}.$$

So

$$\mathbb{P}(\{b_1, b_2\} \cap \{undefined\} \neq \emptyset, N \ge 0.5\mathbb{E}\, N) \le e^{-\frac{0.5\mathbb{E}\, N(0.5\mathbb{E}\, N - 1)}{m}} = e^{-\Omega(n)} \to 0.$$

Notice that $b_1, b_2 \in [N]$ implies that $b'_1, b'_2 \in [N]$. Now recall that when $b_1, b_2, b'_1, b'_2 \in [N]$ (see Lemma 4.5), $b_1 \neq b_2$ and $b'_1 \neq b'_2$. Finally for $i, j \in \{1, 2\}$ using the fact that, conditionally on N, b'_i is uniformly distributed in $[N]$ and independent of b_j,

$$\mathbb{P}(b'_i = b_j, N \ge 0.5\mathbb{E}\, N) \le \frac{2}{\mathbb{E}\, N} \to 0;$$

and it follows by the union bound and the above estimates that $\mathbb{P}(\bar{C}) = o(1)$.

The event D**.** We will define random vertices $u^*, v^* \in \{0\} \cup [n]$. We can assume that b'_1 and b'_2 are generated, given N, as follows. Let b^*_1, b^*_2 be drawn independently and uniformly at random from $\{undefined\} \cup [N]$. If $b^*_1 \neq b^*_2$ or $N \ge 2$ and $b_1, b_2 \notin \{undefined\}$, we set $(b'_1, b'_2) = (b^*_1, b^*_2)$. Otherwise, we let $(b'_1, b'_2) = (b''_1, b''_2)$, where (b''_1, b''_2) is a new uniformly random pair of distinct balls from $[N]$, independent from (b^*_1, b^*_2). For $i \in \{1, 2\}$ set $l^*_i = label_{\mathcal{C}_0}(b^*_i)$ if $b^*_i \in [N]$ and $l^*_i = 0$ otherwise and define $(u^*, v^*) = (l^*_1, l^*_2)$.

Using (50) we get that $(u^*, v^*) = (l'_1, l'_2)$ with probability $O(n^{-1})$.

Conditionally on $X_1, \dots, X_n$, u^* and v^* are independent (and also independent of $G(\mathcal{C}')$), and

$$\mathbb{P}_*(u^* = i) = \mathbb{P}_*(v^* = i) = \frac{X_i}{N+1} \quad \text{for } i = 1, 2, \dots, n.$$

Now we have

$$\begin{aligned} \mathbb{P}(\bar{D}) &\le \mathbb{P}(S_{l'_1}(\mathcal{C}') \cap S_{l'_2}(\mathcal{C}') \neq \emptyset) = \mathbb{P}(u^* v^* \in E(G(\mathcal{C}'))) + O(n^{-1}) \\ &\le \mathbb{E}\, \mathbb{E}_* \left(\mathbb{P}_*(u^* v^* \in E(G(\mathcal{C}')))\mathbb{I}_{N \ge 0.5\mathbb{E}\, N}\right) + \mathbb{P}(N < 0.5\mathbb{E}\, N) + O(n^{-1}). \end{aligned} \tag{51}$$

The second term in the last line is $O(n^{-1})$ by (50). The first term is at most

$$\begin{aligned} &\mathbb{E}\, \mathbb{E}_* \left(\sum_{i \neq j} \frac{X_i X_j}{(N+1)^2} \mathbb{I}_{N \ge 0.5\mathbb{E}\, N} \mathbb{P}_*(ij \in E(G(\mathcal{C}'))) \right) \\ &\le \frac{4n^2}{(\mathbb{E}\, N + 2)^2} \mathbb{E}\, X_1 X_2 \mathbb{P}_*(12 \in E(G(\mathcal{C}'))) \le \frac{4n^2 \mathbb{E}\, X_1^2 X_2^2}{(\mathbb{E}\, N + 2)^2 m} = O(n^{-1}). \end{aligned}$$

This yields that $\mathbb{P}(\bar{D}) = O(n^{-1})$.

The event E. As discussed above, $\mathcal{C}'$, $\mathcal{C}''$ and the configuration used in the proof of Theorem 1.3 all have the same distribution. So $\mathbb{P}(E) \to 1$ by the proof of Theorem 1.3.

The event F_1. Fix $i \in \{1,2\}$. Let F_{1i} be the event that the bin L in $\mathcal{C}_0$ does not contain a ball with label l_i^*. Since $l_i' = l_i^*$ whp, it is enough to show that $\mathbb{P}(\bar{F}_{1i}) = o(1)$. Let $\tilde{F}$ be the event that $N \geq 0.5\mathbb{E}\,N > 2$, $\sum_i X_i = N$, $\sum_i X_i^2 \leq m \ln n$ and $\omega'(\mathcal{C}_0) \leq (\ln n)^2$. Then

$$\mathbb{P}(\bar{F}_{1i}|\tilde{F}) \leq \mathbb{E} \sum_i \frac{X_i \omega'(\mathcal{C}_0)}{N} \frac{X_i}{N+1} \leq \frac{(\ln n)^3}{\mathbb{E}\,Y n} = o(1).$$

By the inequality (40) shown in the proof of Theorem 1.3

$$\mathbb{P}(\omega'(\mathcal{C}_0) > (\ln n)^2) \to 0. \tag{52}$$

By (38) and Markov's inequality, $\mathbb{P}(\sum_i X_i^2 > m \ln n) = o(1)$. Therefore by the union bound

$$\mathbb{P}(\tilde{F}) \geq 1 - \mathbb{P}(\omega'(\mathcal{C}_0) > (\ln n)^2) - \mathbb{P}(\sum_i X_i^2 > m \ln n) - \mathbb{P}(N < 0.5\mathbb{E}\,N) = 1 - o(1).$$

So $\mathbb{P}(\bar{F}_{1i}) \leq \mathbb{P}(\bar{F}_{1i}|\tilde{F}) + o(1) \to 0$.

The event F_2. Fix $j \in \{1,2\}$. The event C implies that $b_j \in T_{w_j'}(\mathcal{C}_0'')$. The event E implies that $l_1 \neq l_2$. It suffices to prove that the following two events occur whp: $\tilde{A}$: in the configuration $\mathcal{C}_0$, the bin w_j' does not contain a ball with label l_j; and $\tilde{B}$: in the second phase no ball with label l_j is added to the bin w_j'. Denote $\tilde{S} = S_{l_j}(\mathcal{C}_0)$ (in the case where $b_j = undefined$, this set is empty). Since $\mathbb{P}(b_j' \neq b_j^*) = O(n^{-1})$ we may replace b_j' with b_j^*. Define $X_0 = X_0(n) = 0$ and recall that $l_j = 0$ when b_j is undefined. Using the conditional independence of b_j^* and $\mathcal{C}_0$,

$$\mathbb{P}_*(b_j^* \in bin_{\mathcal{C}_0}^{-1}(\tilde{S})) \leq (N+1)^{-1}\mathbb{E}_*|bin_{\mathcal{C}_0}^{-1}(\tilde{S})|$$

Given $X_1, \ldots, X_n$, l_j and $\mathcal{C}_0$, such that $\omega'(\mathcal{C}_0) \leq (\ln n)^2$ we have

$$bin_{\mathcal{C}_0}^{-1}(\tilde{S}) \leq X_{l_j} (\ln n)^2.$$

By symmetry, given $X_1, \ldots, X_n$ and $\mathcal{C}_0$ such that $\omega'(\mathcal{C}_0) \geq 2$, the ball b_j is uniformly distributed in $[N]$. We have

$$\mathbb{E}_* X_{l_j} \leq \frac{X_1^2 + \cdots + X_n^2 + 1}{N+1}.$$

Therefore, for n large enough, using (50), (52) and the above estimates

$$\begin{aligned}
\mathbb{P}(\bar{\tilde{A}}) \leq & \mathbb{E}\,\mathbb{I}_{\omega(\mathcal{C}_0) \leq (lnn)^2} \mathbb{I}_{N \geq 0.5N} \mathbb{P}_*(b_j^* \in bin_{\mathcal{C}_0}^{-1}(\tilde{S})) \\
& + \mathbb{P}(b_j^* \neq b_j') + \mathbb{P}(\omega(\mathcal{C}_0) > (\ln n)^2) + \mathbb{P}(N < 0.5\mathbb{E}\,N) \\
\leq & \frac{4n(\ln n)^2 \mathbb{E}\,X_1^2}{(\mathbb{E}\,N + 2)^2} + o(1) = o(1).
\end{aligned}$$

Now consider the event $\tilde{B}$. We have shown in the proof of Theorem 1.3 that there is $\delta_n \to 0$, such that

$$\mathbb{P}\left(N, N', N'' \in ((1-\delta_n)\mathbb{E}\, N, (1+\delta_n)\mathbb{E}\, N)\right) \to 1.$$

The number of balls added in the second phase depends on $\mathcal{C}_0$ and b_1, b_2, b_1', b_2', but suppose $T = 10\delta_n \mathbb{E}\, N$ balls are generated (independently) in advance together with their bin numbers (chosen independently from $[m]$). Then whp, these T balls are enough to complete the construction of $\mathcal{C}'$ and $\mathcal{C}''$, and, since $m = \Omega(n)$, the probability that at least one of these balls falls into the bin w_j' is at most

$$\frac{T}{m} \le \frac{10\delta_n n^{1/2}}{m^{1/2}} \to 0.$$

So

$$\mathbb{P}(\tilde{\bar{B}}) \le \frac{T}{m} + \mathbb{P}((N'-N) + (N''-N) > 4\delta_n \mathbb{E}\, N) \to 0.$$

The event A'. Denote by B_k the event that $S_k(\mathcal{C}') \cap S_{u^*}(\mathcal{C}') \ne \emptyset$ and $S_k(\mathcal{C}') \cap S_{v^*}(\mathcal{C}') \ne \emptyset$. Denote by B_{k1} the event that $S_k(\mathcal{C}') \cap S_{u^*}(\mathcal{C}') \cap S_{v^*}(\mathcal{C}') \ne \emptyset$, and by B_{k2} the event $B_k \setminus B_{k1}$. Since $(l_1', l_2') = (u^*, v^*)$ whp, it suffices to show that

$$\mathbb{P}(\exists k \notin \{u^*, v^*\} : B_k) \le \mathbb{P}(\exists k \notin \{u^*, v^*\} : B_{k1}) + \mathbb{P}(\exists k \notin \{u^*, v^*\} : B_{k2}) \to 0.$$

In the proof we only need to work with the random intersection graph $G(\mathcal{C}')$. Using (51) above, we have

$$\begin{aligned}
&\mathbb{P}(\exists k \notin \{u^*, v^*\} : B_{k1}) \le \mathbb{P}(u^*v^* \in E(G(\mathcal{C}'))) \to 0.\\
&\mathbb{P}(\exists k \notin \{u^*, v^*\} : B_{k2}) \le \mathbb{P}(N < 0.5\mathbb{E}\, N) + \mathbb{P}(u^* = v^*)\\
&\quad + \mathbb{E}\, \mathbb{I}_{N \ge 0.5\mathbb{E}\, N} \mathbb{I}_{u^* \ne v^*} \sum_{k \in [n] \setminus \{u^*, v^*\}} \mathbb{P}_*(B_{k2})
\end{aligned}$$

The first and second terms on the right of the last inequality are $O(n^{-1})$ using (50). The last term is at most

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i \ne j} \mathbb{I}_{N \ge 0.5\mathbb{E}\, N} \frac{X_i}{(N+1)} \frac{X_j}{(N+1)} \sum_{k \in [n] \setminus \{i,j\}} \mathbb{P}_*(B_{k2} | u^* = i, v^* = j)\right)\\
&\le \frac{4}{m^2(\mathbb{E}\, N + 2)^2} \sum_{i,j,k \in [n], i \ne j, i \ne k, j \ne k} \mathbb{E}\, X_i^2 X_j^2 X_k^2 \le \frac{4(\mathbb{E}\, Y^2)^3}{n(\mathbb{E}\, Y)^2} = O(n^{-1})
\end{aligned}$$

Here in the last line we used

$$\mathbb{P}_*(B_{k2} | u^* = i, v^* = j) \le \frac{X_i X_j (X_k)_2}{m^2},$$

which follows using the union bound and the observation that if $S_i(\mathcal{C}')$ and $S_j(\mathcal{C}')$ both intersect $S_k(\mathcal{C}')$, but the intersection of all three is empty, then there must be distinct elements w_1, w_2, such that $w_1 \in S_i(\mathcal{C}') \cap S_k(\mathcal{C}')$ and $w_2 \in S_j(\mathcal{C}') \cap S_k(\mathcal{C}')$. By the union bound we get $\mathbb{P}(\exists k \notin \{u^*, v^*\} : B_k) \to 0$.

□

5 Equivalence between set size and degree parameters

Here we prove Lemmas 1.5 and 1.6. In the proof we write $X = X(n)$, $Y = Y(n)$, and $D_1 = D_1(n)$. We denote $X_1, X_2, \dots$ the sizes of subsets $S_1, S_2, \dots \subseteq W$ prescribed to the vertices $1, 2, \dots \in V = [n]$ of $G(n)$.

Proof of Lemma 1.5 We start by showing that if either $\mathbb{E}\, Y$ or $\mathbb{E}\, D_1$ converges and for some positive sequence $\{a_n\}$ converging to zero (we write $a = a_n$ for short),

$$\mathbb{E}\, Y \mathbb{I}_{\{Y > (an)^{1/2}\}} \to 0 \tag{53}$$

then

$$\mathbb{E}\, Y = (\mathbb{E}\, D_1)^{1/2} + o(1). \tag{54}$$

We note that $\mathbb{E}\, D_1 = (n-1)\mathbb{P}(S_1 \cap S_2 \neq \emptyset)$. We estimate this probability using the inequalities, see Lemma 6 in [6],

$$\frac{X_1 X_2}{m} \geq \mathbb{P}(S_1 \cap S_2 \neq \emptyset | X_1, X_2) \geq \max\left\{0, \left(\frac{X_1 X_2}{m} - \frac{X_1^2 X_2^2}{m^2}\right)\right\} =: Z. \tag{55}$$

Notice that $\mathbb{E}\, Y = \Omega(1)$. This is clear if $\mathbb{E}\, Y \to y \in (0; \infty)$. Otherwise, we have $\mathbb{E}\, D_1 \to d \in (0; \infty)$ and, by the first inequality of (55),

$$(n-1)\frac{(\mathbb{E}\, Y)^2}{n} \geq (n-1)\mathbb{P}(S_1 \cap S_2 \neq \emptyset) = \mathbb{E}\, D_1.$$

Furthermore, from $\mathbb{E}\, Y = \Omega(1)$ and (53) we conclude that $\mathbb{E}\, X \mathbb{I}_{\{X \geq (am)^{1/2}\}} = o(\mathbb{E}\, X)$. Using this bound we estimate $\mathbb{E}\, Z$ from below

$$\begin{aligned} \mathbb{E}\, Z &\geq \mathbb{E}\, Z \mathbb{I}_{X_1 X_2 \leq am} \geq (1-a)m^{-1}\mathbb{E}\, X_1 X_2 \mathbb{I}_{X_1 X_2 \leq am} \\ &\geq (1-a)m^{-1}\mathbb{E}\, X_1 \mathbb{E}\, X_2 - m^{-1}\mathbb{E}\, X_1 X_2 \mathbb{I}_{X_1 X_2 > am}, \end{aligned} \tag{56}$$

where

$$\begin{aligned} \mathbb{E}\, X_1 X_2 \mathbb{I}_{X_1 X_2 > am} &\leq \mathbb{E}\, X_1 X_2 \left(\mathbb{I}_{X_1 > (am)^{1/2}} + \mathbb{I}_{X_2 > (am)^{1/2}}\right) \\ &\leq 2\mathbb{E}\, X \mathbb{E}\, X \mathbb{I}_{X > (am)^{1/2}} \\ &= o((\mathbb{E}\, X)^2). \end{aligned} \tag{57}$$

Hence, $\mathbb{E}\, Z \geq (1 - o(1))(\mathbb{E}\, X)^2$. Combining this inequality with (55) we obtain

$$\mathbb{P}(S_1 \cap S_2 \neq \emptyset) \sim m^{-1}(\mathbb{E}\, X)^2,$$

thus proving (54).

It remains to prove that (2)$\Leftrightarrow$(6). Since both implications are shown in much the same way, we only prove (2)$\Rightarrow$(6). For this purpose we fix $0 < \tilde{\epsilon} < \min\{\epsilon, \epsilon_0\}$ and show that for each $0 < \delta < 1$ and each sequence $\{t_n\}$ with $n^{1/2-\tilde{\epsilon}} \leq t_n \leq n^{1/2+\tilde{\epsilon}}$

$$\liminf_n \big(\mathbb{P}(Y_1(n) \geq t_n)/\mathbb{P}(D_1(n) \geq t_n)\big) \geq (d^{1/2}(1+\delta))^{-\alpha}, \tag{58}$$

$$\limsup_n \big(\mathbb{P}(Y_1(n) \geq t_n)/\mathbb{P}(D_1(n) \geq t_n)\big) \leq (d^{1/2}(1-\delta))^{-\alpha}. \tag{59}$$

Here the random variable $Y_1(n) := (n/m)^{1/2} X_1(n)$ has the same distribution as $Y(n)$. We prove (58) and (59) by contradiction.

Proof of (58). Suppose there is an increasing sequence $\{n_k\}$ of positive integers and a sequence $\{b_k\}$ with $n_k^{1/2-\tilde{\epsilon}} \le b_k \le n_k^{1/2+\tilde{\epsilon}}$ such that, for some $0 < \delta < 1$,

$$\mathbb{P}(Y_1(n_k) \ge b_k) < (d^{1/2}(1+\delta))^{-\alpha} \mathbb{P}(D_1(n_k) \ge b_k), \quad k = 1, 2, \dots. \tag{60}$$

Define $\{l_k\}$ by the relation $b_k = d^{1/2}(1 + \delta/2) l_k$, $k \ge 1$. Introduce events $\mathcal{A}_k = \{D_1(n_k) \ge b_k\}$, $\mathcal{B}_k = \{Y_1(n_k) \ge l_k\}$ and write

$$\mathbb{P}(\mathcal{A}_k) = \mathbb{P}(\mathcal{A}_k \cap \mathcal{B}_k) + \mathbb{P}(\mathcal{A}_k \cap \bar{\mathcal{B}}_k). \tag{61}$$

In what follows we drop the subscript k and write b, l, n, m instead of b_k, l_k, n_k, m_k. We note that (2) together with (60) imply

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) \le \mathbb{P}(\mathcal{B}) \sim d^{\alpha/2}(1+\delta/2)^{\alpha} \mathbb{P}(Y_1(n) \ge b) \le c_1 \mathbb{P}(\mathcal{A}),$$

where the constant $c_1 = ((1+\delta/2)/(1+\delta))^{\alpha} < 1$. Next we show that $\mathbb{P}(\mathcal{A} \cap \bar{\mathcal{B}}) = O(n^{-10})$ thus obtaining a contradiction to (60), (61).

Denote $x = \lfloor (m/n)^{1/2} l \rfloor$. Conditionally, given the event $\mathcal{C} = \{X_1(n) = x\}$, the random variable $D_1(n)$ has binomial distribution $Bin(n-1, p)$ with success probability $p = \mathbb{P}(S_1 \cap S_2 \ne \emptyset \big| |S_1| = x)$ satisfying $p \sim d^{1/2} l/n$. Indeed, the first inequality of (55) implies

$$p \le \frac{x \mathbb{E}\, X_2}{m} = \frac{x (m/n)^{1/2} \mathbb{E}\, Y}{m} \sim d^{1/2} \frac{l}{n}.$$

Here we used $\mathbb{E}\, Y \to d^{1/2} > 0$. The second inequality of (55) implies, see (56),

$$p \ge \frac{1-a}{m} x \mathbb{E}\, X_2 \mathbb{I}_{\{x X_2 < am\}} = \frac{1-a}{m} x (\mathbb{E}\, X_2 - r) \sim \frac{x \mathbb{E}\, X_2}{m}.$$

Here $r = \mathbb{E}\, X_2 \mathbb{I}_{\{x X_2 \ge am\}} = o(\mathbb{E}\, X_2)$, for $a = a(n_k) = \ln^{-1} n_k$, cf. (57).

Next, since $b \sim (1+\delta/2) np$ and $np \sim d^{1/2} l = \Omega(n^{1/2-\tilde{\epsilon}})$ we obtain, by Chernoff's inequality, $\mathbb{P}(\mathcal{A}|\mathcal{C}) = O(n^{-10})$. Now, using the inequality $\mathbb{P}(\mathcal{A}|Y_1(n) = y) \le \mathbb{P}(\mathcal{A}|\mathcal{C})$, for $y \le l$, we obtain

$$\mathbb{P}(\mathcal{A} \cap \bar{\mathcal{B}}) = \mathbb{E}\, \mathbb{P}(\mathcal{A}|Y_1(n)) \mathbb{I}_{\{Y(n) \le l\}} \le \mathbb{P}(\mathcal{A}|\mathcal{C}) = O(n^{-10}). \tag{62}$$

Proof of (59). Suppose there is an increasing sequence $\{n_k\}$ of positive integers and a sequence $\{b_k\}$ with $n_k^{1/2-\tilde{\epsilon}} \le b_k \le n_k^{1/2+\tilde{\epsilon}}$ such that, for some $0 < \delta < 1$,

$$\mathbb{P}(Y_1(n_k) \ge b_k) > (d^{1/2}(1-\delta))^{-\alpha} \mathbb{P}(D_1(n_k) \ge b_k), \quad k = 1, 2, \dots. \tag{63}$$

Define $\{l_k\}$ by the relation $b_k = d^{1/2}(1 - \delta/2) l_k$, $k \ge 1$. We write

$$\mathbb{P}(D_1(n_k) \ge b_k) = \mathbb{P}(Y_1(n_k) \ge l_k) \mathbb{P}(D_1(n_k) \ge b_k | Y_1(n_k) \ge l_k). \tag{64}$$

We note that, by (2) and (63), the first term on the right is at least $(c_2 + o(1))$ $\mathbb{P}(D_1(n_k) \geq b_k)$ where the constant $c_2 = ((1-\delta/2)/(1-\delta))^{\alpha} > 1$. Finally, we obtain a contradiction, by showing that the second term of (64) is $1 - O(n^{-10})$. Here we proceed as in (62) above. We write

$$\mathbb{P}(D_1(n_k) < b_k | Y_1(n_k) \geq l_k) \leq \mathbb{P}(D_1(n_k) < b_k | \mathcal{C})$$

and show that binomial probability on the right-hand side is $O(n^{-10})$ using Chernoff's inequality.

□

Proof of Lemma 1.6 The identity (9) follows from (54) since

$$\mathbb{E}\, Y \mathbb{I}_{Y > \epsilon_n n^{1/2}} \leq (\mathbb{E}\, Y^2 \mathbb{I}_{Y > \epsilon_n n^{1/2}})^{1/2} \to 0.$$

Let us show (10). Denote N the number of 2-stars in $G = G(n)$ centered at vertex $1 \in V = [n]$. Introduce the events $\mathcal{A}_{ij} = \{i \sim j\}$, $i, j \in V$. Write, for short, $\mathcal{A} = \mathcal{A}_{12} \cap \mathcal{A}_{13}$. Let $\tilde{\mathbb{P}}$ denote the conditional probability given the sizes X_1, X_2, X_3 of the random subsets prescribed to vertices $1, 2, 3 \in V$. We remark that (10) follows from (9) combined with the simple identities

$$\mathbb{E}\, D_1(D_1 - 1) = 2\mathbb{E}\, N = (n-1)(n-2)\mathbb{P}(\mathcal{A}),$$

and the inequalities

$$(\mathbb{E}\, Y)^2 \mathbb{E}\, Y^2 \geq n^2 \mathbb{P}(\mathcal{A}) \geq (1 - o(1))(\mathbb{E}\, Y)^2 \mathbb{E}\, Y^2. \tag{65}$$

Let us prove (65). For this purpose we write (using the conditional independence of events $\mathcal{A}_{12}$ and $\mathcal{A}_{13}$, given X_1, X_2, X_3)

$$\mathbb{P}(\mathcal{A}) = \mathbb{E}\, \tilde{\mathbb{P}}(\mathcal{A}) = \mathbb{E}\, \tilde{\mathbb{P}}(\mathcal{A}_{12}) \tilde{\mathbb{P}}(\mathcal{A}_{13})$$

and evaluate conditional probabilities $\tilde{\mathbb{P}}(\mathcal{A}_{ij})$ using (55). From the first inequality of (55) we obtain the first inequality of (65)

$$\mathbb{P}(\mathcal{A}) = \mathbb{E}\, \tilde{\mathbb{P}}(\mathcal{A}_{12}) \tilde{\mathbb{P}}(\mathcal{A}_{13}) \leq \mathbb{E}\, (X_1^2 X_2 X_3)/m^2 = (\mathbb{E}\, Y)^2 \mathbb{E}\, Y^2 / n^2.$$

Thus, even without the assumption (8) (we use this fact this in the proof of Proposition 4.3), we have

$$\mathbb{E}\, D_1 \leq \mathbb{E}\, Y \quad \text{and} \quad \mathbb{E}\, D_1(D_1 - 1) \leq \mathbb{E}\, Y^2 \mathbb{E}\, Y. \tag{66}$$

To show the second inequality of (65) we apply the second inequality of (55) and use truncation. We denote $\mathbb{I}_i = \mathbb{I}_{\{X_i \leq \epsilon_n m^{1/2}\}}$, $\bar{\mathbb{I}}_i = 1 - \mathbb{I}_i$ and write, cf. (56),

$$\begin{aligned} \mathbb{P}(A) &\geq \mathbb{E}\, \tilde{\mathbb{P}}(\mathcal{A}) \mathbb{I}_1 \mathbb{I}_2 \mathbb{I}_3 \geq (1 - \epsilon_n^2)^2 \mathbb{E}\, (X_1^2 X_2 X_3 / m^2) \mathbb{I}_1 \mathbb{I}_2 \mathbb{I}_3 \\ &\geq (1 - \epsilon_n^2)^2 \mathbb{E}\, (X_1^2 X_2 X_3 / m^2)(1 - \bar{\mathbb{I}}_1 - \bar{\mathbb{I}}_2 - \bar{\mathbb{I}}_3) \\ &= (1 - o(1))(\mathbb{E}\, Y)^2 \mathbb{E}\, Y^2 / n^2. \end{aligned}$$

In the last step we used the fact that $\mathbb{E}\, Y^2 \geq (\mathbb{E}\, Y)^2 = \Omega(1)$ and the bounds

$$\begin{aligned} &\mathbb{E}\, X_1^2 \bar{\mathbb{I}}_1 = (m/n)\, \mathbb{E}\, Y^2 \mathbb{I}_{\{Y > \epsilon_n n^{1/2}\}} = o(\mathbb{E}\, X^2), \\ &\mathbb{E}\, X_j \bar{\mathbb{I}}_j = (m/n)^{1/2} \mathbb{E}\, Y \mathbb{I}_{\{Y > \epsilon_n n^{1/2}\}} = o(\mathbb{E}\, X), \qquad j = 2, 3. \end{aligned}$$

□

6 Concluding remarks

In this work we determined the order of the clique number in $G(n, m, P)$ for a wide range of $m = m(n)$ and $P = P(n)$. We saw that in sparse power-law random intersection graphs with unbounded degree variance, the clustering property of $G(n, m, P)$ has little influence in the formation of the maximum clique. This suggests that simpler models, such as the one in [11], may be preferable in the case of very heavy degree tails. However, when the degree variance is bounded, most random graph models, including the Erdős-Rényi graph and the model of [11] have only bounded size cliques whp. In contrast, we showed that in random intersection graphs the clique number can still diverge slowly.

We have a kind of "phase transition" as the tail index α for the random subset size (degree) varies, see (2). Assume, for example that $m = \Theta(n)$. When $\alpha < 2$, the random graph $G(n, m, P)$ whp contains cliques of only logarithmic size. When $\alpha > 2$, it whp contains a 'giant' clique of polynomial size. But what happens when (2) is satisfied with $\alpha = 2$ but the degree variance is unbounded?

We proposed a surprisingly simple algorithm for finding (almost) the largest clique in sparse random intersection graphs with finite degree variance. The performance of both Greedy-Clique and Mono-Clique algorithms can be of further interest, since these algorithms do not use the possibly hidden random subset structure. How well would they perform on arbitrary sparse empirical networks? Can we suspect a hidden intersecting sets structure for networks where the Mono-Clique algorithm performs well?

Another direction of possible future research would be to determine the asymptotic clique number in dense random intersection graphs (alternatively, the order of the largest intersecting set in dense random hypergraphs). For example, even in the random uniform hypergraph case where $m = \Theta(n)$ and the random subset size $X(n) = \Omega(n^{1/2})$ is deterministic, exact asymptotics of the clique number remain open.

References

[1] N. Alon, T. Jiang, Z. Miller and D. Pritkin, Properly coloured subgraphs and rainbow subgraphs in edge-colourings with local constraints, *Random Struct. Algorithms* **23** (2003), 409–433.

[2] J. Balogh, T. Bohman and D. Mubayi, Erdős - Ko - Rado in random hypergraphs. *Combinatorics, Probability and Computing* **18** (2009), 629–646.

[3] M. Behrisch, A. Taraz and M. Ueckerdt, Colouring random intersection graphs and complex networks. *SIAM J. Discrete Math.* **23** (2009), 288–299.

[4] G. Bianconi and M. Marsili, Emergence of large cliques in random scale-free networks. *Europhys. Lett.* **74** (2006), 740–746.

[5] M. Bloznelis, J. Jaworski and V. Kurauskas, Assortativity and clustering coefficient of sparse random intersection graphs, *Electronic Journal of Probability* **18**, No. 38 (2013), 1–24.

[6] M. Bloznelis, Degree and clustering coefficient in sparse random intersection graphs, *Ann. Appl. Probab.*, **23**, No. 3, (2013), 1254–1289.

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

[8] M. Deijfen and W. Kets, Random intersection graphs with tunable degree distribution and clustering, *Probab. Eng. Inf. Sci.* **23** (2009) 661–674.

[9] J. Galambos and E. Seneta, Regularly varying sequences, *Proc. Amer. Math. Soc.* **41**(1973) 110–116.

[10] E. Godehardt and J. Jaworski, Two models of random intersection graphs for classification. In: O. Optiz and M. Schwaiger, Editors, *Studies in Classification, Data Analysis and Knowledge Organization* **22**, Springer, Berlin (2003), 67–82.

[11] S. Janson, T. Łuczak and I. Norros, Large cliques in a power-law random graph, *J. Appl. Probab.* **47** (2010), 1124–1135.

[12] S. Janson, T. Łuczak, and A. Ruciński, Random Graphs, *Wiley-Interscience Series in Discrete Mathematics and Optimization*, Wiley-Interscience, New York, 2000.

[13] M. Karoński, E.R. Scheinerman and K.B. Singer-Cohen, On random intersection graphs: the subgraph problem, *Comb. Probab. Comput.* **8** (1999) 131–159.

[14] V. Kolchin, B. Sevstyanov and V.Chistyakov, Random Allocations, *V.H. Winston and Sons*, 1978, Washington D.C.

[15] C. McDiarmid, Concentration, in *Probabilistic Methods for Algorithmic Discrete Mathematics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed Eds., Springer, New York (1998) 195–248.

[16] S. Nikoletseas, C. Raptopoulos and P. G. Spirakis, Maximum cliques in graphs with small intersection number and random intersection graphs, *Mathematical Foundations of Computer Science 2012*, Springer Berlin Heidelberg, 2012. 728 – 739.

[17] K. Rybarczyk and D. Stark, Poisson approximation of the number of cliques in random intersection graphs, *J. Appl. Probab.*, **47** (2010), 826–840.