Large cliques in sparse random intersection graphs

Mindaugas Bloznelis[*]
Faculty of Mathematics and Informatics
Vilnius University
Lithuania
mindaugas.bloznelis@mif.vu.lt

Valentas Kurauskas[*]
Institute of Mathematics and Informatics[†]
Vilnius University
Lithuania
valentas@gmail.com

Submitted: Jun 18, 2016; Accepted: Mar 28, 2017; Published: Apr 13, 2017
Mathematics Subject Classifications: 05C80; 05C82; 05C85; 05C69

Abstract

An intersection graph defines an adjacency relation between subsets $S_1, \ldots, S_n$ of a finite set $W = \{w_1, \ldots, w_m\}$: the subsets S_i and S_j are adjacent if they intersect. Assuming that the subsets are drawn independently at random according to the probability distribution $\mathbb{P}(S_i = A) = P(|A|)\binom{m}{|A|}^{-1}$, $A \subseteq W$, where P is a probability on $\{0, 1, \ldots, m\}$, we obtain the random intersection graph $G = G(n, m, P)$.

We establish the asymptotic order of the clique number $\omega(G)$ of a sparse random intersection graph as $n, m \to +\infty$. 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 asymptotic degree distribution of G is a power-law with exponent $\alpha \in (1, 2)$. It is of size $\frac{\ln n}{\ln \ln n}(1 + o_P(1))$ if the degree distribution has bounded variance, i.e., $\alpha > 2$. We construct a simple polynomial-time algorithm which finds a clique of the optimal order $\omega(G)(1 - o_P(1))$.

Keywords: clique; random intersection graph; greedy algorithm; complex network; power-law; clustering

1 Introduction

Bianconi and Marsili [5] observed that "scale-free" real networks can have very large cliques; they gave an argument suggesting that the clique number can grow polynomially with graph order n if the degree variance is unbounded. Using a more precise analysis,

*Supported by the Research Council of Lithuania (MIP-052/2010, MIP-067/2013).
†The main contribution of the second author was made during PhD studies at the Faculty of Mathematics and Informatics, Vilnius University.

Janson, Łuczak and Norros [14] established 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 affects 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 [24], [25]. An interesting question is about tracing the relation between the clustering property and the clique number in a power-law network. This question is addressed in the present paper: we show precise asymptotics for the clique number of a related random intersection graph $G(n, m, P)$, which admits a tunable clustering coefficient and power-law degree distribution [6, 7, 8, 9]. We find that the effect of clustering on the clique number only shows up for the degree sequences having finite variance. In particular, for a finite variance power-law (asymptotic) degree, the clique number of $G(n, m, P)$ is $\frac{\ln n}{\ln \ln n}(1 + o_P(1))$ while the clique number of the random graph of [14] is at most 4. We note that the random graph of [14] has conditionally independent edges and does not possess the clustering property. For a power-law (asymptotic) degree with infinite variance the clique number asymptotics for both random graphs are the same.

In the language of hypergraphs, the question considered in the paper is about the size of the largest intersecting family in a random hypergraph on the vertex set $[m]$, where n identically distributed hyperedges are of random sizes distributed according to P.

The paper is organized as follows. In this section we collect several facts about the random intersection graph $G(n, m, P)$ and present our main results on the clique number asymptotics. Proofs of these results are given in Sections 2 and 3. In Section 4 we present and rigorously analyse algorithms for finding large cliques.

Random intersection graphs first studied by Karoński, Scheinerman and Singer-Cohen [16] are convenient models of affiliation networks, a class of social networks where two actors are declared adjacent if they share some common attributes [24]. We denote by $V = \{v_1, \dots, v_n\}$ the set of vertices (actors of the network). Vertices are represented by the collections of attributes $S_1, \dots, S_n$ selected by actors independently at random from the attribute set $W = \{w_1, \dots, w_m\}$ according to the probability distribution $\mathbb{P}(S_i = A) = P(|A|)\binom{m}{|A|}^{-1}$, $A \subseteq W$. Here P is the probability distribution of the sizes $X_i := |S_i|$ of the attribute sets. We interpret $X_1, \dots, X_n$ as weights modelling the actors' activity. This model has been introduced by Godehardt and Jaworski [12]. It is related to the random graph of [14]: in both models, given the weights, edge probabilities are (approximately) proportional to the product of weights of their endpoints. Consequently, their degree sequences behave in a similar way. A distinctive feature of $G(n, m, P)$ is that it admits a nonvanishing tunable clustering coefficient, defined in (5) below. Analysis and detection of large cliques in social networks is another source of motivation of our study.

Formally, 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 $X = X(n)$ denote a random variable distributed according to $P(n)$ and define $Y = Y(n) := \sqrt{\frac{n}{m}} X(n)$. If not explicitly stated otherwise, the limits below will be taken as $n \to +\infty$. We use the standard notation $o()$, $O()$, $\Omega()$, $\Theta()$, $o_P()$, $O_P()$, see, for example, [15]. 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 with high probability (*whp*), if $\mathbb{P}(\mathcal{A}_n) \to 1$.

Assuming that $Y(n)$ converges in L_1 to some random variable Y_0 with $\mathbb{E}\, Y_0 < \infty$ one can show that the degree of the typical vertex of $G(n)$, equivalently, the degree $D_1 = D_1(n)$ of vertex $v_1 \in V$, converges in distribution to a Poisson mixture: $\mathbb{P}(D_1 = k) \to \mathbb{E}\, e^{\lambda}\lambda^k/k!$, $k = 0, 1, \ldots$, see [7]. Here $\lambda = Y_0 \mathbb{E}\, Y_0$ is a random variable. Since the Poisson distribution is tightly concentrated around its mean, one can show that the asymptotic degree distribution is power-law with an exponent $\alpha > 1$ whenever the distribution of Y_0 is. For our purposes the weak convergence to a power law is not sufficient as we need a tight control on the tail $\mathbb{P}(Y(n) \geqslant k)$ for $k \approx n^{1/2}$. To this aim we introduce the following condition: given $\alpha > 0$ and slowly varying function L there is $\epsilon_0 > 0$ such that for each sequence $\{x_n\}$ with $n^{0.5-\epsilon_0} \leqslant x_n \leqslant n^{0.5+\epsilon_0}$ we have

$$\mathbb{P}\left(Y(n) \geqslant x_n\right) \sim L(x_n) x_n^{-\alpha}. \tag{1}$$

Our first result concerns the power-law random intersection graph $G(n, m, P)$ with infinite degree variance. Its conditions are formulated in terms of parameters defining the random graph.

Theorem 1.1. *Let $m, n \to +\infty$. Let $1 < \alpha < 2$. Assume that $\{G(n)\}$ is a sequence of random intersection graphs. Suppose that $\{Y(n)\}$ satisfies condition (1) and*

$$\mathbb{E}\, Y(n) = O(1). \tag{2}$$

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

$$\omega(G(n)) = (1-\alpha/2)^{-\alpha/2}\, L\left((n \ln n)^{1/2}\right) n^{1-\alpha/2} (\ln n)^{-\alpha/2} (1 + o_P(1)). \tag{3}$$

We note that (1) and (3) refer to the same function L. For $L \equiv 1$ we obtain Pareto tails. We remark that condition (1) can be replaced by a related condition (6) on the tail of the degree distribution, see Lemma 1.5 below.

The asymptotics in (3) turn out to be the same as in the model of Janson, Łuczak and Norros [14] with corresponding parameters. Let us mention that the lower bound for $\omega(G(n))$ follows by an argument of [14]: a clique of order (3) is formed by vertices $v \in V$ with largest set sizes $|S_v|$. To show a matching upper bound we developed a method based on a result of Alon, Jiang, Miller and Pritkin [2] in extremal combinatorics. Among several auxiliary combinatorial lemmas, Lemma 2.10 about distinct representatives may be of an independent interest.

In the case where the degree distribution has bounded variance, in addition to determining the asymptotic order of $\omega(G(n))$, we 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 also T_w. We say that every edge of T_w receives colour w and call this clique *monochromatic*. Note that the edges of $G(n)$ are covered by the union of monochromatic cliques T_w, $w \in W$. We denote the size of the largest monochromatic clique by $\omega'(G(n))$. Another, out of many ways, when a clique on $S \subseteq V$ arises, is when $G(n)$ contains a *rainbow* clique on S: for each pair of vertices $\{x, y\} \subseteq S$ we can assign a different attribute $w = w_{\{x,y\}}$ such that $x, y \in T_w$. The size of the largest rainbow clique in $G(n)$ will be denoted by $\omega_R(G(n))$. Clearly, $\omega(G(n)) \geqslant \max\{\omega'(G(n)), \omega_R(G(n))\}$.

Theorem 1.2. *Let $m, n \to +\infty$. Assume that $\{G(n)\}$ is a sequence of random intersection graphs satisfying (2). Suppose $VarY(n) = O(1)$. Then*

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

If, additionally, 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 $\omega(G(n)) = \max\{\omega'(G(n)), \omega_R(G(n))\}$ and $\omega_R(G(n)) \leqslant 3$ whp.

By Markov's inquality, condition (4) is satisfied by a uniformly square integrable sequence $\{Y(n)\}$.

Next, we state a result about the size of the largest monochromatic clique. For this purpose we can relate the random intersection graph to the *balls in bins* model. Let every vertex $v \in V$ throw $|S_v|$ 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. The asymptotics of $M(N, m)$ are well known, see, e.g., Section 6 of Kolchin et al [17].

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

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

$$d_{TV}(\omega'(G(n)), M(\lfloor n\mathbb{E}\,X(n) \rfloor, m)) \to 0.$$

The proof of Lemma 1.3 is straightforward, but technical; it can be found in [18]. The case $m = o(n)$ is omitted, since:

Remark 1.4. No sequence of random intersection graphs $\{G(n)\}$ satisfies $m = o(n)$, $\mathbb{E}\,Y(n) = \Theta(1)$ and $VarY(n) = O(1)$ simultaneously.

Proof. Suppose the above relations are satisfied. Since $X = X(n)$ is a non-negative integer, we have $\mathbb{E}\,X^2 \geqslant \mathbb{E}\,X$. 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. □

Let us discuss the relation between the clique number and the clustering coefficient. We recall that the *(global) clustering coefficient* of a graph G is the conditional probability

$$\mathbb{P}(v_1^* \sim v_2^* | v_1^* \sim v_3^*, v_2^* \sim v_3^*), \tag{5}$$

where (v_1^*, v_2^*, v_3^*) is a triple of vertices of G drawn uniformly at random and $v_i^* \sim v_j^*$ is the event that v_i^* and v_j^* are adjacent in G. It was shown in [7] that the clustering coefficient of $G(n, m, P)$ approximately equals to

$$\frac{\mathbb{E}\, X(n)}{\mathbb{E}\, X(n)^2} = \frac{n^{1/2}\mathbb{E}\, Y(n)}{m^{1/2}\mathbb{E}\, Y^2(n)}.$$

In particular, it attains a non-trivial asymptotic value when $m = \Theta(n)$ and $\mathbb{E}\, Y^2(n) = \Theta(1)$. In the latter case Theorem 1.2 and Lemma 1.3 together with the asymptotics for $M(N, m)$ (Theorem II.6.1 of [17]), imply that

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

It is here where the positive clustering coefficient comes into play: sparse random intersection graphs have cliques of unbounded size even when $Var(D_1(n)) = \Theta(1)$, see Lemma 1.7(i). 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 [14], with bounded degree variance, the largest clique whp has at most 4 vertices.

For both of our main results, Theorem 1.1 and Theorem 1.2, we have corresponding simple polynomial-time 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 [14]: 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 bounded degree variance we propose the following algorithm: for each pair of adjacent vertices, check if their common neighbours form a clique. Output the vertices of a largest such clique together with the corresponding pair. 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 (1) for the tail of the normalised subset size $Y(n)$, we may consider a similar condition for the tail of $D_1(n)$: there are constants $\alpha' > 1, \epsilon' > 0$ and a slowly varying function $L'(x)$ such that for any sequence t_n with $n^{1/2-\epsilon'} \leqslant t_n \leqslant n^{1/2+\epsilon'}$

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

The following two lemmas are proved in [18]. Let $\mathbb{I}_A$ be the indicator of A.

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

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

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 (1) holds. In that case,* $\alpha' = \alpha$ *and* $L'(t) = d^{\alpha/2}L(t)$.

Thus, under a mild additional assumption (7), condition (1) 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*

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

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

Our last lemma exposes a close relationship between $Y(n)$ and $D_1(n)$ for $m = \Theta(n)$ under a natural uniform integrability requirement. A recent further work [19] on general sparse random graphs shows that uniform integrability of an appropriate degree moment is a necessary and sufficient condition to be able to approximate global subgraph count statistics by a small number of local samples.

Lemma 1.7. *Assume that* $\{G(n)\}$ *is a sequence of random intersection graphs,* $m \to +\infty$. *Let* k *be a positive integer.*

(i) Assume $m = \Theta(n)$. *Then* $\mathbb{E}\,D_1(n)^k = \Theta(1)$ *and* $D_1(n)^k$ *is uniformly integrable if and only if* $\mathbb{E}\,Y(n)^k = \Theta(1)$ *and* $Y(n)^k$ *is uniformly integrable.*

(ii) Assume $m = o(n)$. *If* $\mathbb{E}\,D_1(n) = \Theta(1)$ *then* $D_1(n)$ *is not uniformly integrable.*

Let us end the introduction by a discussion of some related literature. The largest intersecting subset problem for *uniform* hypergraphs was considered by Balogh, Bohman and Mubayi [3]. Although the motivation and the approach of [3] are different from ours, their result yields Theorem 1.2 for sparse and dense *uniform* random intersection graphs (where all sets have the same deterministic size).

Small cliques in random intersection graphs were studied in [16], where edge density thresholds for emergence of constant size cliques were determined, and in [23], where the Poisson approximation to the distribution of the number of small cliques was established. The clique number was studied by Nikoletseas, Raptopoulos and Spirakis [21], see also Behrisch, Taraz and Ueckerdt [4], in the case, where $m \approx n^{\beta}$, for some $0 < \beta < 1$. We note that the papers [4, 16, 21, 23] considered a particular random intersection graph with the binomial distribution $P \sim Bin(m, p)$. Sparse graphs $G(n, m, P)$ with $P \sim Bin(m, p)$ and $m = o(n)$ are covered by our Theorem 1.2. However, they are not very interesting: they consist of $n - o_P(n)$ isolated vertices, see the proof of Lemma 1.7(ii). Meanwhile, our results include the case $m = \Omega(n)$. They were obtained independently of [3, 4, 21][1]. In the bounded variance regime we, as well as [3, 4, 21], discovered the same universal phenomenon: the largest clique is whp formed by a single attribute.

Similarly as in [5, 14], we have a kind of "phase transition" as the tail index α for the random subset size (or vertex degree) varies, see (1). Assume, for example, that $m = \Theta(n)$ and P is asymptotically a Pareto distribution. 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. It would be interesting to determine the clique number in the case $\alpha = 2$. Our proofs in Section 2 and the results of [3] show that as $G(n)$ becomes denser, the property that the largest clique is generated by a single attribute ceases to hold. For such dense $G(n)$ the exact asymptotics of $\omega(G(n))$ remains an interesting open problem (but see the recent work [11] and the references therein).

2 Power-law tails

2.1 Proof of Theorem 1.1

We start by introducing some notation. Given a family of subsets $\{S_v, v \in V'\}$ of an attribute set W', we denote by $G(V', W')$ the *intersection graph* on the vertex set V' defined by this family: $u, v \in V'$ are adjacent (denoted $u \sim v$) whenever $S_u \cap S_v \neq \emptyset$. Let $H = (V_H, E_H)$ be a hypergraph where the set of (hyper-)edges E_H can be a multiset. Recall that for $w \in W'$, $T_w = \{v \in V' : w \in S_v\}$. We say that $G = G(V', W')$ *contains a copy* of H on $S \subseteq V'$ if there is a bijection $\sigma : V_H \to S$ and an injection $f : E_H \to W'$ such that $\sigma(e) \subseteq T_{f(e)}$ for each $e \in E_H$. Here $\sigma(e) = \{\sigma(x) : x \in e\}$. A hypergraph H corresponds to a graph $C(H)$ on the vertex set V_H, obtained by replacing each edge $e \in E_H$ with a clique on e (we merge repeated edges). H is called a *clique cover* of $C(H)$. G contains a subgraph isomorphic to G' if and only if it contains a copy of some clique cover of G', see [16]. The *number of copies*, $Q(H, G)$ of H in G is the number of different tuples (S, σ, f) that realise a copy of H in G. For $v \in V_H$, its *degree* in H is $d_H(v) = |\{e \in E_H : v \in e\}|$. If G contains a copy of H, where H is a graph (each edge is a pair), we say that G *contains a rainbow* H; this extends the above definition of a rainbow clique. The next simple bound generalises an estimate obtained by Karoński et al [16].

Lemma 2.1. *Let H be any hypergraph with h vertices of degrees $d_1, \dots, d_h$. Suppose H has k edges. Let $G = G(n, m, P)$ be a random intersection graph, and let X be distributed according to P.*

$$\mathbb{E}\, Q(H, G) \leqslant n^h m^k \prod_{i=1}^{h} \frac{\mathbb{E}\, X^{d_i}}{m^{d_i}}.$$

Proof. Let $S_1, \dots, S_n$ be the random sets of G. Given $|S_i|$, the probability that the set S_i contains d_i fixed attributes is $(|S_i|)_{d_i}/(m)_{d_i}$. Thus, the probability that a copy of H is

[1] They were first presented at the Lithuanian young scientists' conference in February, 2012, see [13].

realised by fixed σ and f is

$$\mathbb{E}\prod_{i=1}^{h}\frac{(|S_i|)_{d_i}}{(m)_{d_i}} \leqslant \prod_{i=1}^{h}\frac{\mathbb{E}\,X^{d_i}}{m^{d_i}}.$$

The claim follows by summing over all $(m)_k = m(m-1)\dots(m-k+1)$ choices of σ and $(n)_h$ choices of f. □

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 $V(G) = V(R)$ unless stated otherwise. Let t be a positive integer and let R be an arbitrary 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 $e = xy$ of the set $E(H) \setminus E(R)$ can be prescribed an attribute w_e such that $x, y \in T_{w_e}$ 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 $|S_v|$ of the attribute set has binomial distribution $P \sim Bin(m, p)$. 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 $|S_v|$ of random sets are concentrated around their mean value $\mathbb{E}\,|S_v| = mp$. An application of Chernoff's bound (see, e.g., [20])

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

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

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

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

Let us prove Theorem 1.1. For a number $\epsilon_1 \in (0, \epsilon_0)$, where ϵ_0 is defined in (1), and each n define subgraphs $G_i \subseteq G(n)$, $i = 0, 1, 2$, induced by the vertex sets

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

respectively. Here

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

with $e_1 = e_1(n) = \max(0, \ln L((n \ln n)^{1/2}))$. Note that $e_1 = 0$ if $L(x) = 1$ is constant. Write

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

We will prove three lemmas. Let β be as in Theorem 1.1.

Lemma 2.2. $\omega(G_2) \geqslant (1 - o_P(1))\,(1-\alpha/2)^{-\alpha/2}\,K$.

Lemma 2.3. *If* $\epsilon_1 < \frac{\beta}{6}$ *then there is* $\delta > 0$ *such that* $\mathbb{P}\left(\omega(G_0) \geqslant n^{1-\alpha/2-\delta}\right) \to 0$.

Lemma 2.4. *If* $\epsilon_1 < \frac{\beta-2+\alpha}{24}$ *then* $\omega(G_1) = o_P(K)$.

Our proof of Lemma 2.3 (Lemma 2.4) works for any $m = \Omega(n^\beta)$ with $\beta > \alpha - 1$ ($\beta > 2 - \alpha$). The proof of Lemma 2.2 works for arbitrary $m = m(n) \geqslant 1$.

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 $\omega(G_2) \leqslant \omega(G) \leqslant \omega(G_0) + \omega(G_1) + \omega(G_2)$ and Lemmas 2.2-2.4. □

2.2 Proof of Lemma 2.2

In this section we use ideas from [14] to give a lower bound on the clique number. We first note the following auxiliary facts.

Lemma 2.5. *Suppose* $a = a(n), b = b(n)$ *are sequences of positive real numbers such that* $0 < \ln b + a \to +\infty$. *Let* $z = z(n)$ *be the positive root of* $a - \ln z - bz^2 = 0$. *Then* $z \sim ((a + 0.5\ln(2b))/b)^{1/2}$.

Proof. Changing the variables $t = 2bz^2$ we get $t + \ln(t) = 2a + \ln(2b)$. From the assumption it follows that $t + \ln t \sim t$, which proves the claim. □

Lemma 2.6 ([10])**.** *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 2.2. Write $N = |V_2|$ and let

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

be the vertices of V_2 listed in an arbitrary order. We consider a greedy algorithm for finding a clique in G proposed by Janson, Łuczak and Norros [14]. 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, so that $\omega(G_2) \geqslant |A^N|$.

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

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

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 (1) we obtain for $N \sim Bin(n, q)$ 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 estimates $L((n/m)^{1/2}\theta) \sim L(\sqrt{n \ln n})$ and $\ln L(\sqrt{n \ln n}) = o(\ln n)$, see Lemma 2.6. 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 follows from the bound $\mathbb{E}\,(L_\theta/N) = o(1)$, shown below (define this ratio to be 0 on the rare event $N = 0$).

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

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

where

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

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

Let us briefly explain the intuition for the choice of θ in the above proof. For simplicity assume $L(x) = 1$ for all x 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

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

After we write the last relation as in Lemma 2.5 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 2.5 that the solution θ satisfies

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

2.3 Proof of Lemma 2.3

Before proving Lemma 2.3 we collect some preliminary results.

Lemma 2.7. *Let $h \geqslant 2$ be an 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 $pn^{1/(h-1)}m^{1/2} \to a \in \{0, +\infty\}$. Then*

$$\mathbb{P}(G \text{ contains a rainbow } K_h) \to \begin{cases} 0, & \text{if } a = 0, \\ 1, & \text{if } a = +\infty. \end{cases}$$

Proof. The case $a = +\infty$ follows from Claim 2 of [16]. For the case $a = 0$ we have, by the first moment method,

$$\mathbb{P}(G \text{ contains a rainbow } K_h) \leqslant \binom{n}{h}(m)_{\binom{h}{2}} p^{2\binom{h}{2}} \leqslant \left(n^{1/(h-1)}m^{1/2}p\right)^{h(h-1)} \to 0. \qquad \square$$

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

Lemma 2.8. *Let $1 < \alpha < 2$. Assume that $\{G(n)\}$ is a sequence of random intersection graphs satisfying (2), (1). Suppose that for some $\beta > \alpha - 1$ we have $m = \Omega(n^\beta)$. Then there is a constant $\delta > 0$ such that whp $\omega'(G(n)) \leqslant n^{1-\alpha/2-\delta}$.*

Proof. Let $X = X(n)$ and $Y = Y(n)$ be as defined above (2). 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 $T_w = \{v : w \in S_v\}$ is binomially distributed

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

we have, for any positive integer k

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

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

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

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

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

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

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

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 (as in our more general definition given above.)

Lemma 2.9 ([2]). *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 K_h.*

Proof of Lemma 2.3. 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 the positive constants δ and c are from Lemmas 2.8 and 2.9, respectively. We first show that

$$\mathbb{P}(G_0 \text{ contains a rainbow } K_h) = o(1). \tag{13}$$

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 2.7 implies

$$\mathbb{P}(\tilde{G} \text{ contains a rainbow } K_h) = o(1). \tag{14}$$

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 $|\tilde{S}_v| > \theta_1$, 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, since $\epsilon_1 < \beta/6$, (13) follows from (14) and the fact that whp $\min_v |\tilde{S}_v| > \theta_1$ by (12) applied to $\tilde{G}$ with $y = m^{1/3}$.

Next, we colour every edge xy of G_0 by an arbitrary element of $S_x \cap S_y$ and observe that the inequality $\omega'(G(n)) \leqslant t$ (which holds with probability $1 - o(1)$, by Lemma 2.8) implies that the obtained colouring is t-good. Furthermore, by Lemma 2.9, every k-clique of G_0 contains a rainbow clique; however the probability of the latter event is negligibly small by (13). We conclude that $\mathbb{P}(\omega(G_0) \geqslant k) = o(1)$ thus proving the lemma. □

After the submission of our paper, we became aware that Nikoletseas, Raptopaulos and Spirakis [21] independently also used the idea that a large clique must contain a large rainbow clique to analyse the clique number in the binomial random intersection graph. The proof of [21] heavily relies on the properties of the binomial distribution, and does not seem to lead to an alternative proof of Lemma 2.3.

2.4 Proof of Lemma 2.4

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

Lemma 2.10. *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 S be a random subset of $[m]$ of size $s \geqslant k$. Suppose that $a_1 + \cdots + a_k \leqslant m$. Then the probability*

$$\mathbb{P}\left(\{S \cap A_1, \ldots, S \cap A_k\} \textit{ has a system of distinct representatives}\right) \tag{15}$$

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

Proof. Call any of $\binom{m}{s}$ 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

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

Suppose the claim is false. Out of all families that maximize (15) pick a family $\mathcal{F}$ with smallest $d(\mathcal{F})$. Then $d(\mathcal{F}) > 0$ and we can assume that there is an element $x \in [m]$ such that $x \in A_1 \cap A_2$. Since $\sum_{i=1}^{k} |A_i| \leqslant m$, 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 $c \in \mathcal{C}$ 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 $|\mathcal{C}_{DR}(\mathcal{F}')| \geqslant |\mathcal{C}_{DR}(\mathcal{F})|$. But $d(\mathcal{F}') \leqslant d(\mathcal{F}) - 1$, which contradicts our assumption about the minimality of $d(\mathcal{F})$. □

The next lemma is a version of a result of Erdős and Rényi about the maximum clique of the random graph $G(n, p)$ (see, e.g., [15]).

Lemma 2.11. *Let $n \to +\infty$. Assume that a sequence p_n with $p_n \in [0, 1]$ converges 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 a sequence of arbitrary graphs $\{R_n\}$ with $V(R_n) = [n]$ and $e(R_n) \leqslant r_n$ the number Q_n of cliques of size $\lfloor \tilde{K}(1 + \delta_n) \rfloor$ in $G(n, p_n) \vee R_n$ satisfies

$$\mathbb{E}\, Q_n \leqslant \epsilon_n.$$

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

$$\mathbb{E}\, Q_n \leqslant \binom{n}{a} p^{\binom{a}{2} - r} \leqslant \left(\frac{en}{a}\right)^a p^{\frac{a(a-1)}{2} - r} = e^{aB}, \tag{16}$$

where, by the inequality $\ln p \leqslant -(1 - p)$, for n large enough,

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

□

Lemma 2.12. *Let $\{G(n)\}$ be a sequence of binomial random intersection graphs $G(n) = G(n,m,p)$, 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. Set $\bar{K} = 2e^{mp^2} \ln n$. Assume that $r_n \ll \bar{K}^2$ and*

$$mp^2 \to +\infty, \qquad \ln n \ll mp, \qquad \bar{K}p \to 0, \qquad \bar{K} \leqslant n/2. \tag{17}$$

There are positive sequences $\{\epsilon_n\}, \{\delta_n\}$ converging to zero such that $\delta_n \bar{K} \to +\infty$ and for an arbitrary sequence of graphs $\{R_n\}$ with $V(R_n) = V(G(n))$ and $e(R_n) \leqslant r_n$

$$\mathbb{P}\left(Rainbow(G(n), R_n, \bar{K}(1+\delta_n))\right) \leqslant \epsilon_n, \qquad n = 1, 2, \ldots \tag{18}$$

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 $\varphi_n \to +\infty$ 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 $G(n)$ to the size $M = \lfloor mp + x_n \rfloor$. Denote

$$\bar{S}_v = \begin{cases} S_v, \text{ if } |S_v| \leqslant M, \\ M \text{ element random subset of } S_v, \text{ otherwise.} \end{cases}$$

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

$$\mathbb{P}(B) = 1 - o(1). \tag{19}$$

Now, let $t \in [\bar{K}, 2\bar{K}]$ and let $T = \{u_1, \ldots, u_t\}$ 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 $i \in \{1, \ldots, t\}$, let $T_i \subseteq \{u_1, \ldots u_{i-1}\}$ denote the subset of vertices which are not adjacent to u_i in R_T. 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 $A_T(i)$, $1 \leqslant i \leqslant t$ hold simultaneously. We shall prove below that whenever n is large enough

$$\mathbb{P}(A_T) \leqslant \left(1 - (1-p)^M\right)^{\binom{t}{2} - e(R_T)}. \tag{20}$$

Next, proceeding as in Lemma 2.11 we find positive sequences $\{\delta'_n\}$, $\{\epsilon'_n\}$ converging to zero such that the number Q'_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 $\mathbb{E}\, Q'_n \leqslant \epsilon'_n$. For this purpose, we apply (16) to a' and $p' = 1 - (1-p)^M$, and use (20). We remark that $a' = \bar{K}(1 + \delta''_n)$, where $\{\delta''_n\}$ converges to zero and $\delta''_n \bar{K} \to +\infty$. Indeed, we have $\delta'_n \ln n/(1-p)^M \to +\infty$, by

Lemma 2.11, and we have $(1-p)^M = e^{-mp^2 - O(px_n + mp^3)}$ with $px_n + mp^3 = o(1)$. In particular, for large n, we have $a' \in [\bar{K}, 2\bar{K}]$.

The key observation of the proof is that the events B and $Rainbow(G(n), R_n, a')$ imply $Q'_n > 0$. Hence, by Markov's inequality,

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

Finally, invoking (19) we obtain (18).

It remains to show (20). 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 $1 \leqslant i \leqslant t$,

$$\mathbb{P}(A_T(i) | A_T(1), \ldots, A_T(i-1)) \leqslant \left(1 - (1-p)^M\right)^{|T_i|}. \tag{21}$$

Now (20) follows from the simple identity $\sum_{1 \leqslant i \leqslant t} |T_i| = \binom{t}{2} - e(R_T)$. Let us prove (21). For this purpose we apply Lemma 2.10. We first condition on $\{\bar{S}_u,\ u \in T_i\}$ and the size $|S_{u_i}|$ of S_{u_i}. By Lemma 2.10 the conditional probability

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

is maximized when the sets $\bar{S}_u$, $u \in T_i$ are pairwise disjoint (at this point we verify that $\bar{K}p \to 0$ and $t \leqslant 2\bar{K}$ implies the condition of Lemma 2.10 that $\sum_{u \in T_i} |\bar{S}_u| \leqslant tM < m$, for large n). Secondly, we drop the conditioning on $|S_{u_i}|$ and allow S_{u_i} to choose its element independently at random with probability p. In this way we obtain (21). □

Lemma 2.13. *Let $\{G(n)\}$ be a sequence of random binomial intersection graphs $G(n) = G(n, m, p)$, 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 (17) 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 $|T_w| \geqslant 3$, 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 (w covers an edge xy if $x, y \in T_w$). 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 $p' = \mathbb{P}(|U| \leqslant 2)$ 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 $|\Delta_z| - |T_z| \geqslant 1$ 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 are independent of G_S, G_B and T'_z, $z \in W_S$. The collection of random sets $\{T'_w, w \in W_B \cup W_S\}$ defines 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 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

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

Invoking the simple bound

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

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

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 $K' > 0$,

$$\mathbb{P}\left(\omega(G) \geqslant K'\right) \leqslant \mathbb{E}\, \mathbb{P}\left(\omega(G) \geqslant K' | G_B\right) \mathbb{I}_{\{e(G_B) \leqslant r_n\}} + \mathbb{P}\left(e(G_B) \geqslant r_n\right). \tag{22}$$

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

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

We choose $K' = \bar{K}(1 + \delta_n)$ and apply Lemma 2.12 to the conditional probability on the right. At this point we specify $\{\delta_n\}$ and find positive $\epsilon_n \to 0$ such that

$$\mathbb{P}\left(Rainbow(G', G_B, K') | G_B\right) \leqslant \epsilon_n$$

uniformly in G_B satisfying $e(G_B) \leqslant r_n$. Hence, (22) implies $\mathbb{P}\left(\omega(G) \geqslant \bar{K}(1 + \delta_n)\right) \leqslant \epsilon_n + o(1) = o(1)$. $\square$

Now we are ready to prove Lemma 2.4.

Proof of Lemma 2.4. Let

$$0 < \epsilon < 2^{-1} \min\{1,\ 1 - 2^{-1}\alpha,\ \beta - 2 + \alpha - 6\alpha\epsilon_1\}, \tag{23}$$

define $\theta > 0$ by $\theta^2 = (1 - \varepsilon - 2^{-1}\alpha)m \ln n$ and let $\bar{G}_1$ be the subgraph of G_1 induced by vertices $v \in V_1$ with $|S_v| \leqslant \theta$. Let $D = |V(G_1) \setminus V(\bar{G}_1)|$ denote the number of vertices of G_1 that do not belong to $\bar{G}_1$.

First, using (1) and Lemma 2.6 we estimate the expected value of D for $n \to +\infty$

$$\mathbb{E}\, D = n\left(\mathbb{P}(|S_v| \geqslant \theta) - \mathbb{P}(|S_v| \geqslant \theta_2)\right) \leqslant (h(\epsilon) + o(1))K. \tag{24}$$

Note that $h(\epsilon) := (1 - \epsilon - 2^{-1}\alpha)^{-\alpha/2} - (1 - 2^{-1}\alpha)^{-\alpha/2} \to 0$ can be made arbitrarily small by taking ϵ small enough. Next, we claim that for any ε satisfying (23),

$$\mathbb{P}\left(\omega(\bar{G}_1) \geqslant 4n^{1-2^{-1}\epsilon-2^{-1}\alpha} \ln n\right) = o(1). \tag{25}$$

Note that $n^{1-2^{-1}\epsilon-2^{-1}\alpha} \ln n \ll K$. Clearly, the lemma follows from (24), Markov's inequality and (23), since $\omega(G_1) \leqslant D + \omega(\bar{G}_1)$ and we can replace ϵ by a sequence $\epsilon'_n \to 0$ so that both terms are $o_P(K)$.

It remains to prove (25). Let $\bar{N}$ be a binomial random variable with distribution $Bin(n, \mathbb{P}(|S_v| > \theta_1))$, 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

$$\mathbb{P}(\bar{G}_1 \subseteq G') = 1 - o(1). \tag{26}$$

Such a coupling is possible because the events $A = \{$every vertex of G' is prescribed at least θ attributes$\}$ and $B = \{|V(\bar{G}_1)| \leqslant \bar{n}\}$ occur whp. 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 $|V(\bar{G}_1)| \sim Bin(n, \mathbb{P}(\theta_1 < |S_v| < \theta))$ and $\bar{N}$ so that $\mathbb{P}(|V(\bar{G}_1)| \leqslant \bar{N}) = 1$ and then invoke the bound $\mathbb{P}(\bar{N} \leqslant \bar{n}) = 1 - o(1)$, which follows from (1) and Chernoff's inequality.

Next we apply Lemma 2.13 to G' (the conditions of the lemma on $\bar{n}$, m and $\bar{p}$ can be easily checked) and obtain the bound

$$\mathbb{P}\left(\omega(G') > 4n^{1-2^{-1}\epsilon-2^{-1}\alpha} \ln \bar{n}\right) = o(1), \tag{27}$$

which together with (26) implies (25). □

One might ask if we could employ the results of [21] in the above proof, since the random intersection graph is binomial. The answer seems to be "no": the condition of $m = \bar{n}^{\beta}$ with $\beta < 1$ required by [21] is not satisfied. In fact $m/\bar{n}$ grows polynomially with n. Theorem 3 of [3] suggests G' is in the range where largest clique is not formed by a single attribute, and the precise asymptotics of $\omega(G')$ for such dense graphs remains an open problem.

3 Bounded variance

Proof of Theorem 1.2. Fix $c > 0$. Define truncated probability distribution $P'_n = P'_{n,c}$ by setting $P'_n(k) = P_n(k)$ if $k < \lfloor c\sqrt{m} \rfloor$ and $P'_n(\lfloor c\sqrt{m} \rfloor) = \sum_{i \geqslant \lfloor c\sqrt{m} \rfloor} P_n(i)$. Let $G_c(n)$ be the random intersection graph $G(n, m, P'_n)$ with sets $S'_1, \dots, S'_n$ defined as follows. Given a realisation of the random sets $S_1, \dots, S_n$ of $G(n)$, for each $v \in V(G(n)) = V(G_c(n))$ with $|S_v| \leqslant c\sqrt{m}$ set $S'_v = S_v$. For those v where $|S_v| > c\sqrt{m}$, let S'_v be a uniformly random subset of size $\lfloor c\sqrt{m} \rfloor$ from S_v, chosen for each v independently (conditioned on $S_1, \dots, S_n$). Write $X = X(n)$, $Y = Y(n)$, and let $X' = X'(n)$ have distribution P'_n.

The graph $G_c(n)$ is a subgraph of $G(n)$ and the edges that differ have at least one endpoint in the set $R_c = \{v \in V(G(n)) : |S_v| > c\sqrt{m}\}$. Also by Markov's inequality $|R_c| = O_P(1)$ since

$$\mathbb{E}\,|R_c| = n\mathbb{P}(X > c\sqrt{m}) \leqslant \frac{n\mathbb{E}\,X^2}{c^2 m} = \frac{\mathbb{E}\,Y^2}{c^2},$$

Let H_1 be the hypergraph on $\{1,2,3,4\}$ with edges $\{\{1,2,3\}, \{1,4\}, \{2,3,4\}\}$. Let H_2 be the hypergraph on $\{1,2,3,4\}$ with edges $\{\{1,2\}, \{1,3\}, \{1,4\}, \{2,3,4\}\}$. By Lemma 2.1 and trivial bounds $\mathbb{E}\,(X')^2 \leqslant \mathbb{E}\,X^2$, $\mathbb{E}\,(X')^3 \leqslant c\sqrt{m}\mathbb{E}\,X^2$, we get

$$\mathbb{E}\,Q(H_1, G_c(n)) \leqslant n^4 m^3 \left(\frac{\mathbb{E}\,(X')^2}{m^2}\right)^4 \leqslant \frac{(\mathbb{E}\,Y^2)^4}{m}; \tag{28}$$

$$\mathbb{E}\,Q(H_2, G_c(n)) \leqslant n^4 m^4 \left(\frac{\mathbb{E}\,(X')^2}{m^2}\right)^3 \frac{\mathbb{E}\,(X')^3}{m^3} \leqslant \frac{c(\mathbb{E}\,Y^2)^4}{\sqrt{m}}; \tag{29}$$

$$\mathbb{E}\,Q(K_4, G_c(n)) \leqslant n^4 m^6 \left(\frac{\mathbb{E}\,(X')^3}{m^3}\right)^4 \leqslant (c\mathbb{E}\,Y^2)^4. \tag{30}$$

Thus $G_c(n)$ whp contains no copy of H_1 or H_2. We claim that whp

$$\omega(G_c(n)) = \max(\omega'(G_c(n)), \omega_R(G_c(n))). \tag{31}$$

Indeed, let S be a set of vertices in $G_c(n)$, such that there is a clique on S, of the maximum size. Let Q be a subset of S, such that $\cap_{v \in Q} S'_v \neq \emptyset$ of the largest cardinality. (31) is trivially true when $|Q| \leqslant 1$, so we can assume $|S| \geqslant |Q| \geqslant 2$. If $|Q| = 2$, $G_c(n)$ contains a rainbow clique on S, so that $\omega(G_c(n)) = \omega_R(G_c(n))$. Suppose $\omega(G_c(n)) > \omega'(G_c(n))$. Then $|Q| < |S|$, and let x be an arbitrary vertex in $S \setminus Q$. The maximality of Q implies that $G_c(n)$ contains a copy of either H_1 or H_2 on a subset of Q (with 1 mapped to x). But we have shown that this does not occur whp. So whp $|Q| \in \{2, |S|\}$. This finishes the proof of (31).

A rainbow K_h in $G_c(n)$ yields $(h)_4$ copies of K_4. Therefore by (30) and Markov's inequality for $h \geqslant 4$

$$\mathbb{P}(\omega_R(G_c(n)) \geqslant h) \leqslant \mathbb{P}\left(Q(K_4, G_c(n)) \geqslant (h)_4\right) \leqslant \frac{c^4(\mathbb{E}\,Y^2)^4(1 + o(1))}{(h)_4},$$

so $\omega_R(G_c(n)) = O_P(1)$. Using (31) and our bound for $|R_c|$, we get

$$\omega'(G(n)) \leqslant \omega(G(n)) \leqslant \omega'(G_c(n)) + \omega_R(G_c(n)) + |R_c| = \omega'(G(n)) + O_P(1).$$

Finally, suppose (4) holds. Our proof above holds for any fixed $c > 0$. By (30) for arbitrarily small $\epsilon > 0$ we can pick $c > 0$ such that $\mathbb{P}(G_c(n)$ contains a copy of $K_4) < \epsilon + o(1)$, $\mathbb{P}(|R_c| > 0) \leqslant \mathbb{E}\,|R_c| \to 0$ and (31) holds. This implies that whp

$$\omega(G(n)) = \max(\omega'(G(n)), \omega_R(G(n))) \text{ and } \omega_R(G(n)) \leqslant 3. \qquad \square$$

If the condition $m \to +\infty$ is dropped in the second part of Theorem 1.2, then we have whp $\omega(G(n)) \leqslant \omega'(G(n)) + c$ for $c > 0$ [18]. It is easy to see that we can take $c = 0$ when $\omega'(G(n)) > 4$ whp. Otherwise, it seems that $c = 2$ is best possible.

4 Algorithms

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, [7]. 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 sparse random intersection graph G.

The Greedy-Clique algorithm of [14] 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 the vertices of G in order of decreasing degree
    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 the 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 (1) 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 2.2, but it does not follow from Lemma 2.2, since Greedy-Clique is not allowed to know the attribute subset sizes. The proof of Proposition 4.1 is given in [18], p. 87.

For random intersection graphs with bounded degree variance we suggest the following simple algorithm, which, as we shall see, can be implemented to run in expected time $O(n)$.

Mono-Clique(G):
    for $uv \in E(G)$
        $D(uv) \leftarrow |\Gamma(u) \cap \Gamma(v)|$
    for $uv \in E(G)$ in order of decreasing $D(uv)$
        $S \leftarrow \Gamma(u) \cap \Gamma(v)$
        if S *is a clique* **then**
            return $S \cup \{u, v\}$
    return $\{1\} \cap V(G)$ (if all fails, return a trivial clique)

Here $\Gamma(v)$ denotes the set of neighbours of v. Note that $D(uv) = |\Gamma(u) \cap \Gamma(v)|$ is the number of triangles that contain the edge uv. Below we also discuss the clique percolation method [22] which achieves a similar performance on sparse random intersection graphs. For alternative algorithms, proved to work under different/stronger conditions (for example, $P \sim Bin(m, p)$), or aiming to reconstruct all of the monochromatic cliques $|T_w|$, see [4, 21].

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 whp* $\omega(G(n)) \geqslant \omega_n$*, 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. Assume for now that

$$\mathbb{E}\, Z = O(1), \tag{32}$$

which will be proved later.

Let $K \subseteq [n]$ be the (lexicographically first) largest clique of $G(n)$. Denote $s = |K|$. If $s \leqslant 2$ or there is a pair $\{x, y\} \subseteq K$, $x \neq y$ such that $G(n)[\Gamma(x) \cap \Gamma(y)]$ is a clique, then the algorithm returns a clique of size s (since we consider edges in the decreasing order of $D(uv)$). 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 $\binom{s}{2} \leqslant Z$. 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

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

and using (32)

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

Also if $\omega(G(n)) \geqslant \omega_n$ whp, then by (32) and Markov's inequality

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

It remains to show (32). Let H_1, H_2, H_3 be the hypergraphs on vertex set $\{1, 2, 3, 4\}$ with edge sets $\{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 1\}\}$, $\{\{1, 2, 3\}, \{3, 4\}, \{4, 1\}\}$ and $\{\{1, 2, 3\}, \{3, 4, 1\}\}$ respectively. It is easy to see that

$$Z \leqslant Q(H_1, G(n)) + Q(H_2, G(n)) + Q(H_3, G(n)).$$

Write $X = X(n)$. By Lemma 2.1 since $m = \Omega(n)$ and $\mathbb{E}\,Y(n)^2 = O(1)$

$$\mathbb{E}\,Z \leqslant n^4 m^4 \left(\frac{\mathbb{E}\,X^2}{m^2}\right)^4 + n^4 m^3 \frac{\mathbb{E}\,X}{m} \left(\frac{\mathbb{E}\,X^2}{m^2}\right)^3 + n^4 m^2 \left(\frac{\mathbb{E}\,X}{m}\right)^2 \left(\frac{\mathbb{E}\,X^2}{m^2}\right)^2 = O(1). \quad \square$$

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

Proof. Consider the running time of the first loop (i.e., precomputing $D(uv)$ for all edges uv). We can assume that each list in the adjacency list structure is 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

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

To see the last bound, notice that the sum of degree squares in a graph counts the number of 2-paths plus the number of edges in the graph. The number of 2-paths in $G(n)$ is bounded by the number of copies of hypergraphs $\tilde{H}_1$ and $\tilde{H}_2$ on $\{1, 2, 3\}$, where $\tilde{H}_1$ has edges $\{\{1, 2\}, \{2, 3\}\}$ and $\tilde{H}_2$ has a single edge $\{1, 2, 3\}$. Using our notation, $\sum_{v \in V} D_v^2 \leqslant Q(\tilde{H}_1, G(n)) + Q(\tilde{H}_2, G(n)) + Q(K_2, G(n))$. Applying Lemma 2.1 and the bound $m = \Omega(n)$, see also [16, 19],

$$\mathbb{E} \sum_{v \in V} D_v^2 \leqslant \frac{n^3 \mathbb{E}\,X^2 (\mathbb{E}\,X)^2}{m^2} + \frac{n^3 (\mathbb{E}\,X)^3}{m^2} + \frac{n^2 (\mathbb{E}\,X)^2}{m} = O(n).$$

The second loop can be implemented so that the edge uv with the next largest value of $D(uv)$ is found at each iteration (we avoid sorting the list of edges in advance to keep the running time linear). In this way picking the next edge requires at most $ce(G(n))$ steps for a constant c. We recall that the number of edges $uv \in E(G)$ with $\Gamma(u, v) := \Gamma(u) \cap \Gamma(v) \neq \emptyset$ 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 $cZe(G(n))$ and

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

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

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

Finally, invoking the simple bound $\mathbb{E}\, e(G(n)) = \binom{n}{2} \mathbb{P}(u \sim v) = O(n)$, and (32) we get

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

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. We can test if a set S is a clique in time proportional to $e(G(n))$ by scanning the edges incident to vertices in S once and verifying that the number of neighbours in S for each vertex $v \in S$ is $|S| - 1$. When $S = \Gamma(u, v)$ is not a clique, i.e. $x, y \in S$ and $xy \notin E(G(n))$, the event $\mathcal{C}(u, x, v, y)$ holds. Thus by the previous bound, the total expected time spent in the second loop is again $O(\mathbb{E}\, Z\, e(G(n))) = O(n)$. □

Combining the next lemma with Lemma 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 Lemma 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 $\{u, v\} \cup (\Gamma(u) \cap \Gamma(v)) = M$.*

The proof can be found in [18] p. 93.

Finally, as an alternative, we show that a popular, robust and simple *clique percolation method* [22] can be used to find all largest cliques in the bounded degree variance case. Following [22], define a *k-clique-community* in a graph G as a union of all k-cliques (complete subgraphs of size k) that can be reached from each other through a series of adjacent k-cliques, where adjacency means sharing $k-1$ vertices.

Call a 3-clique community H of an intersection graph $G = G(V, W)$ *monochromatic* if it is a subgraph of a monochromatic clique in G.

Lemma 4.5. *Let $\{G(n)\}$ be as in Lemma 1.3. The largest 3-clique community which is not monochromatic has size $O_P(1)$.*

Thus if $\omega'(G(n)) \geqslant \omega_n$ whp for some $\omega_n \to \infty$, the largest 3-clique communities are the largest cliques of $G(n)$ (which are monochromatic). The Mono-Clique algorithm can be used to find them efficiently.

Proof of Lemma 4.5. Let C be a 3-clique community which is not monochromatic. We can assume $|V(C)| \geqslant 4$. Note that each monochromatic clique T_w either shares all of its edges or none of its edges with C. By the definition of a 3-clique community and the fact that C is not monochromatic, whenever $T_w \subset V(C)$, C must contain a triangle on vertices $\{x, y, z\}$ with $x, y \in T_w$ and $z \notin T_w$.

Let H_1, H_2, H_3 be as in the proof of Theorem 4.2. Consider a vertex v in C. First suppose there is w such that $T_w \subset V(C)$, $v \in T_w$ and $|T_w| \geqslant 3$. Let x, y, z be a triangle in C such that $x, y \in T_w$ and $z \in V(C) \setminus T_w$. If $v \notin \{x, y\}$ then $G(n)$ contains a copy of H_2 or a copy of H_3 on vertices $\{v, x, y, z\}$. If $v \in \{x, y\}$, we can take another vertex $v' \in T_w \setminus \{x, y\}$ and get a copy of H_2 or H_3 on $\{v', x, y, z\}$.

Now suppose all T_w such that $T_w \subset C$ and $v \in T_w$ have size 2. Since $|C| \geqslant 4$, it is easy to see that for some $x, y, z \in V(G(n))$ there is a copy of H_1 or H_2 on $\{v, x, y, z\}$.

For each $v \in V(C)$ we can assign a copy of H_1, H_2 or H_3 containing v. Since each copy can be obtained at most 4 times, by Markov's inequality and the bounds $\mathbb{E}\, Q(H_i, G(n)) = O(1)$, $i \in \{1, 2, 3\}$ shown in the proof of Theorem 4.2

$$|V(C)| \leqslant 4\,(Q(H_1, G(n)) + Q(H_2, G(n)) + Q(H_3, G(n)) = O_P(1). \qquad \square$$

5 Proof of Lemma 1.7

Proof of Lemma 1.7. Write $X = X(n)$, $Y = Y(n)$, $D_1 = D_1(n)$ and assume $X = |S_1|$. For arbitrary $n, m \to +\infty$ suppose $\mathbb{E}\, D_1 = \Omega(1)$ and D_1 is uniformly integrable. We claim

$$\mathbb{P}(X > 0) = \Omega(1). \tag{33}$$

If this did not hold, then there would be a subsequence of n, such that for any $C > 0$ we would have $\mathbb{E}\, D_1 \mathbb{I}_{D_1 \leqslant C} \leqslant C\mathbb{P}(D_1 > 0) \leqslant C\mathbb{P}(X > 0) \to 0$. This contradicts the assumption about D_1.

Proof of (i). For the "if" part, suppose $\mathbb{E}\, Y^k = O(1)$ and Y^k is uniformly integrable. We claim that D_1^k must be uniformly integrable. If not, from a subsequence of n such that

$\mathbb{E}\, D_1^k \mathbb{I}_{D_1 > \omega_n}$ has a limit in $(0, +\infty]$ with $\omega_n \to +\infty$, pick a subsubsequence A, such that $\mathbb{E}\, Y_n^k \to y > 0$ and Y_n^k converges weakly to a random variable with mean y as $n \to +\infty$, $n \in A$. For $k = 1, 2$ and $m = \Omega(n)$, using Theorem 2.1 of [7], Lemma 1.5 and Lemma 1.6 we get that D_1^k is uniformly integrable on A, a contradiction. For all $k \geqslant 1$ and $m = \Theta(n)$ Lemma 4.7 and Theorem 3.1 of [19] similarly shows that D_1^k is uniformy integrable. The proof that $\mathbb{E}\, D_1^k = \Theta(1)$ follows similarly.

We will prove the "only if" part by contradiction. Suppose $\mathbb{E}\, D_1^k = \Theta(1)$ and D_1^k is uniformly integrable, but Y^k is not. Then, since $m = \Theta(n)$, for some $\omega_n \to +\infty$ and a subsequence A we have

$$\lim_{n \to +\infty, n \in A} \mathbb{E}\, X^k \mathbb{I}_{X \geqslant \omega_n} > 0. \tag{34}$$

Below we will take limits over $n \to +\infty$, $n \in A$. We can assume $\omega_n = o(\sqrt{m})$. Write $p_0 = \liminf \mathbb{P}(X \in [1, \omega_n))$, and let $B \subseteq A$ be a subsequence that realises the infimum. First assume $p_0 = 0$. For $n \in B$ by (33) we have $\mathbb{P}(X \geqslant \omega_n) = \Omega(1)$. The probability $p(m, s, t)$ that two independent random subsets of sizes s and t of $[m]$ intersect satisfies

$$\frac{st}{m}\left(1 - \frac{st}{m}\right) \leqslant p(m, s, t) \leqslant \frac{st}{m}. \tag{35}$$

It follows by monotonicity of $p(m, s, t)$ and the linearity of expectation that

$$\mathbb{E}\, D_1 \geqslant (n-1)\mathbb{P}(X \geqslant \omega_n)^2 p(m, \omega_n, \omega_n) = \Omega\left(\frac{n\omega_n^2}{m}\right),$$

so $\mathbb{E}\, D_1 \to +\infty$ as $n \to +\infty$, $n \in B$. This is a contradiction, so it must be $p_0 > 0$.

Now consider arbitrary $n \in A$. Let $T = T(n)$ be the number of sets in $S_2, \dots, S_n$ of $G(n)$ that are non-empty. Define $Z = 0.2nm^{-1}p_0 X$. Let x be any integer with $x \geqslant \omega_n$ and $\mathbb{P}(X = x) > 0$. Given $X = x$, let $\tilde{D}$ have distribution $Bin(\lceil 0.5p_0 n \rceil, \frac{X}{m})$. We have $Z < 0.5\mathbb{E}\,(\tilde{D}|X = x)$ and $\mathbb{E}\,(\tilde{D}|X = x) \geqslant 0.5p_0 nm^{-1}\omega_n \geqslant c\omega_n$ for a constant $c > 0$ which does not depend on n or x. Applying a Chernoff bound we get

$$\mathbb{P}(\tilde{D} < Z | X = x) \leqslant e^{-\frac{1}{8}\mathbb{E}\,(\tilde{D}|X=x)} \leqslant e^{-\frac{1}{8}c\omega_n}.$$

Since the probability that a uniformly random subset of $[m]$ of size X intersects a fixed non-empty set is at least $\frac{X}{m}$, we get

$$\begin{aligned}
&\mathbb{P}(D_1 \geqslant Z | X = x) \geqslant \mathbb{P}(D_1 \geqslant Z | X = x, T \geqslant 0.5p_0 n)\mathbb{P}(T \geqslant 0.5p_0 n | X = x) \\
&\geqslant \mathbb{P}(\tilde{D} \geqslant Z | X = x)\mathbb{P}(T \geqslant 0.5p_0 n) \geqslant (1 - e^{-\frac{1}{8}c\omega_n})\mathbb{P}(T \geqslant 0.5p_0 n).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{E}\,(D_1^k \mathbb{I}_{D_1 \geqslant 0.2nm^{-1}p_0\omega_n}) \geqslant \mathbb{E}\,\mathbb{E}\,(D_1^k \mathbb{I}_{D_1 \geqslant Z}\mathbb{I}_{X \geqslant \omega_n} | X) \\
&\geqslant \mathbb{E}\, Z^k \mathbb{I}_{X \geqslant \omega_n}(1 - e^{-\frac{1}{8}c\omega_n})\mathbb{P}(T \geqslant 0.5p_0 n) = \Theta(\mathbb{E}\, X^k \mathbb{I}_{X > \omega_n}) = \Omega(1).
\end{aligned}$$

This is a contradiction to the fact that D_1^k is uniformly integrable. Thus (34) does not hold and Y^k is uniformly integrable. Now the fact that $\mathbb{E}\, Y^k = \Theta(1)$ follows by contradiction as we apply Lemma 4.7 and Theorem 3.1 of Kurauskas [19] (or Lemma 1.5 and Lemma 1.6 for $k = 1, 2$) with $\tilde{Y} = Y \mathbb{I}_{Y<C}$ and a large constant C. This finishes the proof of (i).

It is easy to check that the "only if" part fails for $n = o(m)$ and $k \geqslant 1$. To get a counterexample, consider X with $\mathbb{P}(X = m) = (mn)^{-1/2}$ and $\mathbb{P}(X = \lfloor m^{1/2} n^{-1/2} \rfloor) = 1 - (mn)^{-1/2}$.

Proof of (ii). Suppose the contrary, i.e., D_1 is uniformly integrable. By (33), $\mathbb{P}(X > 0) \geqslant a - o(1)$ for some $a > 0$, and by the law of large numbers, there are at least $0.5an$ non-empty sets of $G(n)$ whp. On the latter event, we see that $G(n)$ contains m disjoint cliques that together cover at least $0.5an$ vertices (group the non-empty sets of $G(n)$ according to the smallest attribute they contain). Using a standard convexity argument, the number of edges in $G(n)$ is at least

$$m \binom{\lfloor \frac{0.5an}{m} \rfloor}{2} = n \Omega\left(\frac{n}{m}\right),$$

with probability $1 - o(1)$. Thus $\mathbb{E}\, D_1 = 2\mathbb{E}\, e(G(n))/n \to +\infty$, which contradicts that $\mathbb{E}\, D_1 = \Theta(1)$. □

References

[1] R. Albert and A.L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47, 2002.

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

[3] J. Balogh, T. Bohman, and D. Mubayi. Erdős-Ko-Rado in random hypergraphs. *Comb. Probab. Comp.*, 18:629–646, 2009.

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

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

[6] M. Bloznelis, J. Jaworski, and V. Kurauskas. Assortativity and clustering coefficient of sparse random intersection graphs. *Electron. J. Probab.*, 18(38):1–24, 2013.

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

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

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

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

[11] M.M. Gauy, H. Hàn, and I.C. Oliveira. Erdős-Ko-Rado for random hypergraphs: asymptotics and stability. arXiv:1409.3634, 2014.

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

[13] http://www.infobalt.lt/lt/members/stories/i/7 (in Lithuanian), accessed 2016-06-03.

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

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

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

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

[18] V. Kurauskas. On two models of random graphs. PhD thesis, Vilnius University, 2013. Available at http://vddb.laba.lt/obj/LT-eLABa-0001:E.02~2013~D_20131216_081822-36288.

[19] V. Kurauskas. On local weak limit and subgraph counts for sparse random graphs. arXiv:1504.08103, 2015.

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

[21] 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, 728–739, 2012.

[22] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818, 2005.

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

[24] S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications. *Cambridge University Press*, Cambridge, 1994.

[25] D. J. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.