Theoretical Computer Science 412 (2011) 6750–6760

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

On the independence number and Hamiltonicity of uniform random intersection graphs[☆]

S. Nikoletseas [*], C. Raptopoulos, P.G. Spirakis

Computer Technology Institute, P.O. Box 1122, 26110 Patras, Greece
University of Patras, 26500 Patras, Greece

ARTICLE INFO

Article history:
Received 19 April 2010
Received in revised form 28 August 2011
Accepted 1 September 2011
Communicated by J. Díaz

Keywords:
Random graphs
Intersection graphs
Independent set
Hamiltonian cycle
Probabilistic methods

ABSTRACT

In the uniform random intersection graphs model, denoted by $G_{n,m,\lambda}$, to each vertex v we assign exactly λ randomly chosen labels of some label set $\mathcal{M}$ of m labels and we connect every pair of vertices that has at least one label in common. In this model, we estimate the independence number $\alpha(G_{n,m,\lambda})$, for the wide range $m = \lfloor n^{\alpha} \rfloor$, $\alpha < 1$ and $\lambda = O(m^{1/4})$. We also prove the Hamiltonicity of this model by an interesting combinatorial construction. Finally, we give a brief note concerning the independence number of $G_{n,m,p}$ random intersection graphs, in which each vertex chooses labels with probability p.

1. Introduction

Random intersection graphs $G_{n,m,p}$ were introduced by Karoński et al. [9] and Singer-Cohen [14]. In such graphs, each one of m labels is chosen independently with probability p by each one of n vertices, and there are edges between any vertices with overlaps in the labels chosen. Fill, Sheinerman and Singer-Cohen in [7] proved that the $G_{n,m,p}$ becomes statistically equivalent to an Erdős–Rényi random graph (in which every edge appears independently with some probability $\hat{p}$), when the number of labels m is quite large (in fact for $m \geq n^6$, but it was conjectured that the same holds for smaller m). However, the two models seem to behave quite differently when the number of labels is less than the number of vertices.

Godehardt and Jaworski [8] defined a different model called *uniform random intersection graphs model* $G_{n,m,\lambda}$. In this model, to each of the n vertices of the graph, a random subset of λ elements of a universal set of m elements in total is independently assigned. Two vertices u, v are then adjacent in the $G_{n,m,\lambda}$ graph if and only if their assigned sets of elements have at least one element in common. The $G_{n,m,\lambda}$ seems to behave similarly to $G_{n,m,p}$ when one can show concentration on the number of labels chosen by a vertex in the latter (which can happen for quite large λ). However, notice that for small values of λ such concentration results do not hold, and the statistical behavior of the two models is quite different. The main focus of this work is about Hamiltonicity and independent sets in the uniform random intersection graphs model $G_{n,m,\lambda}$. The last section of the paper will give a brief note on the independence number of $G_{n,m,p}$.

☆ This work was partially supported by the IST Programme of the European Union under contact number IST-2005-15964 (AEOLUS).
* Corresponding author at: Computer Technology Institute, P.O. Box 1122, 26110 Patras, Greece.
E-mail addresses: nikole@cti.gr (S. Nikoletseas), raptopox@ceid.upatras.gr (C. Raptopoulos), spirakis@cti.gr (P.G. Spirakis).

0304-3975/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2011.09.003

Importance and motivation. Random intersection graphs may be used to model several real-life applications characterized by local interactions quite accurately (compared to the $G_{n,\hat{p}}$ model where edges appear independently with probability $\hat{p}$). In particular, the $G_{n,\hat{p}}$ model seems inappropriate for describing some real world networks (like sensor and social networks) because it lacks certain features of those networks, such as a scale free degree distribution and the emergence of local clusters. One of the underlying reasons for this mismatch is its independence of the edges, in other words the missing transitivity that characterizes such networks: if vertices x and y exhibit a relationship of some kind in a real world network and so do vertices y and z, then this suggests a connection between vertices x and z, too.

For example, we consider the following scenario concerning efficient and secure communication in sensor networks: The vertices in our model correspond to sensor devices that blindly choose a limited number of resources among a globally available set of shared resources (such as communication channels, encryption keys etc.). Whenever two sensors select at least one resource in common (e.g. a common communication channel, a common encryption key), a communication link is implicitly established (represented by a graph edge); this gives rise to communication graphs that look like random intersection graphs.

Random intersection graphs in general and in particular the uniform random intersection graphs model $G_{n,m,\lambda}$ are relevant to and capture quite nicely social networking. Indeed, a social network is a social structure made of nodes (individuals or organizations) tied by one or more specific types of interdependency, such as values, visions, financial exchange, friends, conflicts, web links etc. Social network analysis views social relationships in terms of nodes and ties. Nodes are the individual actors within the networks and ties are the relationships between the actors.

Other applications may include oblivious resource sharing in a distributed setting, interactions of mobile agents traversing the web, etc. In fact there are practical situations where each communication agent (e.g. a wireless node) gets access only to some ports (statistically) out of a possible set of communication ports. When another agent also selects a communication port, then a communication link is implicitly established and this gives rise to communication graphs that look like random intersection graphs. Even epidemiological phenomena (like spread of disease) tend to be more accurately captured by those ''interaction-sensitive'' random graph models.

Related work. Uniform random intersection graphs were first considered by Godehardt and Jaworski in [8], where they focused on the distribution of the number of isolated vertices in $G_{n,m,\lambda}$, as well as the distribution of vertex degrees. The vertex degree distribution of general random intersection graphs (where the choice of the label sets S_v is made according to a general distribution) was studied independently by Blonzelis [3] and Deijfen and Kets [5]. Connectivity and communication security aspects of $G_{n,m,\lambda}$ in various important settings is studied in [12,2].

The question of how close $G_{n,m,p}$ and $G_{n,p}$ are for various values of m, p has been studied by Fill, Sheinerman and Singer-Cohen in [7]. In [11], the authors investigate expansion properties of $G_{n,m,p}$ and give tight bounds on the mixing and the cover time of random walks on instances of the random intersection graphs model. Algorithms for finding large independent sets in $G_{n,m,p}$ were proposed in [10] (however, no attempt was made to see how close the independent sets given by those algorithms are to optimal size). The authors of [6] find thresholds (that are optimal up to a constant factor) for the appearance of Hamilton cycles in random intersection graphs. The efficient construction of Hamilton cycles in $G_{n,m,p}$ is studied in [13]. Also, by using a sieve method, Stark [15] gives exact formulas for the degree distribution of an arbitrary fixed vertex of $G_{n,m,p}$ for a quite wide range of the parameters of the model.

Our results. To the best of the authors' knowledge, this is the first work were the independence number of $G_{n,m,\lambda}$ is analyzed. Throughout this paper we assume that the number of vertices n is large, i.e. $n \to \infty$ (and the values of the rest of the parameters of the models are defined in terms of n when needed). Our contribution in this work is the following:

(a) We show that when the number of vertices n is at least $(1+\epsilon)\binom{m}{\lambda}\ln\binom{m}{\lambda}$, for some constant $\epsilon > 0$ arbitrarily small, then $G_{n,m,\lambda}$, with $\lambda \geq 2$, has a Hamilton cycle with high probability (whp),[1] i.e. a very small constant number of labels suffices to yield Hamiltonicity. The proof uses the coupon collector's problem together with an interesting combinatorial construction. It also leads to a polynomial time randomized algorithm for constructing Hamilton cycles whp for this range of values for the parameters of the model.

(b) By applying the first and second moment methods, we show that when $m = \lfloor n^{\alpha} \rfloor$ for some fixed constant $\alpha < 1$ and for all $\lambda = O(m^{1/4})$, the independence number of the $G_{n,m,\lambda}$ model is approximately $\frac{m}{c_0\lambda}$, where c_0 is the smallest real number that satisfies $c_0 \geq 1$ and

$$c_0^{-1}\left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} = c_0^{-1} + (1-c_0^{-1})\ln(1-c_0^{-1}) - \frac{\ln(c_0\lambda)}{c_0\lambda}.$$

(c) Finally we give a note on $G_{n,m,p}$, that provides bounds on its independence number. Note that this was left open in [10].

[1] Unless stated otherwise, in this paper we assume that an event $A = A_n$ happens *with high probability* if its probability goes to 1 as n goes to infinity, i.e. $\lim_{n\to\infty}\Pr(A_n) = 1$.

2. Notation and definition of the models

We now formally define the two models that concern this work. We choose to define them in chronological order.

Definition 1 (*Random Intersection Graph* — $G_{n,m,p}$ *[9,14]*). Consider a universe $\mathcal{M} = \{1, 2, \ldots, m\}$ of elements and a set of vertices $V(G) = \{v_1, v_2, \ldots, v_n\}$. If we assign independently to each vertex v_j, $j = 1, 2, \ldots, n$, a subset S_{v_j} of $\mathcal{M}$ choosing each element $i \in \mathcal{M}$ independently with probability p and put an edge between two vertices v_{j_1}, v_{j_2} if and only if $S_{v_{j_1}} \cap S_{v_{j_2}} \neq \emptyset$, then the resulting graph is an instance of the random intersection graph $G_{n,m,p}$. In this model we also denote by L_l the set of vertices that have chosen label $l \in M$. The degree of $v \in V(G)$ will be denoted by $d_G(v)$. Also, the set of edges of $G_{n,m,p}$ will be denoted by $e(G)$.

Consider now the bipartite graph with vertex set $V(G) \cup \mathcal{M}$ and edge set $\{(v_j, i) : i \in S_{v_j}\} = \{(v_j, i) : v_j \in L_i\}$. We will refer to this graph as the *bipartite random graph* $B_{n,m,p}$ *associated to* $G_{n,m,p}$.

When we assume that every vertex chooses exactly λ labels, a completely different kind of randomization to intersection graphs is introduced. This model is called *uniform random intersection graphs model* $G_{n,m,\lambda}$. This was first mentioned in [8], and (apart from the case where the number of labels chosen by a vertex in $G_{n,m,p}$ are concentrated around their mean value) its probabilistic behavior seems a lot different than the one of $G_{n,m,p}$. Below is the formal definition of this model.

Definition 2 (*Uniform Random Intersection Graph* — $G_{n,m,\lambda}$ *[8]*). Consider a universe $\mathcal{M} = \{1, 2, \ldots, m\}$ of labels and a set of vertices $V(G) = \{v_1, v_2, \ldots, v_n\}$. If we assign independently to each vertex v_j, $j = 1, 2, \ldots, n$, a subset $S_{v_j} \subseteq \mathcal{M}$ of exactly λ randomly chosen distinct labels and put an edge between two vertices v_{j_1}, v_{j_2} if and only if $S_{v_{j_1}} \cap S_{v_{j_2}} \neq \emptyset$, then the resulting graph is an instance of the uniform random intersection graph $G_{n,m,\lambda}$.

The main focus of this paper is about Hamiltonicity and the *independence number* of the uniform random intersection graphs model in the interesting case when the number of labels is much smaller than the number of vertices. In particular, we are interested in the case where the number of vertices n is large and ideally goes to infinity. Parameters m, λ and p are generally defined as functions of n. Therefore, all asymptotics in this paper are with respect to $n \to \infty$.

The independence number of a graph G is denoted by $\alpha(G)$ and is equal to the cardinality of the maximum independent set of G, i.e. the size of the largest set of vertices of G with no edges between them. More specifically, in Section 3 we prove that if the number of labels is small enough then even $\lambda = 2$ suffices to prove that $G_{n,m,\lambda}$ has a Hamilton cycle almost surely. Section 4 concerns the existence of independent sets of various sizes in $G_{n,m,\lambda}$. Finally, in Section 5 we give a note on the independence number for $G_{n,m,p}$, which was left open in [10].

To prove the existence of independent sets of size k with high probability (whp), the second moment method is used, which is briefly described below (see [1] for a nice treatment of probabilistic methods). Suppose that $X = X_1 + \cdots X_k$, where X_i is the indicator random variable of event A_i. For indices i, j, we write $i \sim j$ iff $i \neq j$ and the events A_i, A_j are not independent. Now let

$$\Delta = \sum_{i \sim j} \Pr(A_i \cap A_j)$$

where the sum is over ordered pairs. Now when $i \sim j$, $Cov(X_i, X_j) \leq E[X_i X_j] = \Pr(A_i \cap A_j)$, which gives

$$Var(X) \leq E[X] + \Delta. \tag{1}$$

We say that $X_1, \ldots, X_k$ are *symmetric* if for every $i \neq i'$ there is an automorphism of the underlying probability space that sends A_i to event $A_{i'}$. This means that for *every fixed* i, i', with $i \neq i'$ the following holds

$$\sum_{j \sim i} \Pr(A_j | A_i) = \sum_{j \sim i'} \Pr(A_j | A_{i'}) \stackrel{def}{=} \Delta^*.$$

Hence, Δ^* is independent of the (fixed) choice of i, which gives

$$\Delta = \sum_i \Pr(A_i) \sum_{j \sim i} \Pr(A_j | A_i) = E[X]\Delta^*.$$

Hence, by using (1) and Chebyshev's inequality we get

Theorem 1 (*[1]*). *If* $E[X] \to \infty$ *and* $\Delta^* = o(E[X])$, *then* $X > 0$ *whp. Furthermore* $X \sim E[X]$ *whp.*

3. A coupon-collector result

In this section, we prove that $G_{n,m,\lambda}$ has a Hamilton cycle with high probability (whp) when the number of vertices is large enough, even for a very small constant $\lambda \geq 2$.

Theorem 2. *Let $e_0 = \binom{m}{\lambda}$. If $n \geq (1+\epsilon)e_0 \ln e_0$, for some constant $\epsilon > 0$ arbitrarily small, then $G_{n,m,\lambda}$, with $m \geq \lambda \geq 2$, is Hamiltonian with high probability as n goes to infinity.*

Proof. For simplicity, we will refer to a set of λ labels as a *set element*. The total number of possible set elements in $G_{n,m,\lambda}$ is then obviously e_0. Let $\mathcal{E}_e$ be the event that no vertex chooses e. By independence,

$$\Pr(\mathcal{E}_e) = \left(1 - \frac{1}{e_0}\right)^n.$$

Let X denote the mean number of set elements not chosen by any vertex. Then

$$E[X] = e_0 \Pr(\mathcal{E}_e) \leq e^{\ln e_0 - \frac{n}{e_0}} \to 0$$

for any $n \geq (1+\epsilon)e_0 \ln e_0$. Hence, the vertices of $G_{n,m,\lambda}$ will have chosen *all* available set elements (choosing exactly 1 set element each) whp.

We now complete the proof by showing how this implies the existence of a Hamiltonian cycle in the case $\lambda \geq 2$. Consider an arbitrary ordering of the labels of the graph $\{l_1, l_2, \ldots, l_m\}$ (we can use the canonical ordering implied by the set $\mathcal{M}$, but any ordering will suffice for our purposes) and construct the sets $D_1, D_2, \ldots, D_m$, where $D_i = \{v \in V : l_i \in S_v \text{ and } l_j \notin S_v, \text{ for all } j \leq i-1\}$, i.e. D_i is the set of vertices that have chosen label l_i and none of the labels $l_1, \ldots, l_{i-1}$. Notice that these sets form a partition of the vertex set. We now establish two properties that these sets have.

1. First of all, note that since the vertices of $G_{n,m,\lambda}$ have chosen every available set element, we will have that the only empty sets will be all D_i with $i = m-\lambda+2, \ldots, m$. Indeed, for all $i \leq m-\lambda+1$, there will be at least one vertex u that has $i \in S_u$ and $S_u \subseteq \{l_i, \ldots l_m\}$. Also, since every vertex chooses exactly λ distinct labels, every vertex that has chosen l_i, for $i = m-\lambda+2, \ldots, m$, must belong to *exactly one* of $D_1, \ldots, D_{m-\lambda+1}$.
2. Second, note that by construction (of the D_is), and because of the fact that the vertices of $G_{n,m,\lambda}$ have chosen every available set element, there will be at least one edge between D_i and D_j, for all $i = 1, \ldots, m-\lambda$ and all $j = i+1, \ldots, m-\lambda+1$. Also, for *every* edge $\{x_i, y_{i+1}\}$ between D_i and D_{i+1}, $i = 1, \ldots, m-\lambda-1$, there is an edge $\{x_{i+1}, y_{i+2}\}$ between D_{i+1} and D_{i+2} that satisfies $\{x_i, y_{i+1}\} \cap \{x_{i+1}, y_{i+2}\} = \emptyset$, unless $|D_{i+1}| = 1$, where $y_{i+1} \equiv x_{j+1}$. Indeed, this is a consequence of the fact that the vertices of $G_{n,m,\lambda}$ have chosen every available set element (i.e. every combination of λ labels) whp. Finally, *all* edges $\{x_j, y_{j+1}\}$ between D_j and D_{j+1}, for *every* $j = i+2, \ldots, m-\lambda$, satisfy $\{x_i, y_{i+1}\} \cap \{x_j, y_{j+1}\} = \emptyset$, by the construction of the sets D_i.

These two properties allow us to fix a sequence of pairs $\{x_i, y_{i+1}\}$, for all $i = 1, 2, \ldots, m-\lambda$, that are disjoint, except for the case where some $|D_i| = 1$, which does not change our proof. As a final step, let y_1 be a vertex that satisfies $\{l_1, l_{m-\lambda+1}\} \subseteq S_{y_1}$, and $S_{y_1} \backslash \{l_1, l_{m-\lambda+1}\} \subseteq \{l_{m-\lambda+2}, \ldots, l_m\}$. Such a vertex exists whp and it is connected to all vertices in $D_{m-\lambda+1}$ since $l_{m-\lambda+1} \in S_{y_1}$.

Let now σ_i, $i = 1, \ldots, m-\lambda$ be an arbitrary ordering of the set D_i, that begins with y_i and ends with x_i. Also, let $\sigma_{m-\lambda+1}$ be an arbitrary ordering or the set $D_{m-\lambda+1}$, that begins with $y_{m-\lambda+1}$. Since every D_i is a clique, it is easy to verify that the sequence $\sigma_1\sigma_2\cdots\sigma_{m-\lambda}\sigma_{m-\lambda+1}$ is indeed a Hamilton cycle. □

Note here that $\lambda = 2$ is in fact as small as one can have in order to achieve Hamiltonicity. Indeed, for $\lambda = 1$ the graph is disconnected (see also [2]). In this sense, our result is optimal. Finally, note that our proof leads naturally to a randomized polynomial time (in terms of n and m) algorithm for constructing Hamilton cycles whp in this case.

4. The size of independent sets in $G_{n,m,\lambda}$

In this section, we will use the first and second moment probabilistic methods to approximate whp the size of the largest independent set in $G_{n,m,\lambda}$, for $m = \lfloor n^\alpha \rfloor$, constant $\alpha < 1$ and $\lambda = O(m^{1/4})$.

We will need the following useful lemma, concerning the properties of a specific function that arises in the application of the probabilistic method in our case.

Lemma 1. *Let m be an increasing function of n, let $\alpha < 1$ be a positive constant and let $\lambda = O(m^{1/4})$ be a positive integer. Consider the function $f : (0,1) \to \mathbb{R}$ given by the formula*

$$f(x) = x\left(\frac{1}{\alpha} - 1\right)\frac{\ln m}{\lambda} - x - (1-x)\ln(1-x) + x\frac{\ln\left(\frac{\lambda}{x}\right)}{\lambda}.$$

If ϵ is any sufficiently small constant, then the following hold:

(a) *If $\lim_{x \to 1^-} f(x) = \left(\frac{1}{\alpha} - 1\right)\frac{\ln m}{\lambda} - 1 + \frac{\ln \lambda}{\lambda} \geq 0$, then the equation $f(x) = 0$ does not have any solution in $(0,1)$ and also $f(1-\epsilon) > 0$.*

(b) *If* $\lim_{x\to 1^-} f(x) = \left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} - 1 + \frac{\ln \lambda}{\lambda} < 0$*, then the equation* $f(x) = 0$ *has a unique solution* $x_0 \in (0, 1)$*, for which* $f((1-\epsilon)x_0) > 0$, $x_0 > \frac{\ln\ln m}{\lambda}$ *and, if also* $(1+\epsilon)x_0 < 1$*, then* $-f((1+\epsilon)x_0) = \omega\left(x_0\frac{1}{\lambda}\right) = \omega\left(\frac{\ln\ln m}{\lambda^2}\right)$.

Proof. The first derivative of f is given by

$$\frac{df(x)}{dx} = \left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} + \ln(1-x) + \frac{\ln\left(\frac{\lambda}{x}\right)}{\lambda} - \frac{1}{\lambda} \tag{2}$$

and its second derivative is given by

$$\frac{d^2f(x)}{dx^2} = -\frac{1}{1-x} - \frac{1}{x\lambda}. \tag{3}$$

Moreover, by using (2) and the definition of f, we get that

$$f(x) = x\frac{df(x)}{dx} - x - \ln(1-x) + \frac{x}{\lambda} = x\frac{df(x)}{dx} + \frac{x}{\lambda} + \sum_{i\geq 2}\frac{x^i}{i} \tag{4}$$

where in the second equality we used the Taylor expansion $\ln(1-x) = -\sum_{i\geq 1}\frac{x^i}{i}$, valid for all $|x| < 1$.

From the above, we can see that $\frac{d^2f(x)}{dx^2} < 0$, for all $x \in (0, 1)$, and so $\frac{df(x)}{dx}$ is monotone decreasing in this interval. Moreover, we have that $\lim_{x\to 0^+}\frac{df(x)}{dx} > 0$ and $\lim_{x\to 1^-}\frac{df(x)}{dx} < 0$, which means that the equation $\frac{df(x)}{dx} = 0$ has a unique solution in $(0, 1)$, where also f is maximized. Let $\xi \in (0, 1)$ be the unique solution of $\frac{df(x)}{dx} = 0$. By (4), we conclude that $f(\xi) > 0$.

(a) To prove the first part of the Lemma, note that $\lim_{x\to 0^+} f(x) = 0$ and also $\lim_{x\to 1^-} f(x) = \left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} - 1 + \frac{\ln\lambda}{\lambda} \geq 0$ by assumption. Hence, since $\frac{df(x)}{dx}$ and $f(x)$ are continuous in $(0, 1)$, the equation $f(x) = 0$ cannot have a solution in $(0, 1)$ for the range of values specified in this part of the lemma, because otherwise the equation $\frac{df(x)}{dx} = 0$ should have had at least one more solution in $(0, 1)$, which leads to a contradiction.

Furthermore, for any sufficiently small positive constant ϵ (which is less than 1), we have that $f(1-\epsilon) = (1-\epsilon)\left(\left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} - 1 + \frac{\ln\lambda}{\lambda}\right) - \epsilon\ln\epsilon - (1-\epsilon)\frac{\ln(1-\epsilon)}{\lambda} > 0$.

(b) In this case, the existence and uniqueness of the solution of the equation $f(x) = 0$ in $(0, 1)$ comes from the continuity of $\frac{df(x)}{dx}$ and $f(x)$ in $(0, 1)$, combined with the observations that $\lim_{x\to 0^+} f(x) = 0$, $f(\xi) > 0$ and $\lim_{x\to 1^-} f(x) = \left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} - 1 + \frac{\ln\lambda}{\lambda} < 0$. Indeed, the equation $f(x) = 0$ has no solution in $(0, \xi]$ and exactly one solution in $(\xi, 1)$.

It is now easy to verify from the above that, if x_0 is the unique solution to $f(x) = 0$ in $(0, 1)$, then f is positive in $(0, x_0)$ and hence, if ϵ is a sufficiently small positive constant (which is less than 1), then $f((1-\epsilon)x_0) > 0$.

Moreover, since for this range of values we have $\left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} - 1 + \frac{\ln\lambda}{\lambda} < 0$, it must be that $\lambda = \omega(1)$ and more specifically $\lambda > \left(\frac{1}{\alpha}-1\right)\frac{\ln m}{1-o(1)}$. Hence, by the formula for $\frac{df(x)}{dx}$, the unique solution ξ to $\frac{df(x)}{dx} = 0$ in $(0, 1)$ satisfies the inequality $\xi > \frac{\ln\ln m}{\lambda}$, because $\lim_{x\to 1^-}\frac{df(x)}{dx} < 0$ and $\frac{df\left(\frac{\ln\ln m}{\lambda}\right)}{dx} > 0$. In view of the above, the solution x_0 to $f(x) = 0$ in $(0, 1)$ should satisfy the inequalities $x_0 > \xi > \frac{\ln\ln m}{\lambda}$. Notice now that, for all $x \in (0, 1)$, we have that $\frac{d^2f(x)}{dx^2} < -1$, which (if also $(1+\epsilon)x_0 < 1$ so that f is well defined) means that $\frac{df((1+\epsilon)x_0)}{dx} \leq -\epsilon x_0$ and finally $f((1+\epsilon)x_0) \leq -\epsilon^2 x_0^2$. We conclude then that, in the range of parameters specified in this part of the lemma, we have that $-f((1+\epsilon)x_0) \geq \epsilon^2 x_0^2 = \omega\left(\frac{x_0}{\lambda}\right) = \omega\left(\frac{\ln\ln m}{\lambda^2}\right)$, which completes the proof. □

We are now ready for the application of the probabilistic method. Let $X^{(k)}$ be the number of independent sets of size k in $G_{n,m,\lambda}$. The following theorem concerns the asymptotic behavior of the mean value of $X^{(k)}$.

Theorem 3. *Let* $G_{n,m,\lambda}$ *be a random instance of the uniform random intersection graphs model, with* $m = \lfloor n^\alpha \rfloor$, $\alpha < 1$ *and* $\lambda = O(m^{1/4})$*. Let also* c_0 *be the solution of the equation*

$$c_0^{-1}\left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} = c_0^{-1} + (1-c_0^{-1})\ln(1-c_0^{-1}) - \frac{\ln(c_0\lambda)}{c_0\lambda} \tag{5}$$

if there is such a solution in $(1, \infty)$ *and* $c_0 = 1$ *otherwise.*

Then, for any positive constant ϵ *that can be arbitrarily small, we have that* $E\left[X^{\left((1+\epsilon)\frac{m}{c_0\lambda}\right)}\right] \to 0$ *and* $E\left[X^{\left((1-\epsilon)\frac{m}{c_0\lambda}\right)}\right] \to \infty$.

Proof. Let V' be an arbitrary set of k vertices. In order for V' to be independent, each vertex in it should choose its labels in such a way that $S_v \cap S_u = \emptyset$, for all $u, v \in V'$. So, the probability that V' is an independent set is

$$p^*(k) \equiv \Pr(V' \text{ does not contain any edges}) = \frac{\binom{m}{\lambda}\binom{m-\lambda}{\lambda}\binom{m-2\lambda}{\lambda}\cdots\binom{m-(k-1)\lambda}{\lambda}}{\binom{m}{\lambda}^k} = \frac{\frac{(m-\lambda)!}{(m-k\lambda)!}}{(m(m-1)\cdots(m-\lambda+1))^{k-1}}.$$

We note here that, by the definition of the uniform random intersection graphs model, the size of the largest independent set of vertices cannot be more than m/λ (since every vertex selects *exactly* λ distinct labels in $\mathcal{M}$), and so only values for k such that $k\lambda \le m$ are of interest. However, for technical reasons, we will require that $k\lambda \le m-1$. Since $\lambda = o(m)$, this assumption does not affect the generality of our theorem.

By using now Stirling's approximation and the fact that $\lambda = O(m^{1/4})$, we get that

$$p^*(k) = \frac{\Theta(1)\frac{\sqrt{2\pi(m-\lambda)}\left(\frac{m-\lambda}{e}\right)^{m-\lambda}}{\sqrt{2\pi(m-k\lambda)}\left(\frac{m-k\lambda}{e}\right)^{m-k\lambda}}}{(m(1-O(\lambda/m)))^{\lambda(k-1)}}$$
$$= e^{(m-\lambda)\ln(m-\lambda)+\lambda-(m-k\lambda)\ln(m-k\lambda)-k\lambda-\lambda(k-1)\ln(m(1-O(\lambda/m)))+O(\ln m)}$$
$$= e^{m\ln m-(m-k\lambda)\ln(m-k\lambda)-k\lambda\ln m-k\lambda\pm O(\lambda+\ln m)}$$

where in the last equality we used the fact that $\lambda(k-1)\ln(1-O(\lambda/m)) = O(\lambda)$, since $\lambda = o(m)$ and $k-1 \le m/\lambda$. Moreover, taking into consideration that $k \le \frac{m}{\lambda} = o(n)$ and by using the linearity of expectation, we get that

$$E[X^{(k)}] = \binom{n}{k}p^*(k) = \frac{(n(1-O(k/n)))^k}{k!}p^*(k)$$
$$= \frac{(n(1-O(k/n)))^k}{\Theta(1)\sqrt{2\pi k}\left(\frac{k}{e}\right)^k}p^*(k)$$
$$= e^{k\ln n-k\ln k+k-o(k)}p^*(k)$$
$$= e^{k\ln(n/k)+m\ln m-(m-k\lambda)\ln(m-k\lambda)-k\lambda\ln m-k\lambda+k\pm Res} \tag{6}$$

for some positive $Res = o(k) + O(\lambda + \ln m)$.

We now set $n = m^{1/\alpha}$ and $k \equiv \frac{m}{c\lambda}$, where $c = c(m)$ is a function of m (and λ) that is greater than 1 (but as the reader can see it can be quite close to 1 depending on the values of m and λ). By (6) we then have that

$$E\left[X^{\left(\frac{m}{c\lambda}\right)}\right] = e^{\frac{m}{c\lambda}\ln\left(c\lambda m^{\frac{1}{\alpha}-1}\right)+m\ln m-\frac{m}{c}\ln m-m\left(1-\frac{1}{c}\right)\ln\left(m\left(1-\frac{1}{c}\right)\right)-\frac{m}{c}+\frac{m}{\lambda c}\pm Res}$$
$$= e^{m\left(\frac{1}{c}\left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda}-\frac{1}{c}-\left(1-\frac{1}{c}\right)\ln\left(1-\frac{1}{c}\right)+\frac{\ln(c\lambda)}{c\lambda}\right)+\frac{m}{\lambda c}\pm Res}. \tag{7}$$

By now using Lemma 1, we can see that c_0 is well defined, in the sense that it is uniquely defined and it satisfies the claims of the theorem. Indeed, if ϵ is an arbitrarily small positive constant, we then have that $E\left[X^{\left((1-\epsilon)\frac{m}{c_0\lambda}\right)}\right] = e^{mf\left((1-\epsilon)\frac{1}{c_0}\right)+\frac{m(1-\epsilon)}{\lambda c_0}\pm Res} \to \infty$, since either $c_0 = 1$, or $\frac{1}{c_0} > \frac{\ln\ln m}{\lambda}$, by part (b) of Lemma 1, so $\frac{m}{\lambda(1-\epsilon)c_0} = \omega(Res)$. Moreover, if also $(1+\epsilon)\frac{1}{c_0} < 1$ holds, then $E\left[X^{\left((1+\epsilon)\frac{m}{c_0\lambda}\right)}\right] = e^{mf\left((1+\epsilon)\frac{1}{c_0}\right)+\frac{m(1+\epsilon)}{\lambda c_0}\pm Res} \to 0$, by part (b) of Lemma 1. Finally, we note that, for the case $(1+\epsilon)\frac{1}{c_0} \ge 1$, it suffices to observe that ϵ is arbitrarily small and also that the size of any independent set cannot be larger than $\frac{m}{\lambda}$. This completes the proof of the theorem. □

By now using Markov's inequality together with Theorem 3, we get the following

Corollary 1. *Let $G_{n,m,\lambda}$ be a random instance of the uniform random intersection graphs model, with $m = \lfloor n^\alpha \rfloor$, $\alpha < 1$ and $\lambda = O(m^{1/4})$. Let also c_0 be as in Theorem 3. Then, with probability that tends to 1, as $n \to \infty$, the size of the largest independent set in $G_{n,m,\lambda}$ is no more than $(1+\epsilon)\frac{m}{c_0\lambda}$, for any constant $\epsilon > 0$ that can be arbitrarily small.*

We now continue by applying the second moment method in order to prove the existence of "large" independent sets of vertices whp. More precisely, the following theorem will help us find a quite good approximation of the size of the largest independent set of vertices in $G_{n,m,\lambda}$, with $m = \lfloor n^\alpha \rfloor$, $\alpha < 1$ and $\lambda = O(m^{1/4})$, which holds with high probability.

Theorem 4. *Let $G_{n,m,\lambda}$ be a random instance of the uniform random intersection graphs model, with $m = \lfloor n^{\alpha} \rfloor$, $\alpha < 1$ and $\lambda = O(m^{1/4})$. Let also c_0 be as in Theorem 3. Then, with probability that tends to 1, as $n \to \infty$, there are independent sets in $G_{n,m,\lambda}$, with size at least $(1-\epsilon)\frac{m}{c_0\lambda}$, for any constant $\epsilon > 0$ that can be arbitrarily small.*

Proof. Set $k = \lfloor (1-\epsilon)\frac{m}{c_0\lambda} \rfloor$. Note that, by Lemma 1, we have that $k = \omega(1)$. In order to prove the theorem, we will use the second moment probabilistic method. Let V' be an independent set of k vertices, and let V'' be another set of k vertices that has exactly s vertices in common with V'. Given that V' has no edges, in order for V'' to be an independent set, the $k-s$ vertices of V'' not belonging to V' should choose non-overlapping label sets, that also do not overlap with the $s\lambda$ distinct labels chosen by the s vertices in $V' \cap V''$ (that do not have edges between them by assumption). So, we have that

$$\Pr(V'' \text{ is an independent set} | V' \text{ is an independent set}) = \frac{\binom{m-s\lambda}{\lambda}\binom{m-(s+1)\lambda}{\lambda}\cdots\binom{m-(k-1)\lambda}{\lambda}}{\binom{m}{\lambda}^{k-s}} = \frac{(m-s\lambda)!}{(m-k\lambda)!}\left(\frac{(m-\lambda)!}{m!}\right)^{k-s}.$$

For a set S of vertices, let X_S be an indicator random variable that takes the value 1 if S is an independent set and 0 otherwise. In the current setting, the quantity Δ^* involved in the second moment method (see Section 2) is given by the formula

$$\Delta^* = \sum_{s=2}^{k-1} \binom{n-k}{k-s}\binom{k}{s} \Pr(X_{V''} = 1 | X_{V'} = 1).$$

We now define the following quantity:

$$R \equiv \frac{\Delta^*}{E[X^{(k)}]} = \frac{\sum_{s=2}^{k-1} \binom{n-k}{k-s}\binom{k}{s} \Pr(X_{S'} = 1 | X_S = 1)}{E[X^{(k)}]} = \sum_{s=2}^{k-1} \frac{\binom{n-k}{k-s}\binom{k}{s}}{\binom{n}{k}} \frac{(m-s\lambda)!}{(m-\lambda)!}\left(\frac{m!}{(m-\lambda)!}\right)^{s-1}.$$

Let also

$$a_s \equiv \frac{\binom{n-k}{k-s}\binom{k}{s}}{\binom{n}{k}}.$$

and

$$b_s \equiv \frac{(m-s\lambda)!}{(m-\lambda)!}\left(\frac{m!}{(m-\lambda)!}\right)^{s-1}.$$

By now setting $R_s \equiv a_s b_s$, if we show that $R_s = o\left(\frac{1}{k}\right)$, then $R = o(1)$, which leads to the conclusion of the theorem, by using the second moment probabilistic method (see Section 2). For the proof, we need to find quite tight bounds for the quantities a_s and b_s.

First of all, by using the relation $\frac{n^k}{k^k} \le \binom{n}{k} \le \frac{n^k}{(ek)^k}$, we get that

$$\begin{aligned} a_s &\le \frac{k^{k+s}}{(e(k-s))^{k-s}(es)^s n^s} \\ &= \frac{\left(\frac{k}{n}\right)^s}{\left(e\left(1-\frac{s}{k}\right)\right)^{k-s}\left(e\frac{s}{k}\right)^s} \\ &= \frac{e^{-s\ln\left(\frac{m}{s}\frac{s}{k}m^{\frac{1}{\alpha}-1}\right)}}{\left(e\left(1-\frac{s}{k}\right)\right)^{k-s}\left(e\frac{s}{k}\right)^s} \\ &= \frac{e^{-s\lambda\left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} - s\lambda\frac{\ln\left(\frac{m}{s}\right)}{\lambda} - s\ln\left(\frac{s}{k}\right)}}{\left(e\left(1-\frac{s}{k}\right)\right)^{k-s}\left(e\frac{s}{k}\right)^s} \\ &= \frac{e^{-s\lambda\left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} - s\lambda\frac{\ln\left(\frac{m}{s}\right)}{\lambda}}}{e^k\left(1-\frac{s}{k}\right)^{k-s}\left(\frac{s}{k}\right)^{2s}}. \end{aligned} \quad (8)$$

Moreover, by Stirling's approximation, we get that

$$\begin{aligned}
b_s &\le \Theta(1)\frac{\left(\frac{m-s\lambda}{e}\right)^{m-s\lambda}}{\left(\frac{m-\lambda}{e}\right)^{m-\lambda}}\left(\frac{\sqrt{2\pi m}}{\sqrt{2\pi(m-\lambda)}}\right)^{k-1}\left(\frac{\left(\frac{m}{e}\right)^m}{\left(\frac{m-\lambda}{e}\right)^{m-\lambda}}\right)^{s-1}\\
&\le \Theta(1)\frac{\left(\frac{m-s\lambda}{e}\right)^{m-s\lambda}}{\left(\frac{m-\lambda}{e}\right)^{m-\lambda}}\left(\frac{\left(\frac{m}{e}\right)^m}{\left(\frac{m-\lambda}{e}\right)^{m-\lambda}}\right)^{s-1}\\
&= \Theta(1)\frac{(m-s\lambda)^{m-s\lambda}}{(m-\lambda)^{m-\lambda}}\left(\frac{m^m}{(m-\lambda)^{m-\lambda}}\right)^{s-1}\\
&= \Theta(1)\frac{(m-s\lambda)^{m-s\lambda}}{(m-\lambda)^{m-\lambda}}m^{\lambda(s-1)}\left(1+\frac{\lambda}{m-\lambda}\right)^{(m-\lambda)(s-1)}\\
&\le \Theta(1)\frac{(m-s\lambda)^{m-s\lambda}}{(m-\lambda)^{m-\lambda}}m^{\lambda(s-1)}e^{\lambda(s-1)}\\
&= \Theta(1)e^{(m-s\lambda)\ln(m-s\lambda)-(m-\lambda)\ln(m-\lambda)+\lambda(s-1)\ln m+\lambda(s-1)}\\
&= \Theta(1)e^{m\left(1-\frac{s\lambda}{m}\right)\ln\left(1-\frac{s\lambda}{m}\right)+m\left(1-\frac{s\lambda}{m}\right)\ln m-(m-\lambda)\ln m-(m-\lambda)\ln\left(1-\frac{\lambda}{m}\right)+\lambda(s-1)\ln m+\lambda(s-1)}\\
&\le \Theta(1)e^{m\left(1-\frac{s\lambda}{m}\right)\ln\left(1-\frac{s\lambda}{m}\right)+s\lambda}. \qquad (9)
\end{aligned}$$

We note here that all the $\Theta(1)$ quantities that appear in the above inequalities come from Stirling's approximation and actually depend on the constant ϵ. In particular, for the first inequality we used the fact that $s \le k$ and for the second inequality we used the fact that $k-1 \le \frac{m}{\lambda}$, so $\left(1+\frac{\lambda}{m-\lambda}\right)^{(k-1)/2} \le e^{\frac{\lambda(k-1)}{2(m-\lambda)}}$ is bounded by a constant. Furthermore, in the fifth inequality, we used the inequality $1+x \le e^x$, which means that the $\Theta(1)$ quantity changed by a multiplicative constant. Finally, in the last equality we used the expansion $\ln\left(1-\frac{\lambda}{m}\right) = -\sum_{i\ge1}\frac{1}{i}\left(\frac{\lambda}{m}\right)^i$ (which converges for any $\lambda = o(m)$), in order to show that $-(m-\lambda)\ln\left(1-\frac{\lambda}{m}\right) \le (m-\lambda)\left(\frac{\lambda}{m}+\frac{1}{2}\sum_{i\ge2}\left(\frac{\lambda}{m}\right)^i\right) = \lambda+O(1)$, for any $\lambda = O(m^{1/4})$.

We now set $x = \frac{s\lambda}{m}$. By combining inequalities (8) and (9), we get that

$$R_s \le \Theta(1)\frac{e^{-mf(x)}}{e^k\left(1-\frac{s}{k}\right)^{k-s}\left(\frac{s}{k}\right)^{2s}} = \Theta(1)\frac{e^{-mf(x)}}{e^{k+k\left(1-\frac{s}{k}\right)\ln\left(1-\frac{s}{k}\right)+2k\frac{s}{k}\ln\left(\frac{s}{k}\right)}} \qquad (10)$$

where $f(x) = x\left(\frac{1}{\alpha}-1\right)\frac{\ln m}{\lambda} - x - (1-x)\ln(1-x) + x\frac{\ln\left(\frac{\lambda}{x}\right)}{\lambda}$ is the function of Lemma 1. Moreover, it is straightforward to show that $\left(\frac{s}{k}\right)^{2s} \ge e^{-\frac{2}{e}k}$. Hence, we have that

$$R_s \le \Theta(1)\frac{e^{-mf(x)}}{e^{\left(1-\frac{2}{e}\right)k}e^{k\left(1-\frac{s}{k}\right)\ln\left(1-\frac{s}{k}\right)}}.$$

By now using the expansion $\ln\left(1-\frac{s}{k}\right) = -\sum_{i\ge1}\frac{1}{i}\left(\frac{s}{k}\right)^i$, together with the formula relating f to its first derivative (see proof of Lemma 1), we get that

$$R_s \le \Theta(1)\frac{e^{-m\left(x\frac{df(x)}{dx}-x-\ln(1-x)+\frac{x}{\lambda}\right)}}{e^{\left(1-\frac{2}{e}\right)k}e^{k\left(-\frac{s}{k}+\sum_{i\ge2}\left(\frac{s}{k}\right)^i\frac{1}{i(i-1)}\right)}} = \Theta(1)\frac{e^{-m\left(x\frac{df(x)}{dx}+\sum_{i\ge2}\frac{x^i}{i}+\frac{x}{\lambda}\right)}}{e^{\left(1-\frac{2}{e}\right)k+k\left(-\frac{s}{k}+\sum_{i\ge2}\left(\frac{s}{k}\right)^i\frac{1}{i(i-1)}\right)}}. \qquad (11)$$

Let now ξ be the (unique) solution of the equation $\frac{df(x)}{dx} = 0$ in $(0, 1)$. As we also saw in the proof of Lemma 1, we have that $\frac{df(x)}{dx} > 0$, for all $x \in (0, \xi)$ and $\frac{df(x)}{dx} < 0$, for all $x \in \left(\xi, \frac{1-\epsilon}{c_0}\right)$. By now observing that $s = \frac{mx}{\lambda}$, by (11) we get that

$$R_s \le \Theta(1)\frac{e^{-m\left(x\frac{df(x)}{dx}+\sum_{i\ge2}\frac{x^i}{i}\right)}}{e^{\left(1-\frac{2}{e}\right)k+k\sum_{i\ge2}\left(\frac{s}{k}\right)^i\frac{1}{i(i-1)}}} = o\left(\frac{1}{k}\right)$$

for all $x \in (0, \xi)$.

Setting now $y = \frac{s}{k}$, the exponent of the denominator in the right side of (10) divided by k is given by the function $h(y) = 1 + (1-y)\ln(1-y) + 2y\ln y$. A careful analysis of $h(y)$ can show that $\frac{dh(y)}{dy} = -\ln(1-y) + 2\ln y + 1$ and $\frac{d^2h(y)}{dy^2} > 0$, for any $y \in (0, 1)$. Hence, the function h is minimized at y_0 with $-\ln(1-y_0) + 2\ln y_0 + 1 = 0$, that is

$y_0 = \frac{-1+\sqrt{1+4e}}{2e} \approx 0.44987$. Unfortunately $0 > h(y_0) > -0.2$, but we do have that $h(y) > \frac{1}{7000}$, for all $y \notin \left(\frac{1}{4}, \frac{2}{3}\right)$. This means that $R_s = o\left(\frac{1}{k}\right)$, for every $s \geq \frac{2k}{3}$ and every $s \leq \frac{k}{4}$.

In view of the above and by noting that $\frac{s}{k} = \frac{c_0 x}{1-\epsilon}$, we conclude that we only have to show that $R_s = o\left(\frac{1}{k}\right)$ holds even for s such that $\frac{s}{k} \in \left(\max\left\{\frac{1}{4}, \frac{c_0\xi}{1-\epsilon}\right\}, \frac{2}{3}\right)$, or equivalently, for $x \in \left(\max\left\{\frac{(1-\epsilon)}{4c_0}, \xi\right\}, \frac{2(1-\epsilon)}{3c_0}\right)$. Of course, if the above interval is empty (i.e. its lower bound is greater than the upper bound), then we are done. Such a thing can happen for example if $\lambda = o(\ln m)$, so that $\xi \to 1$. Hence, we only have to deal with the case $\lambda = \Omega(\ln m)$. However, as we saw in the proof of Lemma 1, we have that $\frac{d^2 f(x)}{dx^2} < -1$, which means that $f(x) \geq \frac{1}{9}\frac{1}{c_0^2}$, for every $x < \frac{2(1-\epsilon)}{3c_0}$, by the definition of c_0. By (10) and the fact that $h(y_0) > -0.2$, we then get that

$$R_s \leq \Theta(1)\frac{e^{-m\frac{1}{9}\frac{1}{c_0^2}}}{e^{-0.2k}} = \Theta(1)e^{-k\frac{\lambda}{8c_0}+0.2k}.$$

By the second part of Lemma 1 though, we have that $\frac{1}{c_0} = \omega\left(\frac{1}{\lambda}\right)$, for $\lambda = \Omega(\ln n)$. Hence, we come to the desired result that $R_s = o\left(\frac{1}{k}\right)$, which completes the proof of the theorem. $\square$

By now combining Corollary 1 and Theorem 4 and by pointing out that the positive constant ϵ can be arbitrarily small, we get the following theorem that gives a quite good approximation for the size of the largest independent set in $G_{n,m,\lambda}$:

Theorem 5. *Let $G_{n,m,\lambda}$ be a random instance of the uniform random intersection graphs model, with $m = \lfloor n^\alpha \rfloor$, $\alpha < 1$ and $\lambda = O(m^{1/4})$. Let also c_0 be as in Theorem 3. Then, with probability that tends to 1, when $n \to \infty$, the size k of the largest independent set of vertices in $G_{n,m,\lambda}$ satisfies $k \sim \frac{m}{c_0\lambda}$.* $\square$

As the careful reader may have noticed by now, finding a closed formula for c_0 in Theorem 3 seems quite difficult. In fact it still remains an open problem. However, for some range of values of m, λ, this is still possible. The next result is a direct consequence of Theorem 5 and the definition of function $f(x)$ and it demystifies the nature of c_0 for some (quite wide) range of values for the parameters of $G_{n,m,\lambda}$.

Corollary 2. *Let $G_{n,m,\lambda}$ be a random instance of the uniform random intersection graphs model, with $m = \lfloor n^\alpha \rfloor$, $\alpha < 1$ and let k be the size of the largest independent set of vertices in $G_{n,m,\lambda}$. Then, with probability that tends to 1, as $n \to \infty$, the following are true:*

(i) *If $\left(\frac{1}{\alpha} - 1\right)\frac{\ln m}{\lambda} - 1 + \frac{\ln \lambda}{\lambda} \geq 0$, then $k \sim \frac{m}{\lambda}$. In particular, the above inequality is satisfied for any $\lambda \geq B \ln m$, where B is a positive constant such that $B \leq \frac{1}{\alpha} - 1$.*
(ii) *If $\lambda = B' \ln m$, where $B' > \frac{1}{\alpha} - 1$ is a constant, then c_0 in Theorem 5 is a constant larger than 1, therefore $k = \Theta\left(\frac{m}{\lambda}\right)$.*
(iii) *If $\lambda = \omega(\ln m)$ (but $\lambda = O(m^{1/4})$), then c_0 in Theorem 5 satisfies $c_0 \sim \frac{\alpha\lambda}{2(1-\alpha)\ln m}$, therefore $k \sim 2\left(\frac{1}{\alpha} - 1\right)\frac{m \ln m}{\lambda^2}$.* $\square$

Remark. It is straightforward to verify that the unconditional probability of existence of an edge between two fixed vertices in $G_{n,m,\lambda}$, with $m = \lfloor n^\alpha \rfloor$, $\alpha < 1$ and $\lambda = O(m^{1/4})$, is $\hat{p} = 1 - \frac{\binom{m-\lambda}{\lambda}}{\binom{m}{\lambda}} \sim \frac{\lambda^2}{m}$. We could now witness the effect that labels have on the appearance of edges in $G_{n,m,\lambda}$ by comparing its independence number to the independence number of a Erdős–Rényi random graph $G_{n,\hat{p}}$, the latter being asymptotically equal to $\frac{2\ln n\hat{p}}{\hat{p}} \sim \frac{2m \ln \frac{n\lambda^2}{m}}{\lambda^2}$. Using the results of Corollary 2, we can see that when λ is small (e.g. $\lambda = \Theta(\ln m)$), then the independence number of $G_{n,m,\lambda}$ is much larger (by a factor of λ) than the independence number of $G_{n,\hat{p}}$. On the other hand, when $\lambda = \omega(\ln m)$, the two independence numbers become of the same order.

5. A note on independent sets in $G_{n,m,p}$

In this section we give a brief note on $\alpha(G_{n,m,p})$, i.e. the independence number of random intersection graphs, for some ranges of the parameters m, p. We first prove the following

Lemma 2. *Let $G_{n,m,p}$ be a random instance of the random intersection graphs model, with $p = \omega\left(\frac{\ln n}{m}\right)$ and $p = o\left(\sqrt{\frac{\ln n}{m}}\right) = o(1)$. Then $\alpha(G_{n,m,p}) \leq \frac{(2+o(1))\ln n}{mp^2}$ whp.*

Proof. Let δ be an arbitrarily small positive constant. Note that, for any $k \sim \frac{(2+\delta)\ln n}{mp^2}$, the conditions of the statement of our Lemma imply respectively $kp \to 0$ and $k = \omega(1)$. Furthermore

$$(1-p)^k + kp(1-p)^{k-1} = 1 - \binom{k}{2}p^2(1-p)^{k-2} - \binom{k}{3}p^3(1-p)^{k-3} - \binom{k}{4}p^4(1-p)^{k-4} - o(k^4p^4).$$

Therefore, we get that

$$(1-p)^k + kp(1-p)^{k-1} \sim e^{-\frac{k^2p^2}{2} \pm O(kp^2+k^3p^3)}. \tag{12}$$

Now let $X^{(k)}$ denote the number of independent sets of size k in $G_{n,m,p}$. Then, by using the fact $\binom{n}{k} = e^{k\ln(n/k)+O(k)}$, we have

$$\begin{aligned} E[X^{(k)}] &= \binom{n}{k}\left((1-p)^k + kp(1-p)^{k-1}\right)^m \\ &\sim e^{k\left(\ln(n/k) - \frac{kmp^2}{2}\right)+O(k)\pm O(mkp^2+mk^3p^3)}. \end{aligned} \tag{13}$$

By (13) we now get that $E[X^{(k)}] \to 0$ for any $k \geq \frac{(2+\delta)\ln n}{mp^2}$. By using Markov's inequality we conclude the proof of the lemma. □

We will now apply the second moment method to prove existence of large independent sets of vertices. As a first step we compare Δ^* to $E[X^{(k)}]$, where as before

$$\Delta^* = \sum_{s=1}^{k-1}\binom{n-k}{k-s}\binom{k}{s}\Pr(X_{S'}=1|X_S=1)$$

for $|S'\cap S| = s$, $|V_{S'}| = |V_S| = k$, and S a fixed independent set of vertices. In [10] the authors have proved that

$$E[X^{(k)}] = \binom{n}{k}\left((1-p)^k + kp(1-p)^{k-1}\right)^m$$

and have also provided an *exact* formula for $\Pr(X_{S'}=1|X_S=1) \stackrel{def}{=} \gamma(k,s)$, namely

$$\begin{aligned} \gamma(k,s) &= \left((1-p)^{k-s} + (k-s)p(1-p)^{k-s-1}\left(1-\frac{sp}{1+(k-1)p}\right)\right)^m \\ &= (1-p)^{-sm}\left((1-p)^k + (k-s)p(1-p)^{k-1}\left(1-\frac{sp}{1+(k-1)p}\right)\right)^m \\ &\leq (1-p)^{-sm}\frac{E[X^{(k)}]}{\binom{n}{k}}. \end{aligned}$$

Consequently, we have

$$\begin{aligned} \frac{\Delta^*}{E[X^{(k)}]} &\leq \sum_{s=1}^{k-1}\frac{\binom{n-k}{k-s}\binom{k}{s}}{\binom{n}{k}}(1-p)^{-sm} \\ &\leq \sum_{s=1}^{k-1}\frac{\frac{(n-k)^{k-s}}{(k-s)!}\binom{k}{s}}{\frac{(n-k)^k}{k!}}(1-p)^{-sm} \\ &\leq \sum_{s=1}^{k-1}\frac{k!\binom{k}{s}}{(n-k)^s(k-s)!}\left(1+\frac{p}{1-p}\right)^{sm} \\ &\leq \sum_{s=1}^{k-1}\frac{k!\binom{k}{s}}{(n-k)^s(k-s)!}e^{\frac{smp}{1-p}} \end{aligned} \tag{14}$$

where in the last inequality we used the fact that $1+\frac{p}{1-p} \leq e^{\frac{p}{1-p}}$.

If we now assume that $mp \leq \beta\ln n$, for some constant $\beta < 1$, then for all $k = o\left(n^{\frac{1-\beta}{3}}\right)$ and $p \to 0$, we have by (14) that

$$\begin{aligned} \frac{\Delta^*}{E[X^{(k)}]} &\leq \sum_{s=1}^{k-1}\frac{k^s\binom{k}{s}}{(n-k)^s}e^{\frac{smp}{1-p}} \\ &\leq \sum_{s=1}^{k-1}\left(\frac{k^2}{n}\right)^s e^{\frac{smp}{1-p}} \\ &= \sum_{s=1}^{k-1}e^{s\left(\frac{mp}{1-p}+2\ln k-\ln n\right)} = o(1). \end{aligned} \tag{15}$$

Note now that for the range of parameters of the following theorem, for any $k \sim \frac{2\ln n}{mp^{2-\epsilon}}$, we have that $\omega(1) = k = o\left(n^{\frac{1-\beta}{3}}\right)$, and also $kp \to 0$. This means that we can apply the relation (13) of Lemma 2 to prove that $E[X^{(k)}] \to \infty$. The following theorem then follows by (15) and the second moment method.

Theorem 6. *Let $G_{n,m,}$ be a random instance of the random intersection graphs model. Let also ϵ be an arbitrarily small positive constant and β a positive constant less than 1. Assume that $mp \le \beta \ln n$, $p = \omega\left(\sqrt[1-\epsilon]{\frac{\ln n}{m}}\right)$, $p = o\left(\sqrt[2-\epsilon]{\frac{\ln n}{m}}\right)$ and $p = \omega\left(\sqrt[2-\epsilon]{\frac{\ln n}{n^{\frac{1-\beta}{3}} m}}\right)$. Then $\alpha(G_{n,m,p}) \ge \frac{2\ln n}{mp^{2-\epsilon}}$ whp.* □

6. Conclusions and future work

In this work we studied Hamiltonicity and independent sets in the uniform random intersection graphs model $G_{n,m,\lambda}$. In particular, we approximated the independence number of $G_{n,m,\lambda}$ for $m = \lfloor n^\alpha \rfloor$, $\alpha < 1$ and for all $\lambda = O(m^{1/4})$, in terms of the solution of some function f. An open problem here is the derivation of a closed expression for c_0 of Theorem 3. Moreover, the problem of determining the cardinality of the maximum independent set in $G_{n,m,\lambda}$ for $\lambda \in [m^{1/4}, m^{1/2}]$ remains open. A possible solution to this could be the use of the weighted second moment method used in [4].

The design and analysis of algorithms that construct independent sets (when their input is an instance of $G_{n,m,\lambda}$), whose cardinality approaches the theoretical bounds given here, is a subject of our future work. Finally, we are very interested in determining the chromatic number $\chi(G_{n,m,\lambda})$ of the uniform random intersection graphs model, or finding an upper bound that is as close as possible to the obvious lower bound $\frac{n}{\alpha(G_{n,m,\lambda})}$, where $\alpha(G_{n,m,\lambda})$ is the independence number.

Acknowledgements

We would like to thank the anonymous reviewers whose insightful and very detailed comments contributed toward improving this work, and also the reviewer who proposed using the weighted version of the second moment in order to close the $\lambda \in [m^{1/4}, m^{1/2}]$ gap.

References

[1] N. Alon, J.H. Spencer, The Probabilistic Method, second ed., John Wiley & Sons, Inc., 2000.
[2] S.R. Blackburn, S. Gerke, Connectivity of the uniform random intersection graph, http://www.citebase.org/abstract?id=oai:arXiv.org:0805.2814, (2008).
[3] M. Blonzelis, Degree distribution of a typical vertex in a general random intersection graph, Lithuanian Mathematical Journal 48 (1) (2008) 38–45.
[4] V. Dani, C. Moore, Independent sets in random graphs from the weighted second moment method, arXiv:1011.0180v2 [cs.CC].
[5] M. Deijfen, W. Kets, Random intersection graphs with tunable degree distribution and clustering, http://www2.math.su.se/~mia/rig.pdf.
[6] C. Efthymiou, P.G. Spirakis, On the existence of Hamilton cycles in random intersection graphs, in: The Proceedings of the 32nd International Colloquium on Automata, Languages and Programming, ICALP, 2005, pp. 690–701.
[7] J.A. Fill, E.R. Sheinerman, K.B Singer-Cohen, Random intersection graphs when $m = \omega(n)$: an equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models, http://citeseer.nj.nec.com/fill98random.html.
[8] E. Godehardt, J. Jaworski, Two models of random intersection graphs for classification, in: O. Opitz, M. Schwaiger (Eds.), Studies in Classification, Data Analysis and Knowledge Organisation, Springer Verlag, Berlin, Heidelberg, New York, 2002, pp. 67–82.
[9] M. Karoński, E.R. Scheinerman, K.B. Singer-Cohen, On random intersection graphs: the subgraph problem, Combinatorics, Probability and Computing Journal 8 (1999) 131–159.
[10] S. Nikoletseas, C. Raptopoulos, P. Spirakis, The existence and efficient construction of large independent sets in general random intersection graphs, in: The Proceedings of the 31st International Colloquium on Automata, Languages and Programming, ICALP, in: Lecture Notes in Computer Science, Springer Verlag, 2004, pp. 1029–1040.
[11] S. Nikoletseas, C. Raptopoulos, P.G. Spirakis, Expander properties and the cover time of symmetric random intersection graphs, in: The Proceedings of the 32nd International Symposium on Mathematical Foundations of Computer Science, MFCS 2007, 2007.
[12] R. Di Pietro, L.V. Mancini, A. Mei, A. Panconesi, J. Radhakrishnan, Sensor networks that are provably resilient, in: The Proceedings of IEEE, 2nd International Conference on Security and Privacy in Communication Networks (SecureComm), 2004.
[13] C. Raptopoulos, P.G. Spirakis, Simple and efficient greedy algorithms for Hamilton cycles in random intersection graphs, in: The Proceedings of the 16th International Symposium on Algorithms and Computation, ISAAC, 2005, pp. 493–504.
[14] K.B. Singer-Cohen, Random intersection graphs, Ph.D. Thesis, John Hopkins University, 1995.
[15] D. Stark, The vertex degree distribution of random intersection graphs, Random Structures & Algorithms 24 (3) (2004) 249–258.