Maximum Cliques in Graphs with Small Intersection Number and Random Intersection Graphs

S. Nikoletseas\textsuperscript{1,2}, C. Raptopoulos\textsuperscript{1}, and P. G. Spirakis\textsuperscript{1,2}

\textsuperscript{1} Computer Technology Institute, P.O. Box 1122, 26110 Patras, Greece
\textsuperscript{2} University of Patras, 26500 Patras, Greece

nikole@cti.gr, raptopox@ceid.upatras.gr, spirakis@cti.gr

Abstract. In this paper, we relate the problem of finding a maximum clique to the intersection number of the input graph (i.e. the minimum number of cliques needed to edge cover the graph). In particular, we consider the maximum clique problem for graphs with small intersection number and random intersection graphs (a model in which 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).

We first present a simple algorithm which, on input \(G\) finds a maximum clique in \(O(2^m + O(m) + n^3 \min(2^m, n))\) time steps, where \(m\) is an upper bound on the intersection number and \(n\) is the number of vertices. Consequently, when \(m \leq \ln \ln n\) the running time of this algorithm is polynomial.

We then consider random instances of the random intersection graphs model as input graphs. As our main contribution, we prove that, when the number of labels is not too large (\(m = n^{\alpha}, 0 < \alpha < 1\)), we can use the label choices of the vertices to find a maximum clique in polynomial time whp. The proof of correctness for this algorithm relies on our Single Label Clique Theorem, which roughly states that whp a “large enough” clique cannot be formed by more than one label. This theorem generalizes and strengthens other related results in the state of the art, but also broadens the range of values considered (see e.g. \([20]\) and \([9]\)).

As an important consequence of our Single Label Clique Theorem, we prove that the problem of inferring the complete information of label choices for each vertex from the resulting random intersection graph (i.e. the label representation of the graph) is solvable whp; namely, the maximum likelihood estimation method will provide a unique solution (up to permutations of the labels). Finding efficient algorithms for constructing such a label representation is left as an interesting open problem for future research.

1 Introduction

A clique in an undirected graph \(G\) is a subset of vertices any two of which are connected by an edge. The cardinality of the maximum clique is called the clique number of \(G\). The problem of finding the maximum clique in an arbitrary graph is fundamental in Theoretical Computer Science and appears in many different settings. As an example, consider a social network where vertices represent people and edges represent mutual acquaintance. Finding a maximum clique in this network corresponds to finding the largest subset of people who all know each other. More generally, the analysis of large networks in order to identify communities, clusters, and other latent structure has come to the forefront of much research. The Internet, social networks, bibliographic databases, energy distribution networks, and global networks of economies are some of the examples motivating the development of the field.

It is well known that determining the clique number of an arbitrary graph is NP-complete \([15]\). In fact, the fastest algorithm known today runs in time \(O(1.1888^n)\) \([18]\), where \(n\) is the number of vertices in the graph. Moreover, the best known approximation algorithm for the clique number has a performance guarantee of \(O\left(\frac{n(\log \log n)^{\frac{11}{2}}}{(\log n)^{\frac{1}{2}}}\right)\) \([7]\) (there are algorithms with better approximation ratios for graphs with large clique number; see e.g. \([1]\)). Even though this approximation ratio appears to be weak at first glance,
there are several results on hardness of approximation which suggest that there can be no approximation algorithm with an approximation ratio significantly less than linear (see e.g. [10]). It was also shown in [4] that, if $k$ is the clique number, then the clique problem cannot be solved in time $n^{o(k)}$, unless the exponential time hypothesis fails (note that the brute force search algorithm runs in time $O(n^k k^2)$, which seems quite close).

The intractability of the maximum clique problem for arbitrary graphs lead researchers to the study of the problem for appropriately generated random graphs. In particular, for Erdős-Rényi random graphs $G_{n,\frac{1}{2}}$ (i.e. random graphs in which each edge appears independently with probability $\frac{1}{2}$), there are several greedy algorithms that find a clique of size about $\ln n$ with high probability (whp, i.e. with probability that tends to 1 as $n$ goes to infinity), see e.g. [9,14]. Since the clique number of $G_{n,\frac{1}{2}}$ is asymptotically equal to $2 \ln n$ with high probability, these algorithms approximate the clique number by a factor of 2. In fact, it was conjectured that finding a clique of size $(1 + \epsilon) \ln n$ (for a constant $\epsilon > 0$), with probability at least $\frac{1}{2}$, would require techniques beyond the current limits of complexity theory. This belief was strengthened by the fact that the Metropolis algorithm also fails to find the maximum clique in $G_{n,\frac{1}{2}}$ (see [11]). A more dramatized version of the above conjecture was presented in [11], stating that the problem of finding an $1.01 \ln n$ clique remains hard even if the input graph is a $G_{n,\frac{1}{2}}$ random graph in which we have planted a randomly chosen clique of size $n^{0.49}$. This conjecture has some interesting cryptographic consequences, as shown in [12]. It also seems tight, since finding the maximum clique in the case where the planted clique has size at least $\sqrt{n}$ can be done in polynomial time by using spectral properties of the adjacency matrix of the graph (see [2]). We finally note that there are quite a few nice results concerning generalizations of the planted clique problem in various (quite general) random graphs models (see e.g. [5,6]).

1.1 Our Contribution

In this work, we complement the state of the art by relating the maximum clique problem to the intersection number of the input graph $G$ (i.e. the minimum number of cliques that can edge cover $G$). In particular, we consider the maximum clique problem for graphs with small intersection number and random intersection graphs.

More analytically, we begin by considering arbitrary graphs with small intersection number. We present a simple algorithm which, on input $G$, finds a maximum clique in $O(2^n + O(m) + n^2 \min\{2^m, n\})$ time steps, where $m$ is an upper bound on the intersection number of $G$ and $n$ is the number of vertices. Consequently, when $m \leq \ln \ln n$ the running time of this algorithm is polynomial. We note here that computing the exact value of the independence number of $G$ is itself an NP-complete problem, but this knowledge is only needed in the analysis of the algorithm.

We then consider random instances of the random intersection graphs model (introduced in [13,20]) as input graphs. In this model, denoted by $G_{n,m,p}$, 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. Random intersection graphs are relevant to and capture quite nicely social networking. Indeed, a social network is a 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 include oblivious resource sharing in a (general) distributed setting, efficient and secure communication in sensor networks [16], interactions of mobile
agents traversing the web etc. Even epidemiological phenomena (like spread of disease) tend to be more accurately captured by this “interaction-sensitive” random graph model.

As our main contribution, we prove that, when the number of labels is not too large, we can use the label choices of the vertices to find a maximum clique in polynomial time (in the number of labels $m$ and vertices $n$ of the graph). Most of the work in this paper is devoted in proving our Single Label Clique Theorem (Theorem 3 in Section 4). Our proof technique is original and employs a probabilistic contradiction argument. The theorem states that when the number of labels is less than the number of vertices, any large enough clique in a random instance of $G_{n,m,p}$ is formed by a single label. This statement may seem obvious when $p$ is small, but it is hard to imagine that it still holds for all “interesting” values for $p$ (see also the discussion in Section 2). Indeed, when $p = o\left(\sqrt{\frac{1}{nm}}\right)$, by slightly modifying an argument of [3], we can see that $G_{n,m,p}$ almost surely has no cycle of size $k \geq 3$ whose edges are formed by $k$ distinct labels (alternatively, the intersection graph produced by reversing the roles of labels and vertices is a tree). On the other hand, for larger $p$ a random instance of $G_{n,m,p}$ is far from perfect and the techniques of [3] do not apply (for a more thorough discussion see the beginning of Section 4). By using the Single Label Clique Theorem, we provide a tight bound on the clique number of $G_{n,m,p}$ when $m = n^\alpha, \alpha < 1$. A lower bound in the special case where $mp^2$ is constant, was given in [20]. We considerably broaden this range of values to also include vanishing values for $mp^2$ and also provide an asymptotically tight upper bound.

We claim that our proof also applies for $\alpha < 2$, provided $p$ is not too small. We should note here that in [8] the authors prove the equivalence (measured in terms of total variation distance) of random intersection graphs and Erdős-Rényi random graphs, when $m = n^\alpha, \alpha > 6$. This bound on the number of labels was improved in [19], by showing equivalence of sharp threshold functions among the two models for $\alpha \geq 3$. In view of these results, we expect that our work will shed light also in the problem of finding maximum cliques in Erdős-Rényi random graphs.

Finally, as yet another consequence of our Single Label Clique Theorem, we prove that the problem of inferring the complete information of label choices for each vertex from the resulting random intersection graph (i.e. the label representation of the graph) is solvable whp; namely, the maximum likelihood estimation method will provide a unique solution (up to permutations of the labels) [1]. In particular, given values $m, n$ and $p$, such that $m = n^\alpha, 0 < \alpha < 1$, and given a random instance of the $G_{n,m,p}$ model, the label choices for each vertex are uniquely defined. Finding efficient algorithms for constructing such a label representation is left as an open problem for future research.

1.2 Organization of the paper

In Section 2, we formally define random intersection graphs. We also provide some useful definitions and notation which are used throughout the paper. The relation of the intersection number to the clique number of an arbitrary graph is discussed in Section 3. Section 4 is devoted to the proof of our Single Label Clique Theorem for random intersection graphs. The consequences of our main theorem concerning the efficient construction of a

---

3 A perfect graph is a graph in which the chromatic number of every induced subgraph equals the size of the largest clique of that subgraph. Consequently, the clique number of a perfect graph is equal to its chromatic number.

4 More precisely, if $\mathcal{B}$ is the set of different label choices that can give rise to a graph $G$, then the problem of inferring the complete information of label choices from $G$ is solvable if there is some $B^* \in \mathcal{B}$ such that $\Pr(B^*(G)) > \Pr(B(G))$, for all $\mathcal{B} \ni B \neq B^*$.
maximum clique and the uniqueness of the label representation of $G_{n,m,p}$ are presented in Section 5. Finally, we discuss the presented results and further research in Section 6.

2 Definitions and Preliminaries

The formal definition of the random intersection graphs model is as follows:

**Definition 1 (Random Intersection Graph - $G_{n,m,p}$ [13,20])**. Consider a universe $\mathcal{M} = \{1, 2, \ldots, m\}$ of elements and a set of $n$ vertices $V$. Assign independently to each vertex $v \in V$ a subset $S_v$ of $\mathcal{M}$, choosing each element $i \in \mathcal{M}$ independently with probability $p$ and draw an edge between two vertices $v \neq u$ if and only if $S_v \cap S_u \neq \emptyset$. The resulting graph is an instance $G_{n,m,p}$ of the random intersection graphs model.

In this model we also denote by $L_i$ the set of vertices that have chosen label $i \in \mathcal{M}$. Given $G_{n,m,p}$, we will refer to $\{L_i, i \in \mathcal{M}\}$ as its label representation. Consider the bipartite graph with vertex set $V \cup \mathcal{M}$ and edge set $\{(v, i) : i \in S_v\} = \{(v, i) : v \in L_i\}$. We will refer to this graph as the bipartite random graph $B_{n,m,p}$ associated to $G_{n,m,p}$. Notice that the associated bipartite graph is uniquely defined by the label representation.

It follows from the definition of the model that the edges in $G_{n,m,p}$ are not independent. In particular, the (unconditioned) probability that a specific edge exists is $1 - (1 - p^2)^m$. Therefore, if $mp^2$ goes to infinity with $n$, then this probability goes to 1. In the paper, we will thus consider the “interesting” range of values $mp^2 = O(1)$ (i.e. the range of values for which the unconditioned probability that an edge exists does not go to 1). Furthermore, as is usual in the literature, we will assume that the number of labels is some power of the number of vertices, i.e. $m = n^\alpha$, for some $\alpha > 0$.

The following definitions will also be useful:

**Definition 2 (Intersection number)**. The intersection number of a graph $G$ is the smallest number of cliques needed to cover all of the edges of $G$.

Equivalently, the intersection number is the smallest number of elements in a representation of $G$ as an intersection graph of finite sets.

**Definition 3 (Edge clique cover)**. A set of cliques $\mathcal{C} = \{C_1, \ldots, C_m\}$ is an edge clique cover of a graph $G = (V, E)$ if for every edge $e \in E$ there is at least one clique $C_i$ such that $e \in C_i$ and for every non edge $e' \notin E$, there is no such clique in $\mathcal{C}$.

Therefore, the intersection number of $G$ is the minimum $m$ such that $\mathcal{C} = \{C_1, \ldots, C_m\}$ is an edge clique cover of $G$.

2.1 Notation

We use the convention that the random intersection graphs model is denoted by $\mathcal{G}_{n,m,p}$ (i.e. with a calligraph $\mathcal{G}$), while a specific random instance of the model is denoted by $G_{n,m,p}$ (i.e. with a simple $G$).

For a vertex $v \in V$, we denote by $N_G(v)$ the set of neighbors of $v$ in $G$. We will say that two vertices $v, u \in V$ belong to the same closed neighborhood in $G$ and we will write $v \leftrightarrow_G u$ if and only if $N_G(v) \cup \{v\} = N_G(u) \cup \{u\}$.

Let $\mathcal{C}'$ denote a partition of the vertex set $V$ of a graph $G$ and let $v \in V$. We will denote by $\mathcal{C}'[v]$ the unique set inside $\mathcal{C}'$ that contains $v$, that is $\mathcal{C}'[v] = \{C' \in \mathcal{C}' : v \in C'\}$.

Throughout the paper, we make use of the well known asymptotic notation $O(\cdot), \Omega(\cdot), o(\cdot)$ and $\omega(\cdot)$. Furthermore, we use the relation “$\sim$” for asymptotically equal. In particular, if $f(n), g(n)$ are two functions of $n$, then $f(n) \sim g(n)$ means that $\lim_{n \to \infty} \frac{f(n)}{g(n)} = 1$ or equivalently $f(n) = g(n) + o(g(n))$. 

4
3 An Algorithm for Maximum Clique

In this section we consider arbitrary graphs as input graphs for the maximum clique problem. In particular, we relate the running time of the following algorithm to the intersection number of the input graph $G$.

```
Algorithm FIND_MAX-CLIQUE
Input: $G = (V, E)$
1. Set $U = V$ and $C' = \emptyset$;
2. while $U \neq \emptyset$ do
3. Pick $v \in U$ and let $C' = \{u \in U : u \leftrightarrow_G v\}$;
4. Include $C'$ in $C'$;
5. Set $U = U \setminus C'$; endwhile
6. Let $G' = (V', E')$ be an induced subgraph of $G$ that has exactly one vertex for every set $C' \in C'$;
7. Using exhaustive search, find a clique $S$ in $G'$ such that $|\cup_{v' \in S} C'[v']|$ is maximum;
8. Output $Q = \cup_{v' \in S} C'[v']$;
```

An example of how the graph $G'$ is constructed (in step 6) for a specific graph $G$ is shown in Figure 1. Notice that $G$ has five closed neighborhoods (whereas its intersection number is 3), which are shown in dashed squares, so the graph $G'$ has 5 vertices. The corresponding clique of $G'$ that maximizes $|\cup_{v' \in S} C'[v']|$ is $S = \{4, 6\}$.

![Figure 1](An example of a graph $G$ and corresponding $G'$.)

3.1 Analysis of FIND_MAX-CLIQUE

We first present the following lemma that concerns basic properties of the relation $\leftrightarrow_G$.

**Lemma 1.** The closed neighborhood relation $\leftrightarrow_G$ is an equivalence relation with the following properties:

1. It is an equivalence relation which partitions the vertex set $V$ in equivalence classes called closed neighborhoods.
2. A closed neighborhood is a clique. Two closed neighborhoods either form a clique, or no edge between their vertices exists.

\textbf{Proof.} (1) The fact that $\leftrightarrow_G$ is an equivalence relation follows directly by its definition. Therefore, every vertex belongs to exactly one equivalence class (i.e. exactly one closed neighborhood).

(2) By definition, a closed neighborhood forms a clique. Let now $C'_1, C'_2$ be two distinct closed neighborhoods and let $v \in C'_1, u \in C'_2$. Suppose that there is an edge $(u, v)$ between $u$ and $v$ in $G$, i.e. $u \in N_G(v)$. Consider now any two vertices $v' \in C'_1, u' \in C'_2$ (including $v, u$). By definition of the closed neighborhood relation, we must have that $N_G(v) \cup \{v\} = N_G(v') \cup \{v\}$. Since the closed neighborhoods are disjoint, this means that $u' \in N_G(v')$. Therefore, either every edge between $C'_1$ and $C'_2$ appears in $G$, and $C'_1 \cup C'_2$ forms a clique, or no edge between them exists. This completes the proof. \hfill $\square$

We now prove the following theorem about the correctness of the Algorithm $\text{FIND\_MAX\_CLIQUE}$.

\textbf{Theorem 1 (Correctness).} $\text{FIND\_MAX\_CLIQUE}$ correctly outputs a maximum clique in $G$.

\textbf{Proof.} Notice that, by the second part of Lemma \[1\] and by construction of $G'$, any clique $S$ in $G'$ corresponds to a clique $\cup_{v' \in S} C'[v']$ in $G$.

Therefore, we only need to show that a maximum clique $Q$ of $G$ corresponds to a clique in $G'$, because then the algorithm will be able to find it in step 7. Equivalently, we need to show that there are $k \geq 1$ closed neighborhoods $C'_1, \ldots, C'_k$ which constitute a partition of $Q$, that is $\cup_{i=1}^k C'_i = Q$. Indeed, by construction of $G'$, the vertices in $G'$ that correspond to these closed neighborhoods will form a clique in $G'$ (any choice of two vertices will be connected).

To prove the above, let $C'$ be a closed neighborhood that has at least one common vertex $v'$ with $Q$, i.e. $v' \in C' \cap Q$. Then, by definition of the $\leftrightarrow_G$ relation, every vertex $u' \leftrightarrow_G v'$ is connected to $v'$ and to all the vertices that $v'$ is connected to (including all vertices in $Q$). Therefore, by maximality of $Q$, all the vertices in $C'$ must be contained in the maximum clique, i.e. $C' \subseteq Q$. Consequently, a closed neighborhood is either entirely contained in $Q$, or disjoint from it. By the first part of Lemma \[1\] we can then partition $Q$ using all the closed neighborhoods that have common vertices with $Q$. This completes the proof. \hfill $\square$

The following result relates the running time of Algorithm $\text{FIND\_MAX\_CLIQUE}$ to the intersection number of its input graph $G$.

\textbf{Theorem 2 (Efficiency).} Let $G = (V, E)$ be a graph with intersection number $m$. Then $\text{FIND\_MAX\_CLIQUE}$ on input $G$ finds a maximum clique in $O(2^m + O(m) + n^2 \min\{2^m, n\})$ time steps.

\textbf{Proof.} By definition, since the intersection number of $G$ is $m$, there is a set of cliques $\mathcal{C} = \{C_1, \ldots, C_m\}$ that is an edge clique cover of $G$. For a vertex $v \in V$, we denote by $S_v$ the set of cliques in $\mathcal{C}$ that include $v$. Notice then that if $S_v = S_u$, then not only are $u$ and $v$ connected, but they also have the exact same set of neighbors in $V \setminus \{u, v\}$, i.e. $N_G(v) \cup \{v\} = N_G(u) \cup \{u\}$.

Given now a specific edge clique cover $\mathcal{C}'$, there are at most $2^m$ different ways in which we can construct a set $S_v$. Consequently, there are at most $2^m \leq n$ distinct closed neighborhoods $C'_1, \ldots, C'_{2m}$ in $G$ which constitute a partition of the set of non-isolated vertices.
Note also that determining whether or not \( v \leftrightarrow_G u \) for any two vertices requires \( O(n) \) steps. Therefore, steps 2 to 5 needed for partitioning the vertex set \( V \) in closed neighborhoods in the algorithm require \( O(n^2 \min\{2^m, n\}) \) time.

From the above, we also conclude that the number of vertices in \( G' \) is at most \( 2^m \). Therefore, the time needed to construct \( G' \) in step 6 in the algorithm is \( O(2^{2m}) \). Finally, there are at most \( 2^{2m} \) subsets of vertices in \( G' \), so step 7 in the algorithm takes \( O(2^{2m} + 2m) \) time. This completes the proof.

\[ \square \]

Note that the algorithm does not need the actual value of the independence number. We only use this information for bounding its running time. The following is a direct consequence of Theorem 2.

**Corollary 1.** Let \( m \leq \ln \ln n \) be an upper bound on the independence number of an arbitrary undirected graph \( G \) on \( n \) vertices. Then there is an algorithm that finds the maximum clique of \( G \) in time \( O(n^2 \ln n) \).

As a final remark, since the intersection number of \( G_{n,m,p} \) is at most \( m \) (but could be even less), the above result also holds for any random instance of the random intersection graphs model with at most \( \ln \ln n \) labels.

### 4 Clique number for \( m = n^\alpha, 0 < \alpha < 1 \)

In this section we give a tight bound on the clique number of \( G_{n,m,p} \) when \( m = n^\alpha, \alpha < 1 \). A lower bound in the special case where \( mp^2 \) is constant, was given in [20]. We considerably broaden this range of values to also include vanishing values for \( mp^2 \) and also provide a tight upper bound.

We will also assume, without loss of generality, that \( p = \Omega\left(\sqrt{\frac{1}{nm}}\right) \). Indeed, when \( p = o\left(\sqrt{\frac{1}{nm}}\right) \), by slightly modifying an argument of [3], we can see that \( G_{n,m,p} \) almost surely has no cycle of size \( k \geq 3 \) whose edges are formed by \( k \) distinct labels. Therefore, the maximum clique of \( G_{n,m,p} \) when \( p = o\left(\sqrt{\frac{1}{nm}}\right) \), is formed by exactly one label. As a matter of fact, if \( L_l \) is the set of vertices that have chosen label \( l \in \mathcal{M} \), then the maximum clique is equal to \( L_l \), where \( l \in \arg \max_{i \in \mathcal{M}} |L_i| \). Furthermore, since \( G_{n,m,p} \) is chordal whp (see Lemma 5 in [3]), the maximum clique can be found in polynomial time.

We stress out the fact that the techniques employed to provide the algorithmic and structural results in [3] cannot be used in the case where \( p = \Omega\left(\sqrt{\frac{1}{nm}}\right) \). In particular, \( G_{n,m,p} \) is far from perfect, especially in the the case \( mp = \omega(\ln n) \) (which is included in the range of values that we study here). An intuitive justification is as follows: when \( mp = \omega(\ln n) \), then the size of the label sets of every vertex are highly concentrated around their mean value \( mp \). Therefore, the statistical behavior of \( G_{n,m,p} \) is expected to be similar to the statistical behavior of uniform random intersection graphs \( G_{n,m,\lambda} \), in which each vertex selects exactly \( \lambda = mp \) labels from \( \mathcal{M} \). It was proved in [17] (part (iii) in Corollary 2), that the size of the maximum independent set when \( m = n^\alpha, \alpha < 1 \) and \( \lambda = \omega(\ln n) \), is asymptotically equal to \( 2(1 - \alpha)\frac{\ln n}{\lambda} \). Therefore, when \( mp = \omega(\ln n) \), the size of the maximum independent set in \( G_{n,m,p} \) will be around \( \Theta\left(\frac{\ln n}{mp^2}\right) \), so its chromatic number will be \( \Omega\left(\frac{mp^2}{\ln n}\right) \). However, as can be seen in Corollary [4] (which is a direct consequence of our main theorem), the size of the maximum clique in \( G_{n,m,p} \) when \( m = n^\alpha, \alpha < 1 \) and \( mp^2 = O(1) \) is asymptotically equal to \( np \). This is much smaller than the lower bound
Let $G_{n, m, p}$ be a random instance of the random intersection graphs model with $m = n^\alpha$, $0 < \alpha < 1$ and $p = \Omega\left(\sqrt{\frac{1}{nm}}\right)$. Then the following hold:

A. Let $L_i$ be the set of vertices that have chosen label $i \in M$. Then

$$\Pr(\exists i \in M : |L_i| - np \geq 3\sqrt{np\ln n}) \leq \frac{1}{n^3} \to 0. \quad (1)$$

B. Let also $S_v$ denote the set of labels that were chosen by vertex $v$. Then

$$\Pr(\exists v \in V : |S_v| > mp + 3\sqrt{mp\ln m + \ln n}) \to 0. \quad (2)$$

Proof. For the first part, fix a label $i \in M$. Notice that $|L_i|$ is a binomial random variable with parameters $n, p$, i.e. $|L_i| \sim B(n, p)$. By Chernoff bounds, for any $t \geq 0$, we have that

$$\Pr(|L_i| - np \geq t) \leq e^{-\frac{t^2}{2(np + t) + \frac{3}{2}np}} + e^{-\frac{t^2}{4np}}.$$  

Setting $t = 3\sqrt{np\ln n}$ and noting that $t = o(np)$, we then have that $\Pr(|L_i| - np \geq 3\sqrt{np\ln n}) \leq e^{-4\ln n}$ and the lemma follows from Boole’s inequality.

For the second part, fix a vertex $v$. Notice that $|S_v|$ is a binomial random variable with parameters $m, p$, i.e. $|S_v| \sim B(m, p)$. By Chernoff bounds, for any $\delta \geq 0$, we have that

$$\Pr(|S_v| > (1 + \delta)mp) < \left(\frac{e^\delta}{(1 + \delta)^{(1+\delta)}}\right)^{mp}.$$  

Setting $\delta = \frac{1}{mp}(3\sqrt{mp\ln m + \ln n})$ and using Boole’s inequality we get the desired result. \qed

Notice that the above lemma provides a lower bound on the clique number. However, a clique in $G_{n, m, p}$ can be formed by combining more than one label. Clearly, a clique $Q$ which is not formed by a single label will need at least 3 labels, since 2 labels cannot cover all the edges needed for $Q$ to be a clique. In the discussion below, we will provide a much larger lower bound on the number of labels needed to form a clique $Q$ of size $|Q| \sim np$ which is not formed by a single label. The following definition will be useful.

Definition 4. Denote by $A_{y, x}$ the event that there are two disjoint sets of vertices $V_1, V_2 \subset V$, where $|V_1| = y$ and $|V_2| = x$ such that the following hold:

1. All vertices in $V_1$ have chosen some label $l_0$, i.e. $l_0 \in \cap_{u \in V_1} S_u$.
2. None of the vertices in $V_2$ has chosen $l_0$, i.e. $l_0 \notin \cup_{v \in V_2} S_v$.
3. Every vertex in $V_1$ is connected to every vertex in $V_2$.

As a warm-up, we prove the following technical lemma, which is a first indication that in a $G_{n, m, p}$ graph, whp we cannot have $y$ too large and $x$ too small at the same time. This lemma will also be used as a starting step in the proof of our main theorem.
Lemma 3. Let $G_{n,m,p}$ be a random instance of the random intersection graphs model with $m = n^\alpha$, $0 < \alpha < 1$ and $p = \Omega\left(\frac{1}{nm}\right)$ and $mp^2 = O(1)$. Then, for any $y \geq np(1 - o\left(\frac{1}{\ln n}\right))$, $\Pr(A_{y,1}) = o(1)$.

Proof. Fix a particular label $l_0$, a subset $V_1$ of the vertices having chosen $l_0$ (i.e. $V_1 \subset L_{l_0}$) and a vertex $v \notin L_{l_0}$. The probability that $v$ is connected to all vertices in $V_1$ is exactly

$$p(V_1, v) = \sum_{k=1}^{m-1} \binom{m-1}{k} p^k(1-p)^{m-k-1}(1-(1-p)^k)^y. \quad (3)$$

Indeed, $p^k(1-p)^{m-k-1}$ is the probability that $v$ has chosen $k$ specific labels different from $l_0$ and $(1-(1-p)^k)^y$ the probability that a specific vertex in $V_1$ has chosen at least one of those labels (so that it is connected to $v$).

By Boole’s and Markov’s inequality we then have that

$$\Pr(A_{y,1}) \leq m \left(\frac{|L_{l_0}|}{y}\right)(n - |L_{l_0}|)p(V_1, v) \quad (4)$$

By Lemma 2 for any vertex $v$, we have that $|S_v| \leq (1 + o(1))mp + \ln n$ whp. Since $(1 - (1-p)^k)^y$ is increasing in $k$ and also $(m-1)^{-1}p^k(1-p)^{m-k-1}$ is maximum around $mp$, we conclude that the maximum of $(m-1)^{-1}p^k(1-p)^{m-k-1}(1-(1-p)^k)^y$ for $k \in \{1 \ldots (1 + o(1))mp\}$ is attained at some index $k' = (1 + o(1))mp$. Therefore,

$$\Pr(A_{y,1}) \leq m^2 n \left(\frac{|L_{l_0}|}{y}\right)\left(\frac{m-1}{k'}\right)p^{k'}(1-p)^{m-k'-1}(1-(1-p)^{k'})^y + o(1) \quad (5)$$

$$\leq m^2 n \left(\frac{|L_{l_0}|}{y}\right)\left(1-(1-p)^{k'}\right)^y + o(1) \quad (6)$$

where the $o(1)$ term corresponds to the error term from Lemma 2. Using now the fact that \( (by \ the \ expansion \ of \ the \ natural \ logarithm) (1 - p)^\frac{1}{p} = e^{\frac{1}{p} \ln (1-p)} = e^{-\sum_{j=1}^{\infty} \frac{j^{p-1}}{j} \geq e^{-1 - \sum_{j=2}^{\infty} \frac{1}{j^{1-p}}} \geq e^{-1 - \frac{1}{1-p}} \geq e^{-1}, \) for any $p \to 0$, we have that

$$\Pr(A_{y,1}) \leq m^2 n \left(\frac{|L_{l_0}|}{y}\right)\left(1- e^{-2mp^2}\right)^y + o(1) \quad (7)$$

$$= m^2 n \left(\frac{|L_{l_0}|}{|L_{l_0}| - y}\right)\left(1- e^{-2mp^2}\right)^y + o(1) \quad (8)$$

$$\leq m^2 n (|L_{l_0}|\ln |L_{l_0}|)^y\left(1 - e^{-2mp^2}\right)^y + o(1) \quad (9)$$

For any $y \geq |L_{l_0}| \left(1 - o\left(\frac{1}{\ln n}\right)\right)$, we then have that $\Pr(A_{y,1}) \to 0$. But by Lemma 2 we have also that $|L_{l_0}| \leq np \left(1 + o\left(\frac{1}{\ln n}\right)\right)$, which completes the proof.

The above lemma has the following alternative interpretation, which will be useful in the sequence:

Corollary 2. Let $G_{n,m,p}$ be a random instance of the random intersection graphs model with $m = n^\alpha$, $0 < \alpha < 1$, $p = \Omega\left(\frac{1}{nm}\right)$ and $mp^2 = O(1)$. Let also $Q$ be a clique in $G_{n,m,p}$ that is not formed by a single label and also $|Q| \sim np$. If $l_0 \in \mathcal{M}$ is any label chosen by some vertex $v \in Q$, then there is a positive constant $c' < \frac{1}{2p}$, such that whp there are at least $n^{c'}$ vertices in $Q$ that have not chosen $l_0$.
Proof. Notice that, by assumption, \( np = \Omega(n^{\frac{1-\alpha}{2}}) \). Therefore, for any positive \( \epsilon' < \frac{1-\alpha}{2} \), we have that \( n^{\epsilon'} = o\left(\frac{np}{\ln^2 n}\right) \). The result then follows by Lemma 3.

We now strengthen the above analysis by using the following simple observation: For a set of vertices \( V_2 \) and \( k \geq 2 \), let \( S_{V_2}^{(k)} \subseteq \mathcal{M} \) denote the set of labels that have been chosen by at least \( k \) of the vertices in \( V_2 \). Then the probability that every vertex of a set of vertices \( V_1 \) is connected to every vertex in \( V_2 \) is at most

\[
p(V_1, V_2) \leq \left( |S_{V_2}^{(2)}|p + (1-p)|S_{V_2}^{(2)}| \prod_{v \in V_2} \left( 1 - (1-p)^{|S_v - S_{V_2}^{(2)}|} \right) \right)^y
\]

(10)

\[
\leq \left( |S_{V_2}^{(2)}|p + \prod_{v \in V_2} \left( 1 - (1-p)^{|S_v|} \right) \right)^y
\]

(11)

Indeed, the first of the above inequalities corresponds to the probability that each vertex in \( V_2 \) either choses one of the labels shared by at least two vertices in \( V_2 \), or it is connected to all vertices in \( V_2 \) by using labels chosen by exactly one vertex in \( V_2 \).

Lemma 4. Let \( G_{n,m,p} \) be a random instance of the random intersection graphs model with \( m = n^\alpha \), \( 0 < \alpha < 1 \), \( p = \Omega\left(\frac{1}{\ln m}\right) \) and \( mp^2 = O(1) \). Let also \( x = \frac{1}{p^2} \), for some positive constant \( \epsilon < 1 \) that can be as small as possible. Then, for any \( y \geq np^{1+c} \), where \( 0 < c < \frac{1-\alpha}{1+\alpha} \) is a constant, we have \( \Pr(A_{y,x}) = o(1) \).

Proof. Fix a set \( V_2 \) of \( x \) vertices. We first give an upper bound on the size of \( S_{V_2}^{(2)} \). Towards this end, let \( X = |S_{V_2}^{(2)}| \) and notice that \( X \) is binomially distributed with parameters \( m, \hat{p} = 1 - (1-p)^x - xp(1-p)^{x-1} \). Since, by assumption \( xp \to 0 \), we have that \( \hat{p} \leq \frac{x^2}{p} \).

Therefore \( X \) is stochastically dominated by a binomial random variable \( Y \sim \mathcal{B}\left(m, \frac{x^2}{p}\right) \).

By Chernoff bounds we then have, for any \( t \geq 0 \),

\[
\Pr\left( X > \frac{mx^2p^2}{2} + t \right) \leq e^{-\frac{\left(\frac{mx^2p^2}{2} + t\right)^2}{2\left(\frac{mx^2p^2}{2} + \frac{tx^2}{p}\right)}}
\]

(12)

Set \( t = \frac{1}{p^{3+y+\epsilon'}} \), where \( \epsilon' \) is a positive constant that can be as small as possible. Since \( mp^2 = O(1) \), we have that \( t = o\left(\frac{mx^2p^2}{2}\right) \). By Boole’s inequality then, the probability that there is a subset \( V_2 \) of \( x \) vertices that has \( |S_{V_2}^{(2)}| > \frac{mx^2p^2}{2} + \frac{tx^2}{p} \) is at most

\[
nxe^{-\frac{1}{3yp^{3+y+\epsilon'}}} = o(1).
\]

(13)

Now that we have an upper bound on the size of \( S_{V_2}^{(2)} \) that holds whp, notice that by the second part of Lemma 2 and the fact that \( mp^2 = O(1) \), whp we have

\[
\prod_{v \in V_2} \left( 1 - (1-p)^{|S_v|} \right) \leq \frac{1}{2^{o(x)}} = o(|S_{V_2}^{(2)}|p).
\]

(14)

Therefore, by (11), we have that \( p(V_1, V_2) \leq 2|S_{V_2}^{(2)}|p |V_1| \). By Boole’s and Markov’s inequality we then have that
\[ \Pr(A,y,x) \leq m \left( \frac{|L_{l_0}|}{y} \right)^n p(V_1, V_2) \]
\[ \leq m \left( \frac{|L_{l_0}|}{y} \right)^n (2|S_{l_2}^{(2)}|p)^y + o(1) \]
\[ \leq m \left( \frac{|L_{l_0}|}{y} \right)^n \left( 2p^{1-2\epsilon - \epsilon'} \right)^y + o(1) \]

where the \( o(1) \) term corresponds to the error terms from Lemma 2 and equation 13. Using now the first part of Lemma 4 and an upper bound for the binomial coefficient we have

\[ \Pr(A,y,x) \leq m \left( \frac{8np}{y} \right)^y \left( p^{1-2\epsilon - \epsilon'} \right)^y + o(1). \]

Setting \( y = np^{1+c} \), for any positive constant \( c < \frac{1-\alpha}{1+\alpha} \), we have \( y \geq 1 \) and also that \( \Pr(A,y,x) = o(1) \). This completes the proof.

Lemma 3 has the following interpretation:

**Corollary 3.** Let \( G_{n,m,p} \) be a random instance of the random intersection graphs model with \( m = n^\alpha, 0 < \alpha < 1 \), \( p = \Omega \left( \sqrt{\frac{1}{nm}} \right) \) and \( mp^2 = O(1) \). Let also \( Q \) be a clique in \( G_{n,m,p} \) that is not formed by a single label and also \( |Q| \sim np \). Then whp, for any label \( l_0 \in \mathcal{M} \), we have that \( |Q \cap L_{l_0}| \leq np^{1+c} \), where \( 0 < c < \frac{1-\alpha}{1+\alpha} \) is a constant.

In particular, if \( Q \) is not formed by a single label, then whp it is formed by at least \( \frac{1}{p^{\epsilon}} \) distinct labels.

**Proof.** By Corollary 2 if \( Q \) is not formed by a single label, then given any label \( l_0 \in \mathcal{M} \) which is chosen by some vertex \( v \in Q \), there is a positive constant \( \epsilon' < \frac{1-\alpha}{1+\alpha} \), such that whp there are at least \( n^{\epsilon'} \) vertices in \( Q \) that have not chosen \( l_0 \). Therefore, we can apply Lemma 4 using any \( \epsilon < \frac{2\epsilon'}{1+\alpha} \). More specifically, for any such \( \epsilon \) we have \( \Pr(A_{np^{1+c}, \frac{1}{p^{\epsilon}}}) = o(1) \).

Additionally, this implies that whp if \( Q \) is not formed by a single label, it needs at least \( \frac{np}{mp^{1+c}} = \frac{1}{p^{\epsilon}} \) distinct labels. This is also a lower bound on the number of labels needed by a vertex \( v \) in order to connect to all vertices in \( Q \).

Before presenting the proof of our main theorem, we prove the following useful lemma, which states that if a large clique is not formed by a single label, then it must contain a quite large clique \( Q' \) whose edges are formed by distinct labels.

**Lemma 5.** Let \( G_{n,m,p} \) be a random instance of the random intersection graphs model with \( m = n^\alpha, 0 < \alpha < 1 \), \( p = \Omega \left( \sqrt{\frac{1}{nm}} \right) \) and \( mp^2 = O(1) \). Let also \( Q \) be any clique in \( G_{n,m,p} \) that is not formed by a single label and also \( |Q| \sim np \). Then whp, \( Q \) contains a clique \( Q' \) whose edges are formed by distinct labels and whose size is at least \( p^{-\frac{2}{3}} \), for any positive constant \( c < \frac{1-\alpha}{1+\alpha} \).

**Proof.** Let \( Q' \) be a subset of \( Q \) which is maximal with respect to the following property \( P \): “to each pair of vertices \( u \neq v \) in \( Q' \) we can assign a distinct label \( l \), such that \( l \in S_u \cap S_v \).”

Consider now the set of vertices \( W = \{ w : S_w \cap S_{Q'} \neq 0 \} \), namely the set of vertices that share a label with at least 2 vertices in \( Q' \) (note that \( Q' \subseteq W \), because every pair
of vertices in $Q$ is connected). Since $Q'$ is maximal, the set $Q - W$ must be the empty set. Indeed, if $z \in Q - W$, then (barring in mind that $Q$ is a clique) $z$ can be connected to each vertex in $Q'$ using distinct labels, which are also different from those already used to connect pairs of vertices in $Q'$. Therefore, $Q' \cup \{z\}$ would also have property $P$, which contradicts the maximality of $Q'$.

By Corollary 3 now, we have that $|W| \leq |S_Q^{(2)}| np^{1+c}$, where $0 < c < \frac{1-\alpha}{1+\alpha}$ is a constant. Furthermore, by equation (13), we have that $|S_Q^{(2)}| \leq \frac{m|Q'|^2p^2 + |Q'|^2}{p}$ whp, for any $\epsilon' > 0$ that can be as small as possible. Combining the above, and since $mp^2 = O(1)$, we have that

$$|W| \leq \frac{np^{1+c}|Q'|^2}{(1 + o(1))p^{\epsilon'}}.$$  

(19)

Consequently, the requirement $Q - W = \emptyset$ translates to

$$|Q| - \frac{np^{1+c}|Q'|^2}{(1 + o(1))p^{\epsilon'}} \leq 0$$  

(20)

or equivalently

$$|Q'| \geq \sqrt{\frac{|Q|}{(1 + o(1))np^{1+c-\epsilon'}}}.$$  

(21)

Baring in mind that $|Q| \sim np$ and that $\epsilon' > 0$ can be as small as possible, this completes the proof.

We now present our main theorem.

**Theorem 3 (Single Label Clique Theorem).** Let $G_{n,m,p}$ be a random instance of the random intersection graphs model with $m = n^\alpha, 0 < \alpha < 1$ and $mp^2 = O(1)$. Then whp, any clique $Q$ of size $|Q| \sim np$ in $G_{n,m,p}$ is formed by a single label. In particular, the maximum clique is formed by a single label.

**Proof.** We first note that, as discussed in the beginning of section 4, when $p = o\left(\sqrt{\frac{1}{nm}}\right)$, by slightly modifying an argument of [3] (in particular Lemma 5 there), we can see that $G_{n,m,p}$ almost surely has no cycle of size $k \geq 3$ whose edges are formed by $k$ distinct labels. Therefore, the maximum clique of $G_{n,m,p}$ when $p = o\left(\sqrt{\frac{1}{nm}}\right)$, is formed by exactly one label and our theorem holds. Consequently, we will assume w.l.o.g. for the remainder of the proof that $p = \Omega\left(\sqrt{\frac{1}{nm}}\right)$.

Let $Q$ be a clique of size $|Q| \sim np$ in $G_{n,m,p}$. By Lemma 5 if $Q$ is not formed by a single label, then $G_{n,m,p}$ must contain a clique $Q'$ whose edges are formed by distinct labels and whose size is at least $\beta \overset{def}{=} p^{-\frac{\epsilon'}{2}}$, for any positive constant $c < \frac{1-\alpha}{1+\alpha}$. By Markov’s inequality, the probability that such a $Q'$ exists in $G_{n,m,p}$ is at most

$$n^\beta \prod_{k=1}^{\beta-1} \left(\frac{m}{\beta - k}\right)^{p(\beta - k)}.$$  

(22)

Indeed, we can choose the vertices in $Q'$ arranged in a line in at most $n^\beta$ ways. Then we can select the labels needed for the $k$-th vertex to connect to all vertices to its right in at most $(\frac{m}{\beta - k})$ ways and each such label must be chosen by the $k$-th vertex, as well as another
vertex to its right (hence the term $p^{2(\beta-k)}$ in the product). Upper bounding the binomial coefficients in the above and using the fact $mp^2 = O(1)$, we get

$$\Pr\{Q' \text{ exists in } G_{n,m,p}\} \leq n^\beta \prod_{k=1}^{\beta-1} \left(\frac{em}{\beta - k}\right)^{\beta-k} p^{2(\beta-k)} \leq n^\beta \prod_{k=1}^{\beta-1} \left(\frac{\Theta(1)}{\beta - k}\right)^{\beta-k} = e^{\beta \ln n + O(\beta^2)} \prod_{k=1}^{\beta-1} \left(\frac{1}{\beta - k}\right)^{\beta-k} \leq e^{\beta \ln n + O(\beta^2)} \prod_{k=\beta/2}^{\beta-1} \left(\frac{2}{\beta}\right)^{\beta-k} = e^{\beta \ln n + O(\beta^2)} \frac{1}{\beta^{\Theta(\beta^2)}} = o(1).$$

(23) (24) (25) (26)

Therefore, whp $Q'$ does not exist in $G_{n,m,p}$, which completes the proof.

Notice that, by Theorem 3, the maximum clique in $G_{n,m,p}$ with $m = n^\alpha, 0 < \alpha < 1$ and $mp^2 = O(1)$ must be one of the sets $L_l, l \in M$. Therefore, the clique number of $G_{n,m,p}$ can be bounded using the first part of Lemma 2. In particular

**Corollary 4.** Let $G_{n,m,p}$ be a random instance of the random intersection graphs model with $m = n^\alpha, 0 < \alpha < 1$, $p = \Omega\left(\sqrt{\frac{1}{nm}}\right)$ and $mp^2 = O(1)$. Then, whp the maximum clique $Q$ of $G_{n,m,p}$ satisfies $|Q| \sim np$.

5 **Label Reconstruction**

One of the implications of our main Theorem 3 is that whp we can find the maximum clique in $G_{n,m,p}$ with $m = n^\alpha, 0 < \alpha < 1$ and $mp^2 = O(1)$ in polynomial time, just by looking at the associated bipartite graph $B_{n,m,p}$. In the following algorithm, we denote by $L_i$ the set of neighbors of label $i \in M$ in $B_{n,m,p}$, which can be determined in $O(n)$ time.

| Algorithm MAX-CLIQUE_FROM_LABELS |
|-----------------------------------|
| **Input:** $B_{n,m,p}$ |
| 1. Set $Q = \emptyset$; |
| 2. for $i = 1$ to $m$ do |
| % Check if the clique induced by label $i$ is larger % |
| 3. if $|L_i| > |Q|$ then set $Q = L_i$; endfor |
| 4. Output $Q$; |

By Theorem 3 when $m = n^\alpha, 0 < \alpha < 1$ and $mp^2 = O(1)$, Algorithm MAX-CLIQUE_FROM_LABELS returns the maximum clique of $G_{n,m,p}$ whp, in $O(nm)$ time. Therefore, the randomness of the model works in our favor for this case. Indeed, since any graph can be written as an intersection graph with at most $\binom{n}{2}$ labels, the problem of finding a maximum clique in a graph, given its label representation remains NP-complete.
Furthermore, it remains hard even when the intersection number is \( n^\alpha, 0 < \alpha < 1 \) unless the exponential time hypothesis fails (see e.g. [4]).

This leads to the following natural question: Could one infer any information about the structure of the associated bipartite graph when provided with \( G_{n,m,p} \) (i.e. the vertices and the edges of the graph)? Notice here that a graph \( G_{n,m,p} \) can correspond to more than one associated bipartite graphs. However, we show here that the problem of finding the associated bipartite graph given \( G_{n,m,p} \) and the actual values of \( m, n \) and \( p \) is \emph{solvable} whp when the number of labels is less than the number of vertices; namely, the maximum likelihood estimation method will provide a unique solution (up to permutations of the labels). More specifically, if \( B_{n,m,p} \) is the set of non-isomorphic associated bipartite graphs that give rise to \( G_{n,m,p} \), then there is some \( B^* \in B_{n,m,p} \) such that \( \Pr(B^*|G_{n,m,p}) > \Pr(B|G_{n,m,p}), \) for all \( B_{n,m,p} \ni B \neq B^* \).

**Theorem 4.** Let \( G_{n,m,p} \) be a random instance of the random intersection graphs model with \( m = n^\alpha, 0 < \alpha < 1 \), \( p = \Omega \left( \frac{1}{\sqrt{nm}} \right) \) and \( mp^2 = O(1) \). Then, whp the bipartite graph \( B_{n,m,p} \) associated to \( G_{n,m,p} \) is uniquely determined, up to permutations of the labels.

**Proof.** Let \( L_i \) denote the set of vertices that have chosen label \( i \in \mathcal{M} \). Given \( G_{n,m,p} \), we will refer to \( \{L_i, i \in \mathcal{M}\} \) as its \emph{label representation}. Notice then that the associated bipartite graph is uniquely defined by the label representation.

Suppose now for the sake of contradiction that \( \{L_i^{(1)}, i \in \mathcal{M}\} \) and \( \{L_i^{(2)}, i \in \mathcal{M}\} \) are two distinct label representations (up to permutations of the labels) of \( G_{n,m,p} \), where \( m = n^\alpha, 0 < \alpha < 1 \), \( p = \Omega \left( \frac{1}{\sqrt{nm}} \right) \) and \( mp^2 = O(1) \). Notice that, by the first part of Lemma \( \ref{lem:label-representation} \) whp both of these label representations should satisfy \( |L_i^{(\xi)}| \sim np \), for any \( i \in \mathcal{M} \) and \( \xi = 1, 2 \).

Notice then that there must be a label \( l \), such that \( L_l^{(1)} \notin \{L_i^{(2)}, i \in \mathcal{M}\} \), i.e. the clique induced by label \( l \) can be edge covered by more than one other cliques of size asymptotically equal to \( np \). However, by Theorem \( \ref{thm:max-clique} \) whp no clique \( Q \) of size \( |Q| \sim np \) can be formed by more than one labels, which contradicts the assumption that \( L_l^{(1)} \notin \{L_i^{(2)}, i \in \mathcal{M}\} \).

Consequently, \( \{L_i^{(1)}, i \in \mathcal{M}\} \) and \( \{L_i^{(2)}, i \in \mathcal{M}\} \) must be similar, up to permutations of the labels, i.e. \( L_l^{(1)} \in \{L_i^{(2)}, i \in \mathcal{M}\} \), for every \( l \in \mathcal{M} \). This completes the proof.

Notice that the uniqueness of the bipartite graph can also be proved in the case where \( p = o \left( \frac{1}{\sqrt{nm}} \right) \). Indeed, in this case \( G_{n,m,p} \) almost surely has no cycle of size \( k \geq 3 \) whose edges are formed by \( k \) distinct labels (see also the beginning of Section \( \ref{sec:random-intersection} \)). Therefore, every clique of size at least \( 3 \) is formed by a single label and so the proof of Theorem \( \ref{thm:max-clique} \) applies in this (sparser) case also.

6 Conclusions

In this work, we studied the maximum clique problem by relating it to the intersection number of the input graph. In particular, we first proved that if the intersection number of the graph \( G \) is sufficiently small, then a simple algorithm can find a maximum clique in \( G \) in polynomial time. We then considered random instances of the random intersection graphs model as input graphs. In particular, by proving the Singe Label Clique Theorem, we provided new, more general and asymptotically tight bounds for the clique number of \( G_{n,m,p} \) when \( m = n^\alpha, \alpha < 1 \). We also claim that our proof carries over for \( \alpha < 2\).
provided there is a lower bound on $p$ (in particular, we claim that our analysis can be applied also for $mp^2 = \Theta(1)$). One of the consequences of our theorem is that we can use the label representation of $G_{n,m,p}$ to find a maximum clique in polynomial time whp. This raised the question of whether we could reconstruct the label choices of the vertices in $G_{n,m,p}$ given only the graph structure. We proved here that the label reconstruction problem is solvable whp when the number of labels is less than the number of vertices. Finding efficient algorithms for constructing such a label representation is left as an open problem for future research. In view of the equivalence results between random intersection graphs and Erdős-Rényi random graphs, we expect that our work will shed light also in the problem of finding maximum cliques for input graphs generated by the latter model.

References

1. Alon N. and Kahale N.: Approximating the independence number via the $\theta$-function. Math. Programming 80: 253-264 (1998)
2. Alon N., Krivelevich M. and Sudakov B.: Finding a large hidden clique in a random graph. Random Structures and Algorithms 13: 457-466 (1998)
3. Behrisch M., Taraz A. and Ueckerdt M.: Coloring random intersection graphs and complex networks. SIAM J. Discrete Math. 23, pp. 288-299 (2008)
4. Chen Jianer, Huang Xizhen, Kanj Iyad A., Xia Ge: Strong computational lower bounds via parameterized complexity. J. Comput. Syst. Sci. 72(8): 1346-1367 (2006)
5. Coja-Oghlan A.: A spectral heuristic for bisecting random graphs. Random Structures and Algorithms, 29: 351-398 (2006)
6. Coja-Oghlan A., Lanka A.: Finding planted partitions in random graphs with general degree distributions. SIAM Journal on Discrete Mathematics 23:1682-1714 (2009)
7. Feige, U.: Approximating maximum clique by removing subgraphs. SIAM Journal on Discrete Mathematics 18(2): 219-225 (2004)
8. Fill J.A., Sheinerman E.R., Singer-Cohen K.B: Random intersection graphs when $m = \omega(n)$: an equivalence theorem relating the evolution of the $G(n,m,p)$ and $G(n,p)$ models. Random Struct. Algorithms 16(2), 156-176 (2000)
9. Grimmett G.R. and McDiarmid C.J.H.: On coloring random graphs. Math. Proc. Cambridge Philos. Soc. 77: 313-324 (1975)
10. Hästad, J.: Clique is hard to approximate within $n^{1-\epsilon}$. Acta Mathematica 182: 105-142 (1999)
11. Jerrum M.: Large cliques elude the metropolis process. Random Structures and Algorithms 3: 347-359 (1992)
12. Juels A. and Peinado M.: Hiding cliques for cryptographic security. Proc. of the Ninth Annual ACM-SIAM SODA, ACM Press, 678-684 (1998)
13. Karoński M., Scheinerman E.R. and Singer-Cohen K.B.: On random intersection graphs: the subgraph problem. Combinatorics, Probability and Computing journal 8: 131-159 (1999)
14. Karp R.M.: Probabilistic analysis of some combinatorial search problems. Algorithms and Complexity: New Directions and Recent Results, J.F. Traub, Ed., Academic Press, New York (1976).
15. Karp R.M.: Reducibility among combinatorial problems. Complexity of computer computations, R.E. Miller and J.W. Thatcher (eds.), Plenum Press, New York, 85-103 (1972)
16. Nikoletseas S.E., Raptopoulos C.L., Spirakis P.G.: Communication and security in random intersection graphs models. WOWMOM 2011: 1-6
17. Nikoletseas S.E., Raptopoulos C.L., Spirakis P.G.: On the independence number and Hamiltonicity of uniform random intersection graphs. Theor. Comput. Sci. 412(48): 6750-6760 (2011)
18. Robson, J.M.: Finding a maximum independent set in time $O(2^n/4)$. (2001)
19. Rybarczyk K.: Equivalence of a random intersection graph and $G(n,p)$. Random Structures and Algorithms 38(1-2): 205-234 (2011)
20. Singer-Cohen K.B.: Random intersection graphs. PhD thesis, John Hopkins University (1995)