Selected Combinatorial Problems Through the Prism of Random Intersection Graphs Models

Paul G. Spirakis[1,2][0000−0001−5396−3749], Sotiris Nikoletseas[2,3][0000−0003−3765−5636], and Christoforos Raptopoulos[3][0000−0002−9837−2632]

[1] Department of Computer Science, University of Liverpool, UK
p.spirakis@liverpool.ac.uk
[2] Computer Technology Institute, Greece
[3] Computer Engineering & Informatics Department, University of Patras, Greece
{nikole,raptopox}@ceid.upatras.gr

1 Introduction and Motivation

We discuss a simple, yet general family of models, namely *Random Intersection Graphs (RIGs)*, initially introduced by Karoski et al. [4] and Singer-Cohen [10]. In such models there is a universe $\mathcal{M}$ of *labels* and each one of n vertices selects a random subset of $\mathcal{M}$. Two vertices are connected if and only if their corresponding subsets of labels intersect. A formal definition is given below:

Definition 1 (Random Intersection Graph - $\mathcal{G}_{n,m,p}$ [4, 10]). *Consider a universe $\mathcal{M} = \{1, 2, \ldots, m\}$ of labels 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 $i \in M$. Given a random instance $G_{n,m,p}$ of the random intersection graphs model, we will refer to $\{L_i, i \in \mathcal{M}\}$ as its *label representation*, and the corresponding matrix $\mathbf{R}$ with columns the incidence vectors of label sets assigned to vertices is called the *representation matrix*. Furthermore, we refer to the bipartite graph with vertex set $V \cup \mathcal{M}$ and edge set $\{(v, i) : i \in S_v\} = \{(v, i) : v \in L_i\}$ 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 (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. We can thus restrict the range of the parameters to 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 $\alpha > 0$.

It is worth mentioning that the edges in $G_{n,m,p}$ are not independent. For example, there is a strictly positive dependence between the existence of two edges

that share an endpoint (i.e. $\Pr(\exists\{u,v\}|\exists\{u,w\}) > Pr(\exists\{u,v\})$). This dependence is stronger the smaller the number of labels $\mathcal{M}$ includes, while it seems to fade away as the number of labels increases. In fact, by using a coupling technique, the authors in [3] proved the equivalence (measured in terms of total variation distance) of uniform random intersection graphs and Erdős-Rényi random graphs, when $m = n^\alpha, \alpha > 6$. This bound on the number of labels was improved in [5], where it was proved that the total variation distance between the two models tends to 0 when $m = n^\alpha, \alpha > 4$. Furthermore, [9] proved the equivalence of sharp threshold functions among the two models for $\alpha \geq 3$. Similarity of the two models has been proved even for smaller values of α (e.g. for any $\alpha > 1$) in the form of various translation results (see e.g. Theorem 1 in [8]), suggesting that some algorithmic ideas developed for Erdős-Rényi random graphs also work for random intersection graphs (and also weighted random intersection graphs). These results suggest that random intersection graphs are quite general and that known techniques for random graphs can be used in the analysis of random intersection graphs with a large number of labels.

Motivation. Random intersection graphs may model several real-life applications quite accurately. 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. RIG modeling is useful in the efficient blind selection of few encryption keys for secure communications over radio channels ([1]), as well as in k-Secret sharing between swarm mobile devices (see [2]). Furthermore, random intersection graphs are relevant to and capture quite nicely social networking. Indeed, a social network is a structure made of nodes tied by one or more specific types of interdependency, such as values, visions, financial exchange, friends, conflicts, web links etc. Other applications may include oblivious resource sharing in a distributed setting, interactions of mobile agents traversing the web, social networking etc. Even epidemiological phenomena (like spread of disease between individuals with common characteristics in a population) tend to be more accurately captured by this "proximity-sensitive" family of random graphs.

From an average case analysis algorithmic perspective, the number of labels m may be viewed as a parameter controlling the clique cover size of input graphs. It is worth noting that some combinatorial problems that are considered to be hard when the input is drawn from the Erdős-Rényi random graphs model are easily solved when the input is drawn from the random intersection graphs model and the representation matrix $\mathbf{R}$ is explicitly provided as part of the input (rather than just giving the constructed graph instance as input). One such example is the problem of finding a maximum clique in $G_{n,m,p}$, in the dense case $m = n^\alpha, \alpha < 1$. Furthermore, there are combinatorial optimization problems that can be naturally described as graph theoretical problems in generalizations of the aforementioned model of random intersection graphs. In this talk, we discuss some structural and algorithmic results regarding random inter-

section graphs and we present an interesting connection between the problem of discrepancy in random set systems and the problem of Max Cut in weighted random intersection graphs.

2 Maximum Cliques in Random Intersection Graphs

We discuss the Single Label Clique Theorem (SLCT) from [6], which states that when the number of labels is less than the number of vertices, any large enough clique in a random instance of $\mathcal{G}_{n,m,p}$ 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. Indeed, when $p = o\left(\sqrt{\frac{1}{nm}}\right)$, it can be proved 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 $\mathcal{G}_{n,m,p}$ is far from perfect[4] and thus the proof of the SLCT is based on a careful contradiction argument regarding the non-existence of large multi-label cliques.

3 Maximum Cut and Discrepancy in Random Set Systems

A natural weighted version of the random intersection graphs model was introduced in [7], where to each edge $\{u, v\}$ we assign weight equal to the number of common labels chosen by u and v, namely $|S_u \cap S_v|$. In particular, the weight matrix of a random instance of the weighted random intersection graphs model $\overline{\mathcal{G}}_{n,m,p}$ is equal to $\mathbf{R}^T\mathbf{R}$, where the columns of $\mathbf{R}$ are the incidence vectors of label sets assigned to vertices; we denote the corresponding random instance by $G(V, E, \mathbf{R}^T\mathbf{R})$.

We initially present some results from [7] regarding the concentration of the weight of a maximum cut of $G(V, E, \mathbf{R}^T\mathbf{R})$ around its expected value, and then show that, when the number of labels is much smaller than the number of vertices (in particular, $m = n^\alpha, \alpha < 1$), a random partition of the vertices achieves asymptotically optimal cut weight with high probability. Furthermore, in the case $n = m$ and constant average degree (i.e. $p = \frac{\Theta(1)}{n}$), we show that with high probability, a majority type randomized algorithm outputs a cut with weight that is larger than the weight of a random cut by a multiplicative constant strictly larger than 1.

Finally, we present a connection between the computational problem of finding a (weighted) maximum cut in $G(V, E, \mathbf{R}^T\mathbf{R})$ and the problem of finding a 2-coloring that achieves minimum discrepancy for a set system Σ with incidence matrix $\mathbf{R}$ (i.e. minimum imbalance over all sets in Σ). This connection

[4] 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.

was exploited in [7] by proposing a (weak) bipartization algorithm for the case $m = n, p = \frac{\Theta(1)}{n}$ that, when it terminates, its output can be used to find a 2-coloring with minimum discrepancy in a set system with incidence matrix $\mathbf{R}$. In fact, with high probability, the latter 2-coloring corresponds to a bipartition with maximum cut-weight in $G(V, E, \mathbf{R}^T\mathbf{R})$.

References

1. S. Dolev, S. Gilbert, R. Guerraoui and C.C. Newport: Secure Communication over Radio Channels. In Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC), pp. 105-114 (2008).
2. S. Dolev, L. Lahiani and M. Yung: Secret Swarm UnitReactive k-Secret Sharing. In Proceedings of the 8th International Conference on Cryptology in India (INDOCRYPT), pp. 123-137 (2007).
3. Fill, J., Sheinerman, E., and Singer-Cohen, K.: 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), 156176 (2000).
4. M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen: On random intersection graphs: the subgraph problem, Comb. Probab Comput. 8 (1999), 131-159.
5. Kim, J.H., Lee, S.J., Na, J.: On the total variation distance between the binomial random graph and the random intersection graph. Random Structures and Algorithms 52(4), 662679 (2018).
6. S.E. Nikoletseas, C.L. Raptopoulos, and P.G. Spirakis: Maximum cliques in graphs with small intersection number and random intersection graphs. Comput. Sci. Rev. 39: 100353 (2021).
7. Nikoletseas, S., Raptopoulos, C., Spirakis, P.: Max cut in weighted random intersection graphs and discrepancy of sparse random set systems. In: Proceedings of the 32nd International Symposium on Algorithms and Computation (ISAAC), pp. 116 (2021).
8. Raptopoulos, C., Spirakis, P.: Simple and efficient greedy algorithms for hamilton cycles in random intersection graphs. In: Proceedings of the 16th International Symposium on Algorithms and Computation (ISAAC), pp. 493504 (2005).
9. Rybarczyk, K.: Equivalence of a random intersection graph and $G(n, p)$. Random Structures and Algorithms 38(1-2), 205234 (2011).
10. K.B. Singer-Cohen: Random intersection graphs, Ph.D. thesis, John Hopkins University (1995).