Colouring Non-Sparse Random Intersection Graphs*

S. Nikoletseas[1,2], C. Raptopoulos[3], and P. G. Spirakis[1,2]

[1] Computer Technology Institute, P.O. Box 1122, 26110 Patras, Greece
[2] University of Patras, 26500 Patras, Greece
[3] Heinz Nixdorf Institute, Computer Science Departement, University of Paderborn, 33095 Paderborn, Germany
nikole@cti.gr, raptopox@hni.uni-paderborn.de, spirakis@cti.gr

Abstract. An intersection graph of n vertices assumes that each vertex is equipped with a subset of a global label set. Two vertices share an edge when their label sets intersect. Random Intersection Graphs (RIGs) (as defined in [18, 31]) consider label sets formed by the following experiment: each vertex, independently and uniformly, examines all the labels (m in total) one by one. Each examination is independent and the vertex succeeds to put the label in her set with probability p. Such graphs nicely capture interactions in networks due to sharing of resources among nodes. We study here the problem of efficiently coloring (and of finding upper bounds to the chromatic number) of RIGs. We concentrate in a range of parameters not examined in the literature, namely: (a) $m = n^\alpha$ for α less than 1 (in this range, RIGs differ substantially from the Erdös-Renyi random graphs) and (b) the selection probability p is quite high (e.g. at least $\frac{\ln^2 n}{m}$ in our algorithm) and disallows direct greedy colouring methods.

We manage to get the following results:

- For the case $mp \leq \beta \ln n$, for any constant $\beta < 1 - \alpha$, we prove that np colours are enough to colour most of the vertices of the graph with high probability (whp). This means that even for quite dense graphs, using the same number of colours as those needed to properly colour the clique induced by any label suffices to colour almost all of the vertices of the graph. Note also that this range of values of m, p is quite wider than the one studied in [4].
- We propose and analyze an algorithm CliqueColour for finding a proper colouring of a random instance of $\mathcal{G}_{n,m,p}$, for any $mp \geq \ln^2 n$. The algorithm uses information of the label sets assigned to the vertices of $G_{n,m,p}$ and runs in $O\left(\frac{n^2 mp^2}{\ln n}\right)$ time, which is polynomial in n and m. We also show by a reduction to the uniform random intersection graphs model that the number of colours required by the algorithm are of the correct order of magnitude with the actual chromatic number of $G_{n,m,p}$.

* This work was partially supported by the ICT Programme of the European Union under contract number ICT-2008-215270 (FRONTS). Also supported by Research Training Group GK-693 of the Paderborn Institute for Scientific Computation (PaSCo).

– We finally compare the problem of finding a proper colouring for $G_{n,m,p}$ to that of colouring hypergraphs so that no edge is monochromatic. We show how one can find in polynomial time a k-colouring of the vertices of $G_{n,m,p}$, for any integer k, such that no clique induced by only one label in $G_{n,m,p}$ is monochromatic.

Our techniques are novel and try to exploit as much as possible the hidden structure of random intersection graphs in this interesting range.

1 Introduction

We study random intersection graphs (RIGs), a relatively recent combinatorial model, that nicely captures interactions between nodes in distributed networks. Such interactions may occur for example when nodes blindly select resources (such as frequencies) from a limited globally available domain. For this graph model, we investigate the important combinatorial problem of vertex colouring, namely assigning integers (colours) to the vertices of the graph such that no adjacent vertices get the same colour. Colouring of sparse random intersection graphs was studied in [4]. The range of values that we consider here is different and gives quite denser graphs. Furthermore, our techniques are different than those used by the authors in [4]. Colouring properties provide useful insight to algorithmic design for important problems (like frequency assignment and concurrency control) in distributed networks characterized by dense interactions and resource limitations, such as wireless mobile and sensory networks.

1.1 Importance and Motivation

Random intersection graphs may be used to model several real-life applications characterized by dense, blind, possibly local interactions quite accurately (compared to the well known $G_{n,\hat{p}}$ 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 mobile, 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 the independence between 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. Particularly for security purposes, the random selection of elements in our graphs

can be seen as a way to establish local common keys on-line, without any global scheme for predistribution of keys. In such a case, the set of labels can be a global set of large primes (known to all) but each node selects uniformly at random only a few. Two nodes that have selected a common prime can communicate securely. Notice that no other node can know what numbers a different node has selected. Thus, the local communication is guaranteed to be secure. In the case when the shared resource is the wireless spectrum, then nodes choosing the same label (frequency) may interfere, and the corresponding link in the intersection graph abstracts a conflict, while a colour class (e.g. the vertices with the same colour) corresponds to wireless devices that can simultaneously access the wireless medium.

Random intersection graphs are also 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. In particular, as [8] explicitly suggests proposing RIGs with a power law degree distribution, people with common friends may become friends as well, since they probably share common attributes and attributes can be obtained quite randomly (such as social preference, hobbies etc). So when a person A connects to B and B connects to C, the probability of a connection between A and C is higher because of more probable attributes in common. Thus RIGs can abstract such tendency for triangle clusterings.

Other applications may include oblivious resource sharing in a (general) distributed setting, interactions of mobile agents traversing the web etc. Even epidemiological phenomena (like spread of disease) tend to be more accurately captured by these "interaction-sensitive" random graph models.

1.2 Related Work

Random intersection graphs, denoted by $G_{n,m,p}$, were first defined in [18, 31]. In this model, to each of the n vertices of the graph, a random subset of a universal set of m elements is assigned, by independently choosing elements with the same probability p. Two vertices u, v are then adjacent in the $G_{n,m,p}$ graph if and only if their assigned sets of elements have at least one element in common. Various properties of $G_{n,m,p}$ such as connectivity, degree distribution, independent sets, Hamilton cycles and its relation to the well known Bernoulli random graph model were investigated in [18, 32, 25, 12, 14].

In [4] the authors propose algorithms that whp probability colour sparse instances of $G_{n,m,p}$. In particular, for $m = n^{\alpha}, \alpha > 0$ and $p = o\left(\sqrt{\frac{1}{nm}}\right)$ they show that $G_{n,m,p}$ can be coloured optimally. Also, in the case where $m = n^{\alpha}, \alpha < 1$ and $p = o\left(\frac{1}{m \ln n}\right)$ they show that $\chi(G_{n,m,p}) \sim np$ whp. To do this, they prove that $G_{n,m,p}$ is chordal whp (or equivalently, the label graph does not contain cycles) and so a perfect elimination scheme can be used to find a colouring in

polynomial time. The range of values we consider here is different than the one needed for the algorithms in [4] to work. In particular, we study colouring $G_{n,m,p}$ for the wider range $mp \leq (1-\alpha)\ln n$, as well as the denser range $mp \geq \ln^2 n$. We have to note also, that the proof techniques used in [4] cannot be used in the range we consider, since the properties that they examine do not hold in our case. Hence a completely new approach is needed.

The book [22] contains several techniques for upper bounding the chromatic number $\chi(G)$ of arbitrary graphs G. For general graphs it seems that one cannot easily beat the bound $\chi(G) \leq \Delta$, where Δ is the maximum degree of G (or $\chi(G) \leq \Delta+1$ if G is a clique or an odd cycle). However, assuming that the graph G has some additional structure, many interesting and advanced techniques for bounding the chromatic number exist for proving bounds for the chromatic number. These techniques are in fact algorithms that have (a) an iterative part that can generally be implemented in polynomial time which is followed almost always by (b) an application of the Local Lemma, which does not always lead to an algorithm that runs in polynomial time. Using a technique like that, Johansson (see chapter 13 of [22]) proved that for triangle free graphs we have $\chi(G) \leq \frac{160\Delta}{\ln \Delta}$, which is the strongest result known so far. Using a modification of Johansson's technique, Frieze and Mubayi [15] proved a quite strong bound on the chromatic number of simple Hypergraphs.

Concerning part (b) of the above technique, when some additional assumptions are true (that can in general be thought of as a stricter form of the Local Lemma assumptions), then Beck's technique [3] (see also [24] where the authors put Beck's technique in a more general framework) can be used to actually convert the Local Lemma existential proof into a polynomial running time algorithm! We have to say here that there are not many polynomial time algorithms that colour graphs with a relatively small number of colours. The best known approximation algorithm gives an approximation of $O\left(n\frac{(\log\log n)^2}{(\log n)^3}\right)$. Also, Molloy and Reed [23] used Beck's technique to find a polynomial algorithm to colour optimally graphs whose chromatic number is close to their maximum degree Δ. Another notable algorithm is the one proposed by Alon and Kahale [2] that uses the second to last eigenvalue of a special case of random graphs that are 3-colourable (more specifically they are constructed starting with 3 sets of n vertices each and then drawing edges between any pair of vertices that lie on different sets independently with probability p; of course these graphs are 3-colourable by definition and it is assumed that one is given the graph as it is, but without any information concerning the 3 sets).

Colouring Bernoulli random graphs was considered in [5] and also [21]. As it seems to be implied by these two works, randomness sometimes allows for smaller chromatic number than maximum degree whp. For $G_{n,\hat{p}}$, it is shown that whp $\chi(G_{n,\hat{p}}) \sim \frac{d}{\log d}$, where d is the mean degree. We have to point out here that both [5] and [21] prove that there exists a colouring of $G_{n,\hat{p}}$ using around $\frac{d}{\log d}$, but their proof does not lead to polynomial time algorithms. In fact, to the best of

our knowledge, the problem of constructing a colouring of $G_{n,\hat{p}}$ using $\Theta\left(\frac{d}{\log d}\right)$ colours remains open for non-trivial values of $\hat{p}$.

Distributed Computing Related Work. From a distributed computing perspective, our work is related to collision avoidance and message inhibition methods ([20]) as well as range assignment problems in directional antennas' optimization ([7]). The (distant-2) chromatic number of random proximity and random geometric graphs has been studied in [9]. Furthermore, our colouring results can be applied in coordinating MAC access in sensor networks (see [6]). Our results also relate to distributed colouring and channel utilization ([19, 30]). Finally, the RIG modeling can be useful in the efficient blind selection of few encryption keys for secure communications over radio channels ([10]), as well as in k-Secret sharing between swarm mobile devices (see [11]).

1.3 Our Contribution

In this paper we study the problem of colouring a random instance of the random intersection graphs model $\mathcal{G}_{n,m,p}$, mainly for the interesting range $m = n^{\alpha}, \alpha < 1$, where the model seems to differ the most from Bernoulli random graphs (see [14] and [28]). In particular

- For the case $mp \leq \beta \ln n$, for any constant $\beta < 1-\alpha$, we prove that np colours are enough to colour most of the vertices of the graph with high probability (whp). This means that even for quite dense graphs, using the same number of colours as those needed to properly colour the clique induced by any label suffices to colour almost all of the vertices of the graph. Note also that this range of values of m, p is quite wider than the one studied in [4].
- We propose and analyze an algorithm CliqueColour for finding a proper colouring of a random instance of $\mathcal{G}_{n,m,p}$, for any $mp \geq \ln^2 n$. The algorithm uses information of the label sets assigned to the vertices of $G_{n,m,p}$ and runs in $O\left(\frac{n^2mp^2}{\ln n}\right)$ time, which is polynomial in n and m. We also show by a reduction to the uniform random intersection graphs model that the number of colours required by the algorithm are of the correct order of magnitude with the actual chromatic number of $G_{n,m,p}$.
- We finally compare the problem of finding a proper colouring for $G_{n,m,p}$ to that of colouring hypergraphs so that no edge is monochromatic. We show how one can find in polynomial time a k-colouring of the vertices of $G_{n,m,p}$, for any integer k, such that no clique induced by only one label in $G_{n,m,p}$ is monochromatic.

Our proof techniques try to take advantage of the special randomness of $G_{n,m,p}$ and the way that edges appear in it as part of cliques. Especially in the design of algorithm CliqueColor, by carefully colouring a few vertices we were able to reduce the complex problem of colouring the whole graph, to the problem of colouring a simpler one.

2 Definition of the Model

We now formally define the model of random intersection graphs.

Definition 1 (Random Intersection Graph - $G_{n,m,p}$ [18, 31]). *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 $G_{n,m,p}$.*

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

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 $d_G(v)$.

By the above definition, one can realize that the edges in a random intersection graph appear as parts of cliques. In particular, the sets $L_l, l \in \mathcal{M}$ are in fact a (not necessarily minimal) clique cover of $G_{n,m,p}$. The size of each such clique is a binomial random variable with parameters n, p. Similarly, the number of cliques that a vertex v belongs to (i.e. $|S_v|$) is a binomial random variable with parameters m, p. In general, one can imagine that the smaller p is, the smaller the intersections between different (label) cliques will be in the clique cover implied by the sets $L_l, l \in \mathcal{M}$. In the extreme case where the cliques $L_l, l \in \mathcal{M}$ are disjoint, one could colour $G_{n,m,p}$ optimally by using the sets L_l. The authors in [4] show that one can also optimally colour $G_{n,m,p}$ even in the case where the sets $L_l, l \in \mathcal{M}$ intersect, provided that there is no induced cycle of size more than 3 in the instance graph. In this paper we consider a different range of values for the parameters of $\mathcal{G}_{n,m,p}$ that give whp intersection graphs in which the intersection between the sets $L_l, l \in \mathcal{M}$ is much higher and the clique structure of the graph is more complex. Consequently, the techniques used in [4] cannot be used here.

A closely related model to $\mathcal{G}_{n,m,p}$ is the *Uniform Random Intersection Graphs Model*, denote by $\mathcal{G}_{n,m,\lambda}$, where λ is a positive integer, which was first defined in [17]. In this model, every vertex chooses independently, uniformly at random a set of exactly λ labels and then we connect vertices that have at least one label in common. It is worth mentioning here that, apart from the case where the number of labels chosen by a vertex in $\mathcal{G}_{n,m,p}$ is concentrated around its mean value, the probabilistic behavior of $\mathcal{G}_{n,m,\lambda}$ seems a lot different than the one of $\mathcal{G}_{n,m,p}$.

3 Colouring Almost all Vertices

We are going to consider the case where $m = n^\alpha$, for $\alpha \in (0, 1)$ some fixed constant. The area $mp = o\left(\frac{1}{\ln n}\right)$ gives almost surely instances in which the label graph (i.e. the dual graph where the labels in $\mathcal{M}$ play the role of vertices

and the vertices in V play the role of labels) is quite sparse and can be coloured optimally using $\max_{l\in\mathcal{M}}|L_l|$ colours (see [4]). We will here consider the denser area $mp = \Omega\left(\frac{1}{\ln n}\right)$. In this range of values, it is easy to see that the values of $|L_l|$ are concentrated around np. We were able to prove that even for values of the parameters m, p that give quite denser graphs, we can still use np colours to properly colour most of the graph. [1]. Our proof technique is inspired by analogous ideas of Frieze in [16] (see also [21]). Before presenting the main result, we state an auxiliary lemma that was proved in [25, 26] and will be useful in the proof.

Lemma 1 ([25, 26]). *Let $G_{n,m,p}$ be a random instance of the random intersection graphs model. Then the conditional probability that a set of k_0 vertices is an independent set, given that k_i of them are already an independent set is equal to*

$$\left((1-p)^{k_0-k_i} + (k_0-k_i)p(1-p)^{k_0-k_i-1}\left(1-\frac{k_ip}{1+(k_i-1)p}\right)\right)^m.$$

Proof. See [25, 26]. □

We are now ready to present our theorem.

Theorem 1. *When $m = n^\alpha, \alpha < 1$ and $mp \leq \beta \ln n$, for any constant $\beta < 1-\alpha$. Then a random instance of the random intersection graphs model $G_{n,m,p}$ contains a subset of at least $n - o(n)$ vertices that can be coloured using np colours, with probability at least $1 - e^{-n^{0.99}}$.*

Proof. In what follows, we will denote by $G_{n,m,p}$ an instance of the random intersection graphs model $\mathcal{G}_{n,m,p}$. We also denote by $B_{n,m,p}$ the bipartite graph associated to $G_{n,m,p}$. We prove a slightly stronger property than what the lemma requires.

Assume an arbitrary ordering of the vertices $v_1, v_2, \ldots, v_n$. For $i = 1, 2, \ldots, n$, let B_i be the subgraph of $B_{n,m,p}$ induced by $\cup_{j=1}^{i} v_j \bigcup \mathcal{M}$. We denote by H_i the intersection graph whose bipartite graph has vertex set $V \bigcup \mathcal{M}$ and edge set that is exactly as B_i between $\cup_{j=1}^{i} v_j$ and $\mathcal{M}$, whereas every other edge (i.e. the ones between $\cup_{j=i}^{n} v_j$ and $\mathcal{M}$) appears independently with probability p.

Set $x = np$. Let X denote the size of the largest x-colourable subset of vertices in $G_{n,m,p}$ and let X_i denote the expectation of the largest x-colourable subset in H_i. Notice that X_i is a random variable depending on the overlap between $G_{n,m,p}$ and H_i. Obviously, $X = X_n$ and setting $X_0 = E[X]$, we have $|X_i - X_{i+1}| \leq 1$, for all $i = 1, 2, \ldots, n$. It is straightforward to verify that the sequence $X_0, X_1, \ldots, X_n$ is a Doob Martingale (see also Chapter 9 of [29]). Hence, by applying Azuma's inequality, we have that

$$\Pr(|X - E[X]| \geq t) \leq 2e^{-\frac{t^2}{2n}}.$$

Set now $k_0 = \frac{(1-\epsilon^2)n}{x}$, where ϵ is a positive constant that is arbitrarily close to 0. For $t = \epsilon\frac{k_0 x}{1+\epsilon} = \epsilon(1-\epsilon)n$, Azuma's inequality becomes

[1] Note however, that this does not mean that the chromatic number is close to np, since the part that is not coloured could be a clique in the worst case

$$\Pr\left(|X - E[X]| \geq \epsilon(1-\epsilon)n\right) \leq 2e^{-\frac{\epsilon^2 n}{3}}. \quad (1)$$

Let now Y denote the number of x-colourable subsets of $(1+\epsilon)\frac{xk_0}{1+\epsilon}$ vertices in $G_{n,m,p}$, that can be split in exactly x independent sets (i.e. chromatic classes) of size exactly k_0. We can now verify that, proving that $\Pr(Y > 0)$ is greater or equal to the right hand side of inequality (1), i.e. $2e^{-\frac{\epsilon^2 n}{3}}$, then we will have proven that (a) $E[X] \geq \frac{xk_0}{1+\epsilon}$ and (b) that the values of X are concentrated around something greater than $\frac{xk_0}{1+\epsilon}$ with high probability. More specifically, (a) comes from the observation that the event $\{Y > 0\}$ implies the event $\{X \geq xk_0\}$, hence $\Pr(Y > 0) \leq \Pr(X \geq xk_0) = \Pr\left(X - \frac{xk_0}{1+\epsilon} \geq \frac{\epsilon xk_0}{1+\epsilon}\right) = \Pr\left(X - \frac{xk_0}{1+\epsilon} \geq \epsilon(1-\epsilon)n\right)$. If now $E[X]$ was strictly less than $\frac{xk_0}{1+\epsilon}$, then this would mean that $\Pr(Y > 0) < \Pr\left(X - E[X] \geq \epsilon(1-\epsilon)n\right)$ which by (1) is less than $2e^{-\frac{\epsilon^2 n}{3}}$. Hence, proving that $\Pr(Y > 0) \geq 2e^{-\frac{\epsilon^2 n}{3}}$ could only mean that $E[X] \geq \frac{xk_0}{1+\epsilon}$. Part (b) then follows as well.

The remarks (a) and (b) described above are sufficient to prove the theorem, since ϵ can be as small as possible. Since Y is a nonnegative random variable that takes only integral values, in order to bound $\Pr(Y > 0)$, we will use the well known inequality (see also exercise 1 of Chapter 4 in [1])

$$\Pr(Y > 0) \geq \frac{E^2[Y]}{E[Y^2]}.$$

Since every colour class considered in Y must have exactly k_0 vertices and obviously different colour classes must not overlap, we get that

$$E[Y] = \prod_{i=1}^{x} \binom{n-(i-1)k_0}{k_0} \left((1-p)^{k_0} + k_0 p(1-p)^{k_0-1}\right)^m$$

where the term $\left((1-p)^{k_0} + k_0 p(1-p)^{k_0-1}\right)^m \stackrel{def}{=} p_1$ is the probability that a colour class is indeed an independent set, that is no two vertices in it have a common label. Similarly, we have that

$$E[Y^2] \leq E[Y] \sum_{k_1,\ldots,k_x \leq k_0} \prod_{i=1}^{x} \binom{k_0}{k_i}\binom{n-ik_0}{k_0-k_i} p_2$$

where p_2 is the conditional probability that a colour class of k_0 vertices is an independent set, given that k_i of them are already an independent set. By Lemma 1, we have that

$$p_2 \stackrel{def}{=} \left((1-p)^{k_0-k_i} + (k_0-k_i)p(1-p)^{k_0-k_i-1}\left(1 - \frac{k_i p}{1+(k_i-1)p}\right)\right)^m.$$

Combining the above, we conclude that

$$\frac{1}{\Pr(Y>0)} \leq \frac{E[Y^2]}{E^2[Y]} \leq \sum_{k_1,\ldots,k_x \leq k_0} \prod_{i=1}^{x} \frac{\frac{k_0!}{k_i!(k_0-k_i)!}\frac{(n-ik_0)!}{(k_0-k_i)!(n-(i+1)k_0+k_i)!}}{\frac{(n-(i-1)k_0)!}{k_0!(n-ik_0)!}} \frac{p_2}{p_1}$$

$$\leq \sum_{k_1,\ldots,k_x \leq k_0} \prod_{i=1}^{x} \frac{\left(\frac{k_0!}{(k_0-k_i)!}\right)^2}{k_i!(n-ik_0)^{k_i}} \frac{p_2}{p_1} \tag{2}$$

The fraction $\frac{p_2}{p_1}$ can be bounded in a quite straightforward manner as follows

$$\sqrt[m]{\frac{p_2}{p_1}} \leq \frac{(1-p)^{k_0-k_i} + (k_0-k_i)p(1-p)^{k_0-k_i-1}}{(1-p)^{k_0} + k_0p(1-p)^{k_0-1}}$$

$$= \frac{1-p+(k_0-k_i)p}{1-p+k_0p}(1-p)^{-k_i} = \left(1 - \frac{k_ip}{1-p+k_0p}\right)(1-p)^{-k_i}$$

$$\leq e^{-\frac{k_ip}{1-p+k_0p}+k_ip} = e^{\frac{k_0k_ip^2-k_ip^2}{1-p+k_0p}} \leq e^{k_0k_ip^2}$$

where the last inequality follows since $k_0 \to \infty$ for $mp = O(\ln n)$ and $m = n^\alpha, \alpha < 1$.

For $i = 1, \ldots, x$, let $A_i \stackrel{def}{=} \frac{\left(\frac{k_0!}{(k_0-k_i)!}\right)^2}{k_i!(n-ik_0)^{k_i}} \frac{p_2}{p_1}$, so that $\frac{E[Y^2]}{E^2[Y]} \leq \sum_{k_1,\ldots,k_x \leq k_0} \prod_{i=1}^{x} A_i$. When $k_i = 0$, then trivially $A_i = 1$. On the other hand, when $1 \leq k_i \leq k_0$, using the inequalities $\frac{k_0!}{(k_0-k_i)!} \leq k_0^{k_i}, k_i! \geq \left(\frac{k_i}{e}\right)^{k_i}$ and the fact that $xk_0 = (1-\epsilon^2)n$, we can see that

$$A_i \leq \frac{k_0^{2k_i}}{k_i^{k_i}(n)^{k_i}} e^{mk_0k_ip^2} = e^{2k_i\ln k_0 - k_i\ln k_i - k_i\ln n + mk_0k_ip^2 + O(k_i\ln\ln n)} \tag{3}$$

We now distinguish two cases.

(a) $1 \leq k_i \leq \frac{k_0}{\ln^2 n}$. Then $A_i \leq e^{2k_i\ln n + mk_0k_ip^2} \leq e^{k_i(2\ln n + mp)} = e^{O\left(\frac{k_0}{\ln n}\right)}$, since $mp = O(\ln n)$.

(b) $\frac{k_0}{\ln^2 n} < k_i \leq k_0$. Then $A_i \leq e^{\alpha k_i\ln n - k_i\ln n + mk_0k_ip^2 + O(k_i\ln\ln n)} = o(1)$, since $\beta < 1-\alpha$. We should also mention that the $O(\cdot)$ part of the expontent is different than the $O(\cdot)$ part of the exponent in (3).

The crucial observation now is that, for all values of k_i, $A_i^x \leq e^{O\left(\frac{n}{\ln n}\right)}$. As a final note, the total number of terms in the sum $\sum_{k_1,\ldots,k_x \leq k_0}$ is $(k_0+1)^x = e^{x\ln(k_0+1)} \leq e^{n^{1-\alpha}\ln^2 n}$.

By (2), we then have that

$$\Pr(Y>0) \geq e^{-n^{1-\alpha}\ln^2 n - O\left(\frac{n}{\ln n}\right)} \geq 2e^{-\frac{\epsilon^2 n}{3}}$$

which concludes the proof.

□

It is worth noting here that the proof of Theorem 1 can also be used similarly to prove that $\Theta(np)$ colours are enough to colour $n - o(n)$ vertices even in the case where $mp = \beta \ln n$, for any constant $\beta > 0$. However, finding the exact constant multiplying np is technically more difficult.

4 A Polynomial Time Algorithm for the Case $mp \geq \ln^2 n$

In the following algorithm every vertex chooses i.u.a.r (independently, uniformly at random) a preference in colours, denoted by $shade(\cdot)$ and every label l chooses a preference in the colours of the vertices in L_l, denoted by $c_l(\cdot)$.

Algorithm CliqueColour:
Input: An instance $G_{n,m,p}$ of $\mathcal{G}_{n,m,p}$ and its associated bipartite $B_{n,m,p}$.
Output: A proper colouring $G_{n,m,p}$.

1. for every $v \in V$ choose a colour denoted by $shade(v)$ independently, uniformly at random among those in $\mathcal{C}$;
2. for every $l \in \mathcal{M}$ choose a colouring $c_l(\cdot)$ of the vertices in L_l such that for every colour in $\{c \in \mathcal{C} : \exists v \in L_l$ with $shade(v) = c\}$ there is exactly one vertex in the set $\{u \in L_l : shade(u) = c\}$ having $c_l(u) = c$ while the rest remain uncoloured;
3. set $U = \emptyset$ and $C = \emptyset$;
4. **for** $l = 1$ **to** m **do** {
5. colour every vertex in $L_l \backslash \{U \cup C\}$ according to $c_l(\cdot)$ iff there is no collision with the colour of a vertex in $L_l \cap C$;
6. include every vertex in L_l coloured that way in C and the rest in U; }
7. let $\mathcal{H}$ denote the (intersection) subgraph of $G_{n,m,p}$ induced by the vertices in U;
8. give a proper colouring of $\mathcal{H}$ using a new set of colours $\mathcal{C}'$;
9. **output** a colouring of $G_{n,m,p}$ using $|\mathcal{C} \cup \mathcal{C}'|$ colours;

It is easy to see now that the above algorithm provides a proper colouring of its input graph. The number of colours that it needs (i.e. the cardinality of the sets $\mathcal{C}$ and $\mathcal{C}'$) and the time needed to colour $\mathcal{H}$ in step 8 are considered in Theorem 3.

Theorem 2 (Correctness). *Given an instance $G_{n,m,p}$ of the random intersection graphs model, algorithm CliqueColour always finds a proper colouring.*

Proof. For the sake of contradiction, suppose that in the colouring proposed by the algorithm there are two vertices v_1 and v_2 that are connected and have been assigned to the same colour c. This of course means that these two vertices have at least one label in common. Since the sets $\mathcal{C}$ and $\mathcal{C}'$ are disjoint and the colouring of $\mathcal{H}$ provided at step 8 of the algorithm is proper, the only way that such a collision would arise is if both v_1 and v_2 belong to C. This means that both were coloured by the first pass of the algorithm and also $shade(v_1) =$

$shade(v_2) = c$. Let l be the smallest indexed label in $|S_{v_1} \cap S_{v_2}|$. It is easy to see then that we come to a contradiction, as label l and step 5 will guarantee that at least one of the two vertices lies in U.

□

The following theorem concerns the efficiency of algorithm CliqueColour, provided that additionally $mp \geq \ln^2 n$ and $p = o\left(\frac{1}{\sqrt{m}}\right)$. Notice that for p larger than $\frac{1}{\sqrt{m}}$, every instance of the random intersection graphs model $\mathcal{G}_{n,m,p}$, with $m = n^\alpha, \alpha < 1$, is complete whp.

Theorem 3 (Efficiency). *Algorithm CliqueColour succeeds in finding a proper $\Theta\left(\frac{nmp^2}{\ln n}\right)$-colouring using of $G_{n,m,p}$ in polynomial time, provided that $mp \geq \ln^2 n, p = o\left(\frac{1}{\sqrt{m}}\right)$ and $m = n^\alpha, \alpha < 1$.*

Proof. For $s \in \mathcal{C}$, let Z_c denote the number of vertices $v \in V$ such that $shade(v) = c$. Z_c is a binomial random variable, so by Chernoff bounds we can see that, for any positive constant β_1 that can be arbitrarily small

$$\Pr\left(\left|Z_c - \frac{n}{|\mathcal{C}|}\right| \geq \frac{\beta_1 n}{|\mathcal{C}|}\right) \leq 2e^{-\frac{\beta^2 n}{3|\mathcal{C}|}}.$$

For $|\mathcal{C}| = \Theta\left(\frac{mnp^2}{\ln n}\right)$ and $p = o\left(\frac{1}{\sqrt{m}}\right)$, we can then use Boole's inequality to see that there is no $c \in \mathcal{C}$ such that $\left|Z_c - \frac{n}{|\mathcal{C}|}\right| \geq \frac{\beta_1 n}{|\mathcal{C}|}$, with probability $1 - o(1)$, i.e. almost surely.

Using the same type of arguments, we can also verify that for arbitrarily small positive constants β_2 and β_3, we have that $\Pr(\exists v \in V : ||S_v| - mp| \geq \beta_2 mp) = o(1)$ and $\Pr(\exists l \in \mathcal{M} : ||L_l| - np| \geq \beta_3 np) = o(1)$, for all $mp = \omega(\ln n)$ and $m = n^\alpha, \alpha < 1$.

We will now prove that the maximum degree of the graph $\mathcal{H}$ is small enough to allow a proper colouring of $\mathcal{H}$ using $\mathcal{C}' = \Theta\left(\frac{nmp^2}{\ln n}\right)$ colours. For a label $l \in \mathcal{M}$ let Y_l denote the number of vertices $v \in L_l$ such that $c_l(v) \neq shade(v)$. In order for a label l not to be able to assign colour $shade(v)$ to $v \in L_l$, it should be the case that it has assigned colour $shade(v)$ to another vertex $u \in L_l$ with $shade(u) = shade(v)$. Hence, the only way to have a collision is when two or more vertices with the same shade have all chosen label l. Notice also that in order to have $Y_l \geq k$, the number of different shades appearing among the vertices that have chosen label l should be at most $|L_l| - k$. This means that $\Pr(Y_l \geq k) \leq \binom{|L_l|}{k}\left(\frac{|L_l|-k}{|\mathcal{C}|}\right)^k$. Given the concentration bound for $|L_l|$, we have that $\Pr(\exists l : Y_l \geq k)$ is at most

$$m\binom{(1+\beta_3)np}{k}\left(\frac{(1+\beta_3)np - k}{|\mathcal{C}|}\right)^k + o(1) \leq m\left(\frac{3np}{k}\right)^k\left(\frac{2np}{|\mathcal{C}|}\right)^k + o(1).$$

By now setting $k = \frac{np}{\ln n}$ and for $|\mathcal{C}| \geq 18\frac{mnp^2}{\ln n}$ we then have that, with probability $1 - o(1)$, there is no label $l \in \mathcal{M}$ such that $Y_l \geq \frac{np}{\ln n}$.

For a label $l \in \mathcal{M}$ now let W_l be the number of vertices $v \in L_l$ such that $shade(v) = c_l(v)$ but they remained uncoloured, hence included in $\mathcal{H}$. In order for a vertex $v \in L_l$ to be counted in W_l, there should exist a label j prior to l (i.e. a label among $1, \ldots l-1$) such that $v \in L_j$ and there is another vertex $u \in L_j$ with $shade(u) = shade(v)$. The probability that this happens is at most $p\left(1-(1-p)^{Z_{shade(v)}}\right)\left(1+(1-p)+(1-p)^2+\cdots\right) = 1-(1-p)^{Z_{shade(v)}}$. The crucial observation now is that, because choices of labels by vertices (of the same shade or not) is done independently and because the vertices counted in W_l have (by definition of the colouring $c_l(\cdot)$ in step 2 of the algorithm) different shades, the inclusion in W_l of any vertex $u \in L_l$ with $shade(u) = c_l(u)$ does not affect the inclusion of another $v \in L_l \backslash \{u\}$ with $shade(v) = c_l(v)$. Hence, taking also into account the concentration bound for $Z_{shade(v)}$ and $|L_l|$, we have that

$$\Pr(\exists l : W_l \geq k') \leq m\binom{(1+\beta_3)np}{k'}\left(1-(1-p)^{(1+\beta_1)\frac{n}{|\mathcal{C}|}}\right)^{k'} + o(1).$$

By now setting $k' = \frac{np}{\ln n}$ and using the relation $(1-x)^y \sim 1-xy$, valid for all x, y such that $xy = o(1)$, we have that when $|\mathcal{C}| \geq 18\frac{mnp^2}{\ln n}$, there is no label l such that $W_l \geq \frac{np}{\ln n}$, with high probability.

We have then proved that the number of vertices in U of the algorithm that have chosen a specific label is with high probability at most $\frac{2np}{\ln n}$. Since, for any vertex v in $G_{n,m,p}$ has $|S_v| \leq (1+\beta_2)mp$, we conclude that the maximum degree in $\mathcal{H}$ satisfies $\max_{v\in\mathcal{H}} degree_{\mathcal{H}}(v) \leq (1+\beta_2)mp\frac{2np}{\ln n}$. It is then evident that we can colour $\mathcal{H}$ greedily, in polynomial time, using $\frac{2.1nmp^2}{\ln n}$ more colours, with high probability. Hence, we can colour $G_{n,m,p}$ in polynomial time, using at most $\frac{20.1nmp^2}{\ln n}$ colours in total.

□

It is worth noting here that the number of colours used by the algorithm in the case $mp \geq \ln^2 n, p = O\left(\frac{1}{\sqrt[4]{m}}\right)$ and $m = n^\alpha, \alpha < 1$ is of the correct order of magnitude. Indeed, by the concentration of the values of $|S_v|$ around mp for any vertex v with high probability, one can use the results of [27] for the uniform random intersection graphs model $G_{n,m,\lambda}$, with $\lambda \sim mp$ to provide a lower bound on the chromatic number. Indeed, it can be easily verified that the independence number of $G_{n,m,\lambda}$, for $\lambda = mp \geq \ln^2 n$ is at most $\Theta\left(\frac{\ln n}{mp^2}\right)$, which implies that the chromatic number of $G_{n,m,\lambda}$ (and hence of the $G_{n,m,p}$ because of the concentration of the values of $|S_v|$) is at least $\Omega\left(\frac{nmp^2}{\ln n}\right)$.

5 Colouring Random Hypergraphs

The model of random intersection graphs $\mathcal{G}_{n,m,p}$ could also be though of as generating random Hypergraphs. The Hypergraphs generated have vertex set

V and edge set $\mathcal{M}$. There is a huge amount of literature concerning colouring hypergraphs. However, the question about colouring there seems to be different from the one we answer in this paper. More specifically, a proper colouring of a hypergraph seems to be any assignment of colours to the vertices, so that no monochromatic edge exists. This of course implies that fewer colours than the the chromatic number (studied in this paper) are needed in order to achieve this goal.

We would also like to mention that as far as $\mathcal{G}_{n,m,p}$ is concerned, the problem of finding a colouring such that no label is monochromatic seems to be quite easier when p is not too small.

Theorem 4. *Let $G_{n,m,p}$ be a random instance of the model $\mathcal{G}_{n,m,p}$, for $p = \omega(\frac{\ln m}{n})$ and $m = n^{\alpha}$, for any fixed $\alpha > 0$. Then with high probability, there is a polynomial time algorithm that finds a k-colouring of the vertices such that no label is monochromatic, for any fixed integer $k \geq 2$.*

Proof. By Chernoff bounds and Boole's inequality we can easily show that, for any constant $\epsilon > 0$ that can be arbitrarily small

$$\Pr(\exists l : ||L_l| - np| \geq \epsilon np) \leq 2me^{-\frac{\epsilon^2 np}{3}} \to 0$$

for any $p = \omega(\frac{\ln m}{n})$.

If we were to choose the colour of each vertex independently, uniformly at random among the available colours, then the mean number of monochromatic edges in $G_{n,m,p}$, would be almost surely (given that the above concentration bound holds)

$$E[\#\text{ monochromatic edges}] = \sum_{l \in \mathcal{M}} k^{1-|L_l|} \leq mk^{1-(1-\epsilon)np} < 1.$$

Then, using the method of conditional expectations (see [13, 22]) we can derive an algorithm that finds the desired colouring in time $O(mn^{k+2})$. Indeed, since $E[\#\text{ monochromatic edges}] < 1$, there must be a vertex v and a colour c, such that colouring v with c guarantees that $E[\#\text{ monochromatic edges}|colour(v) = c] < 1$. This, combined with the fact that, given any colouring $\mathcal{C}_S$ of any subset S of vertices, we can compute $E[\#\text{ monochromatic edges}|\mathcal{C}_S]$ in time $O(nm)$, leads to the desired algorithm.

□

6 Future Work

We are currently trying to extend the applicability of our colouring methods to other ranges of p, and to the related model $\mathcal{G}_{n,m,\lambda}$, where each vertex selects u.a.r. λ labels to form its corresponding label set.

References

1. N. Alon and J. Spencer, *"The Probabilistic Method"*, John Wiley & Sons, inc., (2000).
2. N. Alon and N. Kahale, *"A spectral technique for colouring random 3-colourable graphs"*, in the Proceedings of the 26th Annual ACM Symposium on Theory of Computing, pp. 346-355, (1994).
3. J. Beck, *"An Algorithmic Approach to the Lovász Local Lemma"*, Random Structures & Algorithms 2:343-365, (1991).
4. M. Behrisch, A. Taraz and M. Ueckerdt, *"Colouring random intersection graphs and complex networks"*, SIAM J. Discrete Math. Vol 23, Issue 1, pp. 288-299, (2009).
5. B. Bollobás, *"The chromatic number of random graphs"*, in Combinatorica, Vol. 8(1), pp. 49-55, (1988).
6. C. Busch, M. Magdon-Ismail, F. Sivrikaya and B. Yener, *"Contention-free MAC Protocols for Asynchronous Wireless Sensor Networks"*, Distributed Computing 21(1): 23-42, (2008).
7. I. Caragiannis, C. Kaklamanis, E. Kranakis, D. Krizanc, and A. Wiese, *"Communication in Wireless Networks with Directional Antennas"*, in the Proceedings of the 20th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA 08), pages 344-351, (2008).
8. M. Deijfen and W. Kets, *"Random Intersection Graphs With Tunable Degree Distribution and Clustering"*, CentER Discussion Paper Series No. 2007-08. Available at SSRN: `http://ssrn.com/abstract=962359`.
9. J. Diaz, Z. Lotker and M.J. Serna, *"The Distant-2 Chromatic Number of Random Proximity and Random Geometric Graphs"*, Inf. Process. Lett. 106(4): 144-148, (2008).
10. S. Dolev, S. Gilbert, R. Guerraoui and C.C. Newport, *"Secure Communication over Radio Channels"*, in the ACM Symposium on Principles of Distributed Computing (PODC), pp. 105-114, (2008).
11. S. Dolev, L. Lahiani and M. Yung, *"Secret Swarm UnitReactive k-Secret Sharing"*, in the 8th International Conference on Cryptology in India (INDOCRYPT), pp. 123-137, (2007).
12. C. Efthymiou and 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 Programmng (ICALP), pp. 690-701, (2005).
13. P. Erdös and J. Selfridge, *"On a Combinatorial Game"*, J. Combinatorial Th. (A) 14:298-301, (1973).
14. J.A. Fill, E.R. Sheinerman and 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`
15. A. Frieze and D. Mubayi, *"Colouring Simple Hypergraphs"*, `arXiv:0901.3699v1`, (2008).
16. A. Frieze, *"On the Independence Number of Random Graphs"*, Disc. Math., 81, pp. 171-175, (1990).
17. E. Godehardt and J. Jaworski, *"Two models of Random Intersection Graphs for Classification"*, Studies in Classification, Data Analysis and Knowledge Organization, Opitz O., Schwaiger M, (Eds), Springer Verlag, Berlin, Heidelberg, New York, pp. 67-82, (2002).

18. M. Karoński, E.R. Scheinerman and K.B. Singer-Cohen, *"On Random Intersection Graphs: The Subgraph Problem"*, Combinatorics, Probability and Computing journal 8, pp. 131-159, (1999).
19. K. Kothapalli, C. Scheideler, C. Schindelhauer and M. Onus, *"Distributed Coloring in $O(\sqrt{\log n})$ bits"*, in the 20th IEEE International Parallel & Distributed Processing Symposium (IPDPS), (2006).
20. P. Leone, L. Moraru, O. Powell and J.D.P. Rolim, *"Localization Algorithm for Wireless Ad-Hoc Sensor Networks with Traffic Overhead Minimization by Emission Inhibition"*, in ALGOSENSORS 2006, pp. 119-129.
21. T. Łuczak, *"The chromatic number of random graphs"*, in Combinatorica, Vol. 11(1), pp. 45-54, (2005).
22. M. Molloy and B. Reed, *"Graph Colouring and the Probabilistic Method"*, Springer, (2002).
23. M. Molloy and B. Reed, *"Colouring Graphs whose Chromatic Number is almost their Maximum Degree"*, in the Proceedings of Latin American Theoretical Informatics, pp. 216-225, (1998).
24. M. Molloy and B. Reed, *"Further Algorithmic Aspects of the Local Lemma"*, in the Proceedings of the 30th ACM Symposium on Theory of Computing, pp. 524-529, (1998).
25. S. Nikoletseas, C. Raptopoulos and 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), (2004).
26. S. Nikoletseas, C. Raptopoulos and P. Spirakis, *"Large Independent Sets in General Random Intersection Graphs"*, in the Theoretical Computer Science (TCS) Journal, Special Issue on Global Computing, (2008).
27. S. Nikoletseas, C. Raptopoulos and P. Spirakis, *"On the Independence Number and Hamiltonicity of Uniform Random Intersection Graphs"*, accepted in the 23rd IEEE International Parallel and Distributed Processing Symposium (IPDPS), (2009).
28. C. Raptopoulos and 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), pp. 493-504, (2005).
29. S. M. Ross, *"Stochastic Processes"*, John Wiley & Sons, Inc., 2nd edition, (2000).
30. C. Schindelhauer and K. Voss, *"Oblivious Parallel Probabilistic Channel Utilization without Control Channels"*, in the 20th IEEE International Parallel & Distributed Processing Symposium (IPDPS), (2006).
31. K.B. Singer-Cohen, *"Random Intersection Graphs"*, PhD thesis, John Hopkins University, (1995).
32. D. Stark, *"The Vertex Degree Distribution of Random Intersection Graphs"*, in Random Structures & Algorithms 24 (3), pp. 249-258, (2004).