COMBINATORICA 27 (5) (2007) 519–550
DOI: 10.1007/s00493-007-2169-9

COMBINATORICA
Bolyai Society – Springer-Verlag

COMPLETE PARTITIONS OF GRAPHS*

MAGNÚS M. HALLDÓRSSON, GUY KORTSARZ,
JAIKUMAR RADHAKRISHNAN,
SIVARAMAKRISHNAN SIVASUBRAMANIAN

Received November 12, 2004

A complete partition of a graph G is a partition of its vertex set in which any two distinct classes are connected by an edge. Let $\mathsf{cp}(G)$ denote the maximum number of classes in a complete partition of G. This measure was defined in 1969 by Gupta [19], and is known to be NP-hard to compute for several classes of graphs. We obtain essentially tight lower and upper bounds on the approximability of this problem. We show that there is a randomized polynomial-time algorithm that given a graph G with n vertices, produces a complete partition of size $\Omega(\mathsf{cp}(G)/\sqrt{\lg n})$. This algorithm can be derandomized.

We show that the upper bound is essentially tight: there is a constant $C>1$, such that if there is a randomized polynomial-time algorithm that for all large n, when given a graph G with n vertices produces a complete partition into at least $C\cdot\mathsf{cp}(G)/\sqrt{\lg n}$ classes, then $\mathsf{NP}\subseteq\mathsf{RTime}(n^{O(\lg\lg n)})$. The problem of finding a complete partition of a graph is thus the first natural problem whose approximation threshold has been determined to be of the form $\Theta((\lg n)^c)$ for some constant c strictly between 0 and 1.

1. Introduction

A *complete partition* of a graph $G=(V,E)$ is a partition of the vertex set V in which there is an edge connecting every pair of distinct classes. That is, a partition $V_1,\dots,V_t$ of V is complete if, for every i,j, $i\neq j$, there is an edge $\{v_i,v_j\}$ such that $v_i\in V_i$ and $v_j\in V_j$. In the complete partition problem, we wish to find a complete partition of the largest possible size, that is, a complete partition with the maximum number of classes. Let $\mathsf{cp}(G)$ denote the maximum size of a complete partition of G. This function is related to

Mathematics Subject Classification (2000): 68W25, 68W20, 68Q17, 05C70

* The work reported here is a merger of the results reported in [30] and [21].

several other graph-theoretic parameters, and in graph theory, it goes by the name *pseudoachromatic number*. It is known that computing $\mathsf{cp}(G)$ exactly is NP-hard. We survey these results and connections below.

1.1. Our contribution

We consider approximation algorithms for the complete partition problem and show matching upper and lower bounds.

Theorem 1.1 (Upper bound). *There is a randomized polynomial-time approximation algorithm that given a graph G on n vertices produces a complete partition of size $\Omega(\mathsf{cp}(G)/\sqrt{\lg n})$.*

The randomized algorithm promised in this theorem can be derandomized using the method of conditional expectations to get a deterministic polynomial-time algorithm with the same performance guarantee but slightly worse running time. For bounded-degree degree graphs, we present an efficient randomized algorithm, which given a graph with m edges and maximum degree Δ, constructs a complete partition of size $\frac{1}{4}\left\lfloor\sqrt{\frac{m}{\log \Delta}}\right\rfloor$, provided $\Delta^2 \lg \Delta = o(m)$. This immediately implies a $O(\sqrt{\lg \Delta})$-factor approximation algorithm for the complete partition problem, whenever $\Delta^2 \lg \Delta \leq m$. This algorithm was suggested by the referee.

In deriving our randomized algorithm, we work with an easily computable graph parameter closely related to $\mathsf{cp}(G)$. Let $\beta(G)$ be the largest integer t such that there exists a subgraph $H \subseteq G$ (H is not necessarily a vertex induced subgraph of G) with maximum degree at most t and at least $t^2/2$ edges. This number can be computed in polynomial time using an algorithm of Edmonds and Johnson [14] for computing the largest degree-constrained subgraph of a graph. It follows immediately from definition that $\beta(G)+1$ is an upper bound on $\mathsf{cp}(G)$. Interestingly, we show a lower bound on $\mathsf{cp}(G)$ in terms of $\beta(G)$.

Theorem 1.2. *There is a polynomial-time approximation algorithm that given a graph G produces a complete partition of size $\Omega(\beta(G)/\sqrt{\lg \beta(G)})$.*

Note that Theorem 1.1 follows immediately from this. By considering the random graph $G_{n,1/2}$, we show that this lower bound on $\mathsf{cp}(G)$ in terms of $\beta(G)$ is the best possible. Next, we show that Theorem 1.1 is essentially the best we can hope for.

Theorem 1.3. *There exists a constant $C \geq 1$ such that if there exists a randomized polynomial-time algorithm that with high probability computes* *a complete partition in an n-vertex graph G of size more than $C\mathsf{cp}(G)/\sqrt{\lg n}$, then* $\mathsf{NP} \subseteq \mathsf{RTime}(n^{O(\lg\lg n)})$.

This lower bound is obtained by adapting to our setting the arguments based on interactive proof systems previously used for the domatic number and the achromatic number problems [17,31].

1.2. Motivation and related work

The complete partition problem was introduced by Gupta [19] and further studied by Bhave [5], Sampathkumar and Bhave [36], Bollobás, Reed and Thomason [7], Kostochka [27], Yegnanarayanan [38], and Balasubramanian *et al.* [3]. These deal with bounds on special classes of graphs, random graphs and extremal problems; however, almost no general upper or lower bounds have previously been given for constructing a complete partition.

The complete partition problem is related to many well-studied combinatorial problems on graphs, in particular, to the achromatic number problem, where we are required to properly color the vertices of the graph with the maximum number of colors, so that there is an edge between every pair of color classes. See, [12] for a detailed survey. In this work, given a graph G, we denote its achromatic number by $\mathsf{achr}(G)$. The achromatic number problem is thus a variant of the complete partition problem, where we additionally demand that the classes in the partition be independent sets. The current bounds known on the approximation threshold for the achromatic number leave a huge gap. It can be approximated to within a factor $O(n\cdot\lg\lg n/\lg n)$ in general graphs [29], to within a factor of $O(n^{1/3})$ in graphs of girth 5 [29] and to within a factor $O(n^{4/5})$ in bipartite graphs [31]. On the other hand, no approximation algorithm can approximate it to within a factor $\lg^{1/4-\epsilon} n$ (for any $\epsilon > 0$), unless $\mathsf{NP} \subseteq \mathsf{RTime}(n^{\mathbf{poly}\log n})$ [31]; this inapproximability result applies to bipartite graphs.

Complete partitions lie under the surface of several graph theoretic measures. An immediate upper bound on complete partitions follows from the fact that there must be an edge between each pair of classes, or $\mathsf{cp}(G) \leq q(|E(G)|)$, where $q(m) \approx \sqrt{2m}$ is the largest value t such that $\binom{t}{2} \leq m$. Partitions where this bound is tight are vertex colorings where each pair of colors appears on exactly one edge; they are said to be *complete* and *harmonious*. Edwards and McDiarmid [15] showed that it is NP-hard to determine if a tree has a complete harmonious coloring, and Bodlaender [4] showed that it is also NP-hard for graphs that are simultaneously cographs and interval graphs. This implies the NP-hardness for the complete partition problem on these graph classes. Hedetniemi [25] conjectured that the achromatic and pseudoachromatic number are always equal for a tree, but this was disproved by Edwards [13]. On the other hand, Cairnie and Edwards [9] gave an algorithm for forests of bounded degree that finds a complete coloring with at least $q(m)-1$ colors, thus approximating $\mathsf{cp}(G)$ within additive term of 1. They [8] also showed that the achromatic number, and thus also the pseudoachromatic number, of bounded-degree graphs is $(1-o(1))q(|E(G)|)$.

Other complete partitioning problems have been studied. Bollobás, Catlin and Erdős [6] define the *contraction clique number* $\mathsf{ccl}(G)$ as the maximum number of classes in a complete partition where each class induces a *connected* subgraph of a graph G. Equivalently, $\mathsf{ccl}(G)$ is the largest clique minor of G. This is also known as the *Hadwiger number* of a graph, in relation to a much studied conjecture that $\mathsf{ccl}(G) \geq \chi(G)$. The conjecture is equivalent to the Four Color Theorem for $\mathsf{ccl}(G)=4$ and was proven for $\mathsf{ccl}(G) = 5$ by Robertson, Seymour and Thomas [35], but is open for larger values. Clearly, $\mathsf{ccl}(G) \leq \mathsf{cp}(G)$ for any graph G. Tight results are known for random graphs. Bollobás *et al.* [6] showed that for almost all graphs G, $\mathsf{ccl}(G) = n/(\sqrt{\lg n} + \theta(1))$ and McDiarmid [32] showed the same type of bound for achromatic number. Both of these imply that for almost all graphs, $\mathsf{cp}(G) \geq \lfloor n/(\sqrt{\lg n} + \theta(1))\rfloor$.

1.3. Our techniques

We now describe briefly how the parameter $\beta(G)$ described above helps us find a good complete partition. Recall, that the input graph G has large $\beta(G)$. So, there is a subgraph with all degrees bounded by $t=\beta(G)$ and with $\frac{t^2}{2}$ edges. The algorithm of Edmonds and Johnson [14], helps us find such a subgraph. Now, we want to pick $t' \approx t/\sqrt{\lg t}$ disjoint sets $X_1, X_2, \ldots, X_{t'}$, such that $X_i \sim X_j$. The natural idea would be to pick a random partition into t' classes (of size n/t' each) and hope that with constant probability, all of them would be connected. Unfortunately, this does not work, for instance, for the graph consisting one K_t and $n-t$ isolated vertices. Then, a fraction $t\exp(-\sqrt{\lg t})$ of the X_i's will consist of only isolated vertices. Note, however, that the remaining X_i's will be completely connected, and they are a constant fraction of all X_i's. In general, however, the situation is complicated. We need to use a two step process, where we first partition the vertex set into sets of size n/t, and then after identifying the good candidates among them, merge approximately $\sqrt{\lg t}$ of them. We present the argument in detail in Section 2. In the next subsection, we show

how to derandomize our algorithm. We then prove that there exists graphs G such that $\beta(G) = \Omega(\mathsf{cp}(G)\sqrt{\lg n})$.

Our lower bound proof uses some of the ideas from earlier inapproximability papers on covering problems, for example, Lund and Yannakakis [26], Feige [16], Feige, Halldórsson, Kortsarz and Srinivasan [17], but borrows most heavily from Kortsarz and Shende [31], who showed that no randomized algorithm can approximate the achromatic number of a graph to a factor much less that $(\lg n)^{1/4}$ unless $\mathsf{NP} \subseteq \mathsf{RTime}(n^{O(\mathrm{poly}(\lg n))})$. Following their approach, we show how to reduce an NP-complete language to a random instance of the complete partition problem. The distribution of the random instances is such that when the original input is in the language, then with probability one the random graph has a large complete partition, but when the original input is not in the language, with high probability the random graph has no large complete partition. Our argument can be adapted to improve the original inapproximability result in [31]: we can now show that no randomized algorithm can approximate the achromatic number within a factor much less than $\sqrt{\lg n}$, unless $\mathsf{NP} \not\subseteq \mathsf{RTime}(n^{O(\lg\lg n)})$. In Section 3, we present our inapproximability result. It is essentially self-contained, assuming a version of the PCP theorem from [17] and Raz's parallel repetition theorem [34]. In Section 3.4, we present the proof of the improved inapproximability result for the achromatic number problem.

2. Approximation algorithm for complete partition

In this section, we present an $O(\sqrt{\lg n})$-factor approximation algorithm for obtaining a complete partition of a graph. Notation: when we say that $X \sim Y$ for $X, Y \subseteq V(G)$, we mean that there is an edge in G connecting some vertex of X and some vertex of Y.

2.1. Dense degree-bounded subgraphs and t-expanding sets

Let us consider first the case of regular graphs, in order to develop some intuition. Observe that $\sqrt{m/2}$ is an upper bound on $\mathsf{cp}(G)$, since there need to be at least $\binom{\mathsf{cp}(G)}{2}$ edges to separate all the classes. It gives good results to simply partition the vertex set at random. Suppose we partition the set into $\sqrt{m/2}$ classes. The expected number of edges separating a given pair of classes is then around 1. Merging these classes together so that each new set contains around $\sqrt{3\log n}$ classes, means that the expected number of edges between any pair of sets becomes approximately $3\log n$. This can be turned into a high probability statement, so that we can conclude using the union bound that the partition is almost always good. For technical reasons, it may require special treatment for regular graphs of high degree, but in all cases a random partition suffices.

For non-regular graphs, additional ideas are needed. To begin with, the bound $\mathsf{cp}(G) \le \sqrt{m/2}$ is not useful in general; a case in point being the star graph $K_{1,\ell}$, that contains ℓ edges but $\mathsf{cp}(G) = 2$. Instead, we need a new, polynomially computable upper bound measure, which we introduce in the following paragraph. Also, disparate degrees can wreak havoc with a naive partition. Instead, we develop another measure, that induced a large bipartite subgraph, and show how to partition it. The actual partitioning is done in two rounds, one for each sides of the bipartite subgraph.

Definition 2.1. For a graph G, let $\beta(G)$ be the largest integer t such that there exists a subgraph H of maximum degree at most t and with at least $\frac{t^2}{2}$ edges. That is, H satisfies $\forall v \in V(H)$, $\deg_H(v) \le t$ and $|E(H)| \ge \frac{t^2}{2}$.

An algorithm of Edmonds and Johnson [14] for computing the largest degree-constrained subgraph of G can be used to determine $\beta(G)$ efficiently. The most efficient algorithm known for this problem, due to Gabow [18], runs in time $O(m^{\frac{3}{2}})$, where $m = |E(G)|$. Clearly, we need at most n calls to this algorithm to find $\beta(G)$, where $n = |V(G)|$. Now, consider a complete partition $S_1, S_2, \ldots, S_t$ of G. For each pair $\{S_i, S_j\}$, $(i \neq j)$, retain exactly one edge connecting them, and delete all other edges from G. Clearly, the resulting graph has maximum degree at most $t-1$ and at least $\binom{t}{2}$ edges. Thus, $\beta(G) \ge t-1$.

Proposition 2.2. $\mathsf{cp}(G) \le \beta(G) + 1$.

In the next subsection, we will show that $\mathsf{cp}(G) = \Omega\left(\beta(G)/\sqrt{\lg \beta(G)}\right)$. For this, we need to relate $\beta(G)$ to another parameter of G.

Definition 2.3 (Multi-expansion). Let G be a graph and let $\mathcal{C}$ be a collection of disjoint subsets of $V(G)$. We say that $\mathcal{C}$ is t-expanding if (i) $|\mathcal{C}| = t$, and (ii) $\forall S \in \mathcal{C}$, $|N(S) \setminus \bigcup_{T \in \mathcal{C}} T| \ge t$. The *multi-expansion number* of G, denoted by $\mathsf{me}(G)$, is the largest t such that there is a t-expanding collection of subsets of $V(G)$.

Our approach. We wish devise an algorithm that given a graph with large $\beta(G)$, produces a large complete partition. Instead of working directly with $\beta(G)$ it will be convenient for us to work with $\mathsf{me}(G)$. That is, we first show (Lemma 2.4) that if $\beta(G)$ is large, then $\mathsf{me}(G)$ is large; next, we show that if $\mathsf{me}(G)$ is large, then $\mathsf{cp}(G)$ is large (Theorem 2.5).

Lemma 2.4. $\mathsf{me}(G) \geq \frac{\beta(G)}{20}$

Proof. Let $b = \beta(G)$. Then, by the definition of $\beta(G)$, there is a subgraph H of G such that (a) $\forall v \in V(H)$ $\deg_H(v) \leq b$, and (b) $|E(H)| \geq b^2/2$. We can obtain a bipartite subgraph $H_0 = (U_0, V_0, E_0)$ of H with at least $b^2/4$ edges. We will show that there are at least $k \geq b/20$ disjoint subsets $S_1, S_2, \ldots, S_k$ of U_0 such that for $i = 1, 2, \ldots, k$, $|N_{H_0}(S_i)| \geq b/20$. Note that since $N(S_i) \subseteq V_0$, we automatically have $N_{H_0}(S_i) \cap \bigcup_j S_j = \emptyset$.

We construct the sets $S_1, S_2, \ldots, S_k$ sequentially. Suppose $S_1, S_2, \ldots, S_i$ have been constructed. To construct S_{i+1}, we consider the induced bipartite subgraph $H_i = (U_i, V_i, E_i)$ of H_0, where $U_i = U_0 \setminus (S_1 \cup S_2 \cup \cdots \cup S_i)$ and $V_i = V_0$. Now, starting from $S_{i+1} = \emptyset$, repeatedly add vertices from U_i to S_{i+1} until $|N(S_{i+1})| \geq b/20$. At each step, a new vertex v is chosen such that, at that point, at least half of its neighbors are outside $N(S_{i+1})$; that is, $|N(v) - N(S_{i+1})| > \frac{1}{2}\deg_{H_i}(v)$. If no such vertex is available, we stop, set $k = i$, and return the sets $S_1, S_2, \ldots, S_k$.

It only remains to show that $k \geq b/20$. We now examine the vertices in the graph H_0. When the process terminated, the set S_{k+1} was left incomplete. Let F' denote the set of edges of H_0 incident on $S_1 \cup S_2 \cup \cdots \cup S_k \cup N(S_{k+1})$ at this point. We claim that $|F'| \geq \frac{1}{2}|E_0|$. For otherwise, the set $F'' = E_0 \setminus F'$ has size more than $\frac{1}{2}|E_0|$. This implies that there is a vertex in U_0 that has more edges incident on it from F'' than it has from F'. Such a vertex qualifies to be added to S_{k+1}. But since the process terminated, there was no such vertex. We conclude that $|F'| \geq \frac{1}{2}|E_0| \geq \frac{b^2}{8}$. We can also obtain an upper bound for $|F'|$ by bounding the number of edges incident on $S_1, S_2, \ldots, S_k$ and $N(S_{k+1})$ separately.

First consider $E(S_i)$, the set of edges incident on S_i. By our rule for choosing the new vertex, each time a vertex is added, the number of edges added to $E(S_i)$ is at most twice the number of new neighbors added. So, at all times, we have $|E(S_i)| \leq 2|N(S_i)|$. At the moment just before the last vertex was added to S_i, we have $|N(S_i)| < b/20$, so that $|E(S_i)| < 2(b/20)$ at that point. Now, the last vertex has degree at most b. Thus, when the construction of S_i is complete (and we are ready to move on to S_{i+1}) $|E(S_i)| < b + 2(b/20) \leq (11/10)b$. Thus, the total number of edges incident on $S_1, S_2, \ldots, S_k$ is at most $k \times (11/10)b$. Furthermore, since S_{k+1} could not be completed, $|N(S_{k+1})| \leq b/20$, and since all degrees are at most b, the number of edges incident on $N(S_{k+1})$ can be at most $b \times (b/20)$.

Combining the upper and lower bounds for $|F'|$, we obtain

$$k \times (11/10)b + b \times (b/20) \geq |F'| \geq b^2/8.$$

Thus, $k \geq (3/44)b \geq (1/20)b$. ∎

2.2. The randomized approximation algorithm

In this section, we present a randomized polynomial-time algorithm that, given a graph G, produces a complete partition of G whose size is, with high probability, within a factor $O(\sqrt{\lg \mathsf{cp}(G)})$ of the optimum. This follows immediately from the following theorem.

Theorem 2.5. *There is a polynomial-time algorithm that takes as input a graph G on n (assumed to be large) vertices and a t-expanding family of sets $\{S_1, S_2, \ldots, S_t\}$ in G (with $t \geq 10$), and with probability at least $\frac{1}{4}$, delivers a complete partition of G of size at least $t/(4\sqrt{\lg t})$.*

Example. Suppose $G = K_{n,n}$ is the complete bipartite graph with each partition of size n. Color G with two colors, and let $\{A_i : i = 1, 2, \ldots, n\}$ be the collection of singleton sets in the first color class and let $\{B_j : j = 1, 2, \ldots, n\}$ be the collection of singleton sets in the second color class. Note that $\{A_i : i = 1, 2, \ldots, n\}$ is an n-expanding family in G. Furthermore every A_i has an edge to every B_j. So, we immediately obtain the complete partition $\{A_i \cup B_i : i = 1, 2, \ldots, n\}$.

In our algorithm we will try get into a situation similar to the above example, using the assumption that $\mathsf{me}(G)$ is large. Instead of the singleton sets A_i in the above example, we will have sets $\hat{S}_i$, which will be obtained by randomly merging some of the sets in the given t-expanding family $\{S_1, S_2, \ldots, S_t\}$; instead of the singleton sets B_j, we will have sets $\hat{T}_j$ obtained by randomly choosing vertices that belong to no S_i. The final complete partition will be $\{\hat{S}_i \cup \hat{T}_i : i = 1, 2, \ldots, n\}$.

We now describe the algorithm formally. Let $S = \bigcup_{i=1}^{t} S_i$, where $\{S_1, S_2, \ldots, S_t\}$ is the t-expanding family given to us. We know that for all i, $|N(S_i) \setminus S| \geq t$. Our algorithm has two steps.

Step I: We obtain disjoint sets $\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_{\hat{\ell}} \subseteq V \setminus S$, where $\hat{\ell} = \Omega(t/\sqrt{\lg t})$. Each $\hat{T}_j$ will have a neighbor in at least $(1-\epsilon)t$ of the S_i's. We will ensure that $\epsilon = \exp(-\Omega(\sqrt{\lg t}))$.

Step II: We randomly merge groups of $O(\sqrt{\lg t})$ S_i's into one, and get $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_{\hat{\ell}}$. This will ensure that with high probability $\hat{S}_i \sim \hat{T}_j$ for $i, j = 1, 2, \ldots, \hat{\ell}$.

The final solution will be obtained by merging $\hat{S}_i$ with $\hat{T}_i$ (for $i = 1, 2, \ldots, \hat{\ell}$) and putting any remaining vertices into these sets. We will need two lemmas (one for each step) to show that the algorithm outlined above succeeds with constant probability.

Let $\ell = \frac{t}{2\sqrt{\lg t}}$, and let $T_1, T_2, \ldots, T_\ell$ be a random partition of $V \setminus S$ into equal parts (we ignore divisibility issues). We say that T_i is *well-connected* if

$T_i \sim S_j$ for at least $(1-4\exp(-2\sqrt{\lg t}))t$ of the sets S_j. Let $\tilde{\ell}$ be the random variable denoting the number of well-connected T_i's.

Lemma 2.6. $\Pr[\tilde{\ell} \le \frac{\ell}{2}] \le \frac{1}{2}$.

Proof. For $i=1,2,\ldots,\ell$ and $j=1,2,\ldots,t$, let χ_{ij} be the indicator variable for the event $T_i \sim S_j$. We will use the following fact: if A is a subset of $[N]=\{1,2,\ldots,N\}$ of size a and B is a random subset of $[N]$ size b (each of the $\binom{N}{b}$ sets being equally likely), then

$$\Pr[A \cap B = \emptyset] = \frac{(N-a)(N-a-1)\cdots(N-a-b+1)}{N(N-1)\cdots(N-b+1)} \le \left(1-\frac{a}{N}\right)^b \le \exp\left(-\frac{ab}{N}\right).$$

Since S_j has at least t neighbors outside S and T_i is a random subset of $V \setminus S$ of size $|V \setminus S|/\ell = |V \setminus S| \cdot 2\sqrt{\lg t}/t$, we have

$$\Pr[\chi_{ij}=0] \le \left(1-\frac{t}{|V \setminus S|}\right)^{\frac{|V \setminus S|\cdot 2\sqrt{\lg t}}{t}} \le \exp\left(-2\sqrt{\lg t}\right).$$

In particular, the expected number of pairs (T_i, S_j) such that $T_i \not\sim S_j$ is at most $\exp(-2\sqrt{\lg t})\ell t$.

Let $p = \Pr[\tilde{\ell} \le \frac{\ell}{2}]$. Now, whenever T_i is not well-connected, $T_i \sim S_j$ holds for at most $(1-4\exp(-2\sqrt{\lg t}))t$ of the sets S_j, that is, $T_i \not\sim S_j$ for at least $4\exp(-2\sqrt{\lg t})t$ values of j. It follows from the definition of p, that the expected number of pairs (T_i, S_j) such that $T_i \not\sim S_j$ is at least $p \cdot \frac{\ell}{2} \cdot 4\exp(-2\sqrt{\lg t})t$. Thus, we have

$$2p\exp\left(-2\sqrt{\lg t}\right)\ell t \le \exp\left(-2\sqrt{\lg t}\right)\ell t;$$

that is $p \le \frac{1}{2}$. ∎

Assume we have succeeded in Step I, and there are at least $\hat{\ell} = \frac{\ell}{2} = \frac{t}{4\sqrt{\lg t}}$ well-connected sets among the T_i's; of these, we retain exactly $\hat{\ell}$ and refer to them as $\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_{\hat{\ell}}$.

We are now ready to implement Step II. Randomly partition the S_i's into $\hat{\ell}$ blocks each with $t/\hat{\ell}$ sets, and merge the sets in each block to obtain $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_{\hat{\ell}}$.

Lemma 2.7. *With probability at least $\frac{1}{2}$, $\hat{T}_i \sim \hat{S}_j$ for all $i,j=1,\ldots,\hat{\ell}$.*

Proof. Each $\hat{T}_i$ has a neighbor in at least $(1-4\exp(-2\sqrt{\lg t}))t$ of the S_j. So,

$$\Pr[\hat{T}_i \not\sim \hat{S}_j] \leq \left(4\exp(-2\sqrt{\lg t})\right)^{t/\hat{\ell}} \leq \left(\exp(-2\sqrt{\lg t}+2)\right)^{t/\hat{\ell}}.$$

Now, $\frac{t}{\hat{\ell}} = 4\sqrt{\lg t}$. So, $\Pr[\hat{T}_i \not\sim \hat{S}_j] \leq \exp(-8\lg t + 8\sqrt{\lg t}) \leq \frac{1}{t^2}$. The lemma follows from the union bound because the number of (i,j) pairs, $\hat{\ell}^2 \ll t^2$. ∎

Proof of Theorem 2.5. By Lemmas 2.6 and 2.7, with probability at least $\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$, we obtain sets $\hat{S}_1,\hat{S}_2,\ldots,\hat{S}_{\hat{\ell}}$ and $\hat{T}_1,\hat{T}_2,\ldots,\hat{T}_{\hat{\ell}}$, such that $\hat{S}_i \sim \hat{T}_j$. For $i=1,2,\ldots,\hat{\ell}$, let $\hat{V}_i = \hat{S}_i \cup \hat{T}_i$. Clearly, $\hat{V}_1,\hat{V}_2,\ldots,\hat{V}_{\hat{\ell}}$ is a complete partition of G. ∎

Proof of Theorem 1.2. Note that one can, in polynomial-time, compute $\beta(G)$ and obtain the required dense degree-bounded subgraph. Given such a subgraph, by Lemma 2.4, we can, in polynomial-time, obtain a multi-expander of size $\beta(G)/20$. Finally, using Theorem 2.5 we obtain a complete partition of size $\Omega(\beta(G)/\sqrt{\lg\beta(G)})$. ∎

2.3. Derandomization

Both steps in the randomized algorithm can be derandomized using the method of conditional expectation. We present the details only for Step I; for Step II, which is similar, we just sketch the main ideas. In Step I, we randomly partition the set $V\setminus S$ into ℓ equal parts and argue that with high probability many of the sets in the partition will be well-connected. We now show how we can produce many well-connected sets deterministically.

Let $T_1,T_2,\ldots,T_\ell$ be a random partition of $V\setminus S$ into sets of (almost) equal sizes. Recall that the indicator random variable χ_{ij} (used in the proof of Lemma 2.6) takes the value 1 if and only if there is an edge between T_i and S_j. Let $\chi=\sum_{ij}\chi_{ij}$. We have argued that

$$\mathrm{E}[\chi] \geq \left(1-\exp(-2\sqrt{\lg t})\right)\ell t.$$

We now wish to instantiate the sets T_i in such a way that the value of χ is at least its expected value. Once this is done, we immediately see that at least $\frac{3}{4}\ell$ of all T_j's will be well-connected. This instantiation will be done incrementally. At each stage, we will pick a new vertex in $V\setminus S$ and decide which T_i it must go into.

We will use $P_1,P_2,\ldots,P_\ell$ to describe the choices made so far, where P_i consists of those vertices of $V\setminus S$ that will belong to T_i in our instantiation.

Initially, we set $P_1, P_2, \ldots, P_\ell = \emptyset$, and add the vertices of $V \setminus S$ one after another to these sets ensuring that $|P_i| \le |V \setminus S|/\ell$. Let "$P \subseteq T$" denote the event "$P_i \subseteq T_i$ for $i = 1, 2, \ldots, \ell$." Then, initially,

$$\mathrm{E}[\chi \mid P \subseteq T] = \mathrm{E}[\chi].$$

We will ensure that at all times,

$$\mathrm{E}[\chi \mid P \subseteq T] \ge \mathrm{E}[\chi].$$

Suppose we need to insert $v \in V \setminus S$ into one of the P_i's (after having made the choice for the previous vertices). We try each choice of P_i (provided it is not already of the required size), insert v into P_i, compute the conditional expectation $\mathrm{E}[\chi \mid P \subseteq T]$ in each case, and retain the choice that gives the highest conditional expectation. It is straightforward to verify that this method works correctly. We only need to describe how the conditional expectations are evaluated. By linearity of expectation, it is enough to show how $\mathrm{E}[\chi_{ij} \mid P \subseteq T]$ can be computed efficiently. Suppose, S_j and P_i are connected by an edge then we have $\mathrm{E}[\chi_{ij} \mid P \subseteq T] = 1$. Otherwise, let p be the number of vertices of $V \setminus S$, that are still not in any P_i, and q be the number of neighbors of S_j among these p vertices. Then,

$$\mathrm{E}[\chi_{ij} \mid P \subseteq T] = 1 - \frac{\binom{p-q}{k-|P_i|}}{\binom{p}{k-|P_i|}},$$

where $k = |V \setminus S|/\ell$ (the desired size of the sets in the final partition).

The argument for Step II is similar. Suppose we have identified the well-connected sets $\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_{\hat{\ell}}$. In the randomized algorithm, we randomly partition the family $\{S_i : i = 1, 2, \ldots, t\}$ into $\hat{\ell}$ equal parts and argue that with high probability each $\hat{T}_j$ will have a neighbor in some S_j of each block of the partition. In our deterministic algorithm, we start with $\hat{\ell}$ empty blocks and add S_j's one after another to the blocks. Let η_{ij} be the indicator random variable for the event "some neighbor of $\hat{T}_i$ is in some set in block j" (here $i, j = 1, 2, \ldots, \hat{\ell}$) when the blocks are chosen randomly and of equal size; let $\eta = \sum_{ij} \eta_{ij}$. The calculation in Lemma 2.7 shows that $\mathrm{E}[\eta] > \hat{\ell}^2 - 1$. As in the derandomization of Step I, we can build the blocks by adding one set S_i at a time, and ensure that in the end the conditional expectation of η is more than $\hat{\ell}^2 - 1$. But at this point η is an integer with value at most $\hat{\ell}^2$. So, its value must be $\hat{\ell}^2$, implying that each $\hat{T}_i$ has a neighbor in some set of each block.

2.4. The gap between $\beta(G)$ and $\mathsf{cp}(G)$

In this section we will show that the relationship between $\beta(G)$ and $\mathsf{cp}(G)$ is essentially optimal.

Theorem 2.8. *For all large n, for almost all graphs G on n vertices $\beta(G)=\Omega(n)$ and $\mathsf{cp}(G)=O(\frac{n}{\sqrt{\lg n}})$.*

Proof. Consider the random graph G on n vertices obtained by placing each edge independently with probability $\frac{1}{2}$. Using Chernoff bounds, it is easy to verify that with probability $1-o(1)$ $|E(G)|\geq\frac{1}{2}\binom{n}{2}-n\lg n$ and, for all $v\in V(G)$, we have $\deg(v)\leq\frac{n-1}{2}+2\sqrt{n\lg n}$. For large n, $(\frac{n-1}{2}+2\sqrt{n\lg n})^2/2\leq\frac{1}{2}\binom{n}{2}-n\lg n$, implying that with probability at least $1-o(1)$, $\beta(G)\geq\frac{n-1}{2}+2\sqrt{n\lg n}$.

We will show that with probability at least $1-o(1)$, G has no complete partition of size $t=\frac{2n}{\sqrt{\lg n}}+1$. Thus, with probability $1-o(1)$, the events "$\beta(G)\geq\frac{n}{2}$" and "$\mathsf{cp}(G)<t$" hold simultaneously. Our theorem follows from this.

It remains to bound the probability that $\mathsf{cp}(G)\geq t$. Fix a partition $\mathcal{S}=(S_1,S_2,\ldots,S_t)$ of $V(G)$; let $s_i=|S_i|$.

$$\begin{aligned}
\Pr[\mathcal{S}\text{ is a complete partition}] &= \prod_{1\leq i<j\leq t}(1-2^{-s_is_j})\\
&\leq \prod_{1\leq i<j\leq t}\exp(-2^{-s_is_j})\\
&= \exp\left(-\sum_{1\leq i<j\leq t}2^{-s_is_j}\right)\\
&\leq \exp\left(-\binom{t}{2}2^{-\sum_{1\leq i<j\leq t}s_is_j\binom{t}{2}^{-1}}\right)\\
&\qquad\text{(because Geometric Mean} \leq \text{Arithmetic Mean)}\\
&\leq \exp\left(-\binom{t}{2}2^{-\frac{1}{2}(\sum_i^t s_i)^2\binom{t}{2}^{-1}}\right)\\
&\leq \exp\left(-\binom{t}{2}2^{-\frac{1}{2}n^2\binom{t}{2}^{-1}}\right).
\end{aligned}$$

For our choice of t, the last expression is at most $\exp(-n^{1.5})$. Since there are at most n^t possible choices for $\mathcal{S}$,

$$\begin{aligned}
\Pr[G\text{ has a complete partition of size }t] &\leq \sum_{\mathcal{S}}\Pr[\mathcal{S}\text{ is a complete partition of }G]\\
&\leq n^t\times\exp(-n^{1.5}).
\end{aligned}$$

∎

2.5. Bounded degree graphs

The original version of this paper gave an $O(\sqrt{\Delta})$ factor approximation algorithm for graphs with degree bounded by Δ. The following substantially stronger result is due to the referee.

Theorem 2.9. *Let G be a graph with m edges and maximum degree Δ such that $\sqrt{m} > 2000\Delta\sqrt{\log \Delta}$. Then, G has a complete partition with at least $\frac{1}{4}\left\lfloor\sqrt{\frac{m}{\log \Delta}}\right\rfloor$ parts. A randomized polynomial-time algorithm can produce such a partition with high probability.*

Since, $\mathsf{cp}(G) = O\left(\sqrt{|E(G)|}\right)$, we have the following corollary.

Corollary 2.10. *Let G be a graph with m edges and maximum degree Δ such that $m = \omega(\Delta^2 \lg \Delta)$. Then, there is a randomized polynomial-time algorithm that produces a complete partition whose size is smaller than the optimum by a factor at most $O(\sqrt{\log \Delta})$.*

We are now ready to prove Theorem 2.9.

Proof. The overall plan is as follows. We will color the vertices of the graph randomly with $\ell = \theta\left(\sqrt{\frac{m}{\log \Delta}}\right)$ colors, in the hope of finding at least $\frac{\ell}{4}$ color classes that are completely connected. Unfortunately, our argument will only ensure that all but some $\frac{m}{4\Delta}$ of the pairs of color classes are automatically connected. To take care of the missing connections, we will first set aside a matching of about $\frac{m}{4\Delta}$ edges and color only the rest of the graph. These matching edges will then be used to connect the pairs that remain unconnected after the random coloring.

To construct the matching M, greedily pick $\left\lceil\frac{m}{4\Delta}\right\rceil$ non-adjacent edges from the graph G. Set these edges aside for later use. Let G_1 be the graph obtained from G by deleting all vertices belonging to some edge in M. Each edge of M is adjacent to at most $2\Delta - 2$ other edges, so the number of edges that are removed from G is at most

$$\left(\frac{m}{4\Delta} + 1\right)(2\Delta - 1) \le \frac{m}{2} + 2\Delta - 1 - \frac{m}{2\Delta} - 1 \le \frac{m}{2}.$$

Thus, G_1 has at least $\frac{m}{2}$ edges. We are now ready to color the vertices of G_1, but it will be convenient to do this coloring in two phases. First, partition the vertex set into sets V_1 and V_2 such that the induced bipartite subgraph $H = (V_1, V_2, E)$ has at least $\frac{m}{4}$ edges. Let $k = 2\sqrt{\lg \Delta}$ and let $\ell = \left\lfloor\frac{\sqrt{m}}{k}\right\rfloor$. Choose colors for the vertices of V_1 from the set $[\ell]$, uniformly and independently.

To identify the color classes that are likely to get connected easily when we color V_2, we will examine the neighborhoods of the color classes in V_1. The probability that a vertex $w \in V_2$ has a neighbor (in V_1) colored $r \in [\ell]$ is exactly

$$1-\left(1-\frac{1}{\ell}\right)^{d_w} \geq 1-\left(1-\frac{k}{\sqrt{m}}\right)^{d_w} \geq 1-\exp\left(-\frac{kd_w}{\sqrt{m}}\right) \geq \frac{kd_w}{2\sqrt{m}}.$$

Here d_w is the degree of w in the graph H; for the last inequality we used our assumption that $\sqrt{m} \geq 2000\Delta\sqrt{\lg \Delta}$. For $r \in [\ell]$, let N_r be the neighborhood of the color class r, that is, the set of vertices in V_2 that have a neighbor colored r. Thus,

$$\text{(1)} \qquad \mathrm{E}\left[|N_r|\right] \geq \sum_{w \in V_2} \frac{kd_w}{2\sqrt{m}} \geq \frac{k}{2\sqrt{m}} \times \frac{m}{4} \geq \frac{k\sqrt{m}}{8}.$$

We wish to show that $|N_r|$ is usually not much smaller than its expectation. For this, we bound the variance of $|N_r|$. For $v \in V_1$, let ψ_v be the indicator variable for the event "v is colored r." In particular, we have $\Pr[\psi_v = 1] = \mathrm{E}[\psi_i] = \frac{1}{\ell}$. Then, it can be shown that

$$\text{(2)} \quad \mathrm{var}\left[|N_r|\right] \leq \sum_{v \in V_1} d_v^2\, \mathrm{E}[\psi_v] \leq \frac{1}{\ell} \sum_{v \in V_1} d_v^2 \leq \frac{\Delta}{\ell} \sum_{v \in V_1} d_v \leq \frac{\Delta m}{\ell} \leq 1.1\Delta k\sqrt{m}.$$

From (1) and (2), using Chebyshev's inequality, we can conclude that

$$\Pr\left[|N_r| < \frac{k\sqrt{m}}{16}\right] < \frac{\mathrm{var}[|N_r|]}{(\mathrm{E}[|N_r|]/2)^2} < \frac{1}{2}.$$

(We used our assumption that $\sqrt{m} \geq 2000\Delta\sqrt{\lg \Delta}$.) Thus, the expected number of r with $|N_r| \geq \frac{k\sqrt{m}}{16}$ is more than $\frac{\ell}{2}$, and by Markov's inequality, with probability at least $\frac{1}{3}$, the number of such color classes r is more than $\frac{\ell}{4}$. Fix a set of $\ell' = \lceil \frac{\ell}{4} \rceil$ good color classes.

Now, color the vertices of V_2 at random with good colors, chosen uniformly and independently. Fix a pair of good colors (r_1, r_2). We want to estimate the probability that some edge of H connects the color classes r_1 and r_2. The probability that no vertex in N_{r_1} receives the color r_2 is at most

$$\left(1-\frac{1}{\ell'}\right)^{|N_{r_1}|} \leq \left(1-\frac{1}{\ell'}\right)^{\frac{k\sqrt{m}}{16}} \leq \exp\left(-\frac{k\sqrt{m}}{\ell'}\right) \leq \exp\left(-\frac{k^2}{4}\right).$$

The expected number of pairs of good color classes that have no edge connecting them is at most

$$\exp\left(-\frac{k^2}{4}\right)\binom{\ell'}{2} \leq \frac{m}{8\Delta},$$

and the probability that more than $\frac{m}{4\Delta}$ edges are missing is at most $\frac{1}{2}$. Finally, connect these pairs using the edges of the matching M that were set aside in the beginning. This leads to a complete partition with $\frac{\ell}{4} \geq \frac{1}{4}\left\lfloor\sqrt{\frac{m}{\lg \Delta}}\right\rfloor$ parts. ∎

3. Lower bound

The goal of this section is to prove the following.

Theorem 3.1. *Let $L \subseteq \{0,1\}^*$ be a language in* NP. *Then, there is a randomized transformation f that transforms inputs in $\{0,1\}^*$ to graphs, and a polynomial-time computable function $\alpha : \{0,1\}^* \to \mathbb{N}$ such that for all $x \in \{0,1\}^*$,*

(a) if $x \in L$ then $\mathsf{cp}(f(x)) \geq \alpha(x)$ with probability 1;
(b) if $x \notin L$, then with probability at least $\frac{3}{4}$, we have $\mathsf{cp}(f(x)) \leq C\alpha(x)/\sqrt{\lg|V(f(x))|}$, where $C \geq 1$ is an absolute constant independent of L and x.

Furthermore, $f(x)$ can be computed in time $|x|^{O(\lg\lg|x|)}$.

Corollary 3.2. *There exists a constant $C \geq 1$ such that if there exists a randomized polynomial-time algorithm that with high probability computes a complete partition in a graph G of size more than $\frac{C\mathsf{cp}(G)}{\sqrt{\lg|V(G)|}}$, then* $\mathsf{NP} \subseteq \mathsf{RTime}(n^{O(\lg\lg n)})$.

Plan. Our proof is inspired by the result of Feige, Halldórsson, Kortsarz and Srinivasan [17] on the inapproximability of the domatic number of a graph. In that work, they show how one can transform an input instance x of a language $L \in \mathsf{NP}$ into a bipartite graph $G = (V_1, V_2, E)$ with the following properties. If $x \in L$, then V_1 can be partitioned into a large number of disjoint dominating sets; however, if $x \notin L$, then every dominating set of G is large, so that no such large partition can exist. Our construction will closely follow this idea. We will ensure that when $x \in L$, V_1 can be partitioned into $|V_2|$ disjoint dominating sets. Now, a complete partition of size V_2 can easily be obtained from this by adding one vertex of V_2 to each dominating set. We also need to show that when $x \notin L$, the size of the largest complete partition is $O(|V_2|/\sqrt{\lg n})$, where n is the number of vertices in G. The result in [17] can be used directly to ensure that when $x \notin L$, then V_1 cannot be partitioned into many disjoint dominating sets. But this does not immediately imply that there is no complete partition of large size. To prove our result, we need to examine the structure of the graph more closely. In fact, our graph G will be constructed randomly. To show that in the case $x \notin L$ all complete partitions have size $O(|V_2|/\sqrt{\lg n})$, we will show that the probability of any fixed partition with a large number of classes being completely connected is very small, and get our required conclusion by summing over all such potential complete partitions.

3.1. Background

Let $\mathcal{I}_n$ be the set of 5-regular, simple graphs with vertex set $[n] = \{1, 2, \ldots, n\}$. Let $\mathcal{I} = \bigcup_{n \geq 1} \mathcal{I}_n$. We say that a graph $I \in \mathcal{I}$ is ϵ-non-3-colorable if in every 3-coloring of the vertices of I at most $(1-\epsilon)|E(I)|$ edges of I are properly colored. The following consequence of the PCP theorem [2] appears in Feige et al. [17].

Theorem 3.3. *There is an $\epsilon > 0$ such that the following holds. Let $L \subseteq \{0,1\}^*$ be a language in NP. Then, there is a polynomial computable function $f : \{0,1\}^* \to \mathcal{I}$ such that for all $x \in \{0,1\}^*$*

- *if $x \in L$, then $f(x)$ is 3-colorable;*
- *if $x \notin L$, then $f(x)$ is ϵ-non-3-colorable.*

Consider the following two-prover protocol. In this protocol, and later, we will refer to colorings of the form $\chi : e \to \{R, G, B\}$, where e is an edge of a graph. Here e is to be thought of as a set of two vertices, say $\{v, w\}$. Then, the color assigned to v is $\chi(v)$ and the color assigned to w is $\chi(w)$.

Input: $I \in \mathcal{I}$.
Verifier: Picks a $v \in V(I)$ at random. Picks a random pair of edges (e, e'), both incident on v. Sends e to the Prover I and e' to Prover II.
Prover I: Sends a proper 3-coloring $\chi : e \to \{\mathsf{R}, \mathsf{G}, \mathsf{B}\}$ to the verifier.
Prover II: Sends a proper 3-coloring $\chi' : e' \to \{\mathsf{R}, \mathsf{G}, \mathsf{B}\}$ to the verifier.
Verifier: Accepts iff $\chi(v) = \chi'(v)$.

Proposition 3.4. *If I is 3-colorable then there is a strategy for the provers that convinces the verifier with probability 1. If I is ϵ-non-3-colorable, then for every strategy of the provers,*

$$\Pr[\textit{Verifier accepts}] \leq 1 - \frac{\epsilon}{2}.$$

Theorem 3.5 (Raz's parallel repetition theorem). *If we perform ℓ parallel repetitions of a one-round two-prover protocol with probability of acceptance ϵ, and accept if and only if all ℓ runs lead to acceptance, then the probability of acceptance is at most $2^{-c\ell}$ where $c > 0$ is a constant that depends only on ϵ, and the length of the answers of the provers in the original proof system.*

Consider the following ℓ-fold parallel repetition of the protocol above.

Input: $I \in \mathcal{I}$.
Verifier: Picks $v_1, v_2, \ldots, v_\ell \in V(I)$ at random. For each $i = 1, 2, \ldots, \ell$, picks a random pair (e_i, e_i') of edges both incident on v_i. Sends $\langle e_1, e_2, \ldots, e_\ell \rangle$ to the first prover and $\langle e_1', e_2', \ldots, e_\ell' \rangle$ to the second prover.
Prover I: Sends $\langle \chi_1, \chi_2, \ldots, \chi_\ell \rangle$ to the verifier, where $\chi_i : e_i \to \{\mathsf{R}, \mathsf{G}, \mathsf{B}\}$ is a proper coloring of e_i.
Prover II: Sends $\langle \chi_1', \chi_2', \ldots, \chi_\ell' \rangle$ to the verifier, where $\chi_i' : e_i' \to \{\mathsf{R}, \mathsf{G}, \mathsf{B}\}$ is a proper coloring of e_i'.
Verifier: Accepts iff for all $i = 1, 2, \ldots, \ell$, $\chi_i(v_i) = \chi_i'(v_i)$.

Corollary 3.6. *Let the maximum number of properly colored edges in any 3-coloring of I be $(1-\epsilon)|E(I)|$ ($\epsilon > 0$). Then, there exists a constant $c = c(\epsilon) > 0$ such that the following holds for all ℓ. Let Π be a strategy for the provers in the above protocol that makes the verifier accept with the highest probability. Then, under Π,* $\Pr[\textit{Verifier accepts}] \leq 2^{-c\ell}$.

From now on, we will use c for the constant associated in the above corollary with the constant ϵ promised in Theorem 3.3; we may assume $c \leq 1$.

The task of determining the best strategy for the provers in the parallel protocol corresponds to solving a certain covering problem on a graph obtained from the input instance I. We will use the Min-Rep problem formulated in Kortsarz [28].

Definition 3.7 (Min-Rep-Graph [28]). Given a graph $I \in \mathcal{I}$, the bipartite graph $\mathsf{MRG}(I) = (U_1, U_2, E)$, is defined as follows. U_1 and U_2 consist of tuples

$$\langle e_1, e_2, \ldots, e_\ell, \chi_1, \chi_2, \ldots, \chi_\ell \rangle,$$

where for $i=1,2,\dots,\ell$, $e_i \in E(I)$ and $\chi_i : e_i \to \{\mathsf{R},\mathsf{G},\mathsf{B}\}$ is a proper coloring of e_i. The edges of $\mathsf{MRG}(I)$ will be labeled by elements of $V(I)^\ell$, and we will use $(u,u',\overline{v})$ to denote the edge between $u \in U_1$ and $u' \in U_2$ with label $\overline{v} \in V(I)^\ell$. Let $E_{\overline{v}}$ be the edges labeled $\overline{v}$.

$$E_{\overline{v}} = \{(\langle \overline{e},\overline{\chi}\rangle, \langle \overline{e'},\overline{\chi'}\rangle, \overline{v}) : \langle \overline{e},\overline{\chi}\rangle \in U_1 \text{ and } \langle \overline{e'},\overline{\chi'}\rangle \in U_2 \text{ and}$$
$$\text{for } i = 1,2,\dots,\ell,\ v_i \in e_i, e'_i \text{ and } \chi_i(v_i) = \chi'_i(v_i)\}.$$

Finally, the edge set of $\mathsf{MRG}(I)$ is $\bigcup_{\overline{v}\in V(I)^\ell} E_{\overline{v}}$.

Note: In the graph above, and in those we shall encounter below, there may be parallel edges connecting two distinct vertices $\langle \overline{e},\overline{\chi}\rangle$ and $\langle \overline{e'},\overline{\chi'}\rangle$; this happens when there is an i such that $e_i = e'_i$ and $\chi_i = \chi'_i$. However, different edges connecting the same two vertices have different labels.

The vertices of U_1 can be naturally partitioned into sets $(A_{\overline{e}} : \overline{e} \in E(I)^\ell)$, where $A_{\overline{e}}$ consists of vertices of the form $\langle \overline{e},\overline{\chi}\rangle$, all with the same first component. The elements of $A_{\overline{e}}$ correspond to all potential responses of Prover I in the above parallel protocol, when the Verifier sends $\overline{e}$. Similarly, U_2 can be partitioned into sets $(B_{\overline{e}} : \overline{e} \in E(I)^\ell)$. Since there are six ways to properly 3-color an edge, $|A_{\overline{e}}| = |B_{\overline{e}}| = 6^\ell$; since the instance I is a 5-regular graph, there are precisely $q = |E(I)|^\ell = (\frac{5}{2})^\ell |V(I)|^\ell$ parts in each partition. Let $\equiv$ denote the equivalence relation on $U_1 \cup U_2$ induced by the partition above.

Definition 3.8. Let G be a graph and $\equiv$ be an equivalence relation on $V(G)$ (or equivalently a partition of $V(G)$). Let $G/{\equiv}$ be the graph with vertex set $V/{\equiv}$ (i.e. the set of equivalence classes under $\equiv$) and $E(G/{\equiv}) = \{(x/{\equiv}, y/{\equiv}, \overline{v}) : (x,y,\overline{v}) \in E(G)\}$ where $x/{\equiv}$ is the equivalence class of x under $\equiv$. Note that $E(G/{\equiv})$ is a set, that is, there is at most one edge with a given label between two vertices.

Remark. We will refer to the graph $\mathsf{MRG}(I)/{\equiv}$ as the *super min-rep-graph* and refer to its vertices as *super vertices* and its edges as *super edges*. Also, we will write $\mathsf{MRG}_{\equiv}(I)$ instead of $\mathsf{MRG}(I)/{\equiv}$. Note that $\mathsf{MRG}_{\equiv}(I)$ is a regular graph with $2q$ super vertices, each with degree $2^\ell 5^\ell$; thus, the number of super edges in $\mathsf{MRG}_{\equiv}(I)$ is:

$$h = q \cdot 2^\ell 5^\ell = |V(I)|^\ell 5^{2\ell}.$$

Definition 3.9. Let $S \subseteq V(\mathsf{MRG}(I))$ and let G_S be the graph on $V(\mathsf{MRG}(I))$ consisting of those edges of $\mathsf{MRG}(I)$ that have both end points in S. Let $\equiv$ be the equivalence relation on $V(\mathsf{MRG}(I))$ defined above. We say that S covers a super edge $\tilde{e} \in \mathsf{MRG}_{\equiv}(I)$ if $\tilde{e} \in E(G_S/{\equiv})$. We say that S is a cover if it covers all super edges in $\mathsf{MRG}_{\equiv}(I)$.

Remark. Note that S is a cover if and only if $G_S/\equiv$ is the same as $\mathsf{MRG}_{\equiv}(I)$. For example, the entire vertex set $V(G(I))$ is a cover. Note that when S covers the super edge $\tilde{e}=(A_{\overline{e}},B_{\overline{e'}},\overline{v})$, it has a vertex $u_1\in A_{\overline{e}}$ and a vertex $u_2\in B_{\overline{e'}}$ such that $(u_1,u_2,\overline{v})\in E(\mathsf{MRG}(I))$. In particular, since it is not enough if S has only one end point of the edge, this notion of covering, referred to as min-rep-covering by Kortsarz and Shende [31], is different from the usual definition for vertex covers.

We are now ready to state the key properties of $\mathsf{MRG}(I)$.

Lemma 3.10 (Main Lemma).

- *If $I\in\mathcal{I}$ is 3-colorable, then $V(\mathsf{MRG}(I))$ can be partitioned into 6^ℓ disjoint covers each of size $2q$.*
- *If I is ϵ-non-3-colorable, then any set S of size at most $2qk$, covers at most a fraction $(32k^2 2^{-c\ell})^{1/3}$ of the super edges of $\mathsf{MRG}_{\equiv}(I)$.*

Proof. For the first part, assume that $\tau_1,\tau_2,\dots,\tau_\ell$ are ℓ proper 3-colorings of I (not necessarily distinct). Let

$$S\cap U_1=\{\langle\overline{e},\overline{\chi}\rangle:\ \text{for } i=1,2,\dots,\ell,\ \chi_i \text{ is the restriction of } \tau_i \text{ to } e_i\};$$
$$S\cap U_2=\{\langle\overline{e'},\overline{\chi'}\rangle:\text{for } i=1,2,\dots,\ell,\ \chi'_i \text{ is the restriction of } \tau_i \text{ to } e'_i\}.$$

It is easy to verify that S is a cover in $\mathsf{MRG}(I)$. Now, fix a 3-coloring χ of I. Then, there are 6 different variants of χ (we may assume that I has at least one edge, so χ cannot be the constant function). Letting $\tau_1,\tau_2,\dots,\tau_\ell$ each range over these 6 possibilities in all possible ways, we get 6^ℓ versions of the cover S. These versions are pairwise disjoint. Note that each cover has exactly $2q$ vertices.

For the second part, fix a set S of size $2qk$ and let α be the fraction of the super edges of $\mathsf{MRG}_{\equiv}(I)$ covered by S. We will use the two-prover protocol above. The verifier's actions correspond to picking a random super edge $(A_{\overline{e}},B_{\overline{e'}},\overline{v})$ of $\mathsf{MRG}_{\equiv}(I)$, and sending $\overline{e}$ to Prover I and $\overline{e'}$ to Prover II. Consider the following strategy for the provers: on receiving $\overline{e}$ from the verifier, Prover I picks a random vertex $\langle\overline{e},\overline{\chi}\rangle$ in $A_{\overline{e}}\cap S$ (such a vertex exists) and returns $\overline{\chi}$; similarly, on receiving $\overline{e'}$ from the verifier, Prover II picks a random vertex $\langle\overline{e'},\overline{\chi'}\rangle$ in $B_{\overline{e'}}\cap S$ and sends back $\overline{\chi'}$.

Since $\mathsf{MRG}_{\equiv}(I)$ is regular, the expected value (over a random choice of the super edge $(A_{\overline{e}},B_{\overline{e'}},\overline{v})$ of $\mathsf{MRG}_{\equiv}(I)$) of $|S\cap(A_{\overline{e}}\cup B_{\overline{e'}})|$ is $2qk/q=2k$. By Markov's inequality,

$$\Pr\left[|S\cap(A_{\overline{e}}\cup B_{\overline{e'}})|\geq\frac{4k}{\alpha}\right]\leq\frac{\alpha}{2}.$$

Thus, with probability $\alpha/2$ the following two events hold simultaneously: S covers the super edge $(A_{\overline{e}}, B_{\overline{e'}}, \overline{v})$ and $|S \cap (A_{\overline{e}} \cup B_{\overline{e'}})| < \frac{4k}{\alpha}$. If S covers the super edge $(A_{\overline{e}}, B_{\overline{e'}}, \overline{v})$, then there is an edge $(\langle \overline{e}, \overline{\chi} \rangle, \langle \overline{e'}, \overline{\chi'} \rangle, \overline{v})$ between them in $\mathsf{MRG}(I)$ with both end points in S. If $|S \cap (A_{\overline{e}} \cup B_{\overline{e'}})| < \frac{4k}{\alpha}$ the probability that both provers pick their corresponding end point in our randomized strategy is at least $(\frac{\alpha}{4k})^2$. Thus, the overall probability of the verifier accepting is at least $\alpha/2 \cdot (\alpha/4k)^2 = \alpha^3/32k^2$. Our claim now follows from Corollary 3.6. ∎

3.2. The construction

As stated earlier, our final graph $G(I)$ will be generated using the graph $\mathsf{MRG}(I)$. This transformation is similar to constructions used in earlier works (e.g. Lund and Yannakakis [26], Feige [16], and especially Feige, Halldórsson, Kortsarz and Srinivasan [17], and Kortsarz and Shende [31]). However, since we do not assume that the reader is familiar with the details in those papers, we shall present the construction in detail. We will first describe a random graph $\tilde{G}(I)$. The final graph $G(I)$ will be obtained by taking several independent copies of $\tilde{G}(I)$ and identifying some vertices from different copies.

The intermediate random graph $\tilde{G}(I)$. The graph $\tilde{G}(I)$ is a random bipartite graph with vertex sets $\tilde{V}_1$ and $\tilde{V}_2$. The first vertex set $\tilde{V}_1$ will, in fact, be the same as $V(\mathsf{MRG}(I))$, and as before we partition it into $2q$ clusters $(A_{\overline{e}}, B_{\overline{e}} : \overline{e} \in E(I)^\ell)$ each with 6^ℓ vertices. The other vertex set $\tilde{V}_2$ has one set $M_{\tilde{e}}$ of m vertices for each super edges $\tilde{e}$ of $\mathsf{MRG}_\equiv(I)$. Since there are exactly $|V(I)|^\ell 5^{2\ell}$ super edges in $\mathsf{MRG}_\equiv(I)$, $\tilde{V}_2$ has $|V(I)|^\ell 5^{2\ell} m$ vertices.

We now describe how the edges are placed in $\tilde{G}(I)$. The edges connecting $M_{\tilde{e}}$ to the vertices in $\tilde{V}_1$ are determined using a random process. This process is run independently for each edge $\tilde{e}$ of $\mathsf{MRG}(I)$. Consider a super edge $\tilde{e} = (A_{\overline{e}}, B_{\overline{e'}}, \overline{v})$ of $\mathsf{MRG}_\equiv(I)$. The vertices in $M_{\tilde{e}}$ have neighbors only in $A_{\overline{e}}$ and $B_{\overline{e'}}$. To describe how these neighbors are chosen, we partition $A_{\overline{e}}$ into 3^ℓ sets $(A_{\overline{e},\tau} : \tau \in \{\mathsf{R}, \mathsf{G}, \mathsf{B}\}^\ell)$ each of size 2^ℓ, where

$$A_{\overline{e},\tau} = \{\langle \overline{e}, \overline{\chi} \rangle : \chi_i(v_i) = \tau_i\}.$$

The idea is that $A_{\overline{e}}$ has 6^ℓ vertices, each corresponding to assigning a different coloring to the edges involved in $\overline{e}$. Some of these colorings agree on what they assign to the vertices in $\overline{v}$; these vertices are put into a single class. Similarly, we partition $B_{\overline{e'}}$ into 3^ℓ sets of size 2^ℓ each, where

$$B_{\overline{e'},\tau} = \{\langle \overline{e'}, \overline{\chi} \rangle : \chi_i(v_i) = \tau_i\}.$$

Note that in $\mathsf{MRG}(I)$ every vertex in $A_{\overline{e},\tau}$ is connected by an edge to every vertex of $B_{\overline{e'},\tau}$. A vertex $v_2 \in M_{\tilde{e}}$ is connected to either all the vertices in $A_{\overline{e},\tau}$ or all the vertices in $B_{\overline{e},\tau}$, but not to both, that is, if a vertex of $M_{\tilde{e}}$ has a neighbor in $A_{\overline{e},\tau}$, then it has no neighbors in $B_{\overline{e'},\tau}$. Under these conditions the vertex $v_2 \in M_{\tilde{e}}$ has exactly 2^{3^ℓ} possible neighborhoods, we choose one of them randomly. Alternatively, we may use the following randomized procedure to generate the edges connecting $M_{\tilde{e}}$ to $A_{\overline{e}}$ and $B_{\overline{e'}}$.

```
for each v_2 ∈ M_ẽ independently
for each τ ∈ {R, G, B}^ℓ independently
        Pick C ∈ {A_{ē,τ}, B_{ē',τ}} at random (each with probability 1/2);
        Connect v_2 to all vertices in C.
```

Remarks. Note that the appearance of edges in $\tilde{G}(I)$ depends on random choices. However, there are some pairs of vertices that never get connected. We call a pair $(u,v_2) \in \tilde{V}_1 \times \tilde{V}_2$ a *potential edge* if it appears as an edge in some instance of $\tilde{G}$, that is, the event "$(u,v_2) \in E(\tilde{G}(I))$" has non-zero probability; for a potential edge p, let J_p be the indicator random variable for the event $p \in \tilde{G}$. Two potential edges (u,v_2) and (u',v_2') are said to be *related* if the events "$(u,v_2) \in E(\tilde{G})$" and "$(u',v_2') \in E(\tilde{G})$" are dependent; this happens if and only if

- $v_2 = v_2'$ (say $v_2 \in M_{\tilde{e}}$ where $\tilde{e} = (A_{\overline{e}}, B_{\overline{e'}}, \overline{v})$) and
- both $u, u' \in A_{\overline{e},\tau} \cup B_{\overline{e'},\tau}$ for some $\tau \in \{\mathsf{R},\mathsf{G},\mathsf{B}\}^\ell$.

Thus, the set of potential edges related to an $(u,v_2) \in (A_{\overline{e},\tau} \cup B_{\overline{e'},\tau}) \times M_{\tilde{e}}$, where $\tilde{e} = (A_{\overline{e}}, B_{\overline{e'}}, \overline{v})$, is precisely $(A_{\overline{e},\tau} \cup B_{\overline{e'},\tau}) \times \{v_2\}$. Since, $|A_{\overline{e},\tau}|, |B_{\overline{e'},\tau}| = 2^\ell$, (u,v_2) is related to precisely $2^{\ell+1}$ potential edges (including itself). Furthermore, we have the following proposition. [Parts (a) and (b) follow immediately from our construction, part (c) follows from part (b), and part (d) follows from part (c) using Lemma 3.10.]

Proposition 3.11. *(a) Suppose $\mathcal{E}_1$ and $\mathcal{E}_2$ are events and S_1 and S_2 are subsets of potential edges, such that $\mathcal{E}_1$ is fully determined by $(J_p : p \in S_1)$ and $\mathcal{E}_2$ is fully determined by $(J_p : p \in S_2)$. Suppose no potential edge in S_1 is related to a potential edge in S_2. Then, $\mathcal{E}_1$ and $\mathcal{E}_2$ are independent.*

(b) Suppose $(u,u',\overline{v}) \in E(\mathsf{MRG}(I))$. Then, $M_{(u/\equiv, u'/\equiv, \overline{v})} \subseteq N(u) \cup N(u')$.

(c) If S is a cover in $\mathsf{MRG}(I)$, then $N(S) = \tilde{V}_2$ in $\tilde{G}(I)$.

(d) If I is 3-colorable, then $\tilde{V}_1$ can be partitioned into 6^ℓ disjoint sets $S_1, S_2, \ldots$ such that $N(S_i) = \tilde{V}_2$.

The final random graph $G(I)$. The final random graph $G(I)$ is obtained as follows. First, pick r independent copies of $\tilde{G}(I)$; call them $\tilde{G}^1, \tilde{G}^2, \ldots, \tilde{G}^r$, with corresponding vertex sets $(\tilde{V}_1^1, \tilde{V}_2^1), (\tilde{V}_1^2, \tilde{V}_2^2), \ldots, (\tilde{V}_1^r, \tilde{V}_2^r)$; let

$$r = \frac{|\tilde{V}_2|}{6^\ell}.$$

[Our choice of parameters below ensures that r is an integer.] The graph $G(I)$ is obtained by identifying the corresponding vertices in $\tilde{V}_2^1, \tilde{V}_2^2, \ldots, \tilde{V}_2^r$. E.g., for each super edge $\tilde{e}$ of $\mathsf{MRG}_{\equiv}(I)$, there is an instance of $M_{\tilde{e}}$, say $M_{\tilde{e}}^j \subseteq \tilde{V}_2^j$. These r instances of $M_{\tilde{e}}$ are identified to produce just one set $M_{\tilde{e}}$ of size m. Thus, $G(I)$ is a bipartite graph with vertex sets (V_1, V_2) where

$$V_1 = \tilde{V}_1^1 \cup \tilde{V}_1^2 \cup \cdots \cup \tilde{V}_1^r,$$

has $2rq6^\ell = 2q|V_2|$ vertices and V_2 has $|V(I)|^\ell 5^{2\ell} m$ vertices. Note that each cluster of vertices of the form $A_{\overline{e}}$ has r versions in V_1; we refer to them as $A_{\overline{e}}^1, A_{\overline{e}}^2, \ldots, A_{\overline{e}}^r$. Similarly, we have $B_{\overline{e}}^1, B_{\overline{e}}^2, \ldots, B_{\overline{e}}^r$.

We now examine how $\mathsf{cp}(G(I))$ behaves for 3-colorable and ϵ-non-3-colorable instances I.

When I is 3-colorable. Note, that Part (a) of Proposition 3.11 applies to $G(I)$ as well. Also, since $r6^\ell \geq |V_2|$, Part (d) implies that if I is 3-colorable, there are $|V_2|$ disjoint sets $S_1, S_2, \ldots \subseteq V_1$, such that $N(S_i) = V_2$. By adding a different vertex of V_2 to each of these sets, we obtain a complete partition of $G(I)$ of size $|V_2(G(I))|$.

Proposition 3.12. *If I is 3-colorable,* $\mathsf{cp}(G(I)) \geq |V_2| = |V(I)|^\ell 5^{2\ell} m.$

When I is ϵ-non-3-colorable. We wish to show that if I is ϵ-non-3-colorable, then with high probability $\mathsf{cp}(G(I)) \leq D|V_2|/\sqrt{\lg|V_2|}$, for a suitable choice of D, ℓ and m. In the following, we let

$$t = \left\lceil \frac{D|V_2|}{\sqrt{\lg|V_2|}} \right\rceil.$$

We fix the values of the parameters as follows:

$$\ell = \left\lceil \frac{10}{c} \lg\lg|V(I)| \right\rceil;$$
$$m = 6^\ell q;$$
$$D = 10.$$

Note that for this choice of parameters, $\lg|V_2| = \Theta(\lg|V(G)|)$.

Our inapproximability result requires the following two inequalities.

- For a fixed partition $\mathbf{X} = (X_1, X_2, \ldots, X_t)$ of $V(G(I))$,

$$\Pr[\mathbf{X} \text{ is a complete partition}] \leq \exp(-\Omega(|V_2|^{5/3})). \tag{3}$$

- The number of partitions of $V(G(I))$ into t parts is at most

$$t^{|V(G)|} \leq \exp(O(|V_2|^{3/2})). \tag{4}$$

Assuming these inequalities are proved, we have the following lemma, from which Theorem 3.1 follows immediately (using Theorem 3.3).

Lemma 3.13 (Separation). *If I is 3-colorable, then $\mathsf{cp}(G(I)) \geq |V_2|$. If I is ϵ-non-3-colorable, then with high probability $\mathsf{cp}(G(I)) \leq \dfrac{D|V_2|}{\sqrt{\lg |V_2|}}$.*

It remains to show the inequalities (3) and (4). We first outline our argument. In the outline, we will encounter Lemmas 3.14, 3.15 and 3.16. After the outline, we state these lemmas and assuming they hold, we complete the proof of inequalities (3) and (4). In Section 3.3, we prove the three lemmas.

In order to prove (3), let us fix a potential partition $\mathbf{X} = (X_1, X_2, \ldots, X_t)$. For $\mathbf{X}$ to be a complete partition we should have $X_i \sim X_j$ for all pairs $\{X_i, X_j\}$, $(i \neq j)$. Let us examine how connections between X_i and X_j arise in $G(I)$. There are three issues that we need to deal with in our calculations:

X_i covers X_j: Suppose $X_i \cap V_1$ covers some super edge $\tilde{e}$ of $\mathsf{MRG}_{\equiv}$ and there is some vertex $v_2 \in M_{\tilde{e}}$. Now, Proposition 3.11 (b) implies that v_2 has a neighbor in $X_i \cap V_1$ with probability 1. In such a situation, we say that X_i covers v_2, for no matter how the random edges are chosen, v_2 will be connected to some vertex in X_i. We say that X_i covers X_j if X_i covers some vertex of $X_j \cap V_2$. Since such pairs (X_i, X_j) are connected with probability 1, we need to ensure that there are not too many such pairs. Lemma 3.14 below shows that in any fixed partition $\mathbf{X}$ of $V(G)$, for most pairs (X_i, X_j), X_i does not cover X_j. The reason for this is that the size of a typical $X_i \cap V_1$ is $O(q\sqrt{\lg n})$ and, we can use Lemma 3.10 to bound the number of super edges $\tilde{e}$ of $\mathsf{MRG}_{\equiv}(I)$ covered by such small sets.

X_i and X_j are strongly coupled: Suppose for a pair $\{X_i, X_j\}$, neither X_i covers X_j nor X_j covers X_i. Consider the potential edges between $X_i \cap V_1$ and $X_j \cap V_2$. Each such edge has probability exactly $\frac{1}{2}$ of appearing in $G(I)$. Furthermore, since X_i does not cover X_j, the events "$(u, v_2) \in E(G(I))$", for different potential edges (u, v_2) connecting $X_i \cap V_1$

with $X_j \cap V_2$ are either identical or independent. In particular, if there are k_{ij} potential edges between $X_i \cap V_1$ and $X_j \cap V_2$, then

$$\Pr[(X_i \cap V_1) \sim (X_j \cap V_2) \text{ in } G] \leq 1 - 2^{-k_{ij}}.$$

To exploit this bound, we need to ensure that k_{ij} is small. Lemma 3.15 shows that for most pairs (X_i, X_j), we have $k_{ij} \ll \lg |V(G)|$.

Dependent potential edges: We have indicated above how we plan to control the probability that there is an edge between $X_i \cap V_1$ and $X_j \cap V_2$. The events "$(X_i \cap V_1) \sim (X_j \cap V_2)$" and "$(X_j \cap V_1) \sim (X_i \cap V_2)$" are independent because they involve different vertices of V_2 (see the remarks after the definition of $\tilde{G}(I)$). So, using the argument outlined above we can claim that

$$\Pr[X_i \sim X_j \text{ in } G] \leq 1 - 2^{-(k_{ij} + k_{ji})}.$$

This quantity is quite close to 1 for the bound on k_{ij} that we ensure in Lemma 3.15. But, we want our failure probability to be exponentially small. To achieve that we will exploit the fact that $X_i \sim X_j$ for many pairs $\{X_i, X_j\}$. So, we ensure that the events of the form $X_i \sim X_j$ are independent for many pairs $\{X_i, X_j\}$. The event $X_i \sim X_j$ is independent of a set of events of the form "$X_{i'} \sim X_{j'}$", if no potential edge between X_i and X_j is related to a potential edge between any of the $X_{i'}$ and $X_{j'}$. We stated above that we can ensure that the number of potential edges between X_i and X_j are few ($\ll \lg |V_2|$); furthermore, each such edge is related to at most $2^{\ell+1}$ other potential edges. This enables us to argue that for at least $\Omega(t^2/(2^\ell \lg |V_2|))$ pairs $\{X_i, X_j\}$, the events $X_i \sim X_j$ are mutually independent (see Lemma 3.16 below).

Lemma 3.14. *There are at most $\frac{|V_2|^2}{\lg |V_2|}$ pairs $\{X_i, X_j\}$ where X_i covers X_j or X_j covers X_i.*

Lemma 3.15. *There are at most $\frac{32|V_2|^2}{\lg |V_2|}$ pairs $\{X_i, X_j\}$ where there are more than $\frac{1}{4} \lg |V_2|$ potential edges between them.*

Lemma 3.16. *Let P be a subset of pairs $\{X_i, X_j\}$ such that there are at most $\frac{1}{4} \lg |V_2|$ potential edges connecting any pair. Then, there is a subset P' of P of size at least $|P|/(2^\ell \lg |V_2|)$ such that for all pairs $\{X_i, X_j\} \in P'$, the events $X_i \sim X_j$ are mutually independent.*

Proof of (3). Let $\mathbf{X} = (X_1, X_2, \ldots, X_t)$ be a partition of $V(G)$. By Lemmas 3.14 and 3.15, for at least

$$\binom{t}{2} - 33 \frac{|V_2|^2}{\lg |V_2|} \geq \frac{|V_2|^2}{\lg |V_2|}$$

pairs $\{X_i, X_j\}$, we have

- X_i does not cover X_j and X_j does not cover X_i;
- the number of potential edges between X_i and X_j is at most $\frac{1}{4}\lg|V_2|$.

By Lemma 3.16, there is a subset P' of at least $\frac{|V_2|^2}{2^\ell(\lg|V_2|)^2}$ pairs such that

- for all pairs $\{X_i, X_j\} \in P'$ the events $X_i \sim X_j$ are mutually independent.

Then,

$$\begin{aligned}\Pr[\mathbf{X} \text{ is a complete partition}] &\le \Pr[\forall\{X_i, X_j\} \in P' : X_i \sim X_j] \\ &\le \left(1 - \frac{1}{2^{\frac{1}{4}\lg|V_2|}}\right)^{\frac{|V_2|^2}{2^{\ell+2}\lg|V_2|}} \\ &\le \exp\left(-\Omega\left(\frac{|V_2|^{7/4}}{\lg|V_2|}\right)\right) \\ &\le \exp(-\Omega(|V_2|^{5/3})).\end{aligned}$$

∎

Proof of (4). The total number of partitions of $V(G)$ into t parts is at most

$$\begin{aligned}t^{|V(G)|} &\le |V_2|^{|V(G)|} \\ &\le \exp(|V(G)|\lg|V_2|) \\ &\le \exp(O(|V_2|^{3/2})).\end{aligned}$$

To justify the last inequality, observe that

$$\begin{aligned}|V_2|^{3/2} &= |V_2||V_2|^{1/2} \\ &= |V_2|(|V(I)|^\ell 5^{2\ell} m)^{1/2} \\ &= |V_2|(|E(I)|^\ell \cdot 10^\ell \cdot 6^\ell q)^{1/2} \\ &= |V_2| \cdot (q \cdot 10^\ell \cdot 6^\ell q)^{1/2} \\ &= |V_2| q (60)^{\ell/2}.\end{aligned}$$

On the other hand, since $|V_1| = 2q|V_2|$, $q = (\frac{5}{2})^\ell |V(I)|^\ell$ and $\ell = \frac{10}{c}\lg\lg|V(I)|$, we have

$$\begin{aligned}|V(G)|\lg|V_2| &= (2q+1)|V_2|\lg|V_2| \\ &\le (2q+1)|V_2|\lg(|V(I)|^\ell 5^{2\ell} m) \\ &= (2q+1)|V_2|\lg(|V(I)|^\ell 5^{2\ell} 6^\ell q) \\ &\ll |V_2| q (60)^{\ell/2}.\end{aligned}$$

∎

3.3. Proofs of the combinatorial lemmas

Proof of Lemma 3.14. Let $|X_i \cap V_1| = 2k_i q$. Then,

$$\sum_{i=1}^{t} 2k_i q = |V_1| = 2q|V_2|.$$

Thus, $\sum_{i=1}^{t} k_i \le |V_2|$.

Recall that V_1 consists of r copies $\tilde{V}_1^1, \tilde{V}_1^2, \ldots, \tilde{V}_1^r$ of $\tilde{V}_1$. Denote by v_1^j the copy of the vertex $v_1 \in \tilde{V}_1$ that appears in $\tilde{V}_1^j$. If some X_i has many copies of a vertex, we retain exactly one of them in $\hat{X}_i$:

$$\hat{X}_i = \{v_1 \in \tilde{V}_1 : v_1^j \in X_i, \text{ for some } j \in \{1, 2, \ldots, r\}\}.$$

Clearly, $|\hat{X}_i| \le |X_i| = 2k_i q$. Also, if X_i covers $v_2 \in M_{\tilde{e}} \subseteq V_2$, then $\hat{X}_i$ covers the super edge $\tilde{e}$ in $\mathsf{MRG}_{\equiv}(I)$. Furthermore, the fraction of vertices of V_2 covered by X_i is at most the fraction of super edges of $\mathsf{MRG}_{\equiv}(I)$ covered by $\hat{X}_i$. Let $\hat{X}_i$ cover a fraction α_i of the super edges of $\mathsf{MRG}_{\equiv}(I)$. By Lemma 3.10,

$$\alpha_i \le (32k_i^2 2^{-c\ell})^{1/3} \le 4k_i^{2/3} 2^{-c\ell/3}.$$

The number of X_j's covered by X_i is at most the number of vertices of V_2 covered by X_i, that is, $\alpha_i|V_2|$. Thus, the total number of pairs (X_i, X_j) where X_i covers X_j is at most

$$\begin{aligned}
\sum_{i=1}^{t} \alpha_i |V_2| &\le \sum_{i=1}^{t} 4k_i^{2/3} 2^{-c\ell/3} |V_2| \\
&\le 4 \cdot 2^{-c\ell/3} t |V_2| \left(\sum_{i=1}^{t} k_i/t \right)^{2/3} \\
&\le 4 \cdot 2^{-c\ell/3} \cdot \frac{D|V_2|}{\sqrt{\lg |V_2|}} \cdot |V_2| \frac{(\lg |V_2|)^{1/3}}{D^{2/3}} \\
&\le \frac{|V_2|^2}{2 \lg |V_2|}.
\end{aligned}$$

[The second inequality holds because $x^{2/3}$ is a concave function (Jensen's inequality); the last inequality holds because, for our choice of ℓ, $2^{c\ell/3} \gg \lg |V_2|$.]

So, the number of pairs (X_i, X_j) where X_i covers X_j is at most $\frac{|V_2|^2}{2 \lg |V_2|}$. Similarly, the number of such pairs where X_j covers X_i is at most $\frac{|V_2|^2}{2 \lg |V_2|}$. ∎

Notation. Recall that V_1 is the disjoint union of sets $A^1_{\overline{e}}, A^2_{\overline{e}}, \dots, A^r_{\overline{e}}$ and $B^1_{\overline{e}}, B^2_{\overline{e}}, \dots, B^r_{\overline{e}}$. Let

$$\mathcal{A}_{\overline{e}} = \bigcup_{j=1}^{r} A^j_{\overline{e}}, \qquad \mathcal{A} = \bigcup_{\overline{e}\in|E(I)|^\ell} \mathcal{A}_{\overline{e}}, \qquad \mathcal{B}_{\overline{e}} = \bigcup_{j=1}^{r} B^j_{\overline{e}}, \qquad \mathcal{B} = \bigcup_{\overline{e}\in|E(I)|^\ell} \mathcal{B}_{\overline{e}}.$$

We define a version of the super min-rep-graph with vertex sets $\{\mathcal{A}_{\overline{e}} : \overline{e} \in E(I)^\ell\}$ and $\{\mathcal{B}_{\overline{e}} : \overline{e} \in E(I)^\ell\}$. This graph, $\mathsf{MRG}^*_\equiv(I)$, has a super edge of the form $\tilde{\mathbf{e}} = (\mathcal{A}_{\overline{e}}, \mathcal{B}_{\overline{e'}}, \overline{v})$ if and only if $\mathsf{MRG}_\equiv(I)$ has the corresponding super edge $\tilde{e} = (A_{\overline{e}}, B_{\overline{e'}}, \overline{v})$. Clearly, $\mathsf{MRG}^*_\equiv$ is isomorphic to $\mathsf{MRG}_\equiv(I)$ under the correspondence $\mathcal{A}_{\overline{e}} \mapsto A_{\overline{e}}$ and $\mathcal{B}_{\overline{e'}} \mapsto B_{\overline{e'}}$, where the edge $\tilde{\mathbf{e}}$ corresponds to $\tilde{e}$. With this correspondence in mind, we will refer to the sets $M_{\tilde{e}} \subseteq V_2$ as $M_{\tilde{\mathbf{e}}}$.

Proof of Lemma 3.15. Let $\mathcal{H}_A$ be the pairs (i,j) where the number of potential edges between $X_i \cap \mathcal{A}$ and $X_j \cap V_2$ is at least $\frac{1}{16}\lg|V_2|$. Similarly, let $\mathcal{H}_B$ be the pairs (i,j) where the number of potential edges between $X_i \cap \mathcal{B}$ and $X_j \cap V_2$ is at least $\frac{1}{16}\lg|V_2|$. Clearly, if there are at least $\frac{1}{4}\lg|V_2|$ potential edges between X_i and X_j, then either (i,j) or (j,i) is in $\mathcal{H}_A \cup \mathcal{H}_B$. Thus, it suffices to show that

$$|\mathcal{H}_A|, |\mathcal{H}_B| \le 16|V_2|^2/\lg|V_2|.$$

In the following $\overline{e}$ ranges over the set $E(I)^\ell$ and i,j range over the set $\{1,2,\dots,t\}$. Let $t_{\overline{e},i} = |\mathcal{A}_{\overline{e}} \cap X_i|$. Clearly,

$$\sum_{\overline{e},i} t_{\overline{e},i} = \sum_{\overline{e},i} |\mathcal{A}_{\overline{e}} \cap X_i| = |V_1|/2. \tag{5}$$

Now, if X_i has a vertex in $\mathcal{A}_{\overline{e}}$, we consider which all vertices of $X_j \cap V_2$ it can have edges to. These vertices must come from $M_{\tilde{\mathbf{e}}}$, where $\tilde{\mathbf{e}}$ is incident on $\mathcal{A}_{\overline{e}}$ in the graph $\mathsf{MRG}^*_\equiv(I)$. Let

$$n_{\overline{e},j} = \left|\bigcup_{\tilde{\mathbf{e}}} (M_{\tilde{\mathbf{e}}} \cap X_j)\right|,$$

here $\tilde{\mathbf{e}}$ ranges over super edges incident on $\mathcal{A}_{\overline{e}}$ in the min-rep-graph $\mathsf{MRG}^*_\equiv(I)$ defined above. Every $\mathcal{A}_{\overline{e}}$ has the same number d of super edges $(\mathcal{A}_{\overline{e}}, \mathcal{B}_{\overline{e}}, \overline{v})$ incident on it. [Indeed $d = 10^\ell$, because there are 2^ℓ ways for choosing the label $\overline{v}$, and once the label is chosen, there are 5^ℓ ways of choosing $\overline{e'}$.] So,

$$\sum_{j=1}^{t} n_{\overline{e},j} \le dm. \tag{6}$$

We are now ready to estimate the size of $\mathcal{H}_A$. Note that the number of potential edges between $X_i \cap \mathcal{A}_{\overline{e}}$ and $X_j \cap V_2$ is precisely

$$\sum_{\overline{e}} t_{\overline{e},i} \cdot n_{\overline{e},j}.$$

Note that for $(i,j) \in \mathcal{H}_A$, this quantity is at least $\frac{1}{16} \lg |V_2|$. Let us sum this equation over all (i,j) and note that the contribution from the pairs in $\mathcal{H}_A$ itself will be $|\mathcal{H}_A| \cdot \frac{1}{16} \lg |V_2|$. Thus, we have the following sequence of deductions using (5) in the third inequality and (6) in the second inequality:

$$\begin{aligned}
\sum_{i,j} \sum_{\overline{e}} t_{\overline{e},j} \cdot n_{\overline{e},i} &\geq |\mathcal{H}_A| \cdot \frac{1}{16} \lg |V_2| \\
\sum_{\overline{e},i} t_{\overline{e},i} \sum_{j} n_{\overline{e},j} &\geq |\mathcal{H}_A| \cdot \frac{1}{16} \lg |V_2| \\
dm \cdot \sum_{\overline{e},i} t_{\overline{e},i} &\geq |\mathcal{H}_A| \cdot \frac{1}{16} \lg |V_2| \\
dm \cdot \frac{|V_1|}{2} &\geq |\mathcal{H}_A| \cdot \frac{1}{16} \lg |V_2| \\
q|V_2|dm &\geq |\mathcal{H}_A| \cdot \frac{1}{16} \lg |V_2| \quad (\text{using } |V_1| = 2q \cdot |V_2|).
\end{aligned}$$

Finally, since $qdm = |V_2|$, we conclude that $|\mathcal{H}_A| \leq 16|V_2|^2 / \lg |V_2|$. Similarly, $|\mathcal{H}_B| \leq 16|V_2|^2 / \lg |V_2|$. It follows that the number of pairs $\{X_i, X_j\}$ where there are at least $\frac{1}{4} \lg |V_2|$ potential edges between X_i and X_j is at most $32|V_2|^2 / \lg |V_2|$. ∎

Proof of Lemma 3.16. By Proposition 3.11 (a), two events of the form $X_i \sim X_j$ are dependent only if they involve related potential edges. Let P' be a maximal subset of P such that for any two distinct pairs $\{X_{i_1}, X_{j_1}\}$ and $\{X_{i_2}, X_{j_2}\}$ in P', every potential edge connecting X_{i_1} and X_{j_1} is unrelated to every potential edge connecting X_{i_2} and X_{j_2}. Let $|P'| = \alpha |P|$ for some $\alpha \in [0,1]$. Then, for every pair $\{X_{i'}, X_{j'}\}$ in $P \backslash P'$, some potential edge connecting $X_{i'}$ and $X_{j'}$ is related to some potential edge connecting some pair in P' (otherwise, P' would not be maximal). There are at most $\frac{1}{4} \lg |V_2|$ potential edges connecting any one pair in P', each such edge is related to exactly $2^{\ell+1}$ other potential edges. Thus,

$$|P \setminus P'| \leq \left(2^{\ell+1} \cdot \frac{1}{4} \lg |V_2| \right) |P'|,$$

implying

$$|P'| \geq \frac{|P|}{1 + 2^{\ell-1} \lg |V_2|} \geq \frac{|P|}{2^\ell \lg |V_2|}. \qquad \blacksquare$$

3.4. Lower bounds for computing the achromatic number

The goal of this section is to prove the following.

Theorem 3.17. *Let $L \subseteq \{0,1\}^*$ be a language in* NP. *Then, there is a randomized transformation f that transforms inputs in $\{0,1\}^*$ to graphs, and a polynomial-time computable function $\alpha : \{0,1\}^* \to \mathbb{N}$ such that for all $x \in \{0,1\}^*$,*

(a) if $x \in L$ then $\mathsf{achr}(f(x)) \geq \alpha(x)$ with probability $\frac{3}{4}$;

(b) if $x \notin L$, then with probability at least $\frac{3}{4}$, we have $\mathsf{achr}(f(x)) \leq C\alpha(x)/\sqrt{\lg |V(f(x))|}$, where $C \geq 1$ is an absolute constant independent of L and x.

The function f can be computed in time $|x|^{O(\lg \lg |x|)}$.

Corollary 3.18. *There exists a constant $C \geq 1$ such that if there exists a randomized polynomial-time algorithm that with high probability computes a complete partition in a graph G of size more than $\frac{C\,\mathsf{achr}(G)}{\sqrt{\lg |V(G)|}}$, then* $\mathsf{NP} \subseteq \mathsf{RTime}(n^{O(\lg \lg n)})$.

Proof of Theorem 3.17. Consider the bipartite graph $G = (V_1, V_2, E)$ constructed above for showing Theorem 3.1. We will need the following key property of this graph when the input x is in the language L:

> There is a partition $C_1, C_2, \ldots$ of V_1 into $t = |V_2|$ parts such that $N(C_i) = V_2$. Furthermore, each vertex in V_2 has exactly t edges. Thus, for each vertex $v_2 \in V_2$ and class C_i, there is exactly one edge between v_2 and C_i.

Let G' be the random graph obtained from G by randomly and independently deleting one edge from each vertex in V_2.

If $x \in L$: We need to show that in this case, with high probability, G' has a large achromatic partition. For $i = 1, 2, \ldots, t$, let

$$V_{2,i} = \{v_2 \in V_2 : v_2 \notin N_{G'}(C_i)\},$$

and consider the sets $C'_i = C_i \cup V_{2,i}$ and let $C'' = \bigcup_{i : V_{2,i} = \emptyset} C_i$. Then, it is easy to see that

$$\{C'_i : V_{2,i} \neq \emptyset\} \cup \{C''\}$$

is an achromatic partition of G'. We need to argue that there are many i such that $V_{2,i} \neq \emptyset$. Note that the number of such i's has exactly the same distribution as the number of non-empty bins when t balls are thrown into t bins, independently at random. In this case, the expected number of non-empty bins is exactly $(1-1/e)t$. Using the method of bounded differences [33] or by exploiting the negative correlation [11], we can conclude that the Chernoff–Hoeffding bounds hold for this distribution, that is,

$$\Pr[\text{the number of non-empty bins} \leq (1-1/e-\epsilon)t] \leq \exp(-2\epsilon^2 t).$$

Thus, with probability at least 3/4, G' has an achromatic partition of size at least $t/2$, provided that $|V_2| = t \geq 10$.

If $x \notin L$: We know that the size of every complete partition of G is $O(t/\sqrt{\lg |V(G)|})$. Thus, in particular, the achromatic number of G is $O(t/\sqrt{\lg |V(G)|})$. ∎

Acknowledgment. We thank Keith J. Edwards for helpful pointers. We are grateful to the anonymous referee for comments on the earlier version of the paper and permission to include the referee's results in Section 2.5.

References

[1] S. Arora and C. Lund: Hardness of approximations, in: D. Hochbaum, editor, *Approximation Algorithms for NP-Hard Problems*, PWS Publishing Company, 1996.

[2] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy: Proof verification and hardness of approximation problems, *J. ACM* **45(3)** (1998), 501–555.

[3] R. Balasubramanian, V. Raman and V. Yegnanarayanan: On the pseudoachromatic number of join of graphs, *J. Computer Math.* **80** (2003), 1131–1137.

[4] H. L. Bodlaender: Achromatic number is NP-complete for cographs and interval graphs, *Inform. Process. Lett.* **31(3)** (1989), 135–138.

[5] V. N. Bhave: On the pseudoachromatic number of a graph, *Fund. Math.* **102(3)** (1979), 159–164.

[6] B. Bollobás, P. A. Catlin and P. Erdős: Hadwiger's conjecture is true for almost all graphs, *Europ. J. Combin.* **1** (1980), 195–199.

[7] B. Bollobás, B. Reed and A. Thomason: An extremal function for the achromatic number, *Graph structure theory*, 161–165, *Contemp. Math.* **147**, AMS 1993.

[8] N. Cairnie and K. J. Edwards: Some results on the achromatic number, *J. Graph Theory* **26(3)** (1997), 129–136.

[9] N. Cairnie and K. J. Edwards: The achromatic number of bounded degree trees, *Discrete Math.* **188** (1998), 87–97.

[10] J. Chuzhoy, S. Guha, E. Halperin, S. Khanna, G. Kortsarz and J. Naor: Tight lower bounds for the asymmetric k-center problem, in *Proc. 36th Annual ACM Symposium on Theory of Computing* (STOC), 21–27, 2004.

[11] D. Dubhashi and D. Ranjan: Balls and bins: A study in negative dependence, *Random Structures and Algorithms* **13(2)** (1998), 99–124.

[12] K. J. Edwards: The harmonious chromatic number and the achromatic number, in *Surveys in Combinatorics 1997*, Cambridge University Press, pages 13–47, 1997.
[13] K. J. Edwards: Achromatic number versus pseudoachromatic number: A counterexample to a conjecture of Hedetniemi, *Discrete Math.* **219** (2000), 271–274.
[14] J. Edmonds and E. L. Johnson: Matching: A well solved class of integer linear programs, in *Combinatorial Structures and Their Appl. (Proc. Calgary Internat. Conf.)*, pages 89–92, 1970.
[15] K. J. Edwards and C. J. H. McDiarmid: The complexity of harmonious coloring for trees, *Disc. Appl. Math.* **57** (1995), 133–144.
[16] U. Feige: A threshold of $\ln n$ for approximating set cover, *J. ACM* **45(2)** (1998), 634–652.
[17] U. Feige, M. M. Halldórsson, G. Kortsarz and A. Srinivasan: Approximating the domatic number, *SIAM J. Computing* **32(1)** (2002), 172–195.
[18] H. N. Gabow: An efficient reduction technique for degree constrained subgraph and bidirected network flow problems, in *Proc. 15th Annual ACM Symposium on Theory of Computing* (STOC), pages 448–456, 1983.
[19] R. P. Gupta: Bounds on the chromatic and achromatic numbers of complementary graphs, in *Recent Progress in Combinatorics* Academic Press, New York (1969), pp. 229–235.
[20] M. M. Halldórsson: Approximating the minimum maximal independence number, *Inform. Process. Lett.* **46(4)** (1993), 169–172.
[21] M. M. Halldórsson: An approximation algorithm for complete partitions of regular graphs, Unpublished manuscript, http://www.hi.is/~mmh/publications.html, February 2004.
[22] J. Håstad: Clique is hard to approximate within $n^{1-\epsilon}$, *Acta Mathematica* **182** (1999), 105–142.
[23] J. Håstad: Some optimal inapproximability results, *J. ACM* **48** (2001), 798–859.
[24] E. Halperin and R. Krauthgamer: Polylogarithmic inapproximability, in *Proc. 35th Annual ACM Symposium on Theory of Computing* (STOC), 585–594, 2003.
[25] S. Hedetniemi: Open problems in combinatorial optimization, Webpage at www.cs.clemson.edu/~hedet/coloring.html, September 1998.
[26] C. Lund and M. Yannakakis: On the hardness of approximating minimization problems, *J. ACM* **41(5)** (1994), 960–981.
[27] A. V. Kostochka: The minimum Hadwiger number for graphs with a given mean degree of vertices, *Metody Diskret. Analiz.* **38** (1982), 37–58 (in Russian).
[28] G. Kortsarz: On the hardness of approximating spanners, *Algorithmica* **30(3)** (2001), 432–450.
[29] G. Kortsarz and R. Krauthgamer: On the approximation of the achromatic number, *SIAM Journal of Discrete Math.* **14(3)** (2000), 408–422.
[30] G. Kortsarz, J. Radhakrishnan and S. Sivasubramanian: Complete partitions of graphs, in *Proc. 16th Annual SIAM-ACM Symp. Discrete Algorithms (SODA)*, pages 860–869, 2005.
[31] G. Kortsarz and S. Shende: Approximating the achromatic number problem on bipartite graphs, in *Proceedings of 11th European Symposium on Algorithms* (Budapest, 2003), *Lecture Notes in Computer Science* **2832**, Springer, Berlin (2003), pp. 385–396.
[32] C. J. H. McDiarmid: Achromatic numbers of random graphs, *Mathematical Proceedings of Cambridge Philosophical Society* **92** (1982), 21–28.

[33] C. McDiarmid: On the method of bounded differences, *London Mathematical Society Lecture Notes Series* **141** (1989), 148–188.

[34] R. Raz: A parallel repetition theorem, *SIAM J. Comput.* **27(3)** (1998), 763–803.

[35] N. Robertson, P. D. Seymour and R. Thomas: Hadwiger's conjecture for K_6-free graphs, *Combinatorica* **13(3)** (1993), 279–361.

[36] E. Sampathkumar and V. N. Bhave: Partition graphs and coloring numbers of a graph, *Discrete Math.* **16(1)** (1976), 57–60.

[37] V. G. Vizing: On an estimate of the chromatic class of a p-graph, *Diskret. Anal.* **3** (1964), 25–30 (in Russian).

[38] V. Yegnanarayanan: On pseudocoloring of graphs, *Util. Math.* **62** (2002), 199–216.

Magnús M. Halldórsson

Department of Computer Science
University of Iceland
IS-107 Reykjavik
Iceland
mmh@hi.is

Guy Kortsarz

Department of Computer Science
Rutgers University – Camden
Camden, NJ 08102
USA
guyk@camden.rutgers.edu

Jaikumar Radhakrishnan, Sivaramakrishnan Sivasubramanian

School of Technology and Computer Science
Tata Institute of Fundamental Research
Homi Bhabha Road
Mumbai 400005
India
jaikumar@tifr.res.in, skrishna@tcs.tifr.res.in