

Coloring Semirandom Graphs

Amin Coja-Oghlan*

Humboldt-Universität zu Berlin, Institut für Informatik,
Unter den Linden 6, 10099 Berlin, Germany
coja@informatik.hu-berlin.de

Abstract. We study semirandom k -colorable graphs made up as follows. Partition the vertex set $V = \{1, \dots, n\}$ randomly into k classes V_1, \dots, V_k of equal size and include each V_i - V_j -edge with probability p independently ($1 \leq i < j \leq k$) to obtain a graph G_0 . Then, an adversary may add further V_i - V_j -edges ($i \neq j$) to G_0 , thereby completing the semirandom graph $G = G_{n,p,k}^*$. We show that if $np \geq \max\{(1 + \varepsilon)k \ln n, C_0 k^2\}$ for a certain constant $C_0 > 0$ and an arbitrarily small but constant $\varepsilon > 0$, an *optimal* coloring of $G_{n,p,k}^*$ can be found in polynomial time with high probability. Furthermore, if $np \geq C_0 \max\{k \ln n, k^2\}$, a k -coloring of $G_{n,p,k}^*$ can be computed in polynomial expected time. Moreover, an *optimal* coloring of $G_{n,p,k}^*$ can be computed in expected polynomial time if $k \leq \ln^{1/3} n$ and $np \geq C_0 k \ln n$. By contrast, it is NP-hard to k -color $G_{n,p,k}^*$ w.h.p. if $np \leq (\frac{1}{2} - \varepsilon)k \ln(n/k)$.

1 Introduction

1.1 Graph Coloring Heuristics

In the *Graph Coloring Problem* we are given a graph $G = (V, E)$, and the goal is to color the vertices V with as few colors as possible such that adjacent vertices receive distinct colors. The least number of colors so that there exists such a coloring is the *chromatic number* $\chi(G)$.

While the Graph Coloring Problem is of fundamental interest in theoretical computer science as well as in discrete mathematics, the problem is notoriously hard. Indeed, Feige and Kilian [15] proved that no polynomial time algorithm approximates $\chi(G)$ within a factor of $n^{1-\varepsilon}$ for all input graphs G , unless $ZPP=NP$; here $n = \#V$, and $\varepsilon > 0$ is an arbitrarily small constant. Furthermore, Khanna, Linial, and Safra [27] showed that it is NP-hard to color 3-colorable graphs with 4 colors.

Nevertheless, these hardness results merely provide evidence that for every polynomial time algorithm there *exist* hard problem instances. Hence, the hardness results do not rule out the existence of good *graph coloring heuristics* that perform well on “almost all instances” in some meaningful sense. Therefore, the goal of this paper is to analyze graph coloring heuristics rigorously within the framework of the algorithmic theory of random graphs (cf. [20]).

Of course, in order to obtain rigorous results, we need to specify precisely what “almost all instances” is supposed to mean. One possible answer is to consider the well-known *Erdős-Rényi-model* $G_{n,p}$ of random graphs. The random graph $G_{n,p}$ has n vertices $V = \{1, \dots, n\}$, and each of the $\binom{n}{2}$ possible edges is present with probability p independently. Bollobás [5] and Łuczak [31] determined the probable value of $\chi(G_{n,p})$: we have

$$\chi(G_{n,p}) \sim -\frac{n \ln(1-p)}{2 \ln(np)} \quad \text{w.h.p. if } n^{-1} \ll p \leq 0.99. \quad (1)$$

(For small edge probabilities $p = O(1/n)$, Achlioptas and Naor [1] obtained more precise results.) We emphasize that (1) shows that the chromatic number $\chi(G_{n,p})$ is fairly “high”. For if $np = \Omega(\ln n)$, then with probability $1 - o(1)$ as $n \rightarrow \infty$ the maximum degree of $G_{n,p}$ is $O(np)$ (cf. [6, Chapter 3]). Therefore, Brook’s theorem immediately entails that the chromatic number is $O(np)$, and (1) is just by a $O(\ln(np))$ -factor smaller than this trivial upper bound.

* Date: January 10, 2005. An extended abstract version of this paper appeared in Proc. 31st ICALP (2004) 383–395

In order to investigate graphs with a smaller chromatic number than (1), Kučera [29] suggested a random model $G_{n,p,k}$ that has an additional parameter k to control the chromatic number. The random graph $G_{n,p,k}$ is obtained as follows.

- M1.** Partition the vertex set $V = \{1, \dots, n\}$ randomly into k classes V_1, \dots, V_k of equal cardinality (we assume that k divides n).
- M2.** Include every V_i - V_j -edge ($i \neq j$) with probability p independently of all others to obtain $G_0 = G_{n,p,k}$.

Thus, V_1, \dots, V_k is a k -coloring “planted” in $G_{n,p,k}$, so that $\chi(G_{n,p,k}) \leq k$. We say that $G_{n,p,k}$ has some property \mathcal{E} *with high probability* (“w.h.p.”) if the probability that \mathcal{E} holds tends to 1 as $n \rightarrow \infty$.

However, the $G_{n,p}$ and the $G_{n,p,k}$ model share a serious drawback: in both models the instances are purely random. As the theory of random graphs shows (cf. [25]), such instances have a very particular combinatorial structure. Therefore, designing heuristics for $G_{n,p}$ or $G_{n,p,k}$ yields heuristics for a *very special class* of graphs. Consequently, heuristics for purely random instances may lack “robustness”, as even minor changes in the structure of the input may deteriorate the performance.

Therefore, Blum and Spencer [4] suggested a *semirandom* model $G_{n,p,k}^*$ that is inbetween the random graph $G_{n,p,k}^*$ and worst-case instances. The semirandom graph $G_{n,p,k}^*$ is obtained as follows. First, a random graph $G_0 = G_{n,p,k}$ is chosen via M1–M2; let V_1, \dots, V_k signify its planted k -coloring. Then, an adversary completes the problem instance as follows.

- M3.** The adversary may add to G_0 further V_i - V_j -edges ($i \neq j$) to obtain $G = G_{n,p,k}^*$.

Note that V_1, \dots, V_k remains a “planted” k -coloring of $G_{n,p,k}^*$. Hence, $\chi(G_{n,p,k}^*) \leq k$.

Let $\mathcal{I}(G_0)$ signify the set of all graphs that can be obtained from $G_0 = G_{n,p,k}$ via M3. We say that $G_{n,p,k}^*$ has some property \mathcal{E} *with high probability* (“w.h.p.”) if the property \mathcal{E} holds with probability $1 - o(1)$ as $n \rightarrow \infty$ regardless of the adversary’s decisions. That is,

$$\lim_{n \rightarrow \infty} \mathbb{P}[G_0 = G_{n,p,k} \text{ is such that } \mathcal{E} \text{ holds for all } G \in \mathcal{I}(G_0)] = 1.$$

In contrast to $G_{n,p,k}$, the semirandom graph $G_{n,p,k}^*$ does not *consist* of random edges, but *contains* some random edges. Therefore, $G_{n,p,k}^*$ models a somewhat more general type of instances. On the one hand, the adversary can alter certain “statistical” properties of $G_0 = G_{n,p,k}$. For example, the adversary can change the distribution of the vertex degrees or add “dense spots” to the graph, thereby changing also spectral properties. On the other hand, the adversary is just allowed to add edges that “point towards” the hidden coloring V_1, \dots, V_k . Thus, intuitively the adversary just seems to make the problem “easier”. Therefore, it appears natural to require that a “robust” heuristic should not get confused by the adversary’s actions. In other words, the $G_{n,p,k}^*$ model discriminates between heuristics that are robust enough to withstand such an adversarial “help”, and heuristics that are not.

Let us discuss the difference between $G_{n,p,k}$ and $G_{n,p,k}^*$ with a concrete example. Alon and Kahale [2] suggested a spectral heuristic that k -colors $G_{n,p,k}$ w.h.p. if k is fixed and $p > C_k/n$ for a certain constant $C_k > 0$. Given an input instance $G_0 = G_{n,p,k}$, the heuristic first removes all vertices of degree greater than $5np$, thereby obtaining a graph G'_0 . Then, the heuristic computes the $k - 1$ eigenvectors of the adjacency $A(G'_0)$ of G'_0 with the smallest eigenvalues. These eigenvectors yield partition of G'_0 that is “close” to the planted coloring of G_0 w.h.p. Finally, in order to obtain an actual k -coloring of G_0 , the heuristic improves this partition using various combinatorial techniques.

However, this spectral approach breaks down on the $G_{n,p,k}^*$ model. Let us assume for concreteness that $k = 3$, and that $C_3 \leq np = O(1)$. Then a standard argument shows that w.h.p. each of the planted color classes V_1, V_2, V_3 of G_0 contains $\Omega(n)$ isolated vertices. Hence, w.h.p. the adversary can pick disjoint sets $A_1, A_2 \subset V_1, B_1, B_2 \subset V_2$ of isolated vertices such that $\#A_i = \#B_i = 2np/3$. Then, the adversary adds all A_i - B_i -edges to G_0 to obtain a graph G . Thus, in G both (A_1, B_1) and (A_2, B_2) are bipartite cliques. Let G' be the graph obtained by removing all vertices of degree $> 5np$ from G , and let A be the adjacency matrix of G' . Then similar spectral arguments as in [2] show that the two eigenvectors of A with the smallest eigenvalues just represent the bipartite cliques $(A_1, B_1), (A_2, B_2)$,

but do not encode any useful information to 3-color G . Thus, the adversary can jumble up the spectrum of G_0 to render the spectral approach useless. (A similar construction shows that on $G_{n,p,k}^*$ the spectral approach breaks down also for larger values of p – say, $np = n^{o(1)}$.)

1.2 Results

The goal of this paper is to investigate heuristics for coloring $G_{n,p,k}^*$. First, we present a simple heuristic that computes an *optimal* coloring of $G_{n,p,k}^*$ in polynomial time w.h.p. In addition, we suggest heuristics for coloring $G_{n,p,k}^*$ in polynomial expected time. We will compare these results with previous work in Section 1.3.

Coloring $G_{n,p,k}^*$ optimally. While $G_{n,p,k}^*$ is always k -colorable, it might happen that the chromatic number is actually smaller than k . Therefore, we say that a heuristic \mathcal{A} *colors $G_{n,p,k}^*$ optimally w.h.p.* if the following two conditions are satisfied.

Correctness. For all input graphs G the algorithm \mathcal{A} either outputs an optimal coloring or “fail”.

Completeness. On input $G_{n,p,k}^*$, the output is an optimal coloring w.h.p.

Thus, we require \mathcal{A} not only to find a coloring of the input graph, but also to compute a matching *lower* bound on the chromatic number. In other words, \mathcal{A} is supposed to *certify* that its output is an optimal coloring.

Theorem 1. *Let $\varepsilon > 0$ be arbitrarily small but constant. Moreover, suppose that $k = k(n)$ and $p = p(n)$ are such that*

$$np \geq \max\{(1 + \varepsilon)k \ln n, C_0 k^2\} \text{ for a certain constant } C_0 > 0. \quad (2)$$

There is a polynomial time algorithm `Color` that colors $G_{n,p,k}^$ optimally w.h.p.*

Note that for $k = o(\ln n)$ – hence in particular for constant k – the assumption in Theorem 1 just reads $np \geq (1 + \varepsilon)k \ln(n)$. `Color` employs a semidefinite programming (“SDP”) relaxation $\bar{\vartheta}_2$ of the chromatic number (we will recall the definition in Section 2). The basic observation is that on $G = G_{n,p,k}^*$ w.h.p. all optimal fractional solutions to $\bar{\vartheta}_2$ are *integral*, i.e., encode actual colorings of G .

The algorithm `Color` can be considered as a “more robust” version of the spectral heuristic of Alon and Kahale [2]. More precisely, `Color` can cope with the semirandom model $G_{n,p,k}^*$, because we replace the spectral techniques by SDP techniques. Nevertheless, the *proof* that all optimal fractional solutions to the SDP $\bar{\vartheta}_2$ are integral w.h.p. relies on SDP duality and extends the spectral considerations of Alon and Kahale. Extending the spectral techniques to the semirandom model was posed as an open problem by Frieze and McDiarmid [20, Research Problem 19].

The following hardness result complements Theorem 1.

Theorem 2. *Let $3 \leq k \leq n^{1/2}$. There is no polynomial time algorithm that in the case*

$$np \leq (1 - \varepsilon) \frac{k}{2} \ln(n/k) \quad (3)$$

k -colors $G_{n,p,k}^$ w.h.p., unless $NP \subset RP$.*

If $k = o(\ln n)$, then conditions (2) and (3) differ only by a factor of 2. Thus, Theorem 2 implies that for $k = o(\ln n)$ the positive result Theorem 1 is essentially best possible.

Coloring $G_{n,p,k}^*$ in expected polynomial time. Despite Theorem 2, can we push the positive result Theorem 1 any further? The algorithm `Color` for Theorem 1 runs *always* in polynomial time and k -colors $G_{n,p,k}^*$ *with high probability*. One way to strengthen this result is to devise an algorithm that even k -colors *any* k -colorable input graph such that the *expected* running time over $G_{n,p,k}^*$ is polynomial. Here we define the *expected running time* of an algorithm \mathcal{A} on input $G_{n,p,k}^*$ as

$$\sum_{G_0} \mathbb{P}(G_0 = G_{n,p,k}) \cdot \max_{G \in \mathcal{I}(G_0)} R_{\mathcal{A}}(G),$$

where $R_{\mathcal{A}}(G)$ denotes the running time of \mathcal{A} on input G , and the sum ranges over all possible outcomes G_0 of $G_{n,p,k}$. The following theorem shows that there is a coloring algorithm with polynomial expected running time for almost the same range of the parameters as in Theorem 1.

Theorem 3. *Suppose that $k = k(n)$ and $p = p(n)$ are such that*

$$np \geq C_0 \max\{k \cdot \ln n, k^2\} \text{ for a certain constant } C_0 > 0. \quad (4)$$

There is an algorithm `ExpColor` that k -colors any k -colorable input graph and that applied to $G_{n,p,k}^$ has polynomial expected running time.*

To achieve the algorithm `ExpColor` with polynomial expected running time, we need to refine the heuristic `Color` significantly. Indeed, while `Color` may just “give up” if the input lacks certain “typical” properties of $G_{n,p,k}^*$, `ExpColor` must be able to handle *all* k -colorable input graphs. Hence, if we imagine `ExpColor`’s quest for a k -coloring as a search tree, then this search tree can be of polynomial or of exponential size, or anything in between. Therefore, in order to guarantee a polynomial expected running time, we need to extend `Color` and its analysis in two respects.

- We need to improve the algorithm so that the size of the search tree is distributed “smoothly” such that it is small on average. Loosely speaking, this means that `ExpColor` needs to cope with minor “atypical defects” in the input instance in such a way that the running time increases only proportionally to the size of the “defect”.
- We need to invent methods to analyze the average size of the search tree. In particular, we need to quantify how “typical” or “atypical” in terms of the $G_{n,p,k}^*$ model a certain input graph is.

Thus, we need to refine both the heuristic and its analysis, but the enhanced analysis also sheds further light on the algorithmic techniques that the heuristic relies on.

For $k = o(\ln n)$ Theorem 2 shows that the bound (4) on p is best possible up to the precise value of C_0 . However, in contrast to `Color`, `ExpColor` does not certify the optimality of the obtained coloring. Nevertheless, at least for $k \leq \ln^{1/3} n$ (and hence in particular for constant k), it is easy to modify `ExpColor` to obtain an algorithm that certifies the optimality of its output.

Theorem 4. *Suppose that $k = k(n)$ and $p = p(n)$ are such that*

$$np \geq C_0 k \cdot \ln n \text{ for a certain constant } C_0 > 0, \text{ and } k \leq \ln^{1/3} n. \quad (5)$$

There is an algorithm `OptColor` that colors any input graph optimally and that applied to $G_{n,p,k}^$ has polynomial expected running time.*

1.3 Related Work

Blum and Spencer [4] were the first to study the $G_{n,p,k}^*$ model. They showed that a k -coloring of $G_{n,p,k}^*$ can be found in polynomial time w.h.p. if k is constant and

$$np \geq n^{\alpha_k + \varepsilon}, \quad \text{where } \alpha_k = \frac{k^2 - k - 2}{k^2 + k - 2} \quad (6)$$

and $\varepsilon > 0$ is an arbitrarily small constant. This coloring heuristic is purely combinatorial.

Feige and Kilian [16] suggested the strongest previous heuristic for coloring $G_{n,p,k}^*$. The heuristic finds a k -coloring in polynomial time w.h.p. if k is constant and $np \geq (1 + \varepsilon)k \ln n$. Note that for constant k this assumption is identical to (2). In order to k -color $G_{n,p,k}^*$, the heuristic tries to recover the classes of the planted k -coloring one by one. To recover a color class, the heuristic combines SDP techniques for approximating the independence number from Alon and Kahale [3] with the random hyperplane rounding technique from Goemans and Williamson [22]. These SDP techniques are needed to obtain an initial partition of the input graph that consists of relatively “sparse” sets. Then, the heuristic makes use of matching techniques and expansion properties of $G_{n,p,k}^*$ to extract the color class from the initial partition.

Theorem 1 improves on the result of Feige and Kilian in the following respects.

- It is not clear whether the heuristic in [16] is applicable when k grows as a function of n (say, $k \gg \ln n$). For the analysis of the SDP rounding techniques in [16] requires that the initial partition consists of $\exp(\Omega(k))$ classes to guarantee that the classes of the partition are sparse enough. On the other hand, choosing $p = 1/2$ we can make k as large as $\Omega(\sqrt{n})$ in Theorem 1.
- The algorithm `Color` is much simpler. For instance, it needs to solve an SDP only once, whereas the heuristic of Feige and Kilian requires several SDP computations. (Nonetheless, the techniques in [16] apply to further problems that we do not address in this paper.)
- Instead of just producing a k -coloring of $G = G_{n,p,k}^*$ w.h.p., `Color` also provides a *certificate* that its output is indeed optimal.

In addition, Feige and Kilian [16] proved that no polynomial time algorithm k -colors $G_{n,p,k}^*$ w.h.p. if $np \leq (1 - \varepsilon) \ln n$, unless $\text{NP} \subseteq \text{RP}$. Theorem 2 improves this result by a factor of $k/2$, although the proof uses a similar idea.

Theorems 3 and 4 also improve on a coloring algorithm in Coja-Oghlan [10], which is based on similar techniques as the algorithm of Feige and Kilian [16]. The coloring algorithm in [10] k -colors any k -colorable input graph, and the expected running time on $G_{n,p,k}^*$ is $n^{\Theta(k)}$, provided that $np \gg k \ln n$. Hence, the running time of this algorithm becomes superpolynomial if $k = k(n)$ grows as a function of n . By contrast, the expected running time of `ExpColor` is polynomial in both n and k (cf. Theorem 3). Furthermore, in contrast to the algorithm `OptColor` (cf. Theorem 4), even for constant k the coloring algorithm in [10] does not certify the optimality of its output.

Building on [34], Subramanian [33] gave a heuristic for coloring $G_{n,p,k}^*$ optimally in polynomial expected time for constant values of k under the assumption (6). The heuristic is purely combinatorial, and the certificate of optimality is just a clique of size k w.h.p. Theorem 4 extends this result to significantly smaller values of p . In fact, for small edge probabilities $p = C_0 k \ln n$ as in Theorem 4, the clique number of $G_{n,p,k}$ is 3 w.h.p. Hence, $G_{n,p,k}^*$ has no clique of size k (unless the adversary includes one) that yields a certificate that the obtained coloring is optimal. Therefore, in order to certify optimality `OptColor` employs SDP techniques.

With respect to coloring $G_{n,p,k}$, Kučera [29] presented a simple heuristic that for $k = O(\sqrt{n/\ln n})$ and $p = 1/2$ recovers the planted k -coloring of $G_{n,p,k}$ w.h.p. Note that Theorem 1 provides a slightly result: `Color` colors $G_{n,p,k}$ optimally if $p = 1/2$ and $k \leq c\sqrt{n}$ for a certain constant $c > 0$.

Dyer and Frieze [14] showed that an optimal coloring of $G_{n,p,k}$ can be found in polynomial expected time if $p = \Omega(1)$ remains bounded away from 0 as $n \rightarrow \infty$. Moreover, the best previous heuristic for coloring $G_{n,p,k}$ in polynomial expected time is due to Subramanian [33]. The heuristic is combinatorial and colors $G_{n,p,k}$ optimally in polynomial expected time if k is constant and

$$np \geq n^{\gamma(k)+\varepsilon}, \quad \text{where } \gamma(k) = \frac{k^2 - 3k + 2}{k^2 - k + 2}.$$

Theorem 4 provides a coloring heuristic that also applies to significantly smaller values of p . Extending Subramanian's result to smaller values of p was also mentioned as an open problem in the survey of Krivelevich [28, Section 7].

Some heuristics for random instances of more general partitioning problems also entail results on coloring $G_{n,p,k}$. For instance, the heuristic of Subramanian and Veni Madhavan [35], which is based on breadth first search, k -colors $G_{n,p,k}$ in polynomial time w.h.p. if k is constant and $np \geq \exp(C\sqrt{\ln n})$ for a certain constant $C > 0$. Moreover, McSherry's spectral heuristic [32] finds a k -coloring in polynomial time w.h.p. if k is constant and $np \gg \ln^3 n$. Finally, a randomized linear time partitioning heuristic of Bollobás and Scott [7] recovers the hidden coloring w.h.p. if $np \geq Ck^2 \ln n$ for a certain constant $C > 0$. Indeed, Bollobás and Scott conjecture that their heuristic can also handle the semirandom graph $G_{n,p,k}^*$. Nevertheless, Theorem 1 applies to even smaller edge probabilities than [7, 32, 35]. Some further references on coloring random and semirandom graphs can be found in the survey [28].

1.4 Techniques and Outline

The algorithms `Color`, `ExpColor`, and `OptColor` for Theorems 1, 3, and 4 make use of different techniques than the previous algorithms for coloring $G_{n,p,k}$ and $G_{n,p,k}^*$. For instance, `Color` relies on a direct analysis of the optimal solutions to the SDP relaxation $\bar{\vartheta}_2$ on $G_{n,p,k}^*$ (cf. Section 2 for the definition of $\bar{\vartheta}_2$). More precisely, we show that all optimal fractional solutions are in fact integral w.h.p., i.e., correspond to k -colorings of $G_{n,p,k}^*$. While the algorithm for coloring semirandom graphs in [16] is also based on SDP techniques (cf. Section 1.3), `Color` is rather different: the analysis of `Color` shows that there is a single SDP that captures the problem completely.

The techniques in the analysis of `Color` extends previous work of Boppana [8] and Feige and Kilian [16] on the MIN BISECTION problem. More precisely, in [8] it was shown that all optimal fraction solutions to a SDP relaxation of MIN BISECTION correspond to actual bisections w.h.p. on certain random instances; this analysis was extended in [16] to semirandom models. Nevertheless, the analysis of $\bar{\vartheta}_2$ on the $G_{n,p,k}^*$ model turns out to be significantly more involved than the analyses for MIN BISECTION in [8, 16]. One reason is that while in the MIN BISECTION problem the goal is to recover *two* classes, the number $k = k(n)$ of color classes in the $G_{n,p,k}^*$ model may actually grow as a function of n . A further reference is the work of Feige and Krauthgamer [17] on semirandom instances of the Clique problem; the heuristic is based on the integrality of optimal fractional solutions to a certain relaxation of the clique number.

In order to obtain the heuristic `ExpColor` with polynomial expected running time, we need to refine the investigation of $\bar{\vartheta}_2$ on $G_{n,p,k}^*$. While `Color` relies on the fact that all fractional solutions are perfectly integral w.h.p., `ExpColor` is based on the observation that with probability extremely close to 1 all fractional solutions are at least “not too far” from being integral. To prove this statement, we invoke results from Coja-Oghlan, Moore, and Sanwalani [13] on semidefinite relaxations of MAX k -CUT on the Erdős-Rényi model $G_{n,p}$. In addition, to extract the coloring from the fractional solution, `ExpColor` employs network flow techniques from [11], which extend matching techniques from [16]. Finally, `OptColor` combines `ExpColor` with a technique for computing a lower bound on $\chi(G_{n,p,k}^*)$.

The heuristic `Color` and its analysis are the content of Section 3. Moreover, we present `ExpColor` in Section 4. Then, in Section 5 we modify `ExpColor` to obtain the algorithm `OptColor` for Theorem 4. Section 6 is devoted to the proof of Theorem 2. Finally, Section 7 contains the proofs of some technical lemmas.

There are various constants involved in the analyses of the algorithms. Most of the constants are somewhat arbitrary and are only made explicit for concreteness; no attempt has been made to optimize these constants.

2 Preliminaries

Notation. Throughout, we let $V = \{1, \dots, n\}$. Moreover, if X is a set, then we let $\delta_{x,X} = 1$ if $x \in X$ and $\delta_{x,X} = 0$ otherwise.

If G is a graph, then we let $V(G)$ denote the vertex set and $E(G)$ the edges set of G . For a set $A \subset V(G)$, $N(A) = N_G(A) = \{w \in V : \exists a \in A : \{v, w\} \in E(G)\}$ signifies the neighborhood of A . Moreover, $\bar{N}(A) = \bar{N}_G(A) = V \setminus N_G(A)$ denotes the non-neighborhood. Furthermore, by $G[A]$ we denote the subgraph of G induced on A . If $B \subset V(G)$ is a further set, then we let $e(A, B) = e_G(A, B)$ be the number of A - B -edges, i.e.,

$$e(A, B) = e_G(A, B) = \#\{\{v, w\} \in E(G) : v \in A, w \in B\}.$$

In addition, we let $e(A) = e_G(A) = e_G(A, A)$.

If $G = G_{n,p,k}^*$, then we let $G_0 = G_{n,p,k}$ denote the random graph from which G has been obtained via M3. Moreover, we let V_1, \dots, V_k denote the planted color classes of G and G_0 . If $U \subset \{1, \dots, k\}$, then we let $V_U = \bigcup_{u \in U} V_u$.

The scalar product of two vectors $\xi, \eta \in \mathbf{R}^n$ is denoted by $\langle \xi, \eta \rangle$. Moreover, $\|\xi\| = \langle \xi, \xi \rangle^{1/2}$ signifies the L^2 -norm. We let $\mathbf{1} = \mathbf{1}_n \in \mathbf{R}^n$ denote the vector with all entries equal to 1. In addition, if X a set and $A \subset X$, then $\mathbf{1}_A = (e_x)_{x \in X}$ denotes the vector with entries $e_x = 1$ if $x \in A$ and $e_x = 0$ if $x \in X \setminus A$. If $\xi \in \mathbf{R}^n$ is a vector, then $\text{diag}(\xi)$ signifies the $n \times n$ matrix with diagonal ξ whose off-diagonal entries are 0.

The eigenvalues of a real symmetric $n \times n$ matrix A are denoted by $\lambda_1(A) \geq \dots \geq \lambda_n(A)$. If A, B are symmetric $n \times n$ matrices, then we write $A \geq B$ if $\lambda_n(A - B) \geq 0$. Recall that A is positive semidefinite if $\lambda_n(A) \geq 0$, i.e., $A \geq 0$. Furthermore, for a $n \times n$ matrix M we let $\|M\| = \max_{\xi \in \mathbf{R}^n, \|\xi\|=1} \|M\xi\|$. In addition, $\text{diag}(M) \in \mathbf{R}^n$ is the vector consisting of the diagonal entries of M . By \mathbf{J} we denote a matrix with all entries equal to 1 (of any size). Moreover, $\mathbf{E} = \text{diag}(\mathbf{1})$ signifies the matrix with ones on the diagonal and off-diagonal entries equal to 0.

We shall mainly be interested in matrices associated with graphs. The adjacency matrix of a graph G is denoted by $A(G)$. In addition, $L(G) = \text{diag}(A(G)\mathbf{1}) - A(G)$ signifies the Laplacian.

A SDP relaxation of the chromatic number. The coloring heuristics rely on a semidefinite programming (“SDP”) relaxation ϑ_2 of the chromatic number. The semidefinite program was first defined by Goemans and Kleinberg [21] and was further studied by Charikar [9] and Szegedy [36]. Following Charikar, we define ϑ_2 in terms of vector colorings; this approach is related to the work of Karger, Motwani, and Sudan [26].

Let $G = (V, E)$ be a graph with vertex set $V = \{1, \dots, n\}$. Let (v_1, \dots, v_n) be an n -tuple of unit vectors in \mathbf{R}^n , and let $k > 1$. We call (v_1, \dots, v_n) a *rigid vector k -coloring* if

$$\langle v_i, v_j \rangle = (1 - k)^{-1} \quad \text{for all } \{i, j\} \in E, \text{ and } \langle v_i, v_j \rangle \geq (1 - k)^{-1} \quad \text{for all } \{i, j\} \notin E.$$

Now, we define $\bar{\vartheta}_2(G) = \inf\{k > 1 : G \text{ admits a rigid vector } k\text{-coloring}\}$. Since $\bar{\vartheta}_2(G)$ can be stated as a semidefinite program, the number $\bar{\vartheta}_2(G)$ and a rigid vector $\bar{\vartheta}_2(G)$ -coloring can be computed in polynomial time within a tiny numerical error, e.g., via the ellipsoid method (cf. [23, 36]).

Furthermore, we have $\bar{\vartheta}_2(G) \leq \chi(G)$. For assume that G is k -colorable, and let V_1, \dots, V_k be a partition of V into k independent sets. Moreover, let (ξ_1, \dots, ξ_k) be a family of unit vectors in \mathbf{R}^{k-1} such that $\langle \xi_i, \xi_j \rangle = -(k-1)^{-1}$ if $i \neq j$; such a family can be constructed inductively and it is unique up to an orthogonal transformation. Let $v_i = \xi_j$ for all $i \in V_j$. Then $(v_i)_{i \in V}$ is a rigid vector k -coloring of G , whence $\bar{\vartheta}_2(G) \leq k$. Indeed, $\bar{\vartheta}_2$ is a tighter relaxation of χ than both the vector chromatic number from [26] and the Lovász number $\vartheta(G)$ (cf. [23, 36]).

Let $A = A(G) = (a_{ij})_{i,j \in V}$ be the adjacency matrix of G . Moreover, let $L = L(G)$ signify the Laplacian. Let $k \geq 2$. In addition to $\bar{\vartheta}_2(G)$, we also need the following SDP from Frieze and Jerrum [19]:

$$\begin{aligned} \text{SDP}_k(G) = \max \quad & \sum_{1 \leq i < j \leq n} a_{ij} \cdot \frac{k-1}{k} (1 - \langle v_i, v_j \rangle) \\ \text{s.t.} \quad & \|v_i\| = 1 \text{ for } i = 1, \dots, n, \\ & \langle v_i, v_j \rangle \geq (1 - k)^{-1} \text{ for all } 1 \leq i < j \leq n, \\ & v_1, \dots, v_n \in \mathbf{R}^n. \end{aligned} \tag{7}$$

If k is an integer, then $\text{SDP}_k(G)$ is an upper bound on the weight of a MAX k -CUT of G . In particular, $\text{SDP}_2(G)$ equals the MAX CUT relaxation of Goemans and Williamson [22].

An important property of SDP_k is that the semidefinite program is *monotone*:

$$\text{if } G' \text{ contains } G \text{ as a subgraph, then } \text{SDP}_k(G') \geq \text{SDP}_k(G). \tag{8}$$

Furthermore, $\bar{\vartheta}_2(G)$ and $\text{SDP}_k(G)$ are related as follows: if G has a rigid vector k -coloring $(v_i)_{i \in V}$, then $(v_i)_{i \in V}$ is a feasible solution to SDP_k with objective function value

$$\sum_{1 \leq i < j \leq n} a_{ij} \cdot \frac{k-1}{k} (1 - \langle v_i, v_j \rangle) = \#E. \tag{9}$$

As trivially $\text{SDP}_k(G) \leq \#E$, we conclude that $\text{SDP}_k(G) = \#E$. Conversely, if $(v'_i)_{i \in V}$ is a feasible solution to $\text{SDP}_k(G)$ with objective function value $\#E$, then $(v'_i)_{i \in V}$ is a rigid vector k -coloring.

To prove Theorem 3, we need the following result, which is an immediate consequence of [13, Theorems 3 and 4].

Lemma 5. *There exist constants $\zeta_0, \zeta_1 > 0$ such that the following holds. Suppose that $np \geq \zeta_0$. Then for all $k \geq 2$ we have $\mathbb{P}[\text{SDP}_k(G_{n,p}) \leq (1 - k^{-1})\binom{n}{2}p + \zeta_1 n^{3/2} p^{1/2}] \geq 1 - \exp(-300n)$.*

Eigenvalues of random matrices. The proof of Theorem 1 relies on estimates of the eigenvalues of $A(G_{n,p,k})$. In order to estimate these eigenvalues, we employ the following two results.

Lemma 6. *Suppose that $np \geq c_1 \ln n$ for a constant $c_1 > 0$. Then there exists a number $c_2 > 0$ that depends only on c_1 such that with probability $\geq 1 - O(n^{-2}p^{-1})$ the random symmetric matrix $A = A(G_{n,p})$ enjoys the following property: $\forall \mathbf{1} \perp \eta \in \mathbf{R}^n : \|A\eta\| \leq c_2 \sqrt{np} \cdot \|\eta\|$.*

Lemma 7. *Suppose that $np \geq c_1 \ln n$ for a constant $c_1 > 0$. Then there exists a number $c_2 > 0$ that depends only on c_1 such that with probability $\geq 1 - O(n^{-2}p^{-1})$ the following holds. Let $A = (a_{ij})_{i,j=1,\dots,n}$ be a matrix whose entries are mutually independent random variables such that $a_{ii} = 0$ for all i and*

$$p = \mathbb{P}(a_{ij} = 1) = 1 - \mathbb{P}(a_{ij} = 0) \quad (i \neq j).$$

Then $\forall \mathbf{1} \perp \eta \in \mathbf{R}^n : \|A\eta\| \leq c_2 \sqrt{np} \cdot \|\eta\|$.

Lemma 6 is implicit in Feige and Ofek [18], and Lemma 7 in Alon and Kahale [2]. Explicit proofs of both lemmas can also be found in [12].

Chernoff bounds. Assume that X is binomially distributed with parameters (n, p) . Let $\mu = \mathbb{E}(X) = np$. We frequently need the following *Chernoff bounds* on the tails of X (cf. [25, Chap. 2] for proofs):

$$\mathbb{P}(X \geq \mu + t) \leq \exp\left(-\frac{t^2}{2(\mu + t/3)}\right), \quad \mathbb{P}(X \leq \mu - t) \leq \exp\left(-\frac{t^2}{2\mu}\right) \quad (0 < t). \quad (10)$$

Moreover, letting $\phi(x) = (1+x) \ln(1-x) - x$ for $x > -1$, we have

$$\mathbb{P}(X \leq \mu - t) \leq \exp\left(-\mu \phi\left(\frac{-t}{\mu}\right)\right) \quad (0 < t < \mu). \quad (11)$$

3 A Simple Heuristic for Finding an Optimal Coloring

3.1 Outline

We assume that (2) is satisfied with a sufficiently large constant $C_0 > 0$, which will be specified implicitly in the analysis. The algorithm `Color` for Theorem 1 is shown in Figure 1.

In summary, `Color`(G) computes the rigid vector coloring $(x_v)_{v \in V}$. This can be done in polynomial time via semidefinite programming (cf. Section 2). Then, `Color` constructs an auxiliary graph H in which two vertices v, w are adjacent iff their distance $\|x_v - x_w\|$ is at least 0.1, i.e., if x_v and x_w are “far apart”. To this graph H , `Color` applies the simple greedy coloring algorithm. (Recall that the greedy algorithm just goes through the vertices $v = 1, \dots, n$ and colors each v with the least color in $\{1, \dots, n\}$ not yet used by the neighbors of v .)

To show that `Color` either finds an optimal coloring of the input graph G or outputs “fail”, note that the graph H constructed in Step 2 contains G as a subgraph. For if $\{v, w\} \in E$, then $\langle x_v, x_w \rangle \leq 0$. Since $\chi(G) \geq \bar{\vartheta}_2(G)$, \mathcal{C} is an optimal coloring of G if \mathcal{C} uses at most $\lceil \bar{\vartheta}_2(G) \rceil$ colors.

Hence, to prove Theorem 1, it remains to show that `Color`($G = G_{n,p,k}^*$) outputs an optimal coloring w.h.p. Let V_1, \dots, V_k be the k -coloring planted in G . Directed by the proof that $\bar{\vartheta}_2(G) \leq \chi(G)$

Algorithm 8. COLOR(G)

Input: A graph $G = (V, E)$. *Output:* Either a $\chi(G)$ -coloring of G or “fail”.

1. Compute $\bar{\vartheta}_2(G)$ along with a rigid vector $\bar{\vartheta}_2(G)$ -coloring of G $(x_v)_{v \in V}$.
2. Let $H = (V, F)$ be the graph with edge set $F = \{\{v, w\} : \langle x_v, x_w \rangle \leq 0.995\}$. Apply the greedy coloring algorithm to H , and let \mathcal{C} be the resulting coloring.
3. If \mathcal{C} uses at most $\lceil \bar{\vartheta}_2(G) \rceil$ colors, then output \mathcal{C} as a coloring of G . Otherwise, output “fail”.

Fig. 1. the algorithm COLOR.

(cf. Section 2), we call a rigid vector k -coloring $(x_v)_{v \in V}$ *integral* if there are vectors $(x_i^*)_{i=1, \dots, k}$ such that $x_v = x_i^*$ for all $v \in V_i$, and $\langle x_i^*, x_j^* \rangle = (1 - k)^{-1}$ for $i \neq j$. In other words, $(x_v)_{v \in V}$ is integral iff the rigid vector coloring maps each color class onto a single point, and the angle between the points corresponding to V_i and V_j is $\cos^{-1} [(1 - k)^{-1}]$ if $i \neq j$.

If the rigid vector coloring $(x_v)_{v \in V}$ computed in Step 1 is integral, then the graph H constructed in Step 2 is a complete k -partite graph with color classes V_1, \dots, V_k . That is, in H the sets V_1, \dots, V_k are independent, but each $v \in V_i$ is connected with all vertices in $V \setminus V_i$. Consequently, the greedy algorithm finds a k -coloring of H . Hence, if also $\bar{\vartheta}_2(G) = k$, then COLOR finds and outputs an optimal coloring of G . Thus, the remaining task is to establish the following lemma.

Lemma 9. *Let $G = G_{n,p,k}^*$. W.h.p. we have $\bar{\vartheta}_2(G) = k$, and every rigid vector k -coloring of G is integral.*

To prove Lemma 9, we make use of the relationship between SDP_h and $\bar{\vartheta}_2$ (cf. Section 2). With respect to SDP_h , we prove the following in Section 3.2.

Lemma 10. *There is a constant $\zeta > 0$ such that $G = G_{n,p,k}^*$ enjoys the following property.*

$$\begin{aligned} & \text{Let } G' \text{ be a graph obtained by adding an edge } \{v^*, w^*\} \text{ to } G, \text{ where } v^*, w^* \in V_i \text{ for} \\ & \text{some } i. \text{ Let } 2 < h \leq k. \text{ Then } \text{SDP}_h(G') \leq \#E(G) - \zeta \cdot \frac{n^2 p}{hk} \cdot (k - h). \end{aligned} \quad (12)$$

Proof of Lemma 9. To prove that $\bar{\vartheta}_2(G_{n,p,k}^*) = k$ w.h.p., let $G = G_{n,p,k}^*$, and assume that $\bar{\vartheta}_2(G) = h < k$. Let $(x_v)_{v \in V}$ be a rigid vector h -coloring of G . Then $(x_v)_{v \in V}$ is a feasible solution to SDP_h , whence $\text{SDP}_h(G) = \#E(G)$ due to (9). However, by Lemma 10 and the monotonicity property (8) we have $\text{SDP}_h(G) < \#E(G)$ w.h.p. Thus, $\bar{\vartheta}_2(G) = k$ w.h.p.

Finally, to show that any rigid vector k -coloring $(x_v)_{v \in V}$ of $G = G_{n,p,k}^*$ is integral w.h.p., suppose that G has the property stated in Lemma 10. Let $s, t \in V_i^*$, and let G' be the graph obtained from G by adding the edge $\{s, t\}$. Then we have

$$\begin{aligned} \#E(G) &= \frac{k-1}{k} \left[\sum_{\{v,w\} \in E(G)} 1 - \langle x_v, x_w \rangle \right] \\ &\leq \frac{k-1}{k} \left[1 - \langle x_s, x_t \rangle + \sum_{\{v,w\} \in E(G)} 1 - \langle x_v, x_w \rangle \right] \leq \text{SDP}_k(G') \stackrel{\text{Lemma 10}}{\leq} \#E(G). \end{aligned}$$

Therefore, $\langle x_s, x_t \rangle = 1$, whence $x_s = x_t$, because x_s, x_t are unit vectors. Consequently, there are unit vectors x_i^* such that $x_v = x_i^*$ for all $v \in V_i$, $i = 1, \dots, k$.

Furthermore, if $i \neq j$, then $e_G(V_i, V_j)$ is binomially distributed with mean $n^2 k^{-2} p$. Hence, our assumption (2) and the Chernoff bound (10) entail that $e_G(V_i, V_j) > 0$ for all $i \neq j$ w.h.p. Thus, let $v \in V_i$, $w \in V_j$ be vertices such that $\{v, w\} \in E(G)$. Then $\langle x_i^*, x_j^* \rangle = \langle x_v, x_w \rangle = (1 - k)^{-1}$. Hence, the rigid vector coloring $(x_v)_{v \in V}$ is in fact integral. \square

3.2 Proof of Lemma 10

SDP duality provides a powerful tool for proving an upper bound on the optimal solution to a maximization problem such as SDP_h . Let $G = (V, E)$ be a graph. Then the *dual semidefinite program* of $\text{SDP}_h(G)$ reads

$$\begin{aligned} \text{DSDP}_h(G) = \min & \frac{h-1}{2h} \sum_{i=1}^n y_{ii} - \frac{1}{2h} \sum_{i \neq j} y_{ij} \\ \text{s.t. } & L(G) \leq Y, \\ & y_{ij} \leq 0 \text{ for } i \neq j, \\ & Y = (y_{ij})_{i,j=1,\dots,n} \text{ is a real symmetric } n \times n \text{ matrix} \end{aligned}$$

(cf. Helmberg [24, Chap. 2] for a thorough treatment of SDP duality theory). By weak SDP duality (cf. [24, pp. 17–18]), we have $\text{SDP}_h(G) \leq \text{DSDP}_h(G)$. Observe that the set of feasible solutions Y to $\text{DSDP}_h(G)$ is the same for all values of h .

To prove Lemma 10, we exhibit a feasible solution to $\text{DSDP}_h(G)$ for which the desired objective function value claimed in (12) is attained. The construction makes use of Lemmas 6 and 7. Let us first consider a random k -colorable graph $G = G_{n,p,k}$ with planted k -coloring V_1, \dots, V_k . As permuting the vertices does not affect the semidefinite program, we may assume that

$$V_i = \left\{ \frac{(i-1)n}{k} + 1, \dots, \frac{in}{k} \right\} \quad (i = 1, \dots, k). \quad (13)$$

Let G' be the graph obtained from G by adding an edge $\{v^*, w^*\}$, where $v^*, w^* \in V_{i_0}$ for some i_0 . Let $A = A(G)$, $L' = L(G')$, $B = L' - L(G)$. For $v \in V$, let d_v be the degree of v in G , and let

$$d_v^{(i)} = e_G(v, V_i), \quad \bar{d} = \frac{1}{(k-1)n} \sum_{v \in V} d_v.$$

Let $i(V_a) = (a-1)\frac{n}{k} + i$ denote the i 'th vertex in V_a . Moreover, set

$$d_{\min} = \min_{i,j,a \neq b} \frac{n \cdot d_{i(V_a)}^{(b)} \cdot d_{j(V_b)}^{(a)}}{k \cdot e_G(V_a, V_b)}. \quad (14)$$

Further, we define a family of $\frac{n}{k} \times \frac{n}{k}$ -matrices Y'_{ab} as follows: we let $Y'_{aa} = 0$ for $a = 1, \dots, k$, and for $1 \leq a, b \leq k$, $a \neq b$, we let

$$Y'_{ab} = \left[\frac{k}{n} d_{\min} - \frac{d_{i(V_a)}^{(b)} \cdot d_{j(V_b)}^{(a)}}{e_G(V_a, V_b)} \right]_{i,j=1,\dots,n/k}.$$

In addition, we let $Y' = (Y'_{ab})_{a,b=1,\dots,k}$ be the $n \times n$ matrix comprising of the blocks Y'_{ab} . Further, we let $y' = (d_v + d_{\min})_{v \in V} \in \mathbf{R}^n$, and finally $Y = Y' + \text{diag}(y')$. Then Y is a real symmetric $n \times n$ matrix, and the definition (14) of d_{\min} ensures that all off-diagonal entries of Y are ≤ 0 .

We claim that Y is a feasible solution to DSDP_h w.h.p. Thus, we need to show that $L' \leq Y$ w.h.p. Since $L' - Y = -(A - B + Y') - d_{\min} \mathbf{E}$, it suffices to prove that $\lambda_n(A - B + Y') = -d_{\min}$ w.h.p. As a first step, we shall exhibit a subspace $K \subset \mathbf{R}^n$ generated by eigenvectors of $A - B + Y'$ that correspond to the planted coloring V_1, \dots, V_k . To this end, we note that

$$Y_{ab} \mathbf{1} = \left[d_{\min} - \sum_{j=1}^{n/k} \frac{d_{i(V_a)}^{(b)} d_{j(V_b)}^{(a)}}{e_G(V_a, V_b)} \right]_{1 \leq i \leq n/k} = \left[d_{\min} - d_{i(V_a)}^{(b)} \right]_{1 \leq i \leq n/k} \quad (a \neq b),$$

because $\sum_{j=1}^{n/k} d_{j(V_b)}^{(a)} = e_G(V_a, V_b)$. Therefore, for $c = 1, \dots, k$ we have

$$Y' \mathbf{1}_{V_c} = [(1 - \delta_{v, V_c})(d_{\min} - e_G(v, V_c))]_{v \in V}. \quad (15)$$

Further,

$$A \mathbf{1}_{V_c} = (e_G(v, V_c))_{v \in V}, \text{ and } B \mathbf{1}_{V_c} = 0. \quad (16)$$

Combining (15) and (16), we get

$$(A - B + Y') \mathbf{1}_{V_c} = [(1 - \delta_{v, V_c}) d_{\min}]_{v \in V}. \quad (17)$$

Finally, we let $\xi^{(a,b)} = \mathbf{1}_{V_a} - \mathbf{1}_{V_b} \in \mathbf{R}^n$ ($a, b = 1, \dots, k$). Then (17) yields

$$(A - B + Y') \xi^{(a,b)} = -d_{\min} \xi^{(a,b)} \quad (a \neq b), \quad (A - B + Y') \mathbf{1} = (k-1) d_{\min} \mathbf{1}. \quad (18)$$

Let $K \subset \mathbf{R}^n$ be the vector space spanned by $\mathbf{1}$ and the vectors $\xi^{(a,b)}$ ($a \neq b$). Then $\mathbf{1}_{V_1}, \dots, \mathbf{1}_{V_k} \in K$, and therefore $\mathbf{1}_{V_1}, \dots, \mathbf{1}_{V_k}$ generate K . Since by (18) K is generated by eigenvectors of $A - B + Y'$, any eigenvector η of $A - B + Y'$ with eigenvalue $< -d_{\min}$ is perpendicular to K . Thus, the following lemma shows that no eigenvector with eigenvalue $< -d_{\min}$ exists w.h.p., and hence concludes the proof that $L' \leq Y$ w.h.p.

Lemma 11. *Let $G = G_{n,p,k}$. Then w.h.p. we have $d_{\min} = \Omega(np/k)$ and $|\langle (A - B + Y')\eta, \eta \rangle| < d_{\min}$ for all unit vectors $\eta \perp K$ and all possible choices of v^*, w^* .*

We prove Lemma 11 in Section 3.3. Now, suppose that indeed $L' \leq Y$. Then by construction

$$\sum_{i=1}^n y_{ii} = \langle y', \mathbf{1} \rangle = 2\#E(G) + nd_{\min}, \quad (19)$$

$$\sum_{i \neq j} y_{ij} = \sum_{a \neq b} \langle Y'_{ab} \mathbf{1}, \mathbf{1} \rangle \stackrel{(15)}{=} \sum_{a \neq b} \frac{n}{k} - e_G(V_a, V_b) = (k-1)nd_{\min} - 2\#E(G). \quad (20)$$

Combining (19) and (20), we obtain

$$\text{DSDP}_h(G') \leq \frac{h-1}{2h} \sum_{i=1}^n y_{ii} - \frac{1}{2h} \sum_{i \neq j} y_{ij} = \#E(G) - \frac{nd_{\min}}{2h}(k-h).$$

As $d_{\min} = \Omega(np/k)$ w.h.p. by Lemma 11, we conclude that Lemma 10 holds for $G = G_{n,p,k}$.

Finally, let $G = G_{n,p,k}^*$, and let G_0 be the random k -colorable graph contained in G (i.e., $G \in \mathcal{I}(G_0)$). Let G'_0 (resp. G') be obtained from G_0 (resp. from G) by adding an edge $\{v^*, w^*\}$, $v^*, w^* \in V_i$. Since adding a single edge can increase the value of SDP_h by at most 1, w.h.p. we have

$$\text{SDP}_h(G') \leq \text{SDP}_h(G'_0) + \#E(G) - \#E(G_0) \leq \#E(G) - \Omega\left(\frac{np}{2hk}\right)(k-h),$$

as desired.

3.3 Proof of Lemma 11

To prove the lemma, we decompose the adjacency matrix A of $G = G_{n,p,k}$ into blocks $A = (A_{ab})_{a,b=1,\dots,k}$ of size $\frac{n}{k} \times \frac{n}{k}$. Then due to our assumption (13) for any two vertices $v \in V_a$ and $w \in V_b$, the $(v - (a-1)\frac{n}{k}, w - (b-1)\frac{n}{k})$ -entry of A_{ab} is 1 if $\{v, w\} \in E(G)$ and 0 if $\{v, w\} \notin E(G)$. In particular, $A_{ab}^T = A_{ba}$ and $A_{aa} = 0$. Moreover, the entries of each block A_{ab} with $a \neq b$ are mutually independent random variables that attain the value 1 with probability p and the value 0 with probability $1 - p$.

Lemma 12. *If $G = G_{n,p,k}$, then w.h.p. the following statements hold.*

1. $\bar{d} = \Theta(np/k)$.
2. For all unit vectors $\eta \perp K$ we have $|\langle A\eta, \eta \rangle| \leq O(\sqrt{\bar{d}k})$.
3. For all $a, b \in \{1, \dots, k\}$ and all $\mathbf{1} \perp \xi \in \mathbf{R}^{n/k}$ we have $|\|\mathbf{1}\|^{-1} \cdot \langle A_{ab}\xi, \mathbf{1} \rangle| \leq O(\sqrt{\bar{d}})$.

Proof. The first statement is an immediate consequence of the definition of \bar{d} and the Chernoff bound (10). Moreover, the third statement is an immediate consequence of Lemma 7, because $np/k = \Omega(\ln n)$ by (2).

To prove the second statement, let $G_a = G_{n/k,p}$ for $a = 1, \dots, k$ be a family of k mutually independent random graphs. Let $A_a^* = A(G_a)$ be the adjacency matrices ($a = 1, \dots, k$). Moreover, let

$$A_* = \begin{pmatrix} A_1^* & 0 \\ & \ddots \\ 0 & A_k^* \end{pmatrix}$$

be the $n \times n$ matrix with the $\frac{n}{k} \times \frac{n}{k}$ -blocks A_1^*, \dots, A_k^* on the diagonal and 0's elsewhere. In addition, set $A^* = A + A_*$. Then A^* is distributed as the adjacency matrix $A(G_{n,p})$ of a random graph $G_{n,p}$. Observe that $\eta \perp K$ implies that $\eta \perp \mathbf{1}$. Thus, since $np = \Omega(\ln n)$ by (2), Lemma 6 entails that there is a constant $\zeta_1 > 0$ such that w.h.p.

$$\forall \eta \perp K, \|\eta\| = 1 : |\langle A^*\eta, \eta \rangle| \leq \zeta_1 \sqrt{\bar{d}k}. \quad (21)$$

Furthermore, decomposing $\eta \perp K$ into k subsequent pieces $\eta_1, \dots, \eta_k \in \mathbf{R}^{n/k}$, we obtain

$$\begin{aligned} |\langle A_*\eta, \eta \rangle| &= \left| \sum_{a=1}^k \langle A_a^*\eta_a, \eta_a \rangle \right| \leq \sum_{a:\eta_a \neq 0} \|\eta_a\|^2 \cdot \left| \left\langle A_a^* \frac{\eta_a}{\|\eta_a\|}, \frac{\eta_a}{\|\eta_a\|} \right\rangle \right| \\ &\leq \|\eta\|^2 \cdot \max_{a:\eta_a \neq 0} \left| \left\langle A_a^* \frac{\eta_a}{\|\eta_a\|}, \frac{\eta_a}{\|\eta_a\|} \right\rangle \right|. \end{aligned} \quad (22)$$

If $\eta \perp K$, then $\eta_a \perp \mathbf{1}$ for $a = 1, \dots, k$. Therefore, as $A_a^* = A(G_{n/k,p})$ and $np/k = \Omega(\ln n)$, by Lemma 6 there is a constant $\zeta_2 > 0$ such that w.h.p.

$$\forall \eta \perp K, \|\eta\| = 1 : |\langle A_*\eta, \eta \rangle| \leq \max_{a:\eta_a \neq 0} \left| \left\langle A_a^* \frac{\eta_a}{\|\eta_a\|}, \frac{\eta_a}{\|\eta_a\|} \right\rangle \right| \leq \zeta_2 \sqrt{\bar{d}}. \quad (23)$$

Finally, we claim that w.h.p.

$$\forall \eta \perp K, \|\eta\| = 1 : |\langle A\eta, \eta \rangle| \leq (\zeta_1 + \zeta_2) \sqrt{\bar{d}k}. \quad (24)$$

Indeed, suppose that A violates (24). Then there is a unit vector $\eta \perp K$ such that $|\langle A\eta, \eta \rangle| > (\zeta_1 + \zeta_2) \sqrt{\bar{d}k}$. Hence, for all A_* that satisfy (23) we have $|\langle A^*\eta, \eta \rangle| \geq |\langle A\eta, \eta \rangle| - |\langle A_*\eta, \eta \rangle| > \zeta_1 \sqrt{\bar{d}k}$, so that A^* violates (21). Since (21) and (23) hold with probability $1 - o(1)$, we conclude that the probability that (24) is violated is $o(1)$, as desired. \square

Proof of Lemma 11. Since $np \geq (1 + \varepsilon)k \ln(n)$, the fact that $d_{\min} = \Omega(\bar{d})$ w.h.p. follows from the Chernoff bound (11). Furthermore, we claim that

$$|\langle Y'\eta, \eta \rangle| \leq O(\sqrt{\bar{d}k}) \quad \text{w.h.p.} \quad (25)$$

Indeed, consider the following $\frac{n}{k} \times \frac{n}{k}$ matrices Z_{ab} ($a, b = 1, \dots, k$): we let $Z_{aa} = 0$ for all a , and $Z_{ab} = \frac{k}{n} d_{\min} \mathbf{J} - Y'_{ab}$ ($a \neq b$). Moreover, let $Z = (Z_{ab})_{a,b=1,\dots,k}$ be the $n \times n$ matrix consisting of the blocks Z_{ab} . Then for all $\eta \perp K$ we have

$$\langle Z\eta, \eta \rangle = -\langle Y'\eta, \eta \rangle, \quad (26)$$

because $\eta \perp \mathbf{1}_{V_a}$ for all $a \in \{1, \dots, k\}$. Thus, it suffices to estimate $|\langle Z\eta, \eta \rangle|$. Let $\xi = (\xi_i)_{1 \leq i \leq n/k}$, $\eta = (\eta_i)_{1 \leq i \leq n/k} \in \mathbf{R}^{n/k}$ be unit vectors perpendicular to $\mathbf{1}$. Then

$$\begin{aligned} e_G(V_a, V_b) \langle \eta, Z_{ab}\xi \rangle &= \left\langle \eta, \left[\sum_{j=1}^{n/k} d_{i(V_a)}^{(b)} d_{j(V_b)}^{(a)} \xi_j \right]_{1 \leq i \leq n/k} \right\rangle = \left\langle \eta, \left[d_{i(V_a)}^{(b)} \langle A_{ba}\mathbf{1}, \xi \rangle \right]_{1 \leq i \leq n/k} \right\rangle \\ &= \langle A_{ba}\mathbf{1}, \xi \rangle \sum_{i=1}^{n/k} d_{i(V_a)}^{(b)} \eta_i = \langle A_{ba}\mathbf{1}, \xi \rangle \langle A_{ab}\mathbf{1}, \eta \rangle = \langle A_{ab}\xi, \mathbf{1} \rangle \langle A_{ba}\eta, \mathbf{1} \rangle. \end{aligned}$$

By the third part of Lemma 12, w.h.p. we have

$$|\langle A_{ab}\xi, \mathbf{1} \rangle|, |\langle A_{ba}\eta, \mathbf{1} \rangle| \leq O\left(\sqrt{\bar{d}n/k}\right) \quad (27)$$

for all unit vectors $\xi, \eta \perp \mathbf{1}$ and all a, b . Moreover, since $e_G(V_a, V_b)$ is binomially distributed with mean $n^2 k^{-2} p$, the Chernoff bound (10) and the first part of Lemma 12 entail that w.h.p.

$$e_G(V_a, V_b) = \Omega(\bar{d}n/k) \quad (1 \leq a < b \leq k). \quad (28)$$

Combining (27)–(28), we get

$$|\langle Z_{ab}\xi, \eta \rangle| \leq \frac{O(\bar{d}n/k)}{e_G(V_a, V_b)} = O(1) \quad \text{w.h.p.} \quad (1 \leq a < b \leq k). \quad (29)$$

Thus, let $\eta \perp K$ be a unit vector. Decomposing η into k pieces $\eta_1, \dots, \eta_k \in \mathbf{R}^{n/k}$, we get

$$\begin{aligned} |\langle Z\eta, \eta \rangle| &= \left| \sum_{a,b=1}^k \langle Z_{ab}\eta_b, \eta_a \rangle \right| \leq \sum_{a,b:\eta_a \neq 0 \neq \eta_b} \|\eta_a\| \cdot \|\eta_b\| \cdot \left| \left\langle Z_{ab} \frac{\eta_b}{\|\eta_b\|}, \frac{\eta_a}{\|\eta_a\|} \right\rangle \right| \\ &\stackrel{(29)}{\leq} O \left[\sum_{a,b=1}^k \|\eta_a\| \cdot \|\eta_b\| \right] \leq O(1) \left[\sum_{a=1}^k \|\eta_a\| \right]^2 = O(k) \stackrel{(2)}{\leq} O\left(\sqrt{\bar{d}k}\right). \end{aligned} \quad (30)$$

Combining (26) and (30), we obtain (25).

As $\|B\| \leq 2$, the second part of Lemma 12 in combination with (25) yields that w.h.p.

$$\forall \eta \perp K, \|\eta\| = 1 : |\langle (A - B + Y')\eta, \eta \rangle| \leq O\left(\sqrt{\bar{d}k}\right). \quad (31)$$

As $d_{\min} = \Omega(\bar{d})$ and $\bar{d} = \Theta(np/k)$, (2) and (31) give

$$\forall \eta \perp K, \|\eta\| = 1 : |\langle (A - B + Y')\eta, \eta \rangle| < d_{\min},$$

provided that the constant C_0 is sufficiently large. \square

4 Coloring $G_{n,p,k}^*$ in Polynomial Expected Time

In this section we present the algorithm `ExpColor` for Theorem 3. After exhibiting a few properties of $G_{n,p,k}^*$ in Section 4.1, we outline the algorithm `ExpColor` and its subroutines in Section 4.2. Sections 4.3–4.5 contain the technical details of the analysis of `ExpColor`.

Throughout, we $G = G_{n,p,k}^*$. We assume that (4) is satisfied with a sufficiently large constant C_0 .

4.1 Basic Properties of $G_{n,p,k}^*$

Let $U \subset \{1, \dots, k\}$, and consider the graph $G[V_U]$ induced on the color classes V_i with $i \in U$. Let $u = \#U$. Then $G[V_U]$ is u -colorable, so that $\text{SDP}_u(G[V_U]) = \#E(G[V_U])$ (cf. (9)). Now let G' be a graph obtained from G by adding random edges *inside* the color classes V_i , $i \in U$. The following lemma, which we prove in Section 7.1, shows that these additional random edges do not increase the value of SDP_u “too much”. More precisely, we have $\text{SDP}_u(G') - \text{SDP}_u(G) = O(\frac{nu}{k}\sqrt{np})$; note that by (4) $\#E(G[V_U])$ is binomially distributed with mean $\binom{u}{2}n^2k^{-2}p \gg \frac{nu}{k}\sqrt{np}$.

Lemma 13. *Let $U \subset \{1, \dots, k\}$ be a set of cardinality $u = \#U$. With probability $\geq 1 - \exp(-100nu/k)$ the graph $G = G_{n,p,k}^*$ enjoys the following property.*

Let G' be a graph obtained from G by adding each edge inside the color classes V_i with probability p independently. Then for a certain constant $C_1 > 0$ we have

$$\mathbb{P} \left[\text{SDP}_u(G'[V_U]) \leq \#E(G[V_U]) + C_1 \frac{nu}{k} \sqrt{np} \right] \geq 2/3, \quad (32)$$

where probability is taken over the choice of the random edges inside the color classes.

We will use Lemma 13 in Section 4.2 in order to investigate the geometric structure of rigid vector colorings of G .

Now consider a single color class V_i . The subgraph of $G = G_{n,p,k}^*$ consisting only of the V_i - $V \setminus V_i$ -edges is a random bipartite graph. Hence, we expect that this bipartite graph is a good “expanding graph”. To quantify the expansion property of this graph precisely, we need the following concept. Let $T \subset V \setminus V_i$, and let $\eta \geq 0$. A set M of T - V_i -edges of G is a d -fold matching with defect $\leq \eta$ from T to V_i if there exists a set $D \subset T$, $\#D \leq \eta$, such that

- every vertex in $T \setminus D$ is incident with precisely d edges in M , and
- every vertex in V_i is incident with at most one edge in M .

Now, we define the *defect* $\text{def}_G(V_i)$ as follows (cf. [11, Section 2.3]).

- D1.** If there is a subset $U \subset V_i$ of cardinality $\#U \geq \frac{n}{2k}$ such that $\#V \setminus (V_i \cup N_G(U)) > \frac{n}{200k^2}$, then we let $\text{def}_G(V_i) = \frac{n}{2k}$.
- D2.** Otherwise, we let $\text{def}_G(V_i)$ be the least number $0 \leq \eta \leq \frac{n}{2k}$ such that for all $6 \leq d \leq \lceil 50k \rceil$ the following holds: every set $T \subset V \setminus V_i$ of size $\#T \leq \frac{n}{2dk}$ admits a d -fold matching to V_i with defect $\leq \eta$.

Then $\text{def}_G(V_i)$ quantifies the expansion of the bipartite graph consisting of the V_i - $V \setminus V_i$ -edges of G : the smaller the defect is, the better is the expansion.

The following lemma bounds the probability that the defect gets large.

Lemma 14. *Let $\eta_i \geq 0$ for $i = 1, \dots, k$. Then $\mathbb{P}(\text{def}_G(V_i) \geq \eta_i \text{ for } i = 1, \dots, k) \leq \prod_{i=1}^k \left(\frac{n/k}{\eta_i}\right)^{-100}$.*

We prove Lemma 14 in Section 7.2. Furthermore, in Section 7.3 we prove the following lemma, which shows that every sufficiently large independent set of $G_{n,p,k}^*$ consists mainly of vertices from one color class w.h.p.

Lemma 15. *With probability $\geq 1 - \exp(-100n)$ the semirandom graph $G = G_{n,p,k}^*$ enjoys the following property.*

If U is an independent set in G of size $\#U \geq \frac{n}{100k}$, then $\#U \cap V_i > \frac{199}{200}\#U$ for some $1 \leq i \leq k$. (33)

Moreover, with probability $\geq 1 - \exp(-100n/\ln k)$ the graph $G = G_{n,p,k}^$ satisfies the following condition for all $i \in \{1, \dots, k\}$.*

If $U \subset V_i$, $\#U \geq \frac{n}{2k \ln(k)}$, then $\#\bar{N}_G(U) \leq \frac{2n}{k}$. (34)

4.2 Outline

In order to k -color G , $\text{ExpColor}(G, k)$ (cf. Figure 2) runs the procedure Classes , which proceeds recursively in k stages. In each stage, Classes tries to recover one of the color classes V_1, \dots, V_k , and then hands the graph without the recovered color class to the next stage. More precisely, if W_l is the set of vertices that have not yet been colored in the previous stages, then the l 'th stage tries to exhibit a set \mathcal{S}_l of large independent sets of $G[W_l]$. Then for each $S_l \in \mathcal{S}_l$, Classes passes the graph $G[W_l \setminus S_l]$ to stage $l + 1$, which tries to find a $(k - l)$ -coloring of this graph. If G is “typical”, which happens with high probability, then each S_l will consist precisely of one color class, so that a k -coloring will be found immediately.

Algorithm 16. $\text{ExpColor}(G, k)$

Input: A graph $G = (V, E)$, an integer $k \geq \chi(G)$. *Output:* A k -coloring of G .

1. For $T = 1, \dots, \lfloor \exp(n / \ln k) \rfloor$ do
2. Let $\eta = \max\{\xi \in \mathbf{Z} : \exp(\xi), \binom{n/k}{\xi} \leq T\}$.
For each decomposition $\eta = \eta_1 + \dots + \eta_k$ where $0 \leq \eta_i \leq \frac{n}{2k}$ are integers such that $\prod_{i=1}^k \binom{n/k}{\eta_i} \leq T$ do
3. If $\text{Classes}(G, V, k, \eta_1, \dots, \eta_k)$ k -colors G , then output the coloring and halt.
4. For $T = \lceil \exp(n / \ln k) \rceil, \dots, \lfloor \exp(n) \rfloor$ do
If $\text{Exact}(G, k, T)$ k -colors G , then output the resulting coloring and halt.
5. Color G optimally via Lawler's algorithm [30] in time $O(2.443^n)$.

Fig. 2. the algorithm ExpColor .

However, since our goal is an algorithm that k -colors *all* k -colorable graphs, we also have to deal with “atypical” input instances G . To this end, ExpColor uses the variable T , which controls the size of the “search tree” that ExpColor is building, i.e., what amount of running time ExpColor spends in order to k -color G . This amount of time is distributed among the k stages of Classes via the variables η_1, \dots, η_k . The variable η_{k-l+1} determines for how “typical” the l 'th stage takes its input graph: the larger η_{k-l+1} , the less “typical” the graph is assumed to be. In order to (try to) produce a set \mathcal{S}_l that contains one of the hidden color classes, stage l of ExpColor may spend time $(n \binom{n/k}{\eta_{k-l+1}})^{O(1)}$. Thus, as the variable T grows from 1 to $\exp(n)$, the running time increases “smoothly” from polynomial to exponential.

In addition to Classes , ExpColor has a further subroutine Exact . This procedure is used as a fallback if Classes does not k -color G before T exceeds $\exp(n / \ln k)$.

In order to analyze ExpColor , we shall assign to each graph $G = G_{n,p,k}^*$ a value T^* such that ExpColor k -colors G before T exceeds T^* . Then, on the one hand we can bound the running time of $\text{ExpColor}(G, k)$ in terms of T^* . On the other hand, we shall investigate the distribution of T^* to prove that the *expected* running time is polynomial.

The procedure Classes . The input of Classes (cf. Figure 3) consists of the graph G , a set $W \subset V(G)$, the number k , and integers η_1, \dots, η_l . Classes is to find an l -coloring of $G[W]$. In Steps 1–3, Classes computes a set \mathcal{S}_l of independent sets of $G_l = G[W]$, each of cardinality n/k . Then, in Steps 4–5, Classes tentatively colors each of the sets $S_l \in \mathcal{S}_l$ with the l 'th color, and calls itself recursively on input $(G, W \setminus S_l, k, \eta_1, \dots, \eta_{l-1})$ in an attempt to $(l - 1)$ -color $G[W \setminus S_l]$.

Suppose that the input graph G is a semirandom graph $G_{n,p,k}^*$ with hidden coloring V_1, \dots, V_k . Similarly as the heuristic Color in Section 3, Classes employs the relaxation $\bar{\nu}_2$ of the chromatic number (cf. Section 2 for the definition), but in a more sophisticated way. If $\eta_l < \frac{n}{2k}$, then Step 2 of Classes tries to use the rigid vector coloring $(x_v)_{v \in W}$ to recover a large independent set S_v (cf. Lemma 19 be-

Algorithm 17. `Classes`($G, W, k, \eta_1, \dots, \eta_l$)

Input: A graph $G = (V, E)$, a set $W \subset V$, integers k, η_1, \dots, η_l .

Output: Either an l -coloring of $G[W]$ or “fail”.

1. Let $G_l = G[W]$.
If $l = 1$ and G_l is an independent set, then return a 1-coloring of G_l .
If $\bar{\vartheta}_2(G_l) > l$, then return “fail”.
Otherwise, compute a rigid vector l -coloring $(x_v)_{v \in W}$ of G_l .
2. If $\eta_l < \frac{n}{2k}$
If for all $w \in W$ the set $S_w = \{u \in W : \langle x_u, x_w \rangle \geq 0.99\}$ has cardinality $< \frac{199n}{200k}$, return “fail”.
Otherwise, let $v = \min \{w \in W : \#S_w \geq \frac{199n}{200k}\}$.
Let $S_l = \text{Purify}(G, S_v, \eta_l, n/k)$.
3. else
Let $S_l = \emptyset$. For each $U \subset W$, $\#U = \frac{n}{2k \ln(k)}$, do
Let $T = \bar{N}_{G_l}(U)$.
If $\#T \leq 2n/k$, then for all $I \subset T$, $\#I = n/k$, do
If I is an independent set, then add I to S_l .
4. For each $S_l \in \mathcal{S}_l$ do
5. If `Classes`($G, W \setminus S_l, k, \eta_1, \dots, \eta_{l-1}$) $(l-1)$ -colors $G[W \setminus S_l]$, return the l -coloring of G_l obtained by coloring S_l with an l 'th color.
6. Return “fail”.

Fig. 3. the procedure `Classes`.

low). By Lemma 15, with extremely high probability S_v consists mainly of vertices of one color class V_i . Then, to recover V_i from S_v , `Classes` uses a further procedure `Purify` (cf. Corollary 20 below).

However, if $\eta_l \geq \frac{n}{2k}$, then Step 3 of `Classes` assumes that $\bar{\vartheta}_2$ behaves “badly”, so that the aforementioned approach is hopeless. Instead, Step 3 enumerates all subsets U of W of cardinality $\frac{n}{k \ln k}$ and considers their non-neighborhoods. Eventually, Step 3 will encounter a set U that lies entirely inside a color class V_i . By the second part of Lemma 15, we expect that $\#\bar{N}_{G_l}(U) \leq \frac{2n}{k}$. If so, Step 3 adds all independent subsets of $\bar{N}_{G_l}(U)$ of cardinality $\frac{n}{k}$ to S_l . Thus, as U is contained in the independent set V_i , we have $V_i \subset \bar{N}_{G_l}(U)$, so that the color class V_i will be added to S_l . The following proposition summarizes the analysis of `Classes`.

Proposition 18. *To each semirandom graph $G = G_{n,p,k}^*$ that satisfies Properties (33) and (34) we can associate a sequence $(\eta_1^*, \dots, \eta_k^*) \in \{0, 1, \dots, \frac{n}{2k}\}^k$ such that the following two conditions hold.*

1. `Classes`($G, V, k, \eta_1^*, \dots, \eta_k^*$) outputs a k -coloring of G .
2. Let $\eta_1, \dots, \eta_k \geq 0$. Then $\mathbb{P}(\eta_i^* \geq \eta_i \text{ for all } i) \leq \prod_{i=1}^k \binom{n/k}{\eta_i}^{-90}$.

The running time of `Classes`($G, V, k, \eta_1, \dots, \eta_k$) is at most $n^{O(1)} \prod_{i=1}^k \binom{n/k}{\eta_i}^{14}$.

The crucial insight behind `Classes` is that w.h.p. we can use the rigid vector coloring to recover a “large” independent set of size $\frac{199n}{200k}$ (cf. Step 2). By Lemma 15, such an independent set will consist mainly of vertices from one of the planted color classes, i.e., in the case of success we have recovered a huge fraction of one color class. In order to extract a large independent set from the vector coloring, the basic idea is as follows. Imagine that we would throw random edges into the color classes of $G = G_{n,p,k}^*$ by including the edges inside the color classes V_i with probability p independently. (Of course, the algorithm can’t do this, because it does not know the color classes yet.) Let G' be the resulting graph. How do $\text{SDP}_k(G)$ and $\text{SDP}_k(G')$ compare? By Lemma 13, with probability $\geq \frac{2}{3}$ over the choice of the random edges inserted into the color classes, $\text{SDP}_k(G')$ exceeds $\text{SDP}_k(G) = \#E(G)$ by at most $O(n^{3/2}p^{1/2})$. Hence, if $(x_v)_{v \in V}$ is a rigid vector k -coloring of G , then there are only $O(n^{3/2}p^{1/2})$

random edges $\{v, w\}$ inside the color classes V_i whose contribution $1 - \langle x_v, x_w \rangle$ to the sum (cf. (9))

$$\text{SDP}_k(G) \leq \left(1 - \frac{1}{k}\right) \sum_{\{s,t\} \in E(G')} 1 - \langle x_s, x_t \rangle \leq \text{SDP}_k(G')$$

is “large”, say, $1 - \langle x_v, x_w \rangle \geq 1/200$. But then we can derive from our assumption (4) that there is at least one color class such that for almost all vertices v, w in this class the vectors x_v, x_w are “close to each other”, say, $\langle x_v, x_w \rangle \geq 0.99$. In fact, these vertices can be found easily by “guessing” one of them, say w , and considering all the vertices that are close to it, i.e., the set S_w . The following lemma makes this idea rigorous.

Lemma 19. *Let $G = G_{n,p,k}^*$. Assume that Property (32) holds for the set $U \subset \{1, \dots, k\}$, $\#U = u > 1$. Let $(x_v)_{v \in V_U}$ be a rigid vector u -coloring of $G[V_U]$. Then there is a vertex $v \in V_U$ such that $S_v = \{w \in V_U : \langle x_v, x_w \rangle \geq 0.99\}$ is an independent set of cardinality $\geq \frac{199n}{200k}$ in G .*

Proof. Consider the graph $H = (V_U, F)$, where $F = \{\{v, w\} : \langle x_v, x_w \rangle < 0.99\}$. Then $G[V_U]$ is a subgraph of H , because $\langle x_v, x_w \rangle < 0$ for all edges $\{v, w\} \in E(G)$. Let $\mathcal{B} = \bigcup_{i \in U} E(H[V_i])$ be the set of all edges of H that join two vertices that belong to the same color class of G . Let $b = \#\mathcal{B}$.

Furthermore, let G' be the random graph obtained from G by including each V_i - V_i -edge with probability p independently for all $i \in \{1, \dots, k\}$. Note that $(x_v)_{v \in V}$ is a feasible solution to SDP_u . Hence, by Property (32), with probability $\geq 2/3$ over the choice of the random edges inside the color classes we have

$$\sum_{\{v,w\} \in E(G'[V_U])} \frac{u-1}{u} (1 - \langle x_v, x_w \rangle) \leq \text{SDP}_u(G'[V_U]) \leq \#E(G[V_U]) + C_1 \frac{nu}{k} \sqrt{np}. \quad (35)$$

Observe that an edge $e = \{v, w\}$ of $G'[V_U]$ contributes 1 to the sum on the left hand side if $e \in E(G)$, and that e contributes $\geq \frac{1}{200}$ if $e \in \mathcal{B}$. Therefore, (35) entails that

$$\text{P} \left(\#\mathcal{B} \cap E(G'[V_U]) \leq 200C_1 \frac{nu}{k} \sqrt{np} \right) \geq \frac{2}{3}. \quad (36)$$

We claim that $b \leq \frac{u}{401} n^2 k^{-2}$. Indeed, assume for contradiction that $b > \frac{u}{401} n^2 k^{-2}$. Then (4) yields that $bp > 2000C_1 nuk^{-1} \sqrt{np}$, provided that the constant C_0 is large enough. Since $\#\mathcal{B} \cap E(G'[V_U])$ is binomially distributed with mean bp , by the Chernoff bound (10) we obtain

$$\text{P} \left(\#\mathcal{B} \cap E(G'[V_U]) > 200C_1 \frac{nu}{k} \sqrt{np} \right) \geq \text{P} \left(\#\mathcal{B} \cap E(G'[V_U]) \geq \frac{bp}{10} \right) \geq \frac{1}{2},$$

contradicting (36).

Thus, $b \leq \frac{u}{401} n^2 k^{-2}$. Consequently, there is some $i \in U$ and a vertex $v \in V_i$ such that v has degree $< \frac{n}{200k}$ in $H[V_i]$. Hence, $S_v = \bar{N}_H(v)$ has size $\#S_v \geq \frac{199n}{200k}$. Furthermore, as for all $w, w' \in S_v$ we have $\langle x_v, x_w \rangle, \langle x_v, x_{w'} \rangle \geq 0.99$, we obtain that $\langle x_w, x_{w'} \rangle \geq 0$. Therefore, $\{w, w'\} \notin E(G)$, so that S_v is an independent set in G . \square

In addition to the relaxation $\bar{\vartheta}_2$, Step 2 of `Classes` employs a procedure `Purify` from [11]. The following corollary is a reformulation of [11, Proposition 2.6] for the present setting.

Corollary 20. *Let $G = G_{n,p,k}^*$. Let $i \in \{1, \dots, k\}$. Suppose that I is an independent set that satisfies $\#I \cap V_i \geq \frac{99n}{100k}$. Further, assume that $\text{def}_G(V_i) \leq \eta < \frac{n}{2k}$. Then the output \mathcal{S} of `Purify`($G, I, \eta, n/k$) contains V_i as an element. Finally, the running time of `Purify`($G, I, \eta, n/k$) is $\leq n^{O(1)} \binom{n/k}{\eta}^{14}$.*

Thus, suppose that Step 2 of `Classes` recovers a large independent set S_v from the rigid vector coloring such that $\#S_v$ already contains 99% of the vertices of some color class V_i . Then Corollary 20 entails that `Purify` will in fact recover the actual class V_i , provided that the parameter η exceeds the defect $\text{def}_G(V_i)$. Combining Lemma 19 and Corollary 20, we prove Proposition 18 in Section 4.3.

The Procedure `Exact` If `Classes` fails to k -color the input graph $G = G_{n,p,k}^*$, then `ExpColor` calls the procedure `Exact` (cf. Figure 4). The goal of `Exact` is to exhibit a k -coloring of G in expected time $n^{O(1)} \exp(n/\ln k)$. Thus, the expected running time of `Exact` is somewhat smaller than the worst case running time $\exp(\Theta(n))$ of known exact coloring algorithms. In Section 4.4 we prove the following proposition.

Proposition 21. *Let $G = (V, E)$ be a k -colorable graph, and let $\exp(n/\ln k) \leq T \leq \exp(n)$.*

1. *`Exact`(G, k, T) either outputs a k -coloring of G or “fail”.*
2. *The running time of `Exact`(G, k, T) is $\leq n^{O(1)} T^4$.*
3. *If $G = G_{n,p,k}^*$, then the probability that `Exact`($G = G_{n,p,k}^*, k, T$) answers “fail” is $\leq T^{-90}$.*

Algorithm 22. `Exact`(G, k, T)

Input: A graph $G = (V, E)$, an integer $k \geq \chi(G)$, an integer $T \geq 0$.

Output: Either a k -coloring of G or “fail”.

1. Let $0 \leq x \leq n$ be the largest integer such that $\binom{n}{x} k^x k! \leq T$.
For each triple (X, φ, σ) , where
 - $X \subset V$, $\#X = x$,
 - φ is a k -coloring of $G[X]$,
 - σ is a permutation of $\{1, \dots, k\}$
- do the following.
 2. For $l = 1, \dots, k$ do
 - If $V \setminus (S_1 \cup \dots \cup S_{l-1})$ has no independent subset of size n/k , then abort the “for” loop and try the next triple (X, φ, σ) .
 - Otherwise, let S'_l be the lexicographically first subset of $V \setminus (S_1 \cup \dots \cup S_{l-1})$ of size n/k that is independent in G . Then, let $S_l = (S'_l \cup \varphi^{-1}(\sigma_l)) \setminus \varphi^{-1}(\{1, \dots, k\} \setminus \{\sigma_l\})$.
 3. If (S_1, \dots, S_k) is a k -coloring of G , then output this coloring and halt.
 4. Answer “fail”.

Fig. 4. the procedure `Exact`.

The idea behind `Exact` is to “guess” a certain part of the hidden coloring of $G = G_{n,p,k}^*$. More precisely, `Exact` enumerates all sets X of a suitably chosen size x , and all k -colorings φ of X ; the k -coloring $\varphi : X \rightarrow \{1, \dots, k\}$ that `Exact` is really interested in is the one induced by the planted coloring of G . Then, in Step 2, `Exact` tries to find large independent sets S'_l of G . By Property (33), we expect that each of these sets consists mainly of vertices in one color class V_{σ_l} . Using the “guess” (X, φ, σ) , `Exact` tries to correct the set S'_l so that $S_l = V_{\sigma_l}$: Step 2 removes all vertices in $\varphi^{-1}(\{1, \dots, k\} \setminus \{\sigma_l\})$, i.e., all vertices from the other classes that have erroneously ended up in S'_l , and adds all vertices in $\varphi^{-1}(\sigma_l)$, i.e., all missing vertices from V_{σ_l} . The size of the “guess” of `Exact` is ruled by the parameter T . Note that the choice of x ensures that the number of possible triples (X, φ, σ) is $\leq T$.

4.3 Proof of Proposition 18

Given $G = G_{n,p,k}^*$, we define the sequence $\eta^* = (\eta_1^*, \dots, \eta_k^*)$ of numbers $\eta_i^* \in \{0, \dots, \frac{n}{2k}\}$ along with a permutation σ of $\{1, \dots, k\}$ inductively as follows. Having defined η_i^* and σ_i for $i = 1, \dots, l-1$, we let

$$U_l = \{1, \dots, k\} \setminus \{\sigma_1, \dots, \sigma_{l-1}\}. \quad (37)$$

If the graph $G[V_{U_l}]$ does not satisfy Property (32), then we let $\eta_l^* = n/(2k)$, and let $\sigma_l = \min U_l$. Otherwise, let $(x_v)_{v \in V_{U_l}}$ be the rigid vector $(k-l+1)$ -coloring of $G[V_{U_l}]$ computed by Step 1 of Classes on input G and $W = V_{U_l}$. By Lemma 19, there is a vertex w such that $\#S_w \geq \frac{199n}{200k}$. Let

$$v_l = \min \left\{ w \in V_{U_l} : \#S_w \geq \frac{199n}{200k} \right\} \quad (38)$$

be the smallest such vertex (recall that $V_{U_l} \subset V = \{1, \dots, n\}$). As we assume that G has Property (33), there is a unique $\sigma_l \in \{1, \dots, k\}$ such that $\#V_{\sigma_l} \cap S_{v_l} > \frac{99n}{100k}$. Now, we let

$$\eta_l^* = \text{def}_G(V_{\sigma_l}), \quad (39)$$

and proceed inductively.

The following lemma establishes the first part of Proposition 18. Throughout, we assume that G has Properties (33) and (34).

Lemma 23. *Classes($G, V, k, \eta_k^*, \eta_{k-1}^*, \dots, \eta_1^*$) finds a k -coloring of G .*

Proof. We show by induction that eventually $V_{\sigma_l} \in S_l$ for $l = 1, \dots, k$. Assume that the algorithm sets $S_j = V_{\sigma_j}$ for $j = 1, \dots, l-1$. We show that then the set S_l computed by

$$\text{Classes} \left[G, W = V \setminus \bigcup_{j=1}^{l-1} V_{\sigma_j}, k, \eta_k^*, \dots, \eta_l^* \right]$$

contains V_{σ_l} as an element. There are two cases to consider.

1st case: $\eta_l^* < n/(2k)$. Classes executes Step 2. Thus, by (38), Step 2 picks $v = v_l$. Moreover, due to (39) we have $\eta_l^* = \text{def}_G(V_{\sigma_l}) < n/(2k)$. Therefore, Corollary 20 entails that the output S_l of Purify contains V_{σ_l} as an element.

2nd case: $\eta_l^* \geq n/(2k)$. Eventually, Step 3 will encounter some $U \subset V_{\sigma_l}$, $\#U = \frac{n}{2k \ln k}$. By Property (34), we have $\#\bar{N}_{G_l}(U) \leq 2n/k$. As $V_{\sigma_l} \subset \bar{N}_{G_l}(U)$, Step 3 adds V_{σ_l} to S_l .

Thus, in both cases we have $V_{\sigma_l} \in S_l$, so that eventually Step 4 will try $S_l = V_{\sigma_l}$. Then Step 5 calls

$$\text{Classes} \left[G, V \setminus \bigcup_{j=1}^l V_{\sigma_j}, k, \eta_k^*, \dots, \eta_{l+1}^* \right].$$

Hence, proceeding inductively, we conclude that the coloring $(V_{\sigma_1}, \dots, V_{\sigma_k})$ will be recovered. \square

Moreover, the second assertion in Proposition 18 follows from the next lemma.

Lemma 24. *Let $\eta_1, \dots, \eta_k \geq 0$ be integers. Let $\eta_1^*, \dots, \eta_k^*$ be as defined above, with input graph $G = G_{n,p,k}^*$. Then $\mathbb{P}[\eta_i^* \geq \eta_i \text{ for } i = 1, \dots, k] \leq \prod_{i=1}^k \left(\frac{n/k}{\eta_i}\right)^{-90}$.*

Proof. Fix a sequence $\eta_1, \dots, \eta_k \geq 0$ of integers. Let $G = G_{n,p,k}^*$ be a semirandom graph with planted coloring V_1, \dots, V_k . Given integers $1 \leq l, \lambda \leq k$, we define an event $\mathcal{E}(l, \lambda)$ as follows: $G \in \mathcal{E}(l, \lambda)$ iff there exist two disjoint sets $J_1, J_2 \subset \{1, \dots, k\}$ and an injective map $\tau : J_1 \cup J_2 \rightarrow \{1, \dots, k\}$ such that the following conditions are satisfied.

- E1.** $\#J_1 = l$ and $\#J_2 = \lambda$.
- E2.** For $U = V_{\tau(J_1)}$ Property (32) is violated.
- E3.** $\text{def}_G(V_{\tau_i}) \geq \eta_i \geq 1$ for all $i \in J_2$.
- E4.** $\eta_i = 0$ for all $i \in \{1, \dots, k\} \setminus (J_1 \cup J_2)$.

We shall prove below that

$$\mathbb{P}[\mathcal{E}(l, \lambda)] \leq \prod_{i=1}^k \binom{n/k}{\eta_i}^{-98} \quad \text{for all } l, \lambda. \quad (40)$$

Furthermore, we claim that if $G = G_{n,p,k}^*$ is such that $\eta_i^* \geq \eta_i$ for $i = 1, \dots, k$, then there exist l^*, λ^* so that $G \in \mathcal{E}(l^*, \lambda^*)$. For if Property (32) does not hold in G with $U = \{1, \dots, k\}$, then $G \in \mathcal{E}(k, 0)$. Otherwise, we can define

$$l^* = k - \max \{1 \leq l \leq k : \text{Property (32) holds in } G[V_{U_j}] \text{ for all } j \leq l\}, \quad (41)$$

where U_l is defined in (37). In addition, we set

$$J_1 = \{k - l^* + 1, \dots, k\} = \sigma^{-1}(U_{k-l^*+1}), \quad J_2 = \{1 \leq i \leq k - l^* : \eta_i > 0\},$$

and we define $\tau : J_1 \cup J_2 \rightarrow \{1, \dots, k\}$, $j \mapsto \sigma_j$. Then by (39) and (41), J_1, J_2 , and τ satisfy E1–E4 with respect to $\mathcal{E}(l^*, \lambda^* = \#J_2)$. Hence, $G \in \mathcal{E}(l^*, \lambda^*)$. As a consequence, if $\max_{i=1, \dots, k} \eta_i > 0$, then we obtain

$$\mathbb{P}[\eta_i^* \geq \eta_i \text{ for } i = 1, \dots, k] \leq \sum_{l, \lambda=1}^k \mathbb{P}[\mathcal{E}(l, \lambda)] \stackrel{(40)}{\leq} k^2 \cdot \prod_{i=1}^k \binom{n/k}{\eta_i}^{-98} \stackrel{(4)}{\leq} \prod_{i=1}^k \binom{n/k}{\eta_i}^{-90},$$

as desired.

Thus, the remaining task is to prove (40). If we fix sets J_1, J_2 , and an injection $\tau : J_1 \cup J_2 \rightarrow \{1, \dots, k\}$ such that E1 and E4 hold, then by Lemma 13 and Lemma 14 we have

$$\mathbb{P}[\text{E2, E3 occur}] \leq \exp\left(-\frac{100ln}{k}\right) \prod_{i \in J_2} \binom{n/k}{\eta_i}^{-100} \leq \exp\left(-\frac{ln}{k}\right) \prod_{i=1}^k \binom{n/k}{\eta_i}^{-99}. \quad (42)$$

Further, there are

$$\leq \binom{k}{l} \binom{k}{\lambda} k^l \lambda! \leq k^{2l+\lambda} \quad (43)$$

ways to choose J_1, J_2 , and τ subject to E1 and E4. (For there are $\leq \binom{k}{l}$ ways to choose J_1 and $\leq \binom{k}{\lambda}$ ways to choose J_2 . Given J_1 and J_2 , there are $\leq k^l$ ways to choose the restriction of τ to J_1 , and finally $\leq \lambda!$ choices of the restriction of τ to J_2 subject to E4.) Combining (42) and (43), we obtain

$$\begin{aligned} \mathbb{P}[\mathcal{E}(l, \lambda)] &\leq k^{2l+\lambda} \cdot \exp\left(-\frac{ln}{k}\right) \prod_{i=1}^k \binom{n/k}{\eta_i}^{-99} \leq k^\lambda \cdot \left(\frac{k}{n}\right)^\lambda \cdot \prod_{i=1}^k \binom{n/k}{\eta_i}^{-98} \\ &\leq \left(\frac{k^2}{n}\right)^\lambda \cdot \prod_{i=1}^k \binom{n/k}{\eta_i}^{-98} \stackrel{(4)}{\leq} \prod_{i=1}^k \binom{n/k}{\eta_i}^{-98}, \end{aligned}$$

thereby establishing (40). \square

Proof of Proposition 18. Since the first two assertions follow from Lemmas 23 and 24, we just need to bound the running time of $\text{Classes}(G, V, k, \eta_1, \dots, \eta_k)$. Clearly, Step 1 runs in polynomial time. Moreover, by Corollary 20, the total time spend on executing Step 2 (for all k stages) is

$$\leq n^{O(1)} \prod_{i=1}^k \binom{n/k}{\eta_i}^{14}. \quad (44)$$

Further, Step 3 consumes time

$$n^{O(1)} \binom{n}{n/(2k \ln k)} \binom{2n/k}{n/k} \leq n^{O(1)} \binom{n/k}{n/(2k)}^7; \quad (45)$$

for there are $\leq \binom{n}{n/(2k \ln k)}$ ways to choose the set U , and if $\#T \leq 2n/k$, then there are $\leq \binom{2n/k}{n/k}$ ways to choose the set I . Since Step 3 gets executed only if $\eta_l = \frac{n}{2k}$, (44) and (45) entail that the total running time is

$$\leq n^{O(1)} \left[\prod_{i=1}^k \binom{n/k}{\eta_i}^{14} + \prod_{i=1}^k \binom{n/k}{\eta_i}^7 \right] \leq n^{O(1)} \prod_{i=1}^k \binom{n/k}{\eta_i}^{14},$$

so that the proposition follows. \square

4.4 Proof of Proposition 21

Due to Steps 3–4, $\text{Exact}(G, k, T)$ either outputs a k -coloring of its input graph G or “fail”. Thus, the remaining task is to bound the running time of Exact on input $G_{n,p,k}^*$.

Thus, let $G = G_{n,p,k}^*$, and let V_1, \dots, V_k be the planted k -coloring of G . Assuming Property (33), we construct

- sets $X_l^* \subset V$ of cardinality $2x_l^*$ for $l = 1, \dots, k$,
- a permutation σ^* of $\{1, \dots, k\}$,
- independent sets S_1^*, \dots, S_k^*

inductively as follows. Starting with $l = 1$, let $S_l^* \subset V \setminus \bigcup_{j=1}^{l-1} V_{\sigma_j^*}$ be the lexicographically first independent set of cardinality n/k . Then by Property (33), there is an $1 \leq i \leq k$ such that $\#V_i \cap S_l^* > n/(2k)$. Set $\sigma_l^* = i$, let

$$X_l^* = (V_i \setminus S_l^*) \cup (S_l^* \setminus V_i), \quad (46)$$

$$x_l^* = S_l^* \setminus V_i, \quad (47)$$

and proceed inductively. Finally, set $X^* = \bigcup_{l=1}^k X_l^*$ and $x^* = \#X^*$. Then

$$x^* \leq 2 \sum_{i=1}^k x_i^*. \quad (48)$$

Further, let φ^* be the coloring induced on X^* by the k -coloring (V_1, \dots, V_k) of G , and set

$$T^* = \binom{n}{x^*} k^{x^*} k!.$$

If Property (33) is violated in G , then we let $T^* = \lceil \exp(n) \rceil$.

Lemma 25. *If $T^* < \lceil \exp(n) \rceil$, then $\text{Exact}(G, k, T)$ outputs a k -coloring for all $T \geq T^*$.*

Proof. We claim that at the latest when $T = T^*$, $(X, \varphi, \sigma) = (X^*, \varphi^*, \sigma^*)$ Steps 2–3 of Exact will k -color G . The proof is by induction on $l = 1, \dots, k$. Suppose that Exact has set $S_j = V_{\sigma_j}$ for all $1 \leq j < l$. To show that then $S_l = V_{\sigma_l}$, we let $S_l' \subset V \setminus \bigcup_{j=1}^{l-1} S_j = V \setminus \bigcup_{j=1}^{l-1} V_{\sigma_j}$ be the independent set of size n/k computed in Step 2. As S_l' is the lexicographically first independent set of size n/k , by construction we have $S_l' = S_l^*$. Hence, (46) entails that

$$(V_{\sigma_l} \setminus S_l') \cup (S_l' \setminus V_{\sigma_l}) = X_l^* \subset X^* = X. \quad (49)$$

As $\sigma = \sigma^*$ and $\varphi = \varphi^*$ is the k -coloring induced by (V_1, \dots, V_k) on X^* , we get

$$S_l = (S_l' \cup \varphi^{-1}(\sigma_l)) \setminus \varphi^{-1}(\{1, \dots, k\} \setminus \{\sigma_l\}) = (S_l' \cup (X^* \cap V_{\sigma_l})) \setminus (X^* \setminus V_{\sigma_l}) \stackrel{(49)}{=} V_l,$$

as desired. Thus, Steps 2–3 find a k -coloring. \square

Lemma 26. *Let $\exp(n/\ln k) \leq T \leq \lceil \exp(n) \rceil$. Then $\mathbb{P}[T^* > T] \leq T^{-90}$.*

Proof. If G violates Property (33), then $T^* = \lceil \exp(n) \rceil$, and the assertion follows from Lemma 15. Thus, we may assume that Property (33) holds. Set

$$x^{(T)} = \max \left\{ x \geq 0 : \forall 0 \leq y \leq x : T \geq \binom{n}{y} k^y k! \right\},$$

$$\mathcal{X}^{(T)} = \left\{ (x_1, \dots, x_k) \in \{0, \dots, x^{(T)}\}^k : \sum_{i=1}^k x_i = \left\lceil \frac{x^{(T)}}{2} \right\rceil \right\}.$$

Since (4) entails

$$\binom{n}{100k/p} k^{100k/p} k! \leq \binom{n}{100k/p}^2 k^{200k/p} \leq \left(\frac{enp}{100}\right)^{200k/p} \leq \exp\left(\frac{n}{\ln k}\right) \leq T,$$

we conclude that

$$x^{(T)} \geq \frac{100k}{p}. \quad (50)$$

Given a sequence x_1, \dots, x_k of integers ≥ 0 , we consider the following event $\mathcal{E}(x_1, \dots, x_k)$:

there is a permutation σ of $\{1, \dots, k\}$ and a collection of sets S''_1, \dots, S''_k such that $S''_l \subset V_{\sigma_l}$, $\#S''_l \geq \frac{n}{2k}$, and $\#\bar{N}_G(S''_l) \setminus V_{\sigma_l} \geq x_i$ for $l = 1, \dots, k$.

We shall prove below that for all $(x_1, \dots, x_k) \in \mathcal{X}^{(T)}$

$$\mathbb{P}[\mathcal{E}(x_1, \dots, x_k)] \leq \exp\left(2n - \left\lceil \frac{np}{4k} - \ln n \right\rceil \frac{x^{(T)}}{2}\right). \quad (51)$$

Now, let $G = G_{n,p,k}^*$ be such that $T^* > T$. Let $S''_l = S_l^* \cap V_{\sigma_l^*}$. Because the sets S_l^* are independent, we have $S_l^* \setminus V_{\sigma_l^*} \subset \bar{N}_G(S''_l)$ ($l = 1, \dots, k$). Moreover, by construction we have $\#S''_l \geq \frac{n}{2k}$, so that $\mathcal{E}(x_1^*, \dots, x_k^*)$ occurs (cf. (47)). Further, as $T^* > T$, we have

$$x^{(T)} < x^* \stackrel{(48)}{\leq} 2 \sum_{i=1}^k x_i^*.$$

Reducing some of the x_i^* 's if necessary, we obtain a sequence $(x_1, \dots, x_k) \in \mathcal{X}^{(T)}$ such that $\mathcal{E}(x_1, \dots, x_k)$ occurs. Thus,

$$\mathbb{P}[T^* > T] \leq \sum_{(x_1, \dots, x_k) \in \mathcal{X}^{(T)}} \mathbb{P}[\mathcal{E}(x_1, \dots, x_k)] \stackrel{(51)}{\leq} \#\mathcal{X}^{(T)} \cdot \exp\left(2n - \left\lceil \frac{np}{4k} - \ln n \right\rceil \frac{x^{(T)}}{2}\right). \quad (52)$$

Observe that

$$\#\mathcal{X}^{(T)} \leq \binom{x^{(T)} + k - 1}{k - 1} \leq 2^{x^{(T)} + k - 1}, \quad (53)$$

$$T \leq \binom{n}{x^{(T)} + 1} k^{x^{(T)} + 1} k! \stackrel{(50)}{\leq} \exp(4x^{(T)} \ln n). \quad (54)$$

Therefore, continuing (52), we get

$$\begin{aligned} \mathbb{P}[T^* > T] &\stackrel{(53)}{\leq} 2^{x^{(T)} + k - 1} \exp\left(2n - \left\lceil \frac{np}{4k} - \ln n \right\rceil \frac{x^{(T)}}{2}\right) \\ &\stackrel{(50)}{\leq} \exp\left(2x^{(T)} \ln(n) + 2n - \frac{np}{4k} x^{(T)}\right) \stackrel{(4), (50)}{\leq} \exp\left(-\frac{np}{8k} x^{(T)}\right) \stackrel{(4), (54)}{\leq} T^{-90}, \end{aligned}$$

as desired.

Finally, let us prove (51). Let $(x_1, \dots, x_k) \in \mathcal{X}^{(T)}$. Let us fix sets $S'_l \subset V_{\sigma_l}$ of cardinality $\geq \frac{n}{2k}$ for a moment. In addition, consider sets X_1, \dots, X_k such that $X_l \subset V \setminus V_{\sigma_l}$ and $\#X_l = x_l$ for $l = 1, \dots, k$. If $e_G(X_l, S'_l) = 0$ for $l = 1, \dots, k$, then we know of $\geq \frac{1}{2} \sum_{l=1}^k \#X_l \cdot \#S'_l \geq \frac{nx^{(T)}}{4k}$ edges that are not present in $G = G_{n,p,k}^*$, although each of these edges occurs with probability p independently in $G_{n,p,k}$. Hence,

$$\mathbb{P}[e_G(X_l, S'_l) = 0 \text{ for } l = 1, \dots, k] \leq (1-p)^{nx^{(T)}/(4k)} \leq \exp\left(-\frac{x^{(T)}np}{4k}\right). \quad (55)$$

Furthermore, given the permutation σ , there are at most $2^{n/k}$ ways to choose the set S'_l . Moreover, there are at most $\binom{n}{x_l}$ ways to choose the set X_l . Therefore, the union bound and (55) yield

$$\begin{aligned} \mathbb{P}[\mathcal{E}(x_1, \dots, x_k)] &\leq \exp\left(-\frac{x^{(T)}np}{4k}\right) \cdot k! \prod_{l=1}^k \binom{n}{x_l} 2^{n/k} \\ &\leq k! \exp\left(n + \left[\ln(n) - \frac{np}{4k}\right] x^{(T)}\right) \stackrel{(4)}{\leq} \exp\left(2n + \left[\ln(n) - \frac{np}{4k}\right] x^{(T)}\right), \end{aligned}$$

so that (51) follows. \square

Proof of Proposition 21. The first assertion is immediate. For $\exp(n/\ln k) \leq T \leq \lceil \exp(n) \rceil$, Steps 1–3 of `Exact`(G, k, T) consume time

$$\leq n^{O(1)} \left[\binom{n}{x} k^x k! \right] \cdot \binom{n}{n/k} \leq n^{O(1)} T \cdot \exp\left(\frac{2n}{\ln k}\right) \leq n^{O(1)} T^3;$$

for there are $\leq \binom{n}{x} k^x k!$ ways to choose the triple (X, φ, σ) , and Step 2 needs to check $\leq \binom{n}{n/k}$ subsets of V . Hence, the proposition follows from Lemma 25 and Lemma 26. \square

4.5 Proof of Theorem 3

By Proposition 21, `ExpColor` computes a k -coloring of every k -colorable input graph. Thus, the remaining task is to show that `ExpColor`($G_{n,p,k}^*, k$) runs in polynomial expected time.

Given $1 \leq T \leq \exp(n/\ln k)$, we can bound the running time of Steps 2–3 of `ExpColor` as follows. There are at most $z_\eta = \binom{\eta+k-1}{k-1}$ ways to choose the numbers η_1, \dots, η_k . If $\eta \geq k-1$, then

$$z_\eta \leq 2^{\eta+k-1} \leq \exp(2\eta) \leq T^2, \quad (56)$$

by the definition of η . Moreover, if $\eta < k-1$, then due to (4) we have

$$z_\eta = \binom{\eta+k-1}{\eta} \leq \left(\frac{2k}{\eta}\right)^\eta \leq \left(\frac{n}{k\eta}\right)^{2\eta} \leq \left(\frac{n/k}{\eta}\right)^2 \leq T^2. \quad (57)$$

Further, having fixed (η_1, \dots, η_k) , by Proposition 18 Step 3 consumes time $\leq n^{O(1)} \prod_{i=1}^k \binom{n/k}{\eta_i} \leq n^{O(1)} T^{14}$. Thus, the total running time of Steps 2–3 for a given T is

$$R_T \leq z_\eta \cdot n^{O(1)} T^{14} \stackrel{(56), (57)}{\leq} n^{O(1)} T^{16}. \quad (58)$$

Now let $G = G_{n,p,k}^*$. Then we define a number $T^* = T^*(G_{n,p,k}^*)$ as follows. If G violates Property (33), then we set $T_1^* = \exp(n)$. Otherwise, if G violates Property (34), then we let $T_1^* = \exp(n/\ln k)$. Moreover, if G satisfies both Property (33) and (34), then let $(\eta_1^*, \dots, \eta_k^*)$ be as in Proposition 18 and set $T_1^* = \prod_{i=1}^k \binom{n/k}{\eta_i^*}$. In addition, let

$$T_2^* = \min \{ \exp(n/\ln k) \leq T \leq \lceil \exp(n) \rceil : \text{Exact}(G, k, T) \text{ finds a } k\text{-coloring of } G \}.$$

Set $T^* = T_1^*$, if $T_1^* \leq \exp(n/\ln k)$, and $T^* = \max\{T_1^*, T_2^*\}$ otherwise. Then by Proposition 18 and Proposition 21, `ExpColor` finds a k -coloring of G before the variable T exceeds T^* .

Combining Propositions 18 and 21 and Lemmas 13 and 15, we conclude that

$$\mathbb{P}[T^* > T] \leq T^{-80} \quad (59)$$

for all $1 \leq T \leq \lceil \exp(n) \rceil$. Consequently, by (58) and (59) we get

$$\sum_{T=1}^{\lceil \exp(n/\ln k) \rceil} R_T \mathbb{P}[T^* > T] \leq n^{O(1)} \sum_{T=1}^{\infty} T^{16-80} = n^{O(1)},$$

so that the expected time spent on executing Steps 1–3 of `ExpColor` is polynomial. Finally, if $T^* \geq \exp(n/\ln k)$, then by Proposition 21 the expected time spent on executing Steps 4–5 of `ExpColor` is polynomial as well. Thus, `ExpColor`($G_{n,p,k}^*$, k) runs in polynomial expected time.

5 Coloring $G_{n,p,k}^*$ Optimally in Polynomial Expected Time

The goal of this section is to prove Theorem 4. *Throughout, we assume that (5) holds.* The algorithm `OptColor` is based on the following observation.

Lemma 27. *We have $\mathbb{P}\left[\bar{\vartheta}_2(G_{n,p,k}^*) \leq k - \frac{1}{2}\right] \leq \exp(-100n)$.*

Proof. Let $G_0 = G_{n,p,k}$, and let $G = G_{n,p,k}^* \in \mathcal{I}(G_0)$ be a semirandom graph obtained from G_0 . Let $h = k - \frac{1}{2}$. If $\bar{\vartheta}_2(G_0) \leq \bar{\vartheta}_2(G) \leq h$, then G_0 has a rigid vector h -coloring $(x_v)_{v \in V}$. Plugging the feasible solution $(x_v)_{v \in V}$ into the semidefinite program SDP_h , we conclude that

$$\text{if } \bar{\vartheta}_2(G_0) \leq h, \text{ then } \text{SDP}_h(G_0) \geq \#E(G_0) \quad (\text{cf. (9)}). \quad (60)$$

Furthermore, as $\#E(G_0)$ is binomially distributed with mean $(1 - k^{-1})n^2p/2$, (10) entails that

$$\mathbb{P}\left[\#E(G_0) \leq (1 - k^{-1})\frac{n^2p}{2} - Cn^{3/2}p^{1/2}\right] \leq \exp(-101n), \quad (61)$$

where $C > 0$ denotes a suitable constant. Combining (60) and (61), we conclude that

$$\begin{aligned} \mathbb{P}\left[\bar{\vartheta}_2(G_{n,p,k}^*) \leq h\right] &\leq \exp(-101n) + \mathbb{P}\left[\text{SDP}_h(G_{n,p,k}) \geq (1 - k^{-1})\frac{n^2p}{2} - Cn^{3/2}p^{1/2}\right] \\ &\leq \exp(-101n) + \mathbb{P}\left[\text{SDP}_h(G_{n,p,k}) \geq (1 - h^{-1})\frac{n^2p}{2} + \frac{n^2p}{4k^2} - Cn^{3/2}p^{1/2}\right] \\ &\stackrel{(5)}{\leq} \exp(-101n) + \mathbb{P}\left[\text{SDP}_h(G_{n,p,k}) \geq (1 - h^{-1})\frac{n^2p}{2} + \left(\frac{\sqrt{C_0}}{4} - C\right)n^{3/2}p^{1/2}\right]. \end{aligned} \quad (62)$$

Choosing the constant C_0 large enough, we can ensure that $\sqrt{C_0}/4 - C \geq \sqrt{C_0}/8$. Furthermore, as the random graph $G_{n,p}$ can be obtained from $G_{n,p,k}$ by adding random edges inside the planted color classes V_1, \dots, V_k , the monotonicity property (8) entails that $\text{SDP}_h(G_{n,p,k})$ is stochastically dominated by $\text{SDP}_h(G_{n,p})$. Hence, (62) yields

$$\mathbb{P}\left[\bar{\vartheta}_2(G_{n,p,k}^*) \leq h\right] \leq \exp(-101n) + \mathbb{P}\left[\text{SDP}_h(G_{n,p}) \geq (1 - h^{-1})\frac{n^2p}{2} + \frac{\sqrt{C_0}}{8}n^{3/2}p^{1/2}\right]. \quad (63)$$

Finally, Lemma 5 entails that

$$\mathbb{P}\left[\text{SDP}_h(G_{n,p}) \geq (1 - h^{-1})\frac{n^2p}{2} + \frac{\sqrt{C_0}}{4}n^{3/2}p^{1/2}\right] \leq \exp(-101n), \quad (64)$$

provided that the constant C_0 is sufficiently large. Thus, the assertion follows from (63) and (64). \square

Algorithm 28. $\text{OptColor}(G)$

Input: A graph $G = (V, E)$. *Output:* An optimal coloring of G .

1. Let $\kappa = \lceil \bar{\vartheta}_2(G) \rceil$.
2. Call $\text{ExpColor}(G, \kappa)$ and output the resulting coloring.

Fig. 5. the algorithm OptColor .

Given an input graph $G = (V, E)$, the algorithm OptColor just computes the lower bound $\kappa = \lceil \bar{\vartheta}(G) \rceil$ on $\chi(G)$ and calls $\text{ExpColor}(G, \kappa)$ (cf. Figure 5).

Proof of Theorem 4. If $\kappa = k$, then Theorem 3 shows that $\text{ExpColor}(G, \kappa)$ finds a κ -coloring of $G = G_{n,p,k}^*$. Furthermore, if $\kappa < k$, then either Steps 1–4 of $\text{ExpColor}(G, \kappa)$ will color G with κ colors, or Step 5 of ExpColor computes an optimal coloring of G . Hence, as $\kappa \leq \chi(G)$, in any case OptColor outputs an optimal coloring. Finally, the fact that the expected running time of OptColor is polynomial follows from Theorem 3 and Lemma 27. \square

6 Proof of Theorem 2

Consider a random graph $G_0 = G_{n,p,k}$, and let V_1, \dots, V_k be its color classes, where $\#V_i = n/k$. Let $p = dk/n$, $d = (\frac{1}{2} - \varepsilon) \ln(n/k)$. Let $i, j_1, j_2 \in \{1, \dots, k\}$ be distinct. Then for each $v \in V_i$, the number $d_v^{(j_1)}$ of neighbors of v in V_{j_1} has binomial distribution with parameters n/k and p . Therefore,

$$P(d_v^{(j_1)} = 0) = P(d_v^{(j_2)} = 0) = (1 - p)^{n/k} \sim \exp(-d) = \left(\frac{n}{k}\right)^{\varepsilon - 1/2}.$$

Thus, $P(d_v^{(j_1)} = d_v^{(j_2)} = 0) \sim (n/k)^{2\varepsilon - 1}$. Hence, the expected number of vertices $v \in V_i$ satisfying $d_v^{(j_1)} = d_v^{(j_2)} = 0$ is $\sim (n/k)^{2\varepsilon}$. Since edges are chosen independently, by the Chernoff bound (10) the number of such vertices is in fact $\geq \left(\frac{n}{k}\right)^\varepsilon$ w.h.p. Consequently, w.h.p. there are sets $S_i \subset V_i$, $\#S_i = \left(\frac{n}{k}\right)^\varepsilon$, $i = 1, 2, 3$, such that $N_{G_0}(S_i) \cap (V_1 \cup V_2 \cup V_3) = \emptyset$.

Now assume that we had an algorithm \mathcal{A} that can k -color $G_{n,p,k}^*$ w.h.p. Let H be an arbitrary graph that admits a 3-coloring with color classes of cardinality $\left(\frac{n}{k}\right)^\varepsilon$ each. We show how to convert \mathcal{A} into a randomized algorithm that 3-colors H , which is NP-hard. First, randomly partition $V = \{1, \dots, n\}$ into k sets V_1, \dots, V_k of cardinality n/k . Then choose $\left(\frac{n}{k}\right)^\varepsilon$ vertices S_i from V_i at random for $i = 1, 2, 3$. Further, form a complete k -partite graph on the vertices $V_1 \setminus S_1, V_2 \setminus S_2, V_3 \setminus S_3, V_4, \dots, V_k$, and connect $S_1 \cup S_2 \cup S_3$ completely with $V_4 \cup \dots \cup V_k$. Finally, embed a randomly permuted copy of H into the set $S_1 \cup S_2 \cup S_3$ (without taking care of the coloring of H , of course). Let G be the resulting graph. We claim that running $\mathcal{A}(G)$ yields a k -coloring of G w.h.p.

To prove the claim, we volunteer as an adversary that given $G_0 = G_{n,p,k}$ produces the instance G described above. (To this end, we may use unlimited computational power.) Given a random graph $G_0 = G_{n,p,k}$ with color classes V_1, \dots, V_k , we first look for sets $S_i \subset V_i$ such that $\#S_i = \left(\frac{n}{k}\right)^\varepsilon$ and $N(S_i) \cap (V_1 \cup V_2 \cup V_3) = \emptyset$, $i = 1, 2, 3$. As pointed out above, such sets S_1, S_2, S_3 exist w.h.p.—if not, we give up. Then, we turn $G_0 - (S_1 \cup S_2 \cup S_3)$ into a complete k -partite graph, and connect $S_1 \cup S_2 \cup S_3$ completely with $V_4 \cup \dots \cup V_k$. Further, we compute a 3-coloring of the worst-case instance H with color classes of equal size, permute the vertices in each of the three color classes of H randomly, and map the three color classes onto S_1, S_2, S_3 (thus, this time we respect the coloring). The *distribution* of the resulting graph G' coincides with the distribution of the graph G constructed in the previous paragraph, so that \mathcal{A} k -colors G w.h.p. As any k -coloring of G induces a 3-coloring on $G[S_1 \cup S_2 \cup S_3] = H$, we have shown that a polynomial time algorithm for k -coloring $G_{n,p,k}^*$ w.h.p. yields a randomized algorithm for 3-coloring the worst-case instance H .

Remark 29. The only difference between the above construction and the ones given in [16] is that instead of reducing the problem of k -coloring a k -colorable graph to k -coloring $G_{n,p,k}^*$, we reduced the problem of 3-coloring a 3-colorable graph to k -coloring $G_{n,p,k}^*$. The idea of working a worst-case instance into the semirandom instance occurs already in [4].

7 Proofs of Auxiliary Lemmas

Throughout, we assume that (4) holds for some large enough constant $C_0 > 0$.

7.1 Proof of Lemma 13

By Lemma 5 there exists a constant $C_1 > 0$ such that

$$\mathbb{P} \left[\text{SDP}_u(G_{un/k,p}) \leq \left(1 - \frac{1}{u}\right) \binom{un/k}{2} p + \frac{C_1 un}{2k} \sqrt{np} \right] \geq 1 - \exp\left(-\frac{201un}{k}\right). \quad (65)$$

Now consider $G_0 = G_{n,p,k}$, and let G'_0 be a graph obtained from G_0 by adding each edge inside the planted color classes V_i , $i = 1, \dots, k$, with probability p independently. Then G'_0 is distributed as a random graph $G_{n,p}$. Thus, in particular,

$$G'_0[V_U] = G_{un/k,p}. \quad (66)$$

Let \mathcal{B} be the event that $\#E(G_0[V_U]) \geq \left(1 - \frac{1}{u}\right) \binom{un/k}{2} p - \frac{C_1 nu}{2k} \sqrt{np}$. Furthermore, let \mathcal{A} be the event that (32) is violated for V_U in G_0 . Then invoking (65) and (66), we obtain

$$\frac{1}{3} \mathbb{P}(\mathcal{A} \cap \mathcal{B}) \leq \mathbb{P} \left[\text{SDP}_u(G_{un/k,p}) > \left(1 - \frac{1}{u}\right) \binom{un/k}{2} p + \frac{C_1 un}{2k} \sqrt{np} \right] \leq \exp\left(-\frac{201un}{k}\right). \quad (67)$$

Moreover, $\#E(G_0[V_U])$ is binomially distributed with mean $\mathbb{E}[\#E(G_0[V_U])] = \binom{u}{2} \left(\frac{n}{k}\right)^2 p \sim \left(1 - \frac{1}{u}\right) \binom{un/k}{2}$. Hence, choosing the constant $C_1 > 0$ sufficiently large and applying the Chernoff bound (10), we get

$$\mathbb{P}(\mathcal{B}) \geq 1 - \exp\left(-\frac{201nu}{k}\right). \quad (68)$$

Now, (68) implies that $\mathbb{P}(\mathcal{A} \setminus \mathcal{B}) \leq \mathbb{P}(\neg \mathcal{B}) \leq \exp(-201nu/k)$, which in combination with (67) entails

$$\mathbb{P}(\mathcal{A}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\mathcal{A} \setminus \mathcal{B}) \leq 2 \exp\left(-\frac{201nu}{k}\right) \leq \exp\left(-\frac{200nu}{k}\right). \quad (69)$$

Finally, let $G = G_{n,p,k}^*$. Let $G_0 = G_{n,p,k}$ be the random k -colorable graph contained in G . Let $M = \#E(G[V_U]) - \#E(G_0[V_U])$ be the number of edges added by the adversary. Then $\text{SDP}_u(G') \leq \text{SDP}_u(G'_0) + M$, because adding one edge can increase the value of SDP_u by at most 1. Therefore, if $\text{SDP}_u(G'_0) \leq \#E(G'_0[V_U]) + C_1 \frac{nu}{k} \sqrt{np}$, then

$$\text{SDP}_u(G') \leq M + \#E(G'_0[V_U]) + C_1 \frac{nu}{k} \sqrt{np} = \#E(G'[V_U]) + C_1 \frac{nu}{k} \sqrt{np}.$$

Hence, the assertion follows from (69).

7.2 Proof of Lemma 14

Throughout, we fix a partition (V_1, \dots, V_k) of $V = \{1, \dots, n\}$ into k disjoint sets of cardinality n/k . Let $G_0 = G_{n,p,k}$ be a random k -colorable graph with planted coloring V_1, \dots, V_k , and let $G = G_{n,p,k}^*$ be the semirandom graph obtained from G_0 by the adversary. Then $\text{def}_G(V_i) \leq \text{def}_{G_0}(V_i)$ for $i = 1, \dots, k$. Hence, it suffices to show that Lemma 14 holds for $G = G_0 = G_{n,p,k}$.

Since on $G = G_{n,p,k}$ the random variables $(\text{def}_G(V_i))_{i=1,\dots,k}$ are not mutually independent, we decompose G into k mutually independent subgraphs $G^{(i)}$ and investigate the defects $\text{def}_{G^{(i)}}(V_i)$ ($i = 1, \dots, k$). Letting

$$p' = 1 - \sqrt{1-p} \geq p/2, \quad (70)$$

we obtain the graph $G^{(i)}$ by including each of the $(1 - k^{-1})k^{-1}n^2$ possible $(V \setminus V_i)$ - V_i -edges with probability p' independently. Thus, $G^{(i)}$ is a random bipartite graph. Furthermore, in the union

$$H = \bigcup_{i=1}^k G^{(i)} = \left(V, \bigcup_{i=1}^k E(G^{(i)}) \right)$$

every V_i - V_j -edge is present with probability $2p' - p'^2 = p$ independently of all other edges ($i \neq j$). Therefore, H has the same distribution as $G = G_{n,p,k}$. As a consequence, given $\eta_1, \dots, \eta_k \geq 0$, we have

$$\begin{aligned} \mathbb{P}[\text{def}_G(V_i) \geq \eta_i \text{ for } i = 1, \dots, k] &= \mathbb{P}[\text{def}_H(V_i) \geq \eta_i \text{ for } i = 1, \dots, k] \\ &\leq \mathbb{P}[\text{def}_{G^{(i)}}(V_i) \geq \eta_i \text{ for } i = 1, \dots, k] = \prod_{i=1}^k \mathbb{P}[\text{def}_{G^{(i)}}(V_i) \geq \eta_i], \end{aligned}$$

because the graphs $G^{(i)}$ are mutually independent. Hence, our aim is to prove that

$$\mathbb{P}[\text{def}_{G^{(i)}}(V_i) \geq \eta_i] \leq \left(\frac{n/k}{\eta_i} \right)^{-100} \quad \text{for } i = 1, \dots, k. \quad (71)$$

To prove (71), we first bound the probability that Condition D1 in the definition of the defect occurs.

Lemma 30. *With probability $\geq 1 - \exp(-100n/k)$ the random graph $G^{(i)}$ has the following property.*

$$\text{If } U \subset V_i \text{ has cardinality } \geq \frac{n}{2k}, \text{ then } \#V \setminus (V_i \cup N_{G^{(i)}}(U)) \leq \frac{n}{200k^2}. \quad (72)$$

Proof. Assuming that $np \geq C_0 k^2$ for a sufficiently large constant C_0 (cf. (4)), we have $s = 300/p \leq n/(200k^2)$. Fix a set $U \subset V_i$ of cardinality $\geq \frac{n}{2k}$ for a moment. Then

$$\begin{aligned} \mathbb{P}[V \setminus (V_i \cup N_{G^{(i)}}(U)) \geq s] &\leq \binom{n}{s} (1-p')^{ns/k} \stackrel{(70)}{\leq} \left(\frac{en}{s} \right)^s \exp\left(-\frac{nps}{2k}\right) \\ &\leq \exp\left(\frac{300 \ln(np)}{p} - \frac{150n}{k}\right) \stackrel{(4)}{\leq} \exp\left(-\frac{149n}{k}\right). \end{aligned} \quad (73)$$

As there are $\leq 2^{n/k}$ sets $U \subset V_i$ of cardinality $\geq \frac{n}{2k}$, due to the union bound (73) entails that

$$\mathbb{P}[(72) \text{ is violated}] \leq 2^{n/k} \exp\left(-\frac{149n}{k}\right) \leq \exp\left(-\frac{148n}{k}\right),$$

so that the assertion follows. \square

Furthermore, the next lemma regards Condition D2 in the definition of the defect.

Lemma 31. *Let $1 \leq \eta \leq \frac{n}{2k}$. Then with probability $\geq 1 - \left(\frac{n/k}{\eta}\right)^{-101}$ the graph $G^{(i)}$ has the following property.*

Let $6 \leq d \leq \lceil 50k \rceil$. Then every subset $T \subset V \setminus V_i$ of size $\#T \leq \frac{n}{2kd}$ has a d -fold matching to V_i with defect $\leq \eta$.

To prove Lemma 31, we need the following observation.

Lemma 32. *The probability that in $G^{(i)}$ there are $\eta \geq 1$ vertices in $V \setminus V_i$ that have $< np'/(2k)$ neighbors in V_i is $\leq n^{-200\eta}$.*

Proof. Let $v \in V \setminus V_i$. As $e_{G^{(i)}}(v, V_i)$ is binomially distributed with mean np/k , the Chernoff bound (10) yields

$$\mathbb{P} \left[e_{G^{(i)}}(v, V_i) \leq \frac{np'}{2k} \right] \leq \exp \left(-\frac{np'}{8k} \right).$$

Therefore, the probability that there are $\geq \eta \geq 1$ such vertices is

$$\leq \binom{n}{\eta} \exp \left(-\frac{np'\eta}{8k} \right) \stackrel{(70)}{\leq} \exp \left[\left(\ln(n) - \frac{np'}{16k} \right) \eta \right] \stackrel{(4)}{\leq} n^{-200\eta},$$

as claimed. \square

In addition, we need the following lemma from [11, Lemma 4.3].

Lemma 33. *Let $V' \subset V$ be a subset of cardinality n_1 , and let $V'' = V \setminus V'$, $\#V'' = n_2$. Let $\gamma > 0$ be an arbitrary constant, and let $2 \leq d \leq n_2/10$. Then there exists a number $\omega_0 = \omega_0(\gamma)$ such that the following holds. Let $\omega = \omega_0 \max\{d, \ln n\}$, and let H be a random bipartite graph obtained as follows: every vertex in V' chooses a set of at least $\omega = \omega_0 \max\{d, \ln n\}$ neighbors in V'' uniformly at random; these choices occur independently for all vertices in V' . Then for all $\eta \in \{0, 1, \dots, n_2/2\}$ we have*

$$\mathbb{P} \left[\exists T \subset V' : \#T \leq \frac{n_2}{2d} \wedge \#N_H(T) < d\#T - \eta \right] \leq \binom{n_2}{\eta}^{-\gamma}.$$

Proof of Lemma 31. Fix $6 \leq d \leq \lceil 50k \rceil$. We say that $G^{(i)}$ is (d, η) -good if every set $S \subset V \setminus V_i$ admits a d -fold matching to V_i with defect $\leq \eta$. Let $\omega = np'/(2k)$, $W = \{v \in V \setminus V_i : e_{G^{(i)}}(v, V_i) \leq \omega\}$, and $G' = G^{(i)}[V \setminus W]$. Let $\omega_0 = \omega_0(110)$ be the number from Lemma 33. If we choose the constant C_0 large enough, then (4) and (70) yield that $\omega \geq \omega_0 \cdot \max\{d, \ln n\}$.

Letting $0 \leq \eta_1 \leq \eta$, we have

$$\mathbb{P}(\#W = \eta_1) \leq n^{-200\eta_1} \tag{74}$$

by Lemma 32. In addition, set $\eta_2 = \eta - \eta_1$. Let us call a set $T \subset V \setminus (V_i \cup W)$ (d, η_2) -bad if $\#T \leq \frac{n}{2kd}$ and $e_{G^{(i)}}(T, V_i) < d\#T - \eta_2$. Then Lemma 33 entails that

$$\mathbb{P}(G^{(i)} \text{ has a } (d, \eta_2)\text{-bad set} \mid \#W = \eta_1) \leq \binom{n/k}{\eta_2}^{-110}. \tag{75}$$

If $G^{(i)}$ has no (d, η_2) -bad set, then $G^{(i)}$ is (d, η) -good. For if $S \subset V \setminus V_i$ has size $\#S \leq \frac{n}{2dk}$, then Hall's theorem entails that $T = S \setminus W$ has a d -fold matching to T with defect $\leq \eta_2$. As a consequence, (74) and (75) yield

$$\begin{aligned} \mathbb{P} \left[G^{(i)} \text{ is not } (d, \eta)\text{-good} \right] &\leq \sum_{\eta_1=0}^{\eta} \mathbb{P}[\#W = \eta_1] \mathbb{P} \left[G^{(i)} \text{ has a } (d, \eta_2)\text{-bad set} \mid \#W = \eta_1 \right] \\ &\leq \sum_{\eta_1=0}^{\eta} n^{-200\eta_1} \binom{n/k}{\eta_2}^{-110} \leq \binom{n/k}{\eta}^{-109}. \end{aligned} \tag{76}$$

Summing (76) over $6 \leq d \leq \lceil 50k \rceil$, we obtain

$$\mathbb{P} \left[G^{(i)} \text{ is } (d, \eta)\text{-good for all } 6 \leq d \leq \lceil 50k \rceil \right] \geq 1 - 50k \binom{n/k}{\eta}^{-109} \stackrel{(4)}{\geq} 1 - \binom{n/k}{\eta}^{-108},$$

thereby proving the lemma. \square

Combining Lemma 30 and Lemma 31, we conclude that (71) holds, thereby completing the proof of Lemma 14.

7.3 Proof of Lemma 15

To establish the first part of Lemma 15, it suffices to consider random k -colorable graphs $G = G_{n,p,k}$. Let V_1, \dots, V_k be the planted k -coloring. Let $U \subset V$ be any set of cardinality $u = \#U \geq \frac{n}{100k}$ such that

$$u_i = \#U \cap V_i \leq \frac{199}{200}u \text{ for } i = 1, \dots, k. \quad (77)$$

Our goal is to bound the probability that U is independent in G . Clearly, $u = \sum_{i=1}^k u_i$, and the number of possible edges among the vertices of U is $\lambda = \sum_{1 \leq i < j \leq k} u_i u_j = \frac{1}{2} \left[u^2 - \sum_{i=1}^k u_i^2 \right]$. Note that λ is minimized subject to (77) when $\sum_i u_i^2$ is maximized. Thus, λ attains its minimal value for $u_1 = 199u/200$, $u_2 = u/200$, and $u_i = 0$ for $i > 2$. Consequently, there is a constant $C_1 > 0$ such that

$$\lambda \geq \frac{199}{40000}u^2 \geq C_1 \left(\frac{n}{k} \right)^2. \quad (78)$$

Hence,

$$\mathbb{P}[U \text{ is independent}] \leq (1-p)^\lambda \stackrel{(78)}{\leq} \exp\left(-C_1 \frac{n^2 p}{k^2}\right) \stackrel{(4)}{\leq} \exp(-101n), \quad (79)$$

provided that the constant C_0 is large enough. As there are $\leq 2^n$ ways to choose U , the assertion follows from the union bound and (79).

As for the second assertion, let $G = G_{n,p,k}$ be a random k -colorable graph with planted coloring V_1, \dots, V_k . Consider a set $U \subset V_i$, $\#U \geq \frac{n}{2k \ln k}$. Then

$$\begin{aligned} \mathbb{P}\left[\#\bar{N}_G(U) \setminus V_i > \frac{n}{k}\right] &\leq \sum_{T \subset V \setminus V_i, \#T = n/k} \mathbb{P}[e_G(U, T) = 0] \leq \binom{n}{n/k} (1-p)^{n^2/(2k^2 \ln k)} \\ &\leq \exp\left(\frac{2n}{k} \ln(k) + \frac{n}{k} - \frac{n^2 p}{2k^2 \ln k}\right) \stackrel{(4)}{\leq} \exp\left(-\frac{100n}{\ln k}\right), \end{aligned}$$

provided that the constant C_0 is sufficiently large.

References

1. Achlioptas, D., Naor, A.: The two possible values of the chromatic number of a random graph. Proc. 36th STOC (2004) 587–593
2. Alon, N., Kahale, N.: A spectral technique for coloring random 3-colorable graphs. SIAM J. Computing **26** (1997) 1733–1748
3. Alon, N., Kahale, N.: Approximating the independence number via the ϑ -function. Math. Progr. **80** (1998) 253–264
4. Blum, A., Spencer, J.: Coloring random and semirandom k -colorable graphs. J. of Algorithms **19** (1995) 203–234
5. Bollobás, B.: The chromatic number of random graphs. Combinatorica **8** (1988) 49–55
6. Bollobás, B.: Random graphs. 2nd edition. Cambridge University Press (2001)
7. Bollobás, B., Scott, A.D.: Max cut for random graphs with a planted partition. Combinatorics, Probability and Computing **13** (2004) 451–474
8. Boppana, R.: Eigenvalues and graph bisection: an average-case analysis. Proc. 28th FOCS (1987) 280–285
9. Charikar, M.: On semidefinite programming relaxations for graph coloring and vertex cover. Proc. 13th SODA (2002) 616–620
10. Coja-Oghlan, A.: Solving NP-hard semirandom graph problems in polynomial expected time. To appear in the J. of Algorithms. Available at <http://www.informatik.hu-berlin.de/~coja/>
11. Coja-Oghlan, A.: Finding large independent sets in polynomial expected time. To appear in Combinatorics, Probability and Computing. Available at <http://www.informatik.hu-berlin.de/~coja/>
12. Coja-Oghlan, A., Goerdts, A., Lanka, A., Schädlich, F.: Techniques from combinatorial approximation algorithms yield efficient algorithms for random $2k$ -SAT. Theoret. Comp. Sci. **329** (2004) 1–45

13. Coja-Oghlan, A., Moore, C., Sanwalani, V.: MAX k -CUT and approximating the chromatic number of random graphs. Preprint (2003). A preliminary version appeared in the Proc. 30th ICALP (2003) 200–211. Available at <http://www.informatik.hu-berlin.de/~coja/>
14. Dyer, M., Frieze, A.: The solution of some NP-hard problems in polynomial expected time. *J. of Algorithms* **10** (1989) 451–489
15. Feige, U., Kilian, J.: Zero knowledge and the chromatic number. *J. Comp. Syst. Sci.* **57** (1998) 187–199
16. Feige, U., Kilian, J.: Heuristics for semirandom graph problems. *J. Comp. Syst. Sci.* **63** (2001) 639–671
17. Feige, U., Krauthgamer, R.: Finding and certifying a large hidden clique in a semirandom graph. *Rand. Struct. Alg.* **16** (2000) 195–208
18. Feige, U., Ofek, E.: Spectral techniques applied to sparse random graphs. Report MCS03-01, Weizmann Institute (2003)
19. Frieze, A., Jerrum, M.: Improved approximation algorithms for MAX k -CUT and MAX BISECTION. *Algorithmica* **18** (1997) 61–77
20. Frieze, A., McDiarmid, C.: Algorithmic theory of random graphs. *Rand. Struct. Alg.* **10** (1997) 5–42
21. Goemans, M.X., Kleinberg, J.: The Lovasz theta function and a semidefinite programming relaxation of vertex cover. *SIAM J. Disc. Math.* **11** (1998) 1–48
22. Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. of the ACM* **42** (1995) 1115–1145
23. Grötschel, M., Lovász, L., Schrijver, A.: Geometric algorithms and combinatorial optimization. Springer (1988)
24. Helmberg, C.: Semidefinite programming for combinatorial optimization. Habilitationsschrift. Report ZR-00-34, Zuse Institute Berlin (2000)
25. Janson, S., Łuczak, T., Ruciński, A.: Random Graphs. Wiley (2000)
26. Karger, D., Motwani, R., Sudan, M.: Approximate graph coloring by semidefinite programming. *J. of the ACM* **45** (1998) 246–265
27. Khanna, S., Linial, N., Safra, S.: On the hardness of approximating the chromatic number. *Combinatorica* **20** (2000) 393–415
28. Krivelevich, M.: Coloring random graphs – an algorithmic perspective. Proc. 2nd Colloquium on Mathematics and Computer Science (2002) 175–195.
29. Kučera, L.: Graphs with small chromatic number are easy to color. *Inf. Process. Let.* **30** (1989) 233–236
30. Lawler, E.L.: A note on the complexity of the chromatic number problem. *Inf. Process. Let.* **5** (1976) 66–67
31. Łuczak, T.: The chromatic number of random graphs. *Combinatorica* **11** (1991) 45–54
32. McSherry, F.: Spectral partitioning of random graphs. Proc. 42nd FOCS (2001) 529–537
33. Subramanian, C.R.: Minimum coloring random and semirandom graphs in polynomial average time. *J. of Algorithms* **33** (1999) 112–123
34. Subramanian, C.R., Fürer, M., Veni Madhavan, C.E: Algorithms for coloring semi-random graphs. *Rand. Struct. Alg.* **13** (1998) 125–158
35. Subramanian, C.R., Veni Madhavan, C.E.: General partitioning on random graphs. *J. of Algorithms* **42** (2002) 153–172
36. Szegedy, M.: A note on the θ number of Lovász and the generalized Delsarte bound. Proc. 35th FOCS (1994) 36–39