Coloring Sparse Random k-colorable Graphs in Polynomial Expected Time

Julia Böttcher

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

Abstract. Feige and Kilian [5] showed that finding reasonable approximative solutions to the coloring problem on graphs is hard. This motivates the quest for algorithms that either solve the problem in most but not all cases, but are of polynomial time complexity, or that give a correct solution on all input graphs while guaranteeing a polynomial running time on average only. An algorithm of the first kind was suggested by Alon and Kahale in [1] for the following type of random k-colorable graphs: Construct a graph $\mathcal{G}_{n,p,k}$ on vertex set V of cardinality n by first partitioning V into k equally sized sets and then adding each edge between these sets with probability p independently from each other. Alon and Kahale showed that graphs from $\mathcal{G}_{n,p,k}$ can be k-colored in polynomial time with high probability as long as $p \geq c/n$ for some sufficiently large constant c. In this paper, we construct an algorithm with polynomial expected running time for $k = 3$ on the same type of graphs and for the same range of p. To obtain this result we modify the ideas developed by Alon and Kahale and combine them with techniques from semidefinite programming. The calculations carry over to general k.

1 Introduction

The coloring problem on graphs remains one of the most demanding algorithmic tasks in graph theory. Since it is one of the classical $\mathcal{NP}$-hard problems (see [7]) it is unlikely that efficient coloring algorithms exist. If no exact answer to a problem can be found within a reasonable amount of time, one alternative is to search for approximation algorithms. However, for the coloring problem even this suboptimal approach fails. While the standard greedy heuristic with high probability does not use more than $2\chi(G)$ colors on a random input graph G, guaranteeing a similar performance ratio for all graphs is intractable under reasonable computational assumptions. In fact, Feige and Kilian [5] proved that for all $\epsilon > 0$ it is impossible to approximate the coloring problem within $n^{1-\epsilon}$, provided $\mathcal{ZPP} \neq \mathcal{NP}$, where n is the number of vertices of the input graph. Moreover, Khanna, Linial, and Safra [9] showed that coloring 3-colorable graphs with 4 colors is $\mathcal{NP}$-hard.

Accordingly, different approaches must be pursued. One possibility is to ask for algorithms that work with high probability. While finding algorithms of this type is not too difficult for dense k-colorable graphs [4,11,13], it turns out to be harder for sparse k-colorable graphs. For constructing such sparse graphs set $p = c/n$ for a constant c in the following process: Partition the vertex set V into k sets C_i of equal size and allow only edges between these sets, taking each one independently with probability p. We denote graphs obtained in this way by $\mathcal{G}_{n,p,k}$. The sets C_i are also called the *color classes* of $\mathcal{G}_{n,p,k}$.

In 1997, Alon and Kahale [1] established the following result for $\mathcal{G}_{n,p,k}$.

Theorem 1 (Alon and Kahale [1]). *Let $p > c/n$ for some sufficiently large constant c. Then there is a polynomial time algorithm for k-coloring $\mathcal{G}_{n,p,k}$ with high probability.*

But an algorithm that works with high probability has one drawback: For some inputs it does not provide any solution at all. Alternatively, we could require that the algorithm always gives a correct answer to the problem under study but performs well only on average: An algorithm $\mathcal{A}$ with running time $t_{\mathcal{A}}(G)$ on input G has *polynomial expected running time* on $\mathcal{G}_{n,p,k}$ if $\sum_G t_{\mathcal{A}}(G) \cdot \mathbf{P}[\, \mathcal{G}_{n,p,k} = G \,]$ remains polynomial. Here, the sum ranges over all graphs on n vertices. Observe that this is a stronger condition than to work correctly with high probability: An algorithm that k-colors $\mathcal{G}_{n,p,k}$ in polynomial expected running time also solves the k-coloring problem with high probability in polynomial time.

In this paper we present an algorithm for coloring sparse 3-colorable graphs with 3 colors in polynomial expected time.

Theorem 2. *If $p > c/n$ for some sufficiently large constant c, then there is an algorithm COLOR that 3-colors $\mathcal{G}_{n,p,3}$ in polynomial expected time.*

This improves on results of Subramanian [12] and Coja-Oghlan [2] and answers a question of Subramanian [12] and Krivelevich [10]. The calculations carry over to general k. The best known previous algorithm is due to Coja-Oghlan [2] and k-colors graphs from $\mathcal{G}_{n,p,k}$ in polynomial expected time if $np \geq c \cdot \max\bigl(k \ln n, k^2\bigr)$ for a sufficiently large constant c.

The main philosophy of COLOR can be described as follows. On input G, we start by executing a polynomial time algorithm $\mathcal{A}$: In the first step $\mathcal{A}$ determines an initial coloring of G which colors all but a constant fraction of G correctly with high probability. $\mathcal{A}$ then refines this initial coloring by using different combinatorial methods which are modifications of the methods used by Alon and Kahale. With high probability this results in a valid coloring of G. However, since we are interested in returning a valid coloring for all graphs, COLOR also has to take care of exceptional cases. In the case that $\mathcal{A}$ does not produce a valid coloring of G, we therefore proceed by removing a set Y of vertices from G, rerun $\mathcal{A}$ on $G \setminus Y$ and treat Y with brute force coloring methods. In the beginning, Y contains only a single vertex. We repeat this procedure and gradually increase $|Y|$ until G is finally properly colored. We verify that COLOR has polynomial expected running time by showing that $\mathcal{A}$ can handle all but a small number of vertices for most graphs from $\mathcal{G}_{n,p,3}$.

In addition, and in contrast to Alon and Kahale, we apply the concept of semidefinite programming in order to obtain a good initial coloring of $\mathcal{G}_{n,p,3}$ in the first stage of COLOR. To this end we use the semidefinite program $\mathcal{SDP}_3$ introduced by Frieze and Jerrum [6]. The value of $\mathcal{SDP}_3$ on $\mathcal{G}_{n,p}$ has been investigated by Coja-Oghlan, Moore, and Sanwalani [3]. We use their result to show that $\mathcal{SDP}_3$ behaves similarly on $\mathcal{G}_{n,p,3}$. This will then allow us to construct a coloring of $\mathcal{G}_{n,p,3}$ from a solution of $\mathcal{SDP}_3$ that already colors all but a small linear fraction of the input graph correctly. Similar methods have been used by Coja-Oghlan [2].

The remainder of this paper is structured as follows: In Section 2 we investigate the behaviour of $\mathcal{SDP}_3$ on $\mathcal{G}_{n,p,3}$, in Section 3 we give the details of our coloring algorithm, in Section 4 its analysis and in Section 5 some concluding remarks.

2 The Value of $\mathcal{SDP}_3$ on Graphs from $\mathcal{G}_{n,p,3}$

Recall that a *k-cut* of a graph G is a partition of $V(G)$ into k disjoint sets $V_1, \ldots, V_k$, its weight is the total number of edges crossing the cut, and that MAX-k-CUT is the the problem of finding a k-cut of maximum weight.

In the algorithm COLOR we make use of the following SDP relaxation $\mathcal{SDP}_3$ of MAX-3-CUT due to Frieze and Jerrum [6] which provides an upper bound for MAX-3-CUT:

$$\begin{aligned} \max \quad & \sum_{vw \in E(G)} \frac{2}{3}\left(1 - \langle \mathbf{x}_v \,|\, \mathbf{x}_w \rangle\right) & \\ \text{s.t.} \quad & \|\mathbf{x}_v\| = 1 & \forall v \in V, \\ & \langle \mathbf{x}_v \,|\, \mathbf{x}_w \rangle \geq -\frac{1}{2} & \forall v, w \in V. \end{aligned}$$

Here, the maximum runs over all vector assignments $(\mathbf{x}_v)_{v \in V(G)}$ obeying $\mathbf{x}_v \in \mathbb{R}^{|V|}$. Observe that, if G_1 is a subgraph of G_2, then $\mathcal{SDP}_3(G_1) \leq \mathcal{SDP}_3(G_2)$. One way to realize a feasible solution of $\mathcal{SDP}_3$ corresponding to a 3-cut V_1, V_2, V_3 of G is to assign the same vector $\mathbf{s}_i$ to each vertex in C_i in such a way that $\langle \mathbf{s}_i \,|\, \mathbf{s}_j \rangle = -1/2$ for $i \neq j$.

Moreover, there is an obvious connection between maximum 3-cuts and 3-colorings: In the case of a 3-colorable graph G a maximum 3-cut simply contains all edges. Then, we know that each edge contributes exactly 1 to the value of $\mathcal{SDP}_3$. In this case, the special feasible solution to $\mathcal{SDP}_3$ discussed above is optimal. Conversely, if the optimal solution of $\mathcal{SDP}_3(G)$ has this structure, then it is clearly easy to read off a proper 3-coloring of G from this solution. Unfortunately, the position of the vectors $\mathbf{x}_v$ can get "far away" from this ideal picture in general. In this section we show however that with high probability such a scenario does not occur in the case of random 3-colorable graphs from $\mathcal{G}_{n,p,3}$. Although for such graphs the vectors corresponding to vertices of one color class do not necessarily need to be equal, most of them will be comparably close. Similar techniques were used in [2].

In [3] Coja-Oghlan, Moore, and Sanwalani studied the behaviour of $\mathcal{SDP}_3$ on $\mathcal{G}_{n,p}$. They obtained the following result, which will be the key ingredient to our analysis of $\mathcal{SDP}_3$ on graphs from $\mathcal{G}_{n,p,3}$.

Theorem 3 (Coja-Oghlan, Moore & Sanwalani [3]). *If $p \geq c/n$ for sufficiently large c then*

$$\mathcal{SDP}_3(\mathcal{G}_{n,p}) \leq \frac{2}{3}\binom{n}{2}p + \mathcal{O}\Big(\sqrt{n^3 p(1-p)}\Big) \tag{1}$$

with probability at least $1 - \exp(-3n)$.

Note that $2\binom{n}{2}p/3$ is also the size of a random 3-cut in $\mathcal{G}_{n,p}$. So Theorem 3 estimates the difference of the sizes of a maximum 3-cut and a random 3-cut in $\mathcal{G}_{n,p}$.

Let $G = (V, E) \in \mathcal{G}_{n,p,3}$. We can construct a random graph $G^* \in \mathcal{G}_{n,p}$ from G by inserting additional edges with probability p within each color class. The following lemma investigates the effect of this process on the value of $\mathcal{SDP}_3$.

Lemma 1. *Consider a graph $G^* = (V, E^*)$ from $\mathcal{G}_{n,p}$ with $V = [n]$ and let $G = (V, E)$ be the subgraph of G^* with edges $E = E^* \cap \{vw \mid \lceil 3v/n \rceil \neq \lceil 3w/n \rceil\}$. Then for some constant c' not depending on d*

$$\mathcal{SDP}_3(G^*) - \mathcal{SDP}_3(G) \leq c'\sqrt{n^3 p} \tag{2}$$

with probability at least $1 - \exp(-5n/2)$.

Proof. In order to establish this result we prove that

$$\frac{2}{3}\binom{n}{2}p - \frac{c'}{2}\sqrt{n^3 p} \leq \mathcal{SDP}_3(G) \leq \mathcal{SDP}_3(G^*) \leq \frac{2}{3}\binom{n}{2}p + \frac{c'}{2}\sqrt{n^3 p}$$

holds with the same probability. In fact, the second inequality holds by construction since G is a subgraph of G^* and the third inequality is asserted by Theorem 3 if we choose c' accordingly. Thus it remains to show the first inequality. This is obtained by a straightforward application of the Chernoff bound and the fact that $\mathcal{SDP}_3(G) = |E|$ as mentioned earlier.

Equation (2) asserts that the values of $\mathcal{SDP}_3$ for G and G^* are not likely to differ much, if the additional edges within the color classes C_i of G are chosen at random. It follows that in an optimal solution to $\mathcal{SDP}_3(G)$, most of the vectors corresponding to vertices of C_i for a particular i can not be far apart. This is shown in the next lemma.

If not stated otherwise, we consider $\mathcal{SDP}_3$ on input G from now on. Let $(\mathbf{x}_v)_{v \in V(G)}$ be an optimal solution to $\mathcal{SDP}_3(G)$. Then we call

$$\mathbf{N}^\mu(v) := \{v' \in V \mid \langle \mathbf{x}_v \,|\, \mathbf{x}_{v'} \rangle > 1 - \mu\}$$

the *μ-neighborhood* of v.

Lemma 2. *For fixed ϵ with $0 < \epsilon < 1/2$ there is a constant $0 < \mu < 1/2$ such that for any μ' with $\mu \le \mu' < 1/2$ the following holds with probability greater than $1 - \exp(-7n/3)$: For each $i \in \{1, 2, 3\}$ there is a vertex $v_i \in C_i$ such that the set $\mathbf{N}^{\mu}(v_i)$ contains at least $(1-\epsilon)n/3$ vertices of the color class C_i and the set $\mathbf{N}^{\mu'}(v_i)$ contains at most $\epsilon n/3$ vertices from other color classes C_j $(j \neq i)$.*

For proving this lemma, we first observe that the edges uv of G^* with u, v in one color class C_i of G and $\langle \mathbf{x}_u \,|\, \mathbf{x}_v \rangle$ small, form a random subgraph of G^*. From Lemma 1 we can then deduce the first statement of Lemma 2. The second statement follows since $\mathbf{N}^{\mu'}(v)$ induces an empty graph in G for $\mu' < 1/2$. We omit the details.

In the following we will call a vertex $v \in C_i$ obeying the properties asserted by Lemma 2 an (ϵ, μ, μ')-*representative* for color class C_i. In the case $\mu' = \mu$ we omit the parameter μ'.

3 The Algorithm

Roughly speaking, there are two basic principles underlying the mechanisms of COLOR. On the one hand a number of steps, also called the *main steps*, aim at constructing a valid 3-coloring of the input graph with sufficiently high probability. These are the *initial step*, the *iterative recoloring step*, the *uncoloring step* and the *extension step*. However, on atypical graphs this approach might fail. For guaranteeing a valid output for each input, COLOR also has to handle this case; possibly by using computationally expensive methods. Accordingly, the purpose of the remaining operations is to fix the mistakes of the main steps on such atypical graphs. This constitutes the second principle, the so-called *recovery procedure* of COLOR.

The details of COLOR are given in Algorithm 1. We now briefly describe the main steps and then turn to the recovery procedure.

The initial step: (Steps 1, 5, and 6) This step is concerned with finding an initial coloring Υ_0 of the input graph $G = (V, E)$ such that Υ_0 fails on at most ϵn vertices of G. Here, ϵ is small but constant. To obtain Υ_0 we apply the SDP relaxation $\mathcal{SDP}_3$ of MAX-3-CUT. An optimal solution of this semidefinite program can efficiently be computed within any numerical precision (cf. [8]). This solution gives rise to the coloring Υ_0 by grouping vertices whose corresponding vectors have large scalar product into the same color class. In the algorithm COLOR we use a randomized method for this grouping process.

The iterative recoloring step: (Step 8) This step refines the initial coloring in order to find a valid coloring of a much larger vertex set by repeating the following step at most a logarithmic number of times: Assign to each vertex in G the color that is the least favorite among its neighbors. In Section 4 we will show that this approach is indeed successful, in the sense that with sufficiently high probability at most $\alpha_0 n$ vertices are still colored incorrectly after the iterative recoloring step, where α_0 is of order $\exp(-np)$.

The uncoloring step: (Step 9) This step proceeds iteratively as well. In each iteration, the uncoloring step uncolors all vertices that have less than $np/6$

Algorithm 1: COLOR(G)

Input: a graph $\mathcal{G}_{n,p,3} = G = (V, E)$
Output: a valid coloring of $\mathcal{G}_{n,p,3}$

begin
1 let $(\mathbf{x}_v)_{v \in V(G)}$ be an optimal solution of $\mathcal{SDP}_3(G)$;
2 **for** $0 \leq y \leq n$ **do**
3 **foreach** $Y \subset V$ with $|Y| = y$ and **each** valid 3-coloring Υ_Y of Y **do**
4 **for** $\mathcal{O}(n)$ times **do**
/** `The initial step` **/
5 Randomly choose three vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \{\mathbf{x}_v | v \in V(G)\}$;
6 Extend Υ_Y to a coloring Υ_0 of G by setting $\Upsilon_0(v) := i$ for all $v \in G - Y$ where i is such that $\langle \mathbf{x}_v \,|\, \mathbf{x}_i \rangle$ is maximal ;
7 **for** $t = \log n$ **downto** $t = 0$ **do**
/** `The iterative recoloring step` **/
8 **for** $0 \leq s < t$ **do**
Construct a coloring Υ_{s+1} of G with $\Upsilon_{s+1}(v) := \Upsilon_Y(v)$ for $v \in Y$ and $\Upsilon_{s+1}(v) := i$ for $v \notin Y$ where i minimizes $|\mathbf{N}(v) \cap \Upsilon_s^{-1}(i)|$;
Set $\Upsilon' := \Upsilon_t$;
/** `The uncoloring step` **/
9 **while** $\exists v \in G - Y$ with $\Upsilon'(v) = i$ and $|\{w | w \in \mathbf{N}(v), \Upsilon'(w) = j\}| < np/6$ for some $j \neq i$ **do**
uncolor v in Υ' ;
/** `The extension step` **/
10 **if** each component of uncolored vertices is of size at most $\alpha_0 n$ **then**
Extend the partial coloring Υ' to a coloring Υ of G by exhaustively trying each coloring of each component in the set of uncolored vertices ;
11 **if** Υ is a valid coloring of G **then**
return $\Upsilon(G)$ and stop;

end

neighbors of some color other than their own. Observe that in a "typical" graph from $\mathcal{G}_{n,p,3}$ a "typical" vertex and its neighbors will not have this property if they are colored correctly.

The extension step: (Step 10) Here, an exact coloring method is used to extend the partial coloring obtained to the whole graph $\mathcal{G}_{n,p,3}$. In this process the components induced on the uncolored vertices are treated seperately. On each such component K, the algorithm tries all possible colorings until it finds one that is compatible to the coloring of the rest of G.

The main steps are all we need for 3-coloring $\mathcal{G}_{n,p,3}$ with high probability This will be formally proven in Section 4; the analysis of the recoloring, the uncoloring and the extension step is similar to that of Alon and Kahale [1].

If the main steps do not produce a valid coloring of G the recovery procedure (the loops in Steps 2 to 7) comes into play. The concept is as follows. Assume that for an input graph G the main steps of COLOR produce a correct coloring on the subgraph induced by $V \setminus Y$ for some $Y \subset V$ but "fail" on Y. Then an easy way of "repairing" the coloring obtained is to exhaustively test all valid colorings of Y. Of course, we neither know this set Y nor its size $|Y|$. To deal with these two problems, COLOR proceeds by trying all possible subsets Y of V with $|Y| = y$. Here, we start with $y = 0$ and then gradually increase the value of y until a valid coloring of G is determined. This is performed in Steps 2 and 3 of the recovery procedure. We also call Step 3, where all colorings of all vertex sets of size y are constructed, the *brute force coloring method* or *repair mechanism* for these sets.

Since the size of Y in the recovery procedure (Step 3) is increased until a valid coloring is obtained, the correctness of Algorithm 1 is inherent (ultimately, a proper coloring will be found in the last iteration, when Y contains all vertices of G) . An analysis of the expected running time of Algorithm 1 will be presented in the next section.

4 Analysis of the Algorithm

In the following we assume that G is a graph sampled from $\mathcal{G}_{n,p,3}$ where $d := np > c$ for some sufficiently large constant c. Moreover, let $\mathbf{3\text{-}COL}\,(x) := 3^x$ be the time needed to find all 3-colorings of a graph of order x.

4.1 The Initial Step

Lemma 2 guarantees that, given an optimal solution of $\mathcal{SDP}_3(G)$, we can construct a reasonably good initial coloring Υ_0 via the sets $\mathbf{N}^\mu(v_i)$ by choosing an appropriate representative v_i for each color class C_i and setting $\Upsilon_0(v) := i$ for all $v \in N^\mu(v_i)$. Here, ties are broken arbitrarily and vertices not appearing in any of the sets $N^\mu(v_i)$ get assigned an arbitrary color.

Thus it remains to determine appropriate representatives. The easiest way is to simply try all different triples of vertices from G as representatives. This, however, introduces an extra factor of n^3 in the running time. In order to reduce this factor to a linear one, Algorithm 1 proceeds differently. Let v_1, v_2 and v_3 be (ϵ', μ, μ')-representatives of the color classes C_1, C_2, and C_3, respectively. We will make use of the following lemma.

Lemma 3. *If v_i is a (ϵ', μ, μ')-representative for C_i and $\mu' > 4\mu + \sqrt{2\mu}$, then each vertex $v \in \mathbf{N}^\mu(v_i)$ is an $(\epsilon', 4\mu)$-representative of color class C_i.*

By choosing $\mu' > 4\mu + \sqrt{2\mu}$ (and μ sufficiently small such that $\mu' < 1/2$) we therefore get at least $(1 - \epsilon') \cdot n/3$ representatives per color class. But then the probability of obtaining a set of representatives for G by picking three vertices r_1, r_2, r_3 from V at random is at least $(1 - 3 \cdot \epsilon')/9$. Repeating this process raises the probability of success. More specifically, the probability that in $c'n$ trials none yields a triple of $(\epsilon', 4\mu)$-representatives is smaller than $(8/9 + \epsilon'/3)^{c'n}$. Here, $c' > 0$ is an arbitrary constant. Observe additionally that if r_1, r_2, r_3 form a triple of $(\epsilon', 4\mu)$-representatives for G, then $\langle \mathbf{x}_v \,|\, \mathbf{x}_{r_i} \rangle > \left\langle \mathbf{x}_v \,\middle|\, \mathbf{x}_{r_j} \right\rangle$ for at least $(1 - 2\epsilon')n/3$ vertices $v \in C_i$ if $i \neq j$. Let $2 \cdot \epsilon' = \epsilon$. This guarantees a coloring Υ_0 of G that colors at least ϵn vertices of G correctly by assigning each vertex v the color i such that $\langle \mathbf{x}_v \,|\, \mathbf{x}_{r_i} \rangle$ is maximal.

The strategy just described is applied in Step 4 of Algorithm 1. We conclude that this randomized approach gives rise to a valid coloring of an $(1-\epsilon)$-fraction of the graph with probability at least

$$1 - \exp\left(-\frac{7}{3}n\right) - (8/9 + 2\epsilon/3)^{c'n} \geq 1 - 2\exp\left(-\frac{7}{3}n\right) \geq 1 - 10^{-n} \qquad (3)$$

for c' and n sufficiently large and ϵ small enough, e.g. $c' = 30$ and $\epsilon < 1/10$. Here, the probability is taken with respect to the input graphs $\mathcal{G}_{n,p,k}$ and to the random choices of representatives.

Now, consider the coloring Υ_0 constructed in Step 6 of Algorithm 1 and let F_{SDP} be a vertex set of minimal cardinality such that Υ_0 colors at most $\epsilon n/3$ vertices incorrectly in each set $C_i \setminus F_{SDP}$. The following lemma is an immediate implication of Equation (3) and Lemma 2.

Lemma 4. *For all $y > 0$ the following relation holds:* $\mathbf{P}[\,|F_{SDP}| \geq 1\,] \leq 10^{-n}$.

4.2 The Iterative Recoloring Step

After the initial coloring $\Upsilon_0(G)$ is constructed, Step 8 of Algorithm 1 aims at improving this coloring iteratively. We show that this attempt is indeed successful on a large subgraph H of G with high probability.

Let H be the subgraph of G obtained by the following process:

1. Delete all vertices in $\overline{H}^+ := \left\{ v \in V \,\middle|\, v \in C_i, \exists j \neq i :\ deg_{C_j}(v) > (1+\delta)\frac{d}{3} \right\}$
2. Delete all vertices in $\overline{H}^- := \left\{ v \in V \,\middle|\, v \in C_i, \exists j \neq i :\ deg_{C_j}(v) < (1-\delta)\frac{d}{3} \right\}$
3. Iteratively delete all vertices having more than $\delta d/3$ neighbors that were deleted earlier in C_i for some i, i.e., delete all vertices in $\bigcup_{0<l} \overline{H}^l$, where $\overline{H}^0 := \overline{H}^+ \cup \overline{H}^-$ and

$$\overline{H}^l := \left\{ v \in V \,\middle|\, \exists i : \mathbf{N}(v) \cap C_i \cap \bigcup_{l'<l} \overline{H}^{l'} > \delta \frac{d}{3} \right\}$$

for $l > 0$.

We also denote $G - H$ by $\overline{H}$. The lemma below shows that H spans a large subgraph of G with high probability.

Lemma 5. *Let $0 < \alpha < \frac{1}{2}$ and $0 < \delta < \frac{1}{2}$ be constant in the definition of H. Then*

$$\mathbf{P}\left[\,|\overline{H}| \geq \alpha n\,\right] \leq \exp(-\left(\log\alpha + \Omega\left(d\right)\right) \cdot \alpha n) + \exp(\Omega\left(d \cdot \log\alpha \cdot \alpha n\right)). \qquad (4)$$

This lemma follows from the observation that it is unlikely that many vertices are deleted in the first two steps of the construction of H. But then it is also unlikely that many vertices are deleted in the iterative step.

Now, we can use the structural properties of H to show that the algorithm succeeds on H with high probability. For this, we prove that with high probability the number of vertices in H that are colored incorrectly decreases by more than a factor of 2 in each of the iterations of the recoloring step. If this performance is actually achieved, we call the corresponding iteration *successful*, otherwise we say that it *fails*. Algorithm 1 performs at most $\log n$ of these iterations. Afterwards, either the entire graph H is colored correctly or one of the iterations failed. In the latter case Algorithm 1 runs the iterative recoloring step until just before the iteration, say iteration t, when it fails for the first time. The algorithm then proceeds by exhaustively trying all colorings on all subsets of G of size y and thus fixes the coloring of H in this way. However, since Algorithm 1 can not discover whether a particular iteration of the recoloring step succeeds or fails another iteration is necessary at this point. Algorithm 1 applies the strategy of simply trying to repair each of the iterations of the recoloring step subsequently, starting with the last one and proceeding until it reaches iteration t. This explains the innermost loop of the recovery procedure (Step 7 of Algorithm 1). Once the recovery procedure repaired iteration t, all vertices of H are colored correctly.

Lemma 6. *Consider the first iteration of the recoloring step that fails and let $F_H \subseteq H$ denote the set of vertices of H that were colored incorrectly in H before this iteration. Then*

$$\mathbf{P}[\,|F_H| = \alpha n\,] \leq \exp(\Omega\,(d \cdot \log \alpha \cdot \alpha n))$$

for δ sufficiently small but constant in the definition of H.

This follows from the fact that H has good expansion properties and that vertices in H do not have many neighbors outside of H.

4.3 The Uncoloring Step

Note that, if H is colored correctly before the application of the uncoloring step, no vertex of H gets uncolored by this procedure. Indeed, since each vertex v in H has at least $(1-\delta)d/3$ neighbors in $C_i \cap H$ for each i such that $v \notin C_i$ and all these neighbors are colored with color i, v does not get uncolored as long as $\delta < 1/2$.

The following Lemma shows that vertices $v \notin H$ that were not colored correctly by the iterative recoloring step are likely to get uncolored in the uncoloring step.

Lemma 7. *Let $F_\Upsilon \subset G - H$ be the set of vertices that are colored incorrectly and remain colored after the execution of the uncoloring step. Then*

$$\mathbf{P}[\,|F_\Upsilon| = \alpha n\,] \leq \exp(\Omega\,(d \cdot \log \alpha \cdot \alpha n))$$

for $0 < \alpha < \frac{1}{2}$.

Proof. If a vertex v in C_i is colored incorrectly, say with color j, and remains colored after the uncoloring step, v must have at least $d/6$ neighbors of color i. Since v is not adjacent to any vertex in its own color class C_i, all these neighbors are elements of F_Υ as well. Hence, the lemma follows from an estimation of the probability that there is some set $Y \subset V(G)$ with $|Y| = \alpha n$ and minimum degree at least $d/6$.

4.4 The Extension Step

Knowing that the uncoloring step succeeds in uncoloring all vertices of wrong color with high probability, we are now left with the task of assigning a new color to these uncolored vertices. In Algorithm 1 this is taken care of by the extension step (Step 10). Using similar techniques as those developed by Alon and Kahale in [1], we show that all components induced on the set of uncolored vertices are likely to be rather small. Recall that $\alpha_0 = \exp(-\mathcal{O}(d))$.

Lemma 8. *For $\alpha < \alpha_0$,*

$$\mathbf{P}\left[\text{ there is a component of order } \alpha n \text{ in } \overline{H}\ \right] \leq \left(\frac{d}{\exp(\Omega\,(d))}\right)^{\alpha n}.$$

From this lemma it follows that with high probability a valid coloring of H can indeed be extended to the whole graph G by Step 10 of Algorithm 1 as long as $\overline{H}$ does not get too large.

Lemma 9. *The extension step (Step 10) of Algorithm 1 has polynomial expected running time.*

Proof. As explained, in Step 10 of Algorithm 1 an exact coloring method is used to extend the partial coloring obtained in earlier steps to the whole graph G. In this process the components induced on the vertices uncolored by the uncoloring step are considered independently. On each such component the algorithm tries all possible colorings until it finds one that is compatible to the coloring of the rest of G. Trivially there are at most n components in the set of uncolored vertices and so it suffices to show that the probability that a component of $G - H$ has αn vertices multiplied by **3-COL** (αn) remains small for all $\alpha < \alpha_0$ since the extension step is only executed for $\alpha < \alpha_0$:

$$\begin{aligned}&\mathbf{P}[\text{there is a component of order } \alpha n \text{ in } G - H\ ] \cdot \mathbf{3\text{-}COL}\,(\alpha n)\\ &\leq \left(\frac{d}{\exp(\Omega\,(d))}\right)^{\alpha n} \cdot 3^{\alpha n} \leq \left(\frac{3d}{\exp(\Omega\,(d))}\right)^{\alpha n} = \mathcal{O}(1)\,.\end{aligned}$$

4.5 The Expected Running Time of COLOR

All main steps of Algorithm 1, i.e., the construction of the initial coloring, the recoloring step, and the uncoloring step are executed in polynomial time.

Moreover, Lemma 8 guarantees that the extension step of COLOR has polynomial expected running time. It therefore remains to investigate the recovery procedure consisting of the loops in Steps 2, 3, 4 and 7 of Algorithm 1.

The results derived in the last few subsections estimate the probabilities that one of the main steps of Algorithm 1 fails on a vertex set Y of size y. As explained in Section 3, these vertex sets are taken care of by the recovery procedure. The polynomial expected running time of Algorithm 1 is a consequence of the exponentially small probabilities in the previous lemmas. This is shown in Lemma 10 and it immediately implies Theorem 2.

Lemma 10. *The recovery procedure (i.e., Steps 2, 3, 4and 7) of Algorithm 1 has polynomial expected running time.*

Proof. Consider the vertex set Y from Algorithm 1 that is colored correctly in Step 3 of the recovery procedure and let $t(y)$ be the time the algorithm needs to execute this step in the case $|Y| = y$. Further, denote by F the set Y used in the iteration when the algorithm finally obtains a valid coloring. The expected running time $\mathbf{E}[\,t\,]$ of the repair mechanism can then be written as

$$\begin{aligned}\mathbf{E}[\,t\,] &= \sum_{y\leq n} \mathbf{P}[\,|F| = y\,] \cdot t(y) \\ &\leq \mathcal{O}(n) \sum_{y\leq n} \mathbf{P}[\,|F| = y\,] \cdot \binom{n}{y} \mathbf{3\text{-}COL}\,(y)\,.\end{aligned}$$

Recall the definition of H in subsection 4.2 and that F_H are those vertices in H which are colored incorrectly after the recoloring step. F_{SDP} are those vertices that need to be assigned a different color for obtaining a valid coloring on an $(1-\epsilon)$-fraction of G after the inital phase and F_{Υ} are those that are colored incorrectly after the uncoloring step. Moreover, $\alpha_0 = \exp(-\mathcal{O}(d))$.

We bound $\mathbf{P}[\,|F| = y\,]$ by rewriting F as sum of F_{SDP}, F_H, F_{Υ} and possibly $\overline{H}$. For this, observe that $F = F_{SDP} \cup F_H \cup F_{\Upsilon}$. However, we use this partitioning of F only in the case that $|\overline{H}| \leq \alpha_0 n$. If $|\overline{H}| > \alpha_0 n$ we use $F \subset F_{SDP} \cup F_H \cup \overline{H}$. Since all probabilities involved are monotone decreasing, we get

$$\begin{aligned}\mathbf{P}[\,|F| = y\,] \leq &\sum_{\substack{y_1+y_2+y_3=y \\ y_3\leq\alpha_0 n}} \mathbf{P}[\,|F_{SDP}| = y_1, |F_H| = y_2, |F_{\Upsilon}| = y_3\,] \\ &+ \sum_{\substack{y_1+y_2+y_3=y \\ y_3>\alpha_0 n}} \mathbf{P}\big[\,|F_{SDP}| = y_1, |F_H| = y_2, |\overline{H}| = y_3\,\big]\,.\end{aligned}$$

One of the vertex sets involved in the conjunctions of this sum certainly contains more than $y/3$ vertices and so

$$\begin{aligned}&\mathbf{P}[\,|F| = y\,] \\ &\leq n^3 \begin{cases} \max\big(\mathbf{P}\big[\,|F_{SDP}| = \frac{y}{3}\,\big], \mathbf{P}\big[\,|F_H| = \frac{y}{3}\,\big], \mathbf{P}\big[\,|F_{\Upsilon}| = \frac{y}{3}\,\big]\big) & \text{if } \ \frac{y}{3} \leq \alpha_0 n, \\ \max\big(\mathbf{P}\big[\,|F_{SDP}| = \frac{y}{3}\,\big], \mathbf{P}\big[\,|F_H| = \frac{y}{3}\,\big], \mathbf{P}\big[\,|\overline{H}| = \frac{y}{3}\,\big]\big) & \text{otherwise.}\end{cases}\end{aligned}$$

The lemma then follows from Lemmas 4, 5, 6, and 7. We omit the details.

5 Concluding Remarks

We proved that random 3-colorable graphs taken from $\mathcal{G}_{n,p,3}$ can be 3-colored in polynomial expected time if $p \geq c/n$, where c is some sufficiently large constant. The same methods can be used for obtaining a similar result for $\mathcal{G}_{n,p,k}$ with values of k other than 3. More precisely, the calculations carry over directly to arbitrary k for $pn \geq c_k$ where c_k is a constant depending on k.

One remaining question is whether it is possible to design an algorithm for coloring $\mathcal{G}_{n,p,k}$ in polynomial expected time for all values of p. In particular, it is not clear how to deal with the case that pn is constant but much smaller than c_k.

6 Acknowledgements

I am grateful to Amin Coja-Oghlan for many helpful discussions.

References

1. N. Alon and N. Kahale. A spectral technique for coloring random 3-colorable graphs. *SIAM Journal on Computing*, 26(6):1733–1748, 1997. 1, 1, 1, 3, 4.4, A.8
2. A. Coja-Oghlan. Coloring semirandom graphs optimally. In *Proceedings of the 31st International Colloquium on Automata, Languages and Programming*, pages 383–395, 2004. 1, 2
3. A. Coja-Oghlan, C. Moore, and V. Sanwalani. MAX k-CUT and approximating the chromatic number of random graphs. In *Proceedings of the 30th International Colloquium on Automata, Languages and Programming*, pages 200–211, 2003. 1, 2, 3
4. M. E. Dyer and A. M. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10:451–489, 1989. 1
5. U. Feige and J. Kilian. Zero knowledge and the chromatic number. *Journal of Computer and System Sciences*, 57(2):187–199, 1998. 1, 1
6. A. M. Frieze and M. Jerrum. Improved approximation algorithms for MAX k-CUT and MAX BISECTION. *Algorithmica*, 18:61–77, 1997. 1, 2
7. M. R. Garey and D. S. Johnson. *Computers and Intractability*. W.H. Freeman and Company, 1979. 1
8. M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, Berlin, 1993. 3
9. S. Khanna, N. Linial, and S. Safra. On the hardness of approximating the chromatic number. *Combinatorica*, 20(3):393–415, 2000. 1
10. M. Krivelevich. Deciding k-colorability in expected polynomial time. *Information Processing Letters*, 81:1–6, 2002. 1
11. L. Kučera. Expected behavior of graph colouring algorithms. In *Proceedings of the 1977 International Conference on Fundamentals of Computation Theory*, pages 447–451, 1977. 1
12. C. R. Subramanian. Algorithms for coloring random k-colorable graphs. *Combinatorics, Probability and Computing*, 9:45–77, 2000. 1
13. J. S. Turner. Almost all k-colorable graphs are easy to color. *Journal of Algorithms*, 9:253–261, 1988. 1

A Omitted Proofs

A.1 Definitions and Inequalities

For two vertex sets X and Y in a graph $G = (V, E)$ we denote by $\mathbf{e}(X, Y)$ the cardinality of $\{xy \in E | x \in X, y \in Y\}$, by $deg(v)$ the degree of a vertex $v \in V$ and by $mindeg(X)$ the minimum of $\{deg(x) | x \in X\}$.

For bounding binomial coefficients we frequently use the *binary entropy function* $\mathrm{H}(x) = -x \log x - (1-x)\log(1-x)$ where $x \in (0, 1)$. Note that $-x \log x$ has a unique local maximum at $x = 1/2$ and $-(1/4)\log(1/4) > -(1-1/4)\log(1-1/4)$. By symmetry it follows that

$$\mathrm{H}(x) \leq -2 \cdot x \log x \qquad \text{for} \quad x \leq \frac{1}{2}.$$

This implies a bound for binomial coefficients ($x \in (0, 1)$):

$$\binom{n}{xn} \leq \exp(\mathrm{H}(x) \cdot n) \leq \exp(-2x \log x \cdot n)\,. \tag{5}$$

For $X \in \mathrm{Bin}(n, p)$ with expectation $\lambda = np$ and $t \geq 0$, we use the following *Chernoff bounds*:

$$\mathbf{P}[\, X \geq \mathbf{E}[\, X\,] + t\,] \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right), \tag{6}$$

$$\mathbf{P}[\, X \leq \mathbf{E}[\, X\,] - t\,] \leq \exp\left(-\frac{t^2}{2\lambda}\right). \tag{7}$$

A.2 Details of the Proof of Lemma 1

For convenience we restate the outline of the proof given in section 2 and add the missing details.

In order to establish equation (2) we prove that

$$\frac{2}{3}\binom{n}{2}p - \frac{c'}{2}\sqrt{n^3 p} \leq \mathcal{SDP}_3(G) \leq \mathcal{SDP}_3(G^*) \leq \frac{2}{3}\binom{n}{2}p + \frac{c'}{2}\sqrt{n^3 p}$$

holds with probability $1 - \exp(-5n/2)$. In fact, the second inequality holds by construction since G is a subgraph of G^* and the third inequality is asserted by Theorem 3 if we choose c' accordingly. Thus it remains to show the first inequality. This is obtained by a straightforward application of the Chernoff bound (7) and the fact that $\mathcal{SDP}_3(G) = |E|$ as mentioned earlier:

The number of edges $|E|$ is a binomially distributed random variable with expectation $2\binom{n}{2}p/3$ and thus

$$\begin{aligned}
&\mathbf{P}\left[\frac{2}{3}\binom{n}{2}p - \frac{c'}{2}\sqrt{n^3 p} \leq |E|\right] \\
&= \mathbf{P}\left[\mathbf{E}[\,|E|\,] - \frac{c'}{2}\sqrt{n^3 p} \leq |E|\right] = 1 - \mathbf{P}\left[|E| < \mathbf{E}[\,|E|\,] - \frac{c'}{2}\sqrt{n^3 p}\right] \\
&\geq 1 - \exp\left(-\frac{{c'}^2 n^3 p}{2 \cdot \frac{2}{3}\binom{n}{2}p}\right) > 1 - \exp\left(-{c'}^2 \frac{3}{2} n\right).
\end{aligned}$$

This settles the proof since we can certainly choose c' in such a way that $c' > 2$ in Theorem 3.

A.3 Proof of Lemma 2

Let $d := pn > c$. We first show that the first statement of Lemma 2 holds with probability at least $1 - 2\exp(-5n/2)$ for a fixed i. Then, we proceed by proving that for any pair $i \neq j$ and appropriate choices of v_i and μ' the probability that $\mathbf{N}^\mu(v_i)$ contains more than $\epsilon n/6$ vertices from C_j is at most $\exp(-\Omega(dn))$. Since

$$3 \cdot 2 \cdot \exp\left(-\frac{5}{2}n\right) + 6 \cdot \exp(-\Omega(dn)) < 7 \cdot \exp\left(-\frac{5}{2}n\right) \leq \exp\left(-\frac{7}{3}n\right)$$

for d and n sufficiently large this establishes the lemma (the probabilities are multiplied by the number of choices for i and j).

For proving the first part of Lemma 2 let $G_i^\mu = (C_i, E_i^\mu)$ be the graph on one color class C_i of G with edge set $E_i^\mu := \{vw | \langle \mathbf{x}_v \,|\, \mathbf{x}_w\rangle \leq 1-\mu\}$ (recall that $(\mathbf{x}_v)_{v\in V(G)}$ with $\mathbf{x}_v \in \mathbb{R}^{|V|}$ is an optimal solution to $\mathcal{SDP}_3(G)$). Now, consider the edges $E_i^\mu \cap E^*$ that this graph shares with the random graph G^* defined in Lemma 1. Since E_i^μ only depends on the optimal solution to $\mathcal{SDP}_3(G)$ and edges in $E^* \setminus E$ do not influence this solution, $E_i^\mu \cap E^*$ spans a random subgraph of G^*. Therefore, this edge set contains $p|E_i^\mu|$ edges in expectation. Moreover,

$$\begin{aligned}
&\mathbf{P}\Big[\,|E_i^\mu \cap E^*| \leq \mathbf{E}[\,|E_i^\mu \cap E^*|\,] - 3n\sqrt{d}\,\Big] \\
&= \mathbf{P}\Big[\,|E_i^\mu \cap E^*| \leq p \cdot |E_i^\mu| - 3n\sqrt{d}\,\Big] \leq \exp\left(-\frac{3^2 n^2 d}{2p \cdot |E_i^\mu|}\right) \\
&\leq \exp\left(-\frac{9n}{2}\right) \leq \exp(-5n/2)
\end{aligned}$$

by the Chernoff bound (7).

Observe that, $(\mathbf{x}_v)_{v\in V(G)}$ also is a feasible (but not necessarily optimal) solution to $\mathcal{SDP}_3(G^*)$. Consequently, we can apply Lemma 1 for concluding that

$$\begin{aligned}
&\frac{2}{3}\mu \cdot \left(p|E_i^\mu| - 3n\sqrt{d}\right) \leq \frac{2}{3}\mu \cdot |E_i^\mu \cap E^*| \\
&\leq \frac{2}{3} \sum_{vw \in E_i^\mu \cap E^*} (1 - \langle \mathbf{x}_v \,|\, \mathbf{x}_w\rangle) \leq \frac{2}{3} \sum_{vw\in E^*\setminus E} (1 - \langle \mathbf{x}_v \,|\, \mathbf{x}_w\rangle) \leq c' n\sqrt{d}
\end{aligned}$$

holds with probability at least $1 - \exp(-5n/2) - \exp(-5n/2)$ where the second inequality follows from the definition of G_i^μ. By rearranging terms and dividing by n we arrive at a conclusion about the average degree $|E_i^\mu|/n$ of G_i^μ:

$$\frac{|E_i^\mu|}{n} \leq \left(\frac{3c'}{2\mu} + 3\right) \frac{n}{\sqrt{d}} =: \frac{c''}{\sqrt{d}} \cdot \frac{n}{3}.$$

This implies that there is some vertex v_i in C_i with

$$\frac{c''}{\sqrt{d}} \cdot \frac{n}{3} \geq deg_{G_i^{\mu}}(v_i) = \frac{n}{3} - |\mathbf{N}^{\mu}(v_i) \cap C_i|,$$

and thus we establish the first part of Lemma 2 by choosing μ in such a way that $\epsilon \geq c''/\sqrt{d}$ It remains to show that at most $\epsilon n/3$ vertices of other color classes are contained in $\mathbf{N}^{\mu'}(v_i)$ given that $|\mathbf{N}^{\mu}(v_i) \cap C_i| \geq (1-\epsilon)n/3 \geq n/6$. Assume, for contradiction, that for some color class C_j with $i \neq j$ we have $|\mathbf{N}^{\mu}(v_i) \cap C_j| \geq \epsilon/2 \cdot n/3$. Since each edge vw of G contributes exactly one to the optimal value of $\mathcal{SDP}_3(G)$ we know that $\langle \mathbf{x}_v \,|\, \mathbf{x}_w \rangle = -1/2$ and so $\mathbf{x}_v$ and $\mathbf{x}_w$ enclose an angle of 120°. Therefore, the set $\mathbf{N}^{\mu'}(v) \supset \mathbf{N}^{\mu}(v)$ induces an empty graph in G (because $\mu \leq \mu' < 1/2$ and $\arccos(1/2) = 60°$).

Accordingly, we can bound the desired probability by calculating the probability that some vertex sets $Y_i \subset C_i$ and $Y_j \subset C_j$ exist in G with $|Y_i| \geq n/6$, $|Y_j| \geq \epsilon n/6$, and $\mathbf{e}(Y_i, Y_j) = 0$:

$$\begin{aligned}
&\mathbf{P}\Big[\exists Y_i \subset C_i, Y_j \subset C_j : |Y_i| \geq \frac{n}{6}, |Y_j| \geq \epsilon\frac{n}{6}, \mathbf{e}(Y_i, Y_j) = 0\Big] \\
&\leq \binom{\frac{n}{3}}{\frac{n}{6}}\binom{\frac{n}{3}}{\epsilon\frac{n}{6}}\left(1 - \frac{d}{n}\right)^{\epsilon\frac{n^2}{36}} \\
&\leq \exp\Big(\Big(\mathrm{H}\Big(\frac{1}{2}\Big) + \mathrm{H}\Big(\frac{\epsilon}{2}\Big)\Big)\frac{n}{3}\Big) \cdot \exp\Big(-d \cdot \epsilon\frac{n}{36}\Big) \\
&= \exp(-\Omega(dn))
\end{aligned}$$

where the second inequality follows from $(1 - d/n)^n \leq \exp(-d)$ for $d < n$.

A.4 Proof of Lemma 3

This is a consequence of the following lemma. We extend the definition of the *μ-neighborhood* to arbitrary vectors $\mathbf{x} \in \mathbb{R}^n$ in the obvious way:

$$\mathbf{N}^{\mu}(\mathbf{x}) := \{v \in V \mid \langle \mathbf{x} \,|\, \mathbf{x}_v \rangle > 1 - \mu\}.$$

Lemma 11. *Let $\mathbf{x}$, $\mathbf{x}'$, $\mathbf{x}''$, $\mathbf{y}$, and $\mathbf{y}'$ be n-dimensional unit vectors such that $\mathbf{x}', \mathbf{x}'' \in \mathbf{N}^{\mu_1}(\mathbf{x})$ and $\mathbf{y}' \in \mathbf{N}^{\mu_2}(\mathbf{y})$ with $0 < \mu_1, \mu_2 < 1$. If moreover $\mathbf{x}' =: \mathbf{x} + \mathbf{x}^{\mu_1}$ and $\mathbf{y}' =: \mathbf{y} + \mathbf{y}^{\mu_2}$, then*

1. $\|\mathbf{x}^{\mu_1}\| \leq \sqrt{2\mu_1}$,
2. $\mathbf{x}'' \in \mathbf{N}^{4\mu_1}(\mathbf{x}')$, *and*
3. $|\langle \mathbf{x}' \,|\, \mathbf{y}' \rangle - \langle \mathbf{x} \,|\, \mathbf{y} \rangle| \leq \sqrt{2\mu_1} + \sqrt{2\mu_2} + 2\sqrt{\mu_1\mu_2}$.

Proof. By evaluating the norm of $\mathbf{x}'$ we find

$$1 = \|\mathbf{x}'\|^2 = \|\mathbf{x} + \mathbf{x}^{\mu_1}\|^2 = \langle \mathbf{x} + \mathbf{x}^{\mu_1} \,|\, \mathbf{x} + \mathbf{x}^{\mu_1} \rangle = 1 + \langle \mathbf{x}^{\mu_1} \,|\, \mathbf{x}^{\mu_1} \rangle + 2\langle \mathbf{x} \,|\, \mathbf{x}^{\mu_1} \rangle$$

and so

$$\begin{aligned}\langle \mathbf{x}^{\mu_1} \,|\, \mathbf{x}^{\mu_1} \rangle &= -2 \langle \mathbf{x} \,|\, \mathbf{x}^{\mu_1} \rangle = -2 \langle \mathbf{x} \,|\, \mathbf{x}' - \mathbf{x} \rangle \\ &= -2(\langle \mathbf{x} \,|\, \mathbf{x}' \rangle - 1) < -2(1 - \mu_1 - 1) = 2\mu_1.\end{aligned}$$

Similarly, $\langle \mathbf{y}^{\mu_2} \,|\, \mathbf{y}^{\mu_2} \rangle < 2\mu_2$. In addition, the angle enclosed by $\mathbf{x}'$ and $\mathbf{x}''$ is certainly less than $2 \cdot \arccos(1 - \mu_1)$ since these vectors are both in $\mathbf{N}^{\mu_1}(\mathbf{x})$. It follows that

$$\begin{aligned}\langle \mathbf{x}' \,|\, \mathbf{x}'' \rangle &\geq \cos(2 \cdot \arccos(1 - \mu_1)) = 2 \cdot \cos^2(\arccos(1 - \mu_1)) - 1 \\ &= 1 - 4\mu_1 + 2\mu_1^2 \geq 1 - 4\mu_1\end{aligned}$$

implying $\mathbf{x}'' \in \mathbf{N}^{4\mu_1}(\mathbf{x}')$. The third relation is a consequence of the Cauchy-Schwarz inequality which asserts $|\, \langle \mathbf{x}^{\mu_i} \,|\, \mathbf{y} \rangle \,| \leq \|\mathbf{x}^{\mu_i}\| \cdot \|\mathbf{y}\| \leq \sqrt{2\mu_i}$, for $i \in \{1, 2\}$ and $|\, \langle \mathbf{x}^{\mu_1} \,|\, \mathbf{y}^{\mu_2} \rangle \,| \leq \|\mathbf{x}^{\mu_1}\| \cdot \|\mathbf{y}^{\mu_2}\| \leq 2\sqrt{\mu_1 \mu_2}$. Thus,

$$\begin{aligned}|\, \langle \mathbf{x}' \,|\, \mathbf{y}' \rangle - \langle \mathbf{x} \,|\, \mathbf{y} \rangle \,| &= |\, \langle \mathbf{x} + \mathbf{x}^{\mu_1} \,|\, \mathbf{y} + \mathbf{y}^{\mu_2} \rangle - \langle \mathbf{x} \,|\, \mathbf{y} \rangle \,| \\ &= |\, \langle \mathbf{x}^{\mu_1} \,|\, \mathbf{y} \rangle + \langle \mathbf{x} \,|\, \mathbf{y}^{\mu_2} \rangle + \langle \mathbf{x}^{\mu_1} \,|\, \mathbf{y}^{\mu_2} \rangle \,| \\ &\leq \sqrt{2\mu_1} + \sqrt{2\mu_2} + 2\sqrt{\mu_1 \mu_2}.\end{aligned}$$

For $\mathbf{y}' = \mathbf{y}$ and $\mathbf{y} \notin \mathbf{N}^{\mu_1'}(\mathbf{x})$ for some μ_1', the third statement of Lemma 11 implies

$$\langle \mathbf{x}' \,|\, \mathbf{y} \rangle \leq \langle \mathbf{x} \,|\, \mathbf{y} \rangle + \sqrt{2\mu_1} \leq 1 - \mu_1' + \sqrt{2\mu_1}$$

and so $\mathbf{x}' \notin \mathbf{N}^{\mu_1' - \sqrt{2\mu_1}}(\mathbf{x})$. In conjunction with the second statement of Lemma 11, this immediately gives the following result: If $\mathbf{x}$ is an (ϵ', μ, μ')-representative with $\mu' > 4\mu + \sqrt{2\mu}$ and $\hat{\mathbf{x}} \in \mathbf{N}^{\mu}(\mathbf{x})$ then $\hat{\mathbf{x}}$ is an $(\epsilon', 4\mu, \mu' - \sqrt{2\mu})$-representative. This contains Lemma 3 as a special case.

A.5 Proof of Lemma 5

Call steps 1 and 2 of the construction process of H the *initial steps* and step 3 the *iterative deletion process*. For step l of this construction, i.e., the step when $\overline{H}^l$ is deleted, let $\overline{H}^{it} := \bigcup_{0<l'\leq l} \overline{H}^{l'}$. In addition, $\overline{H}^{it}$ may also be used without reference to a particular step l. It then denotes $\bigcup_{0<l'} \overline{H}^{l'}$.

For establishing Equation (4) we distinguish two cases. Either the vertices $\overline{H}^0$ deleted from H initially already contain the majority of vertices from $\overline{H}$, i.e., $|\overline{H}^0| \geq \alpha n/2$. Or $\overline{H}^{it}$ grows beyond the size of $\overline{H}^0$. Following this strategy, we obtain

$$\begin{aligned}\mathbf{P}\Big[\, |\overline{H}| \geq \alpha n \,\Big] &\leq \mathbf{P}\Big[\, |\overline{H}| \geq \alpha n \;\Big|\; |\overline{H}^0| \geq \frac{\alpha}{2} n \Big] \cdot \mathbf{P}\Big[\, |\overline{H}^0| \geq \frac{\alpha}{2} n \Big] \\ &\quad + \mathbf{P}\Big[\, |\overline{H}| \geq \alpha n \;\Big|\; |\overline{H}^0| < \frac{\alpha}{2} n \Big] \cdot \mathbf{P}\Big[\, |\overline{H}^0| < \frac{\alpha}{2} n \Big] \\ &\leq \mathbf{P}\Big[\, |\overline{H}^0| \geq \frac{\alpha}{2} n \Big] + \mathbf{P}\Big[\, |\overline{H}| \geq \alpha n \;\Big|\; |\overline{H}^0| < \frac{\alpha}{2} n \Big],\end{aligned}$$

and so the lemma follows from

$$\mathbf{P}\Big[\,|\overline{H}^0| \geq \frac{\alpha}{2}n\,\Big] \leq \exp(-\,(\log\alpha + \Omega\,(d))\cdot\alpha n)$$

and

$$\mathbf{P}\Big[\,|\overline{H}^{it}| \geq \frac{\alpha}{2}n \;\Big|\; |\overline{H}^0| < \frac{\alpha}{2}n\,\Big] \leq \exp(\Omega\,(d\cdot\log\alpha\cdot\alpha n))\,.$$

These relations will be established subsequently in Lemma 12 and Lemma 13, respectively.

Let $c_{(0)}$ be a constant such that the expression $\log\alpha + \Omega\,(d)$ in Equation (4) can be written as $\log\alpha + c_{(0)}d$. As we will see $c_{(0)}$ depends on the constant δ in the definition of H. If δ is fixed for the construction of H then we will choose $c_{(0)}$ in such a way that Lemma 5 remains valid if we reduce δ slightly to δ'. This will be needed in the proof of Lemma 8, where we employ a modification of the construction process for H. The use of δ' in the definition of $c_{(0)}$ is justified by the fact that we are interested in bounding the size of H from below in this section. This quantity increases with larger δ and so using δ' instead of δ gives a stronger condition for the choice of $c_{(0)}$.

Note further that $\log\alpha + c_{(0)}d$ gets negative for $\alpha < 2^{-c_{(0)}d}$ and consequently the bound (4) on $\mathbf{P}[\,|\overline{H}^0| \geq \alpha n/2\,]$ gets trivial in this case. As we will see in the proof of Lemma 17, however, this bound is small enough for our purposes if

$$\alpha > \alpha_0 := 2^{-c_{(0)}d/10},$$

i.e., the recovery procedure can extend a coloring of H to the whole graph G in polynomial expected running time if $\overline{H} > \alpha_0 n$.

Lemma 12. *For $0 < \alpha < 1/2$ and constant $0 < \delta < 1/2$,*

$$\mathbf{P}\Big[\,|\overline{H}^0| \geq \frac{\alpha}{2}n\,\Big] \leq \exp(-\,(\log\alpha + \Omega\,(d))\cdot\alpha n)\,.$$

Proof. We start by considering the vertices in $\overline{H}^+$ only and give an estimate on $\mathbf{P}[\,|\overline{H}^+| \geq \alpha n/4\,]$. The corresponding probability for $\overline{H}^-$ follows similarly.

Idea: We first calculate the probability $\overline{p}$ that a vertex in C_i has degree greater than $(1+\delta)d/3$ in some partition C_j with $i \neq j$. Then the probability distribution of $\overline{H}_i^+ := |\overline{H}^+ \cap C_i|$ is given by the binomial distribution $\mathrm{Bin}(\overline{p}, n/3)$.

Calculating $\overline{p}$: By symmetry we may consider a vertex $v \in C_1$ without loss of generality. For computing the desired probability we make use of $\mathbf{E}[\,deg_{C_2}(v)\,] = \mathbf{E}[\,deg_{C_3}(v)\,] = d/3$, the fact that δ is constant and d is sufficiently large, and of the Chernoff bound given in (6):

$$\begin{aligned}\overline{p} &\leq \mathbf{P}\left[\,deg_{C_2}(v) \geq (1+\delta)\frac{d}{3}\,\right] + \mathbf{P}\left[\,deg_{C_3}(v) \geq (1+\delta)\frac{d}{3}\,\right]\\ &\leq 2\exp\left(-\frac{\delta^2 d^2}{6(d+\frac{d\delta}{3})}\right) \leq 2\exp\left(-\frac{\delta^2 d}{7}\right) \leq \exp\left(-\frac{\delta^2 d}{8}\right),\end{aligned}$$

for $d = pn$ sufficiently large.

Calculating $\mathbf{P}[\,|\overline{H}^{+}| \geq \alpha n/4\,]$: Without loss of generality, let C_1 be the partition that maximizes $\overline{H}_i^{+}$, i.e., if $|\overline{H}^{+}| \geq \alpha n/4$ then $|\overline{H}_1^{+}| \geq \alpha n/12$. Thus

$$\begin{aligned}
\mathbf{P}\Big[\,|\overline{H}^{+}| \geq \frac{\alpha}{4}n\Big] &\leq 3 \cdot \mathbf{P}\Big[\,\overline{H}_1^{+} \geq \frac{\alpha n}{12}\,\Big] \\
&\leq 3\binom{\frac{n}{3}}{\frac{\alpha}{4}\cdot\frac{n}{3}}\overline{p}^{\,\alpha n/12} \leq 3\exp\Big(-\frac{\alpha}{2}\log\frac{\alpha}{4}\cdot\frac{n}{3}\Big)\cdot\exp\Big(-\frac{\delta^2 d}{8}\cdot\frac{\alpha n}{12}\Big) \\
&\leq \frac{1}{2}\exp\Big(\Big(\ln 6 - \frac{1}{6}\log\frac{\alpha}{4} - \frac{\delta^2 d}{96}\Big)\alpha n\Big) \\
&\leq \frac{1}{2}\exp\Big(-\Big(\log\alpha + \frac{\delta^2 d}{97}\Big)\alpha n\Big) = \frac{1}{2}\exp(-\left(\log\alpha + \Omega\left(d\right)\right)\alpha n)
\end{aligned} \tag{8}$$

for constant δ, $\alpha < 0.5$ and d sufficiently large.

Since $\mathbf{P}[\,|\overline{H}^{0}| \geq \alpha n\,] \leq \mathbf{P}[\,|\overline{H}^{+}| \geq \alpha n\,] + \mathbf{P}[\,|\overline{H}^{-}| \geq \alpha n\,]$ it remains to argue that an analogue of (8) holds for $|\overline{H}^{-}|$ in order to establish Lemma 12. For this purpose, notice that the Chernoff bound (6) exceeds the corresponding lower tail bound (7). Thus, by the definition of $\overline{H}^{-}$ and $\overline{H}^{+}$, the probability that $\overline{H}^{-}$ reaches a certain size can be estimated using the same methods as above for $\overline{H}^{+}$ and so

$$\mathbf{P}\Big[\,|\overline{H}^{-}| \geq \alpha n\Big] \leq \frac{1}{2}\exp(-\left(\log\alpha + \Omega\left(d\right)\right)\alpha n)\,.$$

The iterative deletion process removes vertices having too many neighbors in the set of formerly deleted vertices. The aim of the following calculations is to estimate the probability that the number of these vertices gets considerably bigger than the set of vertices deleted in the initial steps. For this purpose we condition on the event $|\overline{H}^{0}| \leq 3 \cdot |\overline{H}^{it}|$. The constant 3 in this bound is not required for analyzing the performance of the recovery procedure on H. We introduce it for obtaining calculations robust enough to allow certain changes to the construction of H. This will be needed for the proof of Lemma 8.

Lemma 13. *Assume that $0 < \alpha < 1/2$ and that $0 < \delta < 1/2$ is constant in the definition of H. Then*

$$\begin{aligned}
\mathbf{P}\Big[\,|\overline{H}^{it}| \geq \frac{\alpha}{2}n \;\Big|\; |\overline{H}^{0}| \leq \frac{\alpha}{2}n\Big] &\leq \mathbf{P}\Big[\,|\overline{H}^{it}| \geq \frac{\alpha}{2}n \;\Big|\; |\overline{H}^{0}| \leq \frac{3\alpha}{2}n\Big] \\
&\leq \exp(\Omega\left(d\cdot\log\alpha\cdot\alpha n\right))\,.
\end{aligned}$$

Proof. The first inequality holds since edges are chosen independently in G and by the construction of H. For the second inequality, consider the first step in the deletion process, say step k, when more than $\alpha n/2$ vertices have been deleted in addition to $\overline{H}^{0}$ and denote the vertices deleted prior to step k by $\overline{H}^{<k}$. Then

$$\overline{H}^{<k} \leq 2\alpha n$$

since $|\overline{H}^0| \leq 3\alpha n/2$ by assumption. Observe that each vertex in $\overline{H}^{<k} \setminus \overline{H}^0$ was deleted because it had more than $\delta d/3$ neighbours in $\overline{H}^{<k}$ (since $\overline{H}^{<k} \setminus \overline{H}^0 \subset \overline{H}^{it}$). Accordingly the graph induced on $\overline{H}^{<k}$ contains more than $\delta d/3 \cdot \alpha n/2$ edges. We bound the desired probability from above by calculating the probability that there exists any set $Y \subset V$ of size at most $2\alpha n$ in G with $\mathbf{e}(Y,Y) \geq \delta d/3 \cdot \alpha n/2$. We show that with high probability a subgraph of this density does not even exist in $\mathcal{G}_{n,p}$ and therefore the same result follows for $\mathcal{G}_{n,p,3}$:

$$\begin{aligned}
&\mathbf{P}\left[|\overline{H}^{it}| \geq \frac{\alpha}{2}n \;\middle|\; |\overline{H}^0| \leq \frac{3\alpha}{2}n \right] \leq \mathbf{P}\left[\exists Y \subset V : |Y| \leq 2\alpha n, \mathbf{e}(Y,Y) \geq \delta\frac{d}{3}\cdot\frac{\alpha}{2}n \right] \\
&\leq \mathbf{P}\left[\exists Y \subset V : |Y| = 2\alpha n, \mathbf{e}(Y,Y) \geq \delta\frac{d}{3}\cdot\frac{\alpha}{2}n \right] \leq \binom{n}{2\alpha n}\binom{\binom{2\alpha n}{2}}{\delta\frac{d}{3}\cdot\frac{\alpha}{2}n} p^{\delta\frac{d}{3}\cdot\frac{\alpha}{2}n} \\
&\leq \left(\frac{e}{2\alpha}\right)^{2\alpha n} \left(\frac{3e\cdot 2\alpha n(2\alpha n-1)}{\delta d\cdot \alpha n}\right)^{\delta\frac{d}{3}\cdot\frac{\alpha}{2}n} \left(\frac{d}{n}\right)^{\delta\frac{d}{3}\cdot\frac{\alpha}{2}n} \\
&\leq \left(\left(\frac{e}{2\alpha}\right)^{12}\left(\frac{12e\cdot\alpha}{\delta}\right)^{\delta d}\right)^{\frac{\alpha}{6}n} \\
&= \left(\left(\frac{e}{2\alpha}\right)^{12}\left(\frac{12e\cdot\alpha}{\delta}\right)^{12}\left(\frac{12e\cdot\alpha}{\delta}\right)^{\delta d-12}\right)^{\frac{\alpha}{6}n} = \exp(\Omega\left(d\cdot\log\alpha\cdot\alpha n\right)).
\end{aligned}$$

A.6 Proof of Lemma 6

Let $F^s \subseteq H$ be the set of vertices in H colored incorrectly after iteration s of the iterative recoloring step. Consider iteration t of the recoloring step and assume that all previous iterations were successful, but iteration t fails, i.e., $F^{t-1} = F_H$ and $|F^t| \geq |F^{t-1}|/2$. More precisely we just consider an arbitrary subset of F^t of size $|F^{t-1}|/2$. For simplicity we refer to this subset by F^t. Let $i \neq j$ and $v \in F^t$ be a vertex in color class C_i that received color j in iteration t. Observe that, as v remains colored incorrectly after iteration t, v can have at most $deg_G(v)/3$ neighbors that were colored j in the previous iteration. But by construction, H contains only vertices which have at most $(1+\delta)d/3$ neighbors in each color class other than their own and so $deg_G(v) \leq 2\cdot(1+\delta)d/3$ and $deg_{C_j\cap H}(v) \geq (1-\delta)d/3$. This implies that v has at most $(2\cdot(1+\delta)d/9)$ neighbors in $C_j \cap H - F^{t-1}$ and so all other neighbors of v in $C_j \cap H$, namely, at least

$$(1-\delta)\frac{d}{3} - 2\cdot(1+\delta)\frac{d}{9} = (1-5\delta)\,\frac{d}{9}$$

are contained in F^{t-1}. We conclude that the graph induced on $F^{t-1} \cup F^t$ contains at least $(1-5\delta)\,d/9 \cdot |F^t|$ edges. We can therefore bound $\mathbf{P}[\,|F_H| = \alpha n\,]$ by calculating the probability that some vertex set Y exists in V with $|Y| = 3\alpha n/2$

such that $\mathbf{e}(Y,Y) \geq (1-5\delta)d/9 \cdot \alpha n/2$.

$$
\begin{aligned}
&\mathbf{P}\left[\exists Y \subset V : |Y| = \frac{3\alpha}{2}n, \mathbf{e}(Y,Y) \geq (1-5\delta)\,\frac{d}{9}\cdot\frac{\alpha}{2}n\right] \\
&\leq \binom{n}{\frac{3\alpha}{2}n}\binom{\binom{\alpha n}{2}}{(1-5\delta)\frac{d}{9}\cdot\frac{\alpha}{2}n}\cdot p^{(1-5\delta)\frac{d}{9}\cdot\frac{\alpha}{2}n} \\
&\leq \left(\frac{2e}{3\alpha}\right)^{\alpha n}\left(\frac{9e\cdot\alpha n(\alpha n-1)}{(1-5\delta)d\cdot\alpha n}\right)^{(1-5\delta)\frac{d}{9}\cdot\frac{\alpha}{2}n}\left(\frac{d}{n}\right)^{(1-5\delta)\frac{d}{9}\cdot\frac{\alpha}{2}n} \\
&\leq \left(\left(\frac{2e}{3\alpha}\right)^{18}\left(\frac{9e\cdot\alpha}{1-5\delta}\right)^{(1-5\delta)d}\right)^{\frac{\alpha}{18}n} \\
&\leq \left(\left(\frac{2e}{3\alpha}\right)^{18}\left(\frac{9e\cdot\alpha}{1-5\delta}\right)^{18}\left(\frac{9e\cdot\alpha}{1-5\delta}\right)^{(1-5\delta)d-18}\right)^{\frac{\alpha}{18}n} = \exp(\Omega\left(d\cdot\log\alpha\cdot\alpha n\right)),
\end{aligned}
$$

for δ sufficiently small and d sufficiently large. As argued above, this implies

$$\mathbf{P}[\,|F_H| = \alpha n\,] \leq \exp(\Omega\left(d\cdot\log\alpha\cdot\alpha n\right))$$

as desired.

A.7 Details of the Proof of Lemma 7

We determine the probability that there is some set $Y \subset V(G)$ with $|Y| = \alpha n$ and minimum degree at least $d/6$ as follows:

$$
\begin{aligned}
&\mathbf{P}[\,|F_Y| \geq \alpha n\,] \leq \mathbf{P}[\,\exists Y \subset V(G) : |Y| = \alpha n, mindeg(Y) \geq d/6\,] \\
&\leq \mathbf{P}\left[\exists Y \subset V(G) : |Y| = \alpha n, \mathbf{e}(Y,Y) \geq \frac{d}{12}\alpha n\right] \\
&\leq \binom{n}{\alpha n}\binom{\binom{\alpha n}{2}}{\frac{d}{12}\alpha n}p^{d\cdot\alpha n/12} \leq \left(\frac{e}{\alpha}\right)^{\alpha n}\left(\frac{6e(\alpha n-1)}{d}\right)^{d\cdot\alpha n/12}\left(\frac{d}{n}\right)^{d\cdot\alpha n/12} \\
&\leq \left(\frac{e}{\alpha}\left(6e\alpha\right)^{d/12}\right)^{\alpha n} \leq \left(6e^2\left(6e\alpha\right)^{d/12-1}\right)^{\alpha n} = \exp(\Omega\left(d\cdot\log\alpha\cdot\alpha n\right)).
\end{aligned}
$$

A.8 Proof of Lemma 8

Some preparation is needed before we can turn to the proof of Lemma 8. Let αn be the size of the largest component induced on the uncolored vertices. Clearly

$$
\begin{aligned}
&\mathbf{P}[\,\text{there is a component of order } \alpha n \text{ in } G-H\ \,] \\
&\leq \mathbf{P}[\,\exists \text{ a tree } T \subseteq G-H \text{ of order } \alpha n\,] \\
&= \mathbf{P}[\,\exists \text{ a tree } T \text{ with } |T| = \alpha n \,:\, E(T) \subset E(G), T\cap H = \emptyset\,]\,.
\end{aligned}
\tag{9}
$$

Unfortunately the two events occuring on the last line of Equation (9) are not independent of each other since H is not a random subgraph of G. For this reason, we modify the construction of H depending on the tree T under study, resulting in a new subgraph H' of G. For this subgraph the corresponding events can be seperated, giving an upper bound on the probability we are interested in.

So, consider a fixed tree T with $V(T) \subset V(G)$ and all edges between different color classes of G, but not necessarily satisfying $E(T) \subset E(G)$. Denote by $T_{<4}$ all vertices v of T obeying $deg_T(v) < 4$, set $T_{\geq 4} := T - T_{<4}$, and let H' be the graph obtained from G by the following process:

1. Construct a graph $G' = (V, E')$ from G by discarding all edges in $E(T)$ and then reconsidering their occurence by throwing a new die, i.e., $E' = (E \setminus E(T)) \cup E'(T)$ where $E'(T) \subset E(T)$ contains each edge of T with probability p.
2. Delete all vertices in $T_{\geq 4}$ from G'.
3. Apply the procedure for constructing H to G' with δ slightly reduced to δ' (see page 8).

Again, we refer to Steps 1 and 2 of the construction procedure for H within this process as the two *initial steps*. Step 3 of the procedure for H is called the *iterative deletion process*, and the corresponding sets are $\overline{H}^{0'}$ and $\overline{H}^{it'}$.

Note that the graph G' constructed in Step 1 above still is a 3-colorable graph with partitions C_1, C_2, and C_3 since all edges of T run between different partitions of G by definition. When refering to the degree of a vertex $v \in V$ into one of these partitions C_i in G' we write $deg_{C_i'}(v)$.

The next lemma shows that this revised construction does indeed lead to independent events. A similar argument was used in [1].

Lemma 14. *For any fixed tree T,*

$$\mathbf{P}[\, E(T) \subset E(G) \; and \, T \cap H = \emptyset \,] \leq \mathbf{P}[\, E(T) \subset E(G) \,] \cdot \mathbf{P}[\, T \cap H' = \emptyset \,] \,.$$

Proof. By

$$\begin{aligned} &\mathbf{P}[\, E(T) \subset E(G) \, \text{and} \, T \cap H' = \emptyset \,] \\ &\qquad\qquad = \mathbf{P}[\, E(T) \subset E(G) \,] \cdot \mathbf{P}[\, T \cap H' = \emptyset \mid E(T) \subset E(G) \,] \end{aligned}$$

this lemma is a direct consequence of $T \cap H' \subset T \cap H$ since the events $T \cap H' = \emptyset$ and $E(T) \subset E(G)$ are clearly independent due to Step 1 in the construction of H'. Accordingly, it remains to show that $v \notin H \cap T$ implies $v \notin H' \cap T$. We assert this by comparing the different deletion steps in the construction of H versus H'.

First, consider vertices $v \in \overline{H}^0$ deleted in one of the first two steps while constructing H (i.e., vertices v with $deg_{C_i}(v) \lesseqgtr (1 \mp \delta)d/3$ for some i). If $v \notin T$ then v is clearly deleted in the initial steps of the construction of H'. All vertices $v \in T$ with $v \notin T_{<4}$ are deleted from H' as well. For $v \in T_{<4}$ finally we have either $deg_{C_i'}(v) \geq (1+\delta)d - 3 \geq (1+\delta')d$ or $deg_{C_i'}(v) \leq (1-\delta)d + 3 \leq (1-\delta')d$ by the definition of $T_{<4}$ and so $v \in \overline{H}^{0'}$ in this case too.

Now, let us turn to vertices $v \in \overline{H}^{it}$. By the preceeding arguments we know that $\overline{H}^0 \subset \overline{H}^{0'}$. We proceed by induction on the steps in the iterative deletion process. In each of these steps in the construction of H, those vertices v are deleted from H that have more than $2\delta d/3$ formerly deleted neighbors. Again, in the case $v \notin T$ such a deletion carries over to a deletion from H' since $\overline{H}^0 \subset \overline{H}^{0'}$ and by applying the induction hypothesis. The case $v \in T_{\geq 4}$ may be omitted since these vertices were deleted from H' already. Thus, it remains to consider $v \in T_{<4}$ (observe that we can not use the same argument here as in the case $v \notin T$ since the edges on vertices from T are different in H and H'). By induction hypothesis and the definition of $T_{<4}$ we know that v has at least $2\delta d/3 - 3 \leq 2\delta' d/3$ neighbors that were deleted from H' in earlier steps. It follows that v is deleted from H as well.

This concludes the proof of $H' \cap T \subset H \cap T$ and shows more generally that $H' \subset H$. Therefore, the validity of Lemma 14 is verified.

With this we are ready to prove Lemma 8. There, we are interested in trees T with $|T| = \alpha n$. In this case $|T_{<4}| \geq \alpha n/2$ because $|T_{<4}| \geq |T|/2$ follows from

$$|T| - 1 = |E(T)| = \frac{1}{2}\left(\sum_{v \in T_{<4}} deg_T(v) + \sum_{v \in T_{\geq 4}} deg_T(v)\right) \geq \frac{1}{2} \cdot 4|T_{\geq 4}|.$$

The following discussion will partly refer to the proof of Lemma 5.

Proof (of Lemma 8).
By Equation (9) and Lemma 14, we can rewrite the probability we are interested in as follows:

$$\begin{aligned}&\mathbf{P}[\,\exists T \text{ with } |T| = \alpha n \,:\, E(T) \subset E(G), T \cap H = \emptyset\,]\\ &\leq \sum_T \mathbf{P}[\,E(T) \subset E(G)\,] \cdot \mathbf{P}[\,T \cap H' = \emptyset\,].\end{aligned}$$

Since the edges of G (and G') are chosen independently from each other, the probability that $E(T) \subset E(G)$ for a fixed tree T is simply $p^{\alpha n}$. So in the main part of this proof we will investigate the probability $\mathbf{P}[\,T \cap H' = \emptyset\,]$ and show that

$$\mathbf{P}[\,T \cap H' = \emptyset\,] < \exp(-\alpha \cdot \Omega\,(dn)).$$

This suffices for establishing the desired bound: It is well known that the number of labeled trees on y vertices is y^{y-2} and consequently

$$\begin{aligned}&\sum_T \mathbf{P}[\,E(T) \subset E(G)\,] \cdot \mathbf{P}[\,T \cap H' = \emptyset\,]\\ &\leq \binom{n}{\alpha n} (\alpha n)^{\alpha n - 2} \cdot p^{\alpha n - 1} \cdot \exp(-\alpha \cdot \Omega\,(dn))\\ &\leq \left(\frac{e}{\alpha} \cdot \alpha n \cdot \frac{d}{n} \cdot \exp(-\Omega\,(d))\right)^{\alpha n} \leq \left(\frac{d}{\exp(\Omega\,(d))}\right)^{\alpha n}.\end{aligned}$$

Calculating $\mathbf{P}[\,T \cap H' = \emptyset\,]$: H' is obviously not a random subgraph of G. However, the event $T \cap H' = \emptyset$ does not depend on the structure of H', but only on the vertices contained in this subgraph. Moreover, the construction of H' is independent of the labeling of the vertices involved. Now, consider a fixed graph F. Then, by the foregoing remarks, among all graphs from $\mathcal{G}_{n,p,3}$ the event that F is induced on a particular set of vertices and forms the graph H' has the same probability as the event that this happens for any other set of vertices. Consequently, we get an invariance under permutation of the labeling of G (while leaving the labeling of T fixed) and so the only information about H' that is necessary for determining the desired probability is its size. For this reason, we will next try to estimate $|H'|$.

In Lemma 5, the probability that $|G - H|$ exceeds αn was bounded by

$$\exp(-\left(\log \alpha + \Omega\left(d\right)\right) \cdot \alpha n) + \exp(\Omega\left(d \cdot \log \alpha \cdot \alpha n\right)).$$

In this connection we also remarked that the given estimation does not allow for a nontrivial upper bound when $\alpha \ll \alpha_0 = 2^{-c_{(0)} d/10}$, where $c_{(0)}$ is the constant defined on page 17. As mentioned, we will show that we can use the same bound for a corresponding result on $|H'|$ which will, again, only be useful in the case that $|G - H'|$ is at least of size $\alpha_0 n$. This gives a motivation for rewriting $\mathbf{P}[\,T \cap H' = \emptyset\,]$ by conditioning on the event $|H'| \geq (1 - \alpha_0)\, n$:

$$\begin{aligned}
\mathbf{P}[\,T \cap H' = \emptyset\,] = {} & \mathbf{P}[\,T \cap H' = \emptyset \mid |H'| \geq (1-\alpha_0)\, n\,] \cdot \mathbf{P}[\,|H'| \geq (1-\alpha_0)\, n\,] \\
& + \mathbf{P}[\,T \cap H' = \emptyset \mid |H'| < (1-\alpha_0)\, n\,] \cdot \mathbf{P}[\,|H'| < (1-\alpha_0)\, n\,] \\
\leq {} & \mathbf{P}[\,T \cap H' = \emptyset \mid |H'| \geq (1-\alpha_0)\, n\,] + \mathbf{P}[\,|H'| < (1-\alpha_0)\, n\,].
\end{aligned}$$

Since, as discussed above, $\mathbf{P}[\,T \cap H'\,]$ only depends on the size of H' we can bound the first term in this equation in the following way:

$$\begin{aligned}
\mathbf{P}[\,T \cap H' = \emptyset \mid |H'| \geq (1-\alpha_0)\, n\,] &\leq \frac{\binom{\alpha_0 n}{|T|}}{\binom{n}{|T|}} \leq \left(\frac{e \cdot \alpha_0 n}{|T|} \cdot \frac{|T|}{n}\right)^{|T|} \\
&\leq \left(e \cdot 2^{-\frac{1}{10} c_{(0)} \cdot d}\right)^{\alpha n} = \exp(-\alpha \cdot \Omega\left(dn\right)).
\end{aligned}$$

For evaluating the second term, note that the initial steps in the construction of H and H' (i.e., removing vertices of high and low degree in G and G', respectively) are identical apart from the value of δ and δ', respectively. The computations concerning the size of the set $\overline{H}^0 \subset G - H$ of initially deleted vertices have been performed for δ' in the proof of Lemma 12. Since G and G' are both graphs from $\mathcal{G}_{n,p,3}$, we consequently can adopt the results derived there for concluding that $|\overline{H}^{0'}| < \alpha_0 n/4$ holds with probability at least

$$\begin{aligned}
& 1 - \exp\left(-\left(\log \frac{\alpha_0}{2} + c_{(0)} \cdot d\right) \cdot \frac{\alpha_0}{2} n\right) \\
&= 1 - \exp\left(-\left(-\frac{1}{10} c_{(0)} \cdot d - 1 + c_{(0)} \cdot d\right) 2^{-\frac{1}{10} c_{(0)} \cdot d - 1} n\right) \\
&= 1 - \exp\left(-\frac{d}{2^{\mathcal{O}(d)}}\, \Omega\left(n\right)\right).
\end{aligned}$$

$T_{\geq 4}$ is of order at most $\alpha n/2$ and therefore this set contains at most $\alpha_0 n/2$ vertices. So the number $|\overline{H}^{0'} \cup T_{\geq 4}|$ of vertices removed from H' before the iterative deletion process starts is certainly less than $3\alpha_0 n/4$ with the same probability.

Later, in Lemma 13 we calculated $\mathbf{P}\left[\,|\overline{H}^{it}| \geq y \;\middle|\; |\overline{H}^{0}| \leq 3y\right]$ for the number $|\overline{H}^{it}|$ of vertices removed in the iterative deletion process. Recall that the corresponding analysis solely relied on the probability that a set $\overline{H}^{<k} \setminus \overline{H}^{0}$ of vertices deleted in this procedure has got many neighbors in the set $\overline{H}^{<k}$ of all formerly deleted vertices. Here, this probability does not depend on the structure induced on the set $\overline{H}^{0}$ of vertices deleted before the iterative deletion process takes effect. We therefore conclude in accordance to Lemma 13 that given

$$|\overline{H}^{0'} \cup T_{\geq 4}| \leq 3\alpha_0 n/4,$$

the event $|\overline{H}^{it'}| < \alpha_0 n/4$ holds with probability at least

$$\begin{aligned}
&1 - \exp(\Omega\,(d \cdot \log \alpha_0 \cdot \alpha_0 n)) \\
&= 1 - \exp\left(\Omega\left(d \cdot \log 2^{-c_{(0)}d/10} \cdot 2^{-c_{(0)}d/10} n\right)\right) \\
&\geq 1 - \exp\left(\Omega\left(-d \cdot d \cdot 2^{-c_{(0)}d/10} \cdot n\right)\right) = 1 - \exp\left(-\frac{d}{2^{\mathcal{O}(d)}}\,\Omega\,(n)\right).
\end{aligned}$$

It follows that

$$\mathbf{P}[\,|H'| \geq (1-\alpha_0)\,n\,] \geq 1 - \exp\left(-\frac{d}{2^{\mathcal{O}(d)}}\,\Omega\,(n)\right)$$

and so, using the bound on $\mathbf{P}[\,T \cap H' = \emptyset \mid |H'| \geq (1 - 6\cdot\alpha_0)\,n\,]$ calculated earlier, we finally arrive at

$$\begin{aligned}
&\mathbf{P}[\,T \cap H' = \emptyset\,] \\
&\leq \mathbf{P}[\,T \cap H' = \emptyset \mid |H'| \geq (1 - 6\cdot\alpha_0)\,n\,] + \mathbf{P}[\,|H'| < (1 - 6\cdot\alpha_0)\,n\,] \\
&\leq \exp(-\alpha \cdot \Omega\,(dn)) + \exp\left(-\frac{d}{2^{\mathcal{O}(d)}}\,\Omega\,(n)\right) \leq \exp(-\alpha \cdot \Omega\,(dn))\,.
\end{aligned}$$

A.9 Details of the Proof of Lemma 10

Recall that F_H are those vertices in H which are colored incorrectly after the recoloring step of Algorithm 1. F_{SDP} is the set of vertices that need to be assigned a different color for obtaining a valid coloring on an $(1-\epsilon)$-fraction of G after the inital phase. The set of vertices colored incorrectly after the uncoloring step finally is denoted by F_{Υ} and

$$\alpha_0 = 2^{-c_{(0)}d/10}$$

for some constant $c_{(0)}$ (cf. page 17).

Consider the vertex set Y from Algorithm 1 that is colored correctly in Step 3 of the recovery procedure and let $t(y)$ be the time the algorithm needs to execute this step in the case $|Y| = y$. Further, denote by F the set Y used in the iteration when the algorithm finally obtains a valid coloring. The expected running time $\mathbf{E}[\,t\,]$ of the repair mechanism can then be written as

$$\begin{aligned}\mathbf{E}[\,t\,] &= \sum_{y\le n} \mathbf{P}[\,|F| = y\,]\cdot t(y) \\ &= \sum_{y\le n} \mathbf{P}[\,|F| = y\,]\cdot \sum_{y'\le y}\binom{n}{y'}\mathbf{3\text{-}COL}\,(y') \\ &\le \mathcal{O}(n)\sum_{y\le n} \mathbf{P}[\,|F| = y\,]\cdot \binom{n}{y}\mathbf{3\text{-}COL}\,(y)\,.\end{aligned} \tag{10}$$

We proceed by splitting the summands $\mathbf{P}[\,|F| = y\,]\cdot\binom{n}{y}\mathbf{3\text{-}COL}\,(y)$ into different components and show that each of them evaluates to a polynomial. To begin with, we bound $\mathbf{P}[\,|F| = y\,]$ by rewriting F as sum of F_{SDP}, F_H, F_Υ and possibly $\overline{H}$. For this, observe that

$$F = F_{SDP} \cup F_H \cup F_\Upsilon.$$

However, we use this partitioning of F only in the case that $|\overline{H}| \le \alpha_0 n$. If $|\overline{H}| > \alpha_0 n$ we use

$$F \subset F_{SDP} \cup F_H \cup \overline{H}.$$

Since all probabilities involved are monotone decreasing, we get

$$\begin{aligned}\mathbf{P}[\,|F| = y\,] \le &\sum_{\substack{y_1+y_2+y_3=y\\ y_3\le\alpha_0 n}} \mathbf{P}[\,|F_{SDP}| = y_1, |F_H| = y_2, |F_\Upsilon| = y_3\,] \\ &+ \sum_{\substack{y_1+y_2+y_3=y\\ y_3>\alpha_0 n}} \mathbf{P}\big[\,|F_{SDP}| = y_1, |F_H| = y_2, |\overline{H}| = y_3\,\big]\,.\end{aligned}$$

One of the vertex sets involved in the conjunctions of this sum certainly contains more than $y/3$ vertices and so

$$\begin{aligned}&\mathbf{P}[\,|F| = y\,] \\ &\le n^3 \begin{cases}\max\big(\mathbf{P}\big[\,|F_{SDP}| = \frac{y}{3}\,\big], \mathbf{P}\big[\,|F_H| = \frac{y}{3}\,\big], \mathbf{P}\big[\,|F_\Upsilon| = \frac{y}{3}\,\big]\big) & \text{if } \frac{y}{3} \le \alpha_0 n,\\ \max\big(\mathbf{P}\big[\,|F_{SDP}| = \frac{y}{3}\,\big], \mathbf{P}\big[\,|F_H| = \frac{y}{3}\,\big], \mathbf{P}\big[\,|\overline{H}| = \frac{y}{3}\,\big]\big) & \text{otherwise.}\end{cases}\end{aligned}$$

Since the coloring constructed in the initial step of COLOR (possibly with the help of the recovery procedure) fails on at most ϵn vertices, we can assume that the size of F_H does not exceed ϵn either. Therefore, we can refine the last equation in the following way:

$$\mathbf{P}[\,|F|=y\,]$$
$$\le n^3 \begin{cases} \max\big(\mathbf{P}\big[\,|F_{SDP}|=\frac{y}{3}\,\big],\mathbf{P}\big[\,|F_H|=\frac{y}{3}\,\big],\mathbf{P}\big[\,|F_\Upsilon|=\frac{y}{3}\,\big]\big) & \text{if } \frac{y}{3}\le\alpha_0 n,\\ \max\big(\mathbf{P}\big[\,|F_{SDP}|=\frac{y}{3}\,\big],\mathbf{P}\big[\,|\overline{H}|=\frac{y}{3}\,\big]\big) & \text{if } \frac{y}{3}\ge\epsilon n,\\ \max\big(\mathbf{P}\big[\,|F_{SDP}|=\frac{y}{3}\,\big],\mathbf{P}\big[\,|F_H|=\frac{y}{3}\,\big],\mathbf{P}\big[\,|\overline{H}|=\frac{y}{3}\,\big]\big) & \text{otherwise.}\end{cases} \tag{11}$$

Now, let $y=\alpha n$ and recall that

$$\mathbf{P}[\,|F_H|=\alpha n\,]\le(c_H\alpha)^{c'_H d\cdot\alpha n}=:f_H(\alpha),$$

$$\mathbf{P}[\,|F_\Upsilon|=\alpha n\,]\le(c_\Upsilon\alpha)^{c'_\Upsilon d\cdot\alpha n}=:f_\Upsilon(\alpha),$$

and

$$\mathbf{P}\big[\,|\overline{H}|=\alpha n\,\big]$$
$$\le\mathbf{P}\big[\,|\overline{H}|\ge\alpha n\,\big]\le\underbrace{\exp\big(-\alpha n\left(\log\alpha+c_{(0)}\cdot d\right)\big)}_{=:f_{(0)}(\alpha)}+\underbrace{(c_{it}\alpha)^{c'_{it}d\cdot\alpha n}}_{=:f_{it}(\alpha)}=:f_{\overline{H}}(\alpha)$$

by Lemmas 6, 7, and 5, respectively, for appropriate constants c_H, c'_H, c_Υ, c'_Υ, $c_{(0)}$, c_{it}, and c'_{it}. With this we can rewrite (11) as

$$\mathbf{P}[\,|F|=y\,]\le n^3 \begin{cases} \max\big(\mathbf{P}\big[\,|F_{SDP}|=\frac{y}{3}\,\big],\,f_H(\frac{y}{3n}),\,f_\Upsilon(\frac{y}{3n})\big) & \text{if } \frac{y}{3}\le\alpha_0 n,\\ \max\big(\mathbf{P}\big[\,|F_{SDP}|=\frac{y}{3}\,\big],\mathbf{P}\big[\,|\overline{H}|=\frac{y}{3}\,\big]\big) & \text{if } \frac{y}{3}\ge\epsilon n,\\ \max\big(\mathbf{P}\big[\,|F_{SDP}|=\frac{y}{3}\,\big],\,f_H(\frac{y}{3n}),\,\mathbf{P}\big[\,|\overline{H}|=\frac{y}{3}\,\big]\big) & \text{otherwise.}\end{cases} \tag{12}$$

We next exploit the fact that $f_\Upsilon(\alpha)$, $f_H(\alpha)$, and $f_{it}(\alpha)$ are of similar structure.

Lemma 15. *Let $c_{max}:=\max(c_H,c_\Upsilon,c_{it})$. Then there is some constant c'_{max} such that each of the functions $f_\Upsilon(\alpha)$, $f_H(\alpha)$, and $f_{it}(\alpha)$ is less than or equal to*

$$f_{max}(\alpha):=(c_{max}\alpha)^{c'_{max}d\cdot\alpha n}$$

in the whole interval $0<\alpha<1$.

Indeed, our estimates on the three probabilities in question are all of the form $(c\alpha)^{c'd\cdot\alpha n}$ where $c>1$ and $c'>0$. Note that for comparing two functions of this form, we can omit the exponent $d\cdot\alpha n$. So, consider two functions $f_1(y):=(c_1y)^{c'_1}$ and $f_2(y):=(c_2y)^{c'_2}$ with $c_1,c_2>1$, $c'_1,c'_2>0$, and $c_1\ge c_2$. Then $f_1\ge f_2$ for $y<1$ provided that $c'_1\le c'_2$ and $c_1{}^{c'_1}\ge c_2{}^{c'_2}$. These inequalities may be asserted by choosing $c'_2:=c'_1+\nu$ with ν sufficiently small since

$$c_1^{c'_1}\ge c_2^{c'_1+\nu}\quad\Leftrightarrow\quad\nu\le c'_1\frac{\ln c_1-\ln c_2}{\ln c_2}.$$

For $c_1 \geq c_2$ this choice is always possible in such a way that $\nu \geq 0$. If follows that for each function $f \in \{f_\Upsilon, f_H, f_{it}\}$, we can find a constant c'_f such that $(c_{max}\alpha)^{c'_f \alpha n d} \geq f(\alpha)$ for $\alpha < 1$. Then, letting c'_{max} be the maximum of these constants c_f gives a function f_{max} with the required properties.

Accordingly, (12) can be bounded by

$$\begin{aligned} &\mathbf{P}[\,|F| = y\,] \\ &\leq n^3 \begin{cases} \max\left(\mathbf{P}\left[\,|F_{SDP}| = \frac{y}{3}\,\right], f_{max}(\frac{y}{3n})\right) & \text{if} \quad \frac{y}{3} \leq \alpha_0 n, \\ \max\left(\mathbf{P}\left[\,|F_{SDP}| = \frac{y}{3}\,\right], \mathbf{P}\left[\,|\overline{H}| = \frac{y}{3}\,\right]\right) & \text{if} \quad \frac{y}{3} > \epsilon n, \\ \max\left(\mathbf{P}\left[\,|F_{SDP}| = \frac{y}{3}\,\right], f_{max}(\frac{y}{3n}), \mathbf{P}\left[\,|\overline{H}| = \frac{y}{3}\,\right]\right) & \text{otherwise.} \end{cases} \end{aligned}$$

Returning to the evaluation of $\mathbf{P}[\,|F| = y\,] \cdot \binom{n}{y}\textbf{3-COL}\,(y)$ in Equation (10), we therefore can certainly guarantee the required polynomial bound on $\mathbf{E}[\,t\,]$ by asserting the following three relations

$$\mathbf{P}\left[\,|F_{SDP}| = \frac{y}{3}\,\right] \cdot \binom{n}{y}\textbf{3-COL}\,(y) = \mathcal{O}(1) \quad \text{for all} \quad y \leq n, \tag{13a}$$

$$\mathbf{P}\left[\,|\overline{H}| = \frac{y}{3}\,\right] \cdot \binom{n}{y}\textbf{3-COL}\,(y) = \mathcal{O}(1) \quad \text{for all} \quad \frac{y}{3} > \alpha_0 n, \tag{13b}$$

$$f_{max}\left(\frac{y}{3n}\right) \cdot \binom{n}{y}\textbf{3-COL}\,(y) = \mathcal{O}(1) \quad \text{for all} \quad \frac{y}{3} < \epsilon n. \tag{13c}$$

The first of these equations can be derived with the help of Lemma 4: Let $y = \alpha n$. Then

$$\begin{aligned} \mathbf{P}\left[\,|F_{SDP}| = \frac{y}{3}\,\right] \cdot \binom{n}{y}\textbf{3-COL}\,(y) &= \mathbf{P}\left[\,|F_{SDP}| = \frac{\alpha n}{3}\,\right] \cdot \binom{n}{\alpha n}\textbf{3-COL}\,(\alpha n) \\ &\leq 10^{-n} \cdot \left(\frac{e}{\alpha}\right)^{\alpha n} 3^{\alpha n} \leq 10^{-n} \cdot 10^n \end{aligned}$$

since $(e \cdot 3/\alpha)^\alpha < 10$ for all $0 < \alpha \leq 1$.

Moreover, (13b) and (13c) are proven below in Lemma 17 and Lemma 16 respectively. This concludes the proof of Lemma 10. It remains to establishing the second and the third assertion of (13). We start by showing (13c) in Lemma 16. Here we make use of the terminology introduced in the proof of Lemma 10.

Lemma 16. *For $\alpha/3 < \epsilon$,*

$$f_{max}\left(\frac{\alpha}{3}\right) \cdot \binom{n}{\alpha n}\textbf{3-COL}\,(\alpha n)$$

tends to a constant as n goes to infinity.

Proof. A straight forward calculation shows

$$
\begin{aligned}
&f_{max}\left(\frac{\alpha}{3}\right)\cdot\binom{n}{\alpha n}\mathbf{3\text{-}COL}\,(\alpha n)\le (c_{max}\alpha/3)^{c'_{max}d\cdot\alpha n/3}\cdot\binom{n}{\alpha n}\cdot\exp(\ln 3\cdot\alpha n)\\
&\le\ \exp\left(\left((\ln c_{max}+\ln\alpha-\ln 3)\,c'_{max}\frac{d}{3}-2\log\alpha+\ln 3\right)\alpha n\right)\\
&\le\ \exp\left(\left(\ln c_{max}\cdot\frac{c'_{max}}{3}d+\ln\alpha\left(c'_{max}\frac{d}{3}-3\right)\right)\alpha n\right),
\end{aligned}
$$

which is of order $\mathcal{O}(1)$ for

$$\ln\alpha<\frac{\ln c_{max}\cdot c'_{max}\frac{d}{3}}{3-c'_{max}\frac{d}{3}}.$$

This is satisfied for d sufficiently large and

$$\ln\alpha<2\frac{\ln c_{max}\cdot c'_{max}\frac{d}{3}}{-c'_{max}\frac{d}{3}}=\ln\frac{1}{c_{max}^2}$$

since $c_{max}>1$. Thus, Lemma 16 follows if we choose

$$\epsilon\le\frac{1}{3c_{max}^2}.$$

For completing the proof of Lemma 10 it remains to complement the preceeding result by an argument for (13b).

Lemma 17. *For $\alpha/3>\alpha_0$,*

$$\mathbf{P}\Big[\,|\overline{H}|=\frac{\alpha}{3}n\,\Big]\cdot\binom{n}{\alpha n}\mathbf{3\text{-}COL}\,(\alpha n)$$

is of order $\mathcal{O}(1)$.

Proof. Set $\alpha':=\alpha/3$ and recall that

$$\mathbf{P}\big[\,|\overline{H}|=\alpha' n\,\big]\le f_{(0)}(\alpha')+f_{it}(\alpha') \tag{14}$$

with $f_{(0)}(\alpha')=\exp\big(-\alpha' n\left(\log\alpha'+c_{(0)}\cdot d\right)\big)$ and $f_{it}(\alpha')=(c_{it}\alpha')^{c'_{it}d\cdot\alpha' n}$. Notice that $f_{it}(\alpha')\ge 1$ for $\alpha'\ge 1/c_{it}\ge 1/c_{max}>\epsilon$ and so this bound on $\mathbf{P}\big[\,|\overline{H}|=\alpha' n\,\big]$ gets trivial for such α'. Therefore, we proceed by discussing the cases $\alpha_0\le\alpha'<\epsilon$ and $\alpha'\ge\epsilon$ separately. In both cases we will derive a constant bound on

$$\mathbf{P}\big[\,|\overline{H}|=\alpha' n\,\big]\cdot\binom{n}{\alpha n}\mathbf{3\text{-}COL}\,(\alpha n)=\mathbf{P}\Big[\,|\overline{H}|=\frac{\alpha}{3}n\,\Big]\cdot\binom{n}{\alpha n}\mathbf{3\text{-}COL}\,(\alpha n)\,. \tag{15}$$

This establishes the lemma.

First, assume $\alpha' < \epsilon$. We evaluate the two terms $f_{(0)}(\alpha')$ and $f_{it}(\alpha')$ contributing to the bound in (14) independently. By Lemma 15 and Lemma 16, we know that

$$f_{it}\left(\frac{\alpha}{3}\right) \cdot \binom{n}{\alpha n} \textbf{3-COL}\,(3\alpha n) = \mathcal{O}(1)\,.$$

Using (5) we obtain a corresponding result for $f_{(0)}(\alpha')$:

$$\begin{aligned}
&f_{(0)}\left(\frac{\alpha}{3}\right) \cdot \binom{n}{\alpha n} \textbf{3-COL}\,(\alpha n) \\
&\leq \exp\!\left(-\frac{\alpha}{3} n \left(\log \frac{\alpha}{3} + c_{(0)} \cdot d\right)\right) \cdot \binom{n}{\alpha n} \cdot \exp(\ln 3 \cdot \alpha n) \\
&\leq \exp\!\left(-\frac{\alpha}{3} n \left(\log \alpha - \log 3 + c_{(0)} \cdot d + 3 \cdot 2 \log(\alpha) - 3 \ln 3\right)\right) \\
&\leq \exp\!\left(-\frac{\alpha}{3} n \left(6 \log \alpha + \frac{6}{10} c_{(0)} \cdot d\right)\right).
\end{aligned}$$

This term tends to a constant for

$$-\log \alpha < \frac{1}{10} c_{(0)} \cdot d,$$

which holds for $\alpha/3 = \alpha' > \alpha_0$ since we chose $\alpha_0 = 2^{-c_{(0)} d/10}$. This settles the case $\alpha_0 < \alpha' < \epsilon$.

For $\alpha' \geq \epsilon$ we need to follow a different strategy. By monotonicity we have

$$\mathbf{P}\left[\,|\overline{H}| = \alpha' n\,\right] \leq \mathbf{P}\left[\,|\overline{H}| \geq \alpha' n\,\right] \leq \mathbf{P}\left[\,|\overline{H}| \geq \alpha'' n\,\right]$$

for $\alpha'' \leq \alpha'$. Moreover,

$$\binom{n}{\alpha n} \textbf{3-COL}\,(\alpha n) \leq \left(\frac{e}{\alpha}\right)^{\alpha n} 3^{\alpha n} \leq 10^n$$

since $(e \cdot 3/\alpha)^{\alpha} < 10$ for all $0 < \alpha \leq 1$. Accordingly, (15) can be bounded from above by

$$\mathbf{P}\left[\,|\overline{H}| \geq \epsilon n\,\right] \cdot 10^n \leq f_{(0)}(\epsilon) \cdot 10^n + f_{it}(\epsilon) \cdot 10^n.$$

For these two summands we can now easily provide constant bounds:

$$\begin{aligned}
f_{(0)}(\epsilon) \cdot 10^n &= \exp(-\epsilon n \left(\log \epsilon + \Omega\,(d)\right)) \exp(\ln 10 \cdot n) = \exp(n(\mathcal{O}(1) - \Omega\,(d))) \\
f_{it}(\epsilon) \cdot 10^n &= (c_{it} \cdot \epsilon)^{c'_{it} d \cdot \epsilon n} \exp(\ln 10 \cdot n) = \exp(n(\mathcal{O}(1) - \Omega\,(d)))\,.
\end{aligned}$$

Observe that the second part of this proof demonstrates in particular that Algorithm 1 can even afford to use the recovery procedure on the whole graph G if $|\overline{H}|$ grows beyond ϵn.