Why almost all k-colorable graphs are easy to color

Amin Coja-Oghlan * Michael Krivelevich † Dan Vilenchik ‡

May 13, 2007

Abstract

Coloring a k-colorable graph using k colors ($k \geq 3$) is a notoriously hard problem. Considering average case analysis allows for better results. In this work we consider the uniform distribution over k-colorable graphs with n vertices and exactly cn edges, c greater than some sufficiently large constant. We rigorously show that all proper k-colorings of most such graphs are clustered in one cluster, and agree on all but a small, though constant, portion of the vertices. We also describe a polynomial time algorithm that *whp* finds a proper k-coloring of such a random k-colorable graph, thus asserting that most such graphs are easy to color. This should be contrasted with the setting of very sparse random graphs (which are k-colorable *whp*), where experimental results show some regime of edge density to be difficult for many coloring heuristics.

1 Introduction

A *k-coloring* f of a graph $G = (V, E)$ is a mapping from its set of vertices V to $\{1, 2, ..., k\}$. f is a *proper coloring* of G if for every edge $(u, v) \in E$, $f(u) \neq f(v)$. The minimal k s.t. G admits a proper k-coloring is called the chromatic number, commonly denoted by $\chi(G)$. In this work we think of $k > 2$ as some fixed integer, say $k = 3$ or $k = 100$.

1.1 Phase Transitions, Clusters, and Graph Coloring Heuristics

The problem of properly k-coloring a k-colorable graph is one of the most famous NP-hard problems. The plethora of worst-case NP-hardness results for problems in graph theory motivates the study of heuristics that give "useful" answers for "typical" subset of the problem instances, where "useful" and "typical" are usually not well defined. One way of evaluating and comparing heuristics is by running them on a collection of input graphs ("benchmarks"), and checking which heuristic usually gives better results. Though empirical results are sometimes informative, we seek more rigorous measures of evaluating heuristics. Although satisfactory approximation algorithms are known for several NP-hard problems, the coloring problem is not amongst them. In fact, Feige and Kilian [16] prove that no polynomial time algorithm approximates $\chi(G)$ within a factor of $n^{1-\varepsilon}$ for all input graphs G on n vertices, unless ZPP=NP.

*Institute for Informatics, Humboldt-University, Berlin, Germany. Email: `coja@informatik.hu-berlin.de`.

†School Of Mathematical Sciences, Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel. Email: `krivelev@post.tau.ac.il`. Research supported in part by USA-Israel BSF Grant 2002-133, and by grant 526/05 from the Israel Science Foundation.

‡School of Computer Science, Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel. Email: `vilenchi@post.tau.ac.il`.

When very little can be done in the "worst case", comparing heuristics' behavior on "typical", or "average", instances comes to mind. One possibility of rigourously modeling such "average" instances is to use random models. In the context of graph coloring, the $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,m}$ models, pioneered by Erdős and Rényi, might appear to be the most natural candidates. A random graph G in $\mathcal{G}_{n,p}$ consists of n vertices, and each of the $\binom{n}{2}$ possible edges is included w.p. $p = p(n)$ independently of the others. In $\mathcal{G}_{n,m}$, $m = m(n)$ edges are picked uniformly at random. Bollobás [9] and Łuczak [24] calculated the probable value of $\chi(\mathcal{G}_{n,p})$ to be *whp* [1] approximately $n \ln(1-p)/(2\ln(np))$ for $p \in [C_0/n, 0.99]$. Thus, the chromatic number of $\mathcal{G}_{n,p}$ is typically rather high (roughly comparable with the average degree np of the random graph) – higher than k, when thinking of k as some fixed integer, say $k = 3$, and allowing the average degree np to be arbitrarily large.

Remarkable phenomena occurring in the random graph $\mathcal{G}_{n,m}$ are **phase transitions**. With respect to the property of being k-colorable, such a phase transition takes place too. More precisely, there exists a threshold $d_k = d_k(n)$ such that graphs with average degree $2m/n > (1+\varepsilon)d_k$ do not admit any proper k-coloring *whp*, while graphs with a lower average degree $2m/n < (1-\varepsilon)d_k$ will have one *whp* [1]. In fact, experimental results show that random graphs with average degree just below the k-colorability threshold (which are thus k-colorable *whp*) are "hard" for many coloring heuristics. One possible explanation for this, backed up by partially non-rigorous analytical tools from statistical physics [27], is the surmise that k-colorable graphs with average degree just below the threshold show a **clustering phenomenon** of the solution space. That is, typically random graphs with density close to the threshold d_k have an exponential number of **clusters** of k-colorings. While any two k-colorings in distinct clusters disagree on at least εn vertices, any two k-colorings within one cluster coincide on $(1-\varepsilon)n$ vertices. Furthermore, each cluster has a linear number of "frozen" vertices (a subset of vertices $U \subseteq V$ is said to be **frozen** in G if in every proper k-coloring of G all vertices in U receive the same color. A vertex is said to be **frozen** if it belongs to a frozen subset of vertices). Recently some supporting evidence for this theory was proved rigorously for random k-SAT, $k \geq 8$ [25, 5, 26]

Now, the algorithmic difficulty with such a clustered solution space seems to be that the algorithm does not "steer" into one cluster but tries to find a "compromise" between the colorings in distinct clusters, which actually is impossible. By contrast, the recent Survey Propagation algorithm can apparently cope with the existence of a huge number of clusters [11], though no rigorous analysis of the algorithm is known.

In this work we consider the regime of denser graphs, i.e. the average degree will be by a constant factor higher than the k-colorability threshold. In this regime, almost all graphs are not k-colorable, and therefore we shall condition on the event that the random graph is k-colorable. Thus, we consider the most natural distribution on k-colorable graphs with given numbers n of vertices and m of edges, namely, the uniform distribution $\mathcal{G}^{\text{uniform}}_{n,m,k}$. For $m/n \geq C_0$, C_0 a sufficiently large constant, we are able to **rigorously** prove that the space of all proper k-colorings of a typical graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$ has the following structure: an exponential number of proper k-colorings arranged in a **single cluster**. We also describe a polynomial time algorithm that *whp* k-colors $\mathcal{G}^{\text{uniform}}_{n,m,k}$ with $m \geq C_0 n$ edges .

Thus, our result shows that when a k-colorable graph has a single cluster of k-colorings, though its volume may be exponential, then typically the problem is easy. This in some sense complements in a rigorous way the results in [27, 12] (where it is conjectured that when the clustering is complicated, more sophisticated algorithms are needed). Besides, standard probabilistic calculations show that when $m \geq Cn \log n$, C a sufficiently large constant, a random k-colorable graph will have *whp* only one proper k-coloring; indeed, it is known that such graphs are even easier to color than in the case

[1] When writing *whp* ("with high probability") we mean with probability tending to 1 as n goes to infinity.

$m = O(n)$, which is the focus of this paper. A further appealing implication of our result is the fact that almost all k-colorable graphs, sparse or dense, can be efficiently colored. This extends a previous result from [29] concerning dense graphs (i.e., $m = \Theta(n^2)$).

1.2 Results and Techniques

In this section we state our main results precisely. First, we discuss the structure of the solution space (i.e., the set of all proper k-colorings) of $\mathcal{G}_{n,m,k}^{\text{uniform}}$. Formally we prove:

Theorem 1.1. *(clustering phenomena) Let G be random graph from $\mathcal{G}_{n,m,k}^{\text{uniform}}$, $m \geq C_0(k)n$, $C_0(k)$ a sufficiently large constant that depends on k. Then* whp *G enjoys the following properties:*

1. *All but $e^{-\Theta(m/n)}n$ vertices are frozen.*
2. *The graph induced by the non-frozen vertices decomposes into connected components of at most logarithmic size.*
3. *Letting $\beta(G)$ be the number of proper k-colorings of G, we have $\frac{1}{n}\log\beta(G) = e^{-\Theta(m/n)}$.*

Notice that property 1 implies in particular that any two proper k-colorings differ on at most $e^{-\Theta(m/n)}n$ vertices. The above characterization of the solution space of $\mathcal{G}_{n,m,k}^{\text{uniform}}$ leads to the following algorithmic result:

Theorem 1.2. *(algorithm) There exists a polynomial time algorithm that* whp *properly k-colors a random graph from $\mathcal{G}_{n,m,k}^{\text{uniform}}$, $m \geq C_1(k)n$, $C_1(k)$ a sufficiently large constant that depends on k.*

Specifically, we prove that the polynomial time algorithm in Theorem 1.2 is the one presented by Alon and Kahale [6] (more details in Section 4). Our analysis gives for $C_0, C_1 = \Theta(k^4)$, but no serious attempt was made to optimize the power of k.

The Erdős-Rényi graph $\mathcal{G}_{n,m}$ and its well known variant $\mathcal{G}_{n,p}$ are both very well understood and have received much attention during the past years. However, the event of a random graph in $\mathcal{G}_{n,m}$ being k colorable, when k is fixed and the average degree $2m/n$ is above the k-colorability threshold, is very unlikely. Therefore, the distribution $\mathcal{G}_{n,m,k}^{\text{uniform}}$ differs from $\mathcal{G}_{n,m}$ significantly. In effect, many techniques that have become standard in the study of $\mathcal{G}_{n,m}$ just do not carry over to $\mathcal{G}_{n,m,k}^{\text{uniform}}$ – at least not directly. In particular, the contriving event of being k-colorable causes the edges in $\mathcal{G}_{n,m,k}^{\text{uniform}}$ to be dependent. The inherent difficulty of $\mathcal{G}_{n,m,k}^{\text{uniform}}$ has led many researchers to consider the more approachable, but considerably less natural, **planted distribution** introduced by Kučera [23] and denoted throughout by $\mathcal{G}_{n,m,k}^{\text{plant}}$. In this context we can selectively mention [6, 8, 10, 13, 22]. In the planted distribution, one first fixes some k-coloring (that is a partition of the vertices into k color classes), and then picks uniformly at random m edges that respect this coloring. Due to the "constructive" definition of $\mathcal{G}_{n,m,k}^{\text{plant}}$, the techniques developed in the study of $\mathcal{G}_{n,m}$ can be applied to $\mathcal{G}_{n,m,k}^{\text{plant}}$ immediately, whence the model is rather well understood [6].

Of course the $\mathcal{G}_{n,m,k}^{\text{plant}}$ model is somewhat artificial and therefore provides a less natural model of random instances than $\mathcal{G}_{n,m,k}^{\text{uniform}}$. Nevertheless, devising new ideas for analyzing $\mathcal{G}_{n,m,k}^{\text{uniform}}$, in this paper we show that $\mathcal{G}_{n,m,k}^{\text{uniform}}$ and $\mathcal{G}_{n,m,k}^{\text{plant}}$ actually share many structural graph properties such as the existence of a single cluster of solutions. As a consequence, we can prove that a certain algorithm, designed with $\mathcal{G}_{n,m,k}^{\text{plant}}$ in mind, works for $\mathcal{G}_{n,m,k}^{\text{uniform}}$ as well. In other words, presenting new methods for analyzing heuristics on random graphs, we can show that algorithmic techniques invented for the somewhat artificial $\mathcal{G}^{\text{plant}}_{n,m,k}$ model extend to the canonical $\mathcal{G}^{\text{uniform}}_{n,m,k}$.

To obtain these results, we use two main techniques. $\mathcal{G}^{\text{plant}}_{n,m,k}$ (and the analogous $\mathcal{G}^{\text{plant}}_{n,p,k}$ in which every edge respecting the planted k-coloring is included with probability p) is already very well understood, and in particular the probability of some graph properties that interest us can be easily estimated for $\mathcal{G}^{\text{plant}}_{n,m,k}$ using standard probabilistic calculations. It then remains to find a reasonable "exchange rate" between $\mathcal{G}^{\text{plant}}_{n,m,k}$ and $\mathcal{G}^{\text{uniform}}_{n,m,k}$. We use this approach to estimate the probability of "complicated" graph properties, which hold with extremely high probability in $\mathcal{G}^{\text{plant}}_{n,m,k}$. The other method is directly analyzing $\mathcal{G}^{\text{uniform}}_{n,p,k}$, crucially overcoming edge-dependency issues. This method tends to be more complicated than the first one, and involves intricate counting arguments.

1.3 Related Work

The k-colorability problem exhibits a sharp threshold phenomena in the sense that there exists a function $d_k(n)$ s.t. a random graph from $\mathcal{G}_{n,m}$ is *whp* k-colorable if $2m/n < (1-\varepsilon)d_k(n)$ and is *whp* not k-colorable if $2m/n > (1+\varepsilon)d_k(n)$ (cf. [1]). For example, it is known that $d_3(n) \geq 4.03n$ [3] and $d_3(n) \leq 5.044n$ [2]. Therefore, a typical graph in $\mathcal{G}_{n,m}$ with $m = cn$ will not be k-colorable (when thinking of k as a fixed integer, say $k = 3$, and allowing the average degree c to be an arbitrary constant, say $c = 100$, or even a growing function of n). Thus, when considering relatively dense random graphs, one should take care when defining the underlying distribution, e.g. consider $\mathcal{G}^{\text{plant}}_{n,m,k}$ or $\mathcal{G}^{\text{uniform}}_{n,m,k}$.

Almost all polynomial-time coloring heuristics suggested so far were analyzed when the input graph is sampled according to $\mathcal{G}^{\text{plant}}_{n,p,k}$, or various semi-random variants thereof (and similarly for other graph problems such as clique, independent set, and random satisfiability problems). Alon and Kahale [6] suggest a polynomial time algorithm, based on spectral techniques, that *whp* properly k-colors a random graph from $\mathcal{G}^{\text{plant}}_{n,p,k}$, $np \geq C_0k^2$, C_0 a sufficiently large constant. Combining techniques from [6] and [13], Böttcher [10] suggests an expected polynomial time algorithm for $\mathcal{G}^{\text{plant}}_{n,p,k}$ based on SDP (semi-definite programming) for the same p values. Much work was done also on semi-random variants of $\mathcal{G}^{\text{plant}}_{n,p,k}$, e.g. [8, 13, 17, 22].

On the other hand, very little is known on non-planted distributions over k-colorable graph, such as $\mathcal{G}^{\text{uniform}}_{n,m,k}$. In this context one can mention the work of Prömel and Steger [28] who analyze $\mathcal{G}^{\text{uniform}}_{n,m,k}$ but with a parameterization which causes $\mathcal{G}^{\text{uniform}}_{n,m,k}$ and $\mathcal{G}^{\text{plant}}_{n,m,k}$ to coincide, thus not shedding light on the setting of interest in this work. Similarly, Dyer and Frieze [15] deal with very dense graphs (of average degree $\Omega(n)$).

1.4 Paper's Structure

The rest of the paper is structured as follows. We first discuss in Section 2 some general properties that a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$ typically possesses. Then in Section 3 we discuss some more properties that correspond to the clustering phenomena – this in turn will imply Theorem 1.1. The algorithmic perspective is discussed in Section 4 along with a proof of Theorem 1.2. Concluding remarks are given in Section 7. Sections 5 and 6 complete the technical details missing in Sections 2 and 3.

2 General Properties of $\mathcal{G}^{\text{uniform}}_{n,m,k}$

In this section we discuss general properties that a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$ typically possesses. These properties are not particular to $\mathcal{G}^{\text{uniform}}_{n,m,k}$, rather are common (maybe in a slightly different formulation) to many graph distributions, for example $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,m}$.

We start by discussing the discrepancy property (such discussions are ample for $\mathcal{G}_{n,p}$ and $\mathcal{G}^{\text{plant}}_{n,p,k}$, e.g. [6, 19, 20]). This discussion may be of interest of its own, as generally discrepancy properties play a fundamental role in the proof of many important graph properties such as expansion, the spectra of the adjacency matrix, etc, and indeed the discrepancy property plays in our case a major role both in the algorithmic perspective and in the analysis of the clustering phenomena. Therefore, the new approach taken here in establishing the discrepancy property may be of use in other settings where edges are dependent. For another example of proving discrepancy in a model where edges are dependent the reader is referred to [7].

Proposition 2.1. *Let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}, m \geq C_0 k^1 0n, C_0$ a sufficiently large constant. Then* whp *the following holds for every proper k-coloring φ of G. Let $V_1, \ldots, V_k$ be the k color classes of φ, and set $p = p(\varphi)$ s.t. $m = \left(\sum_{i<j} |V_i||V_j|\right) p$ holds. Let G' be the graph obtained from G by removing vertices with degree greater than $10np$. There exists a constant c s.t. for every two sets of vertices A, B, $|A| = a \leq |B| = b$, at least ones of the following two conditions holds for G':*

- $e(A,B) \leq c \cdot \mu(A,B)$,
- $e(A,B) \cdot \ln(\frac{e(A,B)}{\mu(A,B)}) \leq c \cdot b \cdot \ln \frac{n}{b}$,

where $\mu(A,B) = |A||B|p$.

Note that if A and B in Proposition 2.1 are disjoint then $\mu(A,B)$ is the expected number of edges between A and B, had the underlying probability space been $\mathcal{G}^{\text{plant}}_{n,m,k}$ with φ as the planted assignment. Otherwise, $\mu(A,B)$ is an upper bound on that value.

The proof of this proposition is an example of the direct analysis approach. That is, overcoming the edge-dependency issue, using an intricate counting argument, we directly analyze $\mathcal{G}^{\text{uniform}}_{n,m,k}$.

As a corollary of Proposition 2.1 we get the following fact – Corollary 2.2. This fact (in a somewhat different formulation) is proved e.g. in [6] for the planted setting, and is common in the study of random graphs in general.

Corollary 2.2. *Let $\delta \in (0,1]$ be some positive number. Let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$, $m \geq C_0 k^4 n$, $C_0 = C_0(\delta)$ a sufficiently large constant. Then* whp *there exists* no *subgraph of G on at most $\delta n/(1000k)$ vertices in which the average degree is at least $\delta m/(nk)$.*

The next property, whose proof builds upon the discrepancy property just stated, concerns the spectral properties of the adjacency matrix of a typical graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$. Let us start by giving some intuition for considering the spectrum of the graph. For the sake of simplicity assume $k = 3$. Suppose that every vertex in G had exactly d neighbors in every color class other than its own. Let F be the 2-dimensional subspace of all vectors $x = (x_v : v \in V)$ which are constant on every color class, and whose sum is 0. A simple calculation shows that any non-zero element of F is an eigenvector of $A = A(G)$ (A being the adjacency matrix of G) with eigenvalue $-d$. Moreover, if E is the union of random matchings, one can show that $-d$ is *whp* the smallest eigenvalue of A and that

F is precisely the eigenspace corresponding to $-d$. Thus, any linear combination t of e_{n-1} and e_n is constant on every color class (e_{n-1}, e_n being the two smallest eigenvectors of A). If the median of t is 0 and its l_2-norm is $\sqrt{2n}$, then t takes the values 0, 1 or -1 depending on the color class, and this gives a proper coloring of G. In our model these regularity assumptions do not hold, and therefore the spectral step only gives an approximation of some proper k-coloring. A further complication in our setting is the fact that the edges of G are not independent, which is the usual assumption in papers where the spectrum of random graphs is analyzed, for example [6, 19, 20]. Therefore the analysis in our case is more complicated.

Notation. Let $G = (V, E)$ be distributed according to $\mathcal{G}^{\text{uniform}}_{n,m,k}$. Let $d_{avg} = 2m/n$ be the average degree in G, $G' = (V', E')$ be the graph obtained from G by deleting all vertices of degree greater than $2d_{avg}$, and A' be the adjacency matrix of G'. For a symmetric matrix $M \in \mathbb{R}^{q \times q}$, denote by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q$ the eigenvalues of M, by $e_1, e_2, \ldots, e_q$ the corresponding eigenvectors, chosen so that they form an orthonormal basis of $\mathbb{R}^q$, and $\|M\| = \max_i |\lambda_i|$. Given a $n \times n$ matrix M that corresponds in some way to a n-vertexed graph, we usually index the rows and columns of M by the vertices of the graph. For example, given two vertex sets $V_i, V_j \subseteq V$, we let $J_{V_i \times V_j}$ be the $n \times n$ matrix whose entries are $J_{u,v} = 1$ if $(u, v) \in V_i \times V_j$, and $J_{u,v} = 0$ otherwise.

Proposition 2.3. *Let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$, $m \geq C_0 k^1 0n, C_0$ a sufficiently large constant. G' has* whp *a k-coloring $V_1, \ldots, V_k$ such that the following holds. Let A' be the adjacency matrix of G', $p = m^{-1} \cdot \sum_{i<j} V_i V_j$, and $M' = \left(\sum_{i \neq j} p J_{V_i \times V_j}\right) - A'$. Then $\|M'\| \leq (d_{avg}/k)^{0.9}$. Moreover, $|V \setminus V'| \leq n/d_{avg}$.*

Let us sketch how Proposition 2.3 completes the motivation we just gave. For a k-coloring $V_1, \ldots, V_k$ of G we let $\mathbf{1}_{V_i} \in \mathbb{R}^n$ denote the vector whose entries are 1 for V_i and 0 otherwise, $\mathbf{1}$ be the all-one vector, and $\xi^{(i,j)} = \mathbf{1}_{V_i} - \mathbf{1}_{V_j}$. Generalizing the above discussion for any fixed k (assuming again that every vertex in G has exactly d neighbors in every color class other than its own), then one can easily verify that the $\xi^{(i,j)}$'s are eigenvectors of $A(G)$ with eigenvalue $-d$. Furthermore, together with the all-one vector, $\mathbf{1}$, they span a k-dimensional subspace $K \subseteq \mathbb{R}^n$. It is also easy to verify that $K \perp M'$ (M' as defined in Proposition 2.3). Therefore, M' is a shift of A' so that the k eigenvectors corresponding the to largest eigenvalues (in absolute value) – are projected out. If we further assume that $d = d_{avg}/(k-1)$ (that is, every vertex has the same number of vertices in every color class other than its own), then $\|M'\| \leq (d_{avg}/k)^{0.9}$ tells us that the other eigenvalues of A, the ones perpendicular to K, are negligible w.r.t. to the ones corresponding to K. Therefore the "dominant" part of A corresponds to the eigenvalues of the proper coloring, even when these eigenvectors are somewhat distorted due to the "noise" coming from the irregularly of the graph.

3 The Clustering Phenomenon

In this section we analyze the solution space (proper k-colorings) of a typical random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$, $m \geq C_k n$, C_k a sufficiently large constant, and prove Theorem 1.1. Our techniques should be contrasted with the techniques used to analyze the solution space of near-threshold (both above and below) instances. In this context one can mention the work in [25, 4, 5, 26], where the structure of the solution space was analyzed directly (mainly using second moment calculations). This is possible due to the fair simpleness of the underlying probabilistic model (edges are chosen uniformly at random, in $\mathcal{G}_{n,m}$, or independently of each other in $\mathcal{G}_{n,p}$). In our setting, $\mathcal{G}^{\text{uniform}}_{n,m,k}$, the edges are far from being independent of each other, and therefore trying to characterize directly the relations between the different k-colorings may lead to a dead-end. In this paper we take a different approach. We start by studying the structure of a typical graph in $\mathcal{G}_{n,m,k}^{\text{uniform}}$. We are able to characterize such graphs in a manner which reveals the geometrical structure of the solution space. In particular, this suffices to prove Theorem 1.1.

We describe a subset of the vertices, referred to as the *core* vertices, which plays a crucial role in understanding $\mathcal{G}_{n,m,k}^{\text{uniform}}$, both algorithmically and w.r.t. the clustering phenomena. To get intuition, first consider the distribution $\mathcal{G}_{n,m,k}^{\text{plant}}$, and the case $k = 3$ (that is, 3-colorable graphs with exactly m edges). Every vertex v is expected to have m/n neighbors in every color class other than its own. Suppose indeed that this is the case. To complete the discussion we need two extra facts.

Fact 3.1. *Let G be a random graph in $\mathcal{G}_{n,m,3}^{\text{plant}}$, $m/n \geq C_0$, C_0 a sufficiently large constant. Then* whp *there exists* no *subgraph of G containing at most $n/1000$ vertices whose average degree is at least m/n.*

Fact 3.2. *Let G be a random graph in $\mathcal{G}_{n,m,3}^{\text{plant}}$, $m/n \geq C_0$, C_0 a sufficiently large constant. Then* whp *there exists* no *two proper 3-colorings of G at distances at least $n/1000$ from each other.*

"Distance" should be interpreted in the natural sense, a precise definition is given later on. Fact 3.1, with somewhat different constants is proven in [6] (and also in this paper – Corollary 2.2 for the uniform setting), and Fact 3.2 is proven using first moment calculations (similar arguments to Lemma 6.18 ahead).

Now suppose that these two facts are indeed true (which is typically the case), and further assume that every vertex has the expected number of neighbors in every color class (which is typically *not* the case when m/n is constant). Then we claim that the graph is uniquely 3-colorable. If not, then let ψ be a proper 3-coloring of the graph, not equal to the planted 3-coloring φ. Let U be the set of vertices that are colored differently in φ and ψ. Every $u \in U$, say $\psi(u) = c$, must have at least m/n neighbors in $G[U]$ – the neighbors of u in G which are colored c according to φ. However, $|U| \leq n/1000$ due to Fact 3.2, but the minimal degree in $G[U]$ is at least m/n, contradicting Fact 3.1.

Observe that if this is the case, then all vertices of the graph are *frozen.* When $m/n \geq C_0 \log n$, then *whp* every vertex in G has roughly m/n neighbors in every color class other than its own, and combined with the two facts, one derives that typically such graphs in $\mathcal{G}_{n,m,3}^{\text{plant}}$ are uniquely 3-colorable. However, when $m/n = O(1)$ this is *whp* not the case. In particular, *whp* $e^{-\Theta(m/n)}n$ vertices will be isolated (degree 0). Nevertheless, in the case $m/n = O(1)$ there exists a large subgraph of G showing a very similar behavior to the aforementioned one, both in the planted and the uniform setting. The set of vertices inducing this subgraph is called a *core.* A similar notion of core, though in a different context, was first introduced by Alon and Kahale [6].

Definition 3.3. *A set of vertices $\mathcal{H}$ is called a δ-**core** of $G = (V, E)$ w.r.t. a proper k-coloring ψ of the vertices of G with color classes $V_1, \dots, V_k$, if the following properties hold for every $v \in \mathcal{H}$:*

- *v has at least $(1 - \delta)|V_i|p_i$ neighbors in $\mathcal{H} \cap V_i$ for every $i \neq \psi(v)$.*
- *v has at most δr neighbors from $V \setminus \mathcal{H}$,*

where $p_i = \frac{2m}{n} \cdot \frac{1}{n-|V_i|}$ and $r = \max_i |V_i|p_i$.

We proceed by asserting some properties that a core typically possesses. Before doing so, we assert two facts that do not concern directly the core, but play an important role in proving the core's properties. A graph G is said to be ε-**balanced** if it a admits a proper k-coloring in which every color class is of size $(1 \pm \varepsilon)\frac{n}{k}$. We say that a graph is balanced if it is 0-balanced.

In the common definition of $\mathcal{G}_{n,m,k}^{\text{plant}}$ all color classes of the planted k-coloring are of the same cardinality, namely n/k. Therefore, all graphs in $\mathcal{G}_{n,m,k}^{\text{plant}}$ have at least one balanced k-coloring (the planted one). In the uniform setting this need not be the case, at least not a-priori. However, as the following proposition asserts, this is basically the case *whp*.

Proposition 3.4. *Let $m \geq (10k)^4$, then* whp *a random graph in $\mathcal{G}_{n,m,k}^{\text{uniform}}$ is 0.01-balanced.*

Propositions of similar flavor to Proposition 3.4 were proven in similar contexts e.g. [28], and involve rather simple counting arguments. The second property to be established is the following. A graph G in $\mathcal{G}_{n,m,k}^{\text{uniform}}$ is said to be c-**concentrated** w.r.t. a proper k-coloring ψ of G if every coloring at distance at least n/c from ψ leaves at least m/c^2 monochromatic edges.

Proposition 3.5. *Let $\delta \in [0, 1]$ be some positive number. Let G be a random graph in $\mathcal{G}_{n,m,k}^{\text{uniform}}$, $m \geq C_0k^4n$, $C_0 = C_0(\delta)$ a sufficiently large constant. Then* whp *there exists a proper k-coloring φ of G w.r.t. which G is $\delta/(1000k)$-concentrated.*

We now proceed with the core's properties.

Proposition 3.6. *Let $\delta \in (0, 1)$ be some positive number. Let G be a random graph in $\mathcal{G}_{n,m,k}^{\text{uniform}}$, $m \geq C_0k^4n$, $C_0 = C_0(\delta)$ a sufficiently large constant. Then there exist two constants $a_0(\delta), a_1(\delta) > 0$ (independent of m, n) so that* whp *there exists a proper k-coloring φ of G w.r.t. which there exists a δ-core $\mathcal{H}$ satisfying:*

- $|\mathcal{H}| \geq (1 - e^{-m/(a_0nk^9)})n$.
- *The number of edges spanned by $\mathcal{H}$ is at least $(1 - e^{-m/(a_1nk^9)})m$.*
- *Every color class V_i of φ satisfies $0.99n/k \leq |V_i| \leq 1.01n/k$.*

As discussed above for the planted model, if the average degree is sufficiently high (at least logarithmic), then typically $\mathcal{H} = V$. This is also typically the case in $\mathcal{G}_{n,m,k}^{\text{uniform}}$ with $m/n \geq C_0 \log n$. When $m/n = O(1)$, this is no longer true (in either model), as for example *whp* there is a linear number of vertices with degree d for every constant d (in particular $d = 0$).

Proposition 3.7. *Let G be a random graph in $\mathcal{G}_{n,m,k}^{\text{uniform}}, m \geq C_0k^10n, C_0$ a sufficiently large constant. Let $\mathcal{H}$ be some δ-core of G for which Proposition 3.6 holds, and let φ be the underlying k-coloring. If G satisfies Proposition 3.5 w.r.t. φ, and in addition G satisfies Corollary 2.2, then $G[\mathcal{H}]$ is uniquely k-colorable.*

Here and throughout we consider two k-colorings to be the same if one is a permutation of the color classes of the other.

Proposition 3.8. *If $\mathcal{H}, \mathcal{H}'$ are δ-cores of G, and both are uniquely k-colorable, then $\mathcal{H} \cup \mathcal{H}'$ is a δ-core as well. Hence,* whp *there is a unique maximal δ-core.*

Proof. Let $\mathcal{H}$, $\mathcal{H}'$ be two δ-cores of G with corresponding colorings $V_1, \ldots, V_k$ and $V'_1, \ldots, V'_k$. We denote by. By the uniqueness of the coloring it holds that every $V_i \cap \mathcal{H}$ intersects exactly one V'_j. Therefore, w.l.o.g. we may assume that $V_i \cap \mathcal{H} \subseteq V'_i$ for every i. Hence, it is easily verified that $\mathcal{H} \cup \mathcal{H}'$ meets the definition of a core (Definition 3.3) w.r.t. $V_1, \ldots, V_k$ (which equals $V'_1, \ldots, V'_k$). ■

For the rest of the paper, when we refer to a δ-core w.r.t. some coloring, we mean the maximal (unique) one.

Proposition 3.9. *Fix $\delta \in (0,1)$ and let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$, $m \geq C_0 k^4 n$, C_0 a sufficiently large constant. Let $\mathcal{H}$ be a δ-core of G, and let $G[V \setminus \mathcal{H}]$ be the graph induced by the non-core vertices. If $|\mathcal{H}| \geq (1 - e^{-\Theta(m/(nk^9))})n$, then* whp *the largest connected component in $G[V \setminus \mathcal{H}]$ is of size $O(\log n)$.*

Some of the properties discussed in this section were proved in the planted setting $\mathcal{G}^{\text{plant}}_{n,m,k}$, e.g. in [6, 10]. Nevertheless, these proofs use the fact that the edges are chosen uniformly at random. This is of course not the case in the uniform setting (as most choices of m edges uniformly at random result in a graph which is not k-colorable). Therefore, a different approach is needed. One proof technique which we use to prove the core's properties is similar in some sense to the union bound. We first bound the probability that a graph in $\mathcal{G}^{\text{plant}}_{n,m,k}$ does not have the desired property, then we find an exchange rate between the probability of a certain "bad" event occurring in $\mathcal{G}^{\text{plant}}_{n,m,k}$ vs. $\mathcal{G}^{\text{uniform}}_{n,m,k}$. This technique can be applied to "bad" properties that occur with extremely low probability in $\mathcal{G}^{\text{plant}}_{n,m,k}$ (in the order of $e^{-\Theta(n)}$), as the exchange rate that we establish is exponential in n. A detailed exposition of the exchange rate technique is given in Section 5. Unfortunately, some properties, for example Proposition 3.9, hold only with probability $1 - 1/poly(n)$ in $\mathcal{G}^{\text{plant}}_{n,m,k}$. Therefore the exchange rate technique is of no use. Crucially overcoming the edge-dependency issue we directly analyze the uniform distribution. This proof technique, employed e.g. in the proof of Proposition 2.1 and Proposition 3.9, is technically involved, and exemplifies an analysis of a distribution where the events (edge-choice in our case) are dependent, and this dependency seems rather difficult to quantify (and therefore none of the "standard" probabilistic method tools are applicable, at least not immediately).

3.1 Proof of Theorem 1.1

Theorem 1.1 is now an easy consequence of the above discussion. Proposition 3.6 asserts that *whp* a graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$, with the suitable parametrization, will have a big core w.r.t. some proper k-coloring of the vertex set – namely, all but $e^{-\Theta(m/n)}n$ vertices belong to the core. Proposition 3.8 then entails that the core is uniquely k-colorable. Namely, in all proper k-colorings, the core vertices are frozen. Furthermore, this also implies that there is only one cluster of proper k-colorings, in which every two colorings differ on the color of at most $e^{-\Theta(m/n)}n$ vertices. Also, the number of different proper k-colorings is bounded by $\exp\{e^{-\Theta(m/n)}n\}$ (all the possibilities of coloring the non-core vertices). Lastly, Proposition 3.9 asserts the "simpleness" of the subgraph induced by the non-core vertices.

4 The Algorithmic Perspective

In Sections 2 and 3 we implicitly proved that a typical graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$ and in $\mathcal{G}^{\text{plant}}_{n,m,k}$ share many structural properties: spectral properties of the adjacency matrix, the existence of a core, and some properties that it typically enjoys, the non-existence of small yet unexpectedly dense subgraphs

(Corollary 2.2), and so on. In effect, it will turn out that coloring heuristics that prove efficient for $\mathcal{G}^{\text{plant}}_{n,m,k}$ (e.g. [6, 13]) are useful in the uniform setting as well. In particular we shall prove that the coloring algorithm given in [6], designed with the planted distribution in mind, also works in the uniform case. Thus, one merit of our work is justifying the somewhat unnatural usage of planted-solution distributions in average case analysis.

For the sake of completeness we start by giving a short description of Alon and Kahale's algorithm, and discuss the outline of their proof. When describing the algorithm we have a sparse graph in mind, namely $m/n = c$, c a constant satisfying $c \geq C_0 k^4$ (in the denser setting, $m/n = \Omega(\log n)$, matters actually get much simpler).

In the description of the algorithm we use the subprocedure $\mathsf{SpectralApprox}(G, k)$, which we describe in Section 4.1.

$\mathsf{Alon\text{-}Kahale}(G, k)$:
step 1: spectral approximation.
1. $\mathsf{SpectralApprox}(G, k)$.

step 2: recoloring procedure.
2. for $i = 1$ to $\log n$ do:
   2.a for all $v \in V$ simultaneously color v with the least popular color amongst its neighbors.

step 3: uncoloring procedure.
3. while $\exists v \in V$ with less than $m/(n(k-1))$ neighbors colored in some other color do:
   3.a uncolor v.

step 4: Exhaustive Search.
4. let $U \subseteq V$be the set of uncolored vertices.
5. consider the graph $G[U]$.
   5.a if there exists a connected component of size at least $\log n$ – fail.
   5.b otherwise, exhaustively extend the coloring of $V \setminus U$ to $G[U]$.

Figure 1: Alon and Kahale's coloring algorithm

The following theorem is given in [6] (there it is stated with $k = 3$ but the authors point out that it generalizes to any constant k):

Theorem 4.1. *The algorithm* Alon-Kahale whp *properly k-colors a random graph from $\mathcal{G}^{\text{plant}}_{n,m,k}$, $m \geq C_0 k^2 n$, C_0 a sufficiently large constant.*

The algorithm and Theorem 4.1 are originally presented for $\mathcal{G}^{\text{plant}}_{n,p,k}$, however as pointed out by the authors, and as Lemma 5.1 implies, one can safely state it for $\mathcal{G}^{\text{plant}}_{n,m,k}$.

The proof of Theorem 4.1 (according to [6]) proceeds as follows. First, four graph properties are described, and claimed to hold *whp* for a random graph in $\mathcal{G}^{\text{plant}}_{n,m,k}$ with the parametrization of Theorem 4.1. The graph properties are:

P1. The matrix M' defined as in Proposition 2.3 satisfies $\|M'\| \leq d^{0.9}$, where $d = 2m/(nk)$.

P2. There exists *no* subgraph of G containing at most $n/(1000k)$ vertices whose average degree is at least $m/(nk)$.

P3. There exists a 0.99-core $\mathcal{H}$ (w.r.t. the planted coloring) whose size is at least $(1 - e^{-\Theta(m/n)})n$

P4. The largest connected component in the the subgraph induced by the non-core vertices is of size $O(\log n)$.

P2 is stated in [6] in a slightly different formulation, and arguably the proof of Theorem 4.1 is a bit simpler when using **P2** in our formulation.

Now call a graph that possesses **P1-P4** *typical.* Alon and Kahale [6] first prove that indeed *whp* a graph sampled from $\mathcal{G}^{\text{plant}}_{n,m,k}$ is typical. Therefore, one may restrict oneself to typical graphs when proving Theorem 4.1. The proof of the theorem is composed of the following assertions, which are also to be found in [6]. For a planted graph G, we denote by φ its planted k-coloring.

Proposition 4.2. *Assuming G is typical,* SpectralApprox(G, k) *produces a k-coloring which differs from φ on at most $n/(1000k)$ vertices.*

Proposition 4.3. *Assuming G is typical and Proposition 4.2 holds, after the recoloring step ends, the core is colored according to the planted k-coloring φ.*

Proposition 4.4. *Assuming G is typical and Proposition 4.3 holds, the core vertices survive the uncoloring step, and every vertex that survives the uncoloring step is colored according to φ.*

Proposition 4.5. *Assuming G is typical and Proposition 4.4 holds, the exhaustive search completes in polynomial time with a proper k-coloring of the entire graph.*

The proof of Propositions 4.2-4.5, given of course in [6], relies *only* on **P1-P4**. Therefore to prove Theorem 1.2 it suffices to prove that *whp* a graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$ enjoys properties **P1-P4**. One delicate point that needs to be discussed is the fact that an instance from $\mathcal{G}^{\text{uniform}}_{n,m,k}$ does not have a planted coloring. Nevertheless, it suffices to show that there exists a proper k-coloring w.r.t. which **P1-P4** hold (as the algorithm is not required to find any particular coloring, just a proper one).

P1 is given by Proposition 2.3, **P2** by Corollary 2.2, **P3** in Proposition 3.6, and **P4** in Proposition 3.9. Propositions 3.6 and 2.3, as stated, do not guarantee a-priori that **P1** and **P3** should correspond to the same proper k-coloring (which is required to prove Theorem 4.1). Nevertheless, going through the proofs of these propositions it is easily verified that indeed this is the case.

Remark 4.6. *Alon and Kahale analyze only the case $k = 3$, that is 3 colorable graphs. For $k = 3$ one can do with a rather simple procedure in the spectral step. Though the authors of [6] state that their result generalizes to any fixed k, no clue is given as for the extension of the spectral step to the case $k > 3$. One contribution of this work is to explicitly fill out this missing detail – the procedure* SpcetralApprox. *Thus, we also give a full proof of Proposition 4.2 in Section 6.3.*

4.1 The procedure SpectralApprox(G, k).

Before presenting the procedure SpectralApprox(G, k) let us give some motivation. Suppose that G has only one proper k-coloring with color classes $V_1, \ldots, V_k$, and let $\mathcal{E} = \sum_{i \neq j} pJ_{V_i \times V_j}$ ($J_{V_i \times V_j}$ is defined in the notation paragraph before Proposition 2.3 and p satisfies $m = \left(\sum_{i<j} |V_i||V_j|\right) p$). The matrix $\mathcal{E}$ just reflects the coloring $V_1, \ldots, V_k$. Namely, if we think of p as the "edge density" of the bipartite graph consisting of the V_i-V_j-edges ($i \neq j$), then $\mathcal{E}$ reflects the expected edge distribution of the k-partite graph G. In fact, if we could compute $\mathcal{E}$ efficiently then we could easily obtain the

coloring $V_1, \ldots, V_k$ of G using the following simple greedy rule: u and v belong to the same color class iff $\|\mathcal{E}_v - \mathcal{E}_u\| = 0$, where $\|x\|$ denotes the ℓ_2 norm of a vector $x \in \mathbb{R}^n$, and $\mathcal{E}_v$ denotes the v^{th} column of the matrix $\mathcal{E}$. Though we are not given $\mathcal{E}$ we can obtain a fair approximation of it. Specifically, let $\hat{A}$ signify the *rank k approximation* of $A(G)$, obtained as follows. Let $\lambda_1, \ldots, \lambda_k$ be the largest eigenvalues of $A(G)$ in absolute value, and let $e_1, \ldots, e_k$ be corresponding eigenvectors. Then $\hat{A} = \sum_{i=1}^{k} \lambda_i e_i e_i^T$. As we shall prove in Section 6.3, $\hat{A}$ approximates $\mathcal{E}$ is some sense and therefore one can use $\hat{A}$ to compute a good approximation of a proper k-coloring of G. Recall that for a graph G we use G' to denote the graph obtained from G by deleting all vertices of degree greater than $2d_{avg}$ ($d_{avg} = 2m/n$ is the average degree in G).

SpectralApprox(G, k):

1. Compute $\hat{A}$ for $A(G')$.
2. For each $v \in V'$ determine the set $S_v = \{w \in V : \|\hat{A}_v - \hat{A}_w\|^2 \leq 0.01np^2/k\}$.
3. Let $X = \emptyset$.
4. For $i = 1, \ldots, k$ find a vertex x_i such that $X_i = |S_{x_i} \setminus X| \geq (1 - 10^{-10})\frac{n}{k}$; add X_i to X.
5. Output the classes $X_1, \ldots, X_k$.

Figure 2: SpectralApprox(G, k)

5 The Exchange Rate Technique

Let $\mathcal{A}$ be some graph property (it would be convenient for the reader to think of $\mathcal{A}$ as a "bad" property). We start by determining the exchange rate for $Pr[\mathcal{A}]$ between the different distributions. Recall that in the uniform distribution there need not be a balanced k-coloring, as opposed to the common definition of the planted distribution where the planted k-coloring is balanced (i.e. all color classes are of size n/k). Therefore more refined definitions are needed. In addition to the "regular" parameters m, n (or p, n) of the planted/uniform distribution, we introduce k additional parameters $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k \in (-1, k-1]$, $\sum \varepsilon_i = 0$, which characterize the sizes of the different color classes of a proper k-coloring. Specifically, we denote by $\mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}}$, $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$, the distribution where first the vertices are partitioned in to k color classes so that $|V_i| = (1 + \varepsilon_i)n/k$ for every i. Then, every $V_i - V_j$ edge is included w.p. p. Similarly we define $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$. We define $\mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$ to be the uniform distribution over k-colorable graphs that have at least one proper k-coloring where the color classes satisfy $|V_i| = (1 + \varepsilon_i)n/k$.

We use the following notation to denote the probability of $\mathcal{A}$ under the various distributions: $Pr^{\text{uniform,m},\bar{\varepsilon}}[\mathcal{A}]$ denotes the probability of property $\mathcal{A}$ occurring under $\mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$, $Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}]$ for $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$, and $Pr^{\text{planted,n,p},\bar{\varepsilon}}[\mathcal{A}]$ for $\mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}}$.

We shall be mostly interested in the case $m = \left(\sum_{i<j} |V_i||V_j|\right) p$, namely m is the expected number of edges in $\mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}}$. The following lemma, which is proved using rather standard probabilistic calculations, establishes the exchange rate for $\mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}} \rightarrow \mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$.

Lemma 5.1. *($\mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}} \to \mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$) Let $\mathcal{A}$ be some graph property. The following is true when m and p satisfy $m = \left(\sum_{i<j} |V_i||V_j|\right) p$:*

$$Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}] \leq O(\sqrt{m}) \cdot Pr^{\text{planted,n,p},\bar{\varepsilon}}[\mathcal{A}].$$

Proof.(Outline) Let G be a random graph sampled according to $\mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}}$. G has property $\mathcal{A}$ w.p. $Pr^{\text{planted,n,p},\bar{\varepsilon}}[\mathcal{A}]$. Since the distribution of edges in $\mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}}$ is binomial, and m is chosen to be the expected number of edges, standard calculations show that w.p. $\Omega(1/\sqrt{m})$, G has exactly m edges. Also observe that $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}} = \mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}} |$ {The graph has exactly m edges}. Therefore $Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}] = Pr^{\text{planted,n,p},\bar{\varepsilon}}[\mathcal{A}]/\Omega(1/\sqrt{m}) = O(\sqrt{m}) \cdot Pr^{\text{planted,n,p},\bar{\varepsilon}}[\mathcal{A}]$. ■

Next, we obtain $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}} \to \mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$, which is rather involved technically and whose proof embeds interesting results of their own – for example, bounding the expected number of proper k-colorings of a graph in $\mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$. The passage $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}} \to \mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$ is composed of the following two lemmas.

Lemma 5.2. *Let $\mathcal{A}$ be some graph property, then*

$$Pr^{\text{uniform,m},\bar{\varepsilon}}[\mathcal{A}] \leq C_1(n,k,\bar{\varepsilon}) \cdot Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}],$$

where $C_1(n,k,\bar{\varepsilon})$ stands for the expected number of proper k-colorings that a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$ has.

Lemma 5.3. *Let $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ s.t. $\forall i\ |\varepsilon_i| \leq 0.01$, then*

$$C_1(n,k,\bar{\varepsilon}) \leq e^{ke^{-m/(10nk^9)}n}.$$

The following proposition formulates the exchange rate technique in a "practical" way.

Proposition 5.4. *Let $\mathcal{A}$ be some graph property. Then*

$$Pr^{\text{uniform,m}}[\mathcal{A}] \leq o(1) + n^k \cdot e^{ke^{-m/(10nk^9)}n} \cdot \max_{\bar{\varepsilon}:\forall i, |\varepsilon_i| \leq 0.01} Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}]$$

Proof. Let $\mathcal{K}$ be set of all k-colorable graphs with exactly m edges, and let $\mathcal{K}_{\bar{\varepsilon}}$ be all k-colorable graphs that have at least one proper k-coloring with color classes according to $\bar{\varepsilon}$. Proposition 3.4 asserts that

$$\bigcup_{\bar{\varepsilon}:\forall i, |\varepsilon_i| \leq 0.01} \mathcal{K}_{\bar{\varepsilon}} = (1 - o(1))\mathcal{K}.$$

Set

$$\alpha_{\bar{\varepsilon}} = e^{ke^{-m/(10nk^9)}n} \cdot Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}],$$

$$\alpha = \max_{\bar{\varepsilon}:\forall i, |\varepsilon_i| \leq 0.01} \alpha_{\bar{\varepsilon}}.$$

Lemmas 5.2 and 5.3 ensure that at most $\alpha_{\bar{\varepsilon}}$–fraction of the graphs in $\mathcal{K}_{\bar{\varepsilon}}$ have property $\mathcal{A}$. Therefore, the number of graphs in $\mathcal{K}$ that have property $\mathcal{A}$ is at most

$$\left(o(1) + n^k \cdot \alpha\right) |\mathcal{K}|.$$

The n^k factor comes from the fact that there are at most n^k ways to choose $\bar{\varepsilon}$ (that is, at most n^k different $\mathcal{K}_{\bar{\varepsilon}}$'s). ■

Proof.(Lemma 5.2) Fix $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ and let $B(n, k, \bar{\varepsilon})$ be the total number of proper k-colorings on n vertices with the prescribed sizes of the color classes (when we consider two colorings to be the same if one is just a permutation of the color classes of the other). Throughout this section, when referring to a k-coloring, we mean a coloring with the prescribed sizes of the color classes, when $\bar{\varepsilon}$ is clear from the context. Recall that $C_1(n, k, \bar{\varepsilon})$ is defined to be the expected number of proper k-colorings that a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$ has, and $C_2(n, k, \bar{\varepsilon})$ is defined similarly for $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$. Let t_i be the number of graphs on n vertices and m edges which have exactly i proper k-colorings. Let p_i be the probability that a graph with exactly i proper k-colorings is sampled from $\mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$, and let q_i be defined similarly for $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$. For a k-coloring φ, let $\Delta_{n,m,\varphi}$ be the number of graphs on n vertices with m edges for which φ is a proper k-coloring. Observe that due to symmetry $\Delta_{n,m,\varphi}$ is the same for every φ – thus we omit the φ subscript. In the above notation

$$p_i = \frac{t_i}{\sum_{j=1}^{k^n} t_j},$$

$$q_i = t_i \cdot \frac{i}{B(n, k, \bar{\varepsilon})} \cdot \frac{1}{\Delta_{n,m}}.$$

Further observe that

$$B(n, k, \bar{\varepsilon}) \cdot \Delta_{n,m} = \sum_{j=1}^{k^n} j \cdot t_j.$$

This is because every graph with j balanced k-colorings was counted exactly j times in the product $B(n, k, \bar{\varepsilon}) \cdot \Delta_{n,m}$. Lastly,

$$C_1(n, k, \bar{\varepsilon}) = \sum_{i=1}^{k^n} i \cdot p_i = \frac{\sum_{i=1}^{k^n} i \cdot t_i}{\sum_{i=1}^{k^n} t_i},$$

$$C_2(n, k, \bar{\varepsilon}) = \sum_{i=1}^{k^n} i \cdot q_i = \frac{\sum_{i=1}^{k^n} i^2 \cdot t_i}{B(n, k, \bar{\varepsilon}) \cdot \Delta_{n,m}} = \frac{\sum_{i=1}^{k^n} i^2 \cdot t_i}{\sum_{i=1}^{k^n} i \cdot t_i}.$$

Next we obtain the following bound:

$$\frac{Pr^{\text{uniform,m},\bar{\varepsilon}}[\mathcal{A}]}{Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}]} \leq \max_i \frac{p_i}{q_i}.$$

This is established in the following discussion. Let $\mathcal{K}_{\mathcal{A}}$ be the set of graphs in $\mathcal{G}^{\text{uniform}}_{n,m,k,\bar{\varepsilon}}$ for which property $\mathcal{A}$ holds.

$$Pr^{\text{uniform,m},\bar{\varepsilon}}[\mathcal{A}] = \sum_{G \in \mathcal{K}_{\mathcal{A}}} Pr^{\text{uniform,m},\bar{\varepsilon}}[G], \qquad Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}] = \sum_{G \in \mathcal{K}_{\mathcal{A}}} Pr^{\text{planted,m},\bar{\varepsilon}}[G].$$

Now let $b = \max_i \frac{p_i}{q_i}$. For every G in $\mathcal{K}_{\mathcal{A}}$ it holds that $Pr^{\text{uniform,m},\bar{\varepsilon}}[G] \leq b \cdot Pr^{\text{planted,m},\bar{\varepsilon}}[G]$. Therefore,

$$\sum_{G \in \mathcal{K}_{\mathcal{A}}} Pr^{\text{uniform,m},\bar{\varepsilon}}[G] \leq \sum_{G \in \mathcal{K}_{\mathcal{A}}} b \cdot Pr^{\text{planted,m},\bar{\varepsilon}}[G] = b \cdot \sum_{G \in \mathcal{K}_{\mathcal{A}}} Pr^{\text{planted,m},\bar{\varepsilon}}[G].$$

It now remains to estimate $\max_i \frac{p_i}{q_i}$.

$$\frac{Pr^{\text{uniform,m},\bar{\varepsilon}}[\mathcal{A}]}{Pr^{\text{planted,m},\bar{\varepsilon}}[\mathcal{A}]} \leq \max_i \frac{p_i}{q_i} = \max_i \left(\frac{t_i}{\sum_{j=1}^{k^n} t_j}\right) \cdot \left(\frac{B(n,k,\bar{\varepsilon}) \cdot \Delta_{n,m}}{i \cdot t_i}\right)$$
$$= \max_i \left(\frac{1}{i} \cdot \frac{\sum_{j=1}^{k^n} j \cdot t_j}{\sum_{j=1}^{k^n} t_j}\right) = \left(\frac{\sum_{j=1}^{k^n} j \cdot t_j}{\sum_{j=1}^{k^n} t_j}\right) \cdot \left(\max_i \frac{1}{i}\right) = C_1(n,k,\bar{\varepsilon}).$$

■

Since directly estimating $C_1(n,k,\bar{\varepsilon})$ seems an intricate task, the following lemma is very useful.

Lemma 5.5. $C_1(n,k,\bar{\varepsilon}) \leq C_2(n,k,\bar{\varepsilon})$.

Proof. To prove $C_1(n,k,\bar{\varepsilon}) \leq C_2(n,k,\bar{\varepsilon})$, one needs to prove that

$$\left(\sum_{i=1}^{k^n} i \cdot t_i\right)^2 \leq \left(\sum_{i=1}^{k^n} t_i\right) \cdot \left(\sum_{i=1}^{k^n} i^2 \cdot t_i\right).$$

This is just Cauchy-Schwartz, $\left(\sum a_i \cdot b_i\right)^2 \leq \left(\sum a_i^2\right) \cdot \left(\sum b_i^2\right)$, with $a_i = \sqrt{t_i}$ and $b_i = i \cdot \sqrt{t_i}$. ■

The following lemma then finishes the proof of Lemma 5.3.

Lemma 5.6. *Let* $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ *s.t.* $\forall i\ |\varepsilon_i| \leq 0.01$. *Then it holds:*

$$C_2(n,k,\bar{\varepsilon}) \leq e^{ke^{-m/(10nk^9)}n}.$$

Proof. Let G be a graph randomly sampled according to $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$ and let $\varphi = (V_1, V_2, ..., V_k)$ be its planted k-coloring. Let S_k be the group of permutations over the numbers $\{1, .., k\}$. For a k-coloring $\psi = (U_1, U_2, ..., U_k)$, we define the distance between ψ and φ to be

$$dist(\psi, \varphi) = \min_{\sigma \in S_k} \sum_{v \in V} I_v(\psi, \varphi_\sigma), \tag{1}$$

where $I_v(\psi, \varphi_\sigma) = \begin{cases} 1, & \varphi_\sigma(v) \neq \psi(v). \\ 0, & \text{otherwise.} \end{cases}$,

$\varphi_\sigma(v) = \sigma(i)$ for $v \in V_i$, and $\psi(v) = j$ for $v \in U_j$. Put in words, $dist(\psi, \varphi)$ is the number of vertices which belong to different color classes under ψ and φ, when taking the minimum over all possible $k!$ permutations of the color classes in φ.

Let c_r be the probability that a k-coloring (with color classes according to $\bar{\varepsilon}$) at distance r from φ is also a proper coloring of G. Therefore,

$$C_2(n,k,\bar{\varepsilon}) \leq \sum_{r=0}^{n} \binom{n}{r} k^r c_r.$$

Our first task is therefore to upper bound c_r.

Lemma 5.7. *If* $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ *is s.t.* $\forall i\ |\varepsilon_i| \leq 0.01$ *then* $c_r \leq e^{-mr/(10nk^9)}$.

We are now ready to bound $C_2(n,k,\bar{\varepsilon})$.

$$C_2(n,k,\bar{\varepsilon}) \le \sum_{r=0}^{n}\binom{n}{r}k^r c_r = \sum_{r=0}^{n}\binom{n}{r}k^r e^{-mr/(10nk^9)} = \sum_{r=0}^{n}\binom{n}{r}\left(ke^{-m/(10nk^9)}\right)^r 1^{n-r}$$
$$= \left(1+ke^{-m/(10nk^9)}\right)^n \le e^{ke^{-m/(10nk^9)}\cdot n}$$

■

Before proving Lemma 5.7 we establish two more facts.

Lemma 5.8. *Let $\psi = (U_1, U_2, ..., U_k)$ be some k-coloring at distance r from φ. Then there exist i, j, j' s.t. $|U_i \cap V_j|, |U_i \cap V'_j| \ge \frac{r}{3k\cdot(k-1)}$.*

Proof. If not, then for every i there exists some $j = j(i)$ s.t.

$$|U_i \cap V_j| \ge |V_j| - (k-1)\cdot\frac{r}{3k\cdot(k-1)}$$

The last inequality is due to $r \le n$. Observe that this mapping is a bijection, since if $i \ne i'$ and $j(i) = j(i')$ then $|U_i \cap V_j| \ge \frac{0.99n}{k} - \frac{r}{3k} \ge \frac{0.6n}{k}$ and also $|U_{i'} \cap V_j| \ge \frac{0.6n}{k}$, but this implies $U_i \cap U_{i'} \ne \emptyset$, contradicting the definition of ψ. Let σ be the permutation $j(\cdot)$ that was just defined, and consider φ_σ (namely, φ with color-classes permuted according to σ). Since $|U_i \cap V_{\sigma(i)}| \ge |V_{\sigma(i)}| - \frac{r}{3k}$, $dist(\psi,\varphi) \le k\cdot\frac{r}{3k} = \frac{r}{3}$, contradicting the choice of r. ■

Lemma 5.9. *Fix δ, $|\delta| \le 0.01$, and let $r_1 \ge r_2 \ge ... \ge r_k \ge 0$ be a sequence of k integers satisfying $\sum_{i=1}^{k} r_i = \frac{(1+\delta)n}{k}$ and $r_2 \ge \frac{r}{3k\cdot(k-1)}$. Then*

$$\sum_{1\le i<j\le k} r_i\cdot r_j \ge \left(\frac{(1+\delta)n}{k} - \frac{r}{3k\cdot(k-1)}\right)\cdot\frac{r}{3k\cdot(k-1)}.$$

Proof. Let $r = (r_1, r_2, ...r_k)$, and $f(r) = \sum_{1\le i<j\le k} r_i\cdot r_j$. Assuming $r_i \le r_j$, define a new sequence r' by $r'_i = r_i - 1$, $r'_j = r_j + 1$ and $r'_q = r_q$ for $q \ne i, j$. One can verify that $f(r') = f(r) + r_i - r_j - 1$. Since we chose $r_i \le r_j$, $f(r') < f(r)$. It follows that $f(r)$ takes its minimum (under the conditions $r_2 \ge \frac{r}{3k\cdot(k-1)}$ and $\sum_{i=1}^{k} r_i = \frac{(1+\delta)n}{k}$) when $r_3 = r_4 = ... = r_k = 0$, $r_2 = \frac{r}{3k\cdot(k-1)}$ and $r_1 = \frac{(1+\delta)n}{k} - \frac{r}{3k\cdot(k-1)}$. The minimum is then

$$\left(\frac{(1+\delta)n}{k} - \frac{r}{3k\cdot(k-1)}\right)\cdot\frac{r}{3k\cdot(k-1)},$$

as promised. ■

Proof.(Lemma 5.7) Let ψ be a k-coloring coloring at distance r from φ. Let i_0 be the index promised in Lemma 5.8 (the one indexing U_i). Let $r_i = |U_{i_0} \cap V_i|$, and let $f(r) = \sum_{1\le i<j\le k} r_i\cdot r_j$. The conditions of Lemma 5.9 hold due to Lemma 5.8 and $\sum_{i=1}^{k} r_i = \sum_{i=1}^{k}|U_{i_0} \cap V_i| = |U_{i_0}| = \frac{(1+\delta)n}{k}$ (for some $|\delta| \le 0.01$). Lemma 5.9 then implies that

$$f(r) \ge \left(\frac{(1+\delta)n}{k} - \frac{r}{3k\cdot(k-1)}\right)\cdot\frac{r}{3k\cdot(k-1)} \ge \frac{n}{2k}\cdot\frac{r}{3k^2}.$$

The last inequality is due to $r \leq n$ and $|\delta| \leq 0.01$. Further observe that $f(r)$ counts exactly the number of edges in $\{U_{i_0} \cap V_i\} \times \{U_{i_0} \cap V_j\}$ for $i \neq j$, which are all proper under φ but not under ψ. Set $e = \sum_{i<j} |V_i||V_j|$, and observe that $e \geq \binom{k}{2} \left(\frac{0.99n}{k}\right)^2$. Therefore,

$$c_r \leq \binom{e - f(r)}{m} \cdot \binom{e}{m}^{-1} \leq \binom{e - \frac{nr}{6k^9}}{m} \cdot \binom{e}{m}^{-1} \leq e^{-0.99^2 mr/(6k^9 n)} \leq e^{-mr/(10k^9 n)}. \qquad (2)$$

The third inequality is due to:

$$\frac{\binom{a-x}{b}}{\binom{a}{b}} \leq \left(1 - \frac{b}{a}\right)^x \leq e^{-bx/a}.$$

■

6 Complete Proofs for Sections 2 and 3

6.1 Proof of Proposition 2.1 (Discrepancy)

The discrepancy property for random graphs was proven in several occasions. We follow the proof given in [19] (Section 2.2.5 in that paper) for $\mathcal{G}_{n,p}$. We do not give the complete details, just point out how to adjust that proof to fit $\mathcal{G}^{\text{uniform}}_{n,m,k}$. For the sake of clarity of presentation we consider the case where the graph has a proper k-coloring where all color classes are of size n/k. The case where all color classes are nearly balanced is treated very similarly, as in Section 5 (the case where some color class is "far" from being balanced can be disregarded, Proposition 3.4).

The proof branches according to the sizes of the sets A and B. For "big" sets we prove that the first property holds, and for "small" sets – we prove that the second one holds. Throughout the discussion we assume p satisfies $m = \binom{k}{2} \left(\frac{n}{k}\right)^2 p$. Fix two sets of vertices A and B, and first consider the case $|B| \geq n/e$. Observe that $e(A, B) \leq |A| \cdot 10np$ by the bounded-degree property of G'. Therefore,

$$e(A, B) \leq |A| \cdot 10np = (|A||B|p) \cdot (10n/|B|) \leq 30|A||B|p.$$

Thus, the first property holds. Now consider the case $|B| \leq n/e$. The proof in [19] uses some variant of the Chernoff bound to bound the number of edges between A and B. Since the edges in the uniform setting are *not* independent, one needs to reprove the Chernoff bound, or a some variant thereof, in $\mathcal{G}^{\text{uniform}}_{n,m,k}$ (for the case where the random variables are edge indicators). This will be our goal in the next few paragraphs. The crucial step in the proof of the Chernoff bound is restating the expectation of a product of r.v. (random variables) as the product of their expectations (which is possible in the original proof due to independence, but in our setting this is not the case as the edges are not chosen independently of each other). Lemmas 6.1 and 6.3 establish this fact in our setting.

Lemma 6.1. *Let $X_1, X_2, \ldots, X_d$ be d non-negative random variables taking values in Ω, $|\Omega| < \infty$. Then the following holds:*

$$E[X_1 \cdot X_2 \cdots X_d] \leq \max_{i_1, i_2, \ldots, i_{d-1} \in \Omega} E[X_1] \cdot E[X_2 | X_1 = i_1] \cdots E[X_d | X_1 = i_1, \ldots, X_{d-1} = i_{d-1}].$$

Proof. The proof is by induction on d – the number of random variables. The case $d = 1$ is immediate. Now to prove the induction step,

$$E[X_1 \cdot X_2 \cdots X_d] = \sum_{i_1 \in \Omega} Pr[X_1 = i_1] \cdot E[X_1 \cdot X_2 \cdots X_d | X_1 = i_1] = \sum_{i_1 \in \Omega} i_1 E[X_2 \cdots X_d | X_1 = i_1] \cdot Pr[X_1 = i_1]$$

$$\underbrace{\leq}_{\text{induction hyp.}} \sum_{i_1 \in \Omega} i_1 Pr[X_1 = i_1] \cdot \max_{i_2,\ldots,i_{d-1} \in \Omega} E[X_2 | X_1 = i_1] \cdots E[X_d | X_1 = i_1, \ldots, X_{d-1} = i_{d-1}]$$

$$= \left(\max_{i_2,\ldots,i_{d-1} \in \Omega} E[X_2 | X_1 = i_1] \cdots E[X_d | X_1 = i_1, \ldots, X_{d-1} = i_{d-1}] \right) \cdot \sum_{i_1 \in \Omega} i_1 Pr[X_1 = i_1]$$

$$= E[X_1] \cdot \left(\max_{i_2,\ldots,i_{d-1} \in \Omega} E[X_2 | X_1 = i_1] \cdot E[X_3 | X_1 = i_1, X_2 = i_2] \cdots E[X_d | X_1 = i_1, \ldots, X_{d-1} = i_{d-1}] \right)$$

$$= \max_{i_1, i_2,\ldots,i_{d-1} \in \Omega} E[X_1] \cdot E[X_2 | X_1 = i_1] \cdot E[X_3 | X_1 = i_1, X_2 = i_2] \cdots E[X_d | X_1 = i_1, \ldots, X_{d-1} = i_{d-1}].$$

We apply the induction hypothesis to random variables Y_i of the form $X_i | (X_1 = i_1)$, $i \geq 2$, and notice that $Y_3 | (Y_2 = i_2)$ is simply $X_3 | (X_1 = i_1, X_2 = i_2)$. ■

Let X_e be an indicator random variable which is 1 iff the edge $e = (i, j)$ is present in G'. We let $\hat{X}_e = e^{tX_e}$, where t is some fixed positive number. Observe that $\hat{X}_e$ can take two possible values, e^t or 1. The next lemma quantifies in some useful sense the dependency between the edges. We defer its proof to the end of this section.

Lemma 6.2. *Let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$, $m \geq C_0 k^1 0n$, C_0 a sufficiently large constant. Let p be s.t. $m = \binom{k}{2} \left(\frac{n}{k}\right)^2 p$. Let $X_{e_1}, \ldots, X_{e_d}$ be d edge-indicator random variables. Let $b_1, \ldots, b_{d-1}$ take arbitrary values in $\{1, e^t\}$. Then*

$$Pr[\hat{X}_{e_1} = e^t | \hat{X}_{e_1} = b_1, \ldots, \hat{X}_{e_{d-1}} = b_{d-1}] \leq 2p.$$

The next lemma shows how to move from expectation of product to product of expectations.

Lemma 6.3. *Let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$, $m \geq C_0 k^1 0n$, C_0 a sufficiently large constant. Let p be s.t. $m = \binom{k}{2} \left(\frac{n}{k}\right)^2 p$. Let $X_{e_1}, \ldots, X_{e_d}$ be d edge-indicator random variables. Let $\hat{X}_{e_j} = e^{tX_{e_j}}$, let $\mu = p \cdot d$. Then*

$$E[\hat{X}_{e_1} \cdots \hat{X}_{e_d}] \leq \exp\{2\mu(e^t - 1)\}.$$

Proof. By Lemma 6.1,

$$E[\hat{X}_{e_1} \cdots \hat{X}_{e_d}] \leq \max_{b_1,\ldots,b_{d-1} \in \{1, e^t\}} E[\hat{X}_{e_1}] \cdot E[\hat{X}_{e_2} | \hat{X}_{i_1} = b_1] \cdots E[\hat{X}_{e_d} | \hat{X}_{e_1} = b_1, \ldots, \hat{X}_{e_{d-1}} = b_{d-1}]$$

Therefore,

$$E[\hat{X}_{i_j} = e^t | \hat{X}_{i_1} = b_1, \ldots, \hat{X}_{i_{j-1}} = b_{j-1}] \leq 2pe^t + (1 - 2p) = 1 + 2p(e^t - 1) \leq \exp\{2p(e^t - 1)\}.$$

The last inequality is due to $1 + x < e^x$ (Taylor of e^x around 0). Finally,

$$E[\hat{X}_{i_1} \cdots \hat{X}_{i_d}] \leq \max_{b_1,\ldots,b_{d-1} \in \{1, e^t\}} E[\hat{X}_{i_1}] \cdot E[\hat{X}_{i_2} | \hat{X}_{i_1} = b_1] \cdots E[\hat{X}_{i_d} | \hat{X}_{i_1} = b_1, \ldots, \hat{X}_{i_{d-1}} = b_{d-1}]$$

$$\leq \prod_{j=1,\ldots,d} \exp\{2p(e^t - 1)\} = \exp\{2p \cdot d(e^t - 1)\} = \exp\{2\mu(e^t - 1)\}.$$

$\blacksquare$

Let $\mu^* = E[\sum_{j=1..d} X_{i_j}]$ be the expected number of edges under the uniform distribution. Lastly, we make the following observation, which allows us to establish a connection between μ^* and $\mu = p \cdot d$ (which is later used in the proof of the Chernoff bound in the uniform setting).

Lemma 6.4. *Let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}, m \geq C_0 k^1 0n, C_0$ a sufficiently large constant. Let p^* be the probability that an edge (i,j) is present in G, then*

$$p^* = m\binom{n}{2}^{-1}$$

Proof. First observe that indeed p^* is well defined, namely it is the same for every edge $e = (i,j)$. This is of course true due to symmetry. Let X_w be an edge-indicator variable as before. Observe that:

$$m = \sum_{i<j, e=(i,j)} X_e \Rightarrow m = E[m] = E[\sum_{i<j, e=(i,j)} X_e] = \sum_{i<j, e=(i,j)} E[X_e] = \binom{n}{2} p^*,$$

which implies the lemma. $\blacksquare$

An immediate corollary of this lemma is that $\mu^* = p^* d \geq pd/2 = \mu/2$, and that $\mu^* \leq \mu$. This is by the choice of p, and the fact that $k \geq 3$. Now we are ready to re-prove (the variant of) the Chernoff bound in the uniform setting (we defer the proof of Lemma 6.2 to the end of this discussion).

Proposition 6.5. *Let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}, m \geq C_0 k^1 0n, C_0$ a sufficiently large constant. Let p be s.t. $m = \binom{k}{2}\left(\frac{n}{k}\right)^2 p$. Let $X_{e_1}, \ldots, X_{e_d}$ be d edge-indicator random variables, and $X = \sum_{1,\ldots,d} X_{e_i}$. Let $\mu = pd$ as before. Then*

$$Pr[X > (1+\lambda)\mu] \leq \left(\frac{e^{2\lambda}}{(1+\lambda)^{(1+\lambda)/2}}\right)^{\mu}.$$

Corollary 6.6. *For $r \geq 200$, the above inequality reads*

$$Pr[X > r\mu] \leq e^{-\mu(r \ln r)/30}.$$

Proof.(Corollary)

$$\frac{e^{2\lambda}}{(1+\lambda)^{(1+\lambda)/2}} = \exp\{2\lambda - (1/2)(1+\lambda)\ln(1+\lambda)\} \leq \exp\{\lambda(2 - \ln\lambda/2)\} \leq \exp\{-(\lambda \ln \lambda)/20\}$$

The last inequality is true for $\lambda = 199$ for example. Now set $r = 1 + \lambda$. $\blacksquare$

Proof.(Proposition 6.5) For any $t > 0$,

$$Pr[X > (1+\lambda)\mu] \leq Pr[X > (1+\lambda)\mu^*] = Pr[\exp\{tX\} > \exp\{t(1+\lambda)\mu^*\}] \leq \frac{E[\exp\{tX\}]}{\exp\{t(1+\lambda)\mu^*\}}$$
$$\leq \frac{\exp\{2\mu(e^t - 1)\}}{\exp\{t(1+\lambda)\mu^*\}} \leq \frac{\exp\{2\mu(e^t - 1)\}}{\exp\{t(1+\lambda)\mu/2\}}$$

This is true for any t, and in particular for $t = \ln(1+\lambda)$, which gives the desired result. The first inequality is true since $\mu^* \leq \mu$. The second inequality is the Markov inequality. The third inequality is by noticing that $E[\exp\{tX\}]$ is exactly $E[\hat{X}_{e_1} \cdots \hat{X}_{e_d}]$, and using Lemma 6.3. The last inequality is by the fact that $\mu^* \geq \mu/2$.

Comparing our proof with the one in [19]. The case $|B| \geq n/e$ we already proved. The case $|B| \leq n/e$ uses the formulation of the Chernoff bound asserted in Corollary 6.6, though with the constant 3 in the exponent of the right hand side rather than 30. This only affects the constant c in Proposition 2.1. However one is only required to show that such a constant exists, and in this work we make no attempt to optimize any of the constants.

Proof.(Lemma 6.2) Recall that we need to bound $Pr[\hat{X}_{i_j} = e^t | \hat{X}_{i_1} = b_1, \ldots, \hat{X}_{i_{j-1}} = b_{j-1}]$. The fact that $\hat{X}_{i_1} = b_1, \ldots, \hat{X}_{i_{j-1}} = b_{j-1}$ basically implies some constellation of the edges $i_1, \ldots, i_{j-1}$, according to the b-values (if $b_{i_j} = 1$ then the edge i_j was not included since $X_{i_j} = 0$). Consider this constellation of edges, and let s be the number of edges that are present. If $s > m - 1$, then such a graph cannot be sampled, and therefore $Pr[\hat{X}_{i_j} = e^t | \hat{X}_{i_1} = b_1, \ldots, \hat{X}_{i_{j-1}} = b_{j-1}] = 0 \leq 2p$. Thus we are left with the case $s \leq m - 1$.

Let $e = i_j$, a graph G is said to be *e-bad* if it contains e. Furthermore, let $\mathcal{P}_e$ signify the set of all e-bad (balancedly) k-colorable graphs with exactly m edges that also contain the constellation implied by the b values at hand. In addition, denote by $\mathcal{G}$ the set of all (balancedly) k-colorable graphs with exactly m edges that contains this constellation as well. Our objective is to establish the following.

$$\mathcal{P}_e \leq (2p)|\mathcal{G}|. \tag{3}$$

Observe that this immediately implies that the probability of an e-bad graph in $\mathcal{G}_{n,m,k}^{\text{uniform}}$ given the above constellation is at most $2p$. To prove Equation (3) we shall set up a bipartite auxiliary graph $\mathcal{A}$ with vertex set $V(\mathcal{A}) = \mathcal{P}_e \cup \mathcal{G}$. This graph will have the property that the average degree of vertices in $\mathcal{P}_e$ is Δ, while for $\mathcal{G}$ the average degree is Δ', where $\Delta'/\Delta \leq 2p = 2m/E$, where $E = \binom{k}{2}\left(\frac{n}{k}\right)^2$. Since $\Delta|\mathcal{P}_e| = \Delta'|\mathcal{G}|$, by double counting, we thus obtain Equation (3). We describe a procedure that receives a graph $G \in \mathcal{P}_e$ and produces a new graph $G' \in \mathcal{G}$. In our auxiliary graph $\mathcal{A}$, we connect a right-side node G with a left-side one G', if G' can be obtained from G by this procedure. The procedure is the following simple one. Given an e-bad graph G, remove the edge e, and place it instead of a non-edge of G, while respecting at least one balanced proper k-coloring of the graph. The number of possible graphs G' that can be obtained via the above procedure is at least $E - m - s \geq E/2$, thus $\Delta \geq E/2$. This is because we have to choose a place for the displaced edge amongst all possible ones. Conversely, consider the following procedure to recover a graph G from G'. Out of m possible edges, choose one. Remove it and place it between i and j. Therefore $\Delta' \leq m$ (there at most m possibilities to guess that edge). ∎

This concludes the proof of the discrepancy property. ∎

6.2 Proof of Corollary 2.2 (Discrepancy)

Assume that Proposition 2.1 holds with $c \leq 30$ (which is the case *whp*, as implied by the proof of Proposition 2.1), and suppose in contradiction that there exists a subgraph H (on h vertices) of G violating the condition of the corollary. Then for such a graph H, $e(H,H) > h\delta m/(2nk)$. However,

$$c\mu(H,H) = ch^2 p \leq \frac{\delta n}{1000k} cph = \frac{\delta n}{1000k} \cdot \frac{2mk}{n^2(k-1)} \cdot ch < h\delta m/(2nk),$$

contradicting the first condition of Proposition 2.1, and also

$$e(H,H)\ln\frac{e(H,H)}{\mu(H,H)} \geq \frac{\delta hm}{2nk}\ln\left(\frac{\delta hm}{2nk}\cdot\frac{n^2}{2mh^2}\right) \geq \frac{\delta hC_0k^9}{2}\ln\left(\frac{n}{h}\cdot\frac{4k}{\delta}\right) = \frac{\delta hC_0k^9}{2}\left(\ln\frac{n}{h}-\ln\frac{4k}{\delta}\right)$$
$$= hc\ln\frac{n}{h}+h\left(\frac{\delta C_0k^9}{2}-c\right)\ln\frac{n}{h}-\frac{\delta hC_0k^9}{2}\ln\frac{\delta}{4k} \geq$$
$$hc\ln\frac{n}{h}+h\ln(1000k/\delta)^{\frac{\delta C_0k^9}{2}-c}-h\ln(4k/\delta)^{\frac{\delta C_0k^9}{2}} = hc\ln\frac{n}{h}+\ln\left(\frac{1000k\delta}{4k\delta}\right)^{\frac{\delta C_0k^9}{2}}\cdot\left(\frac{\delta}{1000k}\right)^c =$$
$$hc\ln\frac{n}{h}+h\ln\frac{250^{\frac{\delta C_0k^9}{2}}}{(1000k\delta^{-1})^c} > hc\ln\frac{n}{h},$$

Contradicting the second condition of Proposition 2.1. The last (strict) inequality holds since $\frac{250^{\frac{\delta C_0k^9}{2}}}{(\delta/(1000k))^c} > 1$ for a sufficiently large constant C_0.

6.3 Proof of Propositions 2.3 and 4.2 (Spectral Analysis)

We start by analyzing the procedure SepctralApprox – that is proving Proposition 4.2. We assume that Proposition 2.3 holds, which is the case *whp*, and using this fact we show that $\hat{A}$, the rank-k approximation of $A(G')$ (see Section 4.1) approximates $\mathcal{E}$ in some useful sense. Of course, we know the adjacency matrix $A(G')$. Furthermore, we know that $\|M'\| = \|\mathcal{E}-A(G')\|$ is "small" (Proposition 2.3). That is, $A(G')$ is a good approximation of $\mathcal{E}$ *in the operator norm.* However, we can't exploit this fact directly in order to obtain a good entry-wise approximation of $\mathcal{E}$. Indeed, instead of getting a matrix that approximates $\mathcal{E}$ in the operator norm, an approximation B of $\mathcal{E}$ in the *Frobenius norm*

$$\|\mathcal{E}-B\|_F = \sqrt{\sum_{v,w\in V'}(B_{vw}-\mathcal{E}_{vw})^2}$$

would be more useful.

The analysis of SpectralApprox is based on the following lemma, which shows that for *most* vertices v the v-column $\hat{A}_v$ of $\hat{A}$ is close to the v-column $\mathcal{E}_v$ of $\mathcal{E}$.

Lemma 6.7. *Let $Z = \{v \in V' : \|\hat{A}_v - \mathcal{E}_v\|^2 \geq 10^{-10}np^2/k\}$. Then $|Z| \leq nd^{-0.1}$, where $d = d_{avg}/k$, $d_{avg} = 2m/n$.*

Proof.

$$\begin{aligned}\sum_{v\in V'}\|\mathcal{E}_v-\hat{A}_v\|^2 &= \|\mathcal{E}-\hat{A}\|_F^2 \leq 2k\|\mathcal{E}-\hat{A}\|^2 \leq 2k(\|\mathcal{E}-A(G')\|^2+\|\hat{A}(G')-\hat{A}\|^2) \leq 4k\|\mathcal{E}-A(G')\|^2 \\ &= 4k\|M'\|^2 \leq 4kd^{1.8} \leq d^{1.81}.\end{aligned}$$

The first inequality is by the fact that for a matrix B of rank q it holds that $\|B\|^2 \leq q\|B\|_F^2$, and the fact that both $\mathcal{E}$, $\hat{A}$ have rank k and therefore $\mathcal{E}-\hat{A}$ has rank at most $2k$. The second iequality is just the triangle inequalty, and the third inequality is by the fact that $\|\hat{A}(G')-\hat{A}\| \leq \|\mathcal{E}-A(G')|$ because $\hat{A}$ is a rank-k approximation of $A(G')$, and therefore minimizes $\|\hat{A}(G')-B\|$ over all matrices B of rank k. The next-to-last inequality is due to Proposition 2.3.

Finally we derive that $|Z| \cdot 10^{-10}np^2/k \leq d^{1.81}$, so that $|Z| \leq 10^{10}\frac{d^{1.81}k}{np^2} \leq nd^{-0.1}$. Here we assume that the coloring $V_1, \ldots, V_k$ is nearly balanced. That is, for every i, $|V_i - n/k| \leq 0.01n/k$, and therefore $p = \Theta(m/n^2)$. ■

Lemma 6.7 implies that for *most* vertices v, w belonging to the same color class V_i the difference $\|\hat{A}_v - \hat{A}_w\|$ is small, whereas for most $u \in V_j$, $j \neq i$, the distance $\|\hat{A}_v - \hat{A}_u\|$ is large. This implies that the classes $X_1, \ldots, X_k$ provide a good approximation of the coloring $V_1, \ldots, V_k$ (up to a permutation of the indices, of course).

Proposition 6.8. *There is a permutation σ of $\{1, \ldots, k\}$ such that $X_i \triangle V_i' \leq 10^{-9}n/k^2$.*

Proof. We show by induction on i that in each step there is a vertex v_i such that $|S_{v_i}| \setminus X \geq (1 - 10^{-10})\frac{n}{k}$. Moreover, we shall prove that for the vertex v_i chosen by the algorithm there is a class $V'_{\sigma(i)}$ such that $X_i \setminus V'_{\sigma(i)} \subset Z$. Let $1 \leq i \leq k$, and suppose that these statemtents are true for all $1 \leq i' < i$.

Let $j \in \{1, \ldots, k\} \setminus \{\sigma(1), \ldots, \sigma(i-1)\}$. Then by Lemma 6.7 there is a vertex $v^* \in V_j' \setminus Z$. Moreover, since all $u \in V_j' \setminus Z$ we have

$$\|\hat{A}_{v^*} - \hat{A}_u\|^2 \leq 2(\|\hat{A}_{v^*} - \mathcal{E}_{v^*}\|^2 + \|\mathcal{E}_u - \hat{A}_u\|^2) \leq 0.01\frac{np^2}{k}.$$

Hence, $S_{v^*} \supset V_j' \setminus Z$. Furthermore, $V_j' \cap X \subset Z$ by the induction hypothesis. Therefore, $|S_{v^*}| \setminus X \geq |V_j'| \setminus Z \geq (1 - 10^{-10})\frac{n}{k}$. Thus, it is possible for the algorithm to choose a vertex v_i such that $|S_{v_i}| \setminus X \geq (1 - 10^{-10})\frac{n}{k}$.

Now, let v_i be the vertex with this property chosen by the algorithm, and pick some $w \in S_{v_i} \setminus (X \cup Z)$; such a vertex w exists due to the upper bound on $|Z|$ from Lemma 6.7. Then we have

$$\|\hat{A}_{v_i} - \mathcal{E}_w\|^2 \quad \leq \quad \|\hat{A}_{v_i} - \hat{A}_w\|^2 + 2\|\hat{A}_{v_i} + \hat{A}_w\| \cdot \|\mathcal{E}_w - \hat{A}_w\| + \|\mathcal{E}_w - \hat{A}_w\|^2 \leq \frac{0.02np^2}{k}.$$

Further, we have $w \notin \bigcup_{1 \leq j < i} V'_{\sigma(j)}$. For assume that $w \in V'_{\sigma(j)}$ for some $1 \leq j < i$. Then for all $u \in S_{v_i} \setminus V'_{\sigma(j)}$ we have

$$\|\hat{A}_u - \mathcal{E}_w\|^2 \leq (\|\hat{A}_v - \mathcal{E}_w\| + \|\hat{A}_u - \hat{A}_v\|)^2 \leq \frac{0.1np^2}{k}. \tag{4}$$

However, since u, w belong to different color classes, we have $\|\mathcal{E}_u - \mathcal{E}_w\|^2 \geq np^2/k$. Thus, (4) entails that $\|\hat{A}_u - \mathcal{E}_u\|^2 \geq \frac{0.1np^2}{k}$, whence $u \in Z$. Consequently, if $w \in V'_{\sigma(j)}$ for some $1 \leq j < i$, then $S_{v_i} \setminus V'_{\sigma(j)} \subset Z$. As by induction $|V'_{\sigma(j)} \setminus X_j| \leq 0.1\frac{n}{k}$ and $|S_{v_i}| \geq 0.6\frac{n}{k}$, this implies that $|Z| \geq \frac{n}{2k}$, which contradicts Lemma 6.7.

Hence, we have estabished that $w \notin \bigcup_{1 \leq j < i} V'_{\sigma(j)}$, and we let $\sigma(i)$ be such that $w \in V'_{\sigma(i)}$.

Finally, we claim that $S_{v_i} \setminus V'_{\sigma(i)} \subset Z$. For let $u \in S_{v_i} \setminus V'_{\sigma(i)}$. Then $\|\hat{A}_u - \mathcal{E}_w\|^2 \leq \frac{0.1np^2}{k}$ (cf. (4)). Hence, as $\|\mathcal{E}_u - \mathcal{E}_w\|^2 \geq np^2/k$, we conclude that $\|\hat{A}_u - \mathcal{E}_u\|^2 \geq 0.1np^2/k$. Thus, $u \in Z$. ■

This completes the proof of Proposition 4.2. We now proceed with the proof of Proposition 2.3. The proof of Proposition 2.3 is based on a proper modification of techniques developed by Kahn and Szemerédi in [20], where the authors show that the second largest eigenvalue in absolute value of a random d-regular graph is almost surely $O(\sqrt{d})$. Since in our case the graph is not regular, and the edges are not chosen independently, a few modifications are needed.

In what follows, we let $V_1, \ldots, V_k$ be a partition of $V = \{1, \ldots, n\}$ such that $|V_i - \frac{n}{k}| < 0.1\frac{n}{k}$ for all $1 \le i \le k$. Moreover, let $0 < p < 1$ be such that $\sum_{1 \le i < j \le k} V_i V_j p = m$, and set $d = np/k$. Further, let G signify a random graph with planted coloring $V_1, \ldots, V_k$ in which each possible edge compatible with this coloring is present with probability p independently.

In order to prove Proposition 2.3, we shall first analyze the spectral properties of G. Then, we shall combine this information with Proposition 5.4 and the discrepancy property established in Proposition 2.1 in order to obtain the desired result on the spectrum of a uniformly distributed k-colorable graph.

6.3.1 Proof of Proposition 2.3 (Outline)

As the indicator vectors $\vec{1}_{V_1}, \ldots, \vec{1}_{V_k}$ corresponding to the k planted color classes of G play a distinguished role, we shall first analyze the spectral properties of G on the orthogonal complement of the space spanned by these vectors.

Lemma 6.9. *With probability $\ge 1 - \exp(-n)$ the adjacency matrix $A = (a_{vw})_{v,w \in V}$ of G satisfies the following.*

Suppose that $\xi, \eta \in \mathbf{R}^n$ are unit vectors perpendicular to $(\vec{1}_{V_i})_{1 \le i \le k}$. Let

$$L(\xi, \eta) = \left\{ (v, w) \in V \times V : |\xi_v \eta_w| \le \sqrt{p/n} \right\}. \tag{5}$$

Then $\left| \sum_{(v,w) \in L} a_{vw} \xi_v \eta_w \right| \le (np)^{3/4}$.

Furthermore, regarding the vectors $\vec{1}_{V_1}, \ldots, \vec{1}_{V_k}$, we prove the following in Section 6.3.3.

Lemma 6.10. *Let G' be the graph obtained from G by removing all vertices of degree $> 2np$. Then with probability $\ge 1 - \exp(-nd^{-10})$ the matrix $M' = \sum_{i \ne j} p J_{V_i \times V_j \cap V(G')^2} - A'$ satisfies*

$$\|M' \vec{1}_{V_i \cap V(G')}\| \le d^{0.66} \sqrt{n}.$$

Furthermore, we employ the following result, which was established by Kahn and Szemeredi [20] for regular graphs. A proof of the present setting can be found in [19].

Lemma 6.11. *Suppose that $H = (V, E)$ is a graph of maximum degree $\le 2np$ that satisfies the discrepancy property stated in Proposition 2.1. Let $A_H = (a^H_{vw})_{v,w \in V}$ be the adjacency matrix of H. Then for all unit vectors $\xi, \eta \in \mathbf{R}^n$ we have*

$$\sum_{(v,w) \in V^2 \setminus L(\xi,\eta)} a^H_{vw} |\xi_v \eta_w| \le C\sqrt{np}.$$

Proof of Proposition 2.3. Let $G_* = \mathcal{G}^{\text{uniform}}_{n,m,k}$ be a random k-colorable graph, and let G'_* be the subgraph obtained by removing all vertices of degree $> 2np$. Let $V' = V(G'_*)$, and let $A_* = (a_{*\,vw})_{v,w \in V'}$ be the adjacency matrix of G'_*. By Lemmas 5.1 and 5.2, we can infer from Lemma 6.9

that *whp* G'_* has a k-coloring $V'_1, \ldots, V'_k$ such that for all unit vectors $\xi, \eta \in \mathbf{R}^{V'}$ that are perpendicular to $\{\vec{1}_{V'_i} : 1 \le i \le k\}$ we have

$$\left| \sum_{(v,w) \in L'(\xi,\eta)} a_{*\,vw} \xi_v \eta_w \right| \le (np)^{3/4}, \tag{6}$$

where $L'(\xi, \eta) = \left\{ (v, w) \in V' \times V' : |\xi_v \eta_w| \le \sqrt{p/n} \right\}$. In addition, combining Proposition 2.1 and Lemma 6.11, we conclude that

$$\left| \sum_{(v,w) \in (V' \times V') \setminus L'(\xi,\eta)} a_{*\,vw} \xi_v \eta_w \right| \le (np)^{3/4} \tag{7}$$

for all such ξ, η. Thus, combining (6) and (7), we conclude that

$$\forall \xi, \eta \in \mathbf{R}^{V'},\ \|\xi\| = \|\eta\| = 1,\ \xi, \eta \perp \{\vec{1}_{V'_i} : 1 \le i \le k\} : \left|\langle M'\xi, \eta \rangle\right| \le 2(np)^{3/4}. \tag{8}$$

Furthermore, invoking Lemmas 5.1 and 5.2 once more, we conclude from Lemma 6.10 that *whp*

$$\forall 1 \le i \le k : \|M' \vec{1}_{V'_i}\| \le d^{0.66} \sqrt{n} \le \|\vec{1}_{V'_i}\| (np)^{3/4}. \tag{9}$$

Finally, combining (8) and (9), we obtain $\|M'\| \le 8(np)^{3/4} < d^{0.9}$, as claimed. $\square$

6.3.2 Proof of Lemma 6.9

Alon and Kahale [6] established the following estimate.

Lemma 6.12. *Let $1 \le i < j \le k$. Then with probability $\ge 1 - \exp(-2n)$ the adjacency matrix $A = (a_{vw})_{v,w \in V}$ of G satisfies the following.*

Suppose that $\xi \in \mathbf{R}^{V_i}$, $\eta \in \mathbf{R}^{V_j}$ are unit vectors such that $\xi \perp \vec{1}_{V_i}$, $\eta \perp \vec{1}_{V_j}$. Let

$$L_{ij} = \left\{ (v, w) \in V_i \times V_j : |\xi_v \eta_w| \le \sqrt{p/n} \right\}.$$

Then $\left| \sum_{(v,w) \in L_{ij}} a_{vw} \xi_v \eta_w \right| \le c\sqrt{np}$ for a certain constant $c > 0$.

To prove Lemma 6.9, we shall just apply the bound provided by Lemma 6.12 to each pair $1 \le i, j \le k$, $i \ne j$. Thus, let $\xi, \eta \in \mathbf{R}^V$ be such that $\xi, \eta \perp \vec{1}_{V_i}$ for all $1 \le i \le k$. Then Lemma 6.12 entails that with probability $\ge 1 - k^2 \exp(-2n) \ge 1 - \exp(-n)$ we have

$$\left| \sum_{(v,w) \in L} a_{vw} \xi_v \eta_w \right| \quad \le \quad \sum_{1 \le i,j \le k,\, i \ne j} \left| \sum_{(v,w) \in L_{ij}} a_{vw} \xi_v \eta_w \right| \le ck^2 \sqrt{np} \le (np)^{3/4}.$$

6.3.3 Proof of Lemma 6.10

The proof is based on the following Chernoff bound.

Theorem 6.13. *Suppose that X is a binomially distributed random variable with mean μ. Let $\varphi(x) = (1+x)\ln(1+x) - x$. Then*

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

$$P(X \leq \mu - t) \quad \leq \quad \exp\left(-\mu\varphi\left(\frac{-t}{\mu}\right)\right) \leq \exp\left(-\frac{t^2}{2\mu}\right) \quad (0 < t < \mu). \tag{11}$$

The Chernoff bound entails the following result on the degree distribution of G.

Lemma 6.14. *Let $W_{ij} = \{v \in V_i : |e_G(v, V_j) - |V_j|p| > d^{0.51}\}$, where $1 \leq i, j \leq k$ and $i \neq j$. Then $P\left[\exists i, j : |W_{ij}| > nd^{-10}\right] \leq \exp(-nd^{-10})$.*

Proof. Since $\mathrm{E}(e_G(v, V_j)) = |V_j|p$, Theorem 6.13 entails that for any $i \neq j$ and any $v \in V_i$ we have $P\left[|e_G(v, V_j) - |V_j|p| > d^{0.51}\right] \leq n\exp(-d^{\Omega(1)}) \leq d^{-100}$. Therefore, $E(|W_{ij}|) \leq nd^{-100}$. Furhermore, the random variables $(e_G(v, V_j))_{v \in V_i}$ are mutually independent, and thus $|W_{ij}|$ is binomially distributed. Hence, invoking (10) once more, we conclude that $P\left[|W_{ij}| > nd^{-10}\right] \leq \exp(-nd^{-10}\ln d)$. Finally, the union bound entails that with probability $\geq 1 - k^2\exp(-nd^{-10}\ln d) \geq 1 - \exp(-nd^{-10})$ the bound $|W_{ij}| \leq nd^{-10}$ holds for all i, j simultaneously. ■

Corollary 6.15. *With probability $\geq 1 - \exp(-nd^{-10})$ the random graph G has at most nd^{-9} vertices of degree $> 2np$.*

Proof. Any vertex of degree $> 2np$ belongs to $\bigcup_{i \neq j} W_{ij}$, and by Lemma 6.14 with probability $\geq 1 - \exp(-nd^{-10})$ this set has cardinality $\leq k^2 nd^{-10} \leq nd^{-9}$. ■

Lemma 6.16. *With probability $\geq \exp(-nd^{-10})$ the random graph G does not feature two disjoint sets $S, T \subset V$, $|S| \leq nd^{-9} \leq |T|$, such that every vertex in T has at least 100 neighbors in S.*

Proof. Let $s \leq nd^{-9} \leq t$. Since each of the possible $\binom{n}{2}$ possible edges occurs in G with probability $\leq p$ independently, for any set S of size s and any $T \subset V \setminus S$ of size t the probability that all $v \in T$ have 100 neighbors in S is at most $\left[\binom{s}{100}p^{100}\right]^t \leq (sp)^{100t}$. Moreover, there are $\binom{n}{s}$ ways to choose S, and then at most $\binom{n}{t}$ ways to choose T. Hence, the probability $P_{s,t}$ that there exists sets S, T of sizes s resp. t such that $e_G(v, S) \geq 100$ for all $v \in T$ is at most

$$P_{s,t} \quad \leq \quad \binom{n}{s}\binom{n}{t}(sp)^{100t} \leq \exp(-t).$$

Furthermore, as there are at most n^2 ways to choose s and t, we conclude that the probability of the event stated in the lemma is at most $n^2\exp(-t) \leq \exp(-nd^{-10})$. ■

Combining Corollary 6.15 with Lemma 6.16, we obtain the following.

Corollary 6.17. *With probability $\geq 1 - \exp(-nd^{-10})$ the random graph G has at most nd^{-9} vertices v of degree $\leq 2np$ that have at least 100 neighbors of degree $> 2np$.*

Proof of Lemma 6.10. Let G' be the subgraph of G obtained by removing all vertices of edgee $> 2np$. Moreover, let $V' = V(G')$ and $V_i' = V_i \cap V'$. Let $1 \le i \le k$ and set $\eta = M'\vec{1}_{V_i'}$. Then $\eta_v = 0$ for all $v \in V_i'$, and $\eta_v = |V_i'|p - e_G(v, V_i')$ for all $v \in V' \setminus V_i$. Hence,

$$\begin{aligned} \|\eta\|^2 &= \sum_{j \ne i} \sum_{v \in V_j'} (|V_i|p - e_G(v, V_i'))^2 \\ &\le 2\sum_{j \ne i} \sum_{v \in V_j'} (|V_i|p - e_G(v, V_i))^2 + 2\sum_{v \in V'} e_G(v, V \setminus V')^2. \end{aligned} \tag{12}$$

Due to Lemma 6.14, the first sum on the r.h.s. can be estimated as follows:

$$\begin{aligned} \sum_{j \ne i} \sum_{v \in V_j'} (|V_i|p - e_G(v, V_i))^2 &\le d^{1.2}n + 4(np)^2 \sum_{j \ne i} |W_{ji}| \\ &\le d^{1.2}n + 4d^{-10}kn(np)^2 \le 2d^{1.2}n, \end{aligned} \tag{13}$$

because all vertices in V' have degree $\le 2np$. Furthermore, as by Corollary 6.17 there are at most nd^{-9} vertices $v \in V'$ that have > 100 neighbors in $V \setminus V'$, and since all $v \in V'$ satisfy $e_G(v, V \setminus V') \le 2np$, we have

$$\sum_{v \in V'} e_G(v, V \setminus V')^2 \le 10^4 n + 4d^{-9}n(np)^2 \le 10^5 n. \tag{14}$$

Finally, plugging (12) and (13) into (14), we obtain the assertion. □

6.4 Proof of Proposition 3.4 (Balancedness)

Let $\mathcal{M}_{n,m,k}$ be the set of k-colorable graphs on n vertices with m edges. Let $\mathcal{B}_{n,m,k} \subseteq \mathcal{M}_{n,m,k}$ be the set of balancedly k-colorable graphs, and let $\mathcal{N}_{n,m,k} \subseteq \mathcal{M}_{n,m,k}$ be the set of non-0.01-balanced graphs. It suffices to prove that $|\mathcal{N}_{n,m,k}|/|\mathcal{B}_{n,m,k}| = o(1)$. Let φ_0 be a fixed balanced k-coloring, and Φ be the set of all non-0.01-balanced-colorings. For a k-coloring φ (not necessarily balanced) with color classes $V_1, V_2, ..., V_k$, we let $D(\varphi)$ be the number of k-colorable graphs in $\mathcal{M}_{n,m,k}$ which are properly colored by φ. Then

$$D(\varphi) = \binom{\sum_{1 \le i < j \le k} |V_i| \cdot |V_j|}{m}.$$

In the balanced case we get

$$D(\varphi_0) = \binom{\binom{k}{2}\left(\frac{n}{k}\right)^2}{m}.$$

Therefore,

$$|\mathcal{B}_{n,m,k}| \ge D(\varphi_0).$$

Now consider a non-0.01-balanced-coloring φ. φ must have at least one color class whose size is $< 0.99n/k$ or at least one color class whose size is $> 1.01n/k$. Standard calculations show (convexity arguments) that $D(\varphi)$ is maximized when one color class is of size $0.99n/k$ and all the other $(k-1)$ classes are of size $\left(1 + \frac{0.01}{k-1}\right) n/k$ (namely, the $0.01n/k$ vertices lacking in that color class are evenly spread amongst the other color classes), and symmetrically for the case that one color class is of size

$> 1.01n/k$. Therefore,

$$
\begin{aligned}
|\mathcal{N}_{n,m,k}| \leq & |\Phi| \cdot \binom{\binom{k-1}{2}\left(1+\frac{0.01}{k-1}\right)^2\left(\frac{n}{k}\right)^2 + (k-1)0.99\left(1+\frac{0.01}{k-1}\right)\left(\frac{n}{k}\right)^2}{m} + \\
& |\Phi| \cdot \binom{\binom{k-1}{2}\left(1-\frac{0.01}{k-1}\right)^2\left(\frac{n}{k}\right)^2 + (k-1)1.01\left(1-\frac{0.01}{k-1}\right)\left(\frac{n}{k}\right)^2}{m} \leq \\
& 2 \cdot k^n \cdot \binom{\left(1-\frac{1}{5000k^3}\right)\binom{k}{2}\left(\frac{n}{k}\right)^2}{m}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\frac{|\mathcal{N}_{n,m,k}|}{|\mathcal{B}_{n,m,k}|} \leq & 2 \cdot k^n \cdot \binom{\left(1-\frac{1}{5000k^3}\right)\binom{k}{2}\left(\frac{n}{k}\right)^2}{m}\binom{\binom{k}{2}\left(\frac{n}{k}\right)^2}{m}^{-1} \leq \\
& 2 \cdot k^n \left(1 - \frac{m}{\binom{k}{2}\left(\frac{n}{k}\right)^2}\right)^{\binom{k}{2}\left(\frac{n}{k}\right)^2/(5000k^3)} \leq 2 \cdot k^n \cdot e^{-m/(5000k^3)} = o(1).
\end{aligned}
$$

The second and third inequalities are due to

$$
\frac{\binom{a-x}{b}}{\binom{a}{b}} \leq \left(1 - \frac{b}{a}\right)^x \leq e^{-bx/a}.
$$

The last equality is due to the choice of m.

6.5 Proof of Proposition 3.5 (Concentration)

To prove the proposition we employ the exchange rate technique, introduced in Section 5. The first step is to prove the analogue of Proposition 3.5 in the planted model, and show that it holds with extremely hight probability, then use Proposition 5.4. Therefore we first consider $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$ for $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ s.t. $\forall i\ |\varepsilon_i| \leq 0.01$).

Lemma 6.18. *Let $\delta \in (0,1]$ be some positive number. Let G be a random graph in $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$, $m \geq C_0 k^4 n$ $C_0 = C_0(\delta)$ a sufficiently large constant. Then with probability at most e^{-n} every k-coloring at distance $\delta n/(1000k)$ from φ leaves at least $\delta m/(1000k)^2$ monochromatic edges.*

Proof. The basic idea of the proof is to first calculate the expected number of monochromatic edges induced by a k-coloring at distance at least $\delta n/(1000k)$ from φ, and show that this number is "much" higher than $\delta m/(1000k)^2$, then show a concentration result.

Let ψ be an arbitrary k-coloring at distance $r \geq \delta n/(1000k)$ from φ. A very similar argument to Lemma 5.7 gives that the probability that a random edge is monochromatic under ψ is at least $1 - e^{-r/(100nk^9)} \geq r/(100nk^9)$ (if ψ is nearly-balanced then this is exactly the same argument – just set $m = 1$ in equation (2), if ψ is "far" from being balanced, then in particular it is "far" from φ, then this fact is used to lower bound the value $f(r)$ in (2)).

Let X_r be a random variable counting the number of monochromatic edges in G induced by ψ. Then we have:

$$
E[X_r] \geq mr/(100nk^9)
$$

Now set $\alpha = 0.9$ (that is, for $r = n/(1000k)$ it holds that $m/(1000k)^2 \leq (1-\alpha)E[X_r]$). Using for example the Chernoff bound (which is applicable since it is known that X_r is more concentrated than the corresponding quantity if the draws were made with replacement [21] – and then they would have been independent) one obtains that:

$$Pr[X_r \leq m/(1000k)^2] \leq Pr[X_r \leq (1-\alpha)mr/(100nk^9)] \leq Pr[X_r \leq (1-\alpha)E[X_r]] \leq \\ \leq e^{-\alpha^2 E[X_r]/3} \leq e^{-mr/(400nk^9)}.$$

Taking the union bound over all possible k-colorings, one obtains that the probability of a k-colorings at distance greater than $\delta n/(1000k)$ from φ leaving less than than $\delta m/(1000k)^2$ monochromatic edges is at most

$$\sum_{r=\delta n/(1000k)}^{n} \binom{n}{r} k^r e^{-mr/(400nk^9)} \leq \sum_{\beta=\delta n/(1000k)}^{n} \left(\frac{enk}{r}\right)^r e^{-mr/(400nk^9)} \leq \sum_{r=\delta n/(1000k)}^{1} \left(\frac{enk \cdot e^{-C_0k/400}}{r}\right)^r \leq \\ \sum_{r=\delta n/(1000k)}^{1} \left(3000k^2 \cdot \delta^{-1} \cdot e^{-C_0k/400}\right)^r \leq \sum_{r=\delta n/(1000k)}^{n} \left(e^{-C_0k/500}\right)^r \leq e^{-n}.$$

The last inequality is due to the fact that the last sum is a geometric series with quotient $e^{-C_0k/500}$, and the fact that we can take C_0 to be a sufficiently large constant (recall that δ is fixed w.r.t. C_0). ■

We now use Proposition 5.4 to complete the proof of Proposition 3.5. Let $\mathcal{A}$ be the bad event that the sampled graph G is not $\delta/(1000k)$-concentrated for some $\delta \in (0,1]$.

$$Pr^{uniform,m}[\mathcal{A}] \leq o(1) + n^k \cdot e^{ke^{-m/(10nk^9)}n} \cdot e^{-n} = o(1).$$

In the latter we use the fact that k is constant.

6.6 Proof of Proposition 3.6 (Core Size)

To prove this proposition we again employ the exchange rate technique. Thus we first consider $\mathcal{G}^{\text{plant}}_{n,m,k,\bar{\varepsilon}}$ for $\bar{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ s.t. $\forall i\ |\varepsilon_i| \leq 0.01$.

Lemma 6.19. *Let $\delta \in (0,1]$ be some positive number. Let G be a random graph in $\mathcal{G}^{\text{plant}}_{n,p,k,\bar{\varepsilon}}$, $m \geq C_0k^4n$, $C_0 = C_0(\delta)$ a sufficiently large constant. Then there exist a constant $g_0 = g_0(\delta) > 0$ (independent of m, n) so that for every $g \geq g_0$ with probability $(1 - e^{e^{-m/(gnk^9)}n})$ there exists a δ-core $\mathcal{H}$ w.r.t. the planted assignment. Furthermore, $|\mathcal{H}| \geq (1 - e^{-m/(a_0nk^9)})n$ and the number of edges spanned by $\mathcal{H}$ is at least $(1 - e^{-m/(a_1nk^9)})m$, where $a_0(g), a_1(g)$ are two positive monotonically increasing functions of g.*

This lemma, formulated somewhat differently, is proven in [10] for the case $k = 3$, and $\bar{\varepsilon} = 0$. The proof easily generalizes to any constant k, and $\bar{\varepsilon}$ as above. We give its outline here for the sake of completeness.

Proof.(Outline) Recall the definitions $p_i = \frac{2m}{n} \cdot \frac{1}{n-|V_i|}$ and $r = \max_i |V_i|p_i$, where V_i is the i^{th} color class of the planted k-coloring φ.

Consider the following iterative procedure for defining a δ-core w.r.t. φ. Set $H^{(0)}$ to be all vertices that have degree at least $(1-\delta/2)|V_i|p_i$ in every color class V_i of φ other than their own. Iteratively, remove a vertex v from $H^{(i)}$ if either v has less than $(1-\delta)|V_j|p_j$ neighbors in $V_j \cap H^{(i)}$ for some $j \neq \varphi(v)$, or v has more than δr neighbors in $G[V \setminus H^{(i)}]$, to receive H_{i+1}. Let t be the iteration where $H^{(t)} = H^{(t+1)}$, and set $\mathcal{H} = H^{(t)}$.

First observe that the set $\mathcal{H}$ indeed meets the requirements in Definition 3.3. It now remains to prove that the set $\mathcal{H}$ is large. The main idea of the proof is to observe that to begin with very few vertices are eliminated – the degree of a vertex v in every other color class is on average $p_i|Vi|$. Therefore using large-deviation inequalities, one can bound the number of vertices that were removed before the iterative step began. If too many vertices were removed in the iterative step then a small yet dense subgraph exists (as every vertex that is removed contributes at least $\delta|V_i|pi/2$ edges to the subgraph induced on $V \setminus \mathcal{H}$). Corollary 2.2 (which can also be stated in the context of $\mathcal{G}_{n,p,k}^{\text{plant}}$) bounds the probability of the latter occurring. It now remains to do the calculations carefully.

As for the number of edges spanned by the core. Assume that $|\mathcal{H}| \geq (1-e^{-m/(a_0nk^9)})n$. Using the Chernoff bound for example one can prove that there exits a d_0 (specifically, $d_0 = O(m/n)$) s.t. for $d \geq d_0$, $Pr[deg(v) \geq d] \leq e^{-d/100}$. Therefore, the expected number of edges spanned by the non-core vertices is at most

$$e^{-m/(a_0nk^9)}n \cdot d_0 + n\sum_{d=d_0}^{n} de^{-d/100} = e^{-m/(b_0nk^9)}n + e^{-d/200}n \leq e^{-m/(c_0nk^9)}n,$$

where b_0, c_0 are some monastically increasing functions of a_0. The first inequality uses the fact that $d_0 = O(m/n)$, and the fact that the sum is smaller that the sum of a decreasing geometric series with $q = e^{-d/150}$ (for a sufficiently large m/n). Now using large-deviation inequalities one can prove that with sufficiently high probability, this is indeed the case.

Finally, observe that the cardinalates of the color classes of φ meet the third requirement in Proposition 3.6 (that is, they are of size $(1 \pm 0.01)n/k$, by the choice of $\bar{\varepsilon}$). ■

We now use Proposition 5.4 to assert this fact in the uniform case. Let g be s.t. $e^{e^{-m/(gnk^9)}n} \cdot n^k \cdot e^{ke^{-m/(10nk^9)}n} = o(1)$. Let $\mathcal{A}$ be the event that there exists some δ so that the sampled graph G has no proper k-coloring w.r.t. which there exists a δ-core of size at least $(1-e^{-m/(a_0nk^9)})n$ that spans at least $(1-e^{-m/(a_1nk^9)})m$ edges (where a_0, a_1 are chosen according to this g).

$$Pr^{uniform,m}[\mathcal{A}] \leq o(1) + n^k \cdot e^{ke^{-m/(10nk^9)}n} \cdot e^{e^{-m/(gnk^9)}n} = o(1).$$

The last equality is by the choice of g.

6.7 Proof of Proposition 3.7 (Uniqueness of Coloring)

Let $\mathcal{H}$ be some δ-core with φ the underlying k-coloring of G, which meets the requirements of Proposition 3.6. First observe that the k-coloring w.r.t. which G is c-concentrated (in the proof of Proposition 3.5) is the same as the k-coloring w.r.t. which there exists a large core (in the proof of Proposition 3.6) – this is because the proof of both propositions uses the exchange rate technique, and in the planted setting this assignment is the planted one in both cases. Therefore we may assume that G is $(1-\delta)/(1000k)$-concentrated w.r.t. φ (Proposition 3.6 concerns $\delta \in (0,1)$, and therefore $1-\delta \in (0,1)$ as well).

Let ψ be a proper k-coloring of $G[\mathcal{H}]$ so that ψ differs from φ on $\mathcal{H}$ (if no such ψ exists then we are done). By the conditions of Proposition 3.7, $G[\mathcal{H}]$ spans at least $(1-e^{-\Theta(m/(nk^9))})m$ edges. Thus it must be that ψ differs from φ on the coloring of at most $(1-\delta)n/(1000k)$ vertices (otherwise, ψ leaves at least $(1-\delta)m/(1000k)^2 >> (1-e^{-\Theta(m/(nk^9))})m$ monochromatic edges in G – due to concentration, and in particular it does not properly k-color $G[\mathcal{H}]$).

Let $v \in \mathcal{H}$ be some vertex on whose assignment φ and ψ disagree, and w.l.o.g assume that v is colored i in φ and j in ψ. Now consider the neighbors of v in $\mathcal{H}$ which are colored j under φ. It must be that these vertices are not colored j under ψ, but rather some other color j'. Now one can consider the neighbors of a vertex in $N(v)$ which are colored by j' in φ, on which again, ψ and φ must disagree. Put differently, let U be the set of vertices in the core on which ψ and φ disagrees on. By the discussion above and the first requirement in Definition 3.3 it holds that every vertex $v \in U$ has at least

$$\min_i (1-\delta)p_i|V_i| \geq (1-\delta)\frac{2m}{n}\frac{0.99n/k}{n-0.99n/k} \geq (1-\delta)m/(nk)$$

neighbors in U (p_i was defined in Definition 3.3). By our assumption on U, $|U| \leq (1-\delta)n/(1000k)$, this however contradicts Corollary 2.2 (when plugging in $1-\delta$ in Corollary 2.2).

6.8 Proof of Proposition 3.9 (Connected Components)

Let $d = \frac{2m}{kn}$. Let us say that G is *bounded* if the following conditions hold.

B1. For all $X \subset V$ such that $|X| \leq n/d^2$ we have $e(X) \leq 10|X|$.

B2. The maximum degree of G is $\leq \ln^2 n$.

B2. If H is a subgraph of G on $|V(H)| \geq (1-d^{-10})n$ vertices, and if H has a k-coloring $V_1, \ldots, V_k$ such that $e(v, V_j) \geq 0.9d$ for all $v \in V_i$ and all $1 \leq i, j \leq k$, $i \neq j$, then H is uniquely k-colorable.

Moreover, we call G *ε-feasible* if G has an induced subgraph H with the following properties.

F1. $|V(H)| \geq (1-\varepsilon \exp(-\sqrt{d}))n$ and $|E(H)| \geq (1-d^{-1})m$.

F2. There is a k-coloring $V_1, \ldots, V_k$ of G such that $|H| \cap V_i \geq (1-10^{-8}\varepsilon)n/k$ for all i.

F3. Every vertex $v \in H \cap V_i$ satisfies $e(v, V_j \cap H) \geq (1-\varepsilon)d$ for all $j \neq i$.

F4. All $v \in H$ satisfy $e(v, V \setminus H) \leq \varepsilon d$.

F5. H is uniquely k-colorable.

If H, K are two induced subgraphs of G that satisfy F1–F5, then the same is true for $H \cup K$. Therefore, G has a unique maximal induced subgraph that enjoys F1–F5; this subgraph will be denoted by G_δ in the sequel.

Lemma 6.20. *Let $\delta \in (0,1]$ be some positive number. Let G be a random graph in $\mathcal{G}^{\text{uniform}}_{n,m,k}$, $m \geq C_0 k^4 n$, $C_0 = C_0(\delta)$ a sufficiently large constant. Then* whp *G is bounded and δ-feasible.*

This lemma is a direct consequence of Propositions 2.1, 3.6, 3.7.

Let $T \subset V$ be a set of size $t = \lceil \log n \rceil$, and let τ be a tree with vertex set T. Moreover, let us call G *(T,τ)-poor* if

- G is bounded,
- G is 0.01-feasible,
- G contains τ as a subgraph,
- T does not intersect $G_{0.02}$.

Denote by $\mathcal{G}$ the set of all k-colorable graphs with vertex set $V = \{1, \ldots, n\}$ and exactly m edges, and let $\mathcal{P}(T, \tau)$ signify the set of all (T, τ)-poor k-colorable graphs $G \in \mathcal{G}$. Below we shall establish the following.

Lemma 6.21. *We have* $\binom{n}{t} t^{t-2} |\mathcal{P}(T, \tau)| = o(|\mathcal{G}|)$.

Before we prove Lemma 6.21, let us note that it implies Proposition 3.9 immediately.

Proof of Proposition 3.9. Since there are $\binom{n}{t}$ ways to choose a vertex set T of size t, and then t^{t-2} ways to place a tree into that set, Lemmas 6.20 and 6.21 entail that

$$\begin{aligned} & P\left[\mathcal{G}_{n,m,k}^{\text{uniform}} \text{ violates the property stated in Proposition 3.9}\right] \\ & \quad \leq \ P\left[\mathcal{G}_{n,m,k}^{\text{uniform}} \text{ is not } \varepsilon\text{-feasible for some } \varepsilon \in \{0.01, 0.015, 0.02\} \text{ or not bounded}\right] \\ & \qquad + P\left[\exists T, \tau : \mathcal{G}_{n,m,k}^{\text{uniform}} \text{ is } (T, \tau)\text{-poor}\right] \leq o(1) + \sum_{T, \tau} \frac{|\mathcal{P}(T, \tau)|}{|\mathcal{G}|} = o(1), \end{aligned}$$

as claimed. □

Thus, the remaining task is to prove Lemma 6.21. To this end, we fix a set T and a tree τ and set up a bipartite auxiliary graph $\mathcal{A} = \mathcal{A}(T, \tau)$ with vertex set $V(\mathcal{A}) = \mathcal{P}(T, \tau) \oplus \mathcal{G}$; for brevity we set $\mathcal{P} = \mathcal{P}(T, \tau)$. The auxiliary graph will enjoy the following property.

> In $\mathcal{A}$ every vertex $G \in \mathcal{P}_{T,\tau}$ has degree at least Δ, while every vertex $G' \in \mathcal{G}$ has degree at most Δ', where $\binom{n}{t} t^{t-2} \Delta' = o(\Delta)$. (15)

Since $\Delta |\mathcal{P}(T, \tau)| \leq |E(\mathcal{A})| \leq \Delta' |\mathcal{G}|$, Lemma 6.21 follows direcly from (15).

To describe the construction of $\mathcal{G}$, we let I be the set of all $v \in T$ that have degree ≤ 4 in τ; then $|I| \geq t/2$, because τ is a tree. Furthermore, for each $G \in \mathcal{P}(T, \tau)$ we let $V_1(G), \ldots, V_k(G)$ signify the lexicographically first k-coloring of G, and we set

$$\begin{aligned} I_1(G) &= \{v \in I : e_G(v, V \setminus G_{0.02}) \geq 0.001d\}, \\ I_2(G) &= \{v \in I : \exists j : v \notin V_j(G) \wedge e_G(v, V_j(G) \cap G_{0.02}) \leq 0.999d\} \setminus I_1(G). \end{aligned}$$

If G is (T, τ)-poor, then all vertices $v \in I$ are outside of the 0.02-core $G_{0.02}$; hence, due to F3 and F4 we have $I = I_1(G) \cup I_2(G)$. Thus, we decompose $\mathcal{P}$ into two parts $\mathcal{P}_1 = \{G \in \mathcal{P} : |I_1|(G) \geq 0.15t\}$, $\mathcal{P}_2 = \mathcal{P} \setminus \mathcal{P}_1$

As a next step, we will construct two subgraphs $\mathcal{A}_1$, $\mathcal{A}_2$ of $\mathcal{A}$, both of which constists of the $\mathcal{P}_i$-$\mathcal{G}$-edges of $\mathcal{A}$. Thus, $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$, so that (15) will be a consequence of the following statement.

> In $\mathcal{A}_j$ every vertex $G \in \mathcal{P}_j$ has degree at least Δ_j, while every vertex $G' \in \mathcal{G}$ has degree at most Δ'_j, where $\binom{n}{t} t^{t-2} \Delta'_j = o(\Delta_j)$ $(j = 1, 2)$. (16)

In the remainder of this section, we present the constructions of $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ and establish (16). To facilitate these constructions, we say that a pair $\{x, y\}$ of vertices is *compatible* if $\{x, y\} \notin E(G)$, x, y lie in $G_{0.01}$, and x, y belong to different classes of the unique coloring of $G_{0.01}$. Moreover, we say that a set F of pairs of vertices is compatible if every pair in F is compatible and no vertex $v \in V$ occurs in more than one pair.

Lemma 6.22. *Let $G \in \mathcal{P}$ and let $1 \le s \le n^{0.1}$. Then there exist $\binom{n^2/4}{s}$ compatible sets F of size s.*

Proof. Let $Z_1, \ldots, Z_k$ signify the unique k-coloring of $G_{0.01}$, and let $\mathcal{C}$ be a complete k-partite graph with the color classes $Z_1, \ldots, Z_k$. Since G satisfies F2, $\mathcal{C}$ has at least $\sum_{1 \le i < j \le k} \|Z_i\|Z_j| \ge (0.9 - k^{-1})\binom{n}{2}$ edges. Furthermore, let S be a set of s edges of $\mathcal{C}$ chosen uniformly at random. Then the probability that S does not contain an edge of G is

$$\binom{|E(\mathcal{C})| - m}{s}\binom{|E(\mathcal{C})|}{s}^{-1} = \prod_{j=0}^{s-1} 1 - \frac{m}{|E(\mathcal{C})| - j} = 1 - o(1),$$

because $|E(\mathcal{C})| = \Omega(n^2)$, while $ms = o(n^2)$. Moreover, the probability that a specific vertex v occurs twice in S is at most

$$n^2\binom{|E(\mathcal{C})|}{s-2}\binom{|E(\mathcal{C})|}{s}^{-1} \le O(s^2 n^{-2}) = o(n^{-1}).$$

Hence, by the union bound with probability $1 - o(1)$ a randomly chosen S will touch no vertex v more than once. Thus, with probability $1 - o(1)$ a randomly chosen S is compatible, so that the number of compatible sets is $\ge (1 - o(1))\binom{|E(\mathcal{C})|}{s} \ge \binom{n^2/2}{s}$. ■

Construction of $\mathcal{A}_1$. The construction of $\mathcal{A}_1$ is based on the following observation.

Lemma 6.23. *Suppose that $G \in \mathcal{P}_2$. There exist sets $U \subset I_1(G)$, $|U| = \lceil 0.1t \rceil$, and $W \subset V \setminus (\tau \cup G_{0.02})$ such that $e(v, W) \ge 10^{-4}d$ for all $v \in U$, and $e(w, U) \le 10^4$ for all $w \in W$.*

Proof. Let $J \subset I_1(G)$ be a set of size $0.15t$, and let $K \subset J$ be the set of all vertices $w \in V \setminus (G_{0.02} \cup \tau)$ that are adjacent with a vertex in J. Moreover, let $L \subset K$ be the set of all $w \in K$ such that $e(v, J) \ge 10^4$. Then the boundedness property of G implies that $|L| \le 0.01t$. Furthermore, letting $Q = \{v \in J : e(v, L) > 10^4\}$, we have $|Q| \le 0.001t$ (once more due to the boundedness of G). Now, let $U = J \setminus Q$ and $W = K \setminus L$. Then each $w \in W$ has $\le 10^4$ neighbors in U. Moreover, if $v \in U$, then $e(v, W) \ge e(v, V \setminus (G_{0.02} \cup \tau)) - e(v, L) \ge 0.001d - 10 - 10^4 \ge 10^{-4}d$. ■

Our objective is to associate to each $G \in \mathcal{P}_1$ a large number of "target graphs" $G' \in \mathcal{G}$ such that no G' occurs as a target graph too frequently. To this end, we consider the following nondeterministic procedure that maps G to a target graph G'. For each possible outcome G' we include the edge $\{G, G'\}$ into $\mathcal{A}_1$. Set $\gamma = \lceil 10^{-4}d \rceil$ and $u = \lceil 0.1t \rceil$.

C1. Choose a compatible set F of size $t - 1 + \gamma u$.

C2. Choose sets U and W as in Lemma 6.23.

C3. For each $v \in U$ choose a set $\{w_1(v), \ldots, w_\gamma(v)\}$ of neighbors of v in W.

C4. Obtain G' from G by removing the edges of τ along with the edges $\{v, w_i(v)\}$ ($v \in U$, $1 \leq i \leq \gamma$) and adding the edges F.

Lemma 6.22 entails that the number of graphs G' that can be obtained from each G via the above procedure is at least

$$\Delta_1 = \binom{n^2/2}{t-1+\gamma u} \tag{17}$$

(because there are at least this many choices in step C1). Conversely, to recover G from G', we consider the following nondeterministic procedure.

R1. Choose a set F' of $t-1+\gamma u$ edges of G'.

R2. Choose a set $U' \subset T$ of size u.

R3. For each such $v \in U'$ choose a set N'_v of γ vertices outside of the 0.015-core of G'.

R4. Output the graph G'' obtained from G' by removing the edges F' and adding the edges $\{v, w\}$, $v \in U'$, $w \in N'_v$ along with the edges of τ.

Lemma 6.24. *If $\{G, G'\}$ is an edge of $\mathcal{A}_1$, then G' is* 0.015*-feasible and the process R1–R4 applied to G' can yield the output $G'' = G$.*

Proof. Let F, U, W, and $(\{w_1(v), \ldots, w_\gamma(v)\})_{v \in U}$ be the sets chosen by C1–C4 to obtain G' from G. If R1–R4 chooses $F' = F$, $U' = U$, $N'_v = \{w_1(v), \ldots, w_\gamma(v)\}$ for all $v \in U$, then the ourcome will be $G'' = G$. Thus, we just need to show that it is feasible for R1–R4 to choose $N'_v = \{w_1(v), \ldots, w_\gamma(v)\}$, i.e., that G' is 0.015-feasible and the vertices $w_j(v)$ do not belong to the 0.015-core of G'.

To see that G' is 0.015-feasible, let X be the vertex set of $G_{0.01}$. We claim that X satisfies F1–F5 with respect to G' with $\varepsilon = 0.01$. For F1 is an immediate consequence of the fact that G is 0.01-feasible. Moreover, as C4 adds a compatible set F and only removes edges that contain a vertex outside of X, the unique k-coloring of $G_{0.01}$ remains the unique k-coloring of the set X in G', whence F2–F5 follow. Thus, G' is indeed 0.01-feasible, and hence 0.015-feasible as well.

Finally, to show that the vertex set Y of $G'_{0.015}$ is contained in that of $G_{0.02}$, we show that Y is 0.02-feasible in G. For the induced subgraph $G[Y]$ is uniquely k-colorable, because all edges in $E(G') \setminus E(G)$ lie in the uniquely k-colorable subgraph $G_{0.01}$ of G. Hence, Y satisfies F5, and F1–F2 just follow from the fact that Y is 0.015-feasible in G'. Moreover, as no vertex $v \in V$ occurs in the set $E(G) \setminus E(G')$ of edges removed in C4 more than γ times, Y also satisfies F3 and F4 in G with $\varepsilon = 0.02$. ■

Lemma 6.25. *If G' is an outcome of C1–C4 for some $G \in \mathcal{P}_1$, then the number of possibles nondeterministc choices in the R1–R4 is at most $\Delta'_1 = 2^t \binom{m}{t-1+\gamma u} \binom{\exp(-\sqrt{d})n}{\gamma}^u$.*

Proof. The first factor accounts for the number of ways to choose F'. Moreover, there are clearly at most 2^t ways to choose U'. To bound the number of choices of R3, note that for each $v \in U'$ there are at most $\binom{n-|V(G'_{0.015})|}{\gamma}$ ways to choose the set N'_v. As the construction C1–C4 ensures that $G'_{0.015}$ contains the 0.01-core $G_{0.01}$ of G, our assumption that G is 0.01-feasible entails that $|V(G'_{0.015})| \geq n(1 - \exp(-\sqrt{d}))$. ■

Finally, combining (17) with Lemmas 6.24 and 6.25, and observing that $\binom{n}{t} t^{t-2} \Delta'_1 = o(\Delta_1)$, we obtain (16) for $j = 1$.

Construction of $\mathcal{A}_2$. Let $G \in \mathcal{P}_2$, and let $V_1(G), \dots, V_k(G)$ be the lexicographically first k-coloring of G. We split the set $I_2(G)$ into k subsets

$$I_{2j}(G) = \{v \in I_2 : v \notin V_j(G) \wedge e(v, V_j(G) \cap G_{0.02}) \leq 0.999d\} \qquad (1 \leq j \leq k).$$

Moreover, we split $\mathcal{P}_2$ into subsets

$$\mathcal{P}_{2j} = \{G \in \mathcal{P}_2 : |I_{2j}(G)| \geq 0.1t/k\} \setminus \bigcup_{1 \leq i < j} \mathcal{P}_{2i} \quad (1 \leq j \leq k).$$

Without loss of generality, we shall just consider the case $G \in \mathcal{P}_{21}$ in the sequel.

As in the construction of $\mathcal{A}_1$ we consider a nondeterministic procedure that maps $G \in \mathcal{P}_2$ to $G' \in \mathcal{G}$. Let $u = \lceil 0.1t/k \rceil$ and $\gamma = \lceil 10^{-9}d \rceil$.

C1. Choose a compatible set F of size $t-1$.

C2. Choose a subset $U \subset I_{21}(G)$ of size u.

C3. Choose a matching $M \subset E(G_{0.01})$ of size γu such that no vertex v is adjacent to more than 100 vertices that occur in M. Moreover, for each $v \in U$ choose a set $N_v \subset V_1 \cap G_{0.01}$ of size γ such that the sets $(N_v)_{v \in U}$ are pairwise disjoint, $e(v, N_v) = 0$, and no vertex of N_v occurs in M.

C4. Obtain G' from G by removing the edges of τ and the matching M, adding the edges F, and connecting each $v \in U$ with all $w \in N_v$.

For each $G \in \mathcal{P}_{21}$ and each possible outcome G' of C1–C4 we include the edges $\{G, G'\}$ into $\mathcal{A}_2$. The following lemma provides a lower bound on the degree of $G \in \mathcal{P}_{21}$ in $\mathcal{A}_2$.

Lemma 6.26. *Each $G \in \mathcal{P}_{21}$ has at least $\Delta_{21} = \frac{1}{2}\binom{n^2/4}{t-1}\binom{(1-10^{-9})m}{\gamma u}\binom{(1-10^{-9})n/k}{\gamma}^u$ images G'.*

Proof. By Lemma 6.22 there are $\binom{n^2/4}{t-1}$ ways to choose F. Furthermore, F1 implies that $G_{0.01}$ contains at least $(1 - 10{-}9)m$ edges. Moreover, since the maximum degree of G is $\leq \ln^2 n$ by B2, $G_{0.01}$ has at least $(1-o(1))\binom{(1-10^{-9})m}{0.1\delta dt/k}$ matchings of size $0.1\delta dt/k$. Finally, since $|V_1| \geq (1-10^{-9})\frac{n}{k}$ by F2, there are $(1-o(1))\binom{(1-10^{-9})n/k}{\delta d}^{0.1t/k}$ ways to choose the sets $(N_v)_{v \in U}$. ■

Conversely, we consider the following nondeterministic procedure for obtaining a graph G'' from an outcome G' of C1–C4.

R1. Choose a set $F' \subset E(G')$ of size $t-1$.

R2. Determine the unique coloring $V_1', \dots, V_k'$ of $G'_{0.015}$. Then, choose a set $U' \subset T$ of size u and an index l such that each $v \in U'$ has at most $0.9999d$ neighbors in V_l'. Moreover, choose a set M' of γu pairs of vertices such that each $e \in M'$ consists of two vertices belonging to different classes of $V_1', \dots, V_k'$.

R3. For each $v \in U'$ choose a set N_v' of neighbors of v in V_l' such that $|N_v'| = \gamma$.

R4. Obtain a graph G'' from G' by removing F' and all edges $\{v, w\}$ with $v \in U'$, $w \in N_v'$, and adding the edges of τ and M'.

Lemma 6.27. *If $\{G, G'\}$ is an edge of $\mathcal{A}_2$, then G' is 0.015-feasible and the process R1–R4 applied to G' can yield the output $G'' = G$.*

Proof. Suppose that G' has been obtained from G by choosing the matching M, the set U, the sets $(N_v)_{v\in U}$, and the feasible set F. To recover $G'' = G'$, we shall prove that G' is 0.15 feasible and that the process R1–R4 can choose $M' = M$, $F' = F$, and $N'_v = N_v$.

To show that G' is 0.15 feasible, let Z be the set of all vertices that occur in M and $H = V(G_{0.01}) \setminus Z$. We claim that H is 0.015-feasible in G'. For H satisfies the assumption of condition B3 in G, whence $G[H] = G'[H]$ is uniquely k-colorable. Moreover, since $|Z| = O(\ln n)$, H satisfies F1, F2, F3, and F5. Further, since the sets N_v are pairwise disjoint, we have $e_{G'}(v, V \setminus H) \leq e_G(v, V \setminus H) + 1 \leq e_G(v, V \setminus G_{0.01}) + 101$, because no vertex of G has more than 100 neighbors in Z. Therefore, H is 0.015-feasible in G'.

Indeed, we have shown that $V(G'_{0.015}) \supset V(G_{0.01}) \setminus Z$. Hence, as $G'_{0.015}$ is uniquely k-colorable, for a suitable value of l we have $V'_l \supset \bigcup_{v\in U} N_v$. Moreover, since $V(G'_{0.015}) \subset V(G_{0.02})$, all $v \in U$ satisfy $e(v, V'_l) \leq 0.9999d$. Therefore, it is feasible for R1–R4 to choose $M' = M$, $F' = F$, and $N'_v = N_v$, thereby recovering $G'' = G$. ∎

In the light of Lemma 6.27 we can bound the degrees of $G' \in \mathcal{G}$ in $\mathcal{A}_2$ as follows.

Lemma 6.28. *If G' has been obtained from G via C1–C4, then during R1–R4 there are at most $\Delta'_2 = \binom{m}{t-1} 2^t \binom{(1-k^{-1})\binom{n}{2}}{\gamma u} \binom{0.9999d}{\gamma}^u$ ways to choose F', the sets N'_v, and M'. Hence, the degree of any $G' \in \mathcal{G}$ in $\mathcal{A}_2$ is $\leq k\Delta'_2$.*

Proof. There are exactly $\binom{m}{t-1}$ ways to choose F' and at most 2^t ways of choosing U'. Furthermore, by Turan's theorem there are at most $\binom{(1-k^{-1})\binom{n}{2}}{\gamma u}$ ways to choose M'. Finally, since each $v \in U'$ has at most $0.9999d$ neighbors in V'_1, there are at most $\binom{0.9999d}{\gamma}$ ways to choose N'_v. ∎

Combining the bounds from Lemmas 6.26 and 6.28, we obtain

$$\begin{aligned}\frac{\Delta_3}{\Delta'_3} &\geq \Omega(n^{-1}) \left(\frac{n}{4dk}\right)^{t-1} \left(\frac{(1-10^{-9})^2 mn/k}{0.9999(1-k^{-1})n^2 d/2}\right)^{0.1\delta dt/k} \\ &\geq \left(\frac{n}{4dk}\right)^{t-1} \left(\frac{(1-10^{-9})^2}{0.9999}\right)^{\gamma u} \geq \exp(\Omega(\gamma u)) \binom{n}{t} t^{t-2}\end{aligned}$$

Thus, we have established (16) for $j = 3$.

7 Conclusion

In this work we consider the uniform distribution over k-colorable graphs, $\mathcal{G}^{\text{uniform}}_{n,m,k}$, with average degree greater than some sufficiently large constant. We characterize the typical structure of the solution space of such graphs to show that typically there exists only one cluster of proper k-colorings, whose size may be exponential in n, in which almost all vertices are frozen. We also prove that a relatively simple efficient algorithm recovers *whp* a proper k-coloring of such graphs, thus asserting that almost all k-colorable graphs are easy to color.

To obtain our results we had to come up with new analytical tools that apply to a number of further NP-hard problems, including the satisfiability problem. Our result also implies that the algorithmic techniques developed for random formulas from the planted distribution, e.g. [10, 6], can be extended to the significantly more natural uniform distribution [14].

Combining Theorems 1.1 and 1.2 *rigorously* supports the following common thesis: the main key to understanding the hardness (even experimental one) of a certain distribution over k-colorable graphs lies in the structure of the solution space of a typical graph in that distribution. Specifically, our results show (at least in our setting) that typically when a graph has a single cluster of proper k-colorings, though its volume may be exponential in n, then the problem is "easy". On the other hand, when the clustering is complicated, for example in the near threshold regime, experimental results predict that many "simple" heuristics fail, while "heavy machinery" such as Survey Propagation works. Heightening this last point, regard the recent work in [18] which considers the planted 3SAT distribution. There it is proved that the naïve Warning Propagation algorithm works *whp* for planted 3CNF formulas with a suitable parametrization which, amongst other characteristics, typically have one cluster of satisfying assignments. Fitting the result into our perspective – when the clustering is simple, then a simple message passing algorithm works (Warning Propagation), when the clustering is complicated, then only a much more complicated message passing algorithm is known (and even this only experimentally) to work (Survey Propagation).

Acknowledgements: we thank Uriel Feige for useful discussions.

References

[1] D. Achlioptas and E. Friedgut. A sharp threshold for k-colorability. *Random Struct. Algorithms*, 14(1):63–70, 1999.

[2] D. Achlioptas and M. Molloy. Almost all graphs with 2.522n edges are not 3-colorable. *E. Jour. Of Comb.*, 6(1), 1999.

[3] D. Achlioptas and C. Moore. Almost all graphs with average degree 4 are 3-colorable. In *STOC '02: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 199–208, 2002.

[4] D. Achlioptas and C. Moore. Random k-sat: Two moments suffice to cross a sharp threshold. *CoRR*, cond-mat/0310227, 2003.

[5] D. Achlioptas and F. Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 130–139, 2006.

[6] N. Alon and N. Kahale. A spectral technique for coloring random 3-colorable graphs. *SIAM J. on Comput.*, 26(6):1733–1748, 1997.

[7] S. Ben-Shimon and M. Krivelevich. Random regular graphs of non-constant degree: edge distribution and applications. *manuscript*, 2006.

[8] A. Blum and J. Spencer. Coloring random and semi-random k-colorable graphs. *J. of Algorithms*, 19(2):204–234, 1995.

[9] B. Bollobás. The chromatic number of random graphs. *Combinatorica*, 8(1):49–55, 1988.

[10] J. Böttcher. Coloring sparse random k-colorable graphs in polynomial expected time. In *Proc. 30th International Symp. on Mathematical Foundations of Computer Science*, pages 156–167, 2005.

[11] A. Braunstein, M. Mézard, M. Weigt, and R. Zecchina. Constraint satisfaction by survey propagation. *Computational Complexity and Statistical Physics*, 2005.

[12] A. Braunstein, M. Mezard, and R. Zecchina. Survey propagation: an algorithm for satisfiability. *Random Structures and Algorithms*, 27:201–226, 2005.

[13] A. Coja-Oghlan. Coloring semirandom graphs optimally. In *Proc. 31st International Colloquium on Automata, Languages, and Programming*, pages 383–395, 2004.

[14] A. Coja-Oghlan, M. Krivelevich, and D. Vilenchik. Why almost all satifiable k-cnf formulas are easy. In *13th conferene on Analysis of Algorithms*, 2007. to appear.

[15] M. E. Dyer and A. M. Frieze. The solution of some random np-hard problems in polynomial expected time. *J. Algorithms*, 10(4):451–489, 1989.

[16] U. Feige and J. Kilian. Zero knowledge and the chromatic number. *J. Comput. and Syst. Sci.*, 57(2):187–199, 1998.

[17] U. Feige and J. Kilian. Heuristics for semirandom graph problems. *J. Comput. and Syst. Sci.*, 63(4):639–671, 2001.

[18] U. Feige, E. Mossel, and D. Vilenchik. Complete convergence of message passing algorithms for some satisfiability problems. In *RANDOM*, 2006.

[19] Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Struct. Algorithms*, 27(2):251–275, 2005.

[20] Joel Friedman, Jeff Kahn, and Endre Szemerédi. On the second eigenvalue in random regular graphs. In *Proc. 21st ACM Symp. on Theory of Computing*, pages 587–598, 1989.

[21] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[22] M. Krivelevich and D. Vilenchik. Semirandom models as benchmarks for coloring algorithms. In *Third Workshop on Analytic Algorithmics and Combinatorics*, pages 211–221, 2006.

[23] L. Kučera. Expected behavior of graph coloring algorithms. In *Proc. Fundamentals of Computation Theory*, volume 56 of *Lecture Notes in Comput. Sci.*, pages 447–451. Springer, Berlin, 1977.

[24] T. Łuczak. The chromatic number of random graphs. *Combinatorica*, 11(1):45–54, 1991.

[25] M. Mezard, T. Mora, and R. Zecchina. Clustering of solutions in the random satisfiability problem. *Physical Review Letters*, 94:197–205, 2005.

[26] T. Mora, M. Mezard, and R. Zecchina. Pairs of sat assignments and clustering in random boolean formulae, 2005.

[27] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina. Coloring random graphs. *Phys. Rev. Lett.*, 89(26):268701, 2002.

[28] H. Prömel and A. Steger. Random l-colorable graphs. *Random Structures and Algorithms*, 6:21–37, 1995.

[29] J. S. Turner. Almost all k-colorable graphs are easy to color. *J. Algorithms*, 9(1):63–82, 1988.