A spectral technique for coloring random 3-colorable graphs

Noga Alon * Nabil Kahale †

Abstract

Let $G_{3n,p,3}$ be a random 3-colorable graph on a set of $3n$ vertices generated as follows. First, split the vertices arbitrarily into three equal color classes and then choose every pair of vertices of distinct color classes, randomly and independently, to be an edge with probability p. We describe a polynomial time algorithm that finds a proper 3-coloring of $G_{3n,p,3}$ with high probability, whenever $p \geq c/n$, where c is a sufficiently large absolute constant. This settles a problem of Blum and Spencer, who asked if one can design an algorithm that works almost surely for $p \geq \text{polylog}(n)/n$. The algorithm can be extended to produce optimal k-colorings of random k-colorable graphs in a similar model, as well as in various related models. Implementation results show that the algorithm performs very well in practice even for moderate values of c.

1 Introduction

A vertex coloring of a graph G is *proper* if no adjacent vertices receive the same color. The *chromatic number* $\chi(G)$ of G is the minimum number of colors in a proper vertex coloring of it. The problem of determining or estimating this parameter has received a considerable amount of attention in Combinatorics and in Theoretical Computer Science, as several scheduling problems are naturally formulated as graph coloring problems. It is well known (see [13, 12]) that the problem of properly coloring a graph of chromatic number k with k colors is NP-hard, even for any fixed $k \geq 3$, and it is therefore unlikely that there are efficient algorithms for optimally coloring an arbitrary 3-chromatic input graph.

On the other hand, various researchers noticed that *random* k-colorable graphs are usually easy to color optimally. Polynomial time algorithms that optimally color random k-colorable graphs for every fixed k with high probability, have been developed by Kucera [15], by Turner [18] and by Dyer and Frieze [8], where the last paper provides an algorithm whose average running time over all k-colorable graphs on n vertices is polynomial. Note, however, that most k-colorable graphs are quite dense, and hence easy to color. In fact, in a typical k-colorable graph, the number of common neighbors of any pair of vertices with the same color exceeds considerably that of any pair of vertices of distinct colors, and hence a simple coloring algorithm based on this fact already works with high probability. It is more difficult to color sparser random k-colorable graphs. A

A preliminary version of this paper appeared in the Proc. of the 26^{th} ACM STOC, ACM Press (1994), 346-355.

*Institute for Advanced Study, Princeton, NJ 08540, USA and Department of Mathematics, Tel Aviv University, Tel Aviv, Israel. Email: noga@math.tau.ac.il. Research supported in part by the Sloan Foundation, Grant No. 93-6-6 and by a USA-Israel BSF grant.

†ATT Bell Laboratories, Murray Hill, NJ 07974. Email: kahale@research.att.com. This work was done while the author was at DIMACS.

precise model for generating sparse random k-colorable graphs is described in the next subsection, where the sparsity is governed by a parameter p that specifies the edge probability. Petford and Welsh [16] suggested a randomized heuristic for 3-coloring random 3-colorable graphs and supplied experimental evidence that it works for most edge probabilities. Blum and Spencer [6] (see also [3] for some related results) designed a polynomial algorithm and proved that it colors optimally, with high probability, random 3-colorable graphs on n vertices with edge probability p provided $p \geq n^{\epsilon}/n$, for some arbitrarily small but fixed $\epsilon > 0$. Their algorithm is based on a path counting technique, and can be viewed as a natural generalization of the simple algorithm based on counting common neighbors (that counts paths of length 2), mentioned above.

Our main result here is a polynomial time algorithm that works for sparser random 3-colorable graphs. If the edge probability p satisfies $p \geq c/n$, where c is a sufficiently large absolute constant, the algorithm colors optimally the corresponding random 3-colorable graph with high probability. This settles a problem of Blum and Spencer [6], who asked if one can design an algorithm that works almost surely for $p \geq \mathrm{polylog}(n)/n$. (Here, and in what follows, *almost surely* always means: with probability that approaches 1 as n tends to infinity). The algorithm uses the spectral properties of the graph and is based on the fact that almost surely a rather accurate approximation of the color classes can be read from the eigenvectors corresponding to the smallest two eigenvalues of the adjacency matrix of a large subgraph. This approximation can then be improved to yield a proper coloring.

The algorithm can be easily extended to the case of k-colorable graphs, for any fixed k, and to various models of random *regular* 3-colorable graphs.

We implemented our algorithm and tested it for hundreds of graphs drawn at random from the distribution of $G_{3n,p,3}$. Experiments show that our algorithm performs very well in practice. The running time is a few minutes on graphs with up to 100000 nodes, and the range of edge probabilities on which the algorithm is successful is in fact even larger than what our analysis predicts.

1.1 The model

There are several possible models for random k-colorable graphs. See [8] for some of these models and the relation between them. Our results hold for most of these models, but it is convenient to focus on one, which will simplify the presentation. Let V be a fixed set of kn labelled vertices. For a real $p = p(n)$, let $G_{kn,p,k}$ be the random graph on the set of vertices V obtained as follows; first, split the vertices of V arbitrarily into k color classes $W_1, \ldots, W_k$, each of cardinality n. Next, for each u and v that lie in distinct color classes, choose uv to be an edge, randomly and independently, with probability p. The input to our algorithm is a graph $G_{kn,p,k}$ obtained as above, and the algorithm succeeds to color it if it finds a proper k coloring. Here we are interested in fixed $k \geq 3$ and large n. We say that an algorithm colors $G_{kn,p,k}$ *almost surely* if the probability that a randomly chosen graph as above is properly colored by the algorithm tends to one as n tends to infinity. Note that we consider here deterministic algorithms, and the above statement means that the algorithm succeeds to color almost all random graphs generated as above.

A closely related model to the one given above is the model in which we do not insist that the color classes have equal sizes. In this model one first splits the set of vertices into k disjoint color classes by letting each vertex choose its color randomly, independently and uniformly among the k possibilities. Next, one chooses every pair of vertices of distinct color classes to be an edge with probability p. All our results hold for both models, and we focus on the first one as it is more convenient. To simplify the presentation, we restrict our attention to the case $k = 3$ of 3-colorable graphs, since the results for this case easily extend to every *fixed* k. In addition, we make no attempt to optimize the constants and assume, whenever this is needed, that c is a sufficiently large contant, and the number of vertices $3n$ is sufficiently large.

1.2 The algorithm

Here is a description of the algorithm, which consists of three phases. Given a graph $G = G_{3n,p,3} = (V, E)$, define $d = pn$. Let $G' = (V, E')$ be the graph obtained from G by deleting all edges incident to a vertex of degree greater than $5d$. Denote by A the adjacency matrix of G', i.e., the $3n$ by $3n$ matrix $(a_{uv})_{u,v \in V}$ defined by $a_{uv} = 1$ if $uv \in E'$ and $a_{uv} = 0$ otherwise. It is well known that since A is symmetric it has real eigenvalues $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_{3n}$ and an orthonormal basis of eigenvectors $e_1, e_2, \ldots, e_{3n}$, where $Ae_i = \lambda_i e_i$. The crucial point is that almost surely one can deduce a good approximation of the coloring of G from e_{3n-1} and e_{3n}. Note that there are several efficient algorithms to compute the eigenvalues and the eigenvectors of symmetric matrices (cf., e.g., [17]) and hence e_{3n-1} and e_{3n} can certainly be calculated in polynomial time. For the rest of the algorithm, we will deal with G rather than G'.

Let t be a non-zero linear combination of e_{3n-1} and e_{3n} whose median is zero, that is, the number of positive components of t as well as the number of its negative components are both at most $3n/2$. (It is easy to see that such a combination always exists and can be found efficiently.) Suppose also that t is normalized so that it's l_2-norm is $\sqrt{2n}$. Define $V_1^0 = \{u \in V : t_u > 1/2\}$, $V_2^0 = \{u \in V : t_u < -1/2\}$, and $V_3^0 = \{u \in V : |t_u| \leq 1/2\}$. This is an approximation for the coloring, which will be improved in the second phase by iterations, and then in the third phase to obtain a proper 3-coloring.

In iteration i of the second phase, $0 < i \leq q = \lceil \log n \rceil$, construct the color classes V_1^i, V_2^i and V_3^i as follows. For every vertex v of G, let $N(v)$ denote the set of all its neighbors in G. In the i-th iteration, color v by the least popular color of its neighbors in the previous iteration. That is, put v in V_j^i if $|N(v) \cap V_j^{i-1}|$ is the minimum among the three quantities $|N(v) \cap V_l^{i-1}|$, $(l = 1, 2, 3)$, where equalities are broken arbitrarily. We will show that the three sets V_i^q correctly color all but $n2^{-\Omega(d)}$ vertices.

The third phase consists of two stages. First, repeatedly uncolor every vertex colored j that has less than $d/2$ neighbors (in G) colored l, for some $l \in \{1, 2, 3\} - \{j\}$. Then, if the graph induced on the set of uncolored vertices has a connected component of size larger than $\log_3 n$, the algorithm fails. Otherwise, find a coloring of every component consistent with the rest of the graph using brute force exhaustive search. If the algorithm cannot find such a coloring, it fails.

Our main result is the following.

Theorem 1.1 *If $p > c/n$, where c is a sufficiently large constant, the algorithm produces a proper 3-coloring of G with probability $1 - o(1)$.*

The intuition behind the algorithm is as follows. Suppose every vertex in G had exactly d neighbors in every color class other than its own. Then $G' = G$. 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 (as observed in [1]) shows that any non-zero element of F is an eigenvector of A with eigenvalue $-d$. Moreover, if E is the union of random matchings, one can show that $-d$ is almost surely the smallest eigenvalue of A and that F is precisely the eigenspace corresponding to $-d$. Thus, any linear combination t of e_{3n-1} and e_{3n} is constant on every color class. 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 the coloring obtained after phase 1 of the algorithm is a proper coloring. In the model specified in Subsection 1.1 these regularity assumptions do not hold, but every vertex has the same *expected* number of neighbors in every color class other than its own. This is why phase 1 gives only an approximation of the coloring and phases 2 and 3 are needed to get a proper coloring.

We prove Theorem 1.1 in the next two sections. We use the fact that almost surely the largest eigenvalue of G' is at least $(1-2^{-\Omega(d)})2d$, and that its two smallest eigenvalues are at most $-(1-2^{-\Omega(d)})d$ and all other eigenvalues are in absolute value $O(\sqrt{d})$. The proof of this result is based on a proper modification of techniques developed by Friedman, Kahn and Szemerédi in [11], and is deferred to Section 3. We show in Section 2 that it implies that each of the two eigenvectors corresponding to the two smallest eigenvalues is close to a vector which is a constant on every color class, where the sum of these three constants is zero. This suffices to show that the sets V_j^0 form a reasonably good approximation to the coloring of G, with high probability.

Theorem 1.1 can then be proved by applying the expansion properties of the graph G (that hold almost surely) to show that the iteration process above converges quickly to a proper coloring of a predefined large subgraph H of G. The uncoloring procedure will uncolor all vertices which are wrongly colored, but will not affect the subgraph H. We then conclude by showing that the largest connected component of the induced subgraph of G on $V-H$ is of logarithmic size almost surely, thereby showing that the brute-force search on the set of uncolored vertices terminates in polynomial time. We present our implementation results in Section 4. Section 5 contains some concluding remarks together with possible extensions and results for related models of random graphs.

2 The proof of the main result

Let $G = G_{3n,p,3} = (V, E)$ be a random 3-colorable graph generated according to the model described above. Denote by W_1, W_2 and W_3 the three color classes of vertices of G. Let G' be the graph obtained from G by deleting all edges adjacent to vertices of degree greater than $5d$, and let A be the adjacency matrix of G'. Denote by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{3n}$ the eigenvalues of A, and by $e_1, e_2, \ldots, e_{3n}$ the corresponding eigenvectors, chosen so that they form an orthonormal basis of R^{3n}.

In this section we first show that the approximate coloring produced by the algorithm using the eigenvectors e_{3n-1} and e_{3n} is rather accurate almost surely. Then we exhibit a large subgraph H and show that, almost surely, the iterative procedure for improving the coloring colors H correctly. We then show that the third phase finds a proper coloring of G in polynomial time, almost surely. We use the following statement, whose proof is relegated to Section 3.

Proposition 2.1 *In the above notation, almost surely,*
(i) $\lambda_1 \geq (1-2^{-\Omega(d)})2d$,
(ii) $\lambda_{3n} \leq \lambda_{3n-1} \leq -(1-2^{-\Omega(d)})d$ *and*
(iii) $|\lambda_i| \leq O(\sqrt{d})$ *for all* $2 \leq i \leq 3n-2$.

Remark. One can show that, when $p = o(\log n/n)$, Proposition 2.1 would not hold if we were dealing with the spectrum of G rather than that of G', since the graph G is likely to contain many vertices of degree $>> d$, and in this case the assertion of (iii) does not hold for the eigenvalues of G.

2.1 The properties of the last two eigenvectors

We show in this subsection that the eigenvectors e_{3n-1} and e_{3n} are almost constant on every color class. For this, we exhibit two orthogonal vectors constant on every color class which, roughly speaking, are close to being eigenvectors corresponding to $-d$. Let $x = (x_v : v \in V)$ be the vector defined by $x_v = 2$ for $v \in W_1$, and $x_v = -1$ otherwise. Let $y = (y_v : v \in V)$ be the vector defined by $y_v = 0$ if $v \in W_1$, $y_v = 1$ if $v \in W_2$ and $y_v = -1$ if $v \in W_3$. We denote by $||f||$ the l_2-norm of a vector f.

Lemma 2.2 *Almost surely there are two vectors $\epsilon = (\epsilon_v : v \in V)$ and $\delta = (\delta_v : v \in V)$, satisfying $||\epsilon||^2 = O(n/d)$ and $||\delta||^2 = O(n/d)$ so that $x - \epsilon$ and $y - \delta$ are both linear combinations of e_{3n-1} and e_{3n}.*

Proof. We use the following lemma, whose proof is given below.

Lemma 2.3 *Almost surely: $||(A + dI)y||^2 = O(nd)$ and $||(A + dI)x||^2 = O(nd)$.*

We prove the existence of δ as above. The proof of the existence of ϵ is analogous. Let $y = \sum_{i=1}^{3n} c_i e_i$. We show that the coefficients $c_1, c_2, \ldots, c_{3n-2}$ are small compared to $||y||$. Indeed, $(A + dI)y = \sum_{i=1}^{3n} c_i(\lambda_i + d)e_i$, and so

$$\begin{aligned} ||(A+dI)y||^2 &= \sum_{i=1}^{3n} {c_i}^2(\lambda_i + d)^2 \qquad (1) \\ &\geq \Omega(d^2) \sum_{i=1}^{3n-2} {c_i}^2, \end{aligned}$$

where the last inequality follows from parts (i) and (iii) of Proposition 2.1. Define $\delta = \sum_{i=1}^{3n-2} c_i e_i$. By (1) and Lemma 2.3 it follows that $||\delta||^2 = \sum_{i=1}^{3n-2} {c_i}^2 = O(n/d)$. On the other hand, $y - \delta$ is a linear combination of e_{3n-1} and e_{3n}. $\square$

Note that it was crucial to the proof of Lemma 2.2 that, almost surely, $||(A + dI)y||^2$ is $O(nd)$ rather than $\Omega(nd^2)$ as is the case for some vectors in $\{-1, 0, 1\}^{3n}$.

Proof of Lemma 2.3 To prove the first bound, observe that it suffices to show that the sum of squares of the coordinates of $(A + dI)y$ on W_1 is $O(nd)$ almost surely, as the sums on W_2 and W_3 can be bounded similarly. The expectation of the vector $(A + dI)y$ is the null vector, and the expectation of the square of each coordinate of $(A + dI)y$ is $O(d)$, by a standard calculation. This is because each coordinate of $(A + dI)y$ is the sum of n independent random variables, each with mean 0 and variance $O(d/n)$. This implies that the *expected* value of the sum of squares of the coordinates of $(A + dI)y$ on W_1 is $O(nd)$. Similarly, the expectation of the fourth power of each coordinate of $(A + dI)y$ is $O(d^2)$. Hence, the variance of the square of each coordinate is $O(d^2)$. However, the coordinates of $(A + dI)y$ on W_1 are independent random variables, and hence the variance of the sum of the squares of the W_1 coordinates is equal to the sum of the variances, which is $O(nd^2)$. The first bound can now be deduced from Chebyshev's Inequality. The second bound can be shown in a similar manner. We omit the details. □

The vectors $x - \epsilon$ and $y - \delta$ are independent since they are nearly orthogonal. Indeed, if $\alpha(x - \epsilon) + \beta(y - \delta) = 0$, then $\alpha x + \beta y = \alpha\epsilon + \beta\delta$, and so $6n\alpha^2 + 2n\beta^2 = ||\alpha\epsilon + \beta\delta||^2$. But

$$\begin{aligned} ||\alpha\epsilon + \beta\delta|| &\leq |\alpha|\ ||\epsilon|| + |\beta|\ ||\delta|| \\ &= O\left((|\alpha| + |\beta|)\sqrt{n/d}\right). \end{aligned}$$

Thus $6\alpha^2 + 2\beta^2 = O\left((\alpha^2 + \beta^2)/d\right)$, and hence $\alpha = \beta = 0$.

Therefore, by the above lemma, the two vectors $\sqrt{3n}e_{3n-1}$ and $\sqrt{3n}e_{3n}$ can be written as linear combinations of $x - \epsilon$ and $y - \delta$. Moreover, the coefficients in these linear combinations are all $O(1)$ in absolute value. This is because $x - \epsilon$ and $y - \delta$ are nearly orthogonal, and the l_2-norm of each of the four vectors $x - \epsilon, y - \delta, \sqrt{3n}e_{3n-1}$ and $\sqrt{3n}e_{3n}$ is $\Theta(\sqrt{n})$. More precisely, if one of the vectors $\sqrt{3n}e_{3n-1}$, $\sqrt{3n}e_{3n}$ is written as $\alpha(x - \epsilon) + \beta(y - \delta)$ then, by the triangle inequality, $||\alpha x + \beta y|| \leq \Theta(\sqrt{n}) + |\alpha|\ ||\epsilon|| + |\beta|\ ||\delta||$ which, by a calculation similar to the one above, implies that $6\alpha^2 + 2\beta^2 \leq O(1) + O((\alpha^2 + \beta^2)/d)$, and thus α and β are $O(1)$. On the other hand, the coefficients of the vector t defined in subsection 1.2 along the vectors e_{3n-1} and e_{3n} are at most $||t|| = \sqrt{2n}$. It follows that the vector t defined in the algorithm is also a linear combination of the vectors $x - \epsilon$ and $y - \delta$ with coefficients whose absolute values are both $O(1)$. Since both x and y belong to the vector space F defined in the proof of Proposition 2.1, this implies that $t = f + \eta$, where $f \in F$ and $||\eta||^2 = O(n/d)$. Let α_i be the value of f on W_i, for $1 \leq i \leq 3$. Assume without loss of generality that $\alpha_1 \geq \alpha_2 \geq \alpha_3$. Since $||\eta||^2 = O(n/d)$, at most $O(n/d)$ of the coordinates of η are greater than 0.01 in absolute value. This implies that $|\alpha_2| \leq 1/4$, because otherwise at least $2n - O(n/d)$ coordinates of t would have the same sign, contradicting the fact that 0 is a median of t. As $\alpha_1 + \alpha_2 + \alpha_3 = 0$ and ${\alpha_1}^2 + {\alpha_2}^2 + {\alpha_3}^2 = ||f||^2/n = 2 + O(d^{-1})$, this implies that $\alpha_1 > 3/4$ and $\alpha_3 < -3/4$. Therefore, the coloring defined by the sets V_j^0 agrees with the original coloring of G on all but at most $O(n/d) < 0.001n$ coordinates.

2.2 The iterative procedure

Denote by H the subset of V obtained as follows. First, set H to be the set of vertices having at most $1.01d$ neighbors in G in each color class. Then, repeatedly, delete any vertex in H having less than $0.99d$ neighbors in H in some color class (other than its own.) Thus, each vertex in H has roughly d neighbors in H in each color class other than its own.

Proposition 2.4 *Almost surely, by the end of the second phase of the algorithm, all vertices in H are properly colored.*

To prove Proposition 2.4, we need the following lemma.

Lemma 2.5 *Almost surely, there are no two subsets of vertices U and W of V such that $|U| \leq 0.001n$, $|W| = |U|/2$, and every vertex v of W has at least $d/4$ neighbors in U.*

Proof. Note that if there are such two (not necessarily disjoint) subsets U and W, then the number of edges joining vertices of U and W is at least $d|W|/8$. Therefore, by a standard calculation, the probability that there exist such two subsets is at most

$$\begin{aligned}
\sum_{i=1}^{0.0005n} \binom{3n}{i}\binom{3n}{2i}\binom{2i^2}{di/8}\left(\frac{d}{n}\right)^{di/8} &\leq \sum_{i=1}^{0.0005n} \left(\frac{3en}{i}\right)^{3i}\left(\frac{16ei}{n}\right)^{di/8} \\
&\leq \sum_{i=1}^{0.0005n} \left(\frac{3en}{i}\right)^{di/40}\left(\frac{16ei}{n}\right)^{di/8} \\
&= \sum_{i=1}^{0.0005n} \left(48e^2\right)^{di/40}\left(\frac{16ei}{n}\right)^{di/10} \\
&\leq \sum_{i=1}^{0.0005n} \left(48e^2(16e/2000)^2\right)^{di/40}\left(\frac{16ei}{n}\right)^{di/20} \\
&\leq \sum_{i=1}^{0.0005n} \left(\frac{16ei}{n}\right)^{di/20} \\
&= O(1/n^{\Omega(d)}).
\end{aligned}$$

□

If a vertex in H is colored incorrectly at the end of iteration i of the algorithm in phase 2 (i.e. if it is colored j and does not belong to W_j), it must have more than $d/4$ neighbors in H colored incorrectly at the end of iteration $i-1$. To see this, observe that any vertex of H has at most $2(1.01d - 0.99d) = 0.04d$ neighbors outside H, and hence if it has at most $d/4$ wrongly colored neighbors in H, it must have at least $0.99d - d/4 > d/2$ neighbors of each color other than its correct color and at most $d/4 + 0.04d$ neighbors of its correct color. By repeatedly applying the property asserted by the above lemma with U being the set of vertices of H whose colors in the end of the iteration $i-1$ are incorrect, we deduce that the number of incorrectly colored vertices decreases by a factor of two (at least) in each iteration, implying that all vertices of H will be correctly colored after $\lceil \log_2 n \rceil$ iterations. This completes the proof of Proposition 2.4. We note that by being more careful one can show that $O(\log_d n)$ iterations suffice here, but since this only slightly decreases the running time we do not prove the stronger statement here. □

A standard probabilistic argument based on the Chernoff bound (see, for example, [2, Appendix A]) shows that $H = V$ almost surely if $p \geq \beta \log n/n$, where β is a suitably large constant. Thus, it follows from Proposition 2.4 that the algorithm almost surely properly colors the graph by the end of Phase 2 if $p \geq \beta \log n/n$.

For two sets of vertices X and Z, let $e(X, Z)$ denote the number of edges $(u, v) \in E$, with $u \in X$ and $v \in Z$.

Lemma 2.6 *There exists a constant $\gamma > 0$ such that almost surely the following holds.*
(i) For any two distinct color classes V_1 and V_2, and any subset X of V_1 and any subset Y of V_2, if $|X| = 2^{-\gamma d} n$ and $|Y| \leq 3|X|$, then $|e(X, V_2 - Y) - d|X|| \leq 0.001d|X|$.
(ii) If J is the set of vertices having more than $1.01d$ neighbors in G in some color class, then $|J| \leq 2^{-\gamma d} n$.

Proof For any subset X of V_1, $e(X, V_2 - Y)$ is the sum of independent Bernoulli variables. By standard Chernoff bounds, the probability that there exist two color classes V_1 and V_2, a subset X

of V_1 and a subset Y of V_2 such that $|X| = \epsilon n$ and $|Y| \leq 3|X|$ and $|e(X, V_2 - Y) - d|X|| > 0.001d|X|$ is at most

$$6\binom{n}{\epsilon n}\sum_{i=0}^{3\epsilon n}\binom{n}{i}\,2^{-\Omega(\epsilon nd)} = 2^{O(H(\epsilon))n}2^{-\Omega(\epsilon nd)}.$$

Therefore, (i) holds almost surely if γ is a sufficiently small constant. A similar reasoning applies to (ii). Therefore, both (i) and (ii) hold if γ is a sufficiently small constant. □

Lemma 2.7 *Almost surely, H has at least $(1 - 2^{-\Omega(d)})n$ vertices in every color class.*

Proof. It suffices to show that there are at most $7 \cdot 2^{-\gamma d}n$ vertices outside H. Assume for contradiction that this is not true. Recall that H is obtained by first deleting all the vertices in J, and then by a deletion process in which vertices with less than $0.99d$ neighbors in the other color classes of H are deleted repeatedly. By Lemma 2.6 $|J| \leq 2^{-\gamma d}n$ almost surely, and so at least $6 \cdot 2^{-\gamma d}n$ vertices have been deleted because they had less than $0.99d$ neighbors in H in some color class (other than their own.) Consider the first time during the deletion process where there exists a subset X of a color class V_i of cardinality $2^{-\gamma d}n$, and a $j \in \{1, 2, 3\} - \{i\}$ such that every vertex of X has been deleted because it had less than $0.99d$ neighbors in the remaining subset of V_j. Let Y be the set of vertices of V_j deleted so far. Then $|Y| \leq |J| + 2|X| \leq 3|X|$. Note that every vertex in X has less than $0.99d$ neighbors in $V_j - Y$. We therefore get a contradiction by applying Lemma 2.6 to (X, Y). □

2.3 The third phase

We need the following lemma, which is an immediate consequence of Lemma 2.5.

Lemma 2.8 *Almost surely, there exists no subset U of V of size at most $0.001n$ such that the graph induced on U has minimum degree at least $d/2$.*

Lemma 2.9 *Almost surely, by the end of the uncoloring procedure in Phase 3 of the algorithm, all vertices of H remain colored, and all colored vertices are properly colored, i.e. any vertex colored i belongs to W_i. (We assume, of course, that the numbering of the colors is chosen appropriately).*

Proof. By Proposition 2.4 almost surely all vertices of H are properly colored by the end of Phase 2. Since every vertex of H has at least $0.99d$ neighbors (in H) in each color class other than its own, all vertices of H remain colored. Moreover, if a vertex is wrongly colored at the end of the uncoloring procedure, then it has at least $d/2$ wrongly colored neighbors. Assume for contradiction that there exists a wrongly colored vertex at the end of the uncoloring procedure. Then the subgraph induced on the set of wrongly colored vertices has minimum degree at least $d/2$, and hence it must have at least $0.001n$ vertices by Lemma 2.8. But, since it does not intersect H, it has at most $2^{-\Omega(d)}n$ vertices by Lemma 2.7, leading to a contradiction. □

In order to complete the proof of correctness of the algorithm, it remains to show that almost surely every connected component of the graph induced on the set of uncolored vertices is of size at most $\log_3 n$. We prove this fact in the rest of this section. We note that it is easy to replace the term $\log_3 n$ by $O(\frac{\log_3 n}{d})$, but for our purposes the above estimate suffices. Note also that if $p = o(\log n/n)$ some of these components are actually components of the original graph G, as for such value of p the graph G is almost surely disconnected (and has many isolated vertices).

Lemma 2.10 *Let K be a graph, (V_1, V_2, V_3) a partition of the vertices of K into three disjoint subsets, i an integer, and L the set of vertices of K that remain after repeatedly deleting the vertices having less than i neighbors in V_1, V_2 or V_3. Then the set L does not depend on the order in which vertices are deleted.*

Proof Let L be the set of vertices that remain after a deletion process according to a given order. Consider a deletion process according to a different order. Since every vertex in L has at least i neighbors in $L \cap V_1$, $L \cap V_2$ and $L \cap V_3$, no vertex in L will be deleted in the second deletion process (otherwise, we get a contradiction by considering the first vertex in L deleted.) Therefore, the set of vertices that remain after the second deletion process contains L, and thus equals L by symmetry. $\square$

Lemma 2.10 implies that H does not depend on the order in which vertices are deleted.

Proposition 2.11 *Almost surely the largest connected component of the graph induced on $V - H$ has at most $\log_3 n$ vertices.*

Proof. Let T be a fixed tree on $\log_3 n$ vertices of V all of whose edges have their two endpoints in distinct color classes W_i, W_j, $1 \le i < j \le 3$. Our objective is to estimate the probability that G contains T as a subgraph that does not intersect H, and show that this probability is sufficiently small to ensure that almost surely the above will not occur for any T. This property would certainly hold if $V - H$ were a random subset of V of cardinality $2^{-\Omega(d)}n$. Indeed, if this were the case, the probability that G contains T as a subgraph that does not intersect H would be upper bounded by the probability $2^{-\Omega(d|T|)}$ that T is a subset of $V - H$ times the probability $(d/n)^{|T|-1}$ that T is a subgraph of G. This bound is sufficiently small for our needs. Although $V - H$ is *not* a random subset of V, we will be able to show a similar bound on the probability that G contains T as a subgraph that does not intersect H. To simplify the notation, we let T denote the set of edges of the tree. Let $V(T)$ be the set of vertices of T, and let I be the subset of all vertices $v \in V(T)$ whose degree in T is at most 4. Since T contains $|V(T)| - 1$ edges, $|I| \ge |V(T)|/2$. Let H' be the subset of V obtained by the following procedure, which resembles that of producing H (but depends on $V(T) - I$). First, set H' to be the set of vertices having at most $1.01d - 4$ neighbors in G in each color class V_i. Then delete from H' all vertices of $V(T) - I$. Then, repeatedly, delete any vertex in H' having less than $0.99d$ neighbors in H' in some color class (other than its own.)

Lemma 2.12 *Let F be a set of edges, each having endpoints in distinct color classes W_i, W_j. Let $H(F \cup T)$ be the set obtained by replacing E by $F \cup T$ in our definition of H, and $H'(F)$ be the set obtained by replacing E by F in our definition of H'. Then $H'(F) \subseteq H(F \cup T)$.*

Proof. First, we show that the initial value of $H'(F)$, i.e., that obtained after deleting the vertices with more than $1.01d - 4$ neighbors in a color class of G and after deleting the vertices in $V(T) - I$, is a subset of the initial value of $H(F \cup T)$. Indeed, let v be any vertex that does not belong to the initial value of $H(F \cup T)$, i.e. v has more than $1.01d$ neighbors in some color class of $(V, F \cup T)$. We distinguish two cases:

1. $v \in V(T) - I$. In this case, v does not belong to the initial value of $H'(F)$.
2. $v \notin V(T) - I$. Then v is incident with at most 4 edges of T, and so it has more than $1.01d - 4$ neighbors in some color class in (V, F).

In both cases, v does not belong to the initial value of $H'(F)$. This implies the assertion of the lemma, since the initial value of $H'(F)$ is a subgraph of the initial value of $H(F \cup T)$ and hence, by Lemma 2.10, any vertex which will be deleted in the deletion process for constructing H will be deleted in the corresponding deletion process for producing H' as well. $\Box$

Lemma 2.13

$$Pr[T \text{ is a subgraph of } G \text{ and } V(T) \cap H = \emptyset] \leq Pr[T \text{ is a subgraph of } G]\ Pr[I \cap H' = \emptyset].$$

Proof. It suffices to show that

$$\Pr[I \cap H = \emptyset |\ T \text{ is a subgraph of } G] \leq \Pr[I \cap H' = \emptyset].$$

But, by Lemma 2.12,

$$\begin{aligned}
\Pr[I \cap H' = \emptyset] &= \sum_{F: I \cap H'(F) = \emptyset} \Pr[E(G) = F] \\
&\geq \sum_{F: I \cap H(F \cup T) = \emptyset} \Pr[E(G) = F] \\
&= \sum_{F': F' \cap T = \emptyset, I \cap H(F' \cup T) = \emptyset} \Pr[E(G) - T = F'] \\
&= \sum_{F': F' \cap T = \emptyset, I \cap H(F' \cup T) = \emptyset} \Pr[E(G) - T = F' |\ T \text{ is a subgraph of } G] \\
&= \Pr[I \cap H = \emptyset |\ T \text{ is a subgraph of } G],
\end{aligned}$$

where F ranges over the sets of edges with endpoints in different color classes, and F' ranges over those sets that do not intersect T. The third equation follows by regrouping the edge-sets F according to $F' = F - T$ and noting (the obvious fact) that, for a given set F' that does not intersect T, the probability that $E(G) = F$ for some F such that $F - T = F'$ is equal to $\Pr[E(G) - T = F']$. The fourth equation follows from the independence of the events $E(G) - T = F'$ and T is a subgraph of G. $\Box$

Returning to the proof of Proposition 2.11 we first note that we can assume without loss of generality that $d \leq \beta \log n$, for some constant $\beta > 0$ (otherwise $H = V$.) If this inequality holds then, by modifying the arguments in the proof of Lemma 2.7, one can show that each of the graphs H' (corresponding to the various choices of $V(T) - I$) misses at most $2^{-\Omega(d)}n$ vertices in each color class, with probability at least $1 - 2^{-n^{\Theta(1)}}$. Since the distribution of H' depends only on $V(T) - I$ (assuming the W_i's are fixed), it is not difficult to show that this implies that $\Pr[I \cap H' = \emptyset]$ is at most $2^{-\Omega(d|I|)}$. Since $|I| \geq |V(T)|/2$ and since the probability that T is a subgraph of G is precisely $(d/n)^{|V(T)|-1}$ we conclude, by Lemma 2.13, that the probability that there exists some T of size $\log_3 n$ which is a connected component of the induced subgraph of G on $V - H$ is at most $2^{-\Omega(d \log_3 n)}(d/n)^{\log_3 n - 1}$ multiplied by the number of possible trees of this size, which is

$$\binom{3n}{\log_3 n} (\log_3 n)^{\log_3 n - 2}.$$

Therefore, the required probability is bounded by

$$\binom{3n}{\log_3 n}(\log_3 n)^{\log_3 n-2}2^{-\Omega(d\log_3 n)}(\frac{d}{n})^{\log_3 n-1}=O(1/n^{\Omega(d)}),$$

completing the proof. □

3 Bounding the eigenvalues

In this section, we prove Proposition 2.1. Let $G=(V,E),A,p,d,\lambda_i,e_i,W_1,W_2,W_3$ be as in Section 2. We start with the following lemma.

Lemma 3.1 *There exists a constant $\beta>0$ such that, almost surely, for any subset X of $2^{-\beta d}n$ vertices, $e(X,V)\le 5d|X|$.*

Proof As in the proof of Lemma 2.6, the probability that there exists a subset X of cardinality ϵn such that $e(X,V)>5d|X|$ is at most

$$\binom{3n}{\epsilon n}2^{-\Omega(d\epsilon n)}\le 2^{3H(\epsilon/3)n}2^{-\Omega(d\epsilon n)}=2^{-\Omega(d\epsilon n)},$$

if $\log(1/\epsilon)<d/b$, where b is a sufficiently large constant. Therefore, if β is a sufficiently small constant, this probability goes to 0 as n goes to infinity. □

Proof of Proposition 2.1

Parts (i) and (ii) are simple. By the variational definition of eigenvalues (see [19, p. 99]), λ_1 is simply the maximum of $x^tAx/(x^tx)$ where the maximum is taken over all nonzero vectors x. Therefore, by taking x to be the all 1 vector we obtain the well known result that λ_1 is at least the average degree of G'. By the known estimates for Binomial distributions, the average degree of G is $(1+o(1))2d$. On the other hand, Lemma 3.1 can be used to show that $|E-E'|\le 2^{-\Omega(d)}n$, as it easily implies that the number of vertices of degree greater than $5d$ in each color class of G is almost surely less than $2^{-\beta d}n$. Hence, the average degree of G' is at least $(1-2^{-\Omega(d)})2d$. This proves (i).

The proof of (ii) is similar. It is known [19, p. 101] that

$$\lambda_{3n-1}=min_F\ max_{x\in F,x\neq 0}\ \frac{x^tAx}{x^tx},$$

where the minimum is taken over all two dimensional subspaces F of R^{3n}. Let F denote the 2-dimensional subspace of all vectors $x=(x_v:v\in V)$ satisfying $x_v=\alpha_i$ for all $v\in W_i$, where $\alpha_1+\alpha_2+\alpha_3=0$. For x as above

$$x^tAx=2\alpha_1\alpha_2e'(W_1,W_2)+2\alpha_2\alpha_3e'(W_2,W_3)+2\alpha_1\alpha_3e'(W_1,W_3),$$

where $e'(W_i,W_j)$ denotes the number of edges of G' between W_i and W_j. Almost surely $e'(W_i,W_j)\ge(1-2^{-\Omega(d)})nd$ for all $1\le i<j\le 3$, and since $x^tx=n(\alpha_1^2+\alpha_2^2+\alpha_3^2)=-2n(\alpha_1\alpha_2+\alpha_2\alpha_3+\alpha_1\alpha_3)$ it follows that $x^tAx/(x^tx)\le-(1-2^{-\Omega(d)})d$ almost surely for all $x\in F$, implying that $\lambda_{3n}\le\lambda_{3n-1}\le -(1-2^{-\Omega(d)})d$, and establishing (ii).

The proof of (iii) is more complicated. Its assertion for somewhat bigger p (for example, for $p \geq \log^6 n/n$) can be deduced from the arguments of [10]. To prove it for the graph G' and $p \geq c/n$ we use the basic approach of Kahn and Szemerédi in [11], where the authors show that the second largest eigenvalue in absolute value of a random d-regular graph is almost surely $O(\sqrt{d})$. (See also [9] for a different proof.) Since in our case the graph is not regular a few modifications are needed. Our starting point is again the variational definition of the eigenvalues, from which we will deduce that it suffices to show that almost surely the following holds.

Lemma 3.2 *Let S be the set of all unit vectors $x = (x_v : v \in V)$ for which $\sum_{v \in W_j} x_v = 0$ for $j = 1, 2, 3$, then $|x^t Ax| \leq O(\sqrt{d})$ for all $x \in S$.*

The matrix A consists of nine blocks arising from the partition of its rows and columns according to the classes W_j. It is clearly sufficient to show that the contribution of each block to the sum $x^t Ax$ is bounded, in absolute value, by $O(\sqrt{d})$. This, together with a simple argument based on ϵ-nets (see [11], Proposition 2.1) can be used to show that Lemma 3.2 follows from the following statement.

Fix $\epsilon > 0$, say $\epsilon = 1/2$, and let T denote the set of all vectors x of length n every coordinate of which is an integral multiple of $\epsilon/\sqrt{n}$, where the sum of coordinates is zero and the l_2-norm is at most 1. Let B be a random n by n matrix with $0, 1$ entries, where each entry of B, randomly and independently, is 1 with probability d/n.

Lemma 3.3 *If d exceeds a sufficiently large absolute constant then almost surely, $|x^t By| \leq O(\sqrt{d})$ for every $x, y \in T$ for which $x_u = 0$ if the corresponding row of B has more than $5d$ nonzero entries and $y_v = 0$ if the corresponding column of B has more than $5d$ nonzero entries.*

The last lemma is proved, as in [11], by separately bounding the contribution of terms $x_u y_v$ with small absolute values and the contribution of similar terms with large absolute values. Here is a description of the details that differ from those that appear in [11]. Let C denote the set of all pairs (u, v) with $|x_u y_v| \leq \sqrt{d}/n$ and let $X = \sum_{(u,v) \in C} x_u B(u, v) y_v$. As in [11] one can show that the absolute value of the expectation of X is at most $\sqrt{d}$. Next one has to show that with high probability X does not deviate from its expectation by more than $c\sqrt{d}$. This is different (and in fact, somewhat easier) than the corresponding result in [11], since here we are dealing with independent random choices. It is convenient to use the following variant of the Chernoff bound.

Lemma 3.4 *Let $a_1, \ldots, a_m$ be (not necessarily positive) reals, and let Z be the random variable $Z = \sum_{i=1}^m \epsilon_i a_i$, where each ϵ_i is chosen, randomly and independently, to be 1 with probability p and 0 with probability $1 - p$. Suppose $\sum_{i=1}^m a_i^2 \leq D$ and suppose $|Sa_i| \leq ce^c pD$ for some positive constants c, S. Then $Pr[|Z - E(Z)| > S] \leq 2e^{-S^2/(2pe^c D)}$.*

For the proof, one first proves the following.

Lemma 3.5 *Let c be a positive real. Then for every $x \leq c$,*

$$e^x \leq 1 + x + \frac{e^c}{2} x^2.$$

Proof. Define $f(x) = 1 + x + \frac{e^c}{2} x^2 - e^x$. Then $f(0) = 0$, $f'(x) = 1 + e^c x - e^x$ and $f''(x) = e^c - e^x$. Therefore, $f''(x) \geq 0$ for all $x \leq c$ and as $f'(0) = 0$ this shows that $f'(x) \leq 0$ for $x < 0$ and

$f'(x) \geq 0$ for $c \geq x > 0$, implying that $f(x)$ is nonincreasing for $x \leq 0$ and nondecreasing for $c \geq x \geq 0$. Thus $f(x) \geq 0$ for all $x \leq c$, as needed. $\Box$

Proof of Lemma 3.4. Define $\lambda = \frac{S}{e^c pD}$, then, by assumption, $\lambda a_i \leq c$ for all i. Therefore, by the above lemma,

$$\begin{aligned}
E(e^{\lambda Z}) &= \prod_{i=1}^{m}[pe^{\lambda a_i} + (1-p)] \\
&= \prod_{i=1}^{m}[1 + p(e^{\lambda a_i} - 1)] \\
&\leq \prod_{i=1}^{m}[1 + p(\lambda a_i + \frac{e^c}{2}\lambda^2 a_i^2)] \\
&\leq \exp(p\lambda \sum_{i=1}^{m} a_i + p\frac{e^c}{2}\lambda^2 \sum_{i=1}^{m} a_i^2) \\
&\leq e^{\lambda E(Z) + \frac{pe^c}{2}\lambda^2 D}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr\left[(Z - E(Z)) > S\right] &= \Pr\left[e^{\lambda(Z-E(Z))} > e^{\lambda S}\right] \\
&\leq e^{-\lambda S} E(e^{\lambda(Z-E(Z))}) \\
&\leq e^{-\lambda S + \frac{pe^c}{2}\lambda^2 D} \\
&= e^{-\frac{S^2}{2pe^c D}}.
\end{aligned}$$

Applying the same argument to the random variable defined with respect to the reals $-a_i$, the assertion of the lemma follows. $\Box$

Using Lemma 3.4 it is not difficult to deduce that almost surely the contribution of the pairs in C to $|x^t By|$ is $O(\sqrt{d})$. This is because we can simply apply the lemma with $m = |C|$, with the a_i's being all the terms $x_u y_v$ where $(u,v) \in C$, with $p = d/n$, $D = 1$ and $S = ce^c\sqrt{d}$ for some $c > 0$. Since here $|Sa_i| \leq ce^c\sqrt{d}\sqrt{d}/n = ce^c pD$, we conclude that for every fixed vectors x and y in T, the probability that X deviates from its expectation (which is $O(\sqrt{d})$) by more than $ce^c\sqrt{d}$ is smaller than $2e^{-c^2 e^c n/2}$, and since the cardinality of T is only b^n for some absolute constant $b = b(\epsilon)$, one can choose c so that X would almost surely not deviate from its expectation by more than $ce^c\sqrt{d}$.

The contribution of the terms $x_u y_v$ whose absolute values exceed $\sqrt{d}/n$ can be bounded by following the arguments of [11], with a minor modification arising from the fact that the maximum number of ones in a row (or column) of B can exceed d (but can never exceed $5d$ in a row or a column in which the corresponding coordinates x_u or y_v are nonzero). We sketch the argument below. We start with the following lemma.

Lemma 3.6 *There exists a constant C such that, with high probability, for any distinct color classes V_1, V_2, and any subset U of V_1 and any subset W of V_2 such that $|U| \leq |W|$, at least one of the following two conditions hold:*

1. $e'(U,W) \leq 10\mu(U,W)$

2. $e'(U,W)\log(e'(U,W)/\mu(U,W)) \le C|W|\log(n/|W|)$*,*

where $e'(U,W)$ *is the number of edges in* G' *between* U *and* W*, and* $\mu(U,W) = |U|\ |W|d/n$ *is the expected number of edges in* G *between* U *and* W*.*

Proof. Condition 1 is clearly satisfied if $|W| \ge n/2$, since the maximum degree in G' is at most $5d$. So we can assume without loss of generality that $0 < |W| \le n/2$. Give two subsets U and W satisfying the requirements of the lemma, define $\beta = \beta(|U|,|W|)$ to be the unique positive real number such that $\beta\mu(U,W)\log\beta = C|W|\log(n/|W|)$ (the constant C will be determined later.) Condition 2 is equivalent to $e'(U,W) \le \beta\mu(U,W)$. Thus U, W violate Condition 1 as well as Condition 2 only if $e'(U,W) > \beta'\mu(U,W)$, where $\beta' = \max(10,\beta)$. Hence, by standard Chernoff bounds, the probability of this event is at most $e^{-\gamma\beta'\mu(U,W)\log\beta'} \le (|W|/n)^{\gamma C|W|}$, for some absolute constant $\gamma > 0$. Denoting $|W|/n$ by b, the probability that there exist two subsets U and W that do not satisfy either condition is at most

$$\begin{aligned} 6 \sum_{b:\ bn \text{ integer } \le n/2} \binom{n}{bn}^2 b^{\gamma Cbn} &= 6 \sum_{b:\ bn \text{ integer } \le n/2} 2^{O(b\log(1/b)n)} b^{\gamma Cbn} \\ &= n^{-\Omega(1)}, \end{aligned}$$

if C is a sufficiently large constant. □

Kahn and Szemerédi [11] show that for any d-regular graph satisfying the conditions of Lemma 3.6 (without restriction on the ranges of U and W), the contribution of the terms $x_u y_v$ whose absolute values exceed $\sqrt{d}/n$ is $O(\sqrt{d})$. Up to replacing some occurences of d by $5d$, the same proof shows that, for any 3-colorable graph of maximum degree $5d$ satisfying the conditions of Lemma 3.6, the contribution of the terms $x_u y_v$ whose absolute values exceed $\sqrt{d}/n$ is $O(\sqrt{d})$. This implies the assertion of Lemma 3.3, which implies Lemma 3.2.

To deduce (iii), we need the following lemma.

Lemma 3.7 *Let* F *denote, as before, the 2-dimensional subspace of all vectors* $x = (x_v : v \in V)$ *satisfying* $x_v = \alpha_i$ *for all* $v \in W_i$*, where* $\alpha_1 + \alpha_2 + \alpha_3 = 0$*. Then, almost surely, for all* $f \in F$ *we have* $||(A + dI)f||^2 = O(d||f||^2)$*.*

Proof. Let x, y be as in the proof of Lemma 2.3. Note that $x^t y = 0$, and that both $||x||^2$ and $||y||^2$ are $\Theta(n)$. Thus every vector $f \in F$ can be expressed as the sum of two orthogonal vectors x' and y' proportional to x and y respectively. Lemma 2.3 shows that $||(A + dI)y'||^2 = O(d||y'||^2)$, which implies that $||(A + dI)y'||^2 = O(d||f||^2)$, since $||f||^2 = ||x'||^2 + ||y'||^2$. Similarly, it can be shown that $||(A + dI)x'||^2 = O(d||f||^2)$. We conclude the proof of the lemma using the triangle inequality $||(A + dI)f|| \le ||(A + dI)x'|| + ||(A + dI)y'||$. □

We now show that $\lambda_2 \le O(\sqrt{d})$ by using the formula $\lambda_2 = \min_H \max_{x\in H, x\ne 0} x^t Ax/(x^t x)$, where H ranges over the linear subspaces of R^{3n} of codimension 1. Indeed, let H be the set of vectors whose sum of coordinates is 0. Any $x \in H$ is of the form $f + s$, where $f \in F$ and s is a multiple of a vector in S, and so

$$\begin{aligned} x^t Ax &= f^t Af + 2s^t Af + s^t As \\ &= f^t Af + 2s^t (A + dI)f + s^t As \\ &\le -(1 - 2^{-\Omega(d)})d||f||^2 + 2||s||\ ||(A + dI)f|| + O(\sqrt{d})||s||^2 \end{aligned}$$

Number of vertices	d
1000	12
10000	10
100000	8

Figure 1: Implementation results.

$$\begin{aligned} &\leq O(\sqrt{d}||s||||f||) + O(\sqrt{d})||s||^2 \\ &\leq O(\sqrt{d}(||s||^2 + ||f||^2)) \\ &= O(\sqrt{d}||x||^2). \end{aligned}$$

This implies the desired upper bound on λ_2.

The bound $|\lambda_{3n-2}| \leq O(\sqrt{d})$ can be deduced from similar arguments, namely by showing that $x^t Ax/(x^t x) \geq -\Theta(\sqrt{d})$, for any $x \in F^{\perp}$. This completes the proof of Proposition 2.1. □

4 Implementation and Experimental Results.

We have implemented the following tuned version of our algorithm. The first two phases are as described in Section 1. In the third phase, we find the minimum i such that, after repeatedly uncoloring every vertex colored j that has less than i neighbors colored l, for some $l \in \{1,2,3\} - \{j\}$, the algorithm can find a proper coloring using brute force exhaustive search on every component of uncolored vertices. If the brute force search takes more steps than the first phase (up to a multiplicative constant), the algorithm fails. Otherwise, it outputs a legal coloring. The eigenvectors e_{3n} and e_{3n-1} are calculated approximately using an iterative procedure. The coordinates of the initial vectors are independent random variables uniformly chosen in $[0,1]$.

The range of values of p where the algorithm succeeded was in fact considerably larger than what our analysis predicts. Figure 1 shows some values of the parameters for which we tested our algorithm. For each of these parameters, the algorithm was run on more than a hundred graphs drawn from the corresponding distribution, and found successfully a proper coloring for all these tests. The running time was a few minutes on a Sun SPARCstation 2 for the largest graphs. The algorithm failed for some graphs drawn from distributions with smaller integral values of d than the one in the corresponding row. Note that the number of vertices is not a multiple of 3; the size of one color class exceeds the others by one.

5 Concluding remarks

1. There are many heuristic graph algorithms based on spectral techniques, but very few rigorous proofs of correctness for any of those in a reasonable model of random graphs. Our main result here provides such an example. Another example is the algorithm of Boppana [7], who designed an algorithm for graph bisection based on eigenvalues, and showed that it finds the best bisection almost surely in an appropriately defined model of random graphs with a relatively small bisection width. Aspvall and Gilbert [1] gave a heuristic for graph coloring

based on eigenvectors of the adjacency matrix, and showed that their heuristic optimally colors complete 3-partite graphs as well as certain other classes of graphs with regular structure.

2. By modifying some of the arguments of Section 2 we can show that if p is somewhat bigger ($p \geq \log^3 n/n$ suffices) then almost surely the initial coloring V_i^0 that is computed from the eigenvectors e_{3n-1} and e_{3n} in the first phase of our algorithm is completely correct. In this case the last two phases of the algorithm are not needed. By refining the argument in Subsection 2.2, it can also be shown that if $p > 10 \log n/n$ the third phase of the algorithm is not needed, and the coloring obtained by the end of the second phase will almost surely be the correct one.

3. We can show that a variant of our algorithm finds, almost surely, a proper coloring in the model of random *regular* 3-colorable graphs in which one chooses randomly d perfect matchings between each pair of distinct color classes, when d is a sufficiently large absolute constant. Here, in fact, the proof is simpler, as the smallest two eigenvalues (and their corresponding eigenspace) are known precisely, as noted in Subsection 1.2.

4. The results easily extend to the model in which each vertex first picks a color randomly, independently and uniformly, among the three possibilities, and next every pair of vertices of distinct colors becomes an edge with probability p ($> c/n$).

5. If $G = G_{3n,p,3}$ and $p \leq c/n$ for some small positive constant c, it is not difficult to show that almost surely G does not have any subgraph with minimum degree at least 3, and hence it is easy to 3-color it by a greedy-type (linear time) algorithm. For values of p which are bigger than this c/n but satisfy $p = o(\log n/n)$, the graph G is almost surely disconnected, and has a unique component of $\Omega(n)$ vertices, which is called the *giant component* in the study of random graphs (see, e.g., [2], [4]). All other components are almost surely sparse, i.e., contain no subgraph with minimum degree at least 3, and can thus be easily colored in total linear time. Our approach here suffices to find, almost surely, a proper 3-coloring of the giant component (and hence of the whole graph) for all $p \geq c/n$, where c is a sufficiently large absolute constant, and there are possible modifications of it that may even work for all values of p. At the moment, however, we are unable to obtain an algorithm that provably works for all values of p almost surely. Note that, for any constant c, if $p < c/n$ then the greedy algorithm will almost surely color $G_{3n,p,3}$ with a constant number of colors. Thus, our result implies that $G_{3n,p,3}$ can be almost surely colored in polynomial time with a constant number of colors for *all* values of p.

6. Our basic approach easily extends to k-colorable graphs, for every fixed k, as follows. Phase 2 and Phase 3 of the algorithm are essentially the same as in the case $k = 3$. Phase 1 needs to be modified to extract an approximation of the coloring. Let e_i, $i \geq 1$, be an eigenvector of G' corresponding to its ith largest eigenvalue (replace $5d$ by $5kd$ in the definition of G'.) Find vectors $x_1, x_2, \ldots, x_{k+1}$ of norm $\sqrt{kn}$ in $\mathrm{Span}(e_1, e_n, e_{n-1}, \ldots, e_{n-k+2})$ such that $x_1 = \sqrt{kn}e_1$ and $(x_i, x_j) = -n$ for $1 \leq i < j \leq k+1$. For $2 \leq i \leq n$, and any z, let W_ϵ be the set of vertices whose coordinates in x_i are in $(z - \epsilon, z + \epsilon)$. If, for some i and z, both $|W_{\epsilon_k}|$ and $|W_{\epsilon_k/2}|$ deviate from n by at most $\beta_k n/d$, where ϵ_k and β_k are constants depending on k, color the elements in W_{ϵ_k} with a new color and delete them from the graph. Repeat this process

until the number of vertices left is $O(n/d)$, and color the remaining vertices in an arbitrary manner.

7. The existence of an approximation algorithm based on the spectral method for coloring *arbitrary* graphs is a question that deserves further investigation (which we do *not* address here.) Recently, improved approximation algorithms for graph coloring have been obtained using semidefinite programming [14], [5].

Acknowledgement We thank two anonymous referees for several suggestions that improved the presentation of the paper.

References

[1] B. Aspvall and J. R. Gilbert, *Graph coloring using eigenvalue decomposition*, SIAM J. Alg. Disc. Meth., Vol. 5 (1984), 526-538.

[2] N. Alon and J. H. Spencer, **The Probabilistic Method**, Wiley, 1991.

[3] A. Blum, *Some tools for approximate 3-coloring*, Proc. 31^{st} IEEE FOCS, IEEE (1990), 554–562.

[4] B. Bollobás, **Random Graphs**, Academic Press, 1985.

[5] A. Blum and D. Karger, personal communication.

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

[7] R. Boppana, *Eigenvalues and graph bisection: An average case analysis*, Proc. 28^{th} IEEE FOCS, IEEE (1987), 280–285.

[8] M. E. Dyer and A. M. Frieze, *The solution of some random NP-Hard problems in polynomial expected time*, Journal of Algorithms 10 (1989), 451–489.

[9] J. Friedman, *On the second eigenvalue and random walks in random d-regular graphs*, Combinatorica 11 (1991), 331–362.

[10] Z. Füredi and J. Komlós,*The eigenvalues of random symmetric matrices*, Combinatorica 1(1981), 233–241.

[11] J. Friedman, J. Kahn and E. Szemerédi, *On the second eigenvalue in random regular graphs*, Proc. 21^{st} ACM STOC, ACM Press (1989), 587–598.

[12] M. R. Garey and D. S. Johnson, **Computers and intractability: a guide to the theory of NP–completeness**, Freeman and Company, 1979.

[13] R. Karp, **Reducibility among combinatorial problems**, Miller and Thatcher (eds), Plenum Press, 1972.

[14] D. Karger, R. Motwani and M. Sudan, *Approximate graph coloring by semidefinite programming*, Proc. 35^{th} IEEE FOCS, IEEE (1994), 2-13.

[15] L. Kucera, *Expected behavior of graph colouring algorithms*, In *Lecture Notes in Computer Science No. 56*, Springer-Verlag, 1977, pp. 447–451.

[16] A. D. Petford and D. J. A. Welsh, *A randomised 3-colouring algorithm*, Discrete Mathematics, 74 (1989), 253–261.

[17] A. Ralston, **A First Course in Numerical Analysis**, McGraw-Hill, 1985, Section 10.4.

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

[19] J. H. Wilkinson, **The Algebraic Eigenvalue Problem**, Clarendon Press, 1965.