Coloring Random 3-Colorable Graphs with Non-uniform Edge Probabilities[*]

Ulrik Brandes and Jürgen Lerner[**]

Department of Computer & Information Science, University of Konstanz
`lerner@inf.uni-konstanz.de`

Abstract. Random 3-colorable graphs that are generated according to a $G(n, p)$-like model can be colored optimally, if $p \geq c/n$ for some large constant c. However, these methods fail in a model where the edge-probabilities are non-uniform and not bounded away from zero. We present a spectral algorithm that succeeds in such situations.

1 Introduction

Graph coloring [9] is one of the central problems in graph theory and combinatorics. A *(proper) graph coloring* is the assignment of colors to vertices so that adjacent vertices are always colored differently. The problem of coloring graphs with the minimum number of colors is of large theoretical interest. Furthermore, efficient coloring algorithms are important for applications, as many practical problems can be formulated as graph coloring problems. However, even if it is known that a graph G is k-colorable, it is $\mathcal{NP}$-hard to properly color G with k colors, for any fixed $k \geq 3$ [6].

Much research has focused on k-coloring random k-colorable graphs with high probability [11,4,15,1,5,12], see [10] for a survey on random graph coloring. (We say that an algorithm succeeds *with high probability (w. h. p.)* if its failure probability tends to zero as the input size tends to infinity.) There are several models for random k-colorable graphs, all of which have the property in common, that every possible edge (i. e., every pair of differently colored vertices) is included in a sampled graph with non-zero probability.

In this paper we propose a more general model for 3-colorable graphs, where there is no lower bound on the edge probabilities. We show that the algorithms from [1,12] can not color graphs from this model and present a more general spectral algorithm that can cope with these distributions. The assumptions that we need for our algorithm are simultaneously more restrictive and more general than those for known algorithms. Thus, we provide an alternative description for random graphs that are easy to color. We believe that our ideas have similar implications for other spectral algorithms (e. g., [12]) that recover "hidden" combinatorial objects (like cliques, independent sets, or minimum cuts) in random graphs.

Our paper is organized as follows. In Sect. 2, we review known models for random 3-colorable graphs, propose a generalization of these models, and present our algorithm and the assumptions under which it succeeds. Section 3 presents general observations regarding planted partitions and spectral algorithms. The ideas developed there are used in Sect. 4 to prove our main theorem. In Sect. 5, we provide the proofs that traditional methods do not work on our generalized model and Sect. 6 outlines implications for related problems.

Notation. If A is a matrix, then A_{uv} denotes the uv'th entry of A. In this paper, the rows and columns of matrices are often understood as being indexed by the vertices of a graph. We frequently use the notation $A(v)$ to denote the column of A that is indexed by the vertex v. The *transpose* of a matrix A is denoted by A^{T} and is defined by $A_{uv}^{\mathrm{T}} = A_{vu}$.

For vectors $v \in \mathbb{R}^n$ we use the *Euclidean norm*, defined by $\|v\|^2 = \sum_{i=1}^{n} v_i^2$. The *distance* between two vectors u and v is $\|u - v\|$. We recall the definition of two matrix norms: the *2-norm*

$$\|A\|_2 = \max_{\|v\|=1} \|A(v)\| \ ,$$

and the *Frobenius norm* $\|A\|_F^2 = \sum_{u,v=1}^{n} A_{uv}^2$. More background on linear algebra is in [13,8].

2 Background and Results

2.1 Previous Models and Known Results

We review first two random graph models for 3-colorable graphs. Let r be a positive integer and p a real number, $0 \le p \le 1$. The random graph model $\mathcal{G}(r,p,3)$ is a probability distribution for graphs on $n = 3r$ vertices, partitioned into three color classes of size r. The edges between vertices from different color classes are included independently with probability p. The best result for this model (i. e., the algorithm that works for the smallest non-trivial p) is from Alon and Kahale [1], who gave a spectral algorithm that (w. h. p.) 3-colors graphs from $\mathcal{G}(r,p,3)$, if $p \ge c/n$, for a sufficiently large constant c. McSherry [12] described a different spectral algorithm for a more general problem that (w. h. p.) 3-colors graphs from $\mathcal{G}(r,p,3)$, if $p \ge c \log^3(n)/n$.

It has been pointed out (compare [14]), that random graphs from $\mathcal{G}(r,p,3)$ have very special properties that graphs encountered in applications usually do not have. It is more demanding to design algorithms for graph models that mediate between the uniformly structured graphs from $\mathcal{G}(r,p,3)$ and worst-case instances. One possibility to generate such graphs are the so-called *semi-random* graph models. In the semi-random model $\mathcal{G}_S(r,p,3)$, first a "true random" graph is drawn from $\mathcal{G}(r,p,3)$, then an adversary can decide to introduce additional edges between vertices from different color classes. While, at a first glance, it seems to help an algorithm if more bi-colored edges are introduced, this is not the

case. The fact that the random structure of the graph from $\mathcal{G}(r, p, 3)$ is spoiled counts more than the benefit from the additional edges. Feige and Kilian [5] showed that there is a polynomial time algorithm that optimally colors almost all graphs from $\mathcal{G}_S(r, p, 3)$ if p is as large as $p \geq (1+\varepsilon)3 \log n/n$, for every $\varepsilon > 0$. The algorithm from [5] is not based on spectral methods. Instead it uses semidefinite programming, followed by several sophisticated post-processing steps.

It should be noted that graphs from $\mathcal{G}_S(r, p, 3)$ have a substantial subgraph that is a random graph from $\mathcal{G}(r, p, 3)$. The adversary is only allowed to add more edges and cannot force any pair of differently colored vertices to be non-adjacent. In this paper, we consider a different probability distribution, where there is no lower bound on the edge-probabilities.

2.2   A Generalization of $\mathcal{G}(r, p, 3)$

To generalize the random graph model $\mathcal{G}(r, p, 3)$, consider the matrix $A^{(p)}$ in the lefthand-side of (1). (The nine blocks of $A^{(p)}$ are understood as being constant $r \times r$ blocks.) It is easy to see that the sampling process from $\mathcal{G}(r, p, 3)$ can be described as follows: construct a graph on $n = 3r$ vertices, where an edge $\{u, v\}$ is introduced with probability equal to $A_{uv}^{(p)}$. The matrix $A^{(p)}$ is the expected adjacency matrix of the distribution $\mathcal{G}(r, p, 3)$.

$$
A^{(p)} = \begin{bmatrix}
0 \cdots 0 & p \cdots p & p \cdots p \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
0 \cdots 0 & p \cdots p & p \cdots p \\
p \cdots p & 0 \cdots 0 & p \cdots p \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
p \cdots p & 0 \cdots 0 & p \cdots p \\
p \cdots p & p \cdots p & 0 \cdots 0 \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
p \cdots p & p \cdots p & 0 \cdots 0
\end{bmatrix}
\qquad
A[XYZ] = \begin{bmatrix}
[0] & X & Y \\
X^{\mathrm{T}} & [0] & Z \\
Y^{\mathrm{T}} & Z^{\mathrm{T}} & [0]
\end{bmatrix}
\qquad (1)
$$

The fact that the diagonal blocks of $A^{(p)}$ are zero ensures that graphs from $\mathcal{G}(r, p, 3)$ are 3-colorable. The off-diagonal blocks of $A^{(p)}$ describe the expected adjacency structure between two different color classes. In $\mathcal{G}(r, p, 3)$ this structure is uniform. Every vertex has the same probability to connect to every other differently colored vertex. We generalize the model $\mathcal{G}(r, p, 3)$ by allowing arbitrary adjacency structure between different color classes.

Definition 1. *Let r be an integer and $n = 3r$. Further, let X, Y, and Z, be arbitrary real $r \times r$ matrices whose entries are between zero and one and let $A = A[XYZ]$ be defined as in the righthand-side of (1). A graph drawn from the probability distribution $\mathcal{G}(A)$ is a graph on n vertices, where a set of two vertices $\{u, v\}$ is independently chosen to be an edge with probability A_{uv}. The matrix A is the expected adjacency matrix for the distribution $\mathcal{G}(A)$.*

As an example, the distribution $\mathcal{G}(r, p, 3)$ is equivalent to $\mathcal{G}(A^{(p)})$.

The restrictions on the form of A in Def. 1 ensure only that every graph drawn from $\mathcal{G}(A)$ admits a proper 3-coloring whose color classes are all of size r. In particular, the problem of 3-coloring graphs from $\mathcal{G}(A)$ includes the problem of 3-coloring graphs that are 3-colorable with equally sized color classes. Since this problem is $\mathcal{NP}$-complete in general we cannot hope to develop an algorithm that works for all A.

The distribution $\mathcal{G}(A)$ is obviously much more general than $\mathcal{G}(r, p, 3)$. It is simultaneously more restrictive and more general than the semi-random model $\mathcal{G}_S(r, p, 3)$: In $\mathcal{G}(A)$ we do not allow for an adversary to add edges to a sampled graph. On the other hand, in $\mathcal{G}_S(r, p, 3)$, each possible (bi-colored) edge is included with probability *at least* p (independent on the adversary's decisions), whereas in $\mathcal{G}(A)$, the structure of A can force large sets of possible edges to be not included. Thus, the model $\mathcal{G}(A)$ is an alternative possibility to mediate between the uniformly structured graphs from $\mathcal{G}(r, p, 3)$ and worst-case instances.

2.3   Main Results

Our main contribution is a spectral algorithm that 3-colors (w. h. p.) graphs from $\mathcal{G}(A)$ if the matrices X, Y, and Z are regular with a common degree and if the spectral properties of A "do not obfuscate" the planted 3-coloring. In particular, our algorithm succeeds for many matrices X, Y, and Z for which the algorithms from [1,12] do not work. The algorithm is presented below and gets the $n \times n$ adjacency matrix $\hat{A}$ of a sampled graph as input.

`Spectral 3-Coloring Algorithm`$(\hat{A})$

1. Compute $d = \sum_{i,j=1}^{n} \hat{A}_{ij}/(2n)$.
2. Compute (orthonormalized) eigenvectors $\{v_1, v_2, v_3\}$ of $\hat{A}$ associated to those eigenvalues that have the smallest distance to $2d$, $-d$ and $-d$.
3. Let P be the $3 \times n$ matrix whose rows are the v_i and compute $\hat{S} = P^{\mathrm{T}} P$.
4. Compute the pairwise distances of vertices according to the distance between their columns in $\hat{S}$.
5. Successively join vertices with the smallest distance until three color classes are left.

See, e. g., [8] for the efficient computation of eigenvectors. Of course, the computed 3-coloring is not necessarily proper for $\hat{A}$. In the following we clarify the assumptions under which the above algorithm succeeds with high probability.

A matrix is called *regular* (of *degree d*) if the sum of every row and column is equal to d. The first assumption we need for our algorithm is that X, Y, and Z must be regular with a common degree.

We turn our attention to the spectral properties of A: If X, Y, and Z are regular of degree d, then (see Theorem 6) A has the three eigenvalues

$$\lambda_{i_1} = 2d, \text{ and } \lambda_{i_2} = \lambda_{i_3} = -d \qquad (i_2 \neq i_3) \ . \tag{2}$$

It is crucial for our algorithm that the other eigenvalues of A are separated from those specific eigenvalues. The *separation* $\mathrm{sep}_3(A)$ of the planted 3-coloring in A is defined to be the minimal distance from λ_{i_1}, λ_{i_2}, and λ_{i_3} to any other eigenvalue of A. We define the *variance* of the distribution $\mathcal{G}(A)$ to be $\sigma^2 = \max_{u,v}(A_{uv} - A_{uv}^2)$ (i. e., the maximal variance of individual entries). The variance is bounded by $1/4$ and goes to zero if all entries of A go either to zero or to one. For instance, for the distribution $\mathcal{G}(r,p,3)$, if $p = c/n$ (that is the smallest p for which the algorithm from [1] is guaranteed to work), then the variance decreases linearly in n, i. e., σ^2 is in $\mathcal{O}(1/n)$. Our main result is the following.

Theorem 1. *Let X, Y, and Z be regular matrices of degree d and $A = A[XYZ]$. Let σ^2 be the variance of $\mathcal{G}(A)$ and assume that $\mathrm{sep}_3(A)$ is in $\omega(n\sigma)$ and that $\sigma^2 \gg (\log^6 n)/n$. Then, the* Spectral 3-Coloring Algorithm *properly 3-colors graphs from $\mathcal{G}(A)$, with high probability.*

Theorem 1 is proved in Sect. 4.

As a corollary, we get an algorithm that 3-colors (with probability one) a given graph with adjacency matrix $A[XYZ]$ assumed that X, Y, and Z are regular of common degree and that the separation of the planted 3-coloring is not zero.

Corollary 1. *Let X, Y, and Z be regular $\{0,1\}$ matrices of degree d and assume that $\mathrm{sep}_3(A[XYZ]) \neq 0$. Then, the* Spectral 3-Coloring Algorithm *properly 3-colors the graph G with adjacency matrix $A[XYZ]$.*

Interpretation of assumptions. To interpret the assumptions that we make in our theorems, we note first that also the traditional model $\mathcal{G}(r,p,3)$ implicitly makes assumptions on both, the regularity of the block matrices and the separation of certain eigenvalues. The specific form of the matrix $A^{(p)}$ in (1) ensures in particular that the submatrices are regular of degree $d = rp$. In this paper the property of being constant is relaxed to that of being regular. From the point of view of the coloring this means that vertices are no longer required to have the same probability to connect to all vertices of different color, but they are only required to have the same expected number of neighbors of each different color.

Turning to the assumption on the separation, we remark that the specific form of the expected adjacency matrix $A^{(p)}$ implies that $A^{(p)}$ has the three eigenvalues $2d$, $-d$ and $-d$, whereas all other eigenvalues are zero and thus well-separated from the aforementioned. Both previous results [1,12] use this observation and the fact that the random deviation from the expected adjacency matrix has w. h. p. eigenvalues in $\mathcal{O}(\sqrt{d})$. Thus, an assumption on the separation of eigenvalues is also made when assuming that graphs are drawn from the standard model $\mathcal{G}(r,p,3)$. We note however that the assumption on the separation that we make in our paper in not competitive to that of [1,12] when applied to the specific model $\mathcal{G}(r,p,3)$. Currently it is unclear whether the post-processing steps of [1] could be adapted to the more general model $\mathcal{G}(A)$. Similarly, [12] uses the fact that vertices that are in the same color class have identical columns in $A^{(p)}$. Since this is no longer true for our model, it is unclear whether the same bounds could be derived.

2.4 Insufficiency of Traditional Methods

We show here that our algorithm can solve many instances that can not be handled by previous spectral algorithms, e. g., [1,12]. The proofs of the following two Theorems are deferred to Sect. 5.

Theorem 2. *For arbitrary large r, there are $r \times r$ matrices X, Y, and Z such that graphs from $\mathcal{G}(A[XYZ])$ are not colored properly by the algorithms in [1] and [12], but the* `Spectral 3-Coloring Algorithm` *succeeds on these graphs.*

The matrices A that appear in the proof of Theorem 2 also seem to yield instances of $\mathcal{G}(A)$ for which the algorithm from [5], which is designed for the semi-random graph model, does not work. Since this algorithm consists of many randomized sub-procedures, it is more complicated to provide an example of $\mathcal{G}(A)$ for which it fails surely (or with high probability). However, we can show that the proofs in [5] do not generalize to $\mathcal{G}(A)$: A graph G (with a planted coloring) is said to have the *k-collision property*, if for every set U of equally colored vertices and every set T of vertices that are colored differently than those of U, such that $|T|, |U| \geq k$, there is an edge in G joining U and T. It is proved in [5] (by translating Lemma 6 of [5] to the graph coloring problem, compare Section 3 of [5]), that semi-random graphs G have with high probability the k-collision property for $k = \frac{2n \log \log n}{c \log n}$. This does not hold for $\mathcal{G}(A)$. In particular the proofs in [5] do not generalize to $\mathcal{G}(A)$.

Theorem 3. *For arbitrary large r, there are $r \times r$ matrices X, Y, and Z, such that graphs from $\mathcal{G}(A[XYZ])$ do not have the k-collision property for any k that is in $o(n)$, but are properly colored by the* `Spectral 3-Coloring Algorithm`*.*

3   Methodology

The expected adjacency matrix $A^{(p)}$ for the distribution $\mathcal{G}(r, p, 3)$ (see (1)) is so convenient for traditional spectral algorithms since vertices from the same color class have identical columns (neighborhoods) in $A^{(p)}$. Therefore, projecting vertices to the column space of $A^{(p)}$ (that space is spanned by those three eigenvectors that have non-zero eigenvalues) trivially reveals the classes. Moreover, this spectral projection is stable to random noise (given certain assumptions) and the algorithm succeeds also on sampled graphs.

This approach fails for the more general model $\mathcal{G}(A)$. Consider the expected adjacency matrix $A = A[XYZ]$ that arises if the off-diagonal blocks are equal to the matrix shown in (3) and the corresponding graph in Fig. 1(*left*).

$$X = Y = Z = \begin{bmatrix} 0 & 0 & p & p \\ 0 & 0 & p & p \\ p & p & 0 & 0 \\ p & p & 0 & 0 \end{bmatrix} \tag{3}$$

Two vertices that are colored the same may have very different (even disjoint) neighborhoods. In particular, projecting vertices to the column space of A does

not reveal the color classes. Equation (3) and Fig. 1 indicate how to construct distributions $\mathcal{G}(A)$ for which traditional spectral methods [1,12] fail: introduce large (i.e., of linear size) sub-blocks of X, Y, Z that are zero, i.e., prohibit edges between large subsets of differently colored vertices. To cope with the distribution $\mathcal{G}(A[XYZ])$ we have to apply a different projection.
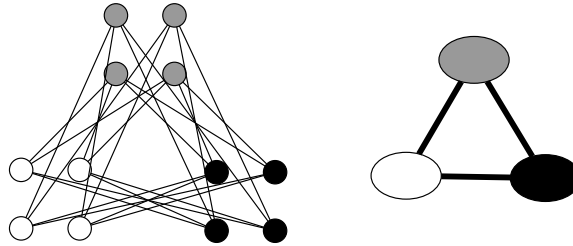

Fig. 1. *Left:* Small example of a non-uniform expected adjacency structure, defined by the block-matrices shown in (3). Edges have weight p. Every white vertex has exactly two black neighbors. *Right:* Quotient induced by the coloring. Edges have weight $2p$.

We represent a vertex k-coloring by a real $k \times n$ matrix P, called the *characteristic matrix* of the coloring, defined by

$$P_{\ell v} = \begin{cases} 1/\sqrt{r} & \text{if vertex } v \text{ is colored } \ell \text{ and } r \text{ is the size of the color class } \ell, \\ 0 & \text{if vertex } v \text{ is not colored } \ell. \end{cases}$$

The characteristic matrix P of the planted 3-coloring projects vertices to 3-dimensional space, such that vertices are mapped to the same point if and only if they are equally colored. Thus, P could be used to determine the planted coloring—we just need a method that identifies the correct P (or a good approximation of it), given only a sample of the distribution.

To derive such a method we observe that, if the block matrices are regular, then the planted 3-coloring satisfies the property of the following definition:

Definition 2. *A coloring is called* structural *(for a symmetric matrix A) if its characteristic matrix P satisfies*

$$\forall u, v \in V : P(u) = P(v) \Longrightarrow PA(u) = PA(v) \ .$$

That is, a coloring is structural if, whenever two vertices are colored the same, then they have the same number of each color in their neighborhoods. In algebraic and spectral graph theory, structural colorings are known under the names of *equitable partitions*, or *divisors of graphs* [7,3]. For example, the coloring of the graph in Fig. 1(*left*) is structural.

The idea of structural colorings serves only as a guide to find a projection that recovers the planted 3-coloring. The property of being a structural coloring is not robust to random noise. However, it can be shown, that a relaxation of

structural colorings is stable. It is noteworthy, that we do not relax the property of being structural but that of being a discrete coloring.

In the remainder of this section we relax the notion of colorings to *projections* and *similarities*, while keeping the property of being structural. In Sect. 4 we show that (w. h. p.) our algorithm computes the appropriate structural similarity for the sampled matrix and recovers the planted 3-coloring.

Structural similarities have been introduced in [2] as a relaxation for role assignments. (Role assignments identify structurally similar vertices in networks.) Here we review some concepts from [2] in a slightly different notation.

Projections and similarities are introduced as relaxations of k-colorings and their associated equivalence relations on the vertex set:

Definition 3. *A real $k \times n$ matrix P with orthonormal rows is called a* projection. *If P is a projection, then the real $n \times n$ matrix $S = P^\mathrm{T}P$ is called the* similarity *associated with P. Let, in addition, be A the adjacency matrix of a graph. Then the real $k \times k$ matrix $B = PAP^\mathrm{T}$ is called the* quotient *induced by A and P.*

The characteristic matrix of a coloring is a special case of a projection. Projections are more general than colorings, since they allow vertices to be members of several color classes: the v'th column of a projection P is a k-dimensional vector that describes the real-valued membership of vertex v to the k color-classes. The entry S_{uv} of the associated similarity S is the dot-product of the u'th and v'th column of P. Thus u and v have high similarity if they are similarly colored. From an algebraic point of view, a similarity is the orthogonal projection to the row-space of P (compare [2]). If P is the characteristic matrix of a coloring, then the quotient $B = PAP^\mathrm{T}$ is the adjacency matrix of the weighted graph that has the k color-classes as vertices and two classes c_1 and c_2 are connected by an edge whose weight is the sum over all edges between c_1 and c_2 divided by $\sqrt{|c_1| \cdot |c_2|}$. For an example, see Fig. 1. The following definition introduces the attribute structural for similarities. It is then noted in Theorem 4 that structural similarities are indeed relaxations of structural colorings.

Definition 4. *Let P be a projection and let S be its associated similarity, then P and S are called* structural *for a matrix A if $SA = AS$.*

Theorem 4 ([2]). *Let P be the characteristic matrix of a vertex coloring $c: V \to \{1, \ldots, k\}$. Then, P is a structural projection if and only if c is a structural coloring.* □

The following Theorem provides the link between spectral techniques and structural similarities. Further it shows how similarities that yield a pre-specified quotient can be efficiently computed.

Theorem 5 ([2]). *Let A be a symmetric $n \times n$ matrix, B a symmetric $k \times k$ matrix, P a projection, and S its associated similarity. Then P and S are structural for A if and only if the image of S is generated by eigenvectors of A.*

Furthermore, P and S are structural for A and the induced quotient equals B if and only if those eigenvectors are associated to the eigenvalues of B.   □

If P and S are structural, we call the eigenvalues of B *associated* to P and S.

Traditional spectral methods typically chose projections associated to the eigenvalues with the largest absolute values (compare [12]). Structural similarities are not restricted to projecting to the largest eigenvalues but can chose all subsets and thereby can recover partitions in more general situations (as demonstrated in this paper). The second part of Theorem 5 is important for determining which eigenvalues have to be chosen for a specific task.

4   Correctness of the Algorithm

Throughout this section, let $A = A[XYZ]$ be a real $n \times n$ matrix as in the righthand side of (1) and let $\mathcal{G}(A)$ be the associated distribution of 3-colorable graphs (compare Def. 1). Let $\hat{A}$ be the adjacency matrix of a sample drawn from $\mathcal{G}(A)$. Further, let σ^2 be the variance of the distribution and assume that $\sigma^2 \gg (\log^6 n)/n$.

The following theorem states that, for regular X, Y, and Z, there is a structural similarity for A, which reveals the planted 3-coloring. Theorem 6 does not rely on any assumptions on $\text{sep}_3(A)$.

Theorem 6. *Let X, Y, and Z be regular $r \times r$ matrices with degree d. Then, there is a structural similarity S that has $2d$, $-d$, and $-d$ as associated eigenvalues and satisfies for all vertices u and v,*

$$\|S(u) - S(v)\| = \begin{cases} 0 & \text{if } u \text{ and } v \text{ are colored the same, and} \\ \sqrt{2/r} & \text{if } u \text{ and } v \text{ are colored differently.} \end{cases} \qquad (4)$$

Proof. Since the matrices X, Y, and Z are regular, the planted 3-coloring is a structural coloring for A. Thus, by Theorem 4 its characteristic matrix P, is a structural projection. Furthermore, the induced quotient $B = PAP^{\mathrm{T}}$ is the adjacency matrix of a triangle whose three edges have weight d. Thus, the associated eigenvalues of P are $2d$, $-d$, and $-d$. Finally, the similarity $S = P^{\mathrm{T}}P$ satisfies (4).   □

We show in Theorem 8 that there is a structural similarity $\hat{S}$ for the *sampled* adjacency matrix $\hat{A}$ that is close enough to S. First we have to recall a well-known bound on the eigenvalues of random matrices.

Theorem 7. *Let F be defined by $F = A - \hat{A}$. Then (w. h. p.) it is $\|F\|_2 \leq 4\sigma\sqrt{n}$ [12]. In particular, the eigenvalues of $\hat{A}$ differ (w. h. p.) from those of A by at most $4\sigma\sqrt{n}$ [13].*

Theorem 8. *Let X, Y, and Z be regular $r \times r$ matrices with common degree d and $A = A[XYZ]$. Further, let S be the similarity from Theorem 6 and assume that $\text{sep}_3(A)$ is in $\omega(\sigma\sqrt{n})$. Then, w. h .p. the similarity $\hat{S}$ that is associated to those three eigenvalues of $\hat{A}$ that have the smallest distance to $2d$, $-d$, and $-d$ satisfies $\|\hat{S} - S\|_2 \in \mathcal{O}(\sigma\sqrt{n}/\text{sep}_3(A))$.*

Proof. (The following assertions hold w.h.p. for sufficiently large n.) By the assumption on $\text{sep}_3(A)$ and Theorem 7, there are three well-defined eigenvalues λ_{i_1}, λ_{i_2}, and λ_{i_3} of $\hat{A}$ that have the smallest distance to $2d$, $-d$, and $-d$. Let v_{i_1}, v_{i_2}, and v_{i_3} be three orthonormal eigenvectors of $\hat{A}$, associated to λ_{i_1}, λ_{i_2}, and λ_{i_3}. Let C be the $n \times 3$ matrix whose columns are the v_{i_j}, $j = 1, 2, 3$. We show that $\hat{S} = CC^{\mathrm{T}}$ satisfies the assertions of the theorem.

By Theorem 5, $\hat{S}$ is structural for $\hat{A}$ and λ_{i_1}, λ_{i_2}, and λ_{i_3} are the eigenvalues associated to $\hat{S}$. To show the bound on $\|\hat{S} - S\|$, let $M = C^{\mathrm{T}}\hat{A}C$, B_1 be an $n \times 3$ matrix whose columns span the image of the similarity S, and B_2 be an $n \times (n-3)$ matrix, such that $(B_1 B_2)$ is an orthogonal $n \times n$ matrix. Let F be defined by $\hat{A} = A - F$ and set $L = B_2^{\mathrm{T}}AB_2$. By definition of M and the fact that $\hat{S}$ commutes with $\hat{A}$ ($\hat{S}$ is structural for $\hat{A}$), it is $0 = \hat{A}C - CM$. By the definition of F it follows $FC = AC - CM$. Let δ be the minimal distance between eigenvalues of M and those of L. By the assumptions on the separation $\text{sep}_3(A)$ and Theorem 7, δ is in $\Omega(\text{sep}_3(A))$ and it follows with Theorems V.3.4 and I.5.5 of [13] that

$$\|S - \hat{S}\|_2 \leq \frac{2\|FC\|_F}{\delta} \ .$$

The 2-norm of F is bounded by $4\sigma\sqrt{n}$ (Theorem 7), the Frobenius norm of the $n \times 3$ matrix FC is at most $\sqrt{3}$-times the 2-norm of FC, and the 2-norm of the matrix C (having orthonormal columns) is 1. Thus, the assertion follows with

$$\|FC\|_F \leq \sqrt{3}\|FC\|_2 \leq \sqrt{3}\|F\|_2 \leq 4\sigma\sqrt{3n} \ .$$

$\square$

To determine $\hat{S}$, the degree d of the block matrices has to be estimated:

Lemma 1. *Let X, Y, and Z be regular $r \times r$ matrices with common degree d and $A = A[XYZ]$. Let $\hat{A}$ be the adjacency matrix of a graph drawn from $\mathcal{G}(A)$ and set $\hat{d} = \sum_{i,j=1}^n \hat{A}_{ij}/(2n)$. Then, with high probability, $\hat{d} - d$ is in $\mathcal{O}(\log n)$.*

Proof. Follows in a straightforward manner from the Hoeffding bound. $\square$

Proof (of Theorem 1). By Lemma 1 and Theorem 7, the similarity $\hat{S}$, as computed by the algorithm, is the similarity from Theorem 8. (Note that $\hat{S} = P^{\mathrm{T}}P$ is independent on orthogonal transformations on the rows of P like, e.g., permutation or reflexion of eigenvectors.) Let v be any vertex and let S be the similarity from Theorem 6. We have by Theorem 8 that w.h.p.

$$\|S(v) - \hat{S}(v)\| \in o(1/\sqrt{n}) \ .$$

Hence, for two vertices u and v it is (applying Theorem 6)

$$\|\hat{S}(u) - \hat{S}(v)\| \in \begin{cases} o(1/\sqrt{n}) & \text{if } u \text{ and } v \text{ are in the same color class,} \\ \Omega(1/\sqrt{n}) & \text{else} \ . \end{cases}$$

Thus for sufficiently large n the clustering procedure in Step 5 yields exactly the planted color classes. $\square$

Proof (of Corollary 1). Follows from Theorem 1 by considering the zero-variance distribution that assigns probability one to G (and probability zero to any other graph). Following the proofs in Sect. 4, it can be seen that the restriction "with high probability" from Theorem 1 can be dropped in this situation.   □

5   Hard Instances for Traditional Methods

Proof (of Theorem 2). The instances for which the algorithm from [1] does not work are essentially a blown-up version of the example in (3) and Fig. 1 with a few added edges. For simplicity we take $p = 1$ in our example. This implies that only one graph has non-zero probability in the distribution. It should be obvious that similar examples with probabilities different from zero or one can be constructed.

Let $r = 2k$ for an integer k. Let H be the graph that is the complete bipartite graph $K_{k,k}$ plus the edges of two cycles of length k, connecting the vertices in the bipartition classes of $K_{k,k}$ (thereby making H non-bipartite). Let $X = Y = Z$ denote the adjacency matrix of H and let $A = A[XYZ]$. Let G be the graph with adjacency matrix A (the unique graph returned by $\mathcal{G}(A)$).

The preprocessing step from [1] is void in this case since G is regular. The last eigenvector v_n has median zero. Further for $t = v_n$, in the first phase of the algorithm from [1] the vertices are colored with only two colors, according to which "bipartition class" they belong to. In particular this coloring is a very bad approximation to the unique proper 3-coloring and it is easy to see that the second and third phase in the proposed algorithm do not overcome this.

Examples of distributions for which the algorithm in [12] does not recover the planted 3-coloring, are quite similar to the one above.

Finally, the above distribution satisfies the assumptions of our theorems and, hence, G can be colored by our algorithm: The matrices X, Y, and Z are d-regular by construction, where $d = k + 2$. By computing the eigenvalues of the complete bipartite subgraphs and applying facts about the eigenvalues of the Kronecker product of matrices (compare [3]), we get that the eigenvalues $2d$, $-d$, and $-d$ have non-zero separation from the others and thus Corollary 1 applies.   □

Proof (of Theorem 3). In the example above there are suitable sets U and T of size linear in n such that there is no edge joining U and T.   □

6   Concluding Remarks

The ideas of this paper are not restricted to graph coloring. Many heuristics for $\mathcal{NP}$-hard graph partitioning problems (like min-bisection, clique, or independent set) are based on spectral techniques that typically chose projections associated to the eigenvalues with the largest absolute values. McSherry [12] showed that these specific spectral projections recover partitions if vertices in the same class have identical columns (neighborhoods) in the expected adjacency matrix. It

seems that the converse is also true: these specific spectral projections recover partitions *only if* vertices in the same class have almost identical neighborhoods in the expected adjacency matrix. We outlined in Sect. 3 that projections associated to eigenvalues that are not necessarily the largest may succeed in more general situations, where vertices from the same class have only same-colored (instead of identical) neighborhoods. It seems to be promising to consider these generalized spectral projections also for the solution of other problems.

References

1. N. Alon and N. Kahale. A spectral technique for coloring random 3-colorable graphs. *SIAM Journal on Computing*, 26:1733–1748, 1997.
2. U. Brandes and J. Lerner. Structural similarity in graphs. In *Proceedings of the 15th International Symposium on Algorithms and Computation (ISAAC'04)*, pages 184–195, 2004.
3. D. M. Cvetković, M. Doob, and H. Sachs. *Spectra of Graphs*. Johann Ambrosius Barth, 1995.
4. M. E. Dyer and A. M. Frieze. Fast solution of some random $\mathcal{NP}$-hard problems. In *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science (FOCS'86)*, pages 331–336, 1986.
5. U. Feige and J. Kilian. Heuristics for semi-random graph problems. *Journal of Computer and System Sciences*, 63(4):639–671, 2001.
6. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-completeness*. Freeman and Company, 1979.
7. C. Godsil and G. Royle. *Algebraic Graph Theory*. Springer, 2001.
8. G. H. Golub and C. F. van Loan. *Matrix Computations*. John Hopkins University Press, 1996.
9. T. R. Jensen and B. Toft. *Graph Coloring Problems*. Wiley, 1995.
10. M. Krivelevich. Coloring random graphs – an algorithmic perspective. In *Proceedings of the 2nd Colloquium on Mathematics and Computer Science (Math-Info'2002)*, pages 175–195, 2002.
11. L. Kučera. Expected behavior of graph coloring algorithms. In *Lecture Notes in Computer Science 56*, pages 447–451. Springer Verlag, 1977.
12. F. McSherry. Spectral partitioning of random graphs. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS'01)*, pages 529–537, 2001.
13. G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.
14. C. R. Subramanian, M. Fürer, and C. E. V. Madhavan. Algorithms for coloring semi-random graphs. *Random Structures and Algorithms*, 13:125–158, 1998.
15. J. Turner. Almost all k-colorable graphs are easy to color. *Journal of Algorithms*, 9:63–82, 1988.