Coloring Bipartite Hypergraphs

Hui Chen Alan Frieze
Carnegie Mellon University,
Pittsburgh PA 15213, USA

June 20, 2000

Abstract

It is NP-Hard to find a proper 2-coloring of a given 2-colorable (*bipartite*) hypergraph H. We consider algorithms that will color such a hypergraph using *few* colors in polynomial time. The results of the paper can be summarized as follows: Let n denote the number of vertices of H and m the number of edges. (i) For bipartite hypergraphs of dimension k there is a polynomial time algorithm which produces a proper coloring using $\min\{O(n^{1-1/k}), O((m/n)^{\frac{1}{k-1}})\}$ colors. (ii) For 3-uniform bipartite hypergraphs, the bound is reduced to $\tilde{O}(n^{2/9})$. (iii) For a class of dense 3-uniform bipartite hypergraphs, we have a randomized algorithm which can color optimally. (iv) For a model of random bipartite hypergraphs with edge probability $p \geq dn^{-2}$, $d > 0$ a sufficiently large constant, we can almost surely find a proper 2-coloring.

1 Introduction

A hypergraph $H = (V, E)$ has vertex set V, $|V| = n$, edge set E, $|E| = m$ and each edge $e \in E$ is simply a subset of V. Its dimension $\dim(H)$ is the size of the largest edge in E. A set $S \subseteq V$ is said to be *independent* if it contains no edge of H. A *proper* k-coloring of the vertex set V is a partition of V into k independent sets. A hypergraph *bipartite* if it admits a proper 2-coloring.

Lovász [9] showed that it is in general NP-Hard to determine whether or not a hypergraph H is bipartite. (There are important special cases where such a coloring can be found, see for example Beck [4], Alon and Spencer [3] or McDiarmid [10]). In such circumstances it is of some interest to see if one can find a proper coloring of a hypergraph H which is known to be bipartite, but for which no proper 2-coloring is given. This problem is very similar in flavour to that of finding a good coloring of a given 3-colorable graph. In this paper we modify recent ideas for tackling this latter problem and apply them in the context of bipartite hypergraphs. The results of this paper can be summarised as follows:

- We first consider an algorithm based on ideas of Wigderson [12] and show how to color a bipartite hypergraph of dimension k with $\min\{O(n^{1-1/k}), O((m/n)^{\frac{1}{k-1}})\}$ colors.
- We then modify the techniques of Karger, Motwani and Sudan [7] to derive a smaller $\tilde{O}(n^{2/9})$[1] upperbound on the number of colors needed to color a 3-uniform bipartite hypergraph in polynomial time.
- We next consider *dense* hypergraphs and show, using similar ideas to those of Edwards [5], that dense 3-uniform bipartite hypergraphs can be 2-colored in polynomial time.
- We then consider the case where H is chosen randomly from some natural distribution. We use a spectral method introduced by Alon and Kahale [2] to show that **whp**[2] we can 2-color a random bipartite hypergraph with edge density $p \geq dn^{-2}$.

2 Approximate Coloring for General Bipartite Hypergraphs

In this section we consider an algorithm for coloring bipartite hypergraphs which is a development of Wigderson's graph coloring algorithm. We assume that the edges of H are pairwise incomparable, if $e \supseteq e'$ then delete edge e.

For $U \subset V$, let $N(U) = \{x \in V : \{x\} \cup U \in E\}$. Note that $N(U) \neq \emptyset$ implies that U is independent. Let $E_U = \{U \cup \{x\} : x \in N(U)\}$ and $H \star U = (V, (E \setminus E_U) \cup \{U\})$. We begin with the following simple lemma.

Lemma 1 *Suppose $U \subseteq V$, H is bipartite and $N(U)$ contains an edge. Then*

(a) *$H \star U$ is bipartite.*

(b) *An s-coloring of $H \star U$ is also an s-coloring of H.*

Proof (a) Let A, B be a partition of V into independent sets. If for example, $U \subseteq A$ then $N(U) \subseteq B$ and so is independent.

[1] $\tilde{O}$ notation suppresses a factor $(\log n)^d$ for some positive constant d.
[2] with high probability i.e. probability $1 - o(1)$

(b) Fix some s-coloring $\mathcal{K}$ of $H \star U$. The edges of H which are not in $H \star U$ all contain U. Since $\mathcal{K}$ properly colors U it properly colors all such edges. □

We can now describe the algorithm

Phase 1

begin

while there exists $U \subseteq V$ with $N(U)$ dependent;

$H \leftarrow H \star U$.

end

This can easily be carried out in time polynomial in m, n.

At the end of Phase 1, H satisfies

$$U \subseteq V,\ N(U) \neq \emptyset \text{ implies } N(U) \text{ is independent.} \tag{1}$$

If $N(U)$ is independent then its elements can be colored with a single color. For $S \subseteq V$ let $E^S = \{e \in E : e \cap S = \emptyset\}$ and $H \setminus S = (V \setminus S, E^S)$.

Phase 2

begin

while there exists $U \subseteq V$ with $|N(U)| \geq n^{1/k}$;

Assign a new color to the elements in $N(U)$;

$H \leftarrow H \setminus N(U)$.

end

Phase 2 is also executable in polynomial time and requires no more than $n^{1-1/k}$ colors.

By the end of Phase 2 the maximum possible number of edges in H has been reduced.

Lemma 2 *At the end of Phase 2, H contains at most $n^{t-1+1/k}$ edges of size t, $2 \leq t \leq k$.*

Proof Consider the set of pairs (U, x) where $|U| = t - 1$ and $x \in N(U)$. Clearly there are at most $\binom{n}{t-1} n^{1/k}$ such pairs. On the other hand, each edge of size t gives rise to t distinct pairs. □

We complete the coloring randomly.

Phase 3

begin

Randomly color each vertex with one of $r = \lceil 4n^{1-1/k} \rceil$ colors;

$E' \leftarrow \{e \in E : e \text{ is not properly colored}\}$;

$S \leftarrow \bigcup_{e \in E'} e$;

$H \leftarrow H \setminus (V \setminus S)$

end

Lemma 3 *Let S be as in Phase 3. Then*

$$\mathbf{Pr}(|S| \geq n/2) \leq 1/2.$$

Proof Let m_t denote the number of edges of size t which remain in H and let $p_t = r^{-(t-1)}$ denote the probability that an edge of size t is not properly colored. Then

$$\begin{aligned} \mathbf{E}(|S|) &\leq \sum_{t=3}^{k} t m_t p_t \\ &\leq \sum_{t=3}^{k} t n^{t-1+1/k} 4^{-t} n^{-(t-1)(k-1)/k} \\ &\leq \sum_{t=3}^{k} t 4^{-t} n^{t/k} \\ &\leq n/7, \end{aligned}$$

for n large. The result follows from the Markov inequality. □

We repeat Phase 3, with the same initial set of edges, until $|S| \leq n/2$. The expected number of repetitions is less than 2. When this happens we have succeeded in reducing the number of vertices in H by a factor of 2 and we have used at most $5n^{1-1/k}$ colors.

We can repeat Phases 1 – 3, using new colors and halving the number of vertices at each iteration until the number of vertices remaining is less than $n^{1-2/k}$, in which case we give each vertex a unique color. The total number of colors used is then at most

$$5 \sum_{i=1}^{\infty} \left(\frac{n}{2^i}\right)^{1-1/k} + n^{1-2/k} \leq 10 n^{1-1/k}.$$

It can also be seen that if the hypergraph does not have many edges, i.e. when $m \leq n^{k-1-1/k}$, the simple random coloring idea will give a better bound (use $r = \lceil 2(m/n)^{\frac{1}{k-1}} \rceil$). To summarise

Theorem 1 *If H is a bipartite hypergraph of dimension k which has m edges then there is a polynomial time algorithm which properly colors the vertices of H in*
$\min\{O(n^{1-1/k}), O((m/n)^{\frac{1}{k-1}})\}$ *colors.*

3 3-Uniform Bipartite Hypergraphs

We now consider coloring a bipartite hypergraph H in which each edge has size 3. The approach here is very similar to that of Karger, Motwani and Sudan [7]. Consequently we will be somewhat brief in our exposition. As H is bipartite there is a partition A, B of V such that each edge of H meets both A and B. Putting $y_i = +1$ for $i \in A$ and $y_i = -1$ for $i \in B$ we see that

$$\{i,j,k\} \in E \text{ implies } y_i y_j + y_i y_k + y_j y_k \leq -1. \tag{2}$$

Arguing as in [7] we consider the semi-definite program

SDP

$$\begin{array}{llll} \text{Minimise} & \alpha & & \\ \text{Subject to} & & & \\ & \mathbf{v}^{(i)} \cdot \mathbf{v}^{(j)} + \mathbf{v}^{(i)} \cdot \mathbf{v}^{(k)} + \mathbf{v}^{(j)} \cdot \mathbf{v}^{(k)} & \leq \alpha, & \forall \{i,j,k\} \in E \\ & \mathbf{v}^{(i)} & \in S_{n-1}, & i \in [n], \end{array}$$

where $S_{n-1} = \{\mathbf{v} \in \boldsymbol{R}^n : |\mathbf{v}| = 1\}$.

We see from (2) that SDP has an optimal solution with $\alpha \leq -1$. We can compute a solution with $\alpha \leq -1 + \epsilon$ in time polynomial n and $\log 1/\epsilon$ – see for example Alizadeh [1]. Thus we can take ϵ as zero on the understanding that the errors introduced are swamped by other errors in the approximation.

So assume that we have computed $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)} \in S_{n-1}$ such that

$$\mathbf{v}^{(i)} \cdot \mathbf{v}^{(j)} + \mathbf{v}^{(i)} \cdot \mathbf{v}^{(k)} + \mathbf{v}^{(j)} \cdot \mathbf{v}^{(k)} \leq -1, \quad \forall \{i,j,k\} \in E \tag{3}$$

We now choose t (t defined later) random vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(t)}$ such that the components $x_j^{(i)}$ are independent standard normal $N(0,1)$ random variables. We say that $\mathbf{x}^{(r)}$ captures $\mathbf{v}^{(i)}$ if

$$\mathbf{x}^{(r)} \cdot \mathbf{v}^{(i)} \geq \mathbf{x}^{(s)} \cdot \mathbf{v}^{(i)}, \quad \forall s \neq r.$$

Let

$$S_r = \{i \in V : \mathbf{x}^{(r)} \text{ captures } \mathbf{v}^{(i)}\}, \quad 1 \leq r \leq t.$$

With probability one, $S_1, S_2, \ldots, S_t$ is a partition of V and so defines a coloring $\mathcal{K}$. Some edges of H may not be properly colored. Let m' be the number of such edges. We say that $\mathcal{K}$ is a semi-coloring if $m' \leq n/4$. In which case, the number of vertices in edges which are monochromatic is at most $3n/4$. Hence we can easily find a set S of at least $n/4$ vertices such that if $e \cap S \neq \emptyset$ then e is properly colored. We remove S, along with the associated colors and apply our algorithm to $H \setminus S$. This yields an $O(t)$ proper coloring of H (The number of colors needed in each next round will be geometrically decreasing).

We show later that for $\Delta = m/n$ and $t = O(\Delta^{1/8}(\log \Delta)^{9/8})$ then

$$\mathbf{Pr}(m' \geq n/4) \leq 1/2. \tag{4}$$

The initial hypergraph could have Δ as large as $n^2/6$ and this would lead to the use of $\tilde{O}(n^{1/4})$ colors. But by applying Phases 1 and 2 of the previous section, with $n^{1/k}$ in A of Phase 2 replaced by $n^{7/9}/\log n$, we can reduce Δ to $n^{16/9} \log n$ at the expense of using at most $n^{2/9} \log n$ colors. This (modulo the proof of (4)) leads to

Theorem 2 *If H is a bipartite hypergraph of dimension 3 then there is a polynomial time algorithm which properly colors the vertices of H in $\tilde{O}(n^{2/9})$ colors.*

Proof of (4)

We need a simple geometric fact:

Lemma 4 *Let $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}, \mathbf{v} \in S_{n-1}$ and suppose*

$$\mathbf{v}^{(1)} \cdot \mathbf{v}^{(2)} + \mathbf{v}^{(1)} \cdot \mathbf{v}^{(3)} + \mathbf{v}^{(2)} \cdot \mathbf{v}^{(3)} \leq -1.$$

Then

$$\min_{i=1,2,3} \mathbf{v}^{(i)} \cdot \mathbf{v} \leq 1/3.$$

Proof Observe first that

$$\begin{aligned}(\mathbf{v}^{(1)} + \mathbf{v}^{(2)} + \mathbf{v}^{(3)})^2 &= (|\mathbf{v}^{(1)}|^2 + |\mathbf{v}^{(2)}|^2 + |\mathbf{v}^{(3)}|^2) \\ &\quad +2(\mathbf{v}^{(1)} \cdot \mathbf{v}^{(2)} + \mathbf{v}^{(1)} \cdot \mathbf{v}^{(3)} + \mathbf{v}^{(2)} \cdot \mathbf{v}^{(3)}) \\ &\leq 1.\end{aligned}$$

So $|\mathbf{v}^{(1)} + \mathbf{v}^{(2)} + \mathbf{v}^{(3)}| \leq 1$. Hence

$$\mathbf{v}^{(1)} \cdot \mathbf{v} + \mathbf{v}^{(2)} \cdot \mathbf{v} + \mathbf{v}^{(3)} \cdot \mathbf{v} \leq 1,$$

and the lemma follows. □

Let now $P(n,t)$ denote the maximum over vectors $\mathbf{v}^{(i)}, i = 1,2,3$ of the probability that the three vectors are captured by the same vector from $\mathbf{x}^{(j)}, 1 \leq j \leq t$. We consider the probability that they are all captured by $\mathbf{x}^{(1)}$ and multiply this by t. By Lemma 4 we can assume without loss of generality that the angle between $\mathbf{x}^{(1)}$ and $\mathbf{v}^{(1)}$ is at least $\theta = \arccos(1/3)$. Put $\zeta = 1/\log \Delta, p = \zeta/\pi$ and $q = 1/\cos(\theta - \zeta)^2$. Consider the 2-dimensional subspace L generated by $\mathbf{v}^{(1)}$ and $\mathbf{x}^{(1)}$ and let R denote the wedge of this plane within an angle ϵ from $\mathbf{v}^{(1)}$. If $\mathbf{x}^{(1)}$ captures $\mathbf{v}^{(1)}$ then, as observed in [7] (proof of Theorem 7.7), the projection of $\mathbf{x}^{(1)}$ onto the nearer of the two lines bounding R exceeds the length of any $\mathbf{x}^{(j)}$ which lies in R. The probability of this event is shown in [7] to be $O((pt)^{-q})$. Thus

$$\begin{aligned} P(n,t) &= O(t(pt)^{-q}) \\ &= O(t(\Delta \log \Delta)^{-\frac{9}{8}+O(\zeta)}) \\ &= O(1/\Delta). \end{aligned} \tag{5}$$

By choosing the constant in the definition of t sufficiently large, the hidden constant in (5) can be made less than 1/8. In which case, the expected number of improperly colored edges will be less than $n/8$. Equation (4) follows from the Markov inequality. □

4 Dense Hypergraphs

Suppose that H is k-uniform. For $X \subset V, |X| = k-1$, let $N(X) = \{v \in V : v \cup X \in E\}$. Let $\alpha > 0$ be fixed. We say that H is α-dense if $|N(X)| > \alpha n$ for all $X \subset V$.

Theorem 3 *If H is bipartite, 3-uniform and α-dense then H can be 2-colored in $n^{O(1/\alpha)}$ time.*

Proof

Let $A \cup B$ be a partition of V into 2 independent sets. Clearly $|A|, |B| \geq \alpha n$. Choose S, $|S| = 3\alpha^{-1} \log n$ randomly from V. Then

$$\begin{aligned} \mathbf{Pr}(\exists x, y \in V : N(x,y) \cap S = \emptyset) &\leq n^2 \left(1 - \frac{|S|}{n}\right)^{\alpha n} \\ &\leq n^{-1}. \end{aligned} \tag{6}$$

So we can assume that $N(x,y) \cap S \neq \emptyset$ for all $x, y \in V$. By considering all $2^{|S|} = n^{O(1/\alpha)}$ possibilities we can guess $S_A = S \cap A$ and $S_B = S \cap B$. Now, see Edwards [5], we construct an instance of 2-SAT with variables $x_v, v \in V \setminus S$ and clauses $\mathcal{C}$. $x_v = 1$ will stand for $v \in A$ and the clauses will be denoted $C_{u,v}$, $u, v \in V \setminus S$ where

$$C_{u,v} = \begin{cases} \{\bar{x}_u, \bar{x}_v\} : & N(u,v) \cap S_A \neq \emptyset,\ N(u,v) \cap S_B = \emptyset, \\ \{x_u, x_v\} : & N(u,v) \cap S_A = \emptyset,\ N(u,v) \cap S_B \neq \emptyset, \\ \text{NO CLAUSE} : & \text{otherwise} \end{cases}$$

Now $\mathcal{C}$ is satisfiable by $x_v = 1$ for $v \in A \setminus S$ and $x_v = F$ for $v \in B \setminus S$. Also if $A' = \{v : x_v = 1\}$ is the 2-SAT solution we construct then $|A' \setminus A| \leq 1$.

For if $v_1, v_2 \in A' \setminus A \subseteq B \backslash S$ then the clause $C_{v_1,v_2} = \{\bar{x}_{v_1}, \bar{x}_{v_2}\}$ is not satisfied. Similarly, $|B' \setminus B| \leq 1$ and a simple brute force final check can correct any errors in $O(n^2)$ time. $\square$

The proof of correctness of the algorithm does not seem to generalize to higher dimensions. Though there seems to be no intrinsic reason why coloring becomes harder when $k \geq 4$, a completely different method seems to be necessary.

5 Random Hypergraphs

We first describe our model of a random bipartite hypergraph. Let $W_1 = \{1, 2, \ldots, n\}$ and $W_2 = \{n+1, n+2, \ldots, 2n\}$. There are $N = 2n\binom{n}{2}$ triples contained in $W_1 \cup W_2$ which contain at least one element from both of W_1 and W_2. We generate $H = H_{2n,3,p} = H(V, \mathcal{E})$ by independently including each possible triple with probability p. This is a natural analogue of the standard model of a random 3-colorable graph.

We will show that there exists a constant $d_0 > 0$ such that if $p \geq d_0 n^{-2}$ then $H = H_{2n,3,p}$ can be properly 2-colored in polynomial time **whp** (without knowledge of the partition W_1, W_2). We only consider the case $p = dn^{-2}$, $d \geq d_0$ constant. Things get easier if $d \to \infty$. The method used is an adaptation of the spectral method of Alon and Kahale [2].

5.1 The reduction

To apply the methodology of [2] we need a graph. So let $G = (V, E)$ where $V = W_1 \cup W_2$ and $e \in E$ if there is some triple t of H with $t \supseteq e$ i.e. make each triple into a triangle in G and merge multiple edges into one. We now proceed more or less as in [2].

1. Construct $G' = (V, E')$ by deleting all edges incident with vertices of degree at least $5d$ in G.
2. Compute the eigenvector v corresponding to the most negative eigenvalue of the adjacency matrix A' of G'.
3. Let $X = \{i \in V : v_i' \geq 0\}$ and $Y = V \setminus X$.
4. Use X, Y as the start of an iterative process to 2-color H.

5.2 Eigenvalues of G'

The intuition behind the approach is as follows: if each vertex of H lies in its expected number of triples with 1 or 2 members of W_1 ($d/2$ and d for a vertex in W_1) then $\mathbf{f} = (-1, -1, \ldots, -1, +1, +1, \ldots, +1)$ (-1 for $i \in W_1$ and $+1$ for $i \in W_2$) is an eigenvector of the adjacency matrix A of G, with eigenvalue $-d$. Since this is approximately true **whp**, there should be an eigenvector close to $\mathbf{f}$ which will give us a good idea of W_1 and W_2.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2n}$ be the eigenvalues of A', and $e_1, e_2, \ldots, e_{2n}$ be corresponding eigenvectors which form an orthonormal basis of $\boldsymbol{R}^{2n}$.

Lemma 5 *The following are true* **whp**:

(i) $\lambda_1 \geq (1 - 2^{-\Omega(d)})3d$,

(ii) $\lambda_{2n} \leq -(1 - 2^{-\Omega(d)})d$,

(iii) $|\lambda_i| = O(\sqrt{d})$ *for all* $2 \leq i \leq 2n - 1$.

Proof The proof is very similar to that of Proposition 2.1 of [2] which is based on ideas of Kahn and Szemerèdi [8]. We give only a bare outline. We use the fact (see for example [11]) that

$$\lambda_i = \min_L \max_{\substack{x \in L \\ x \neq 0}} \frac{x^T A' x}{x^T x}, \tag{7}$$

where L ranges over all subspaces of $\boldsymbol{R}^{2n}$ of dimension $2n - i + 1$.

The matrices A, A' partition naturally into 4 blocks arising from the partition of V into W_1, W_2. The off-diagonal elements of $A_{i,i}$, $i = 1, 2$ (corresponding to edges of G edges within the same W_i) are 0/1 where the 1's occur independently with probability

$$\begin{aligned} p_1 &= 1 - (1-p)^n \\ &= \frac{d}{n} + O\left(\frac{d^2}{n^2}\right) \end{aligned}$$

The off-diagonal elements of $A_{1,2}$ (corresponding to edges of G joining W_1 and W_2) do not occur independently, but this can be sidestepped. Consider $A_{1,2}$. For each of the N triples contained in V there are 2 edges of G which have one end in W_1 and one end in W_2. Randomly color one edge red and the other blue. Let $A_{1,2} = A_{1,2,R} + A_{1,2,B}$ where, for example, $A_{1,2,R}$ is the adjacency matrix of the bipartite grpah $G_R = (W_1, W_2, E_R)$ defined by the red edges. We claim that the edges of G_R occur independently with probability

$$\begin{aligned} p_2 &= (1 - (1-p)^{2(n-1)})/2 \\ &= \frac{d}{n} + O\left(\frac{d^2}{n^2}\right) \end{aligned}$$

Independence comes from the fact that the occurrence or non-occurrence of distinct red edges depends on the occurrence or non-occurrence of disjoint sets of triples.

We are now in good shape to appply the ideas of [2].

(i) Observe first that simple calculations show that **whp**

$$|E| = (1-o(1))3d/2 \tag{8}$$
$$|E \setminus E'| \leq ne^{-\Omega(d)}. \tag{9}$$

Now apply (7) with $L = \boldsymbol{R}^{2n}$ and $x = g = (1,1,1,\ldots,1)$.

(ii) Apply (7) with $L = F = \{\lambda f : \lambda \in \boldsymbol{R}\}$.

(iii) Let S be the set of all unit vectors $x \in \boldsymbol{R}^{2n}$ such that $\sum_{x \in W_j} x_v = 0$, $j = 1,2$. We first need to show that **whp** and uniformly over $x \in S$,

$$|x^T A' x| = O(\sqrt{d}). \tag{10}$$

It is enough to separately bound the contributions of each of the $A_{i,j}$ to $x^T A' x$ by $O(\sqrt{d})$. Further split the contribution of $A_{1,2}$ into that from $A_{1,2,R}$ and $A_{1,2,B}$. Similarly for $A_{2,1}$. We thus have to bound the contribution of 6 $n \times n$ 0/1 matrices with off-diagonal entries occurring independently with probability $(d+o(1))/n$. This is precisely what is done in [2] Lemma 2.4 and the preceding discussion. Thus, we can consider (10) to be proved.

The next two equations are proved in a similar manner to Lemma 2.8 of [2]. It helps to use the above decomposition into 6 matrices, to avoid problems of independence.

$$|(A' + dI)f| = O(|f|\sqrt{d}) \qquad \textbf{whp} \tag{11}$$
$$|(A' - 3dI)g| = O(|g|\sqrt{d}) \qquad \textbf{whp}$$

The proof of (iii) can now be completed. To show $\lambda_2 = O(\sqrt{d})$ we take L in (7) to be the set $\{x \in \boldsymbol{R}^n : x^T g = 0\}$. Write $x \in L$ as $\alpha f + s$ where $s \in S$. Then,

$$\begin{aligned} x^T A' x &= \alpha^2 f^T A' f + 2\alpha s^T A' f + s^T A' s \\ &= \alpha^2 f^T A' f + 2\alpha s^T (A' + dI) f + s^T A' s \\ &\leq -\alpha^2 (1 - e^{-\Omega(d)}) d|f|^2 + 2\alpha |f||s| O(\sqrt{d}) + |s|^2 O(\sqrt{d}) \\ &\leq |x|^2 O(\sqrt{d}). \end{aligned}$$

To show $|\lambda_{2n-1}| = O(\sqrt{d})$ let L be any 2-dimensional subspace of $\boldsymbol{R}^{2n}$. L contains x such that $x^T f = 0$. Write $x = \alpha g + s$ where $s \in S$. Then

$$\begin{aligned} x^T A' x &= \alpha^2 g^T A' g + 2\alpha s^T (A' - 3d) g + s^T A' s \\ &\geq \alpha^2 (1 - e^{-\Omega(d)}) d|g|^2 - 2\alpha |g||s| O(\sqrt{d}) - |s|^2 O(\sqrt{d}) \\ &\geq -|x|^2 O(\sqrt{d}). \end{aligned}$$

□

Let $v_1, v_2, \ldots, v_{2n}$ be an orthonormal set of eigenvectors. If $f = \sum_{i=1}^{2n} c_i v_i$ then **whp**

$$\left| \sum_{i=1}^{2n-1} c_i v_i \right| = O(n/d). \tag{12}$$

To prove (12) we use

$$\begin{aligned} |(A' + dI)f|^2 &= \sum_{i=1}^{2n} c_i^2 (\lambda_i + d)^2 \\ &= \Omega(d^2) \sum_{i=1}^{2n} c_i^2, \end{aligned}$$

from Lemma 5. Applying (11) we get $\sum_{i=1}^{2n-1} c_i^2 = O(n/d)$ which is (12). Let $v_{2n} = (\xi_1, \xi_2, \ldots, \xi_{2n})$ and $U_1 = \{i : \ \xi_i \geq 0\}$ and $U_2 = \{i : \ \xi_i < 0\}$. It follows from (12) that we can assume without loss of generality that

$$|W_j \setminus U_j| = O(n/d) \text{ for } j = 1, 2. \tag{13}$$

5.3 Perfecting the coloring

We can therefore assume that the choice of U_1, U_2 as a coloring leaves all but at most $n/1000$ vertices properly colored. We proceed as in [2] to perfect the coloring.

We first list some properties that that G will have **whp**: $\gamma > 0$ is some small absolute constant.

P1 All but $n(1 - e^{-\gamma d})$ members of W_i have between $.99d$ and $1.01d$ neighbours in W_{3-i} and between $.49d$ and $.51d$ neighbours in W_i for $i = 1, 2$.

P2 For all $A, B \subseteq V$ with $ne^{-\gamma d} = |A| \geq |B|/2$ the number of edges joining A and B is at most $.001d|A|$.

P3 There are no two subsets $U, W \subseteq V$ such that $|U| \leq 0.001n$ and $|W| = |U|/2$, and every vertex of W has at least $d/5$ neighbors in U.

The proofs of these assertions are straightforward and are omitted.

5.3.1 An Iterative Procedure

for $i = 0, 1, \ldots, \lceil \log n \rceil$ **do**

begin

Simultaneously, for all $v \in V$, re-color v with the minority color of its neighbours in the previous round.

end

Analysis

Let H_i be the set of vertices with at most $1.01d$ G-neighbors in W_{3-i} and $0.51d$ G-neighbours in W_i. Let $H = H_0 = H_1 \cup H_2$. Then, while possible, delete from H_i, $i = 1, 2$ a vertex h with at most $0.99d$ H-neighbors in W_{3-i} and $0.51d$ H-neighbours in W_i.

Lemma 6 *H has at least $2n(1 - e^{-\Omega(d)})$ vertices* **whp**.

Proof Property **P1** shows that H_0 is large. Let the vertices removed from H be $h_1, h_2, \ldots, h_m$. Let $m_0 = ne^{-\gamma d}$. If $m \geq m_0$ then there are at least $.002 m_0 d$ edges joining $A = \{h_1, h_2, \ldots, h_{m_0}\}$ and $A \cup (V \setminus H_0)$, contradicting **P2**. □

Lemma 7 *At the end of the iterative procedure H is properly colored,* **whp**.

Proof Let U_i be the set of wrongly colored vertices in H at the start of iteration i. If $v \in U_i$ then by the minority recoloring rule, at least $d/5$ of its neighbors are in U_{i-1}. Since **whp** $|U_1| = e^{-\Omega(d)} n$ we can apply **P3** repeatedly to show that $|U_i| \leq |U_{i-1}|/2$. □

5.3.2 Brute force re-coloring

In this phase, we simply uncolor any vertex in the set $V \setminus H$ to ensure that all the colored vertices are correctly colored. We then use exhaustive search to re-color each component of the graph Γ induced by $V \setminus H$. This generally takes polynomial time as

Proposition 4 **Whp** *the largest connected component of Γ has at most $\lceil \log_2 n \rceil$ vertices.*

Proof We sketch the proof which is similar to Proposition 3.9 of [2].

Let T be a fixed tree on $\log_2 n$ vertices of V. Let $E(T)$ and $V(T)$ denote the the edge set and vertex set of T. Let I be the subset of $V(T)$ all of whose vertices having degree at most 4 in T. So $|I| \geq |V(T)|/2$. Build H' in the following way:

i. Let H' be the set of vertices with at most $1.01d - 4$ G-neighbors in W_{3-i} and $0.51d$ G-neighbors in W_i, $i = 1, 2$.

ii. Delete from H' all vertices of $V(T) - I$.

iii. Repeatedly delete from H' all vertices having at most $0.99d$ H'-neighbors in W_{3-i} and $0.51d$ H'-neighbors in W_i.

The following two claims from [2] are also true in our case. For a set of triples F we let F_e denote the set of edges they induce in the graph G.

CLAIM Let F be any subset of $\mathcal{E}$. Let $H(F \cup T)$ be the value of H in case $E = F_e \cup T$, and $H'(F)$ be the value of H' in the case $E = F_e$. Then $H'(F) \subseteq H(F \cup T)$.

Proof We first show that the initial value of $H'(F)$, obtained after step (i) and (ii), is a subset of $H(F \cup T)$. Let v be any vertex that does not belong to the initial value of $H(F \cup T)$. Then v has more than $1.01d$ neighbors in the opposite color class of $(V, F \cup T)$ or more than $.51d$ neighbors into own color class. Therefore if:

Case 1. $v \in V(T) - I$. Then $v \notin H'(F)$ as it will be deleted at step (ii).

Case 2. $v \notin V(T) - I$. Then v is incident with at most 4 edges of T , thus it either has at least $1.01d - 4$ neightbors in the opposite color class in (V, F) or has at least $.51d$ neighbors in its own color class.

In either case, v does not belong to the initial value of $H'(F)$. By a similar argument to that for Lemma 3.8 from [2](notice that the assumption of a tripartition is not significant here) any vertex which is deleted in the process of constructing H will be deleted in the process of constructing H' as well and this completes the proof.

□

CLAIM

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

Proof It is sufficient to show that

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

Let $\hat{T}$ be the set of triples that contain an edge of T. For a set of triples Z let $Z' = Z \setminus \hat{T}$ and $Z'' = Z \setminus Z'$. By the previous claim, we have

$$\begin{aligned}
\mathbf{Pr}[I \cap H' = \emptyset] &= \sum_{F : I \cap H'(F) = \emptyset} \mathbf{Pr}[\mathcal{E} = F] \\
&\geq \sum_{F : I \cap H(F \cup T) = \emptyset} \mathbf{Pr}[\mathcal{E} = F] \\
&= \sum_{F : I \cap H(F \cup T) = \emptyset} \mathbf{Pr}[F' = \mathcal{E}']\mathbf{Pr}[F'' = \mathcal{E}'']
\end{aligned}$$

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

where F ranges over all the subset of triples , and F' ranges over those that do not contain an edge in T while F'' ranges over those that contain at least an edge in T. □

Assume that $d \le \alpha \log\log n$ for some fixed constant α, as otherwise almost surely $H = V$. One can show that **whp** H' misses at most $2^{-\Omega(d)}n$ vertices in each color classes. Let Φ be the events that there are at most $\log n$ pair of vertices u, v such that the number of triples containing u, v is at least 2. CLAIM

$$\mathbf{Pr}[\overline{\Phi}] = O(\frac{\log\log n}{\log n}) = o(1).$$

The proof is a simple first moment calculation. We omit it here. Now we can delete all pairs of vertices that have at least 2 triples containing them and their neighbors, since there are only $O(\log n)$ of them, a simple brute force coloring will find the correct one.

The intuition now is that since the conditioning between edges is small, we can assume that a large portion (nearly half) of the edges in T can be treated as unconditioned. Given edge $e = \{u, v\}$, let

$$N(e) = \{\{w,u\},\{w,v\}:\ \{w,u,v\}\in\mathcal{E}\}.$$

By the previous claim,**whp** after deleting $O(\log n)$ vertices, $\forall e \in E(T)$, $|N(e)\cap E(T)| \le 1$.

Since the choice of H' is independent of I the probability that there exists some T of size at least $\log_2 n$ is

$$
\begin{aligned}
& \mathbf{Pr}[\text{T is a subgraph of G}|\Phi] \\
\le\ & \frac{\mathbf{Pr}[e_1\in E(G)]}{\mathbf{Pr}[\Phi]}\mathbf{Pr}[E(T)\backslash N(e_1)\subseteq E(G)|\Phi] \\
\le\ & C(d/n^2)^{(|V(T)|-1)/2} \qquad \text{using an inductive argument} \\
\le\ & C'(d/n^2)^{\log_2 n/2}
\end{aligned}
$$

for some large constants C, C' and

$$
\begin{aligned}
\mathbf{Pr}[I\cap H'=\emptyset] &\le \frac{\binom{n-|H'|}{|I|}}{\binom{n}{|I|}} \\
&\le 2^{-\Omega(d|I|)} \\
&\le 2^{-\Omega(d\log_2 n)}
\end{aligned}
$$

Since the total number of possible connected trees of this size or more is at most
,

$$\binom{2n}{\log_2 n}(\log_2 n)^{\log_2 n-2}.$$

The multiplication of the above terms is of $O(n^{-\Omega(d)})$, which is the probability that the algorithm will fail in the third phase. We have thus completed the proof of Proposition 4. □

Acknowledgement We thank Avrim Blum for reminding us of the problem of 2-coloring hypergraphs.

References

[1] F.Alizadeh, *Combinatorial Optimizations with Semi-definite matrices,* 2nd Conference of Integer Programming and Combinatorial Optimizations (1992). pp385-405.

[2] N.Alon, N.Kahale , *A spectral technique for coloring random 3-colorable graphs*, DIMACS TR-94-35.

[3] N.Alon, J.Spencer,*The Probabilistic Method,* John Wiley & Sons (1992).

[4] J.Beck, *An algorithmic approach to Lovász Local Lemma I*, Random Structures & Algorithms (1991). pp343-365.

[5] K.Edwards, *The complexity of coloring problems on dense graphs*, Theoretical Computer Science 43 (1986) 337-343.

[6] M.Goemans, D.Williamson, *.878-Approximation Algorithms for MAX CUT and MAX 2SAT*. Proceedings of the 26th ACM Symposium on Theory of Computing (1994).

[7] D.Karger, R.Motwani, M.Sudan, *Approximation Graph Coloring by Semidefinite Programming.* 35th Foundations of Computer Science (1995). pp2-13.

[8] J.Kahn, Szemerédi,J.Friedman,*On the second eigenvalue in random regular graphs,* Proceedings of the 21st ACM STOC (1989). pp587-598.

[9] L.Lovász, *Covering and coloring of hypergraphs,* Proceding of the 4th Sourtheastern Conference on Combinatorics, Graph Theory and Computing. Utilitas Mathematica Publishing. Winnipeg (1973). pp3-12.

[10] C.J.H McDiarmid, *A random recoloring method for graph and hypergraph,* Combinatorial Probability and Computing 2 (1993). pp363-365.

[11] G.Strang, *Linear algebra and its applications,* Hardcourt Brace Jovanovich Publishing (1988) .

[12] A.Widgerson, *Improving the performance gurantee of approximate graph coloring,* Journal of ACM 30 (1983). pp729-735.