NP-Hardness of Almost Coloring Almost 3-Colorable Graphs

Yahli Hecht
School of Computer Science, Tel Aviv University, Israel

Dor Minzer
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Muli Safra
School of Computer Science, Tel Aviv University, Israel

Abstract

A graph $G = (V, E)$ is said to be (k, δ) almost colorable if there is a subset of vertices $V' \subseteq V$ of size at least $(1-\delta)|V|$ such that the induced subgraph of G on V' is k-colorable. We prove that for all k, there exists $\delta > 0$ such for all $\varepsilon > 0$, given a graph G it is NP-hard (under randomized reductions) to distinguish between:

1. Yes case: G is $(3, \varepsilon)$ almost colorable.
2. No case: G is not (k, δ) almost colorable.

This improves upon an earlier result of Khot et al. [16], who showed a weaker result wherein in the "yes case" the graph is $(4, \varepsilon)$ almost colorable.

2012 ACM Subject Classification Theory of computation → Problems, reductions and completeness

Keywords and phrases PCP, Hardness of approximation

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.51

Category RANDOM

Funding *Yahli Hecht*: Supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 835152).
Dor Minzer: Supported by an Sloan Research Fellowship, NSF CCF award 2227876 and NSF CAREER award 2239160.
Muli Safra: Supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 835152), and by an ISF grant and a BSF grant.

1 Introduction

The graph coloring problem is one of the most basic combinatorial optimization problems studied in theoretical computer science. A graph $G = (V, E)$ is *k-colorable* if there exists a vertex coloring $\mathsf{col}\colon V \to \{1, \dots, k\}$ such that for each edge $e = (u, v) \in E$, it holds that $\mathsf{col}(u) \neq \mathsf{col}(v)$. The chromatic number of G, denoted by $\chi(G)$, is defined to be the smallest integer k so that G is k-colorable. What is the computational complexity of finding the chromatic number of a graph?

It has long been known that computing the chromatic number of a graph is NP-hard [10]. Using the PCP theorem, one can improve this result and show that even approximating the chromatic number of a graph (within any constant factor) is NP-hard [12]. Thus, the next natural question to ask is how hard is it to find somewhat efficient coloring of a graph, provided that a very efficient one exists. Specifically, given a $k \in \mathbb{N}$, what is the smallest k' such that given a k-colorable graph, one may efficiently color it using k'?

As the case of $k = 2$ is the 2-coloring problem which is known to be in P, the first interesting case to consider is that of $k = 3$. Despite significant effort, the best unconditional result along these lines, due to [2], asserts that such graphs are NP-hard to color using 5 colors. In terms of conditional results, based on Khot's 2-to-2 Games Conjecture (with perfect completeness) [13], Dinur et al. [7] proved that it is NP-hard to C-color 4-colorable graphs for all $C > 0$. This result was recently improved by Guruswami and Sandeep [8], who showed that it is NP-hard to C-color 3-colorable graphs for any $C > 0$.[1] The gap between the ratios known to be hard and those known to be efficiently achievable is huge: on the algorithmic front, the state of the art result (due to [11]) asserts that one may efficiently find an $n^{0.2-\varepsilon}$-coloring of a given 3-colorable graph, where n is the number of vertices of the graph and $\varepsilon > 0$ is an absolute constant.

Due to the lack of progress towards unconditional hardness results for the graph coloring problem, one is motivated to consider relaxations of the problem that are easier to work with that still capture its essence. Most relevant to us is the almost coloring relaxation. Here, we say a graph $G = (V, E)$ is (k, δ) almost colorable if there exists a subset of vertices $V' \subseteq V$ with $|V'| \geqslant (1-\delta)|V|$ such that the induced subgraph of G on V' is k-colorable. That is, that the graph $G' = (V', E')$, where $E' = \{e = (u, v) \mid u, v \in V'\}$, is k-colorable. With regards to this notion, based on the hardness of 2-to-2 Games with imperfect completeness [15, 6, 5, 16], one can show that almost coloring almost 4-colorable graphs is NP-hard with any constant number of colors. More specifically, using ideas from [7], one can show that for all $\varepsilon, \delta > 0$, given a graph $G = (V, E)$, it is NP-hard to distinguish between:

1. Yes case: G is $(4, \varepsilon)$ almost colorable.
2. No case: G does not contain an independent set of fractional size δ.

In particular, it is NP-hard to $(k, 1-\varepsilon)$ almost color a given $(4, \varepsilon)$ almost colorable graph. We remark that the "independent set" conclusion in the no case is in fact stronger, and one often manages to achieve it via the typical style of PCP reductions for coloring problems.

The distinction between 3-coloring and 4-coloring may seem minor at first glance. This distinction hides within it a technical barrier related to the difference between Unique-Games and 2-to-2 Games. Indeed, the only type of problems that seem to facilitate results for 3-coloring are Unique-Games [7] (which inherently lack perfect completeness) and a certain variant of 2-to-2 games called Rich 2-to-2 Games [4, 3]. Both problems are conjectured to be NP-hard, but the proof of this assertion seems out of reach of current techniques. Using standard 2-to-2 Games, which are now known to be NP-hard (with imperfect completeness), there seem to have been fundamental difficulties beyond 4-colorable graphs.

Recently, Guruswami and Sandeep [8] observed that there there are transformations from the algebraic CSPs world [17] (which date back to [9]) that reduce the chromatic number of a graph from 4 to 3. Using these ideas, and combining them with the reduction of [7], Guruswami and Sandeep managed to prove the aforementioned conditional result, asserting that assuming the d-to-d conjecture of Khot with perfect completeness, it is NP-hard to C-color a given 3-colorable graph, for all constants $C > 0$. We remark that interestingly their reduction fails to establish the stronger "independent set" type conclusion in the no case.

The main result of this paper is that while this reduction does not preserve the independent set conclusion, it does preserve the intermediate, strictly weaker conclusion of almost coloring. More precisely, we use these ideas to get an (unconditional) NP-hardness result regarding almost coloring almost 3-colorable graphs with a constant number of colors. Specifically, we prove:

▶ **Theorem 1.** *For all $k \in \mathbb{N}$, there exists $\delta > 0$ such that for all $\varepsilon > 0$, given a graph $G = (V, E)$ it is NP-hard (under randomized reductions) to distinguish between:*

1. *Yes case: G is $(3, \varepsilon)$ almost colorable.*
2. *No case: G is not (k, δ) almost colorable.*

[1] In fact, for the result of Guruswami et al./ it suffices to assume that the d-to-d Games Conjecture of Khot [13] holds for some constant $d \in \mathbb{N}$.

Proof Idea

While the reduction of Guruswami and Sandeep [8] does not preserve the notion of almost coloring for all graphs, it turns out that it does so for bounded degree graphs. Indeed, this is the main insight behind the proof of Theorem 1. Thus, our proof of Theorem 1 starts with a hard instance of 2-to-2 Games with imperfect completeness and uses the reduction of [7] to establish hardness of almost coloring almost 4-colorable graphs for instances with certain regularity properties. We then use random sparsification to further reduce such instances to instances with bounded maximal degree. Finally, we use line-graph based reductions as in [8] to reduce this problem to the problem of almost coloring almost 3-colorable graphs.

We remark that just like in the result of Guruswami and Sandeep, our result also fails to establish the stronger "independent set" type conclusion in the no case. Indeed, strengthening Theorem 1 in that manner would automatically improve upon the best known hardness of approximation result for Vertex Cover, which is a long standing challenge. Currently, it is known that it is NP-hard to approximate the size of the smallest vertex cover of a graph within factor $\sqrt{2}-\varepsilon$ for all $\varepsilon > 0$, and an "independent-set"-type strengthening of Theorem 1 would improve this factor to $3/2-\varepsilon$, for all $\varepsilon > 0$. While we do not know how to prove such strengthening, we believe it may be doable and discuss such a possibility in Section 4.

2 Preliminaries

In this section, we present a few notions that will be necessary in the proof of Theorem 1.

2.1 Induced Coloring and Constraint Satisfaction Problems

For an integer $k \in \mathbb{N}$, the k-induced coloring of a graph G is the fractional size of the largest induced subgraph of G which is k-colorable. More precisely:

▶ **Definition 2.** *For a graph $G = (V, E)$, the induced-coloring $\mathsf{iCol}_k(G)$ is the fractional size of the largest induced subgraph that is k-colorable. That is, it is the maximum of $|V'| / |V|$ among all induced sub-graphs (V', E') of G that are k-colorable.*

According to this notation, the fact that a graph G is (k, δ) almost colorable is equivalent to $\mathsf{iCol}_k(G) \geqslant 1-\delta$. Next, we define the label cover problem, followed by the definition of a specific type of label cover instances called 2-to-2 Games.

▶ **Definition 3.** *An instance of Label Cover $\Psi = (G, \Sigma, \Phi)$ consists of a graph $G = (V, E)$, a finite alphabet Σ and a collection of constraints $\Phi = \{\Phi_e\}_{e\in E}$ one for each edge of G. The constraint on an edge $e \in E$ specifies tuples $\Phi_e \subseteq \Sigma \times \Sigma$ that are deemed satisfactory.*

Given an instance of label cover Ψ, the goal is to find an assignment $A\colon V \to \Sigma$ satisfying as many of the constraints as possible. Here, we say A satisfies the constraint on an edge $e = (u, v) \in E$ if $(A(u), A(v)) \in \Phi_e$. We denote by $\mathsf{sat}(\Psi)$ the maximum fraction of constraints that are satisfied in Ψ by any assignment A. We will mostly be interested in a specific type of label cover instances, known as 2-to-2 instances, in which the constraints take a special form.

▶ **Definition 4.** *A label cover instance $(G = (V, E), \Phi, 2R)$ is called a 2-to-2 Games instance if for every edge $e = (u, v) \in E$, the constraint Φ_e is of the form $\cup_{i=1}^{R} A_i \times B_i$, where $\{A_i\}_{i=1}^{R}$ and $\{B_i\}_{i=1}^{R}$ are two partitions of Σ into sets of size 2. Equivalently, the constraint Φ_e takes the form*

$$\left\{(i,j) \in \{1,\dots,2R\} \;\middle|\; (\pi^{-1}(i), \sigma^{-1}(j)) \in \{(2k,2k),(2k,2k+1),(2k+1,2k),(2k+1,2k+1)\}\right\}$$

for some permutations $\pi, \sigma : [2R] \to [2R]$.

Typically, given a 2-to-2 Games instances Ψ, the goal is to find an assignment to the vertices satisfying as many of the constraints as possible. For our purposes, it will be useful for define a variation of this parameter denoted by $\mathsf{iSat}(\Psi)$.

▶ **Definition 5.** *For an integer $m \in \mathbb{N}$ and a 2-to-2 Games instance $\Psi = (G = (V, E), \Sigma, \Phi)$, we define $\mathsf{iSat}(\Psi)$ to be the fractional size of the largest $V' \subseteq V$ for which there is an assignment $A\colon V' \to \Sigma$ satisfying all of the constraints of Ψ inside V'.*

2.2 Fourier Analysis

Our reduction requires a few basic notions from discrete Fourier analysis which we present next. We refer the reader to [19] for a more thorough presentation.

2.2.1 The Fourier Decomposition over Product Spaces

Let $q \in \mathbb{N}$ be an integer, and denote $[q] = \{0, 1, \dots, q-1\}$. We will consider the product space $L_2([q]^n, \mu)$ where μ is the uniform measure over $[q]^n$ and the inner product between $f, g\colon [q]^n \to \mathbb{R}$ is defined as

$$\langle f, g\rangle = \mathop{\mathbb{E}}_{x\sim\mu}[f(x)g(x)].$$

We may pick an orthonormal basis $\alpha_0, \dots, \alpha_{q-1}\colon [q] \to \mathbb{R}$ of $L_2([q], \mu)$ so that $\alpha_0 \equiv 1$, and we do so canonically. Given this basis, we may tensorize it to get an orthonormal basis of $L_2([q]^n, \mu)$ as $\{\alpha_{j_1,\dots,j_n}\}_{j_1,\dots,j_n\in[q]}$ where $\alpha_{j_1,\dots,j_n}$ is defined as

$$\alpha_{j_1,\dots,j_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} \alpha_{j_i}(x_i).$$

Given the orthonormal basis $\{\alpha_{\vec{j}}\}_{\vec{j}\in[q]^n}$, we may decompose any $f\colon [q]^n \to \mathbb{R}$ as a linear combination of the basis functions, and the coefficients in this linear combination are given by the Fourier coefficients:

▶ **Definition 6.** *For $f\colon [q]^n \to \mathbb{R}$ and $\vec{j} \in [q]^n$, we define $\hat{f}(\vec{j}) \stackrel{def}{=} \left\langle f, \alpha_{\vec{j}}\right\rangle$.*

Also, given $\vec{j} \in [q]^n$, we define the degree of $\vec{j}$, $\mathsf{deg}(\vec{j})$, to be the number of indices $i = 1, \dots, n$ such that $j_i \neq 0$. We define the action of a permutation on a function:

▶ **Definition 7.** *For a permutation $\pi\colon [n] \to [n]$ and a function $f\colon [q]^n \to \mathbb{R}$, we define $f^\pi\colon [q]^n \to \mathbb{R}$ by $f^\pi(x_1, \dots, x_n) \stackrel{def}{=} f(x_{\pi(1)}, \dots, x_{\pi(n)}) = f(x^\pi)$.*

2.2.2 Low Degree Influences and a Corollary of the Invariance Principle

Our analysis requires the notion of low degree influences, which is defined in terms of the Fourier coefficients of a function f in the following way:

▶ **Definition 8.** *For a function $f\colon [q]^n \to \mathbb{R}$, a coordinate $i = 1, \dots, n$ and an integer d, the degree d influence of i on f is defined as*

$$I_i^{\leqslant d}[f] \stackrel{def}{=} \sum_{\substack{\vec{j}\in[q]^n\\ \vec{j}_i\neq 0, \mathsf{deg}(\vec{j})\leqslant d}} \hat{f}^2(\vec{j}).$$

The following well known fact asserts that the number of coordinates with significant low degree influence is relatively small.

▶ **Fact 9.** *For a function $f\colon [q]^n \to \mathbb{R}$ with $\|f\|_2 \leqslant 1$ and $\tau > 0$, the number of coordinates $i = 1, \ldots, n$ with $I_i^{\leqslant d}[f] \geqslant \tau$ is at most $\frac{d}{\tau}$.*

In the reduction from 2-to-2 Games to 4-coloring we will have functions operating on the space $[q^2]^n$, and we will want to sometimes view it as $[q]^{2n}$. To do so, we will use the natural identification between $[q]^2$ and $[q^2]$ and thus define an identification between the spaces $[q^2]^n$ and $[q]^{2n}$ in the following way:

▶ **Definition 10.** *For $x = (x_1, \ldots, x_{2n}) \in [q]^{2n}$, we denote $\overline{x} \stackrel{def}{=} ((x_1, x_2), \ldots, (x_{2n-1}, x_{2n})) \in [q^2]^n$.*

Thus, given a function $f\colon [q]^{2n} \to \mathbb{R}$, one may consider the function $\overline{f}\colon [q^2]^n \to \mathbb{R}$ defined by the natural identification above as:

$$\overline{f}((x_1, x_2), \ldots, (x_{2n-1}, x_{2n})) = f(x_1, \ldots, x_{2n}).$$

The following facts from [7] will be used in the analysis of our reduction. The first claim related low-degree influences of f and of $\overline{f}$:

▷ Claim 11. For a function $f\colon [q]^{2n} \to \mathbb{R}$, a coordinate $1 \leqslant i \leqslant n$ and a degree parameter d, we have that

$$I_i^{\leqslant d}[\overline{f}] \leqslant I_{2i-1}^{\leqslant 2d}[f] + I_{2i}^{\leqslant 2d}[f].$$

Next, we need a corollary of the invariance principle which is also used in [7], and towards this end, we define the parameter $\Gamma_\rho(\mu, \tau)$:

▶ **Definition 12.** *Let Φ be the cumulative distribution function of $\mathcal{N}(0,1)$. We denote*

$$\Gamma_\rho(\mu, \tau) \stackrel{def}{=} \Pr\left[X \leqslant \Phi^{-1}(\mu) \wedge Y \geqslant \Phi^{-1}(1-\tau)\right],$$

where (X, Y) are ρ-correlated Gaussian random variables. If $\mu = \tau$, we write $\Gamma_\rho(\mu)$, we also omit ρ when clear from context.

Asymptotics for the value $\Gamma_\rho(\mu, \tau)$ exist (and are not too hard to establish), however, we will only require the fact that $\Gamma_\rho(\varepsilon)$ is a positive constant for all $\rho < 1$ and $\varepsilon > 0$. With the parameter $\Gamma_\rho(\mu, \tau)$ in hand, we now state the corollary of the invariance principle necessary for our analysis which can also be found in [7].

▶ **Theorem 13.** *Let $q \in \mathbb{N}$ be an integer and let T be a connected symmetric Markov chain on $[q]$, with eigenvalues $\lambda_0 = 1 \geqslant \ldots \geqslant \lambda_{q-1}$. If $\rho := \max(|\lambda_1|, |\lambda_{q-1}|) < 1$, then for any $\mu, \tau > 0$ there exist $\delta > 0$ and $d \in \mathbb{N}$, for which the following holds. For all functions $f, g\colon [q]^n \to [0,1]$, if $\mathbb{E}[f] \geqslant \mu, \mathbb{E}[g] \geqslant \tau$, and $\langle f, T^{\otimes n} g\rangle \leqslant \frac{1}{2}\Gamma_\rho(\mu, \tau)$, then there exists a coordinate $i = 1, \ldots, n$ such that $I_i^{\leqslant d}[f], I_i^{\leqslant d}[g] \geqslant \delta$.*

2.2.3 A Markov Chain on $[4]^2$

We end this section with the following claim due to [7] establishing the existence of a Markov chain on $[4]^2$ with certain properties:

▷ Claim 14. There exists a connected symmetric Markov chain T on $\{0,1,2,3\}^2$ with $\max(|\lambda_1|, |\lambda_{q-1}|) < 1$, such that if $T((x_1, x_2) \leftrightarrow (y_1, y_2)) > 0$ then $\{x_1, x_2\} \cap \{y_1, y_2\} = \emptyset$.

3 Hardness of Almost Coloring Almost 3-Colorable Graphs

3.1 The Starting Point: 2-to-2 Games

The starting point of our reduction is the hardness of 2-to-2 Games with imperfect completeness [15, 6, 5, 16] (see [18, 14] for an exposition). Specifically, we will use the following formulation:

▶ **Theorem 15.** *For all $s, \eta > 0$, given a 2-to-2 Games instance Ψ, it is NP-hard to distinguish between:*
- *YES case:* $\mathsf{iSat}(\Psi) \geqslant 1 - \eta$.
- *NO case:* $\mathsf{sat}(\Psi) \leqslant s$.

Additionally, in both the Yes and No cases, the graph G underlying Ψ satisfies the following regularity property: for all $\varepsilon \leqslant \beta \leqslant 1$, any set $S \subseteq V$ of fractional size β contains at least $\Omega(\beta^2)$ fraction of the edges of G.

3.2 A Reduction from 2-to-2 Games to Almost 4-Coloring

In this section, we apply the reduction of [7] on instances from Theorem 15 to get the following hardness result for almost 4-coloring:

▶ **Theorem 16.** *For all $\varepsilon, \eta > 0$, given an edge weighted graph $G = (V, E, w)$, it is NP-hard to distinguish between:*
- *Yes case:* $\mathsf{iCol}_4(G) \geqslant 1 - \eta$
- *No case: any set $S \subseteq V$ of fractional size at least ε contains edges of total weight of at least $\Omega\left(\varepsilon^2 \Gamma_\rho(\varepsilon/2)\right)$, where $\rho \in (0, 1)$ is an absolute constant.*

Proof. Let $\Psi = (G = (V, E), \Sigma, \Phi)$ be a 2-to-2 Games instances from Theorem 15 with completeness $1 - \eta$ and soundness $s > 0$, and let T be the Markov chain from Claim 14. Without loss of generality, we assume that the alphabet Σ is $[2R]$, where $2R = |\Sigma|$. Below, when we write Γ, we mean $\Gamma = \Gamma_\rho$ where ρ is the absolute constant from Claim 14.

We construct a graph $G = (V', E')$ by replacing each vertex $u \in V$ by a block of vertices $\{u\} \times \{0, 1, 2, 3\}^{2R}$, denoted by $B[u]$. Thus, $V' = \bigcup_{u \in V} B[u]$. As for the edges of H, for an edge $\{u, v\} \in E$ in G with a 2-to-2 constraint described by the permutations $\pi, \sigma \in S_{2R}$, we add an edge between (u, x) and (v, y) if $T^{\otimes R}(\overline{x^\pi} \leftrightarrow \overline{y^\sigma}) > 0$, in which case we assign the edge the weight $\frac{1}{|E|} T^{\otimes R}(\overline{x^\pi} \leftrightarrow \overline{y^\sigma})$.

Completeness. If $\mathsf{iSat}(\Psi) \geqslant 1 - \eta$, then there is $S \subseteq V$ of fractional size $1 - \eta$ and a assignment $c \colon S \to [2R]$, satisfying all constraints within S. For each $u \in S$ we assign the block $B[u]$ by $A(u, x) = x_{c(u)}$, and note that this assignment forms a 4-coloring of $\bigcup_{u \in S} B[u]$. Thus, $\mathsf{iCol}_4(G') \geqslant 1 - \eta$.

Soundness. Assume towards contradiction that there is a set of vertices $S' \subseteq V'$ of fractional size ε that contains less than $c\varepsilon^2 \Gamma(\varepsilon/2)$ of the edges (weighted), where $c > 0$ is an absolute constant to be determined. Define

$$S \stackrel{def}{=} \left\{ v \in V \;\middle|\; |B[v] \cap S'| \geqslant \frac{\varepsilon}{2} |B[v]| \right\}.$$

By an averaging argument we have that $|S| \geqslant \frac{\varepsilon}{2} |V|$, and for each $v \in S$ we define $f_v \colon \{0, 1, 2, 3\}^{2R} \to \{0, 1\}$ by $f_v(x) = 1_{(v,x) \in S}$. Thus, $\mathbb{E}[f_v] = \frac{|B[v] \cap S'|}{|B[v]|} \geqslant \frac{\varepsilon}{2}$ for all $v \in S$.

By the regularity condition on Ψ, the set S contains at least $\alpha\varepsilon^2$ fraction of the edges of Ψ, where $\alpha > 0$ is an absolute constant. Fix $u, v \in S$ between which there is an edge and let π, σ be the permutations defining the constraint on it. Note that the weight of edges between the block of u and the block of v is proportional to $\langle \overline{f_u^{\pi^{-1}}}, T^{\otimes R}\overline{f_v^{\sigma^{-1}}}\rangle$. Thus, as the total weight edges covered by S' is at most $c\varepsilon^2\Gamma(\varepsilon/2)$, it follows that for at least half of the edges (u, v) inside S we have that

$$\langle \overline{f_u^{\pi^{-1}}}, T^{\otimes R}\overline{f_v^{\sigma^{-1}}}\rangle \leqslant \frac{2c\varepsilon^2\Gamma(\varepsilon/2)}{\alpha\varepsilon^2} < \frac{1}{2}\Gamma(\varepsilon/2)$$

where we used the fact that c is sufficiently small. We refer to such edges (u, v) as good.

Fix a good edge (u, v). Applying Theorem 13 we find $d \in \mathbb{N}$ and $\delta > 0$ depending only on ε, such that there is $i \in \{1, \ldots, R\}$ for which $I_i^{\leqslant d}[\overline{f_u^{\pi^{-1}}}] \geqslant \delta, I_i^{\leqslant d}[\overline{f_v^{\sigma^{-1}}}] \geqslant \delta$. Using Claim 11 we conclude that

$$I_{\pi(2i-1)}^{\leqslant 2d}[f_u] + I_{\pi(2i)}^{\leqslant 2d}[f_u] = I_{2i-1}^{\leqslant 2d}[f_u^{\pi^{-1}}] + I_{2i}^{\leqslant 2d}[f_u^{\pi^{-1}}] \geqslant I_i^{\leqslant d}[\overline{f_u^{\pi^{-1}}}] \geqslant \delta,$$

$$I_{\sigma(2i-1)}^{\leqslant 2d}[f_v] + I_{\sigma(2i)}^{\leqslant 2d}[f_v] = I_{2i-1}^{\leqslant 2d}[f_v^{\sigma^{-1}}] + I_{2i}^{\leqslant 2d}[f_v^{\sigma^{-1}}] \geqslant I_i^{\leqslant d}[\overline{f_v^{\sigma^{-1}}}] \geqslant \delta.$$

Taking $\mathsf{List}[u] = \left\{ i \in \{1, \ldots, R\} \;\middle|\; I_i^{\leqslant 2d}[f_u] \geqslant \frac{\delta}{2} \right\}$, we conclude that for each good edge (u, v) it holds that the lists $\mathsf{List}[u]$ and $\mathsf{List}[v]$ contain a pair of assignments satisfying the constraint on (u, v). Also, by Fact 9 the size of each one of these lists is at most $\frac{4d}{\delta}$. Thus, if we choose for each $u \in S$ a label from $\mathsf{List}[u]$ uniformly at random, then we get an assignment that satisfies at least $\frac{\delta}{4d}$ of the good edges in expectation. Therefore, in expectation, it satisfies at least $\Omega(\varepsilon^2\delta/d)$ of the constraints of Ψ. In particular, there exists an assignment to Ψ satisfying at least $\Omega(\varepsilon^2\delta/d)$ of the constraints in it, and this is a contradiction provided that the soundness parameter s is sufficiently small. ◀

3.3 Sparsification: Hardness of Almost 4-coloring on Bounded Degree Graphs

In this section, we start with instances produced from Theorem 16, and reduce them to unweighted graph instances in which the maximal degree is bounded. To do so, we apply the random sparsification technique from [1]. More precisely, we show the following result:

▶ **Theorem 17.** *For all $\varepsilon, \eta > 0$, given a graph $G = (V, E)$ it is NP-hard under randomized reductions to distinguish between:*

- *Yes case:* $\mathsf{iCol}_4(G) \geqslant 1 - \eta$
- *No case: G has no independent set of size $\varepsilon\,|V|$.*

Moreover, the maximal degree of a vertex in G is at most $O\left(\frac{1}{\varepsilon^2\Gamma(\varepsilon/2))}\right)$.

Proof. Let $(G = (V, E), w)$ be a weighted graph as in Theorem 16, and construct a graph $G' = (V', E')$ in the following way:

Step 1: reduce the average degree and remove weight. Let $d^{-1} = C^{-1} \cdot \varepsilon^2\Gamma(\varepsilon/2)$ for a large absolute constant $C > 0$ to be determined. Independently sample dn edges from G (allowing repetitions), with probabilities proportional to the weights and include them, unweighted, in the graph G'.

Completeness. It is clear that $\mathsf{iCol}_4(G') \geqslant \mathsf{iCol}_4(G)$ as G' is a subgraph of G.

Soundness. Let $S \subseteq V$ be a set of fractional size ε, and recall that it contained at least $c\varepsilon^2\Gamma(\varepsilon/2)$ of the total weight, for some $c > 0$. Thus, the probability that S is an independent set in G' is at most

$$\left(1 - c\varepsilon^2\Gamma(\varepsilon/2)\right)^{dn} \leqslant e^{-c\varepsilon^2\Gamma(\varepsilon/2)dn} \leqslant e^{-ncC} \leqslant 2^{-2n}$$

for sufficiently large $C > 0$. As there are at most 2^n distinct such sets S, it follows from the union bound that the probability at least one of them is an independent set is at most $2^n \cdot 2^{-2n} \leqslant 2^{-n}$.

Step 2: Average to maximal degree. After step 1 the average degree is at most $2d$. We remove from G' vertices with degree higher than $4d$. By Markov's inequality, at most half of the vertices are removed. Consequently, if an almost coloring for $1 - \eta$ of the vertices existed, the same coloring colors at least $1 - 2\eta$ of the vertices. Similarly, if the largest independent set had fractional size at most ε, after removing vertices, no independent set has a fractional size larger than 2ε. ◀

3.4 Decreasing the Chromatic number

The final step in our proof is to apply a transformation on instances from Theorem 17 that reduces the almost coloring number in the "yes case" to 3, while keeping the soundness in the form of almost coloring. Toward this end, we use the *directed line graph*. The line graph was shown by [9, 20] to reduce the chromatic number (to be roughly logarithmic in the chromatic number of the original graph). This fact was used in [17, 8] to get hardness results for 3-colorable graphs. We also use this property of the line graph. For our purposes though, we show that a stronger guarantee can be made so long as the original graph has a bounded degree.

▶ **Definition 18.** *For a digraph $G = (V, E)$, the directed line graph of G, denoted by $\delta(G) = (V', E')$, is defined in the following way:*

- *The vertices are $V' = E(G)$, the (directed) edges of the original graph G.*
- *The edges are any pair of edges of G that shares a vertex, namely:*

$$E' = \{((u, v), (v, w)) \mid (u, v), (v, w) \in E(G)\}.$$

Applying the line-graph twice is denoted by $\delta^2(G) \overset{def}{=} \delta(\delta(G))$.

We use the following properties of the line graph. The first two lemmas address the completeness of the reduction.

▶ **Lemma 19.** *If G is 4-colorable, then $\chi(\delta^2(G)) \leqslant 3$.*

Proof. Given a 4-coloring $c : V \to \{0, 1, 2, 3\}$ of G, we can color $\delta^2(G)$ in the following way: for every vertex $((i, j), (j, k))$, if $c(j) \in \{0, 1, 2\}$ we will assign the color $c(j)$. Otherwise, we will assign any color in $\{0, 1, 2\} \setminus \{c(i), c(k)\}$. ◀

Next, we show that if G has bounded degree and a large induced subgraph which is 4-colorable, then $\delta^2(G)$ is almost 3-colorable.

▶ **Lemma 20.** *For a directed graph $G = (V, E)$ with bounded degree d, if $\mathsf{iCol}_4(G) \geqslant 1 - \eta$, then $\mathsf{iCol}_3(\delta^2(G)) \geqslant 1 - 6d^2\eta$.*

Proof. Since $\mathsf{iCol}_4(G) \geqslant 1-\eta$, there exists an induced subgraph H of G, where $|V(H)| \geqslant (1-\eta)\,|V(G)|$ and $\chi(H) \leqslant 4$. By Lemma 19 we get that $\delta^2(H)$ is 3-colorable. Therefore, it suffices to bound the fractional number of vertices of $\delta^2(G)$ that are not vertices of $\delta^2(H)$.

The set $V(\delta^2(G)) \setminus V(\delta^2(H))$ contains vertices of the form $((u,v),(v,w))$ when $\{u,v,w\} \not\subseteq V(H)$. The fractional size of $V(G)\setminus V(H)$ in $V(G)$ is at most η, and for each vertex $v \in V(G)$, there exists at most $3d^2$ vertices in $V(\delta^2(G))$ that include v. Therefore, at most $|V(G)|\,3d^2\eta$ of the vertices are not inside $V(\delta^2(H))$, which implies that $\left|V(\delta^2 H)\right| \geqslant (1-6d^2)\left|V(\delta^2 G)\right|$ and so $\mathsf{iCol}_3(\delta^2(G)) \geqslant 1-6d^2\eta$. ◀

The next lemma addresses the soundness of the reduction:

▶ **Lemma 21.** *Let $G=(V,E)$ be a directed graph with a maximal degree of at most d. Then*

$$\mathsf{iCol}_{\lfloor \log(Q) \rfloor}(\delta(G)) \leqslant \frac{d-1}{d} + \frac{\mathsf{iCol}_Q(G)}{d}.$$

Proof. For $S \subseteq V(\delta G) = E(G)$ and a partial $t = \lfloor \log(Q) \rfloor$ coloring $c\colon S \to [t]$, we can construct the following partial coloring of G. Let $H \subseteq V(G)$ be the set of vertices u such that $(u,v) \in S$ for all neighbours v of u in G. Define $f\colon H \to P([t])$ by

$$f(u) = \{c((u,v)) \mid (u,v) \in E(G)\}.$$

The function above is indeed a valid partial coloring, since for each $(u,v) \in E(G)$, such that u, v are both colored, $c((u,v)) \in f(u)$ and $c((u,v)) \notin f(v)$. This partial coloring has at most Q colors, since $|\mathcal{P}([t])| = 2^t \leqslant 2^{\log(Q)} = Q$.

Finally, by definition, we must have that $|H| \leqslant \mathsf{iCol}_Q(G)\,|V|$. On the other hand, if the size of S is denoted by $1-\eta$, then we have that H contains at least $(1-d\eta)\,|V|$ vertices. It follows that $1 - d\eta \leqslant \mathsf{iCol}_Q(G)$, and so $\eta \geqslant \frac{1-\mathsf{iCol}_Q(G)}{d}$. ◀

We are now ready to complete the proof of Theorem 1, restated below:

▶ **Theorem 22.** *For all $k \in \mathbb{N}$ there is $\delta = \delta(k) < 1$ such that the following holds for all $\eta > 0$. Given an undirected graph $G = (V,E)$, it is NP-hard (under randomized reductions) to distinguish between:*
- *Yes case:* $\mathsf{iCol}_3(G) \geqslant 1-\eta$
- *No case:* $\mathsf{iCol}_k(G) \leqslant 1-\delta$

Moreover, the problem is NP-hard on instances with bounded degrees (depending only on k).

Proof. We start with a graph G from Theorem 17 with sufficiently small parameters, construct $\delta^2(G)$, and replace the directed edges with undirected ones. Let d be the bound on the maximal degree of G.

Completeness. If $\mathsf{iCol}_4(G) \geqslant 1-\eta$, then by Lemma 20 $\mathsf{iCol}_3(\delta^2(G)) \geqslant 1-6d^2\eta$, which is at least $1-\sqrt{\eta}$ provided η is small enough (recall that d only depends on the soundness parameter ε in Theorem 17).

Soundness. Suppose that the largest independent set in G has fractional size at most $1/Q$. By Lemma 21 we get that $\mathsf{iCol}_{\lfloor \log(Q/2) \rfloor}(\delta G) \leqslant \frac{2d-1}{2d} + \frac{1}{4d}$. Note that the maximum degree in $\delta(G)$ is at most $4d$, so applying Lemma 21 on $\delta(G)$ we get that

$$\mathsf{iCol}_{\lfloor \log(\lfloor \log(Q/2) \rfloor) \rfloor}(\delta^2(G)) \leqslant \frac{4d-1}{4d} + \frac{4d-1}{16d^2} = 1 - \frac{1}{16d^2}.$$ ◀

4 Discussion

As discussed in the introduction, we believe that it may be possible to improve Theorem 1 so that in the "no case", the graph does not contain an independent set of fractional size ε. More precisely, we believe that the following conjecture should hold:

▶ **Conjecture 23.** *For all $\varepsilon, \eta > 0$, given an undirected graph $G = (V, E)$, it is NP-hard to distinguish between:*
- *Yes case:* $\mathsf{iCol}_3(G) \geqslant 1 - \eta$
- *No case: G does not contain an independent size of fractional size ε.*

If true, Conjecture 23 would be a significant result in our opinion. To start with, it immediately implies an improvement on the best known hardness of approximation result for Vertex Cover (to factor $3/2 - \varepsilon$ for all $\varepsilon > 0$, where the state of the art NP-hardness result stands at $\sqrt{2} - \varepsilon$). Below, we show that a strong enough form of Theorem 1 in which δ is a sufficiently large function of k (more specifically, $\delta \geqslant 2^{-o(k)}$) implies that Conjecture 23 holds. We remark that unfortunately, in our proof $\delta = 2^{2^{-k^{O(1)}}}$, which is quantitatively not strong enough to conclude Conjecture 23.

▶ **Definition 24.** *Given a graph $G = (V, E)$ and $t \in \mathbb{N}$, define $G[t] = (V[t], E[t])$ as:*
- *The vertices are $(a_1, \ldots, a_t) \in V^t$.*
- *There is an edge between $A = (a_1, \ldots, a_t)$ and $B = (b_1, \ldots, b_t)$, if we can move from A to B by replacing one vertex with one of his neighbors.*

$$(b_1, \ldots, b_t) = (a_1, \ldots, a_{i-1}, u, a_{i+1} \ldots, a_t).$$

The following claim shows that the transformation from G to $G[t]$ has an amplification-type effect:

▷ Claim 25. If $\mathsf{iCol}_Q(G) \leqslant 1 - (1 - c)^Q$, then for all $\varepsilon > 0$, there exists $t = t(Q, \varepsilon)$ such that $G[t]$ has no independent set of size $c + \varepsilon$.

Proof. Fix t. For an independent set $I \subseteq V[t]$ we define

$$\mathsf{est}_I(v) := \frac{|\{(a_1, \ldots, a_{t-1}, v) \in V[t]\} \cap I|}{|\{(a_1, \ldots, a_{t-1}, v) \in V[t]\}|}$$

For any independent set I and $v \in V$, it holds that $\mathsf{est}_I(v) \leqslant \mathrm{IS}(G[t-1])$, where $\mathrm{IS}(G)$ denotes the fractional size of the largest IS in G. This holds since $\{(a_1, \ldots, a_{t-1}) \mid (a_1, \ldots, a_{t-1}, v) \in I\}$ is an IS in $G[t-1]$. It also follows that $\mathrm{IS}(G[1]) \geqslant \mathrm{IS}(G[2]) \geqslant \mathrm{IS}(G[3]) \geqslant \ldots$ (as $\mathbb{E}_v[\mathsf{est}(v)] = |I|/|V[t]|$).

Let $\rho > 0$ be a parameter. Since $\mathrm{IS}(G[1]) \geqslant \mathrm{IS}(G[2]) \geqslant \mathrm{IS}(G[3]) \geqslant \ldots$, there exists $t \leqslant \lceil 1/\rho^2 \rceil + 1$, for which $\mathrm{IS}(G[t-1]) - \mathrm{IS}(G[t]) \leqslant \rho^2$. Fixing a maximal independent $I \subseteq V[t]$, it holds that $\mathsf{est}_I(v) \leqslant \mathrm{IS}(G[t]) + \rho^2$ for all $v \in V$. By Markov's inequality, for at most $\rho|V|$ vertices of G, $\mathsf{est}_I(v) < \mathrm{IS}(G[t]) - \rho$. We say a vertex $v \in V$ is good if $\mathsf{est}_I(v) \geqslant \mathrm{IS}(G[t]) - \rho$.

Sampling $a_1, \ldots, a_{t-1} \in V$ uniformly and taking $\{u \mid (a_1, \ldots, a_{t-1}, u) \in I\}$ Q times, we get Q independent sets, which is a partial Q-coloring. For a good vertex v, the probability that it is included in one of the Q samples is at least

$$1 - (1 - \mathsf{est}_I(v))^Q \geqslant 1 - (1 - \mathrm{IS}(G[t]) + \rho)^Q,$$

so the expected fractional size of such Q coloring is at least:

$$(1-\rho)\Big(1-(1-\mathrm{IS}(G[t])+\rho)^Q\Big)\,|V|\,.$$

As this must be smaller than $\mathsf{iCol}_Q(G) \leqslant 1-(1-c)^Q$, it follows that $\mathrm{IS}(G[t]) \leqslant c+O(\rho)$, and the proof is concluded for small enough ρ. ◁

▶ **Corollary 26.** *If $\delta(Q) > \frac{1}{2^{o(Q)}}$ in Theorem 22, then Conjecture 23 holds.*

Proof. Start with a graph $G=(V,E)$ from Theorem 22 and take $G[t]$ for sufficiently large t. If $\mathsf{iCol}_3(G) \geqslant 1-\eta$ and $S \subseteq V$ is a set of fractional size at least $1-\eta$ such that the induced subgraph on S is 3-colorable, then we may color S^t by $(a_1,\dots,a_t) \to \sum_{i=1}^{t} c(a_i) \pmod 3$, when $c: S \to \{0,1,2\}$ is a 3-coloring of S.

The soundness follows from Claim 25. ◀

References

1 Per Austrin, Subhash Khot, and Muli Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. *Theory Comput.*, 7(1):27–43, 2011. doi:10.4086/toc.2011.v007a003.

2 Libor Barto, Jakub Bulín, Andrei A. Krokhin, and Jakub Oprsal. Algebraic approach to promise constraint satisfaction. *J. ACM*, 68(4):28:1–28:66, 2021. doi:10.1145/3457606.

3 Mark Braverman, Subhash Khot, Noam Lifshitz, and Dor Minzer. An invariance principle for the multi-slice, with applications. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 228–236. IEEE, 2021. doi:10.1109/FOCS52979.2021.00030.

4 Mark Braverman, Subhash Khot, and Dor Minzer. On rich 2-to-1 games. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference, ITCS 2021, January 6-8, 2021, Virtual Conference*, volume 185 of *LIPIcs*, pages 27:1–27:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ITCS.2021.27.

5 Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in grassmann graphs. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 940–951. ACM, 2018. doi:10.1145/3188745.3188806.

6 Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 376–389. ACM, 2018. doi:10.1145/3188745.3188804.

7 Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. *SIAM J. Comput.*, 39(3):843–873, 2009. doi:10.1137/07068062X.

8 Venkatesan Guruswami and Sai Sandeep. d-To-1 Hardness of Coloring 3-Colorable Graphs with O(1) Colors. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*, volume 168 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 62:1–62:12, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2020.62.

9 C.C Harner and R.C Entringer. Arc colorings of digraphs. *Journal of Combinatorial Theory, Series B*, 13(3):219–225, 1972. doi:10.1016/0095-8956(72)90057-3.

10 Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972. doi:10.1007/978-1-4684-2001-2_9.

11 Ken-ichi Kawarabayashi and Mikkel Thorup. Coloring 3-colorable graphs with less than n 1/5 colors. *Journal of the ACM (JACM)*, 64(1):1–23, 2017.

12 Sanjeev Khanna, Nathan Linial, and Shmuel Safra. On the hardness of approximating the chromatic number. In *Second Israel Symposium on Theory of Computing Systems, ISTCS 1993, Natanya, Israel, June 7-9, 1993, Proceedings*, pages 250–260. IEEE Computer Society, 1993. doi:10.1109/ISTCS.1993.253464.

13 Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 17th Annual IEEE Conference on Computational Complexity, Montréal, Québec, Canada, May 21-24, 2002*, page 25. IEEE Computer Society, 2002. doi:10.1109/CCC.2002.1004334.

14 Subhash Khot. On the proof of the 2-to-2 games conjecture. *Current Developments in Mathematics*, 2019(1):43–94, 2019.

15 Subhash Khot, Dor Minzer, and Muli Safra. On independent sets, 2-to-2 games, and grassmann graphs. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 576–589. ACM, 2017. doi:10.1145/3055399.3055432.

16 Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 592–601. IEEE Computer Society, 2018. doi:10.1109/FOCS.2018.00062.

17 Andrei Krokhin, Jakub Oprš al, Marcin Wrochna, and Stanislav Živný. Topology and adjunction in promise constraint satisfaction. *SIAM Journal on Computing*, 52(1):38–79, February 2023. doi:10.1137/20m1378223.

18 Dor Minzer. *On Monotonicity Testing and the 2-to-2 Games Conjecture*, volume 49. Association for Computing Machinery, New York, NY, USA, 1 edition, 2022.

19 Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

20 S. Poljak and V. Rödl. On the arc-chromatic number of a digraph. *Journal of Combinatorial Theory, Series B*, 31(2):190–198, 1981. doi:10.1016/S0095-8956(81)80024-X.