Improved Hardness of Approximating Chromatic Number

Sangxia Huang
KTH Royal Institute of Technology
Stockholm, Sweden
sangxia@csc.kth.se

April 13, 2013

Abstract

We prove that for sufficiently large K, it is NP-hard to color K-colorable graphs with less than $2^{\Omega(K^{1/3})}$ colors. This improves the previous result of K versus $K^{\frac{1}{25}\log K}$ in Khot [21].

1 Introduction

A vertex coloring of a graph $G(V, E)$ is an assignment of colors to its vertices such that no two adjacent vertices receive the same color. The minimum number of colors needed for such a coloring is called the chromatic number of G, denoted by $\chi(G)$. As a classical combinatorial optimization problem, graph coloring is closely related to other problems such as finding maximum independent sets and probabilistically checkable proofs (PCPs) with certain special properties. In addition to being an important theoretical challenge, graph coloring also has a number of applications such as scheduling and register allocation.

It is known that determining the chromatic number of a graph exactly is NP-hard [13]. However, in many applications, it suffices to find a good enough approximation. In other words, given a K-colorable graph, we would like to color it with as few color as possible. Wigderson [27] gave an algorithm using $O(n^{1-1/(K-1)})$ colors. This was improved by Berger and Rompel [5] to $O\left((n/\log n)^{1-1/(K-1)}\right)$ colors. Karger, Motwani and Sudan [19] used semi-definite programming to achieve $\tilde{O}(n^{1-3/(K+1)})$, which was adapted in Blum and Karger [6] to an algorithm that colors a 3-colorable graph with $\tilde{O}(n^{3/14})$ colors. For 3-colorable graphs, the best known algorithm is by Kawarabayashi and Thorup [18] which uses $O(n^{0.2038})$ colors, based on results by Arora, Chlamtac and Charikar [2] and Chlamtac [8].

There have been many works on the hardness side as well. It is known that coloring 3-colorable graph with 4 colors is NP-hard, and for general K-colorable graph it is NP-hard to color with $K + 2\lfloor \frac{K}{3} \rfloor - 1$ colors [20, 14]. For sufficiently large K, the best known gap is by Khot [21] which proved that it is NP-hard to color a K-colorable graph with $K^{\frac{1}{25}\log K}$ colors. Assuming a variant of Khot's 2-to-1 Conjecture, Dinur, Mossel and Regev [11] proved that it is NP-hard to K'-color a K-colorable graph for any $3 \leq K < K'$. Guruswami and Sinop [15] proved that assuming the 2-to-1 Conjecture, it is NP-hard to find an independent set with more than $O\left(\frac{n}{\Delta^{1-c/(k-1)}}\right)$ vertices in a k-colorable graph of maximum degree Δ for some absolute constant $c \leq 4$.

Khot's hardness result [21] can be derived either using PCPs from Håstad and Khot [17] or Samorodnitsky and Trevisan [24]. We can view the results in both works as showing approximation resistance for a family of Boolean predicates that has very few accepting inputs — it is NP-hard to approximate Max CSP with those predicates better than just picking random assignments. For each integer $k > 0$, the approximation resistant predicates we get from [17] and [24] has k variables (and thus 2^k possible assignments) but only has $2^{O(\sqrt{k})}$ accepting assignments. The predicate from Håstad and Khot [17] is approximation resistant even on satisfiable instances — or have perfect completeness in PCP language — while the predicate from Samorodnitsky and Trevisan [24] is not. It is noted in Khot [21] that having perfect completeness is not necessary but makes the reduction for coloring easier.

In a recent breakthrough, Chan [7] proved approximation resistance for a family of predicates on k variables but only has $k + 1$ accepting assignments whenever k is of the form $k = 2^r - 1$. Previously, approximation resistance of those predicates are only known assuming the Unique Games Conjecture, proved by Samorodnitsky and Trevisan [25]. Hast [16] proved that predicates on k variables having at most $2\lfloor k/2 \rfloor + 1$ ($= k$ in the current setting) accepting inputs are not approximation resistant, thus these results are almost tight.

In [7], Chan also showed that for any $K \geq 3$, there is $\nu = o(1)$ such that given a graph with an induced K-colorable subgraph of fractional size $1 - \nu$, it is NP-hard to find an independent set of fractional size $1/2^{K/2} + \nu$. Although this gives a larger gap than Khot [21], the result lacks "perfect completeness" and thus is not comparable with Khot [21]. We refer to [10, 22, 7] for additional discussions on Almost-Coloring.

In this paper, we show improved hardness of approximating chromatic number using the above results.

Theorem 1. *For all sufficiently large K, it is NP-hard to color a K-colorable graph with $2^{\Omega(K^{1/3})}$ colors. Moreover, this hardness result holds for graphs that have degree bounded by $O(K2^{K^{1/3}})$.*

Stated in terms of degree, Theorem 1 says that there exists some constant c, such that for all large enough Δ, it is NP-hard to color a $(\log \Delta)^3$-colorable graph of maximum degree bounded by Δ with $O(\Delta^c)$ colors.

2 Preliminaries

In this section we review the basics of Label Cover and PCPs and describe Chan's improved PCP construction.

Let (U, V, E, L, R, Π) be an instance of Label Cover, where $R = dL$ for some constant d, the tuple (U, V, E) is a bipartite graph, vertices in U are assigned labels from $[L]$, and vertices in V are assigned labels from $[R]$. Each edge $e = (u, v)$ is associated with a d-to-1 mapping $\pi_e : [R] \to [L]$. Given a labeling $A : U \to [L], V \to [R]$, we say that the constraint on e is satisfied if $\pi_e(A(v)) = A(u)$. The value of a labeling is the fraction of edges that are satisfied, and the value of a Label Cover instance is the maximum value over all possible labelings of its vertices. The following theorem combines the celebrated PCP theorem [3, 4] with Raz's parallel repetition theorem [23] and shows hardness of Label Cover.

Theorem 2. *For any constant $0 < \sigma < 1$, there are $d, L \leq \mathrm{poly}(1/\sigma)$ such that the problem of deciding satisfiability of a 3-SAT instance with n variables can be Karp-reduced in $\mathrm{poly}(n)$ time to the problem of deciding whether a Label Cover instance of size $\mathrm{poly}(n)$ has value 1 or at most σ. The graph in Label Cover is a bi-regular bipartite graph with left- and right-degrees $\mathrm{poly}(1/\sigma)$.*

As is the case with many inapproximability results, the above Label Cover will be the starting point of our reduction. Formally, let $P : \{-1,1\}^k \to \{-1,1\}$ be a Boolean predicate of arity k, where we follow the convention of having -1 as "True" and 1 as "False". In a Max-P problem, we are given an instance on n Boolean variables $x_1, \cdots, x_n$ with m clauses. All clauses have form $P(l_1, \cdots, l_k)$, where each literal l_i is either a variable or its negation, and the variable of the literals are distinct. The goal of the Max-P problem is to find an assignment to $x_1, \cdots, x_n$ that maximizes the number of clauses satisfied by the assignment. The reduction from Label Cover to Max-P typically translates labelings for $u \in U$ and $v \in V$ to $2^{|L|}$ and $2^{|R|}$ Boolean variables, respectively. These variables are viewed as functions $f^u : \{-1,1\}^{|L|} \to \{-1,1\}$ and $g^v : \{-1,1\}^{|R|} \to \{-1,1\}$. We require that these functions are folded, that is, for any $x \in \{-1,1\}^{|L|}$, $y \in \{-1,1\}^{|R|}$, $f^u(-x) = -f^u(x)$ and $g^v(-y) = -g^v(y)$. For each pair of queries $(x, -x)$, we select one of them. If x is selected, then when $f(-x)$ is needed we return $-f(x)$ instead. Hence in the actual reduction we only use $2^{|L|-1}$ Boolean variables for each $u \in U$ and $2^{|R|-1}$ variables for each $v \in V$. This is also why we need to allow negated literals in the CSP instances. In a correct proof for a satisfiable Label-Cover instance, the functions are long codes for the corresponding labelings of u and v, that is, having $f^u(x) = x_{\sigma_U(u)}$, and $g^v(y) = y_{\sigma_V(v)}$.

Now we describe the clauses in Max-P. For an edge (u, v) in the Label-Cover, we sample *queries*

$$(x^{(1)}, \cdots, x^{(m)}, y^{(m+1)}, \cdots, y^{(k)})$$

according to some carefully chosen *test distribution* $\mathcal{T}$. The distribution $\mathcal{T}$ has the property that for any $l \in L$ and $r \in R$ such that $\pi_{(u,v)}(r) = l$, the predicate P accepts $(f^u(x_l^{(1)}), \cdots, f^u(x_l^{(m)}), g^v(y_r^{(m+1)}), \cdots, g^v(y_r^{(m+1)}))$ with probability 1 (or $1-\varepsilon$ for some constant ε if we are considering non-perfect completeness). One can verify that if the Label Cover instance has value 1 and the test distribution $\mathcal{T}$ satisfies the above property, then any correct proof of a correct labeling has the required completeness. In the soundness analysis, we are given functions f^u and g^v that achieves non-trivial acceptance probability in the above test, and we need to decode those functions and obtain non-trivial labelings of the original Label Cover instance.

In [7], Chan developed a new way of constructing efficient PCPs and proved that the following Hadamard predicate $H_K : \{-1,1\}^K \to \{-1,1\}$ is approximation resistant. For $K = 2^r - 1$, the predicate H_K is on variables $\{x_S\}_{\emptyset \neq S \subseteq [r]}$, defined as

$$H_K(x) = \begin{cases} -1 & \forall S \subseteq [k], |S| > 1, x_S = \prod_{i \in S} x_{\{i\}} \\ 1 & \text{otherwise.} \end{cases}$$

This predicate has $K+1$ accepting assignments. Samorodnitsky and Trevisan [25] showed that H_K is approximation resistance assuming the Unique Games Conjecture — a conjecture stating that finding an approximately optimal solution for a certain special kind of Label Cover is NP-hard. Using his new technique, Chan proved that this is true assuming only $P \neq NP$.

The main idea in Chan's reduction is to consider a direct sum of PCPs. We now sample K edges and run K independent copies of the above test. In the i-th PCP, the i-th query is a uniform random string from $\{-1,1\}^{|L|}$ and all other queries are sampled from $\{-1,1\}^{|R|}$ as described below. In a correct proof, the strategies are expected to be products of long codes encoding the labeling of the vertices.

We now formally define the PCP and how queries are sampled. In the following description, we identify integers from $[K]$ and non-empty subsets of $[r]$ in some canonical way. First we describe the test distribution for a single PCP, indexed by non-empty sets $\emptyset \neq S \subseteq [r]$.

Definition 3. *Let e_S be an edge and π be the constraint on e. Denote the set of possible queries to the T-th position by Q_T, where*

$$Q_T = \begin{cases} \{-1,1\}^{|L|} & T = S \\ \{-1,1\}^{|R|} & T \neq S. \end{cases}$$

The test distribution $\mathcal{T}_{S,e_S}$ is a distribution on $\prod_{T\subseteq[r]} Q_T$. To sample query $(q_T)_{T\subseteq[r]}$ from $\mathcal{T}_{S,e_S}$, first sample q_S from $\{-1,1\}^{|L|}$ uniformly at random. Then, for each $i \in [R]$, let $\{q_{T,i}\}_{T\neq S}$ be a uniformly random accepting assignment of H_K, conditioned on the S-th bit being equal to $q_{S,\pi(i)}$. Finally, independently for each bit, we add noise by resampling from the uniform distribution on $\{-1,1\}$ with probability η.

The final test distribution in the PCP is a product of the above distribution.

Definition 4. *Let (U,V,E,L,R,Π) be a label cover instance. Define $\mathcal{V}_i = V^{i-1} \times U \times V^{K-i}$ for $i \in [K]$. For each $\mathbf{v} \in \mathcal{V}_i$, the proof contains function $\mathbf{f_v} : \left(\{-1,1\}^R\right)^{i-1} \times \{-1,1\}^L \times \left(\{-1,1\}^R\right)^{K-i} \to \{-1,1\}$. The verifier checks the proof as follows:*

1. *Sample independently $K = 2^r - 1$ uniformly random edges $\{e_S\}_{\emptyset\neq S\subseteq[r]}$. Denote $e_S = (u_S, v_S)$.*
2. *Sample queries $\{\mathbf{q}_i\}_{i=1}^K$ from distribution $\prod_{\emptyset\neq T\subseteq[r]} \mathcal{T}_{T,e_T}$.*
3. *Let $\mathbf{v}_i = (v_1,\cdots,v_{i-1},u_i,v_{i+1},\cdots,v_K)$. Accept if $H_K(\mathbf{f}_{\mathbf{v}_1}(\mathbf{q}_1),\cdots,\mathbf{f}_{\mathbf{v}_K}(\mathbf{q}_K)) = -1$.*

In a correct proof, the function $\mathbf{f_v}$ is the product of long codes encoding the labeling of each vertex in $\mathbf{v}$.

Remark. As in the ordinary case, we require that the functions $\mathbf{f_v}$ are folded in the following sense — for any $j \in [K]$, query $\{\mathbf{q}_{j,i}\}_{i\in[K]}$ and $i_0 \in [K]$ we have

$$\begin{aligned} & \mathbf{f_v}(\mathbf{q}_{j,1},\cdots,-\mathbf{q}_{j,i_0},\cdots,\mathbf{q}_{j,K}) \\ = & -\mathbf{f_v}(\mathbf{q}_{j,1},\cdots,\mathbf{q}_{j,i_0},\cdots,\mathbf{q}_{j,K}). \end{aligned}$$

Theorem E.1 along with Theorem A.1, 6.9 and C.2 of Chan [7] shows completeness and soundness of the above reduction and we summarize in the following theorem.

Theorem 5. *Fix some small $\eta, \delta > 0$. Let σ be the soundness of Label Cover, satisfying $\delta = poly(K/\eta) \cdot \sigma^{\Omega(1)}$. Given a Label Cover instance $LC_{L,dL}$, we have the following:*

1. *If $LC_{L,dL}$ has value 1, the above verifier accepts a correct proof with probability at least $1 - K^2\eta$.*
2. *If $LC_{L,dL}$ has value at most σ, then given any proof the verifier accepts with probability at most $(K+1)/2^K + 2\delta$.*

Let $\varepsilon > 0$ be some small constant. In the rest of the paper, let $\delta = \varepsilon \cdot 2^{-K}$, and $\eta = \varepsilon/K^2$. By Theorem 5 of Chan, we require the soundness of Label Cover to be $\sigma = (\delta/\mathrm{poly}(K/\eta))^{O(1)} = 2^{-\Omega(K)}$. This means that the size of the label $L = \mathrm{poly}(1/\sigma) = \exp(\Theta(K))$.

3 Hardness of Approximating Chromatic Number

In this section, we prove Theorem 1 — for sufficiently large K, it is NP-hard to color a K-colorable graph with less than $2^{\Omega(K^{1/3})}$ colors. For convenience of notation, we in fact prove a gap of $O(K^3)$ versus $2^{\Omega(K)}$.

The overall idea is similar to that in Khot [21]. We start by describing the FGLSS graph [12] of the PCP as described in Definition 4. The vertices in the FGLSS graph are function queries and corresponding accepting configurations, denoted as $(\mathbf{f_v}, \mathbf{q}, \mathbf{z})$. The weight of the vertex is the probability that query $(\mathbf{f_v}, \mathbf{q})$ is picked. The total weight of the graph is therefore $K+1$, the number of accepting assignments of H_K. Two vertices are connected if they are clearly inconsistent (returning different answers for the same query to the same function). An independent set in the graph corresponds to a strategy / set of functions, and its weight is the acceptance probability of such strategy. Note that if the maximum weight independent set of the FGLSS graph has weight w, then we need at least $(K+1)/w$ colors to color the whole graph since vertices having the same color must form an independent set.

To use the FGLSS graph for coloring results, we also need to show that if a PCP has acceptance probability $1-\varepsilon$, we can color the FGLSS graph with a small number of colors. Note that in this case, we know that there is an independent set of weight $1-\varepsilon$ in the FGLSS graph, corresponding to a correct proof. Khot's idea in [21] is to modify the definition of the PCP so that the correct proofs are parameterized by some global parameter $\alpha \in \{0,1\}^t$. This gives us 2^t different correct proofs and thus 2^t independent sets of weight $1-\varepsilon$, and by choosing the right t, we expect those independent sets cover most of the FGLSS graph of the modified PCP and thus gives a coloring of at most 2^t colors.

Formally, we modify Definition 4 so that the functions in the proof become $\mathbf{f}_{\mathbf{v}_i} : \left(\{-1,1\}^{R\cdot 2^t}\right)^{i-1} \times \{-1,1\}^{L\cdot 2^t} \times \left(\{-1,1\}^{R\cdot 2^t}\right)^{K-i} \to \{-1,1\}$. Alternatively, we can think of this as modifying Label Cover by appending a t-bit binary string to all the labels and defining the new projection in the Label Cover instance as $\pi'_e(r \circ \alpha) = \pi_e(r) \circ \alpha$ for $r \in R$ and $\alpha \in \{0,1\}^t$, where "$\circ$" denotes string concatenation. The value of this new Label Cover instance is exactly the same as the original setting. Consider the FGLSS graph in this new setting. Soundness is straightforward. If the new proof makes the verifier accept with probability at least $(K+1)/2^K + 2\delta$, then by Theorem 5, the value of the new Label Cover is at least σ and hence the original instance also has value at least σ.

Now let us consider the case of completeness. If the original Label Cover instance has value 1, then extending a valid labeling with any $\alpha \in \{0,1\}^t$ gives us a valid labeling for the modified instance, which corresponds to an independent set of weight at least $1-\varepsilon$ in the modified FGLSS graph. We need to show that the 2^t independent sets corresponding to different $\alpha \in \{0,1\}^t$ cover almost all of the FGLSS graph of the modified PCP. In fact, we can efficiently identify a small fraction of the vertices that contain all vertices that are not covered by any independent sets of the above form and remove them from the FGLSS graph.

To this end, we follow Khot's notation and introduce the following definition characterizing whether we can cover certain vertex with independent sets.

Definition 6. *Consider any K tuples of labelings $\mathbf{l} = \{(l_i, r_i)\}_{i=1}^K$, where $l_i \in [L]$, $r_i \in [R]$ for all $i \in [K]$. Define the i-th mixed labeling $\mathbf{m}_i(\mathbf{l}) = (r_1, \cdots, r_{i-1}, l_i, r_{i+1}, \cdots, r_K)$. Let $\mathbf{f}_{i,\mathbf{l}}$ be the product of long codes encoding the labelings in $\mathbf{m}_i$. Denote by $\mathbf{l}^\alpha := \{(l_i \circ \alpha, r_i \circ \alpha)\}_{i=1}^K$ the labelings extended by α. Define $\mathbf{f}_{i,\mathbf{l}}^\alpha$ similarly.*

A set of queries $\mathbf{q} = (q_1, \cdots, q_K)$ is good if for any K tuples of labelings $\mathbf{l}$ and any accepting assignment $\mathbf{z} = (z_1, \cdots, z_K)$ of the Hadamard predicate, there exists a global *extension α, such that $\mathbf{f}^\alpha_{i,\mathbf{l}}(q_i) = z_i$ for all $i \in [K]$.*

To verify if a set of queries is good, we only need to check all K tuples of labelings and all accepting assignments of the Hadamard predicate H_K. Those are all constants depending only on K (and ε). The following lemma shows that the fraction of bad queries is small.

Lemma 7. *Let t be such that $2^t = C \cdot K^3$ for some large constant C. For large enough K, at most a weighted fraction of $\exp(-O(K))$ of the queries is not good.*

Before proving the lemma, let us see how it leads to our main theorem.

Remove the vertices in the FGLSS graph that correspond to queries that are not good. By Lemma 7, the fraction of vertices removed is bounded by $\exp(-O(K))$. In the soundness case coloring the FGLSS graph still needs at least $(K+1)(1-\exp(-O(K)))/2^{-K} = 2^{\Omega(K)}$ colors. In the completeness case, the Label Cover instance has value 1. Fix a labeling that satisfies all the edges. For a vertex $(\mathbf{f_v}, \mathbf{q}, \mathbf{x})$ in the modified FGLSS graph, let $\mathbf{l_v}$ be the set of K tuples of labelings of the sampled vertices. Each $\alpha \in \{0,1\}^t$ is associated with an independent set consisting of vertices of the form $(\mathbf{f_v}, \mathbf{q}, \mathbf{z})$, where $z_i = \mathbf{f}^\alpha_{i,\mathbf{l_v}}(q_i)$ for all $i \in [K]$.

Consider any vertex $(\mathbf{f_v}, \mathbf{q}, \mathbf{x})$ in the modified FGLSS graph. We know that $\mathbf{q}$ is good so by definition there exists $\alpha_0 \in \{0,1\}^t$ such that $\mathbf{f}^{\alpha_0}_{i,\mathbf{l_v}}(q_i) = x_i$ for all $i \in [K]$. Hence, it is covered by the independent set associated with α_0. Therefore the modified FGLSS graph can be colored with $2^t = O(K^3)$ colors.

Proof of Lemma 7. For query $\mathbf{q}$, let $Q(\mathbf{q})$ be the event that $\mathbf{q}$ is not good in the sense of Definition 6: there exists some labeling $\mathbf{l}$ and some accepting assignment $\mathbf{z}$, such that for any α, there exists $i \in [K]$, $\mathbf{f}^\alpha_{i,\mathbf{l}}(q_i) \neq z_i$. It suffices to bound $\Pr_{\mathbf{q}}[Q(\mathbf{q})]$.

Fix some K tuples of labeling $\mathbf{l}$ of the label cover instance and some accepting assignment $\mathbf{z}$. Consider $\alpha \in \{0,1\}^t$. Over the queries sampled, the probability that $\mathbf{f}^\alpha_{i,\mathbf{l}}(q_i) = z_i$ for all $i \in [K]$ is $1/(K+1)$ before adding noise. To estimate the affect of noise, note that there are K functions, each being a product of K long codes, therefore the answers $\{\mathbf{f}^\alpha_{i,\mathbf{l}}(q_i)\}_{i\in[K]}$ depends on K^2 bits. If none of these K^2 bits are corrupted, then the answer is exactly $\mathbf{z}$. This gives an overall probability of $\Theta(1/K \cdot (1-\eta)^{K^2}) = \Theta(e^{-\eta K^2}/K) = \Theta(1/K)$. The contribution of probability from other sources is negligible.

Note that for different extension α, the bits that $\mathbf{f}^\alpha_{i,\mathbf{l}}$ reads from $\mathbf{q}$ are completely independent, so we have

$$\Pr_{\mathbf{q}}\left[\forall \alpha, \exists i, \mathbf{f}^\alpha_{i,\mathbf{l}}(q_i) \neq z_i\right] = (1-\Theta(1/K))^{2^t} = \exp(-O(2^t/K)).$$

Picking large enough constant C and taking union bound over all possible labelings and accepting configurations, we get that the weighted fraction of $\mathbf{q}$ that are bad is

$$\Pr_{\mathbf{q}}[Q(\mathbf{q})] \leq R^{K-1} \cdot L \cdot (K+1)\exp(-O(2^t/K)) = \exp(-O(K)).$$

□

Now let us consider the degree of the graph produced by the above reduction. Consider a vertex $(\mathbf{f_v}, \mathbf{q}, \mathbf{z})$. Fix some $i \in [K]$. Let $\mathbf{z}'$ be some accepting assignment of H_K with $z'_i \neq z_i$. We first estimate the number of queries $\mathbf{q}'$ with $q'_i = q_i$. Let us consider the i-th test distribution $\mathcal{T}_{i,e_i}$, where e_i is the edge sampled for the i-th test, and denote the constraint on e_i by π. Recall that for each $l \in [L]$ and $r \in \pi^{-1}(l) \subseteq [R]$, the bits $\{q'_{j,r}\}_{j\neq i}$ are sampled by uniformly picking an accepting assignment $\mathbf{x}$ of H_K conditioned on $x_i = q'_{i,l}$. Thus there are at least $((K+1)/2)^{|R|} = 2^{\exp(\Omega(K))}$ possible choices of $\mathbf{q}'$.

Note that for any such $\mathbf{q}'$, there is an edge between $(\mathbf{f_v}, \mathbf{q}', \mathbf{z}')$ and $(\mathbf{f_v}, \mathbf{q}, \mathbf{z})$. Therefore the degree of the graph produced by the above reduction is at least double exponential in K. We now use the approach in Clementi and Trevisan [9] and Trevisan [26] to reduce the degree to $O(K^3 2^K)$.

For ease of presentation, we look at the argument on the original FGLSS graph without removing bad queries. The same argument applies to the graph with bad queries removed because removing vertices from the graph does not increase the maximum degree, and, as argued above, does not significantly affect the soundness and completeness of the reduction.

Denote the FGLSS graph corresponding to the PCP in Theorem 5 as G. We first turn G into an unweighted graph. Let w_{min} be the minimum weight of vertices in G, and λ be the ratio between the minimum and maximum weight of vertices in G. Note that λ depends only on K. Let ξ be some granularity parameter. We obtain an unweighted version G' of G by duplicating vertices — we make $\lfloor w/w_{min} \cdot 1/\xi \rfloor \leq 1/\lambda\xi$ vertices for a vertex of weight w, and connect the duplicated vertices with all the neighbors. This step blows up the size of the graph by $O(1/\lambda^2\xi^2)$, and the fractional size of the maximum independent set in G' is within a multiplicative factor of $O(\xi)$ from that of G due to error introduced by $\lfloor \cdot \rfloor$ when duplicating vertices.

As observed in [26], the graph G' is a union of bipartite complete subgraphs. More precisely, for every index i and i-th query $(\mathbf{f}_{\mathbf{v}_i}, \mathbf{q}_i)$, there is a complete bipartite graph between configurations that answer zero for query $(\mathbf{f}_{\mathbf{v}_i}, \mathbf{q}_i)$ — denoted as $Z_{\mathbf{f}_{\mathbf{v}_i}, \mathbf{q}_i}$ — and configurations that answer one for the same query — denoted as $O_{\mathbf{f}_{\mathbf{v}_i}, \mathbf{q}_i}$. Let l be the maximum size of such sets. We claim that l depends only on K, λ and ξ. To estimate l, consider how many tuples $(\mathbf{f_v}, \mathbf{q}, \mathbf{z})$ can include $(\mathbf{f}_{\mathbf{v}_i}, \mathbf{q}_i)$ on the i-th position. By Theorem 2, the degree of the Label Cover graph is $\text{poly}(1/\sigma) = \exp(\Theta(K))$, thus the $\mathbf{f}_{\mathbf{v}_i}$ coordinate has at most $\exp(\Theta(K^2))$ neighbors. For $\mathbf{q}_i$, consider an edge e the bits in $\mathbf{q}_i$ that are mapped to the same label $l \in [L]$ according to mapping π_e (or a single bit if e is the i-th edge). There are exactly $(K+1)/2$ possible queries. Enumerating over all labels and sampled edges, this gives an upper-bound of $2^{\exp(\Theta(K))}$. Since each of them can be duplicated by at most $1/\lambda\xi$ times, we have $l = 2^{\exp(\Theta(K))}/\lambda\xi$. Also since for each input bit to the predicate H_K, exactly half of the accepting assignments of H_K set that bit to 1 and exactly half to -1 — a property also known as H_K being balanced — we have $|Z_{\mathbf{f}_{\mathbf{v}_i}, \mathbf{q}_i}| = |O_{\mathbf{f}_{\mathbf{v}_i}, \mathbf{q}_i}|$.

We now replace the above bipartite complete graphs in G' with the following construction on the same set of vertices $Z_{\mathbf{f}_{\mathbf{f}_i}, \mathbf{q}_i}$ and $O_{\mathbf{f}_{\mathbf{f}_i}, \mathbf{q}_i}$.

Proposition 8 ([26])**.** *For every $\zeta > 0$ and $b \geq 1$, there is a bipartite graph $([b], [b], E)$ of degree at most $d = 3\zeta^{-1}\log(\zeta^{-1})$ such that for any $A, B \subseteq [b]$, $|A| \geq \lfloor \zeta b \rfloor$, $|B| \geq \lfloor \zeta b \rfloor$, we have $(A \times B) \cap E \neq \emptyset$.*

Trevisan [26] called such graphs (b, ζ)-dispersers, and he used a probabilistic argument to prove the above proposition. As argued above, l is a constant depending only on K, thus we can find the desired disperser by exhaustive search. An important property of bipartite dispersers is that given an independent set I of a (b, ζ)-disperser, we have that either $|I \cap A| \leq \zeta b$ or $|I \cap B| \leq \zeta b$.

Denote the replaced graph by G''. To understand how much the above replacement step increases the size of the maximum independent set, note that for any independent set in a disperser, we can get an independent set in the complete bipartite graph by discarding all vertices on one side, which is at most a ζ fraction if we choose the smaller side. Also, each vertex in the FGLSS graph is involved in at most K complete bipartite graphs of this kind, thus the size of the independent set in the new graph is at most $K\zeta$ larger than G'. By choosing $\zeta = O(2^{-K}/K)$, $\xi = O(2^{-K})$, we have that in the soundness case the maximum independent set G' has size $O(2^{-K})$. The maximum degree of G'' is bounded by $K \cdot 3\zeta^{-1}\log(\zeta^{-1}) = O(K^3 2^K)$.

4 Discussions

In this paper, we proved a gap of K^3 vs. $2^{\Omega(K)}$ for approximating chromatic number. Let us take a closer look at how we get to the power 3 in K^3. The soundness of the Label Cover problem has to be at most $2^{-\Omega(K)}$, which means that the size of the labels are $\exp(\Theta(K))$. Definition 6 involves all possible labelings and accepting assignments of K. The reduction in Definition 4 samples K edges, therefore there would be $R^{K-1}\cdot L = \exp(\Theta(K^2))$ possible labelings and a union bound results in a factor of $\exp(\Theta(K^2))$ in the probability of a query being bad. The other factor of K is due to the fact that the Hadamard predicate H_K has $K+1$ accepting assignments and they are sampled uniformly.

Acknowledgements

The author is grateful to Siu On Chan for pointing out Theorem E.1 in [7], leading to a gap of K vs. $2^{K^{1/3}}$, improving from a gap of K vs. $2^{K^{1/5}}$ in an earlier version of this manuscript.

The author is supported by ERC Advanced Investigator grant 226203.

References

[1] *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*. IEEE Computer Society, 2012.

[2] S. Arora, E. Chlamtac, and M. Charikar. New approximation guarantee for chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 215–224, New York, NY, USA, 2006. ACM.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[4] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998.

[5] B. Berger and J. Rompel. A better performance guarantee for approximate graph coloring. *Algorithmica*, 5(3):459–466, 1990.

[6] A. Blum and D. R. Karger. An $\tilde{O}(n^{3/14})$-coloring algorithm for 3-colorable graphs. *Inf. Process. Lett.*, 61(1):49–53, 1997.

[7] S. O. Chan. Approximation resistance from pairwise independent subgroups. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:110, 2012.

[8] E. Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *FOCS*, pages 691–701. IEEE Computer Society, 2007.

[9] A. E. F. Clementi and L. Trevisan. Improved non-approximability results for minimum vertex cover with density constraints. *Theor. Comput. Sci.*, 225(1-2):113–128, 1999.

[10] I. Dinur, S. Khot, W. Perkins, and M. Safra. Hardness of finding independent sets in almost 3-colorable graphs. In *FOCS*, pages 212–221. IEEE Computer Society, 2010.

[11] I. Dinur, E. Mossel, and O. Regev. Conditional hardness for approximate coloring. *SIAM J. Comput.*, 39(3):843–873, 2009.

[12] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost NP-complete (preliminary version). In *FOCS*, pages 2–12. IEEE Computer Society, 1991.

[13] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[14] V. Guruswami and S. Khanna. On the hardness of 4-coloring a 3-colorable graph. *SIAM J. Discrete Math.*, 18(1):30–40, 2004.

[15] V. Guruswami and A. K. Sinop. The complexity of finding independent sets in bounded degree (hyper)graphs of low chromatic number. In D. Randall, editor, *SODA*, pages 1615–1626. SIAM, 2011.

[16] G. Hast. Beating a random assignment. *PhD Thesis*, 2005.

[17] J. Håstad and S. Khot. Query efficient PCPs with perfect completeness. *Theory of Computing*, 1(1):119–148, 2005.

[18] K. ichi Kawarabayashi and M. Thorup. Combinatorial coloring of 3-colorable graphs. In *FOCS* [1], pages 68–75.

[19] D. R. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998.

[20] S. Khanna, N. Linial, and S. Safra. On the hardness of approximating the chromatic number. *Combinatorica*, 20(3):393–415, 2000.

[21] S. Khot. Improved inaproximability results for maxclique, chromatic number and approximate graph coloring. In *FOCS*, pages 600–609. IEEE Computer Society, 2001.

[22] S. Khot and R. Saket. Hardness of finding independent sets in almost q-colorable graphs. In *FOCS* [1], pages 380–389.

[23] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

[24] A. Samorodnitsky and L. Trevisan. A PCP characterization of NP with optimal amortized query complexity. In F. F. Yao and E. M. Luks, editors, *STOC*, pages 191–199. ACM, 2000.

[25] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. *SIAM J. Comput.*, 39(1):323–360, 2009.

[26] L. Trevisan. Non-approximability results for optimization problems on bounded degree instances. In J. S. Vitter, P. G. Spirakis, and M. Yannakakis, editors, *STOC*, pages 453–461. ACM, 2001.

[27] A. Wigderson. A new approximate graph coloring algorithm. In H. R. Lewis, B. B. Simons, W. A. Burkhard, and L. H. Landweber, editors, *STOC*, pages 325–329. ACM, 1982.