New Tools for Graph Coloring*

Sanjeev Arora
arora@cs.princeton.edu

Rong Ge
rongge@cs.princeton.edu

Department of Computer Science and
Center for Computational Intractability,
Princeton University

Abstract

How to color 3 colorable graphs with few colors is a problem of longstanding interest. The best polynomial-time algorithm uses $n^{0.2072}$ colors. There are no indications that coloring using say $O(\log n)$ colors is hard. It has been suggested that lift and project based SDP relaxations could be used to design algorithms that use n^ϵ colors for arbitrarily small $\epsilon > 0$.

We explore this possibility in this paper and find some cause for optimism. While the case of general graphs is till open, we can analyse the Lasserre relaxation for two interesting families of graphs.

For graphs with low *threshold rank* (a class of graphs identified in the recent paper of Arora, Barak and Steurer on the unique games problem), Lasserre relaxations can be used to find an independent set of size $\Omega(n)$ (i.e., progress towards a coloring with $O(\log n)$ colors) in $n^{O(D)}$ time, where D is the threshold rank of the graph. This algorithm is inspired by recent work of Barak, Raghavendra, and Steurer on using Lasserre Hierarchy for unique games. The algorithm can also be used to show that known integrality gap instances for SDP relaxations like *strict vector chromatic number* cannot survive a few rounds of Lasserre lifting, which also seems reason for optimism.

For *distance transitive* graphs of diameter Δ, we can show how to color them using $O(\log n)$ colors in $n^{2^{O(\Delta)}}$ time. This family is interesting because the family of graphs of diameter $O(1/\epsilon)$ is easily seen to be *complete* for coloring with n^ϵ colors. The distance-transitive property implies that the graph "looks" the same in all neighborhoods.

*Research supported by NSF Grants CCF-0832797, 0830673, and 0528414

1 Introduction

In the graph coloring problem we are given a graph $G = (V, E)$. A *coloring with t colors* is a function $f : V \to [t]$, such that for any $(p, q) \in E$, $f(p) \neq f(q)$. The smallest t such that a coloring exists is called the *chromatic number* of the graph, and the graph is said to be *t-colorable*.

Despite much research we still have no good coloring algorithms even in very restricted cases. This is explained to some extent because it is NP-hard to approximate the chromatic number of a graph up to a factor of $n^{1-\epsilon}$ for any $\epsilon > 0$ ([24], following a long line of work in PCPs). Therefore attention has shifted to the case where the graph is 3-colorable. In this restricted case known algorithms can color the graph using $\tilde{O}(n^c)$[1] colors for some constants c. Wigderson's purely combinatorial algorithm [23] works for $c = 1/2$. Using more combinatorial tools Blum achieved $c = 3/8$ [7]. Karger, Motwani, and Sudan [16] used SDP relaxations to achieve $c = 1/4$, which was combined with combinatorial tools to achieve $c = 3/14$ by Blum and Karger [8]. Arora, Charikar and Chlamtac [3] then carefully analyzed the SDP relaxation to reduce c to 0.2111. Chlamtac [10] further reduced c to 0.2072 using $O(1)$ levels of Lasserre lifting of the basic SDP relaxation (Lasserre lifting is defined in Section 2).

The seeming difficulty in getting even small improvements in c suggests that substantial improvement to c (achieving $c = o(1)$ for example) is intractable, but few lowerbounds are known. Dinur, Mossel and Regev[13] showed that it's hard to color with any constant number of colors (i.e., $O(1)$ colors) based on a variant of Unique Games Conjecture. Some integrality gap results[14, 16, 22] show that the simple SDP relaxation has an integrality gap at least $n^{0.157}$.

Arora et al. [3] suggested that using $O(1)$ or $O(\log n)$ levels of Lasserre lifting on the standard SDP relaxation should allow us to find an n^ϵ-coloring (running time would be $n^{O(k)}$ where k is the number of levels). In general researchers have hoped that a few rounds of lift-and-project strengthening of SDP relaxations (via Lasserre or other methods) should allow better algorithms for many other problems, though few successes have resulted in this endeavor. A notable example is a result of Chlamtac and Singh [11] which allows us to find large independent sets in hypergraphs using $O(1/\gamma^2)$ levels of Lasserre lifting when an independent set of size γn is guaranteed to exist.

The current paper is related to recent developments about the unique games problem. A surprising recent result of Arora, Barak and Steurer [1] showed that unique games can be solved in subexponential time using the idea of *threshold rank*. In the same paper they also gave better algorithms for problems such as MAX-CUT when the threshold rank is low. More recently, Barak, Raghavendra and Steurer [5] showed that the surprising subexponential algorithm for unique games can be rederived using Lasserre lifting. Their rounding algorithm involves a new convex programming relaxation for threshold rank which we also use in a key way. It gives a way to round the SDP solution by showing that the solution vectors exhibit "global correlation."

We extend the techniques of Barak et.al. to show that low threshold rank also helps in coloring 3-colorable graphs with fewer colors. Our algorithm is also derived using Lasserre liftings. In general we think our approach may lead to n^ϵ-coloring in subexponential or even quasi-polynomial time (especially by combining with the Blum coloring tools). We note in Section C that known integrality gap examples for 3-coloring SDPs do not work after polylog levels of Lasserre lifting.

Notice that subexponential time algorithms are believed to not exist (assuming the exponential time conjecture) for *exact* 3-coloring problem because there is a linear size reduction from 3-SAT to 3-$Coloring$. However, conceivably a $O(\log n)$ coloring could be done in polynomial time.

[1] Here and throughout the paper $\tilde{O}$ hides logarithmic factors.

1.1 Our Results

The difficulty in using Lasserre liftings for colorings as well as any other problem is the lack of an obvious rounding algorithm. The paper [5] gives such a rounding algorithm for the unique games problem for graphs of low *threshold rank*. Our first result is a similar algorithm for graph coloring. We state the theorem here and will prove it in Section 4. The hypothesis uses a somewhat different notion of threshold rank than [5].

Theorem 1. *There is a constant $c > 1$ and a randomized rounding algorithm such that the following is true. If a regular 3-colorable graph G has threshold rank $Rank_{-1/16}(G)$ (i.e., the number of eigenvalues less than $-1/16$, where eigenvalues are scaled to lie in $[-1, 1]$) at most D, then the algorithm can find an independent set of size at least $n/12$ in time $n^{O(D)}$ with constant probability. Moreover, if the graph is vertex-transitive (see Section A.2), there is a randomized algorithm that finds a coloring with $O(\log n)$ colors in $n^{O(D)}$ time.*

As a corollary of the above result we can show that existing "counterexamples" for graph coloring algorithms (eg examples that exhibit high integrality gap for strict vector chromatic number [15]) cannot be counterexamples for high level Lasserre liftings since they all have low threshold rank (see Section C). (Note that a similar result for unique games was a precursor to the ultimate subexponential algorithm in ABS[1] which works for all instances.)

When we try to apply similar ideas to general graphs, we quickly realize that the problematic cases (if they exist at all) must be such that different neighborhoods look "different." Of course, this flies against the usual intuition about SDP relaxations: the usual reason for high integrality gaps (at least in explicit examples) is precisely that all neighborhoods look the same and the SDP gives no meaningful clues.

To quantify the notion of all neighborhoods "looking the same," we focus on a specific kind of symmetric graph, the *distance transitive graphs*, which have been well-studied in graph theory (see the book [9]). In fact we restrict attention to such graphs that in addition have low diameter. The reason is that using simple combinatorial arguments one can show that in order to color the graph with n^ϵ colors, it suffices to restrict attention to graphs of diameter $O(1/\epsilon)$. (This is proved similarly to Theorem 15 in Section D). If a distance transitive graph has diameter Δ and $Lasserre_{2^{O(\Delta)}}$ number 3, we show how to find a $O(\log n)$ coloring in $O(n^{2^{O(\Delta)}})$ time. See Section 5.

How can our ideas be generalized to all graphs? In Section D we show that this could be done if we can relate the vertex expansion properties of small sets to notions like threshold rank, and formulate a conjecture which if true would yield subexponential time coloring algorithms that find an n^ϵ-coloring.

2 The SDP and Lasserre Hierarchy

2.1 Lasserre Hierarchy for Coloring SDP

The standard SDP relaxation for graph 3-coloring uses *vector chromatic number*, but it is not amenable to Lasserre lifting. So we start with an equivalent (see [10]) relaxation based upon 0/1 variables. For each vertex p of the graph $G = (V, E)$, we have three variables $x_{p,R}, x_{p,Y}, x_{p,B}$ where $x_{p,C} = 1$ for $C \in \{R, Y, B\}$ "means" the vertex p is colored with color C. Thus exactly one of the

three variables will be 1. The integer program is as follows:

$$\begin{aligned}
\forall p \in V \quad & x_{p,R} + x_{p,B} + x_{p,Y} = 1 \\
\forall p \in V \quad & x_{p,R}x_{p,B} = x_{p,R}x_{p,Y} = x_{p,B}x_{p,Y} = 0 \\
\forall (p,q) \in E, C \in \{R,Y,B\} \quad & x_{p,C}x_{q,C} = 0 \\
\forall p \in V, C \in \{R,Y,B\} \quad & x_{p,C} \in \{0,1\}
\end{aligned}$$

Clearly solutions to this program correspond to valid 3-colorings of the graph G. Now we relax this integer program by replacing each $x_{p,C}$ with a vector $v_{p,C}$. The result is an SDP relaxation. Then we lift this SDP using k levels of Lasserre lifting. (For Lasserre lifting see the surveys[12, 18]).The lifted SDP contains vector variables v_S, where S is a subset of the set $V \times \{R,Y,B\}$ (later denoted by Ω) and has size at most k. The resulting SDP is

$$\begin{aligned}
\forall p \in V \quad & v_{p,R} + v_{p,B} + v_{p,Y} = v_\emptyset && (1) \\
\forall p \in V, C_1 \neq C_2 \quad & \langle v_{p,C_1}, v_{p,C_2} \rangle = 0 && (2) \\
\forall (p,q) \in E, C \in \{R,Y,B\} \quad & \langle v_{p,C}, v_{q,C} \rangle = 0 && (3) \\
\forall P,Q,S,T \subseteq \Omega, P \bigcup Q = S \bigcup T, |P \bigcup Q| \leq k \quad & \langle v_P, v_Q \rangle = \langle v_S, v_T \rangle && (4) \\
& \langle v_\emptyset, v_\emptyset \rangle = 1 && (5)
\end{aligned}$$

In this SDP, Equations (1) to (3) are constraints obtained from the integer program; Equations (4) are the consistency constraints imposed by Lasserre lifting; Equation (13) is just a normalization. Notice that here we are abusing notation a bit: if the set S contains only one event (p,C), we use both $v_{p,C}$ and v_S for the same vector. We call this SDP Las^k and its solution SDP^k.

If we replace the set of colors $\{R,Y,B\}$ with a set of size t, this SDP becomes an SDP relaxation for t-coloring. We define the $Lasserre_k$ number of a graph G as a relaxation of the chromatic number; it is a generalization of existing notions such as *vector chromatic number*.

Definition 1. The $Lasserre_k$ number for a graph G is the smallest number of colors t such that the t-color SDP relaxation for this graph is still feasible after k-level Lasserre lifting.

For any k, the $Lasserre_k$ number of a graph is always at most the true chromatic number, because the integral solution is always feasible for the Lasserre lifting. In particular, if G is 3-colorable, its $Lassere_k$ number is at most 3 for any k. The $Lasserre_1$ number is known to be equivalent to strong vector chromatic number defined in [16] (for proof see [10]).

2.2 Understanding the SDP Solution

It is well known that the vectors in the SDP solution can be viewed as giving locally consistent distributions over solutions (see [17, 12, 10]). Here we discuss how we should interpret the solution of coloring SDP.

Throughout the discussion below, an "atomic event" (abbreviated to just "event" when this causes no confusion) consists of assigning a vertex p some color C. We denote by $\Omega = V \times \{R,Y,B\}$ the set of all atomic events. Our rounding algorithm will iteratively assign colors to vertices. Each step may assign a color C to p, or declare that color C will *never* be assigned to p. In the former case the atomic event (p,C) has happened; in the latter case the complement event happened. It is common to interpret the SDP solution as giving a distribution over these events whose probabilities are equal to the innerproducts of the vectors. We formulate this by the following theorem:

Theorem 2 (Lasserre[17, 10]). *A solution to k-level Lasserre lifting SDP (Las^k) encodes a locally consistent coloring for any set of k vertices. Locally consistent means all colorings with positive probability are valid colorings. If W is a set of atomic events then the probability that they happen is equal to the inner-product of v_S and v_T, where $S \bigcup T = W$. In particular, each vector can be decomposed as $v_W = r_W v_\emptyset + u_W$ where r_W is the probability that all events in W happen and u_W is perpendicular to $v_\emptyset$.*

If $w = (p, C)$ is an atomic event, properties of Lasserre lifting allow us to construct a subsolution in which event w happens (ie vertex p is assigned colored C), and a subsolution in which event w does not happen (ie color C is forbidden for p from now on). We randomly choose one of the subsolutions to preserve the probability of w. That is, if r_w is the probability of event w, we pick the subsolution in which w happens with probability r_w, and pick the subsolution in which w does not happen with probability $1-r_w$. We call this *"conditioning the solution on event w"*. The result of such an operation will be a solution for $k-1$ level of Lasserre lifting, which we call SDP^{k-1}.

The computation needed to compute the new vectors in SDP^{k-1} is simple and follows from the above theorem: the probabilities of the new $k-1$-level solution must be the appropriate conditional probabilities in the locally consistent distributions in the k-level solution.

The atomic event w happens with probability $\langle v_w, v_\emptyset \rangle$ in the above-mentioned locally consistent solution. If w happens then this means we can normalize v_w to be the new $v'_\emptyset$. The new vector for set S is $v'_S = v_{S \bigcup \{w\}} / \|v_w\|_2$. Further, w does not happen probability $1 - \langle v_w, v_\emptyset \rangle$. Thus if w does not happen then one of the events w_1 and w_2, which involve coloring p with the other two colors, will happen. The new vector for a set S will be $v'_S = (v_{S \bigcup \{w_1\}} + v_{S \bigcup \{w_2\}}) / \|v_{w_1} + v_{w_2}\|_2$.

Note that we must use Lasserre instead of weaker relaxations: Sherali-Adams[20] and Lovász-Schrijver[19]. The reason is that we consider solutions as locally consistent solutions (which rules out Lovász-Schrijver) and we use critically that probabilities correspond to inner-products of vectors (which rules out Sherali-Adams). Detailed comparison between the three hierarchies are given in the two surveys [12, 18]

3 Global Correlation, Local Correlation and Rounding

Given a solution to the k-level lifting Las^k of a graph G, we shall define the global correlation of this solution and show how global correlation of order $\Omega(1/k)$ can help to round the solution. Intuitively global correlation measures the average correlation between the colors of two vertices chosen uniformly at random. In general, this correlation may be close to 0: knowing the color of one doesn't give much information about the color of the other. If the global correlation is bounded away from 0 however, then intuitively speaking, fixing the color for a randomly chosen vertex should bias the average remaining vertex a bit towards a particular color. Thus after fixing the colors for a sufficiently large set of vertices, the colors for most of the remaining vertices must get more or less fixed. This the main idea of Barak et.al.[5] in the context of unique games, and Lemma 3 is adapted from there. The amount of variability in the color of the average vertex is quantified using *variance*.

We first examine how conditioning on one atomic event reduces the variance of another. Let w_1, w_2 be two atomic events, r_1, r_2 be their probabilities respectively, and r_{12} be the probability that both of them happen. The variance of the conditional random variable $w_2|w_1$ is given by:

$$\mathrm{Var}[w_2|w_1] = r_1(\frac{r_{12}}{r_1})(1 - \frac{r_{12}}{r_1}) + (1 - r_1)(\frac{r_2 - r_{12}}{1 - r_1})(1 - \frac{r_2 - r_{12}}{1 - r_1}) = \mathrm{Var}[w_2] - \frac{(r_1 r_2 - r_{12})^2}{\mathrm{Var}[w_1]}. \quad (6)$$

By the equation we see that the expected variance always drops, and the drop is proportional to $(r_1 r_2 - r_{12})^2$. Below we call this quantity the *correlation* between the two events.

Correlation has a geometric meaning in Lasserre solutions. Notice that $r_1 = \langle v_{w_1}, v_\emptyset \rangle$, $r_2 = \langle v_{w_2}, v_\emptyset \rangle$, and $r_{12} = \langle v_{w_1} v_{w_2} \rangle$ (by Theorem 2). As in Theorem 2 we express $v_{w_i} = r_i v_\emptyset + u_{w_i}$, then $\langle u_{w_1}, u_{w_2} \rangle = \langle v_{w_1}, v_{w_2} \rangle - r_1 r_2 = r_{12} - r_1 r_2$. Therefore we have $(r_1 r_2 - r_{12})^2 = \langle u_{w_1}, u_{w_2} \rangle^2$.

Definition 2 (Correlation, Global Correlation, Variance)**.** Given a solution SDP^k and two events w_1, w_2, The *correlation* between w_1 and w_2 is defined as (where probabilites r and vectors u are as in Theorem 2):

$$Cor[w_1, w_2] = (r_{w_1} r_{w_2} - r_{\{w_1, w_2\}})^2 = \langle u_{w_1}, u_{w_2} \rangle^2.$$

The *global correlation* of a set of vectors $\{z_p\}$ $(p \in U)$ is just the expected correlation between two randomly picked vectors:

$$GC(\{z_p\}) = \mathop{\mathrm{E}}_{p,q \in U} \langle z_p, z_q \rangle^2.$$

The *global correlation* of the SDP solution is the global correlation of all the u vectors for the set of atomic events (Ω). Intuitively it is the average correlation between a pair of atomic events.

$$GC^k = \mathop{\mathrm{E}}_{w_1, w_2 \in \Omega} \langle u_{w_1}, u_{w_2} \rangle^2. \tag{7}$$

The *variance* of the solution is

$$VAR^k = \mathop{\mathrm{E}}_{w \in \Omega} r_w (1 - r_w). \tag{8}$$

Now we are ready to state the main lemma in this Section.

Lemma 3. *Suppose SDP solution SDP^k has global correlation GC^k and variance VAR^k. Upon picking a random event $w \in \Omega$ and conditioning on that event, the new solution SDP^{k-1} has expected variance at most $VAR^k - 4GC^k$. (the expectation is over the random choice of the event and whether the event happened or not)*

Proof. Pick a random event w_1 in the solution SDP^k, and condition the whole solution on this event. By definition of global correlation (7) we know

$$\mathop{\mathrm{E}}_{w_1} \mathop{\mathrm{E}}_{w_2} \langle u_{w_1}, u_{w_2} \rangle^2 = GC^k.$$

But since w_1 is a $0, 1$ event, $Var[w_1]$ is always at most $1/4$. By Equation (6), for any event w_2, the variance $\mathrm{Var}[w_2|w_1] \leq \mathrm{Var}[w_2] - 4Cor[w_1, w_2]$. Using this we have

$$\mathop{\mathrm{E}}_{w_1} VAR^{k-1} = \mathop{\mathrm{E}}_{w_1} \mathop{\mathrm{E}}_{w_2} \mathrm{Var}[w_2|w_1] \leq \mathop{\mathrm{E}}_{w_1} \mathop{\mathrm{E}}_{w_2} (\mathrm{Var}[w_2] - 4Cor[w_1, w_2]) = VAR^k - 4GC^k.$$

□

Lemma 3 corresponds to a single step in our iterative rounding. So long as the global correlation is substantial —say, at least $10/k$– we can repeat this step up to k times and drive the variance of the solution towards zero. Intuitively, once the variance is small enough, the solution should be almost integral and thus easy to round. Indeed we show the following:

Lemma 4. *Given a vector solution SDP^k ($k \geq 2$) for k-level Lasserre lifting Las^k , we say a vertex p is* determined *if there's a color C such that event $\{p, C\}$ that happens with probability more than $1/2$. Otherwise the vertex is* undetermined*. If SDP^k has variance $VAR^k < 1/8$ then at least $1/4$ of the vertices are determined. Moreover, if we color the determined vertices with the color that makes them determined, then this is a valid partial coloring (i.e., no two adjacent vertices will have the same color).*

Proof. First rewrite the definition of variance as

$$VAR^k = \mathop{\mathrm{E}}_{w\in\Omega} r_w(1-r_w) = \mathop{\mathrm{E}}_{p\in V}\mathop{\mathrm{E}}_{C\in\{R,Y,B\}} r_{(p,C)}(1-r_{(p,C)}).$$

From this formula we know for any vertex p and its 3 events w_1, w_2, w_3, their contribution to VAR^k is proportional to $(Var[w_1]+Var[w_2]+Var[w_3])/3$ (the second expectation in the right hand side). For undetermined vertices, the probabilities for w_1, w_2, w_3 can be more than $1/2$ and they sum up to 1, thus the minimum possible value of the contribution of this vertex p is $(Var[w_1]+Var[w_2]+Var[w_3])/3$ is $(1/4+1/4+0)/3 = 1/6$. If more than $3/4$ of the vertices are undetermined, we would have $VAR^k > 3/4\cdot 1/6 = 1/8$, which contradicts our assumption.

For the moreover part, notice that the solution SDP^k is valid for the second level of Lasserre, which means it induces locally consistent distributions for any two vertices. For any edge (p,q) in the graph, if both p and q have events $\{p,C_1\}$, $\{q,C_2\}$ that happen with probability more than $1/2$, then we show $C_1 \neq C_2$. Suppose for contradition that $C_1 = C_2 = C$. If we look at the distribution that the Lasserre solution induces on these two vertices, with positive probabililty both of them will be colored with color C. This contradicts with the validity of the Lasserre solution. Therefore we must have $C_1 \neq C_2$. □

Local correlation For an SDP solution, we would want to argue either Lemma 4 can be applied or the solution has large global correlation. To show this, we introduce local correlation as an intermediate step. We first show that if we cannot apply Lemma 4, the solution SDP^k always has *local correlation*, then we analyze the relationship between local correlation and global correlation in the next section and show high local correlation implies high global correlation.

For a vertex p, we construct a new vector $z_p = (u_{p,R}, u_{p,B}, u_{p,Y})$ (which means z_p is the concatenation of the 3 vectors, the vector u comes from Theorem 2). It's easy to see that $\langle z_p, z_q\rangle = \sum_{C\in\{R,Y,B\}}\langle u_{p,C}, u_{q,C}\rangle$. Since $\langle z_p, z_q\rangle^2 \leq 27\,\mathrm{E}_{C_1,C_2\in\{R,Y,B\}}\langle u_{p,C_1}, u_{q,C_2}\rangle^2$, we know $GC \geq 1/27\cdot \mathrm{E}_{p,q\in V}\langle z_p, z_q\rangle^2$. Hence the global correlation of the $\{z_p\}$ vectors $\mathrm{E}_{p,q\in V}\langle z_p, z_q\rangle^2$ can be used to lowerbound the global correlation of the solution SDP^k.

Local correlation is the expected correlation between endpoints of edges.

Definition 3 (Local Correlation)**.** Given a graph G and an SDP solution SDP^k, first construct vectors $z_p = (u_{p,R}, u_{p,B}, u_{p,Y})$. Then local correlation for this solution is defined to be

$$LC = \mathop{\mathrm{E}}_{(p,q)\in E}\langle z_p, z_q\rangle. \tag{9}$$

Local correlation depends on both the solution (SDP^k) and the graph G, unlike global correlation which only depends on the solution. Also, local correlation can be negative because we are not taking the squares of inner-products.

We shall prove the following Lemma which ensures high local correlation until we can find a large independent set.

Lemma 5. *If G is a regular graph with Lasserre$_k$ number at most 3, and in an SDP solution SDP^k at most $n/4$ vertices are determined in the sense of Lemma 4 then the local correlation* $\mathrm{E}_{(p,q)\in E}\langle z_p, z_q\rangle \leq -1/8$.

Proof. If (p,q) is an edge, both p and q have no event with probability more than $1/2$ (these vertices were called undetermined in Lemma 4), we shall prove $\langle z_p, z_q\rangle \leq -1/4$. Indeed, since (p,q) is an edge by (2) we have $\langle v_{p,R}, v_{q,R}\rangle = r_{p,R}r_{q,R} + \langle u_{p,R}, u_{q,R}\rangle = 0$. Which means $\langle u_{p,R}, u_{q,R}\rangle$ is always

non-positive. In particular we have $\langle z_p, z_q\rangle = -r_{p,R}r_{q,R} - r_{p,Y}r_{q,Y} - r_{p,B}r_{q,B} \leq -1/4$. The last inequality holds because the r values are all smaller than $1/2$, the worst case happens when the probabilities are $(1/2, 1/2, 0)$ and $(0, 1/2, 1/2)$.

Since only $1/4$ of the vertices are determined (in the sense of Lemma 4), we consider the set S of undetermined vertices. At least $1/2$ of edges of G have both endpoints in S. Therefore $\mathrm{E}_{(p,q)\in E}\langle z_p, z_q\rangle \leq -1/4 * 1/2 = -1/8$. □

Actually a stronger version of the Lemma is true:

Lemma 6. *Suppose G is a regular graph with $Lasserre_k$ number at most 3, and in an SDP solution SDP^k at most $n/4$ vertices are determined in the sense of Lemma 4 . Choose any doubly stochastic matrix M such that $M_{p,q} = 0$ if $(p,q) \notin E$.* [2] *The local correlation respect to this "reweighting" $\mathrm{E}_{(p,q)\in_R M}\langle z_p, z_q\rangle \leq -1/8$.*

The proof is exactly the same as Lemma 5 because we still have at least $1/2$ of the weights between undetermined vertices.

4 Threshold Rank and Global Correlation

In this section we show how local correlation and global correlation are connected through *threshold rank*. Threshold rank of a graph $Rank_C(G)$ is defined by Arora et.al. in [1] as the number of eigenvalues larger than C. As they observed in [1], many problems have subexponential time algorithms when the underlying graph has low (i.e. sublinear) threshold rank. We show that 3-*Coloring* also lies in this category. If the underlying graph (or more precisely, a reweighting of the graph in the sense discussed after Lemma 5) has low threshold rank, then an SDP solution will have high global correlation as long as it has local correlation.

Our definition for threshold rank is different from [1]. We are interested in eigenvalues that are smaller than a certain negative constant $-C$. For a graph G, we use $Rank_{-C}(G)$ to denote the number of eigenvalues of G's normalized ajacency matrix whose value is at most $-C$. In all discussions C should be viewed as a positive constant, and we use negative sign to indicate that we are interested in eigenvalues smaller than $-C$.For a reweighting M of G (in the sense of Lemma 6), we also use $Rank_{-C}(M)$ to denote its threshold rank, which is the number of eigenvalues of M that are at most $-C$.

Consider a convex relaxation of threshold rank given by Barak et.al.[5]. In this relaxation each vertex in the graph has a vector z_p (later we will see that they are indeed related to the vectors $\{z_p\}$ in Lemma 5), and we try to maximize D, the reciprocal of global correlation.

$$\max \quad D$$

$$s.t. \mathop{\mathrm{E}}_{p\in V} \|z_p\|_2^2 = 1 \tag{10}$$

$$\mathop{\mathrm{E}}_{(p,q)\in E} \langle z_p, z_q\rangle \leq -C \tag{11}$$

$$\mathop{\mathrm{E}}_{p,q\in V} (\langle z_p, z_q\rangle)^2 \leq 1/D. \tag{12}$$

Barak et.al.[5] proved the following two Lemmas explaining why this is a relaxation to threshold rank. For completeness we give the proof in Appendix (Section B).

[2] These doubly stochastic matrices are called "reweightings" of G, because they correspond to normalized ajacency matrices of G after we assign weights to edges.

Lemma 7. *If the threshold rank of G is D ($Rank_{-C}(G) = D$), then the optimal value D^* of the convex relaxation is at least D.*

Lemma 8. *If the $Rank_{-C/2}(G) = D$, then the optimal value D^* of the convex relaxation is at most $4D/C^2 = Rank_{-C/2}(G)/(C/2)^2$.*

Lemma 8 is important for our analysis because it implies if the local correlation (left-hand-side of Equation (11)) is smaller than a negative constant, and threshold rank is low, the global correlation (left-hand-side of Equation (12)) must be of order $\Omega(1/D)$. Now we are ready to prove Theorem 1:

Proof. Write the SDP in Section 2 with $c \cdot D$ levels of Lasserre lifting ($Las^{c \cdot D}$), and solve it in time $n^{O(D)}$. We apply the following rounding algorithm inspired by Lemma 3.

1. Initialize SOL to be $SDP^{c \cdot D}$
2. Repeat
3. If at least $n/4$ of the vertices have an event in the SOL with probability more than $1/2$
4. Then apply Lemma 4 to get a partial coloring .
5. Pick a random event w, condition the solution SOL on this event
6. Until SOL is only valid for the first Level of Lasserre

Clearly, if the condition in Step 3 is satisfied and we proceed to Step 4, by Lemma 4 we get a partial coloring for $n/4$ vertices. In particular, one of the colors will have more than $n/12$ vertices, and they form an independent set. Therefore we only need to prove the probability that we reach Step 4 is large.

Let r_i be the probability that Step 4 is reached before iteration i. We would like to prove $r_{c \cdot D} \geq 1/2$. Assume we continue to run the algorithm even if Step 4 is reached (and we have already found an independent set). Let SOL_i be the solution at step i, GC_i be its global correlation and VAR_i be its variance.

We first prove the following Claim:
Claim: If the number of undetermined vertices in SOL_i is smaller than $n/4$, the global correlation GC_i is at least $\Omega(1/D)$.

Proof. Given the assumption, we can apply Lemma 5. From the solution SOL_i, Lemma 5 constructs vectors $\{z_p\}(p \in V)$, and $\mathrm{E}_{(p,q) \in E} \langle z_p, z_q \rangle \leq -1/8$.

We shall normalize these vectors so that they satisfy Equations (10) and (11). The norm of z_p is $\|z_p\|_2^2 = \|u_{p,R}\|_2^2 + \|u_{p,Y}\|_2^2 + \|u_{p,B}\|_2^2 = 1 - r_{p,R}^2 - r_{p,Y}^2 - r_{p,B}^2$. Here $r_{p,X}$ is the probability that p is colored with color X, and the equation follows from Theorem 2. If for vertex p no event has probability more than $1/2$, then $\|z_p\|_2^2$ is a value between $1/4$ and 1. As assumed the number of such vertices is at least $3n/4$ (otherwise Step 4 has already been performed), thus $\mathrm{E}_{p \in V} \|z_p\|_2^2$ is between $3/16$ and 1. We can normalize these vectors by multiplying with $c' = \sqrt{1/\mathrm{E}_{p \in V} \|z_p\|_2^2}$. For the normalized vectors $\{\overline{z_p}\}$, we have $\mathrm{E}_{p \in V} \|\overline{z_p}\|_2^2 = 1$. And since $c' \geq 1$ we still have $\mathrm{E}_{(p,q) \in E} \langle \overline{z_p}, \overline{z_q} \rangle \leq -1/8$.

The vectors $\{\overline{z_p}\}$ satisfy Equation (10) and (11) for $C = -1/8$. Since we know $Rank_{-1/16}(G) = D$, Lemma 8 shows that the left-hand-side of Equation (12) must be at least $(1/16)^2/D = \Omega(1/D)$. That is, the global correlation between vectors $\{\overline{z_p}\}$ is at least $\Omega(1/D)$.

By analysis in Section 3, we know GC_i is within a constant factor of $\mathrm{E}_{p,q} \langle z_p, z_q \rangle^2$. Since the normalization factor c' between z_p and $\overline{z_p}$ is also bounded by a constant, $\mathrm{E}_{p,q} \langle z_p, z_q \rangle^2$ and $\mathrm{E}_{p,q} \langle \overline{z_p}, \overline{z_q} \rangle^2$ are also within a constant factor. Thus $GC_i \geq \Omega(1/D)$. □

The proof proceeds as follows: when r_i, the probability that the solution has more than 3/4 "determined" vertices, is large we can already get a good solution by applying the moreover part of Lemma 4. Otherwise we can apply the claim and Lemma 8 to conclude that the expected global correlation must be high at step i; then Lemma 3 reduces r_i significantly.

In step i, with probability $1-r_i$ the number of "determined" vertices (in the sense of Lemma 4) is smaller than $n/4$. When this happens (number of determined vertices small), Lemma 8 shows the global correlation is at least $\Omega(1/D)$. Therefore the expected global correlation at step i is at least $\mathrm{E}[GC_i] \geq \Omega(1/D) * 1/2 = \Omega(1/D)$ (the expectation is over random choices of the algorithm) just by considering the situations when number of determined vertices is small. By Lemma 3 we know every time Step 5 is applied, the variance is expected to reduce by GC_i. That is, $\mathrm{E}[VAR_{i+1}] \leq \mathrm{E}[VAR_i] - 4\,\mathrm{E}[GC_i] \leq \mathrm{E}[VAR_i] - \Omega(1/D)$. If r_i remains smaller than 1/2 for all the $c \cdot D$ rounds (where c is a large enough constant), we must have $\mathrm{E}[VAR_{c \cdot D}] < 1/16$. By Markov's Inequality with probability at least 1/2 the variance is at most 1/8, in which case Lemma 4 can be applied. That is, $r_{c \cdot D} \geq 1/2$. This is a contradiction and we must have $r_i \geq 1/2$ for some $i \leq c \cdot D$.

Therefore with probability at least 1/2 the rounding algorithm will reach Step 4 and find a large independent set.

For the moreover part, we apply a random permutation π over the vertices before running the whole algorithm. In the permuted graph $n/12$ of the vertices are in the independent set S found by the algorithm above. If we apply the inverse permutation π^{-1} to the independent set found, we claim that any vertex q of the original graph is inside the independent set $\pi^{-1}(S)$ with probability at least 1/12.

Indeed, for any two vertices q_1, q_2 of the original graph, suppose we first apply an automorphism π' such that $\pi'(q_2) = q_1$ (such π' must exist because the graph is vertex-transitive). Now the probability that $q_2 \in \pi'^{-1}\pi^{-1}(S)$ is the same as the probability that $q_1 \in \pi^{-1}(S)$ (they are actually the same event). Also, since π' is an automorphism, the algorithm have no way to tell whether we have applied π' before π, thus we must have $\Pr[q_2 \in \pi^{-1}(S)] = \Pr[q_2 \in \pi'^{-1}\pi^{-1}(S)] = \Pr[q_1 \in \pi^{-1}(S)]$. All probabilities are equal to their average, which is the size of S divided by n, and is at least 1/12.

Repeat this procedure $100 \log n$ times, each vertex is in one of the $100 \log n$ independent sets with probability at least $1 - n^{-2}$. Union bound shows with high probability the union of these independent sets is the vertex set. We use one color for each independent set (if a vertex belongs to multiple sets then choose an arbitrary one among them), which gives a valid $O(\log n)$ coloring. □

Note that the theorem is still true if any reweighting M of G has low threshold rank, the proof is just by replacing Lemma 5 with Lemma 6.

5 Threshold Rank Bound for Distance Transitive Graphs

As we explained in the Introduction, symmetric graphs are a natural class of hard instances for graph coloring problem. Also, by Blum Coloring Tools, it is enough to consider graphs with low diameter for 3-Coloring (the proof is similar to Theorem 15).

In this section we focus on a class of symmetric graphs: distance transitive graphs, and we prove for a distance transitive graph with diameter Δ, the threshold rank $Rank_{-C}(G)$ is at most $(O(1/C^2))^{\Delta}$. We begin by defining distance transitive graphs:

Definition 4 (Distance Transitive Graph)**.** A graph $G = (V, E)$ is distance-transitive if for any pairs of vertices (p, q) and (s, t), where the shortest-path distance between p, q and s, t are the

same, there is always an automorphism that maps p to s and q to t. In other words, pairs of of vertices at a particular distance look identical.

Distance transitive graphs have many nice properties, especially when we look at the neighborhoods of vertices. Define $\Gamma^k(p)$ to be the k-th neighborhood of p (which is the set of vertices at distance k of p), by the distance transitive condition, we know if a pair of vertices p, q have distance k, then $|\Gamma^{k-1}(p) \bigcap \Gamma(q)|$, $|\Gamma^k(p) \bigcap \Gamma(q)|$, $|\Gamma^{k+1}(p) \bigcap \Gamma(q)|$ are three numbers that depend only on k. As a convention, we call these numbers $c_k = |\Gamma^{k-1}(p) \bigcap \Gamma(q)|$, $a_k = |\Gamma^k(p) \bigcap \Gamma(q)|$ and $b_k = |\Gamma^{k+1}(p) \bigcap \Gamma(q)|$ adjacency numbers of the graph G. The size of k-th neighborhood ($|\Gamma^k(p)|$) is represented by n_k. The following is known about spectral properties of distance transitive graphs[6]:

Lemma 9. *A distance transitive graph G has $\Delta+1$ distinct eigenvalues, which are the eigenvalues of the matrix*

$$B = \begin{pmatrix} 0 & 1 & & & & \\ d & a_1 & c_2 & & & \\ & b_1 & a_2 & c_3 & & \\ & & \cdot & \cdot & \cdot & \\ & & & b_{\Delta-2} & a_{\Delta-1} & c_\Delta \\ & & & & b_{\Delta-1} & a_\Delta \end{pmatrix}.$$

Moreover, suppose the i-th eigenvalue is λ_i ($\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_\Delta$), with left eigenvector u_i and right eigenvector v_i ($u_i^T B = \lambda_i u_i^T$, $Bv_i = \lambda_i v_i$), we have $v_i(j) = n_j u_i(j)$. When v_i is normalized so that $v_i(0) = 1$, the multiplicity of λ_i in the original graph is $n/\langle u_i, v_i \rangle$.

Note that the eigenvalues in the above Lemma are for the adjacency matrix of G. There is a normalization factor d (the degree) between these eigenvalues and eigenvalues we were using for threshold rank. Also, there is a simple lemma about the adjacency numbers a_i, b_i and c_i's

Lemma 10. *In distance transitive graphs, $a_i + b_i + c_i = d$ (d is the degree of the graph), $b_i \geq b_{i+1}$, $c_i \leq c_{i+1}$, and $n_{i+1} = n_i b_i / c_{i+1}$.*

Now we are ready to prove the following theorem.

Theorem 11. *If a distance transitive graph has Lasserre$_{2^{O(\Delta)}}$ number 3, then there is an algorithm that colors it with $O(\log n)$ colors in time $n^{2^{O(\Delta)}}$.*

Proof. We first prove that the threshold rank of a distance transitive graph $Rank_{-C}(G)$ is at most $\Delta(10/C^2)^\Delta$. We shall show if i-th eigenvalue of B is a large negative value $\lambda_i < -Cd$, then we have $\langle u_i, v_i \rangle \geq n/(10/C^2)^\Delta$. By Lemma 9 this implies that the multiplicity of this eigenvalue cannot be more than $(10/C^2)^\Delta$. Since B has at most $\Delta + 1$ distinct eigenvalues this is enough to prove the theorem.

For simplicity we'll drop the subscript i. That is, $\lambda = \lambda_i$, $u = u_i$ and $v = v_i$. We write out the recursion for $u(j)$ obtained by the fact that u is a left-eigenvalue of B ($u(0) = 1$, $u(1) = \lambda/d$):

$$u(j) = \frac{(\lambda - a_{j-1})u(j-1) - c_{j-1}u(j-2)}{b_{j-1}} \quad j > 1$$

When $c_j \leq C^2 d/10$, we follow this recursion and show $|u(j+1)| \geq |u(j)|(C/2)$ by induction.

The base case for $j = 0$ and $j = 1$ is obvious because $u(0) = 1$ and $u(1) \leq -C$. Suppose this is true for $j \leq t$, when $j = t + 1$ we have

$$|u(t+1)| = \frac{|(\lambda - a_t)u(t) - c_t u(t-1)|}{b_t} \geq \frac{|\lambda + a_t||u(t)| - |c_t||u(t-1)|}{d} \geq C/2 \cdot |u(t)|.$$

Suppose the l-th level is the first level such that $c_l \geq C^2 d/10$, we know c_{l+j} $(j \geq 0)$ will all be at least $C^2 d/10$ by Lemma 10. Then we have $n_{l+j+1} = n_{l+j} b_{l+j}/c_{l+j+1} \leq 10/C^2 n_{l+j}$. This means $n_l \geq n/(10/C^2)^{\Delta - l}$. By above induction $|u(l)| \geq (C/2)^l$, and by Lemma 9 $v(j) = n_j u(j)$, therefore

$$\langle u, v \rangle = \sum_{j=0}^{\Delta} (u(j))^2 n_j \geq (u(l))^2 n_l \geq n/(10/C^2)^{\Delta}.$$

Thus the threshold rank is always smaller than $\Delta(10/C^2)^{\Delta} = 2^{O(\Delta)}$. Finally using the algorithm in Theorem 1 we can color a distance transitive graph in $O(n^{2^{O(\Delta)}})$ time. □

6 Conclusion

In this paper we explored the relationship between threshold rank and graph coloring. Unlike other problems such as Unique Games and MAX-CUT considered by Arora et.al.[1], we show that 3-Coloring is actually related to the negative side of the spectrum. And we give an algorithm that can find linear size independent set when the graph has threshold rank D and $Lasserre_{c \cdot D}$ number is 3. The efficiency of our algorithm depends on the threshold rank of a graph. Known integrality gap examples [14, 15] of 3-Coloring all have threshold rank that is polylog in the number of vertices (see Section C). Thus our algorithm can detect in quasipolynomial time that they are not 3-Colorable.The relationship between global correlation and rounding and the convex relaxation for threshold rank are inspired by Barak e.al.[5] and we believe these techniques can be useful in other problems.

If our approach is combined with combinatorial tools, it could possibly lead to good subexponential (or even quasipolynomial-time) coloring algorithms. In particular, if the following conjecture is true for any constant C and $D = n^{\delta}$, we get a $\exp(n^{\delta})$ time algorithm for coloring 3-Colorable graph with n^{ϵ} colors (see Section D). We have no counterexamples for the Conjecture when C is a constant and D is more than polylog(n).

Conjecture 12. *There exists an algorithm such that for any graph G, can either*

- *Find a subset S of vertices. The vertex expansion is at most $\Phi_V(S) \leq (n/|S|)^{1/C}$.*
- *Certify the existence of a reweighting M of G such that $Rank_{-1/16}(M) \leq D$.*

We also give efficient algorithm to color distance transitive graphs with low diameter and $Lasserre_k$ number is 3 for large k. These graphs have properties that seem to make it hard for previously known algorithms.

Acknowledgement We thank David Steurer, Grant Schoenebeck and Aravindan Vijayaraghavan for valuable discussions. We especially thank David Steurer for sharing the result [5] with us.

References

[1] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, FOCS '10, pages 563–572, Washington, DC, USA, 2010. IEEE Computer Society.

[2] Sanjeev Arora and Eden Chlamtac. Hard instances for 3-coloring. Unpublished manuscript.

[3] Sanjeev Arora, Eden Chlamtac, and Moses 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.

[4] Sanjeev Arora and Anuradha Venugopalan. Threshold rank bounds for frankl-rodl graphs. Unpublished manuscript.

[5] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation, 2011.

[6] Norman L. Biggs. Intersection matrices for linear graphs. In *Combinatorial Mathematics and its Applications (ed. D.J.A. Welsh)*, pages 15–23. Academic Press, 1971.

[7] Avrim Blum. New approximation algorithms for graph coloring. *J. ACM*, 41:470–516, May 1994.

[8] Avrim Blum and David Karger. An $\tilde{O}(n^{3/14})$-coloring algorithm for 3-colorable graphs. *Information Processing Letters*, 61:49–53, 1996.

[9] A.E. Brouwer, A.M. Cohen, and A. Neumaier. *Distance Regular Graphs.* Springer, August 1989.

[10] Eden Chlamtac. *Non-Local Analysis of SDP-Based Approximation Algorithms.* PhD thesis, Princeton University, 2009.

[11] Eden Chlamtac and Gyanit Singh. Improved approximation guarantees through higher levels of sdp hierarchies. In *Proceedings of the 11th international workshop, APPROX 2008, and 12th international workshop, RANDOM 2008 on Approximation, Randomization and Combinatorial Optimization: Algorithms and Techniques*, APPROX '08 / RANDOM '08, pages 49–62, Berlin, Heidelberg, 2008. Springer-Verlag.

[12] Eden Chlamtac and Madhur Tulsiani. Convex relaxations and integrality gaps. Unpublished manuscript.

[13] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 344–353, New York, NY, USA, 2006. ACM.

[14] Uriel Feige, Michael Langberg, and Gideon Schechtman. Graphs with tiny vector chromatic numbers and huge chromatic numbers. In *Proceedings of the 43rd Symposium on Foundations of Computer Science*, FOCS '02, pages 283–292, Washington, DC, USA, 2002. IEEE Computer Society.

[15] Peter Frankl and Vojtech Rodl. Forbidden intersections. *Transactions of the American Mathematical Society*, 300(1):pp. 259–286, 1987.

[16] David Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45:246–265, March 1998.

[17] Jean B. Lasserre. An explicit exact sdp relaxation for nonlinear 0-1 programs. In *Proceedings of the 8th International IPCO Conference on Integer Programming and Combinatorial Optimization*, pages 293–303, London, UK, 2001. Springer-Verlag.

[18] M. Laurent. A comparison of the sherali-adams, lovasz-schrijver and lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28:470–496, 2001.

[19] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

[20] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxation between the continuous and convex hull representations. *SIAM J. Discret. Math.*, 3:411–430, May 1990.

[21] David Steurer. Subexponential algorithms for d-to-1 two-prover games and for certifying almost perfect expansion. Unpublished manuscript, November 2010.

[22] M. Szegedy. A note on the θ number of lovasz and the generalized delsarte bound. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, pages 36–39, Washington, DC, USA, 1994. IEEE Computer Society.

[23] Avi Wigderson. Improving the performance guarantee for approximate graph coloring. *J. ACM*, 30:729–735, October 1983.

[24] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 681–690, New York, NY, USA, 2006. ACM.

A Preliminaries

In this Section we introduce the Blum Coloring Tools and the symmetry of graphs.

A.1 Blum Coloring Tools

Blum[7] described a series of combinatorial tools to make progress in graph coloring. Most of the coloring algorithms we mentioned used these tools in some ways. In this section we shall state without proof the main theorem of Blum's coloring tools. Before introducing Blum Coloring Tools, first recall the edge expansion and vertex expansion of a set S. For a regular graph $G = (V, E)$ with degree d, The edge expansion of a set of vertices $S \subseteq V$ is the fraction of edges that go out of the set, we denote this by $\Phi(S) = \frac{E(S,V\setminus S)}{d|S|}$ ($E(S,T)$ denotes the number of edges with one endpoint in S and the other endpoint in T). The vertex expansion is the ratio between number of neighbors (excluding the set itself) and number of vertices in the set. We denote this by $\Phi_V(S) = \frac{|\Gamma(S)-S|}{|S|}$. Blum Coloring Tools define three ways to make progress in graph coloring.

Theorem 13 (Blum Coloring Tools)**.** *An algorithm on graph G makes progress towards a $f(n)$ coloring if it finds one of the following:*

- *An independent set S of size $n/f(n)$.*
- *A two-colorable set S with vertex expansion $\Phi_V(S) \leq O(f(n))$.*
- *Two vertices that have the same color in every valid 3-coloring of the graph.*

If there is an efficient algorithm that always makes progress towards a $f(n)$ coloring, there is an efficient algorithm that colors every 3-colorable graph with $\tilde{O}(f(n))$ colors.

A.2 Symmetry of Graphs

Previous improvements over the SDP approach in [16] use global analysis: the worst case in local analysis cannot happen everywhere in the graph. This points us to symmetric graphs when trying to find lowerbounds. Indeed, known integrality gap examples such as the one in [15] have good symmetry properties. However, in this paper we show that if a graph is highly symmetric, then it is actually easier to find a good coloring. We formally define the symmetry properties of graphs in this section.

The symmetry of graphs is characterized by *automorphisms*. An automorphism of a graph G is a permutation $\pi : V \to V$ such that $(\pi(p), \pi(q))$ is an edge if and only if (p, q) is an edge. It is not hard to see that all automorphisms of a particular graph form a group under canonical composition of permutations. We call a graph vertex-symmetric if for any two vertices p and q, there is always an automorphism π such that $\pi(p) = q$. A graph is arc-transitive if for any two arcs (p, q) and (s, t), there is always an automorphism π that maps p to s and q to t. In particular, an arc-transitive graph is always vertex transitive, and they usually have larger automorphism groups.

Distance transtive graphs, as we defined in Section 5, have even larger automorphism groups.

B Relationship Between Threshold Rank and the Relaxation

We restate the convex relaxation for threshold rank here and prove the two Lemmas in [5] about its connection to threshold rank.

$$\begin{aligned} &\max \quad D \\ s.t. &\mathop{\mathrm{E}}_{p\in V} \|z_p\|_2^2 = 1 && (13)\\ &\mathop{\mathrm{E}}_{(p,q)\in E} \langle z_p, z_q\rangle \le -C && (14)\\ &\mathop{\mathrm{E}}_{p,q\in V} (\langle z_p, z_q\rangle)^2 \le 1/D. && (15) \end{aligned}$$

Lemma 7 (Restated)**.** *If the threshold rank of G is D ($Rank_{-C}(G) = D$), then the optimal value D^* of the convex relaxation is at least D.*

Proof. Let w_1, w_2, ..., w_D be the D eigenvectors correspond to eigenvalues at most $-C$ and λ_i's be the corresponding eigenvalues. These vectors are normalized so that $\mathrm{E}_{p\in V} w_{i,p}^2 = 1$ and they are all orthogonal to each other. Now we construct D dimensional vectors z_p: the i-th component of v_p is the p-th component of w_i (and we scale z_p by a factor of $1/\sqrt{D}$). To see how this constuction is done, consider the following matrix

$$\begin{pmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & \dots & w_{2,n} \\ \vdots & \vdots & & \vdots \\ w_{D,1} & w_{D,2} & \dots & w_{D,n} \end{pmatrix}$$

The row vectors are the vectors w_i, the column vectors are the corresponding vectors $\sqrt{D} \cdot z_p$. Now we show that these vectors satisfy the Equations (13) and (14), and the the left-hand-side of Equation (15) is exactly $1/D$. For the Equation (13), we have

$$\mathop{\mathrm{E}}_{p\in V} \|z_p\|_2^2 = \sum_{i=1}^{D} \mathop{\mathrm{E}}_{p\in V} w_{i,p}^2/D = D \cdot 1/D = 1.$$

For Equation (14),

$$\begin{aligned}
\mathop{\mathrm{E}}_{(p,q)\in E}\langle z_p, z_q\rangle &= \sum_{i=1}^{D}\frac{1}{D}\mathop{\mathrm{E}}_{p\in V}\left[w_{i,p}\mathop{\mathrm{E}}_{(q,p)\in E} w_{i,q}\right] \\
&= \frac{1}{D}\sum_{i=1}^{D}\mathop{\mathrm{E}}_{p\in V} w_{i,p}\lambda_i w_{i,p} \\
&= \frac{1}{D}\sum_{i=1}^{D}\lambda_i \\
&\leq -C.
\end{aligned}$$

Now we compute the global correlation (left-hand-side of Equation (15) and show that it is exactly $1/D$.

$$\begin{aligned}
\mathop{\mathrm{E}}_{p,q\in V}(\langle z_p, z_q\rangle)^2 &= \mathop{\mathrm{E}}_{p,q\in V}(\sum_{i=1}^{D} w_{i,p}w_{i,q}/D)^2 \\
&= \sum_{1\leq i,j\leq d}\mathop{\mathrm{E}}_{p,q\in V} w_{i,p}w_{j,p}w_{i,q}w_{j,q}/D^2 \\
&= \frac{1}{D^2}\sum_{1\leq i,j\leq d}\left(\mathop{\mathrm{E}}_{p\in V} w_{i,p}w_{j,p}\right)\left(\mathop{\mathrm{E}}_{q\in V} w_{i,q}w_{j,q}\right) \\
&= \frac{1}{D^2}\sum_{1\leq i,j\leq d}\left(\mathop{\mathrm{E}}_{p\in V} w_{i,p}w_{j,p}\right)^2 \\
&= \frac{1}{D^2}\sum_{1\leq i\leq d}\left(\mathop{\mathrm{E}}_{p\in V} w_{i,p}^2\right)^2 + \frac{1}{D^2}\sum_{i\neq j}\left(\mathop{\mathrm{E}}_{p\in V} w_{i,p}w_{j,p}\right)^2 \\
&= \frac{1}{D^2}\sum_{1\leq i\leq d}(\|w_i\|_2^2/n)^2 + \frac{1}{D^2}\sum_{i\neq j}(\langle w_i, w_j\rangle/n)^2 \\
&= 1/D.
\end{aligned}$$

□

Lemma 8 (Restated)**.** *If the $Rank_{-C/2}(G) = D$, then the optimal value D^* of the convex relaxation is at most $4D/C^2 = Rank_{-C/2}(G)/(C/2)^2$.*

Proof. The main idea is to prove that once Equation (14) is satisfied, a significant portion of the vectors must lie in the space spanned by eigenvectors with eigenvalues at most $-C/2$. To make this precise, we use w_1, w_2, …, w_n to denote the eigenvectors (these vectors are still normalized so that $\mathrm{E}_{p\in V} w_{i,p}^2 = 1$). As the threshold rank $Rank_{-C/2}(G) = D$, the first D have eigenvalues at most $-C/2$. Let $\{z_p\}$ be the set the vectors that optimizes D^*. Since every Euclidean metric over n points has a representation using n dimensions, we can assume without loss of generality that z_p's have dimension n.

As in Lemma 7, we consider matrix $A = (z_1, z_2, \ldots, z_n)$, and take the transpose. We express A^T as $(w'_1, w'_2, \ldots, w'_n)$, and expand each w'_i in the basis $\{w_1, w_2, \ldots, w_n\}$. That is,

$$w'_i = \sum_{j=1}^{n}\alpha_{i,j}w_j.$$

Now we express the inner-product of z_p and z_q in terms of w-vectors.

$$\langle z_p, z_q\rangle = \sum_{i=1}^{n} w'_{i,p} w' i, q = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{i,j}^2 w_{j,p} w_{j,q} + \sum_{i=1}^{n}\sum_{j_1\neq j_2}^{n} \alpha_{i,j_1}\alpha_{i,j_2} w_{j_1,p} w_{j_2,q}. \tag{16}$$

Let $\alpha_j = \sqrt{\sum_{i=1}^{n} \alpha_{i,j}^2}$, we can simplify (16) as

$$\langle z_p, z_q\rangle = \sum_{j=1}^{n} \alpha_j^2 w_{j,p} w_{j,q} + \sum_{i=1}^{n}\sum_{j_1\neq j_2}^{n} \alpha_{i,j_1}\alpha_{i,j_2} w_{j_1,p} w_{j_2,q}.$$

Since w_i's are orthogonal, when computing the left-hand-side of Equation (13) and (14), the expectation is 0 and can be ignored. The left hand side of Equation(13) is just the sum of squares of α_i's. We have:

$$\mathop{\mathrm{E}}_{p\in V} \|z_p\|_2^2 = \sum_{i=1}^{n} \left(\alpha_i^2 \mathop{\mathrm{E}}_{p\in V} w_{i,p}^2\right) = \sum_{i=1}^{n} \alpha_i^2 = 1.$$

The left hand side of Equation (14) is just the sum of squares of α_i's weighted by the corresponding eigenvalues λ_i's. That is

$$\mathop{\mathrm{E}}_{(p,q)\in E} \langle z_p, z_q\rangle = \sum_{i=1}^{n} \left(\alpha_i^2 \mathop{\mathrm{E}}_{(p,q)\in E} w_{i,p} w_{i,q}\right) = \sum_{i=1}^{n} \alpha_i^2 \lambda_i \leq -C.$$

Now by averaging argument we know the contribution of the first D eigenvectors to the left-hand-side of Equation (14) is at least $C/2$ (that is, $\sum_{j=1}^{D} \alpha_i^2 \geq C/2$). Otherwise $\sum_{i=1}^{n} \alpha_i^2 \lambda_i \geq -1 \cdot \sum_{i=1}^{D} \alpha_i^2 - C/2 \sum_{i=D+1}^{n} \alpha_i^2 > -C$.

Finally we look at Equation (15).

$$\mathop{\mathrm{E}}_{p,q\in V} \langle z_p, z_q\rangle^2 = \sum_{i_1,i_2,j_1,j_2,j_3,j_4} \alpha_{i_1,j_1}\alpha_{i_2,j_3}\alpha_{i_1,j_2}\alpha_{i_2,j_4} \mathop{\mathrm{E}}_{p}[w_{j_1,p} w_{j_3,p}] \mathop{\mathrm{E}}_{q}[w_{j_2,q} w_{j_4,q}].$$

In this sum, only the terms with $j_1 = j_3$ and $j_2 = j_4$ can be nonzero by orthogonality of w vectors. When $j_1 = j_3$ and $j_2 = j_4$ the two expectations are equal to 1. Thus the sum can be simplified to

$$\mathop{\mathrm{E}}_{p,q\in V} \langle z_p, z_q\rangle^2 = \sum_{i_1,i_2,j_1,j_2} \alpha_{i_1,j_1}\alpha_{i_2,j_1}\alpha_{i_1,j_2}\alpha_{i_2,j_2} = \sum_{j_1,j_2} \left(\sum_{i=1}^{n} \alpha_{i,j_1}\alpha_{i,j_2}\right)^2 \geq \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \alpha_{i,j}^2\right)^2 = \sum_{j=1}^{n} \alpha_j^4.$$

Applying Cauchy-Schwartz to we lowerbound this by

$$\sum_{i=1}^{n} \alpha_i^4 \geq \sum_{i=1}^{D} \alpha_i^4 \geq \frac{1}{D}(\sum_{i=1}^{D} \alpha_i^2)^2 \geq C^2/4D.$$

Thus the optimal D^* is at most $4D/C^2$. □

C Threshold Rank Bounds for Integrality Gap Examples

The SDP used by Karger, Motwani and Sudan[16] defined the notion of "vector chromatic number". Later integrality gap examples were found by Feige et.al.[14] and Frankl et.al.[15]. They show that there is a graph G with vector vector chromatic number 3, but the largest independent set in the graph G only has size n^c for some c bounded away from 1. Thus it is impossible to round the SDP in [16] and get a coloring with less than n^{1-c} colors.

The graph in Frankl-Rodl[15] has $Lasserre_1$ number 3. They showed the following theorem

Theorem 14 (Frankl-Rodl)**.** *For a graph $G = (\{\pm 1\}^d, E)$ where $(x, y) \in E$ if and only if they differ in exactly $3d/4$ positions[3], the largest independent set of G has size $(2-\epsilon)^d$. Here ϵ is a positive constant that does not depend on d.*

Arora et.al.[4] showed that this graph has low threshold rank. Indeed, since this graph is a Cayley graph on the group $\{\pm 1\}^d$, its eigenvectors correspond to the characters $\chi_S(x) = \prod_{i \in S} x_i$. For any set S, we can compute the corresponding eigenvector. By symmetry it's easy to see the eigenvalue only depends on the size of the set S. When the size of set S is $s \ll n/2$, the eigenvalue is roughly $(-1/2)^s$, and for larger sets the eigenvalues are negligible. Thus $Rank_{-C}(G) = d^{O(\log 1/C)} = \text{polylog}(n)$. The $Lasserre_{\text{polylog}(n)}$ number of the graph must be larger than 3, because otherwise by Theorem 1 we can find an independent set of linear size. Therefore this graph is no longer a valid integrality gap example after $\text{polylog}(n)$ levels of Lasserre lifting. The graphs in [14] come from random sampling and it's impossible to express the eigenvalues explicitly, but random walk analysis similar to [4] can be used to show that the threshold rank is also $\text{polylog}(n)$.

Arora and Chlamtac[2] also constructed a graph which is 3-colorable and has $\text{polylog}(n)$ threshold rank. The graph is also a good vertex expander, thus the existence of such a graph shows that Conjecture 12 cannot be true for $D < \text{polylog}(n)$.

The vertices of this graph are labeled by $\{1, 2, 3\}^d$. Two vertices x, y are adjacent if and only if they differ in all coordinates. For this graph, each eigenvector correspond to a vector $w \in \{-1, 0, 1\}^d$. Let $f(0) = 1$ and $f(1) = f(2) = -1/2$, then the eigenvector correspond to w is just $f(\langle w, x \rangle \mod 3)$. The eigenvalue for the eigenvector corresponding to w only depends on the number of ± 1's in w (which we denote by $|w|$). When $|w| = s$, the eigenvalue is $(-1/2)^s$. Therefore the threshold rank $Rank_{-C}(G)$ is also $\text{poly}(d) = \text{polylog}(n)$.

D Graph Decomposition and Small-Set-Expansion

The Blum Coloring Tools suggest that when a graph has low vertex expansion, we can utilize this property and get an approximate coloring. Indeed, we can get efficient algorithms for coloring by combining expansion properties and threshold rank if the conjecture stated in Conclusion is true. Here we restate the conjecture and prove the theorem.

Conjecture 12 (Restated)**.** *There exists an algorithm such that for any graph G, can either*

- *Find a subset S of vertices. The vertex expansion is at most $\Phi_V(S) \leq (n/|S|)^{1/C}$.*
- *Certify the existence of a reweighting M of G such that $Rank_{-1/16}(M) \leq D$.*

Theorem 15. *If Conjecture 12 is true for constant C and parameter $D(n)$, then there is an algorithm that finds a $O(n^{1/C} \log n)$ coloring in time $n^{O(D(n))}$.*

[3] $3d/4$ has to be even

Proof. By Blum Coloring Tools, it's enough to find an independent set of size $\Omega(n^{1-1/C})$. For any graph G, we apply the algorithm in the Conjecture. If the algorithm certifies the existence of a reweighting, we can find a linear size independent set using Lasserre Rounding (Theorem 1). If the algorithm returns a set S, then we remove the neighbors of S, and recurse on S and $V \backslash (\Gamma(S) \bigcup S)$. We now prove by induction that the number of remaining vertices is always at least $n^{1-1/C}$.

The base case is trivial: if the graph G has only one vertex then the algorithm cannot return such a set. Suppose the number of remaining vertices is always at least $|V|^{1-1/C}$ when the number of vertices is at most k. When the number of vertices is $k+1$, suppose the size of the set found by the algorithm is $|S|$. After recurse on S, by induction hypothesis the number of vertices remaining is $|S|^{1-1/C}$, and the number of vertices we removed is at most $|S| * ((k+1)/|S|)^{1/C}$, the ratio between these two values is exactly $(k+1)^{1/C}$. When we recurse on the set $V \backslash (\Gamma(S) \bigcup S)$, again by induction hypothesis the ratio between total number of vertices and the number of remaining vertices is at most $(k+1)^{1/C}$. Thus at least $(k+1)^{1-1/C}$ fraction of the vertices will remain. By taking the union of all the independent sets we found, we proved that we can always find an independent set of size $\Omega(n^{1-1/C})$. □

We cannot find any counterexamples for Conjecture 12 when the C is a constant and D is more than polylog(n). For $D = \text{polylog}(n)$ Arora and Chlamtac have a example (see Section C). However, it is important that we allow reweightings. Otherwise consider G as a graph with lots of negative eigenvalues (high threshold rank), we construct G' by making two copies of G, and connect them using a random bipartite graph. It is not hard to prove that G' still has high threshold rank, and because random graphs are good vertex expanders there are no sets that satisfy the second condition. This kind of counterexamples can be ruled out because in the reweighting we can set the weights of edges in G to be 0. In this way the reweighting M will correspond to a random bipartite graph which has low threshold rank.

The conjecture possibly true with any constant C and $D = n^{\epsilon}$ for any $\epsilon > 0$, which gives a subexponential time algorithm for coloring a 3-colorable graph with n^{δ} colors for any $\delta > 0$.

The threshold rank of the graph is closely related to small-set-expansion properties. As Steurer showed in [21]

Theorem 16. *Let G be a graph with n vertices such that $Rank_{\lambda}(G) > n^{\beta}/\delta$. Then, G contains a vertex set with volume at most δ and expansion at most $1 - \lambda^{O(1/\beta)}$. Furthermore there exists a polynomial-time algorithm that given G and δ, finds such a vertex set.*

Thus if in a reweighting every small subset is almost perfectly expanding, the graph will have low threshold rank (smaller than n^{β}/δ). The ability to assign different weights for different edges is the main reason that we believe the conjecture is true (and indeed simple counterexamples exist if we are not allowed to reweight the edges). However, this ability to assign weights also makes it difficult to examine the expansion properties of the graph. We observe that for arc-transitive graphs, a small set that is not perfectly expanding implies high threshold rank no matter how the weights are assigned for each edge.

Theorem 17. *If an arc-transitive graph G has a set of vertices S, so that the edge expansion $\Phi(S) < 1 - C$, then $Rank_{C/2}(G) \geq C^2 n/4|S|$.*

Proof. We apply all automorphisms of G to the set S. Suppose S is mapped to S_1, S_2, …, S_N where N is the size of the automorphism group. Since G is arc-transitive, for any edge (p, q), $\Pr_{i \in [N]}[q \in S_i | p \in S_i] = 1 - \Phi(S) > C$. Also, since arc-transitivity implies vertex transitivity, for every vertex p, $\Pr_{i \in [N]}[p \in S_i] = |S|/n$. We construct N dimensional vectors z_1, z_2, …, z_n as

follows: the j-th component ($j \in [N]$) of v_i is $\sqrt{n/N|S|}$ if $i \in S_j$ and 0 otherwise. Clearly we have $\mathrm{E}_{p\in V} \|z_p\|_2^2 = 1$. And we know

$$\mathop{\mathrm{E}}_{(p,q)\in E}[\langle z_p, z_q\rangle] = N \Pr_{i\in[N]}[p \in S_i] \Pr_{i\in[N]}[q \in S_i | p \in S_i] n/|S| N = 1 - \Phi(S) > C.$$

Finally we compute the global correlation between these vectors (we use $S_i(p)$ to denote the indicator variable for the event $p \in S_i$)

$$\begin{aligned}
\mathop{\mathrm{E}}_{p,q\in V}[\langle z_p, z_q\rangle^2] &= \mathop{\mathrm{E}}_{p,q\in V}(\sum_{i\in[N]} S_i(p)S_i(q)n/N|S|)^2 \\
&= \mathop{\mathrm{E}}_{i,j\in[N]} \frac{n^2}{|S|^2} \mathop{\mathrm{E}}_{p}[S_i(p)S_j(p)] \mathop{\mathrm{E}}_{q}[S_i(q)S_j(q)] \\
&= \mathop{\mathrm{E}}_{i,j\in[N]} (n/|S| \Pr[p \in S_i \bigcap S_j])^2 \\
&\leq \mathop{\mathrm{E}}_{i,j} n/|S| \Pr[p \in S_i \bigcap S_j] \\
&= |S|/n.
\end{aligned}$$

The last inequality is simply $\mathrm{E}[X^2] < \mathrm{E}[X]$ when X is always between 0 and 1. Similar to Lemma 8, by the existence of such vectors we can prove $Rank_{C/2}(G) \geq C^2 n/4|S|$. □

Thus a vertex-expanding edge-transitive graph with a small set that has low expansion will disprove the Conjecture. We are also not aware of any such graphs.