2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)

Subexponential LPs Approximate Max-Cut

Samuel B. Hopkins
UC Berkeley
hopkins@berkeley.edu

Tselil Schramm
Harvard University and MIT
tselil@mit.edu

Luca Trevisan
Bocconi University
L.Trevisan@UniBocconi.It

***Abstract*—We show that for every $\varepsilon > 0$, the degree-n^ε Sherali-Adams linear program (with $\exp(\tilde{O}(n^\varepsilon))$ variables and constraints) approximates the maximum cut problem within a factor of $(\frac{1}{2} + \varepsilon')$, for some $\varepsilon'(\varepsilon) > 0$. Our result provides a surprising converse to known lower bounds against all linear programming relaxations of Max-Cut [1], [2], and hence resolves the extension complexity of approximate Max-Cut for approximation factors close to $\frac{1}{2}$ (up to the function $\varepsilon'(\varepsilon)$). Previously, only semidefinite programs and spectral methods were known to yield approximation factors better than $\frac{1}{2}$ for Max-Cut in time $2^{o(n)}$. We also show that constant-degree Sherali-Adams linear programs (with poly(n) variables and constraints) can solve Max-Cut with approximation factor close to 1 on graphs of small threshold rank: this is the first connection of which we are aware between threshold rank and linear programming-based algorithms.**

Our results separate the power of Sherali-Adams versus Lovász-Schrijver hierarchies for approximating Max-Cut, since it is known [3] that $(\frac{1}{2}+\varepsilon)$ approximation of Max Cut requires $\Omega_\varepsilon(n)$ rounds in the Lovász-Schrijver hierarchy.

We also provide a subexponential time approximation for Khot's Unique Games problem [4]: we show that for every $\varepsilon > 0$ the degree-$(n^\varepsilon \log q)$ Sherali-Adams linear program distinguishes instances of Unique Games of value $\geq 1 - \varepsilon'$ from instances of value $\leq \varepsilon'$, for some $\varepsilon'(\varepsilon) > 0$, where q is the alphabet size. Such guarantees are qualitatively similar to those of previous subexponential-time algorithms for Unique Games but our algorithm does not rely on semidefinite programming or subspace enumeration techniques [5], [6], [7].

***Keywords*-Max-Cut; Combinatorial Optimization; Linear Programming; Sherali-Adams; Subexponential Time Algorithms; Global Correlation Rounding; Unique Games Conjecture**

I. Introduction

In the Max-Cut problem, we are given an undirected graph $G = (V, E)$ and are asked to find a partition of V into two sets C and $\overline{C}$ that maximizes the fraction of edges of E having exactly one endpoint in C.

Besides being one of the simplest and most well-studied discrete optimization problems, Max-Cut is held as a hallmark example of the success of semidefinite programs and spectral methods over other algorithms. The $.878\ldots$ approximation algorithm by Goemans and Williamson provided the first application of semidefinite programming (SDP) to the design of approximation algorithms with bounded worst-case approximation ratio. While it is easy to achieve a $\frac{1}{2}$ approximation in polynomial time (a simple greedy algorithm can find a partition that cuts half the edges in linear time), only semidefinite programming [8] and spectral methods [9] have been known to achieve approximations better than $\frac{1}{2}$ in polynomial time, or even subexponential time.

A fundamental difficulty in breaking the $\frac{1}{2}$ barrier for Max-Cut is that an approximation algorithm also provides a *certificate*: if we have, say, a .52-approximation algorithm, and we run it on a graph in which the Max-Cut optimum is less than the approximation, for example $\leq .51$, then the algorithm will output a cut of value $\leq .51$ and the execution of the algorithm, together with its proof of correctness, will provide a certificate that the Max-Cut optimum is $\leq .51/.52 < .99$. This means that the design of a .52-approximation algorithm requires the development of a technique that is able to provide, for every graph whose Max-Cut optimum is $\leq .51$, a certificate that its Max-Cut optimum is $< .99$. To date, only semidefinite programming was known to provide such certificates in sub-exponential time. Even the certificates implied by the spectral algorithm of [9] are dual feasible solutions of the Goemans-Williamson SDP relaxation.

There has been strong evidence that subexponentially-sized linear programming (LP) relaxations of Max-Cut could not provide such certificates. Beyond the failure to obtain LP-based algorithms for Max-Cut, there are many concrete strong lower bounds. Schoenebeck, Tulsiani and Trevisan [3] prove that in order to achieve an integrality gap better than $\frac{1}{2} + \epsilon$ one needs Lovász-Schrijver hierarchy relaxations [10] of size $\exp(\Omega_\epsilon(n))$. For the more powerful Sherali-Adams hierarchy [11], Charikar, Makarychev and Makarychev [1] prove that to achieve such an integrality gap one needs relaxations of size $\exp(n^{\Omega_\epsilon(1)})$. In both proofs, the integrality gap instances are (slightly modified) sparse random graphs, which spectral algorithms can solve almost trivially. Kothari, Meka and Raghavendra [2] show that Sherali-Adams integrality gaps for constraint satisfaction problems imply integrality gaps for *all* linear programs[1] of comparable size, and, in particular, using the integrality gap of [1], show that, in order to achieve approximation $\frac{1}{2} + \epsilon$ one needs an

[1]The result applies to all linear programs obtained in the "extended formulation" framework. In the case of Max-Cut, this applies to all relaxations in which the constraints do not depend on the edges of the graph, and depend only on the number of vertices.

2575-8454/20/$31.00 ©2020 IEEE
DOI 10.1109/FOCS46700.2020.00092

LP of size $\exp(n^{\Omega_\epsilon(1)})$.

These results together made any nontrivial LP-based algorithm for Max-Cut seem quite unlikely. It had been hypothesized that integrality gaps exist for exponential-sized Sherali-Adams LPs, similar to those known for Lovász-Schrijver; the boldest conjecture held that perhaps even random graphs could provide such integrality gaps. But surprisingly, the latter hypothesis was refuted by O'Donnell-Schramm [12], who show that the Sherali-Adams hierarchy can certify that a random graph of average degree d has Max-Cut at most .51 with an LP of size $\exp(n^\alpha)$, where $\alpha \to 0$ as $d \to \infty$. This matches the integrality gap result of Charikar et al., up to the precise dependence of α on d.

The result of O'Donnell and Schramm can be seen as an average-case analysis of the approximation guarantee of subexponential Sherali-Adams relaxations of Max-Cut, and it applies, more generally, to all graphs whose normalized adjacency matrix has bounded spectral radius. But even in the wake of the O'Donnell-Schramm result, it was not clear whether subexponential LPs can provide nontrivial approximations for worst-case Max-Cut instances. Indeed, in the related circumstance of approximating the feasible region of a semidefinite program with $O(n)$ variables and constraints via linear programs, it is known that constant-factor approximation can require $2^{\Omega(n)}$ linear constraints [13].

In this paper we resolve this question by providing a worst-case analysis that applies to all graphs. The graphs considered by O'Donnell and Schramm crucially have Max-Cut and Sherali-Adams values close to $1/2$: their analysis relies on fast mixing of random walks, and larger Max-Cuts preclude this. To overcome this barrier, our key intermediate step (and our main technical contribution) is an analysis of Sherali-Adams relaxations on graphs whose adjacency matrices have both a nontrivial number of large positive eigenvalues and an arbitrary number of large negative ones. In particular, this includes graphs with Max-Cut values close to 1. For this, we distill a simple proof of a strong *local-to-global correlation* lemma (Lemma III.2), with significantly weaker assumptions than existing local-to-global lemmas used to analyze LP and SDP hierarchies [12], [6], [14].

A. *Our Results*

Theorem I.1 (Main Result for Max-Cut)**.** *For every $\alpha > 0$ there is an $\epsilon > 0$ such that a Sherali-Adams LP relaxation of Max-Cut of degree n^α provides an approximation ratio at least $\frac{1}{2} + \epsilon$.*

Up to the precise dependence of ε on α our result is the best possible for Sherali-Adams (and in fact *any* linear program which fits in the *extended formulations* framework) [1], [2].

Our approach also offers insight into polynomial-size LPs from the Sherali-Adams hierarchy. We show that a large class of Max-Cut instances – graphs with *bounded threshold rank* – can be solved almost exactly by LPs of polynomial size. To the best of our knowledge, this is the first known connection between bounded-threshold rank graphs and the guarantees of linear-programming based algorithms; previously such connections were known only for semidefinite programs [6], [7]. We prove Theorem I.1 by combining Theorem I.3 below with a graph decomposition theorem of Steurer [15] (closely related to that of Arora, Barak, and Steurer [5]).

Definition I.2 (Threshold Rank)**.** For $\tau > 0$, we say that a graph G has τ-threshold rank at most k if the normalized adjacency matrix of G has at most k eigenvalues bigger than τ. (Note that we only count the positive eigenvalues larger than τ, not including the number of negative eigenvalues whose absolute value is bigger than τ.)

Theorem I.3 (Max-Cut on Low Threshold-Rank Graphs)**.** *Let G be a graph, let $\tau > 0$, and let $r = \mathrm{rank}_\tau(G)$. The degree-$k$ Sherali-Adams LP relaxation of Max-Cut on G is an ε additive approximation*[2] *to the Max-Cut value of G as long as*

$$k \geq \frac{r}{\varepsilon^{O\left(\log \frac{n}{r} / \log \frac{1}{\tau}\right)}}.$$

As an example, Theorem I.3 says that $O(r)$ levels of the Sherali-Adams hierarchy approximate Max-Cut up to 0.01 additive error on graphs whose normalized adjacency matrices have r eigenvalues larger than $n^{-0.01}$. Theorem I.3 captures the results of [12] as a special case, but it also extends to graphs with nontrivial eigenvalues (both positive and negative): this is crucial to our ability to use it in the proof of Theorem I.1.

Our approach also extends beyond Max-Cut. For a variety of 2-CSPs, including Max-2-Lin and Max-k-Cut, we show that Sherali-Adams LPs of subexponential size obtain non-trivial worst-case approximations. Of particular note is our result for Unique Games, where we show that $O(n^\alpha \log q)$-degree Sherali-Adams relaxations provide a constant-factor approximation for Unique Games on alphabets of size q.[3] This is qualitatively similar in performance to the spectral algorithm of Arora et al. [5] and the semidefinite programming algorithm of Barak et al. [6].

Theorem I.4 (Main Result for Unique Games)**.** *For every $\alpha > 0$ there is an $\varepsilon > 0$ such that a degree-$n^\alpha \log q$ Sherali-Adams relaxation of Unique-Games on an alphabet of size q distinguishes instances of value $\leq \varepsilon$ from instances of value $\geq 1 - \varepsilon$.*

The spirit of these results is that subexponentially-sized LPs can (surprisingly) match the performance of SDPs for

[2]Here we think of the Max-Cut value as normalized to lie in $[0, 1]$.

[3]Recall that the Unique Games Conjecture says that such approximations are NP hard.

some key problems in combinatorial optimization. In the case of Unique Games in the sub-exponential time regime, the performance of the LP is qualitatively similar to what is known for SDPs (which are the best known algorithms) – both achieve constant-factor approximations in subexponential time.[4] In the polynomial-size regime, Theorem I.3 shows that Sherali-Adams LPs match the best known performance guarantees for SDP-based hierarchies for Max-Cut on graphs of low threshold rank, with the exception of a somewhat more stringent requirement on the value of the threshold τ.[5]

One might wonder how universal this phenomenon is, and in particular whether it extends beyond constraint satisfaction problems. We observe that the approximation factor achieved by Sherali-Adams linear programs for the Max-QP problem (also known as the Max-Cut-Gain problem with negative edge weights) is exponentially worse than what is achieved by polynomial-size semidefinite programs, even in the subexponential regime.

Observation I.5. *Degree k Sherali-Adams relaxations provide a $\Theta(n/k)$-approximation to Max-QP.*

Despite the simplicity of the proof of Observtion I.5, we are not aware of a similar statement in the literature. Charikar and Wirth show that a basic (polynomial-size) semidefinite program provides an $O(\log n)$ approximation to this problem [16]. Thus, we have a simple example of an optimization problem for which there is a wide gap between SDP and Sherali-Adams performance even in the subexponential regime.

B. Overview of the proof of our results

The Sherali-Adams hierarchy.: A feasible solution of a degree-r Sherali-Adams Max-Cut LP with value c describes a relaxation of a distribution over cuts that, on average, cut a c fraction of the edges. The solution contains a complete description of the marginals of such a distribution over subsets of at most r vertices. These marginal distributions are locally consistent with one another, in the sense that the distribution for a set A of vertices and the distribution of a set B have the same marginal distribution on $A \cap B$.

Because the LP is a relaxation, there may in fact be no actual distribution over cuts which is consistent with the marginal distributions described by the feasible solution. For example, if the graph underlying the Max-Cut instance is a clique, then if we represent cuts as ± 1 assignments to vertices, a feasible degree-2 solution of value 1 is to define, for every pair u, v of vertices, the distribution that sets $X_u = -1, X_v = 1$ with probability $1/2$ and $X_u = 1, X_v = -1$ with probability $1/2$. These local distributions agree on their intersections, because the local distributions on vertex signs X_u, X_v and on X_v, X_w agree on their marginal distribution on X_v (which, in both cases, is equally likely to be -1 or 1). On the other hand, the max cut value is $\frac{1}{2} + o(1)$.

Local correlation, global correlation, and independent rounding.: We now describe some of the ideas that we build on from prior work on rounding LP or SDP relaxations.

A tempting approach to round a Sherali-Adams Max-Cut relaxation is to assign each vertex randomly according to the local distribution for that vertex, treating each vertex independently. This *independent rounding* approach fails if, for a typical edge (u, v), the local joint distribution for the pair of vertex signs $\{X_u, X_v\}$ differs noticeably from the product distribution $\{X_u\}\{X_v\}$ in which X_u and X_v are assigned independently. In the example of the degree-2 solution for the clique that we discussed above, each edge u, v is cut with probability $1/2$ by independent rounding but with probability 1 by the local distribution on X_u, X_v.

If, however, for a typical edge (u, v) the local distribution on X_u, X_v is close (for example, in ℓ_1 norm) to the product distribution on the marginals (that is, if X_u and X_v have "low correlation" in their local distribution), then independent rounding works. Thus, we would like to take an arbitrary solution and reduce it to a solution in which the correlation of X_u, X_v according to their local distributions is small on average for a random edge (u, v). Such a solution is said to have low *local correlation*. It is not difficult to show that if a solution has value c, and the average, for a random edge (u, v), of the ℓ_1 distance between the local distribution on X_u, X_v and the product distribution on the marginals is at most $\mathbb{E}_{u \sim v} \|\{X_u, X_v\} - \{X_u\}\{X_v\}\|_1 \le \epsilon$, then independent rounding will cut at least a $c - \epsilon$ fraction of edges (in expectation).

The *global correlation* of a feasible solution is the average, over all pairs u, v of the ℓ_1 distance: $\mathbb{E}_{u,v} \|\{X_u, X_v\} - \{X_u\}\{X_v\}\|_1$. The global correlation of a feasible solution can be reduced using an operation called *conditioning*. Starting from a feasible solution of degree r, conditioning over t variables yields a new feasible solution of degree $r - t$, of the same value, and such that the global correlation is at most $O(1/\sqrt{t})$. (See Theorem III.3 for a more precise statement.)

Together these facts imply that, if we could argue that low global correlation implies low local correlation, then we would have a good rounding procedure: first apply conditioning (to reduce global, and therefore local, correlation) and then apply independent rounding. This scheme underlies several approximation algorithms using LP and SDP hierarchies; it was pioneered in [6], [14] and then used in a number of subsequent works.

[4]The precise approximation factors obtained by the algorithms of [5], [6] are better, but the bottom line is that we obtain an approximation ratio independent of the alphabet size in subexponential time, which is the measure of success relevant to the Unique Games Conjecture.

[5]For LPs, our results are meaningful mainly when $\tau \le n^{-\Omega(1)}$, while SDP-based hierarchies require only $\tau \le \mathrm{poly}(\epsilon)$. The lower bounds of [1] show that this n-dependence of the threshold is necessary for LPs; our result matches their lower bound in the case of expander graphs, i.e. for threshold rank 1.

Local to global correlation in low threshold rank graphs.: Our first main result (Lemma III.1) says the following: suppose that, for a random edge (u, v) of G the correlation (as defined above) between X_u and X_v is at least γ. Then, if we pick a random walk $v_0, v_1, \ldots, v_t$ of length t from a random start vertex v_0, the average correlation between X_{v_0} and X_{v_t} is $\gamma^{\Omega(t)}$, provided that the degree of the Sherali-Adams relaxation is at least $O(1/\gamma)^t$.

Next, we argue that in graphs of *bounded threshold rank*, noticeably large correlation along sufficiently long random walks implies noticeably large global correlation.

For a parameter $\tau > 0$, we say that a graph has τ-threshold rank at most k if the normalized adjacency matrix of the graph has at most k eigenvalues bigger than τ. (Note that we only count the positive eigenvalues larger than τ, and we do not count the number of negative eigenvalues whose absolute value is bigger than τ.) Graphs of bounded threshold rank have been studied before from the point of view of the performance of SDPs on such graphs, but not, as far as we know, from the point of view of the performance of LPs.

The key property of graphs of τ-threshold rank at most k is this: for a typical start vertex, the distribution of the last vertex of a random walk of length $t = O\left(\frac{1}{\log 1/\tau} \log n\right)$ has collision probability $O(k/n)$ (see Claim III.7). In particular, if the expected correlation between the endpoints of a t-step random walk is at least δ, then the global correlation is at least $\Omega(\delta/k)$.

In summary, we have that: (i) an upper bound on the correlation along t-steps random walks implies an upper bound on the local correlation; (ii) in graphs of bounded threshold rank, an upper bound on the global correlation implies an upper bound on the correlation along random walks; (iii) one can get a feasible solution of bounded global correlation by applying conditioning. Putting these facts together, if we have a graph of τ-threshold rank k, and we have a Sherali-Adams solution, conditioning on $O(k^2 n^{O((\log(1/\gamma))/\log(1/\tau))})$ variables (which we can do if the degree of the Sherali-Adams solution is larger than the above bound), gives us a solution of local correlation at most γ. The number of variables that we need to condition on is at most the desired bound n^α if, say, $k < n^{\alpha/4}$ and τ is sufficiently small relative to γ and α.

In comparing this approximation to the result of O'Donnell-Schramm [12], we see that they can guarantee a $1 - \gamma$ approximation using a Sherali-Adams relaxation of degree n^α, provided that all the non-trivial eigenvalues of the normalized adjacency matrix are at most τ in magnitude, with similar tradeoffs between γ, α and τ as we have above. The requirement that all the non-trivial eigenvalues are at most τ in magnitude precludes graphs with large cuts (say, Max-Cut value 0.99), meaning that [12] cannot analyze the Sherali-Adams relaxation for all graphs. In the analysis sketched above, we can deal with graphs that have up to $n^{\alpha/4}$ positive eigenvalues larger than τ, instead of just one, and arbitrarily many, even order of n, negative eigenvalues of magnitude larger than τ, instead of none. The fact that we do not put any restriction on the number of large-in-magnitude negative eigenvalues enables us to extend our analysis to general graphs by partitioning general graphs into graphs of bounded threshold rank.

Partitioning into pieces of bounded threshold rank.: Finally, we reduce the case of general graphs to the case of graphs of bounded threshold rank via a decomposition theorem (Theorem III.5) that shows that every graph, after the removal of a bounded number of edges, breaks down into connected components each of bounded threshold rank. We employ such a decomposition theorem proved originally in [15], which is closely related to a graph decomposition theorem proved Arora, Barak and Steurer [5]. The latter concerns thresholds τ close to 1, while we are interested in τ close to zero, because eventually our running time will be exponential in $n^{O(1/\log 1/\tau)}$. Steurer [15] shows that graphs can be decomposed into pieces of small τ-threshold rank for τ close to 0, although one has to remove a large fraction of edges to achieve it.

Obtaining a solution.: We first remove edges to make the residual graph have connected components of bounded threshold rank; under the assumption that the value of the Sherali-Adams solution is sufficiently high to begin with, the resulting components will have Sherali-Adams value close to 1. Then we apply conditioning and rounding independently in each component, obtaining a cut of value close to 1 in each component. Interpreting each cut as a ± 1 assignment to the vertices, we then, independently for each component, either leave the cut as is with probability $\frac{1}{2}$ or "flip" the cut (switching -1s with 1s) with probability $\frac{1}{2}$. This last step ensures that all the edges between components are cut with probability $\frac{1}{2}$. Thus our cut has value at least $\frac{1}{2} + \varepsilon$ for ε proportional to the fraction of edges not cut by the partition.

Unique Games.: Our result for Unique Games has a similar structure: we define a measure of correlation and prove that noticeably large local correlation implies noticeably large correlation between the endpoints of random walks, which implies noticeably large global correlation in graphs of bounded threshold rank. The main difference compared to the Max-Cut analysis is in the measure of correlation between the random variables corresponding to two vertices. We use a notion that we call "permutation correlation," which has previously been used as a measure of correlation in the analysis of SDPs for Unique Games, and we are able to show that if the local permutation correlation along a random edge is γ, then the permutation correlation along the endpoint of a t-step random walk is at least $\gamma^{O(t)}$, where the constant in the big-Oh is an absolute constant that does not depend on the size of the alphabet of the unique game. Crucially, in the case of Unique Games, rounding a Sherali-Adams solution by indpendently sampling from the distributions $\{X_u\}$ succeeds if the solution only has small local permutation correlation (a weaker requirement than small local correlation).

C. Discussion

Our analysis of Sherali-Adams linear programs is similar to the rounding of SDPs in [6]: we partition the graph into components of bounded threshold rank, we show that noticeably large local correlation implies noticeably large global correlation in such graphs, and we observe that large global correlation is a property that can survive only a bounded number of conditionings. However, in the case of SDPs, the spectrahedral constraints allow one to prove a local-to-global correlation lemma directly leveraging spectral properties of the underlying constraint graph in a direct way. For linear programs, it is surprising that we are able to take advantage of spectral properties at all.[6]

The main step is Lemma III.1, which allows us to relate the local correlation along one edge to the correlation between the first and the last vertices of a random walk, and hence allows us to relate local correlation to global correlation in graphs of bounded threshold rank.

The rounding algorithm that we use in this paper could be applied to a feasible Lovász-Schrijver solution: a Lovász-Schrijver solution can be rounded near-optimally if it exhibits low local correlation. The operation of conditioning is well-defined for Lovász-Schrijver solutions, and conditioning on t variables reduces the global correlation to $O(1/\sqrt{t})$ at the cost of reducing the degree of the solution by t. Our algorithm, however, must fail, because of the known integrality gaps, and what fails is Lemma III.1.

It would be interesting to see a similar phenomenon in the SDP setting: could there be a rounding algorithm for a Sum-of-Squares (SoS) relaxation that would be well-defined on weaker relaxations, but whose analysis relies on properties that hold only for SoS relaxations? For example, it is an open problem whether SoS relaxations of polynomial size can refute the Unique Games Conjecture, and it is known that weaker relaxations of polynomial (or even slightly superpolynomial) size cannot refute the Unique Games Conjecture [17]. This is often interpreted as evidence that, in order to refute the Unique Games Conjecture via SoS relaxations, one would have to develop a radically new rounding technique that makes explicit use of SoS constraints. Could it be, instead, that there is a relatively simple rounding scheme for SoS relaxations of Unique Games, that is well defined on weaker relaxations and such that SoS-specific constraints are used only in the analysis?

Finally, we note that our subexponential algorithm for $(1-\varepsilon)$ vs ε Unique Games (that is, distinguishing a $1-\varepsilon$-satisfiable instance from an ε-satisfiable one) requires $n^{\varepsilon'(\varepsilon)} \log q$ rounds of Sherali-Adams, while the prior algorithm based on hierarchies required $n^{\varepsilon'(\varepsilon)} q^3$ rounds. As far as we know, this is the first subexponential algorithm allowing $q \geq n^{\Omega(1)}$. However, our result is not a strict improvement on the prior algorithm [6] (which uses a stronger hierarchy), because that algorithm solves $(1-\varepsilon)$ vs $1/2$ Unique Games.

Open Problems: We mention a few open problems; resolving any of these (affirmatively or negatively) would be of interest.

1) *Beating a random assignment for general* 2*-CSPs?* Hastad [18] shows that for every 2-CSP with alphabet size q, an SDP obtains an approximation ratio that is strictly better than the ratio obtained by randomly sampling assignment in $[q]^n$. Our work shows that subexponential LPs offer similar guarantees for some 2-CSPs including Max-Cut and Unique Games. Do subexponentially-sized linear programs offer a non-trivial approximation for every 2-CSP?
2) *Refined approximations for Max-Cut?* In addition to providing a $0.878\ldots$ approximation to Max-Cut, the Goemans-Williamson SDP also offers more refined guarantees. (1) In a graph with Max-Cut value at least $1-\epsilon$ it finds a cut of size $1-O(\sqrt{\epsilon})$ [8], and (2) via an alternative rounding scheme due to Charikar and Wirth [16], in a graph with Max-Cut value $\frac{1}{2}+\epsilon$ it finds a cut of size $\frac{1}{2}+\Omega(\varepsilon/\log(1/\varepsilon))$. Can these guarantees be matched by subexponentially-sized linear programs? Our analysis cannot be extended to these settings as-is because the graph partitioning scheme we employ forces us to settle for randomly cutting most of the edges of the graph (all but, say, a 0.01 fraction), which would seem to preclude ever finding a cut of size close to 0.99.
3) *LP Certificates of Expansion?* Providing certificates of graph expansion is another combinatorial optimization problem for which spectral methods and SDPs appear to out-perform LPs. For instance, in every graph with expansion 0.99, Cheeger's inequality says that the second eigenvalue of the adjacency matrix certifies that the graph has expansion at least 0.01, while the best similar result for LPs loses a factor of $\log n$ [19]. Can subexponential LPs offer certificates comparable to the second eigenvalue?

Organization

The definitions we need to work with CSPs and the Sherali-Adams hierarchy are in Section II. In Section III we prove Theorem I.1 and Theorem I.3 on Max-Cut. The proofs of Theorem I.4 and Observation I.5, modifications of our arguments to handle a broader class of permutation-symmetric CSPs such as Max-2-Lin and Max-k-Cut, and some technical details are all deferred to the full version of

[6]Though it is less surprising in the wake of the results of O'Donnell and Schramm [12].

our paper, available at https://arxiv.org/pdf/1911.10304.pdf.
.

II. Preliminaries

In this section we provide some preliminaries; the reader may wish to proceed directy to the proof of our main theorem in Section III. For guidance, this section is organized as follows: Section II-A introduces some definitions and notation regarding graphs. Section II-B introduces measures of correlation that we will use to track the local-vs.-global correlation in our pseudodistributions. Section II-C introduces constraint satisfaction problems, and defines Max Cut and Unique Games. Finally, Section II-D introduces the Sherali-Adams LP and the notion of local random variables.

A. Graphs

A graph G on n vertices is a collection of nonnegative *weights* $w_{ij} \geq 0$ for each pair $\{ij\} \in \binom{n}{2}$. In this work all graphs are simple, undirected, and contain no isolated vertices, but may be irregular.

Definition II.1 (Adjacency, Normalized Adjacency, and Walk Matrices)**.** For a graph G we often employ the *adjacency matrix* A with entries $A_{ij} = w_{ij}$, the *degree matrix* with entries $D_{ii} = \sum_j w_{ij}$, the *walk matrix* $D^{-1}A$, and the *normalized adjacency matrix* $N = D^{-1/2}AD^{-1/2}$.

We will use the following notation:

Definition II.2 (Random Walk Endpoints)**.** For a graph G on n nodes, we often use $i \sim_\ell j$ to denote the distribution pairs i, j where i is first chosen according to the stationary measure π of the random walk on G and then j is the result of an ℓ-step random walk initialized at i. For the special case $\ell = 1$, we will use the shorthand $i \sim j$.

Definition II.3 (Threshold Rank)**.** A graph G has τ-*threshold rank* k, or $\operatorname{rank}_\tau(G) = k$, if the normalized adjacency matrix of G has at most k eigenvalues larger than τ.

B. Random Variables and Measures of Correlation

For a random variable X we denote by $\{X\}$ the associated density function.

Definition II.4 (Correlation)**.** If X, Y are jointly distributed taking values in $[q] \times [q]$, we will often be interested in the following notion of correlation between X, Y:

$$\operatorname{Cor}(X, Y) = \sum_{a,b \in [q]} |\Pr(X = a, Y = b) - \Pr(X = a)\Pr(y = b)|$$

Notice that this is exactly the ℓ_1 distance $\|\{X, Y\} - \{X\} \times \{Y\}\|_1$. Pinsker's inequality shows that $\operatorname{Cor}(X, Y) \leq \sqrt{2I(X; Y)}$, where $I(\cdot;\cdot)$ denotes the mutual information between X and Y.

Definition II.5 (Variance for Discrete Random Variables)**.** We also introduce a notion of variance for a $[q]$-valued random variable X. For each $a \in [q]$, let X_a be the $0/1$ variable such that $X_a = \mathbb{I}(X = a)$. Then we let $\mathbb{V}(X) = \sum_{a \in [q]} \mathbb{V}(X_a)$. We observe that since $\mathbb{V}(X_a) \leq \mathbb{E}\, X_a$, we have $\mathbb{V}(X) \leq \sum_{a \in [q]} \mathbb{E}\, X_a = 1$.

C. 2CSPs

An n-variable instance of 2CSP with alphabet size $q \in \mathbb{N}$ consists of a pair (G, Π), where G is an n-node weighted graph with weights w_{ij} having $\sum_{ij} w_{ij} = 1$ and Π is a collection of functions $\Pi_{ij} : [q] \times [q] \to \{0, 1\}$ for every edge in G. Without loss of generality throughout the paper we assume $w_{ij} \in [w_{\max}/n^3, w_{\max}]$, since low-weight edges of weight much less than $w_{\max}/n^2$ can be thrown out.

Definition II.6 (Objective Value)**.** The objective value of an assignment $x \in [q]^n$ for an instance is $\sum_{i \sim j} w_{ij} \Pi_{ij}(x_i, x_j) = \mathbb{E}_{i \sim j} \Pi(x_i, x_j)$, where w_{ij} is the weight of edge i, j.

Definition II.7 (Max-Cut)**.** An instance of Max-Cut is an instance of 2CSP where $q = 2$ and $\Pi_{ij}(x, x') = 1$ if and only if $x \neq x'$.

Definition II.8 (Unique Games)**.** An instance of Unique Games is an instance of 2CSP where $\Pi_{ij}(x, x')$ represents a bijective map from $[q]$ to $[q]$.

D. Local Distributions and Sherali-Adams Linear Programs

We briefly discuss basic definitions involving the Sherali-Adams linear programming hierarchy. For much more detail and proofs, see [20].

Definition II.9 (Local pseudodistribution)**.** For $q, n \in \mathbb{N}$ and $t \leq n$, a q-ary t-local pseudodistribution is a collection $\mu = \{\mu_S\}_{S \in \binom{n}{t}}$ of probability distributions μ_S on $[q]^t$ such that for every $S' \subseteq [n]$ with $|S'| \leq t$, the marginal distributions of μ_S for $S \supseteq S'$ on the variables in S' are all identical.

We often abuse notation and instead write $X_1, \ldots, X_n$ as *t-local random variables* induced by $\{\mu_S\}$, with the understanding that only probabilities and events concerning fewer than t variables at once are well defined. In particular, for two t-local random variables X_i, X_j, when $t \geq 2$ their correlation $\operatorname{Cor}(X_i, X_j)$ is well defined. We often write $\widetilde{\operatorname{Cor}}(X_i, X_j)$ as a reminder that the underlying variables are only t-local.

Definition II.10 (Sherali-Adams Polytope)**.** The set of all q-ary t-local pseudodistributions on n variables is the *degree-t Sherali-Adams polytope* – it is standard (see [20]) that this is a polytope involving $(qn)^{O(t)}$ variables and constraints.

Definition II.11 (Pseudoexpectation)**.** The t-local pseudodistributions are in one-to-one correspondence with *Sherali-Adams pseudoexpectations*, which are linear maps $\widetilde{\mathbb{E}} : \mathbb{R}[x_{ia}]^{\leq t}_{i \in [n], a \in [q]} \to \mathbb{R}$ which satisfy $\widetilde{\mathbb{E}}\, 1 = 1$ and

$\widetilde{\mathbb{E}} f(x) \geq 0$ if f is a nonnegative function depending on $\{x_{ja}\}_{j \in S, a \in [q]}$ for some $S \subseteq [n]$ with $|S| \leq t$. (Here $\mathbb{R}[y]^{\leq t}$ denotes polynomials in variables y with real coefficients and degree at most t.)

Again abusing notation, we often call such a pseudo-expectation $\widetilde{\mathbb{E}}$ or the corresponding pseudodistribution μ (alternatively written $X_1, \dots, X_n$) a *Sherali-Adams pseudodistribution*. If $q = 2$, we call it a *Boolean* Sherali-Adams pseudodistribution.

Definition II.12 (Objective Value of Local Distribution for 2CSP)**.** If (G, Π) is a q-ary 2CSP instance and $\{\mu_S\}$ is t-local pseudodistribution for $t \geq 2$, the objective value of $\{\mu_S\}$ for (G, Π) is given by

$$\sum_{i \sim j} w_{ij} \cdot \Pr_{X_i, X_j \sim \mu_{ij}} (\Pi_{ij}(X_i, X_j) = 1)$$

where μ_{ij} denotes the marginal distribution of any μ_S for $S \supseteq \{i, j\}$ on the indices i, j. If (G, Π) is a q-ary 2CSP, we write $SA_t(G, \Pi)$ for the maximum objective value achieved by any q-ary t-local distribution.

III. Subexponential Linear Programs for Max-Cut

In this section we prove our main result on approximating Max-Cut by subexponential-size LPs from the Sherali-Adams hierarchy. We begin with an overview stating the three main ingredients of the proof.

Local-to-global correlation: First we prove a local-to-global lemma in graphs of low threshold rank. Our first step is to use the "spider random walk" technique of O'Donnell and Schramm to establish that nontrivial correlation across edges implies nontrivial correlation for the endpoints of random walks of length t:

Lemma III.1. *Let G be a graph on n vertices, let t be a power of two, and let $X_1, \dots, X_n$ be $(2\lceil(\frac{1}{\gamma})^t\rceil + 1)$-local Boolean random variables. Then $\mathbb{E}_{i \sim j} \widetilde{\mathrm{Cor}}(X_i, X_j) \geq 16\gamma$ implies that $\mathbb{E}_{i \sim_t j} \widetilde{\mathrm{Cor}}(X_i, X_j) \geq \gamma^t$.*

We then use Lemma III.1 to prove a second statement that allows us to relate the correlation of endpoints of long random walks in the graph to the correlation of a uniformly random pair of vertices. The bound on the correlation crucially depends on the threshold rank of the graph (rather than on the second eigenvalue as a proxy for the mixing time, as is the case in O'Donnell-Schramm).

Lemma III.2. *If G is a graph on n vertices with $\mathrm{rank}_\tau(G) \leq k$, t is a power of two, and $X_1, \dots, X_n$ are $2(\frac{1}{\gamma})^t$-local random variables, then an upper bound on the global squared correlation*

$$\mathbb{E}_{i,j \sim \pi} \widetilde{\mathrm{Cor}}(X_i, X_j)^2 \leq \frac{1}{2} \frac{\gamma^{2t}}{k + n\tau^{2t-1}}$$

for π the stationary measure of G implies an upper bound on the local correlation

$$\mathbb{E}_{i \sim j} \widetilde{\mathrm{Cor}}(X_i, X_j) \leq 16\gamma.$$

This lemma is a composition of Lemma III.6 and Claim III.7, which we prove in Sections III-A and III-B below.

Global correlation rounding: Once we establish a sufficiently strong relationship between local and global correlation in low-threshold rank graphs, we can apply the global correlation rounding technique pioneered by [6]. We will use the following facts which originate in [6], [14] and are by now standard. The first fact says that in expectation, *global* correlation drops under conditioning, while the objective value remains the same.

Theorem III.3. *Let $q, n \in \mathbb{N}$, let $\kappa > 0$, and let μ be a $(\frac{6 \log q}{\kappa} + 2)$-local pseudodistribution over $[q]$-valued random variables $X_1, \dots, X_n$. Let $\pi \in \Delta_n$ be a distribution on $[n]$. Let $i_1, \dots, i_k \in [n]$ be indices chosen i.i.d. from π, with $k = \frac{6 \log q}{\kappa}$. There exists $t \leq k$ such that*

$$\mathbb{E}_{i_1, \dots, i_t} \mathbb{E}_{i,j \sim \pi} \mathbb{E}_{X_{i_1}, \dots, X_{i_t}} \widetilde{\mathrm{Cor}}(X_i, X_j | X_{i_1}, \dots, X_{i_t})^2 \leq \kappa\,.$$

Furthermore, if $X_1, \dots, X_n$ are variables in a 2CSP instance (G, Π), then

$$\mathbb{E}_{i_1, \dots, i_k} \left(\mathbb{E}_{i \sim j} \mathbb{E}_{X_1, \dots, X_k, X_i, X_j} [\Pi(X_i, X_j) \mid X_1, \dots, X_k]\right) = \mathbb{E}_{i \sim j} \mathbb{E}_{X_i, X_j} [\Pi(X_i, X_j)].$$

Note that the expectations taken above over X_i are well-defined for local random variables because each depends on at most $k + 2$ variables.

The second fact states that in 2CSP instances with low *local* correlation, independent rounding produces a solution with high objective value.

Lemma III.4 (Rounding low-correlation Sherali-Adams)**.** *Let (G, Π) be an instance of 2CSP and let $X_1, \dots, X_n$ be q-ary T-local random variables for $T \geq 2$. Suppose that the local correlation $\mathbb{E}_{i \sim j} \widetilde{\mathrm{Cor}}(X_i, X_j) \leq \delta$. Let $Y_i \sim \{X_i\}$ be independent samples from the 1-wise marginals of $X_1, \dots, X_n$. Then $\mathbb{E}_Y \mathbb{E}_{i \sim j} \Pi(Y_i, Y_j) \geq \mathbb{E}_{i \sim j} \mathbb{E}_{X_i, X_j} \Pi(X_i, X_j) - \delta$. Furthermore, this rounding scheme can be derandomized in polynomial time.*

For completeness, proofs of both statements can be found in the full version of this paper.

Graph partitioning: In a graph with high threshold rank, we will perform partitioning into low threshold rank parts in the style of [5]. The partitioning scheme of [5] partitions a graph into parts of expansion at most ϵ when $\mathrm{rank}_{1-\epsilon'}(G)$ is large, for ϵ, ϵ' close to 0. However, their result does not give guarantees for graphs that have large τ-threshold rank when τ is close to 0 rather than 1. A modification of the [5] partitioning scheme for the small-τ regime appears in [15], which shows that if $\mathrm{rank}_\epsilon(G)$ is large, then one can obtain a partition of expansion at most $1-\epsilon'$.

Theorem III.5 (Restatement of [15] Theorem 2.2)**.** *Fix any $\tau, \alpha \in (0,1)$, and take n sufficiently large. Any n-vertex simple graph $G=(V,E)$ admits a partition into components $G_1,\dots,G_m$ such that for all $i \in [m]$, the threshold rank $\operatorname{rank}_\tau(G_i) \le n^\alpha$, with the total fraction of edges cut in the partition bounded by $\frac{1}{|E|}|\cup_{i\neq j\in[m]} E[G_i,G_j]| \le 1 - \exp\left(-O\left(\frac{1}{\alpha^2}\log\frac{1}{\tau}\right)\right)$. Furthermore, there is a polynomial-time algorithm to compute this partition.*

Again, we provide a proof in the full version for completeness.

Putting things together: With these three pieces in place, we can prove Theorem I.1: we apply the partitioning theorem (Theorem III.5) to partition our graph into pieces of small threshold rank, cutting at most a $1-\varepsilon$ fraction of edges in the partition. Then, within each piece, we apply global correlation rounding (Theorem III.3) until the global correlation is low. From our local-to-global lemma (Lemma III.2), we can conclude that the local correlation within each piece is small. Furthermore, under the assumption that the objective value is $\ge 1-\varepsilon'$ for $\varepsilon' \ll \varepsilon$, on average the objective value within the parts will be large and thus from Lemma III.4 independent rounding will return a solution satisfying (say) a $\ge .75$-fraction of edges within each piece, which (when aggregated across the parts) gives objective value $\ge 0.75\varepsilon$. Finally, by applying a random sign change to the solution within each piece, the $(1-\varepsilon)$ fraction of edges crossing the partition are cut with probability $\frac{1}{2}$, for a solution of objective value $\frac{1}{2}+.25\varepsilon$. We give a formal version of this argument in Section III-C.

A. Local-to-Global Lemma for Sherali–Adams

In this section, we prove a lemma in the style of O'Donnell-Schramm [12] that allows us to relate local correlations to the correlations of independently sampled vertices. One key difference between our proof and that of [12] is that we track the distance (in ℓ_1) between joint distribution $\{X,Y\}$ and the product of marginals $\{X\}\{Y\}$, rather than the correlation $\mathbb{E}\,XY$.[7] Another difference is that rather than measuring the distance of the random walk to mixing in terms of the second eigenvalue, we measure it in terms of a trace of a power of the random walk matrix, which allows us to take advantage of graphs of low threshold rank.

Lemma (Restatement of Lemma III.1)**.** *Let G be a graph on n vertices, let $t=2^\ell$ be a power of two, and let $X_1,\dots,X_n$ be $(2\lceil(\frac{16}{\gamma})^t\rceil+1)$-local Boolean variables. Then $\mathbb{E}_{i\sim j}\widetilde{\operatorname{Cor}}(X_i,X_j) \ge \gamma$ implies that $\mathbb{E}_{i\sim_t j}\widetilde{\operatorname{Cor}}(X_i,X_j) \ge (\frac{1}{16}\gamma)^t$.*

[7] It turns out that this is crucial for use with global correlation rounding, as the correlation is on average unaffected by conditioning.

Proof: We will prove the following claim:

Claim. *For any integer s and $(2\lceil\frac{4}{\delta^2}\rceil+1)$-local Boolean random variables $X_1,\dots,X_n$, if $\mathbb{E}_{i\sim_s j}\widetilde{\operatorname{Cor}}(X_i,X_j) \ge \delta$, then $\mathbb{E}_{i\sim_{2s} j}\widetilde{\operatorname{Cor}}(X_i,X_j) \ge \frac{1}{4}\delta^2$.*

We have assumed that $t=2^\ell$, so we now apply this claim recursively ℓ times starting with $\delta=\gamma$, and we end up with a lower bound of

$$\begin{aligned}\mathbb{E}_{i\sim_t j}\widetilde{\operatorname{Cor}}(X_i,X_j) &\ge \frac{1}{4}\left(\frac{1}{4}\left(\cdots\frac{1}{4}\left(\frac{1}{4}\gamma^2\right)^2\cdots\right)^2\right)^2 \\ &= \gamma^{2^\ell}\cdot\frac{1}{4}\cdot\frac{1}{4^2}\cdot\frac{1}{4^4}\cdots\frac{1}{4^{2^\ell}} \ge \gamma^{2^\ell}4^{-2\cdot 2^\ell},\end{aligned}$$

which will give us the lemma.

Now, we prove the claim. Choose a root vertex $r\sim\pi$, and let $i_1,\dots,i_k$ be a set of $k=\lceil\frac{4}{\delta^2}\rceil$ vertices sampled independently, for each by taking an s-step random walk from r, $i_\ell \sim_s r$. Define the vectors $Z^{(1)}, Z^{(-1)} \in \mathbb{R}^{2k+1}$ with the first $2k$ entries indexed by pairs (j,a) for $j\in[k]$ and $a\in\{\pm1\}$ and one singleton entry indexed by r, with $Z^{(b)}_{j,1} = \mathbb{I}[X_{i_j}=1]$ and $Z^{(b)}_{j,-1} = \mathbb{I}[X_{i_j}=-1]$ for $j\in[k]$ and $Z^{(b)}_r = \mathbb{I}[X_r=b]$. Let $C^{(1)}, C^{(-1)} \in \mathbb{R}^{(2k+1)\times(2k+1)}$ be the covariance matrices

$$C^{(b)} = \widetilde{\mathbb{E}}[Z^{(b)}(Z^{(b)})^\top] - \widetilde{\mathbb{E}}[Z^{(b)}]\,\widetilde{\mathbb{E}}[Z^{(b)}]^\top,$$

for $b\in\{\pm1\}$. Since each Z involves only $2k+1$ variables, the C are covariance matrices of a true distribution and are thus positive semidefinite. Further, let $u^{(1)}, u^{(-1)} \in \mathbb{R}^{2k+1}$ be the vectors with

$$u^{(b)}_S = \begin{cases} \operatorname{sign}\left(\Pr[X_{i_j}=a, X_r=b] - \Pr[X_{i_j}=a]\Pr[X_r=b]\right) \\ \qquad \text{if } S=(j,a) \text{ for } j\in[k], a\in\{\pm1\} \\ -\alpha \text{ if } S=r \end{cases}$$

for some $\alpha > 0$ to be chosen later. Then we have that

$$
\begin{aligned}
0 &\le (u^{(1)})^\top C^{(1)} u^{(1)} + (u^{(-1)})^\top C^{(-1)} u^{(-1)} \\
&= \sum_{j,\ell\in[k]} \sum_{b\in\{\pm1\}} \sum_{a,a'\in\{\pm1\}} u^{(b)}_{(j,a)} u^{(b)}_{\ell,a'} \\
&\cdot \left(\Pr[X_{i_j}=a, X_{i_\ell}=a'] - \Pr[X_{i_j}=a]\Pr[X_{i_\ell}=a']\right) \\
&- 2\alpha \sum_{j\in[k]} \sum_{a,b\in\{\pm1\}} \\
&\left|\Pr[X_{i_j}=a, X_r=b] - \Pr[X_{i_j}=a]\Pr[X_r=b]\right| \\
&+ \alpha^2 \sum_{b\in\{\pm1\}} \left|\Pr[X_r=b] - \Pr[X_r=b]^2\right| \\
&\le 2\left(\sum_{j\ne\ell\in[k]} \widetilde{\mathrm{Cor}}(X_{i_j}, X_{i_\ell})\right) \\
&- 2\alpha\left(\sum_{j\in[k]} \widetilde{\mathrm{Cor}}(X_{i_j}, X_r)\right) \\
&+ \left(\alpha^2\,\widetilde{\mathbb{V}}(X_r) + 2\sum_{j\in[k]} \widetilde{\mathbb{V}}(X_{i_j})\right),
\end{aligned}
$$

where the final inequality follows from the fact that the first k entries of $u^{(b)}$ are signs. Each pair X_{i_j}, X_r is distributed identically to a pair from $i_j \sim_s r$, each pair X_{i_j}, X_{i_ℓ} with $j \ne \ell$ is distributed identically to a pair from $i_j \sim_{2s} i_\ell$, and each i_j, r are distributed according to π. Thus, taking the expectation of the above inequality over a random choice of $r, i_1, \dots, i_k$, we have

$$
\begin{aligned}
0 &\le 2k(k-1)\,\mathbb{E}_{i\sim_{2s}j}\,\widetilde{\mathrm{Cor}}(X_i, X_j) \\
&- 2\alpha k\,\mathbb{E}_{i\sim_s j}\,\widetilde{\mathrm{Cor}}(X_i, X_j) + (\alpha^2 + 2k)\,\mathbb{E}_{i\sim\pi}\,\widetilde{\mathbb{V}}(X_i).
\end{aligned}
$$

Rearranging and simplifying, using $\widetilde{\mathbb{V}}(X_i) \le 1$,

$$
\frac{\alpha\delta}{k} - \frac{\frac{1}{2}\alpha^2 + k}{k^2} \le \mathbb{E}_{i\sim_{2s}j}\,\widetilde{\mathrm{Cor}}(X_i, X_j)
$$

And choosing $\alpha = \delta k$ to maximize the left-hand side, as well as the assumption that $k \ge \frac{4}{\delta^2}$, we have our desired bound. ■

Using Lemma III.1 we can lower bound global correlation in terms of the local correlation and the trace.

Lemma III.6. *Let $\tau > 0$, and suppose G is an n-vertex graph with no isolated vertices and with symmetric normalized adjacency matrix $N = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. Let $X_1, \dots, X_n$ be Boolean T-local random variables with $T \ge 2\left(\frac{16}{\delta}\right)^t + 1$ for t some power of two. If the local correlation $\mathbb{E}_{(i,j)\in E}\,\widetilde{\mathrm{Cor}}(X_i, X_j) \ge \delta$, then the global squared correlation is lower bounded by*

$$
\mathbb{E}_{i,j\sim\pi}\,\widetilde{\mathrm{Cor}}(X_i, X_j)^2 \ge \frac{\left(\frac{1}{16}\delta\right)^{2t}}{\mathrm{Tr}(N^{2t})}.
$$

Proof: By Lemma III.1, the correlation of endpoints of t-length random walks is bounded by the local correlation,

$$
\left(\frac{1}{16}\,\mathbb{E}_{i\sim j}\,\widetilde{\mathrm{Cor}}(X_i, X_j)\right)^t \le \mathbb{E}_{i\sim_t j}\,\widetilde{\mathrm{Cor}}(X_i, X_j) \quad (1)
$$

Further, letting $P = D^{-1}A$ be the transition matrix for the random walk on G and letting π be the stationary measure, we have that

$$
\begin{aligned}
\mathbb{E}_{i\sim_t j}\,\widetilde{\mathrm{Cor}}(X_i, X_j) &= \sum_{i,j} \pi_i \cdot (P^t)_{i,j} \cdot \widetilde{\mathrm{Cor}}(X_i, X_j) \\
&\le \sqrt{\left(\sum_{i,j} \frac{\pi_i}{\pi_j} (P^t)^2_{i,j}\right)\left(\sum_{i,j} \pi_i \pi_j\,\widetilde{\mathrm{Cor}}(X_i, X_j)^2\right)} \\
&= \|D^{\frac{1}{2}} P^t D^{-\frac{1}{2}}\|_F \left(\mathbb{E}_{i,j\sim\pi}\,\widetilde{\mathrm{Cor}}(X_i, X_j)^2\right)^{\frac{1}{2}}, \quad (2)
\end{aligned}
$$

where to obtain the first inequality we have applied Cauchy-Schwarz and the assumption that G has no isolated vertices (which ensures $\pi_j \ne 0$). Now, since the symmetric normalized adjacency matrix N has the property that $N^t = D^{\frac{1}{2}} P^t D^{-\frac{1}{2}}$. Therefore, $\|D^{\frac{1}{2}} P^t D^{-\frac{1}{2}}\|_F = \sqrt{\mathrm{Tr}(N^{2t})}$, and we can combine (2) with (1) to deduce that

$$
\left(\frac{1}{16}\,\mathbb{E}_{i\sim j}\,\widetilde{\mathrm{Cor}}(X_i, X_j)\right)^{2t} \le \mathrm{Tr}(N^{2t}) \cdot \mathbb{E}_{i,j\sim\pi}\,\widetilde{\mathrm{Cor}}(X_i, X_j)^2,
$$

from which the lemma follows. ■

B. Trace bound for low threshold rank graphs

In the previous subsection, we obtained a bound on the global correlation in terms of the local correlation and the trace of a power of the normalized adjacency matrix. Now, we will show a bound on the trace of powers of graphs in terms of the threshold rank. Low threshold rank requires that there are few large-magnitude *positive* eigenvalues but does not provide explicit control over large-magnitude negative eigenvalues. This lemma provides a consequence of bounded threshold rank for the trace of powers of the normalized adjacency matrix.

Claim III.7. *Let $N = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ be the symmetric normalized adjacency matrix of a graph G. If $\mathrm{rank}_\tau(G) \le k$, then for any integer $t \ge 1$, $\mathrm{Tr}(N^{2t}) \le 2(k + n\tau^{2t-1})$.*

Proof: Any symmetric matrix can be written as the sum of its projections on to the positive semidefinite and negative definite cones. Let $N = N_+ + N_-$ where N_+ is the projection of N to the positive semidefinite cone. Then,

$$
\mathrm{Tr}(N_+^{2t+1}) \le k + n \cdot \tau^{2t+1},
$$

since there are at most k eigenvalues $\ge \tau$ and $\|N\| \le 1$. We also have that $\mathrm{Tr}(N^\ell) \ge 0$ for any integer ℓ, since the entries of N are nonnegative. So, since $N = N_+ + N_-$, we have that

$$0 \leq \operatorname{Tr}(N^{2t+1}) = \operatorname{Tr}(N_+^{2t+1}) + \operatorname{Tr}(N_-^{2t+1}) \\ \leq (k + n \cdot \tau^{2t+1}) + \operatorname{Tr}(N_-^{2t+1}),$$

which implies $\operatorname{Tr}(N_-^{2t+1}) \geq -(k + n \cdot \tau^{2t+1})$.

Since $\|N\| \leq 1$, $\operatorname{Tr}(N_-^{2t+2}) \leq |\operatorname{Tr}(N_-^{2t+1})|$. The same is true for N_+, and from this we have

$$\operatorname{Tr}(N^{2t+2}) \leq 2(k + n \cdot \tau^{2t+1})$$

as desired. ■

With Lemma III.6, this shows that the local and global correlation are related in low-threshold-rank graphs.

C. Proof of main theorem

With all these pieces we are ready to prove Theorem I.1, restated more formally below. (Theorem I.3 follows from a subset of the proof below, omitting the graph partitioning argument.)

Theorem III.8 (Sherali-Adams for Max-Cut). *For every $\alpha > 0$ there is an $\varepsilon_\alpha = \exp(-O(1/\alpha^3))$ such that if G is an n node graph with n sufficiently large and $\mathrm{SA}_{n^\alpha}(G) \geq 1 - \varepsilon_\alpha$, then there is a cut which cuts at least a $(\frac{1}{2} + \varepsilon_\alpha)$-fraction of the edges of G, and there is a polynomial time rounding algorithm for the degree-n^α Sherali-Adams LP that produces such a cut. Consequently, the degree-n^α Sherali-Adams LP value provides a $(\frac{1}{2} + \varepsilon_\alpha)$-approximation to Max-Cut.*

Proof: Let ε_α be a small number to be set later. Let $X'_1, \ldots, X'_n$ be T-local random variables for $T \geq n^\alpha$, and let $\delta = \frac{1}{64}$ and $\tau = \left(\frac{\delta}{16}\right)^{2/\alpha}$. Assume that the objective value is $\mathrm{SA}(G) = \mathbb{E}_{i\sim j} \mathbb{E}_{X'_i, X'_j} \mathbb{I}(X'_i \neq X'_j) \geq 1 - \varepsilon_\alpha$.

If $\operatorname{rank}_\tau(G) \leq n^{\alpha/2}$, then we may apply global correlation rounding. We sample $i_1, \ldots, i_k \sim \pi$, where π is the stationary measure on G, for some $k \leq n^\alpha/2$, and then sample values for $X'_{i_1}, \ldots, X'_{i_k}$ according to their local distribution. Conditioning on those values, we obtain $(T - k)$-local Boolean random variables $X_1, \ldots, X_n$, we may assume that both $\mathbb{E}_{i,j\sim\pi} \widetilde{\operatorname{Cor}}(X_i, X_j)^2 \leq 12n^{-\alpha}$ and that $\mathbb{E}_{i\sim j} \mathbb{E}_{X_i,X_j} \mathbb{I}(X_i \neq X_j) \geq 1 - 10\varepsilon_\alpha$, as guaranteed by Theorem III.3 together with Markov's inequality. We now apply Lemma III.6 with t the largest power of two such that

$$\frac{(1 - \frac{1}{2}\alpha)\log n}{2\log\frac{1}{\tau}} \leq t \leq \frac{\alpha \log n}{2\log\frac{16}{\delta}}.$$

It may be checked that our choice of δ and τ ensures that such a t exists. Further, this guarantees that $2\left(\frac{16}{\delta}\right)^t \leq n^\alpha$ (so that the locality of our local random variables is high enough to apply III.6) and also $n\tau^{2t-1} \leq n^{\alpha/2}$. From Lemma III.6 we have

$$\left(\frac{1}{16}\mathbb{E}_{i\sim j}\widetilde{\operatorname{Cor}}(X_i, X_j)\right)^{2t} \leq \operatorname{Tr}(N^{2t}) \cdot \mathbb{E}_{i,j}\widetilde{\operatorname{Cor}}(X_i, X_j)^2 \\ \leq \operatorname{Tr}(N^{2t}) \cdot 12n^{-\alpha}.$$

Since $\operatorname{rank}_\tau(G) \leq n^{\alpha/2}$ and since by our choice of t we have $n\tau^{2t-1} \leq n^{\alpha/2}$ we can apply Claim III.7 to the above to obtain

$$\left(\frac{1}{16}\mathbb{E}_{i\sim j}\widetilde{\operatorname{Cor}}(X_i, X_j)\right)^{2t} \leq 2(\operatorname{rank}_\tau(G) + n\tau^{2t-1}) \cdot 12n^{-\alpha} \\ \leq 48n^{-\alpha/2}.$$

This implies that the local correlation is at most

$$\mathbb{E}_{i\sim j}\widetilde{\operatorname{Cor}}(X_i, X_j) \leq 16 \cdot \left(48n^{-\alpha/2}\right)^{1/t} \leq 2\delta,$$

for n sufficiently large. Applying Lemma III.4 we can round to obtain a solution of value at least $\mathrm{SA}(G) - 2\delta$.

Otherwise, if $\operatorname{rank}_\tau(G) > n^{\alpha/2}$, we apply Theorem III.5 to obtain a partition of G into pieces $G_1, \ldots, G_m$ of threshold rank at most $n^{\alpha/2}$, such that the partition has expansion at most $1 - \exp(-O(\frac{1}{\alpha^2}\log\frac{1}{\tau})) = 1 - \exp(-C\frac{1}{\alpha^3})) \leq 1 - \sqrt{\varepsilon_\alpha}$ (where C is a universal constant that comes from the partitioning theorem, and we have chosen the constant in the theorem statement to make this equality hold). Since each piece has threshold rank at most $n^{\alpha/2}$, we may apply global correlation rounding to each piece as above to obtain an assignment $x^{(i)}$ on the variables of G_i which obtains value $\geq \mathrm{SA}(G_i) - 2\delta$ within G_i. Furthermore, because the objective value $\mathrm{SA}(G)$ remains at least $1 - 2\varepsilon_\alpha$ even after conditioning, and since $\cup_{i\in[m]}|E[G_i]|$ accounts for a $\geq \sqrt{\varepsilon_\alpha}$ fraction of the total edges in the graph, on average $\mathrm{SA}(G_i) \geq 1 - 2\sqrt{\varepsilon_\alpha}$ and so we can round the Sherali–Adams solution within each piece to obtain a solution $x^{(i)}$ so that on average, the value of $x^{(i)}$ within G_i is at least $1-2\sqrt{\varepsilon_\alpha}-2\delta$. That is, on the $\Omega(\varepsilon_\alpha)$ fraction of edges *not* cut by the partition, we have produced a max-cut of value close to the original LP value, which itself is close to 1. Now, we will use randomization to make sure that the edges cut by the partition get value at least $\frac{1}{2}$.

Choosing random signs $s_1, \ldots, s_m$ for each piece of the partition and choosing $s_1 x^{(1)}, \ldots, s_m x^{(m)}$ as our global solution will give a solution of expected value $\geq \frac{1}{2} \cdot (1 - \varepsilon_\alpha) + (1 - 2\sqrt{\varepsilon_\alpha} - 2\delta) \cdot \varepsilon_\alpha = \frac{1}{2} + (\frac{1}{2} - 2\delta)\varepsilon_\alpha - 2\varepsilon_\alpha^{3/2}$, since each edge crossing a partition is cut with probability $\frac{1}{2}$ and the value within the union of the parts is at least $1-2\sqrt{\varepsilon_\alpha}-2\delta$. We can think of the partition as defining a new max-cut instance in which the $s^{(i)}$ are the max cut variables, so this can be derandomized in polynomial time by applying standard arguments (e.g. using the greedy algorithm). By our choice of ε_α, we have also that $2\varepsilon_\alpha^{3/2} < \frac{1}{8}\varepsilon_\alpha$, and we chose $\delta \leq \frac{1}{64}$, from which we obtain the objective value promised in the theorem statement (after rescaling ε_α). This completes the proof. ■

Acknowledgment

TS would like to thank Ryan O'Donnell for the collaboration on [12] and the questions and follow-up conversations which were crucial for the present work. SBH thanks Prasad Raghavendra for helpful conversations as this work was being prepared. SBH was supported by a UC Berkeley Miller Fellowship. TS was supported by NSF grants CCF 1565264 and CNS 1618026. LT was supported by the NSF under grant CCF 1815434 and his work on this project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 834861).

References

[1] M. Charikar, K. Makarychev, and Y. Makarychev, "Integrality gaps for Sherali-Adams relaxations," in *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009, pp. 283–292.

[2] P. K. Kothari, R. Meka, and P. Raghavendra, "Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of CSPs," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017*, 2017, pp. 590–603.

[3] G. Schoenebeck, L. Trevisan, and M. Tulsiani, "Tight integrality gaps for Lovasz-Schrijver LP relaxations of vertex cover and max cut," in *Proceedings of the 39th ACM Symposium on Theory of Computing*, 2007, pp. 302–310.

[4] S. Khot, "On the power of unique 2-prover 1-round games," in *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*. ACM, 2002, pp. 767–775.

[5] S. Arora, B. Barak, and D. Steurer, "Subexponential algorithms for unique games and related problems," *J. ACM*, vol. 62, no. 5, pp. 42:1–42:25, 2015.

[6] B. Barak, P. Raghavendra, and D. Steurer, "Rounding semidefinite programming hierarchies via global correlation," in *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, R. Ostrovsky, Ed. IEEE Computer Society, 2011, pp. 472–481. [Online]. Available: https://doi.org/10.1109/FOCS.2011.95

[7] V. Guruswami and A. K. Sinop, "Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives," in *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*. IEEE, 2011, pp. 482–491.

[8] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

[9] L. Trevisan, "Max Cut and the smallest eigenvalue," in *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009, pp. 263–272.

[10] L. Lovász and A. Schrijver, "Cones of matrices and set-functions and 0–1 optimization," *SIAM journal on optimization*, vol. 1, no. 2, pp. 166–190, 1991.

[11] H. D. Sherali and W. P. Adams, "A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems," *SIAM Journal on Discrete Mathematics*, vol. 3, no. 3, pp. 411–430, 1990.

[12] R. O'Donnell and T. Schramm, "Sherali - adams strikes back," in *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA.*, ser. LIPIcs, A. Shpilka, Ed., vol. 137. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2019, pp. 8:1–8:30. [Online]. Available: https://doi.org/10.4230/LIPIcs.CCC.2019.8

[13] G. Braun, S. Fiorini, S. Pokutta, and D. Steurer, "Approximation limits of linear programs (beyond hierarchies)," in *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, ser. FOCS '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 480–489. [Online]. Available: https://doi.org/10.1109/FOCS.2012.10

[14] P. Raghavendra and N. Tan, "Approximating csps with global cardinality constraints using SDP hierarchies," in *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, Y. Rabani, Ed. SIAM, 2012, pp. 373–387. [Online]. Available: https://doi.org/10.1137/1.9781611973099.33

[15] D. Steurer, "Subexponential algorithms for d-to-1 two-prover games and for certifying almost perfect expansion," *Available at the authors website*, vol. 1, pp. 2–1, 2010.

[16] M. Charikar and A. Wirth, "Maximizing quadratic programs: Extending grothendieck's inequality," in *45th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2004, pp. 54–60.

[17] P. Raghavendra and D. Steurer, "Integrality gaps for strong sdp relaxations of unique games," in *Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, ser. FOCS '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 575–585. [Online]. Available: http://dl.acm.org/citation.cfm?id=1747597.1748073

[18] J. Håstad, "Every 2-csp allows nontrivial approximation," *Computational Complexity*, vol. 17, no. 4, pp. 549–566, 2008.

[19] T. Leighton and S. Rao, "Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms," *Journal of the ACM (JACM)*, vol. 46, no. 6, pp. 787–832, 1999.

[20] N. Fleming, P. Kothari, and T. Pitassi, "Semialgebraic proofs and efficient algorithm design," 2019.