Electronic Colloquium on Computational Complexity, Report No. 129 (2009)

Subsampling Semidefinite Programs and Max-Cut on the Sphere

Boaz Barak* Moritz Hardt† Thomas Holenstein‡ David Steurer§

November 29, 2009

Abstract

We study the question of whether the value of mathematical programs such as linear and semidefinite programming hierarchies on a graph G, is preserved when taking a small random subgraph G' of G. We show that the value of the Goemans-Williamson (1995) semidefinite program (SDP) for MAX CUT of G' is approximately equal to the SDP value of G. Moreover, this holds even if the SDP is augmented with any constraint from a large natural family $\mathcal{C}$ of constraints that includes all constraints from the hierarchy of Lasserre (2001). The subgraph G' is up to a constant factor as small as possible.

In contrast, we show that for linear programs this is *not* the case, and there are graphs G that have small value for the MAX CUT LP with triangle inequalities, but for which a random subgraph G' has value $1 - o(1)$ even for the LP augmented with n^{ε} rounds of the Sherali-Adams hierarchy.

Our results yield a general approach to use the Lasserre hierarchy to solve MAX CUT on the average on certain types of input distributions. In particular we show that a natural candidate for hard instances of MAX CUT— the Feige-Schechtman (2002) graphs that are obtained by sampling a random subset S of the sphere and placing an edge between pairs of points in S that are almost antipodal— is actually easy to solve by SDPs with triangle inequalities. Such a result can be considered more or less folklore in the *dense* case, where S is large enough to form a close approximation of the sphere, but requires quite different techniques to demonstrate in the sparse case.

*Department of Computer Science, Princeton University `boaz@cs.princeton.edu`. Supported by NSF grants CNS-0627526, CCF-0426582 and CCF-0832797, US-Israel BSF grant 2004288 and Packard and Sloan fellowships.

†Department of Computer Science, Princeton University, `mhardt@cs.princeton.edu`. Supported by NSF grants CCF-0426582 and CCF-0832797.

‡Department of Computer Science, ETH Zurich, `thomas.holenstein@inf.ethz.ch`. Work done while at Princeton University and supported by NSF grants CCF-0426582 and CCF-0832797.

§Department of Computer Science, Princeton University `dsteurer@cs.princeton.edu`. Supported by NSF grants 0830673, 0832797, 0528414.

ISSN 1433-8092

Contents

1 Introduction **3**

2 Our results **4**
- 2.1 Using SDP hierarchies for average-case problems . . . 5
- 2.2 Tightness of our theorem . . . 6
- 2.3 Related work . . . 7

3 Overview of proof **7**

4 Preliminaries **9**

5 Proof of the Main Theorem **10**
- 5.1 First step — proxy graph theorem . . . 11
- 5.2 Second and third step – subsampling the third power . . . 13
- 5.3 Putting things together . . . 14

6 Reduction to Quadratic Program over the ℓ_∞-Ball **15**

7 Subsampling theorem for quadratic forms **16**
- 7.1 Proof of Concentration Lemma . . . 17
- 7.2 Proof of Structure Lemma . . . 20
- 7.3 Property testing PSD matrices . . . 24

8 Proxy graphs from graph powers **25**
- 8.1 Subsampling the third power . . . 26

9 Feige-Schechtman graphs and triangle inequalities **27**

10 Subsampling edges **31**

11 Sparse subgraphs and Sherali-Adams Lower Bounds **32**
- 11.1 Lower bounds for vertex subsampling linear programs . . . 35

A Deviation bounds **38**
- A.1 Deviation bounds for submatrices . . . 38
- A.2 A Gaussian tail bound . . . 39
- A.3 Chernoff-like tail inequalities . . . 39

1 Introduction

Recent years have seen many works on *property testing*, showing that combinatorial properties of graphs (and other objects) are at least approximately preserved when taking a small random subgraph, sometimes even only of constant size [Ron00, Rub06]. Such results give insights on the properties in question, and can also be used to design more efficient (even sublinear!) algorithms. In this work we study similar questions, but where the property involved is the value of a mathematical program such as a semidefinite program (SDP). There are several motivations for this question:

Sparsification / sublinear algorithms. Evaluating an SDP on a small subgraph G' of G is obviously more efficient than evaluating it on the whole graph G. Generally if G is an n-vertex Δ-regular graph, we'll be interested in a subgraph G' of size $O(n/\Delta)$, which will have constant degree. Thus such results can be seen as a form of *sparsification* of the graph. Compared to *spectral sparsification* [ST04], our notion of sparsifiers is more relaxed, but it's still strong enough to guarantee that the value of the SDP on G' is close to the value on G. The advantage is both a simple and efficient sparsification procedure, and of course allowing to reduce the number of *vertices* in the graph, making the SDP much more efficient (indeed possibly sublinear) if the original degree Δ was large.[1]

Understanding strong SDPs. One can ask why study the behavior of the SDP value if its main purpose is to approximate a combinatorial property such as the maximum cut, which is already known to be preserved under random subsampling (cf. [GGR98]). One answer is that recent works showed the SDP value is important for its own sake, and not just as an approximation for an integral value. For example, the Max Cut SDP (and more generally Unique Games SDP) captures better than the integral value the behavior of the max-cut game under parallel repetition [BHH$^+$08], while several works showed relations between SDPs and *quantum* protocols (see, e.g., [KRT08] and references therein). Another motivation is to help utilize the power of *SDP hierarchies* such as Lasserre [Las01, Sch08]. These stronger SDPs give very appealing candidates for improved approximation algorithms, but very little is known about how to actually take advantage of their power. In fact, our knowledge on these hierarchies is in the interesting state where we have several strong *lower bounds* (i.e., integrality gaps) [Sch08, Tul09, RS09, KS09] but only very few *upper bounds* (i.e., algorithms) [Chl07, CS08]. Subsampling gives a new way to argue about the value of such strong SDPs, at least for average case problems, by deducing the value on the subgraph G' from the value on G. The former might be easier because G is much denser (see below).

Candidate hard instances for unique games. Our original motivation in this work was to study the question of possible hard instances for Max Cut and Unique Games. Khot's *unique games conjecture* [Kho02] asserts that the unique games problem is **NP**-hard, which would imply that it should be hard on the average as well.[2] Surprisingly, at the moment there are not even

[1] Although in a recent beautiful work [BSS09] showed that constant degree spectral sparsifiers exist, they cannot be found by simple sampling and the known algorithm to find them runs in superlinear time (e.g., $O(|V|^3|E|)$). Moreover, since spectral sparsifiers only reduce the number of edges and not vertices, they cannot lead to sublinear algorithms.

[2] This is assuming **NP**-hardness is via a standard many to one Karp reduction, and there is a hard on the average language in **NP**. The latter assumption is implied by, say, the existence of one-way functions.

natural *candidates* for hard input distributions for this problem. (In contrast to problems like 3-SAT and clique, uniformly random instances of unique games are known to be easy instances [AKK⁺08].)

A similar observation applies to approximating the *maximum cut* of a graph better than the Goemans-Williamson (henceforth GW) algorithm [GW94]. This was shown unique-games hard in [KKMO04, MOO05]. Specifically, these works imply that if the unique games conjecture is true then there is some constant $c > 0$ such that for every $\gamma > 0$, there is such a distribution over graphs whose (normalized) max-cut value is at most $1 - \sqrt{\gamma}$, but it is infeasible to *certify* that the value is less than $1 - c\gamma$. Since such a distribution must fool the known SDPs for this problem, a natural choice is the known integrality gap example by Feige and Schechtman [FS02].[3] These are *random geometric graphs*, in which the vertices are identified with random points in the sphere, and an edge is placed between every two vertices that are sufficiently far apart. This distribution fools the GW SDP, but could it be actually hard?

If sufficiently many points are chosen so as to make, say, a $\gamma/10d$-net in the sphere (of dimension d), then it's easy to certify that the max-cut value is $1 - \Omega(\sqrt{\gamma})$ since the graph can be covered uniformly by edge-disjoint odd cycles of length $O(1/\sqrt{\gamma})$. But in fact the integral value remains $1 - \sqrt{\gamma}$ even if much fewer points are chosen. It's fairly straightforward to see that choosing fewer points can't make the problem easier, and intuitively this can make the problem harder. Indeed, the short odd cycles disappear once fewer points can be chosen, and in fact [FS02] did show that that smaller number of points can fool more powerful algorithms in some settings (they reduced the number of points to get an instance for the SDP with triangle inequalities achieving integrality ratio at most 0.891). More generally, starting from an integrality gap instance and obtaining a distribution on inputs by choosing a random subgraph of it seems like a reasonable approach for getting a candidate hard distribution.[4] Nevertheless, we show that we can still certify with high probability that the max-cut value is $1 - \Omega(\sqrt{\gamma})$ for this distribution, as long as the number of points is some constant factor larger than the minimum needed for the average degree to be more than 1.[5] In fact, this result is a special case of a more general argument, see Section 2.1.

2 Our results

Our main result is that subsampling approximately preserves the SDP value for Max Cut, for a large family of semidefinite programs. In what follows, denote by $\mathrm{sdp}_k(G)$ the value of the Goemans-Williamson (GW) SDP for Max Cut, augmented by k levels of the Lasserre hierarchy. (Here, the value of a Max Cut SDP relaxation is normalized to be between 0 and 1.)

Theorem 2.1 (Main). *For every graph G on n vertices and degree Δ, $k \in \mathbb{N}$, and $\varepsilon > 0$, there is $n' = \mathrm{poly}(\varepsilon^{-1})n/\Delta$ such that letting $\gamma = 1 - \mathrm{sdp}_k(G)$, if we choose a random subgraph G' of G with n' vertices then w.h.p.*

$$\frac{\gamma}{3} - \varepsilon \leqslant 1 - \mathrm{sdp}_k(G') \leqslant \gamma + \varepsilon\,. \tag{1}$$

The theorem is actually true for a more general family of constraints than Lasserre constraints—we chose to state in this way for simplicity. We emphasize that the sample complexity is independent of the set of constraints, and hence k can be chosen arbitrarily between 1 and n.

[3] In this work we focus on max-cut for graphs that have cuts that cut a $1 - \gamma$ fraction of the edges for small γ, rather than the problem of maximizing the ratio between the maximum cut and the certified value.

[4] In particular it seems to work for 3SAT [Sch08], see discussion in Section 2.1.

[5] Below that minimum the graph is mostly empty and the max-cut value is close to 1 for trivial reasons.

Subsampling quadratic programs. We also obtain a similar result for *quadratic programs*, where the value $1 - \mathrm{sdp}_k(G)$ is replaced with $\max_{x \in [-1,+1]^n} \langle x, Ax \rangle$. Here, A could be the Laplacian of a graph, but in fact our results for quadratic programs apply to a more general family of matrices, and are also quantitatively tighter in the sense that the constant 3 is replaced with 1 in Equation (1). See Theorem 7.1 for a precise statement.

Property testing for psd matrices. As a corollary of the latter result we give a property testing algorithm for positive semidefiniteness that uses $\mathrm{poly}(\varepsilon^{-1})n/\Delta$ samples to distinguish between positive semidefinite matrices and matrices A for which there is a vector $x \in [-1,+1]^n$ with $\langle x, Ax \rangle \leqslant -\varepsilon$ so long as the matrix A satisfies some natural requirements. The runtime of our tester is polynomial in the size of the sample.

Sherali-Adams lower bound. In contrast we show that such subsampling bounds do *not* hold for *linear programming* hierarchies. Indeed, we show the following theorem:

Theorem 2.2. *For every function $\varepsilon = \varepsilon(n)$ that tends to 0 with n, there exists a function $r = r(n)$ that tends to ∞ with n and family of graphs $\{G_n\}$ of degree $D = D(n)$ such that*

1. *For every n, $\mathrm{lp}_3(G_n) \leqslant 0.8$*
2. *If G' is a random subgraph of G of size $(n/D)^{1+\varepsilon(n)}$ then $\mathbb{E}[\mathrm{lp}_{r(n)}(G')] \geqslant 1 - \frac{1}{r(n)}$.*

where $\mathrm{lp}_k(H)$ denotes the value of k levels of the Sherali-Adams linear program for Max-Cut [SA90, dlVKM07] on the graph H.

Note that Theorem 2.2 shows that subsampling for LPs fails in a very strong way, compared to the bound for SDPs we get from Theorem 2.1: the size of the sample in Theorem 2.2 is $(n/D)^{1+\varepsilon}$ compared to only $O(n/D)$ vertices in Theorem 2.1, and the value on the subgraph in Theorem 2.2 remains very close to 1 even if we allow the number of rounds of the LP to grow from 3 to an unbounded function of n.

Edge subsampling. In addition to the above, we also prove preservation of SDP value under *edge subsampling*, where one reduces the degree of a Δ-regular graph to constant by letting every edge of the graph survive with probability $p \sim 1/\Delta$ independently. The proof in this case is much simpler, and yields tighter quantitative results (replacing 3 with 1 in Equation (1)). But note that many of our intended applications are not implied by such a result.

2.1 Using SDP hierarchies for average-case problems

One corollary of our work is a general way to use semidefinite programming hierarchies for solving Max Cut (and possibly other 2-CSPs) in the *average-case*. Specifically, let $\mathcal{G}$ be some distribution on n-vertex graphs that is obtained by *subsampling* random vertices and/or edges from a denser N-vertex graph G of degree at least εN. One example is the distribution $G_{n,p}$ that is obtained by subsampling edges from the complete graph.

Now Theorem 2.1 implies that Max Cut on $\mathcal{G}$ is well approximated by $\mathrm{poly}(1/\varepsilon)$ rounds of the Lasserre hierarchy. Here we say that an SDP approximates well the max-cut value of a graph G if $1 - \mathrm{sdp}(G) = \Theta(1 - \mathrm{maxcut}(G))$. Note under the unique games conjecture such an approximation is infeasible in the worst-case. The reason is that Max Cut on ε-dense graphs is easy for a poly$(1/\varepsilon)$-round SDP (and even a poly$(1/\varepsilon)$-round LP [dlVKM07]) since the known property testing bounds, e.g. [GGR98], imply that for such graphs the integral value is preserved on a random poly$(1/\varepsilon)$ subsample. Our result shows that the *SDP value* is preserved as well under subsampling. Note that the latter is false for linear programs (see Theorem 2.2). Even for semidefinite programs it's false for 3-CSPs, as the result of [Sch08] shows that random 3SAT formulas are hard for Lasserre SDPs and these can be thought of as subsamples of the trivial 3SAT formula containing all possible clauses.

Our result is more general, and holds even if the original graph has sublinear degree, as long as it is an easy instance for the r-round Lasserre SDP for some r. In particular we apply it to show how to solve Max Cut on *random geometric graphs.* Specifically, we show that the Feige-Schechtman Max-Cut instances [FS02] can in fact be solved (i.e., certified to have small maximum cut) by the GW augmented with triangle inequalities:

Theorem 2.3. *Let $\gamma \in [0,1]$ and $d \in \mathbb{N}$. Then there is a constant $a = a(\gamma)$ such that if we let $n = a/\mu$, where μ is the normalized measure of the $2\sqrt{\gamma}$-ball in the sphere S^{d-1}, and let the n vertex graph G be obtained by picking n random unit vectors $v_1, \ldots, v_n$ and placing the edge $\{i,j\}$ if $\frac{1}{4}\|v_i - v_j\|^2 \geqslant 1-\gamma$, then w.h.p.*

$$\mathrm{sdp}_3(G) = 1 - \Theta(\sqrt{\gamma})\,.$$

This theorem also holds for the GW SDP augmented with the ℓ_2^2 triangle inequalities, and furthermore, from now on we will use $\mathrm{sdp}_3(G)$ to denote this GW SDP augmented with triangle inequalities (This simplifies stating the results somewhat, and comes with almost no loss of generality, as the 3^{rd} level of the Lasserre hierarchy is not often considered by itself).

We note that the [FS02] instances come with a "planted" SDP solution that does not satisfy the triangle inequality constraints. The question here is whether there exists any other solution with approximately the same value that does satisfy these constraints.

If the number of points n is extremely large compared to d, then we believe that such a result is more or less folklore (though we could not find a reference and hence supply a proof in Section 9). But the question becomes very different when the number n of points becomes smaller, or equivalently when the graph becomes *sparser.* Indeed, Feige and Schechtman used such sparse graphs to show that there is some specific constants $\gamma, \varepsilon > 0$ and a choice of n for which $\mathrm{sdp}_3(G) > \mathrm{maxcut}(G) + \varepsilon$ w.h.p for a graph G chosen as in Theorem 2.3. Our result shows that there is a limit to what can be gained from such sparsification. (We note that Theorem 2.3 applies to even sparser graphs than those considered by [FS02] since they used graphs of super-constant degree.)

Under the unique games conjecture, adding triangle inequalities does not help in approximating max cut in the worst case. Yet Theorem 2.3 shows that triangle inequalities do help for this very natural distribution of inputs.

2.2 Tightness of our theorem

Theorem 2.1 is weaker than known results for property testing of *integral* properties in two respects. First, we only get approximation up to a fixed constant (i.e. 3), rather than approaching the value as the sample size grows. Second, it does not generalize to k-CSP's for $k > 2$ (see [AdlVKK03]). As we mentioned above, the second limitation is inherent (e.g., Schoenebeck's result on integrality gaps for random 3SAT [Sch08]). (On a related note, we do suspect that our techniques generalize to arbitrary 2-CSPs.) We believe, though have not yet been able to prove this, that the first limitation is inherent as well. Specifically, while Feige and Schechtman [FS02] showed that there is a random geometric graph in which there is an absolute constant gap between the max-cut and the SDP value (with triangle inequalities), we believe there is no such gap if the random geometric graph is sufficiently dense. This would show that the constant 3 in Equation (1) of Theorem 2.1 cannot be replaced by 1. These observations might also explain why our proof doesn't follow directly from the techniques and results of property testing, but rather we need to use some other tools such as [CW04]'s rounding algorithm (which is specific to 2-CSPs) and consider subsampling of a graph power (resulting in a constant factor loss in the approximation). See our proof overview below.

2.3 Related work

There has been quite some work on estimating graph parameters from random small induced subgraphs. Goldreich, Goldwasser and Ron [GGR98] show that the the Max-Cut value of a dense graph (degree $\Omega(n)$) is preserved when passing to a constant size induced subgraph. (In this and other results, the constants depend on the quality of estimation.) Feige and Schechtman [FS02] showed that the result holds generally for D-regular[6] graphs so long as the degree $D \geqslant \Omega(\log n)$ and the subgraph is of size at least $\Omega(n \log n/D)$. (As a corollary of our results, we slightly strengthen [FS02]'s bounds to hold for any $D > \Omega(1)$ and subgraph size larger than $\Omega(n/D)$.)

Alon et al [AdlVKK03] generalize [GGR98] for k-CSP's and improve their quantitative estimates. See also [RV07] for further quantitative improvements in the case of $k = 2$. Similar techniques were also used in the context of low rank approximation, where the goal is to approximate a matrix A by a low rank matrix D so as to minimize, say, the Frobenius norm $\|A - D\|_F$ [FKV04, RV07]. [AdlVKK03, RV07]'s result is obtained via a subsampling theorem for the *cut norm* which is a special kind of a quadratic program corresponding to a bipartite graph. Thus our subsampling theorem for quadratic programs can be viewed as an extension of these results to the general (non-bipartite) case (see also Remark 7.18). [AN04] gave approximation algorithms to the cut norm, while [CW04, AMMN05] gave such algorithms for more general quadratic programs.

There has also been much work on matrix sparsification by means other than sampling a small induced submatrix, see for instance [ST04, AHK06, AM07, SS08, BSS09]. However, these don't reduce the number of vertices but only the edges. Also, often these methods use non-uniform sampling techniques to achieve sparsification guarantees that would otherwise be impossible. As an example, constant degree spectral sparsification [BSS09] cannot be achieved by sampling vertices (or even edges for that matter) uniformly at random.

3 Overview of proof

In this section we give a broad overview of the proof of Theorem 2.1— our main result. It is obtained in three steps that are outlined in Figure 1.

[6] Their result can be generalized to non-regular graphs as well, but in this work we focus our attention only on regular graphs, as it simplifies many of the calculations.

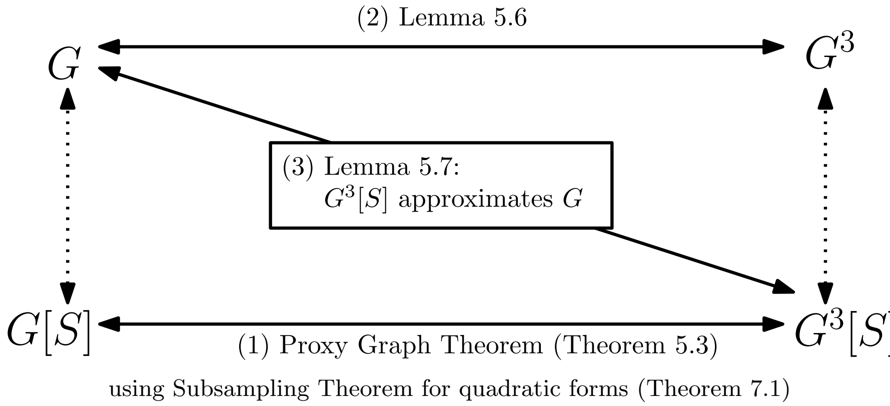

Figure 1: Overview of the proof of Theorem 2.1. $H[S]$ denotes the graph obtained by subsampling H to a random subset S of appropriate size. Arrows indicate that the SDP values of the corresponding graphs are related up to constant factors, dashed arrows being consequences of bold arrows.

Step 1: Proxy graph theorem. The first step (Theorem 5.3) gives a general way to analyze the behavior of the SDP value of an n vertex graph G under subsampling. Here and in the following, subsampling will always refer to choosing an induced subgraph uniformly at random. We show that for every two graphs G and H where G is of degree D and H is at least as dense as G, the following is true: If the Laplacian of H is close the Laplacian of G, then as we subsample $O(n/D)$ vertices from each graph the SDP values of the two induced subgraphs will be close to each other. This reduces the task of analyzing constant degree subsamples of G to the task of analyzing subsamples of some "proxy graph" H of our choosing. This can be very useful, as we can choose H to be much *denser* than G, which means that n/D-sized subgraphs of H will be denser as well. In some sense this is similar to choosing a much larger sample of vertices, and of course the larger the sample, the easier it is to argue that the SDP value is preserved. We stress that this result is very general and works regardless of the additional constraints placed on the SDP.

Tool: Subsampling quadratic programs. The proof of this "proxy graph" step uses the semidefinite rounding algorithm of Charikar and Wirth [CW04] to reduce the problem to analyzing subsampling of a *quadratic program* over $[-1,+1]^n$. (Note that we use rounding here in the "reverse direction"— reducing a question about semidefinite programs into a question about quadratic programs.) We then give in Theorem 7.1 a subsampling theorem for such quadratic programs. To do so we generalize the result of Feige and Schechtman [FS02] (that build on Goldreich, Goldwasser and Ron [GGR98]), that gave such a subsampling theorem for the integral max-cut program. Some complications arise since in our setting the matrices can have negative entries, and also there is no guarantee that there exists an *integral* optimal solution to quadratic program (i.e., a solution in $\{-1,+1\}^n$). Our result is also a bit tighter than [FS02], that required the degree of the subsampled graph to be at least logarithmic, where we only need constant degree. Still the proof follows the general ideas of [FS02, GGR98].

Step 2: Use G^3 as a proxy. To use the first step, we need to come up with an appropriate "proxy graph". This is done in our very simple second step— Lemma 8.1— where we observe that the Laplacian of G^3 approximates the Laplacian of G in the sense needed for Theorem 5.3.

Step 3: Relate subsample of G^3 to $\mathrm{sdp}(G)$. The natural last step would be to show that the value of G^3 is preserved under subsampling. (Note that a random subgraph of G^3 can be much denser than a random subgraph of G, and also much denser than the third power of a random subgraph of G.) However, we don't know how to show that directly. Rather, in our third step—Lemma 5.7— we show that a random subsample of G^3 is likely to have SDP value close to the value of G itself! This is enough, since by transitivity we can argue that all these SDP values (of G, G^3, the subsample of G and subsample of G^3) are close to one another. To prove this step we use the fact that one can choose a random edge in G^3 by first choosing a random edge $e = (u, v)$ in G, and then choosing a random neighbor u' of u and a random neighbor v' of v. We then argue that G^3 is sufficiently dense so that even if we restrict to neighbors u', v' that occur in the subsample, we still get the uniform distribution on the original edge e. This allows us to claim that one can "decode" (in a way slightly reminiscent of Dinur's gap amplification [Din07]) an SDP solution for the subsample of G^3 into an SDP solution for the original graph G. Once more this step applies to a very general family of SDPs, including essentially any reasonable constraint that holds for integral solutions.

4 Preliminaries

Let G be a Δ-regular graph with vertex set $V = [n]$ and edge set E (no parallel edges or self-loops). We give weight ${}^2/_{\Delta n}$ to each edge of G so that every vertex of G has (weighted) degree $2/n$ and G has total edge weight 1. We say a graph is *normalized* if it has total edge weight 1. (We choose this normalization, because we will often think of a graph as a probability distribution over (unordered) vertex pairs.)

We let $L(G)$ denote the (combinatorial) *Laplacian* of G defined by $L(G) := D(G) - A(G)$, where $D(G)$ is the diagonal matrix of the degrees and $A(G)$ denotes the adjacency matrix of G (so that $D(G)_{i,i} = \sum_j A(G)_{i,j}$). For normalized Δ-regular G, we have $D(G) = \frac{2}{n} I$ where I is the identity matrix.

We will use the notation $A \bullet B := \mathrm{tr}(A^T B)$ when A and B are matrices of the same dimension. For a real-valued matrix $A = (a_{ij})$, we denote the *total variation norm* of A by $\|A\|_{\mathrm{TV}} := \sum_{ij} |a_{ij}|$.

We write $A \succeq 0$ to express that A is (symmetric) positive semidefinite, i.e., A is symmetric and $\langle x, Ax \rangle \geqslant 0$ for all $x \in \mathbb{R}^n$. Likewise, the notation $A \succeq B$ and $A \preceq B$ is short for $A - B \succeq 0$ and $B - A \succeq 0$, respectively. (Most matrices we consider will be symmetric.)

Subgraphs and principal submatrices. For a graph G as above and a vertex subset $U \subseteq V$, let $G[U]$ denote the *induced subgraph* on U. To preserve our normalization, we scale the weights of the edges of $G[U]$ such that the total edge weight in $G[U]$ remains 1. We denote by V_δ a random subset of a δ fraction of the vertices in V, and hence $G[V_\delta]$ denotes a random induced subgraph of G of size $\delta|V|$. With our normalization, the typical weight of an edge in $G[V_\delta]$ is ${}^2/_{\delta^2 \Delta n}$.

If B is an $n \times n$ matrix (thought of as a Laplacian of a graph), and $U \subseteq [n]$, then we define (B_U) to be an $n \times n$ matrix such that $(B_U)_{ij} = 0$ if $i \notin U$ or $j \notin U$, $(B_U)_{i,i} = (n/|U|)B_{i,i}$ for every $i \in U$, and $(B_U)_{i,j} = (n/|U|)^2 B_{i,j}$ for every $i \neq j \in U$. Thus if B is the Laplacian of G then B_U roughly corresponds to the Laplacian of $G[U]$ as described above.

Max Cut and its Semidefinite Programming Relaxation. The value of the maximum cut of a graph G is given by $\mathrm{opt}(G) := \max_{x\in\{-1,1\}^n} \langle x, {}^1\!/\!_4 L(G)x\rangle$. The Goemans-Williamson [GW94] relaxation for Max Cut is

$$\mathrm{sdp}(G) = \max\left\{ {}^1\!/\!_4 L(G) \bullet X \mid X \succeq 0, \forall i\colon X_{ii} = 1 \right\} . \tag{2}$$

Note that $\mathrm{opt}(G)$ and $\mathrm{sdp}(G)$ range between 0 and 1, the total edge weight of a normalized graph.

We will consider relaxations obtained by adding valid constraints to the above program. If $\mathcal{M}$ is a convex subset of the set of $n \times n$ positive semidefinite matrices, we let $\mathrm{sdp}_{\mathcal{M}}(G)$ denote the value obtained when taking the maximum in Equation (2) only over matrices in $\mathcal{M}$. We say that $\mathrm{sdp}_{\mathcal{M}}(G)$ is an SDP relaxation of Max Cut, if $\mathcal{M}$ contains all contains all matrices of the form xx^T with $x \in \{\pm 1\}^n$. A specific set of constraints we'll be interested in are the *triangle inequalities* that assert that for all vertices $i, j, k \in V$,

$$X_{ij} + X_{jk} - X_{ik} \leqslant 1 \quad \text{and} \quad - X_{ij} - X_{jk} + X_{ji} \geqslant -1 \,. \tag{3}$$

The relaxation including triangle inequalities will be denoted $\mathrm{sdp}_3(G)$.

Reducing Max Cut to Min UnCut. It will be notationally convenient for us to reduce Max Cut to an equivalent minimization problem. This can be done easily by considering Min UnCut which asks to find a cut that minimizes the number of uncut edges. More formally, the integral value is defined as $\min_{x\in\{-1,1\}^n} \langle x, {}^1\!/\!_4(D(G) + A(G))x\rangle$. We will henceforth denote

$$L_+(G) = D(G) + A(G) \,,$$

A normalized graph G has max-cut value $\mathrm{opt}(G) = 1 - \gamma$ if and only if its min uncut value is γ (since for a normalized regular graph, ${}^1\!/\!_4 L(G) = {}^1\!/\!_n I - {}^1\!/\!_4 L_+(G)$). In the following, we will ignore the factor ${}^1\!/\!_4$ and think of Min UnCut as the problem of minimizing $\langle x, L_+(G)x\rangle$ over $x \in [-1, 1]^n$. The Goemans–Williamson SDP relaxation for Max Cut corresponds to the following SDP relaxation for Min UnCut,

$$\mathrm{sdp}(G) \stackrel{\text{def}}{=} \min L_+(G) \bullet X \,, \tag{4}$$

where the minimum is over all matrices $X \succeq 0$ with $X_{ii} = 1$ for all $i \in V$. Similarly, $\mathrm{sdp}_3(G)$ is defined by restricting the minimum in (4) to matrices X that satisfy in addition the triangle inequalities (3). We have deliberately overloaded notation for the program $\mathrm{sdp}(G)$ ($\mathrm{sdp}_3(G)$) and we will refer to it as the Goemans-Williamson SDP (with triangle inequalities) from here on.

5 Proof of the Main Theorem

In this section, we prove our main theorem, which asserts that the value of "reasonable" SDP relaxations for Max Cut is preserved under vertex-subsampling. This section can also be seen as a formal overview over the remainder of the paper.

Let us first formalize the notion of a "reasonable" SDP relaxation. (We are not aware of any Max Cut relaxation in the literature that is not reasonable according to our definition. In particular, Max Cut relaxations obtained by Lasserre's hierarchy are reasonable.)

Definition 5.1 (Reasonable SDP relaxation for MAX CUT)**.** Let V be a set of vertices and let $\mathcal{M}$ be a convex subset of the set

$$\mathcal{M}_2 \stackrel{\text{def}}{=} \left\{ X \in \mathbb{R}^{V \times V} \mid X \succeq 0, \quad \forall i \in V. \;\; X_{ii} = 1 \right\} .$$

For a graph G with vertex set V, we define $\mathrm{sdp}_{\mathcal{M}}(G)$ by

$$\mathrm{sdp}_{\mathcal{M}}(G) \stackrel{\text{def}}{=} \min_{X \in \mathcal{M}} L_+(G) \bullet X \,.$$

We say that $\mathrm{sdp}_{\mathcal{M}}$ is a *reasonable* relaxation for MAX CUT (or rather MIN UNCUT) if $\mathcal{M}$ is closed under renaming of coordinates in the sense that

$$\forall F \colon V \to V. \quad \forall X \in \mathcal{M}. \quad (X_{F(i),F(j)})_{i,j \in V} \in \mathcal{M} \,.$$

Here, the function F is not required to be bijective. We also say that $\mathcal{M}$ is *reasonable* if it satisfies the condition above.

Theorem 5.2 (Reformulation of Theorem 2.1)**.** *Let $\varepsilon > 0$ and let G be normalized Δ-regular graph with vertex set V. Then, for $\delta = \Delta^{-1} \cdot \mathrm{poly}(1/\varepsilon)$,*

$$\tfrac{1}{3}\mathrm{sdp}_{\mathcal{M}}(G) - \varepsilon \leqslant \mathbb{E}\, \mathrm{sdp}_{\mathcal{M}}(G[V_\delta]) \leqslant \mathrm{sdp}_{\mathcal{M}}(G) + \varepsilon \,,$$

where $\mathrm{sdp}_{\mathcal{M}}$ is any reasonable relaxation. (Recall that the graphs G and $G[V_\delta]$ are normalized so that both have total edge weight 1.)

5.1 First step — proxy graph theorem

Our first step is summarized in the following theorem.

Theorem 5.3 (Proxy Graph Theorem)**.** *Let G and H be two normalized regular graphs with vertex set V and degrees at least Δ such that for a constant $c > 0$ (say $c = 1/3$),*

$$L_+(G) \succeq c L_+(H) \,. \tag{5}$$

Then, for $\delta = \Delta^{-1}\mathrm{poly}\,(1/\varepsilon)$ and any reasonable relaxation $\mathrm{sdp}_{\mathcal{M}}$,

$$\mathbb{E}\, \mathrm{sdp}_{\mathcal{M}}(G[V_\delta]) \geqslant c \cdot \mathbb{E}\, \mathrm{sdp}_{\mathcal{M}}(H[V_\delta]) - \varepsilon \,.$$

To carry out the proof of this theorem, we need the following general technical results (proven in Section 6 and Section 7).

Lemma 5.4 (Reformulation of Lemma 6.1)**.** *Let $\varepsilon > 0$ and let $A \in \mathbb{R}^{V \times V}$ be such that $\max_{x \in [-1,1]^V} \langle x, Ax \rangle \leqslant \varepsilon \|A\|_{\mathrm{TV}}$. Then,*

$$\max A \bullet X \leqslant O\left(\varepsilon \log\left(1/\varepsilon\right)\right) \|A\|_{\mathrm{TV}} \,,$$

where the maximum is over all matrices $X \in \mathbb{R}^{V \times V}$ with $X \succeq 0$ and $X_{ii} \leqslant 1$ for all $i \in V$.

Lemma 5.5 (Reformulation of Theorem 7.1)**.** *Let $\varepsilon > 0$ and let $B \in \mathbb{R}^{V\times V}$ be a matrix with diagonal entries and row and column sums bounded by a constant (say 10). Then for $\delta \geqslant \|B - \mathrm{diag}(B)\|_\infty \cdot \mathrm{poly}\,(1/\varepsilon)$ (here $\mathrm{diag}(B)$ denotes matrix B restricted to its diagonal),*

$$\mathbb{E} \max_{x\in[-1,1]^V} \langle x, B_{V_\delta} x\rangle = \max_{x\in[-1,1]^V} \langle x, Bx\rangle \quad \pm \varepsilon n\,.$$

Assuming these two results we conclude our first step with the proof of the proxy graph theorem.

Proof of Theorem 5.3. Every vertex has weighted degree $2/n$ in both G and H. Hence, the diagonal entries and the row and column sums of $B := cL_+(H) - L_+(G)$ are bounded by $O(1/n)$. Since the graphs have degrees at least Δ, we also have $\|B - \mathrm{diag}(B)\|_\infty \leqslant O(1/\Delta n)$. Thus, by Lemma 5.5 (applied to $n \cdot B$),

$$\mathbb{E} \max_{x\in[-1,1]^V} \langle x, B_{V_\delta} x\rangle \leqslant \max_{x\in[-1,1]^V} \langle x, Bx\rangle + \varepsilon \leqslant \varepsilon\,. \tag{6}$$

For the last inequality, we used that B is negative semidefinite (by assumption (5)).

On the other hand, Lemma 5.4 implies for $A := cL_+(H[V_\delta]) - L_+(G[V_\delta])$,

$$\max_{X\in\mathcal{M}} A \bullet X \leqslant O\left(\varepsilon_{V_\delta} \log\left(\tfrac{1}{\varepsilon_{V_\delta}}\right)\right)\,, \tag{7}$$

where ε_{V_δ} is defined as $\varepsilon_{V_\delta} := \max_{x\in[-1,1]^V} \langle x, Ax\rangle$ (notice that A depends on V_δ). Here, we also used that $H[V_\delta]$ and $G[V_\delta]$ are normalized so that their total edge weight is 1.

Let us estimate the expectation of ε_{V_δ},

$$\begin{aligned}
\mathbb{E}\, \varepsilon_{V_\delta} &= \mathbb{E} \max_{x\in[-1,1]^V} \langle x, Ax\rangle \\
&\leqslant \mathbb{E} \max_{x\in[-1,1]^V} \langle x, B_{V_\delta} x\rangle + \mathbb{E}\|A - B_{V_\delta}\|_{\mathrm{TV}} \\
&\leqslant \varepsilon + \mathbb{E}\|L(G[V_\delta]) - L(G)_{V_\delta}\|_{\mathrm{TV}} + \mathbb{E}\|L(H[V_\delta]) - L(H)_{V_\delta}\|_{\mathrm{TV}} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(using (6) and triangle inequality)} \\
&\leqslant 3\varepsilon\,.
\end{aligned} \tag{8}$$

In the last inequality, we used if G is normalized and regular with degree at least Δ, then the expectation of $\mathbb{E}\|L(G[V_\delta]) - L(G)_{V_\delta}\|_{\mathrm{TV}}$ is bounded by ε (same for H). See Lemma A.2 for a proof.

We can now conclude the proof of the theorem

$$\begin{aligned}
&c \cdot \mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(H[V_\delta]) - \mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(G[V_\delta]) \\
&\leqslant \mathbb{E} \max_{X\in\mathcal{M}} (cL_+(H[V_\delta]) - L_+(G[V_\delta])) \bullet X && \text{(by definition of } \mathrm{sdp}_{\mathcal{M}}) \\
&\leqslant O\left(\mathbb{E} f(\varepsilon_{V_\delta})\right) && \text{(using (7), where } f(x) = x\log(1/x)) \\
&\leqslant O\left(f(3\varepsilon)\right) && \text{(using (8) and concavity of } f) \\
&= O\left(\varepsilon \log(1/\varepsilon)\right)\,.
\end{aligned}$$

$\blacksquare$

5.2 Second and third step – subsampling the third power

In this section, we show that taking $H = G^3$ in Lemma 5.3 allows us to prove our main theorem (Theorem 5.2. (The graph G^3 corresponds to the uniform distribution over walks of length 3 in G.) We need the following properties of G^3.

Lemma 5.6 (Special case of Lemma 8.1)**.** *Let G be normalized regular graph Then,*

$$L_+(G) \succeq \tfrac{1}{3} L_+(G^3)\,.$$

Proof Sketch. Let A be the adjacency matrix of G scaled up by a factor of $n/2$ so that each degree in G is 1 and A is a stochastic matrix. Note that $L_+(G) = 2/n(I + A)$ and $L_+(G^3) = 2/n(I + A^3)$. The matrices $I + A$ and $I + A^3$ have the same eigenvectors. Therefore, to prove the lemma it is enough to verify that $1 + \lambda \geqslant 1/3(1 + \lambda^3)$ for every eigenvalue λ of A. Since A is stochastic, all its eigenvalues lie in the interval $[-1, 1]$. It is easy to verify that $1 + x \geqslant 1/3(1 + x^3)$ for all $x \in [-1, 1]$. (See Lemma 8.1 for details.) ■

Lemma 5.7 (See Section 8.1)**.** *Let G be a normalized Δ-regular graph with vertex set V and let G' be the (weighted) graph on V_δ defined by the following edge distribution:*

- *sample a random edge (i, j) from G,*
- *choose u and v to be random neighbors of i and j in V_δ (if i or j have no neighbor in V_δ, choose a random vertex in V_δ),*
- *output (u, v) as an edge in G'.*

Then for $\delta > \Delta^{-1}\mathrm{poly}(1/\varepsilon)$,

$$\mathbb{E}\|L_+(G^3[V_\delta]) - L_+(G')\|_{\mathrm{TV}} \leqslant \varepsilon\,.$$

Proof Sketch. If every vertex of G has the same number of neighbors in V_δ, then the two graphs $G^3[V_\delta]$ and G' are identical. For $\delta > \Delta^{-1}\mathrm{poly}(1/\varepsilon)$, the following event happens with probability $1 - \varepsilon$: Most vertices of G (all but an ε fraction) have up to a multiplicative $(1 \pm \varepsilon)$ error the same number of neighbors in V_δ. Conditioned on this event, it is possible to bound $\|L_+(G^3[V_\delta]) - L_+(G')\|_{\mathrm{TV}}$ by $O(\varepsilon)$. Assuming this fact, the lemma follows. ■

Lemma 5.8. *Let G be a normalized Δ-regular graph. Then for $\delta > \Delta^{-1} \cdot \mathrm{poly}(1/\varepsilon)$ and for any reasonable relaxation $\mathrm{sdp}_{\mathcal{M}}$,*

$$\mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(G^3[V_\delta]) \geqslant \mathrm{sdp}_{\mathcal{M}}(G) - \varepsilon\,.$$

Proof. Let G' be as in Lemma 5.7 and let X' be an optimal solution for $\mathrm{sdp}_{\mathcal{M}}(G')$ so that $L_+(G') \bullet X' = \mathrm{sdp}_{\mathcal{M}}(G')$. For a mapping $F\colon V \to V_\delta$, let $\mathcal{A}_F(X')$ denote the matrix $(X'_{F(i),F(j)})_{i,j\in V} \in \mathbb{R}^{V\times V}$. (Here, $\mathcal{A}_F$ can be thought of a linear (super-)operator, that maps matrices in $\mathbb{R}^{V_\delta\times V_\delta}$ to matrices in $R^{V\times V}$.) Since $\mathcal{M}$ is reasonable (see Definition 5.1), we have $\mathcal{A}_F(X') \in \mathcal{M}$ for any mapping $F\colon V \to V_\delta$. Let $\mathcal{F}(V_\delta)$ denote the distribution over mappings $F\colon V \to V_\delta$, where for every vertex $i \in V$, we choose $F(i)$ to be a random neighbor of i in V_δ (and if i has no neighbor in V_δ, we choose $F(i)$ to be a random vertex in V_δ). For convenience, we introduce the notation

$N(i,V_\delta)$ for the set of neighbors of i in V_δ (if i has no neighbor in V_δ, we put $N(i,V_\delta)=V_\delta$). We define $\mathcal{A}_{\mathcal{F}(V_\delta)} := \mathbb{E}_{F\sim\mathcal{F}(V_\delta)} \mathcal{A}_F$. Since $\mathcal{M}$ is convex, we also have

$$\mathcal{A}_{\mathcal{F}(V_\delta)}(X') = \mathop{\mathbb{E}}_{F\sim\mathcal{F}(V_\delta)} \mathcal{A}_F(X') = \Big(\mathop{\mathbb{E}}_{F\sim\mathcal{F}(V_\delta)} X'_{F(i),F(j)}\Big)_{i,j\in V} \in \mathcal{M}\,.$$

Let $A=(a_{ij})$ be the adjacency matrix of G and let $A'=(a'_{uv})$ be the adjacency matrix of G'. Recall that $\|A\|_{\mathrm{TV}}=\|A'\|_{\mathrm{TV}}=2$. Hence, a_{ij} is the probability that a random edge of G connects i to j and a'_{uv} is the probability that a random edge of G' connects u to v (see Lemma 5.7 for the edge distribution of G'). We write $ij\sim G$ and $uv\sim G'$ to denote a random edge in G and G', respectively. We claim that $A\bullet\mathcal{A}_{\mathcal{F}(V_\delta)}(X') = A'\bullet X'$,

$$\begin{aligned}
A\bullet\mathcal{A}_{\mathcal{F}(V_\delta)}(X') &= 2\mathop{\mathbb{E}}_{ij\sim G} \mathcal{A}_{\mathcal{F}(V_\delta)}(X')_{ij}\\
&= 2\mathop{\mathbb{E}}_{ij\sim G}\ \mathop{\mathbb{E}}_{F\sim\mathcal{F}(V_\delta)} X'_{F(i),F(j)}\\
&= 2\mathop{\mathbb{E}}_{ij\sim G}\ \mathop{\mathbb{E}}_{u\in N(i,V_\delta)}\ \mathop{\mathbb{E}}_{v\in N(j,V_\delta)} X'_{uv}\\
&= 2\mathop{\mathbb{E}}_{uv\sim G'} X'_{uv} = X'\bullet A'\,.
\end{aligned}\tag{9}$$

Note that the diagonal entries of both X' and $\mathcal{A}_{\mathcal{F}(V_\delta)}(X')$ are all ones. Thus, we also have $L_+(G)\bullet\mathcal{A}_{\mathcal{F}(V_\delta)}(X') = L_+(G')\bullet X'$. It follows that

$$\begin{aligned}
\mathrm{sdp}_{\mathcal{M}}(G) &\leqslant L_+(G)\bullet\mathcal{A}_{\mathcal{F}(V_\delta)}(X') \qquad (\text{using } \mathcal{A}_{\mathcal{F}(V_\delta)}(X')\in\mathcal{M})\\
&= L_+(G')\bullet X' \qquad (\text{using } (9))\\
&= \mathrm{sdp}_{\mathcal{M}}(G')\,.
\end{aligned}\tag{10}$$

We can now finish the proof of the lemma,

$$\begin{aligned}
\mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(G^3[V_\delta]) &\geqslant \mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(G') - \mathbb{E}\|L_{G^3[V_\delta]} - L_{G'}\|_{\mathrm{TV}}\\
&\geqslant \mathrm{sdp}_{\mathcal{M}}(G) - \varepsilon \qquad (\text{using (10) and Lemma 5.7})\,.
\end{aligned}$$

■

5.3 Putting things together

By combining the lemmas in this section, we can prove Theorem 5.2.

Proof of Theorem 5.2. We need to show that

$$\tfrac{1}{3}\mathrm{sdp}_{\mathcal{M}}(G) - O(\varepsilon) \leqslant \mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(G[V_\delta]) \leqslant \mathrm{sdp}_{\mathcal{M}}(G) + O(\varepsilon)\,.$$

The upper bound on $\mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(G[V_\delta])$ is easy to show. We consider an optimal solution $X\in\mathcal{M}$ for G, so that $L_+(G)\bullet X = \mathrm{sdp}_{\mathcal{M}}(G)$. Then, we need to show that the value of X is preserved for $G[V_\delta]$ in expectation, i.e.,

$$\mathbb{E}\,L_+(G[V_\delta])\bullet X \leqslant L_+(G)\bullet X + \varepsilon\,.$$

This follows from Lemma A.1. We leave the details to the reader.

We combine the lemmas in this section to prove the lower bound. Notice that by Lemma 5.6, we can choose $H = G^3$ (for $c = 1/3$) in Lemma 5.3. With this choice of H, we can finish the proof of the theorem,

$$\begin{aligned}\mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(G[V_\delta]) &\geqslant \tfrac{1}{3}\cdot \mathbb{E}\,\mathrm{sdp}_{\mathcal{M}}(G^3[V_\delta]) - \varepsilon \qquad \text{(using Lemma 5.3)}\\ &\geqslant \tfrac{1}{3}\cdot \mathrm{sdp}_{\mathcal{M}}(G) - \tfrac{4}{3}\varepsilon \qquad \text{(using Lemma 5.8)}\,.\end{aligned}$$

■

6 Reduction to Quadratic Program over the ℓ_∞-Ball

Previously, we reduced the problem of arguing about the value of some strong SDP relaxation to a much simpler quadratic form. The key ingredient in this step is the rounding algorithm by Charikar and Wirth [CW04]. The details are given next.

We consider the problem to optimize the quadratic form given by $A \in \mathbb{R}^{n\times n}$ over the ℓ_∞-ball $B_\infty^n = [-1,1]^n$,

$$q(A) \overset{\mathrm{def}}{=} \max_{x_1,\ldots,x_n\in[-1,1]} \sum_{ij} a_{ij}x_ix_j = \max_{x\in B_\infty^n} \langle x, Ax\rangle\,. \tag{11}$$

A natural semidefinite relaxation replaces every x_i by a vector of norm at most 1,

$$\tilde{q}(A) \overset{\mathrm{def}}{=} \max_{\substack{v_1,\ldots,v_n\in\mathbb{R}^d\\ \|v_i\|^2\leqslant 1}} \sum_{ij} a_{ij}\langle v_i, v_j\rangle = \max_{X\succeq 0,\forall i\colon X_{ii}\leqslant 1} A\bullet X\,. \tag{12}$$

It is clear that $\tilde{q}(A) \geqslant q(A)$. The next lemma shows that $\tilde{q}(A)$ cannot be much larger than $q(A)$. The lemma essentially appears in the paper of Charikar and Wirth [CW04]. We will include a proof for completeness.

Lemma 6.1. *Let $A \in \mathbb{R}^{n\times n}$. Suppose $\tilde{q}(A) = \varepsilon\|A\|_{\mathrm{TV}} > 0$. Then,*

$$q(A) \geqslant \Omega\left(\tfrac{\varepsilon}{\log(1/\varepsilon)}\right)\cdot \|A\|_{\mathrm{TV}}\,.$$

Proof. We rescale A such that $\|A\|_{\mathrm{TV}} = 1$. Let $v_1,\ldots,v_n$ be vectors that achieve the maximum in (12). Let $\xi_1,\ldots,\xi_n$ be jointly distributed (symmetric) Gaussian variables such that $\mathbb{E}\,\xi_i\xi_j = \langle v_i, v_j\rangle/t$. Here $t \geqslant 1$ is a parameter that we choose later. Let x_i be the truncation of ξ_i to the interval $[-1,1]$, that is,

$$x_i = \begin{cases} 1 & \text{if } \xi_i > 1,\\ \xi_i & \text{if } \xi_i\in[-1,1],\\ -1 & \text{if } \xi_i < -1.\end{cases}$$

By construction, the variance of ξ is at most $1/t$. Hence, the event $x_i = \xi_i$ has probability at least $1-2^{-\Omega(t)}$ (using standard tail bounds for Gaussian variables). Similarly, the covariances of the ξ_i and x_i variables match up to an error of $2^{-\Omega(t)}$. Specifically, for all $i,j\in[n]$,

$$\begin{aligned}\mathbb{E}\,\xi_i\xi_j &= \mathbb{E}\,x_ix_j + \mathbb{E}(\xi_i - x_i)x_j + \mathbb{E}\,\xi_i(\xi_j - x_j)\\ &= \mathbb{E}\,x_ix_j \pm O\left(\mathbb{E}(\xi_i - x_i)^2\right)^{1/2} \qquad \text{(by Cauchy–Schwartz and } \mathbb{E}\,x_i^2\leqslant 1,\ \mathbb{E}\,\xi_i^2\leqslant 1)\\ &= \mathbb{E}\,x_ix_j \pm 2^{-\Omega(t)} \qquad \text{(by Fact A.3)}\end{aligned}$$

Now we can relate ε to the expected value of the quadratic form A evaluated at the random vector x,

$$\varepsilon = \sum_{ij} a_{ij}\langle v_i, v_j\rangle = t\,\mathbb{E}\sum_{ij} a_{ij}\xi_i\xi_j = t\,\mathbb{E}\sum_{ij} a_{ij}x_ix_j \quad \pm t2^{-\Omega(t)}\,.$$

Hence, $q(M) \geqslant \mathbb{E}\langle x, Ax\rangle \geqslant \varepsilon/t - 2^{-\Omega(t)}$. If we choose t logarithmic in $1/\varepsilon$, we get the desired lower bound on $q(M)$. ■

7 Subsampling theorem for quadratic forms

We now turn to our main technical tool for the proxy graph theorem—a subsampling theorem for quadratic forms. Our theorem will apply to any $n \times n$ matrix $B = (b_{ij})$ which is β-*dense* in the sense that it is the Laplacian of a β-dense graph (i.e., a graph with minimum degree at least $1/\beta$ whose adjacency matrix is normalized to have every row sum to 1). We will actually need a need a slight relaxation of this notion, and hence say B is (c, β)-*dense* if:

1. for all $i \in [n]$, $|b_{ii}| \leqslant c$, and $\sum_{j\neq i} |b_{ij}| \in [1/c, c]$,
2. for all $i, j \in [n], i \neq j$, $|b_{ij}| \leqslant \beta$.

Matrices of this kind are similar to the Laplacian of a regular graph G of total edge weight n. This is a factor n larger than our usual normalization. It will later in the proof be convenient to work with matrices whose rows sum up to 1 rather than $1/n$.

Theorem 7.1. *Let B denote a (c, β)-dense symmetric $n \times n$ matrix for some $\beta > 0$ and constant $c > 0$. Pick $U \subseteq [n]$ at random with $|U| = \delta n$, where $\delta \geqslant C\varepsilon^{-8}\log(1/\varepsilon)^2\beta$ for sufficiently large constant C. Then,*

$$\left|\mathbb{E}\left[\max_{x\in B_\infty^n} \langle x, B_U x\rangle\right] - \max_{x\in B_\infty^n} \langle x, Bx\rangle\right| \leqslant \varepsilon n\,. \tag{13}$$

The theorem will follow from the next two main lemmas. The first lemma shows that the subsampling step is random enough to give a concentration bound for large subsets of B_∞^n.

Lemma 7.2 (Concentration)**.** *Let $X \subseteq B_\infty^n$ of size $|X| \leqslant \exp(\varepsilon^2\delta n/20)$. Then,*

$$\mathbb{E}\left[\max_{x\in X}\langle x, B_U x\rangle\right] \leqslant \max_{x\in X}\langle x, Bx\rangle + \varepsilon n \tag{14}$$

The second main lemma shows that the maximum of the subsample has some structure in the sense that it is characterized by few vectors in B_∞^n.

Lemma 7.3 (Structure)**.** *There is a set $X \subseteq B_\infty^n$ of size $|X| \leqslant \exp(\varepsilon^2\delta n/40)$ such that*

$$\mathbb{E}\left[\max_{x\in B_\infty^n} \langle x, B_U x\rangle\right] \leqslant \mathbb{E}\left[\max_{x\in X}\langle x, B_U x\rangle\right] + \varepsilon n\,. \tag{15}$$

Proof of Theorem 7.1. In one direction,

$$\begin{aligned}
\mathbb{E}\left[\max_{x\in B_\infty^n}\langle x, B_U x\rangle\right] &\leqslant \mathbb{E}\left[\max_{x\in X}\langle x, B_U x\rangle\right] + \varepsilon n && \text{(by Lemma 7.3)}\\
&\leqslant \mathbb{E}\left[\max_{x\in X}\langle x, B x\rangle\right] + 2\varepsilon n && \text{(by Lemma 7.2)}\\
&\leqslant \mathbb{E}\left[\max_{x\in B_\infty^n}\langle x, B x\rangle\right] + 2\varepsilon n\,.
\end{aligned}$$

In the other direction, notice that

$$\max_{x\in B_\infty^n}\langle x, Bx\rangle \leqslant \mathbb{E}\left[\max_{x\in B_\infty^n}\langle x, B_U x\rangle\right] + \varepsilon n\,,$$

simply by considering a maximizing assignment $x \in B_\infty^n$ to the LHS and arguing that $\langle x, B_U x\rangle$ is concentrated around its expectation $\langle x, Bx\rangle$. (More formally, this will be done in the proof of Lemma 7.2.) ■

Remark 7.4. Feige and Schechtman showed a similar theorem for the max-cut value in graphs when subsampled to logarithmic degree. Their proof as well as ours are modifications of the original property testing bound of Goldreich, Goldwasser and Ron [GGR98]. There are a few differences between their setting that might be worth pointing out. First, our maximization problem is defined over the (continuous) ℓ_∞-ball rather than the Boolean hypercube. This results in some discretization issues. Secondly, we consider matrices that could be different from the Laplacian of a graph. In particular, we allow negative diagonal entries. This is the reason the maximum in our case need not be attained at a vertex. Finally, even in the case of maximizing the Laplacian of a graph over cut vectors, we are interested in constant degree graphs where several parameters are no longer concentrated (such as the maximum degree of the graph).

7.1 Proof of Concentration Lemma

The proof starts as follows: We fix a vector $x \in B_\infty^n$ and sample sets $U, V \subseteq [n]$ independently at random each of size $|U| = |V| = \delta n$. We then argue about concentration of $\langle x, A_{U\times V} x\rangle$. The argument is broken up into two steps.

Remark 7.5. To ease notation we will assume that B is (c,β)-dense for $c = 1$. Going from $c = 1$ to a larger constant $c > 1$ is a straightforward modification of the proofs and affects our bounds only by constant factors.

Lemma 7.6. *Let $x \in [-1,1]^n$ and $\varepsilon > 0$. Sample $U \subseteq [n]$ uniformly at random of size $|U| = \delta n$. Then,*

$$\Pr\big(\langle x, A_{U\times[n]}x\rangle > \delta\langle x, Ax\rangle + \varepsilon\delta n\big) \leqslant \exp(-\varepsilon^2\delta n/4). \tag{16}$$

Proof. The proof follows from Azuma's inequality (see Lemma A.4). We define the mapping $f(U) = \sum_{i\in U, j\in[n]} a_{ij}x_i x_j$, where we think of U as a tuple $(i_1,\ldots,i_{\delta n})$ each coordinate being an index in $[n]$. By the normalization of the rows of A, the function has Lipschitz constant at most 2 in the sense that replacing any coordinate i by a $i' \in [n]$ can change the function value by at most 2. Hence, Azuma's inequality implies that $\Pr(f > \mathbb{E} f + t) \leqslant \exp(-t^2/4\delta n)$. Setting $t = \varepsilon\delta n$ thus finishes the proof. ■

Let $U \subseteq [n]$ be a set of cardinality δn. Denote by $A^{\mathrm{trunc}}_{U\times[n]}$ the matrix in which we replace a_{ij} by 0 whenever $\sum_{i'\in U, i'<i} |a_{ij}| > 2\delta$. In other words, we truncate the columns of the matrix $A_{U,[n]}$ at weight 2δ. With some abuse of notation let $A^{\mathrm{trunc}}_{U\times V}$ denote the matrix $A^{\mathrm{trunc}}_{U\times[n]}$ in which we restricted the columns to the set V.

In the following lemma we do need the fact that $\delta > \mathrm{poly}(1/\varepsilon)\beta$.

Lemma 7.7. *Pick $U, V \subseteq [n]$ at random with $|U| = |V| = \delta n$ with $\delta \geqslant C\varepsilon^{-2}\log(1/\varepsilon)\beta$. Then, we have*

$$\mathbb{E}\, \|A_{U\times V} - A^{\mathrm{trunc}}_{U\times V}\|_{\mathrm{TV}} \leqslant \varepsilon\delta^2 n\,. \tag{17}$$

Proof. In order to prove the first claim, we will prove that for every column $j \in [n]$ the expected truncation in column j is bounded by $\varepsilon\delta$. The claim will then follow by linearity of expectation and the fact that $|V| = \delta n$. (We will actually derive a somewhat stronger bound in terms of ε.)

So let us fix a column $j \in [n]$ and consider the random variable $X = \sum_{i\in U} |a_{ij}|$. We can express X as a sum of independent variables $X = \sum_{i=1}^{|U|} X_i$, , where X_i are identically distributed so that X_i assumes each value $|a_{kj}|$ with probability $1/n$. We note that $\mathbb{E}\, X_i = \frac{1}{n}\sum_i |a_{ij}| = \frac{1}{n}$. Let us compute the fourth moment of $X - \mathbb{E}\, X$. First observe that $\mathbb{E}(X_i - \mathbb{E}\, X_i)^4 \leqslant O(\mathbb{E}\,|X_i|^4)$ and

$$\mathbb{E}\,|X_i|^4 \leqslant \frac{1}{n}\sum_{i=1}^{|U|} |a_{ij}|^4 \leqslant \frac{1}{n}\beta^3\,,$$

where we used the fact that $\sum_i |a_{ij}|^4$ is maximized when there are exactly $1/\beta$ terms each of magnitude β. In this case $\sum_i |a_{ij}| \leqslant \beta^3$. Similarly,

$$\mathbb{E}(X_i - \mathbb{E}\, X_i)^2 (X_k - \mathbb{E}\, X_k)^2 \leqslant \frac{O(1)}{n^2}\sum_{i=1}^{|U|} a_{ij}^2 a_{kj}^2 \leqslant \frac{O(\beta^2)}{n^2}\,.$$

By independence and the fact that $\mathbb{E}(X_i - \mathbb{E}\, X_i) = 0$, we therefore have

$$\begin{aligned}
\mathbb{E}(X - \mathbb{E}\, X)^4 &= \sum_i \mathbb{E}(X_i - \mathbb{E}\, X_i)^4 + \sum_{i,k} \mathbb{E}(X_i - \mathbb{E}\, X_i)^2\, \mathbb{E}(X_k - \mathbb{E}\, X_k)^2 \\
&\leqslant \delta n \cdot \frac{O(\beta^3)}{n} + (\delta n)^2 \cdot \frac{O(\beta^2)}{n^2} \\
&= O(\delta\beta^3)\,. && \text{(using } \beta < \delta)
\end{aligned}$$

Thus, by Markov's inequality,

$$\mathbb{P}\mathrm{r}(|X - \mathbb{E}\, X| > t) \leqslant \frac{E(X - \mathbb{E}\, X)^4}{t^4} \leqslant \frac{O(1)\delta\beta^3}{t^4} \leqslant \frac{O(1)\delta^4\varepsilon^8}{t^4}\,, \tag{18}$$

since $\beta \leqslant \varepsilon^2\delta$.

Notice that if $X - \mathbb{E}\, X \leqslant t\delta$, then we truncate at most weight $t\delta$. Therefore we can bound the expected truncation by integrating (18) over $t \geqslant 1$,

$$\int_{t\geqslant 1} t\delta\, \mathbb{P}\mathrm{r}(X - \mathbb{E}\, X > t\delta)\mathrm{d}t \leqslant \int_{t\geqslant 1} t\delta \frac{O(1)\varepsilon^8\delta^3}{t^3\delta^3}\mathrm{d}t \leqslant O(\varepsilon^8\delta)\int_{t\geqslant 1}\frac{1}{t^2}\mathrm{d}t \leqslant O(\varepsilon^8\delta)\,. \tag{19}$$

The claim follows. ■

Lemma 7.8. *Let $x \in [-1,1]^n$ and $\varepsilon > 0$. Let $U \subseteq [n]$ be of a fixed set of cardinality δn. Sample $V \subseteq [n]$ uniformly at random where $|V| = \delta n$. Then,*

$$\Pr\left(\langle x, A^{\text{trunc}}_{U,V} x\rangle > \delta\langle x, A^{\text{trunc}}_{U,[n]} x\rangle + \varepsilon\delta n\right) \leqslant \exp(-\varepsilon^2\delta n/16). \tag{20}$$

Proof. The proof is analogous to that of Lemma 7.6. We consider $f(V) = \sum_{i\in U, j\in V} a_{ij}x_i x_j$ and observe that by the normalization of $A^{\text{trunc}}_{U,[n]}$ (every column has weight at most 2δ) the function has Lipschitz constant at most 4δ. Hence, by Azuma's inequality, $\Pr(f > \mathbb{E}[f] + t) \leqslant \exp(-t^2/16\delta^3 n)$. Setting $t = \varepsilon\delta^2 n$ finishes the proof. ■

We also need a concentration bound for diagonal matrices given in the following.

Lemma 7.9. *Let $x \in B^n_\infty$ and suppose $D = \operatorname{diag}(d_1, \dots, d_n)$. Then,*

$$\Pr(|\delta^{-1}\langle x, D_U x\rangle - \langle x, Dx\rangle| \geqslant \varepsilon n) \leqslant 2\exp(-\varepsilon^2\delta n/4)\,. \tag{21}$$

Proof. The function $f(U) = \delta^{-1} x D_U x = \delta^{-1}\sum_{i\in U} d_i x_i^2$ has Lipschitz constant $2\delta^{-1}$. The claim then follows from Azuma's inequality. ■

The next corollary summarizes what we have shown so far.

Corollary 7.10. *Let $X \subseteq B^n_\infty$ of size $2^{-\varepsilon\delta n/20}$. Pick $U, V \subseteq [n]$ at random with $|U| = |V| = \delta n$ where $\delta > \varepsilon^{-2}\beta$. Then,*

$$\mathbb{E}\left[\max_{x\in X}\langle x, \delta^{-2}A_{U,V}x\rangle\right] \leqslant \max_{x\in X}\langle x, Ax\rangle + \varepsilon n. \tag{22}$$

Proof. Combining Lemma 7.6 with Lemma 7.8 and the union bound over $x \in X$, we obtain that

$$\mathbb{E}\left[\max_{x\in X}\langle x, A^{\text{trunc}}_{U,V}x\rangle\right] \leqslant \max_{x\in X}\mathbb{E}\left[\langle x, A^{\text{trunc}}_{U,V}x\rangle\right] + \varepsilon\delta^2 n. \tag{23}$$

On the other hand, by Lemma 7.7,

$$\mathbb{E}\left[\|A_{U,V} - A^{\text{trunc}}_{U,V}\|\right] \leqslant \varepsilon\delta^2 n$$

which immediately implies that $|\langle x, A_{U,V}x\rangle - \langle x, A^{\text{trunc}}_{U,V}x\rangle| \leqslant \varepsilon\delta^2 n$ for all $x \in B^n_\infty$. In particular together with Equation 23,

$$\mathbb{E}\left[\max_{x\in X}\langle x, A_{U,V}x\rangle\right] \leqslant \max_{x\in X}\mathbb{E}\left[\langle x, A_{U,V}x\rangle\right] + O(\varepsilon\delta^2 n)\,.$$

Noting that $\mathbb{E}\langle x, A_{U,V}x\rangle = \delta^{-2}\langle x, Ax\rangle$ and scaling up the previous bound by δ^{-2} finishes the claim. ■

We are now ready to prove the first main lemma.

Proof of Lemma 7.2. First of all, we can get rid of the matrix D by noting that

$$\mathbb{E}\left|\max_{x\in X}\delta^{-1}\langle x, D_{U\times U}x\rangle - \langle x, Dx\rangle\right| \leqslant \varepsilon n \tag{24}$$

by Lemma 7.9.

It remains to show concentration for $\max_{x \in X} A_{U \times U} x$. The proof is by reduction to the case where we sampled U, V independently of each other. Partition $[n] = S_1 \cup S_2 \cup \cdots \cup S_k$ such that $\|A_{ij}\|_{\mathrm{TV}} \leqslant \varepsilon^2 n$ for all $r \in [k]$. Notice that $k \leqslant O(1/\varepsilon)$ is sufficient. Furthermore, let $U_r = U \cap S_r$. One can show that with probability $1 - o(1)$, for all $r \in [k]$ we have $|U_r| \in [\frac{1}{2k}|U|, \frac{2}{k}|U|]$.

Notice that

$$\sum_{r \in [k]} \|A_{U_r \times U_r}\|_{\mathrm{TV}} \leqslant \varepsilon \|A_{U \times U}\|_{\mathrm{TV}}$$

and hence the LHS negligible in the sense that for any $x \in X$, the contribution to $\langle x, A_{U,U} x\rangle$ coming from diagonal matrices is at most $O(\varepsilon \delta^2 n)$.

On the other hand, whenever $i > j$, we are in the case where we sample sets U_i and U_j independently of each other. Thus we can apply and can apply Corollary 7.10 with error parameter $\varepsilon' = \varepsilon^2$ (forcing us to take $\delta/k > \varepsilon^{-4} \log(1/\varepsilon)^2$). This allows us to sum the expected error over all $i > j$ and thus conclude the theorem. ■

7.2 Proof of Structure Lemma

In this section we will restrict our attention to coordinates in U. Coordinates in $[n] \backslash U$ will play no important role in this part of the argument. Since $|U| = \delta n$, we will think of the entries $(a_{ij})_{j \in U}$ as rescaled by a factor δ^{-1}. Hence, $|a_{ij}| \leqslant \beta/\delta$ (rather than β).

Lemma 7.11. *Fix $x \in B_\infty^n$ and $i \in U$. Suppose $S \subseteq U$ is chosen uniformly at random such that $|S| = \alpha |U|$. Then,*

$$\mathop{\mathbb{E}}_{S} \left| \frac{1}{\alpha} \sum_{j \in S} a_{ij} x_j - \sum_{j \in U} a_{ij} x_j \right| \leqslant \frac{1}{\alpha} \sqrt{\sum_{j \in U} |a_{ij}|^2} \,. \tag{25}$$

Proof. We can write $\frac{1}{\alpha} \sum_{j \in S} a_{ij} x_j$ as a sum $X = \sum_{j \in U} X_j$ so that the X_j's are independent random variables satisfying $|X_j| \leqslant \frac{1}{\alpha}|a_{ij}|$. Moreover, $\mu = \mathbb{E}[X] = \sum_{j \in U} a_{ij} x_j$. We will use the fact $\mathbb{E}|X - \mu| \leqslant \sqrt{\mathbb{E}(X-\mu)^2}$. By independence, $\mathbb{E}(X-\mu)^2 = \sum_{j \in U} \mathbb{E}(X_j - \mu_j)^2$. Notice that $|X_j - \mu_j| \leqslant |X_j| = \frac{1}{\alpha}|a_{ij}|$ and therefore $\mathbb{E}(X_j - \mu_j)^2 \leqslant \frac{1}{\alpha}|a_{ij}|$. Hence,

$$\mathbb{E}|X - \mu| \leqslant \sqrt{\mathbb{E}(X-\mu)^2} \leqslant \frac{1}{\alpha} \sqrt{\sum_{j \in U} |a_{ij}|^2}$$

which is what we claimed. ■

From here on, let

$$Z_i = \sum_{j \in U} |a_{ij}|^2.$$

We will bound the expectation of $\sqrt{Z_i}$ taken over the random choice of U.

Lemma 7.12. *For all $i \in U$,*

$$\mathop{\mathbb{E}}_{U} \sqrt{Z_i} \leqslant O\left(\sqrt{\frac{\beta}{\delta}}\right).$$

Proof. Notice that Z_i is a sum of independent with expectation $\mu = \delta \sum_{j\in[n]} |a_{ij}|^2$. It is easy to see that μ is maximized when there are exactly $1/\beta$ nonzero entries in A each of magnitude β/δ (recall our rescaling by δ^{-1}). Hence, $\mu \leqslant \frac{\delta}{\beta} \cdot \left(\frac{\beta}{\delta}\right)^2 = \frac{\beta}{\delta}$.

On the other hand, $\mathbb{E}\sqrt{X} \leqslant \sqrt{\mathbb{E}\, X}$ by the Cauchy-Schwarz inequality. Hence, $\mathbb{E}\sqrt{Z_i} \leqslant O(\beta/\delta)$. ■

Remark 7.13 (How to think about the parameters)**.** Consider the bound we obtained in (25), i.e., $\sqrt{Z_i}/\alpha$. This bound will be important for us throughout the remainder of the proof. The reader should think of α as a small polynomial in ε, say, ε^3. The exact choice will be clear later. However, at this point it is important to note that we have the freedom to choose $\beta/\delta \ll \alpha^2$ thus making the bound nontrivial by choosing δ some $\mathrm{poly}(1/\varepsilon)$ factor larger than β. This will later determine the sample complexity of our theorem. Let us therefore denote

$$p = \beta/\delta$$

and think of it as some polynomial in ε.

Greedy Assignments. In what follows we consider sets $S \subseteq U$ and $T \subseteq U$ where S is chosen uniformly at random from $U \backslash T$ of size $|S| = \alpha|U|$. We will describe a natural way of extending an assignment for S, i.e., $x \in [-1,1]^S$ to an assignment $\tilde{x} \in [-1,1]^{S\cup T}$ so as to maximize a specific quadratic form we are interested in. For this purpose we define the *greedy assignment* $\tilde{x} \in [-1,1]^T$ (with respect to the coordinates in S) by putting

$$\tilde{x}_i = \arg \max_{y\in[-1,1]} \frac{1}{\alpha} \sum_{j\in S} a_{ij} x_j y - d_i y^2 \,, \tag{26}$$

for all $i \in T$ and $\tilde{x}_i = x_i$ for all $i \notin T$.

Let us point out the following simple observation showing that a small change in the linear term of Equation 26 does not change the maximum by much:

$$|a - a'| \leqslant \varepsilon \quad \Longrightarrow \quad \left| \max_{y\in[-1,1]} ay - dy^2 - \max_{y\in[-1,1]} a'y - dy^2 \right| \leqslant O(\varepsilon)\,. \tag{27}$$

In particular, the maximizer of $a'y - dy^2$ has the same value under $ay - dy^2$ as the maximizer of $ay - dy^2$ up to additive $O(\varepsilon)$ (for any fixed y the difference is at most $(a' - a)y$ and use this for both maxima). The next lemma is a simple consequence of this observation.

Lemma 7.14. *Let $x \in [-1,1]^U$ and let $\tilde{x} \in [-1,1]^T$ denote the greedy assignment obtained from the coordinates in S via (26). Then,*

$$\mathop{\mathbb{E}}_{S\subseteq U\backslash T} \left| \max_{y\in[-1,1]^T} \sum_{i\in T} \sum_{j\in U\backslash T} a_{ij} y_i x_j - d_i y_i^2 - \left(\sum_{i\in T} \sum_{j\in U\backslash T} a_{ij} \tilde{x}_i x_j - d_i \tilde{x}_i^2 \right) \right| \leqslant \sum_{i\in U_l} O\left(\frac{1}{\alpha}\sqrt{Z_i}\right)\,. \tag{28}$$

Proof. From Lemma 7.11 (applied to S and $U\backslash T$) it follows that for all $i \in T$,

$$\mathbb{E} \left| \sum_{j\in U\backslash T} a_{ij} x_j - \sum_{j\in S} a_{ij} x_j \right| \leqslant O\left(\frac{1}{\alpha}\sqrt{Z_i}\right)\,.$$

Hence, the claim follows from the remark after Equation 27. ■

Later, we may also put $T = U$ in which case we extend an assignment $x \in [-1,1]^S$ to all coordinates in U, i.e, $\tilde{x} \in [-1,1]^U$. In this case, we let $\phi(S)$ be the set of all greedy assignments which are obtained as follows: partition U and S into r parts $U_1 \cup \cdots \cup U_r$ and $S_1 \cup \cdots \cup S_r$, then assign coordinates in U_k greedily according to coordinates in S_k.

Since our error bound above only works for coordinates in $U \backslash T$, we cannot use the previous lemma directly to argue about $\phi(S)$. Nevertheless the proof of the next lemma is essentially a reduction to Lemma 7.11 using a partitioning argument similar to the one we saw in the proof of the randomness lemma.

Lemma 7.15. *For every positive integer $r > 0$, we have*

$$\max_{x \in B_\infty^n} \langle x, B_U x\rangle \leqslant \mathbb{E}_S \left[\max_{x \in \phi(S)} \langle x, B_U x\rangle\right] + \frac{1}{\delta}\sum_{i \in U} O\left(\frac{r}{\alpha}\sqrt{Z_i}\right) + \frac{1}{r}\|B_U\|_{\mathrm{tv}}\,. \tag{29}$$

Proof. Let $x \in [-1,1]^n$ be any vector that maximizes $\langle x, A_T x\rangle$. We will exhibit an assignment $\tilde{x} \in \phi(S)$ which demonstrates the claim. To this end we randomly partition U into r pairwise disjoint sets $U = U_1 \cup \cdots \cup U_r$ of equal size. That is, $|U_l| = \frac{1}{r}|U|$. We will also pick S_l uniformly at random from $U \backslash U_l$ of size $|S_l| = \frac{\alpha}{r}|U|$. Let us define a sequence of vectors $x^0 = x, x^1, \ldots, x^r$ and show that

$$\mathbb{E}\langle x^l, B_U x^l\rangle \geqslant \mathbb{E}\langle x^{l-1}, B_U x^{l-1}\rangle - \frac{1}{\delta}\sum_{i \in U_l} O\left(\frac{r}{\alpha}\sqrt{Z_i}\right) - \frac{1}{r^2}\|B_U\|_{\mathrm{tv}}\,, \tag{30}$$

for all $0 < l \leqslant r$. Inductively, let x^l be equal to x^{l-1} in all coordinates except U_l. The coordinates U_l are induced from S_l via equation (26), i.e.,

$$\forall i \in U_l\colon\ x_i^l = \arg\max_{y \in [-1,1]} \frac{1}{\alpha}\sum_{j \in S_l} a_{ij} x_j^{l-1} y - d_i y^2 \tag{31}$$

We observe that $x^r \in \phi(S)$ as desired.

The plan is to apply, at each step l, Lemma 7.14 to the sets S_l, U_l and U.

In order to compare $x^l A_U x^l$ with $x^{l-1} A_U x^{l-1}$, notice that the critical terms are indexed by $i \in U_l, j \in U$ (and by symmetry $i \in U, j \in U_l$). The remaining terms where neither $i \in U_l$ or $j \in U_l$ have not changed. Note that for the moment we ignore all entries a_{ij} of the matrix where both $i \in U_l$, $j \in U_l$ and $i \neq j$ for some $l \in [r]$. Later we will bound the contribution of such terms.

By Lemma 7.14, in expectation taken over S_l,

$$\begin{aligned}
\sum_{i \in U_l}\sum_{j \in U \backslash U_l} a_{ij} x_i^l x_j^l - \sum_{i \in U_l} d_i (x_i^l)^2 &= \sum_{i \in U_l}\sum_{j \in U \backslash U_l} a_{ij} x_i^l x_j^{l-1} - \sum_{i \in U_l} d_i (x_i^l)^2 \\
&\geqslant \max_{x \in [-1,1]^{U_l}} \sum_{i \in U_l}\sum_{j \in U \backslash U_l} a_{ij} x_i x_j^{l-1} - \sum_{i \in U_l} d_i x_i^2 - \sum_{i \in U_l} O\left(\frac{r}{\alpha}\sqrt{Z_i}\right) \\
&\geqslant \sum_{i \in U_l}\sum_{j \in U \backslash U_l} a_{ij} x_i^{l-1} x_j^{l-1} - \sum_{i \in U_l} d_i (x_i^{l-1})^2 - \sum_{i \in U_l} O\left(\frac{r}{\alpha}\sqrt{Z_i}\right)\,.
\end{aligned}$$

On the other hand,

$$\sum_{i,j \in U_l} |a_{ij}| \leqslant \frac{1}{r^2}\sum_{i,j \in U} |a_{ij}| \leqslant \frac{1}{r^2}\delta\|B_U\|_{\mathrm{tv}}\,.$$

Recall $B_U = \delta^{-2} A_{U\times U} - \delta^{-1} dI$, but take into account that we already rescaled a_{ij} by δ^{-1} at the beginning of the section. Therefore we only need to rescale a_{ij} and d_i both by δ^{-1} and we get

$$\mathbb{E}\langle x^l, B_U x^l\rangle \geqslant \mathbb{E}\langle x^{l-1}, B_U x^{l-1}\rangle - \frac{1}{\delta}\sum_{i\in U_l} O\left(\frac{r}{\alpha}\sqrt{Z_i}\right) - \frac{1}{r^2}\|B_U\|_{\text{tv}}\,,$$

and thus by iterating over $l \in [r]$,

$$\mathbb{E}\left[\max_{x\in\phi(S)}\langle x, B_U x\rangle\right] \geqslant \mathbb{E}\langle x^r, B_U x^r\rangle \geqslant \mathbb{E}\langle x^0, B_U x^0\rangle - \frac{1}{\delta}\sum_{i\in U} O\left(\frac{r}{\alpha}\sqrt{Z_i}\right) - \frac{1}{r}\|B_U\|_{\text{TV}}\,.$$

But x_0 does not depend on S and hence

$$\mathbb{E}\langle x^0, B_U x^0\rangle = \max_{x\in B_\infty^n}\langle x, B_U x\rangle$$

proving the claim. ■

Fixing the parameters. Let us fix some of the parameters in the previous lemma and bound the error in expectation taken over $U \subseteq [n]$. This is done next.

Lemma 7.16. *Let $\alpha > 0$. Fix $r = C/\varepsilon$ and $\sqrt{p} = \alpha\varepsilon^2/C$ for sufficiently large constant $C > 0$. Then,*

$$\mathbb{E}\left[\frac{1}{\delta}\sum_{i\in U} O\left(\frac{r}{\alpha}\sqrt{Z_i}\right) + \frac{1}{r}\|B_U\|_{\text{tv}}\right] \leqslant \varepsilon n\,.$$

Proof. Clearly, $\mathbb{E}\left[|U|\right] = \delta n$ and $\mathbb{E}\left[\sqrt{Z_i}\right] = O(\sqrt{p})$, by Lemma 7.12. Also $\mathbb{E}\left[\|B_U\|_{\text{TV}}\right] = O(n)$. Hence, we want

$$O\left(\frac{r\sqrt{p}}{\alpha} + \frac{1}{r}\right) \leqslant \varepsilon.$$

The bound follows by rearrangement. ■

Discretized greedy assignments. We consider a discretized variant of $\phi(S)$, defined as

$$\psi(S) = \{\tilde{x} \mid x \in ([-1,1] \cap \varepsilon^m \mathbb{Z})^S\}$$

for some constant m. Notice that $|\psi(S)| \leqslant \exp(O(|S|\log(1/\varepsilon)))$ where $|S| = \alpha\delta n$. We will thus set

$$\alpha = \varepsilon^2/C\log(1/\varepsilon)$$

for C large enough so that $\alpha\delta n\log(1/\varepsilon) \leqslant \varepsilon^2\delta n/40$. Given Lemma 7.16 this will result in

$$p \leqslant (\alpha\varepsilon^2/C)^2 = \varepsilon^8/C'\log(1/\varepsilon)^2$$

which gives the sample complexity promised in Theorem 7.1. This means we are done if we can show the following lemma.

Lemma 7.17.

$$\mathbb{E}\left[\max_{x\in B_\infty^n}\langle x, B_U x\rangle\right] \leqslant \mathbb{E}\left[\max_{x\in\psi(S)}\langle x, B_U x\rangle\right] + \varepsilon n. \tag{32}$$

Proof. Assume the same setup as in Lemma 7.15 and specifically consider the sequence of assignments $x_0, x_1, \ldots, x_r$ which was used to demonstrate the claim. We will define a new sequence $x_0, \bar{x}_0, x_1, \bar{x}_1, \ldots, \bar{x}_{r-1}, x_r, \bar{x}_r$. Here, $\bar{x}_l$ is obtained from x_l by rounding each coordinate in S to the nearest integer multiple of ε/r in the interval $[-1,-1]$ and inducing the others as previously (but from the new values).

Therefore we are done if we can argue that $\bar{x}^l B_U \bar{x}^l \geqslant x^l B_U x^l - \frac{\varepsilon}{r} n$. However, this is clear since at every step l,

$$\frac{1}{\varepsilon'} \sum_{j \in S_l} a_{ij} \bar{x}_j^{l-1} = \frac{1}{\varepsilon'} \sum_{j \in S_l} a_{ij} \left(x_j^{l-1} \pm \frac{\varepsilon}{r} \right) = \frac{1}{\varepsilon'} \sum_{j \in S_l} a_{ij} x_j^{l-1} \pm \frac{\varepsilon}{r\varepsilon'} \sum_{j \in S_l} |a_{ij}| \,.$$

By observation (27) this bound translates to an additional error of $\frac{\varepsilon}{r} \sum_{r \in S_l} |a_{ij}|$ at step l. Summing up over all l and i this amounts to a total error of $\varepsilon \sum_{i,j \in U} |a_{ij}| = O(\varepsilon \delta^2 n)$ in expectation. ∎

Choosing S independently of U. The reader may have noticed that so far we are choosing S randomly as a subset of U. Hence, the set $\psi(S)$ that we constructed above depends on the choice of U. To finish the proof of the Structure Lemma we need a single set $X \subseteq B_\infty^n$ that is independent of the choice of U. This is easy to accomplish from what we have. Simply pick S and U independently and consider $U' = U \cup S$. Since $|S| \leqslant \varepsilon^2 |U|$, the difference $\|B_U - B_{U'}\|_{\mathrm{TV}}$ is negligible. Therefore, we may exchange U' for U in Lemma 7.17 so that the choice of S and U is independent. Since the lemma is true in expectation taken over S and U, there must exist a fixed choice of S for which the lemma is true in expectation taken over U. But then we may take $X = \psi(S)$ in order to conclude the proof of the Structure Lemma.

Remark 7.18 (Simplifications in the bipartite case)**.** We briefly remark that our proof is somewhat simpler when considering quadratic forms $\max_{x,y \in B_\infty^n} \langle x, Ay \rangle$. In this case, we can get around the partition argument in the proof of Lemma 7.2. Likewise in the proof of Lemma 7.15 we no longer need a partition of size $r \approx 1/\varepsilon$ but instead we can find near optimal greedy assignments $\tilde{x}, \tilde{y}$ in two steps. These simplifications result in polynomial (in $1/\varepsilon$) improvements in the sample complexity of the theorem. The details are left to the reader.

7.3 Property testing PSD matrices

We will illustrate one application of the previous theorem.

Definition 7.19. We say that a matrix B is ε-far from positive semidefinite definite if there exists a vector $x \in B_\infty^n$ such that $\langle x, -Bx \rangle \geqslant \varepsilon$.

Recall that B is positive semidefinite if and only if $x(-B)x \leqslant 0$ for all $x \in B_\infty^n$. Notice we could have defined distance in terms of the operator norm which is to say that there exists an $x \in B_2^n$ such that $\langle x, -Bx \rangle \geqslant -\varepsilon$. However, since $B_2^n \subseteq B_\infty^n$ this would be a stronger notion of "ε-far" thus applying to fewer matrices.

Theorem 7.20. *Let B by a symmetric $(O(1), D/n)$-dense matrix. Then there is a property testing algorithm $\mathcal{A}$ such that:*

"No" case *If B is ε-far from being positive semidefinite, then $\mathcal{A}$ rejects B with probability greater than $2/3$.*

"Yes" case *If B is positive semidefinite, then $\mathcal{A}$ rejects B with probability less than* $1/3$.

Complexity *$\mathcal{A}$ reads only* $\mathrm{poly}(D, \varepsilon^{-1})$ *many entries of B and runs in time* $\mathrm{poly}(D, \varepsilon^{-1})$

Proof. Let $k = \mathrm{poly}(D, \varepsilon^{-1})$. The algorithm will sample a random $k \times k$ minor from B, call it B'. To get an efficient test, i.e., runtime $\mathrm{poly}(D, \varepsilon^{-1})$, we will use the rounding algorithm from Section 6. Specifically, we check if

$$\max_{X \succeq 0, \forall i\colon X_{ii} \leqslant 1} B' \bullet X \leqslant \varepsilon/2\,. \tag{33}$$

"Yes" case If B was positive semidefinite, then $\max_{x \in B_\infty^n} \langle x, -B, x \rangle \leqslant 0$ and hence by Theorem 7.1 we know that with high probability $\max_{x \in B_\infty^n} \langle x, -B'x \rangle \leqslant \varepsilon^2$ (for large enough k). But then, by Lemma 6.1, we know that the LHS in (33) is less than $O(\varepsilon^2 \log(1/\varepsilon)) \leqslant \varepsilon/2$. Hence, our algorithm passes.

"No" case Here we know that there is an $x \in B_\infty^n$ such that $\langle x, -Bx \rangle > \varepsilon$. It is not hard to show that with high probability $\langle x, -B'x \rangle > \varepsilon/2$. This follows, for instance, from Lemma A.1 (an application of Chebyshev's inequality). Since the semidefinite program in (33) is a relaxation of the quadratic form and can therefore only be larger, it follows that the test will fail.

■

8 Proxy graphs from graph powers

In this section, we go back to the construction of proxy graphs from graph powers as was need in the proof of our main theorem. Here, we will fill in the details that were omitted from the proofs of Lemma 5.6 Lemma 5.7.

Our proxy graphs will be a weighted graph power of the original graph. Graph powers are a natural candidates for proxy graphs because they satisfy the assumption of the proxy graph theorem as shown in the next lemma.

Lemma 8.1. *Let G be a normalized Δ-regular graph. Let $k > 0$ be an odd number. Then,*

$$\tfrac{1}{2} L_+(G) \preceq L_+(G^k) \preceq k L_+(G)\,, \tag{34}$$

where G^k is the (normalized) graph corresponding to the uniform distribution over walks of length k in G. (The same claim is true when replacing L_+ by L.)

Proof. For the purpose of this proof, we scale the graphs G and G^k so that every vertex has (weighted) degree 1. The eigenvalues of $L_+(G)$ are given by $1 + \lambda_i(A)$ for $i \in [n]$, whereas the eigenvalues of $L_+(G^k)$ are given by $1 + \lambda_i(A)^k$, where A is the adjacency matrix of G (a stochastic matrix by our scaling). Further, $L_+(G) = I + A$ and $L_+(G^3) = I + A^3$ have a common eigenbasis. Hence, it suffices to verify the bound in Equation (34) for each of the eigenvalues. More precisely, it suffices to show

$$\tfrac{1}{2} - \tfrac{1}{2}x \leqslant 1 - x^k \leqslant k(1 - x).$$

for $x \in [-1, 1]$, since $\lambda_i(A) \in [-1, 1]$.

To see why the first inequality holds, consider $x < 0$. In this case $2(1 - x^k) \geqslant 2$ (for odd k), but $1 - x \leqslant 2$. When $x \geqslant 0$, then $x \geqslant x^k$ and hence $1 - x \leqslant 1 - x^k$. The second inequality follows from the fact that

$$1 - x^k = (1 - x)(1 + x + x^2 + \cdots + x^{k-1}) \leqslant (1 - x)k\,. \qquad ■$$

8.1 Subsampling the third power

In the following let $G = (V, E)$ be a Δ-regular graph and $\delta \geqslant \mathrm{poly}(\varepsilon^{-1})\Delta^{-1}$. Further denote $W = V_\delta$. The following claim implies Lemma 5.7 which was needed in the proof of our main theorem.

Lemma 8.2. *Let D_1 denote the uniform distribution over edges in $G^3[W]$. Let D_2 denote the distribution obtained as follows:*

1. *Pick a random edge $(v, v') \in E$.*
2. *Choose uniformly at random $w \in N_W(v)$ and $w' \in N_W(v')$.*
3. *Output (w, w').*

Then,

$$\mathop{\mathbb{E}}_{W}\left[\mathrm{TV}(D_1, D_2)\right] \leqslant \varepsilon\,.$$

Here and in the following $\mathrm{TV}(D_1, D_2)$ denote the total variation distance between the two distributions D_1 and D_2.

Proof. Let us compare the following two distributions:

P_1 : Pick a uniformly random path $p = (w, v, v', w')$ from the set of all paths of length 3 in G which have $w, w' \in W$.

P_2 : Pick a random edge $v, v' \in E$ and random neighbors $w \in N_W(v), w' \in N_W(v')$ and consider the path (w, v, v', w').

Notice that it suffices to bound the statistical distance between P_1 and P_2. This is because D_1 is just the marginal distribution of P_1 on the endpoints of the path (w, w'). Likewise D_2 is the marginal distribution of P_2 on (w, w').

Now, let $p = (w, v, v', w')$ denote any path of length 3 in G so that $w, w' \in W$. Let N denote the number of such paths. Note that $\mathbb{E}\, N = \delta^2\Delta^3 n$. Let us now compare the probability of this path under the two distributions. For P_1 we get

$$P_1(p) = \frac{1}{N}\,.$$

On the other hand, under P_2,

$$P_2(p) = \frac{1}{|N_W(v)|}\cdot\frac{1}{\Delta n}\cdot\frac{1}{|N_W(v')|}.$$

Note that for every $v \in V$, we have $\mathbb{E}\,|N_W(v)| = \delta\Delta$. It now suffices to argue the bound

$$\mathbb{E}\left[\mathrm{TV}(P_1, P_2)\right] = \mathbb{E}\,\frac{1}{2}\sum_p\left|\frac{1}{N} - \frac{1}{|N_W(v)||N_W(v')|\Delta n}\right| \leqslant \varepsilon. \tag{35}$$

Let us call a path $p = (w, v, v', w')$ *good* if

$$\frac{1}{|N_W(v)|\cdot|N_W(v')|} = \frac{1 \pm \varepsilon'}{\delta^2\Delta^2}.$$

Later we will choose $\varepsilon' = \Omega(\varepsilon)$ to be sufficiently small, say, $\varepsilon' = \varepsilon/100$. We need the following simple concentration bounds.

Claim 8.3. *With probability $1-\varepsilon'$ over the choice of W, we have*

1. $N^{-1} = (1\pm\varepsilon')/(\delta^2\Delta^3 n)\,.$
2. *The fraction of bad paths is less than* $1/O(\varepsilon'^5(\delta\Delta)^3)\,.$

Proof. The first claim follows from Lemma A.1. Regarding the second claim, it is not hard to show for every v, v' that

$$\mathbb{P}\mathrm{r}\left\{\frac{1}{|N_W(v)||N_W(v')|} \notin \frac{1\pm\varepsilon'}{\delta^2\Delta^2}\right\} \leqslant \frac{1}{O(\varepsilon'^4(\delta\Delta)^3)}\,.$$

This can be shown by computing the fourth moment $\mathbb{E}(|N_W(v)| - \delta\Delta)^4$ and bounding the probability of a factor $1+\alpha$ deviation of $|N_W(v)|$ from its mean for small enough $\alpha = \Omega(\varepsilon')$. (The calculation is similar to the proof of Lemma 7.7 and is left to the reader.)

This argument shows that the expected number of bad paths is at most $1/O(\varepsilon'^4(\delta\Delta)^3)$ and the claim is completed by applying Markov's inequality. ■

Given this claim, we can finish the proof of the lemma. Indeed letting Q denote the set of good paths, we have with probability $1-\varepsilon'$,

$$\begin{aligned}
\sum_p \left|\frac{1}{N} - \frac{1}{|N_W(v)||N_W(v')|\Delta n}\right| &\leqslant \sum_{p\in Q} \frac{2\varepsilon'}{\delta^2\Delta^3 n} + \sum_{p\notin Q}\frac{1}{\Delta n}\\
&\leqslant 2\varepsilon' + \frac{N}{O(\varepsilon'^5(\delta\Delta)^3)}\cdot\frac{1}{\Delta n}\\
&= 2\varepsilon' + \frac{1}{O(\varepsilon'^5\delta\Delta)}\cdot\frac{N}{\delta^2\Delta^3 n}\\
&\leqslant O(\varepsilon')\,.
\end{aligned}$$

In the first inequality we used the fact that $|N_W(v)| \geqslant 1$ for any existing path and hence the term $1/|N_W(v)||N_W(v')|\Delta n$ is never larger than $1/\Delta n$. In the last step we used that we may choose $\delta\Delta \geqslant C\varepsilon'^{-5}$ for sufficiently large constant $C > 0$, and that $N \leqslant (1+\varepsilon')\delta^2\Delta^3 n$. Hence,

$$\mathbb{E}\,\mathrm{TV}(P_1, P_2) \leqslant (1-\varepsilon')O(\varepsilon') + \varepsilon' \leqslant \varepsilon\,.$$ ■

9 Feige-Schechtman graphs and triangle inequalities

In this section we discuss the application of our theorem to solving Max Cut in sphere graphs (aka Feige-Schechtman graphs). In the following we will argue directly about Max Cut rather than Min UnCut. Let us first recall some known facts about continuous sphere graphs and then look at dense discretizations of these graphs. We denote by G_γ the graph on the vertex set $V = \mathbb{S}^{d-1}$ with edge set

$$E = \{(u,v)\in V^2 \mid \tfrac{1}{4}\|u-v\|^2 \geqslant 1-\gamma\}\,. \tag{36}$$

The integral value of G_γ, denoted $\mathrm{opt}(G_\gamma)$, is defined as the maximum of $\mu(A,\bar{A}) \stackrel{\mathrm{def}}{=} \mu^2(\{(x,y)\in E\colon x\in A, y\notin A\})$ taken over all measurable subsets $A\subseteq \mathbb{S}^{d-1}$ Here, μ denotes the uniform surface measure of the sphere $\mathbb{S}^{d-1}$ and $\mu^2 = \mu\times\mu$. A theorem of Feige and Schechtman shows that the maximum is attained for any hemisphere.

Theorem 9.1 (Feige-Schechtman [FS02]). *Fix $\gamma \in [0,1]$ and consider the graph G_γ. Then, the maximum of $\mu(A, \bar{A})$ over all measurable subsets $A \subseteq \mathbb{S}^{d-1}$ is attained for any hemisphere $H \subseteq \mathbb{S}^{d-1}$.*

Recall, if A is a hemisphere, $\mu(A, \bar{A}) = 1 - \Theta(\sqrt{\gamma})$. Hence $\mathrm{opt}(G_\gamma) = 1 - \Theta(\sqrt{\gamma})$. At this point we mention that the SDP relaxation for Max Cut is well-defined on infinite graphs though we omit the formal details. In this case it is easiest to think of E as a distribution over edges so that the SDP maximizes the quantity $\mathbb{E}_{(u,v)\sim E}\,\frac{1}{4}\|f(u)-f(v)\|^2$ over all embeddings $f\colon V \to B$ satisfying the usual additional constraints. Here B can be taken to be the unit ball of the infinite dimensional Euclidean space.

The sphere graph itself can then be interpreted as an SDP solution, hence the following fact.

Fact 9.2 (Basic SDP value). $\mathrm{sdp}(G_\gamma) \geqslant 1 - \gamma$.

Proof. The graph itself gives an embedding (the identity embedding) such that for each edge $(u,v) \in E$, $\frac{1}{4}\|u - v\|^2 \geqslant 1 - \gamma$. Since the SDP averages this quantity over all edges in the graph, the claim follows. ■

We will show next that triangle inequalities change the value of the SDP from $1-\gamma$ to $1-\Omega(\sqrt{\gamma})$ thus capturing the integral value up to constant factors in front of γ.

Lemma 9.3 (SDP with triangle inequalities). $\mathrm{sdp}_3(G_\gamma) \leqslant 1 - \Omega(\sqrt{\gamma})$.

We believe this lemma is somewhat folklore, but we will give a proof for lack of a reference. The proof works as follows. First, triangle inequalities are known to imply the odd cycle constraints which means that an SDP with triangle inequalities on an odd cycle of length k has value at most (and, in fact, equal to) $1 - 1/k$.

Lemma 9.4. *Let C be an odd cycle of length k. Then,* $\mathrm{sdp}_3(C) \leqslant 1 - 1/k$.

Second, it follows that if a graph G can be covered uniformly by odd cycles of length k, then its sdp_3-value can be at most $1 - 1/k$.

Lemma 9.5. *Let $G = (V, E)$ be a (possibly infinite) graph. Suppose there exists a distribution $\mathcal{C}$ over odd cycles of length k for some fixed number k such that the marginal distribution on each edge of a random cycle from $\mathcal{C}$ has statistical distance ε to the uniform distribution over edges in G. Then,* $\mathrm{sdp}_3(G) \leqslant 1 - 1/k + \varepsilon$.

Proof. By our assumption we have that for every embedding $f\colon V \to B$,

$$\mathop{\mathbb{E}}_{(u,v)\sim E} \tfrac{1}{4}\|f(u) - f(v)\|^2 \leqslant \mathop{\mathbb{E}}_{C\sim\mathcal{C}} \mathop{\mathbb{E}}_{(u,v)\sim C} \tfrac{1}{4}\|f(u) - f(v)\|^2 + \varepsilon\,.$$

But we know, by Lemma 9.4, that for every $f\colon V \to B$, satisfying the triangle inequalities,

$$\mathop{\mathbb{E}}_{(u,v)\sim C} \frac{1}{4}\|f(u) - f(v)\|^2 \leqslant 1 - \tfrac{1}{k}\,.$$

Hence,

$$\mathop{\mathbb{E}}_{(u,v)\sim E} \frac{1}{4}\|f(u) - f(v)\|^2 \leqslant 1 - \tfrac{1}{k} + \varepsilon\,.$$

■

We will next see that the sphere graph can by uniformly covered by odd cycles of length $O(1/\sqrt{\gamma})$. We begin with the following simple observation.

Lemma 9.6. *For every $l \in [1-\gamma, 1-\gamma/2]$, there exists an odd cycle, denoted $C_l = (v_1, \ldots, v_k)$, in G_γ of length $k = O(\sqrt{\gamma})$ such that $\frac{1}{4}\|v_i - v_{i+1}\|^2 = l$ for all $i \in \{1, \ldots, k-1\}$.*

Proof sketch. Pick an arbitrary great circle around the sphere and place the vertices $v_1, \ldots, v_k$ equally spaced along this circle. For $k = O(\sqrt{\gamma})$ vertices, we can accomplish the Euclidean distance between two consecutive vertices is less than, say, $\sqrt{\gamma}/10$. Now connect each vertex v on the circle to the unique vertex w which maximized $\|v - w\|^2$. This creates an odd cycle and, by our previous observation, it follows that $\frac{1}{4}\|v - w\|^2 \geqslant 1 - \gamma$. Now we can make $\frac{1}{4}\|v - w\|^2 = l$ be walking along the cycle and moving vertices in a direction orthogonal to the plane defined by the circle until all edges have length l. ■

Lemma 9.7. *Let $\gamma > 0$ and let S^{d-1} be the sphere. There exists a distribution $\mathcal{C}$ over odd cycles $C = (v_1, \ldots, v_k)$ for some $k \leqslant \frac{10\pi}{\sqrt{\gamma}}$ such that for all i, the marginal distribution of (v_i, v_{i+1}) has statistical distance $o(1)$ to the uniform distribution over edges in G_γ (as $d \to \infty$).*

Proof. We will describe the distribution $\mathcal{C}$ as follows:

1. Pick a random edge $e = (u, v) \in E$ from G_γ.
2. Let $l = \frac{1}{4}\|u - v\|^2$. If $l \leqslant 1 - \gamma/2$, let $C_l = (v_1, v_2, \ldots, v_k)$ denote the odd cycle given by Lemma 9.6. If $l \geqslant 1 - \gamma/2$, declare "failure".
3. If the previous step succeeded, pick a random rotation R and output $RC = (Rv_1, Rv_2, \ldots, Rv_k)$.

We claim that if the second step succeeds, then indeed every marginal (Rv_i, Rv_{i+1}) is distributed like a uniformly random edge. This is (1) because (u, v) was chosen to be a uniformly random edge and (2) (Rv_i, Rv_{i+1}) is a random rotation of (u, v) and hence, by spherical symmetry, is equally likely to be any edge in E that has the same length as (u, v).

On the other hand, by measure concentration, with probability $1 - \exp(-\Omega(d))$, we have that $\frac{1}{4}\|u-v\|^2 \in [1-\gamma, 1-\gamma/2]$. This completes the claim since the probability of failure only introduces $o(1)$ statistical distance. ■

Using standard discretization arguments all previous lemmas can be transferred to a sufficiently dense discretization of the continuous sphere. Similarly, it is not difficult to show that sufficiently many random points from the sphere will give a good discretization.

Lemma 9.8. *Fix $\gamma \in [0, 1], d \in \mathbb{N}$. Then, there exists an $n_0(d, \gamma) \in \mathbb{N}$ so that if we pick $V \subseteq S^{d-1}$ uniformly at random with $|V| \geqslant n_0$, then the induced subgraph $G_\gamma[V]$ satisfies*

1. $\mathrm{opt}(G_\gamma[V]) = 1 - \Theta(\sqrt{\gamma})$,
2. $\mathrm{sdp}_3(G_\gamma[V]) = 1 - \Theta(\sqrt{\gamma})$.

Proof sketch. The first claim is shown in [FS02]. For the second claim, let us decompose $\mathbb{S}^{d-1}$ into equal volume cells of diameter at most ε. Here, ε is a parameter that we will later take to be very small, say, $\varepsilon \leqslant d/100$. Now pick enough vectors $V \subseteq \mathbb{S}^{d-1}$ uniformly at random such that with probability at least $1-\varepsilon$ every two cells have the same number of vectors up to a factor of $1 \pm \varepsilon$ in it.

We need to show that $\mathrm{sdp}_3(G_\gamma[V]) \leqslant 1 - \Omega(\sqrt{\gamma})$. To this end we first consider a related graph G', which has the same vertex set as G but different edges. A random edge in G' is defined by the following process: Pick first a random edge on the continuous sphere, then for each endpoint pick a random vertex in the equal volume cell containing the endpoint. Finally, normalize the edges such that the total edge weight is the same as in G.

We can use the distribution over odd cycles given by Lemma 9.7 in order to get a distribution for the graph G' as follows: Pick the cycle and map each point to a vertex in the corresponding cell. The resulting marginal distributions will be uniform in G'. Thus, $\mathrm{sdp}_3(G') = 1 - \Theta(\sqrt{\gamma})$.

Finally, we will show that $\mathbb{E}\|L(G') - L(G_\gamma[V])\|_{\mathrm{TV}}$ tends to zero with ε. That is, the two distributions have statistical distance tending to zero. This also shows that for sufficiently small ε, the semidefinite programs also have approximately the same value. Now to argue the above point, consider the process of picking a random edge. Consider first the case that in G', the two cells containing the chosen points have exactly the expected number of vectors in them, and furthermore, suppose that the two cells are *good* in the sense that either none of the vertices in them share an edge or all pairs of vertices between the two cells share an edge in G. In this case, the edges in G going between these two cells have exactly the same probability as under G'.

The first assumption is close enough to the truth, since the number of vertices in different cells differ by at most a factor of $1 \pm \varepsilon$, For the second assumption it suffices to pick ε small enough so that a cap of radius r has the same volume as a cap of radius $r \pm \varepsilon$ up to a factor of $1 \pm o(1)$. This happens for, say, $\varepsilon \ll 1/d$. This will guarantee that the number of bad pairs of cells is small. This argument can be found in [FS02]. ■

It is worth noting that the proof of the previous lemma gives a very weak bound on the number of vertices that we are required to subsample. In particular, it is not difficult to see that the average degree of the graph will be $n^{1-o(1)}$. A priori, it could therefore be the case that the SDP value changes when considering a subsample of the sphere with average degree $\log(n)$ or even $O(1)$. Indeed, [FS02] show that for some fixed γ, a random subsample of the sphere of expected degree $O(\log n)$ will satisfy most triangle inequality constraints with high probability thus exhibiting some integrality gap for sdp_3.[7] However, our main theorem implies that asymptotically sdp_3 behaves like $\sqrt{\gamma}$ rather than γ.

Theorem 9.9. *Fix $\gamma \in [0,1]$ and let $\Delta > \mathrm{poly}(1/\gamma)$. Fix d and choose n such that for n uniformly random points $V \subseteq \mathbb{S}^{d-1}$ the induced graph $G_\gamma[V]$ has expected degree Δ. Then,*

$$\mathrm{sdp}_3(G_\gamma[V]) = 1 - \Theta(\sqrt{\gamma}).$$

Proof. We think of $G_\gamma[V]$ as a uniform vertex subsample of a random dense discretization $G_\gamma[W]$ in d dimension. We then apply Lemma 9.8 and Theorem 2.1 to the graph $G_\gamma[W]$ and conclude the desired claim for $G_\gamma[V]$. ■

[7] In their example the angle between two neighboring vertices is chosen to be more than 60 degrees corresponding to very large γ to which our theorem does not apply due to the constant factor loss in γ.

10 Subsampling edges

In this section, we will prove the analogue of our main theorem in the edge subsampling model. Here, $G = (V, E)$ will always denote a Δ-regular graph on n vertices. Our proof in the case of edge subsampling is much simpler. As it turns out it suffices to bound the cut norm between the original graph and its subsample and to argue that the SDP value is a Lipschitz function of the cut norm. The latter fact is a consequence of Grothendieck's inequality.

We let $E_\delta \subseteq E$ denote a random subset of E of size $\delta|E|$. We'll overload notation slightly by using $G[E_\delta]$ for the graph G restricted to the edge set E_δ.

Definition 10.1. The cut norm of a real valued $n \times n$ matrix A is defined as

$$\|A\|_C = \max_{U,V\subseteq[n]} \left| \sum_{i\in U, j\in V} a_{ij} \right| . \tag{37}$$

It is known that the cut norm is within constant factors of the norm

$$\|A\|_{\infty\mapsto 1} = \max_{x_i,y_j\in\{-1,1\}} \sum_{i,j\in[n]} a_{ij} x_i y_j \, . \tag{38}$$

A natural semidefinite relaxation of (38) replaces every pair x_i, y_j by two unit vectors u_i, v_j, i.e.,

$$\mathrm{sdp}_C(A) = \max_{\|u_i\|=\|v_i\|=1} a_{ij} \langle u_i, v_j \rangle. \tag{39}$$

A theorem of Grothendieck bounds the gap between the cut norm and its relaxation by a multiplicative constant (the Grothendieck constant).

Theorem 10.2. *There is a constant K_G (known to be less than 1.8) such that $\mathrm{sdp}_C(A) \leqslant K_G \|A\|_{\infty\mapsto 1}$.*

The next lemma shows that the cut norm between a graph and its subsample is small.

Lemma 10.3. *Let $\delta \geqslant c\varepsilon^{-2}\Delta^{-1}$. Then,*

$$\mathbb{E} \left\| A(G) - \delta^{-1} A(G[E_\delta]) \right\|_{\infty\mapsto 1} \leqslant \varepsilon.$$

Proof. We can show that with probability $1 - e^{\Omega(n)}$, $|\langle x, Ay\rangle - \delta^{-1}\langle x, A'y\rangle| \leqslant \varepsilon$ simultaneously for all $x, y \in \{-1, 1\}^n$. The proof follows from Hoeffding's bound as in Lemma A.5 and the union bound. The details are straightforward and therefore omitted from this paper. ■

Similarly the following lemma can be shown.

Lemma 10.4. *Let $\delta \geqslant c\varepsilon^{-2}\Delta^{-1}$. Then,*

$$\mathbb{E} \left\| D(G) - \delta^{-1} D(G[E_\delta]) \right\|_{\infty\to 1} \leqslant \varepsilon \, . \tag{40}$$

The previous two lemmas showed that the expected difference in cut norm between the graph G and its edge subsample $G[E_\delta]$ is small.

Corollary 10.5. *For $\delta \geqslant c\varepsilon^{-2}\Delta^{-1}$, we have*

$$\mathbb{E}\,\|L(G) - \delta^{-1}L(G[E_\delta])\|_C \leqslant \varepsilon\,. \tag{41}$$

It turns out that bounding the difference in cut norm is sufficient for bounding the difference in SDP values.

Lemma 10.6. *Let G and G' be any two graphs on n vertices. Let $\mathcal{M} \subseteq \mathcal{M}_2$ (see Definition 5.1) be any set of positive semidefinite $n \times n$ matrices. Suppose $\|L(G) - L(G')\|_C \leqslant t$. Then,*

$$|\mathrm{sdp}_{\mathcal{M}}(G) - \mathrm{sdp}_{\mathcal{M}}(G')| \leqslant O(t)\,.$$

Proof.

$$\begin{aligned}
\left|\mathrm{sdp}_{\mathcal{M}}(G) - \mathrm{sdp}_{\mathcal{M}}(G')\right| &\leqslant \left| \max_{X \in \mathcal{M}_2} (L(G) - L(G')) \bullet X \right| \\
&\leqslant O(1) \cdot \|L(G) - L(G')\|_C && \text{(by Theorem 10.2)} \\
&\leqslant O(t)\,.
\end{aligned}$$

■

Corollary 10.7. *Let G denote a Δ-regular graph and let $\delta \geqslant \mathrm{poly}(1/\varepsilon)\Delta^{-1}$. Then,*

$$\mathbb{E}\,|\mathrm{sdp}_{\mathcal{M}}(G) - \mathrm{sdp}_{\mathcal{M}}(G[E_\delta])| \leqslant \varepsilon \tag{42}$$

for any $\mathcal{M} \subseteq \mathcal{M}_2$.

11 Sparse subgraphs and Sherali-Adams Lower Bounds

We will first show our Sherali-Adams lower bounds in the case of edge subsampling, since here the proof is more transparent and general. For convenience, we sample edges independently from a Δ-regular graph each with probability p. This won't make a difference compared to sampling a fixed number of edges. The proof strategy is to show that a constant degree subsample of any Δ-regular graph G has sufficient small set expansion so that the Sherali-Adams hierarchy on G will have value close to 1 regardless of the integral value of G. If the integral value of G is bounded away from 1, this will result in a gap instance.

We point out that in order to rule out a subsampling theorem for linear programs it would be sufficient to exhibit one graph with this behavior. Our proof is more general in this regard.

Lemma 11.1. *Let G be a Δ-regular graph for $\Delta \geqslant \omega(1)$ and let G' be obtained from G by keeping every edge independently with probability λ/Δ. Then, in expectation we need to remove at most $o(n)$ vertices (or edges) from G' such that the girth of G' is at least $r \geqslant \log \Delta / 3 \log \lambda$.*

Proof. Pick a random sequence of $k+1$ vertices $v_1, \ldots, v_{k+1}$ in G. For this sequence to form a cycle in G' we first need that it forms a cycle in G which means that each v_i is incident to v_{i-1} in G. This happens with probability Δ/n in each of the k steps. Furthermore, we need that all $k+1$ edges are selected into G'. Hence, the probability that they form a cycle in G' is at most

$$\left(\frac{\Delta}{n}\right)^k \left(\frac{\lambda}{\Delta}\right)^{k+1} = \frac{\lambda^{k+1}}{\Delta n^k}$$

Therefore we expect at most

$$\binom{n}{k+1}(k+1)! \cdot \frac{\lambda^{k+1}}{\Delta n^k} \leqslant \frac{(en)^{k+1}(k+1)!}{(k+1)^k} \cdot \frac{\lambda^{k+1}}{\Delta n^k} \leqslant n\frac{(e\lambda)^{k+1}}{\Delta}$$

cycles of length $k+1$. Summing over all lengths up to $r = \log\Delta/3\log\lambda$ gives us

$$\sum_{k=3}^{r} n\frac{(e\lambda)^{k+1}}{\Delta} \leqslant n \cdot \frac{\log\Delta}{2\log\lambda}\frac{\Delta^{1/2}}{\Delta} \leqslant \frac{n}{\Delta^{1/3}} = o(n)$$

cycles over all. ∎

Lemma 11.2. *Let $\lambda > 1, \eta > 0, \theta > 0$. Suppose G is a Δ-regular graph for $\Delta \geqslant n^\theta$ Let G' the graph obtained by sampling edges with probability $\frac{\lambda}{\Delta}$.*

Then, with probability high probability, we can remove $o(n)$ vertices such that all sets of size at most $\beta\Delta$ are $(1+\eta)$-sparse as long as $\beta \leqslant (c\lambda)^{-1/\eta-1/\theta}$ for sufficiently large constant c.

Proof. With every subgraph of size k having $(1+\eta)k$ edges we will associate a fixed spanning tree of $k-1$ edges. As in the proof of Lemma 11.1 each of these edges has a probability of $\Delta/n \cdot \lambda/\Delta = \lambda/n$ of appearing in G', since these are edges in a tree. The remaining ηk edges appear with probability at most λ/Δ. Hence, the expected number of k-subgraphs spanning $(1+\eta)k$ edges is at most

$$\begin{aligned}
\binom{n}{k}\binom{k^2}{(1+\eta)k}\left(\frac{\lambda}{n}\right)^{k-1}\left(\frac{\lambda}{\Delta}\right)^{\eta k} &\leqslant \left(\frac{en}{k}\right)^k (ek)^{(1+\eta)k}\left(\frac{\lambda}{n}\right)^{k-1}\left(\frac{\lambda}{\Delta}\right)^{\eta k} \\
&\leqslant n \cdot e^{3k}\lambda^{2k}\left(\frac{k}{\Delta}\right)^{\eta k} \\
&= n \cdot \left(e^{3/\eta}\lambda^{2/\eta}\frac{k}{\Delta}\right)^{\eta k} \\
&= n \cdot (f(\lambda,\eta)\beta)^k, \qquad (43)
\end{aligned}$$

where $f(\lambda,\eta) \sim (c\lambda)^{1/\eta}$.

At this point we will apply Lemma 11.1 to remove all cycles of length up to $r = \log\Delta/2\log\lambda = \frac{\theta}{2\log\lambda}\log n$. Hence we may assume w.l.o.g. that $k > r$ in (43). It remains to choose β small enough so that (43) simplifies to $o(1)$. This happens for

$$\beta < 1/f(\lambda,\eta)2^{(2\log\lambda)/\theta} = (c'\lambda)^{-1/\eta-1/\theta}.$$

∎

Remark 11.3. In the previous two lemmas, we assumed that edges were sampled independently from G with probability p. However, both Lemma 11.1 and Lemma 11.2 are true in the case where we sample pm edges independently from G. The reason is simply that the probability of seeing a fixed subgraph H on k edges in the second model is bounded by p^k. Conditioned on picking one edge of the subgraph, the second edge has probability $(pm-1)/m < p$ and so forth.

Definition 11.4. We say a graph G is l-path decomposable if every 2-connected subgraph H of G contains a path of length l such that every vertex of the path has degree 2 in H.

The following lemma appears implicitly in [ABLT06] and gives a way of proving that a graph is l-path decomposable.

Lemma 11.5. *Let $l \geqslant 1$ be an integer and $0 < \eta < \frac{1}{3l-1}$, and let G be a $(1+\eta)$-sparse graph which is not a cycle. Then, G is l-path decomposable.*

Combining the above lemma with our previous lemmas, we get the following theorem.

Theorem 11.6. *Let $\lambda \geqslant 1, \theta > 0$. Suppose G is a Δ-regular graph with $\Delta \geqslant n^{\theta}$. Let $G' = G[E_{\lambda/\Delta}]$. Then with high probability we can remove $o(n)$ vertices such that in the remaining graph every set of k vertices induces an $\Omega_{\theta,\lambda}(\log(n/k))$-path decomposable subgraph.*

Proof. By Lemma 11.2, w.h.p we can remove $o(n)$ vertices such that all induced subgraphs of size $\beta\Delta$ are $(1+\eta)$-sparse for $\eta = 1/\Omega_{\theta,\lambda}(\log(1/\beta))$.

By Lemma 11.5, it follows that all these subgraphs are $\Omega(\log(1/\beta))$-path decomposable. The probability that any of these subgraphs is a cycle is negligible.

The claim follows by putting $\beta = k/\Delta$. ∎

The previous estimates can be used to derive Sherali-Adams lower bounds for Max-Cut. The Sherali-Adams LP relaxation for Max-Cut is

$$\begin{aligned} \mathrm{lp}_r(G) = \max \quad & \sum_{(u,v)\in E} x_{uv} \\ \text{s.t.} \quad & \text{the vector } (x_{uv})_{u,v\in V} \text{ lies in the Sherali-Adams relaxation of the cut polytope} \end{aligned}$$

The Sherali-Adams relaxation of the cut polytope is obtained by applying r rounds of lift-and-project operations to the base set of linear inequalities that define the metric polytope, i.e., $\{x_{ij} + x_{jk} \geqslant x_{ik}, x_{ij} + x_{jk} + x_{ik} \leqslant 2, x_{ij} = x_{ji}, 1 \geqslant x_{ij} \geqslant 0\}$. For a formal introduction see, for instance, [CMM09]. This notion can be used to define Sherali-Adams relaxations for various other cut problems, analogously.

The next theorem is then a direct consequence of [CMM09].

Theorem 11.7. *Let $\varepsilon, \lambda > 0$. Suppose G is a Δ-regular graph with $\Delta > n^{\theta}$. Then with high probability over $G' = G[E_{\lambda/\Delta}]$, after removing $o(n)$ vertices,*

$$\mathrm{lp}_r(G') \geqslant 1 - \varepsilon$$

for $r = n^{\alpha}$ where $\alpha(1/\varepsilon, 1/\theta, \lambda)$ tends to zero as any of its arguments grows.

Proof. After removing $o(n)$ vertices, we get from Theorem 11.6 that every set of size k in G is $\Omega_{\theta,\lambda}(\log(n/k))$ path decomposable. Let $\alpha = \Omega_{\theta,\lambda}(\varepsilon^2)$ and $r = n^{\alpha}$.

The main theorem of [CMM09] shows that we have a solution $\{x_{uv}\}_{u,v\in V}$ to the r-th round of the Sherali-Adams relaxation for Max-Cut such that for two adjacent vertices u, v we have

$$x_{uv} \geqslant 1 - \varepsilon\,.$$

∎

As shown in [CMM09], the above approach extends to various other problem via reductions such as Unique Games, Vertex Cover and Balanced Separator.

11.1 Lower bounds for vertex subsampling linear programs

Similar results can be shown in the vertex subsampling model with the same proof strategy. We were unable to show Lemma 11.2 in its general form for any Δ-regular graph. However, it is not difficult to get lower bounds for specific graphs. Notice this is sufficient to rule out a subsampling theorem for linear programs. We will next sketch the proof of Theorem 2.2.

Theorem 2.2 (Restated)**.** *For every function $\varepsilon = \varepsilon(n)$ that tends to 0 with n, there exists a function $r = r(n)$ that tends to ∞ with n and family of graphs $\{G_n\}$ of degree $D = D(n)$ such that*

1. *For every n, $\mathrm{lp}_3(G_n) \leqslant 0.8$*
2. *If G' is a random subgraph of G of size $(n/D)^{1+\varepsilon(n)}$ then $\mathbb{E}[\mathrm{lp}_{r(n)}(G')] \geqslant 1 - \frac{1}{r(n)}$.*

where $\mathrm{lp}_k(H)$ denotes the value of k levels of the Sherali-Adams linear program for Max-Cut [SA90, dlVKM07] on the graph H.

Proof sketch. Let $G_n = G_{n,p}$ for some $p \leqslant \frac{1}{2}$. It is easy to see that three rounds of Sherali-Adams have value at most 0.7 on $G = G_n$ with high probability over G_n itself. This follows by considering triangles in G and arguing that every edge in G occurs in the same number of triangles up to negligible deviation. But 3 rounds of Sherali-Adams have value at most 2/3 on a triangle. Hence, $\mathrm{lp}_3(G) \leqslant 2/3 + o(1)$.

On the other hand let $\delta = \frac{n^\varepsilon}{D}$ where $D = pn$ is the expected degree of G. We observe that $G' = G[V_\delta]$ is exactly distributed like $G' = G_{m,\lambda/m}$ for $m = (n/D)^{1+\varepsilon}$ and $\lambda = m^\varepsilon$. Using standard calculations as in Lemma 11.1 and Lemma 11.2, one can check that such graphs have girth going to infinity, and for some $M \in \omega(1)$, all subsets size M are $(1+\eta)$-sparse, where $\eta \in o(1)$. Hence, we can follow the proof as above and use [CMM09] to argue that $G[V_\delta]$ has Sherali-Adams value larger than $1 - o(1)$ for $\omega(1)$ rounds, and therefore picking $r(n)$ sufficiently small concludes the proof sketch. ∎

Remark 11.8.
1. In the above proof we can also consider k-cliques which would give us a statement about k rounds of SA with the conclusion that $\mathrm{lp}_k(G) \leqslant 1/2 + c(k)$ where $c(k)$ tends to 0 as k grows.
2. We could also choose a subgraph of size $O(n/D)$ in which case we would get polynomial round lower bounds by repeating Lemma 11.1 and Lemma 11.2.
3. It is also easy to see that Lemma 11.1 is true in the vertex subsample model in general for any Δ-regular graph. With this lemma itself one can use the results of [dlVKM07] to show a $O(\log n)$-round lower bound.

References

[ABLT06] S. Arora, B. Bollobás, L. Lovász, and I. Tourlakis. Proving integrality gaps without knowing the linear program. *Theory of Computing*, 2(1):19–51, 2006. 34

[AdlVKK03] N. Alon, W. F. de la Vega, R. Kannan, and M. Karpinski. Random sampling and approximation of MAX-CSPs. *J. Comput. Syst. Sci.*, 67(2):212–243, 2003. 6, 7

[AHK06] S. Arora, E. Hazan, and S. Kale. A fast random sampling algorithm for sparsifying matrices. In *Proc. 10th APPROX*, pages 272–279. Springer, 2006. 7

[AKK+08] S. Arora, S. Khot, A. Kolla, D. Steurer, M. Tulsiani, and N. Vishnoi. Unique games on expanding constraints graphs are easy. In *Proc. 40th STOC*. ACM, 2008. 4

[AM07] D. Achlioptas and F. McSherry. Fast computation of low-rank matrix approximations. *J. ACM*, 54(2), 2007. 7

[AMMN05] N. Alon, K. Makarychev, Y. Makarychev, and A. Naor. Quadratic forms on graphs. In *Proc. 37th STOC*. ACM, 2005. 7

[AN04] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 72–80, New York, NY, USA, 2004. ACM. 7

[BHH+08] B. Barak, M. Hardt, I. Haviv, A. Rao, O. Regev, and D. Steurer. Rounding parallel repetitions of unique games. In *Proc. 49th FOCS*, pages 374–383. IEEE, 2008. 3

[BSS09] J. D. Batson, D. A. Spielman, and N. Srivastava. Twice-Ramanujan sparsifiers. In *Proc. 41st STOC*, pages 255–262. ACM, 2009. 3, 7

[Chl07] E. Chlamtac. Approximation algorithms using hierarchies of semidefinite programming relaxations. In *Proc. 48th FOCS*, pages 691–701. IEEE, 2007. 3

[CMM09] M. Charikar, K. Makarychev, and Y. Makarychev. Integrality gaps for sherali-adams relaxations. In *Proc. 41st STOC*. ACM, 2009. 34, 35

[CS08] E. Chlamtac and G. Singh. Improved approximation guarantees through higher levels of sdp hierarchies. In *Proc. APPROX-RANDOM*, pages 49–62. Springer, 2008. 3

[CW04] M. Charikar and A. Wirth. Maximizing quadratic programs: Extending Grothendieck's inequality. In *Proc. 45th FOCS*, pages 54–60. IEEE, 2004. 7, 8, 15

[Din07] I. Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007. 9

[dlVKM07] W. F. de la Vega and C. Kenyon-Mathieu. Linear programming relaxations of maxcut. In N. Bansal, K. Pruhs, and C. Stein, editors, *SODA*, pages 53–61. SIAM, 2007. 5, 6, 35

[FKV04] A. M. Frieze, R. Kannan, and S. Vempala. Fast monte-carlo algorithms for finding low-rank approximations. *J. ACM*, 51(6):1025–1041, 2004. 7

[FS02] U. Feige and G. Schechtman. On the optimality of the random hyperplane rounding technique for max cut. *Random Struct. Algorithms*, 20(3):403–440, 2002. 4, 6, 7, 8, 28, 29, 30

[GGR98] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998. 3, 6, 7, 8, 17

[GW94] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. Preliminary version in STOC'94. 4, 10

[Kho02] S. Khot. On the power of unique 2-prover 1-round games. In *Proc.* 34*th STOC*, pages 767–775. ACM, 2002. 3

[KKMO04] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput*, 37(1):319–357, 2007. Preliminary version in FOCS' 04. 4

[KRT08] J. Kempe, O. Regev, and B. Toner. Unique games with entangled provers are easy. In *Proc.* 49*th FOCS*. IEEE, 2008. 3

[KS09] S. Khot and R. Saket. SDP integrality gaps with local ℓ_1-embeddability. In *Proc.* 50*th FOCS*. IEEE, 2009. 3

[Las01] J. B. Lasserre. An explicit exact sdp relaxation for nonlinear 0-1 programs. In K. Aardal and B. Gerards, editors, *IPCO*, volume 2081 of *Lecture Notes in Computer Science*, pages 293–303. Springer, 2001. 3

[MOO05] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low in.uences invariance and optimality. In *Proc.* 46*th FOCS*, pages 21–30. IEEE, 2005. 4

[Ron00] D. Ron. Property testing. In *Handbook of Randomized Computing, Vol. II*, pages 597–649. Kluwer Academic Publishers, 2000. 3

[RS09] P. Raghavendra and D. Steurer. Integrality gaps for strong SDP relaxations of unique games. In *Proc.* 50*th FOCS*. IEEE, 2009. 3

[Rub06] R. Rubinfeld. Sublinear time algorithms. In *Proc. International Congress of Mathematicians*, volume 3, pages 1095–1110. European Math. Soc., 2006. 3

[RV07] M. Rudelson and R. Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *J. ACM*, 54(4), 2007. 7

[SA90] H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990. 5, 35

[Sch08] G. Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *FOCS*, pages 593–602. IEEE Computer Society, 2008. 3, 4, 6, 7

[SS08] D. A. Spielman and N. Srivastava. Graph sparsification by effective resistances. In *Proc.* 40*th STOC*. ACM, 2008. 7

[ST04] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proc.* 36*th STOC*, pages 81–90. ACM, 2004. 3, 7

[Tul09] M. Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proc. 41st STOC 2009*, pages 303–312. ACM, 2009. 3

A Deviation bounds

A.1 Deviation bounds for submatrices

The following general lemma is useful in bounding the deviation of expressions like $\langle x, A_{S\times S}x\rangle$ or $\sum_{i,j\in S}|a_{ij}|$ when S denotes a random subset of $[n]$ and A is a $n\times n$ matrix.

Lemma A.1. *Let A denote a symmetric $n\times n$ matrix such that $a_{ii}=0$ for all $i\in[n]$. Suppose there is some $\beta>0$ such that $|a_{ij}|\leqslant\beta$ for all $i,j\in[n]$ and $\sum_j|a_{ij}|\leqslant 1$ for all i. Now, let $S\subseteq[n]$ denote a random subset of $[n]$ of size δn for some $\delta>\beta$. Then, for all $\varepsilon>0$,*

$$\mathbb{Pr}\left(\left|\delta^{-2}\sum_{i,j\in S}a_{ij}-\sum_{i,j\in[n]}a_{ij}\right|>\varepsilon n\right)\leqslant\frac{O(1)}{\varepsilon^2\delta n}\,. \tag{44}$$

Proof. Denote by X_{ij} the random variable which is equal to a_{ij} when both $i\in S$ and $j\in S$ and is zero otherwise. Let $\mu_{ij}=\mathbb{E}\,X_{ij}=\delta^2a_{ij}$. Putting $X=\sum_{i,j\in[n]}X_{ij}$ and $\mu=\mathbb{E}\,X$ we will compute the variance of X. The key fact that we will use is that the selection of i,j and k,l is independent unless either $i=k$ or $j=l$. Pairs where neither is the case will not contribute to the variance. More precisely,

$$\begin{aligned}
\mathbb{E}\,(X-\mu)^2&=\mathbb{E}\left(\sum_{ij}X_{ij}-\mu_{ij}\right)^2\\
&=\mathbb{E}\left[\sum_{i,j,k,l}(X_{ij}-\mu_{ij})(X_{kl}-\mu_{kl})\right]\\
&=\sum_{ij}\mathbb{E}\,(X_{ij}-\mu_{ij})^2+\sum_{ijk}\mathbb{E}\,(X_{ij}-\mu_{ij})\,(X_{kj}-\mu_{kj})\\
&=\sum_{ij}O(\delta^2)a_{ij}^2+\sum_{ijk}O(\delta^3)a_{ij}a_{kj}\,.
\end{aligned}$$

At this point notice that $\sum_{ij}a_{ij}^2$ is maximized when in every row we have $1/\beta$ entries of magnitude β in which case the expression evaluates to $\frac{1}{\beta}\beta^2n=\beta n$. Likewise the second expression $\sum_{ijk}a_{ij}a_{kj}$ is maximized when in every column $j\in[n]$ we have $1/\beta$ nonzero entries of magnitude β. In this case the expression is $(1/\beta)^2\beta^2n=n$. Hence,

$$\sigma^2=\mathbb{E}\,(X-\mu)^2\leqslant O(\delta^2\beta n)+O(\delta^3n)\leqslant O(\delta^3n)\,, \tag{45}$$

where we used that $\delta>\beta$. Hence by Chebyshev's inequality,

$$\mathbb{Pr}(|X-\mu|\geqslant\varepsilon\delta^2n)\leqslant\frac{\sigma^2}{\varepsilon^2\delta^4n^2}=\frac{O(1)}{\varepsilon^2\delta n}\,.$$

This is what we claimed up to scaling. ∎

In the proof of the proxy graph theorem we used the following simple observation relating the Laplacian of a subsample $L(G[V_\delta])$ to the corresponding principal submatrix of the Laplacian $L(G)_{V_\delta}$.

Lemma A.2. *Let G be a Δ-regular graph and let H be a graph of degree at least Δ. Let $\delta \geqslant \mathrm{poly}(1/\varepsilon)\Delta^{-1}$. Then,*

$$\mathbb{E}\|L(G[V_\delta]) - L(G)_{V_\delta}\|_{\mathrm{TV}} \leqslant \varepsilon\,.$$

Proof. By inspection of the two matrices we see that the difference in the entries of the matrix is due to irregularities in the degrees of $G[V_\delta]$. Specifically, the matrix $L(G)_{V_\delta}$ has diagonal entries equal to $1/\delta n$. On the other hand, the i-th diagonal entry of $L(G[V_\delta])$, call it d_i, is equal to $\delta^{-1}\sum_{j\in V_\delta} a_{ij}$. We have that $\mathbb{E}\, d_i = \frac{1}{\delta n}$ and we claim,

$$\sum_{i\in V_\delta} \mathbb{E}\left|d_i - \frac{1}{\delta n}\right| \leqslant \varepsilon\,.$$

This can be derived from Lemma A.1. ■

A.2 A Gaussian tail bound

The following Gaussian tail bound was used in Section 6.

Fact A.3. *Let ξ be a standard Gaussian variable. Then,*

$$\mathbb{E}\max\{\xi - t, 0\}^2 = 2^{-\Omega(t^2)}\,.$$

Proof. The expectation is proportional to the integral

$$\int_{u\geqslant 0} u^2 e^{-(u+t)^2/2}\,\mathrm{d}u \leqslant e^{-t^2/2}\int_{u\geqslant 0} u^2 e^{-u^2/2}\,\mathrm{d}u = O\left(e^{-t^2/2}\right)\,.$$

■

A.3 Chernoff-like tail inequalities

We also needed McDiarmid's large deviation bound (sometimes called Azuma's inequality).

Lemma A.4. *Let $X_1,\ldots,X_m$ be independent random variables all taking values in the set $\mathcal{X}$. Further, let $f\colon \mathcal{X}^m \to \mathbb{R}$ be a function of $X_1,\ldots,X_m$ that satisfies for all $i, x_1, x_2, \ldots, x_m, x_i' \in \mathcal{X}$,*

$$|f(x_1,\ldots,x_i,\ldots,x_m) - f(x_1,\ldots,x_i',\ldots,x_m)| \leqslant c_i.$$

Then, for all $t > 0$,

$$\mathbb{P}\mathrm{r}\{|f - \mathbb{E}\,[f]\,| \geqslant t\} \leqslant 2\exp\left(\frac{-2t^2}{\sum_{i=1}^m c_i^2}\right).$$

When we have independent variables in $[0,1]$ with small expectation we can use the following version of Hoeffding's bound.

Lemma A.5. *Let $X_1,\ldots,X_m$ be independent random variables all taking values in the set $[0,1]$ and let $\mu = \mathbb{E}\left[\sum_{i\in[m]} X_i\right]$. Then, for every $\varepsilon \in (0,1)$,*

$$\mathbb{P}\mathrm{r}\left\{\left|\sum_{i\in[m]} X_i - \mu\right| \geqslant \varepsilon\mu\right\} \leqslant 2\exp(-\varepsilon^2\mu/3).$$

ECCC ISSN 1433-8092

http://eccc.hpi-web.de