Towards an SDP-based Approach to Spectral Methods
A Nearly-Linear-Time Algorithm for Graph Partitioning and Decomposition

Lorenzo Orecchia [*]   Nisheeth K. Vishnoi [†]

Abstract

In this paper, we consider the following graph partitioning problem: The input is an undirected graph $G = (V, E)$, a balance parameter $b \in (0, 1/2]$ and a target conductance value $\gamma \in (0, 1)$. The output is a cut which, if non-empty, is of conductance at most $O(f)$, for some function $f(G, \gamma)$, and which is either balanced or well correlated with all cuts of conductance at most γ. In a seminal paper, Spielman and Teng [16] gave an $\tilde{O}(|E|/\gamma^2)$-time algorithm for $f = \sqrt{\gamma \log^3 |V|}$ and used it to decompose graphs into a collection of near-expanders [18].

We present a new spectral algorithm for this problem which runs in time $\tilde{O}(|E|/\gamma)$ for $f = \sqrt{\gamma}$. Our result yields the first nearly-linear time algorithm for the classic BALANCED SEPARATOR problem that achieves the asymptotically optimal approximation guarantee for spectral methods.

Our method has the advantage of being conceptually simple and relies on a primal-dual semidefinite-programming (SDP) approach. We first consider a natural SDP relaxation for the BALANCED SEPARATOR problem. While it is easy to obtain from this SDP a certificate of the fact that the graph has no balanced cut of conductance less than γ, somewhat surprisingly, we can obtain a certificate for the stronger correlation condition. This is achieved via a novel separation oracle for our SDP and by appealing to Arora and Kale's [3] framework to bound the running time. Our result contains technical ingredients that may be of independent interest.

1 Introduction

1.1 Graph Partitioning. Given a graph $G = (V, E)$, the conductance of a cut $(S, \bar{S})$ is $\phi(S) \overset{\text{def}}{=} |E(S, \bar{S})| / \min\{\text{vol}(S), \text{vol}(\bar{S})\}$, where $\text{vol}(S)$ is the sum of the degrees of the vertices in the set S. A cut $(S, \bar{S})$ is b-balanced if $\min\{\text{vol}(S), \text{vol}(\bar{S})\} \geq b \cdot \text{vol} V$. A graph partitioning problem of widespread interest is the BALANCED SEPARATOR problem: given $G = (V, E)$, a constant[1] balance parameter $b \in (0, 1/2]$, and a conductance value $\gamma \in (0, 1)$, does G have a b-balanced cut S such that $\phi(S) \leq \gamma$?

---
[*] UC Berkeley. Supported by NSF grants CCF-0830797 and CCF-0635401. This work was initiated while this author was visiting Microsoft Research India, Bangalore.
[†] Microsoft Research India, Bangalore.
[1] We will use $O_b(\cdot)$ and $\Omega_b(\cdot)$ in our asymptotic notation when we want to emphasize the dependence of the hidden coefficient on b.

BALANCED SEPARATOR is an intensely studied problem in both theory and practice. It has far-reaching connections to spectral graph theory, the study of random walks and metric embeddings. Besides being a theoretically rich problem, BALANCED SEPARATOR is of great practical importance, as it plays a central role in the design of recursive algorithms, image segmentation and clustering.

Since BALANCED SEPARATOR is an NP-hard problem [6], we seek approximation algorithms that either output a cut of conductance at-most $f(\gamma, \log |V|)$ and balance $\Omega_b(1)$ or a certificate that G has no b-balanced cut of conductance at most γ. In their seminal series of papers [16, 18, 17], Spielman and Teng use an approximation algorithm for BALANCED SEPARATOR as a fundamental primitive to decompose the instance graph into a collection of near-expanders. This decomposition is then used to construct spectral sparsifiers and solve systems of linear equations in nearly linear time. Their algorithm has two crucial features: first, it runs in nearly linear time; second, in the case that no balanced cut exists in the graph, it outputs a certificate of a special form. This certificate consists of an unbalanced cut of small conductance which is well-correlated with all low-conductance cuts in the graph. We prove in Section A.1 in the Appendix that such a cut is indeed a negative certificate for the BALANCED SEPARATOR problem. Formally, they prove the following:

THEOREM 1.1. *[16] Given a graph G, a balance parameter $b \in (0, 1/2]$, $b = \Omega(1)$ and a conductance value $\gamma \in (0, 1)$, PARTITION(G, b, γ) runs in time τ and outputs a cut $S \subseteq V$ such that $\text{vol}(S) \leq 7/8 \cdot \text{vol}(G)$, $\phi(S) \leq f_1$ or $S = \emptyset$, and with high probability, either*

1. *S is $\Omega_b(1)$-balanced, or*

2. *for all $C \subset V$ such that $\text{vol}(C) \leq 1/2 \cdot \text{vol}(G)$ and $\phi(C) \leq O(\gamma)$, $\frac{\text{vol}(S \cap C)}{\text{vol}(C)} \geq 1/2$.*

Originally, Spielman and Teng showed Theorem 1.1 with $f_1 = O\left(\sqrt{\gamma \log^3 n}\right)$ and $\tau = \tilde{O}(m/\gamma^2)$. This was subsequently improved by Andersen, Chung and Lang [1] and then by Andersen and Peres [2] to the current best of $f_1 = O\left(\sqrt{\gamma \log n}\right)$ and $\tau = \tilde{O}(m/\sqrt{\gamma})$. All these results made use of bounds on the convergence of random walk processes on the instance graph, such as the Lovasz-Simonovits bounds [13]. These bounds

yield the log n factor in the approximation guarantee, which appears hard to remove while following this approach.

1.2 Our Contribution In this paper, we use a semidefinite programming approach to design a new spectral algorithm, called BALCUT, that improves on the result of Theorem 1.1. The following is our main result.

THEOREM 1.2. (MAIN THEOREM) *Given a graph $G = (V, E)$, a balance parameter $b \in (0, 1/2]$, $b = \Omega(1)$, and a conductance value $\gamma \in (0, 1)$, BALCUT(G, b, γ) runs in time $\tilde{O}(m/\gamma)$ and outputs a cut $S \subset V$ such that $\mathrm{vol}(S) \leqslant 7/8 \cdot \mathrm{vol}(G)$, if $S \neq \emptyset$ then $\phi(S) \leqslant O_b\left(\sqrt{\gamma}\right)$, and with high probability, either*

1. *S is $\Omega_b(1)$-balanced, or*

2. *for all $C \subset V$ such that $\mathrm{vol}(C) \leqslant 1/2 \cdot \mathrm{vol}(G)$ and $\phi(C) \leqslant O(\gamma)$, $\frac{\mathrm{vol}(S \cap C)}{\mathrm{vol}(C)} \geqslant 1/2$.*

Note that our result improves the parameters of previous algorithms by eliminating the log n factor in the quality of the cut output, making the approximation comparable to the best that can be hoped for using spectral methods [7]. Our result is also conceptually simple: we use the primal-dual framework of Arora and Kale [3] to solve SDPs combinatorially, and we give a new separation oracle that yields Theorem 1.2. Finally, our result implies an approximation algorithm for BALANCED SEPARATOR, as the guarantee of Theorem 1.2 on the cut S output by BALCUT also implies a lower bound on the conductance of balanced cuts of G. The proof can be found in Section A.1 in the Appendix.

COROLLARY 1.1. *Given an instance graph G, a balance parameter $b \in (0, 1/2]$ and a target conductance $\gamma \in (0, 1]$, BALCUT (G, b, γ) either outputs an $\Omega_b(1)$-balanced cut of conductance at most $O_b(\sqrt{\gamma})$ or a certificate that all $\Omega_b(1)$-balanced cuts have conductance at least $\Omega_b(\gamma)$. The running time of the algorithm is $\tilde{O}(m/\gamma)$.*

This is the first nearly-linear-time spectral algorithm for BALANCED SEPARATOR that achieves the asymptotically optimal approximation guarantee for spectral methods.

1.3 Graph Decomposition. The main application of Theorem 1.1 is the construction of a particular kind of graph decomposition. In this decomposition, we wish to partition the vertex set of the instance graph V into components $V_1, \ldots, V_i, \ldots, V_k$ such that the graph induced by G on each V_i has conductance as large as possible, while at most a constant fraction of the edges have endpoints in different components. These decompositions are a useful algorithmic tool in several areas [20, 11, 18].

Kannan, Vempala and Vetta [10] construct such decompositions achieving a conductance value of $\Omega(1/\log^2 n)$. However, their algorithm runs in time $\tilde{O}(m^2)$ on some instances.

Spielman and Teng [18] relax this notion of decomposition by only requiring that each V_i be contained in a superset W_i in G, where W_i has large induced conductance in G. In the same work, they show that this relaxed notion of decomposition suffices for the purposes of sparsification by random sampling. The advantage of this relaxation is that it is now possible to compute this decomposition in nearly-linear time by recursively applying the algorithm of Theorem 1.1.

THEOREM 1.3. *[18] Assume the existence of an algorithm achieving parameters τ and f_1 in Theorem 1.1. Given $\gamma \in (0, 1)$, in time $\tilde{O}(\tau)$, it is possible to construct a decompositions of the instance graph G into components $V_1, \ldots, V_k$ such that:*

1. *for each V_i, there exists $W_i \supseteq V_i$ such that the conductance of the graph induced by G on W_i is $\Omega(\gamma/\log n)$.*

2. *the fraction of edges with endpoints in different components is $O(f_1 \log n)$.*

Using Theorem 1.3, Spielman and Teng showed the existence of a decomposition achieving conductance $\Omega(1/\log^6 n)$. Our improved results in Theorem 1.2 imply that we can obtain decompositions of the same kind with conductance bound $\Omega(1/\log^3 n)$. Our improvement also implies speed-ups in the sparsification procedure described by Spielman and Teng [18]. However, this result has since been superceded by work of Koutis, Miller and Peng [12] that gives a very fast linear equation solver that can be used to compute sampling probabilities for each edge, yielding a spectral sparsifier with high probability [15].

Our work leaves open the important question posed by Spielman [14] of whether stronger decompositions, of the kind proposed by Kannan, Vempala and Vetta [10], can be produced in nearly-linear time.

1.4 Overview of Techniques

Spectral Approach. The simplest algorithm for BALANCED SEPARATOR, also used by Kannan et al. [10], is the recursive spectral algorithm. This algorithm finds the minimum-conductance sweep cut of the second eigenvector of G, removes the cut and all adjacent edges from G, and reiterates on the remaining graph. The algorithm stops when the union of the cuts removed becomes $b/2$-balanced or when the residual graph is found to have spectral gap at least γ, certifying that no more progress can be made. As every cut may only remove $O(1)$ volume and the eigenvector computation takes $\Omega(m)$ time, this algorithm may have quadratic running time. It can be shown using Cheeger's Inequality [5] that the cut this procedure outputs is of conductance at most $O(\sqrt{\gamma})$.

Spielman-Teng Approach. The algorithm of Spielman and Teng which proves Theorem 1.1 is also spectral in nature and uses, as the main subroutine, *local random walks* that run in time proportional to the volume of the output cut to find sparse cuts around vertices of the graphs. These local methods are based on non-trivial random walks on the input graph and aggregation of the information obtained from these walks, all performed while maintaining nearly-linear running time.

Our Approach. We depart from the random-walk paradigm and first consider a natural SDP relaxation for the BALANCED SEPARATOR problem, which BALCUT solves approximately using a primal-dual method. Intuitively, BALCUT manages to maintain the approximation guarantee of the recursive spectral algorithm while running in nearly-linear time by considering a distribution over eigenvectors, represented as a vector embedding of the vertices, rather than a single eigenvector, at each iteration. The sweep cut over the eigenvector is replaced by a sweep cut over the radius of the vectors in the embedding (see Figure 1). Moreover, at any iteration, rather than removing the unbalanced cut found, BALCUT penalizes it by modifying the graph so that it is unlikely but still possible for it to turn up in future iterations. Hence, in both its cut-finding and cut-eliminating procedures, BALCUT tends to "hedge its bets" more than the greedy recursive spectral method. This hedging, which ultimately allows BALCUT to achieve its faster running time, is implicit in the primal-dual framework of Arora and Kale [3].
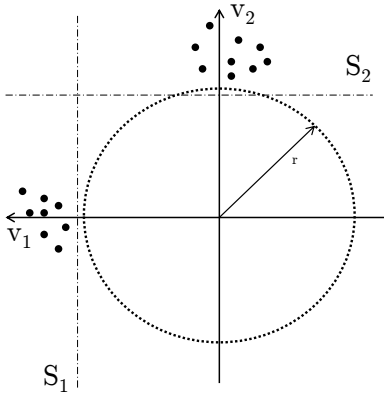

Figure 1: Schematic representation of the speed-up introduced by BALCUT when the instance graph contains many unbalanced cuts of low conductance. Let v_1 and v_2 be the two slowest-mixing eigenvectors of G. Assume that their minimum-conductance sweep cuts S_1 and S_2 are unbalanced cuts of conductance less than γ. If we use the recursive algorithm of Kannan et al. [10], two iterations could be required to remove S_1 and S_2. However, BALCUT considers a multidimensional embedding containing contributions from multiple eigenvectors and performs a radial sweep cut. This allows S_1 and S_2 to be removed in a single iteration.

The SDP relaxation appears in Figure 1.4. We denote by $\mu : V \mapsto \mathbb{R}_{\geq 0}$ the distribution defined as $\mu_i \overset{\text{def}}{=} d_i/\text{vol}(G)$, and by d_i the degree of the i-th vertex. Also, $v_{avg} \overset{\text{def}}{=} \sum_i \mu_i v_i$. Even though our algorithm uses the SDP, at the core, it is spectral in nature, as it relies on the matrix-vector multiplication primitive. Hence, if one delves deeper, a random walk interpretation can be derived for our algorithm.

$$\text{psdp}(G, b, \gamma): \qquad 1/4 \cdot \underset{\{i,j\}\in E}{\mathbb{E}} \|v_i - v_j\|_2^2 \leq \gamma$$

$$\underset{j\sim\mu}{\mathbb{E}} \|v_j - v_{\text{avg}}\|_2^2 = 1$$

$$\forall i \in V \qquad \|v_i - v_{\text{avg}}\|_2^2 \leq \frac{(1-b)}{b}$$

Figure 2: SDP for b-BALANCED SEPARATOR

The Primal-Dual Framework. For our SDP, the method of Arora and Kale can be understood as a game between two players: an embedding player and an oracle player. The embedding player, in every round of this game, gives a candidate vector embedding of the vertices of the instance graph to the oracle player. We show that, if we are lucky and the embedding is feasible for the SDP and, in addition, also has the property that for a large set S, for every $i \in S$, $\|v_i - v_{avg}\|_2 \leq O((1-b)/b)$ (we call such an embedding roundable), then a projection of the vectors along a random direction followed by a sweep cut gives an $\Omega_b(1)$-balanced cut of conductance at most $O(\sqrt{\gamma})$. The difficult case is when the embedding given to the oracle player is not roundable. In this case, the oracle outputs a candidate dual solution along with a cut. The oracle obtains this cut by performing a radial sweep cut of the vectors given by the embedding player. If at any point in this game the union of cuts output by the oracle becomes balanced, we output this union and stop. We show that such a cut is of conductance at most $O_b(\sqrt{\gamma})$. If this union of cuts is not balanced, then the embedding player uses the dual solution output by the oracle to update the embedding. Finally, the matrix-exponential update rule ensures that this game cannot keep on going for more that $O(\log n/\gamma)$ rounds. Hence, if a balanced cut is not found after this many rounds, we certify that the graph does not contain any b-balanced cut of conductance less than γ. To achieve a nearly-linear running time, we maintain only a $\log n$-dimensional sketch of the embedding. The guarantee on the running time then follows by noticing that, in each iteration, the most expensive computational step for each player is a logarithmic number of matrix-vector multiplications, which takes at most $\tilde{O}(m)$ time.

The reason why our approach yields the desired correlation condition in Theorem 1.2 is that, if no balanced cut is found, every unbalanced cut of conductance lower than γ will, at some iteration, have a lot of its vertices mapped to vectors of large radius. At that iteration, the cut output by the oracle player will have a large correlation with the target cut, which implies that the union of cuts output by the oracle player will also display such large correlation. This intuition is formalized in the proof of Theorem 1.2.

Our Contribution. The implementation of the oracle player, specifically dealing with the case when the embedding is not roundable, is the main technical novelty of the paper.

Studying the problem in the SDP-framework is the main conceptual novelty. The main advantage of using SDPs to design a spectral algorithm seems to be that SDP solutions provide a simple representation for possibly complex random-walk objects. Furthermore, the benefits of using a carefully designed SDP formulation can often be reaped with little or no burden on the running time of the algorithm, thanks to the primal-dual framework of Arora and Kale [3].

1.5 Rest of the Paper In Section 2.1, we set the notation for the paper. In Section 2.2, we present our SDP and its dual, and also define the notion of a roundable embedding. In Section 2.3, we present the algorithm BALCUT and the separation oracle ORACLE, and reduce the task of proving Theorem 1.2 to proving statements about the ORACLE. Section 3 contains the proof of the main theorem about the ORACLE used in Section 2.3. For clarity of presentation, several proofs are omitted from the above sections and appear in the appendix.

2 Algorithm Statement and Main Theorems

2.1 Notation

Instance graph and edge volume. We denote by $G = (V, E)$ the unweighted instance graph, where $|V| = n$ and $|E| = m$. We let $d \in \mathbb{R}^V$, be the degree vector of G, i.e. d_i is the degree of vertex i. We mostly work with the edge measure μ over V, defined as $\mu_i \stackrel{\text{def}}{=} \mu(i) \stackrel{\text{def}}{=} d_i/2m$. For a subset $S \subseteq V$, we also define μ_S as the edge measure over S, i.e. $\mu_S(i) \stackrel{\text{def}}{=} \mu(i)/\mu(S)$.

Special graphs For a subset $S \subseteq V$, we denote by K_S the complete graph over S such that edge $\{i, j\}$ has weight $\mu_i \mu_j$ for $i, j \in S$ and 0 otherwise. K_V is the complete graph with weight $\mu_i \mu_j$ between every pair $i, j \in V$.

Graph matrices. For an undirected graph $H = (V, E_H)$, let $A(H)$ denote the adjacency matrix of H and $D(H)$ the diagonal matrix of degrees of H. The (combinatorial) Laplacian of H is defined as $L(H) \stackrel{\text{def}}{=} D(H) - A(H)$. Note that for all $x \in \mathbb{R}^V$, $x^T L(H) x = \sum_{\{i,j\} \in E_H} (x_i - x_j)^2$. By D and L, we denote $D(G)$ and $L(G)$ respectively.

Vector and matrix notation. For a symmetric matrix M, we will use $M \succeq 0$ to denote that it is positive semi-definite and $M \succ 0$ to denote that it is positive definite. The expression $A \succeq B$ is equivalent to $A - B \succeq 0$. For two matrices A, B of equal dimensions, denote $A \bullet B \stackrel{\text{def}}{=} \text{Tr}(A^T B) = \sum_{ij} A_{ij} \cdot B_{ij}$. For a matrix A, we indicate by t_A the time necessary to compute the matrix-vector multiplications Au for any vector u.

Embedding notation. We will deal with vector embeddings of G, where each vertex $i \in V$ is mapped to a vector $v_i \in \mathbb{R}^d$. For such an embedding $\{v_i\}_{i \in V}$, we denote by v_{avg} the mean vector, i.e. $v_{\text{avg}} \stackrel{\text{def}}{=} \sum_{i \in V} \mu_i v_i$. Given a vector embedding of $\{v_i \in \mathbb{R}^d\}_{i \in V}$, recall that $X \succeq 0$, is the Gram matrix of the embedding if $X_{ij} = v_i^T v_j$. For any $X \in \mathbb{R}^{V \times V}, X \succeq 0$, we call $\{v_i\}_{i \in V}$ the *embedding corresponding to X* if X is the Gram matrix of $\{v_i\}_{i \in V}$. For $i \in V$, we denote by R_i the matrix such that

$R_i \bullet X = \|v_i - v_{\text{avg}}\|_2^2$.

Basic facts. We will alternatively use vector and matrix notation to reason about the graph embeddings. The following are some simple conversions between vectors and matrix forms and some basic geometric facts which follow immediately from definitions.

FACT 2.1. $\mathbb{E}_{i \sim \mu} \|v_i - v_{\text{avg}}\|_2^2 = 1/2 \cdot \mathbb{E}_{\{i,j\} \sim \mu \times \mu} \|v_i - v_j\|_2^2 = L(K_V) \bullet X$.

FACT 2.2. *For a subset $S \subseteq V$*, $\sum_{i \in S} \mu_i R_i \succeq \mu(S) L(K_V) - L(K_S)$.

FACT 2.3. *For a subset $S \subseteq V$*, $\mathbb{E}_{\{i,j\} \sim \mu_S \times \mu_S} \|v_i - v_j\|_2^2 = 2 \cdot 1/\mu(S)^2 \cdot L(K_S) \bullet X$.

Modified matrix exponential update. Let $\mathcal{S}_D$ e the subspace of $\mathbb{R}^V$ orthogonal to $\hat{v} \stackrel{\text{def}}{=} 1/\sqrt{2m} \cdot D^{1/2} 1$ and let $\mathcal{I}$ be the identity over $\mathcal{S}_D$, i.e. $\mathcal{I} \stackrel{\text{def}}{=} I - 1/2m \cdot D^{1/2} 1 1^T D^{1/2}$. For a positive ε and a symmetric matrix $M \in \mathbb{R}^{V \times V}$, we define

$$U_\varepsilon(A) \stackrel{\text{def}}{=} 2m \cdot \frac{D^{-1/2} e^{-(2m \cdot \varepsilon) \cdot D^{-1/2} M D^{-1/2}} D^{-1/2}}{\mathcal{I} \bullet e^{-(2m \cdot \varepsilon) D^{-1/2} M D^{-1/2}}}.$$

The following fact about $\mathcal{I}$ will also be needed:

FACT 2.4. $\mathcal{I} = 2m \cdot D^{-1/2} L(K_V) D^{-1/2}$.

2.2 SDP Formulation
We consider an SDP relaxation to the decision problem of determining whether the instance graph G has a b-balanced cut of conductance at most γ. The SDP feasibility program $\text{psdp}(G, b, \gamma)$ appears in Figure 2.2, where we also rewrite the program in matrix notation, using Fact 2.1 and the definition of R_i. psdp can be seen as

$$\text{psdp}(G, b, \gamma) : \quad \mathop{\mathbb{E}}_{\{i,j\} \in E} \|v_i - v_j\|_2^2 \leq 4\gamma$$
$$\mathop{\mathbb{E}}_{j \sim \mu} \|v_j - v_{\text{avg}}\|_2^2 = 1$$
$$\forall i \in V \quad \|v_i - v_{\text{avg}}\|_2^2 \leq \frac{1-b}{b}$$

$$\text{psdp}(G, b, \gamma) : \quad \frac{1}{m} \cdot L \bullet X \leq 4\gamma$$
$$L(K_V) \bullet X = 1$$
$$\forall i \in V \quad R_i \bullet X \leq \frac{1-b}{b}$$

Figure 3: SDP for b-BALANCED SEPARATOR

a scaled version of the balanced-cut SDP of [4], modified by replacing v_{avg} for the origin and removing the triangle-inequality constraints. The first change makes our psdp invariant under translation of the embeddings and makes the connection to spectral methods more explicit. Indeed, the first two constraints of psdpnow exactly correspond to the standard eigenvector problem, with the addition of the R_i

constraint ideally forcing all entries in the eigenvector not to be too far from the mean, just as it would be the case if the eigenvector exactly corresponded to a balanced cut. The removal of the triangle-inequality constraints causes psdp to only deal with the spectral structure of L and not to have a flow component. For the rest of the paper, denote by Δ the set $\{X \in \mathbb{R}^{V \times V}, X \succeq 0 : L(K_V) \bullet X = 1\}$.

The following simple lemma establishes that psdp is indeed a relaxation for the integral decision question and is proved in Section A.2.

LEMMA 2.1. (SDP IS A RELAXATION) *If there exists a b-balanced cut S with $\phi(S) \leqslant \gamma$, then $\mathsf{psdp}(G, b, \gamma)$ has a feasible solution.*

BALCUT will use the primal-dual approach of [3] to determine the feasibility of $\mathsf{psdp}(G, b, \gamma)$. When psdp is infeasible, BAL-CUT will output a solution to the dual $\mathsf{dsdp}(G, b, \gamma)$, shown in Figure 4.

$$\mathsf{dsdp}(G, b, \gamma): \quad \alpha - \frac{1-b}{b} \sum_{i \in V} \beta_i > 4\gamma$$

$$\frac{1}{2m} \cdot L + \sum_{i \in V} \beta_i R_i - \alpha L(K_V) \succeq 0$$

$$\alpha \in \mathbb{R}, \quad \beta \geqslant 0$$

Figure 4: $\mathsf{dsdp}(G, b, \gamma)$ feasibility problem

In the rest of the paper, we are going to use the following shorthands for the dual constraints

$$V(\alpha, \beta) \stackrel{\text{def}}{=} \alpha - \frac{1-b}{b} \sum_{i \in V} \beta_i, \quad M(\alpha, \beta) \stackrel{\text{def}}{=} \frac{L}{2m} + \sum_{i \in V} \beta_i R_i - \alpha L(K_V).$$

Notice that $V(\alpha, \beta)$ is a scalar, while $M(\alpha, \beta)$ is a matrix in $\mathbb{R}^{V \times V}$. Given $X \succeq 0$, a choice of (α, β) such that $V(\alpha, \beta) > 4\gamma$ and $M(\alpha, \beta) \bullet X \geqslant 0$ corresponds to a hyperplane separating X from the feasible region of $\mathsf{psdp}(G, b, \gamma)$ and constitutes a certificate that X is not feasible.

Ideally, BALCUT would produce a feasible solution to psdp and then round it to a balanced cut. However, as discussed in [3], it often suffices to find a solution "close" to feasible for the rounding procedure to apply. In the case of psdp, the concept of "closeness" is captured by the notion of *roundable* solution.

DEFINITION 2.1. (ROUNDABLE EMBEDDING) *Given an embedding $\{v_i\}_{i \in V}$, let $R = \{i \in V : \|v_i - v_{\mathsf{avg}}\|_2^2 \leqslant 32 \cdot {(1-b)}/{b}\}$. We say that $\{v_i\}_{i \in V}$ is a roundable solution to $\mathsf{psdp}(G, b, \gamma)$ if:*

- $\mathbb{E}_{\{i,j\} \in E} \|v_i - v_j\|_2^2 \leqslant 2\gamma$,

- $\mathbb{E}_{j \sim \mu} \|v_j - v_{\mathsf{avg}}\|_2^2 = 1$,

- $\mathbb{E}_{\{i,j\} \sim \mu_R \times \mu_R} \|v_i - v_j\|_2^2 \geqslant {1}/{64}$.

A roundable embedding can be converted into a balanced cut of the conductance required by Theorem 1.2 by using a standard projection rounding, which is a simple extension of an argument already appearing in [4] and [3]. The rounding procedure PROJROUND is described precisely in Section A.4, where the following theorem is proved.

THEOREM 2.1. (ROUNDING ROUNDABLE EMBEDDINGS) *If $\{v_i \in \mathbb{R}^d\}_{i \in V}$ is a roundable solution to $\mathsf{psdp}(G, b, \gamma)$, then PROJROUND($\{v_i\}_{i \in V}, b$) produces a $\Omega_b(1)$- balanced cut of conductance $O_b\left(\sqrt{\gamma}\right)$ with high probability in time $\tilde{O}(nd + m)$.*

2.3 Primal-Dual Framework

Separation Oracle. The problem of checking the feasibility of a SDP can be reduced to that of, given a candidate solution X, to check whether it is close to feasible and, if not, provide a certificate of infeasibility in the form of a hyperplane separating X from the feasible set. The algorithm performing this computation is known as a separation oracle. Arora and Kale show that the original feasiblity problem can be solved very efficiently if there exists a separation oracle obeying a number of conditions. We introduce the concept of *good* separation oracle to capture these conditions for the program $\mathsf{psdp}(G, \beta, \gamma)$.

DEFINITION 2.2. (GOOD SEPARATION ORACLE) *An algorithm is a good separation oracle if, on input some representation of X, the algorithm either finds X to be a roundable solution to $\mathsf{psdp}(G, b, \gamma)$ or outputs coefficents α, β such that $V(\alpha, \beta) \geqslant {3}/{4} \cdot \gamma$, $M(\alpha, \beta) \bullet X \geqslant {1}/{64} \cdot \gamma$ and $-\gamma L(K_V) \preceq M(\alpha, \beta) \preceq 3L(K_V)$.*

Algorithmic Scheme. We adapt the techniques of [3] to our setting, where we require feasible solutions to be in Δ rather than having trace equal to 1. The argument is a simple modification of the anaylsis of [3] and in [19]. The algorithmic strategy of [3] is to produce a sequence of candidate primal solutions $X^{(1)}, \ldots, X^{(T)}$ iteratively, such that $X^{(t)} \in \Delta$ for all t.

Our starting point $X^{(1)}$ will be the solution ${2m}/{n-1} \cdot D^{-1}$. At every iteration, a *good* separation oracle ORACLE will take $X^{(t)}$ and either guarantee that $X^{(t)}$ is roundable or output coefficents $\alpha^{(t)}, \beta^{(t)}$ certifying the infeasiblity of $X^{(t)}$. The algorithm makes use of the information contained in $\alpha^{(t)}, \beta^{(t)}$ by updating the next candidate solution as follows:

$$(2.1) \qquad P^{(t)} \stackrel{\text{def}}{=} -\varepsilon/6 \cdot \left(M(\alpha^{(t)}, \beta^{(t)}) + \gamma L(K_V) \right)$$

$$X^{(t+1)} \stackrel{\text{def}}{=} U_\varepsilon \left(\sum_{i=1}^t P^{(i)} \right)$$

where $\varepsilon > 0$ is a parameter of the algorithm. The following is immediate.

LEMMA 2.2. *For all $t > 0$, $X^{(t)} \in \Delta$.*

Following [3], we prove that, after a small number of iterations this algorithm either yields a roundable embedding or a

Figure 5: The BALCUT Algorithm

feasible solution to $\text{dsdp}(G, b, \Omega(\gamma))$. We present the proof in Section A.3 for completeness.

THEOREM 2.2. (ITERATIONS OF ORACLE, [3]) *Let $\varepsilon = 1/130$. Assume that the procedure ORACLE is a good separation oracle . Then, after $T = O\left(\log n/\gamma\right)$ iterations of the update of Equation 2.1, we either find a roundable solution to $\text{psdp}(G, b, \gamma)$ or the coefficents $\bar{\alpha} = 1/T \sum_{t=1}^{T} \alpha^{(t)}$ and $\bar{\beta} = 1/T \sum_{t=1}^{T} \beta^{(t)}$ are a feasible solution to $\text{dsdp}(G, b, 3/16 \cdot \gamma)$.*

Approximate Computation. Notice that, while we are seeking to construct a nearly-linear-time algorithm, we cannot hope to compute $X^{(t)}$ exactly and explicitly, as just maintaining the full $X^{(t)}$ matrix requires quadratic time in n. Instead, we settle for a approximation $\tilde{X}^{(t+1)}$ to $X^{(t+1)}$ which we define as

$$\tilde{X}^{(t+1)} = \tilde{U}_\varepsilon\left(\sum_{i=1}^{t} P^{(i)}\right).$$

The function $\tilde{U}_\varepsilon$ is a randomized approximation to U_ε obtained by applying the Johnson-Linderstrauss dimension reduction to the embedding corresponding to U_ε. $\tilde{U}_\varepsilon$ is described in full in Section A.5, where we also prove the following lemma about the accuracy and sparsity of the approximation. It is essentially the same argument appearing in [9] applied to our context.

LEMMA 2.3. (APPROXIMATE COMPUTATION) *Let $\varepsilon = 1/130$. For a matrix $M \in \mathbb{R}^{V \times V}$, $M \succeq 0$, let $\tilde{X} \stackrel{\text{def}}{=} \tilde{U}_\varepsilon(M)$ and $X \stackrel{\text{def}}{=} U_\varepsilon(M)$.*

1. *$\tilde{X} \succeq 0$ and $\tilde{X} \in \Delta$.*

2. *The embedding $\{\tilde{v}_i\}_{i \in V}$ corresponding to $\tilde{X}$ can be represented in $d = O(\log n)$ dimensions.*

3. *$\{\tilde{v}_i \in \mathbb{R}^d\}_{i \in V}$ can be computed in time $\tilde{O}(t_M + n)$.*

4. *for any graph $H = (V, E_H)$, with high probability*

$$(1 - 1/64) \cdot L(H) \bullet X - \tau \leqslant L(H) \bullet \tilde{X} \leqslant (1 + 1/64) \cdot L(H) \bullet X + \tau,$$

and, for any vertex $i \in V$,

$$(1 - 1/64) \cdot R_i \bullet X - \tau \leqslant R_i \bullet \tilde{X} \leqslant (1 + 1/64) \cdot R_i \bullet X + \tau,$$

where $\tau \leqslant O(1/\text{poly}(n))$.

This lemma shows that $\tilde{X}^{(t)}$ is a close approximation to $X^{(t)}$. We will use this lemma to show that ORACLE can receive $\tilde{X}^{(t)}$ as input, rather than $X^{(t)}$, and still meet the conditions of Theorem 2.2. In the rest of the paper, we assume that $\tilde{X}^{(t)}$ is represented by its corresponding embedding $\{\tilde{v}_i^{(t)}\}_{i \in V}$.

The Oracle. ORACLE is described in Figure 6. We show that ORACLE on input $\tilde{X}^{(t)}$ meets the condition of Theorem 2.2. Moreover, we show that ORACLE obeys an additional condition, which, combined with the dual guarantee of Theorem 2.2 will yield the correlation property of BALCUT.

THEOREM 2.3. (MAIN THEOREM ON ORACLE) *On input $\tilde{X}^{(t)}$, ORACLE runs in time $\tilde{O}(m)$ and is a good separation oracle for $X^{(t)}$. Moreover, the cut B in Step 4 is guaranteed to exist.*

Proof of Main Theorem. We are now ready to prove Theorem 1.2. To show the overlap condition, we consider the dual condition implied by Theorem 2.2 together with the cut $B^{(t)}$ and the values of the coefficents output by the ORACLE.

Proof. [Proof of Theorem 1.2] If at any iteration t, the embedding $\{\tilde{v}_i^{(t)}\}_{i \in V}$ corresponding to $\tilde{X}^{(t)}$ is roundable, the standard projection rounding PROJROUND produces a cut of bal-

1. **INPUT:** The embedding $\{\tilde{v}_i\}_{i \in V}$, corresponding to $\tilde{X} \in \Delta$. Let $r_i = \|\tilde{v}_i - \tilde{v}_{\mathsf{avg}}\|_2$ for all $i \in V$. Denote $R \stackrel{\text{def}}{=} \{i \in V : r_i^2 \leqslant 32 \cdot (1-b)/b\}$.

2. **CASE 1:** $\mathbb{E}_{\{i,j\} \in E} \|\tilde{v}_i - \tilde{v}_j\|_2^2 \geqslant 2\gamma$. Output $\alpha = \gamma$, $\beta = 0$ and $B = \emptyset$.

3. **CASE 2:** not CASE 1 and $\mathbb{E}_{\{i,j\} \sim \mu_R \times \mu_R} \|v_i - v_j\|_2^2 \geqslant \delta$. Then $\{\tilde{v}_i\}_{i \in V}$ is roundable, as $\tilde{X} \in \Delta$ implies $\mathbb{E}_{j \sim \mu} \ r_j^2 = 1$.

4. **CASE 3:** not CASE 1 or 2. Relabel the vertices of V such that $r_1 \geqslant r_2 \geqslant \ldots \geqslant r_n$ and let $S_i = \{1, \ldots, i\}$ be the j-th sweep cut of r. Let z the smallest index such that $\mu(S_z) \geqslant b/8$. Let B the most balanced sweep cut among $\{S_1, \ldots, S_{z-1}\}$ such that $\phi(B) \leqslant 2048 \cdot \sqrt{\gamma}$. Output $\alpha = 7/8\gamma$, $\beta_i = \mu_i \cdot \gamma$ for $i \in B$ and $\beta_i = 0$ for $i \notin B$. Also output the cut B.

Figure 6: ORACLE

ance $\Omega_b(1)$ and conductance $O_b(\sqrt{\gamma})$ by Theorem 2.1. Similarly, if for any t, $C^{(t)}$ is $b/4$-balanced, BALCUT satisfies the balance condition in Theorem 1.2, as $\phi(C^{(t)}) \leqslant O(\sqrt{\gamma})$ because $C^{(t)}$ is the union of cuts of conductance at most $O(\sqrt{\gamma})$.

Otherwise, after $T = O(\log n/\gamma)$ iterations, by Theorem 2.2, we have that $\bar{\alpha} = 1/T \sum_{t=1}^{T} \alpha^{(t)}$ and $\bar{\beta} = 1/T \sum_{t=1}^{T} \beta^{(t)}$ constitute a feasible solution $\mathsf{dsdp}(G, b, 3/16 \cdot \gamma)$. This implies that $M(\bar{\alpha}, \bar{\beta}) \succeq 0$, i.e.

$$(2.2) \qquad \frac{1}{m} \cdot L + \sum_{i \in V} \bar{\beta}_i R_i - \bar{\alpha} L(K_V) \succeq 0.$$

For any cut C such that $\mu(C) \leqslant 1/2$ and $\phi(C) \leqslant \gamma/16$, let the embedding $\{u_i \in \mathbb{R}\}_{i \in V}$ be defined as $u_i = \sqrt{\mu(\bar{C})/\mu(C)}$ for $i \in C$ and $u_i = -\sqrt{\mu(C)/\mu(\bar{C})}$ for $i \notin C$. Then $u_{\mathsf{avg}} = 0$ and $\mathbb{E}_{i \sim \mu} \|u_i - u_{\mathsf{avg}}\|_2^2 = 1$. Moreover, $\mathbb{E}_{\{i,j\} \in E} \|u_i - u_j\|_2^2 = 1/m \cdot |E(C, \bar{C})|/\mu(C)\mu(\bar{C}) \leqslant 4 \cdot \phi(C) \leqslant \gamma/4$. Let U be the Gram matrix of $\{u_i \in \mathbb{R}\}_{i \in V}$.

We apply the lower bound of Equation 2.2 to U. By Facts 2.1 and 2.2,

$$\mathbb{E}_{\{i,j\} \in E} \|u_i - u_j\|_2^2 + \sum_{i \in V} \bar{\beta}_i \|u_i - u_{\mathsf{avg}}\|_2^2 - \bar{\alpha} \mathbb{E}_{i \sim \mu} \|u_i - u_{\mathsf{avg}}\|_2^2$$

$$= M(\bar{\alpha}, \bar{\beta}) \bullet U \geqslant 0$$

Recall that, by the definition of ORACLE, for all $t \in [T]$, $\alpha^{(t)} \geqslant 7/8 \cdot \gamma$ and $\beta_i^{(t)} = \mu_i \cdot \gamma$ for $i \in B^{(t)}$ and $\beta_i^{(t)} = 0$ for $i \notin B^{(t)}$. Hence,

$$\gamma/4 + \gamma/T \cdot \sum_{t=1}^{T} \left(\mu(B^{(t)} \cap C) \cdot \mu(\bar{C})/\mu(C) + \mu(B^{(t)} \cap \bar{C}) \cdot \mu(C)/\mu(\bar{C}) \right)$$

$$-7/8 \cdot \gamma \geqslant 0$$

Dividing by γ and using the fact that $\mu(C) \leqslant 1/2$ and $\mu(\bar{C}) \leqslant 1$, we obtain

$$1/T \cdot \sum_{t=1}^{T} \left(\frac{\mu(B^{(t)} \cap C)}{\mu(C)} + \frac{\mu(B^{(t)} \cap \bar{C})}{2 \cdot \mu(\bar{C})} \right) \geqslant (7/8 - 1/4) = 5/8.$$

Now,

$$\frac{\mu(S \cap C)}{\mu(C)} + \frac{\mu(S \cap \bar{C})}{2 \cdot \mu(\bar{C})} \geqslant 1/T \cdot \sum_{t=1}^{T} \left(\frac{\mu(B^{(t)} \cap C)}{\mu(C)} + \frac{\mu(B^{(t)} \cap \bar{C})}{2 \cdot \mu(\bar{C})} \right),$$

so that we have

$$\frac{\mu(S \cap C)}{\mu(C)} + \frac{\mu(S \cap \bar{C})}{2 \cdot \mu(\bar{C})} \geqslant 5/8.$$

Moreover, being the union of cuts of conductance $O(\sqrt{\gamma})$, S also has $\phi(S) \leqslant O(\sqrt{\gamma})$. As $\mu(S) \leqslant b/4$, $\mu(S \cap \bar{C})/2 \cdot \mu(\bar{C}) \leqslant \mu(S) \leqslant \frac{b}{4} \leqslant 1/8$. This finally implies that

$$\frac{\mu(S \cap C)}{\mu(C)} \geqslant 1/2.$$

Finally, both PROJROUND and ORACLE run in time $\tilde{O}(m)$ as the embedding is $O(\log n)$ dimensional. By Lemma 2.3, the update at time t can be performed in time t_M, where $M = \sum_{i=1}^{t-1} P^{(i)}$. This is a matrix of the form $a \cdot L + \sum_{i \in V} b_i R_i + c L(K_V)$. The first two terms can be multiplied by a $O(\log n)$ vectors in time $\tilde{O}(m)$, while the third term can be decomposed as $L(K_V) = \mathbb{E}_{i \sim \mu} R_i$ by Fact 2.1 and can therefore be also multiplied in time $\tilde{O}(m)$. Hence, each iteration runs in time $\tilde{O}(m)$, which shows that the total running time is $\tilde{O}(m/\gamma)$ as required.

3 Proof of Main Theorem on ORACLE

3.1 Preliminaries The following is a variant of the sweep cut argument of Cheeger's Inequality [5], tailored to ensure that a constant fraction of the variance of the embedding is contained inside the output cut. For a vector $x \in \mathbb{R}^V$, let $\mathsf{supp}(x)$ be the set of vertices where x is not zero.

LEMMA 3.1. *Let* $x \in \mathbb{R}^V, x \geqslant 0$, *such that* $x^T L x \leqslant \lambda$ *and* $\mu(\mathsf{supp}(x)) \leqslant 1/2$. *Relabel the vertices so that* $x_1 \geqslant x_2 \geqslant \ldots \geqslant x_{z-1} > 0$ *and* $x_z = \ldots = x_n = 0$. *For* $i \in [z - 1]$, *denote by* $S_i \subseteq V$, *the sweep cut* $\{1, 2, \ldots, i\}$. *Further, assume that* $\sum_{i=1}^{n} d_i x_i^2 \leqslant 1$, *and, for some fixed* $k \in [z-1]$, $\sum_{i=k}^{n} d_i x_i^2 \geqslant \sigma$. *Then, there is a sweep cut* S_h *of* x *such that* $z - 1 \geqslant h \geqslant k$ *and* $\phi(S_h) \leqslant 1/\sigma \cdot \sqrt{2\lambda}$.

We will also need the following simple fact.

FACT 3.1. *Given* $v, u, t \in \mathbb{R}^d$, $(\|v - t\|_2 - \|u - t\|_2)^2 \leqslant \|v - u\|_2^2$.

7

3.2 Proof of Theorem 2.3

Proof. Notice that, by Markov's Inequality, $\mu(\bar{R}) \leqslant {}^b/(32\cdot(1-b)) \leqslant {}^b/16$. Recall that $\tau = O\left({}^1/\text{poly}(n)\right)$.

- CASE 1: $\mathbb{E}_{\{i,j\}\in E} \|\tilde{v}_i - \tilde{v}_j\|_2^2 = \frac{1}{2m} \cdot L \bullet \tilde{X} \geqslant 2\gamma$. We have $V(\alpha,\beta) \geqslant \gamma$ and, by Lemma 2.3,

$$M(\alpha,\beta) \bullet X \geqslant (1 - {}^1/64) \cdot 2\gamma - \gamma - \tau \geqslant {}^1/64 \cdot \gamma.$$

- CASE 2: $\mathbb{E}_{\{i,j\}\sim\mu_R\times\mu_R}\|v_i - v_j\|_2^2 \geqslant {}^1/64$. Then $\{\tilde{v}_i\}_{i\in V}$ is *roundable* by Definition 2.1.

- CASE 3: $\mathbb{E}_{\{i,j\}\sim\mu_R\times\mu_R}\|v_i - v_j\|_2^2 < {}^1/64$. This means that, by Fact 2.3, $L(K_R) \bullet \tilde{X} < {}^1/2 \cdot \mu(R)^2 \cdot {}^1/64 < {}^1/128$. Hence, by Fact 2.2,

$$\sum_{i\in\bar{R}} \mu_i R_i \bullet \tilde{X} = \sum_{i\in\bar{R}} \mu_i r_i \geqslant \mu(R) - {}^1/128 \geqslant 1 - {}^1/32 - {}^1/128$$

$$\geqslant 1 - {}^5/128.$$

We then have $\bar{R} = S_g$ for some $g \in [n]$, where we also denote by S_g the g largest coordinates dictated by the sweep cut S_g. Let $k \leqslant z$ be the the vertex in R such that $\sum_{j=1}^{k} \mu_j r_j \geqslant (1 - {}^1/128) \cdot (1 - {}^5/128)$ and $\sum_{j=k}^{g} \mu_j r_j \geqslant {}^1/128 \cdot (1 - {}^5/128)$. By the definition of z, we have $k \leqslant g < z$ and $r_z^2 \leqslant {}^8/b \leqslant 16 \cdot {}^{(1-b)}/b$. Hence, we have $r_z \leqslant {}^1/2 \cdot r_i$, for all $i \geqslant g$. Define the vector x as $x_i \overset{\text{def}}{=} {}^1/2m \cdot (r_i - r_z)$ for $i \in S_z$ and $r_i \overset{\text{def}}{=} 0$ for $i \notin S_z$. Notice that:

$$x^T L x = \sum_{\{i,j\}\in E} (x_i - x_j)^2 \leqslant {}^1/2m \cdot \sum_{\{i,j\}\in E} (r_i - r_j)^2$$
$$\overset{\text{Fact 3.1}}{\leqslant} {}^1/2m \cdot \sum_{\{i,j\}\in E} \|\tilde{v}_i - \tilde{v}_j\|_2^2 \leqslant 2\gamma.$$

Also, $x \geqslant 0$ and $\mu(\text{supp}(x)) \leqslant {}^b/8 \leqslant {}^1/2$, by the definition of z. Moreover,

$$\sum_{i=1}^{n} d_i x_i^2 = {}^1/2m \cdot \sum_{i=1}^{z} d_i (r_i - r_z)^2 \leqslant {}^1/2m \cdot \sum_{i=1}^{z} d_i r_i^2 = 1,$$

and

$$\sum_{i=k}^{n} d_i x_i^2 = {}^1/2m \cdot \sum_{i=k}^{z} d_i (r_i - r_z)^2$$
$$\geqslant {}^1/2m \cdot \sum_{i=k}^{g} d_i (r_i - {}^1/2 \cdot r_i)^2$$
$$= {}^1/2m \cdot {}^1/4 \cdot \sum_{i=k}^{g} d_i r_i^2$$
$$\geqslant {}^1/512 \cdot (1 - {}^5/128) \geqslant {}^1/1024$$

Hence, by Lemma 3.1, there exists a sweep cut S_h with $z > h \geqslant k$, such that $\phi(S_h) \leqslant 2048 \cdot \sqrt{\gamma}$. This shows that

B, as defined in Figure 6 exists. Moreover, it must be the case that $S_h \subseteq B$. As $h \geqslant k$, we have

$$\sum_{i\in B} \mu_i r_i^2 \geqslant \sum_{i\in S_h} \mu_i r_i^2 \geqslant \sum_{i=1}^{k} \geqslant (1-{}^1/128)\cdot(1-{}^5/128) \geqslant 1-{}^3/64.$$

Recall also that, by the construction of z, $\mu(B) \leqslant {}^b/8$. Hence, we have

$$V(\alpha,\beta) = {}^7/8 \cdot \gamma - {}^{(1-b)}/b \cdot \mu(B) \cdot \gamma \geqslant ({}^7/8 - {}^1/8) \cdot \gamma \geqslant {}^3/4\gamma.$$

$$M(\alpha,\beta) \bullet X \geqslant (1 - {}^1/64) \cdot (1 - {}^3/64)\gamma - {}^7/8\gamma - \tau \geqslant {}^1/64 \cdot \gamma.$$

This completes all the three cases. Notice that in every case we have:

$${}^1/2m \cdot L - \gamma L(K_V) \leq M(\alpha,\beta) \leq {}^1/2m \cdot L + \gamma L(K_V).$$

Hence,

$$-\gamma L(K_V) \leq M(\alpha,\beta) \leq 3L(K_V).$$

Finally, using the fact that $\{\tilde{v}_i\}_{i\in V}$ is embedded in $O(\log n)$ dimensions, we can compute $L \bullet \tilde{X}$ in time $\tilde{O}(m)$. $L(K_R) \bullet \tilde{X}$ can also be computed in time $\tilde{O}(n)$ by using the decomposition $\mathbb{E}_{\{i,j\}\sim\mu_R\times\mu_R}\|v_i - v_j\|_2^2 = 2 \cdot \mathbb{E}_{i\sim\mu_R}\|v_i - v_{\text{avg}_R}\|_2^2$, where v_{avg_R} is the mean of vectors representing vertices in R. The sweep cut over r takes time $\tilde{O}(m)$. Hence, the total running time is $\tilde{O}(m)$.

References

[1] Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. Local graph partitioning using pagerank vectors. In *FOCS'06: Proc. 47th Ann. IEEE Symp. Foundations of Computer Science*, pages 475–486, 2006.

[2] Reid Andersen and Yuval Peres. Finding sparse cuts locally using evolving sets. In *STOC '09: Proc. 41st Ann. ACM Symp. Theory of Computing*, pages 235–244, 2009.

[3] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *STOC '07: Proc. 39th Ann. ACM Symp. Theory of Computing*, pages 227–236, 2007.

[4] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. In *STOC '04: Proc. 36th Ann. ACM Symp. Theory of Computing*, pages 222–231, New York, NY, USA, 2004. ACM.

[5] Fan R.K. Chung. *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. American Mathematical Society, 1997.

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

[7] Stephen Guattery and Gary L. Miller. On the performance of spectral graph partitioning methods. In *SODA'95: Proc. 6th Ann. ACM-SIAM Symp. Discrete Algorithms*, pages 233–242, 1995.

[8] G. Iyengar, David J. Phillips, and Clifford Stein. Approximation algorithms for semidefinite packing problems with applications to maxcut and graph coloring. In *IPCO'05: Proc. 11th Conf. Integer Programming and Combinatorial Optimization*, pages 152–166, 2005.

[9] Satyen Kale. Efficient algorithms using the multiplicative weights update method. Technical report, Princeton University, Department of Computer Science, 2007.

[10] R. Kannan, S. Vempala, and A. Vetta. On clusterings-good, bad and spectral. In *FOCS'00: Proc. 41th Ann. IEEE Symp. Foundations of Computer Science*, page 367, Washington, DC, USA, 2000. IEEE Computer Society.

[11] Ioannis Koutis and Gary L. Miller. Graph partitioning into isolated, high conductance clusters: theory, computation and applications to preconditioning. In *SPAA'08: Proc. 20th ACM Symp. Parallelism in Algorithms and Architectures*, pages 137–145, 2008.

[12] Ioannis Koutis, Gary L. Miller, and Richard Peng. Approaching optimality for solving sdd systems. *CoRR*, abs/1003.2958, 2010.

[13] László Lovász and Miklós Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Struct. Algorithms*, 4(4):359–412, 1993.

[14] Daniel A. Spielman. Algorithms, graph theory, and linear equations in laplacian matrices. In *ICM'10: Proc. International Congress of Mathematicians*, 2010.

[15] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. In *STOC '08: Proc. 40th Ann. ACM Symp. Theory of Computing*, pages 563–568, 2008.

[16] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *STOC '04: Proc. 36th Ann. ACM Symp. Theory of Computing*, pages 81–90, New York, NY, USA, 2004. ACM.

[17] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *CoRR*, abs/cs/0607105, 2006.

[18] Daniel A. Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *CoRR*, abs/0809.3232, 2008.

[19] David Steurer. Fast SDP algorithms for constraints satisfaction problems. In *SODA'10: Proc. 21st Ann. ACM-SIAM Symp. Discrete Algorithms*, 2010.

[20] Luca Trevisan. Approximation algorithms for unique games. In *In Proc. of the 46th IEEE Symposium on Foundations of Computer Science*, pages 05–34. IEEE Computer Society, 2005.

A Appendix

A.1 Proof of Corollary 1.1

LEMMA A.1. *If only the second condition in Theorem 1.1 holds, then G has no $\Omega_b(1)$-balanced cut of conductance $O(\gamma)$.*

Proof. We may assume that the cut S output by BALCUT is not $\Omega_b(1)$-balanced. Then, by the second condition, any cut T with $\mathrm{vol}(T) \leqslant 1/2 \cdot \mathrm{vol}(G)$ and $\phi(T) \leqslant O(\gamma)$ must have $\mathrm{vol}(S \cap T)/\mathrm{vol}(T) \geqslant 1/2$. Hence, $\mathrm{vol}(T) \leqslant 2 \cdot \mathrm{vol}(S \cap T) \leqslant 2 \cdot \mathrm{vol}(S) \leqslant \Omega_b(1)$. This implies that there are no $\Omega_b(1)$-balanced cuts of conductance less than $O(\gamma)$.

A.2 Proof of Basic Lemmata

Proof. [Proof of Lemma 2.1] For a b-balanced cut $(S, \bar{S})$ with $\phi(S) \leqslant \gamma$. Without loss of generality, assume $\mu(S) \leqslant 1/2$. Consider the one-dimensional solution assigning $v_i = \sqrt{\mu(\bar{S})/\mu(S)}$ to $i \in S$ and $v_i = -\sqrt{\mu(S)/\mu(\bar{S})}$ to $i \in \bar{S}$. Notice that $v_{\mathrm{avg}} = 0$ and that $\|v_i - v_j\|_2^2 = 1/\mu(S)\mu(\bar{S})$ for $i \in S, j \notin S$. We then have:

–
$$\mathop{\mathbb{E}}_{\{i,j\}\in E} \|v_i - v_j\|_2^2 = \frac{1}{m} \cdot \frac{|E(S, \bar{S})|}{\mu(S)\mu(\bar{S})} \leqslant 2 \cdot \frac{|E(S, \bar{S})|}{2m \cdot \mu(S)\mu(\bar{S})}$$
$$\leqslant 4 \cdot \phi(S) \leqslant 4\gamma.$$

–
$$\mathop{\mathbb{E}}_{i \sim \mu} \|v_i - v_{\mathrm{avg}}\|_2^2 = \mu(S) \cdot \mu(\bar{S})/\mu(S) + \mu(\bar{S}) \cdot \mu(S)/\mu(\bar{S}) = 1.$$

– for all $i \in V$,
$$\|v_i - v_{\mathrm{avg}}\|_2^2 \leqslant \frac{\mu(\bar{S})}{\mu(S)} \leqslant \frac{1 - b}{b},$$

where the last inequality follows as S is b-balanced.

Proof. [Proof of Lemma 3.1] For all $i \in V$, let $y_i = x_i^2$. By Cauchy-Schwarz,

$$\sum_{\{i,j\}\in E} |y_i - y_j| = \sum_{\{i,j\}\in E,\, x_i \geqslant x_j} (x_i - x_j)(x_i + x_j)$$
$$\leqslant \sqrt{\left(\sum_{\{i,j\}\in E} (x_i - x_j)^2\right) \cdot \left(\sum_{\{i,j\}\in E} (x_i + x_j)^2\right)}.$$

Hence,

$$\sum_{\{i,j\}\in E} |y_i - y_j| \leqslant \sqrt{\lambda \cdot 2 \left(\sum_{i \in V} d_i x_i^2\right)} \leqslant \sqrt{2\lambda}.$$

Then, let ϕ be the conductance of the least conductance cut among $S_{k+1}, S_{k+2}, \ldots, S_h$.

$$\sum_{\{i,j\}\in E} |y_i - y_j|$$
$$\geqslant \sum_{i=k+1}^{h} |E(S_i, \bar{S}_i)| \cdot (y_i - y_{i+1}) \geqslant \phi \sum_{i=k+1}^{h} \mathrm{vol}(S_i) \cdot (y_i - y_{i+1})$$
$$\geqslant \phi \sum_{i=k+1}^{h} d_i y_i = \sum_{i=k+1}^{h} d_i x_i^2 = \phi \sum_{i \in V - S_k} d_i x_i^2 = \phi\sigma.$$

Hence, $\phi \leqslant 1/\sigma \cdot \sqrt{2\lambda}$.

A.3 Primal-Dual Framework

A.3.1 Preliminaries Recall that $\mathcal{S}_D$ is the subspace of $\mathbb{R}^V$ orthogonal to $\hat{v} = 1/\sqrt{2m} \cdot D^{1/2} 1$ and that $\mathcal{I}$ be the identity over $\mathcal{S}_D$.

Define

$$E_\varepsilon(Y) \stackrel{\text{def}}{=} \frac{\exp(-\varepsilon Y)}{\mathcal{I} \bullet \exp(-\varepsilon Y)}.$$

The following simple facts will be useful.

Fact A.1. *Let M be a symmetric matrix in $\mathbb{R}^{V \times V}$ such that $M\hat{v} = 0$. Then, $\exp(M)\hat{v} = \hat{v}$.*

Theorem A.1. *Let $\varepsilon > 0$ and let $Y^{(1)}, \ldots, Y^{(T)}$ be a sequence of symmetric matrices in $\mathbb{R}^{V \times V}$ such that, for all $i \in [T]$, $Y_i\hat{v} = 0$ and $0 \leq Y_i \leq I$. Then:*

$$\sum_{y=1}^{T} Y^{(t)} \succeq \left((1-\varepsilon) \sum_{t=1}^{T} Y^{(t)} \bullet Z^{(t)} - \frac{\log n}{\varepsilon} \right) \mathcal{I},$$

where $Z^{t+1} = E_\varepsilon(\sum_{i=1}^{t} Y^{(i)})$.

A.3.2 Proofs In the following, let $M^{(t)} \stackrel{\text{def}}{=} 2m \cdot D^{-1/2} P^{(t)} D^{-1/2}$. Also, let $W^{(t)} = E_\varepsilon\left(\sum_{i=1}^{t-1} M^{(i)}\right)$. Then, we have

$$X^{(t)} = U_\varepsilon\left(\sum_{i=1}^{t-1} P^{(i)}\right) = 2m \cdot D^{-1/2} E_\varepsilon\left(\sum_{i=1}^{t-1} M^{(i)}\right) D^{-1/2}$$
$$= 2m \cdot D^{-1/2} W^{(t)} D^{-1/2}.$$

Proof. [Proof of Lemma 2.2] By Fact 2.4,

$$L(K_V) \bullet X^{(t)} = \left(D^{1/2} \mathcal{I} D^{1/2}\right) \bullet \left(D^{-1/2} W^{(t)} D^{-1/2}\right) = \mathcal{I} \bullet W^{(t)} = 1.$$

We are now ready to complete the proof of Theorem 2.2.

Proof. [Proof of Theorem 2.2] Suppose that Oracle outputs coefficents $\left(\alpha^{(t)}, \beta^{(t)}\right)$ for T iterations. Then for all $t \in [T]$, by the definition of *good* oracle $W^{(t)} \bullet M^{(t)}$

$$= \left(1/2m \cdot D^{1/2} X^{(t)} D^{1/2}\right)) \bullet \left(2m \cdot D^{-1/2} P^{(t)} D^{-1/2}\right) = X^{(t)} \bullet P^{(t)}$$
$$\geq \frac{1}{6} \cdot \left(M\left(\alpha^{(t)}, \beta^{(t)}\right) \bullet X^{(t)} + \gamma L(K_V) \bullet X^{(t)}\right)$$
$$\geq 1/6 \cdot (1/64 \cdot \gamma + \gamma) = 1/6 \cdot 65/64 \cdot \gamma.$$

Notice that, as $\left(\alpha^{(t)}, \beta^{(t)}\right)$ are the output of a *good* oracle, we have $0 \leq P^{(t)} \leq L(K_V)$, which implies $0 \leq M^{(t)} \leq \mathcal{I} \leq I$. Hence, we can apply Theorem A.1 to obtain:

$$\sum_{i=1}^{T} M^{(t)} \succeq \left((1-\varepsilon) \sum_{t=1}^{T} M^{(t)} \bullet W^{(t)} - \frac{\log n}{\varepsilon} \right) \mathcal{I}.$$

This implies

$$\sum_{i=1}^{T} P^{(t)} \succeq \left((1-\varepsilon) \cdot T/6 \cdot 65/64 \cdot \gamma - \frac{\log n}{\varepsilon} \right) L(K_V),$$

and by the definition of $P^{(t)}$,

$$\frac{1}{6} \cdot \sum_{i=1}^{T} \left(M\left(\alpha^{(t)}, \beta^{(t)}\right) + \gamma L(K_V)\right)$$
$$\geq \left((1-\varepsilon) \cdot T/6 \cdot 65/64 \cdot \gamma - \frac{\log n}{\varepsilon} \right) L(K_V).$$

Hence,

$$M(\bar{\alpha}, \bar{\beta}) \geq 1/T \cdot \left((1-\varepsilon) \cdot T \cdot 65/64 \cdot \gamma - \frac{6 \log n}{\varepsilon} - \gamma \cdot T \right) L(K_V)$$

and

$$M(\bar{\alpha}, \bar{\beta}) \geq \left((1-\varepsilon) \cdot 65/64 \cdot \gamma - \frac{6 \log n}{\varepsilon T} - \gamma \right) L(K_V).$$

By picking $\varepsilon \stackrel{\text{def}}{=} 1/130$ and $T \stackrel{\text{def}}{=} 6 \cdot 129 \cdot 1/\varepsilon \cdot \log n/\gamma = O\left(\log n/\gamma\right)$, we obtain $M(\bar{\alpha}, \bar{\beta}) \geq 0 \cdot L(K_V)$. As $M(\bar{\alpha}, \bar{\beta}) 1 = 0$, this also implies

$$M(\bar{\alpha}, \bar{\beta}) \succeq 0.$$

Finally, by the definition of *good* oracle, $V(\bar{\alpha}, \bar{\beta}) \geq 3/4 \cdot \gamma$. Hence, $(\bar{\alpha}, \bar{\beta})$ is a solution to $\mathsf{dsdp}(G, b, 3/16 \cdot \gamma)$.

A.4 Projection Rounding The description of the rounding algorithm ProjRound is given in Figure 7. We remark that during the execution of BalCut the embedding $\{v_i \in \mathbb{R}^d\}_{i \in V}$ will be represented by a projection over $d = O(\log n)$ random directions, so that it will suffice to take a balanced sweep cut of each coordinate vector.

We now present the proof of Theorem 2.1. The constants in this argument were not optimized to preserve the simplicity of the proof.

A.4.1 Preliminaries. We will make use of the following simple facts. Recall that for $y \in \mathbb{R}$, $\text{sgn}(y) = 1$ if $y \geq 0$, and -1 otherwise.

Fact A.2. *For all $y, z \in \mathbb{R}$, $(y + z)^2 \leq 2(y^2 + z^2)$.*

Proof. $2(y^2 + z^2) - (y + z)^2 = (y - z)^2 \geq 0$.

Fact A.3. *For all $y \geq z \in \mathbb{R}$, $\left|\text{sgn}(y) \cdot y^2 - \text{sgn}(z) \cdot z^2\right| \leq (y - z)(|y| + |z|)$.*

Proof.

1. If $\text{sgn}(y) = \text{sgn}(z)$, then $|\text{sgn}(y) \cdot y^2 - \text{sgn}(z) \cdot z^2| = |y^2 - z^2| = (y - z) \cdot |y + z| = (y - z)(|y| + |z|)$ as $y \geq z$.

2. If $\text{sgn}(y) \neq \text{sgn}(y)$, then since $y \geq z$, $(y - z) = |y| + |z|$. Hence, $|\text{sgn}(y) \cdot y^2 - \text{sgn}(z) \cdot z^2| = y^2 + z^2 \leq (|y| + |z|)^2 = (y - z)(|y| + |z|)$.

Fact A.4. *For all $y \geq z \in \mathbb{R}$, $(y - z)^2 \leq 2(\text{sgn}(y) \cdot y^2 - \text{sgn}(z) \cdot z^2)$.*

Proof.

1. If $\mathrm{sgn}(y) = \mathrm{sgn}(z)$, $(y - z)^2 = y^2 + z^2 - 2yz \leqslant y^2 + z^2 - 2z^2 = y^2 - z^2$ as $y \geqslant z$. Since $\mathrm{sgn}(y) = \mathrm{sgn}(z)$, $y^2 - z^2 \leqslant 2(\mathrm{sgn}(y) \cdot y^2 - \mathrm{sgn}(z) \cdot z^2)$.

2. If $\mathrm{sgn}(y) \neq \mathrm{sgn}(z)$, $(y - z)^2 = (|y| + |z|)^2 \leqslant 2(|y|^2 + |z|^2) = 2(\mathrm{sgn}(y) \cdot y^2 - \mathrm{sgn}(z) \cdot z^2)$. Here, we have used Fact A.2.

We also need the following standard facts.

FACT A.5. *Let $v \in \mathbb{R}^d$ be a vector of length ℓ and u a unit vector chosen uniformly at random in $\mathbf{S}^{t-1}$. Then,*

1. $\mathbb{E}_u \left(v^T u \right)^2 = \frac{\ell^2}{d}$, *and*

2. *for $0 \leqslant \delta \leqslant 1$, $\mathbb{P}_u \left[\sqrt{d} \cdot |v^T u| \leqslant \delta \ell \right] \leqslant 3\delta$.*

FACT A.6. *Let Y be a non-negative random variable such that $\mathbb{P}[Y \leqslant K] = 1$ and $\mathbb{E}[Y] \geqslant \delta$. Then,*

$$\mathbb{P}[Y \geqslant \delta/2] \geqslant \frac{\delta}{2K}.$$

The following lemma about projections will be crucial in the proof of Theorem 2.1. It is a simple adaptation of an argument appearing in [4].

LEMMA A.2. (PROJECTION) *Given a roundable embedding $\{v_i \in \mathbb{R}^d\}_{i \in V}$, consider the embedding $x \in \mathbb{R}^n$ such that $x_i \stackrel{\mathrm{def}}{=} \sqrt{d} \cdot u^T v_i$, where $u \in \mathbf{S}^{d-1}$, and assume without loss of generality that $x_1 \geqslant \ldots \geqslant x_n$. Then, there exists $c \in (0, b]$ such that with probability $\Omega_b(1)$ over the choice of $u \in \mathbf{S}^{d-1}$, the following conditions hold simultaneously:*

1. $\mathbb{E}_{\{i,j\} \in E}(x_i - x_j)^2 \leqslant O_b \left(\mathbb{E}_{\{i,j\} \in E} \|v_i - v_j\|^2 \right) = O_b(\gamma)$,

2. $\mathbb{E}_{i \sim \mu}(x_i - x_{\mathsf{avg}})^2 = O_b(1)$, *and*

3. *there exists $1 \leqslant l \leqslant n$ with $\mathrm{vol}(\{1, \ldots, l\}) \geqslant c \cdot \mathrm{vol}(G)$ and, there exists $l \leqslant r \leqslant n$ such that $\mathrm{vol}(\{r, \ldots, n\}) \geqslant c \cdot \mathrm{vol}(G)$ such that $x_l - x_r \geqslant \Omega_b(1)$.*

Proof. We are going to lower bound the probability, over u, of each of (1), (2) and (3) in the lemma and then apply the union bound.

Part (1). By applying Fact A.5 to $v = v_i - v_j$ and noticing $\sqrt{d} \cdot |v^T u| = |x_i - x_j|$, we have

$$\mathbb{E}_u \mathbb{E}_{\{i,j\} \in E}(x_i - x_j)^2 = \mathbb{E}_{\{i,j\} \in E} \|v_i - v_j\|^2.$$

Hence, by Markov's Inequality, for some p_1 to be fixed later

$$\mathbb{P}_u \left[\mathbb{E}_{\{i,j\} \in E}(x_i - x_j)^2 \geqslant 1/p_1 \cdot \mathbb{E}_{\{i,j\} \in E} \|v_i - v_j\|^2 \right] \leqslant p_1.$$

Part (2).

$$\mathbb{E}_u \mathbb{E}_{i \sim \mu}(x_i - x_{\mathsf{avg}})^2 \stackrel{\text{Fact A.5−(1)}}{=} \mathbb{E}_u \mathbb{E}_{i \sim \mu} \|v_i - v_{\mathsf{avg}}\|^2 \stackrel{\text{roundability}}{=} 1.$$

Hence, for some p_2 be fixed later

$$\mathbb{P}_u \left[\mathbb{E}_{i \sim \mu}(x_i - x_{\mathsf{avg}})^2 \geqslant 1/p_2 \cdot \mathbb{E}_{i \sim \mu} \|v_i - v_{\mathsf{avg}}\|^2 \right] \leqslant p_2.$$

Part (3). Let $R \stackrel{\mathrm{def}}{=} \{i \in V : \|v_i - v_{\mathsf{avg}}\|^2 \leqslant 32 \cdot (1-b)/b\}$. Let $\sigma \stackrel{\mathrm{def}}{=} 4 \cdot \sqrt{2} \sqrt{(1-b)/b}$. By Markov's Inequality, $\mu(\bar{R}) \leqslant 1/\sigma^2$. As $\{v_i\}_{i \in V}$ is roundable, for all $i, j \in R$, $\|v_i - v_j\| \leqslant 2\sigma$. Hence, $\|v_i - v_j\| \geqslant 1/2\sigma \cdot \|v_i - v_j\|^2$ for such $i, j \in R$. This, together with the roundability of $\{v_i\}_{i \in V}$, implies that

$$\mathbb{E}_{\{i,j\} \sim \mu_R \times \mu_R} \|v_i - v_j\| \geqslant 1/128\sigma.$$

For any $k \in R$, we can apply the triangle inequality for the Euclidean norm as follows

$$\mathbb{E}_{\{i,j\} \sim \mu_R \times \mu_R} \|v_i - v_j\| \leqslant \mathbb{E}_{\{i,j\} \sim \mu_R \times \mu_R} \left(\|v_i - v_k\| + \|v_k - v_j\| \right)$$
$$\leqslant 2 \cdot \mathbb{E}_{i \sim \mu_R} \|v_i - v_k\|.$$

Hence, for all $k \in R$

$$\mathbb{E}_{i \sim \mu_R} \|v_i - v_k\| \geqslant 1/256\sigma.$$

Let R_k be the set $\{i \in R : \|v_i - v_k\| \geqslant 1/512\sigma\}$. Since $\|v_i - v_k\| \leqslant 2\sigma$, applying Fact A.6 yields that, for all $k \in R$,

$$\mathbb{P}_{i \sim \mu_R} [i \in R_k] \geqslant 1/1024\sigma^2.$$

For all vertices $i \in R_k$, by Fact A.5

$$\mathbb{P}_u [|x_k - x_i| \geqslant 1/9 \cdot 1/512\sigma = 1/4608\sigma] \geqslant \frac{2}{3}.$$

Let $\delta \stackrel{\mathrm{def}}{=} 1/2 \cdot 1/4608\sigma = 1/9216\sigma$. Consider the event $\mathcal{E} \stackrel{\mathrm{def}}{=} \{i \in R_k \wedge |x_i - x_k| \geqslant 2 \cdot \delta\}$. Then,

$$\mathbb{P}_{u, \{i,k\} \sim \mu_R \times \mu_R} [\mathcal{E}] = \mathbb{P}_{\{i,k\} \sim \mu_R \times \mu_R} [i \in R_k] \cdot \mathbb{P}_u[|x_i - x_k| \geqslant 2 \cdot \delta \mid i \in R_k]$$
$$\geqslant \frac{1}{1024\sigma^2} \cdot \frac{2}{3} = \frac{1}{1536\sigma^2} \stackrel{\mathrm{def}}{=} \rho.$$

Hence, from Fact A.6, with probability at least $\rho/2$ over directions u, for a fraction $\rho/2$ of pairs $\{i, k\} \in R \times R$, $|x_k - x_i| \geqslant 2 \cdot \delta$. Let ν be the median value of $\{x_i\}_{i \in V}$. Let $L \stackrel{\mathrm{def}}{=} \{i : x_i \leqslant \nu - \delta\}$ and $H \stackrel{\mathrm{def}}{=} \{i : x_i \geqslant \nu + \delta\}$. Any pair $\{i, j\} \in R \times R$ with $|x_i - x_j| \geqslant 2 \cdot \delta$ has at least one vertex in $L \cup H$. Hence,

$$\mu(L \cup H) \geqslant 1/2 \cdot \rho/2 \cdot \mu(R)^2 \geqslant \rho/4 \cdot (\sigma^2 - 1/\sigma^2)^2 \geqslant \rho/16 = \Omega_b(1).$$

Assume $\mu(L) \geqslant \rho/32$, otherwise, apply the same argument to H. Let l be the largest index in L. For all $i \in L$ and j such that

11

$x_j \geqslant \nu$, we have $|x_i - x_j| \geqslant \delta$. (Similarly, let r be the smallest index in H.) This implies that,

$$|x_l - x_{\lfloor n/2 \rfloor}| \geqslant \delta$$

with probability at least $\rho/2 = \Omega_b(1)$, satisfying the required condition. Let p_3 be the probability that this event does not take place. Then,

$$p_3 \leqslant 1 - \rho/2.$$

To conclude the proof, notice that the probability that all three conditions do not hold simultaneously is, by a union bound, at most $p_1 + p_2 + p_3$. Setting $p_1 = p_2 = \rho/5 = \Omega_b(1)$, we satisfy the first and third conditions and obtain

$$p_1 + p_2 + p_3 \leqslant 1 - \rho \cdot (1/2 - 1/5 - 1/5) \leqslant 1 - \rho/10.$$

Hence, all conditions are satisfied at the same time with probability at least $\rho/10 = \Omega_b(1)$.

From this proof, it is possible to see that the parameter c in our rounding scheme should be set to $\rho/32$.

We are now ready to give a proof of Theorem 2.1. It is essentially a variation of the proof of Cheeger's Inequality, tailored to produce balanced cuts.

Proof. [Proof of Theorem 2.1] For this proof, assume that x has been translated so that $x_{\mathsf{avg}} = 0$. Notice that the guarantees of $A.2$ still apply. Let x, l, r and c be as promised by Lemma A.2. For $z \in \mathbb{R}$, let $\mathrm{sgn}(z)$ be 1 if $z \geqslant 0$ and -1 otherwise. Let

$$y_i \overset{\text{def}}{=} \mathrm{sgn}(x_i) \cdot x_i^2.$$

Hence,

$$
\underset{\{i,j\} \in E}{\mathbb{E}} |y_i - y_j| \overset{\text{Fact A.3}}{\leqslant} \underset{\{i,j\} \in E}{\mathbb{E}} (|x_i - x_j|) \cdot (|x_i| + |x_j|)
$$

$$
\leqslant \sqrt{\underset{\{i,j\} \in E}{\mathbb{E}} (x_i - x_j)^2 \cdot \underset{\{i,j\} \in E}{\mathbb{E}} (|x_i| + |x_j|)^2}
$$

$$
\overset{\text{Fact A.2}}{\leqslant} \sqrt{2 \cdot \underset{\{i,j\} \in E}{\mathbb{E}} (x_i - x_j)^2 \cdot \underset{\{i,j\} \in E}{\mathbb{E}} (x_i^2 + x_j^2)}
$$

$$
= \sqrt{2 \cdot \underset{\{i,j\} \in E}{\mathbb{E}} (x_i - x_j)^2 \cdot \frac{2m}{m} \cdot \underset{i \sim \mu}{\mathbb{E}} x_i^2}
$$

$$
= \sqrt{4 \cdot \underset{\{i,j\} \in E}{\mathbb{E}} (x_i - x_j)^2 \cdot \underset{i \sim \mu}{\mathbb{E}} x_i^2}
$$

$$
\overset{\text{Lemma A.2-(1),(2)}}{\leqslant} O_b\left(\sqrt{\gamma}\right).
$$

Now we lower bound $\mathbb{E}_{\{i,j\} \in E} |y_i - y_j|$. Notice that if $x_i \geqslant x_j$, then $y_i \geqslant y_j$ and vice-versa. Hence,

$$y_1 \geqslant \ldots \geqslant y_n.$$

Let $S_i \overset{\text{def}}{=} \{1, \ldots, i\}$ and let ϕ be the minimum conductance of S_i over all $l \leqslant i \leqslant .r$

$$
\underset{\{i,j\} \in E}{\mathbb{E}} |y_i - y_j| = \frac{1}{|E|} \sum_{i=1}^{n-1} |E(S_i, \bar{S}_i)| \cdot (y_i - y_{i+1})
$$

$$
\geqslant \phi \cdot \sum_{l \leqslant i \leqslant r} \frac{\min\{\mathrm{vol}(S_i), \mathrm{vol}(\bar{S}_i)\}}{|E|} (y_i - y_{i+1})
$$

$$
\overset{\text{Lemma A.2-(3)}}{=} \Omega_b(1) \cdot \phi \cdot \sum_{l \leqslant i \leqslant r} (y_i - y_{i+1})
$$

$$
\geqslant \Omega_b(1) \cdot \phi \cdot (y_l - y_r)
$$

$$
\overset{\text{Fact A.4}}{\geqslant} \Omega_b(1) \cdot \phi \cdot (x_l - x_r)^2
$$

$$
\overset{\text{Lemma A.2}}{\geqslant} \Omega_b(1) \cdot \phi.
$$

Hence, $\phi \leqslant O_b(\sqrt{\gamma})$ with constant probability over the choice of projection vectors u. Repeating the projection $O(\log n)$ times and picking the best balanced cut found yields a high probability statement. Finally, as the embedding is in d dimensions, it takes $\tilde{O}(nd)$ time to compute the projection. After that, the one-dimensional embedding can be sorted in time $\tilde{O}(n)$ and the conductance of the relevant sweep cuts can be computed in time $O(m)$, so that the total running time is $\tilde{O}(nd + m)$.

A.5 Proof of Lemma 2.3

A.5.1 Preliminaries
For the rest of this section the norm notation will mean the norm in the subspace $\mathcal{S}_D$. Hence $\|A\| = \|\mathcal{I} A \mathcal{I}\|$. We will need the following lemmata.

LEMMA A.3. (JOHNSON-LINDENSTRAUSS) *Given an embedding* $\{v_i \in \mathbb{R}^n\}_{i \in V}$, $V = [n]$, *let* $u_1, u_2, \ldots, u_k$, *be vectors sampled independently uniformly from the* $n - 1$*-dimensional sphere of radius* $\sqrt{n/k}$. *Let* U *be the* $k \times t$ *matrix having the vector* u_i *as* i*-th row and let* $\tilde{v}_i \overset{\text{def}}{=} U v_i$. *Then, for* $k_\delta \overset{\text{def}}{=} O(\log n/\delta^2)$, *for all* $i, j \in V$

$$(1 - \delta) \cdot \|v_i - v_j\|^2 \leqslant \|\tilde{v}_i - \tilde{v}_j\|^2 \leqslant (1 + \delta) \cdot \|v_i - v_j\|^2$$

and

$$(1 - \delta) \cdot \|v_i\|^2 \leqslant \|\tilde{v}_i\|^2 \leqslant (1 + \delta) \cdot \|v_i\|^2.$$

LEMMA A.4. ([9]) *There exists an algorithm* EXPV *which, on input of a matrix* $A \in \mathbb{R}^{n \times n}$, *a vector* $u \in \mathbb{R}^n$ *and a parameter* η, *computes a vector* $v \in \mathbb{R}^n$, *such that* $\|v - e^{-A}u\| \leqslant \|e^{-A}\|_2 \cdot \eta$ *in time* $O(t_A \log^3(1/\eta))$.

The algorithm EXPV is described in [9] and [8] .

A.5.2 Proof
We define the $\tilde{U}_\varepsilon$ algorithm in Figure 8 and proceed to prove Lemma 2.3.

Proof. We verify that the conditions required hold.

– By construction, $\tilde{X} \geq 0$, as $\tilde{X} = 1/z \cdot B^T B$, and $L(K_V) \bullet \tilde{X} = 1$.

– $\tilde{X} = (1/\sqrt{z} \cdot B)^T (1/\sqrt{z} \cdot B)$ and B is a $k_\delta \times n$ matrix, with $k_\delta = O(\log n)$, by Lemma A.3.

– We perform $k_\delta = O(\log n)$ calls to the algorithm EXPV, each of which takes time $\tilde{O}(t_A) = \tilde{O}(t_M + n)$. Sampling the vectors $\{u_i\}_{1,\dots,k_\delta}$ and computing Z also requires $\tilde{O}(n)$ time. Hence, the total running time is $\tilde{O}(t_M + n)$.

– Let U be the $k_\delta \times n$ matrix having the sampled vectors $u_1, \dots, u_{k_\delta}$ as rows. Let $\{v_i\}_{i \in V}$ be the embedding corresponding to matrix $Y \stackrel{\text{def}}{=} D^{-1/2} e^{-A} D^{-1/2}$, i.e., v_i is the i-th column of $Y^{1/2}$. Notice that $X = Y/L(K_V) \bullet Y$. Define $\hat{v}_i \stackrel{\text{def}}{=} U v_i$ for all i and let $\hat{Y}$ be the Gram matrix corresponding to this embedding, i.e., $\hat{Y} \stackrel{\text{def}}{=} (Y^{1/2})^T U^T U (Y^{1/2})$. Also, let $\tilde{Y}$ be the Gram matrix corresponding to the embedding $\{\tilde{v}_i\}_{i \in V}$, i.e., $\tilde{Y} = B^T B$ and $\tilde{X} = \tilde{Y}/L(K_V) \bullet \tilde{Y}$. We will relate Y to $\hat{Y}$ and $\hat{Y}$ to $\tilde{Y}$ to complete the proof.

First, by Lemma A.3, applied to $\{v_i\}_{i \in V}$, with high probability, for all H

$$(1 - \delta) \cdot L(H) \bullet Y \leq L(H) \bullet \hat{Y} \leq (1 + \delta) \cdot L(H) \bullet Y$$

and for all $i \in V$

$$(1 - \delta) \cdot R_i \bullet Y \leq R_i \bullet \hat{Y} \leq (1 + \delta) \cdot R_i \bullet Y.$$

In particular, this implies that $(1 - \delta) \cdot \mathcal{I} \bullet Y \leq \mathcal{I} \bullet \hat{Y} \leq (1 + \delta) \cdot \mathcal{I} \bullet Y$. Hence,

$$\frac{1 - \delta}{1 + \delta} \cdot L(H) \bullet X \leq L(H) \bullet \hat{X} \leq \frac{1 + \delta}{1 - \delta} \cdot L(H) \bullet X$$

and for all i

$$\frac{1 - \delta}{1 + \delta} \cdot R_i \bullet X \leq R_i \bullet \hat{X} \leq \frac{1 + \delta}{1 - \delta} \cdot R_i \bullet X.$$

Now we relate $\hat{Y}$ and $\tilde{Y}$. Let $E \stackrel{\text{def}}{=} (\tilde{Y}^{1/2} - \hat{Y}^{1/2}) D^{1/2}$. By Lemma A.4

$$\|E\|_2^2 \leq \|E\|_F^2 = \sum_i \|d_i \tilde{v}_i - \hat{v}_i\|^2 \leq 2m \cdot \|e^{\varepsilon/2 \cdot A}\|_2^2 \cdot \eta^2$$

$$\leq 2m \cdot \|Y^{1/2} D^{1/2}\|_2^2 \cdot \eta^2 \leq (2m)^2 \cdot L(K_V) \bullet Y \cdot \eta^2.$$

This also implies

$$\|E\|_2 \cdot \|\hat{Y}^{1/2} D^{1/2}\|_2 \leq \|E\|_F \cdot \|\hat{Y}^{1/2} D^{1/2}\|_F$$

$$\leq \left(\sqrt{2m} \cdot \|Y^{1/2} D^{1/2}\|_2 \cdot \eta \right) \cdot \sqrt{\sum_i d_i \|\hat{v}_i - \hat{v}_{\text{avg}}\|^2}$$

$$\leq 2m \cdot \eta \cdot (1 + \delta) \cdot L(K_V) \bullet Y.$$

As $D^{1/2} (\tilde{Y} - \hat{Y}) D^{1/2} = E^T E + (\hat{Y}^{1/2})^T E + E^T \hat{Y}^{1/2}$, we have

$$\|D^{1/2} (\tilde{Y} - \hat{Y}) D^{1/2}\|_2$$

$$\leq \|E^T E + (\hat{Y}^{1/2})^T E + E^T \hat{Y}^{1/2}\|_2$$

$$\leq \sqrt{3} \left(\|E\|_2^2 + 2 \cdot \|E\|_2 \|\hat{Y}^{1/2}\|_2 \right)$$

$$\leq 9 \cdot (2m)^2 \cdot (1 + \delta) \cdot L(K_V) \bullet Y \cdot \eta.$$

and

$$|L(K_V) \bullet (\tilde{Y} - \hat{Y})|$$

$$\leq L(K_V) \bullet (E^T E) + 2 \cdot |L(K_V) \bullet (E^T \hat{Y}^{1/2})|$$

$$\leq 1/2m \cdot \|E\|_F^2 + 2/2m \cdot \|E\|_F \|\hat{Y}^{1/2}\|_F$$

$$\leq 3 \cdot 2m \cdot (1 + \delta) \cdot L(K_V) \bullet Y \cdot \eta.$$

Finally, combining these bounds we have

$$\left\| \tilde{X} - \hat{X} \right\|_2 = \left\| \frac{\tilde{Y}}{L(K_V) \bullet \tilde{Y}} - \frac{\hat{Y}}{L(K_V) \bullet \hat{Y}} \right\|_2$$

$$\leq \left\| \frac{\tilde{Y}}{L(K_V) \bullet \tilde{Y}} - \frac{\tilde{Y}}{L(K_V) \bullet \hat{Y}} \right\|_2$$

$$+ \left\| \frac{\tilde{Y}}{L(K_V) \bullet \hat{Y}} - \frac{\hat{Y}}{L(K_V) \bullet \hat{Y}} \right\|_2$$

$$\leq \frac{\|\tilde{Y}\|_2 \cdot |L(K_V) \bullet \tilde{Y} - L(K_V) \bullet \hat{Y}|}{L(K_V) \bullet \tilde{Y} \cdot L(K_V) \bullet \hat{Y}} + \frac{\|\tilde{Y} - \hat{Y}\|_2}{L(K_V) \bullet \hat{Y}}$$

$$\leq \frac{2m \cdot |L(K_V) \bullet \tilde{Y} - L(K_V) \bullet \hat{Y}| + \|\tilde{Y} - \hat{Y}\|_2}{L(K_V) \bullet \hat{Y}}$$

$$\leq \frac{12 \cdot (2m)^2 \cdot (1 + \delta) \cdot L(K_V) \bullet Y \cdot \eta}{(1 - \delta) \cdot L(K_V) \bullet Y}$$

$$\leq 12 \cdot (2m)^2 \cdot 1+\delta/1-\delta \cdot \eta$$

$$\leq O(1/\text{poly}(n))$$

by taking η sufficiently small in $O(1/\text{poly}(n))$.
Hence, as $\|L(H)\|_2 \leq O(m)$ and $\|R_i\|_2 \leq O(m)$

$$|L(H) \cdot \hat{X} - L(H) \cdot \tilde{X}| \leq O(1/\text{poly}(n))$$

and

$$|R_i \bullet \hat{X} - R_i \bullet \tilde{X}| \leq O(1/\text{poly}(n)).$$

This, together with the fact that $1-\delta/1+\delta \geq 1 - 1/64$ and $1+\delta/1-\delta \leq 1 + 1/64$ completes the proof.

1. **Input:** An embedding $\{v_i \in \mathbb{R}^d\}_{i \in V}$, $b \in (0, 1/2]$.

2. Let $c = \Omega_b(1)$ be a constant to be fixed in the proof.

3. For $t = 1, 2, \ldots, O(\log n)$:

   (a) Pick a unit vector u uniformly at random from $\mathbb{S}^{d-1}$ and let $x \in \mathbb{R}^n$ with $x_i \stackrel{\text{def}}{=} \sqrt{d} \cdot u^T v_i$.

   (b) Sort the vector x. Assume w.l.og. that $x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n$. Define $S_i \stackrel{\text{def}}{=} \{j \in [n] : x_j \geqslant x_i\}$.

   (c) Let $S^{(t)} \stackrel{\text{def}}{=} (S_i, \bar{S}_i)$ which minimizes $\phi(S_i)$ among sweep-cuts for which $\mathrm{vol}(S_i) \in [c \cdot 2m, (1 - c) \cdot 2m]$.

4. **Output:** The cut $S^{(t)}$ of least conductance over all choices of t.

Figure 7: ProjRound

- **Input:** A matrix $M \in \mathbb{R}^{n \times n}$.

- Let $\eta \stackrel{\text{def}}{=} O(1/\mathrm{poly}(n))$. Let $\delta \stackrel{\text{def}}{=} 1/512$ and $\varepsilon = 1/130$.

- For k_δ as in Lemma A.3, sample k_δ vectors $u_1, \ldots, u_{k_\delta} \in \mathbb{R}^n$ as in Lemma A.3.

- Let $A \stackrel{\text{def}}{=} (2m \cdot \varepsilon) \cdot D^{-1/2} M D^{-1/2}$.

- For $1 \leqslant i \leqslant k_\delta$, compute vectors $b_i \in \mathbb{R}^n$, $b_i \stackrel{\text{def}}{=} \mathsf{EXPV}(1/2 \cdot A, D^{-1/2} u_i, \eta)$.

- Let B be the matrix having b_i as i-th row, and let $\tilde{v}_i$ be the i-th column of B. Compute $Z \stackrel{\text{def}}{=} \mathbb{E}_{\{i,j\} \in \mu \times \mu} \|\tilde{v}_i - \tilde{v}_j\|^2 = L(K_V) \bullet B^T B$.

- Return $\tilde{X} \stackrel{\text{def}}{=} 1/z \cdot B^T B$, by giving its correspoding embedding, i.e., $\{1/\sqrt{z} \cdot \tilde{v}_i\}_{i \in V}$.

Figure 8: The $\tilde{E}_\varepsilon$ algorithm