Finding Sparse Cuts via Cheeger Inequalities for Higher Eigenvalues

Anand Louis *, Prasad Raghavendra †, Prasad Tetali ‡, Santosh Vempala*

Abstract

Cheeger's fundamental inequality states that any edge-weighted graph has a vertex subset S such that its expansion (a.k.a. conductance of S or the sparsity of the cut $(S, \bar{S})$) is bounded as follows:

$$\phi(S) \stackrel{\text{def}}{=} \frac{w(S, \bar{S})}{\min\{w(S), w(\bar{S})\}} \leqslant \sqrt{2\lambda_2},$$

where w is the total edge weight of a subset or a cut and λ_2 is the second smallest eigenvalue of the normalized Laplacian of the graph. We study three natural generalizations of the sparsest cut in a graph:

- a partition of the vertex set into k parts that minimizes the sparsity of the partition (defined as the ratio of the weight of edges between parts to the total weight of edges incident to the smallest $k-1$ parts);
- a collection of k disjoint subsets $S_1, \ldots, S_k$ that minimize $\max_{i \in [k]} \phi(S_i)$;
- a subset of size $O(1/k)$ of the graph with minimum expansion.

Our main results are extensions of Cheeger's classical inequality to these problems via higher eigenvalues of the graph Laplacian. In particular, for the sparsest k-partition, we prove that the sparsity is at most $8\sqrt{\lambda_k}\log k$ where λ_k is the k^{th} smallest eigenvalue of the normalized Laplacian matrix. For the k sparse cuts problem we prove that there exist ck disjoint subsets $S_1, \ldots, S_{ck}$, such that

$$\max_i \phi(S_i) \leqslant C\sqrt{\lambda_k \log k}$$

where c, C are suitable absolute constants; this leads to a similar bound for the small-set expansion problem, namely for any k, there is a subset S whose weight is at most a $\mathcal{O}(1/k)$ fraction of the total weight and $\phi(S) \leqslant C\sqrt{\lambda_k \log k}$. The latter two results are the best possible in terms of the eigenvalues up to constant factors. Our results are derived via simple and efficient algorithms, and can themselves be viewed as generalizations of Cheeger's method.

*Supported by National Science Foundation awards AF-0915903 and AF-0910584.
†Supported by National Science Foundation Career Award and Alfred P. Sloan Fellowship.
‡Supported by National Science Foundation awards DMS-1101447 and AF-0910584.

1 Introduction

Given an edge-weighted graph $G = (V, E)$, a fundamental problem is to find a subset S of vertices such that the total weight of edges leaving it is as small as possible compared to its size. This latter quantity, called *expansion* or *conductance* of the subset or *sparsity* of the corresponding cut is defined as:

$$\phi_G(S) \stackrel{\text{def}}{=} \frac{w(S, \bar{S})}{\min\{w(S), w(\bar{S})\}}$$

where by $w(S)$ we denote the total weight of edges incident to vertices in S and $w(S,T)$ is the total weight of edges between vertex subsets S and T. The expansion of the graph G is defined as

$$\phi_G \stackrel{\text{def}}{=} \min_{S:w(S)\leqslant 1/2} \phi_G(S).$$

Finding the optimal subset that minimizes expansion $\phi_G(S)$ is known as the sparsest cut problem. The expansion of a graph and the problem of approximating it (sparsest cut problem) have been highly influential in the study of algorithms and complexity, and have exhibited deep connections to many other areas of mathematics. In particular, motivated by its applications and the NP-hardness of the problem, the study of approximation algorithms for sparsest cut has been a very fruitful area of research.

In this line, the fundamental Cheeger's inequality (shown for graphs in [Alo86, AM85]) establishes a bound on expansion via the spectrum of the graph.

Theorem 1.1 (Cheeger's Inequality ([Alo86, AM85]))**.** *For any graph G,*

$$\frac{\lambda_2}{2} \leqslant \phi_G \leqslant \sqrt{2\lambda_2}$$

where λ_2 is the second smallest eigenvalue of the normalized Laplacian [1] *of G.*

The proof of Cheeger's inequality is algorithmic, using the eigenvector corresponding to the second smallest eigenvalue. This theorem and its many (minor) variants have played a major role in the design of algorithms as well as in understanding the limits of computation.

Our work is motivated by extensions of the sparsest cut problem to more than one subset. In this work, we study multiple natural generalizations of sparsest cut problem.

All these generalizations are parametrized by a positive integer k, and reduce to the sparsest cut problem when restricted to the case $k = 2$. A natural question is whether these problems are connected to higher eigenvalues of the graph. We obtain upper and lower bounds for these generalizations of sparsest cut using higher eigenvalues. In the rest of the section, we briefly describe each generalization and present our results.

Sparsest k-partition Given a weighted undirected graph $G = (V, E)$ and an integer $k > 1$, find the k-partition with the least sparsity, where the sparsity of a k-partition $\mathcal{P} = \{S_1, \ldots, S_k\}$ of the vertex set V into k parts is defined as the ratio of the weight of edges between different parts to the sum of the weights of smallest $k-1$ parts in $\mathcal{P}$, i.e.,

$$\phi^{\mathsf{k-sum}}(\mathcal{P}) \stackrel{\text{def}}{=} \frac{\sum_{i\neq j} w(V_i, V_j)}{min_{j\in[k]} w(V \backslash V_j)}.$$

Variants of the sparsest k-partition have been considered in the literature. Closer to ours is the (α, ε)-clustering problem from [KVV04] that asks for a partition where each part has conductance at least α and the total weight of edges removed is minimized.

It is easy to see that the lower bound in Cheeger's inequality implies a lower bound of $\mathcal{P}$,

$$\phi^{\mathsf{k-sum}}(\mathcal{P}) \geqslant \lambda_2/2 \qquad \forall \text{ partitions } \mathcal{P}$$

[1] See Section 1.2 for the definition of the normalized Laplacian of a graph.

for the Sparsest k-partition. As it turns out, this lower bound cannot be strengthened for $k > 2$. To see this, consider the following simple construction: construct a graph G by taking $k-1$ cliques C_1, C_2, ..., C_{k-1} each on $(n-1)/(k-1)$ vertices along with an additional vertex v. Let the cliques $C_1, \ldots, C_{k-1}$ be connected to v by a single edge. Now, the graph G will have $k-1$ eigenvalues close to 1 because of the $k-1$ cuts $(\{v\}, C_i)$ for $i \in [k-1]$. However, the k^{th} eigenvalue will be close to 0, since any other cut which is not a linear combination of these $k-1$ cuts will have to cut through one of the cliques. Therefore, λ_k is a constant smaller than 1/2. But $\min_{\mathcal{P}} \phi^{\mathsf{k-sum}}(\mathcal{P}) = (k-1)/((k-2)(n/k)^2) \approx k^2/n^2$. Thus, $\lambda_k \gg \min_{\mathcal{P}} \phi^{\mathsf{k-sum}}(\mathcal{P})$ for small enough values of k.

Our main result is an upper bound on the Sparsest k-Partition via the higher eigenvalues. Specifically, we show the following.

Theorem 1.2. *For any edge-weighted graph $G = (V, E)$, and any integer $1 \leqslant k \leqslant |V|$, there exists a k-partition $S_1, \ldots, S_k$ of the vertices such that*

$$\phi^{\mathsf{k-sum}}(\{S_1, \ldots, S_k\}) \leqslant 8\sqrt{\lambda_k} \log k$$

where $\lambda_1, \ldots, \lambda_{|V|}$ are the eigenvalues of the normalized Laplacian of G and $c < 1, C$ are absolute constants. Moreover, such a partition can be identified in polynomial time.

The proof of Theorem 1.4 is based on a recursive partitioning algorithm that might be of independent interest.

k-sparse-cuts Given an edge weighted graph $G = (V, E)$ and an integer $k > 1$, find k disjoint subsets $S_1, \ldots, S_k$ of V such that $\max_i \phi_G(S_i)$ is minimized.

Note that the sets $S_1, \ldots, S_k$ need not form a partition of the set of vertices, i.e., there could be vertices that do not belong to any of the sets. Therefore problem models the existence of several well-formed *clusters* in a graph without the clusters being required to form a partition.

Along the lines of lower bound in Cheeger's inequality, it is not hard to show that the k^{th} smallest eigenvalue of the normalized Laplacian of the graph gives a lower bound to the k-sparse cuts problem. Formally, we have the following lower bound.

Proposition 1.3. *For any edge-weighted graph $G = (V, E)$, for any integer $1 \leqslant k \leqslant |V|$, and for any k disjoint subsets $S_1, \ldots, S_k \subset V$*

$$\max_i \phi_G(S_i) \geqslant \frac{\lambda_k}{2}$$

where $\lambda_1, \ldots, \lambda_{|V|}$ are the eigenvalues of the normalized Laplacian of G.

Complementing the lower bound, we show the following upper bound on k-sparse cuts problem in terms of λ_k.

Theorem 1.4. *For absolute constants c, C, the following holds: For every edge-weighted graph $G = (V, E)$, and any integer $1 \leqslant k \leqslant |V|$, there exist $c \cdot k$ disjoint subsets $S_1, \ldots, S_{c \cdot k}$ of vertices such that*

$$\max_i \phi_G(S_i) \leqslant C\sqrt{\lambda_k \log k}$$

where $\lambda_1, \ldots, \lambda_{|V|}$ are the eigenvalues of the normalized Laplacian of G. Moreover, the sets $S_1, \ldots, S_k$ satisfying the inequality can be identified in polynomial time.

The proof of Theorem 1.4 is algorithmic and is based on spectral projection. Starting with the embedding given by the smallest k eigenvectors of the (normalized) Laplacian of the graph, a simple randomized rounding procedure is used to produce k vectors having disjoint support, and then a Cheeger cut is obtained from each of these vectors. The running time is dominated by the time taken to compute the smallest k eigenvectors of the normalized Laplacian.

In general, one can not prove an upper bound better than $\mathcal{O}(\sqrt{\lambda_k \log k})$ for k sparse-cuts. This bound is matched by the family of *Gaussian graphs*. For a constant $\varepsilon \in (-1,1)$, let $N_{k,\varepsilon}$ denote the infinite graph over $\mathbb{R}^k$ where the weight of an edge (x,y) is the probability that two standard Gaussian random vectors X, Y with correlation [2] $1-\varepsilon$ equal x and y respectively. The first k eigenvalues of the Laplacian of $N_{k,\varepsilon}$ are at most ε ([RST10b]). The following Lemma bounds the expansion of small sets in $N_{k,\varepsilon}$.

Lemma 1.5 ([Bor85])**.** *For any set $S \subset \mathbb{R}^k$ with Gaussian probability measure at most $1/k$, $\phi_{N_{k,\varepsilon}}(S) = \Omega(\sqrt{\varepsilon \log k})$.*

For any k disjoint subsets $S_1, \ldots, S_k$ of the Gaussian graph $N_{k,\varepsilon}$, at least one of the sets has measure smaller than $\frac{1}{k}$, thus implying $\max_i \phi_{N_{k,\varepsilon}}(S_i) = \Omega(\sqrt{\epsilon \log k}) = \Omega(\sqrt{\lambda_k \log k})$.

It is natural to wonder if the above bounds extend to the case when the k-sets are required to form a partition. First, it is easy to see that Theorem 1.4 also implies an upper bound of $\mathcal{O}(\sqrt{\lambda_k \log k})$ on $\max_i \phi(S_i)$ for the case when the sets are required to form a partition of the vertex set.

Corollary 1.6. *For any edge-weighted graph $G=(V,E)$ and any integer $1 \leqslant k \leqslant |V|$, there exists a partition of the vertex set V into ck parts $S_1, \ldots, S_{ck}$ such that*

$$\max_i \phi(S_i) \leqslant C\sqrt{\lambda_k \log k}$$

for absolute constants c, C.

Complementing the above bound, we show that for a k- partition $S_1, S_2, \ldots, S_k$, the quantity $\max_i \phi_G(S_i)$ cannot be bounded by $\mathcal{O}(\sqrt{\lambda_k}\mathsf{polylogk})$ in general. We view this as further evidence suggesting that the k-sparse-cuts problem is the right generalization of sparsest cut to multiple subsets.

Theorem 1.7. *There exists a family of graphs such that for any k-partition $\{S_1, \ldots, S_k\}$ of the vertex set*

$$\max_i \phi_G(S_i) \geqslant C \min\left\{\frac{k^2}{\sqrt{n}}, n^{\frac{1}{12}}\right\}\sqrt{\lambda_k}.$$

Small-set expansion Given an edge weighted graph $G=(V,E)$ and $k>1$, find a subset of vertices S such that $w(S) \leqslant w(V)/k$ and $\phi_G(S)$ is minimized.

The small-set expansion problem came up in the context of understanding the Unique Games Conjecture ([RS10, ABS10]). As an immediate consequence of Theorem 1.4, we get the following optimal bound on the small-set expansion problem.

Corollary 1.8. *For any edge-weighted graph $G=(V,E)$ and any integer $1 \leqslant k \leqslant |V|$, there is a subset S with $w(S) = \mathcal{O}(1/k)w(V)$ and $\phi_G(S) \leqslant C\sqrt{\lambda_k \log k}$ for an absolute constant C.*

1.1 Related work

The classic sparsest cut problem has been extensively studied, and is closely connected to metric geometry [LLR95, AR98]. The lower and upper bounds on the sparsest cut given by Cheeger's inequality yield a $\mathcal{O}(\sqrt{\mathsf{OPT}})$ approximation algorithm for the sparsest cut problem. Leighton and Rao [LR99] gave an $\mathcal{O}(\log n)$ factor approximation algorithm via an LP relaxation. The same approximation factor can also be achieved using using properties of embeddings of metrics into Euclidean space [LLR95, AR98]. This was improved to $\mathcal{O}(\sqrt{\log n})$ via a semi-definite relaxation and embeddings of special metrics [ARV04]). In many contexts, and in practice, the eigenvector approach is often preferred in spite of a higher worst-case approximation factor.

For small-set expansion, this quantity was shown to be upper bounded by $\mathcal{O}(\sqrt{\lambda_{k^2} \log k})$ in [LRTV11], and by $\mathcal{O}(\sqrt{\lambda_{k^{100}} \log_k n})$ in [ABS10]. Using a semidefinite programming relaxation, [RST10a] gave an algorithm that outputs a small set with expansion at most $\sqrt{\mathsf{OPT}\log \mathsf{k}}$ where OPT is the sparsity of the optimal set

[2] ε correlated Gaussians can be constructed as follows : $X \sim \mathcal{N}(0,1)^k$ and $Y \sim (1-\varepsilon)X + \sqrt{2\varepsilon - \varepsilon^2}Z$ where $Z \sim \mathcal{N}(0,1)^k$.

of size at most $\mathcal{O}(1/k)$. Bansal et.al. [BFK+11] obtained an $\mathcal{O}(\sqrt{\log n \log k})$ approximation algorithm also using a semidefinite programming relaxation.

A problem closely related to the sparsest k-partition problem is the (α, ε)-clustering problem that asks for a partition where each part has conductance at least α and the total weight of edges removed is minimized. [KVV04] give a recursive algorithm to obtain a bi-criteria approximation to the (α, ε)-clustering problem. Indeed recursive algorithms are one of most commonly used techniques in practice for graph multi-partitioning.

In independent work, [LGT12] have obtained results similar to Theorem 1.4 with different techniques. They also studied a close variant of the problem we consider, and show that every graph G has a k partition such that each part has expansion at most $\mathcal{O}(k^6\sqrt{\lambda_k})$. Other generalizations of the sparsest cut problem have been considered for special classes of graphs ([BLR10, Kel06, ST96]).

A randomized rounding step similar to the one in our algorithm was used previously in the context of rounding semidefinite programs for unique games ([CMM06]).

1.2 Notation

For a graph $G = (V, E)$, we let A be its (weighted) adjacency matrix and d_i be the (weighted) degree of vertex i. We use D to denote the diagonal matrix with $D_{ii} = d_i$. The *normalized* Laplacian of a graph defined as

$$\mathcal{L}_G \stackrel{\text{def}}{=} D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}}$$

We let $0 = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \lambda_n$ denote the eigenvalues of $\mathcal{L}_G$ and $v_1', v_2', \ldots, v_n'$ denote the corresponding eigenvectors. Let $v_i \stackrel{\text{def}}{=} D^{-\frac{1}{2}} v_i'$ for each $i \in [n]$. Therefore,

$$v_i'^T \mathcal{L}_G v_i' = \sum_{u \sim w} (v_i(u) - v_j(w))^2.$$

Since $\forall i \neq j$ $\langle v_i', v_j' \rangle = 0$, $\sum_l d_l v_i(l) v_j(l) = 0$

Given a vector $x \in \mathbb{R}^n$ and an index $i \in [n]$, we define the i^{th} *level set* of x to be the set $\{j \in [n] | x(j) > x_{i,\max}\}$, where $x_{i,\max}$ is the i^{th} largest entry in x.

Given a k-partition $\mathcal{P} = \{S_1, \ldots, S_k\}$ we denote the sum of the weights of the edges with endpoints in different pieces by $\mathsf{E}(\mathcal{P})$. More formally,

$$\mathsf{E}(\mathcal{P}) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{\mathsf{e} \in \mathsf{E}(\mathsf{S_i}, \bar{\mathsf{S}}_\mathsf{i})} \mathsf{w}(\mathsf{e})$$

2 Sparsest k-partition

2.1 Recursive partitioning algorithm

We propose the following recursive algorithm for finding a k-partitioning of G. Use the second eigenvector of $\mathcal{L}$ to find a sparse cut $(C, \bar{C})$. Let $G' = (V, E')$ be the graph obtained by removing the edges in the cut $(C, \bar{C})$ from G and adding self loops at the endpoints of the edges removed. Let $\mathcal{L}'$ be the normalized Laplacian of the graph obtained. The matrix $\mathcal{L}'$ is block-diagonal with two blocks for the two components of G'. The spectrum of $\mathcal{L}'$ (eigenvalues, eigenvectors) is the union of the spectra of the two blocks. The first two eigenvalues of $\mathcal{L}'$ are now 0 and we use the third largest eigenvector of $\mathcal{L}'$ to find a sparse cut in G'. This is the second eigenvector in one of the two blocks and partitions that block. We repeat the above process till we have at least k connected components. This can be viewed as a recursive algorithm, where at each step one of the current components is partitioned into two; the component partitioned is the one that has the lowest second eigenvalue among all the current components. The precise algorithm appears in Figure 1.

2.2 Analysis

In this section, we analyze the recursive partitioning algorithm given in Figure 1. Our analysis will also be a proof of Theorem 1.2. We begin with some monotonicity properties of eigenvalues.

1. Input : Graph $G = (V, E)$, m such that $1 < k < |V|$
2. Initialize $i := 2$, and $G_i = G$, $\mathcal{L}_i$ = normalized Laplacian matrix of G_i
   (a) Find a sparse cut $(C_i, \bar{C}_i)$ in G_i using the i^{th} eigenvector of $\mathcal{L}_i$ (the first $i-1$ are all equal to 0).
   (b) Let $G_{i+1} := \big(G_i \backslash E_{G_i}(C, \bar{C})\big) \cup \{\{v, v\} |\ \exists\, u \text{ such that} \{u, v\} \in E_{G_i}(C, \bar{C})\}$ with $w(\{v, v\}) = \sum_{\{u,v\} \in E_{G_i}(C,\bar{C})} w(\{u, v\})$.
   (c) If $i = k$ then output the connected components of G_{i+1} and End else
   (d) Let $\mathcal{L}_{i+1}$ be the normalized Laplacian matrix of G_{i+1}.

Figure 1: The Recursive k-partition Algorithm

Monotonicity of Eigenvalues. In this section we collect some useful properties about the behavior of eigenvalues upon deleting edges and merging vertices.

Lemma 2.1 (Weyl's Inequality)**.** *Given a Hermitian matrix B with eigenvalues $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$, and a positive semidefinite matrix E, if $\lambda'_1 \leqslant \lambda'_2 \leqslant \ldots \leqslant \lambda'_n$ denote the eigenvalues of $B' \stackrel{\text{def}}{=} B - E$, then $\lambda'_i \leqslant \lambda_i$.*

Proof. The i^{th} eigenvalue of B' can be written as

$$\begin{aligned} \lambda'_i &= \max_{S:rank(S)=i} \min_{x \in S} \frac{x^T B' x}{x^T x} \\ &= \max_{S:rank(S)=i} \min_{x \in S} \frac{x^T B x - x^T E x}{x^T x} \\ &\leqslant \max_{S:rank(S)=i} \min_{x \in S} \frac{x^T B x}{x^T x} \\ &= \lambda_i. \end{aligned}$$

□

Lemma 2.2. *Let $\mathcal{L}$ be the normalized Laplacian matrix of the graph G. Let F be any subset of edges of G. For every pair $(i, j) \in F$, remove the edge (i, j) from G and add self loops at i and j to get the graph G'. Let $\mathcal{L}'$ be the normalized Laplacian matrix of G'. Let the eigenvalues of $\mathcal{L}$ be $0 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$ and let the eigenvalues of $\mathcal{L}'$ be $0 \leqslant \lambda'_2 \leqslant \lambda'_3 \leqslant \ldots \leqslant \lambda'_n$. Then $\lambda'_i \leqslant \lambda_i\ \forall i \in [n]$.*

Proof. Let $C \stackrel{\text{def}}{=} \mathcal{L} - \mathcal{L}'$ is the matrix corresponding to the edge subset F. It has non-negative entries along its diagonal and non-positive entries elsewhere such that $\forall i\ c_{ii} = -\sum_{j \neq i} c_{ij}$. C is symmetric and positive semi-definite as for any vector x of appropriate dimension, we have

$$x^T C x = \sum_{ij} c_{ij} x_i x_j = -\frac{1}{2} \sum_{i \neq j} c_{ij} (x_i - x_j)^2 \geqslant 0.$$

Using Lemma 2.1, we get that $\lambda'_i \leqslant \lambda_i\ \forall i \in [n]$. □

Lemma 2.2 shows that the eigenvalues of $\mathcal{L}_i$ are monotonically non-increasing with i. This will show that $\phi_{G_i}(C_i) \leqslant \sqrt{2\lambda_k}$. We are now ready to prove Theorem 1.2.

Proof of Theorem 1.2 . Let $\mathcal{P}$ be the partition output by the algorithm and let $S(\mathcal{P})$ denote the sum of weights of the smallest $k-1$ pieces in $\mathcal{P}$. Note that we need only the smaller side of a cut to bound the size of the cut : $w(E_G(S,\bar{S})) \leqslant \phi_G\ w(S)$. We define the notion of a $\mathsf{cut-tree}$ $T = (V(T), E(T))$ as follows: $V(T) = \{V\} \cup \{C_i | i \in [k]\}$ (For any cut $(C_i, \bar{C_i})$ we denote the part with the smaller weight by C_i and the part with the larger weight by $\bar{C_i}$. We break ties arbitrarily). We put an edge between $S_1, S_2 \in V(T)$ if $\nexists S \in V(T)$ such that $S_1 \subsetneq S \subsetneq S_2$ or $S_2 \subsetneq S \subsetneq S_1$, (one can view S_1 as a 'top level' cut of S_2 in the former case).

Clearly, T is connected and is a tree. We call V the root of T. We define the *level* of a node in T to be its depth from the root. We denote the level of node $S \in V(T)$ by $L(S)$. The root is defined to be at level 0. Observe that $S_1 \in V(T)$ is a descendant of $S_2 \in V(T)$ if and only if $S_1 \subsetneq S_2$. Now $\mathsf{E}(\mathcal{P}) = \cup_{\mathsf{i}} \mathsf{E}_{\mathsf{G_i}}(\mathsf{C_i}, \bar{\mathsf{C}}_{\mathsf{i}}) = \cup_{\mathsf{i}} \cup_{\mathsf{j:L(C_j)=i}} \mathsf{E}_{\mathsf{G_j}}(\mathsf{C_j}, \bar{\mathsf{C}}_{\mathsf{j}})$. We make the following claim.

Claim 2.3.

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C_j})) \leqslant 2\sqrt{\lambda_k}\ S(\mathcal{P})$$

Proof. By definition of level, if $L(C_i) = L(C_j)$, $i \neq j$, then the node corresponding to C_i in the T can not be an ancestor or a descendant of the node corresponding to C_j. Hence, $C_i \cap C_j = \phi$. Therefore, all the sets of vertices in level i are pairwise disjoint. Using Cheeger's inequality we get that $E(C_j, \bar{C_j}) \leqslant 2\sqrt{\lambda_k} w(C_j)$. Therefore

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C_j})) \leqslant 2\sqrt{\lambda_k} \sum_{j:L(C_j)=i} w(C_j) \leqslant 2\sqrt{\lambda_k} S(\mathcal{P})$$

□

This claim implies that $\phi(\mathcal{P}) \leqslant 2\sqrt{\lambda_k}\ \mathsf{height}(T)$.
The height of T might be as much as k. But we will show that we can assume height(T) to be $\log k$. For any path in the tree $v_1, v_2, \ldots, v_{p-1}, v_p$ such that $\mathsf{deg}(v_1) > 2$, $\mathsf{deg}(v_i) = 2$ (i.e. v_i has only 1 child in T) for $1 < i < k$, we have $w(C_{v_{i+1}}) \leqslant w(C_{v_i})/2$, as v_{i+1} being a child of v_i in the T implies that $C_{v_{i+1}}$ was obtained by cutting C_{v_i} using it's second eigenvector. Thus $\sum_{i=2}^{p} w(C_{v_i}) \leqslant w(C_{v_1})$. Hence we can modify the T as follows : make the nodes $v_3, \ldots, v_p$ children of v_2. The nodes $v_3, \ldots, v_{p-1}$ now become leaves whereas the subtree rooted at v_p remains unchanged. We also assign the level of each node as its new distance from the root. In this process we might have destroyed the property that a node is obtained from by cutting its parent, but we have the property that $w(\cup_{j:L(C_j)=i} E(C_j, \bar{C_j})) \leqslant 4\sqrt{\lambda_k} S(\mathcal{P})\ \forall i$.

Claim 2.4.

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C_j})) \leqslant 4\sqrt{\lambda_k}\ S(\mathcal{P})$$

Proof. If the nodes in level i are unchanged by this process, then the claim clearly holds. If any node v_j in level i moved to a higher level, then the nodes replacing v_j in level i would be descendants of v_j in the original T and hence would have weight at most $w(C_{v_j})$. If the descendants of some node v_j got added to level i, then, as seen above, their combined weight would be at most $w(C_{v_j})$. Hence,

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C_j})) \leqslant 2\left(2\sqrt{\lambda_k} \sum_{j:L(C_j)=i} w(C_j)\right) \leqslant 4\sqrt{\lambda_k}\ S(\mathcal{P})$$

.

□

Repeating this process we can ensure that no two adjacent nodes in the T have degree 2. Hence, there are at most $\log k$ vertices along any path starting from the root which have exactly one child. Thus the height of the new $\mathsf{cut-tree}$ is at most $2\log k$. Thus $\mathsf{E}(\mathcal{P}) \leqslant 8\sqrt{\lambda_k} \log \mathsf{k}\ \mathsf{S}(\mathcal{P})$ and hence $\phi^{\mathsf{k-sum}} \leqslant \frac{\mathsf{E}(\mathcal{P})}{S(\mathcal{P})} \leqslant 8\sqrt{\lambda_k} \log k$.

□

3 k sparse-cuts

3.1 Gaussian Projection Algorithm

Our algorithm for finding $\Theta(k)$ sparse cuts appears in Figure 2.

Input : Graph $G = (V, E)$, parameter k.

1. **Spectral projection.** Let $V = [v_1, \ldots, v_k]$ be an $n \times k$ matrix where v_i is as defined in Section 1.2; let $u_1, \ldots, u_n$ be the rows of V.
2. **Randomized rounding.** Pick k independent Gaussian vectors $g_1, g_2, \ldots, g_k \sim \mathcal{N}(0,1)^k$. Construct vectors $h_1, h_2, \ldots, h_k \in \mathbb{R}^n$ as follows:
$$h_i(j) = \begin{cases} \langle u_j, g_i \rangle \text{ if } i = \mathsf{argmax}_{i \in [k]} \{\langle u_j, g_i \rangle\} \\ 0 \text{ otherwise.} \end{cases}$$
3. **Cheeger cuts.** For $j = 1, \ldots, k$, sort the cooordinates of h_j according to their magnitude, and pick the level set having least expansion .
4. Output all subsets with expansion smaller than $C\sqrt{\lambda_k \log k}$ for an appropriately chosen constant C.

Figure 2: The Many-sparse-cuts Algorithm

The first and third steps are clearly direct generalizations of Cheeger's method for finding a sparse cut. However, the intermediate step of applying a random transformation and rounding appears to be essential to prove a worst-case guarantee.

3.2 Analysis

In this section, we will present the analysis of the Gaussian Projection algorithm presented in Figure 2. We begin with an outline of the argument.

3.2.1 Proof Outline

Notice that the vectors $h_1, h_2, \ldots, h_k$ have disjoint support since for each coordinate j, exactly one of the $\langle u_j, g_i \rangle$ is maximum. Therefore, the Cheeger cuts obtained by the vectors h_i yield k disjoint sets. It is sufficient to show that a constant fraction of the sets so produced have small expansion.

As a first attempt to proving the upper bound in Theorem 1.4, one could first try to bound the Rayleigh quotient of the vectors $\{h_i\}$ by $\mathcal{O}(\lambda_k \log k)$ (say for a constant fraction of vectors h_i). This would imply that the corresponding sets would have value $\mathcal{O}(\sqrt{\lambda_k \log k})$ by following the proof of Cheeger's inequality. Unfortunately, we note that the Rayleigh quotients of the vectors obtained could themselves be as high as $\Omega(\sqrt{\lambda_k \log k})$, and using the proof of Cheeger's inequality this would at best yield a bound of $\mathcal{O}((\lambda_k \log k)^{\frac{1}{4}})$ on the expansion of the sets obtained. Therefore, in our proof we directly analyze the quality of the Cheeger cuts finally output by the algorithm.

We will show that for each $i \in \{1, \ldots, k\}$, the vector h_i has a constant probability of yielding a cut with small expansion. The outline of the proof is as follows. Let f denote the vector h_1. The choice of the index 1 is arbitrary and the same analysis is applicable to all other indices $i \in [k]$.

The quality of the Cheeger cut obtained from f can be upper bounded using the following standard lemma. A proof of this lemma can be found in [Chu97].

Lemma 3.1. *Let $x \in \mathbb{R}^n$ be a vector such that $\frac{\sum_{i\sim j}|x_i - x_j|}{\sum_i d_i x_i} \leqslant \delta$. Then one of the level sets, say S, of the vector x has $\phi_G(S) \leqslant 2\delta$.*

Applying Lemma 3.1, the expansion of the set retrieved from $f = h_1$ is upper bounded by,

$$\frac{\sum_{(i,j)\in E}|f_i^2 - f_j^2|}{\sum_i d_i f_i^2}.$$

Both the numerator and denominator are random variables depending on the choice of random Gaussians $g_1, \ldots, g_k$. It is a fairly straightforward calculation to bound the expected value of the denominator.

Lemma 3.2.

$$2\log k \leqslant \mathbb{E}\left[\sum_i d_i f_i^2\right] \leqslant 4\log k.$$

Bounding the expected value of the numerator is more subtle. We show the following bound on the expected value of the numerator.

Lemma 3.3.

$$\mathbb{E}\left[\sum_{i\sim j}|f_i^2 - f_j^2|\right] \leqslant c(\sqrt{\lambda_k}\log^{\frac{3}{2}} k).$$

where c is an absolute constant.

Notice that the ratio of their expected values is $\mathcal{O}(\sqrt{\lambda_k \log k})$, as intended. To control the ratio of the two quantities, the numerator is to be bounded from above, and the denominator is to be bounded from below. A simple Markov inequality can be used to upper bound the probability that the numerator is much larger than its expectation. To control the denominator, we bound its variance. Specifically, we will show the following bound on the variance of the denominator.

Lemma 3.4.

$$\mathsf{Var}\sum_i d_i f_i^2 \leqslant 28\log^2 k.$$

The above moment bounds are sufficient to conclude that with constant probability, the ratio $\frac{\sum_{(i,j)\in E}|f_i^2 - f_j^2|}{\sum_i d_i f_i^2}$ is within a constant factor of $\mathcal{O}(\sqrt{\lambda_k \log k})$. Therefore, with constant probability over the choice of the Gaussians $g_1, \ldots, g_k$, $\Omega(k)$ of the vectors $h_1, \ldots h_k$ yield sets of expansion $\mathcal{O}(\sqrt{\lambda_k \log k})$.

3.2.2 Technical Preliminaries

Spectral Embedding. Let $\{u_i | i \in V\}$ be the spectral embedding obtained from the spectral projection step of the Algorithm in Figure 2. This embedding satisfies the following properties.

Lemma 3.5. *(Spectral embedding)*

1.

$$\frac{\sum_{i\sim j}\|u_i - u_j\|^2}{\sum_i d_i \|u_i\|^2} \leqslant \lambda_k$$

2.

$$\sum_i d_i \|u_i\|^2 = k$$

3.

$$\sum_{i,j} d_i d_j \langle u_i, u_j\rangle^2 = k.$$

Proof of (3).

$$\begin{aligned}
\sum_{i,j} d_i d_j \langle u_i, u_j \rangle^2 &= \sum_{i,j} d_i d_j \left(\sum_{t=1}^{k} u_i(t) u_j(t) \right)^2 \\
&= \sum_{i,j} d_i d_j \sum_{t_1,t_2} u_i(t_1) u_j(t_1) u_i(t_2) u_j(t_2) \\
&= \sum_{t_1,t_2} \sum_{i,j} d_i d_j u_i(t_1) u_j(t_1) u_i(t_2) u_j(t_2) \\
&= \sum_{t_1,t_2} \left(\sum_i d_i u_i(t_1) u_i(t_2) \right)^2
\end{aligned}$$

Since $\sqrt{d_i}u_i(t_1)$ is the entry to corresponding to vertex i in the t_1^{th} eigenvector, $\sum_i d_i u_i(t_1) u_i(t_2)$ is equal to the inner product of the t_1^{th} and t_2^{th} eigenvectors of $\mathcal{L}$, which is equal to 1 only when $t_1 = t_2$ and is equal to 0 otherwise. Therefore,

$$\sum_{i,j} d_i d_j \langle u_i, u_j \rangle^2 = \sum_{t_1,t_2} \mathbb{I}\left[t_1 = t_2\right] = k$$

□

Next, we recall the one-sided Chebychev inequality.

Fact 3.6 (One-sided Chebychev Inequality)**.** *For a random variable X with mean μ and variance σ^2 and any $t > 0$,*

$$\mathbb{P}\left[X < \mu - t\sigma\right] \leqslant \frac{1}{1+t^2}.$$

Properties of Gaussian Variables. The next few facts are folklore about Gaussians. Let $\mathsf{t}_{1/k}$ denote the $(1/k)^{th}$ cap of a standard normal variable, i.e., $\mathsf{t}_{1/k} \in \mathbb{R}$ is the number such that for a standard normal random variable X, $\mathbb{P}\left[X \geqslant \mathsf{t}_{1/k}\right] = 1/k$.

Fact 3.7. *For a standard normal random variable X and for every $k > 0$,*

$$\mathsf{t}_{1/k} \approx \sqrt{2\log k - \log\log k}$$

Fact 3.8. *Let $X_1, X_2, \ldots, X_k$ be k independent standard normal random variables. Let Y be the random variable defined as $Y \stackrel{\text{def}}{=} \max\{X_i | i \in [k]\}$. Then*

1. $\mathsf{t}_{1/k} \leqslant \mathbb{E}\left[Y\right] \leqslant 2\sqrt{\log k}$
2. $\mathbb{E}\left[Y^2\right] \leqslant 4\log k$
3. $\mathbb{E}\left[Y^4\right] \leqslant 4e\log^2 k$
4. *For any constant ε,* $\mathbb{P}\left[Y \leqslant (1-\varepsilon)\mathsf{t}_{1/k}\right] \leqslant \frac{1}{e^{k^{2\varepsilon}}}$

Proof. For any $Z_1, \ldots, Z_k \in \mathbb{R}^+$ and any $p \in \mathbb{Z}^+$, we have $\max_i Z_i \leqslant (\sum_i Z_i^p)^{\frac{1}{p}}$. Now $Y^4 = (\max_i X_i)^4 \leqslant \max_i X_i^4$.

$$\begin{aligned}
\mathbb{E}\left[Y^4\right] &\leqslant \mathbb{E}\left[\left(\sum_i X_i^{4p}\right)^{\frac{1}{p}}\right] \leqslant \left(\mathbb{E}\left[\sum_i X_i^{4p}\right]\right)^{\frac{1}{p}} \qquad \text{(Jensen's Inequality)} \\
&\leqslant \left(\sum_i (\mathbb{E}\left[X_i^2\right]) \frac{(4p)!}{(2p)!2^{2p}}\right)^{\frac{1}{p}} \leqslant 4p^2 k^{\frac{1}{p}} \quad \text{(using } (4p)!/(2p)! \leqslant (4p)^{2p} \text{)}
\end{aligned}$$

Picking $p = \log k$ gives $\mathbb{E}\left[Y^4\right] \leqslant 4e\log^2 k$.
Therefore $\mathbb{E}\left[Y^2\right] \leqslant \sqrt{\mathbb{E}\left[Y^4\right]} \leqslant 4\log k$ and $\mathbb{E}\left[Y\right] \leqslant \sqrt{\mathbb{E}\left[Y^2\right]} \leqslant 2\sqrt{\log k}$.
And,

$$\mathbb{P}\left[Y \leqslant (1-\varepsilon)\mathsf{t}_{1/k}\right] \leqslant \left(1 - \frac{1}{k^{1-2\varepsilon}}\right)^k \leqslant \frac{1}{e^{k^{2\varepsilon}}}$$

□

Fact 3.9. *Let $X_1, \ldots, X_k$ and $Y_1, \ldots, Y_k$ be i.i.d. standard normal random variables such that for all $i \in [k]$, the covariance of X_i and Y_i is at least $1-\varepsilon^2$. Then*

$$\mathbb{P}\left[\mathsf{argmax_i X_i} \neq \mathsf{argmax_i Y_i}\right] \leqslant c_1\left(\varepsilon\sqrt{\log k}\right)$$

for some absolute constant c_1.

We refer the reader to [CMM06] for the proof of a more general claim.

3.2.3 Main Proofs

Let f denote the vector h_1. The choice of the index 1 is arbitrary and the same analysis is applicable to all other indices $i \in [k]$. We first separately bound the expectations of the numerator and denominator of the sparsity of each cut, and then the variance of the denominator. The proofs of these bounds will follow their application in the proof of our main theorem.

Expectation of the Denominator. Bounding the expectation of the denominator is a straightforward calculation as shown below.

Lemma 3.10 (Restatement of Lemma 3.2)**.**

$$2\log k \leqslant \mathbb{E}\left[\sum_i d_i f_i^2\right] \leqslant 4\log k.$$

Proof of Lemma 3.2. For any $i \in [n]$, recall that

$$f_i = \begin{cases} \|u_i\|\langle \tilde{u_i}, g_1\rangle \text{ if } \langle \tilde{u_i}, g_1\rangle \geqslant \langle \tilde{u_i}, g_j\rangle \ \forall j \in [k] \\ 0 \text{ otherwise.} \end{cases}$$

The first case happens with probability $1/k$ and so $f_i = 0$ with the remaining probability. Therefore, using Fact 3.8,

$$2\|u_i\|^2\log k/k \leqslant \mathbb{E}\left[f_i^2\right] \leqslant 4\|u_i\|^2\log k/k$$

and hence

$$2\log k \leqslant \mathbb{E}\left[\sum_i d_i f_i^2\right] \leqslant 4\log k$$

using $\sum_i d_i\|u_i\|^2 = k$ from Lemma 3.5. □

Expectation of the Numerator. For bounding the expectation of the numerator we will need some preparation. We will make use of the following proposition which relates distance between two vectors to the distance between the unit vectors in the corresponding directions.

Proposition 3.11. *For any two non zero vectors u_i and u_j, if $\tilde{u_i} = u_i/\|u_i\|$ and $\tilde{u_j} = u_j/\|u_j\|$ then*

$$\|\tilde{u_i} - \tilde{u_j}\|\sqrt{\|u_i\|^2 + \|u_j\|^2} \leqslant 2\|u_i - u_j\|$$

Proof. Note that $2\|u_i\|\|u_j\| \leqslant \|u_i\|^2 + \|u_j\|^2$. Hence,

$$\begin{aligned}
\|\tilde{u_i} - \tilde{u_j}\|^2(\|u_i\|^2 + \|u_j\|^2) &= (2 - 2\langle\tilde{u_i}, \tilde{u_j}\rangle)(\|u_i\|^2 + \|u_j\|^2) \\
&\leqslant 2(\|u_i\|^2 + \|u_j\|^2 - (\|u_i\|^2 + \|u_j\|^2)\langle\tilde{u_i}, \tilde{u_j}\rangle)
\end{aligned}$$

If $\langle\tilde{u_i}, \tilde{u_j}\rangle \geqslant 0$, then

$$\|\tilde{u_i} - \tilde{u_j}\|^2(\|u_i\|^2 + \|u_j\|^2) \leqslant 2(\|u_i\|^2 + \|u_j\|^2 - 2\|u_i\|\|u_j\|\langle\tilde{u_i}, \tilde{u_j}\rangle) \leqslant 2\|u_i - u_j\|^2$$

Else if $\langle\tilde{u_i}, \tilde{u_j}\rangle < 0$, then

$$\|\tilde{u_i} - \tilde{u_j}\|^2(\|u_i\|^2 + \|u_j\|^2) \leqslant 4(\|u_i\|^2 + \|u_j\|^2 - 2\|u_i\|\|u_j\|\langle\tilde{u_i}, \tilde{u_j}\rangle) \leqslant 4\|u_i - u_j\|^2$$

□

We will also make use of the following propositions which bounds the expected value of a conditioned random variable.

Proposition 3.12. *For indices $i \neq j$*

$$\mathbb{E}\left[\langle u_i, g_1\rangle^2 | f_j > 0\right] \mathbb{P}\left[f_j > 0\right] \leqslant \frac{4}{k}\|u_i\|^2 \log k$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[\langle u_i, g_1\rangle^2 | f_j > 0\right] \mathbb{P}\left[f_j > 0\right] &\leqslant \mathbb{E}\left[\max_{p\in[k]}\langle u_i, g_p\rangle^2 | \langle u_j, g_1\rangle \geqslant \langle u_j, g_l\rangle \forall l \in [k]\right] \\
&\qquad \mathbb{P}\left[\langle u_j, g_1\rangle \geqslant \langle u_j, g_l\rangle \forall l \in [k]\right] \\
&= \frac{1}{k}\sum_{q\in[k]} \mathbb{E}\left[\max_{p\in[k]}\langle u_i, g_p\rangle^2 | \langle u_j, g_q\rangle \geqslant \langle u_j, g_l\rangle \forall l \in [k]\right] \\
&\qquad \mathbb{P}\left[\langle u_j, g_q\rangle \geqslant \langle u_j, g_l\rangle \forall l \in [k]\right] \\
&= \frac{1}{k}\mathbb{E}\left[\max_{p\in[k]}\langle u_i, g_p\rangle^2\right] \\
&= \frac{4}{k}\|u_i^2\| \log k
\end{aligned}$$

□

Similarly, we also prove the following proposition.

Proposition 3.13. *For indices $i \neq j$*

$$\mathbb{E}\left[\langle u_i, g_1\rangle^2 | f_i > 0 \text{ and } f_j = 0\right] \mathbb{P}\left[f_j = 0\right] \leqslant 4\left(1 - \frac{1}{k}\right)\|u_i\|^2 \log k$$

We will need another lemma that is a direct consequence of Fact 3.9 about the maximum of k i.i.d normal random variables.

Proposition 3.14. *For any $i, j \in [n]$,*

$$\mathbb{P}\left[f_i > 0 \text{ and } f_j = 0\right] \leqslant c_1\left(\|\tilde{u_i} - \tilde{u_j}\|\frac{\sqrt{\log k}}{k}\right).$$

We are now ready to bound the expectation of the numerator, we restate the lemma for the convenience of the reader.

Lemma 3.15. *(Restatement of Lemma 3.3)*

$$\mathbb{E}\left[\sum_{i\sim j}|f_i^2 - f_j^2|\right] \leqslant 8(2c_1+1)(\sqrt{\lambda_k}\log^{\frac{3}{2}} k).$$

Proof of Lemma 3.3. From Fact 3.8, $\mathbb{E}\left[f_i^2|f_i>0\right] \leqslant 4\|u_i\|^2\log k$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[|f_i^2 - f_j^2||f_i,f_j>0\right] &= \mathbb{E}\left[|\langle u_i,g_1\rangle^2 - \langle u_j,g_1\rangle^2||f_i,f_j>0\right] \\
&= \mathbb{E}\left[|\langle u_i-u_j,g_1\rangle\langle u_i+u_j,g_1\rangle||f_i,f_j>0\right] \\
&\leqslant \sqrt{\mathbb{E}\left[\langle u_i-u_j,g_1\rangle^2|f_i,f_j>0\right]}\sqrt{\mathbb{E}\left[\langle u_i+u_j,g_1\rangle^2|f_i,f_j>0\right]}
\end{aligned}$$

Now,

$$\mathbb{E}\left[\langle u_i-u_j,g_1\rangle^2|f_i,f_j>0\right]\mathbb{P}\left[f_i,f_j>0\right] = \int_{f_i,f_j>0}\langle u_i-u_j,g_1\rangle^2$$

$$\leqslant \int_{f_i>0}\langle u_i-u_j,g_1\rangle^2 = \mathbb{E}\left[\langle u_i-u_j,g_1\rangle^2|f_i>0\right]\mathbb{P}\left[f_i>0\right]$$

$$= \frac{1}{k}\sum_{p\in[k]}\mathbb{E}\left[\max_l\langle u_i-u_j,g_l\rangle^2|\langle u_i,g_p\rangle \geqslant \langle u_i,g_l\rangle \forall l\in[k]\right]\mathbb{P}\left[\langle u_i,g_p\rangle \geqslant \langle u_i,g_l\rangle\forall l\in[k]\right]$$

$$= \frac{1}{k}\mathbb{E}\left[\max_l\langle u_i-u_j,g_l\rangle^2\right] \leqslant \frac{4}{k}\|u_i-u_j\|^2\log k$$

Similarly, we get

$$\mathbb{E}\left[\langle u_i+u_j,g_1\rangle^2|f_i,f_j>0\right]\mathbb{P}\left[f_i,f_j>0\right] \leqslant \frac{4}{k}\|u_i+u_j\|^2\log k$$

Therefore, we get

$$\mathbb{E}\left[|f_i^2-f_j^2||f_i,f_j>0\right]\mathbb{P}\left[f_i,f_j>0\right] \leqslant \frac{4}{k}\|u_i-u_j\|\|u_i+u_j\|\log k$$

From Proposition 3.14,

$$\mathbb{P}\left[f_i>0 \text{ and } f_j=0\right] = \mathbb{P}\left[f_j>0 \text{ and } f_i=0\right] \leqslant c_1(\|\tilde{u_i}-\tilde{u_j}\|\sqrt{\log k}/k).$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[f_i^2-f_j^2|f_i>0,f_j=0\right]\mathbb{P}\left[f_i>0,f_j=0\right] &= \mathbb{E}\left[\langle u_1,g\rangle^2|f_i>0,f_j=0\right]\mathbb{P}\left[f_j=0\right]\frac{\mathbb{P}\left[f_i>0,f_j=0\right]}{\mathbb{P}\left[f_j=0\right]} \\
&\leqslant 4(1-\frac{1}{k})\|u_i\|^2\log k\frac{\mathbb{P}\left[f_i>0,f_j=0\right]}{1-1/k} \quad (\text{ Proposition 3.13 }) \\
&= \frac{4c_1}{k}\log^{\frac{3}{2}} k\|u_i\|^2\|\tilde{u_i}-\tilde{u_j}\|
\end{aligned}$$

Similarly,

$$\mathbb{E}\left[f_j^2-f_i^2|f_j>0,f_i=0\right]\mathbb{P}\left[f_j>0,f_i=0\right] \leqslant \frac{4c_1}{k}\log^{\frac{3}{2}} k\|u_j\|^2\|\tilde{u_i}-\tilde{u_j}\|$$

Next,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\sim j}|f_i^2 - f_j^2|\right] &\leqslant \frac{4\log k}{k}\sum_{i\sim j}\left(\|u_i-u_j\|\|u_i+u_j\| + c_1(\|u_i\|^2+\|u_j\|^2)\|\tilde{u_i}-\tilde{u_j}\|\sqrt{\log k}\right) \\
&\leqslant \frac{4\log k}{k}\sum_{i\sim j}\left(\|u_i-u_j\|(\|u_i\|+\|u_j\|) + 2c_1\|u_i-u_j\|\sqrt{\|u_i\|^2+\|u_j\|^2}\sqrt{\log k}\right) \\
&\qquad (\text{ from Lemma 3.11}) \\
&\leqslant \frac{4(2c_1+1)\log^{\frac{3}{2}} k}{k}\sum_{i\sim j}\|u_i-u_j\|(\|u_i\|+\|u_j\|) \\
&\leqslant \frac{4(2c_1+1)\log^{\frac{3}{2}} k}{k}\sqrt{\sum_{i\sim j}(\|u_i-u_j\|)^2}\sqrt{\sum_{i\sim j}(\|u_i\|+\|u_j\|)^2} \quad \text{(Cauchy-Schwartz inequality)} \\
&\leqslant \frac{4(2c_1+1)\log^{\frac{3}{2}} k}{k}\sqrt{\lambda_k\sum_i d_i\|u_i\|^2}\sqrt{4\sum_i d_i\|u_i\|^2} \\
&= \frac{8(2c_1+1)\log^{\frac{3}{2}} k}{k}\sqrt{\lambda_k}\sum_i d_i\|u_i\|^2 \\
&= 8(2c_1+1)\log^{\frac{3}{2}} k\sqrt{\lambda_k}
\end{aligned}$$

$\square$

Variance of the Denominator. Here too we will need some groundwork. Let $\mathcal{G}$ denote the Gaussian space. The Hermite polynomials $\{H_i\}_{i\in\mathbb{Z}_{\geqslant 0}}$ form an orthonormal basis for real valued functions over the Gaussian space $\mathcal{G}$, i.e., $\mathbb{E}_{g\in\mathcal{G}}[H_i(g)H_j(g)] = 1$ if $i=j$ and 0 otherwise. The k-wise tensor product of the Hermite basis forms an orthonormal basis for functions over $\mathcal{G}^k$. Specifically, for each $\alpha\in\mathbb{Z}^k_{\geqslant 0}$ define the polynomial H_α as

$$H_\alpha(x_1,\dots,x_k) = \prod_{i=1}^k H_{\alpha_i}(x_i).$$

The functions $\{H_\alpha\}_{\alpha\in\mathbb{Z}^k_{\geqslant 0}}$ form an orthonormal basis for functions over $\mathcal{G}^k$. The degree of the polynomial $H_\alpha(x)$ denote by $|\alpha|$ is $|\alpha| = \sum_i \alpha_i$.

The Hermite polynomials are known to satisfy the following property (see e.g. the book of Ledoux and Talagrand [LT91], Section 3.2).

Fact 3.16. *Let $(g_i,h_i)_{i=1}^k$ be k independent samples from two ρ-correlated Gaussians, i.e., $\mathbb{E}[g_i^2] = \mathbb{E}[h_i^2] = 1$, and $\mathbb{E}[g_ih_i] = \rho$. Then for all $\alpha\in\mathbb{Z}^k_{\geqslant 0}$,*

$$\mathbb{E}[H_\alpha(g_1,\dots,g_k)H_{\alpha'}(h_1,\dots,h_k)] = \rho^{|\alpha|} \textit{ if } \alpha=\alpha' \textit{ and } 0 \textit{ otherwise}$$

Let $A:\mathcal{G}^k\longrightarrow\mathbb{R}$ be the function defined as follows,

$$A(x) = \begin{cases} x_1^2 & \text{if } (x_1\geqslant x_i\forall i\in[k]) \text{ or } (x_1\leqslant x_i\forall i\in[k]) \\ 0 & \text{otherwise} \end{cases}$$

We know that

$$\mathbb{E}[A]\leqslant\frac{4\log k}{k} \text{ and } \mathbb{E}[A^2]\leqslant\frac{16\log^2 k}{k} \text{ (Fact 3.8).}$$

Lemma 3.17. *Let u, v be unit vectors and $g_1, \ldots, g_k$ be i.i.d Gaussian vectors. Then,*

$$\mathbb{E}[A(\langle u, g_1\rangle, \ldots, \langle u, g_k\rangle)A(\langle v, g_1\rangle, \ldots, \langle v, g_k\rangle)] \leqslant \frac{16\log^2 k}{k}\left(\langle u, v\rangle^2 + \frac{1}{k}\right)$$

Proof. The function A on the Gaussian space can be written in the Hermite expansion $A(x) = \sum_\alpha A_\alpha H_\alpha(x)$ such that

$$\sum_\alpha A_\alpha^2 = \mathbb{E}[A^2] \leqslant \frac{16\log^2 k}{k}.$$

Using Fact 3.16, we can write

$$\mathbb{E}[A(\langle u, g_1\rangle, \ldots, \langle u, g_k\rangle)A(\langle v, g_1\rangle, \ldots, \langle v, g_k\rangle)] = (\mathbb{E}[A])^2 + \sum_{\alpha \in \mathbb{Z}^k_{\geqslant 0}, |\alpha| > 0} A_\alpha^2 \rho^{|\alpha|}$$

where $\rho = \langle u, v\rangle$. Since A is an even function, only the even degree coefficients are non-zero, i.e., $A_\alpha = 0$ for all $|\alpha|$ odd. Along with $\rho \leqslant 1$, this implies that

$$\begin{aligned}\mathbb{E}[A(\langle u, g_1\rangle, \ldots, \langle u, g_k\rangle)A(\langle v, g_1\rangle, \ldots, \langle v, g_k\rangle)] &\leqslant (\mathbb{E}[A])^2 + \rho^2\left(\sum_{\alpha, |\alpha| \geqslant 2} A_\alpha^2\right) \qquad \text{where } \rho = \langle u, v\rangle \\ &\leqslant \frac{16\log^2 k}{k^2} + \langle u, v\rangle^2 \frac{16\log^2 k}{k}\end{aligned}$$

□

Next we bound the variance of the denominator.

Proof of Lemma 3.4.

$$\begin{aligned}\mathbb{E}\left[\sum_{i,j} d_i d_j f_i^2 f_j^2\right] &= \sum_{i,j} d_i d_j \|u_i\|^2 \|u_j\|^2\, \mathbb{E}\left[\frac{f_i^2}{\|u_i\|^2}\frac{f_j^2}{\|u_j\|^2}\right] \\ &\leqslant \sum_{i,j} d_i d_j \|u_i\|^2 \|u_j\|^2\, \mathbb{E}\left[A(\langle \tilde{u}_i, g_1\rangle, \ldots, \langle \tilde{u}_i, g_k\rangle)A(\langle \tilde{u}_j, g_1\rangle, \ldots, \langle \tilde{u}_j, g_k\rangle)\rangle\right] \\ &\leqslant \sum_{i,j} d_i d_j \|u_i\|^2 \|u_j\|^2 \cdot \left(\frac{16\log^2 k}{k}(\langle \tilde{u}_i, \tilde{u}_j\rangle^2 + \frac{1}{k})\right) \qquad \text{(Lemma 3.17)} \\ &\leqslant \frac{16\log^2 k}{k} \cdot \left(\sum_{i,j} d_i d_j \langle u_i, u_j\rangle^2 + \frac{1}{k}(\sum_i d_i \|u_i\|^2)^2\right) \\ &\leqslant 32\log^2 k\end{aligned}$$

Therefore

$$\mathsf{Var}\sum_i d_i f_i^2 = \mathbb{E}\left[\sum_{i,j} d_i d_j f_i^2 f_j^2\right] - \left(\mathbb{E}\left[\sum_i d_i f_i^2\right]\right)^2 \leqslant 28\log^2 k.$$

□

Putting It Together

Proof of Theorem 1.4. Now, for each $l \in [k]$, from Lemma 3.2 we get that $\mathbb{E}\left[\sum_i d_i h_l(i)^2\right] = \Theta(\log k)$ and from Lemma 3.4 we get that $\mathsf{Var} \sum_i h_l(i)^2 = \Theta(\log^2 k)$. Therefore, from the One-sided Chebyshev inequality (Fact 3.6), we get

$$\mathbb{P}\left[\sum_i d_i h_l(i)^2 \geqslant \frac{\mathbb{E}\left[\sum_i d_i h_l(i)^2\right]}{2}\right] \geqslant \frac{\left(\frac{\mathbb{E}\left[\sum_i d_i h_l(i)^2\right]}{2}\right)^2}{\left(\frac{\mathbb{E}\left[\sum_i d_i h_l(i)^2\right]}{2}\right)^2 + \mathsf{Var}\sum_i h_l(i)^2} \geqslant c'$$

where c' is some absolute constant.

Therefore, with constant probability, for $\Omega(k)$ indices $l \in [k]$, $\sum_i d_i h_l(i)^2 \geqslant \frac{\mathbb{E}\left[\sum_i d_i h_l(i)^2\right]}{2}$. Also, for each l, with probability $1 - c'/2$, $\sum_{i \sim j} |h_l(i)^2 - h_l(j)^2| \leqslant 2/c' \, \mathbb{E}\left[\sum_{i\sim j} |h_l(i)^2 - h_l(j)^2|\right]$. Therefore, with constant probability, for a constant fraction of the indices $l \in [k]$, we have

$$\frac{\sum_{i\sim j} |h_l(i)^2 - h_l(j)^2|}{\sum_i d_i h_l(i)^2} \leqslant \frac{4}{c'} \frac{\mathbb{E}\left[\sum_{i\sim j} |h_l(i)^2 - h_l(j)^2|\right]}{\mathbb{E}\left[\sum_i d_i h_l(i)^2\right]} = \mathcal{O}(\sqrt{\lambda_k \log k})$$

Applying Lemma 3.1 on the vectors with those indices will give $\Omega(k)$ disjoint sets $S_1, \ldots, S_{ck}$ such that $\phi_G(S_i) = \mathcal{O}(\sqrt{\lambda_k \log k})$ $\forall i \in [ck]$. This completes the proof of Theorem 1.4. □

3.3 Lower bound for k Sparse-Cuts

In this section, we prove a lower bound for the k-sparse cuts in terms of higher eigenvalues (Proposition 1.3) thereby generalizing the lower bound side of the Cheeger's inequality.

Proposition 3.18 (Restatement of Proposition 1.3)**.** *For any edge-weighted graph $G = (V, E)$, for any integer $1 \leqslant k \leqslant |V|$, and for any k disjoint subsets $S_1, \ldots, S_k \subset V$*

$$\max_i \phi_G(S_i) \geqslant \frac{\lambda_k}{2}$$

where $\lambda_1, \ldots, \lambda_{|V|}$ are the eigenvalues of the normalized Laplacian of G.

Proof. Let α denote $\max_i \phi_G(S_i)$. Let $T \stackrel{\text{def}}{=} V \backslash (\cup_i S_i)$. Let G' be the graph obtained by shrinking each piece in the partition $\{T, S_i | i \in [k]\}$ of V to a single vertex. We denote the vertex corresponding to S_i by s_i $\forall i$ and the vertex corresponding to T by t. Let $\mathcal{L}'$ be the normalized Laplacian matrix corresponding to G'. Note that, by construction, the expansion of every set in G' not containing t is at most α.

Let $U \stackrel{\text{def}}{=} \{D^{\frac{1}{2}} X_{S_i} | i \in [k]\}$. Here X_S is the incidence vector of the set $S \subset V$. Since all the vectors in U are orthogonal to each other, we have

$$\lambda_k(\mathcal{L}) = \min_{S : rank(S) = k} \max_{x \in S} \frac{x^T \mathcal{L} x}{x^T x} \leqslant \max_{x \in span(U)} \frac{x^T \mathcal{L} x}{x^T x} = \max_{y \in \mathbb{R}^k * \{0\}} \frac{\sum_{i,j} w'_{ij} (y_i - y_j)^2}{\sum_i w'_i y_i^2}$$

For any $x \in \mathbb{R}$, let $x^+ \stackrel{\text{def}}{=} \max\{x, 0\}$ and $x^- \stackrel{\text{def}}{=} \max\{-x, 0\}$. Then it is easily verified that for any $y_i, y_j \in \mathbb{R}$, $(y_i - y_j)^2 \leqslant 2((y_i^+ - y_j^+)^2 + (y_i^- - y_j^-)^2)$. Therefore,

$$\begin{array}{rcl} \sum_i \sum_{j>i} w'_{ij} (y_i - y_j)^2 & \leqslant & 2(\sum_i \sum_{j>i} w'_{ij} (y_i^+ - y_j^+)^2 + \sum_i \sum_{j>i} w'_{ij} (y_j^- - y_i^-)^2) \\ & \leqslant & 2(\sum_i \sum_{j>i} w'_{ij} |(y_i^+)^2 - (y_j^+)^2| + \sum_i \sum_{j>i} w'_{ij} |(y_j^-)^2 - (y_i^-)^2)| \end{array}$$

Without loss of generality, we may assume that $y_1^+ \geqslant y_2^+ \geqslant \ldots \geqslant y_k^+ \geqslant y_t = 0$. Let $T_i = \{s_1, \ldots, s_i\}$ for each $i \in [k]$. Therefore, we have

$$\sum_i \sum_{j>i} w'_{ij} |(y_i^+)^2 - (y_j^+)^2| \leqslant \sum_{i=1}^{k} ((y_i^+)^2 - (y_{i+i}^+)^2) w'(E(T_i, \bar{T}_i)) \leqslant \alpha \sum_{i=1}^{k} ((y_i^+)^2 - (y_{i+i}^+)^2) w'(T_i) = \alpha \sum_i^k w'_i (y_i^+)^2$$

Here we are using the fact that $w'(E(T_i, \bar{T}_i)) \leqslant \alpha w'(T_i)$ which follows from the definition of α and that $w'(T_{i+1}) - w'(T_i) = w'_{i+1}$. Similiarly, we get that

$$\sum_i \sum_{j>i} w'_{ij} |(y_j^-)^2 - (y_i^-)^2)| \leqslant \alpha \sum_i w'_i (y_i^-)^2 \ .$$

Putting these two inequalities together we get that

$$\sum_{j>i} w'_{ij} (y_i - y_j)^2 \leqslant 2\alpha \sum_i w'_i y_i^2 \ .$$

Therefore, $\lambda_k(\mathcal{L}) \leqslant 2 \max_i \phi_G(S_i)$. ☐

4 Gap examples

In this section, we present constructions of graphs that serve as lower-bounds against natural classes of algorithms. We begin with a family of graphs on which the performance of recursive partitioning algorithms is poor for the k-Sparse cuts problem.

4.1 Recursive Algorithms

Recursive algorithms are one of most commonly used techniques in practice for graph multi-partitioning. However, we show that partitioning a graph into k pieces using a simple recursive algorithm can yield as many $k(1 - o(1))$ sets with expansion much larger than $\sqrt{\lambda_k}\,\mathsf{polylog}k$. Thus this is not an effective method for finding many sparse cuts.

The following construction (Figure 3) shows that partition of V obtained using the recursive algorithm in Figure 1 can give as many as $k(1 - o(1))$ sets have expansion $\Omega(1)$ while $\lambda_k \leqslant \mathcal{O}(k^2/n^2)$.

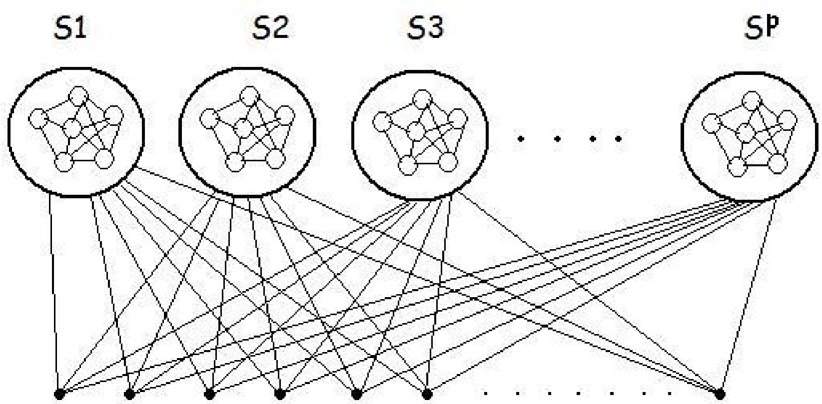

Figure 3: Recursive algorithm can give many sets with very small expansion

In this graph, there are $p \stackrel{\text{def}}{=} k^{\varepsilon}$ sets S_i for $1 \leqslant i \leqslant k^{\varepsilon}$. We will fix the value of ε later. Each of the S_i has $k^{1-\varepsilon}$ cliques $\{S_{ij} : 1 \leqslant j \leqslant k^{1-\varepsilon}\}$ of size n/k which are sparsely connected to each other. The total weight of the edges from S_{ij} to $S_i \backslash S_{ij}$ is equal to a constant c. In addition to this, there are also $k - k^{\varepsilon}$ vertices $v_i : 1 \leqslant i \leqslant k - k^{\varepsilon}$. The weight of edges from S_i to v_j is equal to $k^{-\varepsilon}$.

Claim 4.1. *1.* $\phi(S_{ij}) \leqslant (c+1)k^2/n^2 \ \forall i, j$

2. $\phi(S_i) \leqslant 1/(c+1)\phi(S_{ij}) \ \forall i, j$

3. $\lambda_k = \mathcal{O}(m^2/n^2)$

Proof. 1.

$$\phi(S_{ij}) = \frac{c + \frac{(m-m^{\varepsilon})m^{-\varepsilon}}{m^{1-\varepsilon}}}{(\frac{n}{m})^2 + c + \frac{(m-m^{\varepsilon})m^{-\varepsilon}}{m^{1-\varepsilon}}} \leqslant \frac{(c+1)m^2}{n^2}$$

2. $w(S_i) = \sum_j w(S_{ij})$, but for each S_{ij} only $1/(c+1)$ fraction of edges incident at S_{ij} are also incident at S_i. Therefore, $\phi(S_i) \leqslant 1/(c+1)\phi(S_{ij})$.

3. Follows from (1) and Proposition 1.3.

□

For appropriate values of ε and k, the partition output by the recursive algorithm will be $\{S_i : i \in [k^{\varepsilon}]\} \cup \{v_i : i \in [k - k^{\varepsilon}]\}$. Hence, $k(1 - o(1))$ sets have expansion equal to 1.

4.2 k-partition

In this section, we give a constructive proof of Theorem 1.7, i.e., we construct a family of graphs such that for any k-partition $\{S_1, \ldots, S_k\}$ of the graph, $\max_i \phi(S_i) > \Theta(k^2\sqrt{\frac{p}{n}})$. We view this as further evidence suggesting that the k- sparse-cuts problem is the right generalization of sparsest cut.

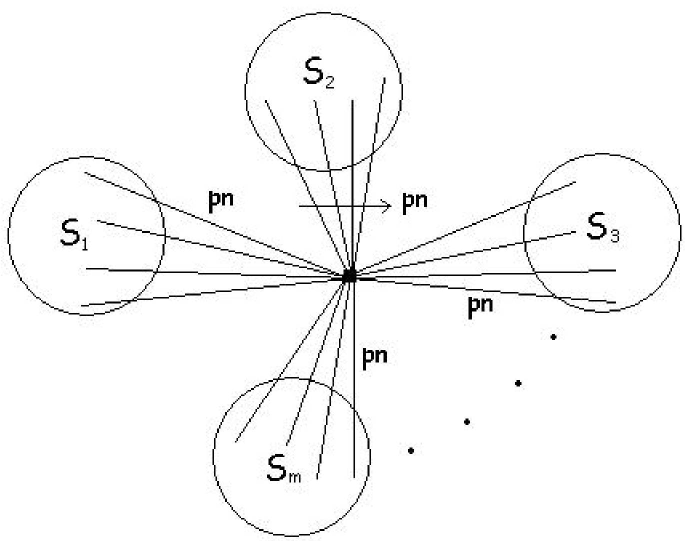

Figure 4: k-partition can have sparsity much larger than $\Omega(\sqrt{\lambda_k}\mathsf{polylogk})$

Lemma 4.2. *For the graph G in Figure 4, and for any k-partition $S_1, \ldots, S_k$ of its vertex set,*

$$\frac{\max_i \phi_G(S_i)}{\sqrt{\lambda_k}} = \Theta(k^2\sqrt{\frac{p}{n}})$$

Proof. In Figure 4, $\forall i \in [k]$, S_i is a clique of size $(n-1)/k$ (pick n so that $k|(n-1)$). There is an edge from central vertex v to every other vertex of weight pn. Here p is some absolute constant. Let $\mathcal{P}' \stackrel{\text{def}}{=} \{S_1 \cup \{v\}, S_2, S_3, \ldots, S_k\}$. For $n > k^3$, it is easily verified that the optimum k-partition is isomorphic to $\mathcal{P}'$. For $k < o(n^{\frac{1}{3}})$, we have

$$\max_{S_i \in \mathcal{P}'} \phi_G(S_i) = \phi_G(S_1 \cup \{v\}) = \frac{pnk}{\left(\frac{n-1}{k}\right)^2 + pnk} = \Theta\left(\frac{pk^3}{n}\right)$$

Applying Proposition 1.3 to $S_1, \ldots, S_k$, we get that $\lambda_k = \mathcal{O}(pk^2/n)$. Thus we have the lemma. □

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer, *Subexponential algorithms for unique games and related problems*, FOCS, 2010. 4

[Alo86] Noga Alon, *Eigenvalues and expanders*, Combinatorica **6** (1986), no. 2, 83–96. 2

[AM85] Noga Alon and V. D. Milman, *λ_1, isoperimetric inequalities for graphs, and superconcentrators*, J. Comb. Theory, Ser. B **38** (1985), no. 1, 73–88. 2

[AR98] Yonatan Aumann and Yuval Rabani, *An $\mathcal{O}$(log k) approximate min-cut max-flow theorem and approximation algorithm*, SIAM J. Comput. **27** (1998), no. 1, 291–301. 4

[ARV04] Sanjeev Arora, Satish Rao, and Umesh V. Vazirani, *Expander flows, geometric embeddings and graph partitioning*, STOC (László Babai, ed.), ACM, 2004, pp. 222–231. 4

[BFK+11] Nikhil Bansal, Uriel Feige, Robert Krauthgamer, Konstantin Makarychev, Viswanath Nagarajan, Joseph Naor, and Roy Schwartz, *Min-max graph partitioning and small set expansion*, FOCS (Rafail Ostrovsky, ed.), IEEE, 2011, pp. 17–26. 5

[BLR10] Punyashloka Biswal, James R. Lee, and Satish Rao, *Eigenvalue bounds, spectral partitioning, and metrical deformations via flows*, J. ACM **57** (2010), no. 3. 5

[Bor85] Christer Borell, *Geometric bounds on the ornstein-uhlenbeck velocity process*, Probability Theory and Related Fields **70** (1985), 1–13, 10.1007/BF00532234. 4

[Chu97] Fan Chung, *Spectral graph theory*, American Mathematical Society, 1997. 8

[CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev, *Near-optimal algorithms for unique games*, STOC, 2006, pp. 205–214. 5, 11

[Kel06] Jonathan A. Kelner, *Spectral partitioning, eigenvalue bounds, and circle packings for graphs of bounded genus*, SIAM J. Comput. **35** (2006), no. 4, 882–902. 5

[KVV04] Ravi Kannan, Santosh Vempala, and Adrian Vetta, *On clusterings: Good, bad and spectral*, J. ACM **51** (2004), no. 3, 497–515. 2, 5

[LGT12] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan, *Multi-way spectral partitioning and higher-order cheeger inequalities*, STOC (2012). 5

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich, *The geometry of graphs and some of its algorithmic applications*, Combinatorica **15** (1995), no. 2, 215–245. 4

[LR99] Frank Thomson Leighton and Satish Rao, *Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms*, J. ACM **46** (1999), no. 6, 787–832. 4

[LRTV11] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala, *Algorithmic extensions of cheeger's inequality to higher eigenvalues and partitions*, APPROX-RANDOM, Lecture Notes in Computer Science, vol. 6845, Springer, 2011, pp. 315–326. 4

[LT91] Michel Ledoux and Michel Talagrand, *Probability in banach spaces*, Springer, 1991. 14

[RS10] Prasad Raghavendra and David Steurer, *Graph expansion and the unique games conjecture*, STOC (Leonard J. Schulman, ed.), ACM, 2010, pp. 755–764. 4

[RST10a] Prasad Raghavendra, David Steurer, and Prasad Tetali, *Approximations for the isoperimetric and spectral profile of graphs and related parameters*, STOC, 2010, pp. 631–640. 4

[RST10b] Prasad Raghavendra, David Steurer, and Madhur Tulsiani, *Reductions between expansion problems*, Manuscript, 2010. 4

[ST96] Daniel A. Spielman and Shang-Hua Teng, *Spectral partitioning works: Planar graphs and finite element meshes*, FOCS, IEEE Computer Society, 1996, pp. 96–105. 5