Finding Sparse Cuts via Cheeger Inequalities for Higher Eigenvalues

Journal:	*Journal of the ACM*
Manuscript ID:	JACM-00148-2012.R1
Manuscript Type:	Paper
Date Submitted by the Author:	08-Sep-2014
Complete List of Authors:	Louis, Anand; Georgia Institute of Technology, School of Computer Science Raghavendra, Prasad; University of California, Berkeley, EECS Tetali, Prasad; Georgia Institute of Technology, School of Computer Science Vempala, Santosh; Georgia Institute of Technology, School of Computer Science
Computing Classification Systems:	G.2.2 Graph Theory, G.1.2 Approximation, F.2.1 Numerical Algorithms and Problems

Finding Sparse Cuts via Cheeger Inequalities for Higher Eigenvalues

Anand Louis[*]
Georgia Tech
anandl@gatech.edu

Prasad Raghavendra[†]
UC Berkeley
prasad@cs.berkeley.edu

Prasad Tetali[‡]
Georgia Tech
tetali@cc.gatech.edu

Santosh Vempala[*]
Georgia Tech
vempala@gatech.edu

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:

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

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 two natural generalizations of the sparsest cut in a graph:

- A collection of k disjoint subsets $S_1, \ldots, S_k$ of the vertex set that minimize $\max_{i \in [k]} \phi(S_i)$;
- A subset of weight $O(1/k)$ of the total weight of the graph having minimum expansion.

Our main results are extensions of Cheeger's classical inequality to these problems via higher eigenvalues of the graph Laplacian. 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 λ_k is the k^{th} smallest eigenvalue of the normalized Laplacian matrix, and 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 $O(1/k)$ fraction of the total weight and $\phi(S) \leqslant C\sqrt{\lambda_k \log k}$. These 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:

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

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 like error correcting codes [SS96], sampling algorithms [SJ89], metric embeddings [LLR95], among others. They are also of interest to practitioners, as algorithms for graph partitioning can be used as fundamental building blocks in many applications such as image segmentation [SM00], clustering [Dhi01] , parallel computation [KGGK94] and VLSI placement and routing [AK95]. 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.

Spectral methods have been very succesfully applied for graph partitioning both in theory and practice. The central object in spectral graph theory – the graph Laplacian is a matrix associated with a graph. Roughly speaking, the Laplacian $\mathcal{L}$ is given by $\mathcal{L} = I - A$ where A is the adjacency matrix whose edges have been reweighed so that the degree of each vertex is 1 (See Section 1.2 for a precise definition).

The eigenvalues of the Laplacian matrix yield insights about structural properties of the graph. Suppose $\lambda_1 = 0 \leqslant \lambda_2 \leqslant \cdots \lambda_n$ are the eigenvalues of the Laplacian matrix $\mathcal{L}$ of a graph G. A basic result in spectral graph theory asserts the following.

Theorem 1.1. *A graph G is disconnected if and only if $\lambda_2 = 0$.*

The fundamental Cheeger's inequality can be thought of as a *robust* version of Theorem 1.1. Specifically, if λ_2 is very small, the graph is close to being disconnected in the sense that it can be disconnected by deleting a small fraction of edges. More precisely, Cheeger's inequality establishes a bound on expansion via the spectrum of the graph. Cheeger's inequality was shown originally for manifolds in [Che70], and was established in the case of graphs in [Alo86, AM85].

Theorem 1.2 (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.

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

The proof of Cheeger's inequality, which can be made algorithmic, uses the eigenvector corresponding to the second smallest eigenvalue. This theorem and its many variants have played a major role in the design of algorithms, for e.g. see [PSL90, DS91, SS96, SM00, ARV09] etc.

In this work, we obtain analogues of Cheeger's inequality for higher eigenvalues along with simple and efficient algorithms for graph partitioning. Towards stating our results, we first recall the following simple generalization of Theorem 1.1.

Theorem 1.3. *A graph G has at least k connected components if and only if $\lambda_k = 0$.*

The above theorem suggests that if G has $\lambda_k \approx 0$ then the graph G must have k subsets of vertices that are only sparsely connected to the rest. We will show a robust version of the above theorem making this intuition precise.

Analogous to ϕ_G, we will define a parameter $\phi_k(G)$ associated with existence of k non-expanding subsets.

Problem 1.4 (k-sparse-cuts)**.** Given an edge weighted graph $G = (V, E)$ and an integer $k > 1$, define

$$\phi_k(G) = \min_{\text{disjoint sets } S_1,\ldots,S_k \subseteq V} \max_{i\in[k]} \phi_G(S_i)$$

We will refer to the problem of finding k disjoint non-empty subsets $S_1, \ldots, S_k \subset V$ such that $\max_i \phi_G(S_i)$ is minimized a.k.a approximating $\phi_k(G)$ as the *k-sparse cuts* problem.

Observe that $\phi_2(G) = \phi(G)$. We stress 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.

First, it is fairly straightforward to show that the k^{th} smallest eigenvalue of the normalized Laplacian of the graph gives a lower bound on $\phi_k(G)$. Formally, we have the following lower bound.

Proposition 1.5. *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 1.6. *For every edge-weighted graph $G = (V, E)$, and any integer $1 \leqslant k \leqslant |V|$, there exist $\Omega(k)$ disjoint subsets $S_1, \ldots, S_{\Omega(k)}$ of vertices such that*

$$\max_i \phi_G(S_i) \leqslant O\left(\sqrt{\lambda_k \log k}\right).$$

where $\lambda_1, \ldots, \lambda_{|V|}$ are the eigenvalues of the normalized Laplacian of G. (Here Ω and O notation only hide fixed absolute constant factors). Moreover, the sets $S_1, \ldots, S_k$ satisfying the inequality can be identified in polynomial time.

An appealing feature of our proof of Theorem 1.6 is that it is constructive. In fact, we present a rather simple algorithm that takes the top k eigen vectors of the graph G as input and produces $\Omega(k)$ sets promised by Theorem 1.6. Roughly speaking, the algorithm can be described as follows. The top k eigen vectors embed the vertices of the graph in $\mathbb{R}^k$. Pick k random directions in $\mathbb{R}^k$ and assign each vertex in the graph to the direction it aligns most with. Now for each of the k-directions, sort the vertices assigned to it by length, and output the prefix in the sorted order with the least expansion. A formal description of the algorithm is presented in Section 3.

The running time is dominated by the time taken to compute the smallest k eigenvectors of the normalized Laplacian. Furthermore, the guarantee of the algorithm holds as long as the embedding of the vertices in to $\mathbb{R}^k$ satisfying two key properties – the embedding has a small Rayleigh coefficient and is *truly* k-dimensional (see Theorem 3.2) . The second requirement is analogous to being in *isotropic position* and forces the embedding to have a variance along $\Omega(k)$ orthogonal directions. In particular, the algorithm does not need the top k-eigenvectors exactly but only a geometric embedding with a few robust properties.

The upper bound obtained in Theorem 1.6 is tight, and is is matched by the family of infinite graphs referred to as *Gaussian graphs*. Specifically, we will show the following in Section 6.

Theorem 1.7. *For every $\epsilon > 0$ and positive integers $2 \leqslant k < K$, there exists families of graphs $N_{K,\epsilon}$ such that,*

$$\phi_k(G) \geqslant O(\sqrt{\lambda_K \log k}) .$$

It is natural to wonder if the above bounds extend to the case when the k-sets are required to form a partition. 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 $O(\sqrt{\lambda_k} \mathsf{polylog} k)$ in general.

Theorem 1.8. *For every k, there exists graphs such that for every k-partition $\{S_1, \ldots, S_k\}$ of the vertex set*

$$\max_i \phi_G(S_i) \geqslant \sqrt{\lambda_k} \cdot \Omega(\sqrt{k}).$$

Our main theorem (Theorem 1.6) immediately yields bounds on expansion of small sets in graphs in terms of their spectra. The small set expansion problem is defined as follows.

Problem 1.9 (Small-set expansion)**.** Given an edge weighted graph $G = (V, E)$ and $\delta \in (0, 1/2)$, let

$$\phi_G(\delta) \stackrel{\text{def}}{=} \min_{\substack{S \subset V, \\ w(S) \leqslant \delta \cdot w(V)}} \phi(S) .$$

The problem of approximating $\phi_G(\delta)$ for a given graph G and a constant δ is known as the small-set expansion problem.

The small-set expansion problem arises naturally in the context of understanding the Unique Games Conjecture (see [RS10, ABS10]). As an immediate consequence of Theorem 1.6, we get the following bound on the small-set expansion of a graph. Again, the bound is tight for the case of Gaussian graphs.

Corollary 1.10. *For any edge-weighted graph $G = (V, E)$ and any integer $1 \leqslant k \leqslant |V|$,*

$$\phi_G\left(\frac{1}{k}\right) \leqslant O(\sqrt{\lambda_k \log k}) .$$

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 $O(\sqrt{\mathsf{OPT}})$ approximation algorithm for the sparsest cut problem. Leighton and Rao [LR99] gave an $O(\log n)$

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 $O(\sqrt{\log n})$ via a semi-definite relaxation and embeddings of special metrics [ARV09]). 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 $O(\sqrt{\lambda_{k^2} \log k})$ in [LRTV11], and by $O(\sqrt{\lambda_{k^{100}} \log_k n})$ in [ABS10]. Using a semidefinite programming relaxation, [RST10] gave an algorithm that outputs a small set with expansion at most $\sqrt{\mathsf{OPT} \log k}$ where OPT is the sparsity of the optimal set of size at most $O(1/k)$. Bansal et. al.[BFK+11] obtained an $O(\sqrt{\log n \log k})$ approximation algorithm also using a semidefinite programming relaxation.

A problem related to the k-sparse cuts 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. Bansal et. al.[BFK+11] study the problem of partitioning a graph into k pieces to minimize the largest edge boundary. They give a bi-criteria approximation algorithm for this problem.

In independent work, [LOGT12] have obtained results similar to Theorem 1.6 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 $O(k^3 \sqrt{\lambda_k})$. 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

Fix $G = (V, E, w)$ to be a finite weighted undirected graph. Let A be its (weighted) adjacency matrix and let $w(\{i, j\}) \in \mathbb{R}^+$ denote the weight of the edge $\{i, j\} \in E$. We let $d_i \stackrel{\text{def}}{=} \sum_{j \sim i} w(\{i, j\})$ denote the (weighted) degree of vertex i. We use D to denote the diagonal matrix with $D_{ii} = d_i$. The Laplacian of the graph G is given by

$$L \stackrel{\text{def}}{=} D - A ,$$

and the *normalized* Laplacian of the graph G is given by

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

Given any non-zero vector $u \in \mathbb{R}^n$, we denote by $\bar{u}$ the unit vector in the direction of u, i.e. $\bar{u} \stackrel{\text{def}}{=} u/\|u\|$. We will denote the standard normal distribution by $\mathcal{N}(0, 1)$. We will use $\mathcal{N}(0, 1)^k$ to denote k-dimensional standard normal distribution where each coordinate is independently sampled from $\mathcal{N}(0, 1)$.

2 Geometric Embeddings

Rayleigh Quotient . For a graph $G = (V, E, w)$, let $f : V \to \mathbb{R}$ be an embedding of G on to $\mathbb{R}$. The Rayleigh quotient of f is given by

$$\mathcal{R}(f) \stackrel{\text{def}}{=} \frac{f^T \mathcal{L}_G f}{f^T f} .$$

If we let $g = D^{-1/2} f$, through a straightforward computation we get that

$$\mathcal{R}(f) = \frac{g^T Lg}{g^T Dg} = \frac{\sum_{i\sim j} w(\{i,j\})\,(g(i)-g(j))^2}{\sum_i d_i g(i)^2}\,.$$

We can generalize this definition to higher dimensional embeddings. Let $f : V \to \mathbb{R}^d$ be an embedding of the graph G in to d-dimensional Euclidean space $\mathbb{R}^d$ for some positive integer d. Then

$$\mathcal{R}(f) \stackrel{\text{def}}{=} \frac{\sum_{i\sim j} w(\{i,j\})\,\|f(i)-f(j)\|_2^2}{\sum_i d_i\,\|f(i)\|^2}\,.$$

It is clear that the Rayleigh quotient of an embedding f measures the ratio between the averaged squared length of the edges to the average squared length of vectors in the embedding.

Dimensionality. As we will be concerned with multi-dimensional embeddings of graphs, we introduce a measure of dimensionality, and show that it nicely captures the intuitive notion of dimension of an embedding.

Definition 2.1. For an embedding $f : V \to \mathbb{R}^d$ we define $\mathcal{D}(f)$ as follows.

$$\mathcal{D}(f) \stackrel{\text{def}}{=} \frac{\sum_{i,j\in E} d_i d_j\,\|f(i)\|_2^2 \cdot \|f(j)\|_2^2}{\sum_{i,j\in E} d_i d_j\,\langle f(i), f(j)\rangle^2}\,.$$

Note that the roles of i, j in the above definition are symmetric and in fact the numerator of $\mathcal{D}(f)$ is equal to $\left(\sum_i d_i\,\|f(i)\|_2^2\right)^2$.

The $\mathcal{D}(f)$ of an embedding f is a measure of the *true-dimensionality* of the embedding. For example, for a one-dimensional embedding $f : V \to \mathbb{R}$ we have

$$\sum_{i,j} d_i d_j\,\langle f(i), f(j)\rangle^2 = \sum_{i,j} d_i d_j f^2(i) f^2(j)\,,$$

implying that its $\mathcal{D}(f) = 1$. The following lemma further supports this intepretation of the measure $\mathcal{D}(f)$.

Lemma 2.2. *Suppose the image of the embedding $f : V \to \mathbb{R}^d$ lies in a k-dimensional subspace of $\mathbb{R}^d$, then $\mathcal{D}(f) \leqslant k$.*

Proof. By an appropriate basis change in $\mathbb{R}^d$, we can assume without loss of generality that for each $i \in V$ all but first k coordinates of the vector $f(i)$ are zero. Let $f(i)[\ell]$ denote the ℓ^{th} coordinate of $f(i)$. Now we will bound the numerator of $\mathcal{D}(f)$ as shown below.

$$\begin{aligned}
\sum_{i,j} d_i d_j \langle f(i), f(j)\rangle^2 &= \sum_{i,j} d_i d_j \left(\sum_{\ell=1}^{k} f(i)[\ell] f(j)[\ell]\right)^2 = \sum_{i,j} d_i d_j \left(\sum_{\ell,\ell'=1}^{k} f(i)[\ell] f(i)[\ell'] f(j)[\ell] f(j)[\ell']\right) \\
&= \sum_{\ell,\ell'=1}^{k} \left(\sum_i d_i f(i)[\ell] f(i)[\ell']\right)\left(\sum_j d_i f(j)[\ell] f(j)[\ell']\right) \\
&= \sum_{\ell,\ell'=1}^{k} \left(\sum_i d_i f(i)[\ell] f(i)[\ell']\right)^2 \qquad (1) \\
&\geqslant \sum_{\ell=1}^{k} \left(\sum_i d_i f(i)[\ell] f(i)[\ell]\right)^2 \\
&\geqslant \frac{1}{k}\left(\sum_i d_i \sum_{\ell=1}^{k} f(i)[\ell] f(i)[\ell]\right)^2 = \frac{1}{k}\left(\sum_i d_i \left\|f(i)\right\|_2^2\right)^2
\end{aligned}$$

The penultimate inequality in the above calculation was obtained via an application of Cauchy-Schwartz inequality. From the above calculation, we have $\mathcal{D}(f) \leqslant k$. □

Conversly, $\mathcal{D}(f)$ for an isotropic embedding is equal to the dimension of the ambient space as shown in the following lemma.

Lemma 2.3. *Suppose $f : V \to \mathbb{R}^k$ be an embedding that places the vertices V in an isotropic position in $\mathbb{R}^k$, i.e.,*

$$\sum_i d_i f(i)[\ell] f(i)[\ell'] = \begin{cases} 1 & \text{if } \ell = \ell' \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

then $\mathcal{D}(f) = k$.

Proof. Borrowing the calculation from (1), the denominator of $\mathcal{D}(f)$ is given by

$$\begin{aligned}
\sum_{i,j} d_i d_j \langle f(i), f(j)\rangle^2 &= \sum_{\ell,\ell'=1}^{k} \left(\sum_i d_i f(i)[\ell] f(i)[\ell']\right)^2 \\
&= k \qquad (\text{ Using equation (2)})\,.
\end{aligned}$$

Further since,

$$\sum_i d_i \left\|f(i)\right\|^2 = \sum_\ell \sum_i d_i (f(i)[\ell])^2 = k\,,$$

we can conclude that

$$\mathcal{D}(f) = \frac{\left(\sum_i d_i \left\|f(i)\right\|_2^2\right)^2}{\sum_{i,j} d_i d_j \langle f(i), f(j)\rangle^2} = \frac{(k)^2}{k} = k\,.$$

□

Spectral Embedding. Let $0 = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$ denote the eigenvalues of $\mathcal{L}_G$. Let $\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_n : V \to \mathbb{R}$ denote the corresponding eigenvectors. The eigenvectors $\mathsf{v}_1, \ldots \mathsf{v}_n$ form an orthonormal set of vectors, i.e.,

$$\langle \mathsf{v}_a, \mathsf{v}_b \rangle = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases} .$$

Let $v_a \stackrel{\text{def}}{=} D^{-\frac{1}{2}} \mathsf{v}_a$ for each $a \in [n]$. Then,

$$\mathsf{v}_a^T \mathcal{L}_G \mathsf{v}_a = \sum_{i \sim j} w(\{i, j\}) (v_a(i) - v_a(j))^2 .$$

By orthonormality of the vectors $\{\mathsf{v}_a\}$ we will have,

$$\sum_{i \in V} d_i v_a(i) v_b(i) = \langle \mathsf{v}_a, \mathsf{v}_b \rangle = \delta_{ab} .$$

Hence for each $\ell \in [n]$, the set of vectors $v_1, \ldots v_\ell$ yield an isotropic embedding of the graph in to $\mathbb{R}^\ell$. For the sake of concreteness, we state this observation formally below.

Lemma 2.4. *For $k \in [n]$, the embedding $f : V \to \mathbb{R}^k$ given by the top k-eigenvectors $\mathsf{v}_1, \ldots, \mathsf{v}_k$,i.e.,*

$$f(i) = \frac{1}{\sqrt{d_i}} (\mathsf{v}_1(i), \ldots, \mathsf{v}_k(i))$$

is an isotropic embedding satisfying

$$\mathcal{D}(f) = k \qquad \textit{and} \qquad \mathcal{R}(f) \leqslant \lambda_k .$$

3 Gaussian Projection Algorithm

In this section, we will present a simple algorithm that takes as input a k-dimensional embedding of a graph G and produces $\Omega(k)$ nonexpanding subsets of G. In particular, given the k-dimensional spectral embedding (Lemma 2.4) the algorithm will produce the $\Omega(k)$-nonexpanding sets, thereby proving the main result of the paper (Theorem 1.6).

The geometric partitioning algorithm that we present is very simple and can be described as follows. The input is a k-dimensional embedding $u : V \to \mathbb{R}^d$ of a graph $G = (V, E, w)$. The algorithm picks k random directions $g_1, \ldots, g_k$ (For technical reasons, we pick $g_1, \ldots, g_k$ to be independent Gaussian vectors). Partition the vertices by assigning a vertex i to the direction g_l along which $u(i)$ has the largest projection. On each subset of the partition, the idea is to sort the vertices by the length of the corresponding vectors. Try all prefixes of the sorted order and output the one with the smallest expansion. A more formal description is presented below.

We will prove the following guarantee for Algorithm 3.1.

Theorem 3.2. *There exists universal constants c_0 and C such that the following holds: given a graph $G = (V, E, w)$, a parameter $k \in \mathbb{Z}_{\geqslant 0}$ and an embedding $u : V \to \mathbb{R}^d$ such that $\mathcal{D}(u) \geqslant k$, with constant probability Algorithm 3.1 outputs $c_0 \cdot k$ non-empty disjoint sets each with expansion at most $C\sqrt{\mathcal{R}(u) \log k}$.*

Notice that Theorem 1.6 follows directly from Theorem 3.2 by using the spectral embedding. The details are as follows.

Theorem 1.6. Invoking Theorem 3.2 with the spectral embedding given by the top k eigenvectors (Lemma 2.4) yields that G has $c_0 k$ non-empty disjoint subsets each having expansion at most $C\sqrt{\lambda_k \log k}$. □

Algorithm 3.1 (Random projection rounding). Input: Graph $G = (V, E, w)$, parameter k and an embedding $u : V \to \mathbb{R}^d$ such that $\mathcal{D}(u) \geqslant k$.

(*Partitioning via Random Projections*)

1. Pick k independent Gaussian vectors $g_1, g_2, \ldots, g_k \sim \mathcal{N}(0, 1)^d$.
2. Partition the vertices in to k sets $T_1, \ldots, T_k$ by assigning each vertex to the direction it aligns most with $g_1, \ldots, g_k$.

   Formally, for each vertex $a \in V$

$$a \in T_\ell \iff \ell = \mathrm{argmax}_{i \in [k]} \{ | \langle u(a), g_i \rangle | \}$$

(*Cheeger Cut*)

For each T_ℓ do

1. Sort the vectors in T_ℓ in decreasing order of their lengths. For $m \in \{1, \ldots, |T_\ell|\}$, let $T_{\ell,m}$ denote the m longest vectors in T_ℓ.
2. Output the set with the least expansion amongst $\{T_{\ell,m}\}_{m=1}^{|T_\ell|}$

Figure 1: The Many-sparse-cuts Algorithm

3.1 Proof Overview

In this section, we will present an overview of the analysis of the Random Projection algorithm (Algorithm 3.1). It is easy to see that all the sets output by the algorithm are pairwise disjoint. This follows from the fact that the sets $\{T_\ell\}_{\ell=1}^k$ form a partition and the algorithm outputs exactly one subset of each T_ℓ.

Towards analyzing the expansion of the sets output by the algorithm, define vectors $h_1, h_2, \ldots, h_k \in \mathbb{R}^V$ as follows:

$$h_i(a) = \begin{cases} \|u(a)\|^2 & \text{if } i = \mathrm{argmax}_{i \in [k]} \{ \langle u(a), g_i \rangle \} \\ 0 & \text{otherwise.} \end{cases}$$

Notice that the vectors $h_1, \ldots, h_k$ are random variables depending on the choice of the random Gaussian vectors $\{g_i\}_{i=1}^k$. Further, the random variables h_i are identically distributed.

Algorithm 3.1 outputs superlevel sets of the vectors h_i, i.e., each subset $T_{\ell,m}$ is of the form,

$$T_{\ell,m} = \{ v | h_\ell(v) \geqslant \theta_{\ell,m} \} ,$$

for some threshold $\theta_{\ell,m}$. In fact, Algorithm 3.1 can be equivalently described as running the standard Cheeger cut algorithm on each of the vectors $\{h_i\}_{i=1}^k$.

We will show that for each $i \in \{1, \ldots, k\}$, the vector h_i has a constant probability of yielding a set with small expansion (less than $O(\sqrt{\mathcal{R}(u) \log k})$. This implies that the algorithm finds in expectation $\Omega(k)$ sets with expansion less than $O(\sqrt{\mathcal{R}(u) \log k})$. In turn, this shows that with constant probability, the algorithm finds $\Omega(k)$ disjoint sets with expansion less than the desired bound.

Since the vectors $\{h_i\}$ are identically distributed, it is sufficient to prove that h_1 yields a set with the desired expansion with constant probability. The quality of the cut obtained from a superlevel set of f can be upper bounded using the following standard lemma. A proof of this lemma can be found in [Chu97] (see Theorem 2.9).

Lemma 3.3. *Suppose $f : V \to \mathbb{R}^+$ is such that its support satisfies $|\mathrm{supp}(f)| \leqslant |V|/2$ and*

$$\frac{\sum_{i\sim j} w(\{i,j\})\,|f(i)-f(j)|}{\sum_i d_i f(i)} \leqslant \varepsilon\,.$$

Then one of the superlevel sets of f, given by

$$S_\theta = \{i | f(i) \geqslant \theta\} \qquad \text{for some } \theta \in \mathbb{R}^+\,,$$

satisfies $\phi_G(S_\theta) \leqslant \varepsilon$.

Applying Lemma 3.3, the expansion of the set retrieved from $f = h_1$ is upper bounded by,

$$\frac{\sum_{i\sim j} w(\{i,j\})\,|f(i)-f(j)|}{\sum_i d_i f(i)}\,. \tag{3}$$

Both the numerator and denominator are random variables depending on the choice of random Gaussians $g_1, \ldots, g_k$. Without loss of generality, we may normalize the embedding so that,

$$\sum_i d_i\, \|u(i)\|^2 = k\,. \qquad \text{(Normalization)} \tag{4}$$

It is a fairly straightforward calculation to bound the expected value of the denominator.

Lemma 3.4. *(Expectation of denominator)*

$$\mathbb{E}\left[\sum_i d_i f(i)\right] = 1\,.$$

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

Lemma 3.5. *(Expectation of numerator)*

$$\mathbb{E}\left[\sum_{i\sim j} w(\{i,j\})\,|f(i)-f(j)|\right] \leqslant O\left(\sqrt{\mathcal{R}(u)\log k}\right)\,.$$

Notice that the ratio of the expected values in Lemma 3.5 and Lemma 3.4 is $O(\sqrt{\mathcal{R}(u)\log k})$, 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 3.6. *(Variance of denominator)*

$$\mathrm{Var}\left[\sum_i d_i f(i)\right] \leqslant 1\,.$$

The above moment bounds are sufficient to conclude that with constant probability, the ratio

$$\frac{\sum_{i\sim j}|f(i)-f(j)|}{\sum_i d_i f(i)} = O\left(\sqrt{\mathcal{R}(u)\log k}\right).$$

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 $O(\sqrt{\mathcal{R}(u)\log k})$. The details of the proof are presented below.

Proof of Theorem 3.2. For each $l \in [k]$, from Lemma 3.4 and Lemma 3.6 we get that

$$\mathbb{E}\left[\sum_i d_i h_l(i)\right] = 1 \qquad \text{and} \qquad \mathsf{Var}\left[\sum_i d_i h_l(i)\right] \leqslant 1\,.$$

Therefore, from the One-sided Chebyshev inequality (Lemma 4.1), we get

$$\mathbb{P}\left[\sum_i d_i h_l(i) \geqslant \frac{1}{2}\right] \geqslant \frac{\left(\frac{\mathbb{E}\left[\sum_i d_i h_l(i)\right]}{2}\right)^2}{\left(\frac{\mathbb{E}\left[\sum_i d_i h_l(i)\right]}{2}\right)^2 + \mathsf{Var}\left[\sum_i d_i h_l(i)\right]} \geqslant c' \tag{5}$$

where c' is some absolute constant. Therefore, with constant probability, for $\Omega(k)$ indices $l \in [k]$, $\sum_i d_i h_l(i) \geqslant 1/2$. Next, for each l, using Markov's inequality

$$\mathbb{P}\left[\sum_{i\sim j} w(\{i,j\})\,|h_l(i)-h_l(j)| \leqslant \frac{2}{c'}\,\mathbb{E}\left[\sum_{i\sim j} w(\{i,j\})\,|h_l(i)-h_l(j)|\right]\right] \geqslant 1-\frac{c'}{2}\,. \tag{6}$$

Therefore, with constant probability, for a constant fraction, say c_0, of the indices $l \in [k]$, we have

$$\frac{\sum_{i\sim j} w(\{i,j\})\,|h_l(i)-h_l(j)|}{\sum_i d_i h_l(i)} \leqslant \frac{4}{c'}\,\frac{\mathbb{E}\left[\sum_{i\sim j} w(\{i,j\})\,|h_l(i)-h_l(j)|\right]}{\mathbb{E}\left[\sum_i d_i h_l(i)\right]} = C\sqrt{\mathcal{R}(u)\log k}$$

for some constant C. Applying Lemma 3.3 on the vectors with those indices will give $c_0 k$ disjoint sets $S_1,\ldots,S_{ck}$ such that $\phi_G(S_i) = C\sqrt{\lambda_k \log k}\ \forall i \in [c_0 k]$. This completes the proof of Theorem 3.2. □

4 Main Proofs

In this section, we will present the proofs of Lemma 3.4, Lemma 3.5 and Lemma 3.6. To this end, we begin by recalling a few simple facts.

4.1 Technical Preliminaries

Fact 4.1 (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 fact is folklore about Gaussians. We refer the reader to the work of Charikar et. al.[CMM06] (see Theorem 4.1) for the proof of a more general claim.

Fact 4.2. *Let $X_1, \dots, X_k$ and $Y_1, \dots, Y_k$ 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 $1 - \varepsilon$. Then*

$$\mathbb{P}\left[\mathrm{argmax}_i X_i \neq \mathrm{argmax}_i Y_i\right] \leqslant c_1 \left(\sqrt{\varepsilon \log k}\right) .$$

for some absolute constant c_1.

4.2 Expectation of the Denominator.

Let f denote the vector h_ℓ. The choice of the index ℓ is arbitrary and the same analysis is applicable to all indices $\ell \in [k]$. Recall that Lemma 3.4 asserts the following bound on the expectation of the denominator in (3).

$$\mathbb{E}\left[\sum_i d_i f(i)\right] = 1 .$$

Proof of Lemma 3.4. For any $i \in [n]$, recall that

$$f(i) = \begin{cases} \|u(i)\|^2 & \text{if } \langle u(i), g_1 \rangle \geqslant \left\langle u(i), g_j \right\rangle \ \forall j \in [k] \\ 0 & \text{otherwise} \end{cases} . \tag{7}$$

The first case happens with probability $1/k$ and so $f(i) = 0$ with the remaining probability. Therefore

$$\mathbb{E}\left[\sum_i d_i f(i)\right] = \sum_i d_i \frac{1}{k} \|u(i)\|^2 = 1 .$$

Here the last equality follows from (4). □

4.3 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 4.3. *For any two non zero vectors u and v,*

$$\|\tilde{u} - \tilde{v}\| \sqrt{\|u\|^2 + \|v\|^2} \leqslant 2 \|u - v\| ,$$

where $\tilde{u} = \frac{u}{\|u\|}$ and $\tilde{v} = \frac{v}{\|v\|}$.

Proof. Note that $2\|u\|\|v\| \leqslant \|u\|^2 + \|v\|^2$. Hence,

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

If $\langle \tilde{u}, \tilde{v} \rangle \geqslant 0$, then using $\|u\|^2 + \|v\|^2 \geqslant 2\|u\|\|v\|$,

$$\|\tilde{u} - \tilde{v}\|^2 (\|u\|^2 + \|v\|^2) \leqslant 2(\|u\|^2 + \|v\|^2 - 2\|u\|\|v\| \langle \tilde{u}, \tilde{v} \rangle) \leqslant 2\|u - v\|^2 .$$

Else if $\langle \tilde{u}, \tilde{v} \rangle < 0$, then

$$\begin{aligned} \|\tilde{u}-\tilde{v}\|^2 (\|u\|^2+\|v\|^2) &\leqslant 2\Big(\|u\|^2+\|v\|^2-2\|u\|\,\|v\|\langle \tilde{u},\tilde{v}\rangle \\ &\qquad -(\|u\|^2+\|v\|^2)\langle \tilde{u},\tilde{v}\rangle+2\|u\|\,\|v\|\langle \tilde{u},\tilde{v}\rangle\Big) \\ &= 2\Big(\|u-v\|^2-(\|u\|^2+\|v\|^2-2\|u\|\,\|v\|)\langle \tilde{u},\tilde{v}\rangle\Big) \\ &\leqslant 4\|u-v\|^2 . \end{aligned}$$

□

We will need another lemma which is a direct consequence of Lemma 4.2 about the maximum of k i.i.d normal random variables.

Corollary 4.4. *There exists an absolute constant c_1 such that for any $i, j \in [n]$,*

$$\mathbb{P}[f(i) > 0 \text{ and } f(j) = 0] \leqslant c_1\left(\|\tilde{u}(i)-\tilde{u}(j)\| \frac{\sqrt{\log k}}{k}\right),$$

where $\tilde{u}(i) = \frac{u(i)}{\|u(i)\|}$ and $\tilde{u}(j) = \frac{u(j)}{\|u(j)\|}$.

Proof. By definition of f (see (7)), we have that $f(i) \neq 0$ if and only if $\mathsf{argmax}_\ell \langle u(i), g_\ell \rangle = 1$. Note that $\mathsf{argmax}_\ell \langle \tilde{u}(i), g_\ell \rangle = \mathsf{argmax}_\ell \langle u(i), g_\ell \rangle$. Let $X_\ell = \langle \tilde{u}(i), g_\ell \rangle$ and $Y_\ell = \langle \tilde{u}(j), g_\ell \rangle$ for $\ell \in [k]$. Note that $\{X_\ell\}_{\ell\in[k]}$ (similarly $\{Y_\ell\}_{\ell\in[k]}$) are i.i.d normal random variables. Moreover, for each ℓ, the covariance of X_ℓ and Y_ℓ is equal to $\langle \tilde{u}(i), \tilde{u}(j) \rangle = 1 - \|\tilde{u}(i)-\tilde{u}(j)\|^2/2$. The corollary follows by applying Lemma 4.2. □

The following lemma is the main technical lemma in bounding the expected value of the numerator.

Lemma 4.5. *For any indices $i, j \in [n]$,*

$$\mathbb{E}[|f(i)-f(j)|] \leqslant O\left(\frac{\sqrt{\log k}}{k}\right) \cdot (\|u(i)\|+\|u(j)\|) \cdot \|u(i)-u(j)\|$$

Proof.

$$\begin{aligned} &\mathbb{E}[|f(i)-f(j)|] \\ &= \|u(i)\|^2\,\mathbb{P}[f(i)>0 \text{ and } f(j)=0] + \|u(j)\|^2\,\mathbb{P}[f(j)>0 \text{ and } f(i)=0] + \left(\|u(i)\|^2-\|u(j)\|^2\right)\mathbb{P}[f(i), f(j)>0] \\ &\leqslant c_1\left(\|\tilde{u}(i)-\tilde{u}(j)\|\frac{\sqrt{\log k}}{k}\right)\left(\|u(i)\|^2+\|u(j)\|^2\right) + \left(\|u(i)\|^2-\|u(j)\|^2\right)\frac{1}{k} && \text{(Using Corollary 4.4)} \\ &\leqslant \frac{2c_1\sqrt{\log k}}{k}\|u(i)-u(j)\|\sqrt{\|u(i)\|^2+\|u(j)\|^2} + \frac{1}{k}\langle u(i)-u(j), u(i)+u(j)\rangle && \text{(Using Proposition 4.3)} \\ &\leqslant \frac{2c_1\sqrt{\log k}}{k}\|u(i)-u(j)\|(\|u(i)\|+\|u(j)\|) + \frac{1}{k}\|u(i)-u(j)\|(\|u(i)\|+\|u(j)\|) && \text{(Using Cauchy-Schwarz)} \end{aligned}$$

As c_1 is an absolute constant arising from Corollary 4.4, the result is immediate. □

We are now ready to bound the expectation of the numerator, namely prove Lemma 3.5 which asserts that,

$$\mathbb{E}\left[\sum_{i\sim j} w(\{i,j\})\,|f(i)-f(j)|\right] \leqslant O\left(\sqrt{\mathcal{R}(u)\log k}\right).$$

Proof of Lemma 3.5.

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i\sim j} w(\{i,j\})\,|f(i)-f(j)|\right] &\leqslant O\left(\frac{\sqrt{\log k}}{k}\right)\cdot \sum_{i\sim j} w(\{i,j\})\,\|u(i)-u(j)\|\,(\|u(i)\|+\|u(j)\|) \qquad \text{(Lemma 4.5)}\\
&\leqslant O\left(\frac{\sqrt{\log k}}{k}\right)\cdot \sqrt{\sum_{i\sim j} w(\{i,j\})\,\|u(i)-u(j)\|^2}\sqrt{\sum_{i\sim j} w(\{i,j\})\,(\|u(i)\|+\|u(j)\|)^2}\\
&\qquad \text{(Using the Cauchy-Schwarz inequality)}\\
&\leqslant O\left(\frac{\sqrt{\log k}}{k}\right)\cdot \sqrt{\mathcal{R}(u)\cdot\left(\sum_i d_i\,\|u(i)\|^2\right)}\sqrt{\sum_{i\sim j} w(\{i,j\})\,2\left(\|u(i)\|^2+\|u(j)\|^2\right)}\\
&\qquad \text{(Using the Cauchy-Schwarz inequality and definition of } \mathcal{R}(u))\\
&\leqslant O\left(\frac{\sqrt{\log k}}{k}\right)\cdot \sqrt{\mathcal{R}(u)}\left(\sum_i d_i\,\|u(i)\|^2\right)\\
&\qquad \left(\text{Using } \sum_{i\sim j} w(\{i,j\})\,\|u(i)\|^2 = \sum_i d_i\,\|u(i)\|^2\right)\\
&= O\left(\sqrt{\mathcal{R}(u)\log k}\right) \qquad \text{(Using (4))}
\end{aligned}
$$

□

4.4 Variance of the Denominator.

We will need some groundwork to bound the variance of the denominator in (3). In particular, we will need to use Hermite polynomial basis for the functions on the Gaussian space.

Hermite polynomials Let $\mathcal{G}$ denote the Gaussian space, i.e., $\mathbb{R}$ with a Gaussian measure. 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

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

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 $H_\alpha(x)$ denoted 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 4.6. *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 $\mathbb{E}[g_i h_i] = \rho$. Then for all $\alpha \in \mathbb{Z}^k_{\geqslant 0}$,*

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

Define a function $B : \mathcal{G}^k \longrightarrow \mathbb{R}$ as follows,

$$B(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } (x_1 \geqslant x_i \forall i \in [k]) \text{ or } (x_1 \leqslant x_i \forall i \in [k]) \\ 0 & \text{otherwise} \end{cases}.$$

Observe that by symmetry of the k coordinates, the function $B(x)$ is 1 with probability exactly $\frac{1}{k}$ and 0 otherwise. Hence,

$$\mathbb{E}[B] = \mathbb{E}[B^2] = \frac{1}{k}.$$

Lemma 4.7. *Let u, v be unit vectors and $g_1, \ldots, g_k$ be i.i.d Gaussian vectors. Then,*

$$\mathbb{E}[B(\langle u, g_1\rangle, \ldots, \langle u, g_k\rangle)B(\langle v, g_1\rangle, \ldots, \langle v, g_k\rangle)] \leqslant \frac{1}{k^2} + \langle u, v\rangle^2 \frac{1}{k}.$$

Proof. The function B on the Gaussian space can be written in the Hermite expansion $B(x) = \sum_\alpha B_\alpha H_\alpha(x)$. By Parseval's identity we have,

$$\sum_\alpha B_\alpha^2 = \mathbb{E}[B^2] = \frac{1}{k}.$$

Using Fact 4.6, we can write

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

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

$$\begin{aligned}\mathbb{E}[B(\langle u, g_1\rangle, \ldots, \langle u, g_k\rangle)B(\langle v, g_1\rangle, \ldots, \langle v, g_k\rangle)] &\leqslant (\mathbb{E}[B])^2 + \rho^2 \left(\sum_{\alpha, |\alpha| \geqslant 2} B_\alpha^2\right) \\ &= \frac{1}{k^2} + \langle u, v\rangle^2 \frac{1}{k}.\end{aligned}$$

□

Now we are ready to wrap up the proof of Lemma 3.6 by showing that,

$$\mathsf{Var}\left[\sum_i d_i f(i)\right] \leqslant 1.$$

Proof of Lemma 3.6.

$$\begin{aligned}
\mathbb{E}\left[\sum_{i,j} d_i d_j f(i) f(j)\right] &= \sum_{i,j} d_i d_j \|u(i)\|^2 \|u(j)\|^2 \,\mathbb{E}\left[\frac{f(i)}{\|u(i)\|^2}\frac{f(j)}{\|u(j)\|^2}\right] \\
&\leqslant \sum_{i,j} d_i d_j \|u(i)\|^2 \|u(j)\|^2 \,\mathbb{E}\left[B(\langle \tilde{u}(i), g_1\rangle, \ldots, \langle \tilde{u}(i), g_k\rangle) B(\langle \tilde{u}(j), g_1\rangle, \ldots, \langle \tilde{u}(j), g_k\rangle)\right] \\
&\leqslant \sum_{i,j} d_i d_j \|u(i)\|^2 \|u(j)\|^2 \cdot \left(\frac{1}{k^2} + \frac{1}{k}\langle \tilde{u}(i), \tilde{u}(j)\rangle^2\right) \qquad (\textit{Lemma } 4.7) \\
&= \left(\frac{1}{k}\sum_{i,j} d_i d_j \langle u(i), u(j)\rangle^2 + \frac{1}{k^2}\left(\sum_i d_i \|u(i)\|^2\right)^2\right) \\
&= \left(\frac{1}{k}\cdot\frac{k^2}{\mathcal{D}(u)} + \frac{1}{k^2}\cdot k^2\right) \qquad (\text{ Using } (4)\ ) \\
&\leqslant 2 \qquad (\text{ since } \mathcal{D}(u) \geqslant k)\,.
\end{aligned}$$

Using Lemma 3.4, we get

$$\mathsf{Var}\left[\sum_i d_i f(i)\right] = \mathbb{E}\left[\sum_{i,j} d_i d_j f(i) f(j)\right] - \left(\mathbb{E}\left[\sum_i d_i f(i)\right]\right)^2 \leqslant 1\,.$$

□

5 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 1.5) thereby generalizing the lower bound side of the Cheeger's inequality.

Proposition 5.1 (Restatement of Proposition 1.5)**.** *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. Set $\alpha \stackrel{\text{def}}{=} \max_i \phi_G(S_i)$. 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 $U \stackrel{\text{def}}{=} \left\{D^{\frac{1}{2}} X_{S_i} : i \in [k]\right\}$. Here X_S is the incidence vector of the set $S \subset V$. Recall the following characterization of the k^{th} eigenvalue λ_k.

$$\lambda_k = \min_{S : \mathsf{rank}(S) = k} \max_{X \in S} \frac{X^T \mathcal{L} X}{X^T X}$$

Since all the vectors in U are orthogonal to each other they span a k-dimensional subspace. This implies that,

$$\lambda_k = \min_{S:\mathsf{rank}(S)=k} \max_{X\in S} \frac{X^T \mathcal{L} X}{X^T X} \leqslant \max_{X\in \mathsf{span}(U)} \frac{X^T \mathcal{L} X}{X^T X}$$

Observe that each vector in the $\mathsf{span} U$ can be thought of as a function over the graph G'. In fact, for every vector $X \in \mathsf{span}(U)$, the corresponding vector $Y \in \mathbb{R}^{G'}$ satisfies $X^T \mathcal{L} X = Y^T \mathcal{L}' Y$. Therefore we can write,

$$\lambda_k \leqslant \max_{Y\in\mathbb{R}^k*\{0\}} \frac{\sum_{i,j} w'(\{i,j\})(Y_i - Y_j)^2}{\sum_i d'_i Y_i^2}, \tag{8}$$

where $w'(\{i,j\})$ and d'_i denotes the edge weights and degrees in graph G'.

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,

$$\begin{aligned}
\sum_i \sum_{j>i} w'(\{i,j\})(Y_i - Y_j)^2 &\leqslant 2\left(\sum_i \sum_{j>i} w'(\{i,j\})(Y_i^+ - Y_j^+)^2 + \sum_i \sum_{j>i} w'(\{i,j\})(Y_j^- - Y_i^-)^2\right) \\
&\leqslant 2\left(\sum_i \sum_{j>i} w'(\{i,j\})\left|(Y_i^+)^2 - (Y_j^+)^2\right| + \sum_i \sum_{j>i} w'(\{i,j\})\left|(Y_j^-)^2 - (Y_i^-)^2\right|\right).
\end{aligned}$$

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

$$\begin{aligned}
\sum_i \sum_{j>i} w'(\{i,j\})\left|(Y_i^+)^2 - (Y_j^+)^2\right| &= \sum_i \sum_{j>i} w'(\{i,j\})\left(\sum_{t=i}^{j-1} (Y_{t+1}^+)^2 - (Y_t^+)^2\right) \\
&\leqslant \sum_{i=1}^{k} ((Y_i^+)^2 - (Y_{i+1}^+)^2) \cdot w'(E(T_i, \bar{T}_i)) \\
&\leqslant \sum_{i=1}^{k} ((Y_i^+)^2 - (Y_{i+1}^+)^2) \cdot \alpha \cdot w'(T_i) \qquad \text{using } w'(E(T_i, \bar{T}_i)) \leqslant \alpha w'(T_i) \\
&= \alpha \sum_i^k d'_i (Y_i^+)^2 \qquad \text{using } w'(T_{i+1}) - w'(T_i) = d'_{i+1}.
\end{aligned}$$

Using a similar calculation, we get that

$$\sum_i \sum_{j>i} w'(\{i,j\})\left|(Y_j^-)^2 - (Y_i^-)^2\right| \leqslant \alpha \sum_i d'_i (Y_i^-)^2.$$

Putting these two inequalities together we get that

$$\sum_{j>i} w'(\{i,j\})(Y_i - Y_j)^2 \leqslant 2\alpha \sum_i d'_i Y_i^2.$$

Together with (8), the above inequality implies that $\lambda_k(\mathcal{L}) \leqslant 2 \max_i \phi_G(S_i)$. □

6 Lower Bound Constructions

This section is devoted to explicit construction of graphs that will serve as tight examples to the Cheeger's inequalities proven in this work.

6.1 k-sparsecuts

We begin by describing the canonical lower bound examples – the Gaussian graphs.

Definition 6.1. (Gaussian graphs) For a constant $\varepsilon \in (-1, 1)$, the Gaussian graph $N_{K,\varepsilon}$ is the infinite graph over $\mathbb{R}^K$ where the weight of an edge (x, y) is the probability density function of two standard Gaussian random vectors X, Y with correlation[2] $1 - \varepsilon$ at (x, y).

Proof. (Proof of Theorem 1.7) The first K eigenvalues of the normalized Laplacian of both $N_{K,\varepsilon}$ are at most ε ([RST12]). In fact, the eigenfunctions associated with the top K eigenvalues are the degree 1 Hermite polynomials given by,

$$\{H(x) = x_i | i \in \{1, \ldots, K\}\},$$

whose eigenvalue is exactly equal to ε. Isoperimetric inequalities on Gaussian space can be used to bound the expansion of small sets in $N_{K,\varepsilon}$.

Lemma 6.2 ([Bor85])**.** *For any set $S \subset \mathbb{R}^K$ with Gaussian probability measure at most δ, $\phi_{N_{K,\varepsilon}}(S) = \Omega(\sqrt{\varepsilon \log(1/\delta)})$.*

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 $\max_i \phi_{N_{k,\varepsilon}}(S_i) = \Omega(\sqrt{\epsilon \log k}) = \Omega(\sqrt{\lambda_K \log k})$. Hence, the Gaussian graph $N_{K,\epsilon}$ satisfies the condition of Theorem 1.7. □

A drawback of the Gaussian graphs construction is that the resulting graphs are infinite. To construct a finite graph with similar properties, the simplest approach is to discretize the Gaussian graphs. By the compactness of Gaussian space, it is straightforward to show that for every fixed K and every fixed error $\eta > 0$, there exists a discretization $\mathcal{H}$ with $C(K, \eta)$ vertices such that

$$\lambda_K \leqslant \epsilon + \eta \text{ and } \phi_k(\mathcal{H}) \geqslant O(\sqrt{\epsilon \log k}) - \eta$$

We omit the details of the discretization argument here.

6.2 k-partition

In this section, we give a constructive proof of Theorem 1.8. In other words, we construct a family of graphs such that for any k-partition $\{S_1, \ldots, S_k\}$ of the graph at least one of the sets S_i has large expansion, i.e.,

$$\max_i \phi(S_i) = \sqrt{\lambda_k} \cdot \Omega(\sqrt{k}).$$

First, we note that a star graph with exactly k leaves satisfies the requirements of the lower bound. However, in the star graph we need $k = \theta(n)$, which is undesirable. Instead, we will construct a graph inspired from the star graph wherein we get a similar bound even for $k = o(n)$.

The graph shown in Figure 2 consists of k cliques of size $(n - 1)/k$ and an additional vertex v connected to each of the cliques with edges of total weight pn. Here p is some absolute constant that we will choose later.

[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 $Z \sim \mathcal{N}(0, 1)^K$.

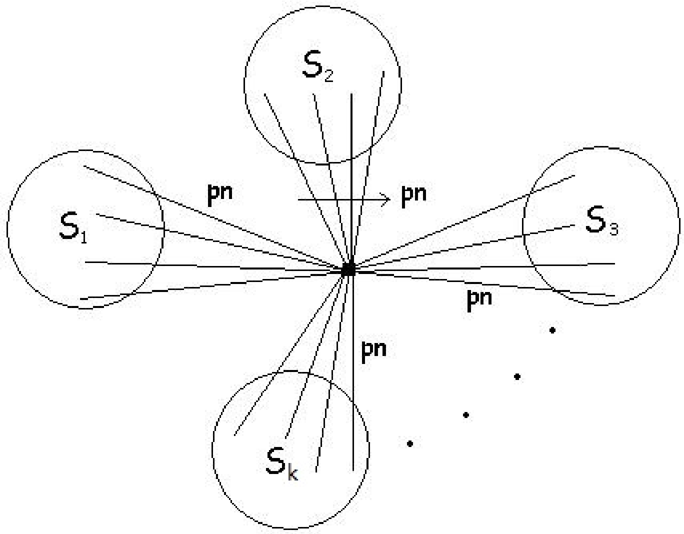

Figure 2: k-partition can have sparsity much larger than $\Omega(\sqrt{\lambda_k}\text{polylog}k)$

Proof. (Proof of Theorem 1.8) In Figure 2, $\forall i \in [k]$, 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. Let $\mathcal{P}' \stackrel{\text{def}}{=} \{S_1 \cup \{v\}, S_2, S_3, \ldots, S_k\}$. For $n > k^3$ and a sufficiently small choice of constant p, it is easily verified that the optimum k-partition is isomorphic to $\mathcal{P}'$. Furthermore, we have

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

Applying Proposition 1.5 to $S_1, \ldots, S_k$, we get that

$$\lambda_k \leqslant 2 \max_i \phi_G(S_i) = \frac{2pn}{\binom{(n-1)/k}{2} + pn} = O(pk^2/n)\,.$$

Picking $k = n^{1/3}/2$ and a sufficiently small constant for p, we have that for every k-partition $T_1, \ldots, T_k$ of the vertex set,

$$\max_i \phi(T_i) \geqslant \sqrt{\lambda_k} \cdot \Omega\left(\frac{k^2}{n^{1/2}}\right) \geqslant \sqrt{\lambda_k} \cdot \Omega(\sqrt{k})$$

□

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer, *Subexponential algorithms for unique games and related problems*, Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on, IEEE, 2010, pp. 563–572. 3, 4

[AK95] Charles J. Alpert and Andrew B. Kahng, *Recent directions in netlist partitioning: A survey*, Integration: The VLSI Journal **19** (1995), 1–81. 1

[Alo86] Noga Alon, *Eigenvalues and expanders*, Combinatorica **6** (1986), no. 2, 83–96. 1

[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. 1

[AR98] Yonatan Aumann and Yuval Rabani, *An $o(\log k)$ approximate min-cut max-flow theorem and approximation algorithm*, SIAM Journal on Computing **27** (1998), no. 1, 291–301. 3, 4

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani, *Expander flows, geometric embeddings and graph partitioning*, Journal of the ACM (JACM) **56** (2009), no. 2, 5. 2, 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*, Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on, IEEE, 2011, pp. 17–26. 4

[BLR10] Punyashloka Biswal, James R Lee, and Satish Rao, *Eigenvalue bounds, spectral partitioning, and metrical deformations via flows*, Journal of the ACM (JACM) **57** (2010), no. 3, 13. 4

[Bor85] Christer Borell, *Geometric bounds on the ornstein-uhlenbeck velocity process*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **70** (1985), no. 1, 1–13. 17

[Che70] Jeff Cheeger, *A lower bound for the smallest eigenvalue of the laplacian*, Problems in analysis **625** (1970), 195–199. 1

[Chu97] Fan Chung, *Spectral graph theory*, American Mathematical Society, 1997. 9

[CMM06] Moses Charikar, Konstantin Makarychev, and Yury Makarychev, *Near-optimal algorithms for unique games*, Proceedings of the thirty-eighth annual ACM symposium on Theory of computing, ACM, 2006, pp. 205–214. 4, 11

[Dhi01] Inderjit S Dhillon, *Co-clustering documents and words using bipartite spectral graph partitioning*, Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2001, pp. 269–274. 1

[DS91] Persi Diaconis and Daniel Stroock, *Geometric bounds for eigenvalues of markov chains*, The Annals of Applied Probability **1** (1991), no. 1, 36–61. 2

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

[KGGK94] Vipin Kumar, Ananth Grama, Anshul Gupta, and George Karypis, *Introduction to parallel computing*, vol. 110, Benjamin/Cummings Redwood City, 1994. 1

[KVV04] Ravi Kannan, Santosh Vempala, and Adrian Vetta, *On clusterings: Good, bad and spectral*, Journal of the ACM (JACM) **51** (2004), no. 3, 497–515. 4

[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. 1, 3, 4

[LOGT12] James R Lee, Shayan Oveis Gharan, and Luca Trevisan, *Multi-way spectral partitioning and higher-order cheeger inequalities*, Proceedings of the 44th symposium on Theory of Computing, ACM, 2012, pp. 1117–1130. 4

[LR99] Tom Leighton and Satish Rao, *Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms*, Journal of the ACM (JACM) **46** (1999), no. 6, 787–832. 3

[LRTV11] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala, *Algorithmic extensions of cheegers inequality to higher eigenvalues and partitions*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, Springer, 2011, pp. 315–326. 4

[LT91] Michel Ledoux and Michel Talagrand, *Probability in banach spaces*, Springer, 1991. 13

[MOR+06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey Steif, and Benny Sudakov, *Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality*, Israel Journal of Mathematics, 2006, pp. 299–336.

[PSL90] Alex Pothen, Horst D Simon, and Kang-Pu Liou, *Partitioning sparse matrices with eigenvectors of graphs*, SIAM Journal on Matrix Analysis and Applications **11** (1990), no. 3, 430–452. 2

[RS10] Prasad Raghavendra and David Steurer, *Graph expansion and the unique games conjecture*, Proceedings of the 42nd ACM symposium on Theory of computing, ACM, 2010, pp. 755–764. 3

[RST10] Prasad Raghavendra, David Steurer, and Prasad Tetali, *Approximations for the isoperimetric and spectral profile of graphs and related parameters*, Proceedings of the 42nd ACM symposium on Theory of computing, ACM, 2010, pp. 631–640. 4

[RST12] Prasad Raghavendra, David Steurer, and Madhur Tulsiani, *Reductions between expansion problems*, Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on, IEEE, 2012, pp. 64–73. 17

[SJ89] Alistair Sinclair and Mark Jerrum, *Approximate counting, uniform generation and rapidly mixing markov chains*, Information and Computation **82** (1989), no. 1, 93–133. 1

[SM00] Jianbo Shi and Jitendra Malik, *Normalized cuts and image segmentation*, Pattern Analysis and Machine Intelligence, IEEE Transactions on **22** (2000), no. 8, 888–905. 1, 2

[SS96] Michael Sipser and Daniel A. Spielman, *Expander codes*, IEEE Transactions on Information Theory **42** (1996), 1710–1722. 1, 2

[ST96] Daniel A Spielmat and Shang-Hua Teng, *Spectral partitioning works: planar graphs and finite element meshes*, Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on, IEEE, 1996, pp. 96–105. 4