Algorithmic Extensions of Cheeger's Inequality to Higher Eigenvalues and Partitions

Anand Louis[1], Prasad Raghavendra[1], Prasad Tetali[2], and Santosh Vempala[1]

[1] College of Computing, Georgia Tech, Atlanta. Email: {anand.louis, raghavendra, vempala}@cc.gatech.edu
[2] School of Mathematics, Georgia Tech, Atlanta. Email: tetali@math.gatech.edu

Abstract. We consider two generalizations of the problem of finding a sparsest cut in a graph. The first is to find a partition of the vertex set into m parts so as to minimize 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 $m-1$ parts). The second, that has appeared in the context of understanding the unique games conjecture, is to find a subset of minimum sparsity that contains at most a $1/m$ fraction of the vertices. 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 m-partition, we prove that the sparsity is at most $8\sqrt{1-\lambda_m}\log m$ where λ_m is the m^{th} largest eigenvalue of the normalized adjacency matrix. For sparsest small-set, we bound the sparsity by $O(\sqrt{(1-\lambda_{m^2})\log m})$. Our results are algorithmic, with the first using a recursive spectral decomposition and the second using a convex relaxation.

1 Introduction

The expansion of a graph is a fundamental and widely studied parameter with many important algorithmic applications [LR99,ARV04,KRV06,She09]. Given an undirected graph $G=(V,E)$, with nonnegative weights $w: E \to \mathbb{R}_+$ on the edges, the *expansion* of a subset of vertices $S \subset V$ is defined as:

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

where by $w(S)$ we denote the total weight of edges incident to vertices in S and for two subsets S,T, we denote the total weight of edges between them by $w(S,T)$. The degree of a vertex v, denoted by d_v is defined as $d_v \stackrel{\text{def}}{=} \sum_{u\sim v} w(u,v)$. The expansion of the graph is $\phi_G \stackrel{\text{def}}{=} \min_{S\subset V} \phi(S)$.

Cheeger's inequality connects this combinatorial parameter to graph eigenvalues. Let λ_i denote the i^{th} largest eigenvalue of the normalized adjacency matrix of G, defined as $B \stackrel{\text{def}}{=} D^{-1}A$ where A is the adjacency matrix of G and D is a diagonal matrix with $D(i,i)$ equal to the (weighted) degree of vertex i (each row of B sums to 1).

Theorem 1 (Cheeger's Inequality ([Alo86,AM85])). *Given a graph G, and its row-normalized adjacency matrix B (each row of B sums to 1), let the eigenvalues of B be $1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n$. Then*

$$2\sqrt{1-\lambda_2} \geq \phi_G \geq \frac{1-\lambda_2}{2}.$$

It is sometimes convenient to re-state this inequality in terms of the eigenvalues of the normalized Laplacian of G, the matrix $L \stackrel{\text{def}}{=} I - B$. The eigenvalues of L are $1-\lambda_i$.

The proof of Cheeger's inequality is algorithmic and uses the second eigenvector of the normalized adjacency matrix. It gives an efficient algorithm for finding an approximate sparsest cut, i.e., a cut whose sparsity is bounded as in the inequality. Finding a sparse cut is a fundamental algorithmic problem, and the algorithm based on the second eigenvector is popular in theory and in practice.

Here we consider two natural generalizations of the sparsest cut problem.

1.1 Generalizations of Sparsest cut

Our first problem is an extension of sparsest cut to partitions with more than two parts.
Sparsest m-partition: Given a weighted undirected graph $G = (V, E)$ and an integer $m > 1$, the sparsity of an m-partition $\mathcal{P} = \{V_1, \ldots, V_m\}$ of the vertex set V into m parts is the ratio of the weight of edges between different parts to the sum of the weights of smallest $m - 1$ parts in $\mathcal{P}$, i.e.,

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

The sparsest m-partition has value $\phi^{sum}_{G,m} \stackrel{\text{def}}{=} min_{\mathcal{P}}\ \phi^{sum}_{G,m}(\mathcal{P})$.

Variants of such a definition have been considered in the literature. The m-cut problem asks for the minimum weight of edges whose deletion leaves m disjoint parts. 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.

The second extension we consider is obtained by restricting the size of the set.
Sparsest Small Set: Given a graph $G = (V, E)$ and an integer $m > 0$, the *small-set sparsity* of G is defined as

$$\phi^{small}_{G,m} \stackrel{\text{def}}{=} \min_{S \subset V, w(S) \leq w(V)/m} \frac{w(S, V \setminus S)}{w(S)}$$

The problem is to find a sparsest small set.
The sparsest small set problem has been shown to be closely related to the Unique Games problem (see [RS10,ABS10]). Recently, Arora et. al. ([ABS10]) showed that $\phi^{small}_{G,m} \leq C\sqrt{(1 - \lambda_{m^{100}}) \log_m n}$ where C is some absolute constant. They also give a polynomial time algorithm to compute a small set with sparsity satisfying this bound.

Both of the problems are NP-hard since they are generalizations of sparsest cut.

1.2 Our results

For sparsest m-partition, we give the following bound using the m^{th} largest eigenvalue of the normalized adjacency matrix of G.

Theorem 2. *For any edge-weighted graph $G = (V, E)$ and integer $|V| \geq m > 0$, there exists an m-partition $\mathcal{P}$ of V such that*

$$\phi^{sum}_{G,m}(\mathcal{P}) \leq 8\sqrt{1 - \lambda_m} \log m,$$

where λ_m is the m^{th} largest eigenvalue of the normalized adjacency matrix of G. Moreover, an m-partition with this sparsity bound can be computed in polynomial time.

The above result is a generalization of the upper bound in Cheeger's inequality (where $m = 2$). Our proof is based on a recursive partitioning algorithm that might be of independent interest. We remark that the dependence on m is necessary and cannot be improved to something smaller than $\sqrt{\log m}$. Moreover, notice that the lower bound of $\Omega(1 - \lambda_2)$ in Cheeger's inequality cannot be strengthened for $m > 2$: Consider the graph G constructed by taking $m-1$ cliques C_1, C_2, ..., C_{m-1} each on $(n-1)/(m-1)$ vertices. Let v be the remaining vertex. Let $C_1, \ldots, C_{m-1}$ be connected to v by a single edge. Now, G will have $m - 1$ eigenvalues close to 1 because of the $m-1$ cuts $(\{v\}, C_i)$ for $i \in [m-1]$, but the m^{th} eigenvalue will be close to 0, as any other cut which is not a linear combination of these $m - 1$ cuts will have to cut through one of the cliques. Therefore, λ_m must be a constant smaller than $1/2$. But $\phi^{sum}_{G,m} = (m-1)/((m-2)(n/m)^2) \approx m^2/n^2$. Thus, $1 - \lambda_m \gg \phi^{sum}_{G,m}$ for small enough values of m.

For the sparsest small-set problem, we present the following bound.

Theorem 3. *Given a graph $G = (V, E)$ and an integer $|V| > m > 1$, there exists a non-empty subset $S \subset V$ such that $|S| \leq \frac{2|V|}{m}$ and*

$$\phi(S) \leq C\sqrt{(1 - \lambda_{m^2}) \log m}$$

where C is a fixed constant. Moreover, such a set can be computed in polynomial time.

The result is a consequence of the rounding technique of [RST10a] and a relation between eigenvalues and the SDP relaxation observed by [Ste10].

A lower bound of $(1-\lambda_2)/2$ for $\phi^{small}_{G,m}$ follows from Cheeger's inequality. Furthermore, it is easy to see that this bound cannot be improved in general. Specifically, consider the graph G constructed by adding an edge between a copy of $K_{\lfloor n/m \rfloor}$ and a copy of $K_{\lceil n(1-1/m) \rceil}$. In this graph, $\phi^{small}_{G,m} \approx 1/(n/m)^2 = m^2/n^2$, whereas G has only 1 eigenvalue close to 1 and $\lambda_m \approx 0$ for $m > 3$. Therefore, $1 - \lambda_m \gg \phi^{small}_{G,m}$ for small enough values of m.

We believe that there is room for improvement in both our theorems, and especially for the sparsest small-set, we believe that the dependence should be on a lower eigenvalue (m instead of m^2). We make the following conjecture:

Conjecture 1. There is a fixed constant C such that for any graph $G = (V, E)$ and any integer $|V| > m > 1$,

$$\phi^{sum}_{G,m}, \phi^{small}_{G,m} \leq C\sqrt{(1-\lambda_m)\log m},$$

where λ_m is the m^{th} largest eigenvalue of the normalized adjacency martix of G.

The bounds in this conjecture are matched by the *Gaussian graphs.* For a constant $\epsilon \in (-1, 1)$, let $N_{k,\epsilon}$ 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 [3] ϵ equal x and y respectively. The first k eigenvalues of $N_{k,\epsilon}$ are at least $1-\epsilon$ (see [RST10b]). The following lemma bounds the expansion of small sets in $N_{k,\epsilon}$.

Lemma 1 ([Bor85,RST10b]). *For $m < k$ we have*

$$\phi^{small}_{N_{k,\epsilon},m} \geq \Omega(\sqrt{\epsilon \log m})$$

Therefore, for any value of $m < k$, $N_{k,\epsilon}$ has $\phi^{small}_{N_{k,\epsilon},m} \geq \Omega(\sqrt{(1-\lambda_m)\log m})$.

2 Monotonicity of Eigenvalues

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

Lemma 2 (Weyl's Inequality). *Given a Hermitian matrix B with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, and a positive semidefinite matrix E, if $\lambda'_1 \geq \lambda'_2 \geq \ldots \geq \lambda'_n$ denote the eigenvalues of $B' \stackrel{\text{def}}{=} B + E$, then $\lambda'_i \geq \lambda_i$.*

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

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

Lemma 3. *Let B be the row normalized 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 B' be the row-normalized matrix of G'. Let the eigenvalues of B be $1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and let the eigenvalues of B' be $1, \lambda'_2, \lambda'_3, \lambda'_4 \geq \ldots \geq \lambda'_n$. Then $\lambda'_i \geq \lambda_i\ \forall i \in [n]$.*

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

Proof. Let $D^{\frac{1}{2}}$ be the diagonal matrix whose $(i,i)^{th}$ entry is $\sqrt{d_i}$. Observe that $DB = B^TD$. Therefore $Q \stackrel{\text{def}}{=} D^{\frac{1}{2}}BD^{-\frac{1}{2}}$ is a symmetric matrix where Moreover, the eigenvalues of Q and B are the same: if ν is an eigenvector of Q with eigenvalue λ, i.e. $D^{\frac{1}{2}}BD^{-\frac{1}{2}}\nu = \lambda\nu$, then $B(D^{-\frac{1}{2}}\nu) = \lambda(D^{-\frac{1}{2}}\nu)$.

Hence, the eigenvalues of Q are $1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and the eigenvalues of $Q' \stackrel{\text{def}}{=} D^{\frac{1}{2}}B'D^{\frac{-1}{2}}$ are $1 \geq \lambda'_2 \geq \lambda'_3 \geq \lambda'_4 \geq \ldots \geq \lambda'_n$.

$C \stackrel{\text{def}}{=} D^{\frac{1}{2}}(B' - B)D^{-\frac{1}{2}}$ 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

$$x^TCx = \sum_{ij} c_{ij}x_ix_j = -\frac{1}{2}\sum_{i \neq j} c_{ij}(x_i - x_j)^2 \geq 0.$$

Using Lemma 2, we get that $\lambda'_i \geq \lambda_i \; \forall i \in [n]$.

Lemma 4. *Let B be the row normalized matrix of the graph G. Let S be a non-empty set of vertices of G. Let G' be the graph obtained from G by shrinking*[4] *S to a single vertex. Let B' be the row normalized adjacency matrix of G'. Let the eigenvalues of B be $1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and let the eigenvalues of B' be $1, \lambda'_2, \lambda'_3, \lambda'_4 \geq \ldots \geq \lambda'_{n-|S|+1}$. Then $\lambda_i \geq \lambda'_i$ for $1 \leq i \leq n - |S| + 1$.*

Proof. Let $D^{\frac{1}{2}}$ be the diagonal matrix whose $(i,i)^{th}$ entry is $\sqrt{d_i}$. Observe that $DB = B^TD$. Therefore $Q \stackrel{\text{def}}{=} D^{\frac{1}{2}}BD^{-\frac{1}{2}}$ is a symmetric matrix where Moreover, the eigenvalues of Q and B are the same: if ν is an eigenvector of Q with eigenvalue λ, i.e. $D^{\frac{1}{2}}BD^{-\frac{1}{2}}\nu = \lambda\nu$, then $B(D^{-\frac{1}{2}}\nu) = \lambda(D^{-\frac{1}{2}}\nu)$. The i^{th} eigenvalue of B can be written as

$$\lambda_i = \max_{S:rank(S)=i} \min_{x \in S} \frac{x^TBx}{x^Tx}$$

and hence

$$\lambda_i = \max_{S:rank(S)=i} \min_{x \in S} 1 - \frac{x^TD^{\frac{1}{2}}(I-B)D^{-\frac{1}{2}}x}{x^Tx} = \max_{S:rank(S)=i} \min_{x \in S} 1 - \frac{\sum_i \sum_{j>i} d_ib_{ij}(x_i - x_j)^2}{\sum_i d_ix_i^2}$$

Let $s = |S|$. Let $v_1, v_2, \ldots, v_n$ be the vertices of G, let $S = \{v_{n-s+1}, v_{n-s+2}, \ldots v_n\}$ and $v_1, v_2, \ldots, v_{n-s}, v'_{n-s+1}$ be the vertices of G' where v'_{n-s+1} is the vertex obtained by shrinking S to a single vertex. If d'_i denotes the degree of i^{th} vertex in G' then $d'_i = d_i$ for $1 \leq i \leq n - s$ and $d'_{n-s+1} = \sum_{i \in S} d_i$.

Let T^k be a variable denoting a subspace of $\mathbb{R}^k$.

$$\begin{aligned}
\lambda'_i &= \max_{T^{n-s+1}:rank(T^{n-s+1})=i} \; \min_{x \in T^{n-s+1}} \frac{x^TB'x}{x^Tx} \\
&= \max_{T^{n-s+1}:rank(T^{n-s+1})=i} \; \min_{x \in T^{n-s+1}} 1 - \frac{\sum_{i=1}^{n-s+1}\sum_{j>i} d'_ib'_{ij}(x_i - x_j)^2}{\sum_i d'_ix_i^2} \\
&= \max_{T^{n-s+1}:rank(T^{n-s+1})=i} \; \min_{x \in T^{n-s+1}} 1 - \frac{\sum_{i=1}^{n-s}\sum_{j>i} d_ib_{ij}(x_i - x_j)^2}{\sum_{i=1}^{n-s} d_ix_i^2 + (\sum_{i=n-s+1}^{n} d_i)x_{n-s+1}^2} \\
&\leq \max_{T^n:rank(T^n)=i} \min_{x \in T^n} 1 - \frac{\sum_{i=1}^{n-s}\sum_{j>i} d_ib_{ij}(x_i - x_j)^2 + \sum_{i=n-s+1}^{n}\sum_{j>i} d_ib_{ij}(x_i - x_j)^2}{\sum_i d_ix_i^2} \\
&= \lambda_i
\end{aligned}$$

[4] A vertex set S is said to be shrunk to a vertex v_S, if all the vertices in S are removed from G and in its place a new vertex v_S is added. All the edges in $E(S, \bar{S})$ are now incident on v_S and all the internal edges in S now become self loops on v_S.

3 Sparsest m-partition

Let A denote the adjacency matrix of the graph. We normalize A by scaling the rows so that the row sums are equal to one. Let B denote this row-normalized matrix.

We propose the following recursive algorithm for finding an m-partitioning of G. Use the second eigenvector of G 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, i.e. $E' = E \backslash E(C, \bar{C})$. We obtain the matrix B' as follows: For all edges $(i, j) \in E'$, $b'_{ij} = b_{ij}$. For all other i, j such that $i \neq j$, $b'_{ij} = 0$. For all i, $b'_{ii} = 1 - \sum_{j \neq i} b'_{ij}$. Note that B' corresponds to the row-normalized adjacency matrix of G', if $\forall (i, j) \in E(C, \bar{C})$ we add self loops at vertex i and vertex j in G'. The matrix B' is block-diagonal with two blocks for the two components of G'. The spectrum of B' (eigenvalues, eigenvectors) is the union of the spectra of the two blocks. The first two eigenvalues of B' are now 1 and we use the third largest eigenvector of G' 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 m 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 highest second eigenvalue among all the current components. The precise algorithm appears in Figure 1.

1. Input : Graph $G = (V, E)$, m such that $1 < m < |V|$
2. Initialize $i := 2$, and $G_i = G$, B_i = row-normalized matrix of G
   (a) Find a sparse cut $(C_i, \bar{C}_i)$ in G_i using the i^{th} eigenvector of B_i (the first $i-1$ are all equal to 1).
   (b) Let $G_{i+1} := G_i \backslash E_{G_i}(C, \bar{C})$
   (c) If $i = m$ then output the connected components of G_{i+1} and End else
      i. Construct B_{i+1} as follows
         A. $\forall (j, k) \in E_{G_{i+1}}$, $B_{i+1}(j, k) = B_i(j, k)$
         B. For all other j, k, $j \neq k$, $B_{i+1}(j, k) = 0$
         C. $\forall j$, $B_{i+1}(j, j) = 1 - \sum_{k \neq j} B_{i+1}(j, k)$
      ii. $i := i + 1$
      iii. Repeat from Step (a)

Fig. 1. The Recursive Algorithm

We now analyze the algorithm. Our analysis will also be a proof of Theorem 2.

The matrix B_i for $i > 2$ is not the row-normalized matrix of G_i, but can be viewed as a row normalized matrix of G_i with a self loop on vertices i and j for each edge $(i, j) \in E_{G_i}(C_i, \bar{C}_i)$. The next theorem is a generalization of Cheeger's inequality to weighted graphs, which relates the eigenvalues of B to the sparsity of G.

Theorem 4 ([KVV04]). *Suppose B is a $N * N$ matrix with nonnegative entries with each row sum equal to 1 and suppose there are positive real numbers $\pi_1, \pi_2, \ldots, \pi_N$ summing to 1 such that $\pi_i b_{ij} = \pi_j b_{ji}$ $\forall i, j$. If v is the right eigenvector of B corresponding to the 2^{nd} largest eigenvalue λ_2 and $i_1, i_2 \ldots, i_N$ is an ordering of $1, 2, \ldots, N$ such that $v_1 \geq v_2 \geq \ldots \geq v_N$, then*

$$2\sqrt{1-\lambda_2} \geq \min_{l:1 \leq l \leq N} \frac{\sum_{1 \leq u \leq l; l+1 \leq v \leq N} \pi_{i_u} b_{i_u i_v}}{\min\{\sum_{1 \leq u \leq l} \pi_{i_u}, \sum_{l+1 \leq v \leq N} \pi_{i_v}\}} \geq \frac{1-\lambda_2}{2}$$

Lemma 2 shows that the eigenvalues of B_i are monotonically nondecreasing with i. This will show that $\phi_{G_i}(C_i) \leq 2\sqrt{1-\lambda_m}$.

We can now prove the main theorem.

Proof (of Theorem 2). Let $\mathcal{P}$ be the set of partitions output by the algorithm and let $S(\mathcal{P})$ denote the sum of weights of the smallest $m-1$ pieces in $\mathcal{P}$. Note that we need only the smaller side of a cut to bound the size of the cut : $|E_G(S, \bar{S})| \leq \phi_G |S|$. We define the notion of a cut-tree $T = (V(T), E(T))$ as follows:

$V(T) = \{V\} \cup \{C_i | i \in [m]\}$ (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 $E(\mathcal{P}) = \cup_i E_{G_i}(C_i, \bar{C}_i) = \cup_i \cup_{j:L(C_j)=i} E_{G_j}(C_j, \bar{C}_j)$. We make the following claim.

Claim.

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leq 2\sqrt{1-\lambda_m} S(\mathcal{P})$$

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 $E(C_j, \bar{C}_j) \leq 2\sqrt{1-\lambda_m} w(C_j)$. Therefore

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leq 2\sqrt{1-\lambda_m} \sum_{j:L(C_j)=i} w(C_j) \leq 2\sqrt{1-\lambda_m} S(\mathcal{P})$$

This claim implies that $\phi(\mathcal{P}) \leq 2\sqrt{1-\lambda_m} height(T)$.

The height of T might be as much as m. But we will show that we can assume height(T) to be $\log m$. For any path in the tree $v_1, v_2, \ldots, v_{k-1}, v_k$ 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 $w(C_{v_{i+1}}) \leq w(C_{v_i})/2$, 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 $\sum_{i=2}^{k} w(C_{v_i}) \leq w(C_{v_1})$. Hence we can modify the T as follows : make the nodes $v_3, \ldots, v_k$ children of v_2. The nodes $v_3, \ldots, v_{k-1}$ now become leaves whereas the subtree rooted at v_k 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 proprety that $w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leq 4\sqrt{1-\lambda_m} S(\mathcal{P})\ \forall i$.

Claim.

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leq 4\sqrt{1-\lambda_m} S(\mathcal{P})$$

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,

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leq 2\left(2\sqrt{1-\lambda_m} \sum_{j:L(C_j)=i} w(C_j)\right) \leq 4\sqrt{1-\lambda_m} S(\mathcal{P})$$

.

Repeating this process we can ensure that no two adjacent nodes in the T have degree 2. Hence, there are at most $\log m$ vertices along any path starting from the root which have exactly one child. Thus the height of the new $\mathsf{cut\text{-}tree}$ is at most $2\log m$. Thus $\mathsf{E}((\mathcal{P})) \leq 8\sqrt{1-\lambda_m} \log m S(\mathcal{P})$ and hence $\phi^{sum}_{G,m} \leq \frac{\mathsf{E}((\mathcal{P}))}{S(\mathcal{P})} \leq 8\sqrt{1-\lambda_m} \log m$.

4 Finding Sparsest Small Sets

Given an integer m and an undirected graph $G = (V, E)$, we wish to find the set $S \subset V$ of size at most $|V|/m$ and having minimum expansion. This is equivalent to finding the vector $x \in \{0,1\}^{|V|}$ which minimizes $\frac{\sum_{i \sim j} w(i,j)(x(i)-x(j))^2}{\sum_i d_i x(i)^2}$ and has at most $|V|/m$ non-zero entries. Ignoring the sparsity constraint, the minimization is equivalent to minimizing $\frac{\sum_{i \sim j} w(i,j)\|v_i - v_j\|^2}{\sum_i d_i \|v_i\|^2}$ over all collections of vectors $\{v_i | i \in [n]\}$. The challenge is to deal with the sparsity constraint. Since any $x \in \{0,1\}^{|V|}$ having at most $|V|/m$ non-zero

$$\min \frac{\sum_{ij} w(i,j)\|v_i - v_j\|^2}{\sum_i d_i \|v_i\|^2}$$
$$\sum_{i,j} (v_i^T v_j)^2 \leq \frac{n^2}{m^2}$$
$$\sum_i \|v_i\|^2 = n$$

Fig. 2. A convex relaxation for sparsest small set

entries satisfies $\sum_{i,j} x(i)x(j) \leq n^2/m^2$ we can relax the sparsity constraint to $\sum_{i,j} < v_i, v_j >^2 \leq n^2/m^2$ while maintaining $\sum_i \|v_i\|^2 = n$. This convex relaxation of the problem is shown in Figure 2.

It was pointed out to us by [Ste10] that eigenvectors of the graph form a feasible solution to this convex relaxation. Here we present a proof of the same.

Let $w_1, w_2, \ldots, w_n$ denote the eigenvectors of DBD^{-1} and let $1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n$ be the respective eigenvalues. Let F be the $m^2 * n$ dimensional matrix which has $w_1, w_2, \ldots, w_{m^2}$ as its row vectors, i.e. $F = [w_1 \ w_2 \ \ldots \ w_{m^2}]^T$. Let $f_1, f_2, \ldots, f_n$ be the columns of F. We define $v_i \stackrel{\text{def}}{=} (\sqrt{\frac{n}{m^2}} f_i)$. We will show that $\{v_i | i \in [n]\}$ forms a feasible solution for the convex program, and that the cost of the solution is bounded by $1 - \lambda_{m^2}$.

Lemma 5. *The vectors v_i, $i \in [n]$ satisfy $\sum_{ij} \langle v_i, v_j \rangle^2 \leq \frac{n^2}{m^2}$.*

Proof.

$$\begin{aligned}
\sum_{i,j} \langle v_i, v_j \rangle^2 &= \frac{n^2}{m^4} \sum_{i,j} (\sum_t f_{it} f_{jt})^2 \\
&= \frac{n^2}{m^4} \sum_{i,j} \sum_{t_1,t_2} f_{it_1} f_{jt_1} f_{it_2} f_{it_2} \\
&= \frac{n^2}{m^4} \sum_{t_1,t_2} \langle w_{t_1} \otimes w_{t_1}, w_{t_2} \otimes w_{t_2} \rangle \\
&= \frac{n^2}{m^4} \sum_{t_1,t_2} \langle w_{t_1}, w_{t_2} \rangle^2 \\
&= \frac{n^2}{m^2}.
\end{aligned}$$

Lemma 6. $\sum_i \|v_i\|^2 = n$

Proof.

$$\begin{aligned}
\sum_i \langle v_i, v_i \rangle &= \frac{n}{m^2} \sum_i \langle f_i, f_i \rangle \\
&= \frac{n}{m^2} \sum_i (\sum_t f_{it}^2) \\
&= \frac{n}{m^2} \sum_t ||w_t||_2^2 \\
&= \frac{n}{m^2} m^2 \\
&= n.
\end{aligned}$$

Lemma 7. $\frac{\sum_{ij} w_{ij} \|v_i - v_j\|^2}{\sum_i d_i \|v_i\|^2} \leq 1 - \lambda_{m^2}$

Proof.

$$\frac{\sum_{ij} w_{ij}\|v_i - v_j\|^2}{\sum_i d_i\|v_i\|^2} = \frac{\sum_l \sum_{ij} w_{ij}\|v_{li} - v_{lj}\|^2}{\sum_l \sum_i d_i\|v_{li}\|^2}$$
$$\leq \max_l \frac{\sum_{ij} w_{ij}\|v_{li} - v_{lj}\|^2}{\sum_i d_i\|v_{li}\|^2}$$
$$\leq 1 - \lambda_{m^2}.$$

Lemmas 5 and 6 show that the $\{v_i | i \in [n]\}$ form a feasible solution to the convex program and Lemma 7 shows that the cost of this solution is at most $\sqrt{1-\lambda_{m^2}}$.

We use the rounding scheme of [RST10a] to round this solution of the convex program to get a set S of size $2n/m$ and $\phi(S) \leq \mathcal{O}(\sqrt{(1-\lambda_{m^2})\log m})$. We give the rounding procedure in Figure 3.

1. For each $i \in [n]$ define functions $f_i \stackrel{\text{def}}{=} \|v_i\|\sqrt{\Phi(x - v_i^*)}$ where $\Phi(x)$ is probability density function of gaussian with mean 0 and variance $1/\sqrt{\log m}$ and v_i^* denotes the unit vector along the direction of v_i.
2. Sample $t \in \mathcal{N}(0,1)^{m^2}$.
3. Compute $\theta = 2m * \sum_i f_i(t)$ and define $x_i \stackrel{\text{def}}{=} max\{f_i(t) - \theta, 0\}$ for each $i \in [n]$.
4. Do a Cheeger rounding on $X \stackrel{\text{def}}{=} [x_1 x_2 \ldots x_n]^T$.

Fig. 3. The Rounding Algorithm

For any f_i and f_j defined as above, their inner product is defined as $\langle f_i, f_j \rangle \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f_i(x) f_j(x) dx$. The following lemma is a slightly modified version of a similar lemma in [RST10a] to suit our requirements. For completeness we give the proof in Appendix A.

Lemma 8. *1.* $\frac{\sum_{i,j} w_{i,j}\|f_i - f_j\|^2}{\sum_i d_i\|f_i\|^2} \leq \frac{\sum_{i,j} w_{i,j}\|v_i - v_j\|^2}{\sum_i d_i\|v_i\|^2}$

2. $\sum_{i,j}\langle f_i, f_j\rangle \leq 2n/m$

Lemma 9 ([RST10a]).

1. $E(support(X)) \leq 2n/m$

2. $\frac{\sum_{i,j} w_{i,j}(x_i - x_j)^2}{\sum_i d_i x_i^2} \leq \frac{\sum_{i,j} w_{i,j}\|f_i - f_j\|^2}{\sum_i d_i\|f_i\|^2}$

Proof (of Theorem 3).

Lemma 8 shows that $\{f_i | i \in [n]\}$ satisfy a stronger sparsity condition than the one in Figure 2 and the value of the objective function of the convex program on $\{f_i | i \in [n]\}$ is at most $\mathcal{O}(\log m)$ times the value of the objective function on $\{v_i | i \in [n]\}$.

Lemma 9 shows that X has at most $2n/m$ non-zero entries and together with Lemma 8 implies that cost of the objective function of the convex program on X is at most $\mathcal{O}(\log m)$ times the cost of the objective function on $\{v_i | i \in [n]\}$.

Performing a Cheeger rounding on X will yield a set of size at most $2n/m$ and expansion $\mathcal{O}(\sqrt{\log m \frac{\sum_{i,j} w_{i,j}\|v_i - v_j\|^2}{\sum_i d_i\|v_i\|^2}}) \leq \mathcal{O}(\sqrt{(1-\lambda_{m^2})\log m})$, where the inequality follows from Lemma 7.

Thus we have Theorem 3.

5 Acknowledgements

We are grateful to David Steurer who first suggested that eigenvectors could be used to construct a feasible solution to the convex program in Figure 2, and for other kind comments on this paper. This work was supported in part by a student fellowship at the Algorithms and Randomness Center (ARC) and by NSF awards AF-0915903 and AF-0910584.

References

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

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

[AM85] Noga Alon and V. D. Milman, *lambda$_1$, isoperimetric inequalities for graphs, and superconcentrators*, J. Comb. Theory, Ser. B **38** (1985), no. 1, 73–88.

[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.

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

[KRV06] Rohit Khandekar, Satish Rao, and Umesh V. Vazirani, *Graph partitioning using single commodity flows*, STOC (Jon M. Kleinberg, ed.), ACM, 2006, pp. 385–390.

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

[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.

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

[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.

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

[She09] Jonah Sherman, *Breaking the multicommodity flow barrier for $\mathcal{O}(\sqrt{\log n})$-approximations to sparsest cut*, FOCS, IEEE Computer Society, 2009, pp. 363–372.

[Ste10] David Steurer, *Personal communication*, 2010.

A Proof of Lemma 8

We first state some known facts about Gaussians.

Fact 5 *Let $\Phi_\sigma(u)$ be the probability function of a multi-dimensional gaussian centered at $u \in \mathbb{R}^n$ and having variance σ in each coordinate. Let δ^n denote the standard Lebesgue measure on $\mathbb{R}^n$. Then*

$$\int \sqrt{\Phi_\sigma(u)\Phi_\sigma(v)}d\delta^n = e^{-\|u-v\|^2/8\sigma^2}$$

The following fact shows that in order for a mapping to preserve distances it is enough to preserve lengths and distances of unit vectors.

Fact 6 *For any two vectors $u, v \in \mathbb{R}^n$, we have*

$$\|u-v\|^2 = (\|u\| - \|v\|)^2 + \|u\|\|v\|\|u^* - v^*\|^2$$

We state Lemma 8 again.

Lemma 10. *1.* $\frac{\sum_{i,j} w_{i,j}\|f_i - f_j\|^2}{\sum_i d_i\|f_i\|^2} \leq \frac{\sum_{i,j} w_{i,j}\|v_i - v_j\|^2}{\sum_i d_i\|v_i\|^2}$

2. $\sum_{i,j}\langle f_i, f_j\rangle \leq 2n/m$

Proof. 1. Since $\|v_i\| = \|f_i\|\ \forall i \in [n]$ it suffices to show that we have $\|f_i - f_j\|^2 \leq \mathcal{O}(\log m\|v_i - v_j\|^2)$ $\forall i, j \in [n]$. Using Fact 5,

$$\|f_i^* - f_j^*\| = 2 - 2e^{-\log m\|v_i^* - v_j^*\|^2/8} \leq \log m\|v_i^* - v_j^*\|^2/4$$

Now, using Fact 6

$$\begin{aligned}\|f_i - f_j\|^2 &= (\|v_i\| - \|v_j\|)^2 + \|v_i\|\|v_j\|\|f_i^* - f_j^*\|^2 \\ &= (\|v_i\| - \|v_j\|)^2 + \log m\|v_i\|\|v_j\|\|v_i^* - v_j^*\|^2/4 \\ &\leq \log m\|v_i - v_j\|^2/4.\end{aligned}$$

The last inequality uses Fact 6 again. This proves the first part.

2. For the second part, we use the fact that $e^{cx} \leq 1 - (1 - e^c)x$.

$$\begin{aligned}\langle f_i^*, f_j^* \rangle &= e^{-\log m(1 - \langle v_i^*, v_j^* \rangle)} \\ &\leq e^{-\log m(1 - |\langle v_i^*, v_j^* \rangle)|} \\ &\leq 1 - (1 - e^{-\log m})(1 - |\langle v_i^*, v_j^* \rangle|) \\ &\leq e^{-\log m} + |\langle v_i^*, v_j^* \rangle|.\end{aligned}$$

Now,

$$\begin{aligned}\sum_{i,j} \langle f_i, f_j \rangle &\leq \sum_{i,j} \|v_i\| \|v_j\| (1/m + \langle v_i^*, v_j^* \rangle) \\ &= 1/m(\sum_i \|f_i\|)^2 + \sum_{i,j} |\langle v_i, v_j \rangle|\end{aligned}$$

By Jensen's inequality, the first term contributes not more than $\sum_i \|f_i\|^2$. The constraints in the convex program imply that $\sum_{i,j} |\langle v_j, v_j \rangle| \leq n/m$. Putting these together we get the second part of the lemma.