

Improved Cheeger’s Inequality: Analysis of Spectral Partitioning Algorithms through Higher Order Spectral Gap

Tsz Chiu Kwok^{*}
CUHK
tckwok@cse.cuhk.edu.hk

Lap Chi Lau[†]
CUHK
chi@cse.cuhk.edu.hk

Yin Tat Lee[‡]
CUHK
yintat@mit.edu

Shayan Oveis Gharan[§]
Stanford University
shayan@stanford.edu

Luca Trevisan[¶]
Stanford University
trevisan@stanford.edu

ABSTRACT

Let $\phi(G)$ be the minimum conductance of an undirected graph G , and let $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2$ be the eigenvalues of the normalized Laplacian matrix of G . We prove that for any graph G and any $k \geq 2$,

$$\phi(G) = O(k) \frac{\lambda_2}{\sqrt{\lambda_k}},$$

and this performance guarantee is achieved by the spectral partitioning algorithm. This improves Cheeger’s inequality, and the bound is optimal up to a constant factor for any k . Our result shows that the spectral partitioning algorithm is a constant factor approximation algorithm for finding a sparse cut if λ_k is a constant for some constant k . This provides some theoretical justification to its empirical performance in image segmentation and clustering problems. We extend the analysis to spectral algorithms for other graph partitioning problems, including multi-way partition, balanced separator, and maximum cut.

Categories and Subject Descriptors

F.2 [Theory of Computation]: Analysis of Algorithms and Problem Complexity

^{*}Supported by Hong Kong RGC grant 2150701.

[†]Supported by Hong Kong RGC grant 2150701.

[‡]Currently a PhD student of MIT.

[§]Supported by a Stanford Graduate Fellowship.

[¶]This material is based upon work supported by the National Science Foundation under grant No. CCF 1017403.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

STOC’13, June 1-4, 2013, Palo Alto, California, USA.

Copyright 2013 ACM 978-1-4503-2029-0/13/06 ...\$15.00.

Keywords

Graph partitioning; Spectral algorithm

1. INTRODUCTION

We study the performance of spectral algorithms for graph partitioning problems. For the moment, we assume the graphs are unweighted and d -regular for simplicity, while the results in the paper hold for arbitrary weighted graphs, with suitable changes to the definitions. Let $G = (V, E)$ be a d -regular undirected graph. The *conductance* of a subset $S \subseteq V$ is defined as

$$\phi(S) = \frac{|E(S, \bar{S})|}{d \min\{|S|, |\bar{S}|\}},$$

where $E(S, \bar{S})$ denotes the set of edges of G crossing from S to its complement. The conductance of the graph G is defined as

$$\phi(G) = \min_{S \subseteq V} \phi(S).$$

Finding a set of small conductance, also called a sparse cut, is an algorithmic problem that comes up in different areas of computer science. Some applications include image segmentation [SM00, TM06], clustering [NJW01, KVV04, Lux07], community detection [LLM10], and designing approximation algorithms [Shm97].

A fundamental result in spectral graph theory provides a connection between the conductance of a graph and the second eigenvalue of its normalized Laplacian matrix. The normalized Laplacian matrix $\mathcal{L} \in \mathbb{R}^{V \times V}$ is defined as $\mathcal{L} = I - \frac{1}{d}A$, where A is the adjacency matrix of G . The eigenvalues of \mathcal{L} satisfy $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{|V|} \leq 2$. It is a basic fact that $\phi(G) = 0$ if and only if $\lambda_2 = 0$. Cheeger’s inequality for graphs provides a quantitative generalization of this fact:

$$\frac{1}{2} \lambda_2 \leq \phi(G) \leq \sqrt{2\lambda_2}. \quad (1.1)$$

This is first proved in the manifold setting by Cheeger [Che70] and is extended to undirected graphs by Alon and Milman [AM85, Alo86]. Cheeger’s inequality is an influential result in spectral graph theory with applications in spectral clustering [ST07, KVV04], explicit construction of expander

graphs [JM85, HLW06, Lee12], approximate counting [SJ89, JSV04], and image segmentation [SM00].

We improve Cheeger’s inequality using higher eigenvalues of the normalized Laplacian matrix.

Theorem 1. *For every undirected graph G and any $k \geq 2$, it holds that*

$$\phi(G) = O(k) \frac{\lambda_2}{\sqrt{\lambda_k}}.$$

This shows that λ_2 is a better approximation of $\phi(G)$ when there is a large gap between λ_2 and λ_k for any $k \geq 3$. The bound is optimal up to a constant factor for any $k \geq 2$, as the cycle example shows that $\phi(G) = \Omega(k\lambda_2/\sqrt{\lambda_k})$ for any $k \geq 2$.

1.1 The Spectral Partitioning Algorithm

The proof of Cheeger’s inequality is constructive and it gives the following simple nearly-linear time algorithm (the *spectral partitioning algorithm*) that finds cuts with approximately minimal conductance. Compute the second eigenfunction $g \in \mathbb{R}^V$ of the normalized Laplacian matrix \mathcal{L} , and let $f = g/\sqrt{d}$. For a threshold $t \in \mathbb{R}$, let $V(t) := \{v : f(v) \geq t\}$ be a threshold set of f . Return the threshold set of f with the minimum conductance among all thresholds t . Let $\phi(f)$ denote the conductance of the return set of the algorithm. The proof of Cheeger’s inequality shows that $\frac{1}{2}\lambda_2 \leq \phi(f) \leq \sqrt{2\lambda_2}$, and hence the spectral partitioning algorithm is a nearly-linear time $O(1/\sqrt{\lambda_2})$ -approximation algorithm for finding a sparse cut. In particular, it gives a constant factor approximation algorithm when λ_2 is a constant, but since λ_2 could be as small as $1/n^2$ even for a simple unweighted graph (e.g. for the cycle), the performance guarantee could be $\Omega(n)$.

We prove **Theorem 1** by showing a stronger statement, that is $\phi(f)$ is upper-bounded by $O(k\lambda_2/\sqrt{\lambda_k})$.

Theorem 2. *For any undirected graph G , and $k \geq 2$,*

$$\phi(f) = O(k) \frac{\lambda_2}{\sqrt{\lambda_k}}.$$

This shows that the spectral partitioning algorithm is a $O(k/\sqrt{\lambda_k})$ -approximation algorithm for the sparsest cut problem, even though it does not employ any information about higher eigenvalues (e.g. the spectral embedding). In particular, spectral partitioning provides a constant factor approximation for the sparsest cut problem when λ_k is a constant for some constant k .

1.2 Generalizations of Cheeger’s Inequality

There are several recent results showing new connections between the expansion profile of a graph and the higher eigenvalues of its normalized Laplacian matrix. The first result in this direction is about the small set expansion problem. Arora, Barak and Steurer [ABS10a] show that if there are k small eigenvalues for some large k , then the graph has a sparse cut S with $|S| \approx n/k$. In particular, if $k = |V|^\epsilon$ for $\epsilon \in (0, 1)$, then the graph has a sparse cut S with $\phi(S) \leq O(\sqrt{\lambda_k})$ and $|S| \approx n/k$. This can be seen as a generalization of Cheeger’s inequality to the small set expansion problem (see [Ste10, OT12, OW12] for some improvements).

Cheeger’s inequality for graph partitioning can also be extended to higher-order Cheeger’s inequality for k -way graph

partitioning [LRTV12, LOT12]: If there are k small eigenvalues, then there are k disjoint sparse cuts. Let

$$\phi_k(G) := \min_{S_1, \dots, S_k} \max_{1 \leq i \leq k} \phi(S_i)$$

where S_1, \dots, S_k are over non-empty disjoint subsets $S_1, \dots, S_k \subseteq V$. Then

$$\frac{1}{2}\lambda_k \leq \phi_k(G) \leq O(k^2)\sqrt{\lambda_k}.$$

Our result can be applied to k -way graph partitioning by combining with a result in [LOT12].

Corollary 1. *For every undirected graph G and any $l > k \geq 2$, it holds that*

(i)

$$\phi_k(G) \leq O(lk^6) \frac{\lambda_k}{\sqrt{\lambda_l}}.$$

(ii) For any $\delta \in (0, 1)$,

$$\phi_{(1-\delta)k}(G) \leq O\left(\frac{l \log^2 k}{\delta^8 k}\right) \frac{\lambda_k}{\sqrt{\lambda_l}}.$$

(iii) If G excludes K_h as a minor, then for any $\delta \in (0, 1)$

$$\phi_{(1-\delta)k}(G) \leq O\left(\frac{h^4 l}{\delta^5 k}\right) \frac{\lambda_k}{\sqrt{\lambda_l}}.$$

Part (i) shows that λ_k is a better approximation of $\phi_k(G)$ when there is a large gap between λ_k and λ_l for any $l > k$. Part (ii) implies that $\phi_{0.9k}(G) \leq O(\lambda_k \log^2 k / \sqrt{\lambda_{2k}})$, and similarly part (iii) implies that $\phi_{0.9k}(G) \leq O(\lambda_k / \sqrt{\lambda_{2k}})$ for planar graphs.

Furthermore, our proof shows that the spectral algorithms in [LOT12] achieve the corresponding approximation factors. For instance, when λ_l is a constant for a constant $l > k$, there is a constant factor approximation algorithm for the k -way partitioning problem.

1.3 Analysis of Practical Instances

Spectral partitioning is a popular heuristic in practice, as it is easy to be implemented and can be solved efficiently by standard linear algebra methods. Also, it has good empirical performance in applications including image segmentation [SM00] and clustering [Lux07], much better than the worst case performance guarantee provided by Cheeger’s inequality. It has been an open problem to explain this phenomenon rigorously [ST07, GM98]. There are some research directions towards this objective.

One direction is to analyze the average case performance of spectral partitioning. A well-studied model is the random planted model [Bop87, AKS98, McS01], where there is a hidden bisection (X, Y) of V and there is an edge between two vertices in X and two vertices in Y with probability p and there is an edge between a vertex in X and a vertex in Y with probability q . It is proved that spectral techniques can be used to recover the hidden partition with high probability, as long as $p - q \geq \Omega(\sqrt{p \log |V| / |V|})$ [Bop87, McS01]. The spectral approach can also be used for other hidden graph partitioning problems [AKS98, McS01]. Note that the spectral algorithms used are usually not exactly the same as the spectral partitioning algorithm. Some of these proofs explicitly or implicitly use the fact that there is a gap between

the second and the third eigenvalues. See the full version for more details.

To better model practical instances, Bilu and Linial [BL10] introduced the notion of stable instances for clustering problems. One definition for the sparsest cut problem is as follows: an instance is said to be γ -stable if there is an optimal sparse cut $S \subseteq V$ which will remain optimal even if the weight of each edge is perturbed by a factor of γ . Intuitively this notion is to capture the instances with an outstanding solution that is stable under noise, and arguably they are the meaningful instances in practice. Note that a planted bisection instance is stable if $p - q$ is large enough, and so this is a more general model than the planted random model. Several clustering problems are shown to be easier on stable instances [BBG09, ABS10b], and spectral techniques have been analyzed for the stable maximum cut problem [BL10, BDLS12]. See the full version for more details.

Informally, the higher order Cheeger's inequality shows that an undirected graph has k disjoint sparse cuts if and only if λ_k is small. This suggests that the graph has at most $k - 1$ outstanding sparse cuts when λ_{k-1} is small and λ_k is large. The algebraic condition that λ_2 is small and λ_3 is large seems similar to the stability condition but more adaptable to spectral analysis. This motivates us to analyze the performance of the spectral partitioning algorithm through higher-order spectral gaps.

In practical instances of image segmentation, there are usually only a few outstanding objects in the image, and so λ_k is large for a small k [Lux07]. Thus Theorem 2 provides a theoretical explanation to why the spectral partitioning algorithm performs much better than the worst case bound by Cheeger's inequality in those instances. In clustering applications, there is a well-known eigengap heuristic that partitions the data into k clusters if λ_k is small and λ_{k+1} is large [Lux07]. Corollary 1 shows that in such situations the spectral algorithms in [LOT12] perform better than the worst case bound by the higher order Cheeger's inequality.

1.4 Other Graph Partitioning Problems

Our techniques can be used to improve the spectral algorithms for other graph partitioning problems using higher order eigenvalues. In the minimum bisection problem, the objective is to find a set S with minimum conductance among the sets with $|V|/2$ vertices. While it is nontrivial to find a sparse cut with exactly $|V|/2$ vertices [FK02, Rac08], it is well known that a simple recursive spectral algorithm can find a *balanced separator* S with $\phi(S) = O(\sqrt{\epsilon})$ with $|S| = \Omega(|V|)$, where ϵ denotes the conductance of the minimum bisection (e.g. [KVV04]). We use Theorem 2 to generalize the recursive spectral algorithm to obtain a better approximation guarantee when λ_k is large for a small k .

Theorem 3. *Let*

$$\epsilon := \min_{|S|=|V|/2} \phi(S).$$

There is a polynomial time algorithm that finds a set S such that $|V|/5 \leq |S| \leq 4|V|/5$ and $\phi(S) \leq O(k\epsilon/\lambda_k)$.

In the maximum cut problem, the objective is to find a partition of the vertices which maximizes the weight of edges whose endpoints are on different sides of the partition. Goemans and Williamson [GW95] gave an SDP-based

0.878-approximation algorithm for the maximum cut problem. Trevisan [Tre09] gave a spectral algorithm with approximation ratio strictly better than $1/2$. Both algorithms find a solution that cuts at least $1 - O(\sqrt{\epsilon})$ fraction of edges when the optimal solution cuts at least $1 - O(\epsilon)$ fraction of edges. Using a similar method as in the proof of Theorem 2, we generalize the spectral algorithm in [Tre09] for the maximum cut problem to obtain a better approximation guarantee when λ_{n-k} is small for a small k .

Theorem 4. *There is a polynomial time algorithm that on input graph G finds a cut (S, \bar{S}) such that if the optimal solution cuts at least $1 - \epsilon$ fraction of the edges, then (S, \bar{S}) cuts at least*

$$1 - O(k) \log \left(\frac{2 - \lambda_{n-k}}{k\epsilon} \right) \frac{\epsilon}{2 - \lambda_{n-k}}$$

fraction of edges.

1.5 More Related Work

Approximating Graph Partitioning Problems: Besides spectral partitioning, there are approximation algorithms for the sparsest cut problem based on linear and semidefinite programming relaxations. There is an LP-based $O(\log n)$ approximation algorithm by Leighton and Rao [LR99], and an SDP-based $O(\sqrt{\log n})$ approximation algorithm by Arora, Rao and Vazirani [ARV04]. The subspace enumeration algorithm by Arora, Barak and Steurer [ABS10a] provides an $O(1/\lambda_k)$ approximation algorithm for the sparsest cut problem with running time $n^{O(k)}$, by searching for a sparse cut in the $(k - 1)$ -dimensional eigenspace corresponding to $\lambda_1, \dots, \lambda_{k-1}$. It is worth noting that for $k = 3$ the subspace enumeration algorithm is exactly the same as the spectral partitioning algorithm. Nonetheless, the result in [ABS10a] is incomparable to Theorem 2 since it does not upper-bound $\phi(G)$ by a function of λ_2 and λ_3 . Recently, using the Lasserre hierarchy for SDP relaxations, Guruswami and Sinop [GS12, GS13] gave an $1 + O(\text{opt}/\lambda_k)$ approximation algorithm for the sparsest cut problem with running time $n^{O(1)}2^{O(k)}$. Moreover, the general framework of Guruswami and Sinop [GS12] applies to other graph partitioning problems including minimum bisection and maximum cut, obtaining approximation algorithms with similar performance guarantees and running times. This line of recent work is closely related to ours in the sense that it shows that many graph partitioning problems are easier to approximate on graphs with fast growing spectrums, i.e. λ_k is large for a small k . Although their results give much better approximation guarantees when k is large, our results show that simple spectral algorithms provide nontrivial performance guarantees.

Higher Eigenvalues of Special Graphs: Another direction to show that spectral algorithms work well is to analyze their performance in special graph classes. Spielman and Teng [ST07] showed that $\lambda_2 = O(1/n)$ for a bounded degree planar graph and a spectral algorithm can find a separator of size $O(\sqrt{n})$ in such graphs. This result is extended to bounded genus graphs by Kelner [Kel06] and to fixed minor free graphs by Biswal, Lee and Rao [BLR10]. This is further extended to higher eigenvalues by Kelner, Lee, Price and Teng [KLPT11]: $\lambda_k = O(k/n)$ for planar graphs, bounded genus graphs, and fixed minor free graphs when the maximum degree is bounded. Combining with a higher order Cheeger inequality for planar graphs [LOT12], this implies that $\phi_k(G) = O(\sqrt{k/n})$ for bounded degree planar graphs.

We note that these results give mathematical bounds on the conductances of the resulting partitions, but they do not imply that the approximation guarantee of Cheeger's inequality could be improved for these graphs, neither does our result as these graphs have slowly growing spectrums.

Planted Random Instances, Semi-Random Instances, and Stable Instances: We have discussed some previous work on these topics, and we will discuss some relations to our results in the full version of this paper.

1.6 Proof Overview

We start by describing an informal intuition of the proof of [Theorem 2](#) for $k = 3$, and then we describe how this intuition can be generalized. For a function $f \in \mathbb{R}^V$, let $\mathcal{R}(f) = f^T L f / (d \|f\|^2)$ be the Rayleigh quotient of f (see [\(2.2\)](#) of [Subsection 2.1](#) for the definition in general graphs). Let f be a function that is orthogonal to the constant function and that $\mathcal{R}(f) \approx \lambda_2$.

Suppose λ_2 is small and λ_3 is large. Then the higher order Cheeger's inequality implies that there is a partitioning of the graph into two sets of small conductance, but in every partitioning into at least three sets, there is a set of large conductance. So, we expect the graph to have a sparse cut of which the two parts are expanders; see [\[Tan12\]](#) for a quantitative statement. Since $\mathcal{R}(f)$ is small and f is orthogonal to the constant function, we expect that the vertices in the same expander have similar values in f and the average values of the two expanders are far apart. Hence, f is similar to a step function with two steps representing a cut, and we expect that $\mathcal{R}(f) \approx \phi(G)$ in this case. Therefore, roughly speaking, $\lambda_3 \gg \lambda_2$ implies $\lambda_2 \approx \phi(G)$.

Conversely, [Theorem 2](#) shows that if $\lambda_2 \approx \phi^2(G)$ then $\lambda_3 \approx \lambda_2$. One way to prove that $\lambda_2 \approx \lambda_3$ is to find a function f' of Rayleigh quotient close to λ_2 such that f' is orthogonal to both f and the constant function. For example, if G is a cycle, then $\lambda_2 = \Theta(1/n^2)$, $\phi(G) = \Theta(1/n)$, and f (up to normalizing factors) could represent the cosine function. In this case we may define f' to be the sine function. Unfortunately, finding such a function f' in general is not as straightforward. Instead, our idea is to find three *disjointly supported* functions f_1, f_2, f_3 of Rayleigh quotient close to λ_2 . As we prove in [Lemma 1](#), this would upper-bound λ_3 by $2 \max\{\mathcal{R}(f_1), \mathcal{R}(f_2), \mathcal{R}(f_3)\}$. For the cycle example, if f is the cosine function, we may construct f_1, f_2, f_3 simply by first dividing the support of f into three disjoint intervals and then constructing each f_i by defining a smooth localization of f in one of those intervals. To ensure that $\max\{\mathcal{R}(f_1), \mathcal{R}(f_2), \mathcal{R}(f_3)\} \approx \lambda_2$ we need to show that f is a "smooth" function, whose values change continuously. We make this rigorous by showing that if $\lambda_2 \approx \phi(G)^2$, then the function f must be smooth. Therefore, we can construct three disjointly supported functions based on f and show that $\lambda_2 \approx \lambda_3$.

We provide two proofs of [Theorem 2](#). The first proof generalizes the first observation. We show that if $\lambda_k \gg k\lambda_2$, then $\phi(G) \approx k\lambda_2$. The main idea is to show that if $\lambda_k \gg k\lambda_2$, then f can be approximated by a k step function g in the sense that $\|f - g\| \approx 0$ (in general we show that any function f can be approximated by a k step function g such that any $\|f - g\|^2 \leq \mathcal{R}(f)/\lambda_k$). It is instructive to prove that if f is exactly a k -step function then $\phi(G) \leq O(k\mathcal{R}(f))$. Our main technical step, [Proposition 2](#), provides a robust

version of the latter fact by showing that for any k -step approximation of f , $\phi(f) \leq O(k(\mathcal{R}(f) + \|f - g\| \sqrt{\mathcal{R}(f)}))$.

On the other hand, our second proof generalizes the second observation. Say $\mathcal{R}(f) \approx \phi(G)^2$. We partition the support of f into disjoint intervals of the form $[2^{-i}, 2^{-(i+1)})$, and we show that the vertices are distributed almost uniformly in most of these intervals in the sense that if we divide $[2^{-i}, 2^{-(i+1)})$ into k equal length subintervals, then we expect to see the same amount of mass in the subintervals. This shows that f is a smooth function. We then argue that $\lambda_k \lesssim k\lambda_2$, by constructing k disjointly supported functions each of Rayleigh quotient $O(k^2)\mathcal{R}(f)$.

2. PRELIMINARIES

Let $G = (V, E, w)$ be a finite, undirected graph, with positive weights $w : E \rightarrow (0, \infty)$ on the edges. For a pair of vertices $u, v \in V$, we write $w(u, v)$ for $w(\{u, v\})$. For a subset of vertices $S \subseteq V$, we write $E(S) := \{\{u, v\} \in E : u, v \in S\}$. For disjoint sets $S, T \subseteq V$, we write $E(S, T) := \{\{u, v\} \in E : u \in S, v \in T\}$. For a subset of edges $F \subseteq E$, we write $w(F) = \sum_{e \in F} w(e)$. We use $u \sim v$ to denote $\{u, v\} \in E$. We extend the weight to vertices by defining, for a single vertex $v \in V$, $w(v) := \sum_{u \sim v} w(u, v)$. We can think of $w(v)$ as the weighted degree of vertex v . For the sake of clarity, we will assume throughout that $w(v) \geq 1$ for every $v \in V$. For $S \subseteq V$, we write $\text{vol}(S) = \sum_{v \in S} w(v)$ to denote the *volume* of S .

Given a subset $S \subseteq V$, we denote the *Dirichlet conductance* of S by

$$\phi(S) := \frac{w(E(S, \bar{S}))}{\min\{\text{vol}(S), \text{vol}(\bar{S})\}}.$$

For a function $f \in \mathbb{R}^V$, and a threshold $t \in \mathbb{R}$, let $V_f(t) := \{v : f(v) \geq t\}$ be a threshold set of f . We let

$$\phi(f) := \min_{t \in \mathbb{R}} \phi(V_f(t)).$$

be the conductance of the best threshold set of the function f , and $V_f(t_{\text{opt}})$ be the smaller side (in volume) of that minimum cut.

For any two thresholds $t_1, t_2 \in \mathbb{R}$, we use

$$[t_1, t_2] := \{x \in \mathbb{R} : \min\{t_1, t_2\} < x \leq \max\{t_1, t_2\}\}.$$

Note that all intervals are defined to be closed on the larger value and open on the smaller value. For an interval $I = [t_1, t_2] \subseteq \mathbb{R}$, we use $\text{len}(I) := |t_1 - t_2|$ to denote the length of I . For a function $f \in \mathbb{R}^V$, we define $V_f(I) := \{v : f(v) \in I\}$ to denote the vertices within I . The volume of an interval I is defined as $\text{vol}_f(I) := \text{vol}(V_f(I))$. We also abuse the notation and use $\text{vol}_f(t) := \text{vol}(V_f(t))$ to denote the volume of the interval $[t, \infty]$. We define the *support* of f , $\text{supp}(f) := \{v : f(v) \neq 0\}$, as the set of vertices with nonzero values in f . We say two functions $f, g \in \mathbb{R}^V$ are disjointly supported if $\text{supp}(f) \cap \text{supp}(g) = \emptyset$.

For any $t_1, t_2, \dots, t_l \in \mathbb{R}$, let $\psi : \mathbb{R} \rightarrow \mathbb{R}$ be defined as

$$\psi_{t_1, \dots, t_l}(x) = \arg\min_{t_i} |x - t_i|.$$

In words, for any $x \in \mathbb{R}$, $\psi_{t_1, \dots, t_l}(x)$ is the value of t_i closest to x .

For $\rho > 0$, we say a function g is ρ -Lipschitz w.r.t. f , if for all pairs of vertices $u, v \in V$,

$$|g(u) - g(v)| \leq \rho |f(u) - f(v)|.$$

The next inequality follows from the Cauchy-Schwarz inequality and will be useful in our proof. Let $a_1, \dots, a_m, b_1, \dots, b_m \geq 0$. Then,

$$\sum_{i=1}^m \frac{a_i^2}{b_i} \geq \frac{\left(\sum_{i=1}^m a_i\right)^2}{\sum_{i=1}^m b_i}. \quad (2.1)$$

2.1 Spectral Theory of the Weighted Laplacian

We write $\ell^2(V, w)$ for the Hilbert space of functions $f : V \rightarrow \mathbb{R}$ with inner product

$$\langle f, g \rangle_w := \sum_{v \in V} w(v) f(v) g(v),$$

and norm $\|f\|_w^2 = \langle f, f \rangle_w$. We reserve $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ for the standard inner product and norm on \mathbb{R}^k , $k \in \mathbb{N}$ and $\ell^2(V)$.

We consider some operators on $\ell^2(V, w)$. The adjacency operator is defined by $Af(v) = \sum_{u \sim v} w(u, v) f(u)$, and the diagonal degree operator by $Df(v) = w(v) f(v)$. Then the *combinatorial Laplacian* is defined by $L = D - A$, and the *normalized Laplacian* is given by

$$\mathcal{L}_G := I - D^{-1/2} A D^{-1/2}.$$

Observe that for a d -regular unweighted graph, we have $\mathcal{L}_G = \frac{1}{d} L$.

If $g : V \rightarrow \mathbb{R}$ is a non-zero function and $f = D^{-1/2} g$, then

$$\begin{aligned} \frac{\langle g, \mathcal{L}_G g \rangle}{\langle g, g \rangle} &= \frac{\langle g, D^{-1/2} L D^{-1/2} g \rangle}{\langle g, g \rangle} = \frac{\langle f, L f \rangle}{\langle D^{1/2} f, D^{1/2} f \rangle} \\ &= \frac{\sum_{u \sim v} w(u, v) |f(u) - f(v)|^2}{\|f\|_w^2} =: \mathcal{R}_G(f) \end{aligned} \quad (2.2)$$

where the latter value is referred to as the *Rayleigh quotient* of f (with respect to G). We drop the subscript of $\mathcal{R}_G(f)$ when the graph is clear in the context.

In particular, \mathcal{L}_G is a positive-definite operator with eigenvalues

$$0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2.$$

For a connected graph, the first eigenvalue corresponds to the eigenfunction $g = D^{1/2} f$, where f is any non-zero constant function. Furthermore, by standard variational principles,

$$\begin{aligned} \lambda_k &= \min_{g_1, \dots, g_k \in \ell^2(V)} \max_{g \neq 0} \left\{ \frac{\langle g, \mathcal{L}_G g \rangle}{\langle g, g \rangle} : g \in \text{span}\{g_1, \dots, g_k\} \right\} \\ &= \min_{f_1, \dots, f_k \in \ell^2(V, w)} \max_{f \neq 0} \left\{ \mathcal{R}(f) : f \in \text{span}\{f_1, \dots, f_k\} \right\}, \end{aligned} \quad (2.3)$$

where both minimums are over sets of k non-zero orthogonal functions in the Hilbert spaces $\ell^2(V)$ and $\ell^2(V, w)$, respectively. We refer to [Chu97] for more background on the spectral theory of the normalized Laplacian. The following proposition is proved in [HLW06] and will be useful in our proof.

Proposition 1 (Hoory, Linial and Wigderson [HLW06]). *There are two disjointly supported functions $f_+, f_- \in \ell^2(V, w)$ such that $f_+ \geq 0$ and $f_- \leq 0$ and $\mathcal{R}(f_+) \leq \lambda_2$ and $\mathcal{R}(f_-) \leq \lambda_2$.*

By choosing either of f_+ or f_- that has a smaller (in volume) support, and taking a proper normalization, we get the following corollary.

Corollary 2. *There exists a function $f \in \ell^2(V, w)$ such that $f \geq 0$, $\mathcal{R}(f) \leq \lambda_2$, $\text{supp}(f) \leq \text{vol}(V)/2$, and $\|f\|_w = 1$.*

Instead of directly upper bounding λ_k in the proof of [Theorem 2](#), we will construct k disjointly supported functions with small Rayleigh quotients. In the next lemma, we show that by the variational principle this gives an upper-bound on λ_k .

Lemma 1. *For any k disjointly supported functions $f_1, f_2, \dots, f_k \in \ell^2(V, w)$, we have*

$$\lambda_k \leq 2 \max_{1 \leq i \leq k} \mathcal{R}(f_i).$$

Many variants of the following lemma are known.

Lemma 2 ([Chu96]). *For every non-negative $h \in \ell^2(V, w)$ such that $\text{supp}(h) \leq \text{vol}(V)/2$, the following holds*

$$\phi(h) \leq \frac{\sum_{u \sim v} w(u, v) |h(u) - h(v)|}{\sum_v w(v) h(v)}.$$

2.2 Energy Lower Bound

We define the *energy* of a function $f \in \ell^2(V, w)$ as

$$\mathcal{E}_f := \sum_{u \sim v} w(u, v) |f(u) - f(v)|^2.$$

Observe that $\mathcal{R}(f) = \mathcal{E}_f / \|f\|_w^2$. We also define the energy of f restricted to an interval I as follows:

$$\mathcal{E}_f(I) := \sum_{u \sim v} w(u, v) \text{len}(I \cap [f(u), f(v)])^2.$$

When the function f is clear from the context we drop the subscripts from the above definitions.

The next fact shows that by restricting the energy of f to disjoint intervals we may only decrease the energy.

Fact 1. *For any set of disjoint intervals I_1, \dots, I_m , we have*

$$\mathcal{E}_f \geq \sum_{i=1}^m \mathcal{E}_f(I_i).$$

The following is the key lemma to lower bound the energy of a function f . It shows that a long interval with small volume must have a significant contribution to the energy of f .

Lemma 3. *For any non-negative function $f \in \ell^2(V, w)$ with $\text{vol}(\text{supp}(f)) \leq \text{vol}(V)/2$, for any interval $I = [a, b]$ with $a > b \geq 0$, we have*

$$\mathcal{E}(I) \geq \frac{\phi^2(f) \cdot \text{vol}_f^2(a) \cdot \text{len}^2(I)}{\phi(f) \cdot \text{vol}_f(a) + \text{vol}_f(I)}.$$

Proof. Since f is non-negative with $\text{vol}(\text{supp}(f)) \leq \text{vol}(V)/2$, by the definition of $\phi(f)$, the total weight of the edges going out the threshold set $V_f(t)$ is at least $\phi(f) \cdot \text{vol}_f(a)$, for any $a \geq t \geq b \geq 0$. Therefore, by summing over these threshold sets, we have

$$\sum_{u \sim v} w(u, v) \text{len}(I \cap [f(u), f(v)]) \geq \text{len}(I) \cdot \phi(f) \cdot \text{vol}_f(a).$$

Let $E' := \{\{u, v\} : \text{len}(I \cap [f(u), f(v)]) > 0\}$ be the set of edges with nonempty intersection with the interval I . Let $\beta \in (0, 1)$ be a parameter to be fixed later. Let $F \subseteq E'$ be the set of edges of E' that are not adjacent to any of the vertices in I . If $w(F) \geq \beta w(E')$, then

$$\begin{aligned} \mathcal{E}(I) &\geq w(F) \cdot \text{len}(I)^2 \geq \beta \cdot w(E') \cdot \text{len}(I)^2 \\ &\geq \beta \cdot \phi(f) \cdot \text{vol}_f(a) \cdot \text{len}(I)^2. \end{aligned}$$

Otherwise, $\text{vol}_f(I) \geq (1-\beta)w(E')$. Therefore, by the Cauchy Schwarz inequality (2.1), we have

$$\begin{aligned} \mathcal{E}(I) &= \sum_{\{u,v\} \in E'} w(u,v) (\text{len}(I \cap [f(u), f(v)]))^2 \\ &\geq \frac{(\sum_{\{u,v\} \in E'} w(u,v) \text{len}(I \cap [f(u), f(v)]))^2}{w(E')} \\ &\geq \frac{(1-\beta) \text{len}(I)^2 \cdot \phi(f)^2 \cdot \text{vol}_f^2(a)}{\text{vol}_f(I)}. \end{aligned}$$

Choosing $\beta = (\phi(f) \cdot \text{vol}_f(a)) / (\phi(f) \cdot \text{vol}_f(a) + \text{vol}_f(I))$ such that the above two terms are equal gives the lemma. \square

3. ANALYSIS OF SPECTRAL PARTITIONING

Throughout this section we assume that $f \in \ell^2(V, w)$ is a non-negative function of norm $\|f\|_w^2 = 1$ such that $\mathcal{R}(f) \leq \lambda_2$ and $\text{vol}(\text{supp}(f)) \leq \text{vol}(V)/2$. The existence of this function follows from Corollary 2. In Subsection 3.1, we give our first proof of Theorem 2 which is based on the idea of approximating f by a $2k+1$ step function g . Our second proof is given in Subsection 3.2.

3.1 First Proof

We say a function $g \in \ell^2(V, w)$ is a l -step approximation of f , if there exist l thresholds $0 = t_0 \leq t_1 \leq \dots \leq t_{l-1}$ such that for every vertex v ,

$$g(v) = \psi_{t_0, t_1, \dots, t_{l-1}}(f(v)).$$

In words, $g(v) = t_i$ if t_i is the closest threshold to $f(v)$.

We show that if there is a large gap between λ_2 and λ_k , then the function f is well approximated by a step function g with at most $2k+1$ steps. Then we define an appropriate h and apply Lemma 2 to get a lower bound on the energy of f in terms of $\|f - g\|_w^2$. One can think of h as a probability distribution function on the threshold sets, and we will define h in such a way that the threshold sets that are further away from the thresholds t_0, t_1, \dots, t_{2k} have higher probability.

Approximating f by a $2k+1$ Step Function

In the next lemma, we show that if there is a large gap between $\mathcal{R}(f)$ and λ_k , then there is a $2k+1$ -step function g such that $\|f - g\|_w^2 = O(\mathcal{R}(f)/\lambda_k)$.

Lemma 4. *There exists a $2k+1$ -step approximation of f , call g , such that*

$$\|f - g\|_w^2 \leq \frac{4\mathcal{R}(f)}{\lambda_k}. \quad (3.1)$$

Proof. Let $M := \max_v f(v)$. We will find $2k+1$ thresholds $0 =: t_0 \leq t_1 \leq \dots \leq t_{2k} = M$, then we let g be a $2k+1$ step approximation of f with these thresholds. Let $C :=$

$2\mathcal{R}(f)/k\lambda_k$. We choose these thresholds inductively. Given t_0, t_1, \dots, t_{i-1} , we let $t_{i-1} \leq t_i \leq M$ to be the smallest number such that

$$\sum_{v: t_{i-1} \leq f(v) \leq t_i} w(v) |f(v) - \psi_{t_{i-1}, t_i}(f(v))|^2 = C. \quad (3.2)$$

Observe that the left hand side varies continuously with t_i : when $t_i = t_{i-1}$ the left hand side is zero, and for larger t_i it is non-decreasing. If we can satisfy (3.2) for some $t_{i-1} \leq t_i \leq M$, then we let t_i to be the smallest such number, and otherwise we set $t_i = M$.

We say the procedure *succeeds* if $t_{2k} = M$. We will show that: (i) if the procedure succeeds then the lemma follows, and (ii) that the procedure always succeeds. Part (i) is clear because if we define g to be the $2k+1$ step approximation of f with respect to t_0, \dots, t_{2k} , then

$$\begin{aligned} \|f - g\|_w^2 &= \sum_{i=1}^{2k} \sum_{v: t_{i-1} \leq f(v) \leq t_i} w(v) |f(v) - \psi_{t_{i-1}, t_i}(f(v))|^2 \\ &\leq 2kC = \frac{4\mathcal{R}(f)}{\lambda_k}, \end{aligned}$$

and we are done. The inequality in the above equation follows by (3.2).

Suppose to the contrary that the procedure does not succeed. We will construct $2k$ disjointly supported functions of Rayleigh quotients less than $\lambda_k/2$, and then use Lemma 1 to get a contradiction. For $1 \leq i \leq 2k$, let f_i be the following function:

$$f_i(v) := \begin{cases} |f(v) - \psi_{t_{i-1}, t_i}(f(v))| & \text{if } t_{i-1} \leq f(v) \leq t_i \\ 0 & \text{otherwise.} \end{cases}$$

We will argue that at least k of these functions have $\mathcal{R}(f_i) < \frac{1}{2}\lambda_k$. By (3.2), we already know that the denominators of $\mathcal{R}(f_i)$ are equal to C ($\|f_i\|_w^2 = C$), so it remains to find an upper bound for the numerators. For any pair of vertices u, v , we show that

$$\sum_{i=1}^{2k} |f_i(u) - f_i(v)|^2 \leq |f(u) - f(v)|^2. \quad (3.3)$$

The inequality follows using the fact that f_1, \dots, f_{2k} are disjointly supported, and thus u, v are contained in the support of at most two of these functions. If both u and v are in the support of only one function, then (3.3) holds since each f_i is 1-Lipschitz w.r.t. f . Otherwise, say $u \in \text{supp}(f_i)$ and $v \in \text{supp}(f_j)$ for $i < j$, then (3.3) holds since

$$\begin{aligned} &|f_i(u) - f_i(v)|^2 + |f_j(u) - f_j(v)|^2 \\ &= |f(u) - g(u)|^2 + |f(v) - g(v)|^2 \\ &\leq |f(u) - t_i|^2 + |f(v) - t_i|^2 \\ &\leq |f(u) - f(v)|^2. \end{aligned}$$

Summing (3.3) we have

$$\begin{aligned} \sum_{i=1}^{2k} \mathcal{R}(f_i) &= \frac{1}{C} \sum_{i=1}^{2k} \sum_{u \sim v} w(u, v) |f_i(u) - f_i(v)|^2 \\ &\leq \frac{1}{C} \sum_{u \sim v} w(u, v) |f(u) - f(v)|^2 = \frac{k\lambda_k}{2}. \end{aligned}$$

Hence, by an averaging argument, there are k disjointly functions f'_1, \dots, f'_k of Rayleigh quotients less than $\lambda_k/2$, a contradiction to Lemma 1. \square

Upper Bounding $\phi(f)$ Using $2k + 1$ Step Approximation

Next, we show that we can use any function g that is a $2k + 1$ approximation of f to upper-bound $\phi(f)$ in terms of $\|f - g\|_w$.

Proposition 2. For any $2k + 1$ -step approximation of f with $\|f\|_w = 1$, called g ,

$$\phi(f) \leq 4k\mathcal{R}(f) + 4\sqrt{2k}\|f - g\|_w \sqrt{\mathcal{R}(f)}.$$

Let g be a $2k + 1$ approximation of f with thresholds $0 = t_0 \leq t_1 \leq \dots \leq t_{2k}$, i.e. $g(v) := \psi_{t_0, t_1, \dots, t_{2k}}(f(v))$. We will define a function $h \in \ell^2(V, w)$ such that each threshold set of h is also a threshold set of f (in particular $\text{supp}(h) = \text{supp}(f)$), and

$$\frac{\sum_{u \sim v} w(u, v)|h(v) - h(u)|}{\sum_v w(v)h(v)} \leq 4k\mathcal{R}(f) + 4\sqrt{2k}\|f - g\|_w \sqrt{\mathcal{R}(f)}. \quad (3.4)$$

We then simply use [Lemma 2](#) to prove [Proposition 2](#).

Let $\mu : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mu(x) := |x - \psi_{t_0, t_1, \dots, t_{2k}}(x)|.$$

Note that $|f(v) - g(v)| = \mu(f(v))$. One can think of μ as a probability density function to sample the threshold sets, where threshold sets that are further away from the thresholds t_0, t_1, \dots, t_{2k} are given higher probability. We define h as follows:

$$h(v) := \int_0^{f(v)} \mu(x) dx$$

Observe that the threshold sets of h and the threshold sets of f are the same, as $h(u) \geq h(v)$ if and only if $f(u) \geq f(v)$. It remains to prove (3.4). We use the following two claims, that bound the denominator and the numerator separately.

Claim 1. For every vertex v ,

$$h(v) \geq \frac{1}{8k} f^2(v).$$

Proof. If $f(v) = 0$, then $h(v) = 0$ and there is nothing to prove. Suppose $f(v)$ is in the interval $f(v) \in [t_i, t_{i+1}]$. Using the Cauchy-Schwarz inequality,

$$\begin{aligned} f^2(v) &= \left(\sum_{j=0}^{i-1} (t_{j+1} - t_j) + (f(v) - t_i) \right)^2 \\ &\leq 2k \cdot \left(\sum_{j=0}^{i-1} (t_{j+1} - t_j)^2 + (f(v) - t_i)^2 \right). \end{aligned}$$

On the other hand, by the definition of h ,

$$\begin{aligned} h(v) &= \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \mu(x) dx + \int_{t_i}^{f(v)} \mu(x) dx \\ &= \sum_{j=0}^{i-1} \frac{1}{4} (t_{j+1} - t_j)^2 + \int_{t_i}^{f(v)} \mu(x) dx \\ &\geq \sum_{j=0}^{i-1} \frac{1}{4} (t_{j+1} - t_j)^2 + \frac{1}{4} (f(v) - t_i)^2, \end{aligned}$$

where the inequality follows by the fact that $f(v) \in [t_i, t_{i+1}]$. \square

And we will bound the numerator with the following claim.

Claim 2. For any pair of vertices $u, v \in V$,

$$\begin{aligned} |h(v) - h(u)| &\leq \frac{1}{2} |f(v) - f(u)| \\ &\quad (|f(u) - g(u)| + |f(v) - g(v)| + |f(v) - f(u)|). \end{aligned}$$

Proof. By the definition of $\mu(\cdot)$, for any $x \in [f(u), f(v)]$,

$$\begin{aligned} \mu(x) &\leq \min\{|x - g(u)|, |x - g(v)|\} \\ &\leq \frac{|x - g(u)| + |x - g(v)|}{2} \\ &\leq \frac{1}{2} \left((|x - f(u)| + |f(u) - g(u)|) \right. \\ &\quad \left. + (|x - f(v)| + |f(v) - g(v)|) \right) \\ &= \frac{1}{2} (|f(u) - g(u)| + |f(v) - g(v)| + |f(v) - f(u)|), \end{aligned}$$

where the third inequality follows by the triangle inequality, and the last equality uses $x \in [f(u), f(v)]$. Therefore,

$$\begin{aligned} h(v) - h(u) &= \int_{f(u)}^{f(v)} \mu(x) dx \\ &\leq |f(v) - f(u)| \cdot \max_{x \in [f(u), f(v)]} \mu(x) \\ &\leq \frac{1}{2} |f(v) - f(u)| \\ &\quad (|f(u) - g(u)| + |f(v) - g(v)| + |f(v) - f(u)|). \end{aligned} \quad \square$$

Now we are ready to prove [Proposition 2](#).

Proof of Proposition 2. First, by [Claim 2](#),

$$\begin{aligned} &\sum_{u \sim v} w(u, v) |h(u) - h(v)| \\ &\leq \sum_{u \sim v} \frac{1}{2} w(u, v) |f(v) - f(u)| \\ &\quad (|f(u) - g(u)| + |f(v) - g(v)| + |f(v) - f(u)|) \\ &\leq \frac{1}{2} \mathcal{R}(f) + \frac{1}{2} \sqrt{\sum_{u \sim v} w(u, v) |f(v) - f(u)|^2} \\ &\quad \sqrt{\sum_{u \sim v} w(u, v) (|f(u) - g(u)| + |f(v) - g(v)|)^2} \\ &\leq \frac{1}{2} \mathcal{R}(f) + \frac{1}{2} \sqrt{\mathcal{R}(f)}. \\ &\quad \sqrt{2 \sum_{u \sim v} w(u, v) (|f(u) - g(u)|^2 + |f(v) - g(v)|^2)} \\ &= \frac{1}{2} \mathcal{R}(f) + \frac{1}{2} \sqrt{\mathcal{R}(f)} \cdot \sqrt{2 \|f - g\|_w^2}, \end{aligned}$$

where the second inequality follows by the Cauchy-Schwarz inequality. On the other hand, by [Claim 1](#),

$$\sum_v w(v)h(v) \geq \frac{1}{8k} \sum_v w(v)f^2(v) = \frac{1}{8k} \|f\|_w^2 = \frac{1}{8k}.$$

Putting above equations together proves (3.4). Since the threshold sets of h are the same as the threshold sets of f , we have $\phi(f) = \phi(h)$ and the proposition follows by [Lemma 2](#). \square

Now we are ready to prove [Theorem 2](#).

Proof of Theorem 2. Let g be as defined in [Lemma 4](#). By [Proposition 2](#), we get

$$\begin{aligned}\phi(f) &\leq 4k\mathcal{R}(f) + 4\sqrt{2k}\|f - g\|_w \sqrt{\mathcal{R}(f)} \\ &\leq 4k\mathcal{R}(f) + 8\sqrt{2k}\mathcal{R}(f)/\sqrt{\lambda_k} \leq 12\sqrt{2k}\mathcal{R}(f)/\sqrt{\lambda_k}.\end{aligned}$$

□

This proof uses [Lemma 3](#) instead of [Lemma 2](#) to prove the theorem.

3.2 Second Proof

Instead of directly proving [Theorem 2](#), we use [Corollary 2](#) and [Lemma 1](#) and prove a stronger version, as it will be used later to prove [Corollary 1](#). In particular, instead of directly upper-bounding λ_k , we construct k disjointly supported functions with small Rayleigh quotients.

Theorem 5. *For any non-negative function $f \in \ell^2(V, w)$ such that $\text{supp}(f) \leq \text{vol}(V)/2$, and $\delta := \phi^2(f)/\mathcal{R}(f)$, at least one of the following holds*

- i) $\phi(f) \leq O(k)\mathcal{R}(f)$;
- ii) *There exist k disjointly supported functions f_1, f_2, \dots, f_k such that for all $1 \leq i \leq k$, $\text{supp}(f_i) \subseteq \text{supp}(f)$ and*

$$\mathcal{R}(f_i) \leq O(k^2)\mathcal{R}(f)/\delta.$$

Furthermore, the support of each f_i is an interval $[a_i, b_i]$ such that $|a_i - b_i| = \Theta(1/k)a_i$.

We will show that if $\mathcal{R}(f) = \Theta(\phi(G)^2)$ (when $\delta = \Theta(1)$), then f is a smooth function of the vertices, in the sense that in any interval of the form $[t, 2t]$ we expect the vertices to be embedded in equidistance positions. It is instructive to verify this for the second eigenvector of the cycle.

Construction of Disjointly Supported Functions Using Dense Well Separated Regions

First, we show that [Theorem 5](#) follows from a construction of $2k$ dense well separated regions, and in the subsequent parts we construct these regions based on f . A region R is a closed subset of \mathbb{R}_+ . Let $\ell(R) := \sum_{v: f(v) \in R} w(v)f^2(v)$. We say R is W -dense if $\ell(R) \geq W$. For any $x \in \mathbb{R}_+$, we define

$$\text{dist}(x, R) := \inf_{y \in R} \frac{|x - y|}{y}.$$

The ϵ -neighborhood of a region R is the set of points at distance at most ϵ from R ,

$$N_\epsilon(R) := \{x \in \mathbb{R}_+ : \text{dist}(x, R) < \epsilon\}.$$

We say two regions R_1, R_2 are ϵ -well-separated, if $N_\epsilon(R_1) \cap N_\epsilon(R_2) = \emptyset$. In the next lemma, we show that our main theorem can be proved by finding $2k, \Omega(\delta/k)$ -dense, $\Omega(1/k)$ well-separated regions.

Lemma 5. *Let R_1, R_2, \dots, R_{2k} be a set of W -dense and ϵ -well separated regions. Then, there are k disjointly supported functions f_1, \dots, f_k , each supported on the ϵ -neighborhood of one of the regions such that*

$$\forall 1 \leq i \leq k, \mathcal{R}(f_i) \leq \frac{2\mathcal{R}(f)}{k\epsilon^2 W}.$$

Proof. For any $1 \leq i \leq 2k$, we define a function f_i , where for all $v \in V$,

$$f_i(v) := f(v) \max\{0, 1 - \text{dist}(f(v), R_i)/\epsilon\}.$$

Then, $\|f_i\|_w^2 \geq \ell(R_i)$. Since the regions are ϵ -well separated, the functions are disjointly supported. Therefore, the end-points of each edge $\{u, v\} \in E$ are in the support of at most two functions. Thus, by an averaging argument, there exist k functions f_1, f_2, \dots, f_k (maybe after renaming) satisfy the following. For all $1 \leq i \leq k$,

$$\sum_{u \sim v} w(u, v) |f_i(u) - f_i(v)|^2 \leq \frac{1}{k} \sum_{j=1}^{2k} \sum_{u \sim v} w(u, v) |f_j(u) - f_j(v)|^2.$$

Therefore, for $1 \leq i \leq k$,

$$\begin{aligned}\mathcal{R}(f_i) &= \frac{\sum_{u \sim v} w(u, v) |f_i(u) - f_i(v)|^2}{\|f_i\|_w^2} \\ &\leq \frac{2 \sum_{j=1}^{2k} \sum_{u \sim v} w(u, v) |f_j(u) - f_j(v)|^2}{k \cdot \min_{1 \leq i \leq 2k} \|f_i\|_w^2} \\ &\leq \frac{2 \sum_{u \sim v} w(u, v) |f(u) - f(v)|^2}{k\epsilon^2 W} = \frac{2\mathcal{R}(f)}{k\epsilon^2 W},\end{aligned}$$

where we used the fact that f_j 's are $1/\epsilon$ -Lipschitz. Therefore, f_1, \dots, f_k satisfy lemma's statement. □

Construction of Dense Well Separated Regions

Let $0 < \alpha < 1$ be a constant that will be fixed later in the proof. For $i \in \mathbb{Z}$, we define the interval $I_i := [\alpha^i, \alpha^{i+1}]$. Observe that these intervals partition the vertices with positive value in f . We let $\ell_i := \ell(I_i)$. We partition each interval I_i into $12k$ subintervals of equal length,

$$I_{i,j} := \left[\alpha^i \left(1 - \frac{j(1-\alpha)}{12k} \right), \alpha^i \left(1 - \frac{(j+1)(1-\alpha)}{12k} \right) \right],$$

for $0 \leq j < 12k$. Observe that for all i, j ,

$$\text{len}(I_{i,j}) = \frac{\alpha^i(1-\alpha)}{12k}. \quad (3.5)$$

Similarly we define $\ell_{i,j} := \ell(I_{i,j})$. We say a subinterval $I_{i,j}$ is *heavy*, if $\ell_{i,j} \geq c\delta\ell_{i-1}/k$, where $c > 0$ is a constant that will be fixed later in the proof; we say it is *light* otherwise. We use H_i to denote the set of heavy subintervals of I_i and L_i for the set of light subintervals. We use h_i to denote the number of heavy subintervals. We also say an interval I_i is *balanced* if $h_i \geq 6k$, denoted by $I_i \in B$ where B is the set of balanced intervals. Intuitively, an interval I_i is *balanced* if the vertices are distributed uniformly inside that interval.

Next we describe our proof strategy. Using [Lemma 5](#) to prove the theorem it is sufficient to find $2k, \Omega(\delta/k)$ -dense, $\Omega(1/k)$ well-separated regions R_1, \dots, R_{2k} . Each of our $2k$ regions will be a union of *heavy* subintervals. Our construction is simple: from each balanced interval we choose $2k$ *separated* heavy subintervals and include each of them in one of the regions. In order to promise that the regions are well separated, once we include $I_{i,j} \in H_i$ into a region R we leave the two neighboring subintervals $I_{i,j-1}$ and $I_{i,j+1}$ unassigned, so as to separate R from the rest of the regions. In particular, for all $1 \leq a \leq 2k$ and all $I_i \in B$, we include the $(3a-1)$ -th heavy subinterval of I_i in R_a . Note that if an interval I_i is balanced, then it has $6k$ heavy subintervals and we can include one heavy subinterval in each of the $2k$

regions. Furthermore, by (3.5), the regions are $(1 - \alpha)/12k$ -well separated. It remains to prove that these $2k$ regions are dense. Let

$$\Delta := \sum_{I_i \in B} \ell_{i-1}$$

be the summation of the mass of the preceding interval of balanced intervals. Then, since each heavy subinterval $I_{i,j}$ has a mass of $c\delta\ell_{i-1}/k$, by the above construction all regions are $c\Delta\delta/k$ -dense. Hence, the following proposition follows from Lemma 5.

Proposition 3. *There are k disjoint supported functions f_1, \dots, f_k such that for all $1 \leq i \leq k$, $\text{supp}(f_i) \subseteq \text{supp}(f)$ and*

$$\forall 1 \leq i \leq k, \mathcal{R}(f_i) \leq \frac{300k^2\mathcal{R}(f)}{(1 - \alpha)^2 c\delta\Delta}.$$

Lower Bounding the Density

So in the rest of the proof we just need to lower-bound Δ by an absolute constant.

Proposition 4. *For any interval $I_i \notin B$,*

$$\mathcal{E}(I_i) \geq \frac{\alpha^6 \phi(f)^2 \ell_{i-1} (1 - \alpha)^2}{24(k\alpha^4 \phi(f) + c\delta)}.$$

Proof. In the next claim, we lower-bound the energy of a light subinterval in terms of ℓ_{i-1} . Then, we prove the statement simply using $h_i < 6k$.

Claim 3. *For any light subinterval $I_{i,j}$,*

$$\mathcal{E}(I_{i,j}) \geq \frac{\alpha^6 \phi(f)^2 \ell_{i-1} (1 - \alpha)^2}{144k(k\alpha^4 \phi(f) + c\delta)}.$$

Proof. First, observe that

$$\ell_{i-1} = \sum_{v \in I_{i-1}} w(v) f^2(v) \leq \alpha^{2i-2} \text{vol}(\alpha^i). \quad (3.6)$$

Therefore,

$$\begin{aligned} \text{vol}(I_{i,j}) &= \sum_{v \in I_{i,j}} w(v) \leq \sum_{v \in I_{i,j}} w(v) \frac{f^2(v)}{\alpha^{2i+2}} \\ &= \frac{\ell_{i,j}}{\alpha^{2i+2}} \leq \frac{c\delta\ell_{i-1}}{k\alpha^{2i+2}} \leq \frac{c\delta\text{vol}(\alpha^i)}{k\alpha^4}, \end{aligned} \quad (3.7)$$

where we use the assumption that $I_{i,j} \in L_i$ in the second last inequality, and (3.6) in the last inequality. By Lemma 3,

$$\begin{aligned} \mathcal{E}(I_{i,j}) &\geq \frac{\phi(f)^2 \cdot \text{vol}(\alpha^i)^2 \cdot \text{len}(I_{i,j})^2}{\phi(f) \cdot \text{vol}(\alpha^i) + \text{vol}(I_{i,j})} \\ &\geq \frac{k\alpha^4 \phi(f)^2 \cdot \text{vol}(\alpha^i) \cdot \text{len}(I_{i,j})^2}{k\alpha^4 \phi(f) + c\delta} \\ &\geq \frac{\alpha^6 \phi(f)^2 \ell_{i-1} (1 - \alpha)^2}{144k(k\alpha^4 \phi(f) + c\delta)}, \end{aligned}$$

where the first inequality holds by (3.7), and the last inequality holds by (3.5) and (3.6). \square

Now, since the subintervals are disjoint, by Fact 1,

$$\begin{aligned} \mathcal{E}(I_i) &\geq \sum_{I_{i,j} \in L_i} \mathcal{E}(I_{i,j}) \geq (12k - h_i) \frac{\alpha^6 \phi(f)^2 \ell_{i-1} (1 - \alpha)^2}{144k(k\alpha^4 \phi(f) + c\delta)} \\ &\geq \frac{\alpha^6 \phi(f)^2 \ell_{i-1} (1 - \alpha)^2}{24(k\alpha^4 \phi(f) + c\delta)}, \end{aligned}$$

where we used the assumption that I_i is not balanced and thus $h_i < 6k$. \square

Now we are ready to lower-bound Δ .

Proof of Theorem 5. First we show that $\Delta \geq 1/2$, unless (i) holds, and then we use Proposition 3 to prove the theorem. If $\phi(f) \leq 10^4 k \mathcal{R}(f)$, then (i) holds and we are done. So, assume that

$$\frac{10^8 k^2 \mathcal{R}^2(f)}{\phi^2(f)} \leq 1, \quad (3.8)$$

and we prove (ii). Since $\|f\|_w^2 = 1$, by Proposition 4,

$$\mathcal{R}(f) = \mathcal{E}_f \geq \sum_{I_i \notin B} \mathcal{E}(I_i) \geq \sum_{I_i \notin B} \frac{\alpha^6 \phi(f)^2 \ell_{i-1} (1 - \alpha)^2}{24(k\alpha^4 \phi(f) + c\delta)}.$$

Set $\alpha = 1/2$ and $c := \alpha^6(1 - \alpha)^2/96$. If $k\phi(f)\alpha^4 \geq c\delta$, then we get

$$\sum_{I_i \notin B} \ell_{i-1} \leq \frac{48k\mathcal{R}(f)}{\alpha^2(1 - \alpha)^2\phi(f)} \leq \frac{1}{2},$$

where the last inequality follows from (3.8). Otherwise,

$$\sum_{I_i \notin B} \ell_{i-1} \leq \frac{48c\delta\mathcal{R}(f)}{\alpha^6(1 - \alpha)^2\phi^2(f)} \leq \frac{1}{2},$$

where the last inequality follows from the definition of c and δ . Since $\ell(V) = \|f\|_w^2 = 1$, it follows from the above equations that $\Delta \geq \frac{1}{2}$. Therefore, by Proposition 3, we get k disjointly supported functions f_1, \dots, f_k such that

$$\mathcal{R}(f_i) \leq \frac{300k^2\mathcal{R}(f)}{(1 - \alpha)^2 c\delta\Delta} \leq \frac{10^8 k^2 \mathcal{R}(f)^2}{\phi(f)^2}.$$

Although each function f_i is defined on a region which is a union of many heavy subintervals, we can simply restrict it to only one of those subintervals guaranteeing that $\mathcal{R}(f_i)$ only decreases. Therefore each f_i is defined on an interval $[a_i, b_i]$ where by (3.5), $|a_i - b_i| = \Theta(1/k)\alpha_i$. This proves (ii). \square

4. EXTENSIONS AND CONNECTIONS

In the full version of this paper, we show how to extend the main result to spectral multiway partitioning, to the balanced separator problem, to the max-cut problem, to the manifold setting, and also discuss the connections of our results to the previous results in the planted random instances, semi-random instances, and stable instances.

5. REFERENCES

- [Alo86] N. Alon. *Eigenvalues and expanders*. *Combinatorica*, 6, 83–96, 1986.
- [AM85] N. Alon, V. Milman. λ_1 , *isoperimetric inequalities for graphs, and superconcentrators*. *Journal of Combinatorial Theory, Series B*, 38(1), 73–88, 1985.
- [AKS98] N. Alon, M. Krivelevich, B. Sudakov. *Finding a large hidden clique in a random graph*. *Random Structures and Algorithms* 13(3-4), 457–466, 1998.
- [ABS10a] S. Arora, B. Barak, D. Steurer. *Subexponential algorithms for unique games and related problems*. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 563–572, 2010.

- [ARV04] S. Arora, S. Rao, U. Vazirani. *Expander flows, geometric embeddings and graph partitioning*. In Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC), 222–231, 2004.
- [ABS10b] P. Awasthi, A. Blum, O. Sheffet. *Stability yields a PTAS for k -median and k -means clustering*. In Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 309–318, 2010.
- [BBG09] M.-F. Balcan, A. Blum, A. Gupta. *Approximate clustering without the approximation*. In Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 1068–1077, 2009.
- [BL10] Y. Bilu, N. Linial. *Are stable instances easy?* In Proceedings of Innovations in Computer Science (ICS), 332–341, 2010.
- [BDLS12] Y. Bilu, A. Daniely, N. Linial, M. Saks. *On the practically interesting instances of MAXCUT*. In arXiv:1205.4893, 2012.
- [BLR10] P. Biswal, J.R. Lee, S. Rao. *Eigenvalue bounds, spectral partitioning, and metrical deformations via flows*. Journal of the ACM 57(3), 2010.
- [Bop87] R. Boppana. *Eigenvalues and graph bisection: An average-case analysis*. In Proceedings of the 28th Annual Symposium on Foundations of Computer Science (FOCS), 280–285, 1987.
- [Che70] J. Cheeger. *A lower bound for the smallest eigenvalue of the Laplacian*. Problems in Analysis, Princeton University Press, 195–199, 1970.
- [Chu96] F. R. K. Chung. *Laplacians of graphs and Cheeger’s inequalities*. Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), volume 2 of Bolyai Soc. Math. Stud., pages 157–172. János Bolyai Math. Soc., Budapest, 1996.
- [Chu97] Fan R. K. Chung. *Spectral graph theory*. volume 92 of CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997. 9
- [FK02] U. Feige, R. Krauthgamer. *A polylogarithmic approximation of the minimum bisection*. SIAM Journal on Computing 31(3), 1090–1118, 2002.
- [Fuj95] K. Fujiwara. *Eigenvalues of Laplacians on a closed Riemannian manifold and its nets*. Proceedings of the American Mathematical Society 123(8), 1995.
- [GW95] M.X. Goemans, D.P. Williamson. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*. Journal of the ACM 42(6), 1115–1145, 1995.
- [GM98] S. Guattery, G.L. Miller. *On the quality of spectral separators*. SIAM J. Matrix Anal. Appl. 19(3), 701–719, 1998.
- [GS12] V. Guruswami, A.K. Sinop. *Faster SDP hierarchy solvers for local rounding algorithms*. In Proceedings of the 53rd IEEE Symposium on Foundations of Computer Science (FOCS), 197–206, 2012.
- [GS13] V. Guruswami, A.K. Sinop. *Approximating non-uniform sparsest cut via generalized spectra*. In Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2013.
- [HLW06] S. Hoory, N. Linial, A. Wigderson. *Expander graphs and their applications*. Bulletin of the American Mathematical Society 43(4), 439–561, 2006.
- [JSV04] M. Jerrum, A. Sinclair, and E. Vigoda. *A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries*. Journal of the ACM, 51(4), 671–697, 2004.
- [JM85] S. Jimbo and A. Maruoka. *Expanders obtained from affine transformations*. In Proceedings of the 17th Annual Symposium on Theory of Computing (STOC), 88–97, 1985.
- [KVV04] R. Kannan, S. Vempala, A. Vetta. *On clusterings: good, bad, and spectral*. Journal of the ACM 51, 497–515, 2004.
- [Kel06] J. Kelner. *Spectral partitioning, eigenvalue bounds, and circle packings for graphs of bounded genus*. SIAM Journal on Computing 35(4), 882–902, 2006.
- [Lee12] J. R. Lee. *On expanders from the action of $GL(2, Z)$* , manuscript, 2013.
- [KLPT11] J. Kelner, J.R. Lee, G. Price, S.-H. Teng. *Metric uniformization and spectral bounds for graphs*. Geom. Funct. Anal., 21(5), 1117–1143, 2011.
- [LLM10] J. Leskovec, K.J. Lang, M.W. Mahoney. *Empirical comparison of algorithms for network community detection*. In Proceedings of the 19th International Conference on World Wide Web (WWW), 631–640, 2010.
- [LOT12] J.R. Lee, S. Oveis Gharan, L. Trevisan. *Multi-way spectral partitioning and higher-order Cheeger inequalities*. In Proceedings of the 44th Annual Symposium on Theory of Computing (STOC), 1117–1130, 2012.
- [LR99] F.T. Leighton, S. Rao. *Multicommodity max-flow min-cut theorem and their use in designing approximation algorithms*. Journal of the ACM 46(6), 787–832, 1999.
- [LRTV12] A. Louis, P. Raghavendra, P. Tetali, S. Vempala. *Many sparse cuts via higher eigenvalues*. In Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC), 1131–1140, 2012.
- [Lux07] U. von Luxburg. *A tutorial on spectral clustering*. Statistics and Computing 17(4), 395–416, 2007.
- [Lux10] U. von Luxburg. *Clustering stability: an overview*. Foundations and Trends in Machine Learning 2(3), 235–274, 2010.
- [McS01] F. McSherry. *Spectral partitioning of random graphs*. In Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS), 529–537, 2001.
- [NJW01] A. Ng, M. Jordan, Y. Weiss. *On spectral clustering: Analysis and an algorithm*. Advances in Neural Information Processing Systems 14, 849–856, 2001.
- [OW12] R. O’Donnell, D. Witmer. *Improved small-set expansion from higher eigenvalues*. CoRR, abs/1204.4688, 2012.
- [OT12] S. Oveis Gharan, L. Trevisan. *Approximating the expansion profile and almost optimal local graph clustering*. In Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2012.
- [Rac08] Harald Räcke. *Optimal hierarchical decompositions for congestion minimization in networks*. In Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC), 255–264, 2008.
- [SM00] J. Shi, J. Malik. *Normalized cuts and image segmentation*. IEEE Pattern Anal. Mach. Intell., 22(8), 888–905, 2000.
- [Shm97] D.B. Shmoys. *Approximation algorithms for cut problems and their applications to divide-and-conquer*. In Approximation Algorithms for NP-hard Problems, (D.S. Hochbaum, ed.) PWS, 192–235, 1997.
- [SJ89] A. Sinclair and M. Jerrum. *Approximate counting, uniform generation and rapidly mixing markov chains*. Inf. Comput., 82(1), 93–133, July 1989.
- [ST07] D.A. Spielman, S.-H. Teng. *Spectral partitioning works: Planar graphs and finite element meshes*. Linear Algebra and its Application 421(2-3), 284–305, 2007.
- [Ste10] D. Steurer. *On the complexity of unique games and graph expansion*. Ph.D. thesis, Princeton University, 2010.
- [Tan12] M. Tanaka. *Higher eigenvalues and partitions of graphs*. In arXiv:1112.3434, 2012.
- [Tre09] L. Trevisan. *Max cut and the smallest eigenvalue*. In Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC), 263–272, 2009.
- [TM06] D.A. Tolliver, G.L. Miller. *Graph partitioning by spectral rounding: Applications in image segmentation and clustering*. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 1053–1060.