GRAPH PARTITIONING AND MULTI-WAY CHEEGER INEQUALITIES

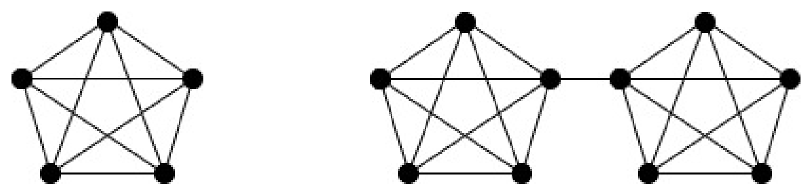
WILLIAM HU

Abstract. A classical result in graph theory states that a graph is disconnected iff its Laplacian matrix has eigenvalue 0, with multiplicity 2. One may be interested in whether a *robust* analogue of this statement exists: is a graph *close* to being disconnected iff the second smallest eigenvalue of its Laplacian is *close* to 0? This is the subject of the discrete Cheeger inequality, which we prove in this paper [1, 2]. We also prove a *higher-order* Cheeger inequality due to [3], which provides a robust analogue of the fact that a graph possesses k disjoint connected components iff its Laplacian has eigenvalue 0 with multiplicity k.

Contents

1. Introduction

1.1. **Conductance of a Graph.** There are several ways to study the connectivity of a graph. One way is by simply determining whether a graph is connected or disconnected. A graph is *connected* iff for every pair of vertices, u and v, there exists a path with u and v as endpoints. However, this notion of connectivity does not help us distinguish how *well-connected* a graph is. In particular, the complete graph, K_n (Figure 1a) seems very well-connected since separating two well-connected components of K_n requires removing many edges. On the other hand, the dumbbell graph, D_n (Figure 1b), which contains two $K_{n/2}$ components connected by a single edge, seems poorly connected as removing a single edge disconnects the graph into two separate well-connected components.



(a) K_5 (b) D_{10}

Figure 1. Comparison of the connectivity of K_5 (a) and D_{10} (b)

Let $G = (V, E)$ be an undirected, unweighted graph. We want a more robust notion of connectivity to distinguish how well-connected G is. One way is to consider an isoperimetric ratio for a cut $S \subset V$ such that the ratio between the size of its boundary and its volume. For any $S \subset V$ we define its boundary size, $E(S)$, as the number of edges with an endpoint in S and an endpoint in $V \setminus S$, i.e. the number of edges cut by S. We define the volume of a cut S as the number of edge-vertex incidence pairs between edges in the graph and vertices in S such that $\mathrm{vol}(S) \stackrel{\mathrm{def}}{=} \sum_{v \in S} d_v$ where d_v is the degree of vertex v. Let us now define the *conductance* of S as the isoperimetric ratio of interest.

Definition 1.1 (Conductance)**.** *Given a graph $G = (V, E)$ that is undirected and unweighted, the* conductance *of a cut $S \subset V$ is defined as the ratio*

$$\phi(S) \stackrel{\mathrm{def}}{=} \frac{E(S)}{\min\{\mathrm{vol}(S), \mathrm{vol}(V \setminus S)\}}.$$

The conductance of a graph G is defined as

$$\phi(G) \stackrel{\mathrm{def}}{=} \min_{S \subset V} \phi(S)$$

Note that if G is d-regular, for every $S \subset V$, $\mathrm{vol}(S) = d|S|$. Thus, the conductance simplifies to $\phi(S) = \frac{E(S)}{d|S|}$. For the rest of this paper, we will focus on proving statements for the case where G is undirected, unweighted, and d-regular. Unless otherwise stated, we will also assume that G is connected. All the results that we state can be generalized to when G is irregular and arbitrarily weighted without significant modification, and we provide some discussion of this in Section 7.

Conductance now allows us to distinguish how well-connected two graphs are. Let's apply this to the two graphs we mentioned earlier: K_n and D_n. We will show K_n has high conductance and D_n has low conductance.

For K_n, choose a cut $S \subset V$ such that $|S| \leq \frac{n}{2}$. We will show that for any such choice of S, $\phi(S) > \frac{1}{2}$. Let $s = |S|$. We have $E(S) = s(n-s)$, since each of the s vertices in S is connected by an edge to all $n-s$ of the vertices in $V \setminus S$. So $\phi(S) = \frac{s(n-s)}{s(n-1)} = \frac{n-s}{n-1}$. Since $s \leq \frac{n}{2}$, setting $s = \frac{n}{2}$ minimizes $\phi(S)$, implying that $\phi(K_n) \geq \frac{n}{2n-2} > \frac{1}{2}$.

On the other hand, we will show that D_n has conductance $O(\frac{1}{k^2})$. Let S be one of the two copies of $K_{n/2}$. We will show that S has conductance $O(\frac{1}{k^2})$. In fact, S is the conductance-minimizing cut, but since $\phi(G) \leq \phi(S)$, it will not be necessary for us to prove this fact. We have $E(S) = 1$ and $\text{vol}(S) = \frac{n}{2} \cdot \frac{n-2}{2} + 1 \sim \frac{n^2}{4} = \Omega(k^2)$. So $\phi(D_n) \leq \phi(S) = O(\frac{1}{k^2})$.

Our calculations demonstrate that K_n has much higher conductance than D_n, indicating that K_n is "much better connected" than D_n.

1.2. **The Spectrum of the Laplacian and Conductance.** The Laplacian matrix of a d-regular graph, $G = (V, E)$, is the $n \times n$ real, positive semi-definite matrix $\mathbf{L} \stackrel{\text{def}}{=} d \cdot \mathbf{I} - \mathbf{A}$ where $\mathbf{A}$ is the adjacency matrix of G and $\mathbf{I}$ is the identity matrix. A fundamental graph theory result is that G can be separated into two disjoint components, i.e. G is disconnected, iff $\lambda_2 = 0$, where λ_2 is the second smallest eigenvalue of $\mathbf{L}$.

Just as how we view conductance as a robust version of connectivity, one might also hope that a robust analogue of the preceding fact holds. Can a graph be "almost" separated into two components when the second eigenvalue is "almost" 0? This is the subject of the discrete Cheeger's inequality, which we state in terms of the *normalized Laplacian*, $\mathcal{L} \stackrel{\text{def}}{=} \mathbf{I} - \frac{1}{d}\mathbf{A}$:

Theorem 1.2 (Cheeger's Inequality)**.** *Let $G = (V, E)$ be a d-regular graph, and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be eigenvalues of $\mathcal{L}$. Then the following holds.*

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

Cheeger's inequality [4, 5, 1, 2, 6] is a fundamental tool in spectral graph theory. The upper bound states that if λ_2 is small, then $\phi(G)$ is also small, and hence a sparse cut exists. On the other hand, when λ_2 is large, the lower bound states that $\phi(G)$ is large, and thus no sparse cuts exist.

Cheeger's inequality was first proved for manifolds by [4]. The theorem for undirected graphs was first proven by [5] and [1]. The inequality has many applications, including in spectral clustering [7, 8], the construction of expander graphs [9, 10], approximate coloring [2, 11], and image segmentation [12].

We provide the proof of Theorem 1.2, given by [13], in Section 3. The lower bound argument, given in Section 3.1, follows by showing that the variational representation of λ_2 is a relaxation of minimum conductance. The upper bound, shown in Section 3.2, is more difficult as it requires proving the existence of a sparse cut. Consequently, the proof for this side is *algorithmic*; i.e., we construct a procedure which, when given the second eigenvector, outputs a cut with small conductance.

Our approach for the upper bound side specifically follows the *spectral partitioning approach*, first provided by [5]. We choose this approach for several reasons. First, the approach will closely resemble the approach we will take to prove the upper bound side of the higher-order Cheeger's inequality in Section 4 and in Section 5. In fact, there will be significant overlap in the methodologies for the upper bound side proofs provided in Section 3 and Section 4.

Second, this spectral partitioning algorithm interestingly has it's own set of applications. Specifically, it provides an efficient algorithm for finding sparse cuts, and consequently has applications in many practical problems such as web search [14, 15], image segmentation [12, 16], graph coloring [17, 18], and the mixing time of random walks [13].

1.3. The Spectrum of the Laplacian and Multi-Way Conductance.

Another interesting graph theory fact is that G has k disjoint connected components iff $\lambda_k = 0$, where λ_k is the k-th smallest eigenvalue of $\mathbf{L}$. A robust connection between k-way connectivity, and the k-th eigenvalue of $\mathbf{L}$ also exists in the form of higher-order Cheeger inequalities. One way to state these is via k-way conductance.

Definition 1.3 (k-Way Conductance)**.** *Given a graph $G = (V, E)$ and $k > 0$, the* k-way conductance *of G is defined as*

$$\phi_k(G) = \min_{\substack{S_1,\ldots,S_k \subset V \\ S_1,\ldots,S_k \text{ are disjoint}}} \max_{i \in [k]} \phi(S_i)\,.$$

In particular, $\phi_k(G)$ is small when there exists k disjoint cuts each with low conductance, and hence G is close to being separable into k components. When $\phi_k(G)$ is large, then any collection of k disjoint cuts must have contain least one cut that is well-connected to all vertices outside of it. We may suspect that when λ_k is "almost 0" G can "almost" be separated into k components and therefore $\phi_k(G)$ will be small. This is the subject of the higher-order Cheeger's Inequalities.

Theorem 1.4 (Higher-Order Cheeger's Inequality)**.** *Let $G = (V, E)$ be a d-regular graph, and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be eigenvalues of $\mathcal{L}$. Then the following holds.*

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

For fixed k, the smaller the value of λ_k, the smaller the upper bound is on $\phi_k(G)$, indicating that when λ_k is "almost" 0, we can find k components of G that are "almost" disconnected from each other.

Our main goal in this paper will be to prove Theorem 1.4. Theorem 1.4 was first given in [3], but was first conjectured by [19] to establish a statement of Simon and Høegh-Krohn [20] conjectured 40 years prior. The algorithm for the upper bound side of Theorem 1.4 is also inspired by spectral partitioning heuristics used when performing clustering of high-dimensional point clouds [21, 22]. The proofs in this paper closely follow [13] to achieve a dependence of $O(k^{3.5})$, and [3] to reduce this dependence to $O(k^3)$ and subsequently $O(k^2)$.

Similar to Theorem 1.2, the proof of the lower bound, provided in Section 4, follows from a relaxation argument. The upper bound is also proven algorithmically: given the smallest k eigenvectors, we will provide a procedure that outputs k disjoint cuts of G each with small conductance.

An important distinction here is that constructing this k-way spectral partitioning algorithm is more difficult than the partitioning algorithm used to prove Theorem 1.2. In Theorem 1.2, the spectral partitioning algorithm performs a line embedding of the vertices in G given by the second eigenvector, $\boldsymbol{x}^{(2)}$, of $\mathcal{L}$, where $i \mapsto \boldsymbol{x}_i^{(2)}$ the i-th coordinate of $\boldsymbol{x}^{(2)}$. The line-embedding then provides a canonical way of distinguishing two well-separated sets: simply select sets of vertices that are far apart on the line. In some sense, a crucial consequence of the upper bound side of Theorem 1.2 is that it shows well-separated sets in the line embedding correspond to minimum conductance cuts in the graph.

In higher dimensions, one can try to perform an embedding of G into $\mathbb{R}^k$ via

$$i \mapsto \begin{pmatrix} \boldsymbol{x}_i^{(1)} & \boldsymbol{x}_i^{(2)} & \cdots & \boldsymbol{x}_i^{(k)} \end{pmatrix}^\top$$

where $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are the smallest k eigenvectors of $\mathcal{L}_G$. We would then like to show that "well-separated" sets in $\mathbb{R}^k$ correspond to k cuts in G that have low k-way conductance. A key distinction is that, in higher-dimensions, using different notions of "well-separated" lead to qualitatively different connections with k-way conductance that manifest via the dependence on k in the upper bound.

What is the right notion of well-separated that results in the tightest dependence on k in the upper bound? This paper will show, in sequence, how to prove the upper bound with an $O(k^{3.5})$, $O(k^3)$, and $O(k^2)$ dependence. In Section 4, we provide the central argument which reduces proving the upper bound down to proving the existence of a partitioning of $\mathbb{R}^k$ with certain "desirable" properties. This is done by providing an algorithm that rounds k cuts with small k-way conductance when given a "desirable" partition of $\mathbb{R}^k$. We provide a simple construction of such a partition to get the initial $O(k^{3.5})$ bound. In Section 5, we formalize the "desirable" properties and show the existence of such a partition using the *Padded Partition Theorem* in a black-box way to improve the dependence on k to $O(k^3)$. Finally, for the $O(k^2)$ bound, we modify the partition properties to better take into account the fact that there are only finitely many embedded vectors in $\mathbb{R}^k$ that we need to focus on. We use the *Lipschitz Partition Theorem*, provided in Section 6, to show the existence of such a partition, and show how it improves the dependence on k to $O(k^2)$.

Finally, we note that there are, in fact, many different "higher-order Cheeger inequalities". These depend both on how the notion of k-way connectivity is defined, and which part of the spectrum is used to relate to k-way conductance. For example, note that in our definition of conductance, Definition 1.3 does not require that the cuts, $S_1, \cdots, S_k$ form a partition of V, i.e. that $S_1 \cup S_2 \cup \cdots \cup S_k = V$. Adding this requirement results in a looser upper bound. Another example is that rather than bounding $\phi_k(G)$ using λ_k, we may instead choose to use λ_{2k}, which would improve the dependence on k. We discuss some of these higher-order Cheeger inequalities in Section 7.

2. Preliminaries

Throughout this paper we will denote two vertices of G, u and v as adjacent via $u \sim v$. We use unbolded x to represent scalars, bold-face lowercase $\boldsymbol{x}$ to represent vectors, and bold-face uppercase $\mathbf{X}$ to represent matrices. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we let $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ denote the standard Euclidean inner product, and $\|\boldsymbol{x}\|$ denote the induced ℓ_2-norm. We will often consider the *indicator vector* of a cut. If $S \subset V$, then its indicator vector is $\mathbf{1}_S$ where

$$(\mathbf{1}_S)_v = \begin{cases} 1 & \text{if } v \in S \\ 0 & \text{otherwise} \end{cases}$$

2.1. **Rayleigh Quotients and Eigenvalues.** Recall that a matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ is symmetric if $\mathbf{M} = \mathbf{M}^\top$. Given a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, its *Rayleigh quotient* with respect to $\boldsymbol{x} \in \mathbb{R}^n$ is defined as

$$R_{\mathbf{M}}(\boldsymbol{x}) \stackrel{\text{def}}{=} \frac{x^\top \mathbf{M} x}{\boldsymbol{x}^\top \boldsymbol{x}} .$$

One property of the Rayleigh quotient is that it is constant under scaling, i.e. for any constant $c \in \mathbb{R}^n$, $R_M(c\boldsymbol{x}) = \frac{c\boldsymbol{x}^\top \mathbf{M} c\boldsymbol{x}}{c\boldsymbol{x} c\boldsymbol{x}^\top} = \frac{\boldsymbol{x}^\top \mathbf{M}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}} = R_M(\boldsymbol{x})$.

An important fact about symmetric matrices is that all their eigenvalues are real and can be expressed variationally using the Courant–Fischer theorem.

Lemma 2.1. *Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be the eigenvalues of a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$. For every $k \in [n]$, we have*

$$\lambda_k = \min_{\substack{U \subset \mathbb{R}^n \\ \dim U = k}} \max_{\boldsymbol{x} \in U \setminus \mathbf{0}} R_{\mathbf{M}}(\boldsymbol{x})$$

For the specific case of λ_2, we have

Lemma 2.2. *We have*

$$\lambda_2 = \min_{\substack{U \subset \mathbb{R}^n \\ \dim U = 2}} \max_{\boldsymbol{x} \in U \setminus \mathbf{0}} R_{\mathbf{M}}(\boldsymbol{x}) = \min_{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \perp \boldsymbol{x}^{(1)}} R_{\mathbf{M}}(\boldsymbol{x})$$

where $\boldsymbol{x}^{(1)}$ is an eigenvector associated with eigenvalue λ_1 of $\mathbf{M}$.

Finally, recall that a symmetric matrix $\mathbf{M}$ is *positive semi-definite* if and only if $\boldsymbol{x}^\top M\boldsymbol{x} \geq 0$ for every $\boldsymbol{x} \in \mathbb{R}^n$. In particular, we will prove in Corollary 2.7 that $\mathcal{L}$ is positive semi-definite.

For the normalized Laplacian, $\mathcal{L}$, the smallest eigenvalue is 0, and $\mathbf{1}$ is an eigenvector of eigenvalue 0. So we get the following corollary.

Corollary 2.3. *Given a d-regular graph $G = (V, E)$, let λ_2 denote the second smallest eigenvalue of $\mathcal{L}$. Then*

$$\lambda_2 = \min_{\boldsymbol{x}\perp\mathbf{1}} \; R_{\mathcal{L}}(\boldsymbol{x}).$$

Corollary 2.3 will be a key fact in proving Theorem 1.2.

2.2. **The Rayleigh Quotient of the Laplacian.** Given a graph $G = (V, E)$, denote $\mathcal{L}$ by its normalized Laplacian. The Rayleigh quotient with respect to $\mathcal{L}$ will be a fundamental tool in studying the conductance of G as it links the eigenvalues of $\mathcal{L}$ and the graph's conductance. For example, a key fact is that the conductance of a cut equals the Rayleigh quotient of the corresponding indicator vector.

Lemma 2.4. *Given a graph $G = (V, E)$ and a vector $\boldsymbol{x} \in \mathbb{R}^k$, we have*

$$R_{\mathcal{L}}(\boldsymbol{x}) = \frac{\sum\limits_{u\sim v} (x_u - x_v)^2}{d \sum\limits_{v} x_v^2}$$

Proof. We start by expanding the Rayleigh quotient via

$$\begin{aligned}
R_{\mathcal{L}}(x) &= \frac{\boldsymbol{x}^\top \mathcal{L} \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}} \\
&= \frac{\boldsymbol{x}^\top \mathbf{I}\boldsymbol{x} - \frac{1}{d}\boldsymbol{x}^\top \mathbf{A}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}} \\
&= \frac{\sum\limits_{v\in V} (x_v)^2 - 2\sum\limits_{u\sim v} \frac{x_u x_v}{d}}{\boldsymbol{x}^\top \boldsymbol{x}} \\
&= \frac{d\sum\limits_{v\in V} (x_v)^2 - 2\sum\limits_{u\sim v} x_u x_v}{d\boldsymbol{x}^\top \boldsymbol{x}} \\
&= \frac{\sum\limits_{v\in V} (x_v)^2 d_v - 2\sum\limits_{u\sim v} x_u x_v}{d\boldsymbol{x}^\top \boldsymbol{x}}.
\end{aligned}$$

Let's calculate the numerator of the Rayleigh quotient. We consider the contribution of each pair of vertices, u and v, to the numerator. If $u \not\sim v$, then there is no contribution to $(x_u)^2 d_u$, $(x_v)^2 d_v$, or $\sum\limits_{u\sim v} x_u x_v$. If $u \sim v$, there is a contribution of $(x_u)^2$ to $(x_u)^2 d_u$, $(x_v)^2$ to $(x_v)^2 d_v$, and $2x_u x_v$ to $2\sum\limits_{u\sim v} x_u x_v$. The total contribution to the numerator is $(x_u)^2 + (x_v)^2 - 2x_u x_v = (x_u - x_v)^2$.

The denominator is $d\boldsymbol{x}^\top \boldsymbol{x} = d\sum_{v\in V} {x_v}^2$. Therefore, we have

$$R_{\mathcal{L}}(x) = \frac{\sum_{u\sim v} (x_u - x_v)^2}{d\sum_{v\in V} (x_v)^2}\,. \qquad \square$$

In the specific case where $\boldsymbol{x}$ is an indicator variable of a cut $S \subset V$, the Rayleigh quotient of $\mathbf{1}_S$ equals the conductance of S.

Corollary 2.5. *Given a graph $G = (V, E)$ and a cut $S \subset V$, we have*

$$R_{\mathcal{L}}(\boldsymbol{x}) = \frac{\sum_{u\sim v} (x_u - x_v)^2}{d|S|} = \phi(S)$$

As a consequence of Corollary 2.5, the conductance of G is equivalent to minimizing the Rayleigh quotient over a portion of the Boolean hypercube, i.e. over $\{0, 1\}$ vectors.

Corollary 2.6. *If $G = (V, E)$ is such that $|V| = n$, then*

$$\phi(G) = \min_{\boldsymbol{x}\in\{0,1\}^n : \|\boldsymbol{x}\|^2 \leq \frac{n}{2}} R_{\mathcal{L}}(x)\,.$$

We also have that $\mathcal{L}$ is positive semidefinite since $\boldsymbol{x}^\top \mathcal{L}\boldsymbol{x} = \sum_{u\sim v} (\boldsymbol{x}_u - \boldsymbol{x}_v)^2 \geq 0$.

Corollary 2.7. *$\mathcal{L}$ is positive semi-definite*

3. Cheeger's Inequality

In this section, we prove Cheeger's Inequality, Theorem 1.2, which states that for a regular, undirected, and unweighted graph, we have $\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$

3.1. Lower Bound.

We first aim to prove the lower bound $\frac{\lambda_2}{2} \leq \phi(G)$. The argument follows by demonstrating that λ_2 is a relaxation of $\phi(G)$.

To begin, note that by Lemma 2.2

$$\lambda_2 = \min_{\substack{U\subset\mathbb{R}^n\\ \dim U = 2}} \max_{\boldsymbol{x}\in U\setminus\mathbf{0}} \{R_{\mathcal{L}}(\boldsymbol{x})\}\,.$$

So it suffices to find a 2-dimensional space, U, where $\max_{\boldsymbol{x}\in U}\{R_{\mathcal{L}}(\boldsymbol{x})\} \leq 2\phi(G)$. For this subsection, we will fix $S \subset V$ to be the minimum conductance cut, i.e.

$$\phi(S) = \phi(G)$$

Now, let $U = \mathrm{Span}(\mathbf{1}_S, \mathbf{1}_{V\setminus S})$. Any vector in U is of the form $a\mathbf{1}_S + b\mathbf{1}_{V\setminus S}$. Since the Rayleigh quotient is constant under scaling,

$$R_{\mathcal{L}}(a\mathbf{1}_S) = R_{\mathcal{L}}(\mathbf{1}_S) = \phi(G)\,,$$

while on the other hand

$$R_{\mathcal{L}}(\mathbf{1}_{V\setminus S}) = \phi(V\setminus S) = \frac{E(V\setminus S)}{\mathrm{vol}(V\setminus S)} = \frac{E(S)}{\mathrm{vol}(V\setminus S)} \leq \frac{E(S)}{\mathrm{vol}(S)} = \phi(S) = \phi(G)$$

Could it be that because any vector, $\boldsymbol{x}$, in U is the sum of two vectors whose individual Rayleigh quotients are at most $\phi(G)$, that $R_{\mathcal{L}}(\boldsymbol{x}) \leq 2\phi(G)$? Since $\mathbf{1}_S$ and $\mathbf{1}_{V\setminus S}$ are orthogonal this is, in fact, the case.

Lemma 3.1. *If $\boldsymbol{x} \perp \boldsymbol{y}$, $R_{\mathcal{L}}(\boldsymbol{x}+\boldsymbol{y}) \leq 2\cdot\max\left\{R_{\mathcal{L}}(\boldsymbol{x}), R_{\mathcal{L}}(\boldsymbol{y})\right\}$*

Proof. We have

$$\begin{aligned}
R_{\mathcal{L}}(\boldsymbol{x}+\boldsymbol{y}) &= \frac{\sum\limits_{u\sim v}(x_u - x_v + y_u - y_v)^2}{||\boldsymbol{x}+\boldsymbol{y}||^2} \\
&= \frac{\sum\limits_{u\sim v}(x_u - x_v + y_u - y_v)^2}{||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2} \\
&= \frac{\sum\limits_{u\sim v}(x_u - x_v)^2 + (y_u - y_v)^2 + 2(x_u - x_v)(y_u - y_v)}{||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2}.
\end{aligned}$$

Since $2(x_u - x_v)(y_u - y_v) \leq (x_u - x_v)^2 + (y_u - y_v)^2$, we also have the following.

$$\begin{aligned}
&\frac{\sum\limits_{u\sim v}(x_u - x_v)^2 + (y_u - y_v)^2 + 2(x_u - x_v)(y_u - y_v)}{||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2} \\
&\leq \frac{\sum\limits_{u\sim v}2(x_u - x_v)^2 + 2(y_u - y_v)^2}{||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2} \\
&\leq 2\cdot\max\{R_{\mathcal{L}}(\boldsymbol{x}), R_{\mathcal{L}}(\boldsymbol{y})\}
\end{aligned}$$

where the last inequality follows by Lemma A.1. □

In fact, as shown in [13], Lemma 3.1 applies to the Rayleigh quotient with respect to any positive semi-definite matrix. In particular, as shown in Corollary 2.7, $\mathcal{L}$ is positive semi-definite.

We now have that every vector in U has Rayleigh quotient at most $2\phi(G)$.

Corollary 3.2. $\forall \boldsymbol{w} \in \mathrm{Span}(\mathbf{1}_S, \mathbf{1}_{V\setminus S})$, $R_{\mathcal{L}}(\boldsymbol{w}) \leq 2\phi(S) = 2\phi(G)$

Proof. This fact follows from Lemma 3.1. Let $\boldsymbol{w} = a\mathbf{1}_S + b\mathbf{1}_{V\setminus S}$, where $a, b \in \mathbb{R}$ and $\boldsymbol{w} \neq \mathbf{0}$. The Rayleigh quotient is constant under scaling, so $R_{\mathcal{L}}(a\mathbf{1}_S) = R_{\mathcal{L}}(\mathbf{1}_S) = \phi(1_S) = \phi(G)$ and $R_{\mathcal{L}}(b\mathbf{1}_{V\setminus S}) = R_{\mathcal{L}}(\mathbf{1}_{V\setminus S}) = \phi(V\setminus S) \leq \phi(G)$. By Lemma 3.1, $R_{\mathcal{L}}(\boldsymbol{w}) \leq 2\phi(G)$ □

Since U is a 2-dimensional space, the lower bound side of Cheeger's inequality follows from Lemma 2.2.

Lemma 3.3 (Cheeger's Lower Bound)**.** $\frac{\lambda_2}{2} \leq \phi(G)$

Proof. By Lemma 2.2, for all 2-dimensional spaces U,

$$\lambda_2 \leq \max_{\boldsymbol{x} \in U \setminus \mathbf{0}} \frac{(x_u - x_v)^2}{d \sum\limits_{v \in V} {x_v}^2}$$

So

$$\lambda_2 \leq \max_{v \in \mathrm{Span}(\mathbf{1}_S, \mathbf{1}_{V \setminus S}) \setminus \mathbf{0}} R_{\mathcal{L}}(\boldsymbol{x}) \leq 2\phi(G)$$

It follows that $\frac{\lambda_2}{2} \leq \phi(G)$, as desired. $\square$

3.2. **Upper Bound.** In this section, we prove that $\phi(G) \leq \sqrt{2\lambda_2}$.

3.2.1. *Description of Algorithm.* Because $\phi(G)$ is the minimum conductance achievable by any cut $S \subset V$, it suffices to find a cut with conductance at most $\sqrt{2\lambda_2}$. We do this via a spectral partitioning algorithm known as *Fiedler's Algorithm*, which we outline below:

1. Let $\boldsymbol{x}$ be an eigenvector of eigenvalue λ_2. Embed the vertices of the graph into $\mathbb{R}$ via its spectral embedding: $v \mapsto x_v$, which embeds each vertex into its corresponding entry of the second eigenvalue.

2. Consider cuts formed by choosing a threshold, $t \in \mathbb{R}$, and then partitioning V into $\Big(\{v \in V : x_v < t\}, \{v \in V : x_v \geq t\}\Big)$.

3. Among the cuts considered in (2), output the one with the lowest conductance.

Why would we think to use the spectral embedding $v \mapsto x_v$? We have already seen a strong connection between the eigenvalues of the Laplacian and the conductance of a graph via the Rayleigh quotient. Additionally, the key idea for the lower bound side was equating a cut, S, with a vector, $\mathbf{1}_S$, to relate the conductance of a cut with the eigenvalues of the Laplacian.

For the upper bound, we want to go the other direction. Given a vector of small Rayleigh quotient, we want to make a cut from that vector with low conductance. Making a cut from a vector is not as straightforward as making a vector out of a cut, as the vector we start with is most likely not a $\{0, 1\}$ vector.

Therefore, the key idea is to use rounding. We assign a threshold value t. We then *round* all entries below t to 0, and all entries above t to 1. In other words, our cut consists of all vertices, v, such that $x_v \geq t$. Since we want a cut with small conductance, we can choose a value of t that minimizes the conductance of the cut.

But why do we use an eigenvector of eigenvalue λ_2? It may seem logical to use a vector that minimizes the Rayleigh quotient, i.e. an eigenvector of eignevalue λ_1. There are a few reasons for using λ_2 over λ_1.

The main reason is that for the Laplacian of every graph, $\lambda_1 = 0$, and the first eigenvector of the Laplacian is always a scalar multiple of $\mathbf{1}$. Therefore, using λ_1 for the embedding would assign the same value to every vertex and not be helpful in forming low-conductance cuts. Additionally, we make critical use of the fact that λ_2 is orthogonal to $\mathbf{1}$ in Claim 3.4.1.

By Corollary 2.3, among vectors orthogonal to $\mathbf{1}$, an eigenvector of eigenvalue λ_2 minimizes the Rayleigh quotient with respect to $\mathcal{L}$. As a result of this choice for $\boldsymbol{x}$, the upper bound of Cheeger's will be in terms of λ_2.

So we now need to show that for any vector, $\boldsymbol{x} : \boldsymbol{x} \perp \mathbf{1}$, Fiedler's algorithm does indeed yield a cut with conductance $\leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{x})}$. It turns out that it is not easy to directly show that the cut found via Fiedler's algorithm has sufficiently small conductance. So our strategy will be as follows.

First, we find a vector, $\boldsymbol{y}$, such that:

(1) $\boldsymbol{y}$ is supported on a set of size $\leq \frac{n}{2}$
(2) $\boldsymbol{y}$ has only nonnegative entries
(3) $\boldsymbol{y}$ has Rayleigh quotient $\leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{x})}$.

Finding such a $\boldsymbol{y}$ will be helpful for two reasons. First, if we let $S_{\boldsymbol{y}}$ be the cut found via Fiedler's algorithm applied to $\boldsymbol{y}$, it is easier to show that $\phi(S_{\boldsymbol{y}}, V \setminus S_{\boldsymbol{y}}) \leq 2R_{\mathcal{L}}(\boldsymbol{y})$ than it is to show the corresponding result for $\boldsymbol{x}$.

Additionally, showing $\phi(S_{\boldsymbol{y}}, V \setminus S_{\boldsymbol{y}}) \leq 2R_{\mathcal{L}}(\boldsymbol{y})$ is sufficient because:

(1) Let $S_{\boldsymbol{x}}$ be the cut found via Fiedler's algorithm applied to $\boldsymbol{x}$. Because of our choice of $\boldsymbol{y}$, the cut $S_{\boldsymbol{y}}$ is one of the cuts considered by Fiedler's algorithm on $\boldsymbol{x}$, i.e. exist, t, such that the cut formed by applying the threshold t on $\boldsymbol{x}$ yields the cut $S_{\boldsymbol{y}}$.
(2) We also have that $R_{\mathcal{L}}(\boldsymbol{y}) \leq R_{\mathcal{L}}(\boldsymbol{x})$. So collectively, $\phi(S_{\boldsymbol{x}}) \leq \phi(S_{\boldsymbol{y}}) \leq R_{\mathcal{L}}(\boldsymbol{y}) \leq R_{\mathcal{L}}(\boldsymbol{x})$.

Secondly, this approach will provide us with the *Vector-Cut Theorem*, Theorem 3.5, which will be a critical tool in the proof of the higher-order Cheeger inequality.

We now begin the proof.

3.2.2. *Rigorous Proof.* The first goal is to find a vector, $\boldsymbol{y}$, such that:

(1) $\boldsymbol{y}$ has nonnegative entries
(2) $\boldsymbol{y}$ is supported on a set of size $\leq \frac{n}{2}$
(3) $R_{\mathcal{L}}(\boldsymbol{y}) \leq R_{\mathcal{L}}(\boldsymbol{x})$.

(1) and (2) ensure that after setting a threshold, t, $\{v : x_v \geq t\}$ is a set of size $\leq \frac{n}{2}$. (3) ensures that finding a cut using Fiedler's Algorithm on $\boldsymbol{y}$ with sufficiently small Rayleigh quotient guarentees that we can do the same on $\boldsymbol{x}$.

Lemma 3.4. *$\forall \boldsymbol{x} \perp \mathbf{1}$, we can find $\boldsymbol{y}$ satisfying (1) and (2) and (3).*

Proof. The idea is to first shift the entries so that half of them are negative and half of them are positive. Then, we either (1) set the negative entries to 0 or (2) set the positive entries to 0 and then take the absolute value of the negative entries. We will show that either (1) or (2) will produce a vector with Rayleigh quotient $\leq R_{\mathcal{L}}(\boldsymbol{x})$.

First, let m be the median value of $\{x_1, \cdots, x_n\}$. Apply a shift of $-m$ to all entries of $\boldsymbol{x}$ to obtain the vector $\boldsymbol{x'} = \boldsymbol{x} - m\mathbf{1}$. If n is even, after the shift, half the entries will be non-positive and half the entries will be non-negative. If n is odd, the median entry will be 0, so less than half the entries will be non-negative and less than half the entries will be non-positive.

Therefore, we can construct two vectors each with support of size $\leq \frac{n}{2}$ as follows:

Let $\boldsymbol{x}^+$ be the vector obtained by setting all negative entries of $\boldsymbol{x'}$ to 0, i.e. for every $i \in [n]$,

$$\boldsymbol{x}_i^+ = \max\{\boldsymbol{x}_i, 0\}$$

Let $\boldsymbol{x}^-$ be the vector obtained by setting all positive entries of $\boldsymbol{x'}$ to 0, and taking the absolute value of all negative entries. In other words, for every $i \in [n]$,

$$\boldsymbol{x}_i^- = |\min\{\boldsymbol{x}_i, 0\}|$$

Claim 3.4.1. *Either $R_{\mathcal{L}}(\boldsymbol{x}^+) \leq R_{\mathcal{L}}(\boldsymbol{x})$ or $R_{\mathcal{L}}(\boldsymbol{x}^-) \leq R_{\mathcal{L}}(\boldsymbol{x})$.*

Proof. First, note that shifting the entries of $\boldsymbol{x}$ by $-m$ does not affect the numerator of the Rayleigh quotient. But because $\boldsymbol{x} \perp 1$, we have that the denominator of the Rayleigh quotient of $\boldsymbol{x'}$ is no smaller than that of $\boldsymbol{x}$. In particular,

$$||\boldsymbol{x'}|| = ||\boldsymbol{x} - m\mathbf{1}|| = ||\boldsymbol{x}|| + ||m|| \geq ||\boldsymbol{x}||.$$

Therefore, the shifted version of $\boldsymbol{x}$, $\boldsymbol{x'}$, has Rayleigh quotient $\leq$ that of $\boldsymbol{x}$. We have

$$R_{\mathcal{L}}(\boldsymbol{x'}) = \frac{\sum\limits_{u\sim v}(x'_u - x'_v)^2}{||\boldsymbol{x'}||^2} = \frac{\sum\limits_{u\sim v}(x_u - x_v)^2}{||\boldsymbol{x'}||^2} \leq \frac{\sum\limits_{u\sim v}(x_u - x_v)^2}{||\boldsymbol{x}||^2} = R_{\mathcal{L}}(\boldsymbol{x})$$

So it suffices to show that $R_{\mathcal{L}}(\boldsymbol{x'}) \geq \min\{(R_{\mathcal{L}}(\boldsymbol{x}^+), R_{\mathcal{L}}(\boldsymbol{x}^-)\}$. We have

$$\begin{aligned} R_{\mathcal{L}}(\boldsymbol{x'}) &= \frac{\sum\limits_{u\sim v}(x_u - x_v)^2}{||\boldsymbol{x'}||^2} \\ &= \frac{\sum\limits_{u\sim v}(x_u^+ - x_v^+ + x_u^- - x_v^-)^2}{||\boldsymbol{x}^+||^2 + ||\boldsymbol{x}^-||^2} \end{aligned}$$

$$= \frac{\sum_{u \sim v} (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2 - 2(x_u^+ - x_v^+)(x_u^- - x_v^-)}{||\boldsymbol{x}^+||^2 + ||\boldsymbol{x}^-||^2}$$

Note that $-2(x_u^+ - x_v^+)(x_u^- - x_v^-) \geq 0$ because $(x_u^+ - x_v^+) \geq 0$ and $(x_u^- - x_v^-) \leq 0$. So

$$\frac{\sum_{u \sim v} (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2 - 2(x_u^+ - x_v^+)(x_u^- - x_v^-)}{||\boldsymbol{x}^+||^2 + ||\boldsymbol{x}^-||^2}$$
$$\geq \frac{\sum_{u \sim v} (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2}{||\boldsymbol{x}^+||^2 + ||\boldsymbol{x}^-||^2}.$$

By Lemma A.1,

$$\frac{\sum_{u \sim v} (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2}{||\boldsymbol{x}^+||^2 + ||\boldsymbol{x}^-||^2} \geq \min\{R_{\mathcal{L}}(\boldsymbol{x}^+), R_{\mathcal{L}}(\boldsymbol{x}^-)\}$$ □

By Claim 3.4.1, we can set $\boldsymbol{y}$ to either $\boldsymbol{x}^+$ or $\boldsymbol{x}^-$, proving Lemma 3.4. □

We now need to show that for the spectral embedding, we can find a cut, $S \subset \operatorname{supp}(\boldsymbol{y})$, such that $\phi(S) \leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{y})}$.

To do this, we first prove the Vector-Cut Theorem.

Theorem 3.5 (Vector-Cut Theorem)**.** *Given $y \in \mathbb{R}^k$ with non-negative entries, there exists a cut, $S \subset \operatorname{supp}(\boldsymbol{y}) : \phi(S) \leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{y})}$.*

Theorem 3.5 will imply applying Fiedler's algorithm on $\boldsymbol{y}$ will yield a cut with conductance that is small relative to $R_{\mathcal{L}}(\boldsymbol{y})$.

Theorem 3.6, will show that $R_{\mathcal{L}}(\boldsymbol{y}) \leq R_{\mathcal{L}}(\boldsymbol{x})$, which will show the conductance of the cut is small relative to $R_{\mathcal{L}}(\boldsymbol{x})$ and complete the proof of Cheeger's inequality.

In addition to proving Cheeger's inequality, Theorem 3.5 will be a critical tool in proving the multi-way Cheeger's inequalities in the later sections.

Proof. We want to prove that using Fiedler's Algorithm, some threshold value for t will provide a sufficiently low-conductance cut. We prove that such a t exists via the probabilistic method.

Pick $t \in (0, \max\{y_v^2\})$ uniformly. WLOG, we may assume $\max_{v \in V}\{y_v\} = 1$ because the Rayleigh quotient is constant under scaling (i.e. we can scale the spectral embedding so that $\max(y_v) = 1$ without affecting the Rayleigh quotient).

Let $S_t = \{v : y_v^2 \geq t\}$. The goal will be to show that $\frac{\mathbb{E}[E(S_t)]}{\mathbb{E}[d|S_t|]} \leq \sqrt{2R_{\mathcal{L}}\boldsymbol{y}}$.

In particular, we will show that the denominator of the LHS is $\geq$ the denominator of the Rayleigh quotient, and the numerator of the LHS is $\leq$ that of the Rayleigh quotient.

For the denominator, by linearity of expectation, we can consider the contribution of each individual vertex to $\mathbb{E}[d|S_t|]$. We have

$$\mathbb{E}[d|S_t|] = d\sum_v \mathbb{P}[y_v^2 \leq t] = d\sum_v y_v^2$$

This is precisely the denominator of the Rayleigh quotient.

For the numerator, we consider the contribution of each individual edge, $\{u,v\}$ to $\mathbb{E}[E(S_t)]$. We have

$$\mathbb{E}[E(S_t)] = \sum_{u\sim v} \mathbb{P}\left(\{u,v\} \text{ is cut}\right) = \sum_{u\sim v} |y_v^2 - y_u^2| = \sum_{u\sim v} |y_u - y_v| \cdot (y_u + y_v)$$

We then apply Cauchy-Schwartz:

$$\mathbb{E}[E(S_t)] = \sum_{u\sim v} |y_u - y_v| \cdot (y_u + y_v) \geq \sqrt{\sum_{u\sim v} |y_u - y_v|^2} \cdot \sqrt{\sum_{u\sim v} (y_u + y_v)^2}$$

Note that $(y_u + y_v)^2 \leq 2(y_u^2 + y_v^2)$, so

$$\begin{aligned}
\mathbb{E}[E(S_t)] &\geq \sqrt{\sum_{u\sim v} |y_u - y_v|^2} \cdot \sqrt{\sum_{u\sim v} (y_u + y_v)^2} \\
&\geq \sqrt{\sum_{u\sim v} |y_u - y_v|^2} \cdot \sqrt{\sum_{u\sim v} 2\left(y_u^2 + y_v^2\right)} \\
&= \sqrt{\sum_{u\sim v} |y_u - y_v|^2} \cdot \sqrt{2d\sum_v y_v^2}
\end{aligned}$$

Dividing, we get

$$\begin{aligned}
\frac{\mathbb{E}[E(S_t)]}{\mathbb{E}[d|S_t|]} &\leq \frac{\sqrt{\sum_{u\sim v} |y_u - y_v|^2} \cdot \sqrt{2d\sum_v y_v^2}}{d\sum_v y_v^2} \\
&= \sqrt{2 \cdot \frac{\sum_{u\sim v} |y_u - y_v|^2}{d\sum_v y_v^2}} = \sqrt{2R_{\mathcal{L}}(\boldsymbol{y})}
\end{aligned}$$

It may seem that we are done, but to finish the proof, we need to show that

$$\frac{\mathbb{E}[E(S_t)]}{\mathbb{E}[d|S_t|]} \leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{y})}$$

implies that the inequality $\frac{E(S_t)}{d|S_t|} \leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{y})}$ is satisfied for some t. This follows from Lemma A.2. $\square$

Finally, to complete the proof of Cheeger's inequality, we will show that $\phi(S_{\boldsymbol{x}}) \leq \phi(S_{\boldsymbol{y}})$ and $R_{\mathcal{L}}(\boldsymbol{y}) \leq R_{\mathcal{L}}(\boldsymbol{x})$ to show that Fiedler's algorithm applied to $\boldsymbol{x}$ yields a cut of low conductance relative to $R_{\mathcal{L}}(\boldsymbol{x})$.

Theorem 3.6 (Vector-Cut Theorem for Vectors $\perp \mathbf{1}$). *Given $\boldsymbol{x} \in \mathbb{R}^k : \boldsymbol{x} \perp \mathbf{1}$, exists a cut, $S \subset \operatorname{supp}(\boldsymbol{x})$, such that $\phi(S) \leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{x})}$*

Theorem 3.6 will imply the upper bound of Cheeger's inequality since the second eigenvector of the Laplacian is $\perp \mathbf{1}$.

Proof. It follows from Lemma 3.4 that $R_{\mathcal{L}}(\boldsymbol{y}) \leq R_{\mathcal{L}}(\boldsymbol{x})$.

So it suffices to show $\phi(S_{\boldsymbol{x}}) \leq \phi(S_{\boldsymbol{y}})$.

This follows because using a threshold of t on $\boldsymbol{y}$ is equivalent to applying a threshold on $\boldsymbol{x}$ of either (1) $t+m$ if $\boldsymbol{y} = \boldsymbol{x}^+$, or (2) $m-t$ if $\boldsymbol{y} = \boldsymbol{x}^-$. The cut found via Fiedler's algorithm on $\boldsymbol{x}$ has conductance at least as low as the cut found by setting a threshold of $t+m$ or $m-t$ on $\boldsymbol{x}$. So it follows that $\phi(S_{\boldsymbol{x}}) \leq \phi(S_{\boldsymbol{y}}) \leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{y})} \leq \sqrt{2R_{\mathcal{L}}(\boldsymbol{x})}$. $\square$

Finally, setting $\boldsymbol{x}$ to an eigenvector of eigenvalue λ_2 yields the upper bound side of Cheeger's inequality.

Theorem 1.2. (Cheeger's Inequality). $\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$

Proof. We proved the lower bound via Lemma 3.3. For the upper bound, use Theorem 3.6, setting $\boldsymbol{x}$ to the second eigenvector of $\mathcal{L}$. $\square$

4. Higher-Order Cheeger, Upper Bound $O(k^{3.5})$

In this section, we will prove higher-order Cheeger, with a dependence of $O(k^{3.5})$ in the upper bound:

$$\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^{3.5})\sqrt{\lambda_k}$$

4.1. **Lower Bound.** We first show that $\frac{\lambda_k}{2} \leq \phi_k(G)$. The proof is very similar to the proof for the lower bound side of Cheeger's inequality, only instead of finding a 2-dimensional space with low Rayleigh quotient and using Lemma 2.2, we find a k-dimensional space using Lemma 2.1.

First, let $S_1, \cdots, S_k \subset V : \phi(S_1), \cdots, \phi(S_k) \leq \phi_k(G)$, and for all $i \in [k]$, let $\boldsymbol{x}^{(i)} = \mathbf{1}_{S_i}$. Fix $U = \mathrm{Span}(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(k)})$.

We hope that, similar to the lower bound side of Cheeger's inequality, since $\forall i \in [k]$, $R_{\mathcal{L}}(\boldsymbol{x}^{(i)}) \leq \phi_k(G)$, all vectors in U have Rayleigh quotient $\leq 2\phi_k(G)$. This would prove that $\lambda_k \leq 2\phi_k(G)$ since $\lambda_k \leq \max_{x \in U}\{R_{\mathcal{L}}(x)\}$.

For a symmetric matrix, $\mathbf{M}$, and orthogonal vectors $\boldsymbol{y}^{(1)}, \cdots, \boldsymbol{y}^{(k)}$, we can do no better than $R_{\mathbf{M}}\big(\boldsymbol{y}^{(1)} + \cdots + \boldsymbol{y}^{(k)}\big) \leq k \cdot \max\{R_{\mathbf{M}}(\boldsymbol{y}^{(1)}), \cdots, R_{\mathbf{M}}(\boldsymbol{y}^{(k)})\}$, where we lose a factor of k rather than a factor of 2.

However, because $\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(k)}$ are not only orthogonal, but also disjointly supported, and because we are dealing specifically with the normalized Laplacian, we only lose a factor of 2 instead of a factor of k.

Lemma 4.1. *If $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}$ are disjointly supported vectors, then*
$$R_{\mathcal{L}}\big(\boldsymbol{x}^{(1)} + \cdots + \boldsymbol{x}^{(k)}\big) \leq 2 \cdot \max\{R_{\mathcal{L}}(\boldsymbol{x}^{(i)})\}$$

Proof. Let $\boldsymbol{x} = \boldsymbol{x}^{(1)} + \cdots + \boldsymbol{x}^{(k)}$.
$$R_{\mathcal{L}}\big(\boldsymbol{x}^{(1)} + \cdots + \boldsymbol{x}^{(k)}\big) = \frac{\sum\limits_{u \sim v} (x_u - x_v)^2}{||\boldsymbol{x}||^2} = \frac{\sum\limits_{u \sim v} \Big(\sum\limits_{i \in [k]} x_u^{(i)} - x_v^{(i)}\Big)^2}{\sum\limits_{i \in [k]} ||\boldsymbol{x}^{(i)}||^2}.$$

Additionally, by Lemma A.1,
$$\max_{i \in [k]}\{R_{\mathcal{L}}(\boldsymbol{x}^{(i)})\} = \max_{i \in [k]} \left\{ \frac{\sum\limits_{u \sim v} (x_u^{(i)} - x_v^{(i)})^2}{||\boldsymbol{x}^{(i)}||^2} \right\} \geq \frac{\sum\limits_{u \sim v} \sum\limits_{i \in [k]} (x_u^{(i)} - x_v^{(i)})^2}{\sum\limits_{i \in [k]} ||\boldsymbol{x}^{(i)}||^2}$$

So it suffices to show that
$$\frac{\sum\limits_{u \sim v} \Big(\sum\limits_{i \in [k]} x_u^{(i)} - x_v^{(i)}\Big)^2}{\sum\limits_{i \in [k]} ||\boldsymbol{x}^{(i)}||^2} \leq 2\, \frac{\sum\limits_{u \sim v} \sum\limits_{i \in [k]} (x_u^{(i)} - x_v^{(i)})^2}{\sum\limits_{i \in [k]} ||\boldsymbol{x}^{(i)}||^2}$$

Our strategy will be to show that $\forall u \sim v$,

$$\left(\sum_{i\in[k]} x_u^{(i)} - x_v^{(i)}\right)^2 \leq 2\sum_{i\in[k]}\left(x_u^{(i)} - x_v^{(i)}\right)^2$$

which will be sufficient.

Note that since the $\boldsymbol{x}^{(i)}$ are disjointly supported, each vertex in the graph is part of the support of at most one of these vectors. If u and v are in the support of the same vector, i.e. for some $i \in [k]$, $u, v \in \mathrm{supp}(x^{(i)})$, the LHS and RHS are equal.

Similarly, if either u or v is not part of the support of any of the vectors, the LHS and RHS are equal.

If u and v are in the support of different vectors, say $\boldsymbol{x}^{(j)}$ and $\boldsymbol{x}^{(k)}$, then

$$\left(\sum_{i\in[k]} x_u^{(i)} - x_v^{(i)}\right)^2 = (x_u^{(j)} - x_v^{(k)})^2 \leq 2\left(\left(x_u^{(j)}\right)^2 + \left(x_v^{(k)}\right)^2\right) = 2\sum_{i\in[k]}\left(x_u^{(i)} - x_v^{(i)}\right)^2$$

as required. □

The lower bound of the higher-order Cheeger's inequality follows.

Corollary 4.2 (Higher-Order Cheeger's Lower Bound)**.** *Let $G = (V, E)$, and $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues of the normalized Laplacian, $\mathcal{L}$. We have*

$$\frac{\lambda_k}{2} \leq \phi_k(G)$$

Proof. We show that $\forall \boldsymbol{w} \in U$, $R_{\mathcal{L}}(\boldsymbol{w}) \leq 2\phi_k(G)$.

Let $\boldsymbol{w} \in U : \boldsymbol{w} = a_1\boldsymbol{x}^{(1)} + \cdots + a_k\boldsymbol{x}^{(k)}$. By Lemma 4.1,
$R_{\mathcal{L}}(\boldsymbol{w}) = R_{\mathcal{L}}\left(a_1\boldsymbol{x}^{(1)} + \cdots + a_k\boldsymbol{x}^{(k)}\right) \leq 2\max\left\{R_{\mathcal{L}}(a_1\boldsymbol{x}^{(1)}), \cdots, R_{\mathcal{L}}(a_k\boldsymbol{x}^{(k)})\right\}$

Note that the Rayleigh quotient is constant under scaling, so

$$\begin{aligned} R_{\mathcal{L}}(\boldsymbol{w}) &\leq 2\max\{R_{\mathcal{L}}(a_1\boldsymbol{x}^{(1)}), \cdots, R_{\mathcal{L}}(a_k\boldsymbol{x}^{(k)})\} \\ &= 2\max\{R_{\mathcal{L}}(\boldsymbol{x}^{(1)}), \cdots, R_{\mathcal{L}}(\boldsymbol{x}^{(k)})\} \\ &\leq 2\phi_k(G) \end{aligned}$$

where the last inequality follows since the $\boldsymbol{x}^{(i)}$ are indicator variables of sets with conductance $\leq \phi_k(G)$. Finally, since U is k-dimensional,

$$\lambda_k \leq \max_{\boldsymbol{x}\in U}\{R_{\mathcal{L}}(\boldsymbol{x})\} \leq 2\phi_k(G)$$

implying $\frac{\lambda_k}{2} \leq \phi_k(G)$, as desired. □

4.2. **The Spectral Embedding.** To prove the upper bound of the higher-order Cheeger's inequality, our goal will be to find k disjoint sets of vertices of low-conductance.

We will attempt to generalize the strategy we applied in Section 3. To prove the upper bound of Cheeger's inequality, the main idea was to map each of our vertices via $v \mapsto x_v$, where $\boldsymbol{x}$ is the second eigenvector of $\mathcal{L}$. We then showed that there exists a threshold, t, such that the cut $S = \{v : (x_v)^2 < t\}$ has sufficiently low conductance.

In this section, we will generalize the embedding $v \mapsto x_v$ to the k-dimensional spectral embedding. We will let $\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(k)}$ be an orthonormal eigenbasis of $\mathcal{L}$, where $\boldsymbol{x}^{(i)}$ is an eigenvector with eigenvalue λ_i. We will then embed the vertices via $v \mapsto (x_v^{(1)}, \cdots, x_v^{(k)})$, which we will define as the k-dimensional spectral embedding.

Definition 4.3 (k-Dimensional Spectral Embedding)**.** *Let G=(V,E) be a regular graph and let $\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^n$ be an orthonormal eigenbasis of $\mathcal{L}$, where $\boldsymbol{x}^i$ is an eigenvector of eigenvalue λ_i.*

The k-dimensional spectral embedding *is defined as*

$$F : V \to \mathbb{R}^k, F(v) \stackrel{\text{def}}{=} \left(\boldsymbol{x}_v^1, \boldsymbol{x}_v^2, \ldots, \boldsymbol{x}_v^k\right)$$

We will use $\hat{F}(v)$ to denote $F(v)$ scaled to unit length.

After applying the spectral embedding, we need to find a way to partition $\mathbb{R}^k$ in a way that will guarantee k relatively low-conductance sets. This is not as straight forward as it was in Section 3. With a 1-dimensional embedding, the logical choice of partition is to choose some threshold value, t, as we did to prove the upper bound. But in k dimensions, there is no obvious or canonical way to partition $\mathbb{R}^k$ into k regions.

To create well-separated sets, we need to have a notion of distance for the embedding. We could choose sets that are well-separated in a Euclidean sense. However, we instead opt for the *Radial Projection Distance* because we already know that F "perfectly" spreads out the mass in an isotropic sense. We will still make use of the Euclidean distance at times, but the *Radial Projection Distance* will be our primary distance metric.

Definition 4.4 (Radial Projection Distance)**.** *$\forall u, v \in V$, define the* Radial Projection Distance *between u and v as:*

$$\text{dist}(u, v) \stackrel{\text{def}}{=} \begin{cases} ||\hat{F}(u) - \hat{F}(v)|| & \text{if } F(u), F(v) \neq 0 \\ 0 & \text{if } F(u) = F(v) = 0 \\ \infty & \text{otherwise} \end{cases}$$

Further, if $S, T \subset V$, then

$$\text{dist}(v, S) \stackrel{\text{def}}{=} \min_{u \in S}\{\text{dist}(u, v)\} \qquad \text{and} \qquad \text{dist}(S, T) \stackrel{\text{def}}{=} \min_{u \in S, v \in T}\{\text{dist}(u, v)\}$$

We also define the *diameter* of a set as follows

Definition 4.5 (Radial Diameter)**.** $\forall S \subset V$, $\operatorname{diam}(S) \stackrel{\text{def}}{=} \max_{u,v \in S}\{\operatorname{dist}(u,v)\}$

Our partitioning strategy will be to first choose k sets that are radially well-separated, i.e. k sets such that if we project the sets onto the unit sphere, the k sets are well-separated from each other.

However, vertices whose corresponding vectors have small ℓ_2 mass can be problematic.

Definition 4.6 (Mass of a Vertex)**.** *$\forall v \in V$, we define the mass of vertex v as $m(v) \stackrel{\text{def}}{=} ||F(v)||^2$. For a set of vertices, $S \subset V$, we define $m(S) \stackrel{\text{def}}{=} \sum_{v \in S} m(v)$.*

Since vectors with small mass are close to the origin, even though the radial distance between these vectors and the vectors in other sets is large, the Euclidean distance may still be small.

So after getting k well-separated sets, we need to find a good way of setting a cutoff value such that all vertices with ℓ_2 mass below that value get removed. Let's first define the ℓ_2 mass of a vertex.

Note that removing vertices of small mass is analogous to the "rounding" we performed in Section 3. In fact, we already have a theorem that can do precisely this, Theorem 3.5, which will ensure that we can remove vertices with small ℓ_2 mass so that our k sets will have sufficiently small conductance.

At this point, it is important to note that we could have used any embedding $V \to \mathbb{R}^k$ and then performed the same procedure of finding well-separated sets and then removing small mass vertices. The spectral embedding is just one possible embedding.

However, there are a few good reasons for using the spectral embedding. First, we have already seen the connection between the eigenvalues of the Laplacian and conductance via the Rayleigh quotient, and the spectral embedding allows us to make use of this connection.

Second, our eventual goal will be to find k well-separated regions each with large ℓ_2 mass. The separation between the regions will give us an upper bound on the numerator of the Rayleigh quotient, and the large ℓ_2 mass will give us a lower bound on the denominator. If we hope to achieve this we need an embedding that spreads out the ℓ_2 mass in all directions. F is such an embedding as it satisfies the following isotropy property which we state without proof.

Lemma 4.7 (Isotropy Condition of F)**.** *For every unit vector $\boldsymbol{u} \in \mathbb{R}^k$,*

$$\sum_{v \in V} \langle \boldsymbol{u}, F(v) \rangle^2 = 1$$

In other words, no matter what direction, $\boldsymbol{u}$, we pick, the sum of the projections of all vectors $F(v)$ onto $\boldsymbol{u}$ is 1. The full proof can be found in [13].

Additionally, note that since the $\boldsymbol{x}^i$ are unit vectors, $m(V) = k$ because

$$m(V) = \sum_{v \in V} m(v) = \sum_{v \in V} \sum_{i \in [k]} (\boldsymbol{x}_v^i)^2 = \sum_{i \in [k]} \sum_{v \in V} (\boldsymbol{x}_v^i)^2 = \sum_{i \in [k]} ||\boldsymbol{x}^i||^2 = k$$

This makes the isotropy condition particularly ideal as it shows that we cannot "concentrate" the mass in fewer than k directions, i.e. F spreads the vertices "perfectly" in k dimensions in an isotropic sense.

As a result of the isotropy condition, a key fact about F is that sets of small radial diameter have small mass. Sets of large mass cannot have small radial diameter as otherwise they would "concentrate" too much of the mass in one direction. We formalize this concept as follows.

Definition 4.8 (Spreading Property of a Function). *For any map $f : V \to \mathbb{R}^k$, f is (R, η)-spreading with respect to G if for any $S \subset V$ with* $\text{diam}(S) \leq R$*, we have $m(S) \leq \eta$.*

So for a function that is well-spreading, for every $R > 0$, the function will have a correspondingly low value of η. Indeed, our spectral embedding, F, satisfies this property.

Lemma 4.9 (Spreading Property of the Spectral Embedding). *Let $F : V \to \mathbb{R}^k$ be the k-dimensional spectral embedding. For every $R > 0$, F is $\left(R, \frac{1}{1-R^2}\right)$-spreading with respect to G. In other words, $\forall S \in \mathbb{R}^k$ such that* $\text{diam}(S) \leq R$, $m(S) \leq \frac{1}{1-R^2}$.

Proof. The main idea for the proof is to use the isotropy property of the spectral embedding along with the law of cosines.

Let $\boldsymbol{s} \in \mathbb{R}^k$, and denote $\hat{\boldsymbol{s}}$ to be $\boldsymbol{s}$ scaled to unit-length. We have

$$\sum_{v \in S} (\langle F(v), \boldsymbol{s} \rangle)^2 \leq \sum_{v \in V} (\langle F(v), \boldsymbol{s} \rangle)^2 = ||\boldsymbol{s}||^2 \sum_{v \in V} (\langle F(v), \hat{\boldsymbol{s}} \rangle)^2 = ||\boldsymbol{s}||^2 .$$

Additionally,

$$\begin{aligned}
(\langle F(v), \boldsymbol{s} \rangle)^2 &= m(v) ||\boldsymbol{s}||^2 (\langle \hat{F}(v), \hat{\boldsymbol{s}} \rangle)^2 \\
&= m(v) ||\boldsymbol{s}||^2 (\langle \hat{F}(v), \hat{\boldsymbol{s}} \rangle)^2 \\
&= m(v) ||\boldsymbol{s}||^2 \cos^2 \theta
\end{aligned}$$

where θ is the angle formed between $\hat{F}(v)$ and $\hat{\boldsymbol{s}}$.

By the law of cosines,

$$\begin{aligned}
\cos \theta &= \frac{||\hat{F}(v)||^2 + ||\hat{\boldsymbol{s}}||^2 - ||\hat{F}(v) - \hat{\boldsymbol{s}}||^2}{2 ||\hat{F}(v)|| \cdot ||\hat{\boldsymbol{s}}||} \\
&= \frac{1^2 + 1^2 - ||\hat{F}(v) - \hat{\boldsymbol{s}}||^2}{2}
\end{aligned}$$

$$= 1 - \frac{1}{2}||\hat{F}(v) - \hat{\boldsymbol{s}}||^2.$$

So

$$\begin{aligned}
(\langle F(v), \boldsymbol{s}\rangle)^2 &= m(v)||\boldsymbol{s}||^2 \cos^2\theta \\
&= m(v)||\boldsymbol{s}||^2 \left(1 - \frac{1}{2}||\hat{F}(v) - \hat{\boldsymbol{s}}||^2\right)^2 \\
&\geq m(v)||\boldsymbol{s}||^2 \left(1 - \frac{1}{2}R^2\right)^2 \\
&\geq m(v)||\boldsymbol{s}||^2 (1 - R^2)
\end{aligned}$$

Now consider the specific case where $\boldsymbol{s} = F(v)$ for some $v \in S$. In this case, we have $||\hat{F}(v) - \hat{\boldsymbol{s}}|| \leq R$, and it follows that

$$\begin{aligned}
||\boldsymbol{s}||^2 &\geq \sum_{v \in S} (\langle F(v), \boldsymbol{s}\rangle)^2 \\
||\boldsymbol{s}||^2 &\geq \sum_{v \in S} m(v)||\boldsymbol{s}||^2 (1 - R^2) \\
1 &\geq \sum_{v \in S} m(v)(1 - R^2) \\
\frac{1}{1 - R^2} &\geq \sum_{v \in S} m(v) = m(S)
\end{aligned}$$

□

Our eventual goal is to bound the conductance by λ_k. One way to do this is to bound the conductance via

$$\text{avg}\{R_{\mathcal{L}}(\boldsymbol{x}^1), R_{\mathcal{L}}(\boldsymbol{x}^2), \ldots, R_{\mathcal{L}}(\boldsymbol{x}^k)\} = \text{avg}\{\lambda_1, \lambda_2, \ldots, \lambda_k\} \leq \lambda_k$$

where avg denotes the arithmetic mean. The purpose of this is that

$$\begin{aligned}
\text{avg}\{R_{\mathcal{L}}(\boldsymbol{x}^1), R_{\mathcal{L}}(\boldsymbol{x}^2), \ldots, R_{\mathcal{L}}(\boldsymbol{x}^k)\} &= \frac{\sum\limits_{i \in [k]} \sum\limits_{u \sim v} ||x_u^i - x_v^i||^2}{k} \\
&= \frac{\sum\limits_{u \sim v} \sum\limits_{i \in [k]} ||x_u^i - x_v^i||^2}{m(v)} \\
&= \frac{\sum\limits_{u \sim v} ||F(u) - F(v)||^2}{m(V)}.
\end{aligned}$$

This expression strongly resembles the Rayleigh quotient of the Laplacian. In fact, we define this quantity as the *Rayleigh quotient of F*.

Definition 4.10. *Let F be the k-dimensional spectral embedding.*

$$R_{\mathcal{L}}(F) \stackrel{\text{def}}{=} \frac{\sum\limits_{u \sim v} ||F(u) - F(v)||^2}{m(V)} = \text{avg}\{R_{\mathcal{L}}(\boldsymbol{x}^1), R_{\mathcal{L}}(\boldsymbol{x}^2), \ldots, R_{\mathcal{L}}(\boldsymbol{x}^k)\}$$

We are now ready to provide an outline for our algorithm.

4.3. **Upper Bound Algorithm.**

(1) Embed the vertices into $\mathbb{R}^k$ using the spectral embedding
(2) Find k well-separated sets, i.e. sets with large mass and large pairwise radial distance, as follows:
   (a) Partition $\mathbb{R}^k$ into parallel cubes. Then shrink all cubes slightly so that the cubes still cover a mass of at least $k - \frac{1}{4}$ (the total mass is k), and are now separated from each other by a Euclidean distance of $\Omega(k^{-3})$.
   (b) To ensure our sets are radially separated by $\Omega(k^{-3})$, we can partition our vertices based on which cube $\hat{F}(v)$ lands in (rather than $F(v)$).
   (c) Since the cubes have bounded radial diameter, by Lemma 4.9, the mass of each cube is at most $1 - \frac{1}{4k}$
   (d) To conslidate our cube sets into sets of large mass, combine any two sets of mass $< \frac{1}{2}$ and repeat this until we have k sets of mass $\geq \frac{1}{2}$ radially separated by $\Omega(k^{-3})$.
(3) From the k well-separated sets, generate k disjointly supported vectors, $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(k)}$ as follows: For each set A_i, let $\boldsymbol{y}^{(i)} = \tau(v) \cdot ||F(v)||$, where $\tau(v) = 1$ for $v \in A_i$, and smoothly decreases to 0 as $\mathrm{dist}(v, A_i)$ increases.
(4) We have $R_{\mathcal{L}}(\boldsymbol{y}^{(i)}) \leq O(k^7) \cdot \max\{R_{\mathcal{L}}(F)\} \leq O(k^7) \cdot \max\{R_{\mathcal{L}}(\boldsymbol{x}^{(k)})\}$. For each $\boldsymbol{y}^{(i)}$ use Theorem 3.5 to find $S \subset \mathrm{supp}(\boldsymbol{y}^{(i)})$ such that
$$\phi(S) \leq \sqrt{2R_{\mathcal{L}}(y)} = O(k^{3.5}) \cdot \sqrt{\lambda_k}$$

4.4. **Rigorous Proof.** Our first goal is to find k well-separated sets. To do this, we first show that we can find t well-separated sets such that, in total, they capture most of the mass, while separately, each has a small amount of mass.

Lemma 4.11 (Well-Separated Sets)**.** *We can find t sets, $A_1, \ldots, A_t$ such that*

(1) $\forall i, j \in [t], i \neq j$, $\mathrm{dist}(A_i, A_j) \geq \Omega(k^{-3})$
(2) $\sum_{i \in [k]} m(A_i) \geq k - \frac{1}{4}$
(3) $\forall i \in [k]$, $m(A_i) \leq 1 + \frac{1}{4k}$

Proof. We will partition $\mathbb{R}^k$ as follows. Tile $\mathbb{R}^k$ with cubes of the form $[x_1, x_1 + L) \times [x_2, x_2 + L) \times \cdots \times [x_k, x_k + L)$, where $x_i \in \mathbb{R}$ and $L = \frac{1}{4k}$ is the side-length of the cubes. To induce separation between the cubes, dilate each cube about its center by a factor of $1 - \frac{1}{8k^2}$ to obtain its *core*. This assures that the Euclidean distance between any two cube cores is at least

$$L \cdot \frac{1}{8k^2} = \frac{1}{4k} \cdot \frac{1}{8k^2} = \frac{1}{32k^3} = \Omega(k^{-3})\,.$$

But we want the *radial distance* to be $\Omega(k^{-3})$, not just the Euclidean distance. This is not a problem as we can project all vectors onto the unit sphere via $F(v) \to \hat{F}(v)$ and then partition each vertex v based on which cube $\hat{F}(v)$ lands in. If $\hat{F}(v)$

does not lie in the core of any cube, then v will not be placed into any set. This will guarantee that the vertices are partitioned into sets with pairwise radial distances of $\Omega(k^{-3})$.

Next, we need to prove there is a way to place our parallel cubes such that they cover most ($\geq k - \frac{1}{4}$) of the mass. We prove this using the probabilistic method and show that for a random placement of the cubes, on expectation, the cubes will capture a mass of $\geq k - \frac{1}{4}$.

To formalize this, we start with a default cube placement, where all cubes are of the form $[x_1, x_1 + L) \times [x_2, x_2 + L) \times \cdots \times [x_k, x_k + L)$; here x_i are multiples of L.

Then we will shift the cubes by a random vector, $\boldsymbol{w}$. Note that the coordinates of $\boldsymbol{w}$ can be considered $\pmod{L}$ since a shift of length L in any of the k directions does not affect the tiling. So we let $\boldsymbol{w} = (w_1, \ldots, w_k)$, where for every $i \in [k]$, w_i is chosen from $[0, L)$ uniformly at random.

Let C be the union of all cube cores. By linearity of expectation, after we perform the shift,

$$\mathbb{E}\Bigg[\sum_{\substack{v \in V: \\ F(v) \in C}} m(v)\Bigg] = \sum_{v \in V} \mathbb{P}[F(v) \in C] \cdot m(v)\,.$$

We claim that for every $v \in V$, $\mathbb{P}[F(v) \in C] > 1 - \frac{1}{4k}$

Claim 4.11.1. *$\forall \boldsymbol{x} \in \mathbb{R}^k$, $\mathbb{P}[\boldsymbol{x} \in C] > 1 - \frac{1}{4k}$*

Proof. Let $I = \left[L \cdot \frac{1}{16k^2}, L \cdot (1 - \frac{1}{16k^2})\right)$. $\mathbb{P}[\boldsymbol{x} \in C]$ after shifting the cubes by $\mathbf{w}$ is the same as $\mathbb{P}[\boldsymbol{x} \in C]$ after shifting $\boldsymbol{x}$ by $-\boldsymbol{w}$. Mod L, after shifting by $-\boldsymbol{w}$, each coordinate of $\boldsymbol{x}$ will land uniformly at random in $[0, L)$.

Let $\bar{x_i} = x_i \pmod{L}$. We then have for all $i \in [k]$:

$$\mathbb{P}[\bar{x_i} \notin I] = \frac{1}{8k^2}$$

By a union bound, $\mathbb{P}\left(\exists\, i \in [k] : \bar{x_i} \notin I\right) \leq \frac{k}{8k^2} = \frac{1}{8k}$. Consequently, $\mathbb{P}[\boldsymbol{x} \in C] \geq 1 - \frac{1}{8k} > 1 - \frac{1}{4k}$, proving the claim. $\square$

By Claim 4.11.1, $\forall v \in V$, $\mathbb{P}[\hat{F}(v) \in C] > 1 - \frac{1}{4k}$, so

$$\begin{aligned}\mathbb{E}\left[\sum_{\hat{F}(v) \in C} m(v)\right] &= \sum_{v \in V} \mathbb{P}\left[\hat{F}(v) \in C\right] \cdot m(v) \\ &> \sum_{v \in V} \left(1 - \frac{1}{4k}\right) \cdot m(v)\end{aligned}$$

$$= \left(1 - \frac{1}{4k}\right) \cdot \sum_{v \in V} m(v)$$
$$= \left(1 - \frac{1}{4k}\right) \cdot k$$
$$= k - \frac{1}{4}$$

Suppose $S_1, \cdots, S_t$ are the cube cores that contain at least one vector $\hat{F}(v)$. For all $i \in [t]$, let $A_i = \{v \in V : \hat{F}(v) \in S_i\}$. In particular, every A_i satisfies:

(1) $\forall v \in V$, $v \in A_i$ if $\hat{F}(v) \in A_i$
(2) $\forall i \in [t]$, A_i is nonempty
(3) $\forall i, j \in [t], \mathrm{dist}(A_i, A_j) = \Omega(k^{-3})$

The final thing we need to show is that the sets A_i have mass $\leq 1 + \frac{1}{4k}$. Note that the cubes have diameter $\leq \frac{\sqrt{k}}{4k} = \frac{1}{4\sqrt{k}}$, and since the A_i are subsets of their respective cubes, for every $i \in [t]$, $\mathrm{diam}(A_i) \leq \frac{1}{4\sqrt{k}}$. Note that for every $i \in [t]$, by Lemma 4.9, $m(A_i) \leq \frac{1}{1-\frac{1}{16k}} = \frac{16k}{16k-1} = 1 + \frac{1}{16k-1} \leq 1 + \frac{1}{4k}$

We conclude that A_i satisfy the conditions of the lemma, as desired. ☐

Because we can produce t well-separated sets such that they capture most of the mass, as guaranteed by (2) in Lemma 4.11, and all have relatively small mass, as guaranteed by (3) in Lemma 4.11, we can consolidate them into k well-separated sets, each of large mass. We will provide an algorithm for this procedure in the next lemma.

Lemma 4.12 (k well-separated sets)**.** *We can pick k sets $S_1, \ldots, S_k \subset V$, such that*

(1) $\forall i, j \in [t], i \neq j$, $\mathrm{dist}(S_i, S_j) \geq \Omega(k^{-3})$
(2) $\forall i \in [k], m(S_i) \geq \frac{1}{2}$

Proof. Pick t sets, $A_1, \ldots, A_t$ as in Lemma 4.11. We consolidate these sets to produce at least k sets that are still well-separated sets and that each have mass $\geq \frac{1}{2}$ as follows:

> As long as there are ≥ 2 sets with mass $< \frac{1}{2}$ combine them. Repeat this until at most one set of mass $< \frac{1}{2}$ remains.

First, note that when the algorithm terminates, the sets produced will still be radially separated from each other by a distance of $\Omega(k^{-3})$. We then claim that, when this algorithm terminates, there will be at least k sets of mass at least $\frac{1}{2}$.

At each step of the algorithm, when two sets are combined, their combined mass is < 1 since their individual masses are $< \frac{1}{2}$. So when the algorithm terminates, each set still has mass $\leq 1 + \frac{1}{4k}$.

Now suppose for the sake of contradiction that the algorithm terminates with $\leq k-1$ sets of mass $\geq \frac{1}{2}$. The combined mass of these $k-1$ sets is at most

$$(k-1)\left(1+\frac{1}{4k}\right) = k-1+\frac{1}{4}-\frac{1}{4k} < k-\frac{3}{4}$$

Besides these $k-1$ sets, there is at most one additional set whose mass is $< \frac{1}{2}$, so the total mass of all sets is $< k-\frac{1}{4}$, which is a contradiction. □

Now that we have k well-separated sets, our plan is to remove vertices with small mass using Fiedler's Algorithm. To show that this yields k disjoint and sufficiently low-conductance cuts, we need to form disjointly supported vectors, $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^k$, from our sets and then use Theorem 3.5.

Logically, we would like to assign entries via

$$y_v^{(i)} = \begin{cases} ||F(v)|| & \text{if } v \in A_i \\ 0 & \text{if } v \notin A_i \end{cases}$$

Our goal is to bound $R_{\mathcal{L}}(\boldsymbol{y}^i)$ by a factor of $R_{\mathcal{L}}(F)$, where we recall that $R_{\mathcal{L}}(F) = \text{avg}\{R_{\mathcal{L}}(x_i)\}$. For every $i \in [k]$, we have

$$R_{\mathcal{L}}(\boldsymbol{y}^i) = \frac{\sum_{u\sim v} \left(y_u^i - y_v^{(i)}\right)^2}{m(\boldsymbol{y}^i)}$$

Since the mass of the well-separated sets is bounded below, it suffices to show that $\sum_{u\sim v} \left(y_u^i - y_v^{(i)}\right)^2 = O\left(R_{\mathcal{L}}(F)\right)$.

Note that

$$R_{\mathcal{L}}(F) = \frac{\sum_{u\sim v} ||F(u)-F(v)||^2}{d\sum_{v} m(v)}$$

The denominator is dk since all vertices are embedded into unit vectors by design, so it suffices to show that

$$\sum_{u\sim v} \left(y_u^i - y_v^{(i)}\right)^2 \leq \sum_{u\sim v} ||F(u)-F(v)||^2$$

We use casework to determine whether this inequality is satisfied.

(1) $u, v \in A_i$. In this case,

$$\left(y_u^{(i)} - y_v^{(i)}\right)^2 = \left(||F(u)|| - ||F(v)||\right)^2 \leq ||F(u)-F(v)||^2$$

where the last inequality follows by Cauchy-Schwartz.

(2) $u, v \notin A_i$. In this case, $\left(y_u^{(i)} - y_v^{(i)}\right)^2 = 0 \leq ||F(u) - F(v)||^2$.

(3) $u \in A_i$, $v \notin A_i$, or vice versa. WLOG, suppose $u \in A_i$, $v \notin A_i$. In this case, $\left(y_u^{(i)} - y_v^{(i)}\right)^2 = ||F(u)||^2$.

Unfortunately, in case (3), even if $||F(u) - F(v)||$ is large, $||F(u)||$ can be arbitrarily large. In particular, it is possible that u and v have large mass, but that the radial distance between $F(u)$ and $F(v)$ may be very small.

However, if the radial distance between $F(u)$ and $F(v)$ is small, i.e. as long as $\text{dist}(u, v)$ is sufficiently large, we should suspect that $||F(u)||$ cannot be much larger than $||F(u) - F(v)||$. This is the case, and we will show that if we fix $\text{dist}(u, v)$, $||F(u)||$ is bounded by a factor of $(||F(u)|| - ||F(v)||)^2$. This follows through a clever use of the triangle inequality.

Lemma 4.13. $||F(u)|| \cdot \text{dist}(u, v) \leq 2 \cdot ||F(u) - F(v)||$

Proof. We have $||F(u)|| \cdot \text{dist}(u, v) = ||F(u) - F(v)|| \cdot \frac{||F(u)||}{||F(v)||}$. If we let $F'(v)$ be the vector $F(v)$ scaled to the same magnitude as $F(u)$, i.e.

$$F'(v) \stackrel{\text{def}}{=} F(v) \cdot \frac{||F(u)||}{||F(v)||}$$

then the LHS of the statement of the lemma is the Euclidean distance between $F(u)$ and $F'(v)$. We can use the triangle inequality with the vector $F(v)$. This gives us

$$||F(u) - F'(v)|| \leq ||F(u) - F(v)|| + ||F(v) - F'(v)|| \,.$$

$||F(v) - F'(v)|| \leq ||F(u) - F(v)||$ since $F(u)$ and $F'(v)$ have the same magnitude and $F'(v)$ and $F(v)$ are scalar multiples of each other. In particular, since $F(v)$ and $F'(v)$ "point in the same direction",

$$||F(v) - F'(v)|| = \Big| ||F(v)|| - ||F(u)|| \Big| \leq ||F(v) - F(u)||$$

Therefore,

$$||F(u) - F(v)|| + ||F(v) - F'(v)|| \leq 2||F(u) - F(v)||$$

and

$$||F(u)|| \text{ dist}(u, v) \leq 2||F(u) - F(v)||$$ □

Let $A = A_1 \cup A_2 \cup \ldots \cup A_k$. Since the A_i are well-separated, $\text{dist}(u, v)$ cannot get arbitrarily small as long as $u, v \in A$, so Lemma 4.13 resolves case (3) for when $u, v \in A$.

However, this still does not resolve the case for when $u \in A_i$, but $v \notin A$ or vice versa. In this case, no matter how close $\hat{F}(u)$ and $\hat{F}(v)$ are to each other, $||F(u)||$ can still be arbitrarily large, so we cannot bound $||F(u)||^2$ by a constant factor of $||F(u) - F(v)||^2$.

To resolve this issue, we need to find a way to deal with vertices that are not contained in any of the well-separated sets.

Rather than assigning $y_v^{(i)} = ||F(v)||$ if $v \in A_i$ and $y_v^{(i)} = 0$ if $v \notin A_i$, we can let $y_v^{(i)}$ decrease linearly toward 0 as $\mathrm{dist}(v, A_i)$ decreases. This will ensure $\left(y_u^{(i)} - y_v^{(i)}\right)^2$ is bounded by a factor of $\mathrm{dist}(v, A_i) \cdot ||F(u)||$, which by Lemma 4.13 means it is also bounded by a factor of $||F(u) - F(v)||^2$. This is given via the *Localization Lemma.*

Lemma 4.14 (Localization Lemma)**.** *Let $A_1, \ldots, A_k$ be k sets such that $m(A_i) \geq \frac{1}{2}$ for all $i \in [k]$, and $\mathrm{dist}(u, v) \geq \delta$ for all u, v in different sets. There exist k disjointly supported vectors, $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^k$, such that for every $i \in [k]$:*

$$R_{\mathcal{L}}(\boldsymbol{y}^i) = O\left(\frac{k}{\delta^2}\right) \cdot R_{\mathcal{L}}(F)$$

Note that Lemma 4.14 gives the main result for this section. By Lemma 4.12, we have k sets where for all u, v in different sets, $\mathrm{dist}(u, v) = \Omega(k^{-3})$. Lemma 4.14 then guarantees we can find k vectors, $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^k$, such that $R_{\mathcal{L}}(\boldsymbol{y}^i) \leq O\left(\frac{k}{k^{-6}}\right) \cdot R_{\mathcal{L}}(F)$. $R_{\mathcal{L}}(F) \leq \max\{R_{\mathcal{L}}(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^k)\} = \lambda_k$, so $R_{\mathcal{L}}(\boldsymbol{y}^i) \leq O(k^7)\, \lambda_k$. From there, Theorem 3.5 provides the desired result.

Proof of Lemma 4.14. We want $y_v^{(i)}$ to decrease linearly as $\mathrm{dist}(v, A_i)$ increases. We also need to make sure that for every $v \in V$, $y_v^{(i)}$ is nonzero for at most one value of i. To ensure this, we let $y_v^{(i)}$ decrease linearly from $||F(v)||$ to 0 as $\mathrm{dist}(v, A_i)$ increases from 0 to $\frac{\delta}{2}$. Since every pair of sets is separated by δ, no vertex can be within $\frac{\delta}{2}$ of more than one set.

So define $y_v^{(i)} = g_i(v) \cdot ||F(v)||$, where $g_i(v)$ is defined as follows.

$$g_i(v) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } v \in A_i \\ 0 & \text{if } \mathrm{dist}(v, A_i) \geq \frac{\delta}{2} \\ 1 - \frac{2}{\delta}\,\mathrm{dist}(v, A_i) & \text{otherwise} \end{cases}$$

We now show that the $\boldsymbol{y}^{(i)}$ have sufficiently low Rayleigh quotient, i.e. that $R_{\mathcal{L}}\left(\boldsymbol{y}^{(i)}\right) \leq O(\frac{1}{\delta^2}) \cdot R_{\mathcal{L}}(F)$. We have

$$R_{\mathcal{L}}(\boldsymbol{y}^i) = \frac{\sum_{u \sim v} \left(y_u^{(i)} - y_v^{(i)}\right)^2}{m(\boldsymbol{y}^{(i)})}$$

$$R_{\mathcal{L}}(F) = \frac{\sum_{u \sim v} ||F(u) - F(v)||^2}{d \sum_{v} m(v)}$$

The denominator of $R_{\mathcal{L}}(\boldsymbol{y}^{(i)})$ is bounded below since $m(\boldsymbol{y}^{(i)}) \geq m(A_i) \geq \frac{1}{2}$. The denominator of $R_{\mathcal{L}}(F)$ is dk, so it suffices to show

$$\sum_{u\sim v} \left(y_u^{(i)} - y_v^{(i)}\right)^2 \leq O\left(\frac{1}{\delta}\right) \sum_{u\sim v} ||F(u) - F(v)||^2$$

We show that for all $u, v \in V$, $y_u^{(i)} - y_v^{(i)} \leq O(\frac{1}{\delta}) \cdot ||F(u) - F(v)||$.

Case 1: $u, v \in A_i$. In this case, $y_u^{(i)} - y_v^{(i)} = ||F(u)|| - ||F(v)|| \leq ||F(u) - F(v)||$

Case 2: $u \notin A_i$, or $v \notin A_i$. WLOG, suppose $v \notin A_i$ and $\text{dist}(u, A_i) \geq \text{dist}(v, A_i)$. First, we claim that we can assume $\text{dist}(v, A_i) \leq \frac{\delta}{2}$ also WLOG.

Claim 4.14.1. *It suffices to show the result for when* $\text{dist}(v, A_i) \leq \frac{\delta}{2}$

Proof. If $\text{dist}(v, A_i) \geq \frac{\delta}{2}$, we may pick $F(v)'$ such that $\text{dist}(v', A_i) = \frac{\delta}{2}$ and such that $F(v)'$ is closer to $F(u)$ than $F(v)$ is, i.e. $||F(v)' - F(u)|| \leq ||F(v) - F(u)||$. Replacing $F(v)$ by $F(v)'$ in the inequality $y_u^{(i)} - y_v^{(i)} \leq ||F(u) - F(v)||$ does not affect the LHS but decreases the RHS, implying that if the inequality holds for $F(v)'$, it holds for $F(v)$. □

Now, to bound $\big|g(u) \cdot ||F(u)|| - g(v) \cdot ||F(v)||\big|$, we use the triangle inequality with $g(u) \cdot ||F(v)||$. This is because we can bound $\big|g(u) \cdot ||F(u)|| - g(u) \cdot ||F(v)||\big|$ by $||F(u) - F(v)||$ and bound $||F(v)|| - g(v) \cdot ||F(v)||$ using Lemma 4.13. We have

$$\begin{aligned}
y_u^{(i)} - y_v^{(i)} &= \big|g(u) \cdot ||F(u)|| - g(v) \cdot ||F(v)||\big| \\
&\leq \big|g(u) \cdot ||F(u)|| - g(u) \cdot ||F(v)||\big| + ||F(v)|| - g(v) \cdot ||F(v)|| \\
&= g(u) \cdot \big|||F(u)|| - ||F(v)||\big| + \big(1 - g(v)\big) \cdot ||F(v)|| \\
&\leq g(u) \cdot ||F(u) - F(v)|| + \big(1 - g(v)\big) \cdot ||F(v)|| \\
&\leq ||F(u) - F(v)|| + \big(1 - g(v)\big) \cdot ||F(v)|| \\
&= ||F(u) - F(v)|| + \frac{2}{\delta} \cdot \text{dist}(u, v) \cdot ||F(v)||
\end{aligned}$$

By Lemma 4.13,

$$\frac{2}{\delta} \cdot \text{dist}(u, v)||F(v)|| \leq \frac{4}{\delta} \cdot ||F(u) - F(v)||$$

giving us

$$\begin{aligned}
y_u^{(i)} - y_v^{(i)} &\leq ||F(u) - F(v)|| + \frac{2}{\delta} \cdot \text{dist}(u, v)||F(v)|| \\
&\leq \left(1 + \frac{\delta}{4}\right) \cdot ||F(u) - F(v)|| \\
&= O\left(\frac{1}{\delta}\right) \cdot ||F(u) - F(v)||
\end{aligned}$$

It follows that $R_{\mathcal{L}}(\boldsymbol{y}^i) \leq O\left(\frac{1}{\delta^2}\right) \cdot R_{\mathcal{L}}(F)$ as required. □

The higher-order Cheeger inequality follows from this result.

Theorem 4.15 (Higher-Order Cheeger's Inequality $O(k^{3.5})$ bound)**.** *Let $G = (V, E)$ be a d-regular graph, and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be eigenvalues of $\mathcal{L}$. Then the following holds.*

$$\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^{3.5})\sqrt{\lambda_k}$$

Proof. We proved the lower bound side in Corollary 4.2.

For the upper bound side, first find k well-separated sets via Lemma 4.12. The well-separated sets satisfy the conditions in Lemma 4.14 with $\delta = \Omega(\frac{1}{k^3})$. Thus, we can use the well-separated sets to find k disjointly supported vectors

$$\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(k)}$$

such that for all $i \in [k]$, we have

$$R_{\mathcal{L}}(\boldsymbol{y}^{(i)}) \leq O(k^7) \cdot R_{\mathcal{L}}(F)\,.$$

Note that

$$\begin{aligned} R_{\mathcal{L}}(F) &= \operatorname{avg}\left\{R_{\mathcal{L}}\big(\boldsymbol{x}^{(1)}\big), R_{\mathcal{L}}\big(\boldsymbol{x}^{(2)}\big), \ldots, R_{\mathcal{L}}\big(\boldsymbol{x}^{(k)}\big)\right\} \\ &= \operatorname{avg}\{\lambda_1, \lambda_2, \ldots, \lambda_k\} \\ &\leq \lambda_k \end{aligned}$$

So $R_{\mathcal{L}}\big(\boldsymbol{y}^{(i)}\big) \leq O(k^7) \cdot \lambda_k$. Finally, by Theorem 3.5, for every $i \in [k]$, we can find a cut S_i such that $S_i \subset \operatorname{supp}(\boldsymbol{y}^{(i)})$ and $\phi(S_i) \leq \sqrt{2 \cdot R_{\mathcal{L}}(\boldsymbol{y}^{(i)})} = O(k^{3.5})\sqrt{\lambda_k}$. This proves the upper bound side and the theorem. □

5. Higher-Order Cheeger, Upper Bound $O(k^3)$

In this section, we provide an overview of how to reduce the upper bound from $O(k^{3.5})\sqrt{\lambda_k}$ to $O(k^3)\sqrt{\lambda_k}$. The details of this improvement can be found in [3].

The success of the cube partition was 2-fold.

(1) The cubes had small diameter, ensuring that they also have small mass and allowing us to consolidate our cube cores to create at least k sets of sufficiently large mass.
(2) After shrinking the cubes sufficiently to create separation, on expectation with respect to the choice of vector shift, $\boldsymbol{w}$, the cube cores still captured the majority of the mass.

We would like to abstract these properties. The first step in abstracting the cube partition is by noting that we proved that there exists a "desirable" cube partition using the probabilistic method via a random vector shift. This can be formalized via a random partition.

Definition 5.1 (Random Partition)**.** *A* random partition*, $\mathcal{P}$, of a metric space X is a probability space, where the sample space is a set of partitions.*

For a partition, $\mathcal{P}$ of X, for all $x, y \in X$, define $x \sim_P y$ if x and y are partitioned into the same set, and $x \not\sim_P y$ otherwise. Similarly, for sets $S, T \subset X$, define $S \sim_P T$ if for all $x, y \in S \cup T$, $x \sim_P y$, and define $S \not\sim_P T$ otherwise.

To satisfy property (1), we want to partition $\mathbb{R}^k$ into low diameter sets. So let us define a partition as *R-bounded* if all sets in the partition have diameter $\leq R$.

Definition 5.2. *A partition, $\mathcal{P} = S_1 \cup S_2 \cup \cdots \cup S_m$, is R-*bounded *if for every $i \in [m]$, $\operatorname{diam}(S_i) \leq m$. A random partition $\mathcal{P}$ is R-bounded if this property holds with probability 1.*

Next, let's generalize property (2). We want a random partition such that after shrinking the sets in our partition via a dilation, on expectation, the majority of the mass remains within the sets.

One way to quantify this is by determining how likely it is that for a given vector, $\boldsymbol{x} \in S_i$, after dilating S_i by a certain factor, $\boldsymbol{x}$ remains in S_i.

But how can we determine whether $\boldsymbol{x}$ stays inside S_i. To do this, we can consider whether vectors "near" $\boldsymbol{x}$ also lie inside S_i. If they do, then $\boldsymbol{x}$ is sufficiently far from the boundary of S_i so that after dilating S_i, $\boldsymbol{x}$ will remain inside S_i. On the other hand, if there are vectors "near" $\boldsymbol{x}$ that get partitioned into different sets, $\boldsymbol{x}$ is near the boundary of S_i and will not remain inside S_i following the dilation.

We can quantify how efficient our random partition is at achieving both of these properties as follows.

Definition 5.3 (Padded Partition)**.** *A random partition, $\mathcal{P}$, of X is (R, α, δ)-*padded *if $\mathcal{P}$ is R-bounded and for every $\boldsymbol{x} \in X$, $\mathbb{P}\big[B_{R/\alpha}(x) \sim \{x\}\big] \geq \delta$.*

The R-bounded condition ensures the sets within the partition have low diameter and thus low mass.

The condition that for every $\boldsymbol{x} \in X$, $\mathbb{P}[B_{R/\alpha}(x) \sim \{x\}] \geq \delta$ ensures that for every vector, $\boldsymbol{x} \in \mathbb{R}^k$, vectors "near" $\boldsymbol{x}$ in a Euclidean sense will "usually" be partitioned into the same set as "$\boldsymbol{x}$". Therefore, if $\boldsymbol{x} \in S_i$, $\boldsymbol{x}$ will "usually" be far enough from the boundary of S_i to ensure that it remains inside S_i following a dilation.

The *Padded Partition Theorem* ensures the existence of such a partition, and a proof is provided in [23].

Theorem 5.4 (Padded Partition Theorem)**.** *For every Euclidean metric space $X \in \mathbb{R}^k$, for every $\Delta > 0$, there exists a random partition of X that is $\big(\Delta, O\big(\frac{k}{\delta}\big), 1-\delta\big)$-padded*

Let's compare how well the cube partition provided in *Section* 4 satisfies property (2) compared to the partition guaranteed by *Theorem* 5.4.

Lemma 5.5. *The cube partition provided in Section 4 is a* $\left(\frac{1}{2\sqrt{k}}, 32 \cdot k^2\sqrt{k}, 1-\frac{1}{8k}\right)$*-padded partition*

Proof. The cubes have side length $\frac{1}{4k}$. Since we dilated each cube by a factor of $1-\frac{1}{8k^2}$ to get the cube cores, we have that for every $\boldsymbol{x} \in \mathbb{R}^k$, $B_{1/64k^3}(\boldsymbol{x}) \sim \boldsymbol{x}$ iff $\boldsymbol{x} \in C$, where C is the union of all cube cores.

It remains to show that $\mathbb{P}[B_{1/64k^3}(\boldsymbol{x}) \sim \boldsymbol{x}] = \mathbb{P}[\boldsymbol{x} \in C] \geq 1-\frac{1}{8k}$. Since we dilated the cubes by a factor of $1-\frac{1}{8k^2}$, we have $\mathbb{P}[\boldsymbol{x} \in C] = (1-\frac{1}{8k^2})^k$. One way to prove that this is $\geq 1-\frac{1}{8k}$ is by using the union bound.

Let $\boldsymbol{x} = (x_1, \cdots, x_k)$. For every $S \in \mathbb{R}^k$, let $\pi_i : \mathbb{R}^k \to \mathbb{R}$ denote the projection of the i-th coordinate of S. If we let S be the cube containing $\boldsymbol{x}$, we have that for every $i \in [k]$, $\mathbb{P}[x_i \notin \pi_i(S)] = \frac{1}{8k^2}$. So $\sum\limits_{i\in[k]} \mathbb{P}[x_i \notin \pi_i(S)] \leq \frac{k}{8k^2} = \frac{1}{8k}$, implying that

$$\mathbb{P}[\boldsymbol{x} \in C] \geq 1 - \frac{1}{8k}$$

Therefore, we have $\mathbb{P}[B_{1/64k^3}(\boldsymbol{x}) \sim \boldsymbol{x}] \geq \frac{1}{8k}$, as desired. □

Note that Theorem 5.4 guarantees a $\left(\frac{1}{2\sqrt{k}}, O(k^2), 1-\frac{1}{8k}\right)$-padded partition, once we have chosen $\alpha = O(k^2)$. This, compared to the $\left(\frac{1}{2\sqrt{k}}, O(k^{2.5}), 1-\frac{1}{8k}\right)$-padded partition in Lemma 5.5 is the key improvement that reduces the dependence on k in the higher-order Cheeger upper bound from $O\big(k^{3.5}\big)$ to $O\big(k^3\big)$.

Because there is a reduction of α by a factor of $k^{1/2}$, rather than shrinking our sets by a factor of $1-\Omega(\frac{1}{k^3})$, we can shrink them by a factor of $1-\Omega(\frac{1}{k^{2.5}})$ and still guarantee, via the probabilistic method, that we can retain at least $k-\frac{1}{4}$ of the total mass.

We can then find well-separated sets where the pairwise distance is $\Omega(k^{-2.5})$ instead of $\Omega(k^{-3})$. This allows us to form vectors with Rayleigh quotient $O(k^6)\lambda_k$ instead of $O(k^7)\lambda_k$. And by exactly the same argument as above, we can round cuts with conductance $O(k^3) \cdot \sqrt{\lambda_k}$ instead of $O(k^{3.5}) \cdot \sqrt{\lambda_k}$.

6. Higher-Order Cheeger's Upper Bound $O(k^2)$

In the previous two sections, our approach involved partitioning $\mathbb{R}^k$ into well-separated sets, consolidating the well-separated sets into k well-separated sets with large mass, and then removing sets of low ℓ_2 mass from each of the k sets via Fiedler's Algorithm. The $O(k^2)$ bound takes a different approach.

The purpose of the *Padded Partition Theorem* from Section 5 was to partition $\mathbb{R}^k$ into sets of low diameter and such that for every $\boldsymbol{x} \in \mathbb{R}^k$, all vectors near $\boldsymbol{x}$ in a Euclidean sense are partitioned into the same set as $\boldsymbol{x}$ with probability close to 1.

But rather than thinking locally around each vector $\boldsymbol{x}$, we may want to take a more pointwise approach. After all, we only have finitely many embedded vectors in $\mathbb{R}^k$ that we are working with. Instead of partitioning $\mathbb{R}^k$ such that for every $\boldsymbol{x}$, vectors locally near $\boldsymbol{x}$ "usually" get partitioned into the same set as $\boldsymbol{x}$, we can instead partition $\mathbb{R}^k$ such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^k$, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are radially close to each other, then with probability "close to 1", they will be partitioned into the same set.

A random partition with this property is called a *Lipschitz Partition.*

Definition 6.1. *A random partition, $\mathcal{P}$, of X is* (R, L)-Lipschitz *if $\mathcal{P}$ is R-bounded and, for all $x, y \in X$*

$$\mathbb{P}[x \not\sim_{\mathcal{P}} y] \leq L \cdot \frac{\operatorname{dist}(x, y)}{R}$$

If $\operatorname{dist}(u, v)$ is small, $L \cdot \frac{\operatorname{dist}(x,y)}{R}$ is small, so $\mathbb{P}[u \sim_{\mathcal{P}} v]$ will be "close to 1". Therefore, in a Lipschitz Random Partition, vertices located near each other will "most likely" be partitioned into the same set.

Additionally, note that we want our partition to be R-bounded. This is because we still need to make sure our sets have sufficiently low diameter so that we can later consolidate them into at least k sets of sufficiently large mass.

To achieve such a partition, will use the *Lipschitz Partition Theorem.* Compared to the *Padded Partition Theorem*, the *Lipschitz Partition Theorem* will both simplify the algorithm and provide a tighter upper bound on the conductance.

Theorem 6.2 (Lipschitz Partition Theorem)**.** *For every Euclidean metric space, $X \in \mathbb{R}^k$, for every $R > 0$, exists a $(R, O(\sqrt{k}))$-Lipschitz Random Partition of X, i.e. a random partition $\mathcal{P}$ that is R-bounded and such that for all $x, y \in X$*

$$\mathbb{P}[x \not\sim_{\mathcal{P}} y] = O(\sqrt{k}) \cdot \frac{\operatorname{dist}(x, y)}{R}$$

The proof is provided in [24]. After partitioning $\mathbb{R}^k$ using a Lipschitz Partition, we will have t sets that collectively cut relatively few edges.

Next, we need to consolidate the sets into sets that each have relatively large mass. We do this using essentially the same method we used in Section 4.

How will we ensure that at least k of these sets have sufficiently low conductance?

The idea will be to use an argument involving Lemma A.1. We will bound the expected number of edges that are cut by at least 1 of the sets in our Lipschitz Partition. So on expectation, the ratio between the number of cut edges and the ℓ_2 mass will be small. In particular, one of the partitions within the sample space must achieve such a small ratio. Then, we will consolidate the sets in the partition to sets of large mass, and then show that at least k of the large mass sets have sufficiently small conductance.

We still want to remove vertices with small mass from our sets. We will prove via the probabilistic method that for each of our k low-conductance sets, there exists some mass threshold that will guarantee that our sets have sufficiently small Rayleigh quotient.

Let us start by choosing a random mass threshold, τ. Pick τ uniformly at random within the interval $(0, \max\{m(v) : v \in V\})$. Then, for every $S \subset V$, let $\hat{S}$ be the threshold cut

$$\hat{S} \stackrel{\text{def}}{=} \{v \in S : m(v) \geq \tau\}$$

Let's now show that there exists some partition, P, such that on expectation with respect to τ, the number of edges cut by at least one of the sets within the partition, i.e. $\mathbb{E}[E(\hat{S}_1) + \cdots + E(\hat{S}_m)]$, is small relative to $\sqrt{R_{\mathcal{L}}(F)} \leq \sqrt{\lambda_k}$.

Lemma 6.3. *Let τ be sampled uniformly from $\big(0, \max\{m(v) : v \in V\}\big)$. For all $R > 0$, there exists a partition, $V = S_1 \cup S_2 \cup \ldots \cup S_m$ such that for all $i \in [m]$, we have $\operatorname{diam}(S_i) \leq R$ and, on expectation with respect to τ,*

$$\mathbb{E}[E(\hat{S}_1) + \cdots + E(\hat{S}_m)] = O\left(\frac{k^{3/2}}{R}\right)\sqrt{R_{\mathcal{L}}(F)}$$

Proof. The proof will closely mirror the proof of Theorem 3.5. First, WLOG, we may assume $\max\{m(v)\} = 1$ because the Rayleigh quotient is constant under scaling (i.e. we can scale the embedding such that $\max\{m(v)\} = 1$ without affecting the Rayleigh quotient).

Let $\mathcal{P} = S_1 \cup S_2 \cup \ldots \cup S_m$ be a Lipschitz Random Partition guaranteed by Theorem 6.2. The goal will be to show that at least one of the partitions within the sample space of $\mathcal{P}$ satisfies the desired property.

Fix $X = E(\hat{S}_1) + \cdots + E(\hat{S}_m)$, and let

$$U = \{u \sim v : \{u, v\} \text{ is cut by at least one of the } \hat{S}_i\}$$

We bound $\mathbb{E}[X]$, by considering the contribution of each edge, $\{u, v\}$, to $\mathbb{E}[X]$. We have $\mathbb{E}[X] = \sum_{u \sim v} \mathbb{P}[\{u, v\} \in U]$. If $\{u, v\}$ is cut, there are two possibilities.

(1) First, suppose $u \not\sim_{\mathcal{P}} v$. Then $\{u, v\}$ will be cut as long as either $m(u) \geq \tau$ or $m(v) \geq \tau$. So we have $\{u, v\}$ cut as long as $m(u) \geq \tau$ or $m(v) \geq \tau$.
(2) Second, suppose $u \sim_{\mathcal{P}} v$, with $u, v \in S_i$. If $\{u, v\}$ is cut by $\hat{S}_i$, it must be that $m(u) < \tau \leq m(v)$, which would imply $u \notin \hat{S}_i$ and $v \in \hat{S}_i$, or $m(v) < \tau \leq m(u)$, which would imply $v \notin \hat{S}_i$ and $u \in \hat{S}_i$.

Note that the two cases provided above are mutually exclusive. So we have

$$\begin{aligned}
&\mathbb{P}\big[\{u, v\} \in U\big] \\
&\leq \mathbb{P}\big[u \not\sim_{\mathcal{P}} v \text{ and } \max\{m(u), m(v)\} \geq \tau\big] \\
&\qquad + \mathbb{P}[u \sim_{\mathcal{P}} v \text{ and } \tau \in [\min\{u, v\}, \max\{u, v\}]\big] \\
&= \mathbb{P}[u \not\sim_{\mathcal{P}} v] \cdot [\mathbb{P}[\max\{m(u), m(v)\} \geq \tau]] \mid (u \not\sim_{\mathcal{P}} v)] + |m(v) - m(u)|.
\end{aligned}$$

The events $(\max\{m(u), m(v)\} \geq \tau)$ and $(u \not\sim_{\mathcal{P}} v)$ are not independent. This is not a problem as we can use the union bound. Given $u \sim_{\mathcal{P}} v$, we still have that

$$\mathbb{P}[m(u) \geq \tau] = m(u)$$

and

$$\mathbb{P}[m(v) \geq \tau] = m(v)$$

So for the probability that either $m(u) \geq \tau$ or $m(v) \geq \tau$, we have

$$\mathbb{P}\left[\,(\max\{m(u), m(v)\} \geq \tau) \mid (u \not\sim_{\mathcal{P}} v)\,\right] \leq m(u) + m(v)$$

So

$$\mathbb{P}\left[u \not\sim_{\mathcal{P}} v\right] \cdot \mathbb{P}\Big[\max\{m(u), m(v)\} \geq \tau \mid (u \not\sim_{\mathcal{P}} v)\Big] \leq \mathbb{P}[u \not\sim_{\mathcal{P}} v]\ (m(u) + m(v))$$

By Theorem 6.2, $\mathbb{P}[u \not\sim_{\mathcal{P}} v] \leq \frac{O(\sqrt{k})}{R}\operatorname{dist}(u, v)$. Implying

$$\begin{aligned}\mathbb{P}[\{u, v\} \in U] &= \mathbb{P}[u \not\sim_{\mathcal{P}} v] \cdot [m(u) + m(v)] + |m(v) - m(u)| \\ &= \frac{O(\sqrt{k})}{R}\operatorname{dist}(u, v) \cdot \Big(m(u) + m(v)\Big) + |m(v) - m(u)|\end{aligned}$$

Additionally, note that WLOG, we may now assume $m(v) \geq m(u)$ and that

$$\mathbb{P}[\{u, v\} \in U] = \frac{1}{k}\left[\frac{O(\sqrt{k})}{R}\operatorname{dist}(u, v)\Big(m(u) + m(v)\Big) + m(v) - m(u)\right]$$

We want to factor out $||F(u)|| + ||F(v)||$, which we do via

$$\begin{aligned}&\frac{\sqrt{k}}{R}\operatorname{dist}(u, v) \cdot \Big(m(u) + m(v)\Big) + m(v) - m(u) \\ &\leq \frac{\sqrt{k}}{R}\operatorname{dist}(u, v) \cdot (||F(u)|| + ||F(v)||)^2 + ||F(v)||^2 - ||F(u)||^2 \\ &= \{||F(u)|| + ||F(v)||\}\frac{\sqrt{k}}{R}\operatorname{dist}(u, v)(||F(u) + ||F(v)||) + ||F(v)|| - ||F(u)||\}\end{aligned}$$

We have

$$\operatorname{dist}(u, v)||F(u)|| \leq 2||F(u) - F(v)||$$

and

$$\operatorname{dist}(u, v)||F(v)|| \leq 2||F(u) - F(v)||$$

by Lemma 4.13.

Additionally, $||F(v)|| - ||F(u)|| \leq ||F(u) - F(v)||$ by Cauchy-Schwartz. Summing these inequalities gives

$$\mathbb{P}[\{u, v\} \in U] \leq \frac{O(\sqrt{k})}{\delta}\big(||F(u)|| + ||F(v)||\big)\big(||F(u) - F(v)||\big)$$

So

$$\mathbb{E}[X] = \frac{O(\sqrt{k})}{R}\sum_{u \sim v}\big(||F(u)|| + ||F(v)||\big)\big(||F(u)|| - ||F(v)||\big)$$

To get the RHS to look like $R_{\mathcal{L}}(F)$, we use Cauchy-Schwartz to get

$$\mathbb{E}[X] \leq \frac{O(\sqrt{k})}{R}\sqrt{\sum_{u\sim v}(||F(u)||+||F(v)||)^2}\cdot\sqrt{\sum_{u\sim v}(||F(u)-F(v)||)^2}$$

Note that

$$\sum_{u\sim v}(||F(u)||+||F(v)||)^2 \leq \sum_{u\sim v}2(||F(u)||^2+||F(v)||^2) = 2k$$

So

$$\begin{aligned}
\mathbb{E}[X] &= O\left(\frac{\sqrt{k}}{R}\right)\cdot\sqrt{2k}\cdot\sqrt{\sum_{u\sim v}||F(u)-F(v)||^2} \\
&= O\left(\frac{k}{R}\right)\sqrt{\sum_{u\sim v}||F(u)-F(v)||^2} \\
&= O\left(\frac{k}{R}\right)\sqrt{\sum_{u\sim v}||F(u)-F(v)||^2} \\
&= O\left(\frac{k^{3/2}}{R}\right)\sqrt{\frac{\sum_{u\sim v}||F(u)-F(v)||^2}{k}} \\
&= O\left(\frac{k^{3/2}}{R}\right)\sqrt{R_{\mathcal{L}}(F)}
\end{aligned}$$

In particular, there must exist a single partition, $P \in \Omega(\mathcal{P})$ such that on expectation with respect to τ, $\mathbb{E}[X] = O(\frac{k^{3/2}}{R})R_{\mathcal{L}}(F)$. □

Next, we consolidate our sets into sets of large mass and prove that at least k of the sets, on expectation with respect to τ, have sufficiently small conductance.

Theorem 6.4. *For any graph G, $\phi_k(G) = O(\frac{\sqrt{k}}{\delta^{3/2}})\sqrt{\lambda_k}$.*

Proof. We first show that WLOG, we may assume δ is not arbitrarily small. This will make it easier for us to show that we can form sufficiently many sets of large mass.

Claim 6.4.1. *It suffices to show the result for when $\delta + \frac{\delta^2}{4} \geq \frac{1}{k}$.*

Proof. Suppose we have proven the theorem for when $\delta + \frac{\delta^2}{4} \geq \frac{1}{k}$. We will show that this implies the result for all $\delta \in (0, 1)$.

Let $\delta + \frac{\delta^2}{4} < \frac{1}{k}$. Pick $\delta' = \frac{1}{k} + \epsilon$ where $\epsilon > 0$ is sufficiently small such that $\delta' + \frac{\delta'^2}{2} \geq \frac{1}{k}$. Note that $r' = (1-\delta')k > (1-\frac{1}{k})k = k-1$, so $r' = \lceil(1-\delta')k\rceil = k$. Since F is $(R, 1+\frac{\delta}{4})$ spreading, F is also $(R, 1+\frac{\delta'}{4})$ spreading as $\delta' > \delta$. So using Theorem 6.4 for δ', we can pick k disjoint sets satisfying the desired property. Since $r = \lceil(1-\delta)k\rceil \leq k$, this proves the claim. □

The proof will strongly resemble the proof of Lemma 4.12. The main idea is to start with the partition guaranteed by Lemma 6.3. We want to consolidate the sets so that their mass is sufficiently large, i.e. $\geq \frac{1}{2}$, which will put a lower bound on the denominator of the Rayleigh quotient. We will then use the bound on the numerator provided by Lemma 6.3 to bound the conductance.

Let P be the partition guaranteed by Lemma 6.3. Before we begin combining sets with low mass, we need to make sure that there is no single set with too much mass, which would prevent us from making $\geq k$ sets with mass $\geq \frac{1}{2}$.

Luckily, our sets have low mass since they are R-bounded. By Lemma 4.9, F is $(R, \frac{1}{1-R^2})$-spreading. Choose $R = \epsilon\sqrt{\delta}$, where ϵ is sufficiently small so that $\frac{1}{1-R^2} \leq 1 + \frac{\delta}{4}$. Using this value for R in Lemma 4.9 gives us that F is $(R, 1 + \frac{\delta}{4})$-spreading. We then let $S_1, \dots, S_m$ be the partition guaranteed by Lemma 6.3. Since the S_i have diameter $\leq R$, they have mass at most $1 + \frac{\delta}{4}$.

Next, we consolidate the sets as follows. Whenever we have at least two sets with mass $\leq \frac{1}{2}$, combine them. Continue this procedure until at most one set remains with mass $\leq \frac{1}{2}$.

When the procedure terminates, all sets have mass $\leq 1 + \frac{\delta}{4}$, because combining two sets of mass $\leq \frac{1}{2}$ creates a set with mass $< 1 + \frac{\delta}{4}$.

Claim 6.4.2. *At least $k(1 - \frac{\delta}{2})$ of the sets have mass $\geq \frac{1}{2}$*

Proof. At most one set has mass $< \frac{1}{2}$. So the number of sets with mass $\in \left(\frac{1}{2}, 1+\frac{\delta}{4}\right)$ is

$$
\begin{aligned}
\geq \frac{k - \frac{1}{2}}{1 + \frac{\delta}{4}} &= k\left(\frac{1 - \frac{1}{2k}}{1 + \frac{\delta}{4}}\right) \\
&\geq k\left(\frac{1 - \frac{1}{2}\left(\delta + \frac{\delta^2}{4}\right)}{1 + \frac{\delta}{4}}\right) \\
&= k\left(\frac{\left(1 + \frac{\delta}{4}\right)\left(1 - \frac{\delta}{2}\right)}{1 + \frac{\delta}{4}}\right) \\
&= k\left(1 - \frac{\delta}{2}\right)
\end{aligned}
$$

where we use the assumption that $\delta + \frac{\delta^2}{4} \geq \frac{1}{k}$ from Claim 6.4.1. □

Let $T_1, \cdots, T_t$ be the sets of mass $\geq \frac{1}{2}$. We claim that, on expectation with respect to τ, at least r of these sets have sufficiently low conductance. First, note that since the mass of these sets is $\geq \frac{1}{2}$, we already have an upper bound for the denominator and suffices to bound the numerator.

Order the t sets in increasing order of the expected number of edges cut, i.e. such that $\mathbb{E}[E(\hat{T}_i)] \leq \mathbb{E}[E(\hat{T}_{i+1})]$. We will bound $\mathbb{E}[E(\hat{T}_r)]$. Note that as $k \to \infty$

there are asymptotically $k\delta$ sets that, on expectation, cut more edges than $\hat{T}_r$. It follows that

$$\mathbb{E}[E(\hat{T}_r)] \leq \frac{\mathbb{E}[E(\hat{S}_1) + \cdots + E(\hat{S}_m)]}{k\delta}$$

By Lemma 6.3, $\mathbb{E}[E(\hat{S}_1) + \cdots + E(\hat{S}_m)] = O(\frac{k^{3/2}}{R})\sqrt{R_{\mathcal{L}}(F)}$, so

$$\frac{\mathbb{E}[E(\hat{S}_1) + \cdots + E(\hat{S}_m)]}{k\delta} \leq \frac{k^{3/2}}{k\delta R}\sqrt{R_{\mathcal{L}}(F)} \leq \frac{\sqrt{k}}{\delta R}\sqrt{R_{\mathcal{L}}(F)}$$

Combining this with $m(T_r) \geq \frac{1}{2}$ gives

$$\frac{\mathbb{E}[E(\hat{T}_r)]}{m(T_r)} \leq O(\frac{\sqrt{k}}{\delta R})\sqrt{R_{\mathcal{L}}(F)}$$

Note that $m(T_i) = \mathbb{E}[\mathrm{vol}(\hat{T}_i)]$ because for every vertex, $\mathbb{P}[v \in \hat{T}_i] = m(v)$. Therefore, for every $i \in [r]$, since $R = \epsilon\sqrt{\delta}$,

$$\frac{\mathbb{E}[E(\hat{T}_i)]}{\mathbb{E}(\mathrm{vol}(\hat{(T_i)}))} = O(\frac{\sqrt{k}}{\delta^{3/2}})\sqrt{R_{\mathcal{L}}(F)}$$

So for all $i \in [r]$, by Lemma A.2 exists some $\tau \in (0,1)$ such that $\phi(\hat{T}_i) = O(\frac{\sqrt{k}}{\delta^{3/2}})\sqrt{R_{\mathcal{L}}(F)}$, as desired. □

We have now essentially proven the desired result.

Theorem 1.4. $\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^2)\sqrt{\lambda_k}$.

Proof. We proved the lower bound side in Corollary 4.2. The upper bound side follows by setting $\delta = \frac{1}{2k}$ in Lemma 6.3. □

7. Conclusion

This paper provided a proof of both the discrete Cheeger inequality, and the higher-order Cheeger inequality concerning k-way conductance $\phi_k(G)$. There are a number of generalizations, and results related to the statements in this paper, which we now discuss.

7.1. Generalization to Irregular, Weighted Graphs.

All results provided in Section 3 through Section 6 also hold for irregular and arbitrarily weighted graphs. We provide an overview of how to generalize the results to irregular graphs. See [3] for details on generalizing the results to weighted graphs and [13] for a more in-depth discussion on generalizing the results to irregular graphs.

To generalize to the case of irregular graphs, we need to relate the Rayleigh quotient with respect to $\mathcal{L}$ to the conductance of the graph in a similar manner to what we did for regular graphs. For irregular graphs, we can define the $\mathcal{L} \stackrel{\text{def}}{=} \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, where $\mathbf{D}$ is the degree matrix and $\mathbf{A}$ is the adjacency matrix

For any $S \subset V$, we would hope that $\phi(S) = R_{\mathcal{L}}(\mathbf{1}_S)$. Although this is not quite the case, the result holds with a slight adjustment.

Lemma 7.1. *Let G=(V,E) be a simple, unweighted, undirected, and not necessarily regular graph. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $\mathcal{L}$. Then for every $S \subset V$,*

$$\phi(S) = R_{\mathcal{L}}(\mathbf{D}^{1/2}\mathbf{1}_S)$$

where $\mathbf{D}$ *is the degree matrix.*

Proof. Let $\boldsymbol{x} = \mathbf{1}_S$.

$$\begin{aligned} R_{\mathcal{L}}(\mathbf{D}^{1/2}x) &= \frac{\mathbf{D}\boldsymbol{x}^\top \boldsymbol{x} - (\mathbf{1}_S)^\top \mathbf{D}^{1/2}\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{D}^{1/2}\mathbf{1}_S}{\boldsymbol{x}^\top \boldsymbol{x}} \\ &= \frac{\mathbf{D}\boldsymbol{x}^\top \boldsymbol{x} - (\mathbf{1}_S)^\top \mathbf{A}\mathbf{1}_S}{\boldsymbol{x}^\top \boldsymbol{x}} \\ &= \frac{\sum\limits_{v \in V} (x_v)^2 d_v - 2\sum\limits_{u \sim v} x_u x_v}{\mathrm{vol}(S)} \end{aligned}$$

We now calculate the numerator. We consider the contribution of each pair of vertices, u and v, to the numerator.

If $u \not\sim v$, then there is no contribution to $(x_u)^2 d_u$, $(x_v)^2 d_v$, or $\sum\limits_{u \sim v} x_u x_v$.

If $u \sim v$, there is a contribution of $(x_u)^2$ to $(x_u)^2 d_u$, $(x_v)^2$ to $(x_v)^2 d_v$, and $2x_u x_v$ to $2\sum\limits_{u \sim v} x_u x_v$. The total contribution to the numerator is

$$(x_u)^2 + (x_v)^2 - 2x_u x_v = (x_u - x_v)^2 .$$

It follows that $R_{\mathcal{L}}(\mathbf{D}^{1/2}x) = \frac{\sum_{u \sim v} (x_u - x_v)^2}{\mathrm{vol}(S)} = \phi(S)$. $\square$

All expressions involving $R_{\mathcal{L}}(x)$ can now be reframed using Lemma 7.1 to generalize our results to the case of irregular graphs.

7.2. **Further Higher-Order Cheeger Inequalities.** There are multiple ways to state higher-order Cheeger inequalities. For example, rather than only stipulating that our k cuts are disjoint, we can also enforce that k-way conductance is quantified over a *partitioning* of the vertices in G. For this notion of conductance, [3] proves there is always a way to find a partition of V with conductance $O(k^4)\sqrt{\lambda_k}$.

Another higher-order Cheeger statement bounds k-way conductance with respect to the Rayleigh quotient of the $2k$-th smallest eigenvalue. For example Theorem 1.2 in [3] demonstrates that

Theorem 7.2. *For every graph G and for every $k \in \mathbb{N}$,*

$$\phi_k(G) \leq O\big(\sqrt{\lambda_{2k} \log k}\big)$$

Notice the dependence on k is significantly tighter. Such an inequality is useful for certifying that a graph is a small-set expander [3].

Finally, there exists a higher-order Cheeger statement that relates standard conductance, rather than k-way conductance, to λ_k [25].

Theorem 7.3. *For every undirected graph G,*

$$\frac{\lambda_2}{2} \leq \phi(G) \leq O\left(\frac{k}{\sqrt{\lambda_k}}\right) \cdot \lambda_2$$

This inequality demonstrates that for constant k and λ_k, the second eigenvector provides a constant factor approximation to the minimum conductance cut, and was motivated by empirical observations in image segmentation that the second eigenvector provides an excellent k-way partition when there's a large gap in the spectrum [12, 26].

7.3. **Tightness of Bounds and Future Research.** It is known that both sides of Cheeger's inequality are tight. For the upper bound, equality is achieved via the path graph, and for the lower bound, equality is achieved via the balanced binary tree graph. The lower bound of the higher-order Cheeger's inequality is also tight.

Improving the dependence on k for the upper bound in Theorem 1.4, as well as some of the higher-order Cheeger inequalities stated in the previous section, remain an interesting open problem. [3] showed that the noisy hypercube graph satisfies $\phi_{k/2}(G) > O\big(\sqrt{k \log k}\big) \cdot \sqrt{\lambda_k}$, which suggests that Theorem 7.2 is tight. Otherwise, it remains admissible that $\phi_k(G) \leq O\big(\log^{O(1)} k\big) \cdot \sqrt{\lambda_k}$.

8. Acknowledgements

I would like to thank Antares Chen, my REU mentor. Antares provided invaluable guidance to me throughout the REU program by introducing me to an interesting topic to study, by providing suggestions on lecture notes and papers that I should read, and by explaining the intuition and technical details to help me understand the material. After the program's conclusion, he continued to guide me in writing a well-structured paper and clarifying remaining areas of confusion.

I would also like to thank Peter May for organizing the UChicago REU program. In addition to learning about spectral graph theory, the various lectures throughout the program have aroused my interest in many mathematical subjects such as fractals, hyperbolic geometry, and number theory.

Through the help of Antares Chen and Peter May, I have found a subject that I enjoy studying in depth and have gotten a taste of research mathematics.

References

[1] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.
[2] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.
[3] James R Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM (JACM)*, 61(6):1–30, 2014.
[4] Jeff Cheeger. A lower bound for the smallest eigenvalue of the laplacian, problems in analysis (papers dedicated to salomon bochner, 1969), 1970.
[5] Noga Alon and Vitali D Milman. $\lambda 1$, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.
[6] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.
[7] Daniel A Spielman and Shang-Hua Teng. Spectral partitioning works: Planar graphs and finite element meshes. *Linear Algebra and its Applications*, 421(2-3):284–305, 2007.
[8] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 51(3):497–515, 2004.
[9] Shuji Jimbo and Akira Maruoka. Expanders obtained from affine transformations. In *Proceedings of the seventeenth annual ACM symposium on Theory of computing*, pages 88–97, 1985.
[10] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.
[11] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM (JACM)*, 51(4):671–697, 2004.
[12] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.
[13] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. *University of California, Berkeley, https://people. eecs. berkeley. edu/luca/books/expanders-2016. pdf*, 2017.
[14] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.
[15] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.
[16] David A Tolliver and Gary L Miller. Graph partitioning by spectral rounding: Applications in image segmentation and clustering. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 1053–1060. IEEE, 2006.
[17] Bengt Aspvall and John R Gilbert. Graph coloring using eigenvalue decomposition. *SIAM Journal on Algebraic Discrete Methods*, 5(4):526–538, 1983.
[18] Noga Alon and Nabil Kahale. A spectral technique for coloring random 3-colorable graphs (preliminary version). In *Proceedings of the twenty-sixth annual ACM symposium on Theory of Computing*, pages 346–355, 1997.
[19] Laurent Miclo. On eigenfunctions of markov processes on trees. *Probability Theory and Related Fields*, 142:561–594, 2008.
[20] Barry Simon and Raphael Høegh-Krohn. Hypercontractive semigroups and two dimensional self-coupled bose fields. *Journal of Functional Analysis*, 9(2):121–180, 1972.
[21] Deepak Verma and Marina Meila. Comparison of spectral clustering methods. *Advances in neural information processing systems*, 15:38, 2003.
[22] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.
[23] Moses Charikar, Chandra Chekuri, Ashish Goel, Sudipto Guha, and Serge Plotkin. Approximating a finite metric by a small number of tree metrics. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No. 98CB36280)*, pages 379–388. IEEE, 1998.
[24] James R Lee and Assaf Naor. Extending lipschitz functions via random metric partitions. *Inventiones mathematicae*, 160(1):59–95, 2005.
[25] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved cheeger's inequality: Analysis of spectral partitioning algorithms through higher

order spectral gap. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 11–20, 2013.

[26] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17:395–416, 2007.

Appendix A. Appendix

In this section, we provide some fundamental facts that we use throughout the paper.

Lemma A.1. *Suppose* $\frac{a_1}{b_1} \leq \frac{a_2}{b_2} \leq \cdots \leq \frac{a_n}{b_n}$.

Then $\frac{a_1}{b_1} \leq \frac{a_1+a_2+\cdots+a_n}{b_1+b_2+\cdots+b_n} \leq \frac{a_n}{b_n}$

The next lemma, Lemma A.2 is used to prove the Theorem 3.5 and Theorem 6.4

Lemma A.2. *Let X and Y be random variables such that $\mathbb{P}[Y > 0] = 1$. Then*

$$\mathbb{P}\left[\frac{X}{Y} \leq \frac{\mathbb{E}X}{\mathbb{E}Y}\right] > 0$$

.

The full proof of Lemma A.2 is provided by [13] in *fact 5.2.*