



Spectral Graph Theory and Higher-order Cheeger Inequalities

Matt Taylor

Supervised by Asma Hassannezhad
Level 6
20 Credit Points

May 9, 2022

Acknowledgements

I am thankful to Asma Hassannezhad, my supervisor, for her guidance throughout this project.

Contents

1	Introduction & Motivation	3
1.1	Heat diffusion	3
1.2	Importance of the second eigenvalue	4
1.3	Identifying bottlenecks	5
1.4	Spectral clustering	6
2	Preliminaries	6
2.1	Spectral graph theory	8
2.2	Properties of the Laplacian	9
2.3	Variational characterisation of eigenvalues	11
2.4	Properties of spectra	13
3	Cheeger's inequality	15
3.1	Lower bound on the Cheeger constant	16
3.2	Upper bound on the Cheeger constant	17
4	Higher-order Cheeger inequalities	20
4.1	Lower bound on the k-way expansion constant	20
4.2	Upper bound on the k-way expansion constant	22
4.2.1	Random partitioning of metric spaces	22
4.2.2	Spreading properties of the embedding	23
4.2.3	Refining the random partition	26
4.3	Further discussion	31
5	Inverse Spectral Theory	32
5.1	Unnormalised Laplacian	32
5.2	Normalised Laplacian	34
6	Spectral Graph Theory in Practice	34
6.1	Spectral clustering	34
6.1.1	Experiments	35
6.1.2	Efficiency	36
6.2	Image segmentation	36
6.3	VLSI design	38
6.4	Vertex centrality	38

1 Introduction & Motivation

How many different friend groups are there at the University of Bristol? How does the shape of an object influence the rate at which heat spreads through it? How does Google order the results of a search? These are among the problems which can be analysed with *spectral graph theory*.

Many important problems in computer science can be expressed in the language of graphs. In *spectral graph theory*, we represent graphs as matrices and study how the eigenvalues of these matrices relate to properties of the graph. Perhaps the most natural way to represent a graph $G = (V, E)$ is the $|V| \times |V|$ adjacency matrix, but we will mostly direct our attention towards the closely related *Laplacian* and *normalised Laplacian* matrices instead. To motivate this decision, we will begin by showing how these matrices arise naturally in the context of heat diffusion.

1.1 Heat diffusion

The analysis of heat diffusion is usually framed in a continuous setting and modelled with partial differential equations (and in fact, the continuous Laplacian operator is analogous to the Laplacian matrix we study). We instead consider a discrete version of this problem where we have a graph $G = (V, E)$ and some amount of ‘heat’ associated with each vertex. The set of edges E determine which vertices exchange heat with each other. Over time, the heat spreads throughout the graph until an equilibrium is reached.

More formally, we can represent the amount of heat in the graph by a vector $u = (u_1, \dots, u_n) \in \mathbb{R}^n$ where $u_i \in \mathbb{R}$ is the amount of heat on the i th vertex. To represent the spread of heat over time, we will define an operator $W : \mathbb{R}^n \rightarrow \mathbb{R}^n$ which takes a heat vector $u \in \mathbb{R}^n$ and gives us a vector $W(u) \in \mathbb{R}^n$ describing the amount of heat on each vertex at the next timestep. This scheme is represented by Figure 1. For simplicity, we consider only d -regular graphs (in which every vertex has d neighbours), but all the results we discuss generalise to irregular graphs.

For a specific vertex $i \in V$ with heat u_i , we would like for $W(u)_i$ (i.e the amount of heat i has after one timestep) to be some kind of weighted average of i ’s heat and the heat of i ’s neighbours. With this in mind, we can define

$$W(u)_i = \frac{1}{2}u_i + \frac{1}{2d} \sum_{j:\{i,j\} \in E} u_j.$$

Note that the value $1/2$ is arbitrary; it simply controls the rate at which heat spreads. The above formulation is in fact equivalent to viewing W as a $n \times n$ real symmetric matrix given by

$$W_{ij} = \begin{cases} \frac{1}{2d} & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}.$$

Thus if the heat on the graph is $u \in \mathbb{R}^n$ at one timestep, then after one timestep the heat will be Wu , and after two it will be $W(Wu) = W^2u$, etc. In particular, the repeated application of W calls for diagonalisation. Let $\psi_1, \dots, \psi_n \in \mathbb{R}^n$ be the eigenvectors of W and let $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ be their associated eigenvalues (which are real because W is symmetric). Writing our initial heat distribution $u \in \mathbb{R}^n$ in the spectral basis (with real coefficients $\{\alpha_i\}_{i \in \mathbb{N}}$) we then have

$$u = \alpha_1\psi_1 + \alpha_2\psi_2 + \dots + \alpha_n\psi_n,$$

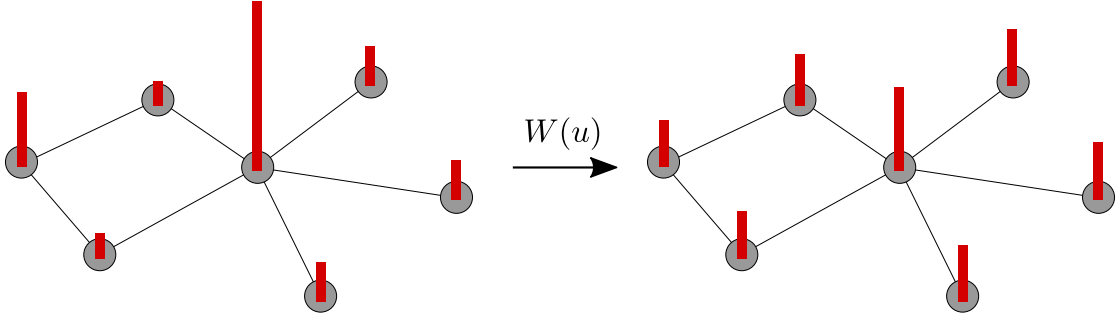


Figure 1: Heat diffusion on a graph, before and after applying W .

$$\begin{aligned}
Wu &= \mu_1 \alpha_1 \psi_1 + \mu_2 \alpha_2 \psi_2 + \cdots + \mu_n \alpha_n \psi_n, \\
W^2 u &= \mu_1^2 \alpha_1 \psi_1 + \mu_2^2 \alpha_2 \psi_2 + \cdots + \mu_n^2 \alpha_n \psi_n, \\
&\dots \\
W^k u &= \mu_1^k \alpha_1 \psi_1 + \mu_2^k \alpha_2 \psi_2 + \cdots + \mu_n^k \alpha_n \psi_n.
\end{aligned}$$

When written in this form we can intuit some properties of the heat flow. One can verify from the definition of W that $\mu_1 = 1$ and $\psi_1 = c\mathbf{1}$ for some $c \in \mathbb{R}$. In a connected graph we will also have $\mu_1 > \mu_2$. Hence over time, as the eigenvalues are exponentiated, all of the terms will go to zero *except* the term $\mu_1 \alpha_1 \psi_1$. This represents the fact that heat will distribute uniformly throughout the graph, since $\psi_1 = c\mathbf{1}$. This exactly matches our physical intuition; that heat applied to one end of an object will spread throughout over time.

In a similar vein, μ_2 is the second largest eigenvalue and thus ψ_2 controls the ‘first order’ heat flow over time. If μ_2 is close to zero, then the $\mu_2^k \alpha_2 \psi_2$ term will decay quickly, and the heat will reach equilibrium very fast. If instead μ_2 is very close to 1, then the heat will reach equilibrium slowly.

1.2 Importance of the second eigenvalue

At this point, it is natural to ponder what properties of the graph influence the value of μ_2 . We will slightly reparameterise things by defining the Laplacian matrix $L = I - W$ and writing the eigenvalues of L by $\lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$. Observe that L has the same eigenvectors as W , and that every eigenvalue μ of W becomes $1 - \mu$ in our new matrix L . So $\lambda_2 \approx 0$ implies that the graph reaches equilibrium slowly, and we want to relate this quantity back to the structure of the graph.

Cheeger’s inequality (Theorem 3.4) addresses precisely this. It was first proved in the continuous setting of manifolds by Jeff Cheeger in 1970 [1], and later a connection was seen to the discrete setting of graphs in 1985 [2]. We will first introduce a few definitions, recalling that we are still talking about d -regular graphs for simplicity. For a subset $S \subseteq V$ with $0 < |S| < n$, we define the *edge expansion* of S by

$$\phi(S) = \frac{\# \text{ edges leaving } S}{d|S|}.$$

Since the number of edges leaving S is surely at most $d|S|$, we can think of $\phi(S)$ as a measure of how much smaller the numerator is than $d|S|$. If many edges leave S , then S is ‘well connected’ with the rest of the graph and $\phi(S)$ will be high.

The edge expansion of a *cut* in the graph is given by

$$\phi(S, V \setminus S) = \max\{\phi(S), \phi(V \setminus S)\}.$$

Finally, the *Cheeger constant* h_G is the minimum edge expansion over all possible cuts, i.e

$$h_G = \min_{\substack{S \subseteq V \\ 0 < |S| < |V|}} \phi(S, V \setminus S).$$

If h_G is small, then there exists a cut $S \subseteq V$ such that there are few edges between S and $V \setminus S$. This represents a ‘bottleneck’ in the graph (which heat should struggle to flow between). We finally arrive at Cheeger’s inequality (Theorem 3.4), which states that

$$\frac{\lambda_2}{2} \leq h_G \leq \sqrt{2\lambda_2}.$$

Qualitatively, this tells us that λ_2 is small *if and only if* there are major bottlenecks in the graph. This again matches our physical intuition: our graph should reach heat equilibrium more slowly if there are major bottlenecks.

1.3 Identifying bottlenecks

The most natural proof of Cheeger’s inequality is algorithmic. It starts from the eigenvector ψ_2 of L corresponding to λ_2 , and applies *Fiedler’s algorithm* (Algorithm 1) to construct a set $S \subseteq V$ for which $\phi(S, V \setminus S)$ is small. Cheeger’s inequality and Fiedler’s algorithm are explored in more detail in Section 3.

Algorithm 1: Fiedler’s algorithm

```

1 Function fiedler( $x$ )
   Input:  $G = (V, E, w)$  and vector  $x \in \mathbb{R}^n$ 
   Output:  $S \subseteq V$  such that  $\phi(S, V \setminus S)$  is small
2    $v_1, \dots, v_n \leftarrow$  vertices  $v \in V$  sorted according to the value  $x_v$ 
3    $k \leftarrow \operatorname{argmin}_{k \in \{1, \dots, n\}} \phi(\{v_1, \dots, v_k\}, \{v_{k+1}, \dots, v_n\})$ 
4   return  $\{v_1, \dots, v_k\}$ 
5 end

```

There are a few ways to see why information about bottlenecks is present in ψ_2 , but one easy way is through Lemma 2.15 which states that

$$\psi_2 = \operatorname{argmin}_{\mathbf{1} \perp f \in \mathbb{R}^n} \frac{\sum_{\{x,y\} \in E} (f_x - f_y)^2}{\sum_{v \in V} d \cdot f_v}.$$

Directing our attention to the numerator, we can see that the value which ψ_2 places on each vertex is such that the squared difference between *neighbouring vertices* is minimised. If we have two poorly connected sets S and $V \setminus S$, then ψ_2 will place a similar value on all vertices in S , and a similar value on all vertices in $V \setminus S$. Since we also require that $\psi_2 \perp \mathbf{1}$, we ensure that these two values are distinct, and thus ψ_2 can be used to identify a given vertex with one of the two sets by inspecting the corresponding component of ψ_2 .

1.4 Spectral clustering

We have seen how Fiedler’s algorithm may be used to find cuts in the graph with low edge expansion. In fact, the problem of finding the cut of minimum edge expansion, also called the *sparsest cut* problem, is known to be NP-Hard [3]. In rough terms, this means that no efficient algorithm solving sparsest cut is known. Improving Fiedler’s algorithm to find the sparsest cut efficiently (for a particular definition of ‘efficiently’) is equivalent to proving $P = NP$ and would earn the author a \$1 million prize [4].

Just as Fiedler’s algorithm may be used to split the graph into *two* non-expanding sets, we may also be interested in finding *more than two* non-expanding sets (referred to as ‘clusters’). Finding clusters in a graph is an interesting problem in its own right. For example, one may represent a social network as a graph $G = (V, E)$ where people are represented as vertices, and friendships between people are represented as edges between vertices. A cluster in this graph represents a friend group or community.

The general approach to producing a k -clustering from spectral information is as follows: instead of considering ψ_2 on its own (as with Fiedler’s algorithm), we consider the matrix with columns ψ_1, \dots, ψ_k (see Definition 4.5). Then geometric clustering algorithms (such as k -means clustering) may be applied to the rows of this matrix to produce k clusters in the graph. These algorithms have been used successfully for many decades; see e.g [5] and [6]. We refer the reader to Section 6 for more details on spectral clustering in practice (in particular Algorithm 5).

Recall that Cheeger’s inequality implies the existence of a good quality 2-clustering *if and only if* λ_2 is small. For many years, no direct relation was known between the existence of a good quality k -clustering and λ_k , until Lee, Oveis Gharan and Trevisan [7] proved in 2014 that

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

where h_k is a natural extension of h_G to cuts consisting of $k > 2$ sets. This is the main theorem to which we will devote most of our attention; see Section 4 for more details. As with Cheeger’s inequality, the proof is constructive, and proceeds along the lines of the generic spectral clustering scheme described above.

In Section 5 we briefly discuss the related problem of constructing a graph whose Laplacian matrix has a particular set of eigenvalues. We will see a few other applications of spectral graph theory in Section 6, e.g in image segmentation and in producing vertex centrality measures.

2 Preliminaries

We begin with some basic definitions and theorems regarding graphs, their Laplacian matrices, and their associated eigenvectors. Many of the theorems and proof ideas are adapted from [8].

Definition 2.1 (Graph). A *weighted undirected* graph is an ordered triple $G = (V, E, w)$. We call V the set of *vertices* and $E \subseteq \{\{x, y\} : x, y \in V\}$ the set of *edges*. For any $x, y \in V$ we say that there is an edge joining x and y if $\{x, y\} \in E$, which we also write as $x \sim y$. Each edge is assigned a positive weight by the (symmetric) function $w : V \times V \rightarrow [0, \infty)$ where $w(x, y) > 0$ if and only if $\{x, y\} \in E$. For two subsets $S_1, S_2 \subseteq V$ we define the set of edges crossing between S_1 and S_2 by

$$E(S_1, S_2) := \{\{x, y\} \in E : x \in S_1, y \in S_2\}.$$

Note that for the purposes of this project we only consider weighted and undirected graphs with finite V , and so we will write $n := |V|$ throughout. Unweighted graphs may be handled by reduction to a weighted graph with weight 1 on all edges. We will see later that directed graphs are problematic because they do not give rise to symmetric operators (e.g rendering Definition 2.4 invalid).

Notation and terminology

Let $G = (V, E, w)$.

1. We say that G is *connected* if there is a path between any two vertices. That is to say, G is connected if for any two vertices $x, y \in V$, there exist vertices $w_1, w_2, \dots, w_k \in V$ such that

$$(x \sim w_1) \wedge (w_1 \sim w_2) \wedge \dots \wedge (w_k \sim y).$$

2. A *component* of a graph is a maximal connected subgraph of G . For example, a connected graph has one component. A disconnected graph consisting of two vertices and no edge between them has two components.
3. For $S \subseteq V$, write $S^c := V \setminus S$. We also write the indicator function $\mathbf{1}_S : V \rightarrow \mathbb{R}$ to mean

$$\mathbf{1}_S(x) := \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \in S^c \end{cases}.$$

4. For $k \in \mathbb{N}$, write $[k] := \{1, 2, \dots, k\}$.
5. For $f : V \rightarrow \mathbb{R}$, define the *support* of f by $\text{supp}(f) := \{v \in V : f(v) \neq 0\}$. We say that a set of functions $\{f_1, \dots, f_n\}$ are *disjointly supported* if the functions have pairwise disjoint supports.
6. Write $A \lesssim B$ to mean $A \in O(B)$, and $A \asymp B$ to mean $O(A) = O(B)$.
7. Let (X, d) be a metric space. For $x \in X$ and $r \in \mathbb{R}$, write $B(x, r)$ to be the open ball of radius r centered on x , so that $B(x, r) = \{y \in X : d(x, y) < r\}$. For a subset $S \subseteq X$, write $\text{diam}(S) = \sup\{d(x, y) : x, y \in S\}$.

We will write $\sum_{x \sim y}$ to denote a sum over all unordered pairs $\{x, y\} \in E$. If $y \in V$ is fixed, then we will write

$$\sum_{\substack{x \\ x \sim y}}$$

to denote a sum over all vertices $x \in V$ such that $x \sim y$ (i.e sum over all neighbours of y). Since we sum over *unordered* pairs, the following equality arises:

Remark 2.2. Let $f : V \times V \rightarrow \mathbb{R}$, then

$$2 \sum_{x \sim y} f(x, y) = \sum_{y \in V} \sum_{\substack{x \\ x \sim y}} f(x, y).$$

For each $x \in V$, we extend the notion of weight to vertices by writing $w(x) := \sum_{y \sim x} w(x, y)$. We can extend this to subsets $S \subseteq V$ by writing $w(S) := \sum_{x \in S} w(x)$.

2.1 Spectral graph theory

We will now introduce *spectral graph theory*. The most basic object of study is a matrix associated with the graph called the *Laplacian* matrix, which is a discrete analogue of the continuous Laplacian operator.

Definition 2.3 (Laplacian). Let D be the diagonal degree matrix where for $x \in V$ the (x, x) th entry is equal to $w(x)$. Let W be the weight matrix where for $x, y \in V$ the (x, y) th entry is equal to $w(x, y)$. Then the *Laplacian matrix* corresponding to a graph G is defined by

$$L = D - W,$$

or in other words for $x, y \in V$ we have

$$L(x, y) = \begin{cases} w(x) & \text{if } x = y, \\ -w(x, y) & \text{otherwise} \end{cases}.$$

Definition 2.4. The *normalised Laplacian* of a graph G is given by

$$\mathcal{L} = D^{-1/2} L D^{-1/2}.$$

We can also write it as

$$\mathcal{L} = D^{-1/2} (D - W) D^{-1/2} = I - D^{-1/2} W D^{-1/2}.$$

We will sometimes use the notation L_G or \mathcal{L}_G to emphasise the graph G to which the matrices correspond.

The normalised Laplacian matrix retains many of the algebraic properties of the Laplacian (e.g Lemma 2.9), except some calculations are simplified by the fact that all its diagonal entries are 1. For some types of graph there is a direct correspondence between eigenvalues of L and \mathcal{L} (see Example 2.6), but in general their eigenvalues (and eigenvectors) behave quite differently.

Note that both L and \mathcal{L} are symmetric, which implies that all their eigenvalues are real (for directed graphs, which we do not study, this does not hold). Later we will show that these matrices are positive semi-definite (see Lemma 2.9), which implies that all their eigenvalues are non-negative. Lemma 2.7 also shows that 0 is an eigenvalue of both. Therefore it makes sense to order the eigenvalues of a given normalised Laplacian as follows:

Definition 2.5 (Spectrum). The *spectrum* of a graph G is the multiset of eigenvalues $\{\lambda_i\}_{i \in [n]}$ of its associated normalised Laplacian \mathcal{L} . Throughout this project we write them in non-decreasing order so that

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

Note that we must represent the spectrum of a graph as a multiset in order to capture the fact that some eigenvalues may be repeated with multiplicity. We will use the word ‘multiplicity’ to mean both geometric and algebraic multiplicity, since they are equal for symmetric matrices such as \mathcal{L} .

Example 2.6. We say a graph is k -regular if each vertex $v \in V$ has k edges leaving it, each with weight 1. This implies $D = kI$ and so

$$\mathcal{L} = D^{-1/2} L D^{-1/2} = \frac{1}{k} L.$$

Thus in a k -regular graph, $f \in \mathbb{R}^n$ is an eigenvector of L with corresponding eigenvalue $\lambda \in \mathbb{R}$ if and only if f is an eigenvector of \mathcal{L} with corresponding eigenvalue λ/k .

Lemma 2.7. 0 is an eigenvalue of both L and \mathcal{L} , with eigenvectors $\mathbf{1}$ and $D^{1/2}\mathbf{1}$ respectively.

Proof. Observe that the sum of any row in L is zero. Thus $L\mathbf{1} = \mathbf{0}$ and

$$\mathcal{L}D^{1/2}\mathbf{1} = D^{-1/2}LD^{-1/2}D^{1/2}\mathbf{1} = D^{-1/2}L\mathbf{1} = \mathbf{0}.$$

□

2.2 Properties of the Laplacian

Next, we will start to examine the properties of these matrices more closely, with the goal of relating facts about the spectrum to facts about the graph's structure.

Recall that $n = |V|$. We emphasise that a function $f : V \rightarrow \mathbb{R}$ may equivalently be viewed as a vector $f \in \mathbb{R}^n$, having one component for each vertex in the graph. Thus we will use these notions interchangeably, often writing the individual components in a more functional notation $f(v) \in \mathbb{R}$ for a given vertex $v \in V$. We also write $\langle f, g \rangle$ and $\|f\|$ to mean the standard inner products and norms on \mathbb{R}^n respectively.

We now prove some useful lemmas which will allow us to relate the spectrum of a graph to arbitrary vectors in \mathbb{R}^n .

Lemma 2.8. Let $f \in \mathbb{R}^n$, then

$$\langle f, Lf \rangle = \sum_{x \sim y} w(x, y)(f(x) - f(y))^2.$$

Proof. Recall that $L = D - W$, then

$$\begin{aligned} \langle f, Lf \rangle &= \sum_{x \in V} f(x)(Lf)(x) \\ &= \sum_{x \in V} f(x) [(Df)(x) - (Wf)(x)] \\ &= \sum_{x \in V} f(x) \left[w(x)f(x) - \sum_{y \sim x} w(x, y)f(y) \right] \\ &= \left[\sum_{x \in V} w(x)f(x)^2 \right] - \left[\sum_{x \in V} f(x) \sum_{y \sim x} w(x, y)f(y) \right] \\ &= \left[\sum_{x \in V} \left[\sum_{y \in V} w(x, y) \right] f(x)^2 \right] - \left[\sum_{x \in V} \sum_{y \in V} w(x, y)f(x)f(y) \right] \\ &= \sum_{x \in V} \sum_{y \in V} [w(x, y)f(x)^2 - w(x, y)f(x)f(y)] \\ &= \frac{1}{2} \sum_{x \in V} \sum_{y \in V} [w(x, y)f(x)^2 - 2w(x, y)f(x)f(y) + w(x, y)f(y)^2] \\ &= \frac{1}{2} \sum_{x \in V} \sum_{y \in V} [w(x, y)(f(x) - f(y))^2] \\ &= \sum_{x \sim y} w(x, y)(f(x) - f(y))^2 \end{aligned}$$

where the last line follows from Remark 2.2.

□

Corollary 2.9. L and \mathcal{L} are positive semi-definite.

Proof. Observe that the right hand sum in the statement of Lemma 2.8 consists of non-negative terms, which implies that L is positive semi-definite. For \mathcal{L} , let $g \in \mathbb{R}^n$ and write $f = D^{-1/2}g$. Then

$$\langle g, \mathcal{L}g \rangle = \left\langle g, D^{-1/2}LD^{-1/2}g \right\rangle = \left\langle D^{1/2}f, D^{-1/2}Lf \right\rangle = \langle f, Lf \rangle \geq 0.$$

□

We now introduce a key quantity called the *Rayleigh quotient*.

Definition 2.10 (Rayleigh quotient). Let $f \in \mathbb{R}^n$ be non-zero and write $g = D^{1/2}f$. Then we call $\mathcal{R}(f)$ the *Rayleigh quotient* of f and define it by

$$\begin{aligned} \mathcal{R}(f) &:= \frac{\langle g, \mathcal{L}g \rangle}{\langle g, g \rangle} \\ &= \frac{\langle g, D^{-1/2}LD^{-1/2}g \rangle}{\langle D^{1/2}f, D^{1/2}f \rangle} \\ &= \frac{\langle D^{1/2}f, D^{-1/2}Lf \rangle}{\sum_{v \in V} [(D^{1/2}f)(v)]^2} \\ &= \frac{\langle f, Lf \rangle}{\sum_{v \in V} [\sqrt{w(v)}f(v)]^2} \\ &= \frac{\sum_{x \sim y} w(x, y)(f(x) - f(y))^2}{\sum_{v \in V} w(v)f(v)^2}. \end{aligned}$$

If instead $f = \mathbf{0}$ then we take $\mathcal{R}(f) = \infty$.

While we prefer to use the final line of Definition 2.10 as the working definition of the Rayleigh quotient, when written in the first form $\langle g, \mathcal{L}g \rangle / \langle g, g \rangle$ it is easy to see a connection to the eigenvalues of \mathcal{L} :

Remark 2.11. Let $g \in \mathbb{R}^n$ be an eigenvector of \mathcal{L} with corresponding eigenvalue λ . Then

$$\mathcal{R}(D^{-1/2}g) = \frac{\langle g, \mathcal{L}g \rangle}{\langle g, g \rangle} = \frac{\langle g, \lambda g \rangle}{\langle g, g \rangle} = \lambda \frac{\langle g, g \rangle}{\langle g, g \rangle} = \lambda.$$

Lemma 2.12. Let ψ_1, \dots, ψ_n be orthonormal eigenvectors of \mathcal{L} corresponding to $\lambda_1, \dots, \lambda_n$. Then for any $f, g \in \mathbb{R}^n$, expressed as $f = \sum_{i=1}^n a_i \psi_i$ and $g = \sum_{i=1}^n b_i \psi_i$ for some coefficients $\{a_i\}_{i \in [n]}$ and $\{b_i\}_{i \in [n]}$, we have

$$\langle f, g \rangle = \sum_{i=1}^n a_i b_i.$$

Proof. We have

$$\langle f, g \rangle = \left\langle \sum_{i=1}^n a_i \psi_i, \sum_{i=1}^n b_i \psi_i \right\rangle = \sum_{j=1}^n \sum_{i=1}^n \langle a_i \psi_i, b_j \psi_j \rangle = \sum_{i=1}^n \langle a_i \psi_i, b_i \psi_i \rangle = \sum_{i=1}^n a_i b_i$$

where in last two equalities we have used the orthonormality of the ψ_1, \dots, ψ_n . □

2.3 Variational characterisation of eigenvalues

In this section, we use a *variational characterisation* of eigenvalues to express them as the solutions to optimisation problems. To see why this is useful, suppose that a statement of the form $\lambda_2 = \min_x f(x)$ holds. Clearly if we have some x for which $f(x) \leq c$, then we have successfully proved that $\lambda_2 \leq c$. We will use these results later when trying to prove bounds on these eigenvalues.

Lemma 2.13 (Courant-Fischer). *For any graph G and $k \in [n]$ we have*

$$\lambda_k = \min_{U: \dim U = k} \left(\max_{\mathbf{0} \neq g \in U} \frac{\langle g, \mathcal{L}g \rangle}{\langle g, g \rangle} \right)$$

where the minimum is taken over all k -dimensional subspaces U of \mathbb{R}^n .

Proof. We will prove this using a similar approach to that of [9]. We start with an orthonormal set of eigenvectors ψ_1, \dots, ψ_n of \mathcal{L} which correspond to the eigenvalues $\lambda_1, \dots, \lambda_n$. Let U be an arbitrary k -dimensional subspace of \mathbb{R}^n , and let $U' = \text{span}\{\psi_k, \dots, \psi_n\}$, noting that $\dim U' = n - k + 1$. Then we have

$$\dim U + \dim U' = k + (n - k + 1) = n + 1.$$

But since these are both subspaces of \mathbb{R}^n , this implies that the set $\{x \neq \mathbf{0} : x \in U \cap U'\}$ is non-empty. One can see that $\psi_k \in U'$ and by Remark 2.11 we have

$$\frac{\langle \psi_k, \mathcal{L}\psi_k \rangle}{\langle \psi_k, \psi_k \rangle} = \lambda_k.$$

Then we have that

$$\begin{aligned} \sup_{\{x \neq \mathbf{0} : x \in U\}} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} &\geq \sup_{\{x \neq \mathbf{0} : x \in U \cap U'\}} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} \geq \inf_{\{x \neq \mathbf{0} : x \in U \cap U'\}} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} \\ &\geq \inf_{\{x \neq \mathbf{0} : x \in U'\}} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} = \lambda_k. \end{aligned}$$

Thus we can take the infimum over all choices of U and the inequality will still hold, yielding

$$\inf_{U: \dim U = k} \sup_{\{x \neq \mathbf{0} : x \in U\}} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} \geq \lambda_k.$$

Finally, note that the subspace $W = \{\psi_1, \dots, \psi_k\}$ is a possible choice for U , and in particular we have $\psi_k \in W$. This vector would achieve the infimum and supremum yielding value λ_k . Hence

$$\lambda_k = \min_{U: \dim U = k} \left(\max_{\mathbf{0} \neq x \in U} \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle} \right).$$

□

We remark that the previous Lemma holds for any Hermitian matrix, not just \mathcal{L} . Equivalently, with a small alteration we may turn Lemma 2.13 into a statement about the Rayleigh quotients ($\mathcal{R}(\cdot)$), which is slightly easier for us to relate back to properties of the graph.

Theorem 2.14 (Courant-Fischer with Rayleigh quotient). *For $k \in [n]$ we have*

$$\lambda_k = \min_{U: \dim U = k} \max_{0 \neq f \in U} \mathcal{R}(f)$$

where the minimum is taken over all k -dimensional subspaces U of \mathbb{R}^n .

Proof. The proof is identical to that of Lemma 2.13, except we now take

$$\begin{aligned} U' &= \text{span}\{D^{-1/2}\psi_k, \dots, D^{-1/2}\psi_n\}, \text{ and} \\ W &= \text{span}\{D^{-1/2}\psi_1, \dots, D^{-1/2}\psi_k\} \end{aligned}$$

where ψ_1, \dots, ψ_n are the eigenvectors of \mathcal{L} as before. We also substitute $\frac{\langle \cdot, \mathcal{L} \cdot \rangle}{\langle \cdot, \cdot \rangle}$ with $\mathcal{R}(\cdot)$, and finish by noting that

$$\mathcal{R}(D^{-1/2}\psi_k) = \frac{\langle \psi_k, \mathcal{L}\psi_k \rangle}{\langle \psi_k, \psi_k \rangle} = \lambda_k.$$

□

The last two inequalities in this section focus specifically on λ_2 and λ_n .

Lemma 2.15. *For any graph G , we have*

$$\lambda_2 = \min_{0 \neq f \perp D\mathbf{1}} \mathcal{R}(f).$$

Furthermore, if $f \in \mathbb{R}^n$ is the vector achieving the minimum above, then $D^{1/2}f$ is the eigenvector of \mathcal{L} corresponding to λ_2 .

Proof. First we substitute $f = D^{-1/2}g$ in the right hand side of the statement, so that by Remark 2.11 we have

$$\min_{0 \neq D^{-1/2}g \perp D\mathbf{1}} \mathcal{R}(D^{-1/2}g) = \min_{0 \neq g \perp D^{1/2}\mathbf{1}} \frac{\langle g, \mathcal{L}g \rangle}{\langle g, g \rangle}.$$

We would like to show that the above quantity is equal to λ_2 . Let ψ_1, \dots, ψ_n be orthonormal eigenvectors of \mathcal{L} corresponding to $\lambda_1, \dots, \lambda_n$. Let $g \in \mathbb{R}^n$ be written as $g = \sum_{i=1}^n a_i \psi_i$ for some $\{a_i\}_{i \in [n]}$. Recall that $\psi_1 = D^{1/2}\mathbf{1}$ by Lemma 2.7 and so if $g \perp D^{1/2}\mathbf{1}$ we will have $a_1 = 0$. Then by Lemma 2.12 we have

$$\langle g, \mathcal{L}g \rangle = \sum_{i=1}^n a_i^2 \lambda_i = \sum_{i=2}^n a_i^2 \lambda_i \geq \lambda_2 \sum_{i=2}^n a_i^2$$

by the ordering of the eigenvalues, and so

$$\frac{\langle g, \mathcal{L}g \rangle}{\langle g, g \rangle} = \frac{\sum_{i=1}^n a_i^2 \lambda_i}{\sum_{i=1}^n a_i^2} \geq \lambda_2 \cdot \frac{\sum_{i=1}^n a_i^2}{\sum_{i=1}^n a_i^2} = \lambda_2.$$

But we also have equality when $D^{1/2}f = g = \psi_2$, which completes the proof. □

Lemma 2.16. *For any graph G , we have*

$$\lambda_n = \max_{0 \neq f \in \mathbb{R}^n} \mathcal{R}(f).$$

Proof. Immediate from Theorem 2.14 applied to $k = n$, seeing that \mathbb{R}^n is the only n -dimensional subspace of itself. □

2.4 Properties of spectra

The remaining theorems in this section will be devoted to some basic properties and useful facts about the spectra of graphs.

Proposition 2.17. *For all $i \in [n]$ we have $\lambda_i \in [0, 2]$.*

Proof. Observe that the lower bound follows from Corollary 2.9. For the upper bound, it suffices to place an upper bound on λ_n (since this is the largest eigenvalue). Let $f \in \mathbb{R}^n$ and note that for any $x, y \in V$ we have

$$\begin{aligned} [f(x) - f(y)]^2 &= [f(x)^2 + f(y)^2 - 2f(x)f(y)] \\ &= 2[f(x)^2 + f(y)^2 - (f(x) + f(y))^2] \\ &\leq 2[f(x)^2 + f(y)^2] \end{aligned}$$

and so by Lemma 2.16 and Remark 2.2 we have

$$\lambda_n = \max_{\mathbf{0} \neq f \in \mathbb{R}^n} \frac{\sum_{x \sim y} w(x, y) (f(x) - f(y))^2}{\sum_{v \in V} w(v) f(v)^2} \leq \max_{\mathbf{0} \neq f \in \mathbb{R}^n} \frac{2 \sum_{x \sim y} w(x, y) (f(x)^2 + f(y)^2)}{\sum_{v \in V} w(v) f(v)^2} \leq 2.$$

□

Lemma 2.18. *The spectrum of G is the union of the spectra of G 's components.*

Proof. Let G_1, G_2, \dots, G_k be the k components of G . Then with a suitable reordering of vertices we can write \mathcal{L} in block matrix form so that

$$\mathcal{L} = \begin{pmatrix} \mathcal{L}_{G_1} & & & \\ & \mathcal{L}_{G_2} & & \\ & & \ddots & \\ & & & \mathcal{L}_{G_k} \end{pmatrix}$$

where all the non-diagonal entries are zero matrices of appropriate size. Note that this form arises because there are no edges between any two components, so all corresponding entries will be zero.

For each $i \in [k]$ let us consider a particular $G_i = (V_i, E_i, w_i)$. Let $f_{G_i} \in \mathbb{R}^{|V_i|}$ be an eigenvector of \mathcal{L}_{G_i} with corresponding eigenvalue λ . Now define $f \in \mathbb{R}^n$ by

$$f(v) = \begin{cases} f_{G_i}(v) & \text{if } v \in V_i \\ 0 & \text{otherwise} \end{cases}. \quad (1)$$

Observe that

$$\mathcal{L}f(v) = \begin{cases} \mathcal{L}_{G_i} f_{G_i}(v) = \lambda f_{G_i}(v) & \text{if } v \in V_i \\ 0 & \text{otherwise} \end{cases}$$

and so $\mathcal{L}f = \lambda f$. We have shown that any eigenvector in G_i corresponds to an eigenvector in G (as given by (1)) with the same eigenvalue, but it remains to show the reverse. This follows from the fact that there are $\sum_i |V_i| = n$ such eigenvectors across all components, and G has n eigenvectors in total. Since there is a bijective mapping between sets of eigenvectors and multisets of eigenvalues, the statement follows. □

Theorem 2.19. *G is connected if and only if the zero eigenvalue of \mathcal{L} has multiplicity one.*

Proof. Suppose that G is connected. Let g be an eigenvector of \mathcal{L} corresponding to the zero eigenvalue (so that $\mathcal{L}g = 0$), whose existence is guaranteed by Lemma 2.7. Writing $f = D^{-1/2}g$, we have

$$0 = \frac{\langle g, \mathcal{L}g \rangle}{\langle g, g \rangle} = \mathcal{R}(f) = \frac{\sum_{x \sim y} w(x, y) (f(x) - f(y))^2}{\sum_{v \in V} w(v) f(v)^2}$$

and in particular $\sum_{x \sim y} w(x, y) (f(x) - f(y))^2 = 0$. As this sum consists of non-negative terms, we must have $f(u) = f(v)$ for all $u \sim v$.

Since G is connected, for any $x, y \in V$ there is a set of vertices $\{w_0, w_1, \dots, w_m\} \subseteq V$ such that $x \sim w_0, w_0 \sim w_1, \dots, w_m \sim y$ all hold, so by transitivity we have $f(x) = f(y)$. Thus f achieves the same value $c \in \mathbb{R}$ for every vertex in the graph, meaning $g = D^{1/2}f = D^{1/2}(c\mathbf{1})$. Hence the eigenspace of the zero eigenvalue has dimension 1.

For the reverse direction, we use the contrapositive. Suppose that G is disconnected. Our goal is to show that the multiplicity of the zero eigenvalue is not equal to one. We can decompose G into its $k > 1$ components G_1, G_2, \dots, G_k , each of which have the value zero in their spectra by Lemma 2.7. Now Lemma 2.18 implies that the spectrum of G contains the value zero at least k times (since the union is taken in the sense of multisets). \square

Corollary 2.20. *G has k components if and only if the zero eigenvalue of \mathcal{L} has multiplicity k .*

Proof. Follows from Lemma 2.18 and Theorem 2.19 applied to each of the components. \square

Proposition 2.21. *We say that a graph G is bipartite if there exists a set $S \subseteq V$ such that $E = E(S, S^c)$. In other words, there are no edges joining any two vertices in S , or joining any two vertices in S^c . Then G is bipartite if and only if for each eigenvalue λ_i of \mathcal{L} , the value $2 - \lambda_i$ is also an eigenvalue of \mathcal{L} .*

Proof. Suppose G is bipartite and connected, with parts S and S^c . It is sufficient for us to show that $I - \mathcal{L}$ and $-(I - \mathcal{L})$ have the same eigenvalues (which implies that the eigenvalues of \mathcal{L} are symmetric about 1).

First note that $I - \mathcal{L}$ has diagonal elements all 0, since the diagonal of \mathcal{L} consists of 1s. Then with a suitable relabeling of the vertices, from bipartiteness there exist matrices A and B such that with block notation we have

$$I - \mathcal{L} = \begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix} = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} (-(I - \mathcal{L})) \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix},$$

and the statement follows from the fact that the matrices on the left and right have determinant one.

In the other case, suppose G is bipartite but not connected. Let G_1, G_2, \dots, G_k be the decomposition of G into each of its $k \in [n]$ components, which must each be connected and bipartite. So by the argument above, if λ is an eigenvalue of some G_i , the value $2 - \lambda$ is also an eigenvalue of G_i . Now apply Lemma 2.18 to transfer this statement to the whole graph G . \square

With the preliminaries out of the way, we will switch focus to examining how spectral information can be used to analyse a specific property of graphs.

3 Cheeger's inequality

Theorem 2.19 tells us that in any graph $\lambda_2 = 0$ if and only if the graph is disconnected. A natural extension of this Lemma would be to consider what happens when $\lambda_2 \approx 0$: perhaps the graph will be *nearly* disconnected? Cheeger's inequality (see Lemma 3.4) resolves this question in the affirmative, providing a *robust* version of Theorem 2.19.

To prove Cheeger's inequality, we must first define what it means for a graph to be *nearly disconnected*, beginning with the following definition:

Definition 3.1 (Expansion). Let $S \subseteq V$ satisfy $w(S) \neq 0$. The *expansion* of S is given by

$$\phi(S) := \frac{w(E(S, S^c))}{w(S)}.$$

For convenience we also take $\phi(S) = \infty$ when $w(S) = 0$.

If a set $S \subseteq V$ has high expansion, it indicates that the S is well-connected with the rest of the graph (i.e the set $E(S, S^c)$ is large).

We briefly note the following connection between Rayleigh quotients and expansions:

Lemma 3.2. Let $S \subseteq V$. Recall that $\mathbf{1}_S \in \mathbb{R}^n$ is the 0/1 vector such that $\mathbf{1}_S(v) = 1$ if and only if $v \in S$. Then we have

$$\mathcal{R}(\mathbf{1}_S) = \phi(S).$$

Proof. Trivial for $w(S) = 0$. Otherwise we have

$$\mathcal{R}(\mathbf{1}_S) = \frac{\sum_{x \sim y} w(x, y) (\mathbf{1}_S(x) - \mathbf{1}_S(y))^2}{\sum_{v \in V} w(v) \mathbf{1}_S(v)^2} = \frac{\sum_{x \in S, y \in S^c} w(x, y)}{\sum_{v \in S} w(v)} = \frac{w(E(S, S^c))}{w(S)} = \phi(S).$$

□

We now return to the problem of quantifying how ‘nearly disconnected’ a graph is. Informally, a graph may be considered nearly disconnected if it can be partitioned into two sets with few edges between them, i.e two sets of low expansion. This notion is captured by the following key definition:

Definition 3.3. The *Cheeger constant* of a graph h_G is defined by

$$h_G := \min_{S \subseteq V} \max \{ \phi(S), \phi(S^c) \}.$$

One can see that h_G is small if and only if there exist two complementary subsets of small expansion, which is exactly what we wanted. Later, we will see a generalisation of the Cheeger constant to graph partitions consisting of more than 2 subsets.

Finally, we arrive at Cheeger's inequality, which provides a quantitative connection between h_G and the spectrum of G .

Theorem 3.4 (Cheeger's inequality). For any graph G , we have

$$\frac{\lambda_2}{2} \leq h_G \leq \sqrt{2\lambda_2}.$$

This implies that h_G is small if and only if λ_2 is small. For this reason, λ_2 is sometimes referred to as the ‘algebraic connectivity’ of a graph.

We will begin by proving the lower bound on h_G .

3.1 Lower bound on the Cheeger constant

In this section, let $S \subseteq V$ be a set of vertices achieving the optimal cut of h_G so that

$$h_G = \max\{\phi(S), \phi(S^c)\} \geq \phi(S).$$

Thanks to our variational characterisation of eigenvalues (Section 2.3), proving upper bounds on eigenvalues (as we now desire) is much easier than proving lower bounds. For example, Lemma 2.15 tells us that

$$\lambda_2 = \min_{\mathbf{1} \perp f \in \mathbb{R}^k \setminus \mathbf{0}} \mathcal{R}(f)$$

and hence the proof of $\lambda_2/2 \leq h_G$ is reduced to finding an appropriate function $f : V \rightarrow \mathbb{R}$ with $\mathcal{R}(f) \leq 2h_G$. One may verify via direct calculation that the function

$$f(v) = \begin{cases} 1/w(S) & \text{if } v \in S \\ 1/w(S^c) & \text{if } v \in S^c \end{cases}$$

is satisfactory (see e.g [8] for details), but we opt instead for a slightly different approach which generalises more intuitively to higher dimensions. We detail this approach in the following remark:

Remark 3.5. Recall Theorem 2.14. We are interested in the $k = 2$ case, which gives us

$$\lambda_2 = \min_{U: \dim U=2} \left(\max_{\mathbf{0} \neq f \in U} \mathcal{R}(f) \right)$$

where the minimum is over 2-dimensional subspaces of \mathbb{R}^n . Now pick $U = \text{span}\{\mathbf{1}_S, \mathbf{1}_{S^c}\}$ so that $\dim U = 2$. We will proceed by showing that any non-zero vector $f \in U$ satisfies $\mathcal{R}(f) \leq 2h_G$, which implies

$$\lambda_2 \leq \max_{\mathbf{0} \neq f \in U} \mathcal{R}(f) \leq 2h_G$$

and hence $\lambda_2/2 \leq h_G$. This is achieved with the aid of the following lemma:

Lemma 3.6. *Let $f, g \in \mathbb{R}^n$ be vectors such that $D^{1/2}f$ and $D^{1/2}g$ are orthogonal. Then*

$$\mathcal{R}(f + g) \leq 2 \cdot \max\{\mathcal{R}(f), \mathcal{R}(g)\}.$$

Proof. Let $x = D^{1/2}f$ and $y = D^{1/2}g$. We will expand x and y as a linear combination of eigenvectors ψ_1, \dots, ψ_n of \mathcal{L} (with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$), so that $x = \sum_i a_i \psi_i$ and $y = \sum_i b_i \psi_i$ for some real numbers $\{a_i\}_{i \in [n]}$ and $\{b_i\}_{i \in [n]}$.

By orthogonality we have $\langle x + y, x + y \rangle = \|x\|^2 + \|y\|^2$. Recalling Definition 2.10 and then applying Lemma 2.12 we have

$$\begin{aligned} \mathcal{R}(f + g) &= \frac{\langle x + y, \mathcal{L}(x + y) \rangle}{\langle x + y, x + y \rangle} \\ &= \frac{\sum_i \lambda_i (a_i + b_i)^2}{\|x\|^2 + \|y\|^2} \\ &\leq \frac{\sum_i 2\lambda_i (a_i^2 + b_i^2)}{\|x\|^2 + \|y\|^2} \\ &= 2 \cdot \frac{\langle x, \mathcal{L}x \rangle + \langle y, \mathcal{L}y \rangle}{\|x\|^2 + \|y\|^2} \end{aligned}$$

$$\begin{aligned}
&\leq 2 \cdot \max \left\{ \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle}, \frac{\langle y, \mathcal{L}y \rangle}{\langle y, y \rangle} \right\} \\
&= 2 \cdot \max \{ \mathcal{R}(f), \mathcal{R}(g) \}.
\end{aligned}$$

□

Now we may complete the proof that $\frac{\lambda_2}{2} \leq h_G$.

Proof of lower bound in Theorem 3.4. Let $U = \text{span}\{\mathbf{1}_S, \mathbf{1}_{S^c}\}$ as before. For any $f \in U$, write $f = a\mathbf{1}_S + b\mathbf{1}_{S^c}$ where $a, b \in \mathbb{R}$. Note that Rayleigh quotients are scale-invariant (so that e.g. $\mathcal{R}(a\mathbf{1}_S) = \mathcal{R}(\mathbf{1}_S)$), and that the vectors $D^{1/2}a\mathbf{1}_S$ and $D^{1/2}b\mathbf{1}_{S^c}$ are orthogonal. So we can apply Lemma 3.6 on these vectors, followed by Lemma 3.2 to obtain

$$\begin{aligned}
\mathcal{R}(f) &= \mathcal{R}(a\mathbf{1}_S + b\mathbf{1}_{S^c}) \\
&\leq 2 \cdot \max \{ \mathcal{R}(a\mathbf{1}_S), \mathcal{R}(b\mathbf{1}_{S^c}) \} \\
&= 2 \cdot \max \{ \mathcal{R}(\mathbf{1}_S), \mathcal{R}(\mathbf{1}_{S^c}) \} = 2 \cdot \max \{ \phi(S), \phi(S^c) \} = 2h_G,
\end{aligned}$$

which by Remark 3.5 is sufficient to prove the desired result. □

3.2 Upper bound on the Cheeger constant

Next we prove the upper bound on h_G , which states that $h_G \leq \sqrt{2\lambda_2}$. To do this, we will prove a slightly stronger statement which is more useful for the purpose of spectral clustering. We will prove in Lemma 3.9 that for any vector $f \in \mathbb{R}^n$ with $D^{1/2}f \perp D^{1/2}\mathbf{1}$, we have $\text{fiedler}(f) \leq \sqrt{2\mathcal{R}(f)}$ where $\text{fiedler}(f)$ is the result of applying Algorithm 1 to f . This implies that even if we only have an approximate eigenvector corresponding to λ_2 , Fiedler's algorithm is still able to produce a cut of good quality.

The following two lemmas will help us achieve this. First, we prove that for our input $f \in \mathbb{R}^n$, we may round this into another vector with smaller Rayleigh quotient. Then we prove that the new vector can be used to find a low-expansion set. These will combine to give us an upper bound on h_G as desired.

Lemma 3.7. *Let $f \in \mathbb{R}^n$ satisfy $D^{1/2}f \perp D^{1/2}\mathbf{1}$. Then there exists a vector $g \in \mathbb{R}^n$ with $w(\text{supp}(g)) \leq w(V)/2$ such that*

$$\mathcal{R}(g) \leq \mathcal{R}(f).$$

Proof. First, we will show that for any $c \in \mathbb{R}$ we have

$$\mathcal{R}(f + c\mathbf{1}) \leq \mathcal{R}(f). \quad (2)$$

This follows from the fact that $\mathcal{R}(f + c\mathbf{1})$ and $\mathcal{R}(f)$ share the same numerator, and the denominator of $\mathcal{R}(f + c\mathbf{1})$ is

$$\|D^{1/2}(f + c\mathbf{1})\|^2 = \|D^{1/2}f\|^2 + \|D^{1/2}\mathbf{1}\|^2 \geq \|D^{1/2}f\|^2$$

where we have used the fact that $D^{1/2}f \perp D^{1/2}\mathbf{1}$.

Next, we relabel the vertices according to f , so that

$$f(v_1) \geq f(v_2) \geq \dots \geq f(v_n).$$

For any $i \in [n]$ let $S_i = \{v_1, \dots, v_i\}$ and take r to be the greatest integer such that $w(S_r) \leq w(V)/2$ (so that $f(v_r)$ is in some sense the ‘median’ value of f). We will translate f to be centered on this value by defining $g \in \mathbb{R}^n$ as

$$g = f - f(v_r).$$

At this point, note that both $w(S_r)$ and $w(S_r^c)$ have value at most $w(V)/2$, by the definition of r . We now claim that one of g^+ or g^- satisfy the statement of the lemma. To see this, observe that $\text{supp}(g^+) = S_r$ and $\text{supp}(g^-) = S_r^c$. It remains to show that one of them has Rayleigh quotient at most $\mathcal{R}(g)$ (and by (2), at most $\mathcal{R}(f)$).

Write $g = g^+ - g^-$. We will quickly prove that for $\{x, y\} \in E$ we have

$$[(g^+(x) - g^+(y)) - (g^-(x) - g^-(y))]^2 \geq (g^+(x) - g^+(y))^2 + (g^-(x) - g^-(y))^2. \quad (3)$$

One can see that if e.g $x \in S_r$ and $y \in S_r^c$ then we have $g^+(y) = g^-(x) = 0$ and the inequality becomes $(g^+(x) + g^-(y))^2 \geq g^+(x)^2 + g^-(y)^2$. Otherwise if e.g we have $x, y \in S_r$ then $g^-(x) = g^-(y) = 0$ and the inequality follows.

To finish, we can use (3) and the fact that g^+ and g^- are disjointly supported to conclude

$$\begin{aligned} \mathcal{R}(g) &= \frac{\sum_{x \sim y} w(x, y) (g(x) - g(y))^2}{\sum_{v \in V} w(v) g(v)^2} \\ &= \frac{\sum_{x \sim y} w(x, y) [(g^+(x) - g^+(y)) - (g^-(x) - g^-(y))]^2}{\sum_{v \in V} w(v) (g^+(x) - g^-(x))^2} \\ &\geq \frac{\sum_{x \sim y} w(x, y) [(g^+(x) - g^+(y))^2 + (g^-(x) - g^-(y))^2]}{\sum_{v \in V} w(v) g^+(x)^2 + \sum_{v \in V} w(v) g^-(x)^2} \\ &= \frac{\sum_{x \sim y} w(x, y) (g^+(x) - g^+(y))^2 + \sum_{x \sim y} w(x, y) (g^-(x) - g^-(y))^2}{\sum_{v \in V} w(v) g^+(x)^2 + \sum_{v \in V} w(v) g^-(x)^2} \\ &= \min\{\mathcal{R}(g^+), \mathcal{R}(g^-)\}. \end{aligned}$$

Hence one of g^+ or g^- satisfy the statement of the lemma. \square

Lemma 3.8. *For any $f : V \rightarrow \mathbb{R}^n$, there exists a subset $S \subseteq \text{supp}(f)$ such that*

$$\phi(S) \leq \sqrt{2\mathcal{R}(f)}.$$

Proof. Suppose that $\text{supp}(f) = V$. Then we have $\phi(V) = 0 \leq \sqrt{2\mathcal{R}(f)}$ so taking $S = V$ completes the proof. Otherwise we have $\text{supp}(f) \neq V$. For each $t \in [0, \infty)$ define a level set

$$S_t := \{u \in V : f(u)^2 > t\}.$$

Note that this definition implies $S_t \subseteq \text{supp}(f)$.

Let $T = \{f(u)^2 : u \in V\}$. Then by considering $w(S_t)$ as a monotonically decreasing step function of t we can see that

$$\begin{aligned} \int_0^\infty w(S_t) dt &= \sum_{t \in T} t \cdot w(\{u \in V : f(u)^2 = t\}) \\ &= \sum_{u \in V} w(u) f(u)^2. \end{aligned}$$

Next, we have

$$\begin{aligned} \int_0^\infty w(E(S_t, (S_t)^c)) dt &= \int_0^\infty \sum_{\substack{x \sim y \\ f(x)^2 < t \leq f(y)^2}} w(x, y) dt \\ &= \sum_{\substack{x \sim y \\ f(x)^2 \leq f(y)^2}} w(x, y) \int_{f(x)^2}^{f(y)^2} dt \end{aligned}$$

$$\begin{aligned}
&= \sum_{x \sim y} w(x, y) |f(x)^2 - f(y)^2| \\
&\leq \sum_{x \sim y} w(x, y) |f(x) - f(y)| \cdot |f(x) + f(y)| \\
&\leq \sqrt{\sum_{x \sim y} w(x, y) |f(x) - f(y)|^2} \sqrt{\sum_{x \sim y} w(x, y) |f(x) + f(y)|^2} \\
&\leq \sqrt{\sum_{x \sim y} w(x, y) |f(x) - f(y)|^2} \sqrt{2 \sum_{v \in V} w(v) f(v)^2}
\end{aligned}$$

where the last inequality holds from Remark 2.2. So putting the inequalities together gives us

$$\begin{aligned}
\frac{\int_0^\infty w(E(S_t, S_t^c)) dt}{\int_0^\infty w(S_t) dt} &\leq \frac{\sqrt{\sum_{x \sim y} w(x, y) |f(x) - f(y)|^2} \sqrt{2 \sum_{v \in V} w(v) f(v)^2}}{\sum_{u \in V} w(u) f(u)^2_{\mathcal{H}}} \\
&\leq \sqrt{\frac{2 \sum_{x \sim y} w(x, y) |f(x) - f(y)|^2}{\sum_{u \in V} w(u) f(u)^2}} \\
&\leq \sqrt{2\mathcal{R}(f)}.
\end{aligned}$$

Finally, we will show by contradiction that there exists some $t \in [0, \infty)$ for which S_t satisfies $\phi(S_t) \leq \sqrt{2\mathcal{R}(f)}$. Suppose that for all $t \in [0, \infty)$ we have

$$w(E(S_t, S_t^c)) > w(S_t) \sqrt{2\mathcal{R}(f)},$$

then clearly $\int_0^\infty w(E(S_t, S_t^c)) dt > \sqrt{2\mathcal{R}(f)} \int_0^\infty w(S_t) dt$ which contradicts the above. \square

Now we return to prove the stronger form of the upper bound.

Lemma 3.9. *Let $f \in \mathbb{R}^n$ be a vector such that $D^{1/2}f \perp D^{1/2}\mathbf{1}$. Then there exists a set $S \subseteq V$ such that*

$$\max\{\phi(S), \phi(S^c)\} \leq \sqrt{2\mathcal{R}(f)}.$$

Proof. Let $f \in \mathbb{R}^n$ be a vector such that $D^{1/2}f \perp D^{1/2}\mathbf{1}$, and let $g \in \mathbb{R}^n$ be the vector with $w(\text{supp}(g)) \leq w(V)/2$ promised to us by Lemma 3.7. By Lemma 3.8, there exists a set $S \subseteq \text{supp}(g)$ such that

$$\phi(S) \leq \sqrt{2\mathcal{R}(g)} \leq \sqrt{2\mathcal{R}(f)}.$$

Now since $w(S) \leq w(\text{supp}(g)) \leq w(V)/2 \leq w(S^c)$, we have $\phi(S) \geq \phi(S^c)$. Then we have

$$\max\{\phi(S), \phi(S^c)\} = \phi(S) \leq \sqrt{2\mathcal{R}(f)}.$$

\square

Using this, we may finally complete the proof that $h_G \leq \sqrt{2\lambda_2}$.

Proof of upper bound in Theorem 3.4. Let $f \in \mathbb{R}^n$ be the eigenvector of \mathcal{L} corresponding to λ_2 , which by Lemma 2.15 is orthogonal to $D^{1/2}\mathbf{1}$. Then by Lemma 3.9, there exists a set $S \subseteq V$ such that

$$\max\{\phi(S), \phi(S^c)\} \leq \sqrt{2\mathcal{R}(f)},$$

but then $\mathcal{R}(f) = \lambda_2$ by Remark 2.11 and hence

$$h_G = \min_{T \subseteq V} \max\{\phi(T), \phi(T^c)\} \leq \max\{\phi(S), \phi(S^c)\} \leq \sqrt{2\mathcal{R}(f)} = \sqrt{2\lambda_2}.$$

\square

Unfortunately this proof can't be generalised to cuts of more than 2 sets without significantly more advanced machinery, which is what we will develop in the following section.

4 Higher-order Cheeger inequalities

We now aim to introduce a generalisation of the Cheeger constant to cuts consisting of more than 2 sets. The main reference for this section, from which we adapt most of the statements and proofs, is [7].

We begin with a higher-order analogue of Definition 3.3.

Definition 4.1 (*k*-way expansion constant). Let $k \in [n]$. The *k*-way expansion constant is defined by

$$h_k := \min_{S_1, \dots, S_k} \max\{\phi(S_1), \dots, \phi(S_k)\}$$

where the minimum is taken over all disjoint, non-empty subsets $S_1, \dots, S_k \subseteq V$ (but the subsets do not have to form a partition of V).

Intuitively, a small value of h_k for some $k \in [n]$ implies that there exists k disjoint, non-empty subsets of V , each with low expansion (i.e well isolated from the rest of the graph). We call such a family of sets a *k*-clustering of G .

Remark 4.2. Note that $h_2 = h_G$: the minimum in h_2 is taken with a looser set of conditions, giving $h_2 \leq h_G$, and for any 2-clustering $S_1, S_2 \subseteq V$ which achieves h_2 we can let S be the set with greater expansion which yields

$$h_2 = \max\{\phi(S_1), \phi(S_2)\} = \phi(S) = \max\{\phi(S), \phi(S^c)\} \geq h_G.$$

We now arrive at the main theorem of this project, to which we will devote much of our attention.

Theorem 4.3 (Higher-order Cheeger inequalities). *Let $k \in [n]$. Then*

$$\frac{\lambda_k}{2} \leq h_k \leq O(k^2) \sqrt{\lambda_k}.$$

The above theorem has a few notable consequences. Firstly, if a particular λ_k is small for some $k \in [n]$, then Theorem 4.3 *guarantees* the existence of a k -way clustering with each cluster having low expansion. Secondly, if $\lambda_k \ll \lambda_{k+1}$ then $h_k \ll h_{k+1}$, hence the graph can be more naturally cut into k parts than $(k+1)$ parts. Unless otherwise clarified, for the rest of this section we will use the variable $k \in [n]$ to represent the number of clusters for which we would like to find a clustering.

As with before, the lower bound on h_k is easier to prove than the upper bound. We will begin by tackling this direction.

4.1 Lower bound on the k-way expansion constant

Our approach will proceed along the lines of Section 3.1. We would like to obtain a higher-order analogue of Lemma 3.6, which states that for two orthogonal vectors f, g with $D^{1/2}f \perp D^{1/2}g$ and $\max\{\mathcal{R}(f), \mathcal{R}(g)\} = c$, we have $\mathcal{R}(f+g) \leq 2c$. We could make a similar argument to show that this generalises to k orthogonal vectors with upper bound kc , but we desire a loss of no more than a factor of 2. Thus we require the stronger condition that the vectors are *disjointly supported*, which gives us the following lemma:

Lemma 4.4. *Let f_1, f_2, \dots, f_k be disjointly supported vectors in \mathbb{R}^n . Then*

$$\mathcal{R} \left(\sum_{i \in [k]} f_i \right) \leq 2 \cdot \max_{i \in [k]} \mathcal{R}(f_i).$$

Proof. Let $\{x, y\} \in E$. Since the f_1, \dots, f_k are disjointly supported, there exists an index $a \in [k]$ such that for all $i \neq a$ we have $f_i(x) = 0$, and an index $b \in [k]$ such that for all $i \neq b$ we have $f_i(y) = 0$. In the case when $a = b$, we have

$$(f_a(x) - f_b(y))^2 = (f_a(x) - f_a(y))^2 = \sum_{i \in [k]} (f_i(x) - f_i(y))^2$$

and in the case where $a \neq b$ we have

$$\begin{aligned} (f_a(x) - f_b(y))^2 &\leq 2(f_a(x)^2 + f_b(y)^2) \\ &= 2(f_a(x) - f_a(y))^2 + 2(f_b(x) - f_b(y))^2 \quad (\text{using } f_a(y) = f_b(x) = 0) \\ &= 2 \sum_{i \in [k]} (f_i(x) - f_i(y))^2. \end{aligned}$$

Thus putting both cases together we have

$$\left(\sum_{i \in [k]} f_i(x) - f_i(y) \right)^2 = (f_a(x) - f_b(y))^2 \leq 2 \sum_{i \in [k]} (f_i(x) - f_i(y))^2. \quad (4)$$

To finish, note that disjoint support implies that for a fixed $v \in V$ we have

$$\left(\sum_{i \in [k]} f_i(v) \right)^2 = \sum_{i \in [k]} f_i(v)^2.$$

Combined with (4), this yields

$$\begin{aligned} \mathcal{R} \left(\sum_{i \in [k]} f_i \right) &= \frac{\sum_{x \sim y} w(x, y) \left[\sum_{i \in [k]} f_i(x) - f_i(y) \right]^2}{\sum_{v \in V} w(v) \left(\sum_{i \in [k]} f_i(v) \right)^2} \\ &\leq 2 \cdot \frac{\sum_{i \in [k]} \sum_{x \sim y} w(x, y) (f_i(x) - f_i(y))^2}{\sum_{v \in V} \sum_{i \in [k]} w(v) f_i(v)^2} \\ &\leq 2 \cdot \max_{i \in [k]} \mathcal{R}(f_i). \end{aligned}$$

□

Now we may complete the proof of $\frac{\lambda_k}{2} \leq h_k$, in much the same fashion as we did in the 2-clustering case with Remark 3.5.

Proof of lower bound in Theorem 4.3. Let S_1, S_2, \dots, S_k be a k -clustering of the graph which achieves h_k , so that

$$h_k = \max_{i \in [k]} \phi(S_i).$$

Theorem 2.14 tells us that

$$\lambda_k = \min_{U: \dim U = k} \left(\max_{\mathbf{0} \neq f \in U} \mathcal{R}(f) \right)$$

where the minimum is taken over all k -dimensional subspaces U of \mathbb{R}^n . Now pick $U = \text{span}\{\mathbf{1}_{S_1}, \mathbf{1}_{S_2}, \dots, \mathbf{1}_{S_k}\}$, noting that these vectors are disjointly supported and $\dim U = k$. Then by Lemma 4.4 and Lemma 3.2 we have

$$\lambda_k \leq \max_{\mathbf{0} \neq f \in U} \mathcal{R}(f) \leq 2 \cdot \max_{i \in [k]} \mathcal{R}(\mathbf{1}_{S_i}) = 2 \cdot \max_{i \in [k]} \phi(S_i) = 2h_k.$$

□

4.2 Upper bound on the k -way expansion constant

Recall our strategy for proving the upper bound of Cheeger's inequality: begin with the eigenvector f_2 corresponding to λ_2 , and 'round' it to produce a new vector which can be used to bisect the vertices, yielding a cut of good quality. Clearly, bisection will only allow us to find clusterings consisting of two satisfactory sets, but to prove the higher order Cheeger inequalities we now need to find k such sets. Hence we cannot simply repeat our argument with f_k instead of f_2 as one may hope.

Instead, to get enough information to form a k -clustering we must look at the bottom k eigenvectors together. Thus we arrive at the following definition:

Definition 4.5 (Spectral embedding). Let g_1, g_2, \dots, g_n be an orthonormal set of eigenvectors of \mathcal{L} which correspond to the eigenvalues $\lambda_1, \dots, \lambda_n$. For $i \in [n]$, write $f_i = D^{-1/2}g_i$. Then the *Spectral embedding* of G is the map $F : V \rightarrow \mathbb{R}^k$ defined by

$$F(v) := (f_1(v), f_2(v), \dots, f_k(v)).$$

Note that f_1 above is simply the constant vector (by Lemma 2.7).

Observe that F maps pairs of 'nearby vertices' in the graph (i.e vertices joined by an edge of high weight) to pairs of nearby points in \mathbb{R}^k , since the eigenvectors f_1, \dots, f_k themselves behave similarly when mapping to \mathbb{R} (this can be demonstrated by Theorem 2.14, noting that the numerator of the Rayleigh quotient will be minimised for such pairs).

Briefly recall Lemma 3.8, which tells us that a vector $f \in \mathbb{R}^n$ with small Rayleigh quotient can be used to find a set $S \subseteq \text{supp}(f)$ with small expansion. Therefore, if we had k *disjointly supported* vectors $\psi_1, \dots, \psi_k \in \mathbb{R}^n$, each with small Rayleigh quotient, then we are guaranteed to be able to find k disjoint sets of low expansion, which is precisely what is needed to bound h_k . Therefore our general approach to proving the upper bound of Theorem 4.3 will be to start with the spectral embedding and 'round' it to produce k such functions with the properties that we want. This approach generalises the idea of Lemma 3.7.

To guarantee that each of our k functions ψ_i have small Rayleigh quotient, we must ensure that $\text{supp}(\psi_i)$ contains a large portion of the 'mass' of F (so that the denominator is large), and that the rounding process is smooth enough to minimize the numerator.

4.2.1 Random partitioning of metric spaces

In order to produce a satisfactory clustering, our first step is to perform random partitioning of the points in the spectral embedding.

Definition 4.6 (Partition of a metric space). Let (X, d) be a metric space. We can represent a *partition* P of (X, d) by a collection $P = \{S_1, S_2, \dots, S_m\}$ such that $\bigcup_{i=1}^m S_i = X$ and the S_i are pairwise disjoint.

Definition 4.7 (Δ -bounded). Let P be a partition of a finite metric space (X, d) . For $\Delta > 0$, we say that P is Δ -bounded if $\text{diam}(S) \leq \Delta$ for every $S \in P$. When talking about random partitions \mathcal{P} , we say that \mathcal{P} is Δ -bounded when it is Δ -bounded with probability one.

Definition 4.8 ((Δ, L) -Lipschitz). Let \mathcal{P} be a random partition of a finite metric space (X, d) . For $\Delta, L > 0$ we say that \mathcal{P} is (Δ, L) -Lipschitz when \mathcal{P} is Δ -bounded, and for each pair $x, y \in X$ we have

$$\mathbb{P}[\mathcal{P}(x) \neq \mathcal{P}(y)] \leq L \frac{d(x, y)}{\Delta}.$$

Intuitively, a partition is bounded when each of the parts are not too large, and a partition is Lipschitz when the probability of two points being in different parts can be bounded by their distance.

We now state (without proof) the main random partitioning theorem which we use to obtain a sufficiently good partition.

Lemma 4.9. *Let (X, d) be a finite metric space where $X \subseteq \mathbb{R}^k$. Then for all $\Delta > 0$, there exists a $(\Delta, O(\sqrt{k}))$ -Lipschitz random partition on (X, d) .*

Proof. A proof of this statement can be found in [10]. □

Definition 4.10 (Radial projection distance). Let $k \in [n]$ and $F : V \rightarrow \mathbb{R}^k$. For all $x, y \in V$, if $\|F(x)\|, \|F(y)\| > 0$ we define the *radial projection distance* by

$$d_F(x, y) := \left\| \frac{F(x)}{\|F(x)\|} - \frac{F(y)}{\|F(y)\|} \right\|.$$

If instead $\|F(x)\| = \|F(y)\| = 0$ we write $d_F(x, y) := 0$, otherwise $d_F(x, y) = 2$.

We pick $d_F(x, y) = 2$ to reflect the fact that the Euclidean distance between two vectors on the unit ball is at most 2.

4.2.2 Spreading properties of the embedding

In order for the random partitioning to give a good quality result, we must first show that F spreads the points sufficiently in \mathbb{R}^k . Figure 2 shows an interpretation of this for the $k = 3$ case.

Definition 4.11. Let F be the spectral embedding as in Definition 4.5. Then we define the *F-mass* of a subset $S \subseteq V$ by

$$\mathcal{M}_F(S) = \sum_{v \in S} w(v) \|F(v)\|^2.$$

Definition 4.12 ((Δ, η) -spreading). Let $k \in [n]$ and $F : V \rightarrow \mathbb{R}^k$. We say that F is (Δ, η) -spreading if for all subsets $S \subseteq V$ we have

$$\text{diam}(S, d_F) \leq \Delta \implies \mathcal{M}_F(S) \leq \eta \cdot \mathcal{M}_F(V).$$

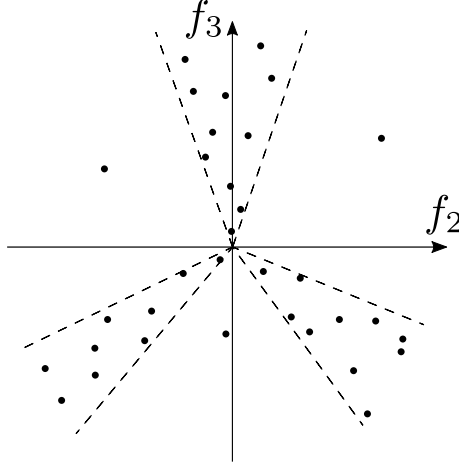


Figure 2: F spreading the vertices for clustering.

Lemma 4.13. *Let $k \in [n]$ and $F : V \rightarrow \mathbb{R}^k$. For all $x, y \in V$ we have*

$$d_F(x, y) \|F(x)\| \leq 2 \|F(x) - F(y)\|.$$

Proof. For any non-zero pair of vectors $x, y \in \mathbb{R}^k$, by the triangle inequality we have

$$\|x\| \left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\| = \left\| x - \frac{\|x\|}{\|y\|} y \right\| \leq \|x - y\| + \left\| y - \frac{\|x\|}{\|y\|} y \right\| \leq 2 \|x - y\|.$$

Thus if $u, v \in V$ with $\|F(u)\|, \|F(v)\| > 0$, then letting $x = F(u)$ and $y = F(v)$ above completes the proof. Otherwise when $\|F(u)\| = 0$ or $\|F(v)\| = 0$, the inequality is trivial. \square

We now briefly show that the f_1, \dots, f_k are orthonormal with respect to a particular inner product on \mathbb{R}^n which we make use of:

Lemma 4.14. *Let $F(v) = (f_1(v), \dots, f_k(v))$ be the spectral embedding of G . Then for any $i, j \in [k]$ we have*

$$\sum_{v \in V} w(v) f_i(v) f_j(v) = \delta_{i,j}$$

where $\delta_{i,j}$ is the Kronecker delta function.

Proof. As in Definition 4.5, let $F(v) = (f_1(v), \dots, f_k(v)) = (D^{-1/2} g_1(v), \dots, D^{-1/2} g_k(v))$. Then we have

$$\begin{aligned} \sum_{v \in V} w(v) f_i(v) f_j(v) &= \sum_{v \in V} \sqrt{w(v)} f_i(v) \sqrt{w(v)} f_j(v) \\ &= \sum_{v \in V} (D^{1/2} f_i)(v) (D^{1/2} f_j)(v) \\ &= \sum_{v \in V} g_i(v) g_j(v) = \delta_{i,j} \end{aligned}$$

where the last equality follows from the orthonormality of the $g_1, \dots, g_k \in \mathbb{R}^n$. \square

The next theorem tells us that F spreads the points isotropically in the space.

Lemma 4.15. *Let $F(v) = (f_1(v), \dots, f_k(v))$ be the spectral embedding of G . Then for any unit vector $\hat{x} \in \mathbb{R}^k$ we have*

$$\sum_{v \in V} w(v) \langle \hat{x}, F(v) \rangle = 1.$$

Proof. For any $x \in \mathbb{R}^k$, define $U : \mathbb{R}^k \rightarrow \mathbb{R}^n$ by

$$(Ux)(v) := \left\langle x, \sqrt{w(v)} F(v) \right\rangle.$$

One can verify that the i th column of U (represented as a matrix in the standard basis) is the vector $(\sqrt{w(v)} f_i(v))_{v \in V}$. Now since the value $(U^T U)_{i,j}$ is the inner product of the i th and j th columns of U , we have

$$(U^T U)_{i,j} = \sum_{v \in V} w(v) f_i(v) f_j(v) = \delta_{i,j}$$

where the last equality follows from Lemma 4.14. Therefore $U^T U = I$. Now let $\hat{x} \in \mathbb{R}^k$ be a unit vector. Then by basic properties of the inner product, we have

$$\sum_{v \in V} w(v) \langle \hat{x}, F(v) \rangle^2 = \sum_{v \in V} (U\hat{x})(v)^2 = \langle U\hat{x}, U\hat{x} \rangle = \langle \hat{x}, U^T U \hat{x} \rangle = \langle \hat{x}, \hat{x} \rangle = 1.$$

□

Finally we arrive at the main spreading theorem.

Lemma 4.16. *Let $F : V \rightarrow \mathbb{R}^k$ be the spectral embedding of G . Then for all $\Delta > 0$, F is $(\Delta, 1/k(1 - \Delta^2))$ -spreading.*

Proof. By rearrangement, for any $x, y \in V$ we have

$$d_F(x, y)^2 = \left\| \frac{F(x)}{\|F(x)\|} \right\|^2 + \left\| \frac{F(y)}{\|F(y)\|} \right\|^2 - 2 \left\langle \frac{F(x)}{\|F(x)\|}, \frac{F(y)}{\|F(y)\|} \right\rangle$$

and thus

$$\left\langle \frac{F(x)}{\|F(x)\|}, \frac{F(y)}{\|F(y)\|} \right\rangle = \frac{1 + 1 - d_F(x, y)^2}{2} = 1 - \frac{d_F(x, y)^2}{2}. \quad (5)$$

Now suppose that we have $S \subseteq V$ such that $\text{diam}(S, d_F) \leq \Delta$ and pick $s \in S$. Applying Lemma 4.15 to the unit vector $F(s)/\|F(s)\|$, and then applying (5) yields

$$\begin{aligned} 1 &= \sum_{v \in V} w(v) \left\langle \frac{F(s)}{\|F(s)\|}, F(v) \right\rangle^2 \\ &= \sum_{v \in V} w(v) \|F(v)\|^2 \left\langle \frac{F(s)}{\|F(s)\|}, \frac{F(v)}{\|F(v)\|} \right\rangle^2 \\ &= \mathcal{M}_F(V) \cdot \left(1 - \frac{d_F(s, v)^2}{2} \right)^2 \\ &\geq \mathcal{M}_F(S) \cdot \left(1 - \frac{\Delta^2}{2} \right)^2 \geq (1 - \Delta^2)^2 \mathcal{M}_F(S) \end{aligned}$$

where in the final line we have used the fact that for all $v \in S$ we have $d_F(s, v) \leq \text{diam}(S, d_F) \leq \Delta$ by assumption.

Finally, by Lemma 4.14 we have

$$\mathcal{M}_F(V) = \sum_{v \in V} w(v) \|F(v)\|^2 = \sum_{v \in V} \sum_{i=1}^k w(v) f_i(v)^2 = \sum_{i=1}^k \delta_{i,i} = k$$

which combines with the above to give

$$(1 - \Delta^2) \mathcal{M}_F(S) \leq \frac{1}{k} \mathcal{M}_F(V).$$

□

4.2.3 Refining the random partition

For the purposes of this section we will quickly fix some notation: let $M = \max\{\|F(v)\|^2 : v \in V\}$ and pick $\tau \in (0, M)$ uniformly at random. For any $S \subseteq V$, we define

$$\hat{S} := \{v \in S : \|F(v)\|^2 \geq \tau\}.$$

We also extend the notion of Rayleigh quotients to arbitrary maps $F : V \rightarrow \mathbb{R}^k$ by writing

$$\mathcal{R}(F) = \frac{\sum_{x \sim y} w(x, y) \|F(x) - F(y)\|^2}{\sum_{v \in V} w(v) \|F(v)\|^2}.$$

Remark 4.17. From the above, we can see that if F is the k -dimensional spectral embedding then $\mathcal{R}(F) \leq \lambda_k$.

The following Lemma is a variant of Lemma 3.8, with stronger requirements to harness the quality of our random partition.

Lemma 4.18. *For all $\Delta > 0$, there exists $r \in [n]$ and a partition $V = S_1 \cup S_2 \cup \dots \cup S_r$ such that for every $i \in [r]$ we have $\text{diam}(S_i, d_F) \leq \Delta$ and*

$$\frac{\mathbb{E} \left[\sum_{i=1}^r w(E(\hat{S}_i, \hat{S}_i^c)) \right]}{\mathbb{E} \left[\sum_{i=1}^r w(\hat{S}_i) \right]} \lesssim \frac{\sqrt{k}}{\Delta} \sqrt{\mathcal{R}(F)}. \quad (6)$$

Proof. By Theorem 4.9 there exists a Δ -bounded random partition \mathcal{P} such that for all $x, y \in V$ we have

$$\mathbb{P}[\mathcal{P}(x) \neq \mathcal{P}(y)] \lesssim \sqrt{k} \cdot \frac{d_F(x, y)}{\Delta}. \quad (7)$$

Write $\mathcal{P} = S_1 \cup S_2 \dots \cup S_r$, noting that r is a random variable. We will assume that $M = 1$ since the Rayleigh quotient is scale-invariant and we may simply scale the input by a constant to achieve the desired (which now implies that $\tau \in (0, 1)$). Then for each i , observe that

$$\mathbb{E} [w(\hat{S}_i)] = \sum_{v \in S_i} w(v) \mathbb{P}[v \in \hat{S}_i] = \sum_{v \in S_i} w(v) \mathbb{P}[\|F(v)\|^2 \geq \tau] = \sum_{v \in S_i} w(v) \|F(v)\|^2 \quad (8)$$

which implies that

$$\mathbb{E} \left[\sum_{i=1}^m w(\hat{S}_i) \right] = \sum_{i=1}^m \mathbb{E} [w(\hat{S}_i)] = \sum_{i=1}^m \sum_{v \in S_i} w(v) \|F(v)\|^2 = \sum_{v \in V} w(v) \|F(v)\|^2. \quad (9)$$

Now suppose that we have $\{x, y\} \in E$. We seek to calculate the probability that the edge $\{x, y\}$ is on the boundary of some \hat{S}_i , i.e

$$\mathbb{P} \left[\{x, y\} \in \bigcup_{i=1}^m E(\hat{S}_i, (\hat{S}_i)^c) \right].$$

Without loss of generality we can take $\|F(x)\|^2 \leq \|F(y)\|^2$. We first consider the case where $\mathcal{P}(x) = \mathcal{P}(y) = S_i$ for some i . Then we have

$$\mathbb{P} \left[\{x, y\} \in E(\hat{S}_i, (\hat{S}_i)^c) \right] = \mathbb{P} [\tau \in [\|F(x)\|^2, \|F(y)\|^2]] .$$

In the other case where $\mathcal{P}(x) \neq \mathcal{P}(y)$, we instead have that

$$\mathbb{P} \left[\{x, y\} \in \bigcup_{i=1}^m E(\hat{S}_i, (\hat{S}_i)^c) \right] = \mathbb{P} [x \in \cup_i \hat{S}_i \text{ or } y \in \cup_i \hat{S}_i] = \mathbb{P} [\|F(x)\|^2 \geq \tau \text{ or } \|F(y)\|^2 \geq \tau] .$$

Putting the two cases together yields

$$\begin{aligned} \mathbb{P} \left[\{x, y\} \in \bigcup_{i=1}^m E(\hat{S}_i, (\hat{S}_i)^c) \right] &= \mathbb{P} [\mathcal{P}(x) \neq \mathcal{P}(y)] \cdot \mathbb{P} [\|F(x)\|^2 \geq \tau \text{ or } \|F(y)\|^2 \geq \tau \mid \mathcal{P}(x) \neq \mathcal{P}(y)] \\ &\quad + \mathbb{P} [\mathcal{P}(x) = \mathcal{P}(y)] \cdot \mathbb{P} [\tau \in [\|F(x)\|^2, \|F(y)\|^2] \mid \mathcal{P}(x) = \mathcal{P}(y)] \\ &\leq \mathbb{P} [\mathcal{P}(x) \neq \mathcal{P}(y)] \cdot \mathbb{P} [\|F(x)\|^2 \geq \tau \text{ or } \|F(y)\|^2 \geq \tau \mid \mathcal{P}(x) \neq \mathcal{P}(y)] \\ &\quad + \mathbb{P} [\tau \in [\|F(x)\|^2, \|F(y)\|^2] \mid \mathcal{P}(x) = \mathcal{P}(y)] . \end{aligned}$$

Next, writing $\alpha := \|F(x)\|$ and $\beta := \|F(y)\|$ for convenience, we have by (7) that

$$\begin{aligned} \mathbb{P} \left[\{x, y\} \in \bigcup_{i=1}^m E(\hat{S}_i, (\hat{S}_i)^c) \right] &\lesssim \sqrt{k} \cdot \frac{d_F(x, y)}{\Delta} (\alpha^2 + \beta^2) + \alpha^2 - \beta^2 \\ &\leq \sqrt{k} \cdot \frac{d_F(x, y)}{\Delta} (\alpha + \beta)^2 + (\alpha + \beta)(\alpha - \beta) \\ &\leq (a + b) \left[\sqrt{k} \cdot \frac{d_F(x, y)}{\Delta} (\alpha + \beta) + \alpha - \beta \right] \\ &\leq (a + b) \left[\frac{4\sqrt{k}}{\Delta} \|F(x) - F(y)\| + \|F(x) - F(y)\| \right] \\ &\leq \frac{5\sqrt{k}}{\Delta} \|F(x) - F(y)\| (\|F(x)\| + \|F(y)\|) \end{aligned}$$

where on the penultimate line we have applied Lemma 4.13 twice, and used the reverse triangle inequality. Therefore by Cauchy-Schwarz we can write

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^m w(\hat{S}_i, (\hat{S}_i)^c) \right] &= \sum_{x \sim y} w(x, y) \cdot \mathbb{P} \left[\{x, y\} \in \bigcup_{i=1}^m E(\hat{S}_i, (\hat{S}_i)^c) \right] \\ &\lesssim \frac{\sqrt{k}}{\Delta} \sum_{x \sim y} w(x, y) \|F(x) - F(y)\| (\|F(x)\| + \|F(y)\|) \\ &\leq \frac{\sqrt{k}}{\Delta} \sqrt{\sum_{x \sim y} w(x, y) (\|F(x)\| + \|F(y)\|)^2} \sqrt{\sum_{x \sim y} w(x, y) \|F(x) - F(y)\|^2} \end{aligned}$$

$$\leq \frac{\sqrt{k}}{\Delta} \sqrt{\sum_{v \in V} w(v) \|F(v)\|^2} \sqrt{\sum_{x \sim y} w(x, y) \|F(x) - F(y)\|^2}.$$

So combining the above with (9) gives

$$\begin{aligned} \frac{\mathbb{E}_{\mathcal{P}} \mathbb{E} \left[\sum_{i=1}^m w(E(\hat{S}, (\hat{S})^c)) \right]}{\mathbb{E}_{\mathcal{P}} \mathbb{E} \left[\sum_{i=1}^m w(\hat{S}_i) \right]} &\lesssim \frac{\sqrt{k}}{\Delta} \frac{\sqrt{\sum_{x \sim y} w(x, y) \|F(x) - F(y)\|^2}}{\sqrt{\sum_{v \in V} w(v) \|F(v)\|^2}} \\ &\lesssim \frac{\sqrt{k}}{\Delta} \sqrt{\mathcal{R}(F)} \end{aligned}$$

where $\mathbb{E}_{\mathcal{P}}$ denotes the expectation over our random choice of \mathcal{P} . This implies that there exists *some* partition in the support of the distribution for which the statement holds. \square

Note that the above Lemma may produce many small non-expanding sets. To fix this, the final step is to merge these small sets together.

Lemma 4.19. *Let $F : V \rightarrow \mathbb{R}^k$. If F is $(\Delta, \frac{1}{k} + \frac{\delta}{4k})$ -spreading, then there exist $r \geq \lceil (1 - \delta)k \rceil$ disjoint sets $S_1, \dots, S_r \subseteq V$ such that for every $i \in [r]$ we have*

$$\phi(S_i) \lesssim \frac{\sqrt{k}}{\delta \Delta} \sqrt{\mathcal{R}(F)}.$$

Proof. Let $V = T_1 \cup T_2 \cup \dots \cup T_m$ be the partition resulting from applying Lemma 4.18 to F . Using the spreading properties of F , we will merge these to obtain the desired sets S_1, S_2, \dots, S_r . Quickly note that since $\lceil (1 - \delta)k \rceil \leq k$, we can assume that

$$\frac{1}{k} \leq \delta + \frac{\delta^2}{2}. \quad (10)$$

First, write $\mathcal{M}_{\text{tot}} = \mathcal{M}_F(V)$. Note that by the definition of F being $(\Delta, \frac{1}{k} + \frac{\delta}{4k})$ -spreading, and that each of our T_i satisfy $\text{diam}(T_i, d_F) \leq \Delta$, we have

$$\mathcal{M}_F(T_i) \leq \frac{\mathcal{M}_{\text{tot}}}{k} \left(1 + \frac{\delta}{4} \right) \quad (11)$$

Next, apply Algorithm 2 to $T = \{T_1, T_2, \dots, T_m\}$ to produce sets $S = \{S_1, S_2, \dots, S_r\}$. Informally, we repeatedly merge the two sets $T_a, T_b \in T$ with smallest $\mathcal{M}_F(\cdot)$ subject to the constraint that $\mathcal{M}_F(T_a) + \mathcal{M}_F(T_b) \leq \frac{\mathcal{M}_{\text{tot}}}{k} \left(1 + \frac{\delta}{4} \right)$ (see line 5 of Algorithm 2). Once no more merging can take place, we exclude any sets T_i for which $\mathcal{M}_F(T_i) \leq \mathcal{M}_{\text{tot}}/4k$ and output the rest as $S = \{S_1, \dots, S_{r'}\}$.

So the end of the process, by 11 and our merging constraint, we will have sets $S = \{S_1, \dots, S_{r'}\}$ which satisfy

$$\frac{\mathcal{M}_{\text{tot}}}{4k} \leq \mathcal{M}_F(S_i) \leq \frac{\mathcal{M}_{\text{tot}}}{k} \left(1 + \frac{\delta}{4} \right). \quad (12)$$

Let r' be the number of sets which result from this process. Observe that the final step may exclude *at most* one set S_i for which $\mathcal{M}_F(S_i) \leq \mathcal{M}_{\text{tot}}/4k$, since we will have merged

any two such sets (one can verify that $2 \cdot \mathcal{M}_{\text{tot}}/4k \leq \frac{\mathcal{M}_{\text{tot}}}{k} (1 + \frac{\delta}{4})$). The remaining sets will then satisfy

$$\sum_{i=1}^{r'} \mathcal{M}_F(S_i) > \mathcal{M}_{\text{tot}} - \frac{\mathcal{M}_{\text{tot}}}{4k}. \quad (13)$$

Since the total mass of the r' sets is at least $\mathcal{M}_{\text{tot}}(1 - 1/4k)$, and each set has mass at most $\frac{\mathcal{M}_{\text{tot}}}{k} (1 + \frac{\delta}{4})$, we can divide these values to conclude that r' satisfies

$$r' > k \cdot \frac{1 - 1/4k}{1 + \delta/4} \geq k \cdot \frac{1 - \frac{1}{4}(\delta + \delta^2/2)}{1 + \delta/4} = k \cdot \frac{(1 + \delta/4)(1 - \delta/2)}{1 + \delta/4} = k(1 - \delta/2) \quad (14)$$

where we have used (10).

Algorithm 2: Iterative Merging

```

1 Function merge( $T$ )
   Input:  $T = \{T_1, T_2, \dots, T_m\}$ ,  $m \in \{2, \dots, k\}$ 
   Output:  $S = \{S_1, S_2, \dots, S_r\}$ ,  $r \in \{2, \dots, m\}$ 
2   while true do
3      $a \leftarrow \text{argmin}_i (\mathcal{M}_F(T_i) : T_i \in T)$ 
4      $b \leftarrow \text{argmin}_i (\mathcal{M}_F(T_i) : T_i \in T \setminus \{T_a\})$ 
5     if  $\mathcal{M}_F(T_a) + \mathcal{M}_F(T_b) \leq \frac{\mathcal{M}_{\text{tot}}}{k} (1 + \frac{\delta}{4})$  then
6        $T' \leftarrow T \setminus \{T_a\} \setminus \{T_b\}$ 
7        $T \leftarrow T' \cup \{T_a \cup T_b\}$ 
8     else
9       return  $S = \{\tau : \mathcal{M}_F(\tau) \geq \mathcal{M}_{\text{tot}}/4k, \tau \in T\}$ 
10    end
11  end
12 end

```

Putting the above argument to one side, we now proceed to handle the denominator of the Rayleigh quotient. Write $r = \lceil (1 - \delta)k \rceil$ so that $r \leq (1 - \delta/2)k \leq r'$ by (14). Reorder the sets so that $\mathbb{E} \left[w(E(\hat{S}_i, (\hat{S}_i)^c)) \right] \leq \mathbb{E} \left[w(E(\hat{S}_{i+1}, (\hat{S}_{i+1})^c)) \right]$ for each $i \in [r' - 1]$. We are left with r ‘small’ sets and $r' - r$ ‘big’ sets. For the $i \in \{r + 1, \dots, r' - 1\}$ of the big sets, we will have

$$(r' - r) \mathbb{E} \left[w(E(\hat{S}_i, (\hat{S}_i)^c)) \right] \leq \mathbb{E} \left[\sum_{j=1}^m w(E(\hat{T}_j, (\hat{T}_j)^c)) \right]$$

since after merging, the amount of weight leaving the sets must be lower. Then all of the smaller sets $i \in [r]$ must satisfy

$$\mathbb{E} \left[w(E(\hat{S}_i, (\hat{S}_i)^c)) \right] \lesssim \frac{1}{r' - r} \mathbb{E} \left[\sum_{j=1}^m w(E(\hat{T}_j, (\hat{T}_j)^c)) \right] \lesssim \frac{1}{\delta k} \mathbb{E} \left[\sum_{j=1}^m w(E(\hat{T}_j, (\hat{T}_j)^c)) \right]$$

But since each T_i satisfies (6), we then have

$$\frac{1}{\delta k} \mathbb{E} \left[\sum_{j=1}^m w(E(\hat{T}_j, (\hat{T}_j)^c)) \right] \lesssim \frac{\sqrt{k}}{\delta k \cdot \Delta} \cdot \sqrt{\mathcal{R}(F)} \cdot \mathbb{E} \left[\sum_{j=1}^m w(\hat{T}_j) \right]$$

$$\lesssim \frac{\sqrt{k}}{\delta k \cdot \Delta} \cdot \sqrt{\mathcal{R}(F)} \cdot \mathcal{M}_{\text{tot}}.$$

Finally, noting that $\mathbb{E}[w(\hat{S}_i)] = \mathcal{M}_F(S_i)$ by (8), and $\mathcal{M}_F(S_i) \asymp \mathcal{M}_{\text{tot}}/k$ by (12), we then have

$$\frac{\mathbb{E}[w(E(\hat{S}_i, (\hat{S}_i)^c))]}{\mathbb{E}[w(\hat{S}_i)]} \asymp \frac{\mathbb{E}[w(E(\hat{S}_i, (\hat{S}_i)^c))]}{\mathcal{M}_{\text{tot}}/k} \lesssim \frac{\sqrt{k}}{\delta \Delta} \cdot \sqrt{\mathcal{R}(F)}.$$

Thus, recalling the definition of \hat{S} , we can conclude that there exists a suitable choice of τ such that the set $S_i^* = \{v \in S_i : \|F(v)\|^2 \geq \tau\}$ satisfies

$$\frac{w(E(S_i^*, (S_i^*)^c))}{w(S_i^*)} = \phi(S_i^*) \lesssim \frac{\sqrt{k}}{\delta \Delta} \cdot \sqrt{\mathcal{R}(F)}.$$

□

Theorem 4.20. *For any $\delta \in (0, 1)$, there exist $r \geq \lceil (1 - \delta)k \rceil$ disjoint, non-empty sets $S_1, \dots, S_r \subseteq V$ such that*

$$\phi(S_i) \lesssim \frac{\sqrt{k}}{\delta^{3/2}} \sqrt{\lambda_k}.$$

Proof. Pick $\Delta \asymp \sqrt{\delta}$ such that $1/(1 - \Delta^2) \leq 1 + \delta/4$. Lemma 4.16 tells us that the spectral embedding F is $(\Delta, \frac{1}{k(1 - \Delta^2)})$ -spreading. But our choice of Δ implies

$$\frac{1}{k(1 - \Delta^2)} \leq \frac{1}{k}(1 + \delta/4) \leq \frac{1}{k} + \frac{\delta}{4k}$$

and hence F is also $(\Delta, \frac{1}{k} + \frac{\delta}{4k})$ -spreading. Applying Lemma 4.19 followed by Remark 4.17 immediately gives us the desired result. □

Now we may finally complete the proof of Theorem 4.3.

Proof of Theorem 4.3. Let $\delta = 1/2k$. Then by Theorem 4.20 there exist

$$r \geq \lceil (1 - 1/2k)k \rceil = k$$

disjoint, non-empty sets $S_1, \dots, S_r \subseteq V$ such that

$$\phi(S_i) \lesssim \frac{\sqrt{k}}{(1/2k)^{3/2}} \sqrt{\lambda_k} = k^2 \sqrt{\lambda_k}.$$

But then

$$h_k = \min_{T_1, \dots, T_k} \max\{\phi(T_1), \dots, \phi(T_k)\} \leq \max\{\phi(S_1), \dots, \phi(S_k)\} = O(k^2) \sqrt{\lambda_k}.$$

The lower bound on h_k was shown in Section 4.1. □

4.3 Further discussion

In this section, we will state (without proof) some other results from [7] and discuss their importance.

We have seen that Theorem 4.3 gives us a loss of $O(k^2)$ in the upper bound (which arises during the random partitioning step). It is natural to ask whether this asymptotic bound is tight or not. The following theorem allows us to improve this bound to $O(\sqrt{\log k})$ for arbitrary graphs, provided that one is instead comparing against λ_{2k} :

Theorem 4.21 (Theorem 1.2 of [7]). *For any $k \leq n/2$, we have*

$$h_k \leq O(\sqrt{\log k})\sqrt{\lambda_{2k}}.$$

We emphasise again that we are now comparing h_k to λ_{2k} , rather than λ_k . In practice this means that the spectral partitioning algorithm performs well for large k given that λ_{2k} is small too. The proof of this theorem continues roughly along the lines of Theorem 4.3, but with an added step: points of the spectral embedding in \mathbb{R}^k are randomly projected into $\mathbb{R}^{O(\log k)}$ in a way which preserves Rayleigh quotient and spreading properties. Random partitioning is then performed in this lower dimensional space, yielding a bound with smaller loss.

Further to this, the authors provide an example of a graph for which Theorem 4.21 is tight, implying that the bound cannot be improved in the general case:

Definition 4.22 (Noisy Hypercube). For any $k \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, let $H_{k,\varepsilon} = (V, E, w)$ be the graph defined as follows: Let $V = \{0, 1\}^k$ so that $n = |V| = 2^k$, and for any $x, y \in V$ we put $w(x, y) = \varepsilon^{\|x-y\|_1}$ where $\|\cdot\|_1$ denotes the L_1 norm.

Theorem 4.23 (Theorem 4.12 of [7]). *For a noisy hypercube $H_{k,\varepsilon} = (V, E, w)$ the following result holds: For any $S \subseteq V$ with $|S| \leq n/k$, we have*

$$\phi_{H_{k,\varepsilon}}(S) \gtrsim \sqrt{\lambda_k \log k}$$

This implies that Theorem 4.21 cannot be improved to λ_k in the general case.

The authors also obtain a bound with *no* dependence on k (for small values of k), subject to some further conditions on the spectrum:

Theorem 4.24 (Theorem 1.3 of [7]). *There is a constant $C > 0$ such that for any $k \leq n/4$, the following holds: If $\lambda_{4k} \geq C(\log k)^2 \lambda_{2k}$, then*

$$h_k \leq O(\sqrt{\lambda_{2k}}).$$

Finally, [11] proves the following interesting connection between λ_2 and λ_k , which is essentially an improved version of Cheeger's inequality:

Theorem 4.25. *For any $k \in [n]$ we have*

$$h_2 \leq \text{fiedler}(f_2) = O(k) \frac{\lambda_2}{\sqrt{\lambda_k}}$$

where f_2 is the eigenvector of \mathcal{L} corresponding to λ_2 .

The implication of this is that Fiedler's algorithm yields a result to within a factor of $O(k/\sqrt{\lambda_k})$ of the optimal solution h_2 , despite not using any information about eigenvalues beyond λ_2 in its execution.

5 Inverse Spectral Theory

In the previous sections, we have analysed graphs through their spectra. Now we will pivot to a related problem: is it possible to construct a graph with an arbitrary spectrum?

We already know that it is not possible for a graph to have negative eigenvalues (see Corollary 2.9) or have $\lambda_1 \neq 0$ (see Lemma 2.7), so we must rule these inputs out. We also know that the case of constructing a disconnected graph can be reduced to the construction of multiple connected graphs (see Lemma 2.18), so we will require $\lambda_2 > 0$ as per Theorem 2.19. First we will look at the problem of constructing L with a given set of eigenvalues, and then we will show that it is not possible to do this for \mathcal{L} in the general case.

The ideas for this for this section are taken from [12].

5.1 Unnormalised Laplacian

We begin with some notation:

Definition 5.1 (Unnormalised Spectrum). Let G be given. Denote the (multiset of) eigenvalues of L_G by $\Lambda(G)$.

Definition 5.2 (Suspension). Let $G = (V, E, w)$ and $c > 0$ be given. Then the *suspension* of G , written $S_G(c)$, is defined to be the graph on $|V| + 1$ vertices formed from G by adding a single vertex and joining it to all other vertices with weight c on each of these edges.

It is easy to see that we can form $L_{S_G(c)}$ from the matrix $L_G + cI$ simply by appending one row and one column. The new row and column both consist of elements of value $-c$, except for the diagonal element which has value $c|V|$. In other words, with a slight abuse of block matrix notation we can write

$$L_{S_G(c)} = \begin{pmatrix} \ddots & & & -c \\ & L_G + cI & & \vdots \\ & & \ddots & -c \\ -c & \dots & -c & c|V| \end{pmatrix}. \quad (15)$$

Next, we will show the main correspondence between eigenvalues of G and eigenvalues of $S_G(c)$:

Lemma 5.3. *Let $G = (V, E, w)$ and $c > 0$ be given. Then $\lambda \in \Lambda(G)$ implies $\lambda + c \in \Lambda(S_G(c))$.*

Proof. Let $f \in \mathbb{R}^{|V|}$ be the eigenvector of L_G corresponding to λ , and write v_0 to denote the new vertex added when taking the suspension of G . Now define the vector $f_0 \in \mathbb{R}^{|V|+1}$ by

$$f_0(x) = \begin{cases} f(x) & \text{if } x \in V \\ 0 & \text{if } x = v_0. \end{cases}$$

Observe that from (15) we have

$$L_{S_G(c)} f_0(x) = \begin{cases} (L_G + cI)f(x) = (\lambda + c)f(x) & \text{if } x \in V \\ 0 & \text{if } x = v_0, \end{cases}$$

which implies that f_0 is an eigenvector of $L_{S_G(c)}$ (with corresponding eigenvalue $\lambda + c$). \square

Observe that in the above proof, all the eigenvectors of $L_{S_G(c)}$ that we found were distinct. Since there are $|V|$ of these, and $L_{S_G(c)}$ has dimension $|V| + 1$, there is another eigenvector to account for. This is reflected by the following lemma:

Lemma 5.4. *Let $G = (V, E, w)$ be given. For all $c > 0$ we have that $c(|V| + 1) \in \Lambda(S_G(c))$.*

Proof. We proceed by directly constructing an appropriate eigenvector $f \in \mathbb{R}^{|V|+1}$, given by

$$f(x) = \begin{cases} -1/|V| & \text{if } x \in V \\ 1 & \text{if } x = v_0. \end{cases}$$

Then simply by applying the definition of matrix-vector multiplication to our matrix (15) we can see that

$$\begin{aligned} L_{S_G(c)}f(x) &= L_{S_G(c)}(x, v_0)f(v_0) + \sum_{y \in V} L_{S_G(c)}(x, y)f(y) \\ &= \begin{cases} (-c)(1) + (-c)(-1/|V|) & \text{if } x \in V \\ (c|V|)(1) + (-c|V|)(-1/|V|) & \text{if } x = v_0, \end{cases} \\ &= \begin{cases} -c(|V| + 1)/|V| & \text{if } x \in V \\ c(|V| + 1) & \text{if } x = v_0, \end{cases} \\ &= c(|V| + 1)f(x). \end{aligned}$$

Thus f is an eigenvector of $L_{S_G(c)}$ with eigenvalue $c(|V| + 1)$. □

Theorem 5.5. *Let $G = (V, E, w)$ and $c > 0$ be given. Then*

$$\Lambda(S_G(c)) = \{\lambda + c : \lambda \in \Lambda(G)\} \cup \{c(|V| + 1)\}.$$

Proof. Immediate by our arguments of Lemma 5.3 and Lemma 5.4. We can be sure that we have correctly accounted for any eigenvalues repeated with multiplicity, since all the eigenvectors involved are orthogonal. □

Theorem 5.6. *For $n \geq 2$, let $0 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n$ be given. Then there exists a complete graph G such that $\Lambda(G) = \{0, \lambda_2, \lambda_3, \dots, \lambda_n\}$.*

Proof. We will provide a constructive proof through finite induction. This procedure is based on Algorithm 3.

Pick c such that $\lambda_2 = c(n + 1)$. If $n = 2$, then the desired result follows from applying Theorem 5.5 to the graph consisting of a single vertex. Otherwise suppose the statement holds for $n - 1$, then we will prove that it holds for n . By assumption, there exists a graph G' on $n - 1$ vertices exists with $\Lambda(G') = \{0, \lambda_3 - c, \dots, \lambda_n - c\}$. Then by Theorem 5.5 we have

$$\Lambda(S_{G'}(c)) = \{0, \lambda_2, \lambda_3, \dots, \lambda_n\}.$$

□

Algorithm 3: Inverse Spectral Construction

```
1 Function construct( $n, \Lambda$ )  
   Input:  $n \geq 1, \Lambda = [\lambda_2, \lambda_3, \dots, \lambda_n]$   
   Output: Graph  $G$  such that  $\Lambda(G) = \{0, \lambda_2, \dots, \lambda_n\}$   
2   if  $n = 1$  then  
3     return graph with single vertex  
4   else  
5      $c \leftarrow \lambda_2 / (n + 1)$   
6      $G' \leftarrow \text{construct}(n - 1, [\lambda_3 - c, \dots, \lambda_n - c])$   
7     return  $S_{G'}(c)$   
8   end  
9 end
```

5.2 Normalised Laplacian

We will provide a quick counterexample to demonstrate that this cannot be done for the normalised Laplacian in general. Let G be a connected graph on n vertices. Note that $\lambda_1 = 0$ by Theorem 2.19. Then we have

$$\sum_{i=2}^n \lambda_i = \text{trace}(\mathcal{L}) = n$$

which implies that $\lambda_2 \leq n/(n-1)$ and $\lambda_n \geq n/(n-1)$. Therefore this cannot be done for an arbitrary set of input eigenvalues in $(0, 2]$.

See [13, Appendix A] for an informal treatment of why such a mapping of eigenvalues from L to \mathcal{L} is hard even allowing for the restrictions on the domain to be overcome.

6 Spectral Graph Theory in Practice

6.1 Spectral clustering

The proof of Theorem 4.3 is constructive in that it provides an algorithm to achieve a good k -clustering. The following steps are performed:

1. Construct the spectral embedding F (Definition 4.5).
2. Randomly partition points of F to produce $k' \geq k$ non-expanding sets (Lemma 4.9).
3. Merge small sets in this partition to produce k clusters (Lemma 4.19).

If one were to implement this algorithm directly, the random partition may be of poor quality in practice. Instead it is common to implement k -means clustering (KMC) on the spectral embedding, as [7] suggests.

KMC takes as input a list of points in \mathbb{R}^m and returns a list of k centers (points in \mathbb{R}^m) so as to minimise the total squared distance between each point and its nearest center. This induces a clustering of the points (by identifying them with their nearest center). In our case, the input is the values of the k -dimensional spectral embedding, so that each point lives in \mathbb{R}^k and there are n of them (one for each vertex in the graph).

Many variants of KMC are known. We implement a simple algorithm often called ‘naive k -means’ (Algorithm 4). The ‘initialise’ step returns an initial list of k centers which are

then refined. One such initialisation method is to simply pick k points uniformly at random from the input. We instead choose to use a method called **k-means++** initialisation, which yields a provably near-optimal solution and performs better in practice [14].

Algorithm 4: Naive k -means clustering

```

1 Function k-means( $X, k$ )
   Input: Data  $X = \{f_1, \dots, f_n\} \subseteq \mathbb{R}^n$ ,  $k \in [n]$ 
   Output:  $k$  centers  $C = \{c_1, \dots, c_k\} \subseteq \mathbb{R}^n$ 
2    $C \leftarrow \text{initialise}(X, k)$ 
3    $C' \leftarrow \emptyset$ 
4   while  $C \neq C'$  do
5      $C' \leftarrow \emptyset$ 
6     for  $c \in C$  do
7        $P \leftarrow \{f \in X : c \text{ is nearest center in } C \text{ to } f\}$ 
8        $c' \leftarrow \text{mean of } P$ 
9        $C' \leftarrow C' \cup \{c'\}$ 
10    end
11  end
12  return k-means( $\{F(v) : v \in V\}$ )
13 end

```

Having described k -means clustering, we can give a practical spectral clustering algorithm (Algorithm 5).

Algorithm 5: Spectral clustering algorithm

```

1 Function spectral-cluster( $G, k$ )
   Input:  $G = (V, E, w)$ ,  $k \in [n]$ 
   Output:  $S = \{S_1, \dots, S_k\}$  such that  $\max_i \phi(S_i)$  is small
2    $L \leftarrow \text{compute Laplacian matrix of } G$ 
3    $\psi_1, \dots, \psi_k \leftarrow k \text{ bottom eigenvectors of } L$ 
4   Let  $F(v) = (\psi_1(v), \dots, \psi_k(v)) \in \mathbb{R}^k$ .
5   return k-means( $\{F(v) : v \in V\}, k$ )
6 end

```

6.1.1 Experiments

We implemented Algorithm 5 in the Python programming language, using the NetworkX library [15]. We show a few inputs and outputs in Figure 3, in which vertices are coloured according to their cluster assignment.

Note that even though in one of the examples we construct a graph whose *unnormalised* Laplacian L has particular eigenvalues, the corresponding normalised Laplacian matrix should not have a particularly different structure. The reason for this is that Cheeger's inequality works for L too (with a slight modification), in which we have

$$\frac{\mu_2}{2} \leq h_2 \leq \sqrt{2\mu_2 \cdot d_{\max}}$$

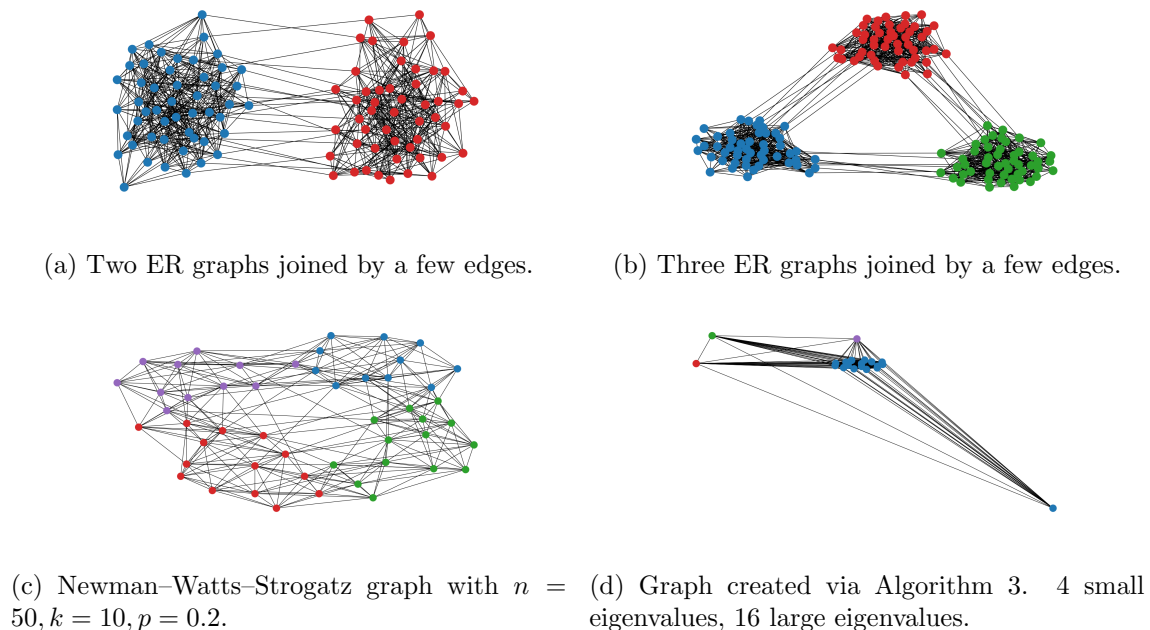


Figure 3: Examples of spectral clustering.

where μ_2 is the second smallest eigenvalue of L and d_{\max} is the maximum value of $w(v)$ over all vertices $v \in V$. Thus the qualitative connection between clusters in the graph and the eigenvalues of L remains.

6.1.2 Efficiency

The popularity of spectral clustering comes from its efficiency. While general linear systems of the form $Ax = b$ can currently be solved for a vector x in time $O(n^{2.37286})$ [16], it is possible to solve systems involving the Laplacian matrix much faster. Laplacian matrices are examples of *symmetric diagonally dominant* (SDD) matrices, in which each diagonal element is greater than or equal to the sum of the corresponding row. Spielman and Teng [17] provide an ε -approximate method to solve SDD systems with m non-zero entries in time

$$m \log^{O(1)}(m) + O\left(\left\lceil \log \frac{\lambda_n}{\varepsilon \lambda_2} \right\rceil \left[m + n 2^{O(\sqrt{\log n \log \log n})} \right]\right),$$

which (with an appropriate degree of eye-squinting) is said to be ‘nearly-linear time’. This can be used to produce a set of orthonormal eigenvectors as required by Algorithm 5. We refer the reader to [3] for a detailed exposition of fast Laplacian solvers.

6.2 Image segmentation

Spectral clustering has been used successfully in computer vision. The first such application was proposed by Shi and Malik [5] to perform image segmentation, in which the goal is to split an image up into multiple segments as would be perceived by a human. The basic idea is as follows:

First, compute a similarity graph $G = (V, E, w)$ where each vertex corresponds to one



Figure 4: Input (top), first and second eigenvectors (left and right). Reproduced from [5].

pixel. For $x, y \in V$ weights are given by

$$w(x, y) = \begin{cases} 0 & \text{if } \|X(i) - X(j)\| > r \\ \exp(-\|F(i) - F(j)\|^2) \cdot \exp(-\|X(i) - X(j)\|^2) & \text{otherwise} \end{cases} \quad (16)$$

where $F(i) \in \mathbb{R}$ and $X(i) \in \mathbb{R}^2$ give the brightness and cartesian coordinates of pixel i respectively, and $r \in \mathbb{R}$ is some particular threshold which helps to limit the amount of computation required. Then, spectral clustering is performed on this graph as usual to obtain k clusters of pixels.

Shi and Malik propose this scheme in [5]. They instead choose to compute eigenvectors of a similar matrix $\mathcal{L}' = I - D^{-1}W$, which differ to the eigenvectors of \mathcal{L} by a factor of $D^{1/2}$. An example input image and the corresponding eigenvectors of \mathcal{L}' are shown in Figure 4.

One may ask: if k -means clustering is used during the spectral clustering algorithm, why not apply k -means clustering directly with the distance function given by (16)? The answer, provided by Ng, Jordan and Weiss [18], is that k -means clustering performs poorly when clusters do not form convex regions in the space of the embedding. This is demonstrated by Figure 5, which shows a set of points in \mathbb{R}^2 being clustered either directly by k -means, or by constructing a graph in which each point is joined to each of its 10 nearest neighbours with weight proportional to the distance, and then performing spectral clustering.



Figure 5: Spectral clustering (left), k -means clustering (right).

6.3 VLSI design

More generally, graph clustering may be used to design efficient divide-and-conquer algorithms where the runtime of the ‘conquer’ step is highly dependent on the number of edges.

For example, spectral clustering has been used successfully in Very Large Scale Integration (VLSI) design. VLSI is the process by which huge numbers of transistors are combined to produce integrated circuits. Since many design tools and algorithms are not practical to be run on the millions or billions of components at once, *circuit partitioning* may be employed to decompose the overall design into a few components which can be analysed and optimised independently, with few wires between them. See [19] for a survey of such techniques, and [20] for a particular example of spectral clustering applied in this area.

Some modifications must be performed to model circuits effectively. Notably, the circuit is represented as a *hypergraph*, since a single wire may join multiple components. This is a particularly crucial fact to capture in the partitioning step, since the number (and weight) of edges leaving a cluster is the exact thing which graph clustering attempts to minimise.

6.4 Vertex centrality

One final application we study is the identification of important or ‘central’ vertices in a graph. For example, consider a graph representation of a social network. An intelligent advertiser may choose the most influential people in the network to promote a product, in the hopes that they will spread it to the largest possible number of their peers, and so on.

One obvious metric for this is the vertex degree, which for a vertex $x \in V$ is given by $w(x)$. However this is a poor estimate when the neighbours of x have low degree themselves, so we need to capture some global information about the position of the vertex in the graph beyond the degree.

Instead we may desire a metric for which x ’s importance is the sum of the importance of its neighbours (weighted by the value on the edge), e.g some vector $f \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$

such that

$$f(x) = \frac{1}{\lambda} \sum_{y \sim x} w(x, y) f(y) = \frac{1}{\lambda} \sum_{y \in V} W_{xy} \cdot f(y) \quad (17)$$

which can be rewritten as the eigenvector problem $Wf = \lambda f$. The Perron-Frobenius theorem tells us that f corresponds to the unique greatest real eigenvalue of W , and moreover that f has strictly positive entries.

This eigenvector can be computed efficiently using the power iteration method (Algorithm 6). At each iteration, the vector is normalised to retain precision during computation (otherwise the components will become too large or too small, and accuracy will suffer).

Algorithm 6: Power iteration

```

1 Function centrality( $G$ )
   Input: Graph  $G$  with weight matrix  $W$ 
   Output:  $f \in \mathbb{R}^n$  giving the eigenvector centrality of each vertex
2    $x \leftarrow$  uniform random vector in  $\mathbb{R}^n$ 
3   change  $\leftarrow \infty$ 
4   while change  $> \varepsilon$  do
5      $x' \leftarrow \frac{Wx}{\|Wx\|}$ 
6     change  $\leftarrow \|x' - x\|$ 
7      $x \leftarrow x'$ 
8   end
9   return  $x$ 
10 end

```

Google’s search engine is based on a similar centrality measure called **PageRank**. The basic idea is to represent the web as a unweighted directed graph $G = (V, E)$ where the vertices correspond to webpages and the edges correspond to links between pages. When a user types a query into Google, the search engine will return a set of matching results ordered by their PageRank (and nowadays, countless other factors), under the assumption that users will want to see the most important or ‘central’ webpages first.

The computation of PageRank can be expressed as an eigenvector problem similar to (17), and is performed via a power iteration method. For this reason, Bryan and Leise [21] describe PageRank as the ‘\$25,000,000,000 eigenvector’. It is safe to say that the applications of spectral graph theory are numerous and profound.

References

- [1] Jeff Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems in analysis*, 625(195-199):110, 1970.
- [2] N Alon and V.D Milman. Isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.
- [3] Nisheeth K Vishnoi et al. $Lx = b$. *Foundations and Trends in Theoretical Computer Science*, 8(1–2):1–141, 2013.
- [4] Stephen Cook. The p versus np problem. *Clay Mathematics Institute*, 2, 2000.

- [5] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.
- [6] William E Donath and Alan J Hoffman. Algorithms for partitioning of graphs and computer logic based on eigenvectors of connection matrices. *IBM Technical Disclosure Bulletin*, 15(3):938–944, 1972.
- [7] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6), Dec 2014.
- [8] Fan R.K Chung. *Spectral graph theory*. American Mathematical Society, 1997.
- [9] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
- [10] M. Charikar, C. Chekuri, A. Goel, S. Guha, and S. 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, 1998.
- [11] 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, STOC ’13*, page 11–20, New York, NY, USA, 2013. Association for Computing Machinery.
- [12] Y. de Colin de Verdière. Sur une hypothèse de transversalité d’arnold. *Commentarii mathematici Helvetici*, 63(2):184–193, 1988.
- [13] JF Lutzeyer and AT Walden. Comparing graph spectra of adjacency and laplacian matrices. *arXiv preprint arXiv:1712.03769*, 2017.
- [14] M Emre Celebi, Hassan A Kingravi, and Patricio A Vela. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert systems with applications*, 40(1):200–210, 2013.
- [15] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
- [16] Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 522–539, 2021.
- [17] Daniel A Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90, 2004.
- [18] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.
- [19] Charles J Alpert and Andrew B Kahng. Recent directions in netlist partitioning: A survey. *Integration*, 19(1-2):1–81, 1995.
- [20] Charles J Alpert and So-Zen Yao. Spectral partitioning: The more eigenvectors, the better. In *Proceedings of the 32nd annual ACM/IEEE design automation conference*, pages 195–200, 1995.

- [21] Kurt Bryan and Tanya Leise. The \$25,000,000,000 eigenvector: The linear algebra behind google. *SIAM review*, 48(3):569–581, 2006.