Chapter 4

Cheeger's Inequality

Cheeger's inequality is the fundamental result in spectral graph theory, which connects a combinatorial property of a graph and an algebraic quantity of its associated matrix. This connection is important in the theory of expander graphs and the theory of random walks, which we will see in later chapters. The proof is also very useful in graph partitioning which we will see soon.

Cheeger [Che70] proved this inequality in the manifold setting, and the inequality in the graph setting was proved in several works in the 80's [AM85, Alo86, SJ89] with motivations from expander graphs and random walks.

4.1 Graph Conductance

Recall from Proposition 3.18 that a graph G is connected if and only if $\lambda_2 > 0$ where λ_2 is the second smallest eigenvalue of the normalized Laplacian matrix. Cheeger's inequality is the robust generalization that λ_2 is large if and only if the graph is well-connected.

To state this formally, we need to define a measure of how well a graph is connected. There are different natural definitions and we state two of them here.

Definition 4.1 (Edge Expansion). *Let $G = (V, E)$ be an undirected graph. The expansion of a subset $S \subseteq V$ and the expansion of the graph G are defined as*

$$\Phi(S) := \frac{|\delta(S)|}{|S|} \qquad \text{and} \qquad \Phi(G) := \min_{S:|S|\leq|V|/2} \Phi(S),$$

where $\delta(S)$ denotes the set of edges with one endpoint in S and one endpoint in $V - S$.

Definition 4.2 (Edge Conductance). *Let $G = (V, E)$ be an undirected graph. The conductance of a subset $S \subseteq V$ and the conductance of the graph G are defined as*

$$\phi(S) := \frac{|\delta(S)|}{\mathrm{vol}(S)} \qquad \text{and} \qquad \phi(G) := \min_{S:\mathrm{vol}(S)\leq|E|} \phi(S),$$

where $\mathrm{vol}(S) := \sum_{v\in S} \deg(v)$ is called the volume of the subset S. Note that for all $S \subseteq V$, $0 \leq \phi(S) \leq 1$, as it is the ratio of the number of edges cut by S to the total degree in S.

For graph partitioning, a subset $S \subseteq V$ corresponds to the partition of the vertex set V into two parts $(S, V - S)$. In the above definitions for $\Phi(G)$ and $\phi(G)$, note that we only consider the part with smaller denominator. Both of these definitions try to identify the "bottleneck" in the graph. When the graph G is d-regular, the two definitions are basically equivalent, with $\Phi(G) = d \cdot \phi(G)$. When the graph G is non-regular, the relation between the edge conductance and the second smallest eigenvalue is more elegant.

Theorem 4.3 (Cheeger's Inequality)**.** *Let $G = (V, E)$ be an undirected graph and let λ_2 be the second smallest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(G)$ in Definition 3.20. Then*

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

The first inequality is called the easy direction, and the second inequality is called the hard direction. We will see that the easy direction corresponds to using the second eigenvalue as a "relaxation" for graph conductance, and the hard direction corresponds to "rounding" a fractional solution to graph conductance to an integral solution.

We say that a graph is an *expander graph* if $\phi(G)$ is large (e.g. $\phi(G) \geq c$ for a constant $0 < c < 1$), and we say that a subset $S \subseteq V$ is a *sparse cut* if $\phi(S)$ is small. Both concepts are very useful.

An expander graph with linear number of edges is an efficient object with diverse applications in theoretical computer science and mathematics; see [HLW06] for an excellent survey. An important implication of Cheeger's inequality is that the second eigenvalue of the normalized Laplacian matrix can be used to certify that a graph is an expander graph, which provides an algebraic way to construct expander graphs that turns out to be very fruitful.

Finding a sparse cut is useful in designing divide-and-conquer algorithms, with applications in image segmentation, data clustering, community detection, VLSI design, among others. The algorithmic implication of Cheeger's inequality is discussed in the following subsection.

Spectral Partitioning Algorithm

A popular heuristic in finding a sparse cut in practice is the following spectral partitioning algorithm.

Algorithm 1 Spectral Partitioning Algorithm

Require: An undirected graph $G = (V, E)$ with $V = [n]$ and $m = |E|$.
1: Compute the second smallest eigenvalue λ_2 of $\mathcal{L}(G)$ and a corresponding eigenvector $x \in \mathbb{R}^n$.
2: Sort the vertices so that $x_1 \geq x_2 \geq \ldots \geq x_n$.
3: Let $S_i = [i]$ if $\mathrm{vol}_G\big([i]\big) \leq m$ and let $S_i = [n] \setminus [i]$ if $\mathrm{vol}_G\big([i]\big) > m$.
4: **return** $\min_{i:1 \leq i \leq n-1}\{\phi(S_i)\}$.

The algorithm is strikingly simple, with only a few lines of code if we use some mathematical software such as MATLAB, which is one reason why this heuristic is popular. The algorithm only checks the linear number of solutions defined by the ordering in a second eigenvector, although there are exponentially many subsets $S \subseteq V$.

There is a near-linear time randomized algorithm to compute an approximate eigenvector of the second eigenvalue, using the power method and a fast Laplacian solver. So, the algorithm is also fast theoretically, but we won't discuss the details in this chapter.

The main reason that it is popular is that it performs very well in various applications, especially in image segmentation and clustering, and it was considered a breakthrough in image segmentation about 20 year ago [SM00].

The proof of Cheeger's inequality will provide a nontrivial performance guarantee of the spectral partitioning algorithm, that it will always output a set S with conductance $\phi(S) \leq \sqrt{2\lambda_2} \leq 2\sqrt{\phi(G)}$. When $\phi(G)$ is a constant, this is a constant factor approximation algorithm. When $\phi(G)$ is small, say $\phi(G) \leq 1/n$, the approximation ratio could be as bad as $\Theta(\sqrt{n})$. It has been an open problem to explain the empirical success rigorously, and we will come back to this in the next chapter.

4.2 Easy Direction

In this section, we prove the easy direction of Cheeger's inequality. For simplicity of exposition, we assume the graph is d-regular, so $\mathcal{L}(G) = \frac{1}{d}L(G)$. We will outline the modifications needed for the analysis of non-regular graphs in a subsection at the end of this chapter, but those are just some additional manipulations where all the main ideas are already in the d-regular case.

First, we start with the nice optimization formulation of the second eigenvalue of the normalized Laplacian matrix using the Rayleigh quotient in Definition 2.9.

Lemma 4.4 (Optimization Formulation for λ_2)**.** *Let $G = (V, E)$ be an undirected d-regular graph with $V = [n]$ and λ_2 be the second smallest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(G)$. Then*

$$\lambda_2 = \min_{x \in \mathbb{R}^n : x \perp \vec{1}} R_{\mathcal{L}}(x) = \min_{x \in \mathbb{R}^n : x \perp \vec{1}} \frac{x^T \mathcal{L} x}{x^T x} = \min_{x \in \mathbb{R}^n : x \perp \vec{1}} \frac{x^T L x}{d \cdot x^T x} = \min_{x \in \mathbb{R}^n : x \perp \vec{1}} \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{d \sum_{i \in V} x(i)^2}.$$

Proof. The first equality is by Lemma 2.11, although it was stated in the maximization form, the same proof works for the minimization form. The last equality is by Lemma 3.17. □

The observation is that the minimization problem of graph conductance can be formulated in a similar way. The following lemma holds for non-regular graphs.

Lemma 4.5 (Optimization Formulation for Graph Conductance)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Then*

$$\phi(G) = \min_{x \in \{0,1\}^n} \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot x(i)^2} \qquad \text{subject to} \qquad \sum_{i \in V} \deg(i) \cdot x(i)^2 \leq |E|.$$

Proof. Each feasible solution $S \subseteq V$ with $\operatorname{vol}(S) \leq |E|$ in the graph conductance problem corresponds to a feasible solution $\chi_S \in \{0,1\}^m$ with $\sum_{i \in V} \deg(i) \cdot x(i)^2 \leq |E|$ in this formulation, and vice versa. Note that the numerator counts the number of edges in $\delta(S)$ and the denominator computes the volume $\operatorname{vol}(S)$ of S. □

Note that for d-regular graphs, the constraint simplifies to $\sum_{i \in V} x(i)^2 \leq n/2$.

Intuition: The main difference between these two formulations is that the former optimizes over the continuous domain $x \in \mathbb{R}^n$, while the latter optimizes over the discrete domain $x \in \{0,1\}^n$. A good way to think of the relation between the two optimization problems is that the former problem is a "relaxation" of the latter problem. This is a common idea in the design of approximation algorithms. The latter problem is an NP-hard optimization problem, because of the discrete domain. The relaxation idea is to optimize over a larger continuous domain, so that the problem can be solved in polynomial time. Since we optimize over a larger domain, the objective value of the former problem could only be smaller than that of the latter problem, and so we expect that $\lambda_2 \leq \phi(G)$. This is the main intuition for the easy direction. For these two formulations, however, there are also some differences between the constraints, and it only holds that $\lambda_2 \leq 2\phi(G)$.

Proof of the Easy Direction in the d-Regular Case. To upper bound λ_2, we just need to find a vector $x \perp \vec{1}$ and compute its Rayleigh quotient $R_{\mathcal{L}}(x)$. Let $S \subseteq V$ be a subset of vertices with $|S| \leq \frac{|V|}{2}$. Consider the following "binary" solution $x \in \mathbb{R}^n$ with

$$x(i) = \begin{cases} +1/|S| & \text{if } i \in S \\ -1/|V-S| & \text{if } i \notin S \end{cases}.$$

By construction, $x \perp \vec{1}$, and so

$$\lambda_2 \leq R_{\mathcal{L}}(x) = \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{d \sum_{i \in V} x_i^2} = \frac{|\delta(S)| \cdot \big(\frac{1}{|S|} + \frac{1}{|V-S|}\big)^2}{d\big(|S| \cdot \frac{1}{|S|^2} + |V-S| \cdot \frac{1}{|V-S|^2}\big)} = \frac{|\delta(S)| \cdot |V|}{d \cdot |S| \cdot |V-S|} \leq 2\phi(S),$$

where the last inequality uses the assumption that $|S| \leq \frac{|V|}{2}$ which implies that $\frac{|V|}{|V-S|} \leq 2$. $\square$

4.3 Hard Direction

By optimizing over a larger domain, however, the objective value of the continuous problem will typically be smaller than that of the discrete problem.

Tight Example

Consider the cycle of $4n$ vertices. One can compute the second eigenvector of the cycle exactly, but we don't do it here. Recall that $\lambda_2 = \min_{x \perp \vec{1}} x^T \mathcal{L} x / x^T x$, so to give an upper bound we just need to demonstrate a solution with small objective value. Consider

$$x = \Big(1, 1 - \frac{1}{n}, 1 - \frac{2}{n}, \ldots, \frac{1}{n}, 0, -\frac{1}{n}, \ldots, -1 + \frac{1}{n}, -1, -1 + \frac{1}{n}, \ldots, -\frac{1}{n}, 0, \frac{1}{n}, \ldots, 1 - \frac{1}{n}\Big).$$

Then $x \perp \vec{1}$, and so

$$\lambda_2 \leq \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{2 \sum_{i \in V} x(i)^2} = \Theta\bigg(\frac{n\big(\frac{1}{n}\big)^2}{n}\bigg) = \Theta\Big(\frac{1}{n^2}\Big).$$

On the other hand, it is easy to verify that the conductance of the cycle of $4n$ vertices is $\Theta(\frac{1}{n})$. This is an example where the hard direction $\phi(G) \leq \sqrt{2\lambda_2}$ is tight up to a constant factor.

Rounding

For the discrete optimization problem of graph conductance, we would like the solution to be a binary solution as in the proof of the easy direction. But once we relax the problem to the continuous domain (so that it becomes polynomial time solvable), the optimal solution x could be very "fractional" or "continuous" as the above example shown. For the hard direction, the task is to prove that there is always a binary solution z with objective value at most the square root of the objective value of the continuous solution x. A typical way in approximation algorithms to do this task is to "round" the "fractional/continuous" solution x to an "integral/binary" solution z and bound the objective value of z in terms of the objective value of x. This is what we will do.

Intuition

Think of the optimizer x to the optimization problem in Lemma 4.4 as embedding the vertices of the graph into the real line, so that for most edges $|x(i)-x(j)|$ is small. A natural strategy is to do a "threshold rounding", where we pick a threshold t and set $z(i) = 0$ if $x(i) < t$ and $z(i) = 1$ if $x(i) \geq t$. A simple analogy is that if most rows/edges have few nonzeros, then there is a column/threshold with few nonzeros. This intuition can be made precise if the optimization problem is of the form

$$\min_{x \in \mathbb{R}^n_+} \frac{\sum_{ij \in E} |x(i) - x(j)|}{d \sum_{i \in V} x(i)},$$

but the optimization problem in Lemma 4.4 is a sum of quadratic terms and this is basically where the square root loss in the hard direction comes from.

Truncation

The proof of the hard direction has two steps. The first step is a preprocessing step that truncates an optimizer $x \in \mathbb{R}^n$ to the continuous problem in Lemma 4.4 to a vector $y \in \mathbb{R}^n$ with $\mathrm{vol}(\mathrm{supp}(y)) \leq |E|$, where $\mathrm{supp}(y) := \{i \mid y(i) \neq 0\}$. This is to ensure that the solution S produced in the second step satisfies $\mathrm{vol}(S) \leq |E|$, satisfying the constraint in the discrete problem in Lemma 4.5.

There are two ways to do this step. The first way requires that x is indeed an eigenvector, and the proof is shorter and is enough for establishing Cheeger's inequality. The second way only requires that x is perpendicular to the first eigenvector, which is important for algorithmic purpose, but the proof is a bit longer. We will present the proof for the first way and outline the proof for the second way in the problem subsection in the end. The following lemma is from [HLW06] and it holds for general undirected graphs.

Lemma 4.6 (Truncating Eigenvector)**.** *Let $G = (V, E)$ be an undirected graph and $x \in \mathbb{R}^n$ be an eigenvector of $\mathcal{L}(G)$ with eigenvalue λ. Let $x_+ \in \mathbb{R}^n$ be the vector with $x_+(i) = \max\{x(i), 0\}$ for $1 \leq i \leq n$, and $x_- \in \mathbb{R}^n$ be the vector with $x_-(i) = \min\{x(i), 0\}$ for $1 \leq i \leq n$. Then $R_{\mathcal{L}}(x_+) \leq R_{\mathcal{L}}(x) = \lambda$ and $R_{\mathcal{L}}(x_-) \leq R_{\mathcal{L}}(x) = \lambda$.*

Proof. For each vertex $i \in \mathrm{supp}(x_+)$, by the definition of normalized Laplacian matrix in Definition 3.20 and the assumption that x is an eigenvector with eigenvalue λ,

$$(\mathcal{L}x_+)(i) = x_+(i) - \sum_{j:ij \in E} \frac{x_+(j)}{\sqrt{\deg(i)\deg(j)}} \leq x(i) - \sum_{j:ij \in E} \frac{x(j)}{\sqrt{\deg(i)\deg(j)}} = (\mathcal{L}x)(i) = \lambda \cdot x(i) = \lambda \cdot x_+(i).$$

This implies that

$$\langle x_+, \mathcal{L}x_+ \rangle = \sum_{i \in \mathrm{supp}(x_+)} x_+(i) \cdot (\mathcal{L}x_+)(i) \leq \sum_{i \in \mathrm{supp}(x_+)} \lambda \cdot x_+(i)^2 = \lambda \cdot \|x_+\|_2^2.$$

Therefore, $R_{\mathcal{L}}(x_+) = \langle x_+, \mathcal{L}x_+ \rangle / \|x_+\|_2^2 \leq \lambda$, and the proof is the same for x_-. □

Let x be an eigenvector with eigenvalue λ_2. Since $x \neq 0$ and x is perpendicular to the first eigenvector which is a positive vector, both $\operatorname{supp}(x_+)$ and $\operatorname{supp}(x_-)$ are non-empty sets. By choosing either x_+ or x_- that has a smaller volume in its support and taking a proper normalization, we arrive at the following corollary.

Corollary 4.7 (Preprocessed Vector)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$ and λ_2 be the second eigenvalue of $\mathcal{L}(G)$. There exists a vector $y \in \mathbb{R}^n$ that satisfies (i) $y \geq 0$, (ii) $R_{\mathcal{L}}(y) \leq \lambda_2$, (iii) $\operatorname{vol}(\operatorname{supp}(y)) \leq \frac{1}{2}\operatorname{vol}(V) = |E|$, and (iv) $\|y\|_2^2 = 1$.*

For algorithmic purpose, we may not be able to compute an eigenvector exactly, but rather a vector x that is perpendicular to the first eigenvector and $R_{\mathcal{L}}(x) \approx \lambda_2$. In the problem subsection in the end, we describe how to truncate the vector to satisfy Corollary 4.7, which is similar but with an additional shifting/centering step.

In the d-regular case, to summarize, the truncation step transforms a vector x with small Rayleigh quotient that satisfies $x \perp \vec{1}$ in the continuous problem in Lemma 4.4 into a vector y with small Rayleigh quotient that satisfies $|\operatorname{supp}(y)| \leq n/2$ that is required in the discrete problem in Lemma 4.5.

Threshold Rounding

The main step in the hard direction is the threshold rounding step hinted earlier, which takes a vector y in Corollary 4.7 and outputs a set $S \subseteq \operatorname{supp}(y)$ with small conductance $\phi(S)$. As described in the spectral partitioning algorithm, we will only consider those threshold/level sets $S'_t := \{i \in V \mid y(i) \geq t\}$ for $t > 0$, as in every proof of Cheeger's inequality. Our proof will follow that of Trevisan [Tre16], whose idea is to choose a random t and considers the level set $S_t := \{i \in V \mid y(i)^2 \geq t\}$, and to bound the conductance of S_t by computing the expectation of the numerator and the expectation of the denominator separately. The idea of choosing a random t is similar to the idea of randomized rounding in approximation algorithms, and his analysis of computing the expectations separately simplifies the proof.

Lemma 4.8 (Threshold Rounding)**.** *Let $G = (V, E)$ be an undirected d-regular graph with $V = [n]$. Let $y \in \mathbb{R}^n_+$ be a vector with non-negative entries. There exists $t > 0$ such that the threshold set $S_t := \{i \in [n] \mid y(i)^2 \geq t\}$ is nonempty and satisfies $\phi(S_t) \leq \sqrt{2R_{\mathcal{L}}(y)}$.*

Proof. For convenience, we scale y so that $\max_i y(i) = 1$. Let $t \in (0, 1]$ be chosen uniformly at random. Note that the set S_t is nonempty by construction. In the following, we compute the expected value of the numerator and of the denominator in Lemma 4.5 separately.

For an edge $ij \in E$, note that the probability that $ij \in \delta(S_t)$ is $|y(i)^2 - y(j)^2|$, when the random threshold t falls between $y(i)^2$ and $y(j)^2$. By linearity of expectation,

$$\mathbb{E}_t\big[|\delta(S_t)|\big] = \sum_{ij \in E} \Pr(ij \in \delta(S_t)) = \sum_{ij \in E} \big|y(i)^2 - y(j)^2\big| = \sum_{ij \in E} \big|y(i) - y(j)\big| \cdot \big|y(i) + y(j)\big|.$$

To relate this expected value to the numerator of the Rayleigh quotient in Lemma 4.4, the Cauchy-Schwarz inequality is used as in every proof of Cheeger's inequality so that

$$\mathbb{E}_t\big[|\delta(S_t)|\big] \leq \sqrt{\sum_{ij \in E} \big|y(i) - y(j)\big|^2} \sqrt{\sum_{ij \in E} \big|y(i) + y(j)\big|^2} \leq \sqrt{\sum_{ij \in E} \big|y(i) - y(j)\big|^2} \sqrt{2d \cdot \sum_{i \in V} y(i)^2},$$

where the second inequality holds because $\sum_{ij\in E}\left|y(i)+y(j)\right|^2 \leq \sum_{ij\in E} 2\big(y(i)^2+y(j)^2\big) = 2d\sum_{i\in V} y(i)^2$ where the assumption that G is d-regular is used.

For a vertex $i \in V$, note that the probability that $i \in S_t$ is $y(i)^2$, when the random threshold t is smaller than $y(i)^2$. By linearity of expectation,

$$\mathbb{E}_t\big[d|S_t|\big] = d \cdot \sum_{i\in V} \Pr(i \in S_t) = d \cdot \sum_{i\in V} y(i)^2.$$

Therefore,

$$\frac{\mathbb{E}_t\big[|\delta(S_t)|\big]}{\mathbb{E}_t\big[d|S_t|\big]} \leq \sqrt{\frac{2\sum_{ij\in E}\left|y(i)-y(j)\right|^2}{d\cdot\sum_{i\in V} y(i)^2}} = \sqrt{2R_{\mathcal{L}}(y)}.$$

Note that we cannot conclude from this that $\mathbb{E}_t[\phi(S_t)] = \mathbb{E}_t\big[|\delta(S_t)|/d|S_t|\big] \leq \sqrt{2R_{\mathcal{L}}(y)}$, but we can conclude from this that

$$\mathbb{E}_t\Big[|\delta(S_t)| - d|S_t|\sqrt{2R_{\mathcal{L}}(y)}\Big] \leq 0 \quad\implies\quad \exists t > 0 \text{ with } \phi(S_t) = \frac{|\delta(S_t)|}{d|S_t|} \leq \sqrt{2R_{\mathcal{L}}(y)}.$$

□

Analysis of the Spectral Partitioning Algorithm

We summarize the proof of the hard direction, which also provides an analysis of the spectral partitioning algorithm.

Proof of the Hard Direction in the d-Regular Case. Let $x \in \mathbb{R}^n$ be an eigenvector of $\mathcal{L}(G)$ with eigenvalue λ_2. First we apply the truncation step in Lemma 4.6 and Corollary 4.7 to obtain a vector $y \in \mathbb{R}^n$ with $R_{\mathcal{L}}(y) \leq R_{\mathcal{L}}(x) = \lambda_2$ and $|\operatorname{supp}(y)| \leq n/2$. Then we apply the threshold rounding step in Lemma 4.8 on y to obtain a nonempty set $S_t = \{i \in [n] \mid y(i)^2 \geq t\}$ with $t > 0$ and $\phi(S_t) \leq \sqrt{2R_{\mathcal{L}}(y)} \leq \sqrt{2\lambda_2}$. Since $S_t \subseteq \operatorname{supp}(y)$, it follows that $0 < |S_t| \leq |\operatorname{supp}(y)| \leq n/2$ and thus $\phi(G) \leq \phi(S_t) \leq \sqrt{2\lambda_2}$. Finally, note that S_t is a threshold set of y, which is also a threshold set of x by the construction in Lemma 4.6, as y is either x_+ or x_-. This implies that the spectral partitioning algorithm has considered this set, and thus it will output a set S with $\phi(S) \leq \sqrt{2\lambda_2}$.
□

4.4 Discussions

We discuss more about the performance of the spectral partitioning algorithm and also outline the modifications needed for general weighted graphs.

More Examples

Both sides of Cheeger's inequality are tight, even the constants are tight. For the easy direction, one can check that it is tight for the hypercubes; see Problem 3.23. For the hard direction, we have already seen that it is tight up to a constant factor for the cycles. It is possible to assign edge weights to the cycle so that even the constant $\sqrt{2}$ is tight, and we leave it as a challenging example to work out.

These are the standard examples to show that both sides of Cheeger's inequality are tight, but they do not yet provide much insights about how the spectral partitioning algorithm only outputs a set S with $\phi(S) \approx \sqrt{\phi(G)}$. For the cycle examples where the $\phi(G) \approx \sqrt{\lambda_2}$, the spectral partitioning actually works perfectly to output a set S with $\phi(S) \approx \phi(G)$, because it outputs a set S with $\phi(S) \approx \sqrt{\lambda_2} \approx \phi(G)$. Indeed, it is a general phenomenon that rounding algorithms work perfectly for the worst integrality gap examples.

So, to find an example where the spectral partitioning algorithm performs poorly, we need to look at the examples where the easy direction is tight but the algorithm outputs a set S with $\phi(S) \approx \sqrt{\lambda_2} \approx \sqrt{\phi(G)}$. Actually, for the hypercube examples where the easy direction is tight, there are vectors in the second eigenspace where the spectral partitioning algorithm performs perfectly and performs poorly.

Problem 4.9 (Spectral Partitioning for Hypercubes)**.** *Let G be the hypercube of dimension d with 2^d vertices and $\mathcal{L}(G)$ be its normalized Laplacian matrix.*

1. *Show that there is an eigenvector vector $x \in \mathbb{R}^{2^d}$ of $\mathcal{L}(G)$ with eigenvalue λ_2 so that the spectral partitioning algorithm applied on x outputs a set S with $\phi(S) = R_{\mathcal{L}}(x) = \frac{1}{2}\lambda_2$.*

2. *Show that there is an eigenvector vector $y \in \mathbb{R}^{2^d}$ of $\mathcal{L}(G)$ with eigenvalue λ_2 so that the spectral partitioning algorithm applied on y outputs a set S with $\phi(S) \approx \sqrt{R_{\mathcal{L}}(y)} = \sqrt{\lambda_2}$.*

Since we do not have control over which eigenvector in the second eigenspace is returned, this provides an example where the spectral partitioning algorithm could perform poorly. But perhaps this example is not so satisfying as we do not see clearly how the spectral partitioning algorithm is fooled.

We construct such an example in the following by tweaking the cycle example. Let G be the weighted graph with vertices $\{v_1, \ldots, v_n, v_{n+1}, \ldots, v_{2n}\}$, and two cycles $(v_1, v_2, \ldots, v_n)$ and $(v_{n+1}, v_{n+2}, \ldots, v_{2n})$ where every edge in these cycles is of weight one, and a "hidden" matching $\{v_1 v_{n+1}, v_2 v_{n+2}, \ldots, v_n v_{2n}\}$ where every edge in the matching is of weight say $100/n^2$. Then it is easy to see that the set of smallest conductance is the set $S := \{v_1, \ldots, v_n\}$ with $\phi(S) = O(1/n^2)$. However, the edges in the hidden matching are so light that the spectral partitioning algorithm did not "feel" them, and still thinks that the embedding of the cycle is the best embedding of the vertices onto the real line. Indeed, one can verify that the second eigenvector x in this example is still the same as that in the cycle of n vertices, with $x(v_i) = x(v_{n+i})$ for $1 \leq i \leq n$. Therefore, λ_2 is still $O(1/n^2)$ which is close to $\phi(G)$, but the cut of smallest conductance is completely lost in x and every threshold set has conductance $\Omega(1/n)$. This is a more insightful example to see how the spectral partitioning algorithm is fooled. This example is a weighted graph, but one can also modify this example slightly to keep the same structure while making the graph unweighted.

Cheeger's Inequality for General Weighted Graphs

Once we understand the proof for the d-regular case, it is not difficult to extend it to general weighted graphs.

Let $G = (V, E)$ be an edge weighted graph with a non-negative weight $w(e) \geq 0$ on each edge $e \in E$. The weighted degree of a vertex i is defined as $\deg_w(i) = \sum_{j: ij \in E} w(ij)$, and the diagonal degree matrix is denoted by D_w. The weighted adjacency matrix A_w is defined so that $(A_w)_{i,j} = w(ij)$ for all $i, j \in V$. The weighted Laplacian matrix L_w is defined as $D_w - A_w$, and the weighted normalized Laplacian matrix $\mathcal{L}_w$ is defined as $D_w^{-\frac{1}{2}} L_w D_w^{-\frac{1}{2}}$. Check that the quadratic form $x^T L_w x = \sum_{ij \in E} w(ij)\big(x(i) - x(j)\big)^2$. The following is a generalization of Lemma 4.4 for λ_2 of $\mathcal{L}_w(G)$, which can be obtained by a change of variable.

Exercise 4.10 (General Optimization Formulation for λ_2). *Let $G = (V, E)$ be a weighted graph with $V = [n]$ and λ_2 be the second smallest eigenvalue of $\mathcal{L}_w(G)$. Show that*

$$\lambda_2 = \min_{y \in \mathbb{R}^n : \sum_{i \in V} \deg_w(i) y(i) = 0} \frac{y^T L_w y}{y^T D_w y} = \min_{y \in \mathbb{R}^n : \sum_{i \in V} \deg_w(i) y(i) = 0} \frac{\sum_{ij \in E} w(ij)\big(y(i) - y(j)\big)^2}{\sum_{i \in V} \deg_w(i) \cdot y(i)^2}.$$

The weighted conductance of a subset is defined naturally as $\phi_w(S) := w(\delta(S)) / \operatorname{vol}_w(S)$ where $w(\delta(S)) := \sum_{e \in \delta(S)} w(e)$ and $\operatorname{vol}_w(S) := \sum_{i \in S} \deg_w(i)$, and the weighted conductance of a graph is defined as $\phi_w(G) := \min_{S : \operatorname{vol}_w(S) \leq \frac{1}{2} \operatorname{vol}_w(V)} \phi(S)$. Choosing an appropriate binary solution involving $\operatorname{vol}_w(S)$, the easy direction can be shown similarly as in the d-regular case.

Exercise 4.11 (Easy Direction for General Weighted Graphs). *Let $G = (V, E, w)$ be an edge weighted graph and λ_2 be the second smallest eigenvalue of $\mathcal{L}_w(G)$. Show that $\frac{1}{2}\lambda_2 \leq \phi_w(G)$.*

The main changes are actually in the easy direction. For the hard direction, the proofs are basically the same. The arguments in Lemma 4.6 and Corollary 4.7 work the same way. The analysis of the threshold rounding is very similar to that in Lemma 4.8, but on the formulation in Exercise 4.10.

Exercise 4.12 (Hard Direction for General Weighted Graphs). *Let $G = (V, E, w)$ be an edge weighted graph and λ_2 be the second smallest eigenvalue of $\mathcal{L}_w(G)$. Prove that $\phi_w(G) \leq \sqrt{2\lambda_2}$.*

4.5 Problems

Problem 4.13 (Truncation). *We outline the truncation step which does not require that the vector is an eigenvector. The following statements are for general weighted graphs. You may specialize the problem to the d-regular case.*

Let $G = (V, E)$ be a weighted undirected graph with $V = [n]$ and $y \in \mathbb{R}^n$ be a vector with $\sum_{i \in V} \deg_w(i) y(i) = 0$. Let $R_w(y) := y^T L_w y / y^T D_w y$ be the weighted Rayleigh quotient.

1. *Let c be a value such that $\operatorname{vol}_w(\{i \mid y(i) < c\}) \leq \frac{1}{2} \operatorname{vol}_w(V)$ and $\operatorname{vol}_w(\{i \mid y(i) > c\}) \leq \frac{1}{2} \operatorname{vol}_w(V)$. Let $z := y - c\vec{1}$. Prove that $R_w(z) \leq R_w(y)$. You may need to use the assumption that $\sum_{i \in v} \deg_w(i) y(i) = 0$.*

2. *Let $z \in \mathbb{R}^n$ be the vector obtained in the previous step. Let $z_+ \in \mathbb{R}^n$ be the vector with $z_+(i) := \max\{z(i), 0\}$ for $1 \leq i \leq n$, and $z_- \in \mathbb{R}^n$ be the vector with $z_-(i) := \min\{z(i), 0\}$ for $1 \leq i \leq n$. Prove that $\min\{R_w(z_+), R_w(z_-)\} \leq R_w(z)$.*

3. *Conclude with the suitable generalization of Corollary 4.7 that allows one to continue to prove the hard direction for general weighted graphs.*

4.6 References

[Alo86] Noga Alon. Eigenvalues and expanders. *Comb.*, 6(2):83–96, 1986. 25

[AM85] Noga Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985. 25

[Che70] J. Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems Anal.*, page 195, 1970. 25

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 26, 29

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.*, 82(1):93–133, 1989. 25

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000. 26

[Tre16] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. 2016. 30