CSE 521: Design and Analysis of Algorithms I **Spring 2016**

Lecture 12: Introduction to Spectral Graph Theory, Cheeger's inequality

Lecturer: Shayan Oveis Gharan *May 4th* *Scribe: Gabriel Cadamuro*

12.1 Overview of spectral graph theory

Spectral graph theory seeks to relate the eigenvectors and eigenvalues of matrices corresponding to a Graph to the combinatorial properties of the graph. While generally the theorems are designed for unweighted and undirected graphs they can be extended to the weighted graph case and (less commonly) the directed graph case. In these notes we only study undirected graphs.

Given a graph G we use A to denote adjacency matrix of the graph and D to denote the diagonal degree matrix of the vertex degrees. Recall these are defined as

$$\begin{aligned} A_{i,j} &= \begin{cases} w_{i,j} & \text{if } i \sim j \\ 0 & \text{otherwise,} \end{cases} \\ D_{i,j} &= \begin{cases} d_w(i) = \sum_j w_{ij} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

If G is unweighted, w_{ij} is 1, when there is an edge between vertex i and j and 0 otherwise. Generally in spectral graph theory we work with a normalization of the adjacency matrix of G, known as graph *Laplacian*. The Laplacian of G denoted by L is defined as follows:

$$L = D - A$$

The main reason for this normalization choice is that the quadratic form of a Laplacian can be nicely written as an energy function. For any graph laplacian L and $x \in R^n$

$$x^T L x = \sum_{i \sim j} (x_i - x_j)^2 w_{ij} \tag{12.1}$$

We will see shortly that this formulation can be used to relate eigenvalues of L, A to the combinatorial properties of G. Let us justify the above identity. Consider a graph G with only two vertices $1, 2$ and an edge $(1, 2)$. In this case

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad L = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

Therefore,

$$x^T L x = \begin{bmatrix} x_1, x_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 - x_1 x_2 - x_1 x_2 + x_2^2 = (x_1 - x_2)^2$$

To prove (12.1), let L_{ij} be the laplacian of the graph with all the vertices in G but only the edge between i to j. We note then that

$$x^T Lx = x^T \left(\sum_{i \sim j} L_{ij} \right) x = \sum_{i \sim j} x^T L_{ij} x = \sum_{i \sim j} (x_i - x_j)^2$$

12.2 Properties of the graph laplacian

One can use the Laplacian of G to state several properties of G. For example, $\det(L + \mathbf{1}\mathbf{1}^T)$ is equal to n times the number of spanning trees of G. In this section we only describe and prove basic properties of the Laplacian. We also use $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ to denote the eigenvalues of L.

Prop 1: For any graph G, $L \succeq 0$. This can be seen by looking at its quadratic form. For any vector $x \in \mathbb{R}^n$,

$$x^T Lx = \sum_{i \sim j} (x_i - x_j)^2 \geq 0$$

Prop 2: For any graph G, $\lambda_1 = 0$. This essentially follows from the variational characterization of eigenvalues. Recall that,

$$\lambda_1 = \min_{\|x\|=1} x^T Lx = \min_{\|x\|=1} \sum_{i \sim j} (x_i - x_j)^2.$$

Since $\sum_{i \sim j} (x_i - x_j)^2 \geq 0$,, to show $\lambda_1 = 0$, it is enough to construct a vectors x such that $\sum_{i \sim j} (x_i - x_j)^2 = 0$. For that matter it is enough to let $x = \mathbf{1}$. This argument also shows that the all-1s vector is always the first eigenvector of Laplacian of any graph.

Prop 3: For any graph G, $\lambda_2 = 0$ iff G is disconnected. We prove this statement in two steps: (i) we first prove that if $\lambda_2 = 0$, then G is disconnected, (ii) in the second step, we show that if G is disconnected, then $\lambda_2 = 0$.

(i) Assume $\lambda_2 = 0$. Let x be the eigenvector corresponding to λ_2. Since $\lambda_2 = 0$, $\sum_{i \sim j} (x_i - x_j)^2 = 0$. Therefore, for all $i \sim j$, we have $x_i = x_j$. But this means that any two vertices i, j which are in the same connected component of G we have $x_i = x_j$ since equality is transitive. So to show that G has two connected components it is enough to show that x assigns two distinct values to the vertices of G.

On the other hand, since x is the second eigenvector, we have $x \perp \mathbf{1}$. Therefore, there is i, j such that $x_i \neq x_j$ as desired.

(ii) Now, assume that G is disconnected, and we show that $\lambda_2 = 0$. Let $S \subseteq V$ be a disconnected set of vertices in G and $\overline{S}$ be its complement. We construct the vector x as follows:

$$x_i \begin{cases} \frac{1}{|S|} & \text{if } i \in S \\ -\frac{1}{|\overline{S}|} & \text{otherwise.} \end{cases}$$

Since x assigns the same number to the vertices of each connected component of G, for all $i \sim j$, we have $x_i = x_j$. So,

$$x^T Lx = \sum_{i \sim j} (x_i - x_j)^2 = 0.$$

Moreover we note that $\langle x, \mathbf{1}\rangle = |S|(\frac{1}{|S|}) - |\overline{S}|(\frac{1}{|\overline{S}|}) = 0$. Therefore,

$$\lambda_2 = \min_{y \perp \mathbf{1}, \|y\|=1} y^T L y = x^T L x = 0.$$

The third property can be naturally extended to any k. In particular, for any graph G, $\lambda_k = 0$ if and only if G has at least k connected components. We leave this as exercise.

12.3 Normalized Laplacian

Let us define the normalized adjacency matrix as $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. That since D is a diagonal matrix that taking the power can be done entrywise and we can simply understand the entires of these matrices as

$$D^{-\frac{1}{2}}_{i,i} = \frac{1}{\sqrt{d_w(i)}}$$
$$\tilde{A}_{ij} = \frac{w_{ij}}{\sqrt{d_w(i) d_w(j)}}$$

Similarly, we can define the normalized Laplacian matrix

$$\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = I - \tilde{A}.$$

The above matrix is closely related to the normalized adjacency matrix, if λ is an eigenvalue of $\tilde{L}$, $1 - \lambda$ is an eigenvalue of $\tilde{A}$. This is because

$$\lambda x = \tilde{L} x = (I - \tilde{A}) x = x - \tilde{A} x \Rightarrow \tilde{A} x = (1 - \lambda) x.$$

The normalized adjacency matrix (and the normalized Laplacian matrix) are closely related to the random walk matrix of G. The matrix $P = D^{-1} A$ is the transition probability matrix of the simple random walk on G. It is not hard to see that any eigenvalue of $\tilde{A}$ is also an eigenvalue of P. We will see more details in the future lectures.

For simplicity of the notation, from now on, we assume that G is d-regular; namely that each vertex in the graph has precisely d edges incident to it. All the following results work for non-regular (weighted) graphs but this assumption considerably simplifies the notation. In this case, the normalized laplacian is just $\tilde{L} = \frac{L}{d} = I - A/d$..

Another reason for working with the normalized Laplacian matrix is that the eigenvalues of $\frac{L}{d}$ are in the range

$$0 \leq \lambda_1 \ldots \lambda_n \leq 2,$$

in other words the eigenvalues are scale free. Note that since $\tilde{L} = L/d$, any eigenvalue λ of $\frac{L}{d}$ corresponds to an eigenvalue $d\lambda$ of L.

12.4 Cheeger's inequality

From now on assume that

$$0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$$

are the eigenvalues of L/d. The following theorem is one of the fundamental results in spectral graph theory with many applications in complexity theory, coding theory, analysis of random walks, approximation algorithms, etc.

Theorem 12.1 (Cheeger's inequality [AM85; Alo86])**.** *For any graph G with normalized Laplacian $\tilde{L}$ and eigenvalues $0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$,*

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

To better understand the above example let us consider extreme cases.

Example 1: Disconnected graphs By property 3 of the Laplacian matrix, a graph G is disconnected if and only if the second smallest eigenvalue of the (normalized) Laplacian matrix is 0. If G is disconnected, then $\phi(G) = 0$ and $\lambda_2 = 0$ so the above theorem is tight. Conversely, if $\lambda_2 = 0$, then G is disconnected and $\phi(G) = 0$. So, the above theorem can be seen as a robust version of this basic fact in spectral graph theory, that is $\lambda_2 \approx 0$ if and only if G is almost disconnected.

Next, we present an example where the right side of Cheeger's inequality is tight.

Example 1: The cycle graph We recall from the previous notes (Fact 11.3) that $\phi(G) = \frac{2}{n}$ when G is a cycle: we will show that in this case $\lambda_2 \lesssim \frac{1}{n^2}$. Consider the following vector x,

$$x_i = \begin{cases} \frac{1}{2} - \frac{2i}{n} & \text{if } i \leq n/2 \\ -\frac{1}{2} + \frac{2i}{n} & \text{otherwise.} \end{cases}$$

This means that for each pair of adjacent vertices $i, i+1$

$$(x_i - x_{i+1})^2 = \frac{4}{n^2}.$$

Since $x \perp \mathbf{1}$, we can write,

$$\lambda_2 \leq \left(\frac{x}{\|x\|}\right)^T \frac{L}{d} \left(\frac{x}{\|x\|}\right) \leq \frac{x^T L x}{d \cdot \|x\|^2} = \frac{n \cdot (4/n^2)}{d \cdot (n/2) \cdot (1/4)^2} = O(1/n^2),$$

where in last inequality we used that $|x_i| \geq 1/4$ for at least half of the vertices.

The above calculations show that (up to a constant factor) the right side of Cheeger's inequality is tight for a cycle. The left side of the Cheeger's inequality is tight for the d-dimensional hypercube $\{0,1\}^d$. In this case $\phi(H_d) = 1/d$ and $\lambda_2 = \Theta(1/d)$. We leave the details of the calculations as an exercise.

We now examine a different class of graphs in which this inequality is important.

Definition 12.2. *An expander graph is one in which the 2nd eigenvalue of the normalized laplacian is lower bounded by an absolute constant $\lambda_2 \geq \Omega(1)$*

The simplest example of an expander graph is the complete graph on n vertices. In this case it is not hard to see that $\lambda_2 = 1$. For an infinite sequence of graphs the above definitions mean that λ_2 remain bounded away from zero as the size of the graph grows. For a single graph it might be more intuitive to think of it as having $\lambda_2 > c$ for some constant c like $\frac{1}{4}$. Sparse expander graphs are important in many fields of computer science including optimization, coding theory and complexity theory. See PS5 for many examples.

Though specific construction of sparse expander graphs are hard to describe we can prove that a random d-regular graph is an expander graph with high probability. In other words, although (sparse) expander graphs are hard to construct explicitly, most of the graphs are expanders.

From a very highlevel point of view, expander graphs can be see as "sparse complete graphs"; for example if G is a d-regular expander, then for all $S \subseteq V$,

$$|E_G(S, \overline{S})| \approx \frac{d}{n} |E_{K_n}(S, \overline{S})|$$

where K_n is the complete graph over n vertices. Expander graphs have very small diameter and fast *mixing time* of random walks.

We now prove the left side of Cheeger's inequality.

Lemma 12.3. *Let L/d be the normalized Laplacian of a d-regular graph G, and λ_2 be the second smallest eigenvalue of L/d. Then*

$$\frac{\lambda_2}{4} \leq \phi(G)$$

Proof. For a vector $x \in \mathbb{R}^n$ let

$$R(x) = \frac{x^T \frac{L}{d} x}{x^T x} = \frac{x^T L x}{d x^T x} = \frac{\sum_{i \sim j} (x_i - x_j)^2}{d \cdot \sum_i x_i^2}$$

By variational characterization we can write,

$$\lambda_2 = \min_{x \perp \mathbf{1}} R(x) = \min_{x \perp \mathbf{1}} \frac{\sum_{i \sim j} (x_i - x_j)^2}{d \cdot \sum_{x_i^2}}. \tag{12.2}$$

Fix a set $S \subseteq V$ such that $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$. Next, we show that the conductance of a set $S \subseteq V$ is nothing but the Rayleigh quotient of the indicator vector of that set. Recall that $\mathbf{1}^S$ is the vector where $\mathbf{1}^S_i = 1$ if $i \in S$ and 0 otherwise. Then,

$$\phi(S) = \frac{|E(S, \overline{S})|}{\mathrm{vol}(S)} = \frac{|E(S, \overline{S})|}{d|S|} = \frac{\sum_{i \sim j} |\mathbf{1}^S_i - \mathbf{1}^S_j|}{d \sum_i \mathbf{1}^S_i},$$

where in the last equality we used that $|\mathbf{1}^S_i - \mathbf{1}^S_j| = 1$ if and only if i, j lie on opposite side of the $(S, \overline{S})$ cut. Now, observe that for all i, $\mathbf{1}^S_i = (\mathbf{1}^S_i)^2$ and for all i, j

$$|\mathbf{1}^S_i - \mathbf{1}^S_j| = (\mathbf{1}^S_i - \mathbf{1}^S_j)^2.$$

since the values can only take 0 or 1. Therefore,

$$\phi(S) = \frac{\sum_{i \sim j} |\mathbf{1}^S_i - \mathbf{1}^S_j|^2}{d \sum_i (\mathbf{1}^S_i)^2} = R(\mathbf{1}^S).$$

Using the above identity, (12.2) seems to imply that $\lambda_2 \leq \phi(S)$ which should directly conclude the proof. The catch is that $\mathbf{1}^S$ is not orthogonal to $\mathbf{1}$, the first eigenvector of L/d. This minor issue can be resolved by projecting $\mathbf{1}^S$ to the space orthogonal to $\mathbf{1}$.

Consider the vector

$$y = \mathbf{1}^S - \frac{\mathbf{1}}{\|\mathbf{1}\|} \langle \mathbf{1}^S, \frac{\mathbf{1}}{\|\mathbf{1}\|} \rangle = \mathbf{1}^S - \frac{|S|}{n}.$$

By construction, we have $\mathbf{1} \perp y$. In addition, the numerator of the Rayleigh quotient is shift-invariant, i.e., for any c,

$$|x_i - x_j|^2 = |(x_i + c) - (x_j + c)|^2,$$

so $|\mathbf{1}_i^S - \mathbf{1}_j^S|^2 = |y_i - y_j|^2$. To prove the lemma we show that $\|y\|^2 \geq \|\mathbf{1}^S\|^2/4$. Then, by (12.2)

$$\lambda_2 \leq R(y) = \frac{\sum_{i\sim j} |\mathbf{1}_i^S - \mathbf{1}_j^S|^2}{d \cdot \sum_i y_i^2} \leq \frac{\sum_{i\sim j} |\mathbf{1}_i^S - \mathbf{1}_j^S|^2}{\frac{d}{4} \cdot \sum_i (\mathbf{1}_i^S)^2} = 4\phi(S) \leq 4\phi(G).$$

which completes the proof.

So, it remains to show that $\|y\|^2 \geq \|\mathbf{1}^S\|^2$. Note that we still have not used the assumption that $\text{vol}(S) \leq \text{vol}(V)/2$, or equivalently that $|S| \leq |V|/2$. This implies that $|S| \leq n/2$. Therefore,

$$\|y\|^2 = \sum_i y_i^2 \geq \sum_{i \in S} y_i^2 = \sum_{i \in S} (1 - |S|/n)^2 \geq |S|/4 = \|\mathbf{1}^S\|/4.$$

This completes the proof of Lemma 12.3. □

The right side of Cheeger's inequality has a more involved proof: please see the "reading" section of the course page.

12.5 Relation to spectral partitioning

The above proof shows why the spectral partitioning algorithm generally does well when it comes to finding $\phi(G)$. Here is the highlevel idea: The second eigenvector of L/d is the vector x which minimizes $R(x)$ among all real-valued vectors (orthogonal to $\mathbf{1}$). On the other hand, as we saw in the proof of Lemma 12.3, $\phi(G)$ is the $\{0,1\}^n$ vector y which minimizes $R(y)$. Therefore, the second eigenvalue of the normalized Laplacian matrix approximates the conductance of G by formulating it as a continuous optimization problem. The spectral partitioning algorithm can be seen as a *rounding* of this continuous optimization problem. Given a real values vector $x \in \mathbb{R}^n$, we would like to construct a vector $y \in \{0,1\}^n$ such that $R(y) \approx R(x)$. Having such a y, the support of y, i.e., $\{i : y_i = 1\}$ would correspond to a low conductance cut. The spectral partitioning algorithm uses the most natural rounding idea: Choose a threshold t and let $y_i = 0$ if $x_i < t$ and let $y_i = 1$ otherwise. The algorithm chooses the best possible threshold by going over all possible thresholds. The following lemma which we state without a proof gives a bound on the performance of the spectral partitioning algorithm:

Lemma 12.4. *For any graph G, the spectral partitioning algorithm returns a set S such that* $\text{vol}(S) \leq \text{vol}(V)/2$ *and*

$$\phi(S) \leq \sqrt{2\lambda_2},$$

where as usual λ_2 is the second smallest eigenvalue of the normalized Laplacian matrix $\tilde{L}$.

Since by Lemma 12.3, $\phi(G) \geq \lambda_2/2$, $\phi(S) \leq O(\sqrt{\phi(G)})$. The proof of the above lemma also implies the right side of the Cheeger's inequality.

A tight example for spectral partitioning. As we mentioned earlier, the right side of the Cheeger's inequality is tight for a cycle of length n. But, it is not hard to see that the spectral partitioning algorithm returns the minimum conductance cut. So, a cycle is not a tight example for the spectral partitioning algorithm.

Let us conclude this lecture by giving a tight example for the spectral partitioning algorithm. Consider the weighted *ladder* graph shown in the left of Figure 12.1. This graph consists of two cycles $1, 2, \ldots, n/2$ and $n/2, n/2+1, \ldots, n$.

In this case the minimum conductance cut is shown at the middle of Figure 12.1 for which $\phi(G) \approx \frac{100}{n^2}$. However, recall that the Rayleigh quotient of the second eigenvector of a cycle of length n is no more than $20/n^2$. Therefore, the second eigenvector of the ladder graph is just two copies of the second eigenvector of a cycle of length $n/2$; in particular the endpoints of each dashed has the same value in the second eigenvector. It follows that the spectral partitioning algorithm cannot distinguish the inner cycle from the outer one, because the second eigenvector assigns the same number to the endpoints of each dashed edge. The cut at the right of Figure 12.1 shows the output of the spectral partitioning algorithm; in this case, although $\phi(G) = \Theta(1/n^2)$, the spectral partitioning algorithm returns a set S with $\phi(S) = \Theta(1/n)$.

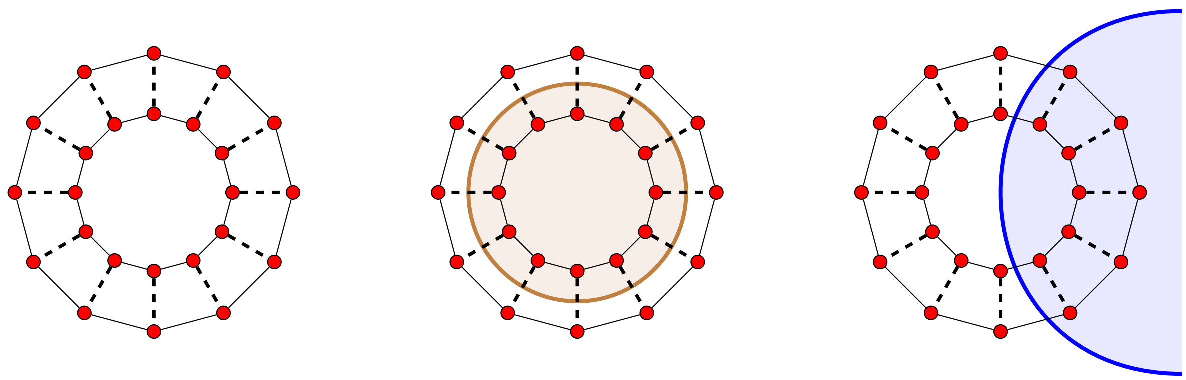

Figure 12.1: An illustration of the Ladder graph. Here each solid edge has weight 1 and each dashed edge has weight $100/n^2$ where $n/2$ is the length of the cycles. The middle figure shows the set with the optimum conductance for a sufficiently large n, and the right figure shows the output of the spectral partitioning algorithm.

In the next lecture we briefly discuss an extension of the spectral partitioning algorithm based on the first k eigenvectors of the normalized Laplacian matrix.

References

[Alo86] N. Alon. “Eigenvalues and expanders”. In: *Combinatorica* 6 (2 Jan. 1986), pp. 83–96 (cit. on p. 12-4).

[AM85] N. Alon and V. Milman. “Isoperimetric inequalities for graphs, and superconcentrators”. In: *Journal of Combinatorial Theory, Series B* 38.1 (Feb. 1985), pp. 73–88 (cit. on p. 12-4).