COMS 4995-3: Advanced Algorithms March 20, 2017

Lecture 15 – Spectral graph algorithms: Cheeger's Inequality

Instructor: *Alex Andoni* Scribes: *Rishabh Dudeja (rd2714)*

1 Introduction

In the last lecture we studied the relation between the combinatorial properties of a graph $G = (V, E)$ with n vertices and m edges. and the spectral properties of the normalized adjacency matrix $\hat{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. Let $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$ be the ordered eigenvalues of the normalized adjacency matrix. We saw that $\lambda_1 = 1$. Furthermore $\lambda_2 = 1$ iff the graph is disconnected. And finally $\lambda_n = -1$ iff the graph is bipartite. In this lecture we will introduce the Laplacian Matrix of a Graph and describe how (and why) it can be used for visualizing graphs. We will then prove a result called the Cheeger's Inequality which connects the spectral properties of the graph Laplacian with a combinatorial notion of degree of connectedness of the graph.

2 The Laplacian Matrix

Let $G = (V, E)$ be a graph with n nodes and m edges. Let A be its adjacency matrix. Let D be the diagonal matrix with $D_{ii} = d_i$, the degree of node i.

Definition 1 (Graph Laplacian). *The Laplacian of a graph $G = (V, E)$ with adjacency matrix A is $L := D - A$.*

We note that $L_{ii} = d_i$ the degree of node i and for any $i \neq j$, $L_{ij} = -1$ iff $(i, j) \in E$. It will be helpful to decompose the laplacian matrix L as a sum of m matrices L_e corresponding to each edge $e \in E$. For any edge $e = (i, j) \in E$ we define L_e as a $n \times n$ matrix with all zero entries except: $(L_e)_{ii} = 1, (L_e)_{jj} = 1, (L_e)_{ij} = (L_e)_{ji} = -1$. Using this decomposition we can write a quadratic form of the Laplacian in a particularly nice form:

Lemma 2 (Quadratic Forms of Laplacian). *For any $x \in \mathbb{R}^n$, $x^T L x = \sum_{(i,j) \in E} (x_i - x_j)^2$*

Proof. We write $L = \sum_{e=(i,j) \in E} L_e$. And hence $x^T L x = \sum_{e=(i,j) \in E} x^T L_e x = \sum_{e=(i,j) \in E} (x_i^2 + x_j^2 - 2x_i x_j) = \sum_{e=(i,j) \in E} (x_i - x_j)^2$ □

Recall that in the last lecture we proved a result relating the combinatorial properties of a graph like connectivity, bipartiteness to the spectrum of the normalized adjacency matrix. An analogous result holds for the spectrum of the Laplacian Matrix.

Theorem 3 (Spectrum of Graph Laplacian). *Let $\mu_1 \leq \mu_2 \cdots \leq \mu_n$ be the eigenvalues of the Laplacian Matrix L. Let $v_1, v_2 \ldots v_n$ be the corresponding eigenvectors. Then, $\mu_1 = 0$ and $v_1 = \frac{1}{\sqrt{n}}(1, 1, \ldots 1)$. Furthermore, $\mu_2 = 0$ iff the graph is disconnected.*

Proof. Skipped, Analogous to the proof from the last lecture. □

3 Laplacian Matrix and Graph Visualization

The second and third eigenvalues of the Laplacian Matrix can be used to create interesting visualizations of the under lying graph. Let v_2 and v_3 denote the eigenvector corresponding to the second and third smallest eigenvalues of the graph laplacian L. A simple heuristic to visualize the graph G in $\mathbb{R}^2$ proceeds as follows: For each node $i \in V$, plot node i at the coordinate $(v_{2,i}, v_{3,i})$ where $v_{2,i}, v_{3,i}$ refer the i-th entry of the n-dimensional vectors v_2, v_3.

Why does this heuristic give interesting visualizations of the graph?

To try and understand this better, let us first imagine that we want to visualize the graph in $\mathbb{R}$ instead of $\mathbb{R}^2$. Let $x \in \mathbb{R}^n$ be the vector representing the position of each node in the graph. A good visualization of the graph would place the connected nodes together and the unconnected nodes further apart. This desired property can be formulated as an optimization problem as follows:

Option 1. *Choose the coordinates as the solution of the optimization problem:* $\min_{x \in \mathbb{R}^n} \sum_{(i,j)\in E} (x_i - x_j)^2$

This option sets $x = 0$. This is not desirable. To avoid this degenerate solution, we can further impose the constraint $\|x\|_2 = 1$.

Option 2. *Choose the coordinates as the solution of the optimization problem:* $\min_{x \in \mathbb{R}^n} \sum_{(i,j)\in E} (x_i - x_j)^2$ *subject to* $\|x\|_2 = 1$.

Using Lemma 2, we know that $\sum_{(i,j)\in E}(x_i - x_j)^2 = x^T Lx$. Furthermore by the theorem on Rayleigh Quotient from the last lecture, we know that v_1 minimizes $x^T Lx$ subject to $\|x\|_2 = 1$. Since $v_1 = \frac{1}{\sqrt{n}}(1, 1 \ldots 1)$, Option 2 sets $x = \frac{1}{\sqrt{n}}(1, 1 \ldots 1)$. Again this isn't desirable. The coordinates of Option 2 are merely a translation of the coordinates of Option 2. To prevent such translation we additionally impose the constraint $\sum_{i=1}^{n} x_i = 0$. Alternatively since $v_1 = \frac{1}{\sqrt{n}}(1, 1, \ldots 1)$, the constraint can be written as $x^T v_1 = 0$:

Option 3. *Choose the coordinates as the solution of the optimization problem:* $\min_{x \in \mathbb{R}^n} \sum_{(i,j)\in E} (x_i - x_j)^2$ *subject to* $\|x\|_2 = 1$, $x^T v_1 = 0$

Again by the Rayleigh Quotient description of the second eigenvector, we know that Option 3 sets $x = v_2$. Thus Option 3 doesn't suffer from the problems of Option 1 and Option 2 which plotted all nodes at the same coordinate. We can try to generalize Option 3 to get a visualization in $\mathbb{R}^2$.

Let $x, y \in \mathbb{R}^n$ represent the vector of x and y coordinates of the nodes. That is, Node i is plotted at the point (x_i, y_i). Taking cue from the 1-Dimensional visualization, we will try to minimize the sum of squared distances between the locations of directly connected nodes. This gives us the following Option:

Option 4. *Choose the coordinates as the solution of the optimization problem:* $\min_{x \in \mathbb{R}^n} \sum_{(i,j)\in E} (x_i - x_j)^2 + (y_i - y_j)^2$ *s.t.* $\|x\|_2 = 1, \|y\|_2 = 1$ *and* $x^T v_1 = 1, y^T v_1 = 0$.

This option sets $x = y = v_2$. This is because $\sum_{(i,j)\in E}(x_i - x_j)^2 + (y_i - y_j)^2 = x^T Lx + y^T Ly$. For all x such that $\|x\|_2 = 1, x^T v_1 = 0$, we know $x^T Lx \geq v_2^T Lv_2$. Likewise, for all y s.t. $\|y\|_2 = 1, y^T v_1 = 0$, $y^T Ly \geq v_2^T Lv_2$. Hence the optimal value of the objective is lower bounded by $2v_2^T Lv_2$. Furthermore the lower bound is attained when $x = y = v_2$. Hence $x = y = v_2$ is an optimal solution. However this is not desirable, because Option 4 doesn't use all of the two dimensions available for visualization. To avoid this we impose the additional constraint $x^T y = 0$ to ensure the second coordinate captures information which is not already captured by the first coordinate.

Option 5. *Choose the coordinates as the solution of the optimization problem:* $\min_{x\in\mathbb{R}^n}\sum_{(i,j)\in E}(x_i - x_j)^2 + (y_i - y_j)^2$ *s.t.* $\|x\|_2 = 1, \|y\|_2 = 1$ *and* $x^T v_1 = 1, y^T v_1 = 0, x^T y = 0$.

It can be shown that Option 5 sets $x = v_2, y = v_3$. This is precisely the heuristic proposed in the beginning of the section.

4 Cheeger's Inequality

We have seen that the second smallest eigenvalue μ_2 of the graph laplacian is 0 iff the graph is disconnected. However μ_2 is a real number. Is it true that if $\mu_2 >> 0$ then the graph is well-connected in some sense and if $\mu_2 \approx 0$ then the graph is not well connected? Cheeger's Inequality is a result of this flavor.

To state Cheeger's Inequality we will first define some new quantities.

Definition 4 (Normalized Laplacian Matrix). $\hat{L} := D^{-\frac{1}{2}}LD^{-\frac{1}{2}} = D^{-\frac{1}{2}}(D-A)D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} = I - \hat{A}$

We recall that $\hat{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the normalized adjacency matrix whose spectral properties we studied in the last lecture. Let $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$ be the eigen-values of $\hat{A}$ and $v_1, v_2 \ldots v_n$ be the corresponding eigenvectors. Since $\hat{L} = I - \hat{A}$, $v_1, v_2 \ldots v_n$ are also the eigenvectors of $\hat{L}$ but with the eigenvalues $\mu_1 = 1-\lambda_1 \leq \mu_2 = 1-\lambda_2 \cdots \leq \mu_n = 1-\lambda_n$. From the last lecture we know that $\lambda_1 = 1$, hence the smallest eigenvalue of $\hat{L}$, $\mu_1 = 0$. Furthermore since $\lambda_2 = 1$ iff G is disconnected. This means that $\mu_2 = 0$ iff G is disconnected. Finally since $\lambda_n \geq -1$ and $\lambda_n = -1$ iff the graph is bipartite, we conclude that $\mu_n \leq 2$ and $\mu_n = 2$ iff the graph is bipartite. These results are summarized in the observation below:

Observation 5. *Let* $\mu_1 \leq \mu_2 \cdots \leq \mu_n$ *be the eigenvalues of* $\hat{L}$ *and* $v_1, v_2 \ldots v_n$ *be the corresponding eigenvectors, then* $\mu_1 = 0$, $\mu_n \leq 2$ *and* $v_1 = (\sqrt{d_1}, \sqrt{d_2} \ldots \sqrt{d_n})$.

Definition 6 (Cuts). *Any* $S \subset V$ *is called a Cut.*

Definition 7 (Boundary of a Cut). $\partial(S) := \{(i,j) \in E : i \in S, j \in \bar{S}\}$

The size of the boundary of a Cut S captures how well connected the set of nodes S is to the rest of the graph. However $|\partial(S)|$ might be small simply because there are very few nodes in S. Hence we will need to normalize the size of the size of the boundary of a cut by some measure of the size of a cut (called volume of a cut). This normalized ratio is called the conductance of a cut.

Definition 8 (Volume of a Cut). $Vol(S) = \sum_{i\in S} d_i$

Definition 9 (Conductance of a Cut). $\phi(S) = \frac{\partial S}{Vol(S)}$

It is easy to see that $\phi(S) = \frac{\text{\# External Edges}}{2(\text{\# Internal Edges }) + \text{\# External Edges}}$. Hence $\phi(S)$ is low if there are very few edges from S to $\bar{S}$ in comparison to the number of edges within S. Hence it captures how isolated a set of nodes is from the remaining graph. Using this notion of conductance of a cut, we can define a notion of global connectivity of a graph:

Definition 10 (Conductance of a Graph G).

$$\phi(G) := \min_{S\subset V, 0<Vol(S)\leq\frac{1}{2}Vol(V)} \phi(S)$$

Intuitively, the conductance of a graph G is small, if there exists extremely isolated (with small conductance) cuts inside the graph. The restriction that $Vol(S) > 0$ is to prevent the case that $S = \{\}$ in which case the conductance is not defined. The constraint $Vol(S) \leq \frac{1}{2}Vol(V)$ is to prevent the cases like $S = V \setminus \{i\}$. For such a cut, $\phi(S) = \frac{d_i}{\sum_{j \neq d_i} d_j}$ which is expected to be small for most reasonable graphs (and hence is not very interesting).

Remark 1. *An equivalent definition of conductance of a cut is* $\phi(S) = \frac{|\partial S|}{\min(Vol(S), Vol(\bar{S}))}$. *Correspondingly the conductance of a graph is then defined as* $\phi(G) = \min_{S \neq \phi} \phi(S)$

Remark 2. *Computing either* $\phi(G)$ *or the* S *that is the solution of the optimization problem defining* $\phi(S)$ *is NP-Hard. Even a constant factor approximation for this problem is not known.*

Remark 3. *One could come up with alternative ways of capturing the intuition behind Volume of a cut, Conductance of a cut etc. The advantage of defining them this way is that results relating combinatorial properties of the graph and linear algebraic properties of the laplacian can be proven.*

We now state Cheeger's Inequality.

Theorem 11 (Cheeger's Inequality)**.**

$$\frac{\mu_2}{2} \leq \phi(G) \leq \sqrt{2\mu_2}$$

Before proving the result we first note that Cheeger's Inequality allows us to tightly bound (both from above and below) $\phi(G)$ the conductance of a graph (a combinatorial notion) using μ_2 (a linear algebraic notion) and vice-versa. Intuitively it states that μ_2 is small iff $\phi(G)$ is small (or the graph is not well-connected). Furthermore if $\mu_2 = 0 \implies \phi(G) = 0$. Said differently there exists a cut with zero conductance. This is possible iff the graph is disconnected (taking S as one of the components gives $\phi(S) = 0$). Hence we recover the result we proved in the last lecture.

In today's lecture we will only prove the lower bound on $\phi(G)$. This is easier to prove than the upper bound.

Proof. We need to show:

$$\phi(G) := \min_{S \subset V, 0 < Vol(S) \leq \frac{1}{2}Vol(V)} \phi(S) \geq \frac{\mu_2}{2}$$

It is sufficient to show for any fixed S such that $0 < Vol(S) \leq \frac{1}{2}Vol(V)$:

$$\phi(S) \geq \frac{\mu_2}{2}$$

Fix such a cut S.

Next we recall the Rayleigh Quotient characterization of μ_2:

$$\mu_2 = \min_{x \neq 0, x^T v_1 = 0} \frac{x^T \hat{L} x}{x^T x} = \min_{x \neq 0, x^T v_1 = 0} \frac{x^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} x}{x^T x}$$

We perform the substitution $y = D^{-\frac{1}{2}} x$:

$$\mu_2 = \min_{y \neq 0, v_1^T D^{\frac{1}{2}} y = 0} \frac{y^T L y}{y^T D y}$$

Hence to prove that $2\phi(S) \geq \mu_2 = \min_{y\neq 0, v_1^T D^{\frac{1}{2}} y = 0} \frac{y^T L y}{y^T D y}$, it is sufficient to find a y satisfying: $y \neq 0, v_1^T D^{\frac{1}{2}} y = 0$ and $2\phi(S) \geq \frac{y^T L y}{y^T D y}$.

A natural choice of $y = 1_S$ that is, the indicator vector for cut S ($y_i = 1$ iff $i \in S$). However this does not satisfy $v_1^T D^{\frac{1}{2}} y = 0$. To fix this we use $y = 1_S - \sigma e$ where $\sigma := \frac{Vol(S)}{Vol(V)}$ where $e = (1, 1, \dots 1)$. Doing so ensures $v_1^T D^{\frac{1}{2}} y = 0$ as can be verified below:

$$\begin{aligned} y^T (D^{\frac{1}{2}} v_1) &= \langle y, D^{1/2} v_1 \rangle \\ &= \langle 1_S, D^{1/2} v_1 \rangle - \frac{Vol(S)}{Vol(V)} \langle e, D^{1/2} v_1 \rangle \end{aligned}$$

Next we recall that $v_1 = (\sqrt{d_1} \dots \sqrt{d_n})$ and $D = Diag(d_1, \dots d_n)$. Hence $D^{\frac{1}{2}} v_1 = (d_1, d_2 \dots d_n)$. Hence:

$$\begin{aligned} y^T (D^{\frac{1}{2}} v_1) &= \langle 1_S, D^{1/2} v_1 \rangle - \frac{Vol(S)}{Vol(V)} \langle e, D^{1/2} v_1 \rangle \\ &= \sum_{i \in S} d_i - Vol(S) \\ &= 0 \end{aligned}$$

Now we are just left to verify $2\phi(S) \geq \frac{y^T L y}{y^T D y}$. We first simplify $y^T L y$ and $y^T D y$ separately.

$$\begin{aligned} y^T L y &= (1_S - \sigma e)^T L (1_S - \sigma e) \\ &= 1_S^T L 1_S && \text{[Since Le = 0, e is an eigenvector with 0 eigenvalue for L]} \\ &= \sum_{(i,j) \in E} ((1_S)_i - (1_S)_j)^2 \\ &= \sum_{(i,j) \in \partial(S)} 1 \\ &= |\partial(S)| \end{aligned}$$

Next we simplify $y^T D y$:

$$\begin{aligned} y^T D y &= 1_S^T D 1_S + \sigma^2 e^T D e - 2\sigma 1_S^T D e \\ &= Vol(S) + \sigma^2 Vol(V) - 2\sigma Vol(S) \\ &= Vol(S) \left(1 - \frac{Vol(S)}{Vol(V)} \right) \end{aligned}$$

Combining the expressions for $y^T L y$ and $y^T D y$:

$$\frac{y^T L y}{y^T D y} = \frac{\phi(S)}{1 - \frac{Vol(S)}{Vol(V)}}$$

The condition we needed to check was $2\phi(S) \geq \frac{y^T L y}{y^T D y} \Leftrightarrow 1 - \frac{Vol(S)}{Vol(V)} \geq \frac{1}{2} \Leftrightarrow Vol(S) \leq \frac{1}{2} Vol(V)$ which was one of the assumptions on S. $\square$