Spectral Graph Theory and Algorithms, Summer 2017 26 June 2017

Lecture 6

Lecturer: David P. Williamson

1 Recap

Let $S \subset V$. Recall that $\delta(S)$ denotes the set of edges with exactly one endpoint in S, and define $\text{vol}(S) \equiv \sum_{i \in S} d(i)$. Last week we defined the **conductance** of $S \subset V$ to be

$$\phi(S) \equiv \frac{|\delta(S)|}{\min\{\text{vol}(S), \text{vol}(V - S)\}},$$

and the conductance of G to be

$$\phi(G) = \min_{S \subset V} \phi(S).$$

We noticed that $0 \leq \phi(S) \leq 1$ for all $S \subset V$.

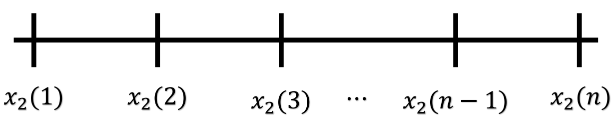
We call a graph G an *expander* if $\phi(G)$ (or $\alpha(G)$) is "large" (i.e. a constant[1]). Otherwise, we say that G has a sparse cut.

We gave an algorithm for finding a sparse cut that works well in practice, but that lacks strong theoretical guarantees, called **spectral partitioning**. In the following, $\mathscr{L}$ is the "normalized Laplacian". We'll define later what that is, but it's related to our usual Laplacian L_G.

Algorithm 1: Spectral Partitioning

Compute x_2 of $\mathscr{L}$ (the eigenvector corresponding to $\lambda_2(\mathscr{L})$);
Sort V such that $x_2(1) \leq \cdots \leq x_2(n)$.;
Define the **sweep cuts** for $i = 1, ..., n-1$ by $S_i \equiv \{1, ..., i\}$.;
Return $\min_{i \in \{1,...,n-1\}} \phi(S_i)$.;

The following picture illustrates the idea of the algorithm; sweep cuts correspond to cuts between consecutive bars:



Today we will prove the following for d-regular graphs. The proof is a bit more complicated for general undirected graphs, but all the essential ideas appear in the proof for d-regular graphs.

[0]This lecture is derived from Lau's 2015 notes, Lectures 3 and 4, `https://cs.uwaterloo.ca/~lapchi/cs798-2015/notes/L03.pdf` and `https://cs.uwaterloo.ca/~lapchi/cs798-2015/notes/L04.pdf`. Scribes were Sam Gutekunst and Victor Reis.

[1]One should then ask "A constant with respect to what?" Usually one defines families of graphs of increasing size as families of expanders, in which case we want the conductance or expansion constant with respect to the number of vertices.

Theorem 1 (Cheeger's Inequality) *Let λ_2 be the second smallest eigenvalue of $\mathscr{L}$. Then:*

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

The theorem proved by Jeff Cheeger actually has to do with manifolds and hypersurfaces; the theorem above is considered to be a discrete analog of Cheeger's original inequality. But the name has stuck.

2 Normalized Adjacency and Laplacian Matrices

To start our proof, we introduce normalized adjacency and Laplacian matrices. We use notation from Lap Chi Lau.

Definition 1 *The* **normalized adjacency matrix** *is*

$$\mathscr{A} \equiv D^{-1/2} A D^{-1/2},$$

where A is the adjacency matrix of G and $D = diag(d)$ for $d(i)$ the degree of node i.

For a graph G (with no isolated vertices), we can see that

$$D^{-1/2} = \begin{pmatrix} \frac{1}{\sqrt{d(1)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{d(2)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{d(n)}} \end{pmatrix}.$$

Definition 2 *The* **normalized Laplacian matrix** *is*

$$\mathscr{L} \equiv I - \mathscr{A}.$$

Notice that $\mathscr{L} = I - \mathscr{A} = D^{-1/2}(D - A)D^{-1/2} = D^{-1/2} L_G D^{-1/2}$, for L_G the (unnormalized) Laplacian.

Recall that for the largest eigenvalue λ of A and Δ the maximum degree of a vertex in a graph, $d_{avg} \leq \lambda \leq \Delta$. "Normalizing" the adjacency matrix makes its largest eigenvalue 1, so the analogous result for normalized matrices is the following:

Claim 2 *Let $\alpha_1 \geq \cdots \geq \alpha_n$ be the eigenvalues of $\mathscr{A}$ and let $\lambda_1 \leq \cdots \lambda_n$ be the eigenvalues of $\mathscr{L}$. Then*

$$1 = \alpha_1 \geq \cdots \geq \alpha_n \geq -1, \qquad 0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2.$$

Proof: First, we show that 0 is an eigenvalue of $\mathscr{L}$ using the vector $x = D^{-1/2} e$. Then

$$\mathscr{L}(D^{1/2} e) = D^{-1/2} L_G D^{-1/2} D^{1/2} e = D^{-1/2} L_G e = 0,$$

since e is a eigenvector of L_G corresponding to eigenvalue 0. This shows that $D^{1/2} e$ is an eigenvector of $\mathscr{L}$ of eigenvalue 0. To show that it's the smallest eigenvalue, notice that $\mathscr{L}$ is positive semidefinite. We use the fact that L_G is positive semidefinite. Thus $L_G = BB^T$ for some B, so that $\mathscr{L} = VV^T$ for $V = D^{-1/2} B$.

To show that $\alpha_1 \leq 1$, we make use of the positive semidefiniteness of $\mathscr{L} = I - \mathscr{A}$. This gives us that, for all $x \in \mathbb{R}^n$:

$$x^T(I - \mathscr{A})x \geq 0 \implies x^T x - x^T \mathscr{A} x \geq 0 \implies 1 \geq \frac{x^T \mathscr{A} x}{x^T x}. \tag{1}$$

This Rayleigh quotient gives us the upper bound that $\alpha_1 \leq 1$. To get equality, consider again $x = D^{1/2}e$. Since, for this x,

$$x^T \mathscr{L} x = 0 \implies x^T(I - \mathscr{A})x = 0.$$

The exact same steps as in Equation 1 yield $\frac{x^T \mathscr{A} x}{x^T x} = 1$, as we now have equality.

By following the proof of the Perron-Frobenius theorem, we can show that $|\alpha_n| \leq \alpha_1 = 1$, so that $\alpha_n \geq -1$. Thus

$$\frac{x^T \mathscr{A} x}{x^T x} \geq -1,$$

which implies that

$$-x^T \mathscr{A} x \leq x^T x \implies x^T I x - x^T \mathscr{A} x \leq 2x^T x \implies \frac{x^T \mathscr{L} x}{x^T x} \leq 2 \implies \lambda_n \leq 2,$$

using the same Rayleigh quotient trick and that λ_n is the maximizer of that quotient. $\square$

Remark 1 *Notice that, given the spectrum of $\mathscr{A}$, we have the following: $-\mathscr{A}$ has spectrum negatives of $\mathscr{A}$, and $I - \mathscr{A}$ adds one to each eigenvalue of $-\mathscr{A}$. Hence, $0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$ follows directly from $1 = \alpha_1 \geq \cdots \geq \alpha_n \geq -1$.*

3 Cheeger's Inequality

We now can prove the ineuqlaity

Theorem 3 (Cheeger's Inequality) *Let λ_2 be the second smallest eigenvalue of $\mathscr{L}$. Then:*

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

Typically, people think of the first inequality being "easy" and the second being "hard." We'll prove the easy inequality first, and then the hard one. In both cases we'll limit ourselves to d-regular graphs. Notice that in the case of d-regular graphs, we have that $\mathscr{L} = \frac{1}{d}L_G$ and $\mathscr{A} = \frac{1}{d}A$.

Proof: Note that

$$\mathscr{L}e = \frac{1}{d}L_G e = 0,$$

so that e is an eigenvector of eigenvalue 0.

Now let S^* be such that $\phi(G) = \phi(S^*)$, and try defining

$$x(i) = \begin{cases} 1, & i \in S^* \\ -1, & \text{else.} \end{cases}$$

Then we would like the proof to proceed as follows:

$$\begin{aligned}\lambda_2 = \min_{z:<z,e>=0} \frac{z^T \mathscr{L} z}{z^T z} &\le \frac{x^T \mathscr{L} x}{x^T x} \\ &= \frac{\frac{1}{d} x^T L_G x}{x^T x} \\ &= \frac{\frac{1}{d}\sum_{(i,j)\in E}(x(i) - x(j)^2}{\sum_{i\in V} x(i)^2} \\ &= \frac{4|\delta(S^*)|}{dn}.\end{aligned}$$

There are two problems with this idea, though. First, we don't know whether $< x, e >= 0$. Second, the denominator is wrong: we'd like it to be $\min(\mathrm{vol}(S^*), \mathrm{vol}(V - S^*))$.

Hence, we redefine

$$x(i) = \begin{cases} \frac{1}{|S^*|}, & i \in S^* \\ -\frac{1}{|V-S^*|}, & \text{else.} \end{cases}$$

Now we notice that

$$< x, e >= |S^*| \cdot \frac{1}{|S^*|} + |V - S^*| \cdot \frac{1}{|V - S^*|} = 0.$$

Furthermore, following the proof above, we get that

$$\begin{aligned}\lambda_2 &\le \frac{\frac{1}{d}|\delta(S^*)| \left(\frac{1}{|S^*|} + \frac{1}{|V-S^*|}\right)^2}{|S^*| \left(\frac{1}{|S^*|}\right)^2 + |V - S^*| \left(\frac{1}{|V-S^*|}\right)^2} \\ &= \frac{\frac{1}{d}|\delta(S^*)| \left(\frac{1}{|S^*|} + \frac{1}{|V-S^*|}\right)^2}{\frac{1}{|S^*|} + \frac{1}{|V-S^*|}} \\ &= \frac{|\delta(S^*)|n}{d|S^*||V - S^*|} \\ &\le \frac{2|\delta(S^*)|}{\min\{\mathrm{vol}(S^*), \mathrm{vol}(V - S^*)\}} \\ &= 2\phi(G),\end{aligned}$$

recalling that for G d-regular, then $\mathrm{vol}(S) = d|S|$. □

This completes the proof of the first inequality.

Recall $\mathrm{supp}(x) = \{i \in V : x(i) \neq 0\}$, $\mathrm{supp}^+(x) = \{i \in V : x(i) > 0\}$, and $\mathrm{supp}^-(x) = \{i \in V : x(i) < 0\}$, and define

$$R(x) \equiv \frac{x^T \mathscr{L} x}{x^T x}.$$

Let x_2 is the eigenvector corresponding to λ_2 of $\mathscr{L}$. Suppose without loss of generality that $|\mathrm{supp}^+(x_2)| \le |\mathrm{supp}^-(x_2)|$, so that $|\mathrm{supp}^+(x_2)| \le n/2$. Define y so that $y(i) = x_2(i)$ when $x_2(i) \ge 0$, and $y(i) = 0$ otherwise. We will show below that $R(y) \le R(x_2)$ and then show that we can find a set $S \subset \mathrm{supp}(y)$ such that

$$\frac{|\delta(S)|}{d|S|} \le \sqrt{2R(y)}.$$

Then since $|S| \le n/2$, we will have that

$$\phi(G) \le \phi(S) = \frac{|\delta(S)|}{d|S|} \le \sqrt{2R(y)} \le \sqrt{2R(x_2)} = \sqrt{2\lambda_2},$$

as desired.

Claim 4 $R(y) \le R(x_2)$

Proof: We note that for all i such that $y(i) > 0$,

$$\begin{aligned}(\mathscr{L}y)(i) = y(i) - \frac{1}{d}\sum_{j:(i,j)\in E} y(j) &\le x_2(i) - \frac{1}{d}\sum_{j:(i,j)\in E} x_2(j)\\ &= (\mathscr{L}x_2)(i)\\ &= \lambda_2 x_2(i).\end{aligned}$$

Therefore we have that

$$\begin{aligned}y^T\mathscr{L}y &= \sum_{i\in V} y(i)\cdot(\mathscr{L}y)(i)\\ &\le \sum_{i:y(i)>0} x_2(i)\cdot(\lambda_2 x_2(i))\\ &= \sum_{i:y(i)>0} \lambda_2(x_2(i))^2 = \sum_{i\in V}\lambda_2(y(i))^2.\end{aligned}$$

Thus

$$\frac{y^T\mathscr{L}y}{y^Ty} \le \lambda_2 = \frac{x_2^T\mathscr{L}x_2}{x_2^Tx_2}.$$

□

Now we prove the following, which, as argued above, implies the second inequality in Cheeger's Inequality.

Lemma 5 *For any y, there is $S \subseteq \mathrm{supp}(y)$ such that $|\delta(S)|/|S| \le \sqrt{2R(y)}$.*

Proof: Without loss of generality, we assume $-1 \le y(i) \le 1$, as we can scale y if not. Our trick (from Trevisan) is to pick $t \in (0,1]$ uniformly at random, and let $S_t = \{i \in V : y(i)^2 \ge t\}$. Notice that:

$$\mathbb{E}[|S_t|] = \sum_{i\in V}\Pr[i\in S_t] = \sum_{i\in V} y(i)^2,$$

and assuming[2] that $(i,j)\in E \implies y(i)^2 \le y(j)^2$,

$$\mathbb{E}[|\delta(S_t)|] = \sum_{(i,j)\in E}\Pr[(i,j)\in\delta(S_t)] = \sum_{(i,j)\in E}\Pr[y(i)^2 < t \le y(j)^2] = \sum_{(i,j)\in E}(y(j)^2 - y(i)^2).$$

[2]We make this assumption without loss of generality because it doesn't matter in the end and is notationally convenient.

Rewriting the above using difference of squares and using Cauchy-Schwarz,

$$\begin{aligned}
\sum_{(i,h)\in E}(y(j)-y(i))(y(j)+y(i)) &\le \sqrt{\sum_{(i,j)\in E}(y(j)-y(i))^2\sum_{(i,j)\in E}(y(j)+y(i))^2} \\
&\le \sqrt{\sum_{(i,j)\in E}(y(j)-y(i))^2}\sqrt{2\sum_{(i,j)\in E}(y(j)^2+y(i)^2)} \\
&= \sqrt{\sum_{(i,j)\in E}(y(j)-y(i))^2}\sqrt{2d\sum_{i\in V}y(i)^2} \\
&= \sqrt{2R(y)}\cdot d\sum_{i\in V}y(i)^2.
\end{aligned}$$

This gives that

$$\frac{\mathbb{E}[|\delta(S_t)|]}{\mathbb{E}[d|S_t|]} \le \sqrt{2R(y)} \implies \mathbb{E}[|\delta(S_t)| - \sqrt{2R(y)}\cdot d|S_t|] \le 0.$$

This means that there exists a t such that

$$\frac{|\delta(S_t)|}{d|S_t|} \le \sqrt{2R(y)}.$$

To derandomize the algorithm, look at each of the n possible cuts S_t by looking at sweep cuts for the order $y(1)^2 \le y(2)^2 \le \cdots \le y(n)^2$. Thus just by looking at the sweep cuts, we will find some set S such that $\phi(S) \le \sqrt{2\lambda_2}$. $\square$

4 Bounds on largest eigenvalue

We proved earlier that $\lambda_n \le 2$ for the normalized Laplacian. Note that

$$\lambda_n = \max_{x\in\mathbb{R}^n}\frac{x^\top \mathscr{L} x}{x^\top x} = \max_{x\in\mathbb{R}^n}\frac{x^\top D^{-1/2}L_G D^{-1/2}x}{x^\top x} = \max_{y\in\mathbb{R}^n}\frac{y^\top L_G y}{y^\top D y},$$

where we take $y = D^{-1/2}x$. We also claim the following

Claim 6 $\lambda_n = 2$ *if and only if* G *has a bipartite component.*

We can easily show the if direction. If G has a bipartite component S with sides L, R, define a vector $y \in \mathbb{R}^n$ as $y(i) = 1$ if $i \in L$, $y(i) = -1$ if $i \in R$ and $y(i) = 0$ otherwise.

If $\delta(A, B)$ denotes the set of edges with one endpoint in A and another in B, we have

$$\frac{y^\top L_G y}{y^\top D y} = \frac{\sum_{(i,j)\in E}(y(i)-y(j))^2}{\sum_{i\in V} d(i)y(i)^2} = \frac{4|\delta(L,R)|}{\mathrm{vol}(S)} = 2.$$

Now we'll show a statement stronger than the converse: G has a bipartite component when $\lambda_n = 2$, and has an "almost" bipartite component when λ_n is close to 2. To make this more precise, consider the quantity

$$\beta(G) = \min_{\substack{S\subseteq V\\ S=L\cup R\\ L\cap R=\emptyset}} \frac{2|E(L)| + 2|E(R)| + |\delta(S)|}{\mathrm{vol}(S)},$$

where $E(X)$ denotes the set of edges with both endpoints in X. Alternatively,

$$\beta(G) = \min_{y \in \{-1,0,1\}^n} \frac{\sum_{(i,j)\in E} |y(i) + y(j)|}{\sum_{i \in V} d(i)|y(i)|},$$

where $L = \{i : y(i) = 1\}, R = \{i : y(i) = -1\}$ and $S = L \cup R$.

Since λ_n is the largest eigenvalue of $\mathscr{L}$, $\beta_n = 2 - \lambda_n$ is the smallest eigenvalue of $2I - \mathscr{L} = 2I - (I - \mathscr{A}) = I + \mathscr{A}$. Hence

$$\beta_n = \min_{x \in \mathbb{R}^n} \frac{x^\top (I + \mathscr{A}) x}{x^\top x} = \min_{x \in \mathbb{R}^n} \frac{x^\top D^{-1/2}(D + \mathscr{A})D^{-1/2} x}{x^\top x} = \min_{y \in \mathbb{R}^n} \frac{y^\top (D + A) y}{y^\top D y};$$

that is,

$$\beta_n = \min_{y \in \mathbb{R}^n} \frac{\sum_{(i,j)\in E} (y(i) + y(j))^2}{\sum_{i \in V} d(i) y(i)^2}.$$

Trevisan proves the following very nice analogy to the Cheeger inequality.

Theorem 7 (Trevisan 2009)

$$\frac{1}{2}\beta_n \le \beta(G) \le \sqrt{2\beta_n}.$$

Proof: For the first inequality, simply note that

$$\begin{aligned}
\beta_n = \min_{y \in \mathbb{R}^n} \frac{\sum_{(i,j)\in E} (y(i) + y(j))^2}{\sum_{i \in V} d(i) y(i)^2} &\le \min_{y \in \{-1,0,1\}^n} \frac{\sum_{(i,j)\in E} (y(i) + y(j))^2}{\sum_{i \in V} d(i) y(i)^2} \\
&\le \min_{y \in \{-1,0,1\}^n} \frac{\sum_{(i,j)\in E} 2|y(i) + y(j)|}{\sum_{i \in V} d(i) y(i)^2} = 2\beta(G),
\end{aligned}$$

by noticing that $(y(i) + y(j))^2 \le 2|y(i) + y(j)|$ for $y(i), y(j) \in \{-1, 0, +1\}$. $\square$

We'll prove the other direction in the next lecture.