CS 798 - Algorithmic Spectral Graph Theory, Fall 2015, Waterloo

Lecture 4 : Last eigenvalue

Last time we saw the connection between the second eigenvalue and graph expansion.

Today we will see how to use the same proof technique to prove a connection between the last eigenvalue and the bipartiteness of a subgraph.

We will then see that this connection can be used to design a non-trivial approximation algorithm for max cut.

Last eigenvalue and bipartiteness

Recall that $\mathcal{L} = I - \mathcal{A} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ is the normalized Laplacian matrix.

Let $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2$ be its eigenvalues (recall that we showed $\lambda_n \leq 2$ last time).

It can be proved that $\lambda_n = 2$ if and only if the graph has a component that is a bipartite graph.

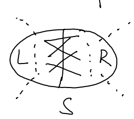
Trevisan [Tre09] proved that λ_n is "close" to 2 iff the graph has a subgraph "close" to a bipartite component, in the same style as in the proof of Cheeger's inequality.

We will first get some intuition and see what is the right generalization to prove.

By the characterization using Rayleigh quotient, we have $\lambda_n = \max_x \frac{x^T \mathcal{L} x}{x^T x} = \max_x \frac{x^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} x}{x^T x}$.

Doing a change of variable by writing $y = D^{-\frac{1}{2}} x$, we have $\lambda_n = \max_y \frac{y^T L y}{y^T D y} = \max_y \frac{\sum_{ij \in E} (y_i - y_j)^2}{\sum_{i \in V} d_i y_i^2}$.

Suppose G has a component S that is bipartite, i.e. there is a bipartition of S into (L, R) such that all the edges in S are between L and R.



Consider the vector $y \in \{-1, 0, 1\}^n$ where $y_i = \begin{cases} +1 & \text{if } i \in R \\ -1 & \text{if } i \in L \\ 0 & \text{otherwise} \end{cases}$.

For $X, Y \subseteq V$, let $E(X, Y)$ denotes the set of edges with one endpoint in X and another endpoint in Y.

Then $\frac{\sum_{ij \in E} (y_i - y_j)^2}{\sum_{i \in V} d_i y_i^2} = \frac{4 \cdot |E(L,R)|}{\sum_{i \in S} d_i} = \frac{2 \operatorname{vol}(S)}{\operatorname{vol}(S)} = 2$, recalling that $\operatorname{vol}(S)$ is defined as $\sum_{i \in S} d_i$,

and the second last inequality holds because S is a bipartite component and so $|E(L,R)| = \frac{\sum_{i \in S} d_i}{2} = \frac{\operatorname{vol}(S)}{2}$.

This shows that if the graph has a bipartite component, then $\lambda_n = 2$.

We won't prove the other direction here, as we will prove a more general statement that would imply that $\lambda_n = 2$ implies the graph has a bipartite component.

Combinatorial definition

We need a combinatorial definition to measure how a subgraph is close to a bipartite component.
Like in graph expansion, there are multiple natural definitions to define bipartiteness.
We find the right definition by looking at the spectral formulation, by finding an analogy with Cheeger's inequality.

The second eigenvalue is $\lambda_2 = \min_{x \perp v_1} \frac{x^T \mathcal{L} x}{x^T x} = \min_{x \perp v_1} \frac{x^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} x}{x^T x} = \min_{D^{\frac{1}{2}} y \perp v_1} \frac{y^T L y}{y^T D y} = \min_{D^{\frac{1}{2}} y \perp v_1} \frac{\sum_{ij \in E} (y_i - y_j)^2}{\sum_{i \in V} d_i y_i^2}$.

The corresponding combinatorial parameter is defined when we restrict the domain to be $\{0,1\}^n$:

$$\phi(G) = \min_{y \in \{0,1\}^n} \frac{\sum_{ij \in E} (y_i - y_j)^2}{\sum_{i \in V} d_i y_i^2} = \min_{y \in \{0,1\}} \frac{\sum_{ij \in E} |y_i - y_j|}{\sum_{i \in V} d_i |y_i|} = \min_{S \subseteq V} \frac{|\delta(S)|}{\mathrm{vol}(S)}.$$

This is why λ_2 is related to the conductance (but not expansion or sparsest cut).

Bipartiteness ratio

To define a perfect analogy with Cheeger's inequality, we want a quantity which is close to zero (instead of close to two) when there is a subgraph close to bipartite.

To do this, we flip the spectrum by using the matrix $2I - \mathcal{L} = 2I - (I - \mathcal{A}) = I + \mathcal{A}$.

Notice that if $\lambda_1 \leq \ldots \leq \lambda_n$ is the spectrum of $\mathcal{L}$, then $2 - \lambda_1 \geq 2 - \lambda_2 \geq \ldots \geq 2 - \lambda_n$ is the spectrum of $I + \mathcal{A}$.

So, if the largest eigenvalue of λ_n is close to 2, then the smallest eigenvalue of $I + \mathcal{A}$ is close to zero.

Let $2 \geq \beta_1 \geq \beta_2 \geq \ldots \geq \beta_n \geq 0$ be the eigenvalues of $I + \mathcal{A}$.

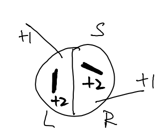
Then $\beta_n = \min_x \frac{x^T (I + \mathcal{A}) x}{x^T x} = \min_x \frac{x^T D^{-\frac{1}{2}} (D + A) D^{-\frac{1}{2}} x}{x^T x} = \min_y \frac{y^T (D + A) y}{y^T D y} = \min_y \frac{\sum_{ij \in E} (y_i + y_j)^2}{\sum_{i \in V} d_i y_i^2}$.

For bipartiteness, we are only interested in those $y \in \{-1, 0, +1\}^n$.

So, let's define $\beta(G) = \min_{y \in \{-1, 0, +1\}^n} \frac{\sum_{ij \in E} |y_i + y_j|}{\sum_{i \in V} d_i |y_i|}$.

By setting L be the vertices with value -1, R be the set of vertices with value 1, and $S = L \cup R$,

we have $\beta(G) = \min_{\substack{S \subseteq V \\ (L,R) \text{ bipartition of } S}} \frac{2|E(L,L)| + 2|E(R,R)| + |\delta(S)|}{\mathrm{vol}(S)}$.



In words, we want to minimize the number of edges not in $E(L,R)$, by putting a penalty of 2 on each edge in $E(L,L) \cup E(R,R)$ and a penalty of 1 for each edge in $\delta(S)$.

A Cheeger's type inequality

Trevisan proved the following analog of Cheeger's inequality for the last eigenvalue.

Theorem [Tre09] $\quad \frac{1}{2}\beta_n \leq \beta(G) \leq \sqrt{2\beta_n}$.

Proofs

Once the theorem is formulated, the proofs are very similar to those for Cheeger's inequality.

The first inequality is the easy direction, and the second inequality is the hard direction.

As in Cheeger's inequality, a good way is to think of the easy direction as showing the spectral formulation is a relaxation of the combinatorial problem, and the hard direction as showing a rounding algorithm that turns fractional solution into an integral solution.

Easy direction

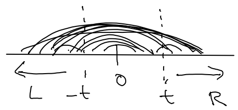
This is very easy given our previous discussion.

$$\beta_n = \min_y \frac{\sum_{ij\in E}(y_i+y_j)^2}{\sum_{i\in V} d_i y_i^2} \leq \min_{y\in\{-1,0,+1\}^n} \frac{\sum_{ij\in E}(y_i+y_j)^2}{\sum_{i\in V} d_i y_i^2} \leq \min_{y\in\{-1,0,+1\}^n} \frac{\sum_{ij\in E} 2|y_i+y_j|}{\sum_{i\in V} d_i|y_i|} = 2\beta(G),$$

where the last inequality is because $(y_i+y_j)^2 \leq 2|y_i+y_j|$ for $y_i, y_j \in \{-1,0,+1\}$. $\square$

Hard direction

Intuition: If β_n is small, then for most edges, both endpoints are of opposite sign with similar absolute values,



and we should find a threshold t and put everything larger than t on one side, everything smaller than $-t$ on the other side, and everything in between to be zero.

Analysis: Assume $\max_i y_i^2 = 1$, by scaling the vector.

Pick a random number t uniformly randomly in $[0,1]$.

Set $x_i = \begin{cases} +1 & \text{if } y_i^2 \geq t \text{ and } y_i \text{ positive} \\ -1 & \text{if } y_i^2 \geq t \text{ and } y_i \text{ negative} \\ 0 & \text{otherwise.} \end{cases}$

As in the proof of Cheeger's inequality, we will prove that $\dfrac{E[\sum_{ij\in E}|x_i+x_j|]}{E[\sum_{i\in V} d_i|x_i|]} \leq \sqrt{2\beta_n}$, and

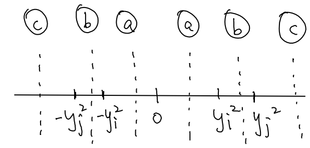
this will imply that there exists t with the corresponding x satisfies $\dfrac{\sum_{ij\in E}|x_i+x_j|}{\sum_{i\in V} d_i|x_i|} \leq \sqrt{2\beta_n}$

In the proof of cheeger's inequality, we show that $E[|x_i-x_j|] = \mathbb{1}(\text{edge } ij \text{ is cut}) = |y_i-y_j|\,|y_i+y_j|$, where the first term is used to relate to the numerator of the optimal solution (using Cauchy-Schwarz) and the second term is used to relate to the denominator.

We prove a similar claim in this setting.

<u>Claim</u> $E[\,|x_i+x_j|\,] \le |y_i+y_j|\,(|y_i|+|y_j|)$ for edge ij.

~~proof~~ There are two cases depending on the signs of y_i and y_j.



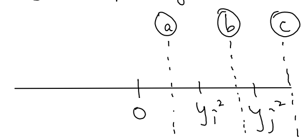
Case 1: y_i and y_j are of the opposite sign. Assume $|y_i| \le |y_j|$.

If $t \le y_i^2$ (see case ⓐ in the figure), then $x_i = 1$ and $x_j = -1$ and so $|x_i+x_j| = 0$.

If $y_i^2 \le t \le y_j^2$, then $x_i = 0$ and $x_j = -1$ and so $|x_i+x_j| = 1$. This happens with probability $y_j^2 - y_i^2$.

Finally, if $y_j^2 \le t$, then $x_i = x_j = 0$ and so $|x_i+x_j| = 0$.

Therefore, $E[\,|x_i+x_j|\,] = y_j^2 - y_i^2 = (y_j - y_i)(y_j+y_i) = (|y_i|+|y_j|)(|y_i+y_j|)$, proving the claim.



Case 2: y_i and y_j are of the same sign. Assume $|y_i| \le |y_j|$

ⓐ If $t \le y_i^2$, then $x_i = x_j = 1$ and so $|x_i+x_j| = 2$ with probability y_i^2.

ⓑ If $y_i^2 \le t \le y_j^2$, then $x_i = 0$ and $x_j = 1$ and so $|x_i+x_j| = 1$ with probability $y_j^2 - y_i^2$.

ⓒ If $y_j^2 \le t$, then $x_i = x_j = 0$ and so $|x_i+x_j| = 0$ with probability $1 - y_j^2$.

Therefore, $E[\,|x_i+x_j|\,] = 2y_i^2 + (y_j^2 - y_i^2) = y_i^2 + y_j^2 \le (|y_i|+|y_j|)(|y_i+y_j|)$. □

With the claim, we finish the proof almost identically to the proof of Cheeger's inequality.

$$
\begin{aligned}
E\Big[\sum_{ij\in E} |x_i+x_j|\Big] &\le \sum_{ij\in E} |y_i+y_j|\,(|y_i|+|y_j|) \\
&\le \sqrt{\sum_{ij\in E} (y_i+y_j)^2}\,\sqrt{\sum_{ij\in E} (|y_i|+|y_j|)^2} && \text{(by Cauchy-Schwarz)} \\
&\le \sqrt{\beta_n \sum_{i\in V} d_i y_i^2}\,\sqrt{\sum_{ij\in E} 2(y_i^2+y_j^2)} && \text{(by definition of } \beta_n \text{ and } (|a|+|b|)^2 \le 2a^2+2b^2) \\
&= \sqrt{\beta_n \sum_{i\in V} d_i y_i^2}\,\sqrt{\sum_{i\in V} 2 d_i y_i^2} \\
&= \sqrt{2\beta_n}\, \sum_{i\in V} d_i y_i^2 .
\end{aligned}
$$

Notice that $E\Big[\sum_{i\in V} d_i |x_i|\Big] = \sum_{i\in V} d_i y_i^2$, and thus we have completed the proof of the theorem.

<u>Maximum cut</u>

Given an undirected graph $G = (V, E)$, the maximum cut problem is to find $S \subseteq V$ such that $|\delta(S)|$ is maximized.

When $G = (X, Y; E)$ is bipartite, then we can cut all the edges as $|\delta(X)| = |\delta(Y)| = |E|$.

It is not difficult to see that we can cut at least 50% of the edges in any graph, and this gives a trivial 0.5-approximation algorithm for the problem.

Linear programming is a popular tool in designing approximation algorithms. but there is no known LP-based approximation algorithm with an approximation ratio better than 0.5.

In 1994 Goemans and Williamson gave a 0.878-approximation algorithm, by introducing semidefinite programming to the design of approximation algorithms.

Interestingly, assuming the "unique games conjecture", this approximation ratio is optimal unless P=NP.

For a long time the SDP approximation algorithm is the only method to do better than 1/2.

Trevisan in 2008 gave a spectral algorithm, proving an approximation ratio 0.531.

The analysis is improved by Sato to show that the approximation ratio is at least 0.614.

Spectral algorithm for maximum cut

Suppose the maximum cut $(S, V-S)$ cuts at least $1-\varepsilon$ fraction of edges.

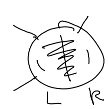
By setting $L=S$ and $R=V-S$, we see that $\beta(G) \leq \varepsilon$, and thus $\beta_n \leq 2\varepsilon$ by the easy direction.

So, if given a graph, we compute that $\beta_n \geq \theta$, then we know that the maximum cut cuts at most $1-\frac{\theta}{2}$ fraction of edges, and thus cutting 50% of edges would be a $\frac{1}{2-\theta}$-approximation.

This is the power of the spectral method, to compute a better upper bound on the optimal value.

What if β_n is very small, say $\beta_n \leq \theta$?

In this case, by the theorem, we know that $\beta(G) \leq \sqrt{2\theta}$, and this means that there is $S = L \cup R$ such that
$$\frac{|E(L,L)| + |E(R,R)| + |\delta(S)|}{vol(S)} \leq \sqrt{2\theta}.$$

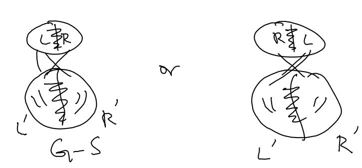


The idea is very natural.

We will commit on putting vertices in L on one side and vertices in R on the other side.

What about the edges in $\delta(S)$?

We will solve the problem on $G-S$ recursively using this algorithm, and then flip (L,R) so that we also cut at least 50% of the edges in $\delta(S)$, i.e.



We are in a win-win situation.

If β_n is small, then we find a good subset $S=(L,R)$, and we cut strictly more than

50% of those edges with at least one endpoint in S.

Clearly, if all β_n we computed are small, then we cut strictly more than 50% of the edges.

On the other hand, if β_n is big in some iteration, we know that there is no good max cut in the remaining graph. and so cutting half the edges in the remaining graph is strictly better than 50% of the optimal.

It should be intuitively clear that setting θ carefully would give us a better than 0.5-approximation.

Formal arguments

The following simple argument is by Nick Harvey.

Set a threshold θ to be determined later.

If $\beta_n \le \theta$, find $S = L \cup R$ with $\frac{2|E(L,L)|+2|E(R,R)|+|\delta(S)|}{vol(S)} \le \sqrt{2\theta}$ using the spectral algorithm.

Then we recurse on $G-S$ and find a partition of $V-S$ into (L', R'), and output the best of $(L \cup L', R \cup R')$ and $(L \cup R', L' \cup R)$.

If $\beta_n > \theta$, cut 50% by the greedy algorithm and return.

In this case, the approximation ratio is at least $\frac{1}{2-\theta}$.

In the first case, the algorithm will cut the edges in $|E(L,R)| + \frac{1}{2}|\delta(S)| + ALG(G-S)$,

where $ALG(G-S)$ is the number of edges that the algorithm cuts in $G-S$.

Note that the optimal solution can cut at most all the edges with one endpoint in S, plus the optimal max cut in $G-S$, which is $|E(L,L)| + |E(R,R)| + |E(L,R)| + |\delta(S)| + OPT(G-S)$

So, the approximation ratio is at least
$$\frac{|E(L,R)| + \frac{1}{2}|\delta(S)| + ALG(G-S)}{|E(L,L)| + |E(R,R)| + |E(L,R)| + |\delta(S)| + OPT(G-S)}$$
$$\ge \min\left\{ \frac{|E(L,R)| + \frac{1}{2}|\delta(S)|}{|E(L,L)| + |E(R,R)| + |E(L,R)| + |\delta(S)|}, \frac{ALG(G-S)}{OPT(G-S)} \right\}.$$

The second term is at least $\frac{1}{2-\theta}$.

The spectral theorem implies that
$$\sqrt{2\theta} \ge \frac{2|E(L,L)| + 2|E(R,R)| + |\delta(S)|}{vol(S)}$$
$$= \frac{2|E(L,L)| + 2|E(R,R)| + |\delta(S)|}{2|E(L,L)| + 2|E(R,R)| + 2|E(L,R)| + |\delta(S)|}$$
$$= 1 - \frac{|E(L,R)|}{|E(L,L)| + |E(R,R)| + |E(L,R)| + \frac{1}{2}|\delta(S)|}$$

$$\geq 1 - \frac{|E(L,R)| + \frac{1}{2}|\delta(S)|}{|E(L,L)| + |E(R,R)| + |E(L,R)| + |\delta(S)|}$$

So, the first ratio is at least $1-\sqrt{2\theta}$, and the approximation ratio is $\geq \min\left\{1-\sqrt{2\theta}, \frac{1}{2-\theta}\right\}$.

By balancing the two terms and setting $\theta = 0.1107$, we get the approximation ratio is at least 0.528.

Open questions: Give a tight analysis of this algorithm (improved analysis or tight example).

More generally, give a tight analysis of the "spectral approach".

References Max cut and the smallest eigenvalue, by Trevisan, 2008.