CMSE 890-001: Spectral Graph Theory and Related Topics, MSU, Spring 2021

Lecture 15: Introduction to Graph Partitioning and Clustering

March 11, 2021

Lecturer: Matthew Hirn

23 The isoperimetric ratio

We now return to standard spectral graph theory and shift into graph partitioning and clustering. As we will see, these results will be intimately related to the second eigenvalue of the graph Laplacian and the to-be-defined normalized graph Laplacian. In short, these results will quantify the intuition we have sometimes already espoused, which is that even if G is connected, λ_2 quantifies how well G is connected. In light of our discussions on eigenvector frequency, this means that it is the low frequency eigenvectors that are good for clustering G.

To begin, let $G = (V, E)$ be an unweighted graph and let $S \subset V$. Let $V - S$ denote all the vertices in the graph that are not in S, i.e.,

$$V - S := \{a \in V : a \notin S\}.$$

One way to measure how well connected S is to the rest of the graph G is to measure the number of edges going from S to $V - S$. These edges are called the *boundary* of S and are collected in the set ∂S:

$$\partial S := \{(a, b) \in E : a \in S \text{ and } b \in V - S\}.$$

Instead of counting the number of edges in ∂S, it is more useful to measure the ratio of edges in ∂S to the size of S. Indeed, for example if ∂S is small but S is also small, then such a set is relatively well connected to the rest of G. We define this ratio as the *isoperimetric ratio of S*:

$$\theta(S) := \frac{|\partial(S)|}{|S|}.$$

The isoperimetric ratio of G is the minimum of $\theta(S)$ over all $S \subset V$ such that $|S| \leq n/2$, i.e.,

$$\theta_G := \min_{\substack{S \subset V \\ |S| \leq n/2}} \theta(S).$$

Intuitively, if the graph G has a "bottleneck" or two well connected parts that are only weakly connected to each other, then θ_G will be small, e.g., the barbell graph from Homework 03. In the more extreme case, if G is disconnected, then $\theta_G = 0$. On the other hand, θ_G will be large if any division of V into S and $V - S$ has many edges between S and $V - S$, e.g., the complete graph. Graphs with large θ_G are well connected in the sense that there is no $S \subset V$ such that $G(S)$ can be removed from G by only cutting a relatively small number of edges. The following theorem shows that λ_2 gives a lower bound for θ_G.

Theorem 39. *Let $G = (V, E)$ and let $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be its graph Laplacian eigenvalues. Then*

$$\theta(S) \geq \lambda_2(1-s)\,, \quad s := \frac{|S|}{n}\,, \tag{51}$$

and as such,

$$\theta_G \geq \frac{\lambda_2}{2}\,. \tag{52}$$

Proof. Since θ_G minimizes over $S \subset V$ with $|S| \leq n/2$ we have that $s \leq 1/2$ and as such $1 - s \geq 1/2$. That proves (52) assuming we can prove (51).

To prove (51), recall from Theorem 13 that

$$\lambda_2 = \min_{\substack{\boldsymbol{x} \in \mathbb{R}^n \\ \langle \boldsymbol{x}, \mathbf{1} \rangle = 0}} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}\,.$$

Thus,

$$\forall\, \boldsymbol{x} \in \mathbb{R}^n \text{ such that } \langle \boldsymbol{x}, \mathbf{1} \rangle = 0\,, \quad \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \lambda_2\,.$$

This is of course the test vector technique we discussed in Section 12. In this case, we need to find a test vector $\boldsymbol{x}$ such that its Rayleigh quotient contains $\theta(S)$.

We would like to pick $\boldsymbol{x} = \mathbf{1}_S$, where

$$\mathbf{1}_S(a) := \begin{cases} 1 & a \in S \\ 0 & a \notin S \end{cases}\,.$$

Indeed, recall $E(S) := \{(a,b) \in E : a, b \in S\}$ and notice that

$$\begin{aligned}
\mathbf{1}_S^T \boldsymbol{L} \mathbf{1}_S &= \sum_{(a,b) \in E} (\mathbf{1}_S(a) - \mathbf{1}_S(b))^2 \\
&= \sum_{(a,b) \in E(S)} (\mathbf{1}_S(a) - \mathbf{1}_S(b))^2 + \sum_{(a,b) \in \partial(S)} (\mathbf{1}_S(a) - \mathbf{1}_S(b))^2 + \sum_{(a,b) \in E(V-S)} (\mathbf{1}_S(a) - \mathbf{1}_S(b))^2 \\
&= 0 + |\partial(S)| + 0 \\
&= |\partial(S)|\,.
\end{aligned}$$

However, $\langle \mathbf{1}_S, \mathbf{1} \rangle > 0$ and so we cannot use $\mathbf{1}_S$ as our test vector. So we use the next best thing, which is

$$\boldsymbol{x} = \mathbf{1}_S - s\mathbf{1}\,,$$

and so

$$\boldsymbol{x}(a) = \begin{cases} 1 - s & a \in S \\ -s & a \notin S \end{cases}\,.$$

We have

$$\begin{aligned}
\langle \boldsymbol{x}, \mathbf{1} \rangle &= \sum_{a \in S} 1 - s + \sum_{a \in V-S} -s \\
&= |S|(1-s) - |V-S|s \\
&= |S|(1-s) - (n - |S|)s \\
&= |S| - s|S| - ns + s|S| \\
&= |S| - |S| \\
&= 0\,.
\end{aligned}$$

Thus $\boldsymbol{x}$ is a test vector. Additionally,

$$\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} = \sum_{(a,b)\in E} ((\mathbf{1}_S(a) - s) - (\mathbf{1}_S(b) - s))^2 = \sum_{(a,b)\in E} (\mathbf{1}_S(a) - \mathbf{1}_S(b))^2 = |\partial(S)|\,.$$

To complete the proof we need to compute $\boldsymbol{x}^T\boldsymbol{x}$:

$$\begin{aligned}
\boldsymbol{x}^T \boldsymbol{x} &= (\mathbf{1}_S - s\mathbf{1})^T(\mathbf{1}_S - s\mathbf{1}) \\
&= \mathbf{1}_S^T \mathbf{1}_S - s\mathbf{1}_S^T \mathbf{1} - s\mathbf{1}^T \mathbf{1}_S + s^2 \mathbf{1}^T \mathbf{1} \\
&= |S| - s|S| - s|S| + s^2 n \\
&= |S| - 2s|S| + s|S| \\
&= |S|(1-s)\,.
\end{aligned}$$

Therefore,

$$\frac{|\partial(S)|}{|S|(1-s)} = \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \lambda_2 \quad \Longrightarrow \quad \theta(S) \geq \lambda_2(1-s)\,.$$

□

Remark 21. Theorem 39 says that graphs G with large λ_2 are well connected.

24 The normalized graph Laplacian

Remark 22. Whenever discussing the normalized graph Laplacian, we will assume G has no isolated vertices. That is, every vertex in G is connected to at least one other vertex and hence $\deg(a) > 0$ for all $a \in V$.

To take our analysis of graph partitioning and clustering further, it will be useful to introduce a new matrix called the *normalized graph Laplacian*. We will denote it by $\boldsymbol{N}$. For a diagonal matrix $\boldsymbol{A}$ with positive entries on its diagonal, we define $\boldsymbol{A}^\alpha$ for any $\alpha \in \mathbb{R}$ as

$$\boldsymbol{A}^\alpha(a,b) := \begin{cases} \boldsymbol{A}(a,a)^\alpha & b = a \\ 0 & b \neq a \end{cases}\,.$$

Now define the normalized graph Laplacian as:

$$\boldsymbol{N} := \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{M}\boldsymbol{D}^{-1/2}\,.$$

Note this definition works for unweighted and weighted graphs. Recall that if $G = (V, E, w)$ is weighted, then the degree of $a \in V$ is defined as

$$\deg(a) := \sum_{b \in N(a)} w(a, b)\,.$$

Recall also that we defined $\boldsymbol{d} : V \to \mathbb{R}$ as the degree vector G, i.e.,

$$\boldsymbol{d}(a) := \deg(a)\,.$$

We also have:

$$d_{\min} := \min_{a \in V} \deg(a) \quad \text{and} \quad d_{\max} := \max_{a \in V} \deg(a)\,.$$

The normalized graph Laplacian provides a degree independent representation of G. Indeed, notice that G and $c \cdot G = (V, E, cw)$ have the same normalized graph Laplacian. We will denote the eigenvalues of $\boldsymbol{N}$ by $\nu_1 \leq \nu_2 \leq \cdots \leq \nu_n$, and, when needed, (orthonormal) eigenvectors of $\boldsymbol{N}$ by $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n$:

$$\boldsymbol{N}\boldsymbol{\phi}_k = \nu_k \boldsymbol{\phi}_k\,, \quad 1 \leq k \leq n\,.$$

Before utilizing $\boldsymbol{N}$ for graph partitioning and clustering, let us first discuss some basic properties of the normalized graph Laplacian. This first theorem will relate the eigenvalues of $\boldsymbol{L}$ to the eigenvalues of $\boldsymbol{N}$.

Theorem 40. *Let $G = (V, E, w)$, let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $\boldsymbol{L}_G$, and let $\nu_1 \leq \nu_2 \leq \cdots \leq \nu_n$ be the eigenvalues of $\boldsymbol{N}_G$. Then*

$$\frac{\lambda_k}{d_{\max}} \leq \nu_k \leq \frac{\lambda_k}{d_{\min}}\,, \quad 1 \leq k \leq n\,.$$

Proof. By the Courant-Fischer theorem (Theorem 3) we know that

$$\nu_k = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S) = k}} \max_{\substack{\boldsymbol{x} \in S \\ \boldsymbol{x} \neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{N} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S) = k}} \max_{\substack{\boldsymbol{x} \in S \\ \boldsymbol{x} \neq \boldsymbol{0}}} \frac{(\boldsymbol{D}^{-1/2}\boldsymbol{x})^T \boldsymbol{L} (\boldsymbol{D}^{-1/2}\boldsymbol{x})}{\boldsymbol{x}^T \boldsymbol{x}}\,. \tag{53}$$

Now make the change of variables $\boldsymbol{y} = \boldsymbol{D}^{-1/2}\boldsymbol{x}$. Since $\boldsymbol{x} = \boldsymbol{D}^{1/2}\boldsymbol{y}$ implies that $\boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}$ and since $\boldsymbol{D}^{-1/2}$ has full rank, we can rewrite (53) as:

$$\nu_k = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S) = k}} \max_{\substack{\boldsymbol{y} \in S \\ \boldsymbol{y} \neq \boldsymbol{0}}} \frac{\boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}}$$

Now observe that

$$\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y} = \sum_{a\in V}\deg(a)|\boldsymbol{y}(a)|^2 \leq d_{\max}\sum_{a\in V}|\boldsymbol{y}(a)|^2 = d_{\max}\boldsymbol{y}^T\boldsymbol{y}\,.$$

Therefore:

$$\nu_k = \min_{\substack{S\subseteq\mathbb{R}^n\\ \dim(S)=k}}\max_{\substack{\boldsymbol{y}\in S\\ \boldsymbol{y}\neq\boldsymbol{0}}}\frac{\boldsymbol{y}^T\boldsymbol{L}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}} \geq \min_{\substack{S\subseteq\mathbb{R}^n\\ \dim(S)=k}}\max_{\substack{\boldsymbol{y}\in S\\ \boldsymbol{y}\neq\boldsymbol{0}}}\frac{\boldsymbol{y}^T\boldsymbol{L}\boldsymbol{y}}{d_{\max}\boldsymbol{y}^T\boldsymbol{y}} = \frac{1}{d_{\max}}\min_{\substack{S\subseteq\mathbb{R}^n\\ \dim(S)=k}}\max_{\substack{\boldsymbol{y}\in S\\ \boldsymbol{y}\neq\boldsymbol{0}}}\frac{\boldsymbol{y}^T\boldsymbol{L}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{y}} = \frac{\lambda_k}{d_{\max}}\,,$$

where we used the Courant-Fischer Theorem one more time in the last equality. The upper bound is proved in a similar fashion. □

Remark 23. In the proof of Theorem 40 we showed that

$$\nu_k = \min_{\substack{S\subseteq\mathbb{R}^n\\ \dim(S)=k}}\max_{\substack{\boldsymbol{y}\in S\\ \boldsymbol{y}\neq\boldsymbol{0}}}\frac{\boldsymbol{y}^T\boldsymbol{L}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}} = \max_{\substack{T\subseteq\mathbb{R}^n\\ \dim(T)=n-k+1}}\min_{\substack{\boldsymbol{y}\in T\\ \boldsymbol{y}\neq\boldsymbol{0}}}\frac{\boldsymbol{y}^T\boldsymbol{L}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}}\,. \tag{54}$$

This formulation of ν_k is an important alternative formula for computing ν_k, so let us collect it separately here.

Theorem 40 almost immediately gives us the following corollary.

Corollary 41. *Let $G = (V, E, w)$. Then the normalized graph Laplacian, $\boldsymbol{N}_G$, is positive semidefinite. Furthermore, $\nu_1 = 0$ and one can take $\boldsymbol{\phi}_1(a) = \boldsymbol{d}^{1/2}(a) := \deg(a)^{1/2}$. Finally, $\nu_2 > 0$ if and only if G is connected.*

Proof. Since $\boldsymbol{L}$ is positive semidefinite and $\lambda_1 = 0$, it is immediate from Theorem 40 that $\boldsymbol{N}$ is positive semidefinite and $\nu_1 = 0$. Additionally, $\nu_2 > 0 \iff G$ is connected, immediately follows by combining Theorem 40 with Theorem 5 (recall the latter says that $\lambda_2 > 0 \iff G$ is connected). Now let us compute $\boldsymbol{N}\boldsymbol{d}^{1/2}$:

$$\boldsymbol{N}\boldsymbol{d}^{1/2} = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}\boldsymbol{d}^{1/2} = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{1} = \boldsymbol{D}^{-1/2}\boldsymbol{0} = \boldsymbol{0}\,.$$

□

Thus the normalized graph Laplacian is in many ways similar to the graph Laplacian. Corollary 41 indicates one difference, which is that $\boldsymbol{\psi}_1 = \boldsymbol{1}$ while $\boldsymbol{\phi}_1 = \boldsymbol{d}^{1/2}$ (although, notice if G is d-regular then $\boldsymbol{\phi}_1$ is constant). The next theorem gives another difference between $\boldsymbol{L}$ and $\boldsymbol{N}$.

Theorem 42. *Let $G = (V, E, w)$. The mean eigenvalue of $\boldsymbol{L}_G$ is the average degree of G, whereas the mean eigenvalue of $\boldsymbol{N}_G$ is 1.*

Proof. Recall from Homework 02, Exercise 02, that the sum of the eigenvalues of a matrix is equal to its trace. Therefore:

$$\frac{1}{n}\sum_{k=1}^{n}\lambda_k = \frac{1}{n}\mathrm{Tr}(\boldsymbol{L}) = \frac{1}{n}\mathrm{Tr}(\boldsymbol{D}-\boldsymbol{M}) = \frac{1}{n}\mathrm{Tr}(\boldsymbol{D}) = \frac{1}{n}\sum_{a\in V}\deg(a)\,,$$

On the other hand,

$$\frac{1}{n}\sum_{k=1}^{n}\nu_k = \frac{1}{n}\mathrm{Tr}(\boldsymbol{N}) = \frac{1}{n}\mathrm{Tr}(\boldsymbol{I}-\boldsymbol{D}^{-1/2}\boldsymbol{M}\boldsymbol{D}^{-1/2}) = \frac{1}{n}\mathrm{Tr}(\boldsymbol{I}) = \frac{1}{n}\cdot n = 1\,.$$

□

Theorem 42 gives another indication that $\boldsymbol{N}$ is a degree independent representation of G, whereas $\boldsymbol{L}$ is not. Here is another fact along similar lines.

Theorem 43. *Let $G=(V,E,w)$. All eigenvalues of $\boldsymbol{N}_G$ lie in the interval $[0,2]$, i.e.,*

$$0=\nu_1\leq\nu_2\leq\cdots\leq\nu_n\leq 2\,.$$

I leave the proof of Theorem 43 to you in your homework. As our last general fact about $\boldsymbol{N}$, we give an analogue of Theorem 13 (recall this theorem gave a formula for λ_2).

Theorem 44. *Let $G=(V,E,w)$ and let $0=\nu_1\leq\nu_2\leq\cdots\leq\nu_n\leq 2$ be the eigenvalues of $\boldsymbol{N}_G$. Then*

$$\nu_2 = \min_{\substack{\boldsymbol{x}\in\mathbb{R}^n\\ \langle\boldsymbol{x},\boldsymbol{d}^{1/2}\rangle=0}}\frac{\boldsymbol{x}^T\boldsymbol{N}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \min_{\substack{\boldsymbol{y}\in\mathbb{R}^n\\ \langle\boldsymbol{y},\boldsymbol{d}\rangle=0}}\frac{\boldsymbol{y}^T\boldsymbol{L}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}}\,.$$

Proof. The proof of the first equality is nearly identical to the proof of Theorem 13, remembering that $\boldsymbol{\phi}_1=\boldsymbol{d}^{1/2}/\|\boldsymbol{d}^{1/2}\|$ and that we can take the eigenvectors $\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\ldots,\boldsymbol{\phi}_n$ of $\boldsymbol{N}$ to be an ONB for $\mathbb{R}^n$.

For the proof of the second equality, make the change of variables $\boldsymbol{y}=\boldsymbol{D}^{-1/2}\boldsymbol{x}$; we know from the proof of Theorem 40 that we get the argument of the minimum. Note also that with this change of variables:

$$0=\langle\boldsymbol{x},\boldsymbol{d}^{1/2}\rangle = \langle\boldsymbol{D}^{1/2}\boldsymbol{y},\boldsymbol{d}^{1/2}\rangle = \langle\boldsymbol{y},\boldsymbol{D}^{1/2}\boldsymbol{d}^{1/2}\rangle = \langle\boldsymbol{y},\boldsymbol{d}\rangle\,.$$

□

References

[1] Daniel A. Spielman. Spectral and algebraic graph theory. Book draft, available at: http://cs-www.cs.yale.edu/homes/spielman/sagt/, 2019.

[2] Michael Perlmutter, Feng Gao, Guy Wolf, and Matthew Hirn. Geometric scattering networks on compact Riemannian manifolds. In *Proceedings of The First Mathematical and Scientific Machine Learning Conference, Proceedings of Machine Learning Research*, volume 107, pages 570–604, 2020.

[3] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

[4] Stéphane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way.* Academic Press, 3rd edition, 2008.