Lecture 18

Taken from [10, Lecture 5] with small modifications.

We just showed that the k^{th} eigenvector of a weighted path graph crosses the origin at most $k-1$ times. For example, Figure 24 shows the second, third, and fourth eigenvectors of the path graph on 13 vertices. At the beginning of today we state another result of Fiedler which says that for every weighted graph G the graph induced on the vertices that are non-negative in the k^{th} eigenvector have at most $k-1$ connected components. In particular, it says that the non-negative vertices of φ_2 are connected. This result generalizes our previous result on path graphs.

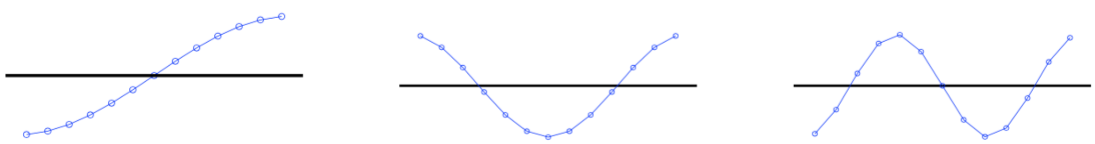

Figure 24: First three non-constant eigenvectors of the path graph on 13 vertices, with a line drawn at the origin.

8.5 Fiedler's Nodal Domain Theorem

Given a graph $G = (V, E, w)$ and a subset of vertices $U \subseteq V$, the *graph induced by G on U* is the graph $G(U)$ with vertex set U and edge set:

$$\{(i, j) \in E : i, j \in U\}.$$

Theorem 8 (Fiedler, 1975). *Let $G = (V, E, w)$ be a weighted connected graph, and let L_G be its graph Laplacian matrix. Let $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of L_G and $\varphi_1, \ldots, \varphi_n$ be the corresponding eigenvectors. For any $k \geq 2$, let*

$$W_k = \{i \in V : \varphi_k[i] \geq 0\}.$$

Then the graph induced by G on W_k, $G(W_k)$, has at most $k-1$ connected components.

A simple example which also illustrates that the theorem is sharp is given by the star graph. In Figure 25 is the star graph S_5 and its second eigenvector φ_2 with eigenvalue $\lambda_2 = 1$

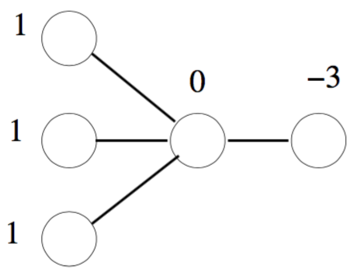

Figure 25: Second eigenvector of L_{S_5} with $\lambda_2 = 1$.

8.6 The second Laplacian eigenvalue

Recall that $\lambda_2 = 0$ if and only if a graph is disconnected. Fiedler observed that as λ_2 becomes further from 0, the graph becomes better connected. This statement of course needs to be clarified, and we shall now go about doing so.

Define the boundary of a set of vertices $S \subset V$ as:

$$\partial(S) = \{(i,j) \in E : i \in S, \;\; j \notin S\}.$$

We then have:

Theorem 9. *Let $G = (V, E)$ be a graph and L_G its graph Laplacian. Let $S \subset V$ and $\sigma = |S|/|V|$. Then:*

$$|\partial(S)| \geq \lambda_2 |S|(1 - \sigma).$$

Proof. Adapting the argument from Lecture 4 about computing PCA, note that for any symmetric matrix M with minimum eigenvalue λ_1,

$$\lambda_1 = \min_{v \neq 0} \frac{v^T M v}{v^T v}.$$

If φ_1 is the eigenvector corresponding to λ_1, then we can compute λ_2 via:

$$\lambda_2 = \min_{v : v^T \varphi_1 = 0} \frac{v^T M v}{v^T v}.$$

Now when $M = L_G$ is a graph Laplacian, we know that $\lambda_1 = 0$ and $\varphi_1 = \mathbf{1}$. So for λ_2 we have:

$$\lambda_2 = \min_{v : v^T \mathbf{1} = 0} \frac{v^T L_G v}{v^T v}$$

and thus:

$$v^T L_G v \geq \lambda_2 v^T v. \tag{54}$$

To utilize (54) we need to construct a vector related to S that is also orthogonal to $\mathbf{1}$. A natural first guess is the vector χ_S, defined as:

$$\chi_S[i] = \begin{cases} 1 & i \in S \\ 0 & i \notin S \end{cases}$$

But χ_S is not orthogonal to $\mathbf{1}$. To fix this, consider:

$$v_S = \chi_S - \sigma \mathbf{1}.$$

Now we have $v_S^T \mathbf{1} = 0$, and

$$\begin{aligned} v_S^T L_G v_S &= (\chi_S - \sigma\mathbf{1})^T L_G (\chi_S - \sigma\mathbf{1}) \\ &= \chi_S^T L_G \chi_S - \sigma \mathbf{1}^T L_G \chi_S - \sigma \chi_S L_G \mathbf{1} - \sigma^2 \mathbf{1}^T L_G \mathbf{1} \\ &= \chi_S^T L_G \chi_S - \sigma (L_G \mathbf{1})^T \chi_S - \sigma \chi_S L_G \mathbf{1} - \sigma^2 \mathbf{1}^T L_G \mathbf{1} \\ &= \chi_S^T L_G \chi_S \\ &= \sum_{(i,j) \in E} (\chi_S[i] - \chi_S[j])^2 \\ &= |\partial(S)|. \end{aligned}$$

To finish the proof, we compute:

$$\begin{aligned} v_S^T v_S &= (\chi_S - \sigma\mathbf{1})^T (\chi_S - \sigma\mathbf{1}) \\ &= |S| - 2\sigma|S| + \sigma^2|V| \\ &= |S| - 2\sigma|S| + \sigma^2|S| + \sigma^2|V| - \sigma^2|S| \\ &= |S| - 2\sigma|S| + \sigma^2|S| + \sigma\frac{|S|}{|V|}|V| - \sigma^2|S| \\ &= |S| - 2\sigma|S| + \sigma^2|S| + \sigma|S| - \sigma^2|S| \\ &= |S| - \sigma|S| \\ &= (1-\sigma)|S|. \end{aligned}$$

□

This theorem says that if λ_2 is big, then G is very well connected: the boundary of every small set of vertices is at least λ_2 times something just slightly smaller than the number of vertices in the set.

8.7 Cheeger's Inequality

Taken from [10, Lecture 7] with small modifications.

We just proved for any $S \subset V$ of a graph $G = (V, E)$ that:

$$|\partial(S)| \geq \lambda_2 |S|(1-\sigma) \Longrightarrow \lambda_2 \leq \frac{|\partial(S)|}{|S|(1-\sigma)}.$$

Cheegers inequality will provide a converse to this theorem.

8.7.1 Conductance

Different versions of Cheegers inequality are related to different measures of the quality of a cut. In Theorem 9, we were concerned with the number of edges cut divided by the number of vertices removed. The sharpest versions of Cheegers inequality hold when edges are treated as the most important objects, rather than vertices. That is, we weight vertices by their degrees. In this case, we will be interested in two measures of the quality of a cut, its *conductance* and its *sparsity*. These are closely related, and their names are often interchanged.

Let $G = (V, E)$. Define the degree of a set of vertices $S \subset V$ as the sum of the degrees $d(i)$ for each vertex $i \in S$:

$$d(S) = \sum_{i \in S} d(i).$$

Define the *conductance* of a set of vertices S by:

$$\phi(S) = d(V) \frac{|\partial(S)|}{d(S) d(\overline{S})}, \tag{55}$$

where $\overline{S}$ is the complement of the vertices S in V, that is:

$$\overline{S} = V \setminus S.$$

The constant $d(V)$ out front of (55) just helps us normalize this quantity. Note that when all vertices have the same degree, so $d(i) = d$ for all $i \in V$, the conductance differs from the quantity measured in Theorem 9 precisely by a factor of d; indeed, in this case:

$$\phi(S) = d|V| \frac{|\partial(S)|}{d|S| d|\overline{S}|} = \frac{|V||\partial(S)|}{d|S|(|V| - |S|)} = \frac{1}{d} \frac{|\partial(S)|}{|S|(1-\sigma)}.$$

We define the *sparsity* of a set S to be:

$$\mathrm{sp}(S) = \frac{|\partial(S)|}{\min(d(S), d(\overline{S}))}.$$

As one of $d(S)$ or $d(\overline{S})$ is always at least half of $d(V)$, we have:

$$\phi(S) \geq \mathrm{sp}(S) \geq \frac{1}{2}\phi(S),$$

so these two quantities will never differ by more than a factor of 2.

The conductance of a graph is defined to be the minimum conductance of all possible cuts, and similarly for the sparsity, which write as (with a slight abuse of notation):

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

8.7.2 The normalized graph Laplacian

The conductance of a graph is best approximated through the eigenvalues of the normalized graph Laplacian, which we define now as N_G where:

$$N_G = D_G^{-1/2} L_G D_G^{-1/2} = I - D_G^{-1/2} A_G D_G^{-1/2}.$$

We will denote the eigenvalues of N_G by $\gamma_1, \ldots, \gamma_n$ to distinguish them from the eigenvalues of L_G. Note that is a simple exercise (based on our previous results) to show that for any graph G, $\gamma_1 = 0$ with eigenvector $\mathbf{d}^{1/2}$, where

$$\mathbf{d}^{1/2}[i] = \sqrt{d(i)},$$

and additionally G is connected if and only if $\gamma_2 > 0$.

The following theorem is an analogue of Theorem 9, but for the normalized graph Laplacian and using the conductance of the graph G.

Theorem 10. *Let $G = (V, E)$ be a graph with normalized graph Laplacian N_G, which as eigenvalues $0 = \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_n$. Then*

$$\phi(G) \geq \gamma_2.$$

Exercises

Exercise 27. From the exposition at the beginning of the proof of Theorem 9, we know that

$$\gamma_2 = \min_{v : v^T \mathbf{d}^{1/2} = 0} \frac{v^T N_G v}{v^T v}.$$

Prove that we can also compute γ_2 as:

$$\gamma_2 = \min_{u : u^T \mathbf{d} = 0} \frac{u^T L_G u}{u^T D_G u}.$$

Exercise 28. Using the previous exercise and the same proof technique of Theorem 9 (with appropriate modifications), prove Theorem 10. That is, prove for any $S \subset V$ that

$$\gamma_2 \leq \frac{d(V)|\partial(S)|}{d(S)d(\overline{S})}.$$

References

[1] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning. The MIT Press, 2002.

[2] Afonso S. Bandeira. Ten lectures and forty-two open problems in the mathematics of data science. MIT course *Topics in Mathematics of Data Science*, 2015.

[3] Jon Shlens. A tutorial on principal component analysis. arXiv:1404.1100, 2014.

[4] Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, Series 6*, 2(11):559–572, 1901.

[5] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72(114):507–536, 1967.

[6] J. Baik, G. Ben-Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

[7] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[8] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer-Verlag New York, 2nd edition, 2009.

[9] Athanasios Tsanas and Angeliki Xifara. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49:560–567, 2012.

[10] Daniel A. Spielman. Spectral graph theory. *Yale Course Notes*, Fall, 2009.