Lecture 19

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

8.8 Cheeger's Inequality (continued)

Recall that $N_G = D_G^{-1/2} L_G D_G^{-1/2}$ is the normalized graph Laplacian of a graph $G = (V, E)$ with eigenvalues $0 = \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_n$. In the previous lecture we proved something similar to the following:

$$\forall\, S \subset V, \quad \gamma_2 \leq \phi(S) = d(V) \frac{|\partial(S)|}{d(S) d(\overline{S})}.$$

Theorem 10 is the easy part. Cheeger's inequality provides the following lower bound:

$$\gamma_2 \geq \phi(G)^2/8.$$

Moreover, an examination of most proofs of Cheeger's inequalities reveal that a cut S for which

$$\gamma_2 \geq \phi(S)^2/8$$

may be found by an examination of φ_2, the second eigenvector of N_G. In fact, we can find such a set S of the form:

$$S = \left\{ d^{-1/2}(i) \varphi_2[i] \geq t \right\},$$

for some $t \geq 0$.

Why is this important? Well, if $\phi(G)$ is small, then there exists some $S \subset V$ such that $\phi(S)$ is small. If $\phi(S)$ is small, then ignoring the normalization $d(V)$, it says that the number of edges going out of S into $\overline{S}$, is small relative to the total volume $d(S)$ of S and $d(\overline{S})$ of $\overline{S}$. Thus, S or $\overline{S}$ may be a cluster or certainly partitioning the vertices V into S and $\overline{S}$ breaks the graph G into two components that are not well connected. Cheeger's inequality shows that γ_2 being small guarantees that $\phi(G)$ is small, and we can find a sparse cut S that yields a value below γ_2 (and hence close to $\phi(G)$) by considering the sign of the second eigenvector of N_G.

Now we state and prove Cheeger's inequality:

Theorem 11 (Cheeger's Inequality). *Let v be a vector orthogonal to $\mathbf{d}^{1/2}$. Then there exists a t for which the set of vertices*

$$S = \left\{ i \in V : d^{-1/2}(i) v[i] \geq t \right\},$$

satisfies

$$\frac{v^T N_G v}{v^T v} \geq \phi(S)^2/8.$$

The proof will use the following three inequalities, stated here without proof.

Lemma 4. *Let $A, B, C, D \geq 0$. Then*

$$\frac{A+B}{C+D} \geq \min\left(\frac{A}{C}, \frac{B}{D}\right).$$

Lemma 5. *For all $0 < \delta < 1$,*

$$(a-b)^2 \geq \delta a^2 - \frac{\delta}{1-\delta} b^2$$

Lemma 6. *Let $z_1, \ldots, z_k \in \mathbb{R}$ such that*

$$z_1 \geq z_2 \geq \cdots \geq z_k \geq 0.$$

If $a_1, \ldots, a_k \in \mathbb{R}$ and $b_1, \ldots, b_k \in \mathbb{R}$ satisfy

$$\sum_{i=1}^{k} a_i z_i \geq \sum_{i=1}^{k} b_i z_i,$$

then there exists a j for which

$$\sum_{i=1}^{j} a_i \geq \sum_{i=1}^{j} b_i.$$

Proof of Theorem 11. Since the proof is long, I drop the subscript G in N_G, L_G, D_G, etc. Now let v be any vector orthogonal to $\mathbf{d}^{1/2}$, and set

$$\nu = \frac{v^T N v}{v^T v}.$$

Also define:

$$u = D^{-1/2}v.$$

Note that this transformation ensures (check on your own)

$$u^T\mathbf{d} = 0.$$

Let c be a number such that

$$\sum_{i:u[i]>c} d(i) \leq d(V)/2$$
$$\sum_{i:u[i]<c} d(i) \leq d(V)/2.$$

That is, c is the $\mathbf{d}$-weighted median of u. Let $z = u - c\mathbf{1}$, so that

$$\sum_{i:z[i]>0} d(i) \leq d(V)/2 \tag{56}$$
$$\sum_{i:z[i]<0} d(i) \leq d(V)/2.$$

One may easily show that

$$z^T Dz \geq u^T Du$$

by considering $f(c) = z^T Dz = (u - c\mathbf{1})^T D(u - c\mathbf{1})$ and showing that $f'(c) = 0 \Rightarrow c = 0$ and $f''(c) > 0$. We thus have:

$$\nu = \frac{v^T Nv}{v^T v} = \frac{u^T Lu}{u^T Du} \geq \frac{z^T Lz}{z^T Dz}.$$

Define:

$$z_+[i] = \begin{cases} z[i] & \text{if } z[i] \geq 0 \\ 0 & \text{otherwise} \end{cases}$$
$$z_-[i] = \begin{cases} z[i] & \text{if } z[i] \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

We have

$$z^T Dz = z_+^T Dz_+ + z_-^T Dz_-.$$

On the other hand, we can show

$$z^T Lz \geq z_+^T Lz_+ + z_-^T Lz_-.$$

To see this, recall that

$$z^T L z = \sum_{(i,j)\in E} (z[i] - z[j])^2.$$

Now note that if $z[i]$ and $z[j]$ have the same sign then $z[i] - z[j] = z_+[i] - z_+[j]$ and $z_-[i] - z_-[j] = 0$, or vice versa, and thus

$$(z[i] - z[j])^2 = (z_+[i] - z_+[j])^2 + (z_-[i] - z_-[j])^2, \quad \text{if } \operatorname{sgn} z[i] = \operatorname{sgn} z[j].$$

If, however, $z[i]$ and $z[j]$ have the opposite sign, then

$$\begin{aligned}(z[i] - z[j])^2 &= z[i]^2 + z[j]^2 - 2z[i]z[j] \\ &\geq z[i]^2 + z[j]^2 \\ &= (z_+[i] - z_+[j])^2 + (z_-[i] - z_-[j])^2.\end{aligned}$$

Using Lemma 4, we then have:

$$\begin{aligned}\nu &\geq \frac{z^T L z}{z^T D z} \\ &\geq \frac{z_+^T L z_+ + z_-^T L z_-}{z_+^T D z_+ + z_-^T D z_-^T} \\ &\geq \min\left(\frac{z_+^T L z_+}{z_+^T D z_+}, \frac{z_-^T L z_-}{z_-^T D z_-}\right) \\ &= \frac{z_s^T L z_s}{z_s^T D z_s},\end{aligned} \tag{57}$$

for one of $s \in \{+, -\}$. Let's assume without loss of generality that $s = +$; we may also assume without loss of generality that

$$z[1] \geq z[2] \geq \cdots z[k] \geq 0$$

are exactly the positive elements of z. In order to simplify the notation, we replace the graph $G = (V, E)$ with $V = \{1, \ldots, n\}$, with the graph $G^+ = (V^+, E^+)$ in which $V^+ = \{1, \ldots, k+1\}$ and

$$E^+ = \{(i,j) \in E : i, j \leq k\} \cup \{(i, k+1) : (i,j) \in E,\ i \leq k < j\}.$$

The vector $z : V \to \mathbb{R}$ is redefined as well, so that $z : V^+ \to \mathbb{R}$ with $z[i]$, $i = 1, \ldots, k$ as before, and $z[k+1] = 0$. This allows us to write:

$$z_+^T L z_+ = \sum_{(i,j)\in E^+} (z[i] - z[j])^2.$$

We will prove the theorem by showing that for one of the sets

$$\forall j = 1, \ldots, k+1, \quad S_j = \{i : i \leq j\},$$

we have

$$\frac{z_+^T L z_+}{z_+^T D z_+} \geq \frac{\mathrm{sp}(S_j)^2}{2}.$$

To this end, set

$$\sigma = \min_j \mathrm{sp}(S_j).$$

By our choice of c above and in particular (56),

$$d(S_j) \leq d(V)/2,$$

and so

$$\mathrm{sp}(S_j) = \frac{|\partial(S_j)|}{d(S_j)} \geq \sigma. \tag{58}$$

Set:

$$\forall 1 \leq i \leq k, \quad a_i = |\{(i,j) \in E^+ : j > i\}|,$$

and

$$\forall 1 \leq i \leq k, \quad b_i = |\{(i,j) \in E^+ : j < i\}|.$$

Note that we have

$$d(i) = a_i + b_i$$

and (you should convince yourself of this one)

$$|\partial(S_j)| = \sum_{i=1}^{j} (a_i - b_i). \tag{59}$$

We thus have

$$d(S_j) = \sum_{i=1}^{j} d(i) = \sum_{i=1}^{j} (a_i + b_i),$$

which together with the previous line implies that

$$\sum_{i=1}^{j} a_i = \frac{d(S_j) + |\partial(S_j)|}{2}.$$

It follows that

$$\sum_{i=1}^{j} b_i = \frac{d(S_j) - |\partial(S_j)|}{2}. \tag{60}$$

By (58) and (59) we have:

$$\sum_{i=1}^{j} (a_i - b_i) \geq \sigma \sum_{i=1}^{j} d(i).$$

Now we compute, using Lemma 5 in the second line:

$$\begin{aligned}
z_+^T L z_+ &= \sum_{(i,j)\in E^+} (z[i] - z[j])^2 \\
&\geq \sum_{(i,j)\in E^+} \left(\sigma z[i]^2 - \frac{\sigma}{1-\sigma} z[j]^2 \right) \\
&= \sigma \sum_{(i,j)\in E^+} z[i]^2 - \frac{\sigma}{1-\sigma} \sum_{(i,j)\in E^+} z[j]^2 \\
&= \sigma \sum_{i=1}^{k} \sum_{\substack{j:(i,j)\in E^+ \\ i<j}} z[i]^2 - \frac{\sigma}{1-\sigma} \sum_{j=1}^{k} \sum_{\substack{i:(i,j)\in E^+ \\ i<j}} z[j]^2 \\
&= \sigma \sum_{i=1}^{k} z[i]^2 \sum_{\substack{j:(i,j)\in E^+ \\ i<j}} 1 - \frac{\sigma}{1-\sigma} \sum_{j=1}^{k} z[j]^2 \sum_{\substack{i:(i,j)\in E^+ \\ i<j}} 1 \\
&= \sigma \sum_{i=1}^{k} a_i z[i]^2 - \frac{\sigma}{1-\sigma} \sum_{j=1}^{k} b_j z[j]^2 \\
&= \sum_{i=1}^{k} \left(a_i \sigma - b_i \frac{\sigma}{1-\sigma} \right) z[i]^2.
\end{aligned}$$

Recalling (57), we thus have:

$$\nu \sum_{i=1}^{k} d(i) z[i]^2 = \nu z_+^T D z_+ \geq z_+^T L z_+ \geq \sum_{i=1}^{k} \left(a_i \sigma - b_i \frac{\sigma}{1-\sigma} \right) z[i]^2.$$

Now we can use Lemma 6, followed by (59), (60), and (58), to conclude there exists a j for which:

$$\begin{aligned}
\nu d(S_j) = \nu \sum_{i=1}^{j} d(i) &\geq \sum_{i=1}^{j} \left(a_i \sigma - b_i \frac{\sigma}{1-\sigma} \right) z[i]^2 \\
&= \sigma \sum_{i=1}^{j} (a_i - b_i) - \frac{\sigma^2}{1-\sigma} \sum_{i=1}^{j} b_i \\
&= \sigma |\partial(S_j)| - \frac{\sigma^2}{1-\sigma} \left[\frac{d(S_j) - |\partial(S_j)|}{2} \right] \\
&\geq \sigma^2 d(S_j) + \frac{\sigma^2 |\partial(S_j)| - \sigma^2 d(S_j)}{2(1-\sigma)} \\
&\geq \frac{2(1-\sigma)\sigma^2 d(S_j) + \sigma^3 d(S_j) - \sigma^2 d(S_j)}{2(1-\sigma)} \\
&= \frac{\sigma^2 (1-\sigma) d(S_j)}{2(1-\sigma)} \\
&= \frac{\sigma^2}{2} d(S_j).
\end{aligned}$$

We thus conclude:

$$\nu \geq \frac{\sigma^2}{2}.$$

As we chose $\sigma = \min_j \mathrm{sp}(S_j)$, this tells us there is a j for which (namely the j such that $\mathrm{sp}(S_j) = \sigma$),

$$\nu \geq \frac{\sigma^2}{2} = \frac{\mathrm{sp}(S_j)^2}{2} \geq \frac{1}{8} \phi(S_j)^2,$$

where the last inequality follows from $\mathrm{sp}(S) \geq \phi(S)/2$ for any $S \subset V$. □

Exercise 29. Examine the proof of Cheeger's Theorem to determine the value of t that determines the set S. Use this to develop an algorithm for computing a set S such that $\gamma_2 \geq \phi(S)^2/8$.

Exercise 30. Implement the algorithm from the previous exercise. Now go back to the MNIST data set. Make a data set out of two of the digits, and construct a graph in which the vertices are the images, and the edges are constructed by some notion of similarity that you deem appropriate. Compute the cut S - does it partition the data into the two digits? If so, you have achieved unsupervised digit recognition!

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.

[11] Jianzhong Wang. *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*. Springer, 2012.