Lecture 10

Introduction to Spectral Graph Theory

Spectral graph theory is the study of a graph through the properties of the eigenvalues and eigenvectors of its associated Laplacian matrix. In the following, we use $G = (V, E)$ to represent a undirected graph with no self-loops, and write $V = \{1, \ldots, n\}$, with the degree of vertex i being denoted by d_i. For undirected graphs our convention will be that if there is an edge then both $(i, j) \in E$ and $(j, i) \in E$. Thus $\sum_{(i,j)\in E} 1 = 2|E|$. If we wish to sum over edges only once, we will write $\{i, j\} \in E$ for the unordered pair. Thus $\sum_{\{i,j\}\in E} 1 = |E|$.

10.1 Matrices associated to a graph

Given an undirected graph G, we define the following matrices:

Definition 10.1 (Adjacency Matrix)**.** The adjacency matrix $A \in \{0, 1\}^{n\times n}$ is defined as

$$A_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in E \\ 0 & \text{otherwise.} \end{cases}$$

Note that A is always a symmetric matrix with exactly d_i ones in the i-th row and the i-th column.

Definition 10.2 (Degree Matrix)**.** The degree matrix $D \in \mathbb{R}^{n\times n}$ is defined as the diagonal matrix with diagonal entries $(d_1, \ldots, d_n)$.

Definition 10.3 (Normalized Adjacency Matrix)**.** The normalized adjacency matrix is defined as

$$\overline{A} = D^{-1}A.$$

Note that this is not necessarily a symmetric matrix. It will be useful when we consider random walks on graphs.

Definition 10.4 (Laplacian and normalized Laplacian Matrix)**.** The Laplacian matrix is defined as

$$L = D - A.$$

The normalized Laplacian is defined as

$$\overline{L} = D^{-1/2} L D^{-1/2} = \mathbb{I} - D^{-1/2} A D^{-1/2}.$$

$\overline{L}$ is always a symmetric matrix. It is the main matrix we will work with.

In spectral graph theory, a basic question is: which properties of G can be inferred from the *eigenvalues* and *eigenvectors* of $\overline{A}$ and $\overline{L}$? Of course, the whole matrices encode the whole graph. But we will see that surprisingly, a lot of information can be read solely from just a few eigenvalues, typically the smallest and largest ones, and the associated eigenvectors. It will be convenient to always order the eigenvalues of A in deceasing order, $\mu_1 \geq \cdots \geq \mu_n$, and those of $\overline{L}$ in increasing order, $\lambda_1 \leq \cdots \leq \lambda_n$.

So why would the eigenvalues of $\overline{A}$ or $\overline{L}$ have anything interesting to say? Let's do some simple examples:

Example 10.5. Consider the graph shown in Figure. 10.1.

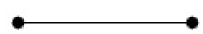

Figure 10.1: A single edge

The adjacency matrix is

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Note that in writing down A we have some liberty in ordering the rows and columns. But this does not change the spectrum as simultaneous reordering of the rows and the columns corresponds to conjugation by a permutation, which is orthogonal and thus preserves the spectral decomposition. We can also compute

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \overline{L}.$$

The spectrum of $\overline{L}$ is given by $\lambda_1 = 0$, $\lambda_2 = 2$. The corresponding eigenvectors are

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Example 10.6. Consider the graph shown in Figure. 10.2.

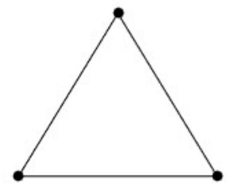

Figure 10.2: The triangle graph

The adjacency matrix is given by

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

We can also compute

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{pmatrix}.$$

The eigenvalues of $\overline{L}$ are $0, 3/2, 3/2$ with corresponding eigenvectors

$$\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

where since the second eigenvalue $3/2$ is degenerate we have freedom in choosing a basis for the associated 2-dimensional subspace.

Example 10.7. As a last example, consider the path of length two, pictured in Figure. 10.3.

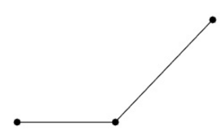

Figure 10.3: The path of length 2

The adjacency matrix is given by

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

We can also compute

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 1 & -1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1 & -1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1 \end{pmatrix}.$$

The eigenvalues of $\overline{L}$ are $0, 1, 2$ with corresponding eigenvectors

$$\frac{1}{2}\begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}, \quad \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \frac{1}{2}\begin{pmatrix} -1 \\ \sqrt{2} \\ -1 \end{pmatrix}.$$

We've seen three examples — do you notice any pattern? 0 seems to always be the smallest eigenvalue. Moreover, in two cases the associated eigenvector has all its coefficients equal. In the case of the path, the middle coefficient is larger — this seems to reflect the degree distribution in some way. Anything else? The largest eigenvalue is not always the same. Sometimes there is a degenerate eigenvalue.

Exercise 1. Show that the largest eigenvalue of the normalized Laplacian $\lambda_n = 2$ if and only if G is bipartite.

We will see that much more can be read about G from $\overline{L}$ in a systematic way. The main connection between eigenvalues of $\overline{L}$ and combinatorial properties of G will rely on the Courant-Fisher theorem:

Theorem 10.8 (Variational Characterization of Eigenvalues)**.** *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues $\mu_1 \geq \cdots \geq \mu_n$, and let the corresponding eigenvectors be $v_1, \cdots, v_n$. Then*

$$\mu_1 = \sup_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{x^T M x}{x^T x}$$

$$\mu_2 = \sup_{\substack{x\in\mathbb{R}^n \\ x\perp v_1}} \frac{x^T M x}{x^T x}$$

$$\vdots$$

$$\mu_n = \sup_{\substack{x\in\mathbb{R}^n \\ x\perp v_1,\dots,v_{n-1}}} \frac{x^T M x}{x^T x} = \inf_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{x^T M x}{x^T x}$$

Proof. By the spectral theorem, we can write

$$M = \sum_{i=1}^{n} \mu_i v_i v_i^T, \tag{10.1}$$

where $\{v_1, \ldots, v_n\}$ are an orthonormal basis of $\mathbb{R}^n$ formed of eigenvectors of M. For $1 \leq k \leq n$, we have

$$\mu_k \leq \sup_{\substack{x \in \mathbb{R}^n \\ x \perp v_1, \ldots, v_{k-1}}} \frac{x^T M x}{x^T x} \tag{10.2}$$

because by taking $x = v_k$ and using (10.1) together with $v_i^T v_k = 0$ for $i \neq k$ we immediately get

$$\frac{x^T M x}{x^T x} = \mu_k.$$

To show the reverse inequality, observe that any x such that $x^T x = 1$ and $x \perp v_1, \ldots, v_{k-1}$ can be decomposed as $x = \sum_{j=k}^n \alpha_j v_j$ with $\sum_j \alpha_j^2 = 1$. Now

$$x^T M x = \sum_{i,j=k}^n \sum_{l=1}^n \mu_l \alpha_i \alpha_j v_i^T v_l v_l^T v_j = \sum_{l=k}^n \mu_l \alpha_l^2 \leq \mu_k$$

since the eigenvalues are ordered in decreasing order. Thus

$$\mu_k \geq \sup_{\substack{x \in \mathbb{R}^n \\ x \perp v_1, \ldots, v_{k-1}}} \frac{x^T M x}{x^T x},$$

which together with (10.2) concludes the proof. □

10.2 The normalized Laplacian

The following gives a very useful interpretation of the quadratic form $x \mapsto x^T \overline{L} x$ associated with the normalized Laplacian.

Claim 10.9. *$\forall x \in \mathbb{R}^n$, we have*

$$x^T \overline{L} x = \sum_{\{i,j\} \in E} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2. \tag{10.3}$$

If G is d-regular, then this simplifies to

$$x^T \overline{L} x = \frac{1}{d} \sum_{\{i,j\} \in E} (x_i - x_j)^2.$$

Proof.

$$\begin{aligned} x^T \overline{L} x &= x^T x - x^T D^{-1/2} A D^{-1/2} x \\ &= \sum_i x_i^2 - \sum_{i,j} \frac{x_i}{\sqrt{d_i}} A_{ij} \frac{x_j}{\sqrt{d_j}} \\ &= \sum_i d_i \left(\frac{x_i}{\sqrt{d_i}} \right)^2 - \sum_{(i,j) \in E} \frac{x_i}{\sqrt{d_i}} \cdot \frac{x_j}{\sqrt{d_i}} \\ &= \sum_{\{i,j\} \in E} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2 . \end{aligned}$$

□

The claim provides the following interpretation of the Laplacian: if we think of the vector x as assigning a weight, or "potential" $x_i \in \mathbb{R}$ to every vertex $v \in V$, then the Laplacian measures the average variation of the potential over all edges. The expression $x^T \overline{L} x$ will be small when the potential x is close to constant across all edges (when appropriately weighted by the corresponding degrees), and large when it varies a lot, for instance when potentials associated with endpoints of an edge have a different sign. Now by the variational characterization of eigenvalues we know how to relate this expression to the eigenvalues of $\overline{L}$, and we get the following corollary.

Claim 10.10. *For any graph G with normalized Laplacian $\overline{L}$, $0 \leq \overline{L} \leq 2\mathbb{I}$. Moreover, if λ_1 is the smallest eigenvalue of $\overline{L}$ then $\lambda_1 = 0$ with multiplicity equal to the number of connected components of G.*

Proof. From (10.3) we see that $x^T \overline{L} x \geq 0$ for any x, and using $(a-b)^2 \leq 2(a^2 + b^2$ we also have $x^T \overline{L} x \leq 2 x^T x$. Using the variational characterization

$$\lambda_1 = \inf_{x \neq 0} \frac{x^T \overline{L} x}{x^T x}, \qquad \lambda_n = \sup_{x \neq 0} \frac{x^T \overline{L} x}{x^T x},$$

where λ_n is the largest eigenvalue, we see that $0 \leq \overline{L} \leq 2\mathbb{I}$.

To see that $\lambda_1 = 0$ alwyas with multiplicity at least 1 it suffices to consider the vector

$$v_1 = \begin{pmatrix} \sqrt{d_1} \\ \vdots \\ \sqrt{d_n} \end{pmatrix},$$

for which $v_1^T \overline{L} v_1 = 0$.

Now suppose G has exactly L connected components. By choosing a vector equal to $\sqrt{d_i}$ for all i that belong to a given connected component and 0 elsewhere we can construct as many orthogonal vectors v such that $v^T \overline{L} v = 0$ as there are connected components. Thus the multiplicity of the eigenvalue 0 is at least as large as the number of connected components.

To show the converse, note that from (10.3) we see that up to normalization any v such that $v^T \overline{L} v = 0$ must be such that $v_i/\sqrt{d_i}$ is constant across each connected component. Thus the dimension of the subspace of all v such that $v^T \overline{L} v = 0$ is effectively at most the number of connected componentens, and there can be at most k linearly independent such vectors: the multiplicity of the eigenvalue 0 is at most the number of connected components. □

An immediate corollary worth stating explicitly is as follows:

Claim 10.11. *For any graph G, the second smallest eigenvalue $\lambda_2(\overline{L}) > 0$ if and only if G is connected.*

These claims show that the small eigenvalues of $\overline{L}$ tells us whether the graph is connected or not. We will make this statement more quantitative by showing that, not only is the question of connectedness related to the question of λ_2 being equal to 0, but in fact the magnitude of λ_2 can be used to quantify, in a precise way, how "well-connected" the graph is. So let us look at a natural measure of connectedness of a graph, its *conductance*. Given a set of edges $S \subsetneq V$, and $S \neq \emptyset$, the boundary of S is defined as

$$\partial S = \big\{\{i,j\} \in E : i \in S, j \notin S\big\}.$$

The conductance of S is

$$\phi(S) = \frac{|\partial S|}{\min\left(d(S), d(V \backslash S)\right)},$$

where $d(S) := \sum_{i \in S} d_i$. If G is d-regular, then this simplifies to

$$\phi(S) = \frac{|\partial S|}{d \cdot \min(|S|, |V \backslash S|)}.$$

Definition 10.12 (Conductance)**.** The conductance of a graph G is defined as

$$\phi(G) = \min_{S : S \neq \emptyset, S \neq V} \phi(S).$$

If G is d-regular, this simplifies to

$$\Phi(G) = \min_{S,\, 1 \leq |S| \leq n/2} \frac{|\partial S|}{d \cdot |S|},$$

the fraction of edges incident on S that have one endpoint outside of S.

The conductance is a measure of how well connected G is. Here are some examples demonstrating this point.

Example 10.13.

- Clearly G is disconnected if and only if there exists a set $S \neq \emptyset$, $S \neq V$ such that $|\partial S| = 0$, i.e. if and only if $\phi(G) = 0$.

- If G is a clique, then

$$\phi(G) = \min_{1 \le k \le n/2} \frac{k(n-k)}{(n-1)k} = \frac{n}{2(n-1)} \approx \frac{1}{2}.$$

- If G is a cycle, then

$$\phi(G) = \min_{1 \le k \le n/2} \frac{2}{2k} = \frac{2}{n}.$$

Exercise 2. Compute the conductance of the hypercube $G = (V, E)$ where $V = \{0,1\}^n$ and $E = \{\{u,v\} \in V : d_H(u,v) = 1\}$, where d_H is the Hamming distance.

The following theorem is the fundamental result relating conductance and the second smallest eigenvalue of the normalized Laplacian.

Theorem 10.14 (Cheeger's inequality)**.** *Let G be an undirected graph with normalized Laplacian $\overline{L} = \mathbb{I} - D^{-1/2}AD^{-1/2}$. Let $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ be the eigenvalues of $\overline{L}$. Then*

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

Remark 10.15.

- Both sides of the inequality are interesting. The left-hand side says that if there is a good cut, that is a cut of small conductance, then there is an eigenvector orthogonal to the smallest eigenvector with small eigenvalue. This is called the "easy" side of Cheeger.

- The right-hand side says that if λ_2 is small, then there must exist a poorly connected set. This is called the "hard" side of Cheeger.

- We will give "algorithmic" proofs of both inequalities: for the left-hand side, given a set S of low conductance we will show how to construct a vector $v \perp v_1$ that achieves a low value in (10.3). For the right-hand side, given a vector $v_2 \perp v_1$ achieving a low value in (10.3) we will construct a set S of low conductance.

- The next exercise shows that both sides of the inequality are tight.

Exercise 3. Show that the left-hand side of Cheeger's inequality is tight by computing the eigenvalues and eigenvectors of the hypercube (hint: Fourier basis). Show that the right-hand side is also tight by considering the example of the n-cycle.

Proof of Cheeger's inequality. We first prove the "*easy side*". Let S be a set of vertices such that $\phi(S) = \phi(G)$. We claim

$$\lambda_2 = \min_{\substack{x \in \mathbb{R}^n \\ x \perp v_1 = (\sqrt{d_1}, \cdots, \sqrt{d_n})}} \frac{x^T \overline{L} x}{x^T x} \le 2\phi(G).$$

To see this, define

$$x = (\underbrace{\sqrt{d_i}, \dots}_{\text{vertices in } S}, \underbrace{-\sigma\sqrt{d_j}, \dots}_{\text{vertices in } \overline{S}}),$$

where $\sigma \frac{d(S)}{d(V\setminus S)}$ is defined so that

$$x^T v_1 = \sum_{i \in S} d_i - \sigma \sum_{i \in \overline{S}} d_i = 0.$$

We then have

$$x^T x = \sum_{i \in S} d_i + \sigma^2 \sum_{i \in V\setminus S} d_i = d(S) + \frac{d(S)^2}{d(V\setminus S)^2} \cdot d(V\setminus S) = \frac{d(S)d(V)}{d(V\setminus S)},$$

and

$$\begin{aligned} x^T \overline{L} x &= \sum_{\{i,j\} \in E} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2 \\ &= \sum_{\{i,j\} \in \partial S} (1+\sigma)^2 = \sum_{\{i,j\} \in \partial S} \left(\frac{d(V\setminus S) + d(S)}{d(V\setminus S)} \right)^2 \\ &= |\partial S| \frac{d(V)^2}{d(V\setminus S)^2}. \end{aligned}$$

This finally implies

$$\frac{x^T \overline{L} x}{x^T x} = \frac{|\partial S| d(V)}{d(S)d(V\setminus S)} \le \frac{2|\partial S|}{\min(d(S), d(V\setminus S))} = 2\phi(G),$$

where the inequality can be seen by considering the cases $d(S) \le d(V\setminus S)$ and $d(S) > d(V\setminus S)$ separately.

Now let's turn to the "*hard side*" of the inequality. Let $y \in \mathbb{R}^n$ be such that

$$\frac{y^T \overline{L} y}{y^T y} \le \lambda_2 \tag{10.4}$$

and $y \perp v_1 = (\sqrt{d_1}, \dots, \sqrt{d_n})^T$. We start with a few convenient normalization manipulations. Let $z = D^{-1/2} y - \sigma \mathbf{1}$ for some σ to be determined soon and $\mathbf{1} = (1, \dots, 1)^T$. Since $\mathbf{1}^T L \mathbf{1} = 0$, we see that $z^T L z = y^T \overline{L} y$. Moreover, $D\mathbf{1} = \sigma_1$, thus

$$z^T D z = y^T y - 2\sigma \underbrace{v_1^T \cdot y}_{=0} + \sigma^2 d(V) \ge y^T y,$$

and using (10.4) we get that for any σ,

$$\frac{z^T L z}{z^T D z} \le \lambda_2.$$

We make the following conventions, without loss of generality:

- Order the coordinates of z so that $z_1 \leq \ldots \leq z_n$.
- Choose σ such that $z_{i_0} = 0$, where i_0 is such that

$$\sum_{i<i_0} d_i < \frac{d(V)}{2} \qquad \sum_{i\leq i_0} d_i \geq \frac{d(V)}{2}. \tag{10.5}$$

- Scale z so that $z_1^2 + z_n^2 = 1$.

Let $t \in [z_1, z_n]$ be chosen according to the distribution with density $2|t|$ (the scaling on z assumed above ensures that this is a properly normalized probability density). Observe that for any $a < b$,

$$\begin{aligned}
\mathbf{Pr}(t \in [a,b]) &= \int_a^b 2|t|dt \\
&= \begin{cases} b^2 + a^2 & \text{if } a < 0 < b \\ b^2 - a^2 & \text{if } b > a > 0 \\ a^2 - b^2 & \text{otherwise} \end{cases} \\
&\leq |b-a|\left(|a| + |b|\right),
\end{aligned}$$

an inequality that is easily verified in all three cases. For any t, let $S_t = \{i : z_i \leq t\}$. Then

$$\mathop{\mathbf{E}}_t d(S_t) = \sum_i \mathbf{Pr}(i \in S_t) d_i = \sum_i \mathbf{Pr}(z_i \leq t) d_i.$$

Our choice of the index i_0 in (10.5) ensures that, if $t < 0$ then $\min(d(S_t), d(V \backslash S_t)) = d(S_t)$, while if $t \geq 0$ then $\min(d(S_t), d(V \backslash S_t)) = d(V \backslash S_t)$. Thus

$$\begin{aligned}
\mathop{\mathbf{E}}_t \min\left(d(S_t), d(V \backslash S_t)\right) &= \sum_{i<j} \mathbf{Pr}(z_i \leq t \wedge t < 0) d_i + \sum_{i \geq j} \mathbf{Pr}(z_i > t \wedge t \geq 0) d_i \\
&= \sum_{i<j} z_i^2 d_i + \sum_{i\geq j} z_i^2 d_i \qquad (10.6) \\
&= z^T D z. \qquad (10.7)
\end{aligned}$$

Next we compute

$$\begin{aligned}
\mathbf{E}\,|\partial S_t| &= \sum_{\{i,j\}\in E} \mathbf{Pr}\left(z_i \leq t \leq z_j\right) \\
&\leq \sum_{\{i,j\}\in E} |z_j - z_i|(|z_i| + |z_j|) \\
&\leq \underbrace{\sqrt{\sum_{\{i,j\}\in E} (z_i - z_j)^2}}_{\sqrt{z^T L z}} \quad \underbrace{\sqrt{\sum_{\{i,j\}\in E} (|z_i| + |z_j|)^2}}_{\leq \sqrt{2 \sum_{\{i,j\}\in E} (|z_i|^2 + |z_j|^2)} = \sqrt{2 z^T D z}} \\
&\leq \sqrt{2\lambda_2} z^T D z, \qquad (10.8)
\end{aligned}$$

where the second inequality is Cauchy-Schwarz. Combining (10.6) and (10.8),

$$\mathbf{E}\,|\partial S_t| \leq \sqrt{2\lambda_2}\,\mathop{\mathbf{E}}_{t} \min\left(d(S_t), d(v \backslash S_t)\right),$$

which we can rewrite as

$$\mathop{\mathbf{E}}_{t}\left[\sqrt{2\lambda_2}\,\min\left(d(S_t), d(v \backslash S_t)\right) - |\partial S_t|\right] \,\geq\, 0.$$

From there we deduce that there exists a choice of t such that

$$\Phi(S_t) \leq \sqrt{2\lambda_2},$$

which immediately gives us that $\Phi(G) \leq \sqrt{2\lambda_2}$, as desired. □

We note that the proof given above is *algorithmic*, in that it describes an efficient algorithm that, given a graph, will produce a set with conductance at most $\sqrt{2\lambda_2}$: simply compute an eigenvector associated with the second smallest eigenvalue (using, e.g., the power method), and output the set S_t which has smallest conductance among the n possibilities: this can be checked efficiently.