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 an undirected $n$-vertex graph with no self-loops, and write $V = \{1, \ldots, n\}$, with the degree of vertex $i$ denoted $d_i$. 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 $G$, the most natural matrix associated to it is its adjacency matrix:

**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. While $A$ is a natural representation of $G$ when we think of a matrix as a table of numbers used to store information, it is less natural if we think of a matrix as an operator, a linear transformation which acts on vectors. The most natural operator associated with a graph is the diffusion operator, which spreads a quantity supported on any vertex equally onto its neighbors. To introduce the diffusion operator, first consider the degree matrix:

**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

$$\overline{A} = AD^{-1}.$$
Note that this is not necessarily a symmetric matrix. But it is a row-stochastic matrix: each row has non-negative entries that sum to 1. This means that if \( p \in \mathbb{R}^n_+ \) is a probability vector defined over the vertices, then \( Ap \) is another probability vector, obtained by “randomly walking” along an edge, starting from a vertex chosen at random according to \( p \). We will explore the connection between \( A \) and random walks on \( G \) in much more detail in the next lecture.

Finally we introduce the Laplacian matrix, which will provide us with a very useful quadratic form associated to \( G \):

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

\[
L = D - A.
\]
The normalized Laplacian is defined as

\[
\overline{L} = D^{-1/2}LD^{-1/2} = \mathbb{I} - D^{-1/2}AD^{-1/2}.
\]

Note that \( L \) and \( \overline{L} \) are always symmetric. They are best thought of as quadratic forms: for any \( x \in \mathbb{R}^n \),

\[
x^T L x = \sum_i d_i x_i^2 - \sum_{(i,j) \in E} x_i x_j = \sum_{\{i,j\} \in E} (x_i - x_j)^2.
\]

For the normalized Laplacian, we have the following claim.

**Claim 10.5.** \( \forall x \in \mathbb{R}^n, \) we have

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

(10.1)

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.**

\[
x^T \overline{L} x = x^T x - x^T D^{-1/2}AD^{-1/2} x
\]

\[
= \sum_i x_i^2 - \sum_{i,j} \frac{x_i}{\sqrt{d_i}} A_{ij} \frac{x_j}{\sqrt{d_j}}
\]

\[
= \sum_i d_i \left( \frac{x_i}{\sqrt{d_i}} \right)^2 - \sum_{(i,j) \in E} \frac{x_i}{\sqrt{d_i}} \cdot \frac{x_j}{\sqrt{d_i}}
\]

\[
= \sum_{\{i,j\} \in E} \left( \frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2.
\]

\( \square \)
Claim 10.5 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 $x^T L x$ 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.

We will return to this interpretation soon. Let’s first see some examples. It will be convenient to always order the eigenvalues of $A$ in deceasing order, $\mu_1 \geq \cdots \geq \mu_n$, and those of $L$ in increasing order, $\lambda_1 \leq \cdots \leq \lambda_n$. So what do the eigenvalues of $A$ or $L$ have to say?

**Example 10.6.** Consider the graph shown in Figure 10.1.

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

The adjacency matrix is

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 10.7.** Consider the graph shown in Figure 10.2.

\[ L = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \]

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 \( L \) 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 10.8.** As a last example, consider the path of length two, pictured in Figure 10.3.

![Figure 10.3: The path of length 2](image)

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 \( L \) are 0, 1, 2 with corresponding eigenvectors

\[
\frac{1}{2} \begin{pmatrix} \sqrt{2} \\ 1 \end{pmatrix}, \quad \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \frac{1}{2} \begin{pmatrix} -1 \\ \sqrt{2} \\ -1 \end{pmatrix}.
\]
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 combinatorial properties of \( G \) from \( L \) in a systematic way. The main connection is provided by the Courant-Fisher theorem:

**Theorem 10.9** (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 \( v_1, \ldots, v_n \). Then

\[
\mu_1 = \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{x^T M x}{x^T x}, \quad \mu_2 = \sup_{x \in \mathbb{R}^n, x \perp v_1} \frac{x^T M x}{x^T x}, \quad \ldots \quad \mu_n = \sup_{x \in \mathbb{R}^n, x \perp v_1, \ldots, v_{n-1}} \frac{x^T M x}{x^T x} = \inf_{x \in \mathbb{R}^n, x \neq 0} \frac{x^T M x}{x^T x}.
\]

**Proof.** By the spectral theorem, we can write

\[
M = \sum_{i=1}^n \mu_i v_i v_i^T,
\]

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

\[
\mu_k \leq \sup_{x \in \mathbb{R}^n, x \perp v_1, \ldots, v_{k-1}} \frac{x^T M x}{x^T x} \tag{10.3}
\]

because by taking \( x = v_k \) and using (10.2) 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

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

since the eigenvalues are ordered in decreasing order. Thus

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

which together with (10.3) concludes the proof. \( \square \)
Using this variational characterization of eigenvalues, we can connect the quadratic form associated with the normalized Laplacian in Claim 10.5 to the eigenvalues of $\mathcal{L}$.

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

**Proof.** From (10.1) we see that $x^T \mathcal{L} x \geq 0$ for any $x$, and using $(a - b)^2 \leq 2(a^2 + b^2$ we also have $x^T \mathcal{L} x \leq 2x^T x$. Using the variational characterization

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

where $\lambda_n$ is the largest eigenvalue, we see that $0 \leq \mathcal{L} \leq 2\mathbb{I}$. To see that $\lambda_1 = 0$ always 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 $v_1^T \mathcal{L} v_1 = 0$.

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 $v^T \mathcal{L} v = 0$ 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 (10.1) we see that up to normalization any $v$ such that $v^T \mathcal{L} v = 0$ 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 $v^T \mathcal{L} v = 0$ is effectively at most the number of connected components, 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. \qed

An immediate corollary worth stating explicitly is as follows:

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

These claims show that the small eigenvalues of $\mathcal{L}$ tell 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 $\lambda_2$ being equal to 0, but in fact the magnitude of $\lambda_2$ can be used to quantify, in a precise way, how “well-connected” the graph is. Let us look at a natural measure of connectedness of a graph, its conductance. Given a set of vertices $\emptyset \subset S \subset V$, the boundary of $S$ is defined as

$$
\partial S = \{ \{i, j\} \in E : i \in S, j \notin S \}.
$$
The conductance of $S$ is
\[ \phi(S) = \frac{|\partial S|}{\min(d(S), d(V\setminus S))}, \]
where $d(S) := \sum_{i \in S} d_i$ is a natural measure of volume: the total number of edges incident on vertices in $S$. If $G$ is $d$-regular, then this simplifies to
\[ \phi(S) = \frac{|\partial S|}{d \cdot \min(|S|, |V\setminus S|)}. \]

**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
  \[ \phi(G) = \min_{1 \leq k \leq n/2} \frac{k(n-k)}{(n-1)k} = \frac{n}{2(n-1)} \approx \frac{1}{2}. \]
- If $G$ is a cycle, then
  \[ \phi(G) = \min_{1 \leq k \leq n/2} \frac{2}{2k} = \frac{2}{n}. \]

**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 a 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 $\mathcal{L} = \mathbb{I} - D^{-1/2}AD^{-1/2}$. Let $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $\mathcal{L}$. Then
\[ \frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}. \]
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 $\lambda_2$ 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 (10.1). For the right-hand side, given a vector $v_2 \perp v_1$ achieving a low value in (10.1) 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.

We’ll prove the inequality in the next lecture.