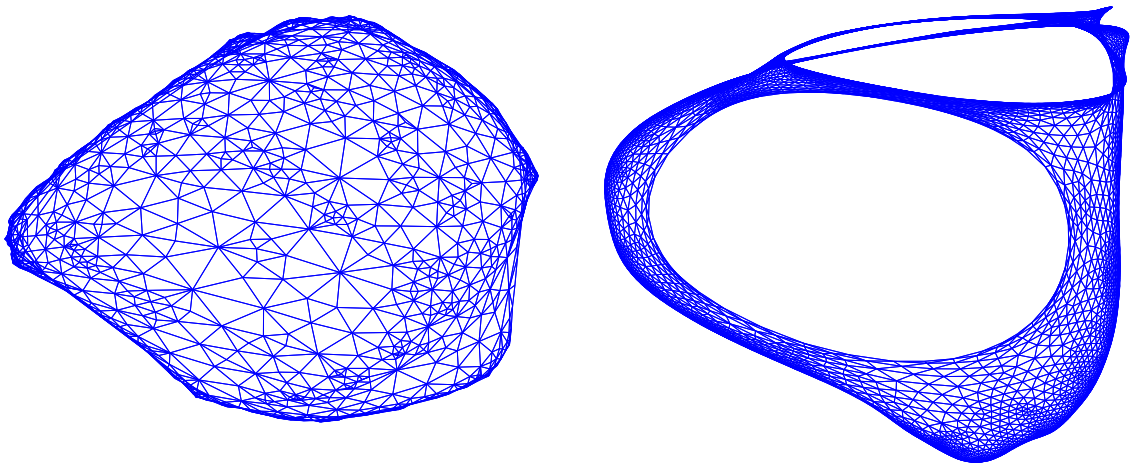
Algorithms II December 8, 2025

Lecture 24: Introduction to Spectral Graph Theory

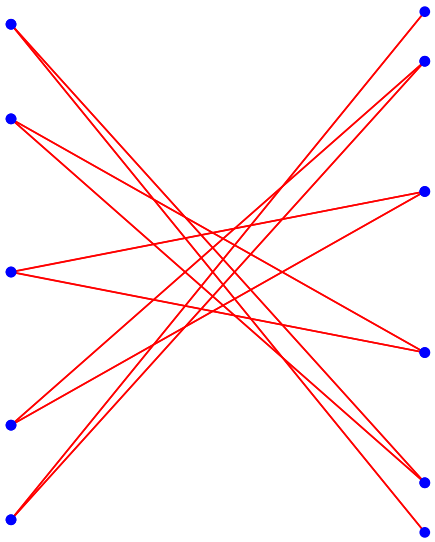
Notes by Ola Svensson[1]

Spectral Graph Theory studies how the eigenvalues and eigenvectors of the adjacency matrix (or related matrices), which are purely algebraic quantities, relate to combinatorial properties of the graph. This is a very active research area with numerous modern applications such as image segmentation, finding min/max cuts, and clustering (community detection).

Let us start with some images to increase our appetite for learning more about this area. The following two graphs are drawn using the coordinates of two eigenvectors corresponding to *large* eigenvalues of the adjacency matrix



and the following graph is drawn using the coordinates of two eigenvectors corresponding to *small* eigenvalues of the adjacency matrix



[1]**Disclaimer:** These notes were written as notes for the lecturer. They have not been peer-reviewed and may contain inconsistent notation, typos, and omit citations of relevant works.

Notice that when we draw graphs using eigenvectors corresponding to large eigenvalues of the adjacency matrix (first two drawings), we have that adjacent vertices are depicted nearby. When we select eigenvectors corresponding to small eigenvalues, the opposite seems to happen (last drawing). In this lecture, our goal is to explain some of these behaviors.

Before continuing, let me remark that there is a lot to say about spectral graph theory (and we have only two lectures). Two courses devoted to the subject are [1] and [2]. The first two drawings are taken from [1] and the last one from [3].

1 The (Normalized) Adjacency Matrix and Eigenvalues/Eigenvectors

Recall the definition of the adjacency matrix of an undirected graph:

Definition 1 *The adjacency matrix A of graph $G = (V, E)$ of $|V| = n$ vertices is a matrix in $\mathbb{R}^{n \times n}$ defined by*

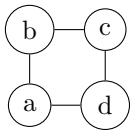
$$A_{ij} = 1 \text{ if and only if } \{i, j\} \in E$$

for every two vertices $i, j \in V$.

It will also be convenient to work with the so-called normalized adjacency matrix. **To simplify notation, we assume throughout the lecture that all graphs that we work with are d-regular**. That is, the degree of each vertex equals d. However, all definitions and results can be generalized to the case when vertices have different degrees.

Definition 2 *The normalized adjacency matrix M of a d-regular graph is equal to $\frac{1}{d}A$, with A being the adjacency matrix.*

Example (2-regular graph):



$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \qquad M = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

The normalized adjacency matrix M is also called the random walk matrix of a graph. The reason for this becomes clear if we consider the above example graph. Suppose you start at vertex a corresponding. Hence, the distribution over the vertices where you are standing can be represented by the vector $(1, 0, 0, 0)$ you are with probability 1 at vertex a and probability 0 at any other vertex. Now suppose you go to a random neighbor in your graph, then with probability $1/2$ you end up at vertex b and with probability $1/2$ you end up at vertex d. Thus after one random step, the distribution of your location is represented by the vector $(0, 1/2, 0, 1/2)$. Now notice that in the example above

$$M \begin{pmatrix} 1 \\ 0 \\ \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1/2 \\ 0 \\ 1/2 \end{pmatrix}$$

More generally, for any starting distribution $p \in \mathbb{R}^n$ (where p_i denotes the probability to be at vertex $i \in V$ initially), Mp is the probabilitydistribution after a single random step. Repeating this argument gives that $M^k p$ is the distribution after k random steps.

1.1 Basic properties of eigenvectors/eigenvalues of M

The normalized adjacency matrix M has some special structure, which we will exploit:

Observation 3 *M is a real symmetric matrix.*

Recall the definition of eigenvalue and eigenvector of a matrix.

Definition 4 *A vector v is an eigenvector of a matrix M with eigenvalue λ if*

$$Mv = \lambda v$$

The following is a fact derived using standard linear algebra (and not proved in this course):

Fact 5 *If $M \in \mathbb{R}^{n \times n}$ is symmetric, then:*

1. *M has n non-necessarily distinct real eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$.*
2. *If $v_1, v_2, \ldots, v_{i-1}$ are eigenvectors for $\lambda_1, \lambda_2, \ldots, \lambda_{i-1}$, then λ_i equals the maximum value λ such that there is a vector v_i orthogonal to $v_1, \ldots, v_{i-1}$ such that $Mv_i = \lambda v_i$. Moreover, any such vector v_i can be selected to be the eigenvector corresponding to λ_i.*
   This means in particular that no matter how the first eigenvector v_1 is chosen, we can always find an orthonormal basis (corresponding to eigenvectors).

Example: (using the same 4-cycle)

$$\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = 0, \lambda_4 = -1$$

$$v_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} v_2 = \begin{pmatrix} 0 \\ -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} v_3 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} v_4 = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

In the next section, we will get a better intuition regarding the eigenvalues of the normalized adjacency matrix and relate them to basic combinatorial properties.

2 Relating the eigenvalues to basic combinatorial properties

The following gives a very good intuition to keep in mind when dealing with spectral graph theory. It already tells you that for a large eigenvalue, we would like adjacent vertices to have similar values (as we saw in the drawings).

Observation 6 *Consider $x \in \mathbb{R}^n$ which assigns a value $x(i)$ to each vertex $i \in V$ and let $y = Mx$, where M is the normalized adjacency matrix of a graph $G = (V, E)$. Then*

$$y(i) = \sum_{\{i,j\} \in E} \frac{x(j)}{d},$$

which is the average value according to x of v's neighbours.

Using this observation, the following properties become quite easy to prove

Lemma 7 *Let M be the normalized adjacency matrix of a d-regular graph G and let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be its eigenvalues. Then:*

1. $\lambda_1 = 1$.
2. $\lambda_2 = 1 \iff$ *G is disconnected.*
3. $\lambda_n = -1 \iff$ *one component of G is bipartite.*

Proof of 1: Since $M \begin{pmatrix} 1 \\ 1 \\ ... \\ 1 \end{pmatrix} = 1 \times \begin{pmatrix} 1 \\ 1 \\ ... \\ 1 \end{pmatrix}$, 1 is an eigenvalue and therefore λ_1, the greatest of all eigenvalues, must be greater or equal to 1.
Additionally, if we consider any eigenvector x and $i \in V$ such that $x(i)$ is maximized and $y = Mx$, we have $y(i) = \sum_{\{i,j\} \in E} \frac{x(j)}{d} \leq \sum_{\{i,j\} \in E} \frac{x(i)}{d} = x(i)$. Therefore, $\lambda_1 \leq 1$. ■

Proof of 2: By the proof of property 1 we saw that we can select the first eigenvector v_1 corresponding to $\lambda_1 = 1$ to be the all 1's vector. We will show that there is a vector $v_2 \perp v_1$ so that $Mv_2 = v_2$ if and only if G is disconnected. In other words, $\lambda_2 = 1$ if and only if G is disconnected.

Suppose first that the graph G is disconnected and so that there is a subset $S \subsetneq V$ of vertices that are not connected to vertices in $V \setminus S$. define v_2 by

$$v_2(i) = \begin{cases} 1/|S| & \text{if } i \in S \\ -1/|V \setminus S| & \text{if } i \in V \setminus S \end{cases}$$

Notice that v_2 is perpendicular to the all 1's vector v_1, i.e., $v_2 \perp v_1$. We now show that $Mv_2 = v_2$. Fix a vertex i and let $y = Mv_2$. We have that

$$y(i) = \frac{1}{d} \sum_{\{j,i\} \in E} v_2(j) = \frac{1}{d} \sum_{\{j,i\} \in E} v_2(i) = v_2(i)\,,$$

where the second inequality follows from that every neighbor of i has the same value (with respect to v_2) by definition. Hence, we have $\lambda_2 = 1$ if G is disconnected.

Now suppose that G is connected. Now let v_2 be an eigenvector corresponding to the second eingevalue λ_2. We have that v_2 is is perpendicular to the all 1's vector v_1, We now show that $\lambda_2 < 1$. Indeed, since $v_2 \perp v_1$, v_2 cannot assign the same value to all vertices. Therefore, as G is connected, there must be a vertex i that has at least one neighbor j for which $v_2(i) \neq v_2(j)$. Select such a vertex i that maximizes $v_2(i)$. By the selection of i we have that for any $\{i,j\} \in E$ we have $v_2(i) \geq v_2(j)$ and for at least one neighbor j^* we have $v_2(i) > v_2(j^*)$. Again let $y = Mv_2$. It follows that

$$y(i) = \frac{1}{d} \sum_{\{j,i\} \in E} v_2(j) \leq \frac{1}{d} \left(\sum_{\{j,i\} \in E : j \neq j^*} v_2(i) + v_2(j^*) \right) < v_2(i)$$

and thus $\lambda_2 < 1$. ■

The proof of property 3 is left as an exercise.
Remark The second property can be generalized: $|\{i|\lambda_i = 1\}|$ is the number of connected components in G.

3 Eigenvalues as solutions to optimization problems

In the next lecture, we will see how the eigenvalues gives us quantifiable bounds about the connectivity of G. (Instead of only the qualitively bounds we gave above, saying the graph was disconnected etc..) For those proofs, it will be convenient to consider an alternative way to define the eigenvalues of a real symmetric matrix M. Namely as the solution to the problem of maximizing the Rayleigh quotient

$$\frac{x^T M x}{x^T x}$$

Lemma 8 *Given real summetric matrix $M \in \mathbb{R}^{n \times n}$, $\lambda_1 = \max_{x \in \mathbb{R}^n} \frac{x^T M x}{x^T x}$.*

Proof

1. First, let's prove $\lambda_1 \leq \max_{x \in \mathbb{R}^n} \frac{x^T M x}{x^T x}$:

   Indeed, $\frac{v_1^T M v_1}{v_1^T v_1} = \frac{v_1^T \lambda_1 v_1}{v_1^T v_1} = \lambda_1 \frac{v_1^T v_1}{v_1^T v_1} = \lambda_1$

2. Next, we need to prove $\lambda_1 \geq \max_{x \in \mathbb{R}^n} \frac{x^T M x}{x^T x}$

   Let y be the vector that attains the maximum value.

   Since $(v_1, v_2, ..., v_n)$ is a basis, $\exists(\alpha_1, \alpha_2, ..., \alpha_n) : y = \sum_{i=1}^{n} \alpha_i v_i$

   Then $\frac{y^T M y}{y^T y} = \frac{\sum_{i=1}^{n} \alpha_i^2 \lambda_i}{\sum_{i=1}^{n} \alpha_i^2} \leq \lambda_1 \frac{\sum_{i=1}^{n} \alpha_i^2}{\sum_{i=1}^{n} \alpha_i^2} = \lambda_1.$

■

Lemma 9 *Given real summetric matrix $M \in \mathbb{R}^{n \times n}$, $\lambda_2 = \max_{x \in \mathbb{R}^n : x \perp v_1} \frac{x^T \cdot M \cdot x}{x^T \cdot x}$.*

Proof Exercise (very similar) to the last lemma. ■

4 The mixing time of random walks

Let $G = (V, E)$ be a d-regular graph and M its normalized adjacency matrix. Recall that M is also called the random walk matrix. This is because, if we let p be an initial distribution over the vertices, then $q = Mp$ is the distribution over vertices where $q(v)$ equals the probability that vertex v is output by the following process:

1. Selecting at random an initial vertex v_1 with probability $p(v_1)$.
2. Take a single random step: move to a random neighbor v_2 of v_1.
3. Output v_2.

Similarly $M^k p$ is the distribution over vertices obtained by doing the above process with k random steps instead of one.

Suppose we start our random walk from a single vertex. That is $p = (1, 0, \ldots, 0)$ puts all its probability mass on a single vertex, say s. A natural question is then how many steps k we need to take in order for $M^k p$ to be close to the uniform distribution. This is called the mixing time of the graph G. Some observations are as follows:

- If the graph is not connected, we will only reach vertices in the same component as s and we will thus never (no matter the choice of k) become "close" to the uniform distribution of vertices.
- If the graph is bipartite, then the random walk will alternate between the left-hand and right-hand side of vertices. And thus again it will not converge to a uniform distribution (no matter k)[2].

Notice that by Lemma 7, the above cases correspond to when $\lambda_2 = 1$ (the graph is disconnected) and $\lambda_n = -1$ (the graph is bipartite). The next result shows that the quantity $\max(|\lambda_2|, |\lambda_n|)$ are exactly the quantities that tell us how fast the random walk is mixing:

Lemma 10 *Consider a d-regular graph $G = (V, E)$ and let $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1$ be the eigenvalues of its normalized adjacency matrix M. If $\max(|\lambda_2|, |\lambda_n|) \leq 1 - \epsilon$, then no matter from which vertex s we start, after $O(\frac{1}{\epsilon} \log n)$ steps we will be at any vertex with probability $\approx \frac{1}{n}$. More precisely, if we let p be the vector that is equal to 1 on the vertex where the random walk starts, then*

$$\left|\left| M^k p - \left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \right|\right|_2^2 \leq o\left(\frac{1}{n^2}\right)$$

when $k = \frac{c}{\epsilon} \log n$ for some constant c.

Proof Let's assume, for the sake of simplicity, that the vertices are ordered so that our start vertex s is first. That is, $p^\top = (1, 0, \ldots, 0)$. Let $(v_1, \ldots, v_n)$, where v_i is an eigenvector for the eigenvalue λ_i, be an orthonormal basis. It means we can decompose p on it (we can write p in the eigenvector basis):

$$p = \sum_{i=1}^{n} \alpha_i v_i$$

where $\alpha_i = \langle p, v_i \rangle$. In particular, $\alpha_1 = \langle p, v_1 \rangle = \frac{1}{\sqrt{n}}$. Since we also have $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$, we can write:

$$\begin{aligned}
\left|\left| M^k x - \left(\tfrac{1}{n}, \ldots, \tfrac{1}{n}\right) \right|\right|_2^2 &= \left|\left| \sum_{i=2}^{n} \alpha_i \lambda_i^k v_i \right|\right|_2^2 \\
&= \sum_{i=2}^{n} |\lambda_i|^k ||\alpha_i v_i||_2^2 \qquad \text{(since } v_1, \ldots, v_n \text{ are orthogonal)} \\
&\leq (1-\epsilon)^k \sum_{i=2}^{n} ||\alpha_i v_i||_2^2 \qquad \text{(since } |\lambda_i| \leq (1-\epsilon) \text{ for } i \geq 2) \\
&= (1-\epsilon)^k \left|\left| \sum_{i=2}^{n} \alpha_i v_i \right|\right|_2^2 \qquad \text{(again by orthogonality of } v_i\text{'s)}
\end{aligned}$$

So if we take $k = \frac{c}{\epsilon} \log n$ and let $A = ||\sum_{i=2}^{n} \alpha_i v_i||_2^2 \leq 1$, we have:

$$\left|\left| M^k x - \left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \right|\right|_2^2 \leq A(1-\epsilon)^{\frac{\log n}{\epsilon} c} \leq A \left(\frac{1}{e}\right)^{c \log n} \leq \frac{1}{n^c}$$

If we take $c > 2$, we obtain the result that we wanted. ■

[2]To overcome this, one often does a "lazy" random walk that with probability 1/2 stays at the same vertex and with probability 1/2 moves to a random neighbor.

References

[1] Daniel Spielman: *Spectral Graph Theory.* http://www.cs.yale.edu/homes/spielman/561/

[2] Luca Trevisan: *CS294: Graph Partitioning, Expanders and Spectral Methods.* https://people.eecs.berkeley.edu/~luca/expanders2016/index.html

[3] Tim Roughgarden: *CS168: The Modern Algorithmic Toolbox (Lectures 11 and 12).* http://theory.stanford.edu/~tim/s17/l/l11.pdf