

Lecture 25: Spectral Graph Theory and Connectivity

Notes by Ola Svensson¹

In this lecture, we continue to study spectral graph theory. We will in particular, see the connection between the second largest eigenvalue and the connectivity of the graph.

To begin with, we recall the basic notions from the last lecture.

(Normalized) adjacency matrix.

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 $i, j \in V$.

As in the last lecture, we work with d -regular graphs (all vertices have degree d) to simplify notation.

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

Eigenvectors and basic combinatorial properties. In the last lecture, we related the eigenvalues of M to basic combinatorial properties of G . Recall the definition of an eigenvalue and an eigenvector:

Definition 3 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 4 If $M \in \mathbb{R}^{n \times n}$ is symmetric (as e.g. the normalized adjacency matrix), then:

1. M has n non-necessarily distinct real eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.
2. If v_1, v_2, \dots, v_{i-1} are eigenvectors for $\lambda_1, \lambda_2, \dots, \lambda_{i-1}$, then λ_i equals the maximum value λ such that there is a vector v_i orthogonal to v_1, \dots, 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).

The main result of last lecture was the following:

Lemma 5 Let M be the normalized adjacency matrix of a d -regular graph G and let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ be its eigenvalues. Then:

1. $\lambda_1 = 1$ and the all one vector is a corresponding eigenvector.
2. $\lambda_2 = 1 \iff G$ is disconnected.
3. $\lambda_n = -1 \iff$ one component of G is bipartite.

In the proof, we repeatedly used the following observation that also gives good intuition:

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.

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.

1 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 $p = (1, 0, \dots, 0)$ 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 5, 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 7 Consider a d -regular graph $G = (V, E)$ and let $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq -1$ 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

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

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, $p^\top = (1, 0, \dots, 0)$. Let (v_1, \dots, v_n) , where v_i is an eigenvector for the eigenvalue λ_i ,

²To 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.

be an orthonormal basis (which exists by Fact 4). 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 $(\frac{1}{n}, \dots, \frac{1}{n}) = \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)$, we can write:

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

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:

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

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

2 Graph connectivity and Cheeger's inequalities

In the previous section, we saw that the mixing time is very related to the second eigenvalue λ_2 of the normalized adjacency matrix. It is also easy to imagine that the more “well-connected” a graph is the more rapidly it is mixing: a random walk on a cycle would take much longer time to mix than a random walk on the complete graph.

Moreover, in Lemma 5, we saw that $\lambda_2 = 1$ if and only if the graph is disconnected. This means that if λ_2 then we can cut the graph into two parts without removing *any* edges. But what happens when λ_2 is close but not equal to 1? In this section, we shall see quantitative bounds that further strengthens the relationship between λ_2 and the connectivity of a graph. Informally we shall see that

- If λ_2 is small then there is no small cut of G (and as we saw in last section a random walk is rapidly mixing).
- If λ_2 is close to 1 then there is a sparse cut that we can find efficiently (decompose the graph into communities).

When talking about the connectivity of a graph, we wish to understand how much we need to “change” the graph so as to partition the graph into two pieces. One way to find such a partitioning is to simply calculate a min cut. As we have seen, this can be done in polynomial time, but simply minimizing the number of edges crossing the cut can also lead to silly solutions. For instance, it may be optimal to split off one vertex from everything else. Often a more reasonable approach would be to attempt to minimize the number of edges while simultaneously making sure each side of the partition has a lot of edges. In the d -regular case that we study, we simply want to promote more balanced cuts. To this end, we define the conductance of a cut and a graph.

Definition 8 Let $G = (V, E)$ be a d -regular graph with n vertices. We define the conductance $h(S)$ of a cut $S \subseteq V$:

$$h(S) = \frac{|\delta(S)|}{d \cdot \min\{|S|, |V \setminus S|\}},$$

where $\delta(S)$ denotes the set of edges crossing the cut S .

We also define the conductance $h(G)$ of the graph G :

$$h(G) = \min_{S \subseteq V} h(S)$$

Notice that the conductance $h(S)$ of a cut roughly measures the fraction of edges that leaves the set S . So if $h(S) = 0.1$ this would approximately mean that only 10 – 20% of the edges leave S and 80 – 90% stay inside S . Such a set S would indicate a very strong community in e.g. Facebook. Finding such sets efficiently is therefore intensely studied and spectral graph theory has turned out to be very successful. Indeed, the conductance of a graph is closely related to λ_2 :

Theorem 9 (Cheeger’s Inequalities)

$$\frac{1 - \lambda_2}{2} \leq h(G) \leq \sqrt{2(1 - \lambda_2)}$$

We remark that both bounds are tight. A tight example for the lower bound is the hypercube and a tight example for the upper bound is a cycle.

In this lecture, we first prove the “easy” direction that $\frac{1 - \lambda_2}{2} \leq h(G)$ and if time we then prove the other “harder” direction. Before giving the proof of the lower bound, we would like to emphasize that the upper bound is in fact given by a very fast algorithm, called spectral partitioning:

Algorithm Spectral-Partitioning:

1. Input: graph $G = (V, E)$ and eigenvector $v_2 \in \mathbb{R}^n$ corresponding to second largest eigenvalue λ_2 of the normalized adjacency matrix M of G .
2. Sort the vertices of V in non-decreasing order of values in v_2 , that is let $V = \{1, \dots, n\}$ where $v_2(1) \leq v_2(2) \leq \dots \leq v_2(n)$.
3. Let $i \in \{1, \dots, n - 1\}$ be such that $h = (\{1, \dots, i\})$ is minimal.
4. Output $S = \{1, \dots, i\}$.

The proof of the upper bound given by Cheeger’s inequalities shows that the returned set S satisfies $h(S) \leq \sqrt{2(1 - \lambda_2)}$. The intuition is as follows. If $\lambda_2 = 1$, then we proved the last time that v_2 will have no edges between vertices with different values in v_2 . Hence, the above algorithm will output a perfect cut (cutting no edges) in this case. Now in the case, when λ_2 is close but not equal to 1 the algorithm will still try all cuts with the intuition that nearby vertices have a similar value in v_2 and take the best. Apart from the excellent guarantee of the above algorithm, it is very fast. It runs in $O(|V| \log |V| + |E|)$ time given the eigenvector v_2 . And if computing v_2 is too costly, one can also see that an approximate eigenvector will be fine.

Let us now return to the proof that $\frac{1 - \lambda_2}{2} \leq h(G)$. For this, it will be convenient to consider the eigenvalues of the normalized adjacency matrix M as solutions to optimization problems.

Eigenvalues as solutions to optimization problems It will be convenient to consider an alternative way to define the eigenvalues of M . Namely as the solution to the problem of maximizing the Rayleigh quotient

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

Lemma 10 $\lambda_1 = \max_{x \in \mathbb{R}^n} \frac{x^T M x}{x^T x}$.

Proof First, let's prove $\lambda_1 \leq \max_{x \in \mathbb{R}^n} \frac{x^T M x}{x^T x}$. Indeed, $\frac{v_1^T M v_1}{v_1^T v_1} = \frac{v_1^T \lambda_1 v_1}{v_1^T v_1} = \lambda_1 \frac{v_1^T v_1}{v_1^T v_1} = \lambda_1$

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

$$(v_1, v_2, \dots, v_n) \text{ is a basis, } \exists(\alpha_1, \alpha_2, \dots, \alpha_n) : y = \sum_{i=1}^n \alpha_i v_i. \text{ Then } \frac{y^T M y}{y^T y} = \frac{\sum_{i=1}^n \alpha_i^2 \lambda_i}{\sum_{i=1}^n \alpha_i^2} \leq \lambda_1 \frac{\sum_{i=1}^n \alpha_i^2}{\sum_{i=1}^n \alpha_i^2} = \lambda_1. \blacksquare$$

Lemma 11 $\lambda_2 = \max_{x \in \mathbb{R}^n : x \perp v_1} \frac{x^T M x}{x^T x}$, where v_1 is the eigenvector corresponding to λ_1 .

Proof Exercise (very similar) to the last lemma. \blacksquare

Proving the “easy direction” of Cheeger’s inequalities. By Lemma 11, we have

$$\begin{aligned} 1 - \lambda_2 &= 1 - \max_{x \in \mathbb{R}^n : x \perp v_1} \frac{x^T M x}{x^T x} \\ &= \min_{x \in \mathbb{R}^n : x \perp v_1} \left(1 - \frac{x^T M x}{x^T x} \right) \\ &= \min_{x \in \mathbb{R}^n : x \perp v_1} \frac{x^T (I - M) x}{x^T x} \end{aligned}$$

The matrix $L := I - M$ is called the normalized Laplacian matrix (recall last week’s exercises) and we have the identity

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

We have thus

$$1 - \lambda_2 = \min_{x \in \mathbb{R}^n : x \perp v_1} \frac{\sum_{\{i,j\} \in E} (x(i) - x(j))^2}{d \cdot x^T x}.$$

This looks like a continuous relaxation of the cut problem and it is indeed the intuition behind both Cheeger’s inequalities. Now for the lower bound, consider a cut S . We assume without loss of generality that $|S| \leq |V|/2$ since we otherwise could consider $|V \setminus S|$. Now define $y \in \mathbb{R}^n$:

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

It is an easy exercise to see that y is orthogonal to the all ones vector v_1 . That is $y \perp v_1$. Hence,

$$1 - \lambda_2 = \min_{x \in \mathbb{R}^n : x \perp v_1} \frac{\sum_{\{i,j\} \in E} (x(i) - x(j))^2}{d \cdot x^T x} \leq \frac{\sum_{\{i,j\} \in E} (y(i) - y(j))^2}{d \cdot y^T y}$$

Notice that the numerator equals $|\delta(S)|$:

$$\sum_{\{i,j\} \in E} (y(i) - y(j))^2 = \sum_{\{i,j\} \in E} (1 - |S|/|V| + |S|/|V|)^2 = |\delta(S)|.$$

And the denominator:

$$y^T y = \sum_{i \in V} y(i)^2 = |S| (1 - |S|/|V|)^2 + (|V| - |S|) (|S|/|V|)^2 = \frac{|S||V \setminus S|}{|V|} \geq |S|/2,$$

where the last inequality holds because $|V \setminus S| \geq |V|/2$ by assumption. Combining the above,

$$\frac{1 - \lambda_2}{2} \leq \frac{1}{2} \frac{\sum_{\{i,j\} \in E} (y(i) - y(j))^2}{d \cdot y^T y} \leq \frac{1}{2} \frac{|\delta(S)|}{d \cdot |S|/2} = \frac{|\delta(S)|}{d \cdot |S|} = h(S).$$

It follows that the conductance of every cut is at least $(1 - \lambda_2)/2$ and so $(1 - \lambda_2)/2 \leq h(G)$ as required.

2.1 Proof of the hard direction

Let v_2 be the second eigenvector. Then our goal is to prove that

$$h(G) \leq \sqrt{2(1 - \lambda_2)} \text{ where } 1 - \lambda_2 = \frac{\sum_{\{i,j\} \in E} (v_2(i) - v_2(j))^2}{d \cdot v_2^T v_2}.$$

We do so in two steps. First we show that we can extract a vector x from v_2 that has support at most $n/2$ and $\frac{\sum_{\{i,j\} \in E} (x(i) - x(j))^2}{d \cdot x^T x} \leq \frac{\sum_{\{i,j\} \in E} (v_2(i) - v_2(j))^2}{d \cdot v_2^T v_2}$. Second, we use x to extract a cut S such that $h(S) \leq \sqrt{2(1 - \lambda_2)}$.

Lemma 12 *There is a non-negative vector $x \in \mathbb{R}^n$ of support at most $n/2$ such that*

$$\frac{\sum_{\{i,j\} \in E} (x(i) - x(j))^2}{d \cdot x^T x} \leq \frac{\sum_{\{i,j\} \in E} (v_2(i) - v_2(j))^2}{d \cdot v_2^T v_2}.$$

Proof Consider what happens to expression if we shift v_2 by the unit vector times a constant c :

$$\frac{\sum_{\{i,j\} \in E} (v_2(i) + c - v_2(j) - c)^2}{d \cdot (v_2 + \mathbf{1}c)^T (v_2 + \mathbf{1}c)}$$

The numerator does not change and the denominator only increases using that $v_2 \perp \mathbf{1}$. Hence, we can obtain a new vector $y = v_2 + v_1$ where y has equal number of positive entries as negative entries and

$$\frac{\sum_{\{i,j\} \in E} (y(i) - y(j))^2}{d \cdot y^T y} \leq \frac{\sum_{\{i,j\} \in E} (v_2(i) - v_2(j))^2}{d \cdot v_2^T v_2}.$$

Now define a and b by $a(i) = \max(0, y(i))$ and $b(i) = \min(0, y(i))$. Then

$$\begin{aligned} \frac{\sum_{\{i,j\} \in E} (y(i) - y(j))^2}{d \cdot y^T y} &= \frac{\sum_{\{i,j\} \in E} (a(i) - a(j) + b(i) - b(j))^2}{d \cdot (a^T a + b^T b)} \\ &\geq \frac{\sum_{\{i,j\} \in E} (a(i) - a(j))^2 + \sum_{\{i,j\} \in E} (b(i) - b(j))^2}{d \cdot (a^T a + b^T b)}. \end{aligned}$$

The statement now follows because $\min A/B, C/D \leq (A + C)/(B + D)$ for positive numbers. So x in the statement can either be selected to be a or $-b$. ■

From now on, we only need to round a nonnegative vector x to a set of small conductance.

Lemma 13 *For any non-negative vector x of support at most $n/2$, there is a set S in the support of x such that*

$$h(S)^2 \leq 2 \frac{\sum_{\{i,j\} \in E} (x(i) - x(j))^2}{d \cdot x^T x}.$$

Before proving this lemma, we prove an easier statement. We show that if we drop all squares then we get an exact characterization of the conductance.

Lemma 14 *For any non-negative vector x of support at most $n/2$, there is a set S in the support of x such that*

$$h(S) \leq \frac{\sum_{\{i,j\} \in E} |x(i) - x(j)|}{d \cdot \sum_{i \in V} x(i)}.$$

Proof Since the right hand side is homogeneous in x we may assume that $\max_{i \in V} x(i) \leq 1$. Let $t \sim [0, 1]$ be chosen uniformly at random, and let $S_t := \{i : x(i) > t\}$. Then by linearity of expectation,

$$\mathbb{E}[|S_t|] = \sum_{i \in V} \Pr[i \in S_t] = \sum_{i \in V} x(i).$$

Also for any edge (i, j) the probability that this edge is cut is exactly $|x(i) - x(j)|$. Therefore,

$$\mathbb{E}[E(S_t, \bar{S}_t)] = \sum_{\{i,j\} \in E} \Pr[(i, j) \text{ is cut}] = \sum_{\{i,j\} \in E} |x(i) - x(j)|.$$

Therefore,

$$\frac{\mathbb{E}[E(S_t, \bar{S}_t)]}{d \mathbb{E}[|S_t|]} = \frac{\sum_{\{i,j\} \in E} |x(i) - x(j)|}{\sum_{i \in V} x(i)}.$$

Note that in general if $\frac{\mathbb{E}[A]}{\mathbb{E}[B]} \leq c$ it may be that $A/B > c$ with probability 1. For example, if $A = B = -1$ with probability 1/2 and $A = 1, B = 2$ with probability 1/2. Then $\mathbb{E}[A] = \mathbb{E}[A]/\mathbb{E}[B] = 0$ but $A/B > 1/2$ with probability 1.

However, if A, B are non-negative (with probability 1) then $\mathbb{E}[A]/\mathbb{E}[B] \leq c$ implies that $A/B \leq c$ with a positive probability. This follows from the following simple inequality. For any set of non-negative numbers $a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_m \geq 0$ we have

$$\min_{1 \leq i \leq m} \frac{a_i}{b_i} \leq \frac{a_1 + \dots + a_m}{b_1 + \dots + b_m} \leq \max_{1 \leq i \leq m} \frac{a_i}{b_i}.$$

Therefore there is a set S_t that satisfies the conclusion of the lemma. Note that $|S_t| \leq n/2$ since x has support at most $n/2$. ■

By the above lemma, to prove Lemma 13, all we need to do is to construct a vector y such that

$$\left(\frac{\sum_{\{i,j\} \in E} |y(i) - y(j)|}{d \cdot \sum_{i \in V} y(i)} \right)^2 \leq 2 \frac{\sum_{\{i,j\} \in E} (y(i) - y(j))^2}{d \cdot y^T y}. \quad (1)$$

Proof (proof of Lemma 13) Define $y(i) = x(i)^2$. To show (1) we use Cauchy-Schwarz inequality:

$$\begin{aligned}
\frac{\sum_{\{i,j\} \in E} |y(i) - y(j)|}{d \cdot \sum_{i \in V} y(i)} &= \frac{\sum_{\{i,j\} \in E} |x(i)^2 - x(j)^2|}{d \cdot \sum_{i \in V} x(i)^2} \\
&= \frac{\sum_{\{i,j\} \in E} |x(i) - x(j)| |x(i) + x(j)|}{d \cdot \sum_{i \in V} x(i)^2} \\
&\leq \frac{\sqrt{\sum_{\{i,j\} \in E} |x(i) - x(j)|^2} \sqrt{\sum_{\{i,j\} \in E} |x(i) + x(j)|^2}}{d \cdot \sum_{i \in V} x(i)^2} \\
&\leq \sqrt{2 \frac{\sum_{\{i,j\} \in E} |x(i) - x(j)|^2}{d \cdot \sum_{i \in V} x(i)^2}},
\end{aligned}$$

where for the last inequality we used that $(a + b)^2 \leq 2a^2 + 2b^2$ and so

$$\sum_{\{i,j\} \in E} |x(i) + x(j)|^2 \leq \sum_{\{i,j\} \in E} (2x(i)^2 + 2x(j)^2) = \sum_{i \in V} 2dx(i)^2.$$

■