Topics in Theoretical Computer Science March 11, 2014

Lecture 4

Lecturer: Ola Svensson *Scribes: Pierre-Alain Dupont, Cyprien Mangin*

1 Introduction

In the last lecture, we discussed bipartite expanders and their applications (error correcting codes, saving randomness) from a graph-theoretic point of view. In this lecture, we will approach expanders in a different way, through the graph's underlying adjacency matrix. This approach will prove advantageous in many applications.

2 Algebraic point of view of expanders

Let us remember the combinatorial definition of expanders:

Definition 1 *(combinatorial point of view)*
An n-vertex, d-regular graph $G = (V, E)$ is an ϵ-edge expander if $\forall S \subset V : |S| \leq n/2, |E(S, \overline{S})| \geq \epsilon d|S|$, where $E(S, \overline{S}) = \{(u, v) \in E : u \in S, v \in \overline{S}\}$.

There is an alternative definition for expander graphs, based on the algebraic properties of their random walk matrix(to be defined later):

Definition 2 *(algebraic point of view)*
An n-vertex, d-regular graph G is an ϵ-edge expander if $\lambda_2 \leq 1 - \epsilon$ where λ_2 is the second largest eigenvalue of the random walk matrix of G.

In order to be able to analyze λ_2, we need to know a bit more about what are the eigenvalues of an adjacency matrix. The following section will therefore focus on common spectral graph theory results before we can go back to expanders.

3 Spectral graph theory

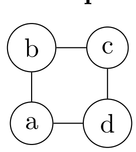
3.1 Basics

Let us give some basic definitions:

Definition 3 *The adjacency matrix A of graph G is a matrix such that $A_{ij} = 1$ if and only if $(i, j) \in E$.*

Definition 4 *The normalized adjacency matrix(or random walk matrix) M of a d-regular graph is equal to $\frac{1}{d}A$, with A being the adjacency matrix.*

Example:



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

Now, M has some special structure, which we will exploit:

Observation 5 *M is a real symmetric matrix.*

Observation 6 *M is a doubly stochastic matrix, i.e., each of its rows sums to 1 and each of its columns sums to 1.*

The following is a fact derived using standard Linear Algebra:

Fact 7 *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 is an eigenvector for λ_1 of length 1, then there exist $(v_2, v_3, ...v_n)$ such that v_i is an eigenvector of λ_i and $(v_1, v_2, ..., v_n)$ are orthogonal.*
   This means that no matter how the first eigenvector v_1 is chosen, we can always find an orthonormal basis.

Example: (using the same 4-cycle)

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

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

Observation 8 *Consider $x \in \mathbb{R}^n$ which assigns a value $x(v)$ to each vertex $v \in V$ and let $y = Mx$. In general, $y(v) = \sum_{(u,v)\in E} \frac{x(u)}{d}$, which is the average value according to x of v's neighbours. This is a very good intuition to keep in mind when dealing with spectral graph theory.*

Lemma 9 *Let M be the normalized adjacency matrix of a d-regular graph G. Then:*

0. *$\lambda_1 = 1$.*

1. *$\lambda_2 = 1 \iff G$ is disconnected.*
   More generally, $|\{i|\lambda_i = 1\}|$ is the number of connected components in G.

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

Proof

0. 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 $v \in V$ such that $x(v)$ is maximized and $y = Mx$, we have $y(v) = \sum_{(u,v)\in E} \frac{x(u)}{d} \leq \sum_{(u,v)\in E} \frac{x(v)}{d} = x(v)$. Therefore, $\lambda_1 \leq 1$.

1. Not proved.

2. See exercise 1.

■

3.2 Eigenvalues as solutions to optimization problems

Next, we will see an alternative way to define the eigenvalues of a real symmetric matrix M, as the solution to the problem of maximizing the Rayleigh quotient

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

Lemma 10 *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 11 *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 Similar to the last lemma. ■

Note: Given a random walk matrix M of a d-regular n-vertex graph, if we pick $v_1 = \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} \\ ... \\ \frac{1}{\sqrt{n}} \end{pmatrix}$, $x \perp v_1 \iff \sum_{i=1}^{n} x_i = 0$.

4 Cheeger's Inequalities

Now that we have seen some basic properties about adjencency matrices and eigenvalues, we shall show that the second largest eigenvalue is very much related to the expansion properties of the graph. To do so, we will prove Cheeger's Inequalities, which provide bounds on the expansion of the graph.

But first of all, let's define the expansion of a graph. This is just a convenient way to write the combinatorial definition of an expander.

Definition 12 *Let $G = (V, E)$ be a d-regular graph with n vertices. We define the expansion $h(S)$ of a cut $(S, \bar{S})$:*

$$h(S) = \frac{E(S, \bar{S})}{d \cdot \min\{|S|, |\bar{S}|\}}$$

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

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

Theorem 13 (Cheeger's Inequalities)

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

Since being an ϵ-expander is equivalent to $\epsilon \leq h(G)$, these inequalities are a useful tool to prove that a given graph is (or not) an expander, through algebraic methods. Namely, the first inequality says that G is an expander if $\lambda_2 < 1$. These inequalities are a link between the algebraic definition of expanders and its combinatorial counterpart.

In this lecture, we will prove only the "easy" direction, that is, the first inequality:

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

To achieve this, we introduce the sparsity of a cut $(S, \bar{S})$:

Definition 14 *Let $G = (V, E)$ be a d-regular graph with n vertices. We define the sparsity $\phi(S)$ of a cut $(S, \bar{S})$:*

$$\phi(S) = \frac{E(S, \bar{S})}{\frac{d}{n}|S||\bar{S}|}$$

We also define the sparsity $\phi(G)$ of the graph G:

$$\phi(G) = \min_{S \subset V} \phi(S)$$

Note that for any S, either $|S| \geq \frac{n}{2}$ or $\left|\bar{S}\right| \geq \frac{n}{2}$. If, for instance, $\left|\bar{S}\right| \geq \frac{n}{2}$, then:

$$\frac{1}{n}|S|\left|\bar{S}\right| \geq \frac{1}{2}|S| \geq \frac{1}{2}\min\{|S|, \left|\bar{S}\right|\}$$

This implies that $\frac{\phi(S)}{2} \leq h(S)$ for any S. We shall therefore prove this direction of the inequality by proving the following claim.

Claim 15

$$1 - \lambda_2 \leq \phi(G)$$

Proof Let S be the set of vertices that minimizes $\phi(S)$ and x be the characteristic vector of S, defined by:

$$x_i = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}$$

The value $E(S, \bar{S})$ is the number of edges between S and its complementary. An edge (i, j) is between S and its complementary if and only if ($x_i = 1$ and $x_j = 0$) or ($x_i = 0$ and $x_j = 1$). Therefore, we can write:

$$E(S, \bar{S}) = \sum_{(i,j) \in E} |x_i - x_j| = \sum_{(i,j) \in E} (x_i - x_j)^2$$

The second equality comes from the fact that $|x_i - x_j|$ can only be 0 or 1. We can introduce the adjacency matrix in this expression; we just have to be careful because we will count each edge twice.

$$E(S, \bar{S}) = \frac{1}{2}\sum_{i,j} A_{ij}(x_i - x_j)^2 = \frac{d}{2}\sum_{i,j} M_{ij}(x_i - x_j)^2$$

Similarly, we can view the product $|S||\bar{S}|$ as the number of edges in the complete bipartite graph $S \cup \bar{S}$ and write:

$$\frac{d}{n}|S||\bar{S}| = \frac{d}{2n}\sum_{i,j}(x_i - x_j)^2$$

Since the difference $x_i - x_j$ does not change if we add a constant vector to x, let's define y to be $x + c$ where c is chosen so that $\sum_{i=1}^{n} y_i = 0$. The intention behind this choice is to make sure y is orthogonal to v_1.

Now recall the value we are interested in is:

$$\phi(S) = \frac{E(S,\bar{S})}{\frac{d}{n}|S||\bar{S}|}$$

We will simplify separately the denominator and the numerator.

The denominator

$$\begin{aligned}
\frac{d}{n}|S||\bar{S}| &= \frac{d}{2n}\sum_{i,j}(x_i - x_j)^2 \\
&= \frac{d}{2n}\sum_{i,j}(y_i - y_j)^2 \\
&= \frac{d}{2n}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} y_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} y_j^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j\right] \\
&= \frac{d}{2n}\left[\sum_{i=1}^{n} n y_i^2 + \sum_{j=1}^{n} n y_j^2 - 2\left(\sum_{i=1}^{n} y_i\right)^2\right]
\end{aligned}$$

We chose y specifically so that $\sum_{i=1}^{n} y_i = 0$, therefore:

$$\frac{d}{n}|S||\bar{S}| = d\sum_{i=1}^{n} y_i^2 = dy^T y$$

The numerator

$$\begin{aligned}
E(S,\bar{S}) &= \frac{d}{2}\sum_{i,j} M_{ij}(x_i - x_j)^2 \\
&= \frac{d}{2}\sum_{i,j} M_{ij}(y_i - y_j)^2 \\
&= \frac{d}{2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} y_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} y_j^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} y_i y_j\right]
\end{aligned}$$

Since M_{ij} is a doubly-stochastic matrix, we know that the sum of M_{ij} along a row or a column is equal to 1.

$$\begin{aligned}
E(S,\bar{S}) &= \frac{d}{2}\left[\sum_{i=1}^{n} y_i^2 + \sum_{j=1}^{n} y_j^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij} y_i y_j\right] \\
&= d\left(y^T y - y^T M y\right)
\end{aligned}$$

We finally get a nice expression:

$$\phi(S) = \frac{d\left(y^T y - y^T My\right)}{dy^T y} = 1 - \frac{y^T My}{y^T y}$$

Since we know that y is orthogonal to v_1, we have the following:

$$\frac{y^T My}{y^T y} \leq \max_{x \in \mathbb{R}^n : x \perp v_1} \frac{x^T Mx}{x^T x} = \lambda_2$$

We can now conclude:

$$\phi(G) = \phi(S) \geq 1 - \lambda_2$$

■

5 Exercises

The first exercise will show the relation between the smallest eigenvalue and the bipartition of a graph. The second exercise will prove that a random walk on a graph is rapidly mixing, if we can bound the absolute value of every eigenvalue but λ_1 away from 1.

Exercise 1 Given a connected d-regular graph $G = (V, E)$ with n vertices, show that it is bipartite if and only if $\lambda_n = -1$.

Solution First, let's assume that G is bipartite, $V = V_1 \cup V_2$. We consider the vector x defined by:

$$x_i = \begin{cases} 1 & \text{if } i \in V_1 \\ -1 & \text{if } i \in V_2 \end{cases}$$

Since every vertex in V_1 has its neighbours in V_2, and conversely, we have $Mx = -x$. Therefore, -1 is an eigenvalue for M and it has to be the smallest, since the absolute value of every eigenvalue of M is bounded by 1. Hence $\lambda_n = -1$.

We now assume $\lambda_n = -1$ and we consider an eigenvector x associated with λ_n. Let $x_i = D$ be the largest component of x in absolute value. Since $(Mx)_i = -x_i$ is the average of the value of the d neighbours of i, it means that those neighbours have to be associated with the value $-D$. The same goes for the neighbours of the neighbours of i which have to be associated with the value D. Since the graph is connected, we can extend the reasoning to every vertex.

We just proved that each component of x is either D or $-D$. Let $V_1 = \{i : x_i = D\}$ and $V_2 = \{i : x_i = -D\}$. We also proved that every neighbour of a vertex in V_1 has to be in V_2, and conversely. Therefore, G is bipartite with $V = V_1 \cup V_2$.

■

Exercise 2 Consider a d-regular graph G with $|\lambda_i| \leq 1 - \epsilon$ for all $i \geq 2$. Show that a random walk is rapidly mixing, i.e., that no matter from which vertex we start, after $O(\log n)$ steps we will be at any vertex with probability $\approx \frac{1}{n}$. More formally, show that if we let x be the vector that is equal to 1 on the vertex where the random walk starts, then

$$\left|\left| M^k x - \left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \right|\right|_1 \leq o\left(\frac{1}{n^2}\right)$$

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

What happens if it is a bipartite graph?

Solution Let's assume, for the sake of simplicity, that we start from the vertex 1. Let $(v_1, \dots, v_n)$, where v_i is an eigenvector for the eigenvalue λ_i, is an orthonormal basis. It means we can decompose x on it:

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

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

$$\left|\left|M^k x - \left(\frac{1}{n}, \dots, \frac{1}{n}\right)\right|\right|_1 = \left|\left|\sum_{i=2}^{n} \alpha_i \lambda_i^k v_i\right|\right|_1 \leq (1-\epsilon)^k \left|\left|\sum_{i=2}^{n} \alpha_i v_i\right|\right|_1$$

So if we take $k = \frac{c}{\epsilon} \log n$ and we write $A = ||\sum_{i=2}^{n} \alpha_i v_i||_1$, we have:

$$\left|\left|M^k x - \left(\frac{1}{n}, \dots, \frac{1}{n}\right)\right|\right|_1 \leq A(1-\epsilon)^{\frac{\log n}{\epsilon} c} \leq A \left(\frac{1}{e}\right)^{c \log n} = \frac{A}{n^c}$$

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

In the case of a bipartite graph, the problem is that the probabilities are very different depending on whether we did an even number of steps or not. To get back a random walk that will be rapidly mixing, one solution is to use a *lazy* random walk, that is the same random walk, but with a prior probability of $\frac{1}{2}$ to stay where we are.

■