CS-774 Spectral Graph Theory	Aug 03, 2023

Lecture 1: Random Walks, Expanders and Cheeger's Inequality

Lecturer: Akash Kumar — *Scribe: Akash Kumar*

1 What is a Random Walk?

A random walk is a very simple process that takes place on a graph $G = (V, E)$. To specify the process, you first pick an initial vertex from which the walk begins. Let $\boldsymbol{X}_0 = u$ denote this initial vertex. Fix $t \in \mathbb{N}$ and let the random vertex visited by the process at time $t-1$ be denoted as $\boldsymbol{X}_{t-1}$. You choose the next vertex $\boldsymbol{X}_t$ uniformly at random from among the neigbhors of the current vertex, $\boldsymbol{X}_{t-1}$. This is written as

$$\boldsymbol{X}_t \sim \textsc{Unif}(N(\boldsymbol{X}_{t-1})).$$

To develop a rich theory of random walks, it would be helpful to express the distribution from which $\boldsymbol{X}_t$ is sampled using tools from linear algebra. To this end, let us write $\boldsymbol{M} = \boldsymbol{D}^{-1}\boldsymbol{A}$ where $\boldsymbol{A}$ denotes the Adjacency Matrix of G and $\boldsymbol{D}$ is a diagonal matrix where $\boldsymbol{D}(i,i) = deg(i)$. We call the matrix $\boldsymbol{M}$ a Markov Matrix or a Random Walk Matrix. Consider a random walk that begins at vertex u. That is $\boldsymbol{X}_0 = \mathbf{1}_u$. Notice that the vertex $\boldsymbol{X}_1$ is a uniform sample from the distribution $\boldsymbol{p}_1^T = \mathbf{1}_u^T \cdot \boldsymbol{M}$. More generally, suppose you are given a distribution $\boldsymbol{p}_0$ supported over the vertices and the start vertex $\boldsymbol{X}_0 \sim \boldsymbol{p}_0$. In this case, you should verify that $\boldsymbol{X}_1 \sim \boldsymbol{p}_1$ where $\boldsymbol{p}_1^T = \boldsymbol{p}_0^T \cdot \boldsymbol{M}$. Indeed, if the start distribution is denoted $\boldsymbol{p}_0$ and the distribution after t steps is denoted $\boldsymbol{p}_t$, you indeed have $\boldsymbol{p}_t^T = \boldsymbol{p}_0^T \boldsymbol{M}^t$ and $\boldsymbol{X}_t \sim \boldsymbol{p}_t$ (as you should verify). You should think of the ability to perform random walks as a helpful primitive which allows you to explore a graph (just like DFS and BFS help you explore a graph). And just like DFS and BFS allow you to get valuable insights to attack various graph problems, you will see that the random walk primitive also leads to algorithmic insights on graphs. Before we proceed, it will be helpful to verify a few simple facts about random walks.

Claim 1.1. *Let $G = (V, E)$ be a graph and let $\boldsymbol{M} = \boldsymbol{M}(G)$ denote the associated random walk matrix. Let $\boldsymbol{p}$ denote an arbitrary distribution supported on the vertex set of G. Then $\boldsymbol{p}^T\boldsymbol{M}$ is also a distribution supported on vertices of G.* (just saying that it is a valid distribution)

Proof of Claim 1.1. Let $\boldsymbol{q}^T = \boldsymbol{p}^T\boldsymbol{M}$. To verify that $\boldsymbol{q}$ is a bonafide distribution, you need to check that all entries in the vector $\boldsymbol{q}$ are non-negative which is clear (think why). Additionally, you need to show that all the entries in $\boldsymbol{q}$ add up to 1. To do this, note that

$$\sum \boldsymbol{q}(i) = \boldsymbol{p}^T\boldsymbol{M}\mathbf{1} = 1.$$

□

Okay, so that checks out. Markov matrices map distributions to distributions. You want to understand how can you meaningully explore a graph by randomly crawling along its edges. To this end, one of the first things you would like to have is some meaningful notion of how to take limits of power iterations on $\boldsymbol{M}^t$ as $t \to \infty$. In some sense, you should anticipate a limit like this should exist (under some hopefully mild conditons). Indeed, consider how you would model diffusions in a porous media if you were a physicist. You typically model a porous object, say a wet sponge, as a finite graph with some edge relations between "nearby" points in the sponge. Physicists believe that many macrolevel physical properties of a wet sponge result from some interactions between nearby vertices in this graph. These interactions are modeled randomly. You imagine total amount of water held inside this sponge is 1 unit (and you think of this amount of water as inducing a probability distribution). Physical operations on this sponge result in some random walks on the graph you obtain for this sponge. And the limit theorems for these walks typically explain the phenomena you wish to understand. So, let us now make an effort to understand when can you prove a limiting result for power iterations on the random walk matrix.

1.1 Power Iterations on Random Walk Matrices

Let us start with a simple enough example to think about limits for powers of a square matrix. I have not even told you what such a limit might mean but let us ignore that for now. You will see, our explorations make sense even though I have not defined such a limit yet. So, let us consider the matrix $\boldsymbol{M} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$. What is $\boldsymbol{M}^t$? Well, $\boldsymbol{M}$ being a diagonal matrix, you immediately see

$$\boldsymbol{M}^t = \begin{pmatrix} 2^t & 0 \\ 0 & 3^t \end{pmatrix}.$$

Unfortunately, this means power iterations on $\boldsymbol{M}$ "blow-up" as $t \to \infty$. Thus, it seems that if your matrix has eigenvalues which exceed 1 in absolute value, you are in trouble and your power iterations might not converge. To fix this, it is perhaps helpful to restrict our attention to matrices all of whose eigenvalues are at most 1 in absolute value. Even before that you would like to have the assurance that your matrix $\boldsymbol{M}$ has eigenvalues all of which are real numbers. This is indeed true for a rich class of matrices, namely symmetric matrices as can be seen from the Spectral Theorem.

Theorem 1.2. *Let $\boldsymbol{M} \in Sym(\mathbb{R}^{n \times n})$. Then all eigenvalues of $\boldsymbol{M}$ are real and there exists an orthonormal basis of $\mathbb{R}^n$ which consists of the eigenvectors of $\boldsymbol{M}$. Equivalently, doing a SVD reveals $\boldsymbol{M} = \boldsymbol{U}\Sigma\boldsymbol{U}^T$ where $\boldsymbol{U} \in \mathbb{R}^{n \times n}$ is an orthonormal matrix.*

For a symmetric matrix $\boldsymbol{M}$ note that $\boldsymbol{M}^t = \boldsymbol{U}\Sigma^t\boldsymbol{U}$[1]. This confirms if the eigenvalues of $\boldsymbol{M}$ exceed 1 in absolute value, power iterations on $\boldsymbol{M}$ will not converge. Your intuition from physics of diffusions should suggest that random walk matrices better have all eigenvalues between -1 and 1. We will prove this shortly for the special case where random-walk matrix is symmetric but we note that it holds more generally. I state a simple fact below you should verify.

Fact 1.3. *Let $G = (V, E)$ be a graph where the random walk matrix $\boldsymbol{M} = M(G)$ is symmetric. Then all connected components of G are regular.*

This takes us to the graph class we will mostly think about in the rest of these lectures.

[1]This is kind of a cool thing to reflect upon. Powers of a symmetric matrix are symmetric. Can you see why this should hold without doing the SVD?

Let $G = (V, E)$ denote a connected graph all vertices in which have degree d.

Now, let us show that all eigenvalues of a connected d-regular graph lie in the interval $[-1, 1]$.

Lemma 1.4. *Let $G = (V, E)$ be a connected d-regular graph. Let $\boldsymbol{M} = M(G)$ denote the random-walk matrix of G. Then all eigenvalues of $\boldsymbol{M}$ lie in the interval $[-1, 1]$.*

The proof is deferred to §1.1.1. Before we prove this, let me assign two exercises out of the way. These execrcises demonstrate that both of these extreme eigenvalues − 1 and −1 − are achievable.

Fact 1.5. *Let $\boldsymbol{M}$ denote a Markov Matrix. Then 1 is one of its eigenvalues.*

Fact 1.6. *Let $G \cong K_2$ denote a two vertex graph which is just an edge. Let $\boldsymbol{M}$ denote the Markov Matrix for G. Then -1 is one of the eigenvalues of $\boldsymbol{M}$.*

We will need some more elaborate machinery to verify our compelling physical intuition which suggests Lemma 1.4 holds. This is done in the following section.

1.1.1 Enter The Laplacian

In some sense, you need to show that $\boldsymbol{M}$ is "trapped" between the identity matrix $\boldsymbol{I}$ and negative identity, $-\boldsymbol{I}$. One classic approach in linear algebra to these situations is to consider whether the matrix $\boldsymbol{I} - \boldsymbol{M}$ is *positive semi-definite*. The claim below asserts just that.

Claim 1.7. *The matrix $\boldsymbol{I} - \boldsymbol{M}$ is positive semi-definite.*

Proof of Claim 1.7. Write $\overline{\boldsymbol{L}} = \boldsymbol{I} - \boldsymbol{M} = \boldsymbol{I} - \boldsymbol{A}/d = \dfrac{d\boldsymbol{I} - \boldsymbol{A}}{d}$. Let $\boldsymbol{L} = d\boldsymbol{I} - \boldsymbol{A}$. Showing $\overline{\boldsymbol{L}}$ is psd is equivalent to showing $\boldsymbol{L}$ is psd. Since $\boldsymbol{A}$ is symmetric and $d\boldsymbol{I}$ is symmetric, the difference $\boldsymbol{L}$ is also symmetric. For an edge $e = (i, j) \in E(G)$, define a matrix $\boldsymbol{L}_e \in Sym(\mathbb{R}^{n \times n})$ where $\boldsymbol{L}_e(i, i) = -1 = \boldsymbol{L}_e(j, j)$, $\boldsymbol{L}_e(i, j) = 1 = \boldsymbol{L}_e(j, i)$ and $\boldsymbol{L}_e(a, b) = 0$ whenever $a \neq i$ or $b \neq j$. Note that $\boldsymbol{L} = \sum_{e \in E(G)} \boldsymbol{L}_e$ (think why). For any vector $\boldsymbol{x} \in \mathbb{R}^n$, we note $\boldsymbol{x}^T \boldsymbol{L}_e \boldsymbol{x} = (\boldsymbol{x}_i - \boldsymbol{x}_j)^2 \geq 0$. This means for any $\boldsymbol{x} \in \mathbb{R}^n$, we also have

$$\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} = \boldsymbol{x}^T \sum_{e=(i,j) \in E(G)} \boldsymbol{L}_e \boldsymbol{x} = \sum_{(i,j) \in E(G)} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2 \geq 0.$$

This means the matrix $\boldsymbol{L}$ is psd as desired. □

Claim 1.8 follows using arguments almost identical to the ones used for proving Claim 1.7. The proof is left as an exercise.

Claim 1.8. *The matrix $\boldsymbol{I} + \boldsymbol{M}$ is positive semi-definite.*

The matrix $\boldsymbol{L}$ is one of the most important matrices you can associate with a graph. It is called the *Laplacian* of a graph. It is the "analog" of a second derivative operation over graphs. This is because the quadratic forms, $\boldsymbol{x}^T \boldsymbol{L}\boldsymbol{x}$ equals the sum of squared differences between the end-points of all the edges in G. Indeed, most of this series will explore how Laplacian (and thus the Markov Matrix) offers insights about a wide variety of problems in spectral graph theory. The following fact is left as an exercise.

Fact 1.9. *The maximum eigenvalue of the matrix $\boldsymbol{A}$ is d and the minimum eigenvalue of $\boldsymbol{A}$ is at least $-d$.*

With this fact in our hand, we can finish Lemma 1.4.

Proof of Lemma 1.4. Recall $\boldsymbol{M} = \boldsymbol{A}/d$. Proof over. □

Note that this means all eigenvalues of $\overline{\boldsymbol{L}}$ lie between 0 and 2 (think why).

1.2 The stationary distribution

In this section, we will talk about the limiting behavior of a long power iteration on the random walk matrix of a connected d-regular graph. To kick things off, let us start simple and consider what happens when you do random walk on a graph which is just a single edge. On even steps, you occupy the start vertex. On odd steps, you occupy the other. So the walk keeps flip-flopping and there is no stationary distribution that is ever reached. Indeed, this behavior is what you would expect on any bipartite graph. Looks like we are off to a bad start. What went wrong? If you dig into the math for the case where the graph is a single edge, you note that the Markov Matrix is given as

$$\boldsymbol{M} = \boldsymbol{M}_{\text{EDGE}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

One eigenvalue of this matrix is 1 and since the trace is 0, the other eigenvalue must equal -1. Indeed, you will show in one of your exercises that Markov Matrices for a (regular) bipartite graph always have -1 as an eigenvalue. And this is precisely why the power iterations do not converge. Indeed, for $t \in \mathbb{N}$, $\boldsymbol{M}^t = \boldsymbol{U}\Sigma^t\boldsymbol{U} = 1^t \cdot \boldsymbol{u}_1\boldsymbol{u}_1^T + (-1)^t\boldsymbol{u}_2\boldsymbol{u}_2^T$ (where $\boldsymbol{u}_1, \boldsymbol{u}_2$ are the eigenvectors corresponding to eigenvalues $1, -1$ respectievely). For odd t, this equals $\boldsymbol{u}_1\boldsymbol{u}_1^T - \boldsymbol{u}_2\boldsymbol{u}_2^T$ and for even t, it is the sum of these two outer products. This explains why the power iteration does not converge. It would be convenient if we could make all the eigenvalues other than the leading eigenvalue strictly smaller than 1 in absolute value. Luckily, you can do this very easily. Recall from Fact 1.5, Fact 1.6 all eigenvalues of $\boldsymbol{M}$ are trapped between -1 and 1. Note that adding an identity, $\boldsymbol{I}$ to $\boldsymbol{M}$ shifts all eigenvalues by $+1$. So, $\boldsymbol{I} + \boldsymbol{M}$ is psd. But this is not a Markov Matrix, the row sums all equal 2 and the leading eigenvalue is 2 as well. To fix this, we instead consider the matrix $\boldsymbol{W} = \dfrac{\boldsymbol{I} + \boldsymbol{M}}{2}$.

Fact 1.10. *Verify that all eigenvalues of $\boldsymbol{W}$ lie between 0 and 1. In particular $\boldsymbol{W}$ is psd. Additionally $\boldsymbol{W}$ is also a Markov Matrix.*

I claim that this matrix, $\boldsymbol{W}$, is the random walk matrix for a graph obtained by doing a simple modification to the underlying graph $\boldsymbol{M} = \boldsymbol{M}(G)$ – namely it just adds d loops to each vertex.

From now on, I will abuse notation and I will use G to refer to this graph with d loops at each vertex.

In what follows, we will need the following theorem as well.

Theorem 1.11. *Let G be a connected d-regular graph. Let $\boldsymbol{M} = \boldsymbol{M}(G)$ denote the Markov Matrix for G. Let $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n \geq -1$ denote the eigenvalues of $\boldsymbol{M}$. Then $\lambda_1 - \lambda_2 \geq 1/n^2$. In other words, λ_2 is strictly smaller than λ_1.*

We will prove this theorem in a later section. We proceed assuming this claim and finally we present the main result of this section. It asserts that indeed, there is some limiting behavior which emerges on doing power iteration on the matrix $\boldsymbol{W}$ above.

regular-connected

Theorem 1.12. *Let $G = (V, E)$ denote a connected d-regular graph. Let G' denote the graph obtained by adding d loops at every vertex. Let $\boldsymbol{W}$ denote the random walk matrix on G' (recall, $\boldsymbol{W} = \frac{\boldsymbol{I}+\boldsymbol{M}}{2}$ where $\boldsymbol{M}$ is the random walk matrix on G). It holds that*

$$\text{For all distributions } \boldsymbol{p} \text{ on vertices, } \lim_{t\to\infty} \boldsymbol{p}^T \boldsymbol{W}^t = \underbrace{\begin{bmatrix} 1/n & 1/n & 1/n & \cdots & 1/n \end{bmatrix}}_{n \text{ entries}}.$$

Proof of Theorem 1.12. Fix an arbitrary vertex $u \in V$. It suffices to show the above theorem for distributions of the form $\boldsymbol{p} = \mathbf{1}_u$ (why?). For ease of notation, write $\mathbf{1} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$. Recall that $\boldsymbol{W}$ being symmetric, it is as good if we show $\lim_{t\to\infty} \boldsymbol{W}^t \mathbf{1}_u = \mathbf{1}/n$. Indeed, we will show this convergence in the Euclidean-norm, namely

$$\lim_{t\to\infty} \|\boldsymbol{W}^t \mathbf{1}_u - \mathbf{1}/n\|_2 \to 0.$$

$p(\text{general}) = \frac{\sum c_u 1_u}{\sum c_u}$, if holds for 1_u, p follows.

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_n$ denote the eigenvectors of $\boldsymbol{W}$ which correspond to the eigenvalues

$$1 = \lambda_1 \overset{\text{Theorem 1.11}}{>} \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n \geq 0.^2$$ (since W is pos.d)

we assume $\varepsilon > 0$, will be shown by thm 1.11

Write $\lambda_1 - \lambda_2 = 1 - \lambda_2 = \varepsilon$ and note that for any vector $\boldsymbol{v} \perp \mathbf{1}$, we have $\|\boldsymbol{W}\boldsymbol{v}\|_2 \leq (1-\varepsilon)\|\boldsymbol{v}\|_2$. (since v = lc of other eigenvecs) Next up, note that $\boldsymbol{v}_1 = \mathbf{1}/\sqrt{n}$ and $c_1 = \boldsymbol{v}_1 \cdot \mathbf{1}_u = 1/sqrtn$ which means $c_1 \boldsymbol{v}_1 = \mathbf{1}/n$. Finally, on expressing $\mathbf{1}_u$ as a linear combination of the vectors in the eigenbasis, we get $\mathbf{1}_u = \sum c_i \boldsymbol{v}_i$ where $\sum c_i^2 = 1$. (orthonormal eigenbasis)

Thus,

[2] Since $\boldsymbol{W}$ is psd by Fact 1.10

Note: $W \cdot \frac{\mathbb{1}}{n} = \frac{\mathbb{1}}{n}$. So, limit dist is "stationary" (equilibrium)

$$
\begin{aligned}
\|\boldsymbol{W}^t \mathbf{1}_u - \mathbf{1}/n\|_2 &= \|\boldsymbol{W}^t \mathbf{1}_u - c_1 \boldsymbol{v}_1\|_2 \\
&= \|\boldsymbol{W}^t \sum_{i=2}^{i=n} c_i \boldsymbol{v}_i\| \\
&= \|\sum_{i\geq 2} \lambda_i^t c_i \boldsymbol{v}_i\|_2 \\
&\leq \sum_{i\geq 2} \|\lambda_i^t \boldsymbol{v}_i\|_2 && \text{By triangle inequality and noting for all } i, |c_i| \leq 1 \\
&\leq \sum_{i\geq 2} (1-\varepsilon)^t \\
&\leq (1-\varepsilon)^t \cdot n
\end{aligned}
$$

On taking limit as $t \to \infty$, one notes this converges to zero which finishes the proof. □

Remark 1.13. *The proof above actually shows a bit more. You do not need to wait too long for walks to reach the stationary distribution. You can in fact good convergence time bounds. Namely, in* $t = O\left(\frac{\log n}{\varepsilon}\right)$ *steps the walk almost reaches the limiting distribution. It is also called the* stationary distribution *for* $\boldsymbol{W}$ *(or for random walks on G').*

The existence of this limiting distribution leads to a curiousity. Let $t = C \log n/\varepsilon$ for some large constant $C \gg 1$. You see, if a walk is run for t or more steps, the distribution stays more or less invariant. We make a few remarks about the stationary distribution.

non-regular connected

Remark 1.14. *Let $G = (V, E)$ be a connected graph (which is not necessarily regular). Let $\boldsymbol{M} = \boldsymbol{M}(G)$ denote the Markov Matrix for random walks on G. We call the matrix $\boldsymbol{W} = \frac{\boldsymbol{I}+\boldsymbol{M}}{2}$ the matrix for* lazy random walks *on G.*

- We state without proof that such matrices always admit a unique stationary distribution which we denote as $\boldsymbol{\pi}$.
- Let $\boldsymbol{p}_0 = \boldsymbol{\pi}$ denote the current distribution. By definition of stationary distribution, we know $\boldsymbol{p}_1^T = \boldsymbol{p}_0^T \boldsymbol{W} = \boldsymbol{\pi}^T$. Note that this means for any cut $(S, \overline{S})$ in the graph, the probability mass that goes from S to $\overline{S}$ is the same as the probability mass going from $\overline{S}$ to S. This is called a *balance condition* which says mass flowing across either sides of any cut is balanced. Curiously enough, lazy random walks on graphs obey an even more intricate balanced condition, the so called *detailed balance condition* which states the following: take any edge $(u, v) \in E(G)$ and let m_{uv} denote the mass that flows from u to v and m_{vu} denotes the mass that flows from v to u. Detailed balance condition asserts that $m_{uv} = \frac{\boldsymbol{\pi}(u)}{deg(u)} = \frac{\boldsymbol{\pi}(v)}{deg(v)} = m_{vu}$.

since prob. dist doesn't change (balance condⁿ)

- According to item 1, the stationary distribution is unique for $\boldsymbol{W}$ (even in the irreguar case). Also, according to item 2, any distribution which satisfies detailed balance condition is *a stationary distribution*. You can readily verify that the distribution $\boldsymbol{\pi}$ which has $\boldsymbol{\pi}(u) = deg(u)/2|E|$ satisfies the detailed balance condition for $\boldsymbol{W}$ and therefore, it is *the unique stationary distribution* for lazy random walks on G.

x y

$m(x \to y) = \frac{\pi(x)}{deg(x)}$
$m(y \to x) = \frac{\pi(y)}{deg(y)}$
for stationary dist

balance at each edge. (detailed balance condⁿ)

#ques: How to show unique stationary dist exist?

#ques: How to show detailed balance condition for lazy random walks?

Finally, we come to an important definition.

Definition 1.15. *Let $G = (V, E)$ be a connected (not necessarily regular) graph. Let $\boldsymbol{M} = \boldsymbol{M}(G)$ denote the Markov Matrix for G. Let $\boldsymbol{W} = \frac{\boldsymbol{I}+\boldsymbol{M}}{2}$ denote the lazy random walk matrix for G. Let $\boldsymbol{\pi}$ denote the stationary distribution for lazy random walks on G. The mixing time for lazy random walks on G (or power iterations on $\boldsymbol{W}$) is:*

#ques

$$\tau_{mix} = \min\{t \in \mathbb{N} \colon \|\mathbf{1}_u^T \boldsymbol{W} - \boldsymbol{\pi}^T\|_1 \leq 1/4 \quad \forall u \in V\}$$

not getting this, ask sir.

Remark 1.16. *The proof of Theorem 1.12 essentially shows that the mixing time of lazy random walks on a connected d-regular graph is $O(\log n/(1-\lambda_2))$.*

Now, we are ready to take a deeper dive into the main content of this lecture series. This is the content of the next section. It begins by trying to unravel a mystery: what is the connection between the graph being connected and $\lambda_1(\boldsymbol{W}) - \lambda_2(\boldsymbol{W})$ being strictly positive (see Theorem 1.11).

2 Flesh, Blood & Bone: Basics of Spectral Graph Theory

We begin this section with the following fundamental result. It shows that the eigenvalues of the Laplacian do sense some combinatorial properties of the given graph.

Theorem 2.1. *Let $G = (V, E)$ be a d-regular graph. Let $\overline{\boldsymbol{L}} = \boldsymbol{I} - \boldsymbol{A}/d$ denote the Laplacian of this graph with eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$. Then $\lambda_k = 0$ if and only if G has at least k connected components.*

take "basis" (modified vecs) over conn. comp

Proof of Theorem 2.1. Suppose G has exactly k connected components. We will show that $\lambda_k = 0$. To this end, let us write $\boldsymbol{L} = d\boldsymbol{I} - \boldsymbol{A} = d\overline{\boldsymbol{L}}$. Denote the components as $C_1, C_2, \ldots C_k$. For $i \in [k]$, note that $\mathbf{1}_i \in \mathbb{R}^n$ is an eigenvector of the Laplacian where $\mathbf{1}_i(u) = 1$ if $u \in C_i$ and 0 otherwise. All $\{\mathbf{1}_i\}$'s have disjoint supports and are therefore independent. Further, all of them are eigenvectors with eigenvalue 0.

dim(subsp) = dim(vecs)

For the other direction, suppose $\lambda_k = 0$. We will show the graph has at least k connected components. Let $S = \{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x} = 0\}$. For each $\boldsymbol{x} \in S$, $\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x} = \sum_{(u,v)\in E} (\boldsymbol{x}_u - \boldsymbol{x}_v)^2 = 0$ which means $\boldsymbol{x}$ is constant over each connected component. By variational characterization of eigenvalues, we know $\lambda_k = 0$ implies that there is a k-dimensional subspace S' all vectors in which satisfy $\boldsymbol{x}^T\boldsymbol{L}\boldsymbol{x} = 0$. This means $S' \subseteq S$ and thus $dim(S) \geq k$. This means G has at least k components as desired. ◻

We now show a "robust" version of this result. The version above states that $\lambda_2 = 0$ implies the graph is disconneted. The robust version asserts that λ_2 being close to zero implies that the graph is almost disconnected in a sense we will make precise now. We first make an important definition.

Definition 2.2. *Let $G = (V, E)$ be a d-regular graph. For a set non-empty S of vertices ($S \neq V$) we define $\varphi(S) = \frac{|E(S,\overline{S})|}{d|S|}$. This quantity can be given a probabilistic interpretation: it measures the probability that a random edge (u, v) incident on a random $u \in S$ has the other end point $v \notin S$.*

Definition 2.3. *For a d-regular graph $G = (V, E)$, we define $\varphi(G) = \min_{0<|S|\leq n/2} \varphi(S)$.* Expansion/conductance / sparsity

• small $\varphi(G)$ facilitates divide & conquer, since, cross-edges can be nuked

Note : $\frac{x^T L x}{x^T x} = R_L(x) = \frac{\sum_{(u,v)\in E} (x_u - x_v)^2}{\sum_{u \in [n]} x_u^2}$

#ques: Calculation of expansion of hypercube?

$\lambda_k = \min_{\substack{V \subseteq \mathbb{R}^n \\ \dim V = k}} \max_{\substack{v \in V \\ \|v\|=1}} v^T M v$

$(\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n)$

Remark 2.4. Let us make sure we understand the point of the above definition. Saying $\varphi(G)$ is large is basically the same thing as saying your graph is well connected. At this point, you might wonder why on earth would one think that any 3-regular graph, expander or not, is well connected. After all, we can disconnect vertex #1 from everything else by just deleting all the three edges incident on this vertex. To understand this new notion of connectivity, note take a graph with $\varphi(G) \ge \alpha$ and consider some small subset of vertices, S. Note that to disconnect S from rest of the graph you must delete at least $\alpha \cdot d|S|$ edges incident on S – which is large as a fraction of the total number of edges S is incident with. And this comment applies to "all small sets". Another comment to make concerns the expansion of the bigger side. To this end, suppose $|S| \le n/2 \le |\overline{S}|$. Clearly, $\varphi(\overline{S}) \le \varphi(S)$. To get a meaningfully rich theory we restrict ourselves to thinking about expansion of only the smaller side.

connected, m-edges, if S has one outedge, $|S| \le m$

so, $\frac{1}{|S|} \ge \frac{1}{m}$

Example 2.5. Note that a graph has expansion strictly greater than zero iff it is connected. Any graph with m edges has expansion at least $1/m$. In particular, a connected d-regular graph has expansion at least $2/dn$. You might wonder, do expanders exist? Given the title of this lecture, they clearly should (and do) exist but then you might ask can we construct them? This is a very deep question with a remarkably rich history parts of which we might see towards the end of this class.

For connected d-regular graph, $|edge| = \frac{dn}{2}$, hence the bound.

As an exercise, you should try computing the expansion of a cycle and a hypercube by hand. Now, let me make good on my promise from before and explain in what sense λ_2 being small leads to a robust notion of connectivity (or disconnectivity).

formally relates λ_2 to $\varphi(G)$

Theorem 2.6 (Cheeger's Inequality). *Let $G = (V, E)$ be a d-regular graph. Let $\overline{\boldsymbol{L}} = \boldsymbol{I} - \boldsymbol{A}/d$ denote the normalized Laplacian of G with eigenvalues being $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \le 2$. Then, we have the following:* $(L = I - W)$

$$\frac{\lambda_2}{2} \le \varphi(G) \le 4\sqrt{\lambda_2}.$$

$(\varphi(G) \le \sqrt{2\lambda_2})$

Proof : (Alon–Milman '85)

Let us setup some of the machinery we will use before we proceed with the proof. It is instructive to compare variational characterization of λ_2 with $\varphi(G)$. First, recall that for a set S we have,

Rayleigh Quotient

$$\mathcal{R}_{\overline{\boldsymbol{L}}}(\mathbf{1}_S) = \frac{\sum_{(u,v)\in E}(\mathbf{1}_S(u) - \mathbf{1}_S(v))^2}{d\sum_{v\in V} \mathbf{1}_S(v)^2} = \frac{|E(S, \overline{S})|}{d|S|} = \varphi(S)$$

1_S = vector with 1 if $u \in S$, 0 if $u \notin S$

This means you can view Rayleigh Quotients as a continuous analog of conductance, a fact that will serve us well as we proceed with the argument. We now prove the following claim.

Lemma 2.7. $\lambda_2/2 \le \varphi(G)$.

Proof of Lemma 2.7. We will show this by producing a vector orthogonal to $\mathbf{1}$ which has small Rayleigh Quotient. Let S be the set which minimizes $\varphi(S)$. Let $\boldsymbol{x} = \frac{\mathbf{1}_S}{|S|} - \frac{\mathbf{1}_{\overline{S}}}{|\overline{S}|}$. Note that $\boldsymbol{x} \perp \mathbf{1}$. This means $\lambda_2 \le \mathcal{R}_{\overline{\boldsymbol{L}}}(\boldsymbol{x})$. Also, you can verify

$x \perp 1$, 1 is an eigenvector of L max value of

$$\lambda_2 \le \mathcal{R}_{\overline{\boldsymbol{L}}}(\boldsymbol{x}) = \frac{|E(S,\overline{S})|}{d|S|} \cdot \left(\frac{|S| + |\overline{S}|}{|\overline{S}|}\right) \le 2\varphi(S) \le 2\varphi(G).$$

$1 \cdot 1_S = |S|$

$x \cdot 1 = 1 - 1 = 0$

(since, $R_L(x)$ is rayleigh quotient corresp to another space)

#ques: why is $\lambda_2 \le R_L(x)$? → it is rayleigh quotient of a different n-1 dimensional space!

• if $\varphi(G) > \alpha$, then G is an expander graph. [Lubotzky – Philips – Sarnak '84] came up with expander graphs

Given a d, α construct ∞-family of graphs $\{G_n\}$, G_n is d-regular and $\varphi(G) > \alpha$

□

While this argument is fairly simple, I would like to point out it still says something remarkable. Namely, if you believe a graph is an expander, there is a fairly easy way to certify that. Up to some loss, the Rayleight quotient of any vector in the span of the first two eigenvectors is a certificate that the expansion of the graph is kind of large. Next up, we prove the following lemma.

(Harder direction) **Lemma 2.8.** *For any vector $\boldsymbol{x} \perp \mathbf{1}$, we have $\varphi(G) \leq 4\sqrt{\mathcal{R}_{\overline{\boldsymbol{L}}}(\boldsymbol{x})}$. In particular, this means $\varphi(G) \leq 4\sqrt{\lambda_2}$.*

We note that Theorem 1.11 immediately follows from Theorem 2.6. This is because any connected graph has expansion at least $1/|E(G)| = 1/m$. And thus we have $\lambda_2 \geq 1/m^2$ as desired. For bounded degree graph, this bound equals $1/n^2$. Note that Theorem 2.6 immediately follows from Lemma 2.7 and Lemma 2.8. Thus, our task now reduces to proving Lemma 2.8. The proof is fairly intricate and algorithmic. Below, we present the algorithm.

fast

One of these cuts is the best cut (min. $\varphi(S)$) → shady motivation

sweep factor based approach

FindSparseCut(G)

1. Let $\boldsymbol{x}$ be the eigenvector with eigenvalue λ_2. (negative, +ve entries, since $\sum x_i = 0$)
2. Sort $\boldsymbol{x}$ as $\boldsymbol{x}_1 \leq \boldsymbol{x}_2 \leq \cdots \leq \boldsymbol{x}_n$.
3. Assign $S_1 = \{1\}$ and $X = S_1$
4. For $i = 2$ *to* $n - 1$.
   (a) Consider sweep cut $(S_i, \overline{S_i})$ where S_i consists of the first i vertices. → prefix cuts
   (b) Assign $S = \mathrm{argmin}(|S_i|, |\overline{S_i}|)$.
   (c) Assign $X =$ better cut between X *and* S.
5. Return X

The above algorithm just sorts the vertices in increasing order of values along $\boldsymbol{x}$ and then takes the best sweep cut. Note that this way the algorithm only optimizes over n cuts which is way better than the number of cuts you would consider in a brute force approach. Let us now roll our sleeves and clear out the path ahead. First things first. We need to show that you can find a set containing at most half the vertices which has small expansion. This is done in the claim below. For ease of notation, we will now write $\mathcal{R}$ to denote Rayleigh Quotient with respect to $\mathcal{R}_{\overline{\boldsymbol{L}}}$.

Claim 2.9. *For any vector $\boldsymbol{x} \perp \mathbf{1}, \boldsymbol{x} \in \mathbb{R}^n$, there are vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ with disjoint supports such that $\mathcal{R}(\boldsymbol{y}), \mathcal{R}(\boldsymbol{z}) \leq 4 \cdot \mathcal{R}(\cancel{\boldsymbol{z}})$.*

set of entries with non-zero values are disjoint

Proof of Claim 2.9. Define $\boldsymbol{y}$ as a truncation of $\boldsymbol{x}$ which is obtained by copying over all positive entries in $\boldsymbol{x}$ and zeroing out the negative ones. Similarly define $\boldsymbol{z}$ which is obtained by a similar truncation done on $-\boldsymbol{x}$. Formally,

$$\boldsymbol{y}_u = \begin{cases} \boldsymbol{x}_u \text{ if } \boldsymbol{x}_u \geq 0 \\ 0 \text{ otherwise} \end{cases} \qquad \boldsymbol{z}_u = \begin{cases} -\boldsymbol{x}_u \text{ if } -\boldsymbol{x}_u \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

→ leads to disjoint supports

Let $(\boldsymbol{x}_u - \boldsymbol{x}_v)^2$ denote $\boldsymbol{x}$-length of edge (u, v). Truncations can never elongate an edge. Thus, for any edge

(trivial by cases)
$$(\boldsymbol{y}_u - \boldsymbol{y}_v)^2, (\boldsymbol{z}_u - \boldsymbol{z}_v)^2 \leq (\boldsymbol{x}_u - \boldsymbol{x}_v)^2.$$

$$R_{\overline{L}}(x) = x^T \overline{L} x = \frac{\sum_{(u,v)\in E} (y_u - y_v)^2}{d \sum y_i^2}$$

y, z have disjoint support, hence are orthogonal.

Idea: truncate bigger x' to obtain y', z', with x'-length of (u,v) $\leq$ x-length of (u,v)

So, the numerators for Rayleigh Quotients of $\boldsymbol{y}$ and $\boldsymbol{z}$ are smaller than the numerator for Rayleigh Quotients of $\boldsymbol{x}$. The main term in the denominator of these Rayleigh Quotients is the Euclidean length of these vectors. For $\boldsymbol{x}$, we have $\|\boldsymbol{x}\|^2 = 1$. If $\boldsymbol{y}, \boldsymbol{z}$ have squared lengths at least $1/4$, we are done. Note that $\|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2 = \|\boldsymbol{x}\|^2 = 1$. Without loss of generality, suppose $\|\boldsymbol{y}\|^2 \leq 1/4$ which means $\|\boldsymbol{z}\|^2 \geq 3/4$. In this case, we will modify $\boldsymbol{y}$ and $\boldsymbol{z}$ to obtain non-negative vectors with small Rayleigh Quotients. So, the difficulty in the current argument is that the vector $\boldsymbol{y}$, which is obtained by truncating $\boldsymbol{x}$, is too small. To fix this, consider $\boldsymbol{x}' = \boldsymbol{x} + \frac{\mathbf{1}}{2\sqrt{n}}$. Notice that the lengths of all edges according to $\boldsymbol{x}$ and $\boldsymbol{x}'$ are the same. Also, note that since $\boldsymbol{x} \perp \mathbf{1}$ we have $\|\boldsymbol{x}'\|^2 = \|\boldsymbol{x}\|^2 + 1/4 = 5/4$. We now truncate this longer vector to obtain $\boldsymbol{y}'$ and $\boldsymbol{z}'$. We will show that being truncations of this slightly longer vector, $\boldsymbol{y}'$ and $\boldsymbol{z}'$ both will have squared length at least $1/8$.

put in orignal expr & check.

edge lengths $(x_u - x_v)^2$ are unchanged

$y' = $ pos trun(x')

$z' = $ neg trun(y')

Let us show that $\|\boldsymbol{y}'\|^2 \geq 1/8$. To this end, note that

$$\|\boldsymbol{y}'\|^2 + \|\boldsymbol{z}'\|^2 = \|\boldsymbol{x}'\|^2 = 5/4.$$

Moreover, the vector $\boldsymbol{z}'$ is a truncation of $\boldsymbol{x}' = \boldsymbol{x} + \frac{\mathbf{1}}{2\sqrt{n}}$ which retains only absolute values of negative entries in $\boldsymbol{x}'$. This means,

$$\|\boldsymbol{z}'\|^2 \leq \|\boldsymbol{z}\|^2 \leq \|\boldsymbol{x}\|^2 = 1.$$

negative entries now lesser in magnitude

In turn, this means $\|\boldsymbol{y}'\|^2 \geq 1/4 \geq 1/8$.

Finally, we show that $\|\boldsymbol{z}'\|^2 \geq 1/8$. To this end, note that

$(a+b)^2 \leq 2a^2 + 2b^2$

$$3/4 \leq \|\boldsymbol{z}\|^2 = \sum \boldsymbol{z}_u^2 \overset{why?}{\leq} \sum_{u \in V} \left(\boldsymbol{z}'_u + \frac{1}{2\sqrt{n}}\right)^2 \overset{\text{see below}}{\leq} \sum_{u \in V} \left(2\boldsymbol{z}'^2_u + \frac{1}{4n}\right) = 2\|\boldsymbol{z}'\|^2 + \frac{1}{2}.$$

-ve vals of z_u eq., 0 vals of z_u compensated

The "see below" inequality uses $(a+b)^2 \leq 2a^2 + 2b^2$. Note that from the above, it follows that $\|\boldsymbol{z}'\|^2 \geq 1/8$ as desired. □

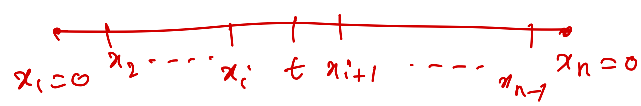
From now on, we just need to "round" this non-negative vector to find a low conductance cut. Luckily for us, results for analogous cut problems have been (re)discovered multiple times. We state and prove this result which basically asserts that you can find a set S all of whose vertices are a subset of non-zero coordinates in $\boldsymbol{y}$ (or $\boldsymbol{z}$) above.

Claim 2.10. *For every non-negative vector $\boldsymbol{x} \in \mathbb{R}^n$, there is a set $S \subseteq supp(\boldsymbol{x})$ which satisfies*

$$\varphi(S) \leq \frac{\sum_{(u,v)\in E} |\boldsymbol{x}_u - \boldsymbol{x}_v|}{d \cdot \sum \boldsymbol{x}_v}.$$

One useful approach to these problems is to in fact imagine that your graph is laid out on the real line where the ith vertex i situated on point $\boldsymbol{x}_i$. An edge $(i,j) \in E(G)$ is drawn so that it connects $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. It is convenient to proceed assuming $\max \boldsymbol{x}_i = 1$. With this setup behind you, all you need to do to specify a cut is to just pick a random $t \in [0,1]$. The vertices to the left of t form the left side of the cut and the others form the right side (which we denote as S_t). You want to pick t from a well chosen distribution which ensures that in expectation, the cut satisfies the properties you want. Details follow.

i.e. u s.t. $x_u < t$ form left side of cut



Proof of Claim 2.10. Since the right hand side is homogeneous in $\boldsymbol{x}$, we can always rescale $\boldsymbol{x}$ and assume without loss of generality that $\max \boldsymbol{x}_i = 1$. We pick $t \sim \textsc{unif}[0,1]$. Let

$$S_t = \{v \in V : \boldsymbol{x}_v \geq t\}.$$

We will show the following two items.

- $\mathbf{E}[|S_t|] = \sum \boldsymbol{x}_v$.
- $\mathbf{E}[\#\text{ edges crossing } (S_t, \overline{S_t})] = \sum_{(u,v)\in E} |\boldsymbol{x}_u - \boldsymbol{x}_v|$

} proof done in notes

We now show how to finish the proof assuming both of these items. After that is done, we will prove both of these items. Nevertheless, the reader is invited to prove it on their own. Going forward, let us assume both of the items above indeed hold. Thus, we have a ratio of expectations:

Note: $|S_t|$ can exceed $n/2$, so we choose vector with smaller support!

$$\frac{\mathbf{E}[|E(S_t, \overline{S_t})|]}{d \cdot \mathbf{E}[|S_t|]} = \frac{\sum_{(u,v)\in E} |\boldsymbol{x}_u - \boldsymbol{x}_v|}{d \cdot \sum \boldsymbol{x}_v}.$$

$\mathbb{E}[X] \leq \alpha$, $\mathbb{E}[Y] = \beta$ $\Rightarrow \mathbb{E}\left[X - \frac{\alpha Y}{\beta}\right] \leq 0$ $\Rightarrow \exists X, Y$ s.t. $\frac{X}{Y} \leq \frac{\alpha}{\beta}$

While this ratio matches the right hand side, we are interested in finding one single S_t which simultaneously satisfies both the items above. In other words, we are interested in the expectation of the ratio and not the ratio of expectations. But notice that numerator and denominator are both non-negative quantities. For two non-negative random variables $\boldsymbol{X}, \boldsymbol{Y}$ note that $\frac{\mathbf{E}[\boldsymbol{X}]}{\mathbf{E}[\boldsymbol{Y}]} \leq \alpha$ implies that the event $\boldsymbol{X}/\boldsymbol{Y} \leq \alpha$ occurs with a positive probability(!) In turn, this means that for our problem, there indeed exists a cut $(S_t, \overline{S_t})$ for which $\varphi(S_t)$ is at most $\frac{\sum_{(u,v)\in E} |\boldsymbol{x}_u - \boldsymbol{x}_v|}{d \cdot \sum \boldsymbol{x}_v}$ as desired. □

NB: $\frac{a_1+a_2+\dots+a_n}{b_1+\dots+b_n} = r$

$\min \frac{a_i}{b_i} \leq r \leq \max \frac{a_i}{b_i}$

If all $a_i > r b_i$ $\Rightarrow \sum a_i > r \sum b_i$ (contradiction)

Finally, we collect all these ingredients and finish proving Cheeger's bound.

Proof of Lemma 2.8. We would like to find the "correct" non-negative vector on which we can apply Claim 2.10 to find the low conductance cut we want. We use Claim 2.9 and obtain vectors non-negative vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ both of which satisfy $\mathcal{R}(\boldsymbol{y}), \mathcal{R}(\boldsymbol{z}) \leq 4\mathcal{R}(\boldsymbol{x})$. Say $\boldsymbol{y}$ is the vector with smaller support among these. We finally define the non-negative vector we want: let $\boldsymbol{w} \in \mathbb{R}^n$ be the following non-negative vector: for $v \in V$, we have $\boldsymbol{w}_v = \boldsymbol{y}_v^2$.

We will show

$$\frac{\sum_{(u,v)\in E} |\boldsymbol{w}_u - \boldsymbol{w}_v|}{d \sum \boldsymbol{w}_v} \leq 4\sqrt{\mathcal{R}(\boldsymbol{x})}.$$

The proof is a slick application of Cauchy-Schwartz.

$$
\begin{aligned}
\frac{\sum_{(u,v)\in E} |\boldsymbol{w}_u - \boldsymbol{w}_v|}{d \sum \boldsymbol{w}_v} &= \frac{\sum_{(u,v)\in E} (\boldsymbol{y}_u^2 - \boldsymbol{y}_v^2)}{d \sum \boldsymbol{y}_v^2} \\
&= \frac{\sum_{(u,v)\in E} |\boldsymbol{y}_u - \boldsymbol{y}_v| \cdot |\boldsymbol{y}_u + \boldsymbol{y}_v|}{d \sum \boldsymbol{y}_v^2} \\
&\le \frac{\sqrt{\sum_{(u,v)\in E} |\boldsymbol{y}_u - \boldsymbol{y}_v|^2} \cdot \sqrt{\sum_{(u,v)\in E} |\boldsymbol{y}_u + \boldsymbol{y}_v|^2}}{d \cdot \sum \boldsymbol{y}_v^2} \\
&\le \sqrt{\mathcal{R}(\boldsymbol{y})} \cdot \sqrt{\frac{\sum_{(u,v)\in E} |\boldsymbol{y}_u + \boldsymbol{y}_v|^2}{d \sum \boldsymbol{y}_v^2}}
\end{aligned}
$$

(holy-shit!)

cauchy-schwartz

It remains to upper bound $\sum_{(u,v)\in E} |\boldsymbol{y}_u + \boldsymbol{y}_v|^2$. Using the inequality $(a+b)^2 \le 2a^2 + 2b^2$, it follows that $\sum_{(u,v)\in E} |\boldsymbol{y}_u + \boldsymbol{y}_v|^2 \le 2d \sum \boldsymbol{y}_v^2$. Using $\mathcal{R}(\boldsymbol{y}) \le 8\mathcal{R}(\boldsymbol{x})$, the result follows. $\square$

$\le 2\sqrt{2}\sqrt{R(x)} \cdot \sqrt{2}$

$= 4\sqrt{R(x)} = 4\sqrt{\lambda_2}$

$x = 2^{nd}$ eigenvector !

<u>Note</u>: Alon-Millman algorithm still not proved, we just got bounds for $\varphi(G)$ as $\frac{\lambda_2}{2} \le \varphi(G) \le 4\sqrt{\lambda_2}$