Lecture 18-19: Eigenvalues of Graphs, Expander Graphs

Instructor: Anand Louis *Scribe: Aditya Subramanian, Sruthi Gorantla*

Last week we studied the Singular Value Decomposition of rectangular matrices. This week we shift our focus to the spectrum of square matrices, and in particular the symmetric square matrices related to graphs.

1 Gershgorin Circle Theorem

This discussion here is based on the lecture notes [Tul19]. We begin with a simple yet powerful theorem that lets us give bounds on the eigenvalues of matrices.

Theorem 1. *Let $M \in \mathbb{C}^{n\times n}$. Let $R_i = \sum_{i\neq j} |M_{ij}|$. Define the set*

$$Disc(M_{ii}, R_i) := \{z \mid z \in \mathbb{C}^n, |z - M_{ii}| \leq R_i\}$$

If λ is an eigenvalue of M, then

$$\lambda \in \bigcup_{i=1}^{n} Disc(M_{ii}, R_i)$$

Proof. Let $x \in \mathbb{C}^n$ be an eigenvector corresponding to the eigenvalue λ. Let $i_0 = \text{argmax}_{j\in[n]}\{|x_j|\}$. Since x is an eigenvector we have

$$Mx = \lambda x \implies \forall i \in [n], \sum_{j=1}^{n} M_{ij}x_j = \lambda x_i$$

In particular, for $i = i_0$ we have,

$$\sum_{j=1}^{n} M_{i_0 j} = \lambda x_{i_0} \implies \sum_{j=1}^{n} M_{i_0 j}\frac{x_j}{x_{i_0}} = \lambda \implies \sum_{j\neq i_0} M_{i_0 j}\frac{x_j}{x_{i_0}} = \lambda - M_{i_0 i_0}$$

Thus we have,

$$|\lambda - M_{i_0 i_0}| \leq \sum_{j\neq i_0} |M_{i_0 j}|\left|\frac{x_j}{x_{i_0}}\right| \leq \sum_{j\neq i_0} |M_{i_0 i_0}| = R_{i_0}$$

□

Corollary 2. *For graphs of maximum degree d, all eigenvalues of the adjacency matrix lie in $[-d, d]$.*

This follows simply from the theorem as the non-diagonal entries in every row sum to the degree of the vertex, d. Since the diagonal entry is always zero, all eigenvalues lie in a disc of radius d centered at 0. Since A is symmetric, the eigenvalues are real, and the intersection of the real line with the disc gives us the range $[-d, d]$.

Corollary 3. *For graphs with self loops such that a vertex i, with degree d_i, also has a self loop of weight d_i, all the eigenvalues of the adjacency matrix lie in $[0, 2\max_i d_i]$.*

In this case we get a disc centered around d_i (since the diagonal entries are now d_i) of radius d_i corresponding to each vertex. Taking union of all such discs gives us the required result.

Corollary 4. *A symmetric diagonally dominant matrix ($|M_{ii}| \geq \sum_{i\neq j} |M_{ij}|$) with non-negative diagonal entries is positive semidefinite i.e $M \succcurlyeq 0$.*

Similar to corollary 3 we get a union of discs all of which are non-negative. This tells us that all eigenvalues are ≥ 0, or in other words that the matrix is positive semidefinite.

2 Cholesky Decomposition of Semidefinite Matrices

Theorem 5. *For a symmetric matrix M, $M \succcurlyeq 0$ iff there exists a matrix A such that $M = A^T A$*

We already proved in the previous lecture that $M = A^T A \implies M \succcurlyeq 0$ (when we showed that eigenvalues of $A^T A$ are squares of singular values of A). Here we show only the other direction.

Proof. We know that a symmetric matrix can be represented as $M = V\Lambda V^T$, where columns of V are the orthonormal eigenvectors, and Λ is the diagonal matrix of corresponding eigenvalues (which are ≥ 0) of M. Rewriting this, we get

$$M = V\Lambda V^T = V\Lambda^{1/2}\Lambda^{1/2}V^T = (\Lambda^{1/2}V^T)^T(\Lambda^{1/2}V^T)$$

We have successfully decomposed M as $A^T A$ where $A = \Lambda^{1/2}V^T$. □

We note that this decomposition is not unique. We can multiply A by any rotation matrix R to still get a valid solution.

3 Courant Fischer Theorem

The discussion here is based on [AS18].

Theorem 6. *Let A be a symmetric matrix with eigenvalues $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ and corresponding orthonormal eigenvectors $v_1, v_2, \ldots, v_n$ respectively. Then,*

$$\sigma_k = \max_{\substack{S \subseteq \mathbb{R}^n \\ rank(S)=k}} \min_{x \in S} \frac{x^T A x}{x^T x} = \min_{\substack{T \subseteq \mathbb{R}^n \\ rank(T)=n-k+1}} \max_{x \in T} \frac{x^T A x}{x^T x}$$

Proof. First we verify that for the subspace $S = span\{v_1, v_2, \ldots, v_k\}$, the minimum over $\frac{x^T A x}{x^T x}$ is at least σ_k . For every $x \in S$, we can write,

$$x = \sum_{i=1}^{k} c_i v_i$$

so,

$$\frac{x^T A x}{x^T x} = \frac{\sum_{i \in [k]} \sigma_i c_i^2}{\sum_{i \in [k]} c_i^2} \geq \frac{\sum_{i \in [k]} \sigma_k c_i^2}{\sum_{i \in [k]} c_i^2} = \sigma_k$$

Now we need to verify that this is indeed the maximum. Let $T_k = span\{v_k, v_{k+1} \ldots, v_n\}$. As T_k has dimension $n-k+1$, for any subspace S of dimension k, $S \cap T_k \neq \emptyset$. The minima over the restricted subspace $S \cap T_k$ has to be atleast as large as the minima over S, so

$$\min_{x \in S} \frac{x^T A x}{x^T x} \leq \min_{x \in S \cap T_k} \frac{x^T A x}{x^T x}$$

any $x \in S \cap T_k$ can be expressed as,

$$x = \sum_{i=k}^{n} c_i v_i$$

and so,

$$\frac{x^T A x}{x^T x} = \frac{\sum_{i=k}^{n} \sigma_i c_i^2}{\sum_{i=k}^{n} c_i^2} \leq \frac{\sum_{i=k}^{n} \sigma_k c_i^2}{\sum_{i=k}^{n} c_i^2} = \sigma_k$$

Hence we can say that for any subspace S of dimension k,

$$\min_{x \in S} \frac{x^T A x}{x^T x} \leq \sigma_k$$

so we can conclude that,

$$\sigma_k = \max_{\substack{S \subseteq \mathbb{R}^n \\ rank(S)=k}} \min_{x \in S} \frac{x^T A x}{x^T x}$$

□

Definition 7. *The Rayleigh quotient of a vector x w.r.t matrix M is defined as $R_M(x) \stackrel{def}{=} \frac{x^T M x}{x^T x}$.*

Remark For an eigenvector, the Rayleigh quotient is it's corresponding eigenvalue.
For a eigenvector λ (with eigenvalue σ) of matrix M,

$$R_M(\lambda) = \frac{\lambda^T M \lambda}{\lambda^T \lambda} = \frac{\sigma \lambda^T \lambda}{\lambda^T \lambda} = \sigma$$

Courant-Fischer theorem tells us that σ_k is the minimum value that the rayleigh quotient can take among vectors in a subspace of dimension k, and from the remark above we know that this extrema is achieved by the corresponding eigenvector.

4 Eigenvalues of Graphs

4.1 Adjacency Matrix

We know that for a d-regular graph, every row and column of the Adjacency matrix A sums to d. So, if we multiply A by the all 1s vector $\mathbf{1} = (1, 1, \ldots, 1)^T$ we get $A\mathbf{1} = d\mathbf{1}$ which implies that d is an eigenvalue of A with the corresponding eigenvector $\mathbf{1}$. From Corollary 2 we know that all eigenvalues of A are in the range $[-d, d]$, which means that this is the maximum eigenvalue of A.

Let $\lambda_1, \lambda_2, \ldots \lambda_n$ be the eigenvalues of A. We can further show that

Lemma 8. *Second largest eigenvalue of $A = d$ iff the graph is disconnected.*

Proof. **disconnected graph** $\implies \lambda_2 = d$
Since graph is disconnected we can appropriately label the vertices to write the adjacency matrix as

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

where A_1, A_2 are the corresponding adjacency matrices of the components. Let $\mathbf{1}_{a1}, \mathbf{1}_{a2}$ represent the column vectors with 1 only in positions corresponding to the vertices in A_1, A_2 respectively; and zero otherwise. Now,

$$\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} \mathbf{1}_{a1} \\ 0 \end{bmatrix} = d \begin{bmatrix} \mathbf{1}_{a1} \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{1}_{a2} \end{bmatrix} = d \begin{bmatrix} 0 \\ \mathbf{1}_{a2} \end{bmatrix}$$

So we have 2 orthogonal eigenvectors both corresponding to the eigenvalue d i.e. $\lambda_2 = d$.
$\lambda_2 = d \implies$ **disconnected graph**
If $\lambda_2 = d$, then there exist 2 orthogonal eigenvectors with eigenvalue d. Using these eigenvectors, we shall show that the graph has at least 2 components and hence is disconnected. Consider the matrix $L = dI - A$. Note that any eigenvector x of A with eigenvalue λ, is also an eigenvector of L with eigenvalue $d - \lambda$, since

$$Lx = (dI - A)x = dIx - Ax = dx - \lambda x = (d - \lambda)x$$

Let x be one of the eigenvectors of A with eigenvalue d. It satisfies $Lx = 0$. Now,

$$0 = x^T L x = x^T (dI - A)x = \sum_i dx_i^2 - 2\sum_{ij \in E} x_i x_j = \sum_{ij} x_i^2 - 2\sum_{ij \in E} x_i x_j = \sum_{ij \in E} (x_i - x_j)^2$$

$$\therefore 0 = \sum_{ij \in E} (x_i - x_j)^2$$

Thus for each pair of vertices (i,j) connected by an edge, we have $x_i = x_j$. We can inductively apply this fact to all vertices connected by a path (i.e. vertices in one component, say C_1 of the graph) to get that $\forall i, j \in C_1, x_i = x_j$.

But now, if the graph had only one component we would have $\forall i, j \in V, x_i = x_j$ which tells us that any eigenvector x with eigenvalue d is spanned by the all 1s vector $\mathbf{1}$. This is a contradiction since there are at least 2 orthogonal vectors with eigenvalue d, and so there are at least 2 components in the graph. □

Lemma 9. *Multiplicity of the eigenvalue d = number of components in the graph.*

Proof-sketch. We can generalize the proof of Lemma 8 for a graph with k components. We re-label the vertices so that we can write the Adjacency martix A as a block diagonal martix with the blocks A_i being the adjacency matrices of the individual components, C_i. Now, if we construct eigenvectors e_i with $e_i(j) = 1$ iff $j \in V(C_i)$ representing the vertices of each component, we see that we get the required k orthogonal eigenvectors each corresponding to the eigenvalue d. So, multiplicity of eigenvalue $d \geq$ number of components.

Similarly extending the proof in the other direction, if we had multiplicity of eigenvalue d as k but only $k-1$ components, we would have that all eigenvectors of eigenvalue d can be spanned by the set of $k-1$ orthogonal vectors $\{e_1, e_2, \ldots, e_{k-1}\}$ (e_i as defined in the last paragraph) . This gives us a contradiction since there are k such orthogonal eigenvectors. Hence, the number of components $\geq$ multiplicity of eigenvalue d. □

Now, if we are given an arbitrary weighted graph, we know that in the adjacency matrix, A_{ij} = weight of edge $\{i,j\}$. Let us define the diagonal matrix D where $D_{ii} = d_i = \sum_{j \neq i} A_{ij}$.

Definition 10. *The Normalized Adjacency Matrix of a graph is defined as*

$$\mathcal{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

Lemma 11. *The eigenvalues of the normalized adjacency matrix $\mathcal{A}$ lie in the range $[-1, 1]$. (HW problem)*

4.2 Laplacian Matrix

Here we define another matrix associated with a graph

Definition 12. *The Laplacian Matrix of a graph is defined as*

$$L \stackrel{def}{=} D - A$$

where A is the adjacency matrix of the graph and D is the diagonal matrix with $D_{ii} = d_i$.

By construction L is symmetric, diagonally dominant, and has positive diagonal entries. Hence by Corollary 4 it is a positive semidefinite matrix i.e. $L \succcurlyeq 0$.

Similarly we also define

Definition 13. *The Normalized Laplacian Matrix of a graph is defined as*

$$\mathcal{L} \stackrel{def}{=} D^{-\frac{1}{2}} (D - A) D^{-\frac{1}{2}} = I - \mathcal{A}$$

where I is the identity matrix.

Lemma 14. *$\mathcal{L}$ is positive semidefinite i.e. $\mathcal{L} \succcurlyeq 0$.*

Proof. It is sufficient to show that the smallest eigenvalue of $\mathcal{L}$ is non-negative.

$$\min_x \frac{x^T \mathcal{L} x}{x^T x} = \frac{(D^{-\frac{1}{2}}x)^T L (D^{-\frac{1}{2}}x)}{x^t x} = \frac{y^T L y}{x^T x}$$

Since L is a positive semidefinite matrix, the numerator $y^T \mathcal{L} y \geq 0$. Also $x^T x = ||x||^2 \geq 0$. Hence the expression is always ≥ 0 i.e.

$$\min_x \frac{x^T \mathcal{L} x}{x^T x} \geq 0$$

□

Lemma 15. *λ is an eigenvalue of $\mathcal{A}$ $\iff$ $1 - \lambda$ is an eigenvalue of $\mathcal{L}$.*

Proof. Let λ, x be any eigenvalue, eigenvector pair of $\mathcal{A}$. It follows that $\mathcal{A}x = \lambda x$. Now consider,

$$\mathcal{L}x = (I - \mathcal{A})x = Ix - \mathcal{A}x = x - \lambda x = (1 - \lambda)x$$

So x is also an eigenvector of $\mathcal{L}$ and it's corresponding eigenvalue is $1 - \lambda$. □

Lemma 16. *The smallest eigenvalue of $\mathcal{L}$ is always zero.*

Proof. Consider the vector $x = D^{1/2}\mathbf{1}$,

$$\mathcal{L} \cdot x = \mathcal{L}(D^{1/2}\mathbf{1}) = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} (D^{1/2}\mathbf{1}) = D^{1/2} L \mathbf{1} = 0$$

Hence we see that zero is an eigenvalue of $\mathcal{L}$ (corresponding to the eigenvector x). By Lemma 14 all eigenvalues of $\mathcal{L}$ are non-negative. Hence, zero is the smallest eigenvalue. □

Lemma 17. *For regular graphs, $\lambda_2 = 0$ iff the graph is disconnected. Further, the multiplicity of the eigenvalue 0 is the number of components in the graph.*

Proof-sketch. We note that in $\mathcal{L}$ the rows sum to zero. Hence the eigenvectors that we described for A in Lemma 8 are also eigenvectors of $\mathcal{L}$, but now correspond to the eigenvalue of zero. Rest of the proof follows similarly to the proof of Lemma 8. □

4.3 Expansion and Sparse Cuts

We noted in the last section that the graph is disconnected iff $\lambda_2 = 0$. We further ask the question, if a graph will be *close to disconnected* if λ_2 is *close* to zero? To study this we introduce the notion of expansion of a vertex set, and a sparse cut.

Definition 18. *Expansion of a subset of vertices $S \subseteq V$ is*

$$\phi(S) \stackrel{def}{=} \frac{\sum_{i \in S, j \in V \setminus S} w_{ij}}{\sum_{i \in S} d_i}$$

where w_{ij} is the weight of edge $\{i, j\}$ and d_i is the degree of vertex i.

For regular unweighted graphs, $\phi(S) = \frac{|E(S,\overline{S})|}{d|S|}$

Alternately the expansion is also defined in terms of the *volume* of S, $\mathrm{vol}(S) \stackrel{def}{=} \sum_{i \in S} d_i$ as,

$$\phi(S) \stackrel{def}{=} \frac{\sum_{i \in S, j \in V \setminus S} w_{ij}}{\min\{\mathrm{vol}(S), \mathrm{vol}(V \setminus S)\}}$$

These 2 definitions are equivalent when $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$

Definition 19. *The expansion of a graph G is defined as the minimum expansion of a set S (among sets with volume lesser than half the total volume).*

$$\phi_G = \min_{\substack{S \subseteq V \\ vol(S) \leq vol(V)/2}} \phi(S)$$

Cheeger's inequality gives us a bound on the expansion of a graph in terms of the second eigenvalue. This is very useful to get an approximate value of ϕ_G since it's exact computation is NP-hard.

Theorem 20 (Cheeger's Inequality)**.**

$$\frac{\lambda_2}{2} \leq \phi_G \leq \sqrt{2\lambda_2}$$

we shall study the inequality in more detail next week.

4.4 Expander Graphs

Definition 21. *A graph G is said to be an α-expander if $\phi_G \geq \alpha$.*

Informally we say a graph is an expander if it's expansion is an absolute constant i.e. $\Omega(1)$. Complete graphs K_n, random graphs $G(n, p)$ with a value of p close to 1, are some examples of expanders. Expanders can also be explicitly constructed. Cycle graphs and path graphs are not expanders since their expansion decays with the number of vertices.
Another notion of expansion that is studied is,

Definition 22. *A d-regular graph is called a β-spectral expander if $\max_{i \in \{2,\ldots n\}} |\sigma_i| \leq \beta$, where σ_i are the eigenvalues of the adjacency matrix.*

This notion of expansion is closely related to the first since it can be shown by Cheeger's inequality that such graphs have expansion at least $\frac{d-\beta}{2d}$. For more discussions on expander graphs and Cheeger's inequality see [HLW06, MKO13, AS15, Tre16a, Tre16b, Tre11b, Tre11a].

5 Random Walks on Graphs

The discussion here is based on [TL13, Mah15]. The idea is to start with an vertex chosen uniformly at random and iteratively choose a neighbor of the current vertex uniformly at random. This results in a random walk on a graph G.

Let A be the adjacency matrix of a graph G. Let μ_0 be the starting probability distribution over the vertices of G to start the random walk with. Then, μ_1 is the probability distribution over the neighbours adjacent to the vertex chosen in the first step, μ_2 is the probability distribution over the neighbours adjacent to the vertex chosen in the second step, and so on.

Lemma 23. *For a d-regular graph with adjacency matrix A, if μ_t is the probability distribution of the random walk at time t, then,*

$$\mu^{t+1} = \frac{1}{d} A \cdot \mu^t.$$

Proof. We prove this claim by induction on t. Let μ_0 be the initial probability distribution over all the vertices of the graph G.
Base step. For $t = 1$, let v_i be the vertex chosen in the first iteration. Then there are d choices for the second vertex since the graph is d-regular. Then the probability of landing at a neighbour vertex v_j in the second step is $\frac{1}{d}\mu^0(i)$. Hence we have,

$$\mu^1 = \frac{1}{d} A \cdot \mu^0.$$

Induction step. By the induction hypothesis, $\mu_t(j)$ is the probability of the random walk landing at vertex v_j at time t. Let $\mu^{t+1}(i)$ be the probability of the random walk landing at vertex v_i at time $t+1$. Then,

$$\mu^{t+1}(i) = \sum_{j:(i,j)\in E(G)} \mu^t(j)\frac{1}{d}.$$

Therefore,

$$\mu^{t+1} = \frac{1}{d}A \cdot \mu^t.$$

□

The matrix $\frac{1}{d}A$ is usually called as a *random walk matrix.*

Definition 24 (Stationary distribution)**.** *If one step of a random walk on a distribution μ^* results in the same distribution, then μ^* is said to be a stationary distribution.*

$$\mu^* = \frac{A}{d}\mu^*.$$

Remark μ^* is the eigenvector of the matrix A with eigenvalue d.

Let us consider a connected undirected graph G. We know that the top eigenvalue of its adjacency matrix is d with multiplicity 1. Therefore $\mathbf{1}/n$ is a unique stationary distribution of G.

Definition 25 (δ-mixing time)**.** *The smallest time t such that $d_{TV}\left(\mu^t, \mu^*\right) \leq \delta$ for any starting distribution μ_0. Here, the total variation distance d_{TV} is defined as follows,*

$$d_{TV}\left(\mu^t, \mu^*\right) = \frac{1}{2}\|\mu^t - \mu^*\|_1.$$

5.1 Applications of Random Walks on Expanders

In the following example, we will use random walks to reduce the amount of randomness required by a randomized algorithm.

Consider the set B of binary strings of length n. Let **ALG** be a randomized algorithm to decide whether an input string x belongs to B or not. Suppose ALG used R bits of randomness and satisfies the following properties,

$$\text{If } x \in B, \quad \Pr_r\left[\text{ALG}\left(x, r\right) = \text{yes}\right] \geq \frac{1}{2}.$$
$$\text{If } x \notin B, \quad \Pr_r\left[\text{ALG}\left(x, r\right) = \text{no}\right] = 1.$$

We can easily boost the probability of success by repeating the algorithm multiple times. Consider a second algorithm.

ALG2. Run ALG with k independent $r_1, \ldots, r_k$, output *yes* if any of the runs return *yes*. ALG2 satisfies the following properties,

$$\text{If } x \in B, \quad \Pr_{r_1,\ldots,r_k}\left[\text{ALG2}\left(x, r\right) = \text{yes}\right] \geq 1 - \frac{1}{2^k}.$$
$$\text{If } x \notin B, \quad \Pr_{r_1,\ldots,r_k}\left[\text{ALG2}\left(x, r\right) = \text{no}\right] = 1.$$

The first inequality follows from the independence of $r_1, \ldots, r_k$. Number of bits of randomness needed $=$ kR. In order to reduce the number of bits of randomness, we use a third algorithm that does a random walk on a d-regular spectral expander graph to choose r_i in the ith iteration instead of choosing $r_1, \ldots, r_k$ independently.

ALG3.

1. Construct a d-regular expander graph G on 2^R vertices for some $d = \mathcal{O}(1)$ and $\beta \leq 0.8d$. Then, jth vertex can represented by its R bit string r_j, for all $j \in 2^R$.

2. Sample r_1 to be a uniform random vertex in G.

3. Do $l-1$ steps of random walk starting from r_1, i.e., sample r_{i+1} to be a uniform random neighbor of r_i for all $1 \leq i \leq l-1$.

4. Output *yes* if any of the vertices visited by the random walk output *yes*. Otherwise output *no*.

Clearly Step 2 requires $\log_2 2^R = R$ bits of randomness to choose a starting vertex uniformly at random. Each of the subsequent $l-1$ steps of random walk in Step 3 take $\lceil \log_2 d \rceil$ bits of randomness. Hence, number of random bits used $= R + \lceil \log_2 d \rceil (l-1)$.

Lemma 26. *Total number of walks of length l between i and $j = \left(A^l\right)_{ij}$.*

Proof. We will prove this by induction on l.
Base step. For $l = 1$, walk of length 1 between two vertices i and j exists only when both the vertices adjacent to each other. Hence, total number of walks of length 1 between i and $j = A_{ij}$.
Induction step. By the induction hypothesis, $(A^l)_{ik}$ is the number of walks of length l between vertices i and k, and A_{kj} is the number of walks of length 1 between the vertices k and j. Therefore, for all $k \in [n]$ such that $(A^l)_{ik} \neq 0$ and $A_{kj} \neq 0$, there exists a walk of length $l+1$ between i and j, and the walk is $i \rightsquigarrow k \rightsquigarrow j$. Therefore, the total number of walks of length $l+1$ between i and $j = \sum_{k \in [n]} \left(A^l\right)_{ik} A_{kj} = \left(A^{l+1}\right)_{ij}$. ☐

Remark Total number of walks of length $l = \mathbf{1} A^l \mathbf{1} = d^l n$.

Theorem 27. *The error probability of ALG3 is at most $\frac{1}{2^k}$.*

Proof. Since ALG outputs *no* for a *no* instance with probability 1, we have the following,

$$\text{If } x \notin B, \quad \Pr_{r_1,\ldots,r_k}\left[\text{ALG3}(x,r) = \text{no}\right] = 1.$$

We now analyse the case when $x \in B$.

Fix an $x \in B$. Let $S = \{r : \text{ALG}(x,r) = \text{yes}\}$. Since ALG outputs *yes* with probability at least $\frac{1}{2}$, there are at least $2^R/2$ vertices in the expander graph such that ALG with the random string at those vertices outputs *yes*. Hence, $|S| \geq \frac{n}{2}$ where $n = 2^R$ is the total number of vertices in G. Since $|S|$ is "large", probability of a random walk avoiding S is small.
Let us define $\overline{A}$ as follows,

$$\overline{A}_{ij} = \begin{cases} 0 & \text{if } i \text{ or } j \in S, \\ A_{ij} & \text{otherwise.} \end{cases}$$

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\overline{A}$ with corresponding eigenvectors $v_1, \ldots, v_n$. Then $\lambda_{\max}^l$ is the largest eigen value of $\overline{A}^l$. From Lemma 26 we have that the total number of walks of length l that avoid $S = \mathbf{1}^\top \left(\overline{A}\right)^l \mathbf{1} \leq \lambda_{\max}^l n$.
If we show that $\lambda_{\max} < d$, it follows that the probability of avoiding S is at most $\frac{\lambda_{\max}^l n}{d^l n} = \left(\frac{\lambda_{\max}}{d}\right)^l$, which is very small. Let us look at an example. Consider the all 1 s vector $\mathbf{1}$.

$$\frac{\mathbf{1}^T \overline{A} \mathbf{1}}{\mathbf{1}^T \mathbf{1}} = \frac{\sum_i \sum_j \overline{A}_{ij}}{n} = \frac{\sum_{i \in V \setminus S} \sum_{j \in V \setminus S} A_{ij}}{n} \leq \frac{\sum_{i \in V \setminus S} d}{n} = \frac{n - |S|}{n} d \leq \frac{d}{2}.$$

The last inequality follows from the fact that $|S| \geq \frac{n}{2}$. Therefore, for the vector $\mathbf{1}$, the probability of avoiding S is at most $\frac{1}{2^l}$.

Now, for any vector $x \in \mathbb{R}^n$, define $z_i = \begin{cases} 0 & \text{if } i \in S, \\ x_i & \text{if } i \notin S. \end{cases}$

Then we have,

$$x^\top \overline{A} x = \sum_{ij} \overline{A}_{ij} x_i x_j = \sum_{ij} A_{ij} z_i z_j = z^\top A z.$$

Write $z = c_1 \mathbf{1} + z_\perp$. Here $\mathbf{1}$ is the top eigenvector of A. And $c_1 \mathbf{1}$ is the component of the vector z parallel to $\mathbf{1}$, and $z_\perp$ is the component of z perpendicular to $\mathbf{1}$. Then,

$$\begin{aligned} z^T A z &= (c_1 \mathbf{1} + z_\perp)^T A (c_1 \mathbf{1} + z_\perp) = (c_1 \mathbf{1} + z_\perp)^T (d c_1 \mathbf{1} + A z_\perp) \\ &= d \|c_1 \mathbf{1}\|^2 + c_1 \mathbf{1}^T A z_\perp + 0 + z_\perp^T A z_\perp && \left(\because \quad \mathbf{1}^\top z_\perp = 0\right) \\ &= d \|c_1 \mathbf{1}\|^2 + z_\perp^T A z_\perp && \left(\because \quad \mathbf{1}^\top A z_\perp = \left(A^\top \mathbf{1}\right)^\top z_\perp = d \mathbf{1}^\top z_\perp = 0\right). \end{aligned}$$

Since G is a β-spectral expander,

$$\frac{z_\perp^T A z_\perp}{z_\perp^T z_\perp} \leq \max_{y \perp 1} \frac{y^T A y}{y^T y} \leq \beta. \tag{1}$$

We know that $\langle z, \mathbf{1} \rangle = c_1 \langle \mathbf{1}, \mathbf{1} \rangle + 0$. Therefore, $c_1 = \langle z, \mathbf{1} \rangle / n$. Then,

$$\begin{aligned} \|c_1 \mathbf{1}\|^2 = c_1^2 n = n \frac{1}{n^2} \langle z, \mathbf{1} \rangle^2 = \frac{1}{n} \left(\sum_i z_i \right)^2 &= \frac{1}{n} \left(\sum_{i \in S} 0 + \sum_{i \notin S} z_i \cdot 1 \right)^2 \\ &\leq \frac{1}{n} |\overline{S}| \sum_i z_i^2 && \text{(Cauchy-Schwarz inequality)} \\ &= \frac{1}{n} (n - |S|) \sum_i z_i^2 \leq \frac{1}{2} z^T z && (\because \quad |S| \geq \frac{n}{2}). \end{aligned}$$

Therefore,

$$\|c_1 \mathbf{1}\|^2 \leq \frac{1}{2} \|z\|^2. \tag{2}$$

Therefore,

$$\begin{aligned} z^T A z &\leq d \|c_1 \mathbf{1}\|^2 + \beta \|z_\perp\|^2 = d \|c_1 \mathbf{1}\|^2 + \beta \left(\|z\|^2 - \|c_1 \mathbf{1}\|^2 \right) && \text{(from Equation (1))} \\ &= (d - \beta) \|c_1 \mathbf{1}\|^2 + \beta \left(\|z\|^2 \right) \leq (d - \beta) \left(\frac{1}{2} \|z\|^2 \right) + \beta \|z\|^2 && (\because \quad \beta < d, \text{ using Equation (2)}) \\ &= \frac{d + \beta}{2} \|z\|^2. \end{aligned}$$

Hence,

$$x^T \overline{A} x = z^T A z \leq \frac{d + \beta}{2} \|z\|^2 \leq \frac{d + \beta}{2} \|x\|^2,$$

where the last inequality follows from the definition of z. Since A is a symmetric matrix, $\overline{A}$ is also a symmetric matrix. Recall that $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\overline{A}$ with corresponding eigenvectors $v_1, \ldots, v_n$. Then, for any arbitrary vector $x = c_1 v_1 + \cdots + c_n v_n$, where $c_1, \ldots, c_n$ are non-negative numbers such that $\sum_{i \in [n]} c_i = 1$,

$$R_{\overline{A}}(x) = \frac{x^\top \overline{A} x}{x^\top x} = \frac{\sum_{i \in [n]} c_i^2 \lambda_i}{\sum_{i \in [n]} c_i^2}.$$

Therefore,

$$\frac{d+\beta}{2} \geq \max_x R_{\overline{A}}(x) = \lambda_{\max}.$$

Hence,

$$\Pr\left[\text{Avoiding } S\right] \leq \left(\frac{\lambda_{\max}}{d}\right)^l \leq (0.9)^l \qquad\qquad (\text{using } \beta \leq 0.8d).$$

Therefore, for $l = \mathcal{O}(k)$, we showed that using $R + \mathcal{O}(k)$ bits of randomness, we can reduce the error probability to $\frac{1}{2^k}$. ◻

5.2 Mixing Time of Random Walks on Graphs

In this section we look at the upper bound on mixing time of random walks on graphs.

Theorem 28. *For a d-regular β-spectral expander graph G, for $t = \mathcal{O}\left(\frac{\log(n/\delta)}{\log(d/\beta)}\right)$, we have $d_{TV}\left(\mu^t, \mu^*\right) \leq \delta$.*

Proof. Let μ_0 be the starting distribution over the vertices of the graph. We know that $\mathbf{1}/\sqrt{n}$ is the top eigenvector of A with eigenvalue d, and hence its a stationary distribution μ^* for a random walk on the matrix A/d. Then,

$$\mu^0 = \left\langle \mu^0, \frac{\mathbf{1}}{\sqrt{n}} \right\rangle \frac{\mathbf{1}}{\sqrt{n}} + \mu_\perp^0 = \left(\Sigma_i \mu_i^0\right) \frac{\mathbf{1}}{n} + \mu_\perp^0 = \frac{\mathbf{1}}{n} + \mu_\perp^0.$$

and

$$\begin{aligned}
\mu^t &= \left(\frac{A}{d}\right) \mu^{t-1} && \text{(from Lemma 23)}\\
&= \left(\frac{A}{d}\right)\left(\frac{A}{d}\right) \mu^{t-2} = \cdots = \left(\frac{A}{d}\right)^t \mu^0 = \left(\frac{A}{d}\right)^t \left(\frac{\mathbf{1}}{n} + \mu_\perp^0\right)\\
&= \frac{\mathbf{1}}{n} + \left(\frac{A}{d}\right)^t \mu_\perp^0 && \left(\because \frac{A}{d}\mathbf{1} = \mathbf{1}\right).
\end{aligned}$$

If a graph is β-spectral expander, then

$$\begin{aligned}
\left\|\mu^t - \frac{\mathbf{1}}{n}\right\|_2^2 = \left\|\left(\frac{A}{d}\right)^t \mu_\perp^0\right\|_2^2 &= \left(\mu_\perp^0\right)^T \left(\frac{A}{d}\right)^{2t} \left(\mu_\perp^0\right)\\
&\leq \left(\frac{\beta}{d}\right)^{2t} \left\|\mu_\perp^0\right\|_2^2 && (\because \text{ eigenvalues of } A^k \text{ are } k\text{th powers of eigenvalues of } A)\\
&\leq \left(\frac{\beta}{d}\right)^{2t}.
\end{aligned}$$

The last inequality is because $\left\|\mu_\perp^0\right\|_2 \leq \left\|\mu^0\right\|_2 \leq \left\|\mu^0\right\|_1 = 1$. Then we have,

$$\left\|\mu^t - \frac{\mathbf{1}}{n}\right\|_1 \leq \sqrt{n} \left\|\mu^t - \frac{\mathbf{1}}{n}\right\|_2 \leq \sqrt{n}\left(\frac{\beta}{d}\right)^t \qquad \text{(using Cauchy-Schwarz inequality).}$$

Therefore, $d_{TV}(\mu^t, \mu^*) = \frac{1}{2}\|\mu^t - \mu^*\|_1 \leq \frac{\sqrt{n}}{2}\left(\frac{\beta}{d}\right)^t$. Then for $t \leq \frac{\log(n/\delta)}{\log(d/\beta)}$, we have,

$$t \leq \frac{\log(n/\delta)}{\log(d/\beta)} = \frac{\log(\delta/n)}{\log(\beta/d)}$$

$$\implies t\log(\beta/d) \leq \log(\delta/n) \implies (\beta/d)^t \leq (\delta/n) \implies \frac{\sqrt{n}}{2}(\beta/d)^t \leq \frac{\delta}{2\sqrt{n}} < \delta.$$

◻

Theorem 29. *In terms of the eigenvalues of the adjacency matrix of the d regular graph, the mixing time of a random walk is $t = \mathcal{O}\left(\frac{\log(n/\delta)}{1-\frac{\sigma^*}{d}}\right)$, where $\sigma^* = \max\{|\sigma_2|, |\sigma_n|\}$.*

Proof. Similar to the previous theorem, start with the initial distribution μ_0. Then after t steps of random walk,

$$\left\|\mu^t - \frac{\mathbf{1}}{n}\right\|_2^2 \leq \left(\mu_\perp^0\right)^T \left(\frac{A}{d}\right)^{2t} \left(\mu_\perp^0\right). \tag{3}$$

Since $\mu_\perp^0$ is perpendicular to the eigenvector corresponding to the largest eigenvalue of A, we have that

$$\left(\mu_\perp^0\right)^T \left(\frac{A}{d}\right)^{2t} \left(\mu_\perp^0\right) \leq \left(\frac{\sigma^*}{d}\right)^{2t} \|\mu_\perp^0\|_2^2 \leq \left(\frac{\sigma^*}{d}\right)^{2t},$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ are eigenvalues of the adjacency matrix A and $\sigma^* = \max\{|\sigma_2|, |\sigma_n|\}$. Therefore from Equation (3) we have,

$$\left\|\mu^t - \frac{\mathbf{1}}{n}\right\|_2^2 \leq \left(\frac{\sigma^*}{d}\right)^{2t}.$$

Now using the Cauchy-Schwarz inequality,

$$d_{TV}(\mu^t, \mu^*) = \frac{1}{2}\left\|\mu^t - \frac{\mathbf{1}}{n}\right\|_1 \leq \frac{\sqrt{n}}{2}\left\|\mu^t - \frac{\mathbf{1}}{n}\right\|_2 \leq \sqrt{n}\left(\frac{\sigma^*}{d}\right)^t.$$

Let $\sqrt{n}\left(\frac{\sigma^*}{d}\right)^t \leq \frac{\delta}{\sqrt{n}}$, then,

$$\left(\sigma^*/d\right)^t < \delta/n \implies t\log\left(\sigma^*/d\right) < \log(\delta/n) \implies t < \frac{\log(\delta/n)}{\log\left(\sigma^*/d\right)} \implies t < \frac{\log(n/\delta)}{\log\left(d/\sigma^*\right)}.$$

For any $x > 0, \log\left(\frac{1}{x}\right) \geq 1 - x$. Therefore,

$$t < \frac{\log(n/\delta)}{\log\left(d/\sigma^*\right)} \implies t < \frac{\log(n/\delta)}{1-\frac{\sigma^*}{d}}. \tag{4}$$

$\square$

Note that we can't upper bound this quantity with $\frac{\log(n/\delta)}{\lambda_2}$ since $1 - \frac{|\sigma_n|}{d}$ could be smaller than λ_2 (for eg., for a bipartite graph, $\sigma_n = -d$). Therefore, we use the following random walk matrix instead of A/d.

Lazy Random Walk. To avoid the problem mentioned above, consider the lazy random walk matrix $W = \frac{1}{2}(I + A/d)$. That is, at every vertex v, the random walk stays at the vertex v with probability $1/2$ or chooses one of its neighbours uniformly at random with probability $1/2d$. Let v_i be the eigenvector of A corresponding to the eigenvalue σ_i. Then,

$$Wv_i = \frac{1}{2}\left(I + \frac{A}{d}\right)v_i = \frac{1}{2}\left(v_i + \frac{\sigma_i}{d}v_i\right) = \frac{1}{2}\left(1 + \frac{\sigma_i}{d}\right)v_i.$$

Therefore, v_i is also an eigenvetor of W with eigenvalue $\frac{1}{2}\left(1+\frac{\sigma_i}{d}\right)$, for all $i \in [n]$. Let $\hat{\sigma}_i \stackrel{\text{def}}{=} \frac{1}{2}\left(1+\frac{\sigma_i}{d}\right)$, for all $i \in [n]$. We know that $d \geq \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq -d$. Therefore, $1 \geq \hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \cdots \geq \hat{\sigma}_n \geq 0$. Therefore, the second largest eigenvalue of W in absolute value is always $\hat{\sigma}_2$. We can now relate this to λ_2, the second (smallest) eigenvalue of the normalised Laplacian of the graph G, as follows,

$$\lambda_2 = 1 - \frac{\sigma_2}{d} = 2(1 - \hat{\sigma}_2),$$

to use the Cheeger's inequality. Note that $\mathbf{1}/n$ is a stationary distribution of the random walk matrix W as well. Hence using a similar analysis we can show the following result,

Theorem 30. *For a d-regular graph G with expansion ϕ_G, a lazy random walk on G has mixing time $t = \mathcal{O}\left(\frac{\log(n/\delta)}{\phi_G^2}\right)$, such that $d_{TV}(\mu^t, \mu^*) \leq \delta$.*

Proof-Sketch. Using the lazy random walk matrix W with the second largest eigenvalue $\hat{\sigma}_2$, similar to the result in Theorem 29 we can show that,

$$t < \frac{\log(n/\delta)}{1 - \hat{\sigma}_2}.$$

From Cheeger's inequality we have, $\sqrt{2\lambda_2} \geq \phi_G \implies 4(1 - \hat{\sigma}_2) \geq \phi_G^2$. Therefore,

$$t < \frac{4\log(n/\delta)}{\phi_G^2}.$$

□

References

[AS15] Daniel A. Spielman. Conductance, the normalized laplacian, and cheeger's inequality. `https://www.cs.yale.edu/homes/spielman/561/lect06-15.pdf`, Fall, 2015. Yale - Applied Mathematics 561/ Computer Science 662, Spectral Graph Theory, Fall 2015, Lecture 6.

[AS18] Daniel A. Spielman. Essential spectral theory, hall's spectral graph drawing, the fiedler value. `https://www.cs.yale.edu/homes/spielman/561/lect02-18.pdf`, Fall, 2018. Yale - CPSC 662 / AMTH 561: Spectral Graph Theory, Lecture 2.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, August 2006.

[Mah15] Michael Mahoney. Expanders, in theory and in practice (1 of 2). `https://www.stat.berkeley.edu/~mmahoney/s15-stat260-cs294/Lectures/lecture08-17feb15.pdf`, Spring, 2015. U.C. Berkeley — Stat260/CS294: Spectral Graph Methods, Lecture 8.

[MKO13] Aleksandr M. Kazachkov and Ryan O'Donnell. Expanders. `http://www.cs.cmu.edu/~odonnell/toolkit13/lecture12.pdf`, Fall, 2013. A Theorist's Toolkit (CMU 18-859T, Fall 2013), Lecture 12.

[TL13] Madhur Tulsiani and Wenjie Luo. Lecture 10. `https://ttic.uchicago.edu/~madhurt/courses/toolkit2013/lecture10.pdf`, Spring, 2013. TTIC/CMSC 31150 - Mathematical Toolkit.

[Tre11a] Luca Trevisan. Lecture 10. `https://people.eecs.berkeley.edu/~luca/cs278-08/lecture10.pdf`, 2011. U.C. Berkeley — CS278: Computational Complexity, Handout N10.

[Tre11b] Luca Trevisan. Lecture 6. `http://theory.stanford.edu/~trevisan/cs359g/lecture06.pdf`, 2011. Stanford University — CS359G: Graph Partitioning and Expanders, Handout 6.

[Tre16a] Luca Trevisan. Cheeger's inequalities. `https://people.eecs.berkeley.edu/~luca/expanders2016/lecture03.pdf`, 2016. U.C. Berkeley — CS294: Spectral Methods and Expanders, Lecture 3.

[Tre16b] Luca Trevisan. Cheeger's inequalities contd. `https://people.eecs.berkeley.edu/~luca/expanders2016/lecture04.pdf`, 2016. U.C. Berkeley — CS294: Spectral Methods and Expanders, Lecture 4.

[Tul19] Madhur and Tulsiani. Lecture 10. `https://ttic.uchicago.edu/~madhurt/courses/toolkit2019/l10.pdf`, Autumn, 2019. TTIC 31150/CMSC 31150 - Mathematical Toolkit.