Brouwer Spectral Graph Theory Notes

Evan Hansen

February 16, 2021

1 Introduction

The adjacency matrix A of a simple graph Γ is the matrix where $A_{xy} = 1$ iff $(x, y) \iff x$ has an edge going to y and is 0 otherwise. If Γ is a multigraph then we let this be the number of edges.

Suppose Γ is undirected without loops. Then the incidence matrix of Γ, M, has rows indexed by $V\Gamma$ and columns indexed by $E\Gamma$, and $M_{xe} = 1$ when x is an endpoint of edge e.

The directed incidence matrix N of a directed graph Γ is similarly the matrix where $N_{xe} = -1, 1, 0$ when x is the head, tail, or not on e respectively.

For a graph with no loops, the Laplace matrix, L, has $L_{xy} = -A_{xy}$ for $x \neq y$ and $L_{xx} = d_x$, the degree of x. So the row sums of L are all zero, and if D is the diagonal matrix with $D_{xx} = d_x$ we have $L = D - A$. The signless Laplace matrix Q is simiilarly defined by $Q = D + A$.

Lemma 1. *If Γ is a graph without loops then the associated Laplace matrices satisfy $Q = MM^{\intercal}$ and $L = NN^{\intercal}$ where here N is given by giving any orientation to the edges.*

Proof. For $x \neq y$,

$$(MM^{\intercal})_{xy} = \sum_e [x \in e][y \in e] = A_{xy}$$

which is the number of edges between x and y, i.e. it is A_{xy}. Also

$$(MM^{\intercal})_{xx} = \sum_e [x \in e]^2 = \#\text{edges containing } e$$

so that $Q = MM^{\intercal}$. To see that $L = NN^{\intercal}$ note that everything is the same except that one of the terms in the summand of the first equation will have been made negative. $\square$

Corollary 1. *L and Q are positive semidefinite, and*

$$u^{\intercal} L u = \sum_{(x,y) \in E\ \Gamma} (u_x - u_y)^2$$

and

$$u^{\intercal} Q u = \sum_{(x,y) \in E\ \Gamma} (u_x + u_y)^2$$

noting that the sum is over edges, not the ordered pairs, so that this term is included just once per pair, not twice.

Proof.

$$|(u^{\mathsf{T}}N)_e| = |\sum_x u_x N_{xe}| = |u_x - u_y| \text{ where } e = (x, y)$$

and similarly

$$(u^{\mathsf{T}}M)_e = u_x + u_y$$

. □

The ordinary spectrum of a finite graph Γ is the spectrum (set of eigenvalues) of the adjacency matrix A. Similarly, the Laplace spectrum and signless Laplace spectrums of an (undirected, loop-free) graph is defined as the spectrum of the matrices L and Q.

Let $p_A(\theta) = \det(\theta I - A)$ denote the characteristic polynomial of A, and we also call this the characteristic polynomial of the graph Γ.

Claim 1. *If A is the adjacency matrix of a directed graph Γ without loops, then $p_A(\theta) = \sum_i c_i \theta^{n-i}$ where $c_i = \sum_C (-1)^{c(C)}$ where C ranges over all subgraphs of cardinality i that are disjoint unions of directed cycles, and $c(C)$ is the number of cycles in this decomposition.*

Proof. Recall the Leibniz formula

$$\det M = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) M_{1\sigma(1)} \dots M_{n\sigma(n)}$$

and that $\operatorname{sgn}(\sigma) = (-1)^e$ where e is the number of even cycles in σ so that if σ has $n - i$ fixed points then $\operatorname{sgn}(\sigma) = (-1)^{i+c(\sigma)}$ where $c(\sigma)$ is the number of non-identity cycles of σ.

Now, considering the matrix $\theta I - A$, if σ has $n - i$ fixpoints, we find

$$(\theta I - A)_{1\sigma(1)} \dots (\theta I - A)_{n\sigma(n)} = \left(\prod_{j \text{ is a fixpoint}} (\theta I - A)_{j,j} \right) \left(\prod_{\substack{j_0,\dots,j_{\ell-1} \text{ is a cycle of } \sigma \\ k=0,\dots,\ell-1}} -A_{j_k, j_{k+1} \mod \ell} \right)$$

$$= \begin{cases} (-1)^i \theta^{n-i} & \text{every cycle in the permutation } \sigma \text{ is also a directed cycle of the graph} \\ 0 & \text{otherwise} \end{cases}$$

(1)

and thus

$$\det(\theta I - A) = \sum_C (-1)^{c(C)} \theta^{n-i}$$

as required. □

Now we state the following claim about $p'_A(t)$.

Claim 2. *If A is any matrix with characteristic polynomial $p_A(t) = \det(tI - A)$ then*

$$p'_A(t) = \sum_x \det(tI - A_x)$$

where A_x is the $n-1 \times n-1$ matrix A with the xth row and column deleted.

Proof. Consider the determinant as a function that takes in n vectors in $\mathbb{R}^n$, the columns of the matrix. Then $\det : (\mathbb{R}^n)^n \to \mathbb{R}$ is multilinear, and so by basic calculus we know that

$$\frac{d}{dt} \det(v_1, v_2, \dots, v_n) = \det(\frac{d}{dt}v_1, v_2, \dots, v_n) + \det(v_1, \frac{d}{dt}v_2, \dots, v_n) + \dots + \det(v_1, v_2, \dots, \frac{d}{dt}v_n)$$

2

which is to say that

$$p'_A(t) = \sum_x \det(M_x)$$

where M_x is the $n \times n$ matrix that is the same as $tI - A$ except the xth column is set to zero, except there is a one in the xth row. Since the only non-zero entry in the xth column is in the xth row, which is one, we see, by the Leibniz formula, that $\det(M_x) = \det(tI - A_x)$ □

In particular, setting $t = 0$ gives us a formula between the derivative of the characteristic polynomial at zero and the determinants of the principal submatrices.

If Γ is undirected and simple with n vertices then A is real and symmetric so it has real eigenvalues and the eigenvectors can be taken orthogonal. Since $\operatorname{Tr} A = 0$ the sum of the eigenvalues is zero.

L is also real and symmetric, so the Laplace spectrum is real with orthogonal eigenvectors. L is positive semidefinite and has zero row sums, so is singular, and so its eigenvalues are $0 = \mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$. Also $\operatorname{Tr} L = 2|E| = \sum_i \mu_i$. Q is not necessarily singular but also has real, non-negative eigenvalues.

Γ is regular of degree k exactly when $A\mathbf{1} = k\mathbf{1}$. If Γ is regular of degree k then $tI - A$ is strictly diagonally dominant if $t > k$ so that $|\theta| \leq k$ for all eigenvalues θ. Also, in this case $L = kI - A$ so that if the normal spectrum is $k = \theta_1 \geq \ldots \geq \theta_n$ and the Laplace eigenvalues are $0 = \mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$, we find $\mu_i = k - \theta_i$ for all i (which justifies $\theta_1 = k$) and similarly the eigenvalues of $Q = kI + A$ are $2k, k + \theta_2, \ldots, k + \theta_n$.

The complement of a graph Γ, $\overline{\Gamma}$, is the graph with the complement edge set. Then if $\overline{A}$ is the adjacency matrix of $\overline{\Gamma}$, we see that $\overline{A} = J - I - A$ and similarly $\overline{L} = nI - J - L$.

Note that because $\mathbf{1}$ and any vector $v \perp \mathbf{1}$ are eigenvectors of J, and because L has rowsums zero, every eigenvector of L is also an eigenvector of J. It follows that the eigenvalues of $\overline{L}$ are $0, n - \mu_n, \ldots, n - \mu_2$. In particular, we see that $\mu_n \leq n$ for all graphs.

If Γ is regular with degree k and normal spectrum $k = \theta_1 \geq \ldots \geq \theta_n$ then we similarly find that the complement $\overline{\Gamma}$ has normal spectrum $n - k - 1, -1 - \theta_n, \ldots, -1 - \theta_2$.

Claim 3. $(A^\ell)_{xy}$ *is the number of walks of length ℓ from $x \rightsquigarrow y$. As a consequence, $(A^2)_{xx}$ is the degree of x and $\operatorname{Tr} A^2$ is twice the number of edges.*

Proof. Just look at the definition of matrix multiplication. □

Claim 4. *If Γ is undirected then its (normal, signed and unsigned Laplace) eigenvalues are zero if and only if it has no edges.*

More generally,

Claim 5. *If Γ has diameter d then Γ has at least $d + 1$ distinct eigenvalues for $A, L,$ and Q.*

Proof. Proof by contradiction. Let M be any nonnegative, real symmetric matrix with $M_{xy} > 0 \iff (x, y) \in E$, except possibly $M_{xx} > 0$ even if $(x, x) \notin E$. Suppose there are only $t \leq d$ distinct eigenvalues, $\theta_1, \ldots, \theta_t$. Then

$$(M - \theta_1 I) \ldots (M - \theta_t I) = 0$$

so that M^t is a linear combination of $I, M, \ldots, M^{t-1}$. By similar logic as in Claim 3, $(M^\ell)_{xy} > 0 \iff$ there is a path of length ℓ $x \rightsquigarrow y$. Thus the statement above says that for every pair of vertices (x, y), if there is a path $x \rightsquigarrow y$ of length t then there is also a path of length at most $t - 1$. But this shows there are no vertices x, y with distance exactly $t \leq d$ between them, which also shows that the diameter of the graph is $< d$, contradicting our assumption. □

3

Definition 1. *The (x, y)-cofactor a matrix M is defined as $\det M(x, y)$ where $M(x, y)$ is the matrix M but with row x and column y deleted.*

We will want this in the proof of the next theorem.

Claim 6. *If L is any $n \times n$ matrix over any field with characterstic that does not divide n and zero row and column sums, then all its cofactors are equal.*

Proof. Compute $\det(L + J)$. Do the following row and column operations first.

1. Add all the rows the the first.

2. Add all the columns to the first.

3. Then the $(1, 1)$ entry is now n^2, the first row is otherwise n, and the first column is all n.

4. Divide the first row by n.

5. Subtract the first row from all other rows.

6. Then the $(1, 1)$ entry is now n, the rest of the first row is 1, the rest of the first column is zero, and all other entries are the same as the principal submatrix $L(1, 1)$ where its first row and column are removed.

Since row and column operations preserve the determinant, we see that $\det(L + J) = n \det(L(1, 1))$. Now replace "first row" and "first column" with x^{th} row and y^{th} column to see that $\det(L + J) = n \det(L(x, y))$ for all x, y. $\qquad\square$

The following theorem gives an interesting way to compute the number of spanning trees given the Laplace spectrum or Laplace matrix.

Theorem 1. ***Kirchoff's Matrix Tree Thoerem*** *Let Γ be an undirected (multi-)graph with at least one vertex and Laplace matrix L. Let the Laplace spectrum be $0 = \mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$. Let ℓ_{xy} be the (x, y)-cofactor of L. Then the number of spanning trees of Γ, N, satisfies*

$$N = \ell_{xy} = \det\left(L + \frac{1}{n^2} J\right) = \frac{1}{n} \mu_2 \ldots \mu_n \quad \forall x, y \in V\ \Gamma$$

Proof. Let L^S denote L with all the rows and columns indicated by the set S removed. So $\ell_{xx} = \det L^{\{x\}}$.

Proof by induction on n. If $n = 1$ then $\ell_{xx} = 1$ ($L^{\{x\}}$ is the empty matrix). If x has degree 0 then $\ell_{xx} = 0$ since $L^{\{x\}}$ has zero row sums (being the Laplacian of another graph) and this would also show that the graph Γ is not connected, so that it has no spanning trees. If instead xy is an edge to another node $y \neq x$, then consider the Leibniz formula for the determinant to see that removing the edge xy causes ℓ_{xx} to decrease by $\det L^{\{x, y\}}$ (this is because the only entry of $L^{\{x\}}$ that is effected is (y, y)). On the other hand, we note that the matrix $L^{\{x, y\}}$ is the same as the Laplace matrix of Γ/xy (contraction over the edge xy) with the node $\{x, y\}$ corresponding to the contraction having had its row/column deleted. Thus by the inductive hypothesis, $\det L^{\{x, y\}} = \#$ spanning trees in Γ/xy, which is also the number of spanning trees that contain the edge xy. Summing over y gives the result $N = \ell_{xx}$.

Now, because the eigenvectors of L can be taken to be orthogonal, $p_L(t) = \det(tI - L) = t \prod_{i=2}^{n}(t - \mu_i)$. We see that the coefficient of t in the polynomial must be $(-1)^{n-1} \mu_2 \ldots \mu_n$, which

4

must be its derivative at $t = 0$. By setting $t = 0$ in Claim 2 we see that $(-1)^{n-1}\mu_2 \ldots \mu_n = p'_L(0) = \sum_x \det(0I - L^{\{x\}}) = (-1)^{n-1} \sum_x \ell_{xx}$, and hence $\mu_2\mu_3 \ldots \mu_n = \sum_x \ell_{xx} = nN$.

Now because L is a matrix with zero row and column sums, Claim 6 shows that $\ell_{xy} = \ell_{xx}$ for all x, y.

Finally, because L and J are simultaneously diagonalizable, we see that $\det(L + \frac{1}{n^2}J)$ is the product of the eigenvalues, i.e. it is $\frac{1}{n}\mu_2 \ldots \mu_n$.

$\square$

Applying this to K_n gives a proof of the well-known formula n^{n-2} for the number of spanning trees on n (labeled) nodes.

The Matrix Tree Theorem can also be proven from the Cauchy Binet formula very directly.

It is easy to see a graph is bipartite iff its adjacency matrix $A = \begin{bmatrix} 0 & B \\ B^\mathsf{T} & 0 \end{bmatrix}$ for some matrix B. It follows that if $\begin{pmatrix} u \\ v \end{pmatrix}$ is an eigenvector with eigenvalue θ then $\begin{pmatrix} u \\ -v \end{pmatrix}$ is an eigenvector with eigenvalue $-\theta$ so that the spectrum is symmetric w.r.t. 0. It turns out the converse also holds, which will follow from the Perron-Frobenius Theorem. Because the row and column ranks are the same, $\mathrm{rk}\, A = 2\,\mathrm{rk}\, B$. If X_1, X_2 is the partition of the biparite graph required by the definition and $n_i = |X_i|$ then we see that $\mathrm{rk}\, B \leq \min(n_1, n_2)$ so that $\mathrm{rk}\, A \leq 2\min(n_1, n_2)$, and hence Γ has eigenvalue 0 with multiplicity at least $|n_1 - n_2|$.

Note that bipartiteness cannot be detected by looking at the Laplace or signless Laplace spectrum. For instance, $K_{1,3}$ and $K_1 + K_3$ have the same signless Laplace spectrum but only one is bipartite. However, we will see that the Laplace and signless Laplace spectra are equal iff Γ is bipartite.

Claim 7. *The spectrum (normal, signed or unsigned Laplace) of a graph Γ is simply the union of those of its connected components.*

Proof. Order the vertices so that the components form blocks on the diagonal. $\square$

Claim 8. *The multiplicity of 0 in the Laplace spectrum equals the number of connected components.*

Proof. By the last claim, it suffices to show that every graph Laplacian L of a connected graph has the eigenvalue 0 with multiplicity 1. First, since L has 0 row sums, $L\mathbf{1} = 0$, so that $\mathbf{1}$ is one such eigenvector. On the other hand, because $u^\mathsf{T} L u = \sum_{xy \in E}(u_x - u_y)^2$ we see that if $u^\mathsf{T} L u = 0$ then u has the same value on the endpoints of each edge, and hence is constant on the entire graph because it is connected. $\square$

Claim 9. *If Γ is undirected and regular of degree d then d is the largest eigenvalue of Γ and its multiplicity equals the number of connected components.*

Proof. Apply the last claim and note that $L = dI - A$. $\square$

Claim 10. *The multiplicity of 0 in the signless Laplace spectrum is the number of bipartite connected components.*

Proof. It suffices to show this in the connected case. So, assuming Γ is connected, we show that if it has 0 as a signless Laplace eigenvalue it is bipartite, and also that if a graph is bipartite then it has 0 as a signless Laplace eigenvalue of multiplicity 1.

If $u^\mathsf{T} Q u \sum_{xy \in E}(u_x + u_y)^2 = 0$ then $\forall xy \in E, u_x = -u_y$. By connectedness, we see that this implies there is a constant c and a partition of the vertex set V into $X_1 \sqcup X_2$ such that $u_x = c \iff$

$x \in X_1$, $u_y = -c$ $\iff$ $y \in X_2$, and there are X_1, X_2 is the partition required to show the graph is bipartite. I.e. all such eigenvectors are multiples of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ (where the top half is X_1 and bottom X_2). This shows that if a connected graph has a signless Laplace eigenvalue 0 then it is bipartite and the multiplicity of 0 is 1.

On the other hand, given such a partitioning X_1, X_2 we can construct this vector, and it is easy to see that it will have eigenvalue 0, so that all bipartite graphs have this 0 eigenvalue. A visual way to see this is to write

$$Q = \begin{bmatrix} D_1 & A_2 \\ A_1 & D_2 \end{bmatrix}$$

where we have ordered the basis to do X_1 first and X_2 second. Then for any row, the row sum in D_1 minus that in A_2 will be zero, and similarly for A_1 and D_2. $\square$

Claim 11. *The signed and signless Laplace spectra are equal iff the graph is bipartite.*

Proof. Suppose Γ is biparite, and write

$$Q = \begin{bmatrix} D_1 & A_2 \\ A_1 & D_2 \end{bmatrix}$$

as in the proof of the last claim. Then

$$\begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} D_1 & A_2 \\ A_1 & D_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} = \begin{bmatrix} D_1 & -A_2 \\ -A_1 & D_2 \end{bmatrix} = L$$

So that Q and L are similar by the matrix

$$D = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} = D^{-1}$$

and hence have the same spectrum.

On the other hand, if Q and L have the same spectrum, then, in particular, the multiplicity of 0 is the same in both so that by the past few claims we see the number of components equals the number of bipartite components. $\square$

Now we compute the spectra of some special graphs. When discussing an eigenvalue, we will write its multiplicity as an exponent. For instance, we say "the spectrum of J is $n^1, 0^{n-1}$".

K_n, the complete graph on n vertices, has adjacency matrix $A = J - I$, and hence spectrum $(n-1)^1, (-1)^{n-1}$. $L = nI - J$ has spectrum $0^1, n^{n-1}$. $Q = (n-2)I + J$ has spectrum $(n-2)^1, (n-1)^{n-1}$.

The complete bipartite graph $K_{m,n}$ has adjacency matrix $A = \begin{bmatrix} 0 & J \\ J & 0 \end{bmatrix}$ with spectrum $\pm\sqrt{mn}, 0^{m+n-2}$. Here is how you see this. First off, we can write

$$A = \begin{bmatrix} 0 & J \\ J & 0 \end{bmatrix}$$

where the first block width and height is n and the other m. Then a vector $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ has $Av = 0$ iff $Jv_1 = 0$ and $Jv_2 = 0$. Then since we know that J (with width n and height m) has spectrum $0^{m-1}, n$. So by unioning the bases we construct a basis for the $m + n - 2$ dimensional 0 eigenspace.

Next, the obvious things to try for other eigenvectors are something of the form $v = \begin{pmatrix} a\mathbf{1} \\ b\mathbf{1} \end{pmatrix}$. By rescaling we can assume $a = 1$. Suppose v has eigenvalue λ. Then we must have $\lambda = bm$ and $\lambda b = n$. We find that $b = \pm\sqrt{n/m}$ is a solution with associated eigenvalues $\lambda = \pm\sqrt{mn}$.

The Laplace spectrum of $K_{n,m}$ is $0^1, m^{n-1}, n^{m-1}, (m+n)^1$.

Proof. Write out

$$L = \begin{bmatrix} mI & -J \\ -J & nI \end{bmatrix}$$

Now take any vector in the kernel of A which has its last m components equal to zero, i.e. a vector $v = \begin{pmatrix} v_1 \\ 0 \end{pmatrix}$ such that the sum of the entries of v_1 is zero, and see that v will be an eigenvector with eigenvalue m. Symmetrically, we see that a vector $v = \begin{pmatrix} 0 \\ v_2 \end{pmatrix}$ where v_2 has zero sum is an eigenvector with eigenvalue n. This gives us the m^{n-1}, n^{m-1} part of the spectrum. We already knew that 0 should be an eigenvalue with multiplicity 1 because the graph is connected (corresponding to the eigenvector $\mathbf{1}$). Now we just want to find an eigenvector corresponding to the $m+n$ part. We need to pick a vector that is orthogonal to all of the previous eigenvectors we found. We can do this by setting $u = \begin{pmatrix} \mathbf{1} \\ (-n/m)\mathbf{1} \end{pmatrix}$. This works because the 0 eigenvector's positive dot product with the first half cancels with the second, and all the other vectors are 0 on one part and have row sum zero on the other. Plugging in we compute

$$\begin{bmatrix} mI & -J \\ -J & nI \end{bmatrix} \begin{pmatrix} \mathbf{1} \\ (-n/m)\mathbf{1} \end{pmatrix} = \begin{pmatrix} (m+n)\mathbf{1} \\ -(n+n^2/m)\mathbf{1} \end{pmatrix}$$

and because $(-n/m)(n+m) = -(n+n^2/m)$ this is indeed an eigenvector with eigenvalue $n+m$. $\square$

Because the graph is bipartite, the signless Laplace spectrum is identical with the signed Laplace spectrum.

The directed n-cycle D_n has eigenvectors $\begin{pmatrix} 1 & \zeta & \zeta^2 & \dots & \zeta^{n-1} \end{pmatrix}^\mathsf{T}$ with associated eigenvalue ζ where ζ is any n^{th} root of unity. Thus the spectrum is the set of nth roots of unity.

Now consider the undirected n-cycle C_n. Letting B be the adjacency matrix of D_n and A that of C_n, we see that $A = B + B^\mathsf{T}$. Because $B^\mathsf{T} = B^{-1}$ we see that A has the same eigenvectors, now with eigenvalues $\zeta + \zeta^{-1}$. When $\zeta = e^{2\pi ij/n}$, $\zeta + \zeta^{-1} = 2\cos(2\pi j/n)$ so that this, for any $j \in \{0, \dots, n-1\}$ are the eigenvalues of C_n. Since these graphs are regular, we get the Laplace spectrums for free. The Laplace spectrum of C_n is $\{2 - 2\cos(2\pi j/n) : j = 0, \dots, n-1\}$ and the Laplace spectrum of D_n is $\{1 - \zeta : \zeta \text{ is an } n^{th} \text{ root of unity}\}$.

Now consider the undirected path P_n on n vertices. The ordinary spectrum is $2\cos(\pi j/(n+1))$ for $j = 1, \dots, n$. The Laplace spectrum is $2 - 2\cos(\pi j/n)$.

Proof. Consider C_{2n+2}. Let $u(\zeta) = \begin{pmatrix} 1 & \zeta & \dots & \zeta^{2n+1} \end{pmatrix}^\mathsf{T}$ for ζ a $(2n+2)^{th}$ root of unity. We note that $u(\zeta)$ and $u(\zeta^{-1})$ are both eigenvectors of C_{2n+2} with eigenvalue $\zeta + \zeta^{-1} = 2\cos(\pi j/(n+1))$. Thus $v(\zeta) = u(\zeta) - u(\zeta^{-1})$ is also an eigenvector with this eigenvalue. Additionally, we see that $v(\zeta)$ is zero at two points that are separated by n points on either side of the cycle, where it is non-zero and assumes the same values, (assuming $\zeta \neq \pm 1$). Then for $\zeta \neq \pm 1$, $v(\zeta)$ induces a vector on P_n given by removing the two points where it is zero and looking at the values on the two copies of P_n leftover. Becuase $v(\zeta)$ was an eigenvector of C_{2n+2} with eigenvalue $2\cos(\pi j/(n+1))$, we see that these induced vectors are eigenvectors of P_n with the same eigenvalue. The induced

vector is $\begin{pmatrix} \zeta - \zeta^{2n+1} & \dots & \zeta^j - \zeta^{2n+2-j} & \dots & \zeta^n - \zeta^{n+2} \end{pmatrix}^{\mathsf{T}}$ with eigenvalue $2\cos(\pi j/(n+1))$ ($j \neq 0, n+1$). $\qquad\square$

The key part of the analysis seems like a special case of this more general claim.

Claim 12. *Suppose that Γ is any graph and that u is an eigenvector of Γ with eigenvalue θ. Now let $S = \{v \ : \ u_v = 0\}$ and suppose that $\Gamma^{\dagger}$ is a connnected component of $G - S$. Then $u|_{\Gamma^{\dagger}}$ is an eigenvector of $\Gamma^{\dagger}$ with the same eigenvalue θ.*

Proof. Any node v outside of but adjacent to $\Gamma^{\dagger}$ has $u_v = 0$, so that it's like it doesn't exist. For the sake of computing Av. Any other node cannot effect $Av|_{\Gamma^{\dagger}}$ $\qquad\square$

The line graph $L(\Gamma)$ has as vertex set the edge set of Γ, and two nodes are adjacent if the corresponding edges share an endpoint. Letting M be the incidence matrix of Γ, $M^{\mathsf{T}}M - 2I$ is the adjacency matrix of $L(\Gamma)$.

Proof. $(M^{\mathsf{T}}M)_{ij} = \sum_k M^{\mathsf{T}}_{ik}M_{kj} = \sum_k M_{ki}M_{kj} =$ the number of endpoints that the edges i and j share. Subtracting $2I$ removes the 2's on the diagonal that correspond to the fact that any edge shares two endpoints with itself. $\qquad\square$

Claim 13. *Let A be any matrix. Then the non-zero eigenvalues of $A^{\mathsf{T}}A$ and AA^{T} are the same, including multiplicities. As a consequence, any additional eigenvalues of AA^{T} are zero. Specifically, if v is an eigenvector of $A^{\mathsf{T}}A$ with eigenvalue λ then Av is an eigenvector of AA^{T} with the same eigenvalue λ.*

Proof. It suffices to show that every non-zero eigenvalue of $A^{\mathsf{T}}A$ is an eigenvalue of AA^{T}. The opposite direction follows by symmetry. If v is an eigenvector with eigenvalue λ, then

$$(AA^{\mathsf{T}})(Av) = A(A^{\mathsf{T}}Av) = \lambda Av$$

Secondly, if v is not a multiple of w and both are eigenvectors with eigenvalue λ then Av is not a multiple of Aw because then $A^{\mathsf{T}}(Av) = \lambda v$ would be a multiple of $A^{\mathsf{T}}(Aw) = \lambda w$. Thus the transformation $x \mapsto Ax$ preserves multiplicities. $\qquad\square$

Claim 14. *Suppose Γ has m edges and let $\rho_1 \geq \dots \geq \rho_r$ be the positive signless Laplace eigenvalues of Γ. Then the eigenvalues of $L(\Gamma)$ are $\theta_i = \rho_i - 2$ for all $i \in [r]$, and $\theta_i = -2$ for all $i > r$.*

Proof. The signless Laplace matrix Q of Γ and the adjacency matrix B of $L(\Gamma)$ satisfy $Q = MM^{\mathsf{T}}$ and $B + 2I = M^{\mathsf{T}}M$. By the last claim the non-zero eigenvalues of $M^{\mathsf{T}}M$ and MM^{T} are the same, and the claim follows. $\qquad\square$

Corollary 2. *The signed and signless Laplace eigenvalues of P_n (the path on n vertices) are $2 + 2\cos(\pi i/n)$ for $i \in [n]$.*

Proof. Use the last claim and our prior derivation of the normal spectrum of P_n to get the result on the signless eigenvalues. Use bipartiteness to see the signed eigenvalues are the same. $\qquad\square$

Corollary 3. *If Γ is d-regular ($d \geq 2$) with n vertices and $m = kn/2$ edges, and has eigenvalues θ_i for $i \in [n]$, then $L(\Gamma)$ is $(2d-2)$-regular with eigenvalues $\theta_i + k - 2$ for $i \in [n]$ and -2 for the other $m - n$ eigenvalues.*

Proof. Use our prior relation $\mu_i = \theta_i + k$ for the signless Laplace eigenvalues of Γ. Then use the claim to find the normal eigenvalues of $L(\Gamma)$. $\qquad\square$

Definition 2. *The graph $L(K_n)$ is known as the* triangular graph *and denoted $T(n)$.*

Corollary 4. *$T(n)$ has spectrum $2(n-2)^1, (n-4)^{n-1}, (-2)^{n(n-3)/2}$.*

Definition 3. *The graph $L(K_{m,m})$ for $m \geq 2$ is known as the* lattice graph *$L_2(m)$.*

Corollary 5. *The spectrum of $L_2(m)$ is $2(m-1)^1, (m-2)^{2m-2}, (-2)^{(m-1)^2}$.*

Both $T(n)$ and $L_2(m)$ are families of *strongly regular graphs* which will be the subject of chapter 9. The complement of $T(5)$ is the *Petersen graph*, with spectrum $3^1, 1^5, (-2)^4$.

Definition 4. *Given Γ and Δ graphs on vertex sets V and W, the* Cartesian product *$\Gamma \square \Delta$ is defined as the graph on vertex set $V \times W$ where $(v, w) \sim (v', w')$ whenever $v = v'$ and $w \sim w'$ or $w = w'$ and $v \sim v'$. We have $A_{\Gamma \square \Delta} = A_\Gamma \otimes I + I \otimes A_\Delta$.*

Claim 15. *If u and v are eigenvectors for Γ and Δ with ordinary or Laplace eigenvalues θ and η, then w, defined by $w_{(x,y)} = u_x v_y$ (or equivalently $w = u \otimes v$) is an eigenvector of $\Gamma \square \Delta$ with ordinary or Laplace eigenvalue $\theta + \eta$.*

Example 1. *$L_2(m) = K_m \square K_m$.*

Proof. Let the vertex set of $K_{m,m}$ be $L \sqcup R$. Then $L_2(m)$ is the graph on $L \times R$ where $(v, w) \sim (v', w')$ exactly when one of the coordinates are equal, i.e. it is $K_m \square K_m$. $\square$

Definition 5. *The n-dimensional* boolean hypercube, *denoted Q_n, is the graph on the vertex set $\{-1, 1\}$ (or sometimes $\{0, 1\}$) where $v \sim w$ if v and w differ in exactly one coordinate.*

Example 2. *$Q_n = K_2 \square \ldots \square K_2$ (n times) and so since the spectrum of K_2 is $-1^1, 1^1$, the spectrum of Q_n is $n - 2i$ with multiplicity $\binom{n}{i}$ for $i \in \{0, \ldots, n\}$.*

Definition 6. *If again Γ and Δ are graphs with vertex sets V and W then their* Kronecker product *(also* direct product *or* conjunction*) $\Gamma \otimes \Delta$ is the graph on vertex set $V \times W$ defined by $(v, w) \sim (v', w')$ if $v \sim v'$ and $w \sim w'$. It has adjacency matrix $A = A_\Gamma \otimes A_\Delta$. It follows that for v and w eigenvectors of Γ and Δ with eigenvalues θ and η, $v \otimes w$ is an eigenvector of $\Gamma \otimes \Delta$ with eigenvalue $\theta \eta$*

Definition 7. *The strong product $\Gamma \boxtimes \Delta$ is defined by $(v, w) \sim (v', w')$ whenever $v = v'$ or $v \sim v'$ and $w = w'$ or $w \sim w'$. The adjacency matrix is $A = ((A_\Gamma + I) \otimes (A_\Delta + I)) - I$. It follows the eigenvalues are $(\theta + 1)(\eta + 1) - 1$ where θ and η are some eigenvalues of Γ and Δ.*
Note that the edge set of $\Gamma \boxtimes \Delta$ is simply the union of the edgesets of $\Gamma \square \Delta$ and $\Gamma \otimes \Delta$.

Definition 8. *If G is an abelian group and $S \subseteq G$ then the* Cayley graph *on G with difference set S is the (directed) graph Γ with vertex set G and edge set $E = \{(x, y) : y - x \in S\}$. Γ is regular with degree $|S|$. The graph is undirected when $S = -S$.*

Claim 16. *The spectrum of a finite Cayley graph is as follows. Let χ be a character of G, i.e. a homomorphism from $G \to \mathbb{C}^\times$. Then χ is an eigenvector of the Cayley graph on G with any difference set, and its eigenvalue is $\chi(S) = \sum_{s \in S} \chi(s)$. Furthermore, there are $n = |G|$ such characters, and they are an orthogonal basis when viewed as vectors in $G^\mathbb{C}$, so that this accounts for all the eigenvectors.*

Proof. The last claim is known from algebra (and is non-trivial). For the other part, compute $\sum_{y \sim x} \chi(y) = \sum_{s \in S} \chi(y + s) = \chi(S)\chi(y)$ $\square$

Here are some example applications of this theory

Example 3. *Writing K_{10} as the sum of Petersen Graphs*
Can the edges of K_{10} be decomposed into the union of three copies of the Petersen graph? K_{10} has the correct number of vertices and edges, so it seems possible. But it is not.

Suppose that P_1, P_2, P_3 are the adjacency matrices of these copies of the Petersen graph, such that $P_1 + P_2 + P_3 = J - I$. We know that P_1 and P_2 will have eigenvectors of eigenvalue 1 with multiplicity 5, and that these eigenvectors lie in the space perpendicular to $\mathbf{1}$. But this space is only of dimension 9, so there is some eigenvector $u \perp \mathbf{1}$ that is of eigenvalue 1 for both P_1 and P_2. Then we find that $P_3 u = (J - I)u - P_1 u - P_2 u = -3u$, but -3 is not an eigenvalue of the Petersen graph, so that P_3 is not the adjacency matrix of the Petersen graph. In fact, as we will see from the Perron-Frobenius theorem, because the Petersen graph is regular of degree 3, this would imply that P_3 would have to be bipartite.

Claim 17. *Due to Wittenhausen; Graham & Pollak*
Suppose Γ, with adjacency matrix A, has an edge decomposition into r complete bipartite graphs. Then $r \geq n_+(A)$ and $r \geq n_-(A)$ where $n_+(A)$ and $n_-(A)$ are the numbers of positive and negative eigenvalues of A respectively.

Proof. Let the ith complete bipartite graph have a partition into left and right sides such that u_i and v_i are the characteristic vectors of the left and right sides. I.e. the adjacency matrix of the ith complete bipartite graph is $u_i v_i^\intercal + v_i u_i^\intercal$ and $A = \sum_i D_i$.

Then let w be a vector orthogonal to all u_i, there are $n - r$ dimensions of such vectors because this is the space perpendicular to the span of the r u_i vectors. Then $w^\intercal A w = 0$ so that w does not lie in the span of the positive eigenvectors or the span of the negative eigenvectors. Thus $n_\pm(A) \leq r$ because otherwise there would be a non-zero intersection of the subspace spanned by the all positive (negative) eigenvalue eigenvectors and the space perpendicular to all u_i. $\square$

Corollary 6. *K_n has -1 as an eigenvalue with multiplicity $n - 1$ so it cannot be decomposed into any fewer than $n - 1$ complete bipartite graphs.*

Definition 9. *An* isomoprhism *$\pi : \Gamma \to \Gamma'$ is a function from the vertex set V of Γ to the vertex set V' of Γ' such that $\pi(x) \sim \pi(y) \iff x \sim y$. If $\Gamma = \Gamma'$ then π is called an* automorphism. *The automorphisms of a graph Γ form a group under composition, called the* automorphsim group *of Γ and denoted* $\mathrm{Aut}\,\Gamma$.

Given an automorphism of Γ, π, we can define the linear transformation on $\mathbb{R}^V$ given by $(P_\pi(u))_x = u_{\pi(x)}$. The statement that π is an automorphism is equivalent to $AP_\pi = P_\pi A$. It follows that if $Av = \lambda v$ then $AP_\pi v = P_\pi A v = \lambda P_\pi v$ so that P_π has the eigenspace V_λ has an invariant subspace for each eigenvalue λ of A. Also note that P_π is unitary because it is just a permutaiton matrix.

More generally, if $G \leq \mathrm{Aut}\,\Gamma$ then G has a linear representation (an embedding into a group of linear transformations) of degee $m(\theta) = \dim V_\theta$.

One might expect that when $\mathrm{Aut}\,\Gamma$ is large $m(\theta)$ will be large as well to accomodate this embedding, so that Γ has few distinct eigenvalues.

Definition 10. *An eigenvalue θ that occurs with geometric multiplicity 1 is called* simple.

If θ is a simple eigenvalue with eigenvector u and π an automorphism of Γ, then $P_\pi u = \pm u$ because $\langle u \rangle$ is the only place it can send it and it is unitary.

Corollary 7. *If all eigenvalues are simple, then $\mathrm{Aut}\,\Gamma$ is a 2-group (i.e. all elements have order 2).*

Corollary 8. *Let* $\operatorname{Aut}\Gamma$ *be transitive on* V *(the vertex set of* Γ*). I.e. for all* $v, w \in V$, *there exists some automorphism* π *such that* $\pi(v) = w$. *Then* Γ *is regular, say of degree* d *(since the graph must "look the same" locally at each vertex). Further,* ... *some other stuff I don't really care about.*

Definition 11. *Let* Γ *be a graph with at least two vertices. The second smallest Laplace eigenvalue* $\mu_2(\Gamma)$ *is called the* algebraic multiplicity *of* Γ.

Recall that $\mu_2(\Gamma) \geq 0$ iff Γ is connected. Furthermore, the algebraic connectivity is monotone, it can only increase when edges are added, as shown in this claim.

Claim 18. *If* Γ *and* Δ *are two edge-disjoint graphs on the same set of vertices, then*

$$\mu_2(\Gamma \cup \Delta) \geq \mu_2(\Gamma) + \mu_2(\Delta) \geq \mu_2(\Gamma)$$

Proof. Recall that $\mathbf{1}$ is a 0-eigenvector, so that

$$\mu_2(\Gamma) = \min_{\substack{u \perp \mathbf{1} \\ |u|=1}} u^{\mathsf{T}} L u$$

Letting u_Γ and u_Δ be the eigenvectors achieving this minimum for Γ and Δ we see that

$$\mu_2(\Gamma \cup \Delta) = \min_{\substack{u \perp \mathbf{1} \\ |u|=1}} u^{\mathsf{T}} L u = u^{\mathsf{T}} L_\Gamma u + u^{\mathsf{T}} L_\Delta u \geq u_\Gamma^{\mathsf{T}} L_\Gamma u_\Gamma + u_\Delta^{\mathsf{T}} L u_\Delta = \mu_2(\Gamma) + \mu_2(\Delta)$$

$\square$

Definition 12. *The* vertex connectivity *of a graph* Γ *on vertex set* V *is the minimum cardinality of a set* D *such that the induced graph on* $V - D$ *is disconnected.*

Claim 19. *Algebraic connectivity is a lower bound for vertex connectivity.*

Proof. By monotonicity we may assume that Γ contains all possible edges between D and $V - D$.

Then note any non-zero vector u that is zero on D and constant on each component of the graph induced and satisfies $u \perp \mathbf{1}$ is a Laplace eigenvector and it has eigenvalue $|D|$.

This is easy to see by writing out the Laplacian as a block matrix, where the first block corresponds to D and the latter $V - D$. We have

$$L = \begin{bmatrix} L_D & -J \\ -J & L_{V-D} + DI \end{bmatrix}$$

Where L_{V-D} is the Laplacian of the graph induced on $V - D$, and L_D is just whatever needs to be there. So if $u = \begin{pmatrix} 0 \\ v \end{pmatrix}$ then

$$Lu = \begin{bmatrix} L_D & -J \\ -J & L_{V-D} + DI \end{bmatrix} \begin{pmatrix} 0 \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ |D|v \end{pmatrix} = |D|u$$

where the zero in the result is because $u \perp \mathbf{1} \implies v \perp \mathbf{1}$ and the $|D|v$ is because $L_{V-D}v = 0$ because v is constant on the components of the induced graph on $V - D$. Hence $\mu_2(\Gamma) \leq |D|$ as desired. $\square$

Definition 13. *As noted when we discussed bipartiteness, non-isomorphic graphs can still have the same spectrum. Graphs with the same (normal) spectrum are called* cospectral. *If they have the same spectrum for some other matrix, you can just say they are cospectral for that matrix.*

For $n < 4$, the hypercube Q_n is uniquely determined by its spectrum, but not for $n \geq 4$. Indeed, Hoffman showed there are exactly two graphs with spectrum $4^1, 2^4, 0^6, (-2)^4, (-4)^1$. The graph that is not Q_n is called the Hoffman graph. I did not understand the description of the graph given in the book.

Definition 14. *The* Seidel adjacency matrix *S of a graph Γ with adjacency matrix A is defined by*

$$S_{uv} = \begin{cases} 0 & u = v \\ -1 & u \sim v \\ 1 & u \not\sim v \end{cases}$$

so that $S = J - I - 2A$. The spectrum of S is called the Seidel spectrum *of the graph Γ.*

Claim 20. *If Γ is d-regular then its Seidel spectrum consists of $n - 1 - 2d$ and $-1 - 2\theta$ where θ ranges through the (ordinary) eigenvalues of Γ other than d.*

Proof. Since Γ is d-regular, it has $\mathbf{1}$ as an eigenvector with eigenvalue d. Then $\mathbf{1}$ is an eigenvector of S with eigenvalue $n - 1 - 2d$. All other eigenvectors of A are orthogonal to $\mathbf{1}$, so are also eigenvectors of J (with eigenvalue 0) and I, and so the vector with eigenvalue θ (w.r.t. A) will have eigenvalue $-1 - 2\theta$ (w.r.t. S) as desired. $\qquad\square$

Now, let Γ have vertex set V, and let $Y \subseteq V$. Let D be the square matrix where $D_{xx} = -1$ for $x \in Y$ and $D_{xx} = 1$ otherwise. Then DSD has the same spectrum as S (because they are similar). It is the Seidel adjacency matrix of the graph obtained from Γ obtained by leaving edges as they were inside Y and $V - Y$, and flipping adjacency and non-adjacency between nodes in Y and $V - Y$.

Definition 15. *The graph obtained by this method is said to be obtained by* Seidel switching *w.r.t. the set Y.*

Being related by Seidel switching is an equiv. relation, and the classes are called *switching classes*.

Note that the Seidel matrix of $\overline{\Gamma}$ is $-S$, so that taking complements flips the Seidel eigenvalues.

Note that by our earlier Claim 20, two regular graphs that are Seidel cospectral will also be cospectral. This also gives a way of finding cospectral graphs by performing Seidel switching on regular graphs such that the result will also be regular.

Problems

1. More generally, we show that if A is a matrix with integer entries and $\lambda \in \mathbb{Q}$ is an eigenvalue then $\lambda \in \mathbb{Z}$.

   First note that v is an eigenvector with eigenvalue $\lambda \in \mathbb{Q}$ if and only if $(A - \lambda I)v = 0$.

   When you do Gaussian elimination to solve the equation for the span of vectors v that satisfy the equation, you can work over the field $\mathbb{Q}$ so that you never need to introduce irrationals. The equation is solvable by assumption, since λ is an eigenvalue. Then we can assume that v has all rational entries. By rescaling, we can assume that v has all integer entries with gcd 1. Then we have $Av = \lambda v$, but A and v have all integer entries, and hence Av has integer entries as well. But then if $\lambda = p/q$ for $p \perp q$ then q must divide all the entries of v, i.e. $q = 1$.

   $\sqrt{2 + \sqrt{5}}$ has $\sqrt{2 - \sqrt{5}}$ as a conjugate becuase both have minimal polynomial $(x^2 - 2)^2 - 5 = x^4 - 4x^2 - 3$ (over $\mathbb{Q}$). So this would imply that this second number is also an eigenvalue, but this number is imaginary, so this contradicts our assumption that Γ is undirected, as this would imply A is symmetric.

2. We don't need to use the fact that Γ is undirected, so long as its eigenvectors form a basis. Recall that the number of walks of length h from $a \rightsquigarrow b$ is $(A^h)_{ab}$. Then note that this is equivalently $e_a^\intercal A^h e_b$. Now write out $e_b = \sum \beta_i v_i$ where the v_i are the eigenvectors with eignevalues θ_i. Then the number of walks becomes

$$\sum_i \beta_i e_a^\intercal A^h v_i = \sum_i \beta_i e_a^\intercal v_i \theta_i^h \tag{2}$$

Thus we get the desired conclusion by setting $c_i = \beta_i e_a^\intercal v_i$.

3. This is an easy consequence of the Perron-Frobenius theorem. The graph must contain some loops (take a random walk until you repeat yourself, there is always an edge to take), and the gcd of these loops will be some number $g > 2$, which means that if θ is the maximal real eigenvalue the graph will also have the eigenvalue $\theta\zeta$ for all ζ a gth root of unity.

4. (a) Note that $A^h = J$. Then since the spectrum of J is $n^1, 0^{n-1}$ we see that the (algebraic) spectrum of A is $\sqrt[h]{n}^1, 0^{n-1}$. If the eigenvectors formed a basis this would also be the gemoetric spectrum. Specifically, by looking at the Jordan Normal Form we see that A has $\mathbf{1}$ as a left and right eigenvector of eigenvalue $\sqrt[h]{n}$.

   With this in hand, note that $(A\mathbf{1})_v = \sqrt[h]{n} = \#$ edges leaving v. Similarly, because $\mathbf{1}$ is the left eigenvector for this eigenvalue, this is also the number of edges entering v. Since v is arbitrary we get that all in- and out-degrees are $k = \sqrt[h]{n}$, as desired.

   (b) Well, we see that for any two nodes v, w of the de Bruijn graph of order m, there is a unique path of length m $v \rightsquigarrow w$, where we replace the last element of the word of v by the elements of w in order. Thus the spectrum is $2^1, 0^{2^m-1}$ by the last part.

   (c) We use the fact that a de Bruijn cycle is just a Hamiltonian cycle in the de Bruijn graph.

2 Linear Algebra

We skip the proof of the Perron-Frobenius Theorem.

Suppose A is a real, symmetric matrix with rows and columns indexed by the set $X = [n]$. Let $X_1, \ldots, X_m$ partition X. Then we can consider the block matrix for A that this partition gives us, and let $A_{i,j}$ denote the block matrix in the X_i row and X_j column.

We define *the characteristic matrix S* to be the $n \times m$ matrix whose j^{th} column is the characteristic vector of X_j. We also let $n_i = |X_i|$ and $K =$ the diagonal matrix with the n_is as its entries.

Now let $b_{i,j}$ denote the average row sum in $A_{i,j}$, and define *the quotient matrix B* as the matrix with (i,j) entry $b_{i,j}$.

Then it is easily seen that

$$KB = S^\intercal AS \quad \text{and} \quad S^\intercal S = K$$

If the row sums of the $A_{i,j}$ are each constant, then the partition is called *equitable*. In this case, we see that $A_{i,j}\mathbf{1} = b_{i,j}\mathbf{1}$. Thus we see that $AS = SB$ in this case.

Putting this together, we prove

Lemma 2. *If we have an equitable partition of A and $Bv = \lambda v$ then Sv is an eigenvector of A, also of eigenvalue λ.*

Proof.

$$ASv = SBv = \lambda Sv$$

$\square$

Intuitively, the idea is that when you have an equitable partition of A, some of the eigenvectors are those that are constant on each block of the partition, which are those given by the eigenvectors of B in the lemma above, and the rest must be orthogonal to these eigenvectors, so sum to zero on every block of the partition. It is worth noting that since $Jv = 0$ for any v with zero sum, we see that these latter eigenvectors/eigenvalues do not change if we add a constant to every entry in the same block of A.

If we do this with the adjacency matrix A of a graph Γ, then an equitable partition of A is a partition of the vertex set of Γ, $X_1, \ldots, X_m$, such that every vertex in X_i has a fixed number, $b_{i,j}$, of neighbors in X_j.

If this instead holds only when $i \neq j$ then the partition is set to be *almost equitable*.

As an example, we can compute the spectrum of $K_{p,q}$ very easily using this method. The equitable partition given by L, R has quotient matrix $B = \begin{bmatrix} 0 & p \\ q & 0 \end{bmatrix}$ with eigenvalues $\pm\sqrt{pq}$. Since the adjacency matrix has rank 2, these are the only non-zero eigenvalues.

More generally, if Γ_1, Γ_2 are d_1, d_2 regular with n_1, n_2 vertices respectively and we take their *join*, defined as the graph obtained by adding all possible edges between Γ_1 and Γ_2, then we get an equitable partition by taking the vertex sets of the original graphs. The quotient matrix is then

$$B = \begin{bmatrix} d_1 & n_2 \\ n_1 & d_2 \end{bmatrix}$$

so that, letting k', k'' be the eigenvalues of B, and recalling that all other eigenvalues must be orthogonal to the blocks of the partition, we see that the spectrum of the join is the union of the spectra of Γ_1, Γ_2 but with d_1, d_2 removed (because these correspond to eigenvectors with non-zero sum) and k', k'' added.

Definition 16. *Suppose A is $n \times n$, real and symmetric. Then for non-zero u we define the Rayleigh quotient*

$$R(u) = \frac{u^{\mathsf{T}} A u}{u^{\mathsf{T}} u}$$

We prove that real symmetric matrices are orthogonally diagonalizable, and prove the Courant-Fischer inequalities.

Proof. Let A be real and symmetric.

Then the Rayleigh quotient $R : \mathbb{R}^n - \{0\} \to \mathbb{R}$ is continuous. Moreover, it is easily seen that $R(\alpha u) = R(u)$ for all $\alpha \neq 0$. Hence we may view R as a function on S^{n-1}, the $(n-1)$-dimensional sphere, with the canonical embedding as the unit sphere in $\mathbb{R}^n$. S^{n-1} is compact, so R attains some maximum θ_1 at some point u_1 of S^{n-1}.

Now we recall that if A is any matrix then $\frac{d}{du} u^{\mathsf{T}} A u = Au + A^{\mathsf{T}} u$ is the derivative w.r.t. u and so for a symmetric matrix the derivative is $2Au$. Note: usually I would use the transpose of this, but this notation is easier in this instance.

Similarly, we recall that the derivative of $u^{\mathsf{T}} u$ is $2u$. Then since u_1 is a maximum of R, the derivative of R at u_1 is zero, and we have

$$\frac{(u_1^{\mathsf{T}} u_1)(2 A u_1) - (2 u_1)(u_1^{\mathsf{T}} A u_1)}{(u_1^{\mathsf{T}} u_1)^2} = 0$$

by the quotient rule. We then conclude that

$$A u_1 = R(u_1) u_1$$

so that u_1 is an eigenvector with eigenvalue $R(u_1) = \theta_1$.

Now suppose that u is any eigenvector, of eigenvalue θ, say, and let $V^\perp$ denote the set of vectors orthogonal to u. If $v \in V^\perp$ then $u^\intercal(Av) = (u^\intercal A)v = \theta u^\intercal v = 0$ so that $v \in V^\perp \implies Av \in V^\perp$.

Now let $V_1^\perp$ denote the set $\{v \in \mathbb{R}^n \ : \ v^\intercal u_1 = 0\}$ of vectors orthogonal to u_1. Then by the above, $V_1^\perp$ is an invariant subspace w.r.t. A, so that we can let A act on it. Then the same argument as above shows that there exists $u_2 \in S^{n-2} \subseteq V_1^\perp$ that maximizes $R|_{V_1^\perp}$. Examining the above formula for the derivative of R at u_1, you will see that substituting in u_2 and applying the fact that $V_1^\perp$ is an invariant subspace, the derivative would have to be in $V_1^\perp$. Then this derivative must be zero or else we could move along it and stay within $V_1^\perp$, contradicting the assumption that u_2 maximizes the Rayleigh quotient for all vectors in S^{n-2}. Then the same argument shows that u_2 is an eigenvector of A with eigenvalue $R(u_2) = \theta_2$. Then since $u_2 \in V_1^\perp$, $u_2 \perp u_1$.

Continuing in this way, we obtain a sequence $u_1, u_2, \ldots, u_n$ of orthonormal eigenvectors with eigenvalues $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_n$.

Moreover, we saw that, by construction,

$$\theta_i = R(u_i) = \max_{\substack{u \in \mathrm{span}(u_i, \ldots, u_n) \\ u \neq 0}} R(u)$$

and, if we had instead done the argument by taking the minimums of R, then we would have constructed the sequence backwards, and thereby obtained

$$\theta_i = R(u_i) = \min_{\substack{u \in \mathrm{span}(u_1, \ldots, u_i) \\ u \neq 0}} R(u)$$

If you reformulate these equations as inequalities, they are called the Courant-Fischer inequalities. Note that equality in the Courant-Fischer inequality can only be attained at a θ_i-eigenvector.

$\square$

Now we prove interlacing.

Definition 17. *Two sequence of real numbers $\theta_1 \geq \ldots \geq \theta_n$ and $\eta_1 \geq \ldots \geq \eta_m$, where $m < n$, interlace if*

$$\theta_i \geq \eta_i \geq \theta_{n-m+i}$$

for all i.

The name comes from the case $n - m = 1$ in which case you get

$$\theta_1 \geq \eta_1 \geq \theta_2 \geq \eta_2 \geq \ldots \geq \eta_{n-1} \geq \theta_n$$

In the general case you can think of this as meaning that you get this picture except there is a "slack" of $n - m$ that allows η_i to end up farther down the list.

If there exists $0 \leq k \leq m$ such that for all $i \leq k$ $\eta_i = \theta_i$ and for all $i > k$ $\eta_i = \theta_{n-m+i}$ then the interlacing is said to be tight.

Theorem 2. *Cauchy Interlacing*
Suppose that S is an $n \times m$ matrix, where $n > m$, and $S^\intercal S = I$. For A a real, symmetric, $n \times n$ matrix, let $B = S^\intercal A S$.

As a special case, if $S = \begin{bmatrix} I & 0 \end{bmatrix}^\intercal$ then B is a principal submatrix of A.

Let the eigenvalues of A be $\theta_1 \geq \ldots \geq \theta_n$ and those of B $\eta_1 \geq \ldots \geq \eta_m$. Let the corresponding eigenvectors of B be $v_1, \ldots, v_m$.

Then the following hold

1. The eigenvalues of B interlace those of A.

2. If $\eta_i = \theta_i$ or $\eta_i = \theta_{n-m+i}$ then there exists a η_i-eigenvector of B, v, such that Sv is a η_i-eigenvector of A.

3. If $\eta_i = \theta_i$ for all $i \leq k$ for some k then Sv_i is an η_i-eigenvector for A for each $i \leq k$. Similarly, if $\eta_i = \theta_{n-m+i}$ for all $i > k$, Sv_i is then an η_i-eigenvector for each $i > k$.

4. If the interlacing is tight then $SB = AS$.

Proof. Let $u_1, \ldots, u_n$ be an orthonormal set of eigenvectors for A, where $Au_i = \theta_i$.

Then for each $i \in [m]$, let s_i be a non-zero vector in the span of $v_1, \ldots, v_i$ that is also orthogonal to all vectors in the span of $S^\mathsf{T} u_1, \ldots, S^\mathsf{T} u_{i-1}$. Such a vector exists because the dimension of the first span is i and the dimension of the second is at least $m - i + 1$, so that their sum is $m + 1 > m$. Then note that Ss_i is orthogonal to the span of $u_1, \ldots, u_{i-1}$, and hence by Courant-Fischer, we see that

$$\theta_i \geq \frac{(Ss_i)^\mathsf{T} A(Ss_i)}{(Ss_i)^\mathsf{T}(Ss_i)} = \frac{s_i^\mathsf{T} B s_i}{s_i^\mathsf{T} s_i} \geq \eta_i$$

and by negating all terms, we see that the order of the eigenvalues flips, so the same logic gives

$$-\theta_{n-i+1} \geq \frac{(Ss_i)^\mathsf{T}(-A)(Ss_i)}{(Ss_i)^\mathsf{T}(Ss_i)} = \frac{s_i^\mathsf{T}(-B)s_i}{s_i^\mathsf{T} s_i} \geq -\eta_{m-i+1}$$

so that by re-indexing $i \mapsto m - i + 1$ and negating again we see

$$\theta_{n-m+i} \leq \frac{(Ss_{m-i+1})^\mathsf{T} A(Ss_{m-i+1})}{(Ss_{m-i+1})^\mathsf{T}(Ss_{m-i+1})} = \frac{s_{m-i+1}^\mathsf{T} B s_{m-i+1}}{s_{m-i+1}^\mathsf{T} s_{m-i+1}} \leq \eta_i$$

so that $\theta_{n-m+i} \leq \eta_i$. This proves 1.

If $\theta_i = \eta_i$, then s_i and Ss_i are η_i-eigenvectors of B and A by the first inequality and similarly if $\theta_{n-m+i} = \eta_i$ then the second inequality shows that s_{m-i+1} and Ss_{m-i+1} are η_i eigenvectors. Here we have used the fact that equality holds in Courant-Fischer iff the vector is an eigenvector of that eigenvalue. Thus 2 is proven.

Now suppose that $\theta_i = \eta_i$ for all $i \leq k$. We use induction on k. Clearly the $k = 0$ case is trivial. Then suppose the $k - 1$ case is proven, so that $Sv_i = u_i$ for all $i = 1, \ldots, k - 1$. Then in the proof of 1, we could take $s_k = v_k$. Then applying the first inequality shows that Sv_k is a θ_i-eigenvector of A.

The other side is proven similarly using the second inequality. Thus 3 is proven.

So suppose the inequality is tight. By 3, $Sv_1, \ldots, Sv_m$ is an orthonormal set of eigenvectors of A for the eigenvalues $\eta_1, \ldots, \eta_m$. Then $SBv_i = \eta_i Sv_i = ASv_i$ so that equality holds for the v_i, but this is sufficient because the v_i span the space. Thus 4 is proven. $\square$

As noted in the statement of the theorem, letting $S = \begin{bmatrix} I0 \end{bmatrix}^\mathsf{T}$ (and possibly conjugating A by a permutation matrix) gives us a principal submatrix, so that as a corollary we have

Corollary 9. *If B is a principal submatrix of a symmetric matrix A, then the eigenvalues of B interlace those of A.*

3   Expander Graphs

This section is not in the book, I modified the way that some of it was presented. Note that it is standard in expander graphs to denote vectors as row vectors, because this way you can write the transition matrix as being a re-scaled version of the adjacency matrix.

Definition 18. *A digraph Γ is a γ-spectral expander if it is regular (in- and out-degree) and*

$$\lambda(G) = \max_{\substack{v \perp \mathbf{1} \\ v \neq 0}} \frac{||vM||}{||v||} \leq \lambda$$

where $\lambda = 1 - \gamma$ and $M = \frac{1}{d}A$ is the transition matrix for a random walk on Γ.

γ is called the spectral gap. Typically, you are most interested in the case where Γ is d-regular for some small d.

Note that if $v \perp \mathbf{1}$ then $vM \perp \mathbf{1}$ so that any other eigenvector other than the Perron-Frobenius eigenvector $\mathbf{1}$ would have to perpendicular to $\mathbf{1}$, so that the largest absolute value of such an eigenvalue would be λ. Since undirected d-regular graphs have bases of eigenvectors, the converse holds for them, and so this coincides with saying that $|\theta_i| \leq \lambda d$ for $i > 1$.

We note that if Γ is d-regular then u, the unif. dist. on $V \Gamma$, is clearly a stationary distribution of a random walk on Γ.

It is also easily proven that

$$\lambda(G) = \max_{\substack{v \perp \mathbf{1} \\ v \neq 0}} \frac{||vM||}{||v||} = \max_{\pi \text{ a prob. dist. on } V} \frac{||\pi M - u||}{||\pi - u||}$$

whenever Γ is d-regular.

Lemma 3.   *The Expander Mixing Lemma*
If Γ is a d-regular $\gamma = (1 - \lambda)$-two-sided-spectral expander, then for all $S, T \subseteq V \Gamma$, letting $e(S,T)$ again denote the # of $E \rightsquigarrow T$ edges, we have

$$\left| \frac{e(S,T)}{nd} - \alpha\beta \right| \leq \lambda\sqrt{\alpha(1-\alpha)\beta(1-\beta)} \leq \lambda\sqrt{\alpha\beta} \leq \lambda$$

where $|S| = s$, $|T| = t$ and $\alpha = s/n$, $\beta = t/n$ are the densities of S, T.

Proof. We note that $e(S,T) = \chi_S^\mathsf{T} A \chi_T$. $\chi_S = \alpha\mathbf{1} + \chi_S^\perp$ and similarly $\chi_T = \beta\mathbf{1} + \chi_T^\perp$ and note that $\chi_S^\perp, \chi_T^\perp$ both sum to zero. We let $M = \frac{1}{d}A$ be the transition matrix.

$$\begin{aligned}
\frac{e(S,T)}{d} &= \chi_S^\mathsf{T} M \chi_T \\
&= \alpha\mathbf{1}^\mathsf{T} M \beta\mathbf{1} + \chi_S^{\perp\mathsf{T}} M \chi_T^\perp \quad \text{the cross terms are zero, due to orthogonality} \\
&= n\alpha\beta + \chi_S^{\perp\mathsf{T}} M \chi_T^\perp
\end{aligned} \tag{3}$$

and so

$$\left| \frac{e(S,T)}{nd} - \alpha\beta \right| = \frac{1}{n} \left| \chi_S^{\perp\mathsf{T}} M \chi_T^\perp \right|$$

We can then use the fact that Γ is a γ-expander to write

$$\left| \chi_S^{\perp\mathsf{T}} M \chi_T^\perp \right| \leq \lambda \chi_S^{\perp\mathsf{T}} \chi_T^\perp$$

and so our problem comes down to bounding $\langle \chi_S^\perp, \chi_T^\perp \rangle$.

For this, we use the Pythagoras' Theorem and the Cauchy-Schwarz inequality. We note that $s = n\alpha = ||\chi_S||^2 = ||\alpha \mathbf{1}||^2 + ||\chi_S^\perp||^2$. Thus since $||\alpha \mathbf{1}||^2 = \alpha^2 n$ we find $||\chi_S^\perp|| = \sqrt{n\alpha - \alpha^2 n} = \sqrt{n\alpha(1-\alpha)}$. Similarly we find $||\chi_T^\perp|| = \sqrt{n\beta(1-\beta)}$. Thus Cauchy-Schwarz gives $\langle \chi_S^\perp, \chi_T^\perp \rangle \leq n\sqrt{\alpha(1-\alpha)\beta(1-\beta)}$, and we get our result,

$$\left| \frac{e(S,T)}{nd} - \alpha\beta \right| \leq \lambda\sqrt{\alpha(1-\alpha)\beta(1-\beta)}$$

$\square$

Lemma 4. *Matrix Decomposition Lemma*
If Γ is a γ-two-sided expander with transition matrix M, then it is possible to write

$$M = \frac{\gamma}{n}J + \lambda E$$

where $||E|| \leq 1$.

Proof. Clearly if this is true we must have $E = \frac{M - \frac{\gamma}{n}J}{\lambda}$. Then $E\mathbf{1} = \frac{\mathbf{1}M - \frac{\gamma}{n}\mathbf{1}J}{\lambda} = \frac{\mathbf{1} - \gamma\mathbf{1}}{\lambda}$ and by definition $1 - \gamma = \lambda$ so this is just $\mathbf{1}$. Thus $||E\mathbf{1}|| = ||\mathbf{1}||$.

Now suppose that $v \perp \mathbf{1}$, i.e. v sums to zero. Then $vE = \frac{vM - \frac{\gamma}{n}vJ}{\lambda} = \frac{vM}{\lambda}$. Then since $v \perp \mathbf{1}$ we have $||vM|| \leq \lambda||v||$.

Now let v be an arbitrary non-zero vector. Then we write $v = \alpha\mathbf{1} + v^\perp$ for some α such that $v^\perp \perp \mathbf{1}$. Then recall that $v^\perp M \perp \mathbf{1}$ (due to regularity of Γ). Then $||vM||^2 = ||\alpha\mathbf{1}M||^2 + ||v^\perp M||^2 \leq ||\alpha\mathbf{1}||^2 + ||v^\perp||^2 = ||v||^2$ as desired. $\square$

Morally speaking, we can think of this as meaning that we can treat a walk in the expander graph almost as though it was a walk in the complete graph with prob. γ and some other step with prob. λ. But E is not in general stochastic, so this is not quite true.

Theorem 3. *Hitting Property of Random Walks in an Expander*
If Γ is a γ-expander and $\lambda = 1 - \gamma$ then, if you pick a uniform-at-random start node v_1 and then take a random walk $v_1 \rightsquigarrow v_2 \rightsquigarrow \ldots \rightsquigarrow v_\ell$ then the probability that all of the random walk lies in some bad set B is small, so long as B is not (very) large. Specifically, if $|B| = b$ and $b/n = \mu$ is the density of B in Γ then

$$\Pr\left[\forall i \; v_i \in B\right] \leq (\mu + \lambda(1-\mu))^\ell$$

Proof. We define P to be the diagonal matrix where $P_{ii} = [i \in B]$, so that for π a prob. dist. on V, we have $|\pi P|_1 = $ probability (over π) of picking a node in B, where $|\cdot|_1$ denotes the ℓ_1 norm.

Now we note that the probability of the whole walk staying in B is simply $|uP(MP)^{\ell-1}|_1$ where u is the unif. distribution.

Lemma 5. $||PMP||_2 \leq \mu + \lambda(1-\mu)$.

Proof. Write $M = \frac{\gamma}{n}J + \lambda E$, as guaranteed by the Matrix Decomposition Lemma. Then

$$||PMP|| = ||P(\frac{\gamma}{n}J + \lambda E)|| \leq \frac{\gamma}{n}||PJP|| + \lambda||PEP|| \leq \frac{\gamma}{n}||PJP|| + \lambda$$

Now we bound $||PJP||$. Let x be given. Let $y = xP$. Note that $||y|| \leq ||x||$. Also note that the support of y is of size at most b. Then

$$||xPJP|| = ||yJP|| = |y|_1\mathbf{1}P \leq (\sqrt{b}||y||)\sqrt{b} \leq b||x||$$

18

where here we have used the inequality $|y|_1 \leq \sqrt{|Supp(y)|}||y||_2$ to go from the ℓ_1 norm to the ℓ_2 norm. This inequality can be prove by noting that $|y|_1 = \langle y, \chi_{Supp(y)} \rangle$ and applying the Cauchy-Schwarz inequality.

Plugging back into our first inequality we get

$$||PMP|| \leq \gamma\mu + \lambda = \mu + \lambda(1 - \mu)$$

as desired. $\qquad \square$

Now we use this lemma to prove the statement.

$$
\begin{aligned}
\Pr\left[\forall i \ v_i \in B\right] &= |uP(MP)^{\ell-1}|_1 \\
&\leq \sqrt{b}||uP(MP)^{\ell-1}|| \\
&\leq \sqrt{b} \cdot ||uP|| \cdot ||PMP||^{\ell-1} \\
&\leq \sqrt{b}\left(\frac{\sqrt{b}}{n}\right)\left(\mu + \lambda(1 - \mu)\right)^{\ell-1} \\
&\leq \left(\mu + \lambda(1 - \mu)\right)^{\ell}
\end{aligned}
\tag{4}
$$

where we have also used the fact that repeating P multiple times in a row does not effect anything, and again used the bound to go from ℓ_1 to ℓ_2. $\qquad \square$

Thus you could use a random walk in a (family of) expander(s) to reduce the randomness used in bringing down the error probability of an RP (or co-RP) algorithm (a random algorithm that has one-sided error). More generally, we might want to use this on a BPP algorithm (two-sided error), so would be interested in whether the expected amount of time spent in a subset of the graph is roughly proportional to its density, so that we could take a majority vote.

This is proven by the following theorem.

Theorem 4. *Strong Chernoff Bound for Expander Walks*
Let Γ be a $(1 - \lambda)$-spectral expander that is d-regular on n nodes and let $\gamma = 1 - \lambda$. Fix a sequence of functions $f_1, \ldots, f_k$, $f_i : V \rightarrow [0,1]$ for each i. Let v_1 be drawn uniformly at random and $v_1 \rightsquigarrow v_2 \rightsquigarrow \ldots \rightsquigarrow v_k$ be a random walk on Γ starting at v_1. Let $\mu_i = \mathbb{E}_{v \leftarrow V}\left[f(v)\right]$ (drawn uniformly from V). Let $\mu = \sum_i \mu_i$. Then for all $0 < \varepsilon \leq 1$,

$$
Pr\left[\left|\sum_i f_i(v_i) - \mu\right| > \varepsilon k\right] \leq 2e^{-\frac{\varepsilon^2 \gamma k}{4}}
\tag{5}
$$

Proof. Let $X = \sum_i f_i(v_i)$.

We will simply bound $Pr\left[X \geq \mu + \varepsilon k\right]$ because then one can apply the bound to $1 - f_i$ and take the union bound to get the two-sided bound.

We will be using the bound

$$
Pr\left[X \geq \mu + \varepsilon k\right] = Pr\left[e^{rX} \geq e^{r\mu + r\varepsilon k}\right] \leq \frac{\mathbb{E}\left[e^{rX}\right]}{e^{r\mu + r\varepsilon k}}
$$

which is simply Markov's bound applied to the moment generating function e^{rX}.

For each i, let $E_i = \text{diag}_v(e^{rf_i(v)})$. Then let M be the transition probability matrix of the graph Γ. Then letting $u = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{pmatrix}$ we easily see that

$$
\mathbb{E}\left[e^{rX}\right] = uE_1 M E_2 \ldots M E_k \mathbf{1} = u M E_1 M E_2 \ldots M E_k \mathbf{1}
$$

For $v \in \mathbb{R}^n$, let $v^{\parallel}$ and $v^{\perp}$ be the parts parallel and perpendicular to $\mathbf{1}$. Let $z_0 = u$, $z_1 = z_0 M E_1$, $z_{i+1} = z_i M E_{i+1}$. Then we simply want to bound $|z_k|_1$. Recall the bound $|z_k|_1 \leq \sqrt{n}\|z_k^{\parallel}\|_2$, noting equality holds because $z_k \parallel \mathbf{1}$, so we can work in the more convenient ℓ_2 norm instead.

We will prove a bound on $\|z_i^{\parallel}\|$ by induction on i. The idea of the proof is to first give a bound on $\|z_i^{\parallel}\|$ in terms of $\|z_{i-1}^{\parallel}\|$ and $\|z_{i-1}^{\perp}\|$, then give a bound on $\|z_{i-1}^{\perp}\|$ in terms of $\|z_0^{\parallel}\|, \|z_1^{\parallel}\|, \ldots, \|z_{i-2}^{\parallel}\|$. The inductive hypothesis then gives a bound for $\|z_{i-1}^{\perp}\|$. Intuitively, if the mean of f_i is large, and thus the bound needs to be small, then $E_i \approx I$, and M is the matrix of an expander so should push z_{i-1} back towards u. In the following, we think of r as being small, so that $e^r - 1 \approx r$ is also small.

We separate the bounds off into a few lemmas.

Lemma 6. *(Bounding z_{i+1} in terms of z_i)*
Suppose that $0 \leq r \leq \log(1/\lambda)/2$, $f : V \to [0,1]$ is a function with mean ρ, and let $E = \mathrm{diag}_v(e^{rf(v)})$. Then the following hold

1. $\|(z^{\parallel}ME)^{\parallel}\| \leq (1 + (e^r - 1)\rho)\|z^{\parallel}\|$

2. $\|(z^{\parallel}ME)^{\perp}\| \leq \frac{e^r - 1}{2}\|z^{\parallel}\|$

3. $\|(z^{\perp}ME)^{\parallel}\| \leq \frac{e^r - 1}{2}\lambda\|z^{\perp}\|$

4. $\|(z^{\perp}ME)^{\perp}\| \leq \sqrt{\lambda}\|z^{\perp}\|$

Proof. 1. $(z^{\parallel}ME)^{\parallel} = (z^{\parallel}E)^{\parallel} = (\mathbf{1}\cdot(z^{\parallel}E))u = (\mathbf{1}\cdot uE)z^{\parallel}$ and by construction $\mathbf{1}\cdot(uE) = \mathbb{E}_v[e^{rf(v)}]$

Now note that by convexity of the exponential, $e^{rx} \leq 1 + (e^r - 1)x$ for $x \in [0,1]$. So since $f(v) \in [0,1]$, we find $\mathbb{E}_v[e^{rf(v)}] \leq \mathbb{E}_v[1 + (e^r - 1)f(v)] = 1 + (e^r - 1)\rho$.

2. Use linearity and the fact that $v^{\parallel \perp} = 0$ for all v to write

$$(z^{\parallel}ME)^{\perp} = (z^{\parallel}E)^{\perp} = (z^{\parallel}(E - \alpha I))^{\perp} + (z^{\parallel}(\alpha I))^{\perp} = (z^{\parallel}(E - \alpha I))^{\perp}$$

where α is a scalar that we can pick. We pick $\alpha = \frac{e^r + 1}{2}$. Then $E - \alpha I$ is diagonal and its entries lie between $1 - \alpha = -\frac{e^r - 1}{2}$ and $e^r - \alpha = \frac{e^r - 1}{2}$, so in absolute value are at most $\frac{e^r - 1}{2}$ each. The bound follows immediately.

3. Similarly, here we use linearity and the fact that $v^{\perp \parallel} = 0$ to write

$$(z^{\perp}ME)^{\parallel} = (z^{\perp}M(E - \alpha I))^{\parallel}$$

and again set $\alpha = \frac{e^r + 1}{2}$ s.t. $\|E - \alpha I\| \leq \frac{e^r - 1}{2}$. Since Γ is a γ-expander, $\|z^{\perp}M\| \leq \lambda\|z^{\perp}\|$ and the bound follows.

4. This one's easy.
$$\|(z^{\perp}ME)^{\perp}\| \leq \|z^{\perp}ME\| \leq e^r\|z^{\perp}M\| \leq \lambda e^r\|z^{\perp}\|$$

Then we use the fact that $0 \leq r \leq \log(1/\lambda)/2$ to see $1 \leq e^r \leq \frac{1}{\sqrt{\lambda}}$ and find the bound.
$\qquad\square$

Now we show that $\|z_i^{\perp}\|$ can be bounded in terms of the preceding $\|z_0^{\parallel}\|, \|z_1^{\parallel}\|, \ldots, \|z_{i-1}^{\parallel}\|$ as we said before, which will set us up to prove the bound on $\|z_i^{\parallel}\|$ by induction.

Lemma 7. *If $i = 0$ then $\|z_0^{\perp}\| = 0$. If $i > 0$ then $\|z_i^{\perp}\| \leq \frac{e^r - 1}{1 - \lambda} \max_{j<i}\|z_j^{\parallel}\|$.*

Proof. Note that if $i = 0$ then $z_0 = u$ has $z_0^\perp = 0$ which is trivial. Assume that $i > 0$.

$$\|z_i^\perp\| = \|(z_{i-1} M E_i)^\perp\| \leq \|(z_{i-1}^{\parallel} M E_i)^{\parallel}\| + \|(z_{i-1}^\perp M E_i)^\perp\|$$

by the triangle inequality.

We can use the last lemma to bound these two parts, finding

$$\|z_i^\perp\| \leq \frac{e^r - 1}{2} \|z_{i-1}^{\parallel}\| + \sqrt{\lambda} \|z_{i-1}^\perp\|$$

Now we recursively apply this bound to $\|z_{i-1}^\perp\|$, expanding out a geometric sum that can be bounded directly

$$\|z_i^\perp\| \leq \frac{e^r - 1}{2} \sum_{0 < j \leq i} (\sqrt{\lambda})^j \|z_{i-j}^{\parallel}\| \leq \frac{e^r - 1}{2} \sum_{0 < j \leq i} (\sqrt{\lambda})^j \max_{j < i} \|z_j^{\parallel}\| \leq \frac{e^r - 1}{2(1 - \sqrt{\lambda})} \max_{j < i} \|z_j^{\parallel}\|$$

Finally, note that

$$\frac{1}{1 - \sqrt{\lambda}} = \frac{1 + \sqrt{\lambda}}{1 - \lambda} \leq \frac{2}{1 - \lambda}$$

because $\lambda \in [0, 1]$. $\qquad\square$

Using these two lemmas, we can can bound $\|z_i^{\parallel}\|$, which basically finishes the proof once we do some cleanup.

Lemma 8.

$$\|z_i^{\parallel}\| \leq \exp\left((e^r - 1)\mu_i + \frac{\lambda(e^r - 1)^2}{2(1 - \lambda)} \right) \max_{j < i} \|z_j^{\parallel}\|$$

Proof. The first lemma we proved shows that

$$\|z_i^{\parallel}\| \leq (1 + (e^r - 1)\mu_i) \|z_{i-1}^{\parallel}\| + \frac{e^r - 1}{2} \lambda \|z_{i-1}^\perp\|$$

Next, we apply the last lemma to the $\|z_{i-1}^\perp\|$ term, finding

$$\|z_i^{\parallel}\| \leq (1 + (e^r - 1)\mu_i) \|z_{i-1}^{\parallel}\| + \frac{(e^r - 1)^2}{2(1 - \lambda)} \lambda \max_{j < i-1} \|z_j^{\parallel}\| \leq \left(1 + (e^r - 1)\mu_i + \frac{\lambda(e^r - 1)^2}{2(1 - \lambda)} \right) \max_{j < i} \|z_j^{\parallel}\|$$

Now we use the inequality $1 + x \leq \exp(x)$ for positive x to conclude

$$\|z_i^{\parallel}\| \leq \exp\left((e^r - 1)\mu_i + \frac{\lambda(e^r - 1)^2}{2(1 - \lambda)} \right) \max_{j < i-1} \|z_j^{\parallel}\|$$

$\qquad\square$

Now we use this lemma to prove the theorem. Since $\|z_0^{\parallel}\| = \|u\| = 1/\sqrt{n}$, we see by induction that for all $j \geq 0$,

$$\begin{aligned}
\|z_j^{\parallel}\| &\leq \frac{1}{\sqrt{n}} \prod_{i=1}^{j} \exp\left((e^r - 1)\mu_i + \frac{\lambda(e^r - 1)^2}{2(1 - \lambda)} \right) \\
&= \frac{1}{\sqrt{n}} \exp\left((e^r - 1) \sum_{\ell=1}^{j} \mu_\ell + j \frac{\lambda(e^r - 1)^2}{2(1 - \lambda)} \right)
\end{aligned} \tag{6}$$

21

Now we set $j = k$ in the above equation and find

$$\mathbb{E}\big[e^{rX}\big] \leq \sqrt{n}\|z_k^\|\| \leq \exp\left((e^r - 1)\mu + k\frac{\lambda(e^r - 1)^2}{2(1 - \lambda)}\right)$$

Now assume that $0 \leq r \leq 1/2$. Then the following bounds hold:

1. $e^r - 1 \leq 4r/3$

2. $e^r - 1 \leq r + r^2$

Noting that $\mu \leq k$, we can then simplify the expression to

$$\mathbb{E}\big[e^{rX}\big] \leq \exp\left((r + r^2)\mu + k\frac{\lambda(4r/3)^2}{2(1 - \lambda)}\right) \leq \exp\left(r\mu + r^2(1 + \frac{\lambda}{1 - \lambda})k\right) = \exp\left(r\mu + \frac{r^2 k}{1 - \lambda}\right)$$

Then we plug back in to find

$$Pr\big[X \geq \mu + \varepsilon k\big] \leq \frac{\mathbb{E}\big[e^{rX}\big]}{e^{r\mu + r\varepsilon k}} \leq \exp\left(k\big(\frac{r^2}{1 - \lambda} - r\varepsilon\big)\right)$$

Now one plugs in $r = (1 - \lambda)\varepsilon/2$ to minimize the quantity. This is valid because clearly $0 \leq r \leq 1/2$ (recall $\varepsilon \leq 1$) and $\log(1/\lambda)/2 \geq 1/2$ so that also $r \leq \log(1/\lambda)/2$.

The expression then reduces to

$$Pr\big[X \geq \mu + \varepsilon k\big] \leq e^{-\varepsilon^2 \gamma k/4}$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Note that this implies what we wanted by taking all f_i to be the characteristic function of some set S, but it is also much more general.

It is possible to get similar bounds in general primitive (ergodic) Markov chains.

Sometimes, one is interested in different notions of expansion.

Definition 19. *If Γ is a d-regular graph and its probability transition matrix has eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ and $1 - \lambda_2 \geq \gamma$ then Γ is said to have one-sided spectral expansion γ.*

One-sided spectral expansion is strictly weaker than normal, two-sided spectral expansion because $|\lambda_n|$ could be large, and if Γ is bipartite, could even be 1.

For an undirected, loopless graph Γ, suppose that Γ has positive weights assigned to its edges, say edge e has weight w_e. For sake of simplicity, if there is no edge between u and v, we say $w_{uv} = 0$. Then for $x \in V$, define $w_x = \sum_{x \in e} w_e$. The weights induce a random walk on Γ by taking the probability of a transition $x \rightsquigarrow y$ to be w_{xy}/w_x. Then clearly $\pi = (w_x)_x$ is a stationary distribution of the random walk on Γ.

For u in the vertex set of such a graph, we define $d(u) = w_x$ to be its weighted degree. Then for $S \subseteq V$, we define $d(S) = \sum_{u \in S} d(u)$. For a set of edges B we define $d(B) = \sum_{e \in B} w_e$. We also define the "edge boundary" $\partial S = \{ (u, v) \in E : \text{ exactly one of } u, v \text{ belongs to } S\}$. Intuitively, a graph is a good expander if the edge boundaries most sets are reasonably large.

Definition 20. *Sparsity and Conductance*
Sparsity is

$$\sigma = \min_S \left\{ d(V) \cdot \frac{w(\partial S)}{d(S)d(\overline{S})} \right\}$$

and conductance is

$$\phi = \min_S \left\{ \frac{w(\partial S)}{\min\{d(S), d(\overline{S})\}} \right\}$$

One can also define vertex expansion.

Definition 21. *Vertex Expansion*
A graph Γ is an (α, h) vertex expander if for all subsets $S \subseteq V$ such that $|S| \leq \alpha|V|$, we have $|N(S)| \geq h|S|$ where $N(S)$ is the neighborhood of S, defined by the set of all nodes reachable by one edge out of S.

We see that a d-regular, unweighted graph (without multiple edges) is a $(1/2, h)$-vertex expander iff $h \leq d\phi$, where ϕ is the conductance, so that the conductance and vertex expansion of graphs are very related notions.

We note the following relationship between σ and ϕ. For any $S \subseteq V$, we have

$$\frac{d(V)}{d(S)d(\overline{S})} = \frac{1}{\min\{d(S), d(\overline{S})\}} \frac{d(V)}{\max\{d(S), d(\overline{S})\}}$$

and since the last fraction is in $[1, 2]$, we immediately find

Claim 21. *Sparsity and conductance are 2-approximations for each other, i.e.*

$$\phi \leq \sigma \leq 2\phi$$

We define the weighted Laplace matrix of Γ by

$$L_{uv} = \begin{cases} d(u) & u = v \\ -w_{uv} & \text{otherwise} \end{cases}$$

Clearly the weighted Laplace matrix shares some properties with the Laplace matrix of an unweighted graph, namely

$$x^\mathsf{T} L x = \sum_{uv \in E} w_{uv}(x_u - x_v)^2$$

by the same proof as before, where you now consider the "weighted directed incidence matrix" in the proof, and so L is symmetric and PSD.

We also define $\overline{L}$, the normalized Laplacian, by $\overline{L} = D^{-1/2}LD^{-1/2}$ where $D = \text{diag}_v(d(v))$. The "meaning" of the normalized Laplacian is best explained by the case that Γ is d-regular, in which case $L = dI - A$ and so $\overline{L} = I - M$, although the real importance is that its eigenvalues are meaningful. Let $0 = \nu_1 \leq \nu_2 \leq \ldots \leq \nu_n$ be the eigenvalues of $\overline{L}$. Recall that for the unnormalized Laplacian in the unweighted case, we say that the second eigenvalue was the "algebraic connectivity", and we will think of ν_2 being a similar measure. Again, in the case that Γ is d-regular, we have $\nu_2 = \lambda_2$, the second largest eigenvalue (this time, not by absolute value) of the transition matrix M.

We show that ν_2 is similarly related to the expansion / connectivity of Γ. In the case that Γ is d-regular, $L = A - dI$ and so ν_2 is related to the spectral expansion of Γ. We will be interested in connected graphs here, and recalling that $L\mathbf{1} = 0$, we see that $\overline{L}D^{1/2}\mathbf{1} = 0$ so that for a connected graph $D^{1/2}\mathbf{1}$ spans the kernel of $\overline{L}$.

Theorem 5. *Cheeger's Inequality*
For any weighted, undirected graph Γ, ν_2, σ and ϕ approximate each other well. Specifically,

$$\nu_2 \leq \sigma \leq 2\phi \leq \sqrt{8\nu_2}$$

Note that $0 \leq \nu_2 \leq 1$ so that $\sqrt{\nu_2} \geq \nu_2$.

Proof. First we show that $\nu_2 \le \sigma \le 2\phi$, which is the "easy direction". First we use Courant-Fischer to write

$$\nu_2 = \min\left\{\frac{x^\mathsf{T}\overline{L}x}{x^\mathsf{T}x} \ : \ x \ne 0, x^\mathsf{T}D^{1/2}\mathbf{1} = 0\right\} = \min\left\{\frac{y^\mathsf{T}Ly}{y^\mathsf{T}Dy} \ : \ y \ne 0, y^\mathsf{T}D\mathbf{1} = 0\right\}$$

Note the following fact: adding any multiple of $\mathbf{1}$ to y does not change the expression being minimized on the right because it changes neither the numerator nor denominator.

Also, let $\mathbf{d} = (d(v))_v$. Then the condition $y^\mathsf{T}D\mathbf{1} = 0$ is clearly equivalent to $y^\mathsf{T}\mathbf{d} = 0$.

Now let y' be the characteristic vector of some set S. Set $s = d(S)/d(V)$ and let $y = y' - s\mathbf{1}$. Then

$$y^\mathsf{T}Ly = y'^\mathsf{T}Ly' = \sum_{uv \in E}(y'(u) - y'(v))^2 = |\partial S|$$

Next we compute

$$y^\mathsf{T}Dy = \sum_v y(v)^2 d(v) \sum_{v \in S}(1-s)^2 d(v) + \sum_{v \in \overline{S}}(-s)^2 d(v)$$

$$= (1-s)^2 d(S) + s^2 d(V) = (1-s)d(S) + s\left(\frac{d(V)}{d(S)}d(V) - d(S)\right) = (1-s)d(S) \tag{7}$$

and finally we compute

$$y^\mathsf{T}\mathbf{d} = \sum_{v \in S} 1 d(v) + \sum_{v \in V} -s d(v) = d(S) - s d(V) = 0$$

Then plugging y into our earlier formula, we find that

$$\nu_2 \le \frac{y^\mathsf{T}Ly}{y^\mathsf{T}Dy} = \frac{|\partial S|}{(1-s)d(S)} = \frac{d(V)|\partial S|}{d(\overline{S})d(S)} \le \sigma \le 2\phi$$

Now we prove the other direction.

First we need to prove a lemma that we could have used during the proof of the first direction, but was not really necessary.

Lemma 9. *For all y, we have*

$$y^\mathsf{T}Dy \ge \frac{1}{2d(V)}\sum_{u \ne v}d(u)d(v)(y(u) - y(v))^2$$

and equality holds iff $y \perp \mathbf{d}$. Note that the sum considers u and v as an ordered pair.

Proof. Just re-arrange the sums.

$$\frac{1}{2}\sum_{u \ne v}d(u)d(v)\big(y(u) - y(v)\big)^2 = \frac{1}{2}\sum_{u \ne v}d(u)d(v)\big[y(u)^2 + y(v)^2\big] - \sum_{u \ne v}d(u)d(v)y(u)y(v)$$

$$= \sum_{u \ne v}d(u)d(v)y(u)^2 - \sum_{u \ne v}d(u)d(v)y(u)y(v)$$

$$= \sum_{u,v}d(u)d(v)y(u)^2 - \sum_{u,v}d(u)d(v)y(u)y(v) \tag{8}$$

$$= d(V)\sum_u d(u)y(u)^2 - \left(\sum_u d(u)y(u)\right)^2$$

$$= d(V)y^\mathsf{T}Dy - \big(y^\mathsf{T}\mathbf{d}\big)^2$$

$\square$

Corollary 10.

$$\nu_2 = \min\left\{d(V)\frac{\sum_{uv\in E}w_{uv}(y(u)-y(v))^2}{\sum_{\{u,v\}\subseteq V}d(u)d(v)(y(u)-y(v))^2} : \textit{the denominator is non-zero}\right\}$$

Note that now we are considering the sum in the denominator to be over unordered pairs, which halves the value from the expression in the lemma.

Proof. As noted in the first half of the proof, adding a multiple of $\mathbf{1}$ to y does not change the value on the far right side of our expression for ν_2. By the lemma, the value in this expression can be made as small as that at the top, and by adding a multiple of $\mathbf{1}$ we can turn any solution of this minimization into a solution satisfying $y \perp \mathbf{d}$ without changing the value. $\square$

It will be convenient to take $\phi(S) = \frac{w(\partial S)}{\min\{d(S), d(\overline{S})\}}$.

Define

$$Q(y) = d(V)\frac{\sum_{uv\in E}w_{uv}(y(u)-y(v))^2}{\sum_{\{u,v\}\subseteq V}d(u)d(v)(y(u)-y(v))^2}$$

We will show that, given any y such that $Q(y)$ is defined, we can find a cut $\langle S, \overline{S}\rangle$ such that $\phi(S) \leq \sqrt{2Q(y)}$. Choosing y that minimizes Q will then prove Cheeger's inequality.

Let y be given, and re-number the vertices of Γ $v_1, v_2, \ldots, v_n$ such that $y_1 \leq y_2 \leq \ldots \leq y_n$. We show there exists k such that $S = \{y_1, \ldots, y_k\}$ is such a cut.

Note that adding a scalar multiple of $\mathbf{1}$ to y does not change the value of Q, and neither does rescaling y by a non-zero scalar. Thus, WLOG, we can assume that

$$\sum_{i\;:\;y_i<0}d(v_i) \leq \sum_{i\;:\;y_i\geq 0}d(v_i)$$

and that

$$\sum_{i\;:\;y_i\leq 0}d(v_i) \geq \sum_{i\;:\;y_i>0}d(v_i)$$

This can be accomplished by a shift such that $y_k = 0$ where k is the minimal number such that $\sum_{i\leq k}d(v_i) \geq d(V)/2$.

Here comes the weird part of the proof. Let z by

$$z_i = \begin{cases} -y_i^2 & y_i < 0 \\ y_i^2 & y_i \geq 0 \end{cases}$$

By rescaling y before doing this, we can assume that $z_n - z_1 = 1$. Now draw a threshold value $t \leftarrow [z_1, z_n]$ uniformly at random and define $S = \{v_i \;:\; z_i < t\}$. We claim that

$$\mathbb{E}[w(\partial S)] \leq \sqrt{2Q(y)}\,\mathbb{E}[\min\{d(S), d(\overline{S})\}]$$

from which it follows there exists at least one S such that $\phi(S) \leq \sqrt{2Q(y)}$.

First we evaluate $\mathbb{E}[\min\{d(S), d(\overline{S})\}]$. If v_i belongs to the smaller set, it contributes $d(v_i)$. Otherwise, it contributes nothing. It lands in the smaller set iff t lands between 0 and z_i (because

25

of our shift of y to get the inequalities on the sums of $d(v_i)$'s. This event occurs with probability $|z_i|$. Hence,

$$\mathbb{E}[\min\{d(S), d(\overline{S})\}] = \sum_u d(u)|z(u)| = \sum_u d(u)y(u)^2 = y^\mathsf{T}Dy$$

Next, we bound $\mathbb{E}[w(\partial S)]$. An edge uv contributes $w(u,v)$ to this value if it is cut. This occurs with probability $|z(u) - z(v)|$. We see that

$$|z(u) - z(v)| \leq |y(u) - y(v)|(|y(v)| + |y(v)|)$$

because, if $y(u)$ and $y(v)$ have the same sign then equality holds, and if they have opposite signs then the right hand side is $y(u)^2 + y(v)^2$ while the right hand side is $(|y(u)| + |y(v)|)^2$. Now we compute

$$\mathbb{E}[w(\partial S)] \leq \sum_{uv \in E} w_{uv}|y(u) - y(v)|(|y(u)| + |y(v)|)$$

$$\leq \sqrt{\sum_{uv \in E} w_{uv}(y(u) - y(v))^2} \cdot \sqrt{\sum_{uv \in E} w_{uv}(|y(u)| + |y(v)|)^2} \quad \text{by Cauchy-Schwarz}$$

$$= \sqrt{\frac{Q(y)}{d(V)} \sum_{\{u,v\} \subseteq V} d(u)d(v)(y(u) - y(v))^2} \cdot \sqrt{\sum_{uv \in E} w_{uv}(|y(u)| + |y(v)|)^2} \quad \text{by definition of } Q$$

$$\leq \sqrt{\frac{Q(y)}{d(V)} \sum_{\{u,v\} \subseteq V} d(u)d(v)(y(u) - y(v))^2} \cdot \sqrt{\sum_{uv \in E} w_{uv}(2y(u)^2 + 2y(v)^2)}$$

$$\leq \sqrt{Q(y)y^\mathsf{T}Dy} \cdot \sqrt{2 \sum_{v \in V} d(v)y(v)^2} \quad \text{by the earlier bound on } y^\mathsf{T}Dy$$

$$= \sqrt{2Q(y) \cdot (y^\mathsf{T}Dy)}$$

$$(9)$$

which completes the proof.

$\square$

Corollary 11. *Vertex- and One-Sided-Spectral- Expansion are Similar for Regular Graphs*
If Γ is a d-regular, undirected, unweighted graph with one-sided expansion γ and and it is a $(1/2, d\phi)$ expander (and ϕ is the largest value that can be put here) then

$$\frac{1}{2}\gamma \leq \phi \leq \sqrt{2\gamma}$$

Proof. In this case, the conductance is ϕ. Because Γ is d-regular, we find $L = dI - A$ and $\overline{L} = I - M$. Then $\nu_2 = \gamma$ and we apply Cheeger's inequality. $\square$

Corollary 12. *One can compute a sparse cut of a graph by computing the ν_2-eigenvector of $\overline{L}$ and then testing which of the threshold S's used in the proof of the Cheeger inequality is the sparsest, of which there are only n. At least one of them will have $\phi(S) \leq \sqrt{2\nu_2}$ and this is a good approximation, because we know that we cannot do any better than $\nu_2/2$. When ν_2 is very large (close to 1), which will happen in very dense graphs where $m \gg n$, the approximation will be especially good because $\sqrt{\nu_2} \approx \nu_2$ and so the worst-case approximation factor is $\approx 2\sqrt{2}$.*