Spectral Graph Theory

Max Hopkins
nmhopkin@eng.ucsd.edu

Daniël Kroes
dkroes@ucsd.edu

Jiaxi Nie
jin019@ucsd.edu

Jason O'Neill
jmoneill@ucsd.edu

Andrés Rodríguez Rey
a3rodrig@ucsd.edu

Nicholas Sieger
nsieger@ucsd.edu

Sam Sprio
sspiro@ucsd.edu

Winter 2020 Quarter

Abstract

Spectral graph theory is a vast and expanding area of combinatorics. We start these notes by introducing and motivating classical matrices associated with a graph, and then show how to derive combinatorial properties of a graph from the eigenvalues of these matrices. We then examine more modern results such as polynomial interlacing and high dimensional expanders.

1 Introduction

These notes are comprised from a lecture series for graduate students in combinatorics at UCSD during the Winter 2020 Quarter. The organization of these expository notes is as follows. Each section corresponds to a fifty minute lecture given as part of the seminar. The first set of sections loosely deals with associating some specific matrix to a graph and then deriving combinatorial properties from its spectrum, and the second half focus on expanders.

Throughout we use standard graph theory notation. In particular, given a graph G, we let $V(G)$ denote its vertex set and $E(G)$ its edge set. We write $e(G) = |E(G)|$, and for (possibly non-disjoint) sets A and B we write $e(A, B) = |\{uv \in E(G) : u \in A,\ v \in B\}|$. We let $N(v)$ denote the neighborhood of the vertex v and let d_v denote its degree. We write $u \sim v$ when $uv \in E(G)$.

For a matrix M, we let $\sigma(M)$ denote the multi-set of eigenvalues of M, and we will often write this as $\{\lambda_1^{(a_1)}, \ldots\}$ to indicate that the eigenvalue λ_i appears with multiplicity a_i. We let $\mathbf{1}$ denote the all 1's vector. Whenever we say that λ is the eigenvalue of some graph, we mean that λ is an eigenvalue of M_G where M_G is the relevant matrix associated to G for that section.

Contents

1 Introduction **1**

2 Spectral Graph Theory and the Adjacency Matrix **3**
2.1 Combinatorial Information from Eigenvalues 3
2.2 Computing Eigenvalues of Graphs 5

3 The Laplacian Matrix of a Graph **7**
3.1 The Matrix Tree Theorem 9

4 Towards a Theoretical Foundation for Laplacian-Based Manifold Methods **9**
4.1 Differential Geometry 10

5 The Normalized Laplacian **12**

6 Strongly Regular Graphs **15**

7 (n, d, λ) Graphs and Ramsey numbers **19**
7.1 Ramsey numbers 20

8 Ramanujan Graphs **22**
8.1 Lifts 23
8.2 Interlacing Families 24
8.3 Interlacing and Real Stable Polynomials 26
8.4 Further Directions 28

9 High Dimensional Expanders **28**
9.1 Weighted Expansion 31
9.2 Weighted Simplicial Complexes 32
9.3 Spectral Link Expansion and the Trickle Down Theorem 33
9.4 The Trickling Down Theorem 34

2 Spectral Graph Theory and the Adjacency Matrix

2.1 Combinatorial Information from Eigenvalues

Given an n-vertex graph G, we define its adjacency matrix $A = A_G$ with rows and columns indexed by $V(G)$ by $A_{uv} = 1$ if $uv \in E(G)$ and $A_{uv} = 0$ otherwise. In addition to being a compact way to define the graph G, it turns out that the adjacency matrix encodes combinatorial information about G through its eigenvalues. We first recall the following fact from linear algebra.

Theorem 2.1. *Let M be a Hermitian matrix. Then M has real eigenvalues $\lambda_1(M) \geq \cdots \geq \lambda_n(M)$. Moreover,*

$$\lambda_1(M) = \max_{x \neq 0} \frac{x^* M x}{x^* x},$$

and any x achieving equality is an eigenvector corresponding to $\lambda_1(M)$. Moreover, if the entries of M are non-negative, then this x can be chosen to have non-negative entries.

Because A is a real symmetric matrix, the theorem applies. For the rest of the section we define $\lambda_i(G) := \lambda_i(A_G)$, and we simply write λ_i whenever G is understood. With this we can deduce a number of results relating the eigenvalues of A to combinatorial properties of G.

Theorem 2.2. *Let G be a graph with maximum degree Δ and minimum degree δ. Then*

$$\delta \leq \lambda_1 \leq \Delta.$$

Proof. For the lower bound, let $\mathbf{1}$ be the all 1's vector. Then

$$\lambda_1 = \max_{y \neq 0} \frac{y^* A y}{y^* y} \geq \frac{\mathbf{1}^* A \mathbf{1}}{\mathbf{1}^* \mathbf{1}} = \frac{\sum d_v}{n} \geq \delta.$$

For the upper bound, let x be an eigenvector of A corresponding to λ_1 and let $v \in V(G)$ be such that $|x_v|$ is maximized. Then we have

$$|\lambda_1 x_v| = |(Ax)_v| = |\sum_u A_{v,u} x_u| \leq \sum_{u \sim v} |x_u| \leq d_v |x_v| \leq \Delta |x_v|.$$

Thus $|\lambda_1| \leq \Delta$. Because all the eigenvalues of A are real and $\mathrm{Tr}(A) = 0$, we must have $\lambda_1 \geq 0$, giving the desired result.

□

The above proof easily generalizes to the following.

Theorem 2.3. *Let G be an n-vertex graph and M a hermitian matrix such that $|M_{u,v}| = 1$ whenever $u \sim v$ and $M_{u,v} = 0$ otherwise. Then*

$$\lambda_1(M) \leq \Delta.$$

We note that this theorem is not just generalization for generalization's sake. Indeed, a special case of Theorem 2.3 was used by Hao Huang to prove the sensitivity conjecture [14]. This illustrates the general principle of spectral graph theory that choosing different matrices to associate to your graph G leads to different information being captured by your eigenvalues.

We continue our exploration of combinatorial properties implied by the eigenvalues of A.

Lemma 2.4. *Let G' be a subgraph of G and $\lambda'_i = \lambda_i(A_{G'})$. Then $\lambda'_1 \leq \lambda_1$.*

Proof. Let x be a unit eigenvector for λ_1' with non-negative entries. Then

$$\lambda_1' = x^* A_{G'} x \leq x^* A x \leq \lambda_1.$$

□

Theorem 2.5. $\chi(G) \leq \lambda_1 + 1$.

Proof. This holds when $n = 1$. Inductively assume the result holds up to n. Let v be a vertex of minimum degree δ and $G' = G - x$. If $\lambda_1' = \lambda_1(A_{G'})$, then by induction and the previous lemma we have $\chi(G - x) \leq \lambda_1' + 1 \leq \lambda_1 + 1$. Thus we can color $G - x$ in at most $\lambda_1 + 1$ colors, and as $\delta \leq \lambda_1$ by Theorem 2.2, we can give x a color that is not used by any of its neighbors. □

Note that this bound is often stronger than Brook's Theorem $\chi(G) \leq \Delta$. We also have the following lower bound on $\chi(G)$, whose proof we omit.

Theorem 2.6. [23]

$$\chi(G) \geq 1 - \frac{\lambda_1}{\lambda_n}.$$

Recall that a walk of length k is a sequence of (not necessarily distinct) vertices $x_1, \ldots, x_{k+1}$ such that $x_i \sim x_{i+1}$ for all $1 \leq i \leq k$. A walk is said to be closed if $x_{k+1} = x_1$.

Lemma 2.7. *The number of walks of length k from u to v is $A^k_{u,v}$.*

Proof. By definition of matrix multiplication, we have

$$A^k_{u,v} = \sum A_{uw_1} \cdots A_{w_{k-1}v},$$

where the sum ranges over all sequences $w_1, \ldots, w_{k-1}$. The term will be 1 if this sequence defines a walk and will be 0 otherwise. □

Corollary 2.8. *The number of closed walks of length k is $\mathrm{Tr}(A^k) = \sum \lambda_i^k$.*

Corollary 2.9. $e(G) = \sum \lambda_i^2$.

Corollary 2.10. *A graph G is bipartite iff $\sigma(A)$ is symmetric about 0.*

Proof. If G is bipartite with bipartition $U \cup V$ and x is such that $Ax = \lambda x$, then define y such that $y_u = x_u$ if $u \in U$ and $y_u = -x_u$ if $u \in V$. Then for $u \in U$ we have

$$(Ay)_u = \sum_{v \sim u} A_{u,v} y_v = -\sum_{v \sim u} A_{u,v} x_v = -\lambda x_u = -\lambda y_u.$$

The same conclusion holds if $u \in V$, so $Ay = -\lambda y$. We conclude that the spectrum of A is symmetric about 0. Conversely, if the spectrum of A is symmetric about 0, then G has 0 closed walks of length $2k + 1$ for all k. In particular, G contains no odd cycles. □

Proposition 2.11. *The number of distinct eigenvalues of A is larger than the diameter D of G.*

Proof. By Lemma 2.7, if $k \leq D$, then A^k will contain some non-zero entry which is 0 for all A^ℓ and $\ell < k$. Thus the minimum polynomial of A must have degree at least $D + 1$. □

Theorem 2.12. *If G is d-regular then $\alpha(G) \leq \frac{-n\lambda_n}{d - \lambda_n}$).*

Proof. Let S be an independent set and let $s = |S|/n$. Let x be the vector which has $x_v = -s$ if $v \notin S$ and $x_v = 1 - s$ otherwise. Because the graph is d-regular and S is an independent set,

$$x^t Ax = -d|S|(1-s)s + (dn/2 - d|S|)s^2 = -dns^2(1-s) + dn(.5-s)s^2 = -dns^2/2.$$

Also

$$x^t x = |S|(1-s)^2 + (n-|S|)s^2 = ns(1-s),$$

so by the Raleigh quotient we have $\lambda_n \leq \frac{-ds}{2(1-s)}$, which implies $s \leq \frac{-\lambda_n}{d-\lambda_n}$. □

We note that there exists a generalization of this result with a somewhat clearer proof using the Laplacian matrix. This is the specific case of a general phenomenon: for regular graphs, many reasonable choices for M_G all have "equivalent" spectrums so one can deduce information using any of these matrices. Moreover, the answer you get at the end may suggest which M_G you should use for general graphs (e.g. $d - \lambda_n$ is an eigenvalue of the Laplacian of a d-regular graph, so one might suspect that this is the right way to approach the problem).

We've now seen a couple of combinatorial properties that follow from the eigenvalues of A, but what about its limitations? To this end we say that two graphs G_1, G_2 are cospectral (with respect to A) if $\lambda_i(G_1) = \lambda_i(G_2)$ for all i.

Example 2.13. Let $G_1 = K_{2,2} \sqcup K_1$ and $G_2 = K_{4,1}$. Then the eigenvalues for both of these graphs are $\{-2, 0, 0, 0, 2\}$, so G_1 and G_2 are cospectral.

Corollary 2.14. *From the eigenvalues of A it is impossible to determine if G is connected, contains a C_4, etc.*

In general, when studying some matrix M associated to a graph, looking for cospectral graphs is a good way to understand the limitations of a spectral theory using M. There are a number of tools that can be used to construct cospectral graphs, or simply to verify that two graphs are cospectral.

Theorem 2.15. [19] *Let T be a forest with n vertices and let a_k be the number of k-element matchings in T. Then the characteristic polynomial $p(x)$ of A_T satisfies*

$$x^n - a_1 x^{n-2} + \cdots + (-1)^{\lfloor n/2 \rfloor} a_{\lfloor n/2 \rfloor} x^{n-2\lfloor n/2 \rfloor}.$$

In particular, two forests T_1 and T_2 are cospectral iff they have the same number of k-matchings for all k.

2.2 Computing Eigenvalues of Graphs

We close by showing off various techniques one can use to actually compute the spectrum of various graphs. The simplest is to just use linear algebra.

Proposition 2.16. *If K_n is the complete graph on n vertices, then $\sigma(A) = \{n-1, (-1)^{(n-1)}\}$.*

Proof. Indeed, it is not difficult to see that $\mathbf{1}$ is an eigenvector corresponding to $n-1$. Further, the vectors which have a 1 in position 1, a -1 in some other position, and 0's everywhere else are linearly independent and correspond to an eigenvalue of -1. □

In principle this method always works, but guessing the eigenvectors may not be obvious in general. We can compute eigenvalues in a more combinatorial way by using a converse of Corollary 2.8.

Proposition 2.17. *If G is a graph and there exist real numbers $\{\alpha_i\}_{i=1}^n$ such that G has $\sum \alpha_i^k$ closed walks for all k, then the $\{\alpha_i\}$ are a permutation of the eigenvalues of A.*

One proof of this uses generating functions [24]. Here we present a more direct argument.

Proof. Let the α_i be as stated, and without loss of generality assume that $\alpha_1 \geq \cdots \geq \alpha_n$. By assumption and Corollary 2.8, we have $\sum \alpha_i^k = \sum \lambda_i^k$ for all k. Let ℓ be the number of i such that $\alpha_1 = |\alpha_i|$ and $\epsilon = \max_{|\alpha_i| \neq \alpha_1} |\alpha_i|/\alpha_1$. Then, for all k,

$$\sum (\lambda_i/\alpha_1)^k = \sum (\alpha_i/\alpha_1)^k \leq \ell + \epsilon^k n,$$

and similarly for even k we find $\sum (\lambda_i/\alpha_1)^k \geq \ell$. From these inequalities it is not too difficult to see that we must have $|\lambda_i| \leq \alpha_1$ for all i, and further that equality must hold for exactly ℓ choices of i. Further, if ℓ_+ is the number of i such that $\alpha_1 = \alpha_i$ and $\ell_- = \ell - \ell_+$, then for odd k one can argue that

$$\ell_+ - \ell_- - \epsilon^k n \leq \sum (\lambda_i/\alpha_i)^k \leq \ell_+ - \ell_- + \epsilon^k n.$$

From this we conclude that there are exactly ℓ_+ choices of i with $\lambda_i = \alpha_1$ and ℓ_- choices with $\lambda_i = -\alpha_1$. The result follows by iterating this argument. □

Corollary 2.18. *If $K_{m,n}$ is the complete bipartite graph, then we have $\sigma(A) = \{\sqrt{mn}, -\sqrt{mn}, 0^{(m+n-2)}\}$.*

Proof. It is not difficult to prove that the number of closed walks of length k of $K_{m,n}$ is 0 if k is odd and $2(mn)^{k/2}$ otherwise, and the result follows from Proposition 2.17. □

Representation theory gives another way of giving us eigenvalues of graphs.

Theorem 2.19. *Let Γ be an abelian group and $S \subseteq \Gamma$ a set closed under inverses. Define the Cayley graph $G = G(\Gamma, S)$ by $V(G) = \Gamma$ and $uv \in E(G)$ iff $u - v \in S$.*
If $\chi : \Gamma \to \mathbb{C}$ is a character of Γ, then the vector with $x_a = \chi(a)$ is an eigenvector of A with eigenvalue $\frac{1}{|S|}\sum_{s \in S} \chi(s)$

We omit the proof of this result, see [25].

Corollary 2.20. *If C_n is the cycle of length n, then its eigenvalues are $\cos(2\pi r/n)$ for all $0 \leq r \leq n-1$.*

Proof. C_n is a Cayley graph with $\Gamma = \mathbb{Z}/n\mathbb{Z}$ and $S = \{-1, +1\}$ and has the characters $\chi(x) = e^{2\pi i r x/n}$ for $0 \leq r \leq n-1$. Thus its eigenvalues are

$$\frac{1}{2}e^{2\pi i r/n} + \frac{1}{2}e^{-2\pi i r/n} = \cos(2\pi r/n).$$

□

Corollary 2.21. *If Q_n is the n-dimensional cube, then its eigenvalues are $\binom{n}{i}$ with multiplicity $n - 2i$ for all $0 \leq i \leq n$.*

Proof. Q_n is a Cayley graph with $\Gamma = (\mathbb{Z}/2\mathbb{Z})^n$ and $S = \{e^1, \ldots, e^n\}$, where e^i has a 1 in position i and 0's everywhere else. For each $r \in \{0,1\}^n$ we have the character $\chi(x) = (-1)^{\sum r_i x_i}$. If we let $|r| = \sum r_i$, then we have the eigenvalues

$$\frac{1}{n}\sum \chi(e^i) = \frac{1}{n}\sum (-1)^{r_i} = \frac{1}{n}(-|r| + (n - |r|)) = 1 - \frac{2|r|}{n}.$$

□

Finally, we demonstrate how eigenvalues of larger graphs can be used to find eigenvalues of smaller graphs.

Proposition 2.22. *If P_n is the path graph on n vertices, then its eigenvalues are $2\cos(\pi r/(n+1))$ for all $1 \leq r \leq n$.*

Proof. Consider C_{2n+2} and its eigenvector $x^{(r)} = (\omega, \omega^2, \ldots, \omega^{2n+2})$ with $\omega = e^{\pi ir/(n+1)}$, which has eigenvalue $2\cos(\pi r/(n+1))$. Note that $x^{(-r)}$ has the same corresponding eigenvalue, and hence so does the vector $y^{(r)} := x^{(r)} - x^{(-r)}$. Note that $y^{(r)}_{n+1} = y^{(r)}_{2n+2} = 0$. Upon deleting these two vertices, we are left with the disjoint union of two P_n's and $y^{(r)}$ restricted to this graph is an eigenvector with the same corresponding eigenvalue. □

For more on this, and the idea of an "equitable partition," see [10].

3 The Laplacian Matrix of a Graph

In this section, we will introduce another commonly studied matrix related to a graph known as its Laplacian matrix $L(G)$. Although it is directly related to the adjacency matrix $A(G)$ and the degree matrix $\Delta(G)$, we will give a more round about definition to motivate its definition. To this end, we first examine the incidence matrix of a graph:

Definition 3.1. Let $G = (V, E)$ be a graph. Then let $B(G) \in \{0,1\}^{V \times E}$ be defined by

$$(B(G))_{v,e} = 1 \iff v \in e.$$

For example, if $V(G) = \{1,2,3,4\}$ and $E(G) = \{13, 34, 12, 23\}$, then the incidence matrix is as follows:

$$B(G) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

This is actually a particular case of the incidence matrices we saw last quarter in the proof of Frankl-Wilson. We will now consider an arbitrarily orientation σ of the edges. It turns out that what we will do will be independent of our choice of orientation.

Definition 3.2. Let $G = (V, E)$ be a graph. Then let $D_\sigma(G) \in \{-1, 0, 1\}^{V \times E}$ be defined by

$$(D_\sigma(G))_{v,e} = 1 \iff e = (u \to v)$$

$$(D_\sigma(G))_{v,e} = -1 \iff e = (v \to u).$$

For example, if $V(G) = \{1,2,3,4\}$ and $E_\sigma(G) = \{1 \to 3, 4 \to 3, 1 \to 2, 2 \to 3\}$, then the incidence matrix is as follows:

$$D_\sigma(G) = \begin{bmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

We will consider $D_\sigma(G)D_\sigma(G)^T \in \mathbb{R}^{V \times V}$ where all of the entries will be integer-valued, but we will think of them as real-valued for spectral reasons.

Our above example then yields that

$$D_\sigma(G)D_\sigma(G)^T = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{and}$$

$$D_\sigma(G)D_\sigma(G)^T = \Delta(G) - A(G).$$

It is not too difficult to see that this holds regardless of our orientation σ of the edges. This is most easily seen by example and trying to understand each entry of $D_\sigma(G)D_\sigma(G)^T$. As a result of this invarience, we make the following definition.

Definition 3.3. Let $G = (V, E)$ be a graph. Then, the Laplacian of G is defined as

$$L(G) := \Delta(G) - A(G).$$

Lemma 3.4. *Let $G = (V, E)$ be an n-vertex graph. Then, for any $x \in \mathbb{R}^n$,*

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

where x_u is the $u \in V$ coordinate of the vector x.

Proof. Take some arbitrary orientation σ and consider $D_\sigma(G)$. Then

$$x^T D_\sigma(G)D_\sigma^T(G)x = \langle D_\sigma^T x, D_\sigma^T x\rangle = \sum_{uv \in E} (x_u - x_v)^2$$

as the only terms which remain in the inner product correspond to edges in the graph G. □

Lemma 3.4 is in some sense an l^2 type result and gives many useful corollaries. We list a few below.

Corollary 3.5. *$L(G)$ is positive semi-definite. In particular, all of its eigenvalues are non-negative.*

Corollary 3.6. *The vector $\mathbf{1}$ is an eigenvector with eigenvalue 0.*

Corollary 3.7. *Let $c(G)$ be the number of connected components in the graph G. Then*

$$rank(L(G)) = n - c(G).$$

Corollary 3.8. *The eigenvalues of the Laplacian $\lambda_1 \leq \cdots \leq \lambda_n$ are monotone in the the sense that if $G \subset G'$, then $\lambda_i(G) \leq \lambda_i(G')$.*

Definition 3.9. Let $G = (V, E)$ be a graph. Then $\kappa(G)$ is the minimum $r \geq 0$ so that there exists a set S of r-vertices such that $G[V \setminus S]$ is not connected.

Corollary 3.10. $\lambda_2(L(G)) \leq \kappa(G)$

The second smallest eigenvalue λ_2 is often referred to as the spectral gap.

3.1 The Matrix Tree Theorem

In this subsection, we will consider the problem of determining the number of spanning tree's that a graph has. To this end, given a graph $G = (V, E)$, let $\tau(G)$ denote the number of spanning trees of G. Given a vertex $u \in V$, we let $L(G)[u]$ denote the Laplacian matrix of G where we delete the row and column which correspond to the vertex u.

Theorem 3.11. *Let $G = (V, E)$ be an n vertex graph with Laplacian $L(G)$ which has eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then for any vertex $u \in V$,*

$$\tau(G) = \det(L(G)[u]) = \frac{1}{n} \prod_{i=2}^{n} \lambda_i$$

The proof of Theorem 3.11 is by induction and by using deletion and contraction. Given a graph $G = (V, E)$ and an edge $e \in E$, we define deletion of an edge by $G \setminus e = (V, E \setminus e)$ to be the graph obtained by simply deleting the edge e. Given a graph G and an edge $e \in E$, we define G/e to be the graph obtained by identifying the vertices u and v where $e = uv$ into one vertex u which is adjacent to any vertex adjacent to u or v in G. Then, as any spanning tree of G either contains e or does not contain e,

$$\tau(G) = \tau(G \setminus e) + \tau(G/e).$$

As we have that $L(G) = \Delta(G) - A(G)$, in the case where G is k-regular (and hence $\Delta(G) = kI$) and the adjacency matrix has eigenvalues $\lambda_1, \ldots, \lambda_n$, then the Laplacian matrix has eigenvalues $k - \lambda_1, \ldots, k - \lambda_n$. This allows us to use many of the calculations from Section 2 to compute eigenvalues for the Laplacian.

For the complete graph, a $n-1$-regular graph, the spectrum of the Laplacian is $\{0, (n-1)^{n-1}\}$ and hence Theorem 3.11 yields

$$\tau(K_n) = n^{n-2}.$$

For the cube, which is n-regular, the spectrum of the Laplacian consists of the eigenvalues $2i$ with multiplicity $\binom{n}{i}$.

4 Towards a Theoretical Foundation for Laplacian-Based Manifold Methods

Most of the classical techniques in supervised learning consist of approximation of our data with linear spaces(for example principal component analysis). As you can imagine, approximating data linearly is not always a good approach. There is a trend in the area where instead of assuming our data is close to a plane, it is assumed to lie on a Riemannian manifold. At first it is not entirely obvious what the advantages of doing this are. Sure, we can all agree that our manifold might be a more accurate description of the solution to the problem, but how does such a complicated structure will allow us to get a concrete and easily computable solution.

It turns out that in a Riemannian manifold the heat equation allows us to see how a function evolves locally around a point. Technically speaking, the values of the function around our point diffuse respecting the solution to the heat equation. This means that a good understanding of the solutions of the heat equation on our manifold, might be the key to understand evolution of our function around a point. Now, our heat equation on a Riemannian manifold is defined in terms of the Laplace-Beltrami operator. Then, if we want to have a discrete description of the

solutions, these have to come from a discrete description of the heat equation and, particularly, of the Laplace-Beltrami operator.

In order to approximate our manifold from our data cloud we build a graph G, where the edges correspond to some weights that mimic our notion of distance in the manifold. It turns out that for our Laplace-Beltrami operator we pick the normalized Laplacian of G. Finally, the analogue to our diffusion process on the manifold is the approximations given by our diffusion maps. In [**?**], Belkin and Niyogi show why our choice of the Laplacian of G as our Laplace-Beltrami operator is correct: the graph Laplacian converges[1] to the Laplace-Beltrami operator.

4.1 Differential Geometry

We will begin by giving the basic definitions related to differential geometry. A *smooth manifold* M of dimension d is a topological space with five properties:

4.1. For every two points x, y I can separate them with two open sets; in other words, I can find disjoint open sets U_x and U_y so that $x \in U_x$ and U_y.

4.2. It is second countable, which only means that your open sets can be built from a countable collection of them.

4.3. It can be covered by open sets $\{U_\alpha\}_\alpha$ such that for each α there is a homeomorphism ϕ_α between U_α and some open subset of $\mathbb{R}^d$.

4.4. For each $\alpha \neq \beta$ the map $\phi_\alpha \circ \phi_\beta^{-1}$ is smooth.

4.5. The collection of (U_α, ϕ_α) is maximal with respect to the latter two requirements.

What all this is saying is that locally, up to a homeomorphism, a manifold looks like euclidean space. More explicitly, for every $p \in M$ one can define its *tangent space at* p, T_pM, by $\mathbb{R}^d$ (the tangent space can be thought of as the local approximation of the manifold by a flat space). It was showed by Nash that every smooth manifold can be embedded into a euclidean space. For the rest of this talk we will assume our manifold M is embedded on some euclidean space $\mathbb{R}^N$ and think of its tangent space as an affine subspace of $\mathbb{R}^N$.

A *Riemannian manifold* is a smooth manifold together with a metric dist_M which gives us a notion of distance, and fortunately that of geodesics. It is of great interest to us to study the relationship between lines on the tangent space that go through the origin and geodesics on M, fortunately the *exponential map* $\exp_p : T_pM \to M$ allows us to do so. Here we will not be explicit on what the exponential map is and rely only on the property aforementioned.

Finally, since our manifold was embedded on $\mathbb{R}^N$ we can compare the two metrics, in the case where M is compact we have the following

$$\text{dist}_M(x, y) = \|x - y\| + O(\|x - y\|^3).$$

In $\mathbb{R}^N$ we can define the *Laplace operator* as

$$\Delta f(x) = \sum \frac{\partial^2 f}{\partial x_i^2}(\mathbf{x}).$$

[1] It will become clear what we mean by convergence.

We say that a function $u(\mathbf{x}, t)$ satisfies the *heat equation* if

$$\frac{\partial}{\partial t}u(\mathbf{x}, t) - \Delta u(\mathbf{x}, t) = 0.$$

The latter, describes a diffusion of heat with initial distribution u. The solution to the heat equation is given by a semi-group of heat operators $\mathbf{H}^t$. Given an initial heat distribution f, $\mathbf{H}^t(f)$ is the heat distribution at time t. The explicit formulation of $\mathbf{H}^t(f)$ is given by

$$\mathbf{H}^t f(\mathbf{x}) = \int_{\mathbb{R}^N} f(\mathbf{y}) H^t(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$$
$$H^t(\mathbf{x}, \mathbf{y}) = \frac{1}{(4\pi t)^{\frac{N}{2}}} e^{\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{4t}}.$$

Which has as a consequence that

$$f(\mathbf{x}) = \lim_{t\to 0} \mathbf{H}^t f(\mathbf{x}).$$

In the case of a Riemannian manifold a similar formulation is achieved with the second order differential operator Δ_M, the Laplace-Beltrami operator. As before, we have the heat equation

$$\Delta_M \mathbf{H}^t_M(f) = \frac{\partial}{\partial t}\mathbf{H}^t_M(f)$$

and the approximation

$$f = \lim_{t\to 0} \mathbf{H}^t f.$$

We should remark that computing an exact form of the solution to the heat equation in a manifold is really difficult.

Now, after being able to clarify the intuition on manifolds, we are ready to set up the setting for the problem in question. Let $S = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \subset \mathbb{R}^N$, we construct a graph G whose vertices are data points. Let $W^t_n(i, j) = e^{\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{4t}}$ be the entries of the matrix W^t_n, the adjacency matrix of G. The corresponding Laplacian of G is given by $L^t_n = D^t_n - W^t_n$ where $D^t_n(i, i) = \sum W^t_n(i, j)$. Here we should think of the Laplacian as an operator on functions defined on our data set where for each $f : V \to \mathbb{R}$,

$$L^t_n f(\mathbf{x}_i) = f(\mathbf{x}_i) \sum_j e^{\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{4t}} - \sum_j f(\mathbf{x}_j) e^{\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{4t}}.$$

This operator can be naturally extended to an integral operator $\mathbf{L}^t_n$ on functions in $\mathbb{R}^N$ with respect to the empirical measure of the data set, where for $f : \mathbb{R}^N \to \mathbb{R}$ we have

$$L^t_n f(\mathbf{x}) = f(\mathbf{x}) \sum_j e^{\frac{-\|\mathbf{x} - \mathbf{x}_j\|^2}{4t}} - \sum_j f(\mathbf{x}_j) e^{\frac{-\|\mathbf{x} - \mathbf{x}_j\|^2}{4t}}.$$

Therefore, by construction,

$$\mathbf{L}^t_n f(\mathbf{x}_i) = L^t_n f(\mathbf{x}_i).$$

We call $\mathbf{L}^t_n$ the *Laplacian operator associated to the point cloud* S_n.

Notice that by our construction of $\mathbf{L}^t_n$, we get that we can approximate $\Delta(f)$ by

$$\hat{\Delta}(f) = \frac{(4\pi t)^{-\frac{k+2}{2}}}{n} \mathbf{L}^t_n f(\mathbf{x}).$$

This tells us that we can approximate numerically the value of the Laplace operator on $\mathbb{R}^N$ with the graph Laplacian of a big enough data set.

Finally I present the main result in [2], where the authors are able to give an accurate approximation of the Laplace-Beltrami operator of any manifold through Laplacians of graphs:

Theorem 4.1. [2] *Let data points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_1$ be sampled from a uniform distribution on a manifold $M \subset \mathbb{R}^N$. Put $t_n = n^{-\frac{1}{k+2+\alpha}}$, where $\alpha > 0$ and let $f \in C^\infty(M)$. Then there is a constant C, such that in probability,*

$$\lim_{n\to\infty} C\frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n}\boldsymbol{L}_n^t f(\boldsymbol{x}) = \Delta_M f(\boldsymbol{x}).$$

5 The Normalized Laplacian

Last week we saw that the Raleigh quotient for the Laplacian matrix was essentially

$$\frac{\sum_{u\sim v}(x_u - x_v)^2}{\sum x_u^2}$$

and that this was a useful tool in determining the eigenvalues of the Laplacian. Observe that the denominator of this quotient has each x_u appearing exactly once, while the numerator has each appearing d_u times. To balance things out, consider a matrix which gives the quotient

$$\frac{\sum_{u\sim v}(x_u - x_v)^2}{\sum d_u x_u^2},$$

so now each vertex is given weight according to its degree.

To achieve this quotient, let G be a graph without isolated vertices[2], D its diagonal matrix of degrees, and L its Laplacian matrix. Define the normalized Laplacian

$$\mathcal{L} := D^{-1/2}LD^{-1/2} = I - D^{-1/2}AD^{-1/2}.$$

Equivalently $\mathcal{L}_{u,u} = 1$, $\mathcal{L}_{u,v} = 1/\sqrt{d_u d_v}$ if $u \sim v$ and $\mathcal{L}_{u,v} = 0$ otherwise. Becuase $\mathcal{L}$ is real symmetric, it has real eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. Moreover, we have the following Raleigh quotient to work with.

Lemma 5.1. *Let y be a real vector and $x = D^{1/2}y$. Then*

$$\frac{y^T\mathcal{L}y}{y^Ty} = \frac{\sum_{u\sim v}(x_u - x_v)^2}{\sum d_u x_u^2}.$$

Proof. The numerator is simply x^TLx, so that follows from the equation we worked out last week, and the denominator is easy to work out. □

Because of this formulation, we often find it easiest to work with the normalized x vectors. Thus whenever y is an eigenvector, we say that $x := D^{1/2}y$ is a corresponding harmonic eigenvector.

Corollary 5.2. *It follows that $\lambda_1 = 0$ and the all 1's vector $\mathbf{1}$ is a corresponding harmonic eigenvector and $\lambda_n \leq 2$ with equality iff G contains a bipartite component. Lastly,*

$$\lambda_2 = \max_{x:\sum d_u x_u = 0} \frac{\sum_{uv\in E(G)}(x_u - x_v)^2}{\sum d_u x_u^2},$$

and this is positive iff G is connected.

[2]This is done so that D^{-1} is well defined. One can easily modify definitions to allow for isolated vertices if desired.

Proof. It's easy to see that $\frac{\sum_{uv\in E(G)}(x_u-x_v)^2}{\sum d_u x_u^2}$ is minimized by the all 1's vector, so this is a harmoic eigenvector corresponding to $\lambda_1 = 0$.

The second bit follows by trying to maximize the Raleigh quotient and using $(x_u-x_v)^2 \leq 2(x_u^2+x_v^2)$ with equality holding iff $x_u = -x_v$ for all $u \sim v$, which defines a bipartition of the component.

For the last part, note that $D^{1/2}\mathbf{1}$ is an eigenvector corresponding to λ_1, so we can choose an eigenvector $D^{1/2}x$ corresponding to λ_2 such that $\langle D^{1/2}x, D^{1/2}\mathbf{1}\rangle = 0$. Further, if $\lambda_2 = 0$ and $x_u \neq 0$, then for the Raleigh quotient to be 0 we must have $x_v = x_u$ for every v in a component containing u, and this will not satisfy $\sum d_u x_u = 0$ if G is connected. □

So this is all well and good, but maybe defining a matrix with that Raleigh quotient wasn't motivating for you, or at the very least you might be asking what this matrix is good for. The answer to both of these quandries lie within random walks.

Again let G be an n-vertex graph without isolated vertices and $x \in \mathbb{R}^n$ be a non-negative vector with $\sum x_i = 1$. We define the simple random walk on G with initial distribution x as follows. For the 0th step of the walk, start at vertex u with probability x_u. Given that you are at vertex u at step k, uniformly at random choose a neighbor of u and walk to that vertex. That is, move to vertex v with probability $1/d_u$ if $u \sim v$. We let $x_u^{(k)}$ denote the probability that you are at vertex u after the kth step of the walk.

We wish to understand when and how quickly the $x^{(k)}$ vectors converge. It is not too difficult to see that $x^{(k+1)} = AD^{-1}x^{(k)}$ where A is the adjacency matrix of G and D is the diagonal matrix of degrees. We let $P := AD^{-1}$ be this probability transition matrix. If P had n orthogonal eigenvectors, then we could study how x converges by expanding it in a basis of eigenvectors. This turns out to be true, and an easy way to see this is with the normalized Laplacian.

Lemma 5.3. *Let $x_1, \ldots, x_n$ be a set of harmonic eigenvectors of $\mathcal{L}$ corresponding to λ_i. Then $Dx_1, \ldots, Dx_n$ are eigenvectors of P corresponding to the eigenvalues $1-\lambda_i$.*

Proof. The key idea is that P is similar to the matrix

$$M = D^{-1/2}PD^{1/2} = D^{-1/2}AD^{-1/2} = I - \mathcal{L}.$$

Thus if x_i is a harmonic eigenvector corresponding to λ_i, we have

$$PDx_i = D^{1/2}(I-\mathcal{L})D^{1/2}x = D^{1/2}(1-\lambda_i)D^{1/2}x_i = (1-\lambda_i)Dx_i.$$

□

We note that one could just as well study M and recover all the properties that the normalized Laplacian controls. However, the normalized Laplacian is somewhat more standard, has the benefit of the Raleigh quotient, and some statements are clener using the eigenvalues of $\mathcal{L}$.

Corollary 5.4. *Let $\pi := (d_1/\sum d_i, \cdots)$ and $D^{1/2}x_1, \ldots, D^{1/2}x_n$ an orthogonal set of eigenvectors of $\mathcal{L}$. Given a distribution x, let $c_i = \frac{\langle x, x_i\rangle}{\langle D^{1/2}x_i, D^{1/2}x_i\rangle}$. Then $x^{(k)} = P^k x = \pi + \sum_{i\neq 1} c_i(1-\lambda_i)^k Dx_i$. Further, $x^{(k)} \to \pi$ for all x iff G is connected and not bipartite.*

Proof. Because the $D^{1/2}x_i$ vectors are orthogonal, we have $D^{-1/2}x = \sum c_i D^{1/2}x_i$, and thus $x = \sum c_i Dx_i$. As each Dx_i vector is an eigenvector of P corresponding to eigenvalue $(1-\lambda_i)$, we have $x^{(k)} = P^k x = \sum c_i(1-\lambda_i)^k Dx_i$. In particular, for $i = 1$ we can take $x_1 = \mathbf{1}$ correponding to $1-\lambda_1 = 1$. In this case we have $\langle x, x_i\rangle = \sum x_u = 1$ and $\langle x_1, x_1\rangle = \sum d_u$, and putting this together gives $c_1 Dx_1 = \pi$. Finally, we have $|1-\lambda_i| < 1$ for all $i > 1$ iff G is connected and not bipartite by the corollary. □

Here one might argue that the better eignenvalues to use would be for M to get rid of the $1-\lambda_i$ terms. However, using $(1-\lambda_i)^k \approx e^{-k\lambda_i}$ lets us state the following result concerning the speed of convergence.

Corollary 5.5. *Let $\lambda' = \lambda_2$ if $|1-\lambda_2| \geq |1-\lambda_n|$ and $\lambda' = 2-\lambda_n$ otherwise. Then for any distribution x, we have $\left\|x^{(k)} - \pi\right\|_2 \leq e^{-k\lambda'} \frac{\max \sqrt{d_v}}{\min \sqrt{d_u}}$, i.e. the walk converges to π in the L_2 norm in roughly $k = \log(\Delta^{1/2}/\delta^{1/2})/\lambda'$ steps.*

One can also consider convergence in various other norms, and basically all of these are controlled by λ'. The definition for λ' is somewhat annoying, and one can get around this by using a lazy random walk. This is defined by performing a simple random walk where at each step you have a fifty percent probability of staying in place. Equivalently one can add d_u loops to each vertex.

Morally speaking this walk behaves in the same way as the original random walk but is twice as slow. If $\tilde{L}$ is the appropriate analog of the Laplacian matrix for this non-simple graphs, then it turns out $\tilde{L} = \frac{1}{2}\mathcal{L}$ so all its eigenvalues are halved. In particular, $\tilde{\lambda}_n \leq 1$, which means (1) it won't be the most influential eigenvalue and (2) we no longer have a periodicity issue with bipartite graphs.

If you've talked to Fan for more than five minutes, you're probably very aware of the existence of something called "the Cheeger inequality," which is an isopermietric inequality that is intimately related to the normalized Laplacian.

Recall that for $S, T \subseteq V(G)$ we let $e(S,T)$ denote the number of edges uv with $u \in S$ and $v \in T$. We let $\overline{S} = V(G) \setminus S$ and define the volume of a set $\mathrm{vol}S = \sum_{u \in S} d_u$, which one can think of as the size of a set weighted by the degrees of its vertices. We define $h_G(S) = \frac{e(S,\overline{S})}{\min\{\mathrm{vol}S, \mathrm{vol}\overline{S}}$. That is, this measures how many edges you have to delete in order to separate S from its complement, where you normalize by the total number of edges involving S (typically we think of choosing S so that $\mathrm{vol}S \leq \mathrm{vol}\overline{S}$). We define the Cheeger ratio by $h_G = \min_{S \neq \emptyset, V(G)} h_G(S)$. Thus roughly h_G measures the fewest number of edges you need to delete to separate the graph into two components.

Theorem 5.6 (Cheeger Inequality)**.**

$$\frac{h_G^2}{2}\lambda_1 \leq 2h_G.$$

Proof. For the upper bound, let S be a set and consider the vector $x_u = 1/\mathrm{vol}S$ if $u \in S$ and $x_v = -1/\mathrm{vol}\overline{S}$ otherwise. Note that $\sum d_u x_u = 1 - 1 = 0$, so $\lambda_1 \leq \sum_{u \sim v}(x_u - x_v)^2 / \sum d_u x_u^2$. Each term in the numerator will be 0 unless $u \in S$ and $v \in \overline{S}$ or the other way around. Further, $\sum_{u \in S} d_u x_u^2 = 1/\mathrm{vol}S$ and $\sum_{v \in \overline{S}} d_v x_v^2 = 1/\mathrm{vol}\overline{S}$. Define $\mathrm{vol}V = \mathrm{vol}V(G) = \mathrm{vol}S + \mathrm{vol}\overline{S}$. Thus the Raleigh quotient will be

$$\frac{e(S,\overline{S})(1/\mathrm{vol}S - 1/\mathrm{vol}\overline{S})^2}{1/\mathrm{vol}S + 1/\mathrm{vol}\overline{S}} = \frac{e(S,\overline{S})(\mathrm{vol}S - \mathrm{vol}\overline{S})^2}{\mathrm{vol}S\mathrm{vol}\overline{S}(\mathrm{vol}\overline{S} + \mathrm{vol}S)} = \frac{e(S,\overline{S})(\mathrm{vol}V - 2\mathrm{vol}S)^2}{\mathrm{vol}S\mathrm{vol}\overline{S}\mathrm{vol}V}.$$

If $\mathrm{vol}S = \min\{\mathrm{vol}S, \overline{S}\} \leq \frac{1}{2}\mathrm{vol}V$, then $\overline{S} \geq \frac{1}{2}\mathrm{vol}V$ and this inequality is at most $\frac{e(S,\overline{S})(\mathrm{vol}V)^2}{\frac{1}{2}\mathrm{vol}S(\mathrm{vol}V)^2} = 2h_G(S)$. Taking S so that $h_G = h_G(S)$ gives the result.

The lower bound is significantly more involved and we omit its proof in full. The rough idea is to let x be a harmonic eigenvector corresponding to λ_2 and then partition your vertices into those with x_u large and those with x_u small. Because x minimizes the Raleigh quotient, we should have few of the $(x_u - x_v)^2$ terms with u, v being in opposite sets (as this would make the term large), which is precisely the number of edges in a cut. □

6 Strongly Regular Graphs

Among all graphs, regular graphs are often a topic of interest, and are both fairly well studied, and make various situations easier. A special class of regular graphs are the strongly regular graphs. In some sense they are extremal examples of regular graphs, since their spectrum has only three values.

Definition 6.1. Let (n, k, λ, μ) be nonnegative integers. A graph G is a strongly regular graph with parameters (n, k, λ, μ) if G is a k-regular graph on n vertices such that

- Every pair of adjacent vertices has λ common neighbours.
- Every pair of non-adjacent vertices has μ common neighbours.

Note that this is a very strong condition, hence one should expect to be able to say a lot more about such graphs. However, in order for this to be useful, there should be a fairly large number of graphs satisfying this. So let us give some examples.

- The complete graph K_m is a strongly regular graph with $n = m$, $k = m-1$, $\lambda = m-2$. (And μ really can be anything)
- The complete bipartite graph $K_{m,m}$ works with $n = 2m$, $k = m$, $\lambda = 0$ and $\mu = m$.
- The line graphs of K_m and $K_{m,m}$ satisfy with $(n, k, \lambda, \mu) = (\binom{m}{2}, 2(m-2), m-2, 4)$ and $(n, k, \lambda, \mu) = (m^2, 2(m-1), m-2, 2)$ respectively.
- The Petersen graph is a strongly regular graph with $n = 10$, $k = 3$, $\lambda = 0$ and $\mu = 1$.
- The complement of any strongly regular graph is strongly regular with parameters $(n, n-k-1, n-2-2k+\mu, n-2k+\lambda)$.

Strongly regular graphs do not exist for any set of parameters. In particular, we have the relation

$$(n-k-1)\mu = k(k-\lambda-1).$$

To see this relation, pick any vertex v and let $S = V(G)\backslash(N(u) \cup \{u\})$. We count the number of edges between $N(v)$ and S.

- On the one hand, let $u \in S$. Since u and v are not adjacent, they have μ common neighbours, that all belong to $N(v)$. Therefore, every $u \in S$ contributes μ edges between $N(v)$ and S, for a total of $(n-k-1)\mu$.
- On the other hand, let $w \in N(v)$. Since w and v are adjacent, they have λ common neighbours in $N(v)$. Therefore, w contributes $k-\lambda-1$ edges between $N(v)$ and S, for a total of $k(k-\lambda-1)$.

Let A be the adjacency matrix of a strongly regular graph and let J be the all ones matrix. As usual, the all ones vector is an eigenvector of A with eigenvalue k. We claim that we have the relation

$$A^2 = kI + \lambda A + \mu(J - I - A).$$

Indeed, let u, v be vertices of G, then A^2_{uv} is the number of length 2 paths from u to v.

- When $u = v$, this equals $d(u) = k$ explaining the kI term.
- When $u \sim v$ this equals the number of common neighbours of u and v, which is λ, explaining the λA term.
- When $u \not\sim v$ this again equals the common neighbours of u and v, which is μ, explaining the $\mu(J - I - A)$ term.

Now, we know that $(1, 1, \ldots, 1)$ is an eigenvector of A with eigenvalue k. Let x be any eigenvector with eigenvalue $\theta \neq k$, then $x \perp (1, 1, \ldots, 1)$, hence $Jx = 0$. Therefore,

$$\theta^2 x = A^2 x = (kI + \lambda A + \mu(J - I - A))x = (k + \lambda\theta - \mu(1 + \theta))x,$$

so

$$\theta^2 = (\lambda - \mu)\theta + (k - \mu),$$

showing that θ can take only one of two values, $\rho = \frac{\lambda-\mu+\sqrt{(\lambda-\mu)^2+4(k-\mu)}}{2}$ or $\sigma = \frac{\lambda-\mu-\sqrt{(\lambda-\mu)^2+4(k-\mu)}}{2}$. In fact, this is an equivalence; the connected k-regular graphs with only 3 eigenvalues (say k, ρ, σ) are all strongly regular graphs. Indeed, for any such graph, if $x \perp (1, 1, \ldots, 1)$ we have

$$(A - \rho I)(A - \sigma I)x = 0,$$

hence $(A - \rho I)(A - \sigma I) = \tau J$, which expands as $A^2 = (\tau - \rho\sigma)I + (\rho + \sigma + \tau)A + \tau(J - I - A)$, and we can run the inverse argument as above.
One example of a strongly regular graph is the Clebsch graph, which for one appears in the proof of $R(3, 3, 3) = 17$.

Definition 6.2. The Clebsch graph is the graph formed by taking the 4 dimensional hypercube and adding edges between pairs of opposite vertices (that is, pairs of vertices of Hamming distance 4). This graph is a strongly regular graph with parameters $(16, 5, 0, 2)$. It is in fact the only one.

Theorem 6.3. *The Clebsch graph is the only strongly regular graph with parameters* $(16, 5, 0, 2)$.

The proof of this theorem uses a concept called "local eigenvalues". To introduce those, let $u \in V(G)$ and note that we can write

$$A = \begin{pmatrix} 0 & \mathbf{1}^T & 0 \\ \mathbf{1} & A_1 & B^T \\ 0 & B & A_2 \end{pmatrix}$$

Using the equation $A^2 - (\lambda - \mu)A - (k - \mu)I = \mu J$ and the fact that

$$A^2 = \begin{pmatrix} k & \mathbf{1}^T A_1 & \mathbf{1}^T B^T \\ A_1\mathbf{1} & J + A_1^2 + B^T B & A_1 B^T + B^T A_2 \\ B\mathbf{1} & BA_1 + A_2 B & A_2^2 + BB^T \end{pmatrix}$$

Therefore,

$$\begin{aligned} A_1^2 - (\lambda - \mu)A_1 - (k - \mu)I + B^T B &= (\mu - 1)J; \\ A_2^2 - (\lambda - \mu)A_2 - (k - \mu)I + BB^T &= cJ; \\ BA_1 + A_2 B &= (\lambda - \mu)B + cJ. \end{aligned}$$

Definition 6.4. An eigenvalue θ of A_i is *local* if $\theta \neq k, \rho, \sigma$, and θ has an eigenvector orthogonal to $\mathbf{1}$.

Local eigenvalues satisfy the following property.

Lemma 6.5. *Let G be a strongly regular graph with eigenvalues $k > \rho > \sigma$. Suppose that x is an eigenvector of A_1 with eigenvalue θ and $\langle x, \mathbf{1} \rangle = 0$. If $Bx = 0$ then $\theta \in \{\rho, \sigma\}$ and else we have $\rho > \theta > \sigma$.*

Proof. Since $Jx = 0$ we see that

$$(\theta - \rho)(\theta - \sigma)x = (\theta^2 - (\lambda - \mu)\theta - (k - \mu))x = \left(A_1^2 - (\lambda - \mu)A_1 - (k - \mu)I\right)x = -B^T Bx.$$

If $Bx = 0$ it is clear that $\theta \in \{\rho, \sigma\}$. Else, $(\theta - \rho)(\theta - \sigma)$ is a nonzero eigenvalue of the negative semi-definite matrix $-B^T B$, hence $(\theta - \rho)(\theta - \sigma) < 0$, proving $\rho > \theta > \sigma$. □

Of course a similar result holds for the eigenvalues of A_2. It turns out that we can actually relate the eigenvalues of A_1 and A_2.

Lemma 6.6. *Let G be a strongly regular graph with parameters (n, k, λ, μ). Let θ be a local eigenvalue of A_1 or A_2. Then $(\lambda - \mu) - \theta$ is a local eigenvalue of the other, with the same multiplicity.*

Proof. Suppose θ is a local eigenvalue of A_1 with eigenvector x. As $\mathbf{1}^T x = 0$ we have

$$A_2 Bx = (\lambda - \mu)Bx + cJx - BA_1 x = ((\lambda - \mu) - \theta)\, Bx.$$

As $\mathbf{1}^T B = (k - 1 - \lambda)\mathbf{1}^T$ we have $\mathbf{1}^T Bx = 0$, hence $(\lambda - \mu) - \theta$ is a local eigenvalue with eigenvector Bx. A similar argument holds for local eigenvalues of A_2. As $B^T B$ and BB^T are positive semi-definite, and we work with vectors such that $Bx \neq 0$ and $B^T y \neq 0$, the maps B and B^T are injections from one eigenspace to another, hence the multiplicities coincide. □

Proof of Theorem 6.3. A graph with these parameters must have eigenvalues $5, 1, -3$. Let $u \in V(G)$ and let G_2 the induced subgraph on $V(G) \backslash (\{u\} \cup N(u))$. As 0 is the only eigenvalue of $G_1 = G[N(u)] = 5K_1$, G_2 can only have eigenvalues 3 (because G_2 is cubic), $1, -3$ (as those are eigenvalues of G) and -2 (as this is $(\lambda - \mu) - 0$. If G_2 would be disconnected it would have a component isomorphic to K_4 and hence an eigenvalue -1. Hence G_2 is connected and as a result (as G_2 is not complete) it must have three eigenvalues. This implies that the spectrum of G_2 cannot be symmetric around 0, hence G_2 is not bipartite, and so -3 is not an eigenvalue of G_2. So G_2 has eigenvalues 3, 1 and -2, and hence must be strongly regular with parameters $(10, 3, 0, 1)$. It is now easy to see that G_2 is the Petersen graph. Now, each vertex in G_1 is adjacent to an independent set of size 4 in G_2 (never picking the same one twice). Since the Petersen graph has exactly 5 such independent sets we conclude that G is uniquely determined by the set of parameters. □

Strongly regular graphs are heavily intertwined with combinatorial geometry. One such example are the generalized quadrangles.

Definition 6.7. A *generalized quadrangle* is a point-line incidence structure such that

1. Any two points are on at most one line, or equivalently any two lines meet in at most one point.

2. If P is a point not on a line ℓ, then there is a unique point Q on ℓ such that there is a line through P and Q.

It is clear that the second condition implies that the structure has no triangles. Besides the exceptional case of just two lines, this also implies that we have a lot of quadrangles. Indeed, let ℓ and m be two lines intersecting in P_1 and let P_2 be a point on neither of those lines. Then there exist (unique) points $Q_1 \in \ell$ and $Q_2 \in m$ such that P_2 is collinear with these points. Then $P_1Q_1P_2Q_2$ is a quadrangle.
A special type of generalized quadrangles are ones with some regularity. In particular, a generalized quadrangle has order (s,t) if every line contains $s+1$ points and every point belongs to $t+1$ lines. The *point graph* of a generalized quadrangle is the graph whose vertices are the points and where two vertices are adjacent if the points are collinear.

Lemma 6.8. *Let G be the point graph of a generalized quadrangle of order (s,t), then it is a strongly regular graph with parameters $((s+1)(st+1), s(t+1), s-1, t+1)$.*

Proof. As every point lies on $t+1$ lines, and each of these lines contains exactly s other points, G is $s(t+1)$-regular. Additionally, if (u,v) is an edge in G, then u and v correspond to collinear points P and Q, and there are exactly $s-1$ other points on this line, and since the structure is triangle-free these are the only other points that are collinear with both P and Q. Also, if u and v are not adjacent, then they correspond to nonadjacent points P and Q. For each of the $t+1$ lines through P there is a unique point on it that is also collinear with Q, so u and v have $t+1$ common neighbours.
To count the number of vertices of G, consider a line ℓ, which has $s+1$ points on it. Through every point there are t other lines, each with s other points on it. Note that all these points are different since the structure is triangle-free, giving $st(s+1)$ more points. Finally, we can have no more other points, since every point not on ℓ is collinear with some point on ℓ. Therefore, we have $(s+1)+st(s+1) = (st+1)(s+1)$ vertices. □

We can use the framework of regular graphs to impose some conditions on s and t.

Lemma 6.9. *Let G be a connected strongly regular graph with parameters (n,k,λ,μ). Then G has eigenvalues k, ρ, σ with multiplicities 1 and*

$$m_\rho = -\frac{(n-1)\sigma + k}{\rho - \sigma}$$
$$m_\sigma = \frac{(n-1)\rho + k}{\rho - \sigma}.$$

Proof. It is clear that k has multiplicity 1, hence $m_\rho + m_\sigma = n-1$. Using the fact that the sum of the eigenvalues equals the trace of A, which is 0, this implies that $m_\rho\rho + m_\sigma\sigma + k = 0$. Solving for m_ρ and m_σ yields the result. □

Now, a generalized quadrangle of order (s,t) has eigenvalues $\frac{(s-t-2)\pm\sqrt{\Delta}}{2}$ where

$$\Delta = (s-t-2)^2+4(st+s-t-1) = (s-t)^2-4(s-t)+4+4st+4(s-t)-4 = (s-t)^2+4st = (s+t)^2,$$

so the eigenvalues are $s-1$ and $-t-1$, with multiplicities

$$\frac{st(s+1)(t+1)}{s+t} \quad \text{and} \quad \frac{s^2(st+1)}{s+t},$$

respectively. The fact that these numbers must be integers impose some conditions on $\{s,t\}$. For example, we can consider the situation where each line contains exactly 3 points.

Lemma 6.10. *If a generalized quadrangle of order $(3,t)$ exists, then $t \in \{1,2,4\}$. In particular, the total number of points is 9, 15 or 27.*

Proof. Note that $-t-1$ has multiplicity $\frac{4(2t+1)}{t+2}$, so $t+2 \mid 8t+4$. Therefore, $t+2$ also divides $8t+4-8(t+2) = -12$, showing $t \in \{1,2,4,10\}$. One can exclude $t=10$ by doing some more general bounds on the eigenvalues and multiplicities of strongly regular graphs. □

7 (n,d,λ) Graphs and Ramsey numbers

Speaking very informally, a *pseudo-random graph* $G=(V,E)$ is a graph that behaves like the random graph $G(|V|,p)$ with $p = |E|/\binom{|V|}{2}$. A frequently used approach to pseudo-randomness is to consider the eigenvalues of a graph, which leads us to a concept of (n,d,λ)*-graphs*. For more detailed introduction on pseudo-random graphs, readers are referred to the survey by Krivelevich and Sudakov [17].

Definition 7.1. An (n,d,λ)-graph is a d-regular graph on n vertices in which all eigenvalues, but the first one, are at most λ in their absolute value.

The following well known theorem provide a relation between eigenvalues and pseudo-randomness.

Theorem 7.2. *Let $G=(V,E)$ be an $(n,d,\lambda)-$graph, and $B \subset V$ such that $|B| = bn$. Then,*

$$\sum_{v \in V} (|N_B(v)| - bd)^2 \leq \lambda^2 b(1-b)n,$$

where $N_B(v)$ denote the set of all neighbors of v in B.

Proof. Let A be the adjacency matrix of G and define a vector f_B such that $f_B(v) = 1-b$ when $v \in B$ and $f_B(v) = -b$ otherwise. Note that $f_B \cdot \mathbf{1} = 0$; that is, f is orthogonal to the all 1 vector, which is the eigenvector of the largest eigenvalue of A. Therefore,

$$f^T A^2 f \leq \lambda^2 f^T f.$$

The left-hand side is

$$\|Af\|_2^2 = \sum_{v \in V} (|N_B(v)|(1-b) - (d - |N_B(v)|)b)^2 = \sum_{v \in V} (|N_B(v)| - bd)^2.$$

And the right-hand side is $\lambda^2 b(1-b)$. The desired result follows. □

Given a graph $G=(V,E)$. For two (not necessarily disjoint) subsets B and C of V, let $e(B,C)$ denote the number of ordered pairs (u,v) such that $u \in B$ and $v \in C$ and $uv \in E$.

Corollary 7.3. *Let $G=(V,E)$ be an (n,d,λ)-graph, then for every two subsets B and C of V,*

$$|e(B,C) - (d/n)|B||C|| \leq \lambda\sqrt{|B||C|}$$

Proof. Let $|B| = bn$. Then

$$\begin{aligned} |e(B,C) - (d/n)|B||C|| &\leq \sum_{v \in C} ||N_B(v)| - bd| \\ &\leq \sqrt{|C| \sum_{v \in C} (|N_B(v)| - bd)^2} \quad \text{(Cauchy's inequality)} \end{aligned}$$

By Theorem 1,

$$\sum_{v\in C}(|N_B(v)|-bd)^2 \le \sum_{v\in V}(|N_B(v)|-bd)^2 \le \lambda^2 b(1-b)n \le \lambda^2|B|$$

This completes the proof. □

Now consider the random graph $G(n,p)$, it is not hard to show by Chernoff bound that a.a.s.

$$|e(B,C)-p|B||C|| = O(\sqrt{pn|B||C|})$$

Comparing the preceding two bounds, we can see that when λ is small, especially when $\lambda = O(\sqrt{d})$, the $(n,d,\lambda)-$graphs behave like the random graph $G(n,d/n)$. In fact, this is the best possible λ we can obtain.

Proposition 7.4. *Let G be an $(n,d,\lambda)-$graph, and $d \ll n$, then $\lambda = \Omega(\sqrt{d})$.*

Hence, the $(n,d,\lambda)-$graphs with $\lambda = \Theta(\sqrt{d})$ are known as *optimally pseudo-random*.

7.1 Ramsey numbers

The *Ramsey number* $r(F,t)$ is the minimum number n such that every n-vertex F-free graph has an independent set of size t. When $F = K_s$ we simply write $r(s,t)$ instead of $r(F,t)$. For the upper bounds, Erdős and Szekeres [12] (1935) show, by a nice inductive proof, that $r(s,t) \le \binom{s+t-2}{s-1}$. In particular, if we fix s, this gives $r(s,t) \le c_s t^{s-1}$. The best upper bound is not far away from this, which is shown by Ajtai, Komlós and Szemerédi (1980):

$$r(s,t) \le c_s \frac{t^{s-1}}{(\log t)^{s-2}}$$

The best lower bound so far is shown by Bohman and Keevash [5] (2010) using a natural random K_s-free process:

$$r(s,t) \ge c_s \frac{t^{(s+1)/2}}{(\log t)^{(s+1)/2-1/(s-2)}}$$

A recent work by Mubayi and Vertraete [21] (2019) offers a new way to improve this lower bound.

Theorem 7.5. *Let F be a graph, n,d,λ be positive numbers with $d \ge 1$ and $\lambda > 1/2$ and let $t = \lceil 2n\log^2 n/d\rceil$. If there exists an F-free $(n,d,\lambda)-$graph, then*

$$r(F,t) > \frac{n\log^2 n}{20\lambda}$$

Remark 7.6. For K_s-free $(n,d,\lambda)-$graphs with $\lambda = \Theta(\sqrt{d})$, it is knonw that $d = O(n^{1-\frac{1}{2s-3}})$. If this upper bound is obtainable, that is, if there exist a $(n,d\lambda)-$graph with $\lambda = \Theta(\sqrt{d})$ and $d = \Theta(n^{1-\frac{1}{2s-3}})$, then by the theorem above, $r(s,t) = \Omega(t^{s-1}/\log^{2s-4} t)$, which is tight up to log factors. The best construction of K_s-free pseudo-random graph so far is given by Bishnoi, Ihringer and Pepe [6] with $d = \Omega(n^{1-1/(s-1)})$.

The key of the proof of Theorem 4, is the following Theorem by Alon and Rödl [1]:

Theorem 7.7. *Let G be an $(n,d,\lambda)-$graph with $d \ge 1$ and $\lambda > 1/2$. Then for any integer $t \ge 2n\log^2 n/d$, the number of independent sets of size t in G is at most $(2e^2\lambda/\log^2 n)^t$.*

To prove Theorem 5, we use the following lemma:

Lemma 7.8. *Let $G = (V, E)$ be an $(n, d, \lambda)-$graph, and let B be a subset of V such that $|B| = bn$. Define $C := \{u \in V : |N_B(u)| < bd/2\}$. Then*

$$|B \cap C| \leq 2\lambda n/d.$$

Proof. By Theorem 1, we have

$$\sum_{v \in V} (|N_B(v)| - bd)^2 \leq \lambda^2 b(1-b)n.$$

Every vertex in C contribute at least $b^2d^2/4$ to the left-hand side, hence we have

$$|C|\frac{b^2d^2}{4} \leq \lambda^2 b(1-b)n \leq \lambda^2 bn,$$

which is equivalent to

$$|B||C| \leq \frac{4\lambda^2 n^2}{d^2}$$

Note that $|B \cap C| \leq \min\{|B|, |C|\}$, the desired result follows. □

Proof of Theorem 5. Consider the following process of picking an independent set of size t: Begin with $B_0 = V$, for each $B_i \subset V$ such that $0 \leq i \leq t-1$, pick a vertex v in B_i and then let $B_{i+1} = B_i \backslash (v \cup N(v))$. Let $C_i = \{u \in V : |N_{B_i}(u)| < |B_i|d/2n\}$. Note that if in step i, the selected vertex v that is not in C_i, then we must have $|B_{i+1}| \leq (1 - d/2n)|B_i|$. We call such a choice a *bad pick*, and otherwise a *good pick*. The number of bad picks cannot be more than $2n \log n/d$, otherwise B_i will be empty before the process terminates. Let $l = t/\log n$, then we count the independent sets of size t as following: we first choose an index set of size l where the picks can be either good or bad. Then for the remaining $t - l$ picks, by Lemma 6 each of them have at most $2\lambda n/d$ choices. In this way, the number Z of independent set of size t is at most

$$\frac{1}{t!}\binom{t}{l} n^l \left(\frac{2\lambda n}{d}\right)^{t-l}$$

Using $\binom{t}{l} \leq 2^t$ and $t! \geq (t/e)^t$,

$$Z \leq \left(\frac{4e\lambda n}{td}\right)^t \left(\frac{d}{2\lambda}\right)^l$$

Since $\lambda > 1/2$, we have $d/2\lambda < d < n$ and hence $(d/2\lambda)^l < n^{t/\log n} = e^t$. Using $t \geq 2n(\log^2 n)/d$,

$$Z \leq \left(\frac{4e^2\lambda n}{dt}\right)^t \leq \left(\frac{2e^2\lambda}{\log^2 n}\right).$$

This completes the proof. □

With Theorem 5, we are ready the prove Theorem 4. The idea is to randomly sampled vertices with some suitable probability and then remove a vertex from each independent set of size t.

Proof of Theorem 1. Let G be an F-free $(n, d, \lambda)-$graph. Randomly sample a subset U of V with probability $p = (\log n)^2/2e^2\lambda$. Let Z be the number of independent sets of size t in the subgraph of G induced by U. Then we can obtain an F-free graph with no independent set of size t by removing one vertex from each independent set of size t. By Theorem 5 and the choice of p, the expected number of remaining vertices is at least

$$\mathbb{E}[|U| - |Z|] \geq pn - p^t \left(\frac{2e^2\lambda}{\log^2 n}\right)^t = pn - 1$$

Therefore, we have

$$r(F, t) \geq pn > \frac{n \log^2 n}{20\lambda}.$$

□

8 Ramanujan Graphs

Given the numerous applications of expander graphs, one might wish to know what the "best" expanders are; these graphs are called Ramanujan Graphs. In this section, we will present a proof of the existence of these graphs following the novel treatment of [18].

Next, we consider a random walk on the vertices of G, where we move from v to one of its neighbors with probability $1/\deg(v) = 1/d$. Then the transition matrix for this Markov chain is precisely $\frac{1}{d}A(G)$, and by the basic theory of Markov chains, the largest eigenvector of $\frac{1}{d}A$ is the stationary distribution of the random walk. In our case, that stationary distribution assigns weight $1/n$ to each vertex. If we start with a distribution with all its weight on one vertex, and run the Markov chain, how quickly does the distribution converge to the stationary distribution? As it turns out, the spectrum of a our graph determines this rate of convergence.

Theorem 8.1. *Given the random walk on G as above, let 1_v be a starting distribution with weight only on the vertex v, and let $\frac{1}{n}\mathbf{1}$ be the stationary distribution. Then, the distribution of the random walk after t steps is $\left(\frac{1}{d}A\right)^t 1_v$ and*

$$\|d^{-t}A^t 1_v - \frac{1}{n}\mathbf{1}\|_2 \leq C\left(\frac{\max\{|\lambda_2|, |\lambda_n|\}}{d}\right)^t$$

for some constant C depending on the graph.

Proof. The first claim can be seen from the definition of the transition matrix, so we do not prove it here. For the second claim, let $\mathbf{1}, v_2, \ldots, v_n$ be the eigenvectors of A and let $1_v = \sum_{i=1}^n \alpha_i v_i$.

$$\|d^{-t}A^t 1_v - \frac{1}{n}\mathbf{1}\|_2 = \|d^{-t}A^t\left(\sum_{i=1}^n \alpha_i v_i\right) - \frac{1}{n}\mathbf{1}\|_2 = \|\sum_{i=1}^n \left(\frac{\lambda_i}{d}\right)^t \alpha_i v_i - \frac{1}{n}\mathbf{1}\|_2$$

Noting that $\alpha_1 = \langle 1_v, \mathbf{1}\rangle = \frac{1}{n}$,

$$\|\sum_{i=1}^n \left(\frac{\lambda_i}{d}\right)^t \alpha_i v_i - \frac{1}{n}\mathbf{1}\|_2 \leq \|\left(\frac{\max_{i\neq 1}\lambda_i}{d}\right)^t \sum_{i=2}^n \alpha_i v_i\|_2 \leq C\left(\frac{\max_{i\neq 1}|\lambda_i|}{d}\right)^t$$

□

Thus we define the following

Definition 8.2. For a graph G, its *spectral gap* $\lambda(G)$ is $\max_{i \neq 1} |\lambda_i| = \max\{|\lambda_2|, |\lambda_n|\}$.

Thus, the smaller the spectral gap, the quicker a random walk on G will converge to its limiting distribution.

Now that we have isolated the spectral gap as an interesting graph property, it is natural to ask how small can the spectral gap can be relative to λ_1. The complete graph on n vertices has $\lambda = 1$ and $d = n - 1$. But the degree of this example grows with the size of the graph, so a better question is how small can λ be for a fixed value of λ_1, i.e. for a family of d-regular graphs? Here we have a much more interesting result

Theorem 8.3. [22] *For an infinite family of d-regular graphs,*

$$\lim_{n \to \infty} \lambda(G) \geq 2\sqrt{d-1} - o(1)$$

Definition 8.4. A d-regular graph is *Ramanujan* if

$$\lambda(G) = 2\sqrt{d-1}$$

As always, we want to know if this bound is tight, or equivalently, if Ramanujan graphs exist. Thankfully, the answer is yes.

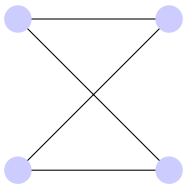
Theorem 8.5. [16] *For every prime p, there exists an infinite family of $p+1$-regular Ramanujan graphs.*

Unfortunately, the analysis of these constructions involve some rather non-trivial number theory results, and the degree restriction is unnatural for many applications in combinatorics and computer science.

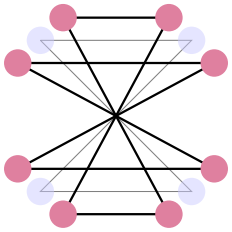
8.1 Lifts

In the years of work following [16], one particular result from [3] proposed a novel construction for Ramanujan graphs. They considered random *lifts* of a graph, illustrated as follows.

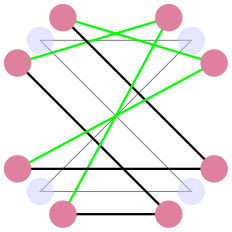
Start with a fixed graph G



Double every vertex, and replace each edge with a matching



Switch each matching at random



More formally,

Definition 8.6. Given a graph G, a *signing* of G is a map

$$s : E(G) \to \{\pm 1\}$$

Definition 8.7. Given a graph $G = (V, E)$ and a signing s of G, the 2-lift of G corresponding to s is the graph $G_2 = (V_2, E_2)$ where $V_2 = V \times \{\pm 1\}$ and

$$(v, s_v) \sim (u, s_u) \iff v \sim u \text{ and } s_v = s((u, v)) * s_u$$

.

How can these lifts give a Ramanujan graph? Bilu and Linial prove the following

Theorem 8.8. [3] *Given a d-regular graph G, there exists a 2-lift of G, G_2 such that*

$$\lambda(G_2) \leq 2\sqrt{d-1} + O(\sqrt{d \log^3(d)})$$

and furthermore they conjectured

Conjecture 8.9. [3] Given a d-regular graph G, there exists a 2-lift of G, G_2 such that

$$\lambda(G_2) = 2\sqrt{d-1}$$

Finally, we get to the main result of today's talk:

Theorem 8.10. [18] *Given a d-regular graph G, there exists a 2-lift of G, G_2 such that*

$$\lambda_2(G_2) \leq 2\sqrt{d-1}$$

Note that this theorem gives no control over λ_n. However, by as bipartite graphs have symmetric spectrum, it follows that:

Corollary 8.11. *Given a d-regular bipartite G, there is a 2-lift of G, G_2 such that*

$$\lambda(G_2) \leq 2\sqrt{d-1}$$

By iteratively applying Corollary 8.11, we achieve our desired result:

Theorem 8.12. *For any $d \geq 3$, there exists an infinite family of d-regular bipartite Ramanujan graphs.*

8.2 Interlacing Families

Recall our definitions of signings (8.6) and 2-lifts (8.7).

Definition 8.13. Given a graph G and a signing s, the *signed adjacency matrix* A_s is defined by

$$(A_s)_{ij} = \begin{cases} s((i,j)) & i \sim j \\ 0 & i \not\sim j \end{cases}$$

Lemma 8.14. [3] *Let G_2 be the 2-lift of a graph G defined by the signing s. Then,*

$$\sigma(A(G_2)) = \sigma(A) \cup \sigma(A_s)$$

where the union is taken with multiplicity.

Thus, if G is a Ramanujan Graph, in order to analyze its 2-lifts, we simply have to analyze the eigenvalues of A_s for every signing s. To do so, we introduce the following polynomial.

Definition 8.15. Given a graph G, its *matching polynomial* $\mu_G(x)$ is

$$\mu_G(x) = \sum_{i=0}^{n/2} x^{n-2i}(-1)^i m_i(G)$$

where $m_i(G)$ is the number of matchings with i edges.

The matching polynomial is rather well-studied, and in particular we need these facts

Theorem 8.16. [11] *All the roots of $\mu_G(x)$ are real.*

Theorem 8.17. [11] *If G is d-regular, the roots of μ_G lie in the interval*

$$[-2\sqrt{d-1}, 2\sqrt{d-1}]$$

The $2\sqrt{d-1}$ is exactly the bound on the second largest root that we want to show! So it would be nice if we could relate the eigenvalues of A_s for some signing to the roots of μ_G. The next lemma shows that, in expectation, A_s and μ_G have exactly the relationship we might want.

Lemma 8.18. *Let G be a graph, and let s be a uniformly random signing. Then,*

$$\mathsf{E}_s\left[\det(xI - A_s)\right] = \mu_G(x)$$

Proof. We expand the determinant as a sum over permutations and apply linearity of expectation:

$$\mathsf{E}_s\left[\det(xI - A_s)\right] = \sum_{\sigma \in \mathfrak{S}_n} (-1)^{\mathrm{sign}(\sigma)} \mathsf{E}_s\left[\prod_{i=1}^{n} (xI - A_s)_{i,\sigma(i)}\right]$$

As the signing on each edge is chosen independently

$$= \sum_{\sigma \in \mathfrak{S}_n} (-1)^{\mathrm{sign}(\sigma)} \prod_{i=1}^{n} \mathsf{E}_s\left[(xI - A_s)_{i,\sigma(i)}\right]$$

As the signing of each edge is 0 in expectation, the only permutations with a nonzero contribution are composed of cycles of length at most 2. These permutations thus correspond to matchings of size i. Finally, note that the sign such a permutation is precisely i

$$= \sum_{i=0}^{n} x^{n-2i}(-1)^i m_i(G)$$
$$= \mu_G(x)$$

□

So all we have to do is show that some A_s has its largest root bounded by the largest root of $\mathsf{E}_s[\det(xI - A_s)]$. However, there is no reason to expect such a relationship as the roots of polynomials do not behave nicely under addition.

Example 8.19. Consider $p(x) = x^2 - 3x + 1$ and $q(x) = x^2 + 3x + 1$, both of which are real-rooted. Their sum, $2x^2 + 2$ has only imaginary roots!

Example 8.20. Consider $r(x) = x^3 - 2x^2 + 3x + 3$ and $s(x) = x^3 - 4x^2 - 1$ with roots $-0.63, 1.31 \pm 1.171i$ and $4.066, 0.03 \pm 0.49i$ respectively. Their sum has roots $2, \frac{1}{2} \pm \frac{\sqrt{3}}{2}$.

To gain some control over the roots of sums of polynomials, we have the following key idea

Definition 8.21. Given a degree n real-rooted polynomial $f = (x-\alpha_1)(x-\alpha_2)\dots(x-\alpha_n)$ and a degree $n-1$ real-rooted polynomial $g = (x-\beta_1)(x-\beta_2)\dots(x-\beta_{n-1})$, we say that g *interlaces* f if

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \cdots \geq \beta_{n-1} \geq \alpha_n$$

The prototypical example to keep in mind that of the Cauchy Interlacing Theorem, which states that for any matrix A and any principal submatrix B of A, the characteristic polynomial of B interlaces the characteristic polynomial of A.

Definition 8.22. Given a collection of degree n real-rooted polynomials $\{f_i\}$, we say that they have a *common interlacing* if there is a degree $n-1$ real-rooted polynomial g such that g interlaces each f_i.

The upshot of interlacing is the following lemma:

Lemma 8.23. *let $\{f_i\}_{i\in I}$ be a collection of degree-n real-rooted polynomials with positive leading coefficient and a common interlacer g. Then the following hold:*

- *$F := \mathsf{E}_{i\in I}[f_i]$ is real-rooted and g interlaces F.*
- *There exists an $i \in I$ such that the largest root of f_i is at most the largest root of F*

Proof. We will only treat the case where g has no common roots with any of the f_i. The general case follows by a limiting argument. Let the roots of g be $\beta_1, \dots, \beta_{n-1}$. As the f_i all have the same degree and have positive leading coefficients, for every $i \in I$, $f_i(\beta_k) < 0$ if k is odd and $f_i(\beta_k) > 0$ if k is even. Thus, $F(\beta_k) > 0$ if k is odd and $F(\beta_k) < 0$ if k is even. Thus, by the intermediate value theorem, F must be real-rooted and g interlaces F.

For the second assertion, let α be the largest root of F, and assume for contradiction that the largest root of any f_i is strictly larger than α. As all the f_i's have positive leading coefficients, if follows that $f_i(\alpha) < 0$ for every $i \in I$. Hence, $F(\alpha) < 0$, a contradiction. □

To summarize, if we can show that $\{\det(xI - A_s)\}_{s:E(G)\to\{\pm 1\}}$ has a common interlacer, then the lemma (8.23) allows us to conclude that there is a 2-lift of G which is Ramanujan.

Question 8.24. Note that the above argument can easily generalize to show that there must be an f_i whose k-largest root is at most the k-largest root of F. Can this be furthermore generalized to show that there is an f_i such that its kth and lth largest roots are simultaneously at most the kth and lth largest roots of F?

8.3 Interlacing and Real Stable Polynomials

How do we show that a collection of polynomials has a common interlacer? Thankfully, interlacing is a well-studied concept, and we have a nice characterization

Lemma 8.25. [13] *Let $\{f_i\}_{i\in I}$ be a collection of degree n real-rooted polynomials. Then the following are equivalent*

- *$\{f_i\}_{i\in I}$ has a common interlacer.*
- *For every choice of $\lambda_i \geq 0$ with $\sum_{i\in I}\lambda_i = i$, the polynomial $\sum_{i\in I}\lambda_i f_i$ is real-rooted.*

To apply this lemma in our setting, we will show the following

Theorem 8.26. *For every assignment of weights $p_e : E(G) \to [0,1]$ to the edges of G, the following polynomial is real-rooted*

$$\sum_{s:E(G)\to\{\pm 1\}} \prod_{e:s(e)=1} p_e \prod_{e:s(e)=-1} (1-p_e) \det(xI - A_s)$$

By setting the p_e appropriately, we can then recover any convex combination $\{\lambda_e\}_{e\in E(G)}$.
The proof is based on the theory of Real-Stable Polynomials.

Definition 8.27. A polynomial $p(z_1,\dots,z_n) \in \mathbb{R}[z_1,\dots,z_n]$ is called *real-stable* if

$$f(z_1,\dots,z_n) \neq 0$$

whenever $\Im(z_i) > 0$ for every i.

Rather than dive into the deep theory of real-stable polynomials, we will instead point the interested reader to [4] for a proper treatment of the subject, and we will simply note the following facts without proof.

8.1. A real-stable polynomial in one variable is real-rooted.

8.2. If p, q are real-stable and $\alpha, \beta \geq 0$, then $\alpha p + \beta q$ is real-stable

8.3. If p is real-stable, then $\frac{\partial}{\partial z_i} p$ is real-stable for any i.

8.4. If $p(z_1,\dots,z_n)$ is real stable and $c \in \mathbb{R}$, then $p(z_1,\dots,z_{i-1},c,z_{i+1},\dots,z_n)$ is real-stable for any $i \in [n]$.

8.5. If $A_1,\dots,A_n$ are positive semidefinite matrices, then $det(z_1A_1 + \dots z_nA_n)$ is real-stable.

All of the above facts (and much more) are proven in [4]. Our first key takeaway is the following. Define Z_i to be the operator which maps $p(z_1,\dots,z_n) \to p(z_1,\dots,z_{i-1},0,z_{i+1},\dots,z_n)$

Claim 8.28. *For variables u, v and $p_{u,v} \in [0,1]$ the operator $T_{u,v} := Z_u Z_v(1 + p_{u,v}\partial_u + (1-p_{u,v})\partial_v)$ preserves real stability.*

Lemma 8.29. *Let A be an invertible $n \times n$ matrix, let u, v be variables, and let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then,*

$$T_{u,v} \det(A + u\vec{x}\vec{x}^\top + v\vec{y}\vec{y}^\top) = p_{u,v} \det(A + \vec{x}\vec{x}^\top) + (1-p_{u,v}) \det(A + \vec{y}\vec{y}^\top)$$

Proof. By the matrix determinant lemma, $\det(A+u\vec{x}\vec{x}^\top) = \det(A)(1+u\vec{x}^\top A^{-1}\vec{x})$, Thus, $\partial_u \det(A + u\vec{x}\vec{x}^\top) = \det(A)(\vec{x}^\top A^{-1}\vec{x})$. Finally, we note that $Z_v\partial_u = \partial_u Z_v$. So, expanding and applying Z_u, Z_v wherever possible.

$$T_{u,v} \det(A + u\vec{x}\vec{x}^\top + v\vec{y}\vec{y}^\top) = \det(A) + p_{u,v} Z_u \partial_u \det(A + u\vec{x}\vec{x}^\top) + (1-p_{u,v}) Z_v \partial_v \det(A + v\vec{y}\vec{y}^\top)$$

By the above

$$\begin{aligned} &= \det(A) + p_{u,v} \det(A)(\vec{x}^\top A^{-1}\vec{x}) + (1-p_{u,v}) \det(A)(\vec{y}^\top A^{-1}\vec{y}) \\ &= p_{u,v} \det(A)(1 + \vec{x}^\top A^{-1}\vec{x}) + (1-p_{u,v}) \det(A)(1 + \vec{y}^\top A^{-1}\vec{y}) \end{aligned}$$

By the matrix determinant lemma in reverse

$$= p_{u,v} \det(A + \vec{x}\vec{x}^\top) + (1-p_{u,v}) \det(A + \vec{y}\vec{y}^\top)$$

□

With this lemma in hand, we can finally prove the main theorem

Proof. We wish to show that

$$\sum_{s:E(G)\to\{\pm 1\}} \prod_{e:s(e)=1} p_e \prod_{e:s(e)=-1} (1-p_e) \det(xI - A_s)$$

is real-rooted for any assignment $p_e : E(G) \to [0,1]$. Clearly, this will follow if we show that

$$\sum_{s:E(G)\to\{\pm 1\}} \prod_{e:s(e)=1} p_e \prod_{e:s(e)=-1} (1-p_e) \det(xI + dI - A_s)$$

is real-rooted where d is the degree of G. To that end, we define for each edge $e = (a,b)$ the matrices $L_e^1 = (e_a - e_b)(e_a - e_b)^\top$ and $L_e^{-1} = (e_a + e_b)(e_a + e_b)^\top$ where e_a, e_b are standard basis vectors. Note that L_e^1, L_e^{-1} are positive semidefinite. Furthermore, for any signing $s : E(G) \to \{\pm 1\}$,

$$\det(xI + D - A_s) = \det(xI + \sum_{e\in E(G)} L_e^{s(e)})$$

Finally, consider the polynomial Q in variables $x, \{u_e\}_{e\in E(G)}, \{v_e\}_{e\in E(G)}$ defined by

$$Q := \det(xI + \sum_{e\in E(G)} u_e L_e^1 + v_e L_e^{-1})$$

By a fact above, Q is real-stable. By applying (8.29) for each edge in G, we see that

$$(T_{e_1} \dots T_{e_m})(Q) = \sum_{s:E(G)\to\{\pm 1\}} \prod_{e:s(e)=1} p_e \prod_{e:s(e)=-1} (1-p_e) \det(xI + \sum_{e\in E(G)} L_e^{s(e)})$$

As T_e preserves real stability and our final polynomial only depends on x, the theorem follows. □

8.4 Further Directions

The proof given is probabilistic in nature, and for some applications, a polynomial-time construction is necessary. Thankfully, [7] gives a polynomial-time constructive proof of the existence of a 2-lift which is Ramanujan. Another relevant issue is that 2-lifts double of the number of vertices at each step, and so the size of our Ramanujan graphs grow exponentially. This issue will be rectified in Interlacing Families 4, the next talk in the ABACUS seminar. Finally, these graphs must be bipartite, and it is a major open problem find nice constructions of non-bipartite Ramanujan graphs.

9 High Dimensional Expanders

In previous weeks, we have seen several different (but morally equivalent) types of expansion on standard graphs, and seen numerous applications of these objects throughout combinatorics. It is natural then to wonder whether these definitions generalize to hypergraphs, the standard higher dimensional variant of graphs, and whether such objects might be equally useful. In these notes we will study one such natural combinatorial generalization: Spectral Link HDX's. This version of high dimensional expansion has come onto the scene relatively recently, and has already shown a number of important applications in computer science and combinatorics (e.g. for counting matroid bases or independent sets, building locally testable codes) and seems like a promising direction to make advances on other important problems like the Unique Games Conjecture.

In these notes we will review the definition of Spectral HDX's, and cover Izhar Oppenheim's Trickle Down Theorem. This structural theorem not only simplifies analysis and understanding of spectral HDX, but also nicely exhibits their "local to global" behavior, a useful paradigm that shows how one can leverage local structure to prove global results.

We will consider a special class of hypergraphs known as pure simplicial complexes as Misha Gromov's seminal work on the topological overlapping property resulted in further studies of high dimensional expanders.

Definition 9.1 (Pure Simplicial Complex)**.** A d-dimensional pure simplicial complex is a hypergraph $H(V, E)$ satisfying three conditions:

9.1. E is downward closed

9.2. The largest edges (faces) in E are of size $d + 1$

9.3. Every face $e \in E$ is contained in at least one face of size $d + 1$.

Equivalently, one can think of a d-dimensional pure simplicial complex as the result of the downward closure of some $(d + 1)$-uniform hypergraph. Let's look at an example of 2-dimensional simplicial complex called the fish complex:

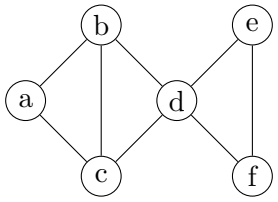


Figure 1: The Fish Complex

It is common to write a simplicial complex X by partitioning it into its i-dimensional faces for $-1 \leq i \leq d$, denoted $X(i)$:

$$X(i) = \{S \subset V \cup E \ : \ |S| = i + 1\}.$$

In this case we think of the entire complex as the union of these sets:

$$X = \bigcup_{i=-1}^{d} X(i).$$

Note that $X(-1)$ is $\{\varnothing\}$, a notational convenience that often is useful when working on simplicial complexes. Let's write these sets for the fish complex as an example:

The Fish Complex may be written as:

$$\begin{aligned} X(-1) &= \{\varnothing\} \\ X(0) &= \{a, b, c, d, e, f\} \\ X(1) &= \{(ab), (ac), (bc), (bd), (cd), (de), (df), (ef)\} \\ X(2) &= \{(abc), (bcd), (def)\} \end{aligned}$$

Perhaps a more intuitive or visual way to draw a simplicial complex is through a Hasse diagram ordered by inclusion. In fact while we will not touch on this in these notes, there is a way to extend this material to graded posets such as the Grassmann poset. Let's take a look at the fish complex from this viewpoint:

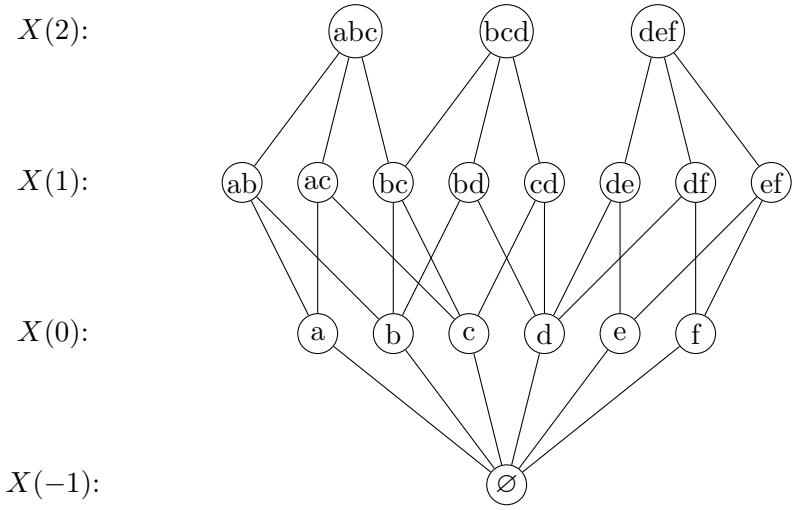


Figure 2: The Fish Complex as a Hasse Diagram.

This viewpoint gives a nice visual overview of the global structure of our simplicial complex. However, our definition of expansion is going to focus on *local* properties of the complex, so let's take a look at breaking down X into local structures. In a graph, the most fundamental local structure at a vertex v is its neighborhood, the set of vertices it is connected to. In a simplicial complex, the corresponding structure is called a *link*. Links can be defined for a face s of any dimension in the simplicial complex, and corresponds to higher dimensional faces which include s.

Definition 9.2. The link of a face $s \in X(i)$ is:

$$X_s = \{T \ : \ T \cup \{s\} \in X\}$$

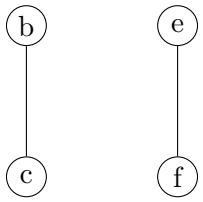


Figure 3: Link X_d of the Fish Complex

While not strictly necessary, it is convenient to lower the dimension of the link by removing s itself from each edge as we have done in the above definition. This allows us to talk about the *underlying graph* of the link, which is just $G(V, E)$ s.t. $V = X_s(0)$ and $E = X_s(1)$. At this point we've pretty much shown our cards: we would like to define global expansion through the spectral expansion of these local links across the complex.

9.1 Weighted Expansion

There is, however, an issue with applying this definition naively. Our simplicial complexes are not restricted to being regular–and at this point we have only defined spectral expansion for regular graphs, as well as how to correctly weight simplicial complexes.

A weighted graph is a graph $G(V, E)$ in which each edge is endowed with some positive real weight. These weights may be renormalized to give a distribution π_E over E (this can be seen as a generalization of the non-weighted case, which would be endowed with the uniform distribution).

We can define a random walk operator on $G(V, E)$ based on π_E that is analogous to the normalized adjacency matrix of an unweighted graph.

Definition 9.3 (Weighted Adjacency Matrix)**.** The weighted adjacency matrix (from here on just adjacency matrix) of a weighted graph $G(V, E)$ with induced distribution π_E is simply the natural random walk defined by π_E, that is:

$$A_{v,w} = \frac{\pi_E((v, w))}{\sum_{(u,w)\in E} \pi_E((u, w))}$$

We will sometimes refer to $A_{v,w}$ as $P(v|w)$, the probability of picking w given v.

It's clear that A has some of the same properties as our standard normalized adjacency matrix simply due to being stochastic. For instance we have:

9.1. The all 1's vector is an eigenvector with eigenvalue 1

9.2. All eigenvalues λ satisfy $-1 \leq \lambda \leq 1$

To define expansion analogously to the case of regular graphs, the only remaining property we need is that A is diagonalizable. As a low-dimensional warmup for Oppenheim's theorem, we will prove this with a "local to global" method. The edge distribution π_E induces a distribution over vertices π_V in the following sense:

$$\pi_V(v) = \frac{1}{2} \sum_{(w,v)\in E} \pi_E(w, v).$$

In other words, π_V can be described as the process of picking a random edge from π_E, and then choosing uniformly a vertex from that edge. To show that A is diagonalizable, we define an inner product space with respect to this distribution, and show that A is self-adjoint with respect to this inner product. In particular, for functions $f, g : V \to \mathbb{R}$, define their inner product by:

$$\langle f, g\rangle = \sum_{v\in V} \pi_V(v) f(v) g(v)$$

The key to understanding why the adjacency matrix is self adjoint is noticing that picking an edge from π_E can be done locally. In particular, drawing a random edge from π_E is the same as drawing a random vertex v from π_V, and then drawing a random neighbor w based on $P(v|w)$. Keeping

this in mind, let's expand $\langle f, Ag \rangle$:

$$\begin{aligned}
\langle f, Ag \rangle &= \sum_{v \in V} \pi_V(v) f(v) Ag(v) && \text{(Global view)} \\
&= \sum_{v \in V} \pi_V(v) f(v) \sum_{(w,v) \in E} Pr(w|v) g(w) \\
&= \mathop{\mathbb{E}}_{v}[f(v)] \mathop{\mathbb{E}}_{w|v}[g(w)] && \text{(Local view from node } v\text{)} \\
&= \mathop{\mathbb{E}}_{v,w}[f(v)g(w)] \\
&= \langle Af, g \rangle
\end{aligned}$$

Since A is self-adjoint, it is diagonalizable by the spectral theorem, and thus we have recovered all the properties of our standard spectral expansion for regular graphs. Thus we can safely define spectral expansion analogously to this case:

Definition 9.4 (Weighted Spectral Expansion)**.** Let A be the adjacency matrix of a weighted graph $G(V, E)$, then A has eigenvalues:

$$1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq -1.$$

We say $G(V, E)$ is a λ spectral expander if:

$$\max(|\lambda_2|, |\lambda_n|) \leq \lambda.$$

We say $G(V, E)$ is a *one-sided* λ spectral expander if:

$$\lambda \leq \lambda_2$$

9.2 Weighted Simplicial Complexes

We will endow each level of our simplicial complex with a distribution π_i analogous to how we defined the distributions π_E and π_V for a standard graph. Just as we chose a distribution for the top level faces (edges) of the standard graph and used it to induce a distribution over vertices, in a simplicial complex we pick an arbitrary distribution π_d across the top level faces and induce distributions across all lower levels. If a simplicial complex X is endowed with such probabilistic structure, we call it a weighted simplicial complex.

Definition 9.5 (Weighted Simplicial Complex)**.** A Weighted Pure Simplicial Complex (X, Π) is a Pure Simplicial Complex X endowed with a joint distribution $\Pi = (\pi_{-1}, \ldots, \pi_d)$ over each level such that:

9.1. π_d is arbitrary

9.2. For $-1 \leq i < d$, π_i is given by the following process: draw a d-face s from π_d, and pick uniformly at random some i-face $t \subset s$

Since our definition of spectral expansion will rely on the local link structure of the complex, we need to be careful as well about how we induce distributions over our links. To this end, we will define the weighted links associated to a weighted complex:

Definition 9.6 (Weighted Link)**.** Let X, Π be a d-dimensional weighted simplicial complex. The weighted link of a face $s \in X(i)$ is the weighted complex:

$$X_s = \{T \ : \ T \cup \{s\} \in X\},$$

with distribution Π^s induced from the distribution across top level faces π^s_{d-i}:

$$\pi^s_{d-i}(t) = \frac{\pi_d(t \cup s)}{\sum_{w \in X_s} \pi_d(w \cup s)}.$$

Notice that π^s_{d-1} is not necessarily equivalent to the normalization of π_{d-i} restricted to the link. The reason for our choice of distribution is that it will allow us to decompose X at any level into a convex combination of its links. This is a key technique when trying to connect local and global properties on an HDX. Going back to our initial example, let's take a look at how these distributions play out if we endow the Fish Complex with a uniform π_2.

9.3 Spectral Link Expansion and the Trickle Down Theorem

Now that we have developed background on weighted expansion and taken a look at some basic examples of our structures, we can define high dimensional expansion in a single sentence.

Definition 9.7 (Spectral Link Expansion)**.** A weighted simplicial complex (X, Π) is a (one-sided) λ Spectral Link Expander if the underlying graph of every link of co-dimension at least 2 is a (one-sided) λ spectral expander.

We only consider links of co-dimension at least 2 since the underlying graphs of the highest two dimensions are respectively the emptyset, and a disjoint set of vertices. Note that the emptyset is always of co-dimension at least 2 (even in standard graphs), and so we always require that the underlying graph of X itself is a spectral expander. Indeed, this is the only requirement for a 1-dimensional graph, which makes this definition a direct generalization of standard spectral expansion.

Now that we've managed to define our objects of interest, the next question to tackle is whether or not such objects even exist. Let's examine the high dimensional analogue of a classic example from standard graphs, the complete graph.

The d-dimensional Complete Complex (also known as the Johnson Complex) on n vertices is the complex made up of all $\binom{n}{[d+1]}$ subsets of $[n]$. The Complete Complex is a $\frac{1}{n-d}$ spectral link expander. This follows from the fact that every link the complete complex is itself a lower dimensional complete complex, and thus the underlying graph of every link is the complete graph. Dense random graphs are also good spectral link expanders, but unlike standard expansion, this is no longer true in the sparse case, as it is likely some link won't even be connected, much less an expander.

To my knowledge, there are currently only a few known examples of constant degree spectral link expanders, and they come from algebraic constructions. Covering even one of these constructions would take more than a full lecture, so instead we will instead focus on a theorem that simplifies the definition of spectral link expanders, and provides a nice intuition for their local-to-global phenomenon.

9.4 The Trickling Down Theorem

In particular, we will take a look at Izhar Oppenheim's Trickling Down Theorem. In the standard definition of expansion, we need every link to be a spectral expander. Oppenheim's theorem states that in fact we only need a much more local condition to be true: as long as the underlying graph of each link is connected and each link of co-dimension 2 is a good enough expander, then the entire complex is an expander:

Theorem 9.8 (Trickling Down Theorem)**.** *Let (X, Π) be a d-dimensional weighted simplicial complex such that the underlying graph of every link of co-dimension at least two is connected, and every link of exactly co-dimension two is a λ expander. The entire complex (X, Π) is a $\frac{\lambda}{1-(d-1)\lambda}$ expander.*

We will sketch the easiest version of this theorem, for one-sided two dimensional expanders, focusing in particular on the local-to-global flavor of the result. The higher dimensional argument follows immediately from repeated application, but two-sided expansion requires some additional work.

Theorem 9.9 (Trickling Down Theorem (2-dim))**.** *Let (X, Π) be a 2-dimensional weighted simplicial complex with an underlying connected graph such that the link of every vertex $v \in X(0)$ is a λ one-sided spectral expander. Then (X, Π) is a $\frac{\lambda}{1-\lambda}$ expander.*

Proof. Let A be the adjacency matrix of the underlying graph of X, and $f : X(0) \to \mathbb{R}$ some unit norm eigenfunction perpendicular to the all 1's vector $\mathbf{1}$. Our goal is to bound the eigenvalue corresponding to f, which we denote by γ. Our first step will be to look at γ from a local view, in particular as an expectation over the links of X:

$$\begin{aligned}
\gamma &= \langle f, Af \rangle && (1)\\
&= \mathop{\mathbb{E}}_{(v,w)} [f(v)f(w)] && (2)\\
&= \mathop{\mathbb{E}}_{v \in X(0)} \left[\mathop{\mathbb{E}}_{(u,w) \in X_v} [f(u)f(w)] \right] && (3)
\end{aligned}$$

Here, (2) follows by the same reasoning as in our previous analysis of a standard weighted graph. (3) follows from our definition of π_1^v, which ensures that picking an edge from π_1 is equivalent to first picking a vertex v from π_0, and then picking an edge from π_1^v. This expression for γ is particularly useful since the information we are trying to leverage is local to the links themselves.

Since we are focusing on links, let's define the restriction of f to the link X_v as f^v, and A_v to be the weighted adjacency matrix corresponding to X_v. Notice that since our expectation is already over links, we can re-write everything in terms of f^v:

$$\begin{aligned}
\gamma &= \mathop{\mathbb{E}}_{v \in X(0)} \left[\mathop{\mathbb{E}}_{(u,w) \in X_v} [f(u)f(w)] \right]\\
&= \mathop{\mathbb{E}}_{v \in X(0)} \left[\mathop{\mathbb{E}}_{(u,w) \in X_v} [f^v(u)f^v(w)] \right]\\
&= \mathop{\mathbb{E}}_{v \in X(0)} [\langle f^v, A_v f^v \rangle]
\end{aligned}$$

Remember that each link X_v is an λ spectral expander. If f^v were perpendiculuar to $\mathbf{1}$, we could immediately use this fact to bound γ given the above form. Unfortunately, while this is true for f itself, it is not necessarily true upon restricting to links. Instead, we define $g^v = f^v - \gamma f(v)\mathbf{1}$ to

be the projection of f^v onto the orthogonal component of $\mathbf{1}$. Since g^v is orthogonal to $\mathbf{1}$, we can massage the above slightly to get our desired bound:

$$\begin{aligned}
\gamma &= \mathop{\mathbb{E}}_{v \in X(0)} \left[\langle f^v, A_v f^v \rangle\right] \\
&= \mathop{\mathbb{E}}_{v \in X(0)} \left[\langle g^v + \gamma f(v)\mathbf{1}, A_v g^v + \gamma f(v)\mathbf{1} \rangle\right] \\
&= \mathop{\mathbb{E}}_{v \in X(0)} \left[\langle g^v, A_v g^v \rangle\right] + \gamma^2 \mathop{\mathbb{E}}_{v \in X(0)} \left[f(v)\right]^2 \\
&= \mathop{\mathbb{E}}_{v \in X(0)} \left[\langle g^v, A_v g^v \rangle\right] + \gamma^2
\end{aligned}$$

Using the standard argument for the adjacency matrices of spectral expanders:

$$\begin{aligned}
\gamma - \gamma^2 &= \mathop{\mathbb{E}}_{v \in X(0)} \left[\langle g^v, A_v g^v \rangle\right] \\
&\leq \lambda \mathop{\mathbb{E}}_{v \in X(0)} \left[\langle g^v, g^v \rangle\right] \\
&= \lambda(1 - \gamma^2)
\end{aligned}$$

Noting that $\gamma < 1$ due to our assumption that the underlying graph is connected finishes the proof. We have focussed mostly on finding and applying the local view in this proof sketch and brushed under the rug the analysis of the other terms. □

For a full explanation of this proof, see this lecture from Yotam Dikstein.

References

[1] N. Alon and V. Rödl, Sharp bounds for some multicolor Ramsey numbers, Combinatorica 25, 125–141, 2005.

[2] M. Belkin and P. Niyogi, Towards a theoretical foundation for laplacian-based manifold methods, Journal of Computer and System Sciences, 74(8), 1289–1308, 2008.

[3] Y. Bilu and N. Linial, Lifts, discrepancy and nearly optimal spectral gap. Combinatorica, 26(5):495–519, 2006.

[4] J. Borcea, and P. Brändén, Multivariate Polya-Schur classification problems in the Weyl algebra, 606–660, 2008.

[5] Bohman, T., Keevash, P. The early evolution of the H-free process, Invent. math., 181, 291–336, link, 2010.

[6] Bishnoi, A., Ihringer, F., Pepe, V., A construction for clique-free pseudorandom graphs ,link, 2019.

[7] M. Cohen, Ramanujan Graphs in Polynomial Time, Proceed- ings, Annual IEEE Symposium on Foundations of Computer Science, FOCS, 276-–281, 2016.

[8] F. Chung, Spectral graph theory, American Mathematical Soc., 92, 1997.

[9] I. Dinur, Spectral High Dimensional Expanders (HDX), link, 2018.

[10] C. Godsil, G. Royle. Algebraic graph theory. Vol. 207. Springer Science & Business Media, 2013.

[11] C. Godsil, I. Gutman. On the theory of the matching polynomial, Journal of Graph Theory, 5(2), 137–144, 1981.

[12] P. Erdős, G. Szekeres, A combinatorial problem in geometry, Compositio Math, 2, 463–470, 1935.

[13] H. J. Fell. On the zeros of convex combinations of polynomials, Pacific Journal of Mathematics, 89(1), 43–50, 1980.

[14] H. Huang, Induced subgraphs of hypercubes and a proof of the Sensitivity Conjecture, Annals of Mathematics 190.3, 949-955, 2019.

[15] Z. Jiang, J. Tidor, Y. Yao, S. Zhang, Y. Zhao, Equiangular lines with a fixed angle, link, 2019.

[16] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. Combinatorica, 8(3), 261–277, 1988.

[17] M. Krivelevich, B. Sudakov, Pseudo-random graphs: More sets, graphs and numbers, Springer, Berlin, Heidelberg, 199–262, 2006.

[18] A. Marcus, D. Spielman, and N. Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees, Annals of Mathematics, 182(1), 307–325, 2015.

[19] Lovász, László. Combinatorial problems and exercises. Vol. 361. American Mathematical Soc., 2007.

[20] A. Lubotzky, High dimensional expanders., link, 2017.

[21] D. Mubayi, J. Verstraete, A note on Pseudorandom Ramsey graphs, link, 2019.

[22] A. Nilli. On the second eigenvalue of a graph. Discrete Mathematics, 91(2), 207–210, 1991.

[23] D. Spielman. Spectral and Algebraic Grpah Theory. link

[24] R. Stanley, Algebraic combinatorics, Springer 20, 22, 2013.

[25] L. Trevisan, Graph Partitioning and Expanders, Lecture Notes. link, 2011.