Spectral Graph Theory, Expanders, and Ramanujan Graphs

Christopher Williamson
2014

Abstract

We will introduce spectral graph theory by seeing the value of studying the eigenvalues of various matrices associated with a graph. Then, we will learn about applications to the study of expanders and Ramanujan graphs, and more generally, to computer science as a whole.

Contents

1 Spectral graph theory introduction **3**
1.1 Graphs and associated matrices . . . 3
1.2 Eigenvalue invariance under vertex permutation . . . 3
1.3 Using the Rayleigh quotient to relate eigenvalues to graph structure . . . 3
1.3.1 Motivating the use of the Rayleigh quotient . . . 5

2 What a graph's spectrum can tell us **6**
2.1 Connectivity . . . 7
2.2 Bipartiteness . . . 8
2.3 A Cheeger Inequality . . . 9

3 The adjacency matrix and the regularity assumption **10**
3.1 Applications of the adjacency matrix without regularity . . . 10
3.2 Relating the adjacency and normalized Laplacian matrices using regularity . . . 11
3.3 Independent sets . . . 11
3.4 The chromatic number . . . 12
3.4.1 The Mycielski construction . . . 13
3.4.2 Chromatic upper bounds . . . 14
3.4.3 Chromatic lower bounds . . . 15

4 Expander graphs **17**
4.1 Definitions . . . 17
4.1.1 Combinatorial definition . . . 17
4.1.2 Algebraic definition . . . 17
4.2 Expanders are sparse approximations of a complete graph . . . 18

5 Ramanujan graphs **20**
5.1 Definition of Ramanujan graph. . . . 20
5.2 Motivation of the definition . . . 20

6 Existence of bipartite Ramanujan graphs **22**
6.1 2-Lifts . . . 22
6.2 The matching polynomial . . . 25
6.2.1 Relationship to characteristic polynomials of A_s . . . 25
6.2.2 Bounding the roots of the matching polynomial . . . 27
6.3 The difficulty with averaging polynomials . . . 29
6.4 Interlacing . . . 30

7 Pseudorandomness and applications to theoretical CS **32**
7.1 Expander mixing lemma . . . 32
7.2 Expansion as pseudorandomness . . . 32
7.3 Extractors via expander random walks . . . 33
7.3.1 Min-entropy of random variables, statistical distance, and extractors . . . 33
7.3.2 The random walk . . . 33
7.4 Other applications to complexity . . . 34
7.4.1 Using expanders to collapse complexity classes . . . 34
7.4.2 Randomness reduction in algorithms . . . 34

8 Survey of constructions of families of graphs **35**
8.1 Expanders . . . 35
8.1.1 Margulis-Gabber-Galil expanders . . . 35
8.1.2 Combinatoric constructions . . . 35
8.2 The LPSM Ramanujan graphs . . . 35
8.2.1 Cayley graphs . . . 35
8.2.2 The Lubotzsky-Phillips-Sarnak construction . . . 36

9 Adknowledgements **36**

10 References **36**

1 Spectral graph theory introduction

1.1 Graphs and associated matrices

We will define a graph to be a set of vertices, V, and a set of edges, E, where E is a set containing sets of exactly two distinct vertices. In this thesis, we will be considering undirected graphs, which is why E is defined in this way rather than as a subset of $V \times V$. A k-regular graph is a graph such that every vertex (sometimes called node) has exactly k edges touching it (meaning every vertex has degree k). More formally, and in the context of undirected graphs, this means that $\forall v \in V$,

$$|\{\{v, u\} | u \in V\}| = k$$

Definition. *(adjacency matrix)* The adjacency matrix A associated with a graph G of nodes ordered $v_1, ..., v_n$ has entries defined as:

$$A_{ij} = \begin{cases} 1 & \text{if } \{v_i, v_j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

We will also consider the diagonal matrix D, such that D_{ii} =deg(v_i). Note that for any k-regular graph G, $D_G = kI$, where I is the appropriately sized identity matrix.

Definition. *(Laplacian matrix)* The Laplacian matrix of a graph G is defined as $L_G := D_G - A_G$. This is equivalent to:

$$L_{ij} = \begin{cases} \deg(v_i) & \text{if } i = j \\ -1 & \text{if } \{v_i, v_j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

Definition. *(normalized Laplacian)* Finally, the normalized Laplacian, $\mathcal{L}$, is defined as: $\mathcal{L} := D^{-1/2} L D^{-1/2}$, which is:

$$\mathcal{L}_{ij} = \begin{cases} 1 & \text{if } i = j \\ \frac{-1}{\sqrt{\deg(v_i)\deg(v_j)}} & \text{if } \{v_i, v_j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

Note that we usually only use these definitions for graphs without self-loops, multiple edges, or isolated vertices ($\mathcal{L}$ isn't defined on graphs with a degree zero vertex). These matrices are useful due to how they act with respect to their Rayleigh quotient, which we will see in the following sections.

1.2 Eigenvalue invariance under vertex permutation

We are going to concern ourselves with the eigenvalues of these matrices that we just defined. However, it is apparent in the definition of these matrices that each vertex had to be assigned a number. Thus it is fair to ask whether renaming the vertices will affect the eigenvalues that we will study. Of course, we will find exactly what is expected – that these values are intrinsic to the graph and not the choice of vertex naming. Following is a sketch of how we can know this. Suppose that M is some matrix defined in terms of a graph G and that a permutation $\sigma \in S_n$ is applied to the vertex names. Define M' the same way that M was defined – just applied to the newly named vertices. Then, $M_{ij} = M'_{\sigma(i)\sigma(j)}$. Let P_σ be the permutation matrix for σ. Then, $P_\sigma^T M P_\sigma = M'$. This is because the multiplication where M is on the right will swap the columns accordingly, and the other multiplication will swap the rows. It is easy to check that for permutation matrices, $P^T = P^{-1}$, meaning that M and M' are similar. Then we only need recall the linear algebra fact that similar matrices have the same characteristic polynomial and thus the same eigenvalues and multiplicities.

1.3 Using the Rayleigh quotient to relate eigenvalues to graph structure

First, we will prove a theorem, and then see how this theorem relates to what is called the Rayleigh quotient of a matrix.

Theorem 1.1. *For all* $\boldsymbol{h} \in \mathbb{R}^{|V|}$*, and if* $\mathbf{f} := D^{-1/2}\mathbf{h}$*:*

$$\frac{\boldsymbol{h}^T \mathcal{L}_G \boldsymbol{h}}{\boldsymbol{h}^T \boldsymbol{h}} = \frac{\sum\limits_{\{i,j\}\in E} \left(\boldsymbol{f}(i) - \boldsymbol{f}(j)\right)^2}{\sum\limits_{v\in V} deg(v)\boldsymbol{f}(v)^2} = \frac{\sum\limits_{\{i,j\}\in E} \left((D^{-1/2}\boldsymbol{h})(i) - (D^{-1/2}\boldsymbol{h})(j)\right)^2}{\sum\limits_{v\in V} deg(v)(D^{-1/2}\boldsymbol{h})(v)^2}$$

The first step to proving the theorem is the following lemma.

Lemma 1.3.1. *For all* $\boldsymbol{x} \in \mathbb{R}^{|V|}$*:*

$$\boldsymbol{x}^T L_G \boldsymbol{x} = \sum_{\{i,j\}\in E} \left(\boldsymbol{x}(i) - \boldsymbol{x}(j)\right)^2$$

Note that for brevity, we are representing the edge $\{v_i, v_j\}$ as $\{i, j\}$, and $\mathbf{x}(i)$ is the ith entry in the vector $\mathbf{x}$.

Proof. Note that $(L\mathbf{x})_i$, the ith entry in the vector $L\mathbf{x}$, equals $L_{\text{row } i} \cdot \mathbf{x}$, which in turn equals

$$\sum_{j=1}^{n} L(i,j)\mathbf{x}(j) = \sum_{j=1,\ j\neq i}^{n} L(i,j)\mathbf{x}(j) + L(i,i)\mathbf{x}(i)$$

$$= \deg(i)\mathbf{x}(i) + \sum_{j=1,\ j\neq i}^{n} L(i,j)\mathbf{x}(j) = \deg(i)\mathbf{x}(i) - \sum_{j:\{i,j\}\in E}^{n} \mathbf{x}(j)$$

Thus,

$$\mathbf{x}^T L_G \mathbf{x} = \sum_{i=1}^{n} \mathbf{x}(i)\left(\deg(i)\mathbf{x}(i) - \sum_{j:\{i,j\}\in E} \mathbf{x}(j)\right) = \sum_{i=1}^{n} \left[\mathbf{x}(i)^2 \deg(i)\right] - \sum_{i=1}^{n} \sum_{j:\{i,j\}\in E} \mathbf{x}(i)\mathbf{x}(j)$$

Since each edge in the second summation is counted twice, we reach:

$$\sum_{i=1}^{n} \left[\mathbf{x}(i)^2 \deg(i)\right] - 2 \sum_{\{i,j\}\in E} \mathbf{x}(i)\mathbf{x}(j)$$

Since $\deg(i) = \sum\limits_{j:\{i,j\}\in E} 1$, we have:

$$\sum_{i=1}^{n} \left[\mathbf{x}(i)^2 \sum_{j:\{i,j\}\in E} 1\right] - 2 \sum_{\{i,j\}\in E} \mathbf{x}(i)\mathbf{x}(j)$$

$$= \sum_{i=1}^{n} \sum_{j:\{i,j\}\in E} \mathbf{x}(i)^2 - 2 \sum_{\{i,j\}\in E} \mathbf{x}(i)\mathbf{x}(j)$$

Note that in the double summation, we will run through each edge twice, but only consider one node of the edge. Instead, we can sum over each edge only once but consider both nodes that form that edge.

$$\sum_{\{i,j\}\in E} \left(\mathbf{x}(i)^2 + \mathbf{x}(j)^2\right) - 2 \sum_{\{i,j\}\in E} \mathbf{x}(i)\mathbf{x}(j)$$

$$= \sum_{\{i,j\}\in E} \left(\mathbf{x}(i) - \mathbf{x}(j)\right)^2$$

□

Proof. (of Theorem 1.1)

The numerator:
Note that $\mathbf{h}^T\mathcal{L}\mathbf{h} = \mathbf{h}^T D^{-1/2}LD^{-1/2}\mathbf{h} = (D^{1/2}\mathbf{f})^T D^{-1/2}LD^{-1/2}(D^{1/2}\mathbf{f}) = \mathbf{f}^T L\mathbf{f} = \sum_{\{i,j\}\in E}(\mathbf{f}(i)-\mathbf{f}(j))^2$, where the last step follows due to Lemma 1.3.1

The denominator:
We have $\mathbf{h}^T\mathbf{h} = \mathbf{f}^T(D^{1/2})^T D^{1/2}\mathbf{f} = \mathbf{f}^T D\mathbf{f} = \sum_{v\in V}\deg(v)\mathbf{f}(v)^2$ □

1.3.1 Motivating the use of the Rayleigh quotient

The Rayleigh quotient of a vector $\mathbf{h}$ with respect to a matrix M is $\frac{\mathbf{h}^T M\mathbf{h}}{\mathbf{h}^T\mathbf{h}}$. This is a useful tool for considering eigenvalues of M – note that if $\mathbf{h}$ is an eigenvector of M with eigenvalue λ, then the Rayleigh quotient equals $\frac{\mathbf{h}^T\lambda\mathbf{h}}{\mathbf{h}^T\mathbf{h}} = \lambda$. This is why Theorem 1.1 is important – it relates the Rayleigh quotient (and thus the eigenvalues of $\mathcal{L}$) to a quotient of sums that are closely related to the edges and vertices in the graph. We can immediately see that any eigenvalues of $\mathcal{L}$ are non-negative, since the Rayleigh quotient is equal to a sum of squares over a sum of positive values. We will show precisely how the Rayleigh quotient can be used to find eigenvalues. One should first recall that for symmetric $n\times n$ matrices, there exists an orthonormal basis of eigenvectors $\psi_0, ..., \psi_{n-1}$, each with eigenvalue $\lambda_0, ..., \lambda_{n-1}$ (we order the eigenvalues so that $\lambda_i \le \lambda_{i+1}$).

Theorem 1.2. *For every symmetric matrix M,*

$$\lambda_0 = \min_{\mathbf{h}\in\mathbb{R}^n}\frac{\mathbf{h}^T M\mathbf{h}}{\mathbf{h}^T\mathbf{h}} \text{ and for } k\ge 1,\ \lambda_k = \min_{\mathbf{h}\perp\psi_0,\ldots,\psi_{k-1}}\frac{\mathbf{h}^T M\mathbf{h}}{\mathbf{h}^T\mathbf{h}}$$

Proof. We will closely follow the proof given in Dan Spielman's spectral graph theory lecture notes [28]. First, note that

$$\psi_j^T\left(\sum_{i=0}^{n-1}(\psi_i^T\mathbf{x})\psi_i\right) = \sum_{i=0}^{n-1}(\psi_i^T\mathbf{x})\psi_j^T\psi_i = (\psi_j^T\mathbf{x})\psi_j^T\psi_j = \psi_j^T\mathbf{x}$$

This means that $\sum_{i=0}^{n-1}(\psi_i^T\mathbf{x})\psi_i = \mathbf{x}$. The reason for this is that $\sum_{i=0}^{n-1}(\psi_i^T\mathbf{x})\psi_i - \mathbf{x}$ is a vector orthogonal to every vector in the orthonormal basis, implying that it is the zero vector. Now, let's minimize the Rayleigh quotient. Note that scaling $\mathbf{h}$ by a constant doesn't change the quotient value, so we can assume $\mathbf{h}^T\mathbf{h} = 1$.

$$\begin{aligned}
\frac{\mathbf{h}^T M\mathbf{h}}{\mathbf{h}^T\mathbf{h}} &= \left(\sum_{i=0}^{n-1}(\psi_i^T\mathbf{h})\psi_i\right)^T M\left(\sum_{j=0}^{n-1}(\psi_j^T\mathbf{h})\psi_j\right)\\
&= \left(\sum_{i=0}^{n-1}(\psi_i^T\mathbf{h})\psi_i\right)^T\left(\sum_{j=0}^{n-1}(\psi_j^T\mathbf{h})\lambda_j\psi_j\right)\\
&= \left(\sum_{i=0}^{n-1}(\psi_i^T\mathbf{h})\psi_i^T\right)\left(\sum_{j=0}^{n-1}(\psi_j^T\mathbf{h})\lambda_j\psi_j\right)\\
&= \left(\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(\psi_i^T\mathbf{h})(\psi_j^T\mathbf{h})\lambda_j\psi_i^T\psi_j\right)\\
&= \left(\sum_{i=0}^{n-1}(\psi_i^T\mathbf{h})(\psi_i^T\mathbf{h})\lambda_i\psi_i^T\psi_i\right) \qquad (\psi_i^T\psi_j = 0 \text{ if } i\ne j)\\
&= \left(\sum_{i=0}^{n-1}(\psi_i^T\mathbf{h})^2\lambda_i\right) \qquad \text{(we are using an orthonormal basis)}
\end{aligned}$$

$$\geq \lambda_0 \sum_{i=0}^{n-1} (\psi_i^T \mathbf{h})^2$$

Note that doing the same calculation, but neglecting the M term, we see that $\mathbf{h}^T\mathbf{h} = 1 = \sum_{i=0}^{n-1} (\psi_i^T \mathbf{h})^2$. So, we have:

$$\frac{\mathbf{h}^T M \mathbf{h}}{\mathbf{h}^T \mathbf{h}} \geq \lambda_0$$

But note that $\frac{\psi_\mathbf{0}{}^T M \psi_\mathbf{0}}{\psi_\mathbf{0}{}^T \psi_\mathbf{0}} = \lambda_0$, implying that $\lambda_0 = \min_{\mathbf{h} \in \mathbb{R}^n} \frac{\mathbf{h}^T M \mathbf{h}}{\mathbf{h}^T \mathbf{h}}$.
The above argument needs little modification to prove that $\lambda_k = \min_{\mathbf{h} \perp \psi_0, \ldots, \psi_{k-1}} \frac{\mathbf{h}^T M \mathbf{h}}{\mathbf{h}^T \mathbf{h}}$. We can skip directly to the fact that

$$\frac{\mathbf{h}^T M \mathbf{h}}{\mathbf{h}^T \mathbf{h}} = \left(\sum_{i=0}^{n-1} (\psi_i^T \mathbf{h})^2 \lambda_i \right)$$

and then apply the orthogonality condition to get

$$\frac{\mathbf{h}^T M \mathbf{h}}{\mathbf{h}^T \mathbf{h}} = \left(\sum_{i=k}^{n-1} (\psi_i^T \mathbf{h})^2 \lambda_i \right) \geq \lambda_k$$

Since the ψ_i form an othonormal basis, we have that the minimum is attained by ψ_k. □

We want to apply these findings to $\mathcal{L}$. We start with a simple fact – that every normalized Laplacian matrix has $\lambda_0 = 0$.

Corollary 1.3.1. *$\lambda_0 = 0$, where λ_0 is the smallest eigenvalue of any normalized Laplacian matrix.*

Proof. Notice that in the proof of Theorem 1.2, the inequality $\left(\sum_{i=0}^{n-1} (\psi_i^T \mathbf{h})^2 \lambda_i \right) \geq \lambda_0 \sum_{i=0}^{n-1} (\psi_i^T \mathbf{h})^2$ holds as an equality if and only if $\mathbf{h}$ is orthogonal to every eigenvector besides ψ_0. Applying this fully to the λ_k case, we have that

$$\arg \min_{\mathbf{h} \perp \psi_0, \ldots, \psi_{k-1}} \frac{\mathbf{h}^T \mathcal{L} \mathbf{h}}{\mathbf{h}^T \mathbf{h}}$$

is a scalar multiple of ψ_k, and thus an eigenvector. So, only eigenvectors minimize the Rayleigh quotient. We can use this fact, along with Theorem 1.1 to find an explicit eigenvector that applies to all normalized Laplacian matrices. If we set $\mathbf{h}$ in the statement of Theorem 1.1 to be $\frac{D^{1/2}\mathbb{1}}{||D^{1/2}\mathbb{1}||}$, then we have that $\mathbf{f}$ is a constant function, and thus the Rayleigh quotient is 0. But then since $\frac{D^{1/2}\mathbb{1}}{||D^{1/2}\mathbb{1}||}$ is minimizing the Rayleigh quotient (the quotient is obviously non-negative), we know that $\frac{D^{1/2}\mathbb{1}}{||D^{1/2}\mathbb{1}||}$ is an eigenvector of $\mathcal{L}$, with eigenvalue 0. □

2 What a graph's spectrum can tell us

Now we will finally use all this to say something combinatorially related to the graph. We start with a quick linear-algebraic lemma.

Lemma 2.0.2. *For square matrices A and B (note that the zero-blocks are not necessarily square),*

$$\det \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = (\det A)(\det B)$$

Proof. Fix the dimensions of B to be $n \times n$. We induct on r, where the dimensions of A are $r \times r$. The base case is clear, and we suppose that the result holds for $r \leq n-1$. Using Laplace's formula for the determinant,

we have (let M_{ij} (resp. A_{ij}) be the minor of the entire block matrix (resp. A) formed by excluding the ith row and jth column):

$$\det \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = a_{11} \det M_{11} + ... + a_{1n} \det M_{1n}$$

By the inductive hypothesis, we have

$$= a_{11} \det A_{11} \det B + ... + a_{1n} \det A_{1n} \det B = (\det A)(\det B)$$

□

2.1 Connectivity

Now for our first combinatorial result:

Theorem 2.1. $\max\{k \mid \lambda_k = 0\} = K_0 \iff$ *G has precisely K_0 disjoint connected components, where λ_j is the $j+1^{st}$ smallest eigenvalue of the normalized Laplacian matrix.*

Proof. We will denote C as the precise number of connected components of G, and show that $K_0 = C$.

Showing that $K_0 \geq C$: Break G into its connected components, $G_1, ..., G_C$. Then, the normalized Laplacian of G is of the form: $\mathcal{L}_G = \begin{bmatrix} \mathcal{L}_{G_1} & 0 & 0 & 0 \\ 0 & \mathcal{L}_{G_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathcal{L}_{G_C} \end{bmatrix}$, since there are no edges between the components. This means that

$$\det(\mathcal{L}_G - \lambda I) = \det \begin{bmatrix} \mathcal{L}_{G_1} - \lambda I & 0 & 0 & 0 \\ 0 & \mathcal{L}_{G_2} - \lambda I & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathcal{L}_{G_C} - \lambda I \end{bmatrix} = \prod_{i=1}^{C} \det(\mathcal{L}_{G_i} - \lambda I)$$

where the last equality is due to Lemma 2.0.2 being repeated inductively. Thus, considering the characteristic polynomial, we see that the eigenvalue 0 occurs in $\mathcal{L}_G$ with multiplicity equal to the sum of the multiplicities of the eigenvalue 0 in each of the $\mathcal{L}_{G_i}$. Since Corollary 1.3.1 tells us that each of these pieces has eigenvalue 0 with multiplicity at least 1, we have $K_0 \geq C$.

Showing that $K_0 \leq C$: It suffices to show that for a connected graph H, $\lambda_1 \neq 0$. This implies that each of the $\mathcal{L}_{G_i}$ (which are connected) has the multiplicity of eigenvalue 0 equal to at most 1. Then, the multiplicity of the eigenvalue 0 in $\mathcal{L}_G$ is at most C. Suppose that H is a connected graph and $\lambda_1 = 0$. Then, we have the existence of two orthogonal eigenvectors, ψ_0, ψ_1 each with eigenvalue 0. Using the Rayleigh quotient, we have that for $k = 0, 1$:

$$0 = \psi_k^T \mathcal{L}_H \psi_k = \sum_{\{i,j\} \in E} \Big((D^{-1/2}\psi_k)(i) - (D^{-1/2}\psi_k)(j) \Big)^2$$

This implies that $D^{-1/2}\psi_k$ is constant on all the vertices in H. Note that connectivity is crucial here, since $D^{-1/2}\psi_k$ is not necessarily constant otherwise. If there existed vertices i and j such that $D^{-1/2}\psi_k(i) \neq D^{-1/2}\psi_k(j)$, then there must exist (due to connectivity) an edge that connects vertices where $D^{-1/2}\psi_k$ takes distinct values. (Otherwise, it would be possible to disconnect the graph into two parts such that the vertices i and j were in different parts.) Then, the Rayleigh quotient could not be zero. So, we know that $D^{-1/2}\psi_0 = \delta D^{-1/2}\psi_1$ for some constant δ. But this is a contradiction, since this implies that $\psi_0 = \delta\psi_1$, and those eigenvectors are orthogonal. □

2.2 Bipartiteness

We will now prove that all eigenvalues of the normalized Laplacian are at most 2. Eventually, we will see that this upper bound is attained if and only if G is bipartite.

Corollary 2.2.1. *$\lambda_{n-1} \leq 2$, where λ_{n-1} is the largest eigenvalue of a normalized Laplacian matrix.*

Proof. First, note that for all real numbers x and y,

$$0 \leq (x+y)^2 \implies -2xy \leq x^2 + y^2 \implies (x-y)^2 \leq 2(x^2+y^2)$$

We use this to obtain an upper bound on the Rayleigh quotient numerator:

$$\sum_{\{i,j\}\in E} (\mathbf{f}(i) - \mathbf{f}(j))^2 \leq 2 \sum_{\{i,j\}\in E} \mathbf{f}(i)^2 + \mathbf{f}(j)^2 = 2 \sum_{v\in V} \deg_v \mathbf{f}(v)^2$$

Then, by Theorem 1.1, we have $\frac{\mathbf{h}^T \mathcal{L} \mathbf{h}}{\mathbf{h}^T \mathbf{h}} \leq 2$, which implies that any eigenvector can have eigenvalue at most 2. □

Now, we can prove the equivalency between bipartiteness and $\lambda_{n-1} = 2$.

Theorem 2.2. *G is bipartite $\iff \lambda_{n-1} = 2$*

Proof. $\Rightarrow$ direction: G is bipartite means that we can split the vertices into subsets $A, B \subset V$ such that every edge starts in A and ends in B. Define

$$\mathbf{f}(v) = \begin{cases} 1 & \text{if } v \in A \\ -1 & \text{otherwise} \end{cases}$$

As always, we take $\mathbf{h} = D^{1/2}\mathbf{f}$. Then, we have

$$\frac{\mathbf{h}^T \mathcal{L}_G \mathbf{h}}{\mathbf{h}^T \mathbf{h}} = \frac{\sum\limits_{\{i,j\}\in E} (\mathbf{f}(i) - \mathbf{f}(j))^2}{\sum\limits_{v\in V} \deg(v)\mathbf{f}(v)^2} = \frac{\sum\limits_{\{i,j\}\in E} 4}{\sum\limits_{v\in V} \deg(v)} = \frac{4|E|}{2|E|} = 2$$

This means that the maximum possible value of the Rayleigh quotient is attained by $\mathbf{h}$. In Corollary 1.3.1, we demonstrated that only eigenvectors minimize the Rayleigh quotient. The argument to show that only eigenvectors maximize the Rayleigh quotient is virtually identical. Using that fact, we have that $\mathbf{h}$ is an eigenvector with eigenvalue 2. Therefore, $\lambda_{n-1} = 2$.

$\Leftarrow$ direction: We assume that $\lambda_{n-1} = 2$ and show that G is bipartite. Note that $\lambda_{n-1} = 2$ implies that (by plugging into Theorem 1) for $\mathbf{f} := D^{-1/2}\psi_{n-1}$,

$$\begin{aligned} \sum_{\{i,j\}\in E} (\mathbf{f}(i) - \mathbf{f}(j))^2 &= 2 \sum_{v\in V} \deg(v)\mathbf{f}(v)^2 \\ &= 2 \sum_{\{i,j\}\in E} \mathbf{f}(i)^2 + \mathbf{f}(j)^2 \end{aligned}$$

This implies that

$$0 = \sum_{\{i,j\}\in E} 2\mathbf{f}(i)^2 + 2\mathbf{f}(j)^2 - \big(\mathbf{f}(i)^2 + \mathbf{f}(j)^2 - 2\mathbf{f}(i)\mathbf{f}(j)\big) = \sum_{\{i,j\}\in E} (\mathbf{f}(i) + \mathbf{f}(j))^2$$

Since all the terms are positive, we can say that every individual term is zero, which means that for all edges $\{i,j\}$, $\mathbf{f}(i) = -\mathbf{f}(j)$. This in turn means that there are no odd cycles in G. If the vertices $v_1, v_2, ..., v_{2n+1}$ form a cycle, then v_1 and v_{2n+1} are adjacent, meaning that $\mathbf{f}(v_1) = -\mathbf{f}(v_{2n+1})$. But we also have $\mathbf{f}(v_1) = \mathbf{f}(v_3) = ... = \mathbf{f}(v_{2n+1})$. This means that $\mathbf{f}(v) = 0$ for all v, which is a contradiction, since $\mathbf{f} = 0 \implies D^{1/2}\mathbf{f} = 0 \implies \psi_{n-1} = 0$. Of course, ψ_{n-1} is not the zero vector because it is an eigenvector.

Now that we know that G has no odd cycles, it follows that G is bipartite (in fact, these are equivalent conditions). One can quickly see the reason for this by considering an algorithm where we determine which vertex set each vertex should be assigned. Without loss of generality, assume that G is connected and start from an arbitrary vertex. Move along edges, assigning each vertex an alternating vertex set each time. If the algorithm succeeds then G is bipartite because we just found the vertex sets. If the algorithm fails, then there exists adjacent vertices that are also connected by a disjoint path of even length. This means that there is a odd cycle, contradicting our earlier findings. So, G is bipartite. □

2.3 A Cheeger Inequality

We want to show that λ_1, the second smallest $\mathcal{L}$ eigenvalue, is a measure of how well connected a graph is. This will be accomplished by defining a constant associated with each graph (one that clearly is related to how well connected a graph is) and then putting bounds on it in terms of λ_1.

Definition. *(volume)* We define the volume of a subset of the vertices $X \subset V$ to be:

$$\text{vol}(X) = \sum_{x \in X} \deg(x)$$

Definition. *(Cheeger constant on a vertex subset)* We also define:

$$h_G(X) = \frac{|E(X, \overline{X})|}{\min\{\text{vol}(X), \text{vol}(\overline{X})\}}$$

Definition. *(Cheeger constant)* Finally, the Cheeger constant of a graph is defined as:

$$h_G = \min_X h_G(X)$$

Note that the volume of some subset of the vertices basically represents how important or large that subset is, taking into account differences in degree (since different vertices may have different degree, we do not simply take $\text{vol}(X) = |X|$). We can consider the Cheeger constant, h_G, as representing how difficult it is to disconnect G. A small value of h_G means that there is a way to divide up the vertices into sets X and $\overline{X}$ so that there are few edges leading between the sets, relative to the size of the sets.

Example. If G is already disconnected, then h_G is zero (take X to be a connected component). If G is complete (without self-loops), then consider a partition $X, \overline{X}$, such that $|X| = n, |\overline{X}| = m$, and without loss of generality, $n \geq m$. Then,

$$h_G(X) = \frac{|E(X, \overline{X})|}{\min\{\text{vol}(X), \text{vol}(\overline{X})\}} = \frac{mn}{\text{vol}(\overline{X})} = \frac{mn}{\sum_{x \in \overline{X}} \deg(x)} = \frac{mn}{(n+m-1)\sum_{x \in \overline{X}} 1} = \frac{n}{n+m-1} = \frac{n}{|V|-1}$$

Since $n \geq m$, we know that $n \geq \lceil |V|/2 \rceil$. Thus,

$$h_G = \min_X h_G(X) = \frac{\lceil |V|/2 \rceil}{|V|-1}$$

Let's find bounds in terms of λ_1.

Theorem 2.3. *$2h_G \geq \lambda_1$, the second smallest $\mathcal{L}$ eigenvalue.*

Proof. Since the Cheeger constant of G equals h_G, there exists a partition of the vertices A, B ($B = \overline{A}$) such that $h_G(A) = h_G$. Define a function over the vertices:

$$f(v) = \begin{cases} 1/\text{vol}(A) & \text{if } v \in A \\ -1/\text{vol}(B) & \text{otherwise} \end{cases}$$

Note that

$$\frac{\sum_{\{i,j\} \in E} (f(i) - f(j))^2}{\sum_{v \in V} \deg(v) f(v)^2} = \frac{\sum_{\{i,j\} \in E(A,B)} \left(\frac{1}{\text{vol}(A)} + \frac{1}{\text{vol}(B)} \right)^2}{\frac{1}{\text{vol}^2(A)} \sum_{v \in A} \deg(v) + \frac{1}{\text{vol}^2(B)} \sum_{v \in B} \deg(v)}$$

$$= \frac{\left(\frac{1}{\text{vol}(A)} + \frac{1}{\text{vol}(B)}\right)^2 \sum_{\{i,j\} \in E(A,B)} 1}{\frac{1}{\text{vol}(A)} + \frac{1}{\text{vol}(B)}} = |E(A,B)| \left(\frac{1}{\text{vol}(A)} + \frac{1}{\text{vol}(B)}\right) \leq \frac{2|E(A,B)|}{\min\{\text{vol}(A), \text{vol}(B)\}} = 2h_G(A) = 2h_G$$

Now, it is important to realize that without loss of generality, we can assume that $\psi_0 = \frac{D^{1/2}\mathbb{1}}{||D^{1/2}\mathbb{1}||}$, by considering the eigenspace decomposition of $\mathcal{L}$. Also, it is very easy to check that $\frac{D^{1/2}\mathbb{1}}{||D^{1/2}\mathbb{1}||} \perp f$. Thus, we have:

$$\lambda_1 = \min_{h \perp \psi_0} \frac{h^T \mathcal{L} h}{h^T h} \leq \frac{f^T \mathcal{L} f}{f^T f} = 2h_G$$

□

We will just state the upper bound, the proof of which is somewhat longer and can be found in [9].

Theorem 2.4. $\lambda_1 \geq \frac{h_G^2}{2}$

Combining the bounds, we have that $\frac{h_G^2}{2} \leq \lambda_1 \leq 2h_G$.

3 The adjacency matrix and the regularity assumption

3.1 Applications of the adjacency matrix without regularity

The adjacency matrix, although typically used when considering regular graphs, can also tell us interesting information about non-regular graphs. We denote the adjacency matrix as A with eigenvectors $\phi_0, ..., \phi_{n-1}$ and corresponding eigenvalues $\mu_0 \geq ... \geq \mu_{n-1}$. The reason for ordering the eigenvalues this way (before, we had λ_0 as the smallest eigenvalue) will be discussed shortly.

First, we very clearly have:

$$(Ax)(u) = \sum_{\{u,v\} \in E} x(v)$$

In particular, we have that $(A\mathbb{1})(u) = \sum_{\{u,v\} \in E} \mathbb{1} = \deg_u$.
Additionally, all of the arguments made about minimizing the Rayleigh quotient can be transformed into statements about maximizing the quotient with only trivial changes. Thus, we have:

$$\mu_0 = \max_{\mathbf{h} \in \mathbb{R}^n} \frac{\mathbf{h}^T A \mathbf{h}}{\mathbf{h}^T \mathbf{h}} \text{ and for } k \geq 1,\ \mu_k = \max_{\mathbf{h} \perp \phi_0, ..., \phi_{k-1}} \frac{\mathbf{h}^T A \mathbf{h}}{\mathbf{h}^T \mathbf{h}}$$

So, putting together these last two statements, we have that:

$$\mu_0 = \max_{\mathbf{h} \in \mathbb{R}^n} \frac{\mathbf{h}^T A \mathbf{h}}{\mathbf{h}^T \mathbf{h}} \geq \frac{\mathbb{1}^T A \mathbb{1}}{\mathbb{1}^T \mathbb{1}} = \frac{\sum_{v \in V} \deg_v}{n} = \deg_{\text{ave}}$$

We will now put a similar upper bound on μ_0, the largest adjacency eigenvalue. Let ϕ_0 be our eigenvector with eigenvalue μ_0, and say that $v = \arg\max_u \phi(u)$. Without loss of generality, we have $\phi_0(v) \neq 0$, since we could take $-\phi_0$ if necessary. So, we have:

$$\mu_0 = \frac{(\mu_0 \phi_0)(v)}{\phi_0(v)} = \frac{(A\phi_0)(v)}{\phi_0(v)} = \frac{\sum_{\{v,u\} \in E} \phi_0(u)}{\phi_0(v)} \leq \sum_{\{v,u\} \in E} 1 = \deg_v \leq \deg_{\max}$$

Moreover, if $\mu_0 = \deg_{\max}$, then $\frac{\sum_{\{v,u\} \in E} \phi_0(u)}{\phi_0(v)} = \sum_{\{v,u\} \in E} 1$ and $\deg_v = \deg_{\max}$. But then, ϕ_0 is a constant vector with value $\deg_v$ for all vertices u that are connected to v. The way that v was chosen then applies to u for all such vertices u. Repeating this argument (on a connected graph) yields the result that G is $\deg_{\max}$-regular. Combining all of the above into a single theorem found in [28], we have:

Theorem 3.1. *$deg_{ave} \leq \mu_0 \leq deg_{max}$ and, for connected graphs, $\mu_0 = deg_{max} \implies G$ is deg_{max}-regular.*

3.2 Relating the adjacency and normalized Laplacian matrices using regularity

Assuming G is regular allows us to very quickly find bounds on the adjacency eigenvalues. Crucial to this is the observation that in the d-regular case, we have:

$$\mathcal{L} = I - \frac{1}{d}A$$

Then, assume we have $\mathcal{L}\psi_i = \lambda_i\psi_i$. Then, we have:

$$\lambda_i\psi_i = (I - \frac{1}{d}A)\psi_i = I\psi_i - \frac{1}{d}A\psi_i = \psi_i - \frac{1}{d}A\psi_i$$

This implies that:

$$\frac{1}{d}A\psi_i = (1-\lambda_i)\psi_i \implies A\psi_i = d(1-\lambda_i)\psi_i$$

So, the eigenvalue λ_i of $\mathcal{L}$ corresponds to an eigenvalue $d(1-\lambda_i)$ of A. (This fact, that large $\mathcal{L}$ eigenvalues correspond to small eigenvalues of A, is why we ordered the μ_j in the opposite direction.) Now, some facts follow easily from the bounds we proved on the λ_i.

- $\mu_0 = d$ (analog of Corollary 1.3.1) Note that this proves a partial converse to Theorem 3.1. For regular graphs, $\mu_0 = d = \deg_{\max}$.
- $\max\{k \mid \mu_k = d\} = K_0 \iff G$ has precisely K_0 disjoint connected components (analog of Theorem 2.1)
- $-d \leq \mu_{n-1}$ (analog of Corollary 2.2.1)
- G is bipartite $\iff \mu_{n-1} = -d$ (analog of Theorem 2.2)

Notice that the above facts imply that $\mu_0 \geq -\mu_{n-1}$. It turns out that this is still true if the regularity assumption is removed (but we omit proof). Also, in the non-regular case, bipartiteness is equivalent to whether or not $\mu_0 = -\mu_{n-1}$.

Example. As an example, let's consider the complete graphs on n vertices. Then, the adjacency matrix takes the form of $J_n - I_n$, where J_n is the $n \times n$ all-ones matrix. It is easy to see that the all-ones vector is an eigenvector with eigenvalue n. Clearly, the rank of J_n is 1 and thus the nullity is $n-1$ (by the rank-nullity theorem). Therefore, J_n has eigenvalue 0 with multiplicity $n-1$. Since the identity matrix has eigenvalue 1 with multiplicity n, then the adjacency matrix A, which equals $J_n - I_n$, has eigenvalues $n-1$ (multiplicity 1) and -1 (multiplicity $n-1$). This means that the normalized Laplacian eigenvalues of the complete graph are 0 (multiplicity 1) and $\frac{n}{n-1}$ (multiplicity $n-1$). Similarly, the non-normalized Laplacian of the complete graph has eigenvalues 0 (multiplicity 1) and n (multiplicity $n-1$).

3.3 Independent sets

We can also use the adjacency eigenvalues to put bounds on the size of an independent set in G. An independent set of vertices is a set of vertices such that no edges connect any of the vertices in the set. First, we will give a proof of a slightly different statement, which then implies an adjacency matrix eigenvalue bound on the size of independent sets in G. We follow the proof of Spielman given in [28]. Note that for this next lemma, we use the eigenvalues of L, the non-normalized Laplacian matrix, and the result holds for non-regular graphs.

Lemma 3.3.1. *If S is an independent set in G, and the average degree of vertices in S is $d_{ave}(S)$, then:*

$$|S| \leq n\left(1 - \frac{d_{ave}(S)}{\lambda_{n-1}}\right)$$

Here, λ_{n-1} is the largest eigenvalue of the non-normalized Laplacian.

Proof. By what is said in section 2.1, we know that

$$\lambda_{n-1} = \max_x \frac{x^T L x}{x^T x}$$

Let χ_S be the indicator vector for S, and define $x := \chi_S - \frac{|S|}{n}\mathbb{1}$. By Lemma 1.3.1, we have that

$$x^T L x = \sum_{\{i,j\}\in E} \left(\chi_S(i) - \frac{|S|}{n} - \chi_S(j) + \frac{|S|}{n}\right)^2 = \sum_{v\in S} \deg(v) = d_{\text{ave}}(S)|S|$$

The second to last equality follows from the fact that S is an independent set. Thus $\chi_S(i)$ and $\chi_S(j)$ are never 1 at the same time – and the sum just counts the number of edges that touch the independent set. Also, we have:

$$x^T x = \sum_{v\in V} \left(\chi_S(v) - \frac{|S|}{n}\right)\left(\chi_S(v) - \frac{|S|}{n}\right) = \sum_{v\in V} (\chi_S(v))^2 + \sum_{v\in V} \left(-2\frac{|S|}{n}\chi_S(v)\right) + \sum_{v\in V} \left(\frac{|S|}{n}\right)^2$$

$$= |S| - 2\frac{|S||S|}{n} + \frac{|S||S|}{n^2}n = |S|\left(1 - 2\frac{|S|}{n} + \frac{|S|}{n}\right) = |S|\left(1 - \frac{|S|}{n}\right)$$

Putting this together, we have that

$$\lambda_{n-1} \geq \frac{d_{\text{ave}}(S)|S|}{|S|(1-|S|/n)} \implies |S| \leq n\left(1 - \frac{d_{\text{ave}}(S)}{\lambda_{n-1}}\right)$$

□

Using the regularity assumption allows us to say something about all independent sets of a regular graph, in general.

Corollary 3.3.1. *For a regular graph G, denote $\alpha(G)$ as the size of the largest independent set. Then, $\alpha(G) \leq n\frac{-\mu_{n-1}}{d-\mu_{n-1}}$.*

Proof. A very similar argument as done in section 3.2 shows that an eigenvalue λ_i of L corresponds to an eigenvalue $d - \lambda_i$ of A for d-regular graphs. So, assuming regularity tells us that $d_{\text{ave}}(S) = d$ and that $\lambda_{n-1} = d - \mu_{n-1}$. So, we have, for any independent set S, $|S| \leq n\left(1 - \frac{d}{d-\mu_{n-1}}\right) = n\left(\frac{-\mu_{n-1}}{d-\mu_{n-1}}\right)$ □

Example. Let's apply this bound to a complete graph. We know from the previous example that $\mu_{n-1} = -1$ and thus the bound yields $\alpha(G) \leq \frac{n}{d+1} = 1$. Of course, in a complete graph, all the independent sets have size 1.

3.4 The chromatic number

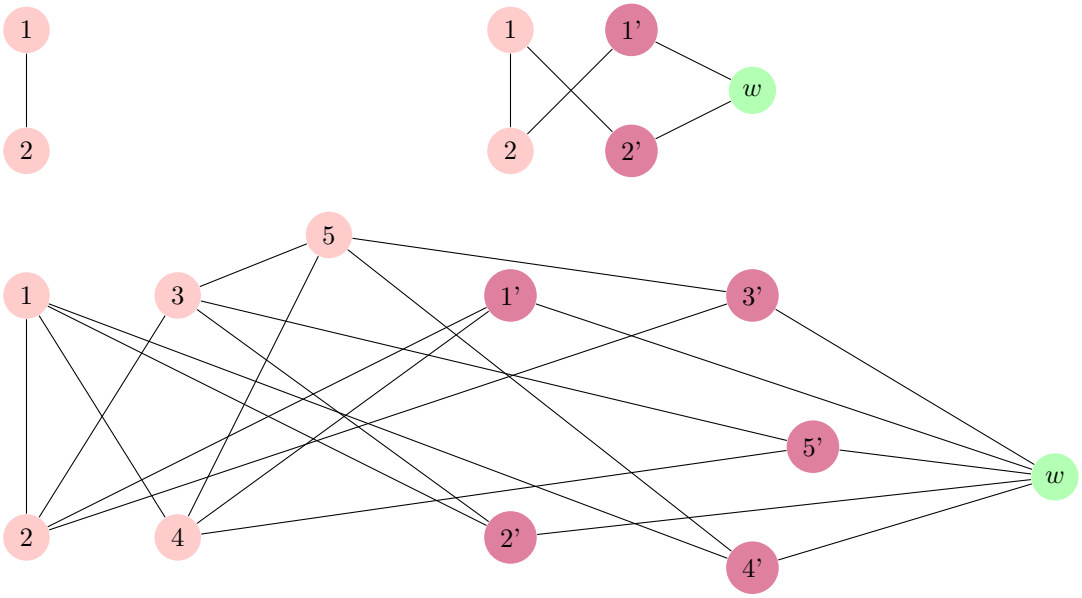
A clique is a set of vertices such that every vertex in the set is connected to every other vertex in the set. Clearly, it is the opposite of an independent set, and an independent set in G is a clique in $\overline{G}$, the dual of G. (The dual of G is the graph formed by taking G, removing all of its edges, and then placing edges wherever edges were not in the original graph.) Clearly, the chromatic number of a graph is at least equal to the size of the largest clique of the graph since every vertex in the clique must have its own color. Thus, we cannot use the previous corollary to obtain chromatic number bounds, as we only gave an upper bound on the size of cliques in $\overline{G}$ (the size of independent sets in G). In fact, the relationship between the size of the largest clique in G and the chromatic number is about as weak as one could possibly expect. We will see this in the next section, where we give a graph construction that illustrates this weakness of this relationship.

3.4.1 The Mycielski construction

This method for constructing graphs, discussed in [21], allows us to obtain graphs that have a chromatic number that is as large as desired, all while ensuring that the graph is triangle-free (which means that the clique number must be less than 3). The construction is defined as follows: take a graph G with vertices $V = \{v_1, ..., v_n\}$. Then, create a duplicate node u_j for each of the v_j and connect each of the u_j to the nodes in V that the corresponding v_j is connected to. Note that there are no edges between vertices in $U = \{u_1, ..., u_n\}$. Then, add one more vertex, w, which is connected to every vertex u_j.

Supposing that one starts with a triangle-free graph with chromatic number c, then we will show that transforming the graph in this way conserves the triangle-free property but adds exactly one to the chromatic number. Then, iterating this repeatedly yields the fact that the chromatic number can be as large as one wants, without the clique number growing whatsoever. A formal proof is omitted due to its length and simplicity, but the reasons can be expressed well in English. By assumption there are no triangles among the vertices in V. If there is a triangle among vertices in V and in U, then this implies a triangle among vertices only in V, a contradiction (substitute the u_j in the triangle with v_j). But a triangle cannot involve w, as w is only connected to vertices in U, none of which are connected to each other.
So we have seen that the new graph is also triangle-free. We will now observe that if the original graph has chromatic number k, then the new graph has chromatic number $k + 1$. First, note that by coloring each u_j the same as the corresponding v_j, and giving the w vertex a new color, we obtain a legal $k+1$ coloring of the new graph, implying that the new chromatic number is at most $k + 1$. To show that the chromatic number is greater than k, suppose that we have a k-coloring of the new graph. Note that at most $k - 1$ colors can occur among the vertices in U, since there exists a vertex (w) that touches every vertex in U. But then, there exists a color, called c such that c occurs among the V vertices but not among the U vertices. Then, one can re-color each of the v_j that are color c to be the same color as the corresponding u_j color, obtaining a $k - 1$ coloring of the original graph, a contradiction. To check that this is a valid coloring, notice that if v_j is color c and u_j is another color c', then none of v_j's neighbors are color c', since u_j touches all of those neighbors. Thus, we can recolor v_j to color c'. Repeating this for each colored c vertex yields the contradiction.
Therefore, there exist triangle-free (meaning the largest clique is of size 2 or less) graphs with chromatic number as large as one pleases. □

An example iteration of the Mycielski construction. Red nodes are the original vertices, purple ones are the duplicated versions, and the green one is the node that is connected to all vertex duplicate (purple) nodes.



3.4.2 Chromatic upper bounds

A very simple first bound that we can place on the chromatic number is that the chromatic number of G, denoted $\chi(G)$ or χ, satisfies $\chi(G) \leq \max_{v \in V}\{\deg(v)\} + 1$. This is obvious, because the worst case scenario is that one is trying to color a vertex of maximum degree, and all of the neighbors have already been colored. In that case, one simply uses the final color (since there is exactly one more color than the maximum degree). However, it is rare for equality to happen – in 1941 Brooks showed that the chromatic number is less than $\max_{v \in V}\{\deg(v)\} + 1$ whenever G is not complete or not an odd cycle [6]. We can improve on this bound using spectral methods. First, a lemma from [28] that will allow us to prove what is called Wilf's Theorem.

Lemma 3.4.1. *If A and B are symmetric matrices, where B is obtained by removing the last row and column from A, then the largest eigenvalue of A is at least the largest eigenvalue of B.*

Proof. Let α and β denote the largest eigenvalues of A, and B, respectively. Here, we assume that A is a $n \times n$ matrix and B is a $n-1 \times n-1$ matrix. Then, let y denote a vector in $\mathbb{R}^{n-1}$ and y^0 denote a vector in $\mathbb{R}^n$, where y^0 has the same entries as y, except for an additional zero as the final entry. Then, we have:

$$y^T B y = {y^0}^T A y^0$$

This is very easy to confirm, but in a nutshell, Ay^0 is equal to By, with the exception that there is a extra value at the end of the vector. Then, the multiplication by ${y^0}^T$ removes this final deviation. But now, we are basically done by considering the Rayleigh quotient. We have:

$$\beta = \max_{y \in \mathbb{R}^{n-1}} \frac{y^T B y}{y^T y} = \max_{y \in \mathbb{R}^{n-1}} \frac{{y^0}^T A y^0}{{y^0}^T y^0} \leq \max_{x \in \mathbb{R}^n} \frac{x^T A x}{x^T x} = \alpha$$

□

Notice that although this lemma states that we are removing the last row and column, by the discussion in section 1.2, we can rename the vertices so that the last row an column represents any vertex we choose. So, if we choose this vertex to be the node of least degree, then when we remove it from A to obtain B, we know that the new graph given by B has a larger $\deg_{\text{ave}}$ value. However, μ_0 decreases. This is interesting in light of Theorem 3.1, which states that $\mu_0 \geq \deg_{\text{ave}}$.

Now, we can prove an improved version of the trivial chromatic bound just discussed.

Theorem 3.2. $\chi(G) \leq \lfloor \mu_0 \rfloor + 1$

Note that in the regular case, we have $\mu_0 = d$, and this theorem becomes equivalent to the very weak previous bound.

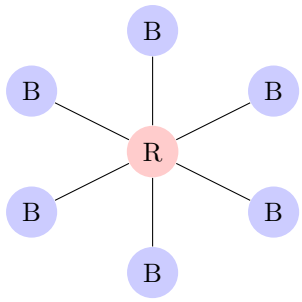
Proof. Induction on the number of vertices in the graph. If we have a single vertex graph, then the only eigenvalue is zero, meaning that $\mu_0 = 0$. Then, clearly the chromatic number is 1, which is less than or equal to $\lfloor \mu_0 \rfloor + 1$. Assume that the theorem holds for all graphs on $n-1$ vertices. Then, take G as a n vertex graph and obtain G' from it by removing the vertex with minimum degree (to be called v). Then, we have:

$$\chi(G') \leq \lfloor \mu_0' \rfloor + 1 \leq \lfloor \mu_0 \rfloor + 1$$

Where the first inequality holds from the induction hypothesis and the second inequality holds from the previous lemma. Now apply this coloring to the analogous vertices of G, which leaves only v to be colored. We chose v to have the fewest neighbors, which means that $\deg(v) \leq \deg_{\text{ave}}(G)$. But Theorem 3.1 tells us that $\deg_{\text{ave}}(G) \leq \mu_0$, which implies the stronger statement $\deg_{\text{ave}}(G) \leq \lfloor \mu_0 \rfloor$ simply because $\deg_{\text{ave}}(G)$ is always an integer. So this means that we can successfully color G, since we only need to assign a color to v, v has at most $\lfloor \mu_o \rfloor$ neighbors, and we have $\lfloor \mu_o \rfloor + 1$ colors at our disposal. □

Example. An example where this bound significantly outperforms the trivial bound is the star graph (one central node connected to some number of vertices that are only connected to the central vertex). If we take a star graph with one central node and 6 points (6 nodes connected to the center), then we have that $\chi(G) = 2$ (the central node is blue and all the other nodes are red) , $\deg_{\text{ave}} = \frac{12}{7} \approx 1.71$, $\deg_{\text{max}} = 6$, and $\lfloor \mu_0 \rfloor = 2$. The trivial bound gives 7 as an upper bound, the new bound gives 3, and the actual answer is 2.

A plot of the star graph just described:



3.4.3 Chromatic lower bounds

Not surprisingly, we can use spectral graph theory to obtain lower bounds. One short lemma will allow us to use Corollary 3.3.1.

Lemma 3.4.2. *A k-colorable graph has an independent set of size at least $\left\lceil \frac{|V|}{k} \right\rceil$.*

Proof. Consider the color classes in G. Of course, each color class is an independent set, and the sum of the sizes of the color classes is $|V|$. Of course, not all of the color classes can have below average size, so there must exist a color class of size at least $\left\lceil \frac{|V|}{k} \right\rceil$. □

We apply this finding to Lemma 3.3.1.

Theorem 3.3. *For a regular graph G, $\chi(G) \geq 1 + \frac{\mu_0}{-\mu_{n-1}}$*

Proof. Take G to be d-regular, and assume that it has chromatic number $\chi(G) = k$. Then, by Lemma 3.4.2, we may assume that there exists an independent set S, where $|S| \geq \left\lceil \frac{|V|}{k} \right\rceil$. Plugging this into Corollary 3.3.1, we have ($|V| = n$):

$$\left\lceil \frac{n}{k} \right\rceil \leq \alpha(G) \leq n \left(\frac{-\mu_{n-1}}{d - \mu_{n-1}} \right)$$

By simple rearranging, we arrive at

$$k \geq 1 - \frac{d}{\mu_{n-1}} = 1 + \frac{\mu_0}{-\mu_{n-1}}$$

The theorem follows since $k = \chi(G)$ □

Example. If we apply the result to the complete graph, we obtain $\chi(G) \geq 1 + \frac{n-1}{1} = n$. This is the chromatic number of the n-vertex complete graph.

In fact, Theorem 3.3 holds even for non-regular graphs (see [1,7,14,17,24]). We will partially prove this with the method of interlacing eigenvalues (we will also see interlacing again later). We start with a definition of interlacing.

Definition. *(sequence interlacing)* Let $\alpha_0, ..., \alpha_{n-1}$ and $\beta_0, ..., \beta_{m-1}$ be two sequences of real numbers, where $n > m$. Then, we say that the two sequences interlace if, for all $i \in 0, .., m$, we have:

$$\alpha_i \geq \beta_i \geq \alpha_{n-m+i}$$

Next, we state a theorem that we will take for granted. Its proof is located in [7].

Theorem 3.4. *Take A to be a $n \times n$ symmetric real matrix. Also, take a collection of m non-zero orthogonal vectors $x_j \in \mathbb{R}^n$. Define a matrix C, where $C_{ij} = \frac{1}{||x_i||_2} x_i^T A x_j$. Notice that C is an $m \times m$ matrix. Then, we have:*

- *The eigenvalues of C interlace the eigenvalues of A.*

- *Let $s = \sum_j x_j$. Then, the number $\frac{s^T A s}{s^T s}$ lies between the smallest and largest eigenvalues of C. If s is an eigenvalue of A with eigenvalue θ, then θ is also an eigenvalue of C, for the eigenvector $\mathbb{1}$.*

Corollary 3.4.1. *Define m to be the chromatic number of G, possibly non-regular. Then, if μ_j are the eigenvalues of A, the adjacency matrix of G, then, $m \geq 1 - \frac{\mu_0}{\mu_{n-1}}$.*

Proof. Again, let ϕ_j denote the eigenvector of A for eigenvalue μ_j. G is m-colorable, which means that we can partition the vertices into m color classes, denoted $V_0, ..., V_{m-1}$. We now define m vectors in $\mathbb{R}^n$, named $x_0, ..., x_{m-1}$. For all j in $0, .., m-1$, define the i^{th} entry of x_j by $(x_j)_i = (\phi_0)_i(\text{char}(V_j))_i$. Here, we are denoting the indicator vector that a vertex is in V_j as $\text{char}(V_j)$. Also, notice that

$$\left(\sum_j x_j\right)_i = \sum_j (x_j)_i = \sum_j (\phi_0)_i(\text{char}(V_j))_i = (\phi_0)_i$$

Note that the last equality holds because the i^{th} vertex is in only one of the sets V_j, and that this statement means that $\sum_j x_j = \phi_0$. Similarly, we know that all the x_j vectors are pairwise orthogonal. To prove this, consider the following:

$$x_j^T x_k = \sum_i (x_j)_i(x_k)_i = \sum_i (\phi_0)_i^2(\text{char}(V_j))_i(\text{char}(V_k))_i = 0 + ... + 0 = 0$$

Note that some of the x_j vectors could be 0. In this case, we remove them and thus may have fewer than m such vectors. This doesn't matter, but for explicitness, we will now say that we have m' x_j vectors (clearly, $1 \leq m' \leq m$). We can now apply the previous corollary, and with its full strength since $s = \sum_j x_j$ is an eigenvalue of A. Therefore, we know that the eigenvalues of the matrix C, as defined in Theorem 3.4 interlace the eigenvalues of A. Also, by the second bullet point in the theorem, we know that s, which in this case equals ϕ_0, is an eigenvalue of C and $C\mathbb{1} = \mu_0\mathbb{1}$. Putting this together, we have

$$\beta_0 + \beta_1 + ... + \beta_{m'-1} \geq \mu_0 + \mu_{n-m'+1} + \mu_{n-m'+2} + ... + \mu_{n-1}$$

We know that $\beta_0 \leq \mu_0$ due to the interlacing. Also, $\beta_0 \geq \mu_0$ since μ_0 is an eigenvalue of both C and A. Thus, $\beta_0 = \mu_0$. Then, we only need to see that $\beta_1 + ... + \beta_{m'-1} \geq \mu_{n-m'+1} + \mu_{n-m'+2} + ... + \mu_{n-1}$, which follows directly from the interlacing. Recall that the trace of a matrix is the sum of the diagonal entries, and that it is well know to also be the sum of the eigenvalues. Thus, we have that

$$\text{tr}(C) = \beta_0 + \beta_1 + ... + \beta_{m'-1} \geq \mu_0 + \mu_{n-m'+1} + \mu_{n-m'+2} + ... + \mu_{n-1}$$

We now want to show that $\text{tr}(C) = 0$. To do this, recall from section 3.1 that

$$(Ax)(u) = \sum_{\{u,v\} \in E} x(v)$$

To show that the trace of C is 0, we will consider an arbitrary entry of C on the diagonal, $\frac{1}{||x_j||_2} x_j^T A x_j$. We obviously only need to show that $x_j^T A x_j = 0$. Using the fact from section 3.1, we see that the i^{th} entry of Ax_j is $\sum_{\{i,v\} \in E} x_j(v)$. Therefore, the sum given by $x_j^T(Ax_j)$ is

$$\sum_i \left[(x_j)_i \sum_{\{i,v\} \in E} x_j(v) \right] = \sum_i \left[(\phi_0)_i(\text{char}(V_j))_i \sum_{\{i,v\} \in E} (\phi_0)_v(\text{char}(V_j))_v \right]$$

Each term in the sum equals 0. Consider the i^{th} term. If vertex i is not in V_j, then the term is already zero because of the $(\text{char}(V_j))_i$ term. If this is not the case, then vertex i is in V_j. But then, no v sharing an edge with vertex i would also be in V_j, since the V_j are color classes. Thus, every one of the $(\text{char}(V_j))_v$ terms is 0. We are finally getting close. We have that

$$0 = \text{tr}(C) = \beta_0 + \beta_1 + ... + \beta_{m'-1} \geq \mu_0 + \mu_{n-m'+1} + \mu_{n-m'+2} + ... + \mu_{n-1}$$

Now, notice that $0 = \text{tr}(A)$ (there are no self-loops) and that this equals the sum of the adjacency eigenvalues. We know from Theorem 3.1 that as long as G has an edge, $\mu_0 > 0$. Then, since the sum of the eigenvalues is 0, this guarantees that there exists a negative eigenvalue, and thus that μ_{n-1} is negative. With this in mind, we say

$$0 \geq \mu_0 + \mu_{n-m'+1} + \mu_{n-m'+2} + ... + \mu_{n-1} \geq \mu_0 + (m'-1)\mu_{n-1}$$

Rearranging, and knowing that $\mu_{n-1} \leq 0$ yields:

$$m' \geq 1 - \frac{\mu_0}{\mu_{n-1}} \implies \chi(G) = m \geq m' \geq 1 - \frac{\mu_0}{\mu_{n-1}}$$

□

4 Expander graphs

4.1 Definitions

We will present two commonly used definitions of regular expander graphs, mention a relationship between then, and then quickly move further using just one of the definitions.

4.1.1 Combinatorial definition

A graph is said to be a (n, d, δ)-expander if for a n vertex, d-regular graph, then for all sets of vertices S, we have:

$$|S| \leq \frac{|V|}{2} \implies |E(S, \overline{S})| \geq \delta d|S|$$

(We consider only sets of size less than or half of all vertices because then no graph ever satisfies any δ value greater than 0; take $S = V$.) This definition means that all small sets of vertices are connected somewhat robustly to the rest of the graph, thus higher δ values correspond to better expansion. All connected graphs are expanders for some $\delta \geq 0$, and non-connected graphs are not expanders as taking S to be the smallest connected component requires δ to be zero. Notice that the Cheeger inequality is closely related to this definition:

$$\begin{aligned}
h_G = \delta &\implies \forall X, h_G(X) \geq \delta \\
&\implies \forall X, \frac{|E(X, \overline{X})|}{\min\{\text{vol}(X), \text{vol}(\overline{X})|\}} \geq \delta \\
&\implies \forall X \text{ such that } |X| \leq \frac{|V|}{2}, \frac{|E(X, \overline{X})|}{\text{vol}(X)} \geq \delta \\
&\implies \frac{|E(X, \overline{X})|}{d|X|} \geq \delta \implies |E(X, \overline{X})| \geq \delta d|X| \\
&\implies G \text{ is an } (n, d, \delta) \text{ expander.}
\end{aligned}$$

Then, using the Cheeger inequality to relate h_G to λ_1 (the second smallest $\mathcal{L}$ eigenvalue) and the fact that $\lambda_i = d(1 - \mu_i)$, we can relate combinatorial expansion to the spectral quantities of G. Also, all of the above steps are reversible, allowing for if and only if statements that relate the spectral quantities to the combinatorial expansion coefficient.

4.1.2 Algebraic definition

Recall that for any d-regular graph G, the eigenvalues of its adjacency matrix will lie in $[-d, d]$. Such a graph is said to be a (n, d, ϵ)-expander if it is connected and all of the non-trivial eigenvalues lie in $[-d+\epsilon, d-\epsilon]$. Since d is always an eigenvalue of G (with multiplicity one, since G is assumed to be connected), d is considered a trivial eigenvalue. Sometimes $-d$ is also considered trivial since every bipartite graph has this eigenvalue. In what follows, we will assign an expansion coefficient, $\hat{\mu}(G)$ to each n-vertex, d-regular graph, and say that G is a spectral expander if $\hat{\mu} < d$ and an ϵ-expander if $\hat{\mu}(G) \leq d - \epsilon$. However, we will define $\hat{\mu}$ in two different ways, which will affect what graphs are considered to be expanders. The two $\hat{\mu}$ definitions are:

1. $\hat{\mu}(G) = \max\{|\mu_1|, |\mu_{n-1}|\}$
2. $\hat{\mu}(G) = \max_{|\mu_i| \neq d} |\mu_i|$

The only discrepancy arises in the case of bipartite graphs. The first definition assigns bipartite graphs an expansion coefficient that is d (due to the $-d$ eigenvalue) and thus says that no bipartite graph is ever an expander. The second definition allows for bipartite graphs. Note that in the case of definition 1, $\hat{\mu}$ is the second largest absolute value of any eigenvalue of G, and in definition 2, $\hat{\mu}$ is the second largest absolute value except for bipartite graphs, in which case it is the third largest. We will specify which definition of $\hat{\mu}$ we are using when needed.

4.2 Expanders are sparse approximations of a complete graph

Here, we will prove a lemma in [2] and [19] that demonstrates that the better an expander is, the more quickly a random walk over its vertices mixes. This highlights a similarity between expanders and a complete graph (where the first step of a random walk is already uniformly distributed). Of course, unlike complete graphs, expanders are more sparse (which is most meaningful in the context of expander families of fixed degree).

Lemma 4.2.1. *Let p be a probability distribution over the vertices. Then, for a d-regular, non-bipartite, and connected graph G, $|| \left(\frac{1}{d}A\right)^s p - \omega||_2 \leq \left(\frac{\hat{\mu}}{d}\right)^s$.*

Note that there is no need to specify which definition of $\hat{\mu}(G)$ we are using, since they agree in the non-bipartite case. Also, note that A is the adjacency matrix, s is the number of steps taken in the walk, and ω is the vector representing the uniform distribution $(1/|V|, ..., 1/|V|)$. Observe that we start the random walk at a vertex distributed according to p, and then proceed to neighbors uniformly at random. Thus, the distribution over the vertices after s steps is $\left(\frac{1}{d}A\right)^s p$. So, this lemma demonstrates that the distribution over the vertices approaches uniform more quickly the better the expansion constant is.

Proof. Recall that for the normalized Laplacian graph, we can assume that $\psi_0 = \frac{D^{1/2}\mathbb{1}}{||D^{1/2}\mathbb{1}||}$, an eigenvector with eigenvalue 0. Also, note that $A\mathbb{1} = d\mathbb{1}$, implying that we can similarly assume that ω is an eigenvector of A with eigenvalue d (recall that G is a d-regular graph). It is easy to check that $\left\langle \omega, \left(\frac{1}{d}A\right) v \right\rangle = \left\langle \left(\frac{1}{d}A\right) \omega, v \right\rangle$. (Since $A = A^T$ and $\sum_i \sum_j A_{ij} v_j = \sum_i \sum_j A_{ij} v_i$.) Earlier, we mentioned that $A\omega = d\omega$, which means that $\left(\frac{1}{d}A\right)\omega = \omega$. So, we have $\left\langle \omega, \left(\frac{1}{d}A\right) v \right\rangle = \left\langle \left(\frac{1}{d}A\right) \omega, v \right\rangle = \langle \omega, v \rangle$. So, $v \perp \omega \implies \left(\frac{1}{d}A\right) v \perp \omega$.

Recall that the orthogonal basis of eigenvectors for A is $\phi_0, ..., \phi_{n-1}$. We can assume that (despite having norm not equal to 1) that $\phi_0 = \omega$. We also take G to be connected and non-bipartite, which means that there is only one eigenvalue with absolute value d.
Let v be in the space of all vectors perpendicular to ω, as above. Then, v is in the span of eigenvectors with eigenvalues less than or equal to $\hat{\mu}$, in absolute value, since $\hat{\mu}$ is the second largest eigenvalue in absolute value. So, we have that $||Av||_2 \leq \hat{\mu}||v||_2$, or equivalently, $|| \left(\frac{1}{d}A\right) v||_2 \leq \frac{\hat{\mu}}{d}||v||_2$. Note that the vector $\left(\frac{1}{d}A\right) v$ is still orthogonal to ω, which means that we can induct on this argument to say that: $|| \left(\frac{1}{d}A\right)^s v||_2 \leq \left(\frac{\hat{\mu}}{d}\right)^s ||v||_2$, for $s \in \mathbb{Z}^+$.

Now let p be our probability distribution. Then, decompose p into parts in the direction of, and orthogonal to ω: $p = \alpha\omega + p'$. Since the entries of ω are constant, and p' is orthogonal to ω, we know that the sum of the entries of p' is 0. Also, we know that the sum of the entries of p is 1, and the sum of entries of $\alpha\omega$ is α. This means that α must equal 1. We then have $p = \omega + p'$.
We need a couple more facts:
$\left(\frac{1}{d}A\right)^s p = \left(\frac{1}{d}A\right)^s (\omega + p') = \left(\frac{1}{d}A\right)^s \omega + \left(\frac{1}{d}A\right)^s p' = \omega + \left(\frac{1}{d}A\right)^s p'$
Also, $\omega \perp p'$ and $p = \omega + p' \implies ||p||_2^2 = ||\omega||_2^2 + ||p'||_2^2 \implies ||p'||_2^2 \leq ||p||_2^2 \leq 1$, where the last inequality follows from the fact that p has entries that are positive and sum to 1. So, we have $||p'||_2^2 \leq 1$ and thus:

$$\left|\left|\left(\frac{1}{d}A\right)^s p - \omega\right|\right|_2 = \left|\left|\left(\frac{1}{d}A\right)^s p'\right|\right|_2 \leq \left(\frac{\hat{\mu}}{d}\right)^s ||p'||_2 \leq \left(\frac{\hat{\mu}}{d}\right)^s$$

The last wo inequalities follow from the facts that $p' \perp \omega$ and $||p'||_2^2 \leq 1$. □

Note that since $\frac{\hat{\mu}}{d} < 1$ for any connected d-regular graph, we have just proven that the distribution of final vertex location over a s-step random walk converges to the uniform distribution as s gets large for any such graph. However, the statement is not true for bipartite graphs (which bounce back and forth forever) or non-regular graphs (which converge to a different stationary distribution).

We can now prove a bound on the diameter of graphs using Lemma 4.2.1.

Corollary 4.2.1. *For a connected, n-vertex, d-regular graph G, we have diam*$(G) \leq \frac{\log(n)}{\log(d)-\log(\hat{\mu})}$

Proof. Recall that the diameter of a graph G is the length of the longest shortest path between any two vertices. This means that:

$$s \geq \text{diam}(G) \iff \forall p, \text{ the number of zeroes in } \left(\frac{1}{d}A\right)^s p \text{ is } 0$$

$\implies$
This is because no matter where the walk started, the walk has been long enought such that, if just the right choices were made, any vertex could have been reached (recall that $\left(\frac{1}{d}A\right)^s p$ is the probability distribution over the vertices after a s step walk).

$\impliedby$
Suppose not. We can assume that there exist α and β such that the distance between them is diam(G) and that $s < \text{diam}(G)$. But then we can choose p so that p pick α with probability 1. But then, there is no way for β to be reached, meaning that there is a zero entry in $\left(\frac{1}{d}A\right)^s p$, a contradiction.

Now that we have that result, we can notice that

$$\left(\frac{\hat{\mu}}{d}\right)^s < \frac{1}{n} \implies \forall p, \text{ the number of zeroes in } \left(\frac{1}{d}A\right)^s p \text{ is } 0 \implies s \geq \text{diam}(G)$$

The last implication we just proved. For the first implication, recall that Lemma 4.2.1 tells us that

$$\left|\left|\left(\frac{1}{d}A\right)^s p - \omega\right|\right|_2 < \frac{1}{n}$$

Assume for the sake of contradiction that there exists a p such that there exists a zero entry in $\left(\frac{1}{d}A\right)^s p$. Then, if we want to minimize $\left|\left|\left(\frac{1}{d}A\right)^s p - \omega\right|\right|_2$ as much as possible, we should assume that $\left(\frac{1}{d}A\right)^s p$ consists of a single zero entry with the remaining entries being $\frac{1}{n}$. But then, we have $\left|\left|\left(\frac{1}{d}A\right)^s p - \omega\right|\right|_2 = \frac{1}{n}$, a contradiction. So we now have established that $\left(\frac{\hat{\mu}}{d}\right)^s < \frac{1}{n} \implies s \geq \text{diam}(G)$. Finally, set $\theta = \frac{\hat{\mu}}{d}$ and we have:

$$\left(\frac{\hat{\mu}}{d}\right)^s < \frac{1}{n} \iff s > \log_\theta \frac{1}{n}$$

Thus, we have proven that

$$s > \log_\theta \frac{1}{n} \implies s \geq \text{diam}(G)$$

which implies that

$$\text{diam}(G) \leq \log_\theta \frac{1}{n} = \frac{\log 1/n}{\log \theta} = \frac{\log n}{\log d - \log \hat{\mu}}$$

□

Example. Note that this means for infinite families of d-regular graphs, all with $\hat{\mu}(G)$ bounded away from d, then $\text{diam}(G) = O(\log n)$.

Example. For the family of complete graphs on n vertices (all of which have a diameter of 1), this bound gives $\text{diam}(G) \leq \frac{\log(n)}{\log(n-1)}$.

For another bound on the diameter (this time in terms of any of the normalized Laplacian eigenvectors, and for regular and non-regular graphs), consult [12].

5 Ramanujan graphs

5.1 Definition of Ramanujan graph.

A graph is said to be a Ramanujan graph if $\hat{\mu}(G) \leq 2\sqrt{d-1}$.
Recall that $\hat{\mu}(G) = \max\{|\mu_1|, |\mu_{n-1}|\}$, unless we want to allow for bipartite Ramanujan graphs, in which case we have $\hat{\mu}(G) = \max_{|\mu_i| \neq d} |\mu_i|$. This definition means that Ramanujan graphs are good expanders. We will see that this is the strictest definition we can have, while still allowing for infinite families of such graphs to exist.

5.2 Motivation of the definition

Theorem 5.1. *For any infinite family of d-regular graphs on n nodes, denoted $\{G_n\}$, we have $\hat{\mu}(G_n) \geq 2\sqrt{d-1} - o(1)$, where the last term is some number that goes to zero as n gets large.*

Proof. We will closely follow the proof of Nilli [25]. Fix two edges in G, $e_1 = \{v_1, v_2\}$ and $e_2 = \{u_1, u_2\}$ such that the distance between these edges is at least $2k+2$. Note that distance between an edge and itself is 0, the distance between two edges that share exactly one node in common is 1, and so on.
For $1 \leq i \leq k$, define:

$$V_i = \{v \in V | \min\{\text{distance}(v, v_1),\ \text{distance}(v, v_2)\} = i\}$$
$$U_i = \{v \in V | \min\{\text{distance}(v, u_1),\ \text{distance}(v, u_2)\} = i\}$$

Let $a, b \in \mathbb{R}$. Define a function f over the vertices as:

$$f(v) = \begin{cases} a(d-1)^{-i/2} & \text{if } v \in V_i \\ b(d-1)^{-i/2} & \text{if } v \in U_i \\ 0 & \text{otherwise} \end{cases}$$

Important observations are as follows:

- $|V_i| \leq (d-1)|V_{i-1}|$ and $|U_i| \leq (d-1)|U_{i-1}|$
  Since G is d-regular, the size of $|V_i|$ is maximized by sending all edges from a vertex in V_{i-1} to V_i. However, for every node in V_{i-1}, one edge must remain that connects it to V_{i-2}, hence $d-1$ edges per node.
- $\left(\cup_i^k V_i\right) \cap \left(\cup_i^k U_i\right) = \varnothing$
  This simply follows from the fact that the sets V_i and U_i are defined with respect to the edges e_1 and e_2, which are a distance at least $2k+2$ apart. (And the sets V_i are only defined for i up to k.)
- All edges touching a vertex in V_i must lead to a vertex in V_{i-1} or V_{i+1}.

Note that:

$$\sum_{v \in V} f(v) = a\left(2 + \frac{|V_1|}{(d-1)^{1/2}} + ... + \frac{|V_k|}{(d-1)^{k/2}}\right) + b\left(2 + \frac{|U_1|}{(d-1)^{1/2}} + ... + \frac{|U_k|}{(d-1)^{k/2}}\right)$$

This means that we can clearly pick a and b so that the sum is zero (and we define them as such). Also note that:

$$f^T f = \sum_{i=0}^{k} \sum_{v \in V_i} \left(\frac{a}{(d-1)^{i/2}}\right)^2 + \sum_{i=0}^{k} \sum_{u \in U_i} \left(\frac{b}{(d-1)^{i/2}}\right)^2 = a^2 \sum_{i=0}^{k} \frac{|V_i|}{(d-1)^i} + b^2 \sum_{i=0}^{k} \frac{|U_i|}{(d-1)^i}$$

Later, we will denote $a^2 \sum_{i=0}^{k} \frac{|V_i|}{(d-1)^i}$ as $f^T f_V$ and the analogous term as $f^T f_U$.
By Lemma 1.3.1 (notice we are using the non-normalized Laplacian),

$$f^T L f = \sum_{\{u,v\} \in E} (f(u) - f(v))^2$$

The definition of f allows us to disregard a lot of the edges. Note that any edges that connect two vertices, neither of which are in one of the V_j or U_j sets will add zero to the sum, just as any edges that connect two vertices in the same V_j set will. So, we only need to consider edges that travel between sets V_i and V_{i+1}, between U_i and U_{i+1}, and between V_k and U_k to nodes in no such set. Also note that there are no edges from a vertex in any V_j set to any U_j set.

This means that the sum can effectively be broken into two parts – one where we consider edges touching vertices in V_j sets and the other where we do the same thing for U_j sets. These parts are completely analogous, so we just look at the first part of the sum. If we say that $f^T LF = V^* + U^*$, where V^* is this first part of the sum, then we can say that:

$$V^* \leq \sum_{i=0}^{k-1} \left[|V_i|(d-1) \left(\frac{a}{(d-1)^{i/2}} - \frac{a}{(d-1)^{(i+1)/2}} \right)^2 \right] + |V_k|(d-1) \left(\frac{a}{(d-1)^{k/2}} \right)^2$$

Notice that the ith term of the sum represents the edges that travel from V_i to V_{i+1}, and the term outside the sum represents the edges that travel from a vertex in V_k to a vertex not in any of the V_j, U_j sets.
Rearranging the right side yields:

$$V^* \leq a^2 \left(\sum_{i=0}^{k-1} \left[|V_i|(d-1) \left(\frac{1}{(d-1)^{i/2}} - \frac{1}{(d-1)^{(i+1)/2}} \right)^2 \right] + |V_k| \frac{d-1}{(d-1)^k} \right)$$

Since $(d - 2\sqrt{d-1}) + (2\sqrt{d-1} - 1) = d - 1$, we can separate the last term as follows:

$$V^* \leq a^2 \left(\sum_{i=0}^{k-1} \left[|V_i|(d-1) \left(\frac{1}{(d-1)^{i/2}} - \frac{1}{(d-1)^{(i+1)/2}} \right)^2 \right] + \frac{|V_k|}{(d-1)^k}(d - 2\sqrt{d-1}) + \frac{|V_k|}{(d-1)^k}(2\sqrt{d-1} - 1) \right)$$

Also, we have:

$$\begin{aligned}
&(d-1) \left(\frac{1}{(d-1)^{i/2}} - \frac{1}{(d-1)^{(i+1)/2}} \right)^2 \\
&= (d-1) \left(\frac{1}{(d-1)^{i/2}} \left(1 - \frac{1}{(d-1)^{1/2}} \right) \right)^2 \\
&= \frac{1}{(d-1)^i} \sqrt{d-1}\sqrt{d-1} \left(1 - \frac{1}{\sqrt{d-1}} \right) \left(1 - \frac{1}{\sqrt{d-1}} \right) \\
&= \frac{1}{(d-1)^i} \left(\sqrt{d-1} - 1 \right) \left(\sqrt{d-1} - 1 \right) \\
&= \frac{1}{(d-1)^i} (d - 2\sqrt{d-1})
\end{aligned}$$

This means that we can further simplify our upper bound on V^*:

$$V^* \leq a^2 \left(\sum_{i=0}^{k-1} \left[|V_i| \frac{1}{(d-1)^i}(d - 2\sqrt{d-1}) \right] + \frac{|V_k|}{(d-1)^k}(d - 2\sqrt{d-1}) + \frac{|V_k|}{(d-1)^k}(2\sqrt{d-1} - 1) \right)$$

Now the $\frac{|V_k|}{(d-1)^k}(d - 2\sqrt{d-1})$ term is in the correct form to be pushed into the summation:

$$V^* \leq a^2 \left(\sum_{i=0}^{k} \left[|V_i| \frac{1}{(d-1)^i}(d - 2\sqrt{d-1}) \right] + \frac{|V_k|}{(d-1)^k}(2\sqrt{d-1} - 1) \right)$$

Now, note that since $|V_i| \leq (d-1)|V_{i+1}|$, $\frac{|V_k|}{(d-1)^k} \leq \frac{|V_j|}{(d-1)^j}$ for $0 \leq j \leq k-1$. Therefore, the $\frac{|V_k|}{(d-1)^k}$ term is less than the average of those terms:

$$\frac{|V_k|}{(d-1)^k} \leq \frac{1}{k+1} \sum_{i=0}^{k} \frac{|V_i|}{(d-1)^i}$$

So, now we have:

$$V^* \leq a^2(d-2\sqrt{d-1})\sum_{i=0}^{k}\left[|V_i|\frac{1}{(d-1)^i}\right] + a^2(2\sqrt{d-1}-1)\frac{1}{k+1}\sum_{i=0}^{k}\frac{|V_i|}{(d-1)^i}$$

$$\implies V^* \leq (d-2\sqrt{d-1})f^T f_V + \frac{2\sqrt{d-1}-1}{k+1}f^T f_V$$

$$\implies \frac{V^*}{f^T f_V} \leq d-2\sqrt{d-1}+(2\sqrt{d-1}-1)\frac{1}{k+1} := \alpha$$

So, we have $\frac{V^*}{f^T f_V} \leq \alpha$ and similarly, $\frac{U^*}{f^T f_U} \leq \alpha$. This means that:

$$\frac{f^T L f}{f^T f} = \frac{V^*+U^*}{f^T f_V + f^T f_U} \leq \frac{\alpha f^T f_V + \alpha f^T f_U}{f^T f_V + f^T f_U} = \alpha$$

Finally, let κ be the second smallest eigenvalue of the non-normalized Laplacian. Then, just like the normalized Laplacian, all eigenvalues are non-negative and 0 is always an eigenvalue given by some eigenvector with constant entries (called C). Then:

$$\kappa = \min_{h \perp C}\frac{h^T L h}{h^T h} \leq \frac{f^T L f}{f^T f} \leq \alpha = d-2\sqrt{d-1}+(2\sqrt{d-1}-1)\frac{1}{k+1}$$

The first inequality follows form the fact that $f \perp C$ because we picked f so that the sum of its entries over all the vertices equals zero. Also, note that κ, the second smallest eigenvalue of L, is the second largest eigenvalue of A (see Corollary 3.3.1 where we mention that an eigenvalue v of the non-normalized Laplacian is an eigenvalue $d-v$ of the adjacency matrix). Therefore,

$$\kappa \leq d-2\sqrt{d-1}+(2\sqrt{d-1}-1)\frac{1}{k+1} \implies \mu_1 \geq d-\alpha = 2\sqrt{d-1}-(2\sqrt{d-1}-1)\frac{1}{k+1}$$

Finally, we have:

$$2\sqrt{d-1}-(2\sqrt{d-1}-1)\frac{1}{k+1} \leq \mu_1 \leq \hat{\mu}(G)$$

In order to reach the final result, consider a graph with n nodes that is d-regular. Pick an arbitrary vertex S and observe that, roughly speaking, at most d vertices are connected to S. Then, following this reasoning, at most d^2 vertices are connected to those vertices and so on. This means that at most d^k vertices are within k steps of S, and that $\text{diam}(G) \geq \log_d n$. Therefore, any d-regular family of graphs has a growing diameter. As n gets large, the diameter gets large, and we can set k to be a larger and larger value. Therefore, for any d-regular graph family, G_n, $\hat{\mu}(G_n) \geq 2\sqrt{d-1}-o(1)$. This means that there doesn't exist an infinite family of graphs for unbounded n such that for all graphs G in the family, $\hat{\mu}(G) \leq 2\sqrt{d-1}-\epsilon$, for any fixed $\epsilon > 0$. □

6 Existence of bipartite Ramanujan graphs

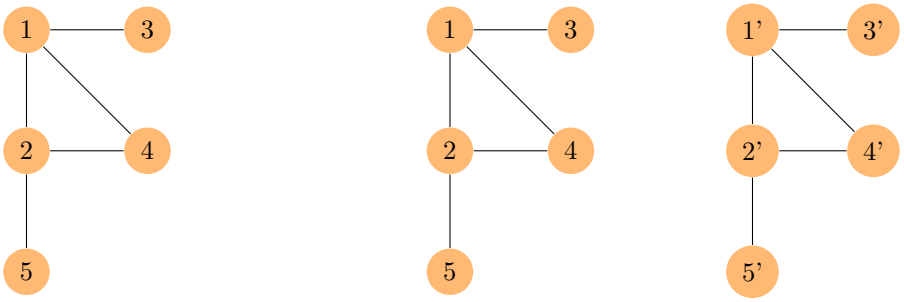
In this section, we give an overview of the arguments in [20], where it is proven that there exist infinite families of bipartite Ramanujan graphs of all degrees at least 3.

6.1 2-Lifts

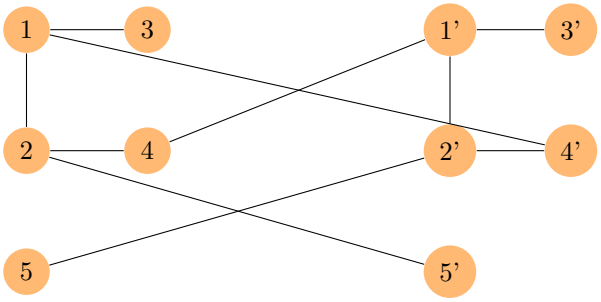
Definition. *(2-lifts)* A 2-lift is a process that acts on a graph, turning it into a graph with twice as many vertices. The process is as follows. Take a graph, and draw a duplicate graph next to it. Then, every original edge $\{u, v\}$ now has two analogs: $\{u_0, v_0\}$ and $\{u_1, v_1\}$. Then, select some subset of the edge pairs $\{u_0, v_0\}$ and $\{u_1, v_1\}$ and transform the edges to: $\{u_0, v_1\}$ and $\{u_1, v_0\}$.

The process can be visualized below.

Step 1 (original graph) **Step 2 (after node duplication)**



Step 3 (cross some of the edges)



Let us call the original graph G, and the resulting graph after completing the 2-lift G'. We also say that G has a $n \times n$ adjacency matrix $A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$. However, rather than considering A', the $2n \times 2n$ adjacency matrix of G', we will notice that we can capture the structure of G' with a signed version of the $n \times n$ matrix A. Recall that in constructing the 2-lift, some of the original edges in G ended up being crossed with the analogous edges among the new vertices.

Definition. *(signed adjacency matrix)* We create the A_s matrix (the signed adjacency matrix for G') by first copying the matrix A. Then, if the edge $\{u, v\}$ was one of the crossed edges in the 2-lift, we replace the 1 at $(A_s)_{uv}$ and $(A_s)_{vu}$ with a -1.

In this small example, we have $A_s = \begin{bmatrix} 0 & 1 & 1 & -1 & 0 \\ 1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \end{bmatrix}$

(Notice that our following discussion in this section is for regular graphs, although this example is not regular.)
It is not immediately clear why it is useful to create such a matrix. However, this matrix is important because its eigenvalues are closely related to the eigenvalues of G' (which are the A' eigenvalues).

Theorem 6.1. *The eigenvalues of G' are the union of the eigenvalues of A and the eigenvalues of A_s (counting multiplicities).*

Proof. We present the proof of Bilu and Linial [4]. As before, let A be the adjacency matrix of G (the graph before the 2-lift), let A' be the adjacency matrix of G', the 2-lift. We also define the zero/one matrices A_1

and A_2:

$$(A_1)_{ij} = \begin{cases} 1 & \text{if } A_{ij} = 1 \text{ and } (A_s)_{ij} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$(A_2)_{ij} = \begin{cases} 1 & \text{if } A_{ij} = 1 \text{ and } (A_s)_{ij} = -1 \\ 0 & \text{otherwise} \end{cases}$$

We can say that A_1 is the restriction of A, only with edges that were not crossed, and that A_2 is the restriction, leaving only the crossed edges (but still representing them with a 1). Clearly, this means that $A = A_1 + A_2$, and $A_s = A_1 - A_2$. It is also quite clear that $A' = \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_2 & A_1 \end{array}\right]$

We will now define a notation for concatenating vectors. If v_1 and v_2 are vectors in $\mathbb{R}^n$, then $[v_1, v_2]$ denotes the vector in $\mathbb{R}^{2n}$ that results from copying v_2 onto the end of v_1. In light of the fact that $A' = \left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_2 & A_1 \end{array}\right]$ it is very easy to confirm that v is an eigenvector of A, with eigenvalue μ implies that $[v, v]$ is an eigenvector of A', with the same eigenvalue. Similarly, u is an eigenvector of A, with eigenvalue λ implies that $[u, -u]$ is an eigenvector of A', with the same eigenvalue. It is also clear that $v_1 \perp v_2 \implies [v_1, v_1] \perp [v_2, v_2]$ and $u_1 \perp u_2 \implies [u_1, -u_1] \perp [u_2, -u_2]$. Therefore, we know that the elements of the set $\{[v, v] | v \text{ an eigenvector of } A\}$ are pairwise orthogonal, and the same can be said of $\{[u, -u] | u \text{ an eigenvector of } A_s\}$. Also, since $[v, v] \cdot [u, -u] = v \cdot u + v \cdot -u = v \cdot u - v \cdot u = 0$, we can say that the entire set $S = \{[v, v] | v \text{ an eigenvector of } A\} \cup \{[u, -u] | u \text{ an eigenvector of } A_s\}$ is pairwise orthogonal. Since A and A_s are symmetric matrices, each matrix has n eigenvectors and therefore, $|S|$ has $2n$ (orthogonal) eigenvectors of A'. Therefore, not only are all the eigenvalues of A and A_s eigenvalues of A', but these are all of the eigenvalues of A'. □

Already, we can significantly reduce the problem of finding bipartite Ramanujan graphs to finding appropriate 2-lifts. Suppose that we want to create an infinite family of d-regular bipartite Ramanujan graphs. We start with $K_{d,d}$, the complete d-regular bipartite graph (with $2d$ nodes). The adjacency matrix looks like $\left[\begin{array}{c|c} 0 & 1 \\ \hline 1 & 0 \end{array}\right]$, and so the eigenvalues of this graph are d and $-d$, as we already know, and the remaining eigenvalues are 0. This makes intuitive sense, as a complete graph is as well connected as possible. Thus, $K_{d,d}$ is Ramanujan. Note that every graph with m edges has 2^m 2-lifts, since for each edge we choose whether to cross it or not. Then, we want to prove that every regular bipartite graph has some 2-lift that preserves the Ramanujan property. However, if the starting graph is Ramanujan, we only need to look at the A_s eigenvalues to determine that the new lifted graph is also Ramanujan (by Theorem 7.1). Then, we can apply 2-lifts repeatedly to $K_{d,d}$ to obtain an infinite bipartite Ramanujan family (the 2-lift procedure clearly preserves d-regularity and bipartiteness). The family we will find consists of graphs where $|V| = 2^j d$, for $j \in \mathbb{Z}^+$.

By the basic eigenvalue definitions, we know that the largest eigenvalue of A_s is the largest root of the characteristic polynomial for A_s. Here, we represent the polynomial as $\chi A_s(x)$ and the largest root as $\lambda_{\max}(\chi A_s(x))$. We are not only concerned about the largest root of $\chi A_s(x)$ but also the smallest one. However, the following theorem shows us that we need only concern ourselves with the largest root (meaning the largest eigenvalue of A_s).

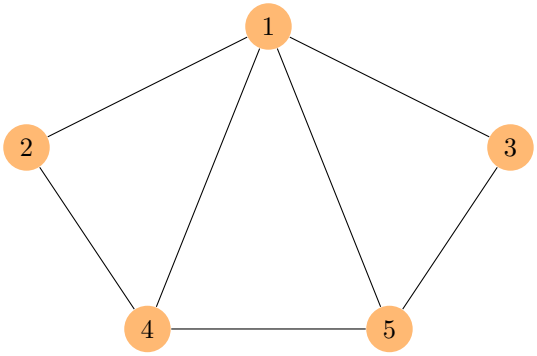
Theorem 6.2. *The adjacency eigenvalues for bipartite graphs are symmetric around zero.*

Proof. The adjacency matrix of a bipartite graph can be given in the form $A = \left[\begin{array}{c|c} 0 & B \\ \hline B^T & 0 \end{array}\right]$. In the rest of the proof, we will assume that the blocks are of the same dimension, although this is not necessary and the same proof suffices. (However, note that a d-regular bipartite graph must have a vertex partition that has equally many vertices on each side, meaning that these blocks are the same dimension.) Suppose that a vector $\hat{v} = [v_1, v_2]$ is a λ-eigenvector for A, where the length of v_1 equals the length of v_2 equals n, if A is a $2n \times 2n$ matrix. We have $\begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$. But we will see that $\hat{v}' = [v_1, -v_2]$ is an eigenvector for eigenvalue $-\lambda$: $\begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ -v_2 \end{bmatrix} = \lambda \begin{bmatrix} -v_1 \\ v_2 \end{bmatrix} = -\lambda \begin{bmatrix} v_1 \\ -v_2 \end{bmatrix}$ □

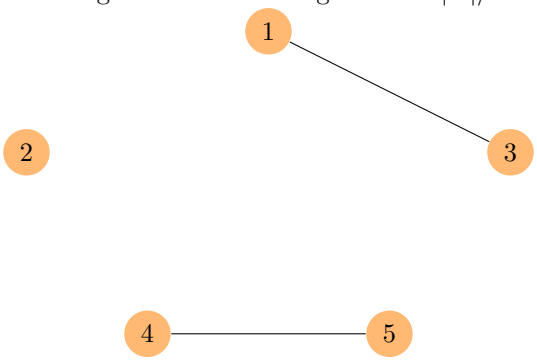
6.2 The matching polynomial

Definition. *(graph matching)* A graph matching for a graph G is a set of the edges of G so that none of the selected edges share a vertex with any other edge.

For example, take the following graph:



Then, there are 7 matchings of G that use only one edge (one matching per edge), and many other matchings that use different numbers of edges. Note that it is trivial that a graph with n nodes can only have matchings with at most $n/2$ edges, as this will use up all of the edges and further edge inclusion will mean that there are some shared vertices. As an example, we will see a matching of G with two edges. Since $|V|/2 < 3$, we know that this is a maximal matching (one of several).



We now introduce an important graph polynomial.

Definition. *(matching polynomial)* The matching polynomial for G is defined as

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

where $m_i(G)$ gives the number of matchings on G with exactly i edges ($m_0 := 1$). Note that $m_i(G) = 0 \ \forall i > \frac{n}{2}$.

6.2.1 Relationship to characteristic polynomials of A_s

We showed in section 6.1 that we need to concern ourselves with the characteristic polynomials of the A_s matrices. Here, we will prove a theorem, originally from [13], that relates these polynomials to the matching polynomial.

Theorem 6.3. $\mathbb{E}_s\left[\chi A_s(x)\right] = \mu_G(x)$

Proof. Using the definition of the characteristic polynomials of a matrix, and the Leibniz formula for the determinant, we obtain:

$$\mathbb{E}_s\left[\chi A_s(x)\right] = \mathbb{E}_s\left[\det(xI - A_s)\right] = \mathbb{E}_s\left[\sum_{\sigma \in \mathrm{sym}([n])} (-1)^{|\sigma|} \prod_{i=1}^{n} (xI - A_s)i, \sigma(i)\right]$$

Rather than simply sum over every permutation σ, we can consider the subset of $[n]$ that σ does not act as the identity on, and consider σ as a permutation over that subset of $[n]$ of size k. It is easy to check that this

reordering yields equality as every permutation sigma is represented in the following sum and no permutation is double counted. Note that the identity permutation on $[n]$ is counted when $k = 0$ as we say that there is one permutation over the empty set (and this permutation never acts as the identity).

$$= \mathbb{E}_s \left[\sum_{k=0}^{n} x^{n-k} \sum_{\substack{S \subseteq [n] \\ |S|=k}} \sum_{\substack{\pi \in \text{sym}(S) \\ \pi(i) \neq i}} (-1)^{|\pi|} \prod_{i \in S} (xI - A_s)i, \pi(i) \right]$$

Notice that the x^{n-k} term after the first sum comes from extending the permutation π from being over a subset f $[n]$ to being over all of $[n]$, and setting π as the identity on the new elements. The diagonal of $xI - A_s$ is equal to x, and the product will have $n - k$ powers of x for a permutation π. Now that the diagonal will never be reached by a permutation π, we can remove the xI portion of the matrix and push the expected value further inside. We also take the string $s \in \{\pm 1\}^{|E|}$ and allow ourselves to use a double subscript with it by saying that s_{ij} equals the value of s on the edge that connects vertex i to vertex j. (If the vertices are not connected, we extend s and set it equal to zero. This means that now, s has length $n(n-1)$.) Now, we have the following:

$$= \sum_{k=0}^{n} x^{n-k} \sum_{\substack{S \subseteq [n] \\ |S|=k}} \sum_{\substack{\pi \in \text{sym}(S) \\ \pi(i) \neq i}} \mathbb{E}_s \left[(-1)^{|\pi|} \prod_{i \in S} (-s)_{i,\pi(i)} \right]$$

Now, let us take a closer look at the $(-1)^{|\pi|} \prod_{i \in S} (-s)_{i,\pi(i)}$ term. By the definition of the extension of s, we know that this product equals zero for every s unless every vertex pair i, $\pi(i)$ is connected. This means that we can consider only permutations π such that, for all $i \in S$, i and π(i) are connected. In this light, we can think of each such π as a union of disjoint cycles over the vertices in S. If we decompose π into its cycles, we have:

$$= \sum_{k=0}^{n} x^{n-k} \sum_{\substack{S \subseteq [n] \\ |S|=k}} \sum_{\substack{\pi \in \text{sym}(S) \\ \pi(i) \neq i}} \mathbb{E}_s \left[(-1)^{|\pi|} \prod_{\text{cyc} \in \pi} \prod_{i \in \text{cyc}} (-s)_{i,\pi(i)} \right]$$

Since the cycles are disjoint, we know that we can use independence to say:

$$= \sum_{k=0}^{n} x^{n-k} \sum_{\substack{S \subseteq [n] \\ |S|=k}} \sum_{\substack{\pi \in \text{sym}(S) \\ \pi(i) \neq i}} (-1)^{|\pi|} \prod_{\text{cyc} \in \pi} \mathbb{E}_s \left[\prod_{i \in \text{cyc}} (-s)_{i,\pi(i)} \right]$$

Now suppose that π contains a 3-cycle (or any cycle larger than 2), which we will denote (abc). Then, $\prod_{i \in \text{cyc}} (-s)_{i,\pi(i)} = ((-s)_{ab})((-s)_{bc})((-s)_{ca})$. But independence again allows us to say that the expected value of the whole product is the product of the expected values, and we have:

$$\mathbb{E}_s \left[\prod_{i \in \text{cyc}} (-s)_{i,\pi(i)} \right] = \mathbb{E}_s \left[((-s)_{ab}) \right] \mathbb{E}_s \left[((-s)_{bc}) \right] \mathbb{E}_s \left[((-s)_{ca}) \right] = 0 * 0 * 0 = 0$$

This follows because s takes values in $\{\pm 1\}$ uniformly at random. The 2-cycles in π remain, however, since

$$\mathbb{E}_s \left[\prod_{i \in \text{cyc}} (-s)_{i,\pi(i)} \right] = \mathbb{E}_s \left[((-s)_{ab})((-s)_{ba}) \right] = \mathbb{E}_s \left[1 \right] = 1$$

(This is because those two random variables are equal and thus not independent. Also, they both equal 1 or -1.) So, we can now assume that π contains only 2-cycles. We can reduce the formula to:

$$= \sum_{k=0}^{n} x^{n-k} \sum_{\substack{S \subseteq [n] \\ |S|=k}} \sum_{\pi \text{ perfect matching on } S} \mathbb{E}_s \left[(-1)^{|\pi|} \prod_{\text{cyc} \in \pi} \prod_{i \in \text{cyc}} (-s)_{i,\pi(i)} \right]$$

Since any perfect matching can only occur when k is even, we have:

$$= \sum_{k=0}^{n} x^{n-k} \sum_{\substack{S\subseteq[n] \\ |S|=k \\ k \text{ even}}} \sum_{\pi \text{ perfect matching on } S} (-1)^{\frac{k}{2}} \prod_{\text{cyc}\in\pi} 1$$

$$= \sum_{k=0}^{n} x^{n-k} \sum_{\substack{S\subseteq[n] \\ |S|=k \\ k \text{ even}}} \sum_{\pi \text{ perfect matching on } S} (-1)^{\frac{k}{2}} = \sum_{k=0}^{n} x^{n-k} \sum_{\substack{S\subseteq[n] \\ |S|=k \\ k \text{ even}}} (-1)^{\frac{k}{2}} m_{\frac{k}{2}}(G|_S)$$

Note that $G|_S$ denotes the subgraph of G containing only vertices in S. Reindexing yields:

$$\sum_{k=0}^{n} x^{n-2k}(-1)^k \sum_{\substack{S\subseteq[n] \\ |S|=2k}} m_k(G|_S)$$

However, $\sum_{\substack{S\subseteq[n] \\ |S|=2k}} m_k(G|_S) = m_k(G)$, and we are done:

$$\mathbb{E}_s\left[\chi A_s(x)\right] = \sum_{k=0}^{n} x^{n-2k}(-1)^k \sum_{\substack{S\subseteq[n] \\ |S|=2k}} m_k(G|_S) = \sum_{k=0}^{n} x^{n-2k}(-1)^k m_k(G) = \mu_G(x)$$

□

6.2.2 Bounding the roots of the matching polynomial

The purpose of this section is to show that for any graph G, the real roots of $\mu_G(x)$ (in fact, all roots are real) are in the interval
$[-2\sqrt{\triangle(G)-1}, 2\sqrt{\triangle(G)-1}]$, where $\triangle(G)$ is the largest degree of any vertex in G. We need two lemmas first, which come from [5]. The main result of Theorem 6.4 was originally proved in [15] and is also located in [10].

Lemma 6.2.1. *Denote the graph G, with the vertex i removed (and all edges touching it) as $G\backslash i$. Accordingly, $G\backslash i\backslash j := (G\backslash i)\backslash j$ and is the graph with the vertices i, j removed (and all edges touching them). Then,*

$$\mu_G(x) = x\mu_{G\backslash i}(x) - \sum_{\{i,j\}\in E} \mu_{G\backslash i\backslash j}(x)$$

Proof. Clearly, the number of matchings of size k over G, that do not involve the vertex i is $m_k(G\backslash i)$. It is also clear that the number of matchings over G that involve vertex i is $\sum_{\{i,j\}\in E} m_{k-1}(G\backslash i\backslash j)$. Since the number of matchings of size k over G is the number of matchings that include i plus that number of matchings that don't, we have:

$$m_k(G) = m_k(G\backslash i) + \sum_{\{i,j\}\in E} m_{k-1}(G\backslash i\backslash j)$$

Now, we can plug this identity into the full equations. We have:

$$\mu_G(x) = \sum_{k\geq 0}(-1)^k x^{|V|-2k} m_k(G) = \sum_{k\geq 0}(-1)^k x^{|V|-2k}\left(m_k(G\backslash i) + \sum_{\{i,j\}\in E} m_{k-1}(G\backslash i\backslash j)\right)$$

$$= \sum_{k\geq 0}(-1)^k x^{|V|-2k} m_k(G\backslash i) + \sum_{k\geq 0}(-1)^k x^{|V|-2k} \sum_{\{i,j\}\in E} m_{k-1}(G\backslash i\backslash j)$$

$$= x\sum_{k\geq 0}(-1)^k x^{|V|-1-2k} m_k(G\backslash i) + \sum_{k\geq 0}\left[(-1)^{k+1} x^{|V|-2(k+1)} \sum_{\{i,j\}\in E} m_k(G\backslash i\backslash j)\right]$$

Note that the index in the second summation can be increased by 1 because the only neglected term has value 0 (we can define m_{-1} to be 0). Continuing, we have:

$$x\mu_{g\setminus i}(x) - \sum_{k\geq 0}\left[(-1)^k x^{|V|-2-2k}\sum_{\{i,j\}\in E} m_k(G\setminus i\setminus j)\right]$$

$$= x\mu_{g\setminus i}(x) - \sum_{\{i,j\}\in E}\left[\sum_{k\geq 0}\left[(-1)^k x^{|V|-2-2k} m_k(G\setminus i\setminus j)\right]\right]$$

$$= x\mu_{g\setminus i}(x) - \sum_{\{i,j\}\in E}\mu_{G\setminus i\setminus j}(x)$$

□

Now, we need another lemma relating the ratios of matching polynomials.

Lemma 6.2.2. *Take $\delta \geq \triangle(G) > 1$. If $\deg(i) < \delta$, then*

$$x > 2\sqrt{\delta-1} \implies \frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > \sqrt{\delta-1}$$

Proof. We start with the base case for an inductive argument (inducting over $|V|$). Say that $|V| = 1$. Then, there can be no edges, and we have that $\mu_G(x) = x$ since there is one matching with 0 edges and zero matchings with 1 edge. Similarly, we have that $\mu_{G\setminus i}(x) = 1$. Clearly the lemma holds in this case since $x > 2\sqrt{\delta-1} \implies \frac{\mu_G(x)}{\mu_{G\setminus i}(x)} = x = 2\sqrt{\delta-1} > \sqrt{\delta-1}$. So now we suppose that the lemma holds for graphs on n vertices. By the previous lemma, we have:

$$\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} = x - \sum_{\{i,j\}\in E}\frac{\mu_{G\setminus i\setminus j}(x)}{\mu_{G\setminus i}(x)}$$

Plugging in our assumption about x and invoking the inductive hypothesis yields:

$$\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > 2\sqrt{\delta-1} - \frac{1}{\sqrt{\delta-1}}\sum_{\{i,j\}\in E} 1 = 2\sqrt{\delta-1} - \frac{\deg(i)}{\sqrt{\delta-1}}$$

Note that we assumed that $\deg(i) \leq \delta - 1$. So we continue:

$$\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > 2\sqrt{\delta-1} - \frac{\delta-1}{\sqrt{\delta-1}} = \sqrt{\delta-1}$$

□

Now, we are finally ready to prove a bound on the roots of the matching polynomial.

Theorem 6.4. *All the real roots of $\mu_G(x)$ are in the interval $[-2\sqrt{\triangle(G)-1}, 2\sqrt{\triangle(G)-1}]$, where $\triangle(G)$ is the largest degree of any vertex in G.*

Proof. First, note that the powers of x of the matching polynomial are centered about 0, meaning that the polynomial is either even or odd. This means that it is sufficient to show that there are no roots for $x > 2\sqrt{\triangle(G)-1}$. As a base case for induction over $|V|$, we know that the graph on one vertex has $\mu_G(x) = x$, and the only zero of x is at zero which is in the desired interval.

We know from Lemma 6.2.1 that

$$\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} = x - \sum_{\{i,j\}\in E}\frac{\mu_{G\setminus i\setminus j}(x)}{\mu_{G\setminus i}(x)}$$

Our assumption on the value of x means that

$$\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > 2\sqrt{\triangle(G)-1} - \sum_{\{i,j\}\in E} \frac{\mu_{G\setminus i\setminus j}(x)}{\mu_{G\setminus i}(x)}$$

If there are no edges between i and j in G, then we are done. The reason for this is that the sum is empty, meaning we have $\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > 2\sqrt{\triangle(G)-1}$. But then by the induction hypothesis, $\mu_{G\setminus i}(x)$ has no zeroes for the $x > 2\sqrt{\triangle(G)-1}$. This means that $\mu_G(x)$ has no zeroes for the $x > 2\sqrt{\triangle(G)-1}$ as otherwise the quotient on the left would be 0, which is less than $2\sqrt{\triangle(G)-1}$.

So we can suppose that there is an edge between i and j. In the notation of Lemma 6.2.2, set $\delta := \triangle(G)$. Notice that we now know that

$$\deg_{G\setminus i}(j) \leq \triangle(G) - 1 < \triangle(G) = \delta$$

This is where the first inequality follows from the fact that there is an edge between i and j – an edge that is no longer present in $G\setminus i$. This string of inequalities allows us to invoke Lemma 6.2.2, where $G\setminus i$ is the original graph. We already knew

$$\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > 2\sqrt{\triangle(G)-1} - \sum_{\{i,j\}\in E} \frac{\mu_{G\setminus i\setminus j}(x)}{\mu_{G\setminus i}(x)}$$

and now we can simplify to

$$\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > 2\sqrt{\triangle(G)-1} - \sum_{\{i,j\}\in E} \frac{1}{\sqrt{\delta-1}}$$

$$\implies \frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > 2\sqrt{\triangle(G)-1} - \frac{\deg_G(i)}{\sqrt{\triangle(G)-1}} \geq 2\sqrt{\triangle(G)-1} - \frac{\triangle(G)}{\sqrt{\triangle(G)-1}}$$

However, one can show easily that $2\sqrt{\triangle(G)-1} - \frac{\triangle(G)}{\sqrt{\triangle(G)-1}} \geq 0$ (for $\triangle(G) \geq 2$) using basic calculus or some other method. So, we now know that $\frac{\mu_G(x)}{\mu_{G\setminus i}(x)} > 0$ whenever $x > 2\sqrt{\triangle(G)-1}$. But now we can make the same inductive argument again. The denominator of the left hand side is not zero, so we know that the numerator is also never zero. Therefore, no real zeroes of any matching polynomial fall outside of the interval $[-2\sqrt{\triangle(G)-1}, 2\sqrt{\triangle(G)-1}]$. □

6.3 The difficulty with averaging polynomials

At this point in the existence proof of bipartite Ramanujan graphs, it might seem like we are done. After all, we know that $\mathbb{E}_s[\chi A_s(x)] = \mu_G(x)$ and thus

$$\lambda_{\max}(\mathbb{E}_s[\chi A_s(x)]) = \lambda_{\max}(\mu_G(x)) \leq 2\sqrt{d-1}$$

With a finite set of numbers, we can say that not all elements of the set are above average (or all below). Thus, we might be led to presume that we can extend this reasoning to polynomials and then already know that there exists some $s \in \{1,-1\}^{|E|}$ such that $\lambda_{\max}(\chi A_s(x)) \leq 2\sqrt{d-1}$. After all, how could the maximum root of all the polynomials $\chi A_s(x)$ be above the maximum root of the average? We will see that in general, this can certainly happen. This means that we need some further analysis. Here is the example given in the original paper [20]. Take two polynomials $f_0(x) = (x+5)(x-9)(x-10)$ and $f_1(x) = (x+6)(x-1)(x-8)$. Then, the largest roots of these polynomials are 10 and 8. We will see that both of these roots are larger than the largest root of the average polynomial, $\frac{1}{2}(f_0(x)+f_1(x))$. The average polynomial is $x^3 - \frac{17}{2}x^2 - \frac{51}{2}x + 249$, which has its largest root before 7.4. We will now see that the crucial fact that allows us to continue the proof is that the polynomials $\chi A_s(x)$ are an interlacing family, which means that this problem cannot arise.

6.4 Interlacing

Definition. *(polynomial interlacing)* We say that a real rooted degree $n-1$ polynomial g with roots $\alpha_1, ..., \alpha_{n-1}$ interlaces a real-rooted degree n polynomial f with roots $\beta_1, ..., \beta_n$ if

$$\beta_1 \leq \alpha_1 \leq ... \leq \alpha_{n-1} \leq \beta_n$$

Definition. *(common interlacing)* If there exists a polynomial g that interlaces each of the polynomials $f_1, ..., f_k$, then we say that the $f_1, ..., f_k$ have a common interlacing.

We will now prove some lemmas that will eventually allow us to show that the averaging polynomials problem can be resolved by considering interlacing. First, we will state a lemma that relates the existence of a common interlacer to real-rootedness of certain polynomials. We will present this lemma without proof, but the proofs are located in [8,11]. The arguments are fairly long, but straightforward and involving only basic polynomial arguments.

Lemma 6.4.1. *If $f_1, ..., f_k$ are univariate polynomials of the same degree and with positive leading coefficient, then $f_1, ..., f_k$ have a common interlacing $\iff$ all convex combinations of the $f_1, ..., f_k$ are real rooted. (This means that $\sum_i \eta_i f_i$ is real rooted for all settings of the η_i where all $\eta_1 \geq 0$ and the η_i sum to 1).*

Now, we will show that a common interlacing resolves the problem of averaging polynomials in a specific case.

Lemma 6.4.2. *If $f_1, ..., f_k$ are of the same degree, are real rooted, have a positive coefficient on the highest degree term, and have a common interlacing, then if we define*

$$f_\varnothing = \sum_{i=1}^{k} f_i$$

we can say that there exists an i so that f_i has a largest root that is at most the largest root of $f_\varnothing$.

Proof. We know that each of the f_i tends to infinity (has positive leading coefficient) and has exactly one root at or after α_{n-1}. (The α_j are the roots of the common interlacer.) Therefore, $\forall i$, $f_i(\alpha_{n-1}) \leq 0$. Thus, the sum of the f_i, which is $f_\varnothing$, is less than or equal to zero at α_{n-1}. Since $f_\varnothing$ also tends to infinity, we know that it has a root at or after α_{n-1}, which we will denote as γ. Since $f_\varnothing(\gamma) = 0$, we know that either all the f_i have a zero at γ (meaning that we are done) or there exists some i' so that $f_{i'}(\gamma) > 0$. But we already know that $f_{i'}$ was non-positive at α_{n-1} and thus has a root between α_{n-1} and γ. We also know that this must be the largest root of $f_{i'}$. So, the largest root of $f_{i'}$ is less than or equal to the largest root of $f_\varnothing$ □

Now, we will define a new term, one that is closely related to the common interlacing, and use the above lemma to see that it also allows us to say what we want about the largest root of an average of polynomials.

Definition. *(interlacing family)* We start with some finite sets $S_1, ..., S_m$ and define a degree n, real-rooted (with positive leading coefficient) polynomial for every choice of element from each set. This means that we now have functions $f_{s_1,...,s_m}$ where $s_1 \in S_1, ..., s_m \in S_m$. For every partial assignment of the set elements, we define a function as follows:

$$f_{s_1,...,s_k} = \sum_{s_{k+1} \in S_{k+1}, ..., s_m \in S_m} f_{s_1,...,s_k,s_{k+1},s_m}$$

Also, in analogy with the previous lemma, we define

$$f_\varnothing = \sum_{s_1 \in S_1, ..., s_m \in S_m} f_{s_1,...,s_m}$$

We say that the polynomials $\{f_{s_1,...,s_m} | s_1 \in S_1, ..., s_m \in S_m\}$ form an interlacing family if for all $k = 0, ..., m-1$ and for all correlating partial assignments $s_1, ...s_k$, we have that the polynomials $\{f_{s_1,...,s_k,t} | t \in S_{k+1}\}$ have a common interlacing.

Theorem 6.5. *If we have an interlacing family of polynomials, then there exists some choice of elements $s_1, ..., s_m$ from the $S_1, ..., S_m$ such that the largest root of $f_{s_1,...,s_m}$ is at most the largest real root of $f_\varnothing$.*

Proof. We will induct on k, the size of the partial assignment. If $k = 0$, then we have, straight from the definition of interlacing family, that the polynomials $\{f_t | t \in S_1\}$ have a common interlacing. Notice that the polynomials in this set are of the same degree (since all the $f_{s_1,...,s_m}$ are) and have positive leading coefficient (for the same reason). Now we can apply Lemma 6.4.1 to show that the polynomials here are real rooted (since all convex combinations are real-rooted, we can look at the combinations that set some η_i to 1 to show that each individual polynomial is real-rooted). Now, we satisfy the requirements of Lemma 6.4.2, and can say that there exists a t so that the largest root of f_t is at most the largest root of the sum, which equals $\sum_{s_1 \in S_1} \sum_{s_2 \in S_2,...,s_m \in S_m} f_{s_1,...,s_m} = f_\varnothing$. To induct, we can suppose that for some partial assignment $s_1^*, ..., s_k^*$, the largest root of $f_{s_1^*,...,s_k^*}$ is at most the largest root of $f_\varnothing$. Then, since $\{f_{s_1^*,...,s_k^*,t} | t \in S_{k+1}\}$ has a common interlacer, we know by Lemmas 6.4.1 and 6.4.2 that there is a $t^* \in S_{k+1}$ such that the largest root of $f_{s_1^*,...,s_k^*,t^*}$ is at most the largest root of $\sum_t f_{s_1^*,...,s_k^*,t^*} = f_{s_1^*,...,s_k^*}$, a polynomial with largest root at most the largest root of $f_\varnothing$, by the inductive hypothesis. Continuing this argument until $k = m - 1$, we achieve the result. □

We now just need to show that the polynomials $\{\chi_{A_s}(x)\}_{s \in \{0,1\}^m}$ are an interlacing family.

Theorem 6.6. *The polynomials $\{\chi_{A_s}(x)\}_{s \in \{0,1\}^m}$ are an interlacing family.*

Proof. Since we are considering signings again, we can take our sets $S_1, ..., S_m$ to be $\{1, -1\}$. We set $f_{s_1,...,s_m}$ to equal χ_{A_s} where the signing $s = s_1, ..., s_m$. Let us consider, for any $\lambda \in [0, 1]$, any k such that $0 \leq k \leq m-1$, and any partial setting of length k, the polynomial

$$\lambda f_{s_1,...,s_k,1} + (1 - \lambda) f_{s_1,...,s_k,-1}$$

Let's assume for a moment that all polynomials of this form are real-rooted. We have

$$f_{s_1,...,s_m} = \chi_{A_s}(x) = \det(xI - A_s) = \sum_{\sigma \in \mathrm{sym}([n])} (-1)^{|\sigma|} \prod_{i=1}^{n} (xI - A_s) i, \sigma(i)$$

Considering the identity permutation, we know that all these polynomials have largest term x^n. Thus, the partial assignment functions, which are just sums of the $f_{s_1,...,s_m}$ have positive leading coefficient and are all of the same degree. So, (under our real-rootedness assumption), they satisfy the conditions of Lemma 6.4.1. Then, Lemma 6.4.1 tells us that $f_{s_1,...,s_k,1}$ and $f_{s_1,...,s_k,-1}$ have a common interlacer. This means that the $\{f_{s_1,...,s_m} | s_1 \in S_1, ..., s_m \in S_m\}$, which are the $\{\chi_{A_s}(x)\}_{s \in \{0,1\}^m}$ are an interlacing family.

To deal with the real-rootedness assumption we just made, we refer the reader to the original paper [20] where this was proven. The proof is rather long and uses the concept of real stability, which is a generalization of the notion of real-rootedness to the complex plane. □

We are now ready to prove the main existence proof for Ramanujan graphs.

Theorem 6.7. *There exist infinite families of bipartite Ramanujan graphs of all degrees at least 3.*

Proof. Select an arbitrary degree d at least 3. Then, we can start with $K_{d,d}$, the complete bipartite graph of degree d. This graph is Ramanujan. Therefore, when considering a two-lift of the graph, the new graph is Ramanujan if the eigenvalues of the signed adjacency matrix, A_s are in the appropriate interval, by Theorem 6.1. But we proved that the χ_{A_s} are an interlacing family, meaning that there is some signing s^* such that the largest root of $\chi_{A_s^*}$ is at most the largest root of the sum $\sum_{s \in \{0,1\}^m} \chi_{A_s}$. Clearly, the largest root of that sum is the same as the largest root of $\mathbb{E}_s [\chi A_s(x)]$, which equals $\mu_G(x)$, by Theorem 6.3. Theorem 6.4 tells us that the largest root of $\mu_G(x)$ is at most $2\sqrt{d-1}$. Therefore, the largest root of $\chi_{A_s^*}$ is at most $2\sqrt{d-1}$. But taking a 2-lift of a bipartite graph yields another bipartite graph, meaning that by Theorem 6.4, the smallest eigenvalue is at least $-2\sqrt{d-1}$. Therefore, we know that the new graph, after taking the 2-lift, is still Ramanujan. Continuing this process means that we get an infinite family of Ramanujan graphs of degree d. □

7 Pseudorandomness and applications to theoretical CS

7.1 Expander mixing lemma

This well-known theorem shows that the better an expander is (the smaller $\hat{\mu}(G)$ is, the more it appears to be random. We will first state and prove the theorem and then see why the theorem sheds light on the pseudorandomness of the graph.

Theorem 7.1. *(expander mixing lemma) Let S and T be subsets of the vertices of a d-regular graph G. Then,*

$$\left||E(S,T)| - \frac{d}{n}|S||T|\right| \leq \hat{\mu}(G)\sqrt{|S||T|}$$

Proof. We follow the arguments made in [16]. Note that in this context, we define $\hat{\mu}(G)$ as $\max\{|\mu_1|, |\mu_{n-1}|\}$. This means that connected bipartite graphs have $\hat{\mu} = d$.
Define 1_S and 1_T to be the indicator vectors for S and T. Also, define α_j and β_j so that the following hold: $1_S = \sum_{i=0}^{n-1} \alpha_i v_i$ and $1_T = \sum_{i=0}^{n-1} \beta_i v_i$ (the v_i are the orthonormal basis of A). It is easy to check that $1_S A 1_T = \sum_{v \in S} \sum_{i=0}^{n-1} A_{vj}(1_T)_j = |E(S,T)|$.
Clearly then, we have $|E(S,T)| = \left(\sum_{i=0}^{n-1} \alpha_i v_i\right) A \left(\sum_{i=0}^{n-1} \beta_i v_i\right)$, which equals the following, by distributing A and using the fact that the v_j are orthogonal: $\sum_{i=0}^{n-1} \mu_i \alpha_i \beta_i$. Next, note that since α_0 is the length of the projection of 1_S onto v_0, we have $\alpha_0 =< 1_S, v_0 >= \frac{|S|}{\sqrt{n}}$. Similarly, $\beta_0 = \frac{|T|}{\sqrt{n}}$. (Note that v_0 is the constant vector with norm 1, or $(1/\sqrt{n}, ..., 1/\sqrt{n})$.) Therefore, we have:

$$|E(S,T)| = \mu_0\alpha_0\beta_0 + \sum_{i=1}^{n-1} \mu_i\alpha_i\beta_i = \frac{d|S||T|}{n} + \sum_{i=1}^{n-1} \mu_i\alpha_i\beta_i$$

That implies:

$$\left||E(S,T)| - \frac{d|S||T|}{n}\right| = \left|\sum_{i=1}^{n-1} \mu_i\alpha_i\beta_i\right| \leq \sum_{i=1}^{n-1} |\mu_i\alpha_i\beta_i|$$

Since $\hat{\mu}(G)$ is defined to be the second largest eigenvalue, in absolute value, we have:

$$\left||E(S,T)| - \frac{d|S||T|}{n}\right| \leq \hat{\mu}(G) \sum_{i=1}^{n-1} |\alpha_i\beta_i|$$

Application of Cauchy-Schwartz (and realizing that adding in the first element in the α and β vectors only increases the norm) yields:

$$\left||E(S,T)| - \frac{d|S||T|}{n}\right| \leq \hat{\mu}(G)||\alpha||_2||\beta||_2 = \hat{\mu}(G)||1_S||_2||1_T||_2 = \hat{\mu}(G)\sqrt{|S||T|}$$

where the second to last equality follows from the n-dimensional Pythagorean theorem, which applies to the eigenvector decomposition of 1_S and 1_T since the v_j are orthogonal. □

We note that a kind of converse of this theorem holds, the exact statement of which and proof we omit. Simply put, knowing that $\left||E(S,T)| - \frac{d}{n}|S||T|\right| \leq \gamma\sqrt{|S||T|}$ allows us to put bounds on $\hat{\mu}(G)$ in terms of γ.

7.2 Expansion as pseudorandomness

Clearly, better expanders give a better bound on the difference between the number of edges between any two subsets of the vertices and the quantity $\frac{d|S||T|}{n}$. This quantity is significant because it represents the expected number of edges between S and T in a random graph. This is because the number of edges leaving a vertex set of size $|S|$ in a d-regular graph is $d|S|$, and if the edges lead to a vertex chosen uniformly at random, then each edge would lead to a node in T with probability $\frac{|T|}{n}$. Hence, the expected number of edges leading between S and T is $d|S|\frac{|T|}{n}$ (note that this protocol for creating a random graph could lead to multiple edges between the same nodes). Therefore, very good expanders will ensure that the number of edges between any two vertex subsets is close to this quantity.

7.3 Extractors via expander random walks

Before getting into the definition of extractors and how we can create them via random walks on expanders, we need some background information.

7.3.1 Min-entropy of random variables, statistical distance, and extractors

Definition. *(min-entropy)* We define the min-entropy of a random variable X as

$$\sup\{k \mid \forall x \in \text{range}(X), \Pr[X = x] \leq 2^{-k}\}$$

Notice that if X has no randomness (meaning there exists an x such that $\Pr[X = x]$=1), then the min-entropy is equal to 0. Also, if X has finite range with n values and is uniform, then the min-entropy is n.

Definition. *(statistical difference)* We define the statistical distance between two random variables with finite and identical range as: $\frac{1}{2}|X - Y|_1$, which equals $\frac{1}{2}\sum_{\omega \in \text{range}(X)} |\Pr[X = \omega] - \Pr(Y = \omega]|$.

Definition. *(extractor)* A (k, ϵ)-extractor is a function Ext:$\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ if for any random variable X with range $\{0,1\}^n$ and min-entropy at least k, then the statistical difference between the uniform distribution over $\{0,1\}^m$ (denoted U_m) and $\mathsf{Ext}(X, U_d)$ is less than ϵ.

7.3.2 The random walk

We now want to show how to use a random walk over an expander graph to obtain an extractor. We follow the description given in [2]. Fix $\epsilon > 0$ and d and for all $n, k \in \mathbb{Z}^+$ with $k \leq n$, let X be an arbitrary random variable over n bits with min-entropy at least k. Also, let G be a d-regular expander graph (from an infinite family of d-regular graphs) with 2^n vertices, where $\hat{\mu}(G) < \frac{d}{2}$. Let s be a sample from X and have z be a uniformly random string of length $t = (n/2 - k/2 + \log 1/\epsilon + 1)\log d = O(n - k + \log 1/\epsilon)$. Note that, by associating each of the vertices in G with a bit string label of length n, we can pick the vertex in G that is represented by the sample s and take a walk from there. This walk is defined by z in the following way: at each step we have to choose an edge from among d choices, a task that requires $\log d$ bits. We do this process $n/2 - k/2 + \log 1/\epsilon + 1$ times, using up all of the bits in z. Then, we output the label (a bit string of length n) of the final vertex visited in the walk.

Theorem 7.2. *The process described above yields a (k, ϵ)-extractor.*

Proof. Recall from Lemma 4.2.1 that:

$$\left|\left|\left(\frac{1}{d}A\right)^t p - \omega\right|\right|_2 \leq \left(\frac{\hat{\mu}}{d}\right)^t ||p'||_2 = \left(\frac{\hat{\mu}}{d}\right)^t ||p - \omega||_2$$

Now, since p represents the probability distribution of a random variable with min-entropy 2^{-k}, we know that $||p||_2^2 \leq 2^{-k}$. This fact is basically the result of noticing that the L_2 norm of a vector where the entries sum to 1 is maximized by putting all of the weight on as few entries as possible (think of how the L_1 norm is always less than or equal to the L_2 norm). So, we can assume that all the weight is on the first 2^k terms and that all entries have weight 2^{-k} (due to the min-entropy constraint): $||p||_2^2 \leq \sum_{i=1}^{2^k} (2^{-k})^2 = 2^{-k} \implies ||p||_2 = 2^{-k/2}$. Clearly, $||\omega||_2 = 2^{-n/2}$. This means that we have:

$$||p - \omega||_2 \leq ||p||_2 + ||\omega||_2 \leq 2^{-k/2} + 2^{-n/2}$$

Since $k \leq n$, we have that $2^{-k/2} + 2^{-n/2} \leq 2^{-k/2}(1 + 1) = 2^{-\frac{k}{2}+1}$ So, we have

$$\left|\left|\left(\frac{1}{d}A\right)^t p - \omega\right|\right|_2 \leq \left(\frac{\hat{\mu}}{d}\right)^t ||p - \omega||_2 \leq \left(\frac{1}{2}\right)^t 2^{-k/2+1} = 2^{-(n/2+\log 1/\epsilon)} = \epsilon 2^{-n/2}$$

The distribution of the output of the extractor is given by $\left(\frac{1}{d}A\right)^t p$. This means that the L_2 norm of the difference betweeh the extractor output and the uniform distribution is less than $\epsilon 2^{-n/2} = \frac{\epsilon}{\sqrt{2^n}}$. It is well-known that if a vector v is in $\mathbb{R}^j$, then $|v|_1 \leq \sqrt{j}||v||_2$. Since $\left(\frac{1}{d}A\right)^t p - \omega \in \mathbb{R}^{2^n}$, it immediately follows that

$$\left|\left(\frac{1}{d}A\right)^t p - \omega\right|_1 \leq \epsilon$$

This means that the statistical difference between $\left(\frac{1}{d}A\right)^t p$ and ω is less than ϵ, which means that the function described is the desired extractor. □

Very many other extractor constructions exist, as one can see in [26,29,31,32,33]. Also notice that extractors can be considered from the perspective as a bipartite graph. An extractor $\mathsf{Ext}{:}\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ takes two arguments and gives a single output, so we can associate left vertices with all 2^n potential first arguments and right vertices as all 2^m potential outputs, and connect vertices x and y if there exists a second argument, $a \in \{0,1\}^d$ such that $\mathsf{Ext}{:}(x,a) = y$. There are interesting connections between how well the extractor works and the expansion properties of the graph (see [29]). Finally, there exist functions, essentially weakened extractors, which are called dispersers. Just as extractors can be thought of as bipartite expanders, dispersers give rise to Ramsey graphs (graphs with no large clique or independent set). Confer [3] for more about this.

Extractors are closely related to pseudorandom generators, which in turn have consequences in complexity theory. For example, Nisan's pseudorandom generator for space-bounded computation relies on extractor functions, and was used to prove relationships between complexity classes. In this case, Nisan's pseudorandom generator, described in [2], shows that BPL $\subseteq$SPACE($\log^2 n$) (BPL is two-sided error, poly-time, and log-space). It was later shown that BPL $\subseteq$ SPACE($\log^{1.5} n$), leading to the conjecture that BPL =L. (Since L $\subseteq$ P and RL, co-RL $\subseteq$ BPL, where RL is one-sided error, this statement would mean that access to randomness does not fundamentally aid computations that can be completed in logspace.)

7.4 Other applications to complexity

7.4.1 Using expanders to collapse complexity classes

In the previous section, we mentioned the conjecture that BPL = L. Here, we will briefly mention the 2004 result of Omer Reingold that shows that SL = L. The proof followed immediately by showing that the **UPATH** problem was in L. Since L $\subseteq$ SL and **UPATH** is SL-complete, that suffices to show equality between those classes. We won't define the class SL, but just mention how expanders played a role in this result. The **UPATH** problem asks that given an undirected graph and two vertices, whether they are connected by some path. Clearly, one can just take a random walk starting at one vertex and see whether they reach the other vertex. However, the length of the path required could become a problem. This issue was handled by Reingold by breaking a graph into pieces that are expander graphs. Then, recall that we saw in Corollary 4.2.1 that expanders have a logarithmic diameter. This means that a logarithmic length random walk suffices to check connectivity, and this is crucial to showing that the calculation can be done in logspace (i.e. in L).

7.4.2 Randomness reduction in algorithms

Expander graphs can be used to reduce the amount of randomness needed by randomized algorithms. For example, take an algorithm in RP. Such an algorithm returns 0 when it should with probability 1 and returns 1 when it should at least half of the time (say with probability $1-p$). Thus, the false positive rate is 0 and the false negative rate is p. Clearly, if we run the algorithm once, and the output is 0, then the probability that the algorithm made a mistake (the correct answer is 1) is p. Of course, we can run the algorithm k times (this would still be running in poly time) to reduce the failure chance to p^k. But it seems like if the algorithm requires m random bits to run, then this operation would require mk bits. Expander graphs allow us to circumvent this difficulty. We associate every length m bit string with each vertex. Instead, one can choose a starting vertex uniformly at random (requiring $\log |V|$ bits) and then take a k step random walk in the graph.

We use the bit string at each vertex as the source of randomness to run the algorithm. To motivate this, first note that expanders are sparse but resemble the complete graph. The sparsity means that taking a random walk over the vertices requires few bits (if the graph is degree d, then each step requires a selection among d choices, which requires $\log d$ random bits). The fact that expanders are like the complete graph (they mix quickly) means that, although the vertices visited on the path are not completely random as they are in a complete graph, they are close enough to random. Again, see [2,16,33] for a more formal explanation of this idea, and the proof that this scheme actually works in that the error is small enough to be ignored, and we can still obtain exponentially small error with a number of random bits that is only linear in the number of runs.

8 Survey of constructions of families of graphs

A family of graphs is usually considered to be some sequence of graphs (with increasingly many vertices) that have a desired property. We will concern ourselves with how to create families of graphs that are expanders (or Ramanujan), where each family has fixed degree.

8.1 Expanders

Any connected and non-bipartite d-regular graph is a spectral expander in that its second largest adjacency matrix eigenvalue (in absolute value) is less than d. Therefore, any sequence of such graphs makes a trivial expander family. Therefore, we want to construct families such that the eigenvalues are all bounded away from d (meaning less than some fixed number that is less than d).

8.1.1 Margulis-Gabber-Galil expanders

First, we denote the group $\mathbb{Z}/n\mathbb{Z}$ as $\mathbb{Z}_n$. Then, for each n, we have a graph G_n with a vertex set of size n^2 that we associate with $\mathbb{Z}_n \times \mathbb{Z}_n$. Thus, each vertex is represented as a pair (a, b), where a and b are in $\{0, ..., n-1\}$. The vertex (a, b) is then connected to: $(a+1, b), (a-1, b), (a, b+1), (a, b-1), S(a, b), S^{-1}(a, b), T(a, b)$, and $T^{-1}(a, b)$, where we define $S(a, b)$ as $(a, a+b)$ and $T(a, b)$ as $(a+b, b)$. Addition in this group is coordinate-wise addition modulo n. Then, by the arguments presented in [30], the graph family $\{G_n\}_{n\in\mathbb{Z}^+}$ is a non-trivial infinite family of expanders. (Note that these graphs are sparse in that they all have degree at most 8.)

8.1.2 Combinatoric constructions

We simply note here that another, and non-algebraic, technique that is used to construct expander families is to use various graph operations, some of which are the matrix product, tensor product, replacement product, and zig-zag product. Precise definitions of these operations, and proofs regarding how they can be used to transform small expanders into larger ones are given in [2].

8.2 The LPSM Ramanujan graphs

8.2.1 Cayley graphs

Definition. *(Cayley graph)* A Cayley graph is defined in terms of a group G and a subset S, where S is a generating set (meaning that $< S >= G$). The vertices of the Cayley graph are associated with the group elements, and for all $g \in G$, we have the following edges:

$$\{\{g, gs\} | s \in S\}$$

In the following construction, we will take $G = \mathrm{PGL}(2, q)$, where that group is the group of 2×2 nonsingular matrices with entries in $\mathbb{F}_q = \mathbb{Z}/q\mathbb{Z}$.

8.2.2 The Lubotzsky-Phillips-Sarnak construction

The construction created in [18] (and described in [23]) has us choosing primes p and q, where $p \neq q$ and both p and q are congruent to 1 mod 4. Pick $u \in \mathbb{Z}$ such that $u \equiv -1 (\text{mod } q)$. By a Jacobi formula, there are exactly $p+1$ integer solutions to the equation $a^2+b^2+c^2+d^2 = p$ where a is odd and greater than zero and the other integers are even. Associate every such solution with a matrix defined as $\begin{bmatrix} a+ub & c+ud \\ -c+ud & a-ub \end{bmatrix}$ Now, take S to be the $p+1$ matrices of this form. This, along with the group G allows us to define a Cayley graph, and it was shown that all such graphs $\text{CAY}(G,S)$ are Ramanujan. Note that these graphs are $p+1$-regular since $|S| = p+1$, and the number of vertices is $q = |\mathbb{F}_q|$. Fixing p and letting q get large yields infinite families of Ramanujan graphs. Margulis [22] generalized this result by giving constructions of Ramanujan graphs for degrees that are prime powers plus one, not just primes plus one. Other Cayley graph expanders are constructed in [19,27,33].

9 Adknowledgements

The author is pleased to thank Ralph Greenberg (UW), Franklin Kenter (Rice), James Lee (UW), James Morrow (UW), Mary Radcliffe (UW), and Anup Rao (UW) for their time, advice, and aid in understanding these fascinating mathematical concepts.

10 References

[1] Apkarian, Nanneh. Lecture. CS262. UCSD, San Diego, CA.
`http://www.math.ucsd.edu/~c1woods/262notes/Apkarian_Midterm.pdf`

[2] Arora, Sanjeev, and Boaz Barak. Computational complexity: a modern approach. Cambridge University Press, 2009.

[3] Barak, Boaz, et al. "2-source dispersers for sub-polynomial entropy and Ramsey graphs beating the Frankl-Wilson construction." Proceedings of the thirty-eighth annual ACM symposium on Theory of computing. ACM, 2006.

[4] Bilu, Yonatan, and Nathan Linial. "Lifts, discrepancy and nearly optimal spectral gap*." *Combinatorica* 26.5 (2006): 495-519.
`http://www.cs.huji.ac.il/~nati/PAPERS/raman_lift.pdf`

[5] Brändén, Petter. "Lecture Notes For Interlacing Families." Royal Institute of Technology, Stockholm, Sweden, 2013.
`http://www.math.kth.se/~pbranden/notes.pdf`

[6] Brooks, R. L. (1941), "On colouring the nodes of a network", Proc. Cambridge Philosophical Society, Math. Phys. Sci. 37: 194197, doi:10.1017/S030500410002168X.

[7] Brouwer, Andries E., and Willem H. Haemers. Spectra of graphs. Springer, 2011.
`http://www.win.tue.nl/~aeb/2WF02/spectra.pdf`

[8] Chudnovsky, Maria, and Paul Seymour. "The roots of the independence polynomial of a clawfree graph." Journal of Combinatorial Theory, Series B 97.3 (2007): 350-357.

[9] Chung, Fan RK. "Laplacians of graphs and Cheegers inequalities." Proc. Int. Conf. *Combinatorics, Paul Erdős is Eighty*, Keszthely (Hungary). Vol. 2. 1993.
`http://www.math.ucsd.edu/~fan/wp/cheeger.pdf`

[10] Csikvári, Péter. Graph Polynomials and Graph Transformations in Algebraic Graph Theory. Diss. PhD thesis. Eötvös Loránd University, 2011.:
http://teo.elte.hu/minosites/ertekezes2011/csikvari_p.pdf,2011

[11] Fisk, Steve. "Polynomials, roots, and interlacing." arXiv preprint math/0612833 (2006).

[12] Gharan, Shayan Oveis, and Luca Trevisan. "A Universal upper bound on Graph Diameter based on Laplacian Eigenvalues." arXiv preprint arXiv:1212.2701 (2012).
http://arxiv.org/pdf/1212.2701v1.pdf

[13] Godsil, Chris D., and Ivan Gutman. "On the theory of the matching polynomial." Journal of Graph Theory 5.2 (1981): 137-144.

[14] Haemers, Willem H. "Interlacing eigenvalues and graphs." Linear Algebra and its Applications 226 (1995): 593-616.
http://members.upc.nl/w.haemers/interlacing.pdf

[15] Heilmann, O.J. and Lieb, E.H. Theory of monomer-dimer systems. Comm. Math. Phys. zs (1972), pp. 190-232
http://projecteuclid.org/download/pdf_1/euclid.cmp/1103857921

[16] Hoory, Shlomo, Nathan Linial, and Avi Wigderson. 2006. "Expander Graphs and Their Applications." Bulletin of the American Mathematical Society, 43(4): 439-561.
http://www.cs.huji.ac.il/~nati/PAPERS/expander_survey.pdf

[17] László, Lovász. Lecture. Eötvös Loránd University, Budapest.
http://www.cs.elte.hu/~lovasz/eigenvals-x.pdf

[18] Lubotzky, Alexander, Ralph Phillips, and Peter Sarnak. "Ramanujan graphs." Combinatorica 8.3 (1988): 261-277.

[19] Lubotzky, Alexander. "Expander graphs in pure and applied mathematics." Bulletin of the American Mathematical Society 49.1 (2012): 113-162.
http://arxiv.org/pdf/1105.2389v1.pdf

[20] Marcus, Adam, Daniel A. Spielman, and Nikhil Srivastava. "Interlacing families I: Bipartite ramanujan graphs of all degrees." *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on.* IEEE, 2013.
http://arxiv.org/pdf/1304.4132v1.pdf

[21] Mitchell, Kim. *Bounds on the chromatic number.* Thesis, University of Colorado at Denver. 2007.
http://www.ucdenver.edu/academics/colleges/CLAS/Departments/math/students/alumni/Documents/Student%20Theses/Mitchell_MSThesis.pdf

[22] Morgenstern, Moshe. "Existence and Explicit Constructions of $q + 1$ Regular Ramanujan Graphs for Every Prime Power q." Journal of Combinatorial Theory, Series B 62.1 (1994): 44-62.

[23] Murty, M. Ram. "Ramanujan graphs." Journal-Ramanujan Mathematical Society 18.1 (2003): 33-52.
http://www.mast.queensu.ca/~murty/ramanujan.pdf

[24] Nikiforov, Vladimir. "Chromatic number and spectral radius." *Linear Algebra and its Applications* 426: 810-814. Print.
http://www.sciencedirect.com/science/article/pii/S0024379507002704

[25] Nilli, Alon. "On the second eigenvalue of a graph." Discrete Mathematics 91.2 (1991): 207-210.
http://www.sciencedirect.com/science/article/pii/0012365X9190112F

[26] Rao, Anup. An exposition of Bourgain's 2-source extractor. Electronic Colloquium on Computational Complexity (ECCC) 14(034) (2007).
http://homes.cs.washington.edu/~anuprao/pubs/bourgain.pdf

[27] Rozenman, E., Shalev, A., Wigderson, A.: Iterative construction of Cayley expander graphs. Theory Comput. 2, 91120 (2006)
http://www.math.ias.edu/~avi/PUBLICATIONS/MYPAPERS/RSW05/rsw05.pdf

[28] Spielman, Daniel. "Spectral graph theory." Lecture Notes, Yale University (2009): 740-0776.

[29] Sun, He. Lecture. Max Planck Institut Informatik, Saarbrucken, Germany.
http://www.mpi-inf.mpg.de/departments/d1/teaching/ws10/EG/notes9.pdf

[30] Trevisan, Luca. Lecture. CS366. Stanford University, Stanford, CA. 4 March 2013.
http://cs.stanford.edu/people/trevisan/cs366/lecture13.pdf

[31] Trevisan, Luca. Extractors and pseudorandom generators. J. ACM, 48 (4) (2001), pp. 860879
http://theory.stanford.edu/~trevisan/pubs/extractor-full.pdf

[32] Trevisan, Luca. Lecture. CS254-12. Stanford University, Stanford, CA.
http://cs.stanford.edu/people/trevisan/cs254-12/prg.pdf

[33] Vadhan, Salil. "Randomness Extractors." Pseudorandomness. : monograph in preparation for FnTTCS, 2013. Print.
http://people.seas.harvard.edu/~salil/pseudorandomness/extractors.pdf