AN INTRODUCTION TO EXPANDERS II

JAEHOON KIM

Abstract. In this note, we learn about expander graphs. The materials on this note are based on the following sources: [3, 4, 8, 9]

1. Recapitulation

Last year, we learned about expander graphs. In short, expanders are d-regular graphs which satisfies the following loosely equivalent properties.

(1) Combinatorial property: Any vertex set U has comparatively large boundary (either the edge boundary $|E(U,\overline{U})|$ or the vertex boundary $N_G(U)$).
(2) Probabilistic property: Random walk mixes fast, i.e. k-th vertex on the random walk on the graph is almost as if we choose a vertex at random.
(3) Algebraic property: All eigenvalues of the adjacency matrix except the largest one are small.

Definition 1.1. *For $S \subseteq V(G)$, the edge boundary $\partial S = E(S,\overline{S})$ is the set of edges from S to $\overline{S} = V \setminus S$. The edge expansion ratio of G, denoted by $h(G)$, is defined as*

$$h(G) = \min_{S,|S|\leq n/2} \frac{|\partial S|}{|S|}.$$

Definition 1.2. *The adjacency matrix of a graph G, denoted by $A = A(G)$ is an $n \times n$ matrix whose (u,v)-entry is the number of edges in G between vertex u and vertex v. As it is real symmetric matrix, it has n real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. If a d-regular graph G has one eigenvalue d and rest of the eigenvalues in $[-\alpha d, \alpha d]$, it is called a (n,d,α)-graph, or an α-expander.*

We prove that $\frac{d-\lambda_2}{2} \leq h(G) \leq \sqrt{2d(d-\lambda_2)}$ holds, showing the combinatorial property and algebraic property are related.

Also, we consider the random walks. It starts at a vertex v chosen according to the initial probability distribution $\mathbf{p}_0 \in \mathbb{R}^n$ over the vertex set $[n] = V(G)$ of G, and it moves to a neighbor chosen uniformly at random. This yields a probability distribution of the t-th vertex on the walk as $(\frac{1}{d}A)^t\mathbf{p}_0$. We showed that

$$\|\mathbf{p}_t - \mathbf{1}\|_1 = \|(\frac{1}{d}A)^t\mathbf{p} - \mathbf{1}\|_1 \leq \alpha^t\sqrt{n}$$

holds for a random walk over an (n,d,α)-graph where $\mathbf{1} = (1,1,\ldots,1)^T$. This shows that the second aspect is also related with others.

We also showed that such an expander is useful for constructions of good error correcting codes, and error reduction on randomized algorithms. We also see some ways to construct expander graphs.

This year, we will introduce the (normalized) Laplacian matrix of graphs which enables us to deal with non-regular graphs. Using this concept, we can extend the definition of expanders to more general notion of graph approximation. We will also learn more ways to construct almost best possible expanders using some concept called eigenvalue interlacing.

Date: August 5, 2022.

2. Laplacian

While learning about the adjacency matrix of graphs and its spectral properties, two questions naturally arise.

- Is the adjacency matrix of graph the most natural notion?
- Is the assumption of graph being regular important?

The theories we introduced in the last year's lecture were built on the assumption that our graph is regular. So, the second question, whether we can also deal with non-regular graphs, is very natural. However, if we try to deal with non-regular graphs, the analysis we did before does not work. For example, once we consider the random walks over non-regular graphs, then the adjacency matrix A_G no longer captures how the random walk behaves. To overcome this (and other) issues, we introduce the following notion of graph Laplacians. Also, we will consider more general notion of edge-weighted graphs where each edge ij of the graph G has a positive edge weight $w_{i,j} \in \mathbb{R}_+$ on each edge ij.

Definition 2.1. *Let $D = D_G$ be an n by n diagonal matrix where $D_{i,i} = d_G(i)$ is the degree of the vertex i, and let $A = A_G$ be the adjacency graph of a weighted graph G, where $A_{i,j} = w_{i,j}$ where $w_{i,j}$ is the weight on the edge ij in G. The matrix $L = L_G = D - A_G$ is called* the Laplacian matrix *of G, and the matrix $\widetilde{L} = \widetilde{L}_G = D^{-1/2}LD^{-1/2}$ is called* the normalized Laplacian matrix *of G.*

There are several reasons why these matrices are useful and important. In particular, we can list the following two reasons.

- This is an analogue of the Laplacian over continuous spaces, measuring 'smoothness of functions' defined over graphs/spaces.
- Together with the normalized Laplacian, many theories generalize to non-regular graphs.

Let's convince ourselves that Laplacian is a very natural concept.

Why do we want to measure 'smoothness' of functions defined over graphs, and how do we define this 'smoothness'? Let's consider one example.

In recent years, the need for analyzing high dimensional data sets have increased a lot. Here, high dimensional data are defined as data in which the number of features (variables observed), are close to or larger than the number of observations (or data points). For example, when we have patients in hospital, there are a lot of data we need from a patient, including blood pressure, heart rate, BMI, Cholesterol level, Body fat, respiratory rate, red cell mass, etc. If the number of features is small, it is relatively easier to analyze the data. However, if the number of features is large, it becomes difficult to analyze this data set.

On the other hand, in many cases, these features are not completely independent of each other. For example, blood pressure level is somehow predictable from cholesterol level in blood, age, and BMI measurements. So, there are some relations between these features, which allow us to analyze the data more efficiently. These relations between the features can be captured using (edge-weighted) graph having the features as vertices, and edge representing some relations between the features. High weights represents strong relations while small weight represents weak relations. In this aspect, assuming the features have real values, each data correspond to a function $f \in \mathbb{R}^V$.

Let's assume that we want to do image-filtering. We have an original image of, say, n pixels, but some noise has been added to the image. How can we filter this image to obtain a 'better-looking' image? Of course, 'better-looking' is not well-defined term. If you think about it, the noise introduces some unnatural "sharpness" to the image, which makes us to notice the noise. So, if we make the picture somehow "smooth", then it

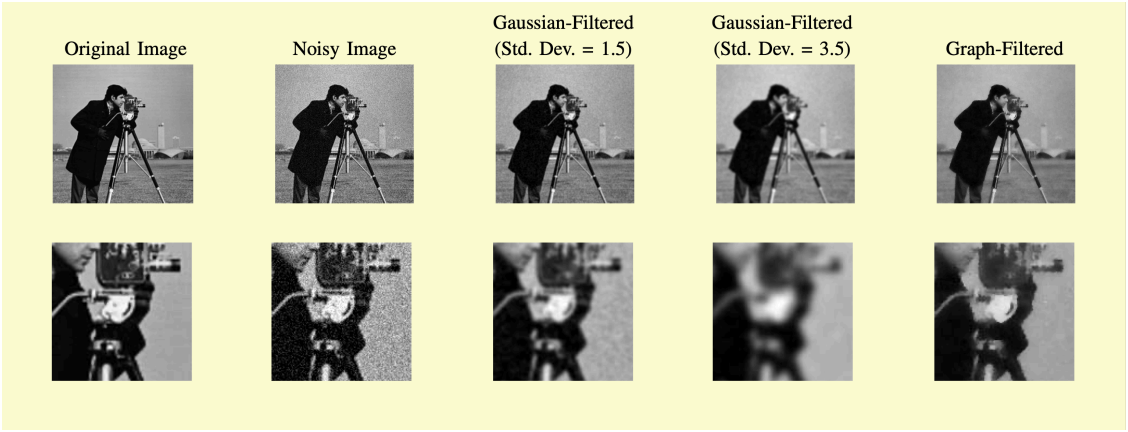

FIGURE 1. An example by Shuman, Narang, Frossard, Ortega and Vandergheynst[8]

becomes better. This is what image filtering is for. So, a most basic way to filter an image is to replace a pixel by the average of its neighborhood. For example, if we have a pixel of value 2 surrounded by eight pixels of value 1, then we replace the value 2 by the average $(2 + 8 * 1)/9 = 10/9$. This will results in a bit blurred picture, but the effects of noises are somewhat smoothed out. More generally, we can improve the quality of the image by considering some weighted average putting more weights on closer pixel according to the normal distribution. This is called the Gaussian filter.

However, this approach does not take into account on the more information we have. For example, a pixel can come from the face of a person, while a pixel right next to it comes from the background sky. We do not want to average out these two pixels, which introduces too much undesired blurr. How one can more efficiently deal with this? We can consider a graph over the vertex set V of pixels. Each pixel has a real value, which represents the color. (Usually a color requires three real numbers to represent, but assume for simplicity that the color can be represented by one number). Then the original image $\mathbf{x} \in \mathbb{R}^V$ is a function over the vertex set, and image with noise $\mathbf{y} \in \mathbb{R}^V$ is also a function where $\mathbf{y} - \mathbf{x}$ is the noise added. Based on $\mathbf{y}$, one can consider a weighted graph G. We put edges if two pixels i, j are close. More precisely, put an edge between a pixel i and eight other pixels surrounding i. Also, if adjacent vertices have similar values, we put high weights $w_{i,j}$ on the edge between them. If they have different values, then we put low weight $w_{i,j}$ on the edge between them. (There are several ways to determine the weight $w_{i,j}$ according to the value difference, but we will not specify them.) Then, depending on how to determine $w_{i,j}$,

$$\sum_{ij \in G} w_{i,j}(\mathbf{y}(i) - \mathbf{y}(j))^2 = \mathbf{y}^T L \mathbf{y}$$

somehow measures how 'smooth' the given image (function) $\mathbf{y}$ is. As the original image is very likely to be more smooth than the noised image, we seek to find a function $\mathbf{z} \in \mathbb{R}^V$ where $\mathbf{z}^T L\mathbf{z}$ is smaller or at least not too large. Of course, we want the $\mathbf{z}$ to be not too far from $\mathbf{y}$, so one can use some way called Tikhonov regularization to find a function $\mathbf{z}$ satisfying

$$\text{argmin}_{\mathbf{z}}\{\|\mathbf{z}-\mathbf{y}\|_2^2 + \gamma \mathbf{z}^T L\mathbf{z}\},$$

where γ is a chosen constant. See the above Figure 1 from [8]. This shows that the graph-based filtering produces a sharper image than traditional Gaussian filter method.

Moreover, if we solve this optimization problem, then the optimal solution is given by

$$\sum_{\ell=0}^{N-1} \left[\frac{1}{1+\gamma\lambda_\ell}\right] \langle \mathbf{y}, \mathbf{u}_\ell\rangle \mathbf{u}_\ell$$

where λ_ℓ is the ℓ-th eigenvalue of the Laplacian L and $\mathbf{u}_\ell$ is the corresponding eigenvector.

Also, Laplacian of a graph is an analogue of the Laplacian over a smooth space $\mathbb{R}^n$. It is roughly a generalization of the second derivatives of a function. In many cases, one can consider a graph embedded on a given space and consider this graph as an discretized approximation of the space. So, if we have an analogue of the Laplacian in graphs, then one can analyze the graphs instead of the space to learn more about the space.

As the Laplacian is the divergence of the gradient of the function, let's figure out what those are in graphs. For given a function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient is

$$\nabla f = (\partial_1 f, \partial_2 f, \dots, \partial_n f).$$

A concept similar to the partial derivative ∂_1 at i is the difference $\mathbf{f}(i) - \mathbf{f}(j)$ along the edge ij divided by some 'distance' $d_{i,j}$ between two vertices i and j. So, one can consider the following incidence matrix K. The columns of K correspond to the vertices in order (we assume that the vertices in $V(G) = [n]$ are ordered) and the rows of K corresponded to edges of G with order (this order can be arbitrary). For each edge $e_i = kh$ with $k < h$, the entry $K_{i,j}$ is defined as

$$K_{i,j} := \begin{cases} 1/d_{k,h} & \text{if } j = k, \\ -1/d_{k,h} & \text{if } j = h, \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Then $K\mathbf{f}$ is an analogue of ∇f.

Also, recall that the divergence on a multivariate differentiable vector field $\mathbf{F}$ measures the amount of 'flow' coming out/into a point when the flow is determined by the vector field. When $\mathbf{F} = (\mathbf{F}_1, \dots, \mathbf{F}_n)$, we have

$$\text{div}\mathbf{F} = \partial_1 \mathbf{F}_1 + \cdots + \partial_n \mathbf{F}_n$$

The vector field should correspond to a function defined over edges (which is same as 'direction'). As the values flows linearly over edges, the $\partial_e \mathbf{F}_e$ for an edge e should be the value of e divided by the length of e (which measures how fast the flows are coming into the vertex). Then divergence at a vertex is simply the summation of the values on the edges divided by $d_{i,j}$, i.e. $\text{div}\mathbf{g}(i) = \sum_{j\in N(i)} \mathbf{g}(ji)/d_{i,j}$, where $\mathbf{g}(ji) = -\mathbf{g}(ij)$ if the edge ij appears in the function $\mathbf{g}$ as the order (j, i). Indeed, this is precisely the matrix $K^T\mathbf{g}$. Hence, we can see that it is very natural to define the Laplacian as

$$K^T K = D - A_G,$$

where G has the weights $w_{i,j} = \frac{1}{d_{i,j}^2}$ for each edge ij. Hence, Laplacian matrix is a graph analogue of the Laplacian of continuous functions. Again, this is measuring how smooth a function defined over the graph is.

Lastly, let's see how normalized Laplacian helps us to deal with non-regular graphs in some situations. We will consider the random walks. Recall the definition of the random walk. For a given (not necessarily regular) edge-weighted graph G, we begin at some vertex. In each time step, we move to another vertex. When it moves, it moves to a neighbor with probability proportional to its weight, i.e. it moves from i to j with probability $\frac{w_{i,j}}{d(i)}$ where $d(i) = \sum_{j \in N(i)} w_{i,j}$ is the weighted degree of the vertex. (If the graph is not weighted, we treat each edge as an edge with weight 1 and moves from i to j with probability $\frac{w_{i,j}}{2d(i)}$.)

Let $\mathbf{p}_t \in \mathbf{R}^n$ be the probability distribution at time t where $\mathbf{p}_t(i)$ is the probability of being at the vertex i at time t. As this is a probability distribution, we have $\sum_i \mathbf{p}_t(i) = 1$. If the walk starts at i, then we consider $\mathbf{p}_0 = \mathbf{1}_{\{i\}}$ being the characteristic vector where all probability is concentrated on the vertex i. The probability of being at a vertex i at time $t+1$ is the sum over the neighbors j of i of the probability that the walk was at j at time t times the probability that it moves from j to i in time $t+1$. Hence

$$\mathbf{p}_{t+1}(i) = \sum_{ij \in E(G)} \frac{w(i,j)}{d(j)} \mathbf{p}_t(j). \tag{2.2}$$

Here $d(j)$ is the weighted degree of j. This yields the walk matrix $W := A_G D^{-1}$. Let $\widetilde{W} = \frac{1}{2}I + \frac{1}{2}W$ be the walk matrix for the lazy walk. (For lazy walk, we have an additional assumption that we stay at the vertex with probability $1/2$ in each step.)

It is natural to analyze the random walks by studying the spectral aspects of W or $\widetilde{W}$. However, one problem is that this matrix is not symmetric, so we can't apply spectral theorem to obtain real eigenvalues and orthonormal eigenvectors. However, we can still show that this matrix has n real eigenvalues and real eigenvectors, while their eigenvectors might not be mutually orthogonal. For this, we consider the following the normalized adjacency matrix instead. This is a symmetric matrix, so it has real eigenvalues and orthonormal eigenvectors. Let

$$\widetilde{A} = D^{-1/2} W D^{1/2} = D^{-1/2} A D^{-1/2}.$$

In random graphs, what we are interested in is the stationary distrubition. At the stationary distribution $\widetilde{A}$ acts as if it is the identity I. So, considering the following matrix makes sense.

$$I - \widetilde{A} = I - D^{-1/2} A D^{-1/2} = D^{-1/2}(D - A)D^{-1/2} = \widetilde{L}_G.$$

Moreover, $I - \widetilde{A}$ is a positive semidefinite matrix while $\widetilde{A}$ is not. So analysis are more convenient in many cases.

Up to multiplying $D^{-1/2}$ on both sides, the normalized Laplacian captures the essence of the behaviour of random walks.

Proposition 2.2. *The vector* $\mathbf{u}$ *is an eigenvector of* $\widetilde{L}$ *of the eigenvalue* λ *if and only if* $D^{1/2}\mathbf{u}$ *is an eigenvector of* W *of the eigenvalue* $1 - \lambda$.

Proof. Note that $D^{1/2}\widetilde{L} = (D - A_G)D^{-1/2} = (I - W)D^{1/2}$, so $D^{1/2}\widetilde{L}\mathbf{u} = (I - W)D^{1/2}\mathbf{u}$. Hence, $\widetilde{L}\mathbf{u} = \lambda\mathbf{u}$ if and only if

$$D^{1/2}\widetilde{L}\mathbf{u} = D^{1/2}\lambda\mathbf{u} = \lambda(D^{1/2}\mathbf{u}) = (I - W)D^{1/2}\mathbf{u}.$$

□

Also, it is easy to see that $\widetilde{W}$ has the same eigenvectors with W. Moreover, assuming G being connected the degree vector $\mathbf{d}$ is the Perron vector (the eigenvector for the largest eigenvalue whose all entries are positive) of W of eigenvalue 1 as

$$AD^{-1}\mathbf{d} = A\mathbf{1} = \mathbf{d}.$$

So, the Peron-Frobenius theorem yields that the eigenvalues of W lie between -1 and 1, and the eigenvalues of $\widetilde{L}$ lies between 0 and 2. It is also not difficult to prove that the graph is bipartite if and only if 2 is an eigenvalue of $\widetilde{L}$.

For lazy random walks, $\widetilde{W} = \frac{1}{2}I + \frac{1}{2}W$ has eigenvalues

$$1 = \omega_1 \geq \cdots \geq \omega_n \geq 0.$$

By the above proposition, we know that $\widetilde{A}$ has the Perron vector $\mathbf{d}^{1/2}$ where $\mathbf{d}^{1/2}(i) = \mathbf{d}(i)^{1/2}$.

One reason why we consider a lazy random walk is that it has an advantage over random walk. The lazy walk on a connected graph always converges to one distribution: *the stationary distribution*. For non-lazy one, the distribution might not converge on bipartite graphs. We already see that the following is an eigenvector for $W = AD^{-1}$.

$$\pi := \frac{1}{\mathbf{1}^T\mathbf{d}}\mathbf{d}.$$

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the orthonormal eigenvectors of $\widetilde{L}$ corresponding to eigenvalues $0 = \lambda_1 = 2 - 2w_1 \leq \cdots \leq \lambda_n = 2 - 2w_n$. If we just consider the eigenvectors of $\widetilde{W}$, they are not orthogonal each other, so we consider the eigenvectors of $\widetilde{L}$. Then $\widetilde{W}$ has eigenvalues $\omega_1, \ldots, \omega_n$. For an initial distribution $\mathbf{p}_0$, we can find constants c_i such that

$$D^{-1/2}\mathbf{p}_0 = \sum_i c_i\mathbf{v}_i.$$

Here, $c_i = \mathbf{v}_i^T D^{-1/2}\mathbf{p}_0$. Especially for $i = 1$, we have

$$c_1 = \mathbf{v}_1^T D^{-1/2}\mathbf{p}_0 = \frac{(\mathbf{d}^{1/2})^T}{\|\mathbf{d}^{1/2}\|}(D^{-1/2}\mathbf{p}_0) = \frac{\mathbf{1}^T\mathbf{p}_0}{\|\mathbf{d}^{1/2}\|} = \frac{1}{\|\mathbf{d}^{1/2}\|}.$$

The last equality holds as $\mathbf{p}_0$ is a probability distribution. Then

$$\begin{aligned}
\mathbf{p}_t = \widetilde{W}^t\mathbf{p}_0 &= D^{1/2}D^{-1/2}\widetilde{W}^t D^{1/2}D^{-1/2}\mathbf{p}_0 \\
&= D^{1/2}(D^{-1/2}\widetilde{W}D^{1/2})^t D^{-1/2}\mathbf{p}_0 \\
&= D^{1/2}(\frac{1}{2}I + \frac{1}{2}\widetilde{A})^t \sum_i c_i\mathbf{v}_i \\
&= D^{1/2}\sum_i \omega_i^t c_i\mathbf{v}_i \\
&= D^{1/2}c_1\mathbf{v}_1 + D^{1/2}\sum_{i\geq 2} w_i^t c_i\mathbf{v}_i.
\end{aligned}$$

We have the final equality as we have $\omega_1 = 1$. As $0 \leq \omega_i < 1$ for all $i \geq 2$, the right-hand term go to zero as t goes to infinity. Here, as $D^{1/2}\mathbf{v}_1$ is a scalar multiple of the Perron vector $\mathbf{d}$ of W, $\mathbf{v}_1 = \frac{\mathbf{d}^{1/2}}{\|\mathbf{d}^{1/2}\|}$, so

$$D^{1/2}c_1\mathbf{v}_1 = D^{1/2}(\frac{1}{\|\mathbf{d}^{1/2}\|})\frac{\mathbf{d}^{1/2}}{\|\mathbf{d}^{1/2}\|} = \frac{\mathbf{d}}{\|\mathbf{d}^{1/2}\|^2} = \frac{\mathbf{d}}{\sum_i \mathbf{d}(i)} = \pi.$$

From this, we can suspect that how fast the convergence happens depends on $\omega_2 = 1 - \lambda_2/2$. So, as λ_2 is away from 0, the convergence get faster. There are several possible ways to measure the speed of convergence. We will consider pointwise distance as follows.

Theorem 2.3. *For all i, j and t, if $\mathbf{p}_0 = \delta_i = \mathbf{1}_{\{i\}}$, then*

$$|\mathbf{p}_t(j) - \pi(j)| \leq \sqrt{\frac{d(j)}{d(i)}}\omega_2^t.$$

Proof. Note that $\mathbf{p}_t(j) = \delta_j^T \mathbf{p}_t$. Then as above, we know

$$\mathbf{p}_t(j) = \delta_j^T \mathbf{p}_t = \pi(j) + \delta_j^T D^{1/2} \sum_{\ell \geq 2} \omega_\ell^t c_\ell \mathbf{v}_\ell.$$

So, we want the upper bound on the size of $\delta_j^T D^{1/2} \sum_{\ell \geq 2} \omega_\ell^t c_\ell \mathbf{v}_\ell$. Recall that

$$c_i = \mathbf{v}_i^T D^{1/2} \delta_i = \frac{1}{\sqrt{\mathbf{d}(i)}} \mathbf{v}_i^T \delta_i.$$

So, $\delta_j^T D^{1/2} \sum_{\ell \geq 2} \omega_\ell^t c_\ell \mathbf{v}_\ell = \sqrt{\frac{\mathbf{d}(j)}{\mathbf{d}(i)}} \delta_j^T \sum_{\ell \geq 2} w_\ell^t \mathbf{v}_\ell (\mathbf{v}_\ell^T \delta_\ell)$. Then we have

$$\begin{aligned} |\delta_j^T \sum_{\ell \geq 2} \omega_\ell^t \mathbf{v}_\ell \mathbf{v}_i^T \delta_i| &= |\sum_{\ell \geq 2} \omega_\ell^t (\delta_j^T \mathbf{v}_\ell)(\mathbf{v}_i^T \delta_i)| \leq \omega_2^t \sum_{\ell \geq 1} |\delta_j^T \mathbf{v}_\ell||\mathbf{v}_i^T \delta_i| \\ &\leq \omega_2^t \sqrt{\sum_{i \geq 1} (\delta_j^T \mathbf{v}_i)^2} \sqrt{\sum_{i \geq 1} (\delta_i^T \mathbf{v}_i)^2} \\ &= \omega_2^T \|\delta_j\| \|\delta_i\| = \omega_2^t. \end{aligned}$$

□

3. Spectral sparsifiers

As we saw before, the Laplacian captures many properties of the graph G. In the example of image processing, the computation of the quadratic form $\mathbf{x}^T L_G \mathbf{x}$ was important. Also, the following theorem shows that this quadratic form captures the essence of the eigenvalues of the Laplacian matrix. Here, the expression $\frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$ is called the *Rayleigh quotient* of $\mathbf{x}$ with respect to M.

Theorem 3.1 (Courant-Fischer Theorem)**.** *Let M be a symmetric matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. Then*

$$\lambda_k = \max_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S) = n-k+1}} \min_{x \in S - \{0\}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S) = k}} \max_{x \in S - \{0\}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}},$$

where the maximization and the minimization are over subspaces S and T of $\mathbb{R}^n$.

Similarly, in many cases, computing the quadratic form $\mathbf{x}^T L_G \mathbf{x}$ is useful and it requires $\Theta(n^2)$ multiplications of numbers in general.

However, if L_G has $O(n)$ many non-zero entries, then we can do this computation more efficient way in $O(n)$ multiplications. This motivates to find a graph H with small number of edges, where L_H can 'approximates' L_G for a given graph G. Of course, finding such a graph H can take quite a time, but if we have to compute $\mathbf{x}^T L_G \mathbf{x}$ many times for different vectors $\mathbf{x}$, finding one such a graph H and compute $\mathbf{x}^T L_H \mathbf{x}$ many times can be much more efficient.

For a d-regular graph G, when the adjacency matrix of G has eigenvalues $\mu_1 \geq \cdots \geq \mu_n$, the Laplacian eigenvalues are exactly $\lambda_i = d - \mu_i$. So being an expander is equivalent to $|d - \lambda_i| \leq \varepsilon d$ for $i \geq 2$. As L_{K_n} has eigenvalues $0, n, n, \ldots, n$ this is also equivalent to

$$\|L_G - \frac{d}{n} L_{K_n}\| \leq \varepsilon d$$

where we write $\|M\| = \max_{\mathbf{x}} \frac{\|M\mathbf{x}\|}{\|\mathbf{x}\|}$ for the operator norm. This is just the largest eigenvalue of the matrix if M is a symmetric matrix.

Definition 3.2. *We say that G is an ε-approximation of a graph H if*

$$(1-\varepsilon)H \preccurlyeq G \preccurlyeq (1+\varepsilon)H,$$

where $H \preccurlyeq G$ means $\mathbf{x}^T L_H \mathbf{x} \leq \mathbf{x}^T L_G \mathbf{x}$ for all x.

We can check the following proposition.

Proposition 3.3. *A d-regular graph G is an ε-expander if and only if it is an ε-approximation of $\frac{d}{n} K_n$.*

Proof. If G is an ε-expander, then for all $x \in \mathbb{R}^n$ that are orthogonal to $\mathbf{1} = (1, 1, \ldots, 1)^T$, we have

$$(1-\varepsilon) d \mathbf{x}^T \mathbf{x} \leq \mathbf{x}^T L_G \mathbf{x} \leq (1+\varepsilon) d \mathbf{x}^T \mathbf{x}.$$

However, as we have $\mathbf{x}^T L_{\frac{d}{n} K_n} \mathbf{x}^T = d \mathbf{x}^T \mathbf{x}$, so G is an ε-approximation of $\frac{d}{n} K_n$.

The other direction also follows as

$$(1-\varepsilon) L_H \preccurlyeq L_G \preccurlyeq (1+\varepsilon) L_H$$

implies

$$-\varepsilon L_H \preccurlyeq L_G - L_H \preccurlyeq \varepsilon L_H.$$

With $H = \frac{d}{n} K_n$, L_H has all eigenvalues d or 0, so we have $\|L_G - L_H\| \leq \varepsilon d$, hence G is an ε-expander. □

So, when the graph being regular is not so important, one can consider a (not necessarily regular) graph which is an approximation of a complete graph instead of a regular expander. This also suggests that we can consider sparse approximation of graphs other than the complete graph. For example, a regular bipartite graphs approximating a complete bipartite graphs are called bipartite expanders, and we saw last year that they are useful.

What about more general graphs? Can we find a sparse graph approximating any given graph G? If H is an ε-approximation of G, then they share many characteristics.

- the eigenvalues of the graphs are similar.
- the edge-boundaries $\sum_{ij\in E(S,\overline{S})} w_{i,j}$ of all sets are similar, as these are given by $\mathbf{1}_S L_G \mathbf{1}_S$.
- all solutions of linear equations in two matrices L_H, L_G are similar.

3.1. **Random subgraph.** The proof in this subsection is from [**?**]. One of the most natural approaches to construct a sparse approximation of G is that, for each edge ab, we independently at random choose the edge with some probability p_{ab}. If ab is included, then we give it weight $w_{a,b}/p_{a,b}$. By dividing by $p_{a,b}$, we preserves the matrix 'in expectation'. To see this, let $L_{a,b}$ be the Laplacian of the single edge ab, then we have $L_G = \sum_{ab\in E} w_{a,b} L_{a,b}$ while

$$\mathbb{E}[L_H] = \sum_{ab\in E} p_{a,b}(w_{a,b}/p_{a,b}) L_{a,b} = L_G.$$

We collect the following matrix Chernoff bound. For a given matrix X, we write $\lambda_k(X)$ to denote the k-th smallest eigenvalue, and $\lambda_{\min}$ and $\lambda_{\max}$ to denote the smallest and largest eigenvalues.

Theorem 3.4. *Let $X_1, \ldots, X_m$ be independent random n-dimensional symmetric positive semidefinitie matrices so that $\|X_i\| \leq R$ almost surely holds. Let $X = \sum_i X_i$ and let $\mu_{\min}$ and $\mu_{\max}$ be the minimum and maximum eigenvalues of $\mathbb{E}[X] = \sum_i \mathbb{E}[X_i]$. Then we have*

$$\mathbf{Pr}[\lambda_{\min}(X) \leq (1-\varepsilon)\mu_{\min}] \leq n \exp(-\frac{\varepsilon^2 \mu_{\min}}{2R}) \text{ for } 0 < \varepsilon < 1,$$

$$\mathbf{Pr}[\lambda_{\max}(X) \leq (1+\varepsilon)\mu_{\max}] \leq n \exp(-\frac{\varepsilon^2 \mu_{\max}}{3R}) \text{ for } 0 < \varepsilon.$$

Note that the matrices $X_1, \ldots, X_m$ can have different distributions. Before applying the matrix Chernoff bound, we take a transformation so that $\mu_{\min} = \mu_{\max} = 1$. This will make analysis easier, and allow us to deal with eigenvalues other than the largest and the smallest ones. For positive definite matrices A and B, we know that $A \preccurlyeq (1+\varepsilon)B$ is equivalent with $B^{-1/2} A B^{-1/2} \preccurlyeq (1+\varepsilon) I$. Even for positive semi-definite matrices, if A and B has the same null spaces, we can do the same thing in their column space. So, we have

$$L_H \preccurlyeq (1+\varepsilon) L_G \iff L_G^{+/2} L_H L_G^{+/2} \preccurlyeq (1+\varepsilon) L_G^{+/2} L_G L_G^{+/2}.$$

where $L_G^{+/2}$ is the square root of the Moore-Penrose pseudo-inverse of L_G. Let $\Pi = L_G^{+/2} L_G L_G^{+/2}$, which is the projection matrix on to the range of L_G. So it suffices to find a graph H where $L_G^{+/2} L_H L_G^{+/2}$ is an ε-approximation of Π. As multiplication by a fixed matrix is a linear operation and expectation commutes with linear operations, we know

$$\mathbb{E}[L_G^{+/2} L_H L_G^{+/2}] = L_G^{+/2} \mathbb{E}[L_H] L_G^{+/2} = \Pi.$$

We will project all the vectors to the span of Π and carry out the analysis, then we can assume that Π is in fact an identity.

Let

$$X_{a,b} = \begin{cases} \frac{w_{a,b}}{p_{a,b}} L_G^{+/2} L_{ab} L_G^{+/2} & \text{with probability } p_{a,b} \\ 0 & \text{otherwise.} \end{cases}$$

so that $L_G^{+/2} L_H L_G^{+/2} = \sum_{ab \in E} X_{a,b}$. We want to choose the probability $p_{a,b}$ so that

$$p_{a,b} = \frac{1}{R} w_{a,b} \| L_G^{+/2} L_{ab} L_G^{+/2} \|$$

holds for some R which we determine later. Then when an edge ab is chosen, we have $\|X_{a,b}\| = R$. We will see that for most of the edges ab, $p_{a,b}$ will be smaller than 1. If it is at least 1, then we will later partition this edge into multi-edges of probability at most 1.

We define the leverage score of edge $\ell_{a,b}$ be the weight of the edge times the effective resistance $R_{\text{eff}}(a,b) = (\mathbf{1}_a - \mathbf{1}_b)^T L_G^+ (\mathbf{1}_a - \mathbf{1}_b)$ between its endpoints, i.e.

$$\ell_{a,b} = w_{a,b} (\mathbf{1}_a - \mathbf{1}_b)^T L_G^+ (\mathbf{1}_a - \mathbf{1}_b).$$

Note that

$$\|L_G^{+/2} L_{ab} L_G^{+/2}\| = \|L_G^{+/2} (\mathbf{1}_a - \mathbf{1}_b)^T (\mathbf{1}_a - \mathbf{1}_b) L_G^{+/2}\| = \|(\mathbf{1}_a - \mathbf{1}_b) L_G^{+/2} L_G^{+/2} (\mathbf{1}_a - \mathbf{1}_b)^T\| = R_{\text{eff}}(a,b).$$

It is known that the the leverage score of an edge equals the probability that the edge appears in a random spanning tree chosen proportional to the weight $\prod_{ij \in E(T)} w_{i,j}$ of the trees, so the sum of leverage score is $n-1$. So, we have $\sum_{ab} p_{a,b} = \frac{n-1}{R}$. We choose $R = \frac{\varepsilon^2}{3.5 \ln n}$. Hence, there are at most $3.5\varepsilon^{-2} n \ln n$ edges ab with $p_{a,b} \geq 1$. For those edges, split $X_{a,b}$ into $k = \lfloor \frac{\ell_{a,b}}{R} \rfloor$ random variables which appear as $R L_G^{+/2} L_{a,b} L_G^{+/2}$ with probability 1 and one more that appears as $\frac{w_{a,b} - Rk}{p_{a,b} - k} L_G^{+/2} L_{a,b} L_G^{+/2}$ with the probability $p_{a,b}/R - k$. By using Chernoff bound on real numbers, the number of edges in H is at most $4\varepsilon^{-2} n \ln n$ with high probability. Also we have

$$\sum_{ab \in E} \mathbb{E}[X_{a,b}] = \Pi.$$

Using matrix Chernoff bound, we have

$$\mathbf{Pr}[\lambda_{\max}(\sum_{ab} X_{a,b}) > 1 + \varepsilon] \leq n \exp(-\frac{\varepsilon^2}{3R}) \leq n \exp(-1.1 \ln n) \leq n^{-1/10}.$$

For the lower bound, we can just consider on the vector spaces consisting of all vectors orthogonal to the constant vector, so treat the smallest eigenvalue of Π to be 1. We then find that

$$\mathbf{Pr}[\lambda_{\min}(\sum_{ab} X_{a,b}) < 1 - \varepsilon] \leq n \exp(-\frac{\varepsilon^2}{2R}) \leq n \exp(-1.7 \ln n) \leq n^{-1/10}.$$

Union bound yields the desired result.

3.2. **Interlacing polynomials.** The multiplicative $\log n$ term on the number of edges naturally arises in the applications of the Chernoff bound. In order to remove this term, we tackle this problem by using the method of eigenvalue interlacing. What does 'interlacing' means? Basic idea is to observe how eigenvalues changes when we add an edge to H. Once we add an edge to H, then eigenvalues of H march forward together in a well-coordinated way, which we call 'interlacing'.

Definition 3.5. *Let $p \in \mathbb{R}[x]$ be a polynomial of degree $k \in \{d-1, d\}$ with all real roots $\alpha_1 \leq \cdots \leq \alpha_k$ and $q \in \mathbb{R}[x]$ be a polynomial of degree d with all real roots $\beta_1 \leq \cdots \leq \beta_d$. If $k = d-1$, then we say p interlaces q if $\beta_1 \leq \alpha_1 \leq \cdots \leq \alpha_{d-1} \leq \beta_d$. If $k = d$, then we say p interlaces Q if $\alpha_1 \leq \beta_1 \leq \cdots \leq \alpha_d \leq \beta_d$.*

We say that a set $\{p_i : i \in [n]\}$ of real-rooted polynomials have *a common interlacing* if there is a polynomial $q \in \mathbb{R}[x]$ that interlaces each p_i. One of the key reason why we consider this concept is the following lemma. Note that adding $(\delta_i - \delta_j)^T(\delta_i - \delta_j)$ to the Laplacian matrix L_H yields the Laplacian L_{H+e} where e is an edge of weight 1. So, studying the eigenvalue of $A + vv^T$ in terms of eigenvalues of A provides some understanding on how the eigenvalues evolve when we add or remove edges.

Lemma 3.6. *Let A be a symmetric matrix and let $\mathbf{v}$ be a vector. For a real number t let $p_t(x) = \chi(A + t\mathbf{v}\mathbf{v}^T)$. Then for $t > 0$, $p_0(x)$ interlaces $p_t(x)$ and there is a monic polynomial $q(x)$ of degree $n-1$ so that for all t, $p_t(x) = \chi(A) - tq(x)$.*

Proof. Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of A and $\mu_1 \leq \cdots \leq \mu_n$ be the eigenvalues of $A + tvv^T$. The Rayleigh quotient of $A + t\mathbf{v}\mathbf{v}^T$ is always greater than or equal to the one of A, so Courant-Fischer Theorem implies $\lambda_i \leq \mu_i$. On the other hand, Courant-Fischer theorem states that

$$\begin{aligned}
\mu_i &= \max_{\dim(S)=n-i+1} \min_{\mathbf{x}\in S-\{0\},\|\mathbf{x}\|=1} \mathbf{x}^T(A + t\mathbf{v}\mathbf{v}^T)\mathbf{x} \\
&\leq \max_{\dim(S)=n-i+1} \min_{\substack{\mathbf{x}\in S,\mathbf{x}\perp\mathbf{v}=0,\\ \|\mathbf{x}\|=1}} \mathbf{x}^T(A + \mathbf{v}\mathbf{v}^T)\mathbf{x} \\
&\leq \max_{\dim(S)=n-i+1} \min_{\substack{\mathbf{x}\in S,\mathbf{x}\perp\mathbf{v}=0,\\ \|\mathbf{x}\|=1}} \mathbf{x}^T A\mathbf{x} \\
&\leq \max_{\dim(S)=n-i\, x\in S-\{0\},\|\mathbf{x}\|=1} \min \mathbf{x}^T A\mathbf{x} = \lambda_{i+1}.
\end{aligned}$$

The final inequality holds as $S \cap \mathbf{v}^\perp$ has dimension either $n-i+1$ or $n-i$. So, $p_0(x)$ interlaces $p_t(x)$.

By rescaling t if necessary, we can assume that v is a unit vector. Note that we have

$$\chi(A + te_1e_1^T) = \det(xI - A - te_1e_1^T) = \det(xI - A) - t\det(xI^{(1)} - A^{(1)}) = \chi(A) - t\chi(A^{(1)}),$$

where $A^{(1)}$ is the submatrix of A by removing its first column and row, and the polynomial $\chi(A^{(1)})$ has degree $n-1$.

Consider a rotation matrix Q with $Qe_1 = v$ with $|\det(Q)| = 1$. Then

$$\begin{aligned}
\chi(A + tvv^T) &= \chi(Q(A + tvv^T)Q^T) = \chi(QAQ^T + te_1e_1^T) \\
&= \chi(QAQ^T) - tq(x) = \chi(A) - tq(x)
\end{aligned}$$

for some $q(x)$ of degree $n-1$. □

Why is interlacing important? As we mentioned, this describes how the eigenvalues evolve when we add an edge by edge to a graph. Also, this concept is helpful to study *real-rooted polynomials*, i.e. the polynomials whose roots are all real. As the Laplacian and adjacency matrix are both symmetric, the characteristic polynomials of graphs are real-rooted, hence this concept of real-rootedness is related to what we are studying. In general, if we consider two real-rooted polynomials, it is not guaranteed that we can get a real-rooted polynomial. However, if we have two polynomials having a common interlacing, then we can say more about their sum. This can be captured by the following theorem.

Theorem 3.7 (Common interlacing theorem)**.** *Let $p_1, \ldots, p_m \in \mathbb{R}[z]$ be a set of real-rooted polynomials of degree n with positive leading coefficient and let $p_\varnothing = \sum_{i\in[m]} p_i$. If $\{p_i : i \in [m]\}$ have a common interlacing, then we have the following.*

- *$p_\varnothing$ is real-rooted.*
- *for each $k \in [n]$, $\min_i \lambda_k(p_i) \leq \lambda_k(p_\varnothing) \leq \max_i \lambda_k(p_i)$.*

Proof. Let $a_1 \leq \cdots \leq a_n$ be a common interlacing sequence for the polynomials $p_1, \ldots, p_m$, meaning that $a_k \leq \lambda_k(p_i) \leq a_{k+1}$, and let a_{n+1} be a large enough number so that $p_i(a_{n+1}) > 0$ for all i. As each p_i has positive leading coefficient, they all have the same sign at a_1 and the same opposite sign at a_2 etc so that it has negative sign at a_n and positive sign at a_{n+1}. This implies that $p_\varnothing$ changes sign in the interval $[a_1, a_2]$ so it has to vanish in the interval by the intermediate value theorem.

So, it has at n real roots. Moreover, it is easy to see that such roots are sandwiched by the smallest and the largest root of the polynomials p_i's on the interval $[a_k, a_{k+1}]$. □

The following lemma swill be useful for us later.

Lemma 3.8. *Let p, q be polynomials of degree n and $n-1$, and let $p_t(x) = p(x) - tq(x)$. If p_t is real-rooted for all $t \in \mathbb{R}$, then q interlaces p.*

Proof. If they have common roots α we divide $(x-\alpha)$ on both polynomial to assume that their roots are distinct. If q does not interlace p, then p has two roots $\lambda_i < \lambda_{i+1}$ where q has no roots in $[\lambda_i, \lambda_{i+1}]$. As t can be both positive and negative, we lose no generality by assuming that p and q both are positive on the interval $(\lambda_i, \lambda_{i+1})$. For small t, $p_t(x)$ has two roots in the interval, but for large t, it has no real roots in the interval. As t increase, the roots continuously vary and do not cross λ_i and λ_{i+1}, this yields a contradiction to the real-rootedness of $p_t(x)$.

□

Lemma 3.9. *If p, q are polynomials with positive leading coefficients of degree n and $t \cdot p(x) + (1-t)q(x)$ is real-rooted for all $t \in [0,1]$, then p and q have common interlacing.*

Proof. Assume that p, q have no common roots. Let $\lambda_1 \leq \cdots \leq \lambda_n$ and $\mu_1 \leq \cdots \leq \mu_n$ be the roots of $p(x)$ and $q(x)$, respectively. If they don't have a common interlacing, up to swapping p and q, we have $\mu_i < \lambda_{i-1} \leq \lambda_i < \mu_{i+1}$ for some i. Without loss of generality, assume $n-i$ is odd. Then $p(x)$ is positive on $(\lambda_{i-1}, \lambda_i)$ and $q(x)$ is negative on the interval $[\lambda_{i-1}, \lambda_i]$. Similarly as before, continuously decreasing t from 1 to 0 will move the roots of $tp(x) + (1-t)q(x)$ in a continuous way. And these roots lie within the interval $[\lambda_{i-1}, \lambda_i]$. However, when $t = 0$, there are not roots in this interval, so these roots become complex at some point, a contradiction that this is real-rooted for all t. □

More generally, the following theorem characterizes the polynomials with common interlacing.

Theorem 3.10 (Dedieu, 1992)**.** *Let $p_1, \ldots, p_m \in \mathbb{R}[x]$ be polynomials of degree n with positive leading coefficients. Then $p_1, \ldots, p_m$ have a common interlacing if and only if all convex combinations of $p_1, \ldots, p_m$ are real-rooted polynomials, i.e. $\sum_{i\in[m]} \lambda_i p_i$ is real-rooted for all $\lambda_i \geq 0$ with $\sum_{i\in[m]} \lambda_i = 1$.*

The following fact is also useful for us later.

Lemma 3.11. *Let p, q be polynomials of degree n and $n-1$ such that q interlaces p and both have positive leading coefficients. For every $t > 0$, define $p_t(x) = p(x) - tq(x)$. Then $p_t(x)$ is real-rooted and $p(x)$ interlaces $p_t(x)$.*

Proof sketch of the lemma. As p, q are interlacing, we know what sign does $q(x)$ have at the i-th root λ_i of $p(x)$. This yields a root in $[\lambda_i, \lambda_{i+1}]$ for every i, and yields the desired statement. □

Also, the following theorem provides particular instances where the interlacing occur, which follows from Lemma 3.6.

Theorem 3.12 (Cauchy interlacing theorem)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $v \in \mathbb{R}^n$. Then the characteristic polynomial $\chi_A(x)$ interlaces $\chi_{A+vv^T}(x)$.*

Moreover, we can express $\chi_{A+vv^T}(x)$ in terms of $\chi_A(x)$. For this, we need the following lemma.

Lemma 3.13 (Matrix-determinant lemma)**.** $\det(A+uv^T) = (1+v^TA^{-1}u)\det(A)$.

Proof. We know that

$$\begin{pmatrix} I & 0 \\ v^T & 1 \end{pmatrix}\begin{pmatrix} I+uv^T & u \\ 0 & 1 \end{pmatrix}\begin{pmatrix} I & 0 \\ -v^T & 1 \end{pmatrix} = \begin{pmatrix} I & u \\ 0 & 1+v^Tu \end{pmatrix}$$

Taking determinant, we obtain $\det(I+uv^T) = 1+v^Tu$. Hence, we have $\det(A+uv^T) = \det(A)\det(I+A^{-1}uv^T) = \det(A)(1+v^TA^{-1}u)$. □

By this lemma, we have

$$\det(xI-A-vv^T) = (1-v^T(xI-A)^{-1}v)\det(xI-A).$$

Assume $v \neq 0$ as this is trivial otherwise. Let $\lambda_1,\dots,\lambda_n$ be the eigenvalues of A and $u_1,\dots,u_n$ be orthonormal eigenvectors corresponding to $\lambda_1,\dots,\lambda_n$. Then we have $A = \sum \lambda_i u_i u_i^T$, and orthonormality implies $\sum u_iu_i^T = I$. Thus we have $xI-A = \sum_i (x-\lambda_i)u_iu_i^T$ so we have

$$(xI-A)^{-1} = \sum_{i\in[n]} \frac{1}{x-\lambda_i} u_iu_i^T.$$

Hence, we have

$$1-v^T(xI-A)^{-1}v = 1-\sum_i \frac{1}{x-\lambda_i} v^Tu_iu_i^Tv = 1-\sum_i \frac{|\langle u_i,v\rangle|^2}{x-\lambda_i}.$$

Thus, we have

$$\chi_{A+vv^T}(x) = \chi_A(x)\cdot\left(1-\sum_i \frac{|\langle u_i,v\rangle|^2}{x-\lambda_i}\right).$$

This roughly says that the eigenvalues μ of $A+vv^T$ is either an eigenvalue of A, or when $f(\mu) = \sum_i \frac{|\langle u_i,v\rangle|^2}{x-\lambda_i}$ becomes 1. As this function maps $(\lambda_i,\lambda_{i+1})$ to $\mathbb{R}$ bijectively for $i\in[n-1]$ (it also maps (λ_n,∞) to $\mathbb{R}_+$), we can also use this to show the Cauchy-interlacing theorem. Note that the value of $f(\cdot)$ indicates some bound on the new eigenvalues.

3.3. **Linear size sparsifiers.** Using eigenvalue interlacing techniques, we can prove that there exists a graph H with linearly many edges approximating a given graph G.

Theorem 3.14. [1] *For every $d>1$ and every undirected weighted graphs $G=(V,E,w)$ with $n=|V|$, $m=|E|$ there is a weighted subgraph H with $d(n-1)$ edges that satisfies*

$$L_G \preccurlyeq L_H \preccurlyeq \frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}} L_G.$$

If we choose $d = \frac{100}{\varepsilon^2}$ then we have $\frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}} \leq 1+\varepsilon$. This proves that there is a graph H with $d(n-1) = O(\frac{n}{\varepsilon^2})$ edges such that $L_G \preccurlyeq L_H \preccurlyeq (1+\varepsilon)L_G$. Note that the best ε we can get here is roughly $\frac{4\sqrt{d-1}}{d}$, which is roughly twice of $\frac{2\sqrt{d-1}}{d}$. As we saw last year, Alon-Boppana theorem yields that this is roughly twice of best possible value.

We will prove the following theorem.

Theorem 3.15. *Let $d>1$ and $v_1,\dots,v_m\in\mathbb{R}^n$ with $\sum_i v_iv_i^T = I$. Then there is a subset $S\subseteq[m]$ with $|S|\leq dn$ and nonzero scalars $s_i>0$ for all $i\in S$, such that*

$$I \preccurlyeq \sum_{i\in S} s_iv_iv_i^T \leq (\frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}})I.$$

Let's first see how this theorem imply what we want. Assume that G is a connected graph because we can find approximations for each component and take a disjoint union. As we know $L_G = K^T K$ for edge-vertex signed incidence weighted graph K. Consider the matrix $M = L_G^{+/2} K^T$ and its column vectors $v_1, \dots, v_m$. (This is exactly same as taking the vectors $L_G^{+/2} L_{ab} L_G^{+/2}$ for each edge $ab \in E(G)$.) Then we have

$$\sum_i v_i v_i^T = L_G^{+/2} K^T K (L_G^{+/2})^T = L_G^{+/2} L_G L_G^{+/2} = \Pi.$$

As Π is an identity on the image of L_G (which is isomorphic to $\mathbb{R}^{n-1}$), the above theorem yields a subset $S \subseteq [m]$ of size at most $d(n-1)$ and numbers s_i for $i \in S$ such that

$$\Pi \preccurlyeq \sum_{i \in S} s_i v_i v_i^T \preccurlyeq (\frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}})\Pi.$$

Let M_S be the m by m diagonal matrix with i-th component s_i for $i \in S$ and 0 for $i \notin S$. Note that $M_S^{1/2} K$ is an edge-vertex incidence matrix obtained by deleting some some edges and rescaling some edge weights from G. So $L_H = K^T M_S K$ is the Laplacian of the subgraph H with edge weights $w_i s_i$. As we know $\sum_{i \in S} s_i v_i v_i^T = L_G^{+/2} L_H L_G^{+/2}$, we have $L_G \preccurlyeq L_H \preccurlyeq (\frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}) L_G$.

Now, we aim to show that Theorem 3.15 holds. We say the vectors are in *isotropic position* if their outer products sums to the identity. In order to prove the theorem, we will choose v_i one by one. In fact, Cauchy's interlacing theorem states that all eigenvalues march forward when we add vv^T to a matrix A.

Assume A is the matrix we have by adding the chosen vectors so far. For eigenvalues $\lambda_1, \dots, \lambda_n$ and corresponding orthonormal eigenvectors $u_1, \dots, u_n$, if we choose v_i uniformly at random, then we have

$$\mathbb{E}[\sum_j (v_i^T u_j)^2] = \frac{1}{m} \sum_j u_j^T v_i v_i^T u_j = \frac{1}{m} \sum_j u_j^T u_j = \frac{1}{m}.$$

This shows that, in expectation, all eigenvalues of A will shift by the same amount $1/m$. So we can expect that all eigenvalues are more or less t/m after adding t vectors randomly chosen from $v_1, \dots, v_m$.

We will prove that this indeed happens. Let's compute the expected characteristic polynomial for $A + vv^T$. Recall Lemma 3.13.

Let A_j be the matrix we obtain by adding vv^T for j random vectors v chosen from $\{v_i\}$. Using matrix-determinant lemma, we have

$$\chi_{A+vv^T}(x) = (1 - v^T (xI - A)^{-1} v) \chi_A(x).$$

As $A = \sum \lambda_i u_i u_i^T$ and $I = \sum u_i u_i^T$ by orthonormality of $u_1, \dots, u_n$, so we have $xI - A = \sum (x - \lambda_i) u_i u_i^T$. Hence, $(xI - A)^{-1} = \sum_{i=1}^n \frac{1}{x - \lambda_i} u_i u_i^T$. Thus

$$1 - v^T (xI - A)^{-1} v = 1 - \sum_{i=1}^n \frac{1}{x - \lambda_i} v^T u_i u_i^T v = 1 - \sum_{i=1}^n \frac{(u_i^T v)^2}{x - \lambda_i}.$$

So,

$$\chi_{A+vv^T}(x) = \chi_A(x)(1 - \sum_{i=1}^n \frac{(u_i^T v)^2}{x - \lambda_i}).$$

Hence, in expectation, we have

$$\mathbb{E}\chi_{A+vv^T}(x) = \chi_A(x)(1 - \frac{1}{m} \sum_j \frac{1}{x - \lambda_j}).$$

As $\det(xI - A) = \prod_i (x - \lambda_i)$, we know that

$$\mathbb{E}\chi_{A_{i+1}}(x) = \chi_{A_i}(x) - \frac{1}{m}\chi'_{A_i}(x).$$

So, if we actually choose 'average of v' every time, then we have

$$\chi_{A_j}(x) = (1 - \frac{1}{m}\frac{d}{dx})^j \chi_{A_0}(x) = (1 - \frac{1}{m}\frac{d}{dx})^j x^n.$$

the set of such polynomials forms a family of orthonormal polynomials known as Laguerre polynomials. It is known that when $j = dn$, we have $\frac{\lambda_{\max}(A_j)}{\lambda_{\min}(A_j)} = \frac{d+1+2\sqrt{d}}{d+1-2\sqrt{d}}$. This is precisely the bound we hope for. However, as we are not choosing 'the average' vector from $v_1, \ldots, v_m$, we actually have to make a specific choice. So we need more complicated argument with the roles of weight s_i.

We now start the proof of Theorem 3.15. The following well-known lemma is useful.

Lemma 3.16 (Sherman-Morrison)**.** *For nonsingular symmetric matrix A and a vector v, we have*

$$(A - vv^T)^{-1} = A^{-1} + \frac{A^{-1}vv^TA^{-1}}{1 - v^TA^{-1}v}.$$

For a symmetric matrix A with eigenvalues $\lambda_1 \le \cdots \le \lambda_n$ and values $\ell, u \in \mathbb{R}$, we define lower/upper barrier function as

$$\Phi_\ell(A) = \mathrm{tr}((A - \ell I)^{-1}) = \sum_{i=1}^n \frac{1}{\lambda_i - \ell}, \text{ and } \Phi^u(A) = \mathrm{tr}((uI - A)^{-1}) = \sum_{i=1}^n \frac{1}{u - \lambda_i}.$$

This is roughly the expression we saw before, which provides some bound on the eigenvalues of the new matrix $A + vv^T$. When u is an upper bound on the largest eigenvalue λ_n of A, $\Phi^u(A)$ being small means that it is an upper bound with a room to spare, while $\Phi^u(A)$ being large means that it slightly over λ_n. At the time, we have an upper bound u_{old} with $\lambda_n \le u_{\text{old}}$. We want to choose s, v and u_{new} so that u_{new} is a better (or at least not worse) upper bound of the largest eigenvalue of $A + svv^T$ than u_{old} is of A. In other words, we want $\Phi^{u_{\text{new}}}(A + svv^T) \le \Phi^{u_{\text{old}}}(A)$. The following lemma tells us for which choices of s, v we can get such a conclusion.

Lemma 3.17 (Shifting upper barrier)**.** *Let A be a symmetric n by n matrix and let $v \in \mathbb{R}^n$ be any vector. Let $u_{\text{old}} > 0$ satisfy $\lambda_{\max}(A) < u_{\text{old}}$ and $u_{\text{new}} = u_{\text{old}} + \delta$ for some $\delta > 0$. Let $s > 0$. If*

$$U_A(v) := \frac{v^T(u_{\text{new}}I - A)^{-2}v}{\Phi^{u_{\text{old}}}(A) - \Phi^{u_{\text{new}}}(A)} + v^T(u_{\text{new}}I - A)^{-1}v \le \frac{1}{s},$$

then we have

$$\Phi^{u_{\text{new}}}(A + svv^T) \le \Phi^{u_{\text{old}}}(A) \text{ and } \lambda_{\max}(A + svv^T) \le u_{\text{new}}.$$

Proof. By Sherman-Morrison lemma, we have

$$\begin{aligned}
\Phi^{u_{\text{new}}}(A + svv^T) &= \mathrm{tr}((u_{\text{new}}I - A - svv^T)^{-1}) \\
&= \mathrm{tr}\left[(u_{\text{new}}I - A)^{-1} + \frac{s(u_{\text{new}}I - A)^{-1}vv^T(u_{\text{new}}I - A)^{-1}}{1 - sv^T(u_{\text{new}}I - A)^{-1}v}\right] \\
&= \mathrm{tr}\left[(u_{\text{new}}I - A)^{-1}\right] + \frac{s \cdot \mathrm{tr}\left[(u_{\text{new}}I - A)^{-1}vv^T(u_{\text{new}}I - A)^{-1}\right]}{1 - sv^T(u_{\text{new}}I - A)^{-1}v} \\
&= \Phi^{u_{\text{new}}}(A) + \frac{v^T(u_{\text{new}}I - A)^{-2}v}{1/s - v^T(u_{\text{new}}I - A)^{-1}v} \\
&= \Phi^{u_{\text{old}}}(A) - (\Phi^{u_{\text{old}}}(A) - \Phi^{u_{\text{new}}}(A)) + \frac{v^T(u_{\text{new}}I - A)^{-2}v}{1/s - v^T(u_{\text{new}}I - A)^{-1}v}.
\end{aligned}$$

Here, $1/s \geq U_A(v) > v^T(u_{\text{new}}I - A)^{-1}v$ and $\Phi^{u_{\text{old}}}(A) - \Phi^{u_{\text{new}}}(A) > 0$, we know that the denominator $1/s - v^T(u_{\text{new}}I - A)^{-1}v$ is positive. However, we know that

$$-(\Phi^{u_{\text{old}}}(A) - \Phi^{u_{\text{new}}}(A)) + \frac{v^T(u_{\text{new}}I - A)^{-2}v}{1/s - v^T(u_{\text{new}}I - A)^{-1}v} \leq 0$$

from the assumption that $U_A(v) \leq \frac{1}{s}$. This proves $\Phi^{u_{\text{new}}}(A + svv^T) \leq \Phi^{u_{\text{old}}}(A)$.

From the above, for any $t \leq s$, we have that $\Phi^{u_{\text{new}}}(A + tvv^T) \leq \Phi^{u_{\text{old}}}(A)$ is finite, as $1/s \leq 1/t$. If $\lambda_{\max}(A + svv^T) \geq u_{\text{new}}$, then there exists some $0 \leq s' \leq s$ with $\lambda_{\max}(A + s'vv^T) = u_{\text{new}}$, then we have $\Phi^{u_{\text{new}}}(A + s'vv^T) = \infty$, a contradiction. Hence $\lambda_{\max}(A + svv^T) \leq u_{\text{new}}$ holds. □

Similarly, we can obtain a condition for finding a good new lower bound.

Lemma 3.18 (Shifting lower Barrier)**.** *Let A be a symmetric n by n matrix and let $v \in \mathbb{R}^n$ be any vector. Let $\ell_{\text{old}} > 0$ satisfy $\lambda_{\min}(A) > \ell_{\text{old}}$ and $\ell_{\text{new}} = \ell_{\text{old}} + \delta$ for some $0 < \delta \leq \frac{1}{\Phi_{\ell_{\text{old}}}(A)}$. Let $s > 0$. If*

$$L_A(v) := \frac{v^T(\ell_{\text{new}}I - A)^{-2}v}{\Phi_{\ell_{\text{new}}}(A) - \Phi_{\ell_{\text{old}}}(A)} - v^T(A - \ell_{\text{new}}I)^{-1}v \geq \frac{1}{s} > 0,$$

then we have

$$\Phi_{\ell_{\text{new}}}(A + svv^T) \leq \Phi_{\ell_{\text{old}}}(A) \text{ and } \lambda_{\min}(A + svv^T) > \ell_{\text{new}}.$$

Proof. As $\lambda_{\min}(A) > \ell_{\text{old}}$, all eigenvalues of A are strictly larger than ℓ_{old}. Hence $\frac{1}{\lambda_i - \ell_{\text{old}}} > 0$ for all $i \in [n]$. As $\Phi_{\ell_{\text{oold}}}(A) = \sum \frac{1}{\lambda_i - \ell_{\text{old}}} \leq \frac{1}{\delta}$, we have $\frac{1}{\lambda_{\min}(A) - \ell_{\text{old}}} < \frac{1}{\delta}$. This shows that $\lambda_{\min}(A) > \ell_{\text{old}} + \delta = \ell_{\text{new}}$. Hence, for every $s > 0$, we have $\lambda_{\min}(A + svv^T) > \lambda_{\min}(A) > \ell_{\text{new}}$. This shows the second inequality.

To prove the first inequality, use Sherman-Morrison lemma, we have

$$\begin{aligned}
\Phi_{\ell_{\text{new}}}(A + svv^T) &= \text{tr}((A + svv^T - \ell_{\text{new}}I)^{-1}) \\
&= \text{tr}\left[(A - \ell_{\text{new}}I)^{-1} - \frac{s(A - \ell_{\text{new}}I)^{-1}vv^T(A - \ell_{\text{new}}I)^{-1}}{1 - sv^T(A - \ell_{\text{new}}I)^{-1}v}\right] \\
&= \text{tr}\left[(\ell_{\text{new}}I - A)^{-1}\right] - \frac{s \cdot \text{tr}\left[(A - \ell_{\text{new}}I)^{-1}vv^T(A - \ell_{\text{new}}I)^{-1}\right]}{1 - sv^T(A - \ell_{\text{new}}I)^{-1}v} \\
&= \Phi_{\ell_{\text{new}}}(A) - \frac{v^T(A - \ell_{\text{new}}I)^{-2}v}{1/s - v^T(A - \ell_{\text{new}}I)^{-1}v} \\
&= \Phi_{\ell_{\text{old}}}(A) + (\Phi_{\ell_{\text{new}}}(A) - \Phi_{\ell_{\text{old}}}(A)) + \frac{v^T(A - \ell_{\text{new}}I)^{-2}v}{1/s - v^T(A - \ell_{\text{new}}I)^{-1}v}.
\end{aligned}$$

Thus, it suffices to prove

$$(\Phi_{\ell_{\text{new}}}(A) - \Phi_{\ell_{\text{old}}}(A)) + \frac{v^T(A - \ell_{\text{new}}I)^{-2}v}{1/s - v^T(A - \ell_{\text{new}}I)^{-1}v} \leq 0,$$

which comes from the assumption that $L_A(v) \geq 1/s$. □

The eigenvalues all will increase in the process of adding $v_i v_i^T$, and the upper bound and lower bound all will increase in the process. The upper bound and the largest eigenvalues both increase, so we want the upper bound to not increase too little. On the other hand, the lower bound and the smallest eigenvalues both increase, so we want the lower bound to not increase too much. This asymmetry causes some problem in the later proof, and the following lemma is useful to deal with the issue.

Lemma 3.19. *Suppose that* $\lambda_{\min}(A) > \ell_{\text{old}}$ *and* $0 \le \Phi_{\ell_{\text{old}}}(A) \le \varepsilon$ *and* $1/\delta - \varepsilon \ge 0$. *Then*

$$\frac{\sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-2}}{\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1}} - \Phi_{\ell_{\text{new}}}(A) \ge \frac{1}{\delta} - \Phi_{\ell_{\text{old}}}(A).$$

Proof. We have $\delta \le 1/\varepsilon$. As $\frac{1}{\lambda_{\min}(A)-\ell_{\text{old}}} < \sum \frac{1}{\lambda_i - \ell_{\text{old}}} = \Phi_{\ell_{\text{old}}}(A) \le \varepsilon$, we have $\lambda_{\min}(A) - \ell_{\text{old}} > 1/\varepsilon \ge \delta$. Using this, we have $\lambda_{\min}(A) > \ell_{\text{old}} + \delta = \ell_{\text{new}}$. Thus for each i, $(\lambda_i - \ell_{\text{new}})^{-1}$ is positive.

Rearranging the desired result yields that

$$\begin{aligned}
\sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-2} &\ge \left(\frac{1}{\delta} + \Phi_{\ell_{\text{new}}}(A) - \Phi_{\ell_{\text{old}}}(A)\right)\left(\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1}\right) \\
&= \left(\frac{1}{\delta} + \delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1}\right)\left(\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1}\right) \\
&= \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1} + \left(\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1}\right)^2.
\end{aligned}$$

This is equivalent to the following.

$$\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-2}(\lambda_i - \ell_{\text{old}})^{-1} \ge \left(\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1}\right)^2. \tag{3.1}$$

So, it suffices to prove (3.1).

Let $x, y \in \mathbb{R}^n$ be the vectors with entries $x_i = (\lambda_i - \ell_{\text{old}})^{-1/2}$ and $y_i = (\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1/2}$ respectively. Then the right side becomes $\delta^2 (x^T y)^2$. Using Cauchy-Schwarz inequality and the fact that $1 \ge \delta\varepsilon$, we have

$$\begin{aligned}
\delta^2 (x^T y)^2 \le \delta^2 \|x\|^2 \|y\|^2 &= \delta \left(\sum_{i\in[n]}(\lambda_i - \ell_{\text{old}})^{-1}\right)\left(\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-2}(\lambda_i - \ell_{\text{old}})^{-1}\right) \\
&= \delta \Phi_{\ell_{\text{old}}}(A)\left(\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-2}(\lambda_i - \ell_{\text{old}})^{-1}\right) \\
&\le \delta\varepsilon \left(\delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-2}(\lambda_i - \ell_{\text{old}})^{-1}\right) \\
&\le \delta \sum_{i\in[n]}(\lambda_i - \ell_{\text{new}})^{-2}(\lambda_i - \ell_{\text{old}})^{-1}.
\end{aligned}$$

This proves the lemma. □

Now, we prove the following key lemma.

Lemma 3.20 (Barrier shifting lemma)**.** *Let* A *be an* n *by* n *symmetric real matrix. Let* $u_{\text{old}}, \ell_{\text{old}} > 0$ *be numbers satisfying* $\lambda_{\max}(A) < u_{\text{old}}$ *and* $\lambda_{\min}(A) > \ell_{\text{old}}$. *Let* $\varepsilon_u, \varepsilon_\ell, \delta_u, \delta_\ell$ *satisfy* $\Phi^{u_{\text{old}}}(A) \le \varepsilon_u, \Phi_{\ell_{\text{old}}}(A) \le \varepsilon_\ell$ *and*

$$0 \le \frac{1}{\delta_u} + \varepsilon_u \le \frac{1}{\delta_\ell} - \varepsilon_\ell.$$

Let $u_{\text{new}} = u_{\text{old}} + \delta_u$ and $\ell_{\text{new}} = \ell_{\text{old}} + \delta_\ell$. Then there exists $i \in [m]$ and $s > 0$ such that the following holds.

(1) $U_A(v_i) \le \frac{1}{s} \le L_A(v_i)$.
(2) $\lambda_{\max}(A + sv_iv_i^T) < u_{\text{new}}$.
(3) $\lambda_{\min}(A + sv_iv_i^T) > \ell_{\text{new}}$.
(4) $\Phi^{u_{\text{new}}}(A + sv_iv_i^T) \le \varepsilon_u$.
(5) $\Phi_{\ell_{\text{new}}}(A + sv_iv_i^T) \le \varepsilon_\ell$.

Proof. In order to use the previous two lemmas, we want to find s and i satisfying

$$U_A(v_i) \le \frac{1}{s} \le L_A(v_i).$$

To make sure such a choice exists, we aim to show that $\sum_{i\in[m]} U_A(v_i) \le \sum_{i\in[m]} L_A(v_i)$. Note that for a matrix M, we have $\operatorname{tr}(M) = \operatorname{tr}[\sum_i (Mv_i)v_i^T] = \operatorname{tr}[\sum_i v_i^T(Mv_i)]$. Using this, we have

$$\begin{aligned}
\sum_{i\in[m]} U_A(v_i) &= \frac{\sum_{i\in[m]} v_i^T(u_{\text{new}}I - A)^{-2}v_i}{\Phi^{u_{\text{old}}}(A) - \Phi^{u_{\text{new}}}(A)} + \sum_{i\in[m]} v_i^T(u_{\text{new}}I - A)^{-1}v_i \\
&= \frac{\operatorname{tr}[(u_{\text{new}}I - A)^{-2}]}{\Phi^{u_{\text{old}}}(A) - \Phi^{u_{\text{new}}}(A)} + \operatorname{tr}[(u_{\text{new}}I - A)^{-1}] \\
&= \frac{\sum_{i\in[m]}(u_{\text{new}} - \lambda_i)^{-2}}{\sum_{i\in[m]}((u_{\text{old}} - \lambda_i)^{-1} - (u_{\text{new}} - \lambda_i)^{-1})} + \Phi^{u_{\text{new}}}(A) \\
&= \frac{\sum_{i\in[m]}(u_{\text{new}} - \lambda_i)^{-2}}{\delta_u \sum_{i\in[m]}((u_{\text{old}} - \lambda_i)^{-1}(u_{\text{new}} - \lambda_i)^{-1})} + \Phi^{u_{\text{new}}}(A) \\
&< \frac{1}{\delta_u} + \Phi^{u_{\text{new}}}(A) \le \frac{1}{\delta_u} + \Phi^{u_{\text{old}}}(A) \le \frac{1}{\delta_u} + \varepsilon_u.
\end{aligned}$$

We have the final line as $u_{\text{new}} > u_{\text{old}}$ implies $\sum_{i\in[m]}(u_{\text{new}} - \lambda_i)^{-2} < \sum_{i\in[m]}((u_{\text{old}} - \lambda_i)^{-1}(u_{\text{new}} - \lambda_i)^{-1})$. Similarly, we have

$$\begin{aligned}
\sum_{i\in[m]} L_A(v_i) &= \frac{\sum_{i\in[m]} v_i^T(A - \ell_{\text{new}}I)^{-2}v_i}{\Phi_{\ell_{\text{new}}}(A) - \Phi_{\ell_{\text{old}}}(A)} + \sum_{i\in[m]} v_i^T(A - \ell_{\text{new}}I)^{-1}v_i \\
&= \frac{\operatorname{tr}[(A - \ell_{\text{new}}I)^{-2}]}{\Phi_{\ell_{\text{new}}}(A) - \Phi_{\ell_{\text{old}}}(A)} + \operatorname{tr}[(A - \ell_{\text{new}}I)^{-1}] \\
&= \frac{\sum_{i\in[m]}(\lambda_i - \ell_{\text{new}})^{-2}}{\sum_{i\in[m]}((\lambda_i - \ell_{\text{new}})^{-1} - (\lambda_i - \ell_{\text{old}})^{-1})} + \Phi_{\ell_{\text{new}}}(A) \\
&= \frac{\sum_{i\in[m]}(\lambda_i - \ell_{\text{new}})^{-2}}{\delta_\ell \sum_{i\in[m]}((\lambda_i - \ell_{\text{new}})^{-1}(\lambda_i - \ell_{\text{old}})^{-1})} + \Phi_{\ell_{\text{new}}}(A).
\end{aligned}$$

Here, unlike the case before, the number ℓ_{new} is bigger than ℓ_{old}, so there's no guarantee that ℓ_{new} is smaller than λ_i, so we don't know if $(\lambda_i - \ell_{\text{new}})^{-1} > (\lambda_i - \ell_{\text{old}})^{-1}$ holds. However, Lemma 3.19 still yields that the above number is at least $\frac{1}{\delta_\ell} - \Phi_{\ell_{\text{old}}}(A) \ge \frac{1}{\delta_\ell} - \varepsilon_\ell$.

Thus, we have

$$\sum_{i\in[m]} U_A(v_i) \le \frac{1}{\delta_u} + \varepsilon_u \le \frac{1}{\delta_\ell} - \varepsilon_\ell \le \sum_{i\in[m]} L_A(v_i).$$

This shows that there exists an $i \in [m]$ such that $0 < U_A(v_i) \le L_A(v_i)$, so choose s so that $U_A(v_i) \le \frac{1}{s} \le L_A(v_i)$. For this choice of i and s, we have

(1) $\lambda_{\max}(A + sv_iv_i^T) < u_{\text{new}}$ and $\Phi^{u_{\text{new}}}(A + sv_iv_i^T) \leq \Phi^{u_{\text{old}}}(A) \leq \varepsilon_u$ by Shifting upper Barrier Lemma.
(2) $\lambda_{\min}(A + sv_iv_i^T) > \ell_{\text{new}}$ and $\Phi_{\ell_{\text{new}}}(A + sv_iv_i^T) \leq \Phi_{\ell_{\text{old}}}(A) \leq \varepsilon_\ell$ by Shifting lower Barrier Lemma.

This finishes the proof. □

Finally, we prove Theorem 3.15.

Let

$$\delta_\ell = 1, \delta_u = \frac{\sqrt{d}+1}{\sqrt{d}-1}, \varepsilon_\ell = \frac{1}{\sqrt{d}}, \varepsilon_u = \frac{\sqrt{d}-1}{d+\sqrt{d}}, \ell_0 = \frac{-n}{\varepsilon_\ell}, u_0 = \frac{n}{\varepsilon_u}.$$

Then we have $\frac{1}{\delta_u} + \varepsilon_u = \frac{1}{\delta_\ell} - \varepsilon_\ell$.

Let A_0 be the empty matrix. For each $0 \leq j \leq dn-1$, we will choose v_i and s and let $A_{j+1} = A_j + sv_iv_i^T$. Initially, we have $\ell_0 < \lambda_{\min}(A_0) = 0 = \lambda_{\max}(A_0) < u_0$ and

$$\Phi^{u_0}(A_0) = \operatorname{tr}[(u_0 I)^{-1}] = \frac{n}{u_0} = \varepsilon_u,$$
$$\Phi_{\ell_0}(A_0) = \operatorname{tr}[(-\ell_0 I)^{-1}] = \frac{n}{-\ell_0} = \varepsilon_\ell.$$

For each $i \in [dn]$, let $u_i = u_0 + i\delta_u$ and $\ell_i = \ell_0 + i\delta_\ell$. Assume that in jth turn, we have

$$\ell_j < \lambda_{\min}(A_j) \leq \lambda_{\max}(A_j) < u_j$$
$$\Phi^{u_j}(A_j) \leq \varepsilon_u, \Phi_{\ell_j}(A_j) \leq \varepsilon_\ell.$$

By Lemma 3.20, there exists $i \in [m]$ and $s_j > 0$ such that $A_{j+1} = A_j + s_j v_i v_i^T$ which yields

$$\ell_{j+1} < \lambda_{\min}(A_{j+1}) \leq \lambda_{\max}(A_{j+1}) < u_{j+1}$$
$$\Phi^{u_{j+1}}(A_{j+1}) \leq \varepsilon_u, \Phi_{\ell_{j+1}}(A_{j+1}) \leq \varepsilon_\ell.$$

By running this until we have A_{dn}, we obtain a set $S \subseteq [m]$ of size at most dn. Then we have

$$(d-\sqrt{d})n = \ell_0 + dn\delta_\ell = \ell_{dn} \leq \lambda_{\min}(A_{dn}) \leq \lambda_{\max}(A_{dn}) < u_{dn} = u_0 + dn\delta_u = \frac{\sqrt{d}(\sqrt{d}+1)^2}{\sqrt{d}-1}n.$$

Hence, $A = \frac{\sqrt{d}-1}{d\sqrt{dn}} A_{dn}$ satisfies

$$\frac{(\sqrt{d}-1)^2}{d} \leq \lambda_{\min}(A) \leq \lambda_{\max}(A) \leq \frac{(\sqrt{d}+1)^2}{d}.$$

This proves what we want.

Note that this proof naturally yields an algorithm for constructing a spectral sparsifier which runs in $O(dn^3m)$ time.

4. Bipartite Ramanujan graphs via lifts

Recall that last year we proved Alon-Boppana theorem, which states that any graph has its second largest eigenvalue at least $2\sqrt{d-1}-\varepsilon$. This motivates the following definition regarding how strong an expander can be.

Definition 4.1 (Ramanujan graph)**.** *A d-regular graph is Ramanujan if every eigenvalue of its adjacency matrix except the largest one is between $-2\sqrt{d-1}$ and $2\sqrt{d-1}$. A d-regular bipartite graph is Ramanujan if every eigenvalue of tis adjacency matrix except the largest and smallest one is between $-2\sqrt{d-1}$ and $2\sqrt{d-1}$.*

It is widely believed that a random d-regular graph has a good probability to be a Ramanujan graph.

In 2004, Friedman proved that $\lambda_i(G) \le 2\sqrt{d-1}+\varepsilon$ holds for all $i>1$ with probability $1-O(n^{\lceil(\sqrt{d-1}+1)/2\rceil-1})$. Miller-Novikoff-Sabelli empirically showed that the largest nontrivial positive eigenvalue and the smallest nontrivial negative eigenvalue of a d-regular graph follows some distribution called a Tracy-Widom distribution. It is conjectured that approximately 52% of d-regular bipartite graphs are Ramanujan and approximately 27% of d-regular non-bipartite graphs should be Ramanujan.

Also, d-regular graphs of degree d are known to exist for infinitely many d. Lubotzky, Phillips, Sarnak, 1986 and Margulis 1988, Morgenstern 1994 showed them. They used number theoretic constructions. Here, we will provide a more flexible approach to build a bipartite Ramanujan graphs. The result is from [5] and the proof uses techniques form [7].

4.1. Interlacing families.

We will consider a notion called interlacing families. Ideally, we want to show that out of whole choices of graphs, a random graph is Ramanujan with a positive probability. In some sense, there are huge set of characteristic polynomials where the expectation of them is like a characteristic polynomial of a Ramanujan graph. In such a situation, we want to be able to choose one polynomial from the set which acts like the expectation. This is possible if all such polynomials have a common interlacing.

However, hoping that all of them has a common interlacing is too much. Still, we want to somehow have control on the roots of (at least) one of the polynomials. More precisely, we want to conclude that the maximum root of one of the polynomials is upper bounded by the maximum root of the average (equivalently the sum) of the polynomials. We can achieve this by enforcing some layered structures among the polynomials as follows.

Definition 4.2 (Interlacing families)**.** *Let $S_1,\dots,S_m$ be finite sets and $S=\prod_i S_i$. For each $\mathbf{s}=(s_1,\dots,s_m)\in S$, we have a polynomial $p_{\mathbf{s}}$ with positive leading coefficient. For each $k\in[m]$ and partial tuple $\mathbf{s}=(s_1,\dots,s_k)\in\prod_{i\in[k]}S_i$, let*

$$p_{\mathbf{s}}=\sum_{(s_{k+1},\dots,s_m)\in\prod_{i>k}S_i} p_{(s_1,\dots,s_m)} \quad \text{and} \quad p_\varnothing=\sum_{\mathbf{s}\in S} p_{\mathbf{s}}.$$

We say that the polynomials $\{p_{\mathbf{s}}:\mathbf{s}\in S\}$ form an interlacing family if for each $0\le k\le m-1$ and $(s_1,\dots,s_k)\in\prod_{i\in[k]}S_i$, the polynomial $\{p_{s_1,\dots,s_k,t}:t\in S_{k+1}\}$ have a common interlacing.

With this definition, we can prove the following theorem.

Theorem 4.3. *Let $S_1,\dots,S_m$ be finite sets and let $\{p_{\mathbf{s}}:\mathbf{s}\in\prod_{i\in[m]}S_i\}$ be an interlacing family of polynomials. Then for every $0\le k\le m$, there exist $(s_1,\dots,s_k)\in\prod_{i\in[m]}S_i$ such that the largest root of $p_{s_1,\dots,s_k}$ is upper bounded by the largest root of $p_\varnothing$.*

Proof. When $k=0$, the statement is trivial as the largest root of $p_\varnothing$ is at most the largest root of $p_\varnothing$.

Assume that there exists $0 \le k \le m-1$ such that the largest root of $p_{s_1,\dots,s_k}$ is at most the largest root of $p_\varnothing$. Then we have

$$p_{s_1,\dots,s_k} = \sum_{t \in S_{k+1}} p_{s_1,\dots,s_k,t}.$$

Since $\{p_{s_1,\dots,s_k,t} : t \in S_{k+1}\}$ has a common interlacing, there exists $s_{k+1} \in S_{k+1}$ such that the largest root of $p_{s_1,\dots,s_{k+1}}$ is at most the largest root of $p_{s_1,\dots,s_k}$, which is at most the largest root of $p_\varnothing$ by the induction hypothesis. □

As a corollary, we obtain the following.

Corollary 4.4. *Let $\mathcal{D}_1,\dots,\mathcal{D}_m$ be independent probability distribution with the outcome of $\mathcal{D}_i$ in S_i. Let $\mathcal{P} = \{p_{s_1,\dots,s_m} : s_i \in S_i\}$ be an interlacing family. Then there exists $(s_1,\dots,s_m) \in \prod_{i\in[m]} S_i$ such that the largest root of $p_{s_1,\dots,s_m}$ is at most the largest root of*

$$\mathbb{E}_{s_i\sim\mathcal{D}_i}[p_{s_1,\dots,s_m}(x)].$$

Proof of this corollary is straightforward. Let $s \sim \mathcal{D}_1 \times \cdots \times \mathcal{D}_m$. Note that the $\mathbb{P}[s = (s_1,\dots,s_m)] \cdot p_{s_1,\dots,s_m}$ has the same root as $p_{s_1,\dots,s_m}$. So, the summation can be converted into weighted average, so direct application of the previous theorem yields this corollary. Expression the interlacing families in terms of probability distribution makes analysis more convenient.

Together with Theorem 3.10, the interlacing condition on the above $\mathcal{P}$ can be restated as follows.

Corollary 4.5. *Let $\mathcal{D}_1,\dots,\mathcal{D}_m$ be independent probability distribution with the outcome of $\mathcal{D}_i$ in S_i. Let $\mathcal{P} = \{p_{s_1,\dots,s_m} : s_i \in S_i\}$. If*

$$\mathbb{E}_{x_i\sim\mathcal{D}_i}[p_{x_1,\dots,x_m}(x)]$$

is real-rooted for any independent probability distribution $\mathcal{D}_i$ on S_i, then there exists $(s_1,\dots,s_m) \in \prod_{i\in[m]} S_i$ such that the largest root of $p_{s_1,\dots,s_m}$ is at most the largest root of

$$\mathbb{E}_{s_i\sim\mathcal{D}_i}[p_{s_1,\dots,s_m}(x)].$$

Note that $\mathbb{E}_{\substack{x_i\sim\mathcal{D}_i \\ \text{for } i>k}}[p_{s_1,\dots,s_k,x_{k+1},\dots,x_m}(x)]$ can be expressed as $\mathbb{E}[p_{x_1,\dots,x_m}(x)]$ where x_i is s_i with probability 1 for $i \le k$, and $x_i \sim \mathcal{D}_i$ for $i > k$. So, the condition on the above corollary ensures that every convex combination of the polynomials on the children of any internal vertex in the interlacing tree is real-rooted. Hence, Theorem 3.10 implies that they form an interlacing family.

4.2. **2-covers.** Bilu and Linial in 2006 used the idea of lifting a graph to construct large expanders. This lifting can be formalized by the following notion of covers.

Definition 4.6. *Let $G = (V,E)$ be a graph with $|E| = m$. Let $s : E \to \{-1,+1\}$ be a signing of the edges of G. Then the 2-cover G_s of G associated with s is the graph $G_s = (V_s,E_s)$ whose vertices and edges can be labeled as $V_s = \{u_L,u_R : u \in V\}$ and $E_s = \{(u_L,v_L)(u_R,v_R) : uv \in E, s(uv) = 1\} \cup \{(u_L,v_R)(u_R,v_L) : uv \in E, s(uv) = -1\}$.*

It is easy to see that when G is a bipartite graph, G_s is also a bipartite graph for any signing s. It is known that the eigenvalues of a 2-cover G_s can be related to the eigenvalues of the base graph G.

Lemma 4.7. *Let G be a graph and A be its adjacency matrix, and A_s be the signed adjacency matrix of A with respect to a signing s of the edges. Let G_s be the 2-cover of G and let B be its adjacency matrix. Then every eigenvalue of A and every eigenvalue of A_s is an eigenvalue of B. Furthermore, the multiplicity of each eigenvalue of B is the sum of its multiplicities in A and A_s.*

The eigenvalues of A are called 'old eigenvalues' and the eigenvalues of A_s are called 'new eigenvalues'. It was conjectured by Bilu and Linial that every for every d-regular graph G, one can find a signing s where all new eigenvalues of G_s are in $[-2\sqrt{d-1}, 2\sqrt{d-1}]$. We will prove this conjecture for bipartite graph G. This will allow us to obtain another d-regular Ramanujan graph on $2n$ vertices from a d-regular Ramanujan graph on n vertices.

In order to find a desired signing, we will prove the following.

- $\{\det(xI - A_s)\}_{s\in\{\pm 1\}^m}$ forms an interlacing family.
- the maximum root of $\sum_s \det(xI - A_s)$ (equivalently $\mathbb{E}_s[\det(xI - A_s)]$) is at most $2\sqrt{d-1}$.

Note that this provides an upper bound on the maximum root, but not a lower bound on the smallest root. However, if the original graph we start with is bipartite, then upper bound on the smallest root automatically imply the lower bound as the eigenvalues of bipartite graphs are symmetric around zero. Hence, this proves what we want.

We will do the second job first. Consider the matching polynomial $M_G(x)$ of G defined by

$$M_G(x) = \sum_{0\le k\le n/2} (-1)^k m_k x^{n-2k},$$

where m_k is the number of matchings with k edges in G with $m_0 = 1$. We can prove the following theorem.

Theorem 4.8. *When $s \in \{\pm 1\}^m$ is chosen uniformly at random, we have*

$$\mathbb{E}[\det(xI - A_s)] = M_G(x).$$

Proof. For a permutation $\sigma \in S_n$, let $\mathrm{Fix}(\sigma)$ be the set of fixed points of σ and let $\mathrm{fix}(\sigma) = |\mathrm{Fix}(\sigma)|$. We know that

$$\det(xI - A_s) = \sum_{\sigma\in S_n} \mathrm{sign}(\sigma) \prod_{i=1}^{n} (xI - A_s)_{i,\sigma(i)} = \sum_{\sigma\in S_n} \mathrm{sign}(\sigma) x^{\mathrm{fix}(\sigma)} \prod_{i\notin \mathrm{Fix}(\sigma)} A_s(i,\sigma(i)).$$

So, we have

$$\mathbb{E}[\det(xI - A_s)] = \sum_{\sigma\in S_n} \mathrm{sign}(\sigma) x^{\mathrm{fix}(\sigma)} \mathbb{E}\left[\prod_{i\notin \mathrm{Fix}(\sigma)} A_s(i,\sigma(i))\right].$$

Here, we know that $\mathbb{E}[S(i,\sigma(i))] = 0$ for every i where $\sigma(i) \neq i$. So, if the term inside the expectation is linear in terms of $A_s(i,\sigma(i))$, then this contributes zero to the computation. Only possible way to make a nonzero contribution is when $i\sigma(i) \in E$ and $\sigma(\sigma(i)) = i$. In this case, we know that $S(i,\sigma(i))S(\sigma(i),i) = 1$. Thus the only permutations that count are the involutions. As an involution with $n-2k$ fixed points have a sign $(-1)^k$, we have

$$\mathbb{E}[\det(xI - A_s)] = \sum_{\sigma\in S_n} \mathrm{sign}(\sigma) x^{\mathrm{fix}(\sigma)} \mathbb{E}\left[\prod_{i\notin \mathrm{Fix}(\sigma)} A_s(i,\sigma(i))\right] = \sum_k (-1)^k m_k x^{n-2k}$$

□

On the other hand, it is known that the following theorem holds. We omit the proof.

Theorem 4.9 (Godsil 93)**.** *Let G be a graph with the maximum degree at most d. Then $M_G(x)$ is real-rooted and the largest root is at most $2\sqrt{d-1}$.*

Now, we prove that the polynomials $\{\det(xI - A_s)\}_{s\in\{\pm 1\}^m}$ forms an interlacing family. Choose an ordering on the m edges of the graph. We can now associate $s \in \{\pm 1\}^k$ with $p_s = \mathbb{E}_{s'\sim\{\pm 1\}^{m-k}}[\chi(A_{s,s'})]$.

Theorem 4.10. *Let $v_1, \dots, v_k$ be independently distributed random n-dimensional vectors and let A be a symmetric n-dimensional matrix. Then the polynomial*

$$\mathbb{E}\left[\chi(A + \sum_{i\in[k]} v_i v_i^T)\right]$$

is real-rooted. Moreover, for every vector u in the support of v_k, all polynomials

$$\mathbb{E}\left[\chi(A + uu^T + \sum_{i\in[k-1]} v_i v_i^T)\right]$$

have a common interlacing

Proof. We prove this by induction on k. Assuming that we have proved it for k, we now prove it for $k+1$. Let u be any vector and $t \in \mathbb{R}$. Let $p_t(x) = \mathbb{E}[\chi(A+tuu^T+\sum_{i\in[k]} v_i v_i^Y)]$. By Lemma 3.6, we have $p_t(x) = p_0(x) - tq(x)$ for some polynomial q of degree $n-1$. By induction hypothesis applied to $A + tuu^T$, the polynomial $p_t(x)$ is real-rooted for all t. Thus Lemma 3.8 implies that $q(x)$ interlaces $p_0(x)$.

On the other hand, Lemma 3.11 tells that $p_0(x)$ interlaces $p_t(x)$. So, we may conclude that for any $u \in \mathbb{R}^n$,

$$\mathbb{E}[\chi(A + \sum_{i\in[k]} v_i v_i^T)] \quad \text{interlaces} \quad \mathbb{E}[\chi(A + uu^T + \sum_{i\in[k]} v_i v_i^T)].$$

Now, choose u from the support of v_{k+1} and apply Theorem 3.7 to conclude that (if p interlaces q_i where all has positive leading coefficients, then p also interlaces a convex combination of q_i)

$$\mathbb{E}[\chi(A + \sum_{i\in[k]} v_i v_i^T)] \quad \text{interlaces} \quad \mathbb{E}[\chi(A + v_{k+1}v_{k+1}^T + \sum_{i\in[k]} v_i v_i^T)]$$

and that the latter polynomial is real-rooted. □

For every edge ab of G, let $v_{a,b}$ be the random vector that is $\mathbf{1}_a - \mathbf{1}_b$ with probability $1/2$ and $\mathbf{1}_b + \mathbf{1}_a$ with probability $1/2$. Then the random matrix A_s is distributed according to

$$\sum_{ab\in E} v_{a,b} v_{a,b}^T - dI.$$

Subtracting dI shifts the roots by d, and so does not impact any results we have proved about interlacing of real-rootedness. Also the expectation of the characteristic polynomial is the matching polynomial, whose largest eigenvalue is at most $2\sqrt{d-1}$. This shows that for a given d-regular bipartite Ramanujan graph on n vertices, we can find another d-regular bipartite Ramanujan graph on $2n$ vertices.

However, if we also want to better control the number of vertices in such a Ramanujan graph, then we can generalize the 2-cover to an r-cover. As ± 1 can be identified with the permutation S_2, we can consider a choice of permutation in S_r. Using this, we can find an n-vertex d-regular Ramanujan graph for all d and all n which has only small prime factors.

One shortcoming of this proof is that it does not yield an explicit graph and does not yield an efficient algorithm.

5. Bipartite Ramanujan graphs via unions of matchings

Another way of proving the existence of Ramanujan graphs is to take a union of d random perfect matchings. This was done by Marcus, Srivastava, Spielman in 2015. Of course, to obtain a bipartite graph, we have to take a bipartition $A \cup B$ and take random perfect matching between A and B. However, for simplicity, we just take d random perfect matchings, and show that the second eigenvalue is smaller than $2\sqrt{d-1}$. These analysis can be extended to obtain a bipartite Ramanujan graph. The result in this section comes from [6].

5.1. **Interlacing.** We first show the following theorem. In the end, we want to keep adding $\Pi M \Pi^T$ to our matrix where M is an adjacency matrix of a perfect matching and Π is a random permutation matrix. For this, let

$$p_{\Pi_1,\dots,\Pi_d}(x) = \chi(\sum_{i\in[d]} \Pi_i M \Pi_i^T).$$

We wish to show that these polynomials form an interlacing family. We take a detour so that we can use more elementary techniques. We will break the one choice of Π_i into many simpler choices so that it would be easier to analyze.

For $i, j \in [n]$, let $\Gamma_{i,j}$ be a permutation matrix swapping i and j. Let S be a random matrix that is equal to $\Gamma_{i,j}$ with probability s and equal to the identity with probability $1-s$. We call this a random swap. Let $\sigma : S_n \to \mathbb{R}_{\geq 0}$ be a probability distribution over the permutations and consider the sum $\sum_{i\in[m]} \sum_{\Pi} \sigma(\Pi)\chi(A_i + \Pi B \Pi^T)$. Let σ be realizable by swaps if this distribution can be achieve by multiplication of random swaps.

We will choose some random swaps $X_1^1, \dots, X_k^1, X_1^2, \dots, X_k^d$. Fix n by n matrices $M_1, \dots, M_d$. For each $B_b^a \in \mathrm{supp}(X_b^a)$, we let

$$p_{B_1^1,\dots,B_k^d}(x) = \chi(\sum_{i\in[d]} (B_k^i \dots B_1^i) M_i (B_k^i \dots B_1^i)^T.$$

In order to be able to apply Corollary 4.5 to these polynomials, it suffices to prove the following.

Lemma 5.1. *For any symmetric matrices $M_1, \dots, M_d$ and for any random swaps $S_1^1, \dots, S_{k_1}^1, \dots, S_{k_d}^d$, the polynomials*

$$\mathbb{E}[p_{S_1^1,\dots,S_{k_d}^d}(x)]$$

is real-rooted.

The following lemma allows us to compute the expectation when we multiply a random swap to a given matrix on both side.

Lemma 5.2. *Let A be a symmetric matrix. Then for all i and j, there exists vectors u and v such that*

$$\Gamma_{i,j} A \Gamma_{i,j}^T = A - uu^T + vv^T.$$

Proof. We may assume $i = 1, j = 2$. As $A - \Gamma_{i,j} A \Gamma_{i,j}^T$ is symmetric, it suffices to prove that it has rank 2. It is easy to see that such a matrix has form

$$\begin{pmatrix} \alpha & \beta & \mathbf{y}^T \\ -\beta & -\alpha & -\mathbf{y}^T \\ \mathbf{y} & -\mathbf{y} & 0 \end{pmatrix}$$

If $\alpha \neq \beta$, then every row is a linear combination of the first two rows. Otherwise, then the first two rows are linearly dependent and all other rows are in the span of $(1, -1, 0, \dots, 0)$. Hence, it has rank 2. □

In order to prove what we want, we define two terminologies.

Definition 5.3. *Let $p(A_1, \dots, A_d)$ be a degree n homogeneous polynomial of the entries of the matrices as variables. We call this polynomial $p(A_1, \dots, A_d)$ be 'good' if $p(xI - A_1, \dots, xI - A_d)$ is real-rooted for any symmetric matrices $A_1, \dots, A_d$. We call a degree n polynomial $p(Y_1, \dots, Y_d)$ be 'nice' if for any t and $v \in \mathbb{R}^n$ and $i \in [d]$, $p(xI - A_1, \dots, xI - A_{i-1}, xI - A_i - tvv^T, xI - A_{i+1}, \dots, xI - A_d) = p(xI - A_1, \dots, xI - A_d) - tq(x)$ holds for some polynomial q of degree $n-1$ with a positive leading coefficient.*

The 'goodness' is referred as the *hyperbolicity* of the polynomial $p(xI - A_1, \dots, xI - A_d)$ with respect to $(I, I, \dots, I)$. For more on this concept, see [10]. The 'niceness' is essentially the *rank*-1 *linearity*.

Note that by Lemma 3.6, $p(A_1, \dots, A_d) = \det(A_1 + \cdots + A_d)$ is a polynomial satisfying both of the above two properties. We show the following.

Lemma 5.4. *Let X be an arbitrary random swap. If the polynomial $p(A_1, \dots, A_d)$ is good and nice, then $\mathbb{E}_X[p(A_1, \dots, A_{i-1}, XA_{i-1}X^T, A_{i+1}, \dots, A_d)]$ is also good and nice.*

Proof. Let X be $\Gamma_{a,b}$ with probability s and I with probability $1-s$. Then as $\Gamma_{a,b}A_i\Gamma_{a,b}^T = A + -uu^T + vv^T$, we have

$$\begin{aligned}&\mathbb{E}_X[p(A_1, \dots, A_{i-1}, XA_{i-1}X^T, A_{i+1}, \dots, A_d)]\\ &= s \cdot p(A_1, \dots, A_i - uu^T - vv^T, A_d) + (1-s)p(A_1, \dots, A_d).\end{aligned}$$

As this is a convex combination of two nice polynomials, it is easy to see that it is also nice.

Fix an arbitrary vector v and u. Let $r_t(x) = p(xI - A_1, \dots, xI - A_i - tvv^T, \dots xI - A_d)$. As p is nice, there exists a degree $n-1$ polynomial $q(x)$ with positive leading coefficient such that

$$r_t(x) = r_0(x) - tq(x)$$

where both r_0 and q has positive leading coefficients. As p is good, we know that $r_t(x)$ is real-rooted for all $t \in \mathbb{R}$. So, Lemma 3.8 implies that $q(x)$ interlaces $r_0(x)$. Lemma 3.11 now implies that $r_0(x)$ interlaces $r_1(x)$.

By the same argument, we know that $p(xI - A_1, \dots, xI - A_i + uu^T - vv^T, \dots, xI - A_d)$ interlaces $r_1(x)$.

Hence, both $p(xI - A_1, \dots, xI - A_i, \dots, xI - A_d)$ and $p(xI - A_1, \dots, xI - A_i + uu^T - vv^T, \dots, xI - A_d)$ interlaces $r_1(x)$, so they have common interlacing. As for certain u and v, we have $\Gamma_{a,b}A_i\Gamma_{a,b}^T = A_i - uu^T + vv^T$, by Theorem 3.7, any convex combination of them is real-rooted, this shows that the polynomial $\mathbb{E}_X[p(A_1, \dots, A_{i-1}, XA_{i-1}X^T, A_{i+1}, \dots, A_d)]$ is good. □

Hence, repeatedly applying this shows that $\mathbb{E}[\Pi_1 X_1 \Pi_1^T, \dots, \Pi_d X_d \Pi_d^T]$ is real-rooted where each Π_i is a random matrix with distributions realizable by swaps.

Lastly we can show that a uniform random permutation can be generated from random swaps.

Lemma 5.5. *For every n, there exists a finite sequence of random swaps $S_1, \dots, S_k$ so that $S_1 \dots S_k$ is a uniform random permutation.*

Proof. We use induction on n. When $n = 1$, it is trivial. By induction hypothesis, we have $S_1, \dots, S_k$ so that we have $S_1 \dots S_k$ is a uniform random permutation on $[n-1]$ which fixes n. Let S be a random swap that swaps 1 and n with probability $1 - 1/n$. Then it is easy to see that

$$S_1 \dots S_k S S_1 \dots S_k$$

has the uniform distribution on S_n. □

Combining these results, we obtain that there exists a choice of $\Pi_1, \dots, \Pi_d$ where the second largest eigenvalue of $p_{\Pi_1, \dots, \Pi_d}$ is at most the second largest eigenvalue of $\mathbb{E}[p_{\Pi_1, \dots, \Pi_d}]$.

5.2. **Quadrature for the finite free convolution.** We want to analyze what $\mathbb{E}_{\Pi\in S_n}[\chi(A+\Pi B\Pi^T)]$ is. This finite sum can be expressed by using the following integral.

Definition 5.6. *Let $p(x), q(x)$ are characteristic polynomials of A and B. Then their finite free convolution is the polynomial*

$$p(x)\boxplus_n q(x) := \mathbb{E}_{Q\in\mathcal{O}(n)}[\chi(A+QBQ^T)].$$

Here, $Q\in\mathcal{O}(n)$ is a random orthonormal matrix chosen according to the Haar measure. The Haar measure here is the probability distribution over $\mathcal{O}(n)$ that is invariant under group operations, which are multiplication by orthonormal matrices. Note that the expectation above does not depend on the eigenvectors of A and B, but only depends on the eigenvalues (thus the characteristic polynomials). To see this, recall that one can find an orthonormal matrix R mapping an orthonormal basis of A to another orthonormal basis of A' with the same eigenvalues, so that $A=RA'R^T$. As $Q\in\mathcal{O}(n)$ is chosen according to the Haar measure, Q and QT have the same distribution. So

$$\mathbb{E}[\chi(A+QBQ^T)] = \mathbb{E}[\chi(RA'R^T+RQB(RQ)^T)] = \mathbb{E}[\chi(A'+QBQ^T)].$$

Similarly, B can be replaced with B' having the same eigenvalues. With this, we know the following. We will omit the proofs.

Theorem 5.7. *Let A and B be symmetric matrices with constant row sums. If $A\mathbf{1}=a\mathbf{1}$ and $B\mathbf{1}=b\mathbf{1}$, then we know that the characteristic polynomials are*

$$\chi(A)=(x-a)p(x) \quad \textit{and} \quad \chi(B)=(x-b)q(x).$$

Then we have

$$\mathbb{E}_{\Pi\in S_n}[\chi(A+\Pi B\Pi^T)] = (x-a-b)(p(x)\boxplus_{n-1} q(x)).$$

Moreover,

Theorem 5.8. *Let $p(x)=\sum_{i=0}^n(-1)^i a_i x^{n-1}$ and $q(x)=\sum_{i=0}^n(-1)^i b_i x^{n-i}$. Then*

$$p(x)\boxplus_n q(x) = \sum_{k=0}^n x^{n-k}(-1)^k \sum_{i+j=k}\frac{(n-i)!(n-j)!}{n!(n-i-j)!}a_i b_j.$$

5.3. **An upper bound on the roots.** Now, we need to prove that the polynomial

$$\mathbb{E}_\Pi[\chi(A+\Pi B\Pi^T)]$$

has its root bounded. For given polynomial $p(x)=\prod_{i\in[d]}(x-\lambda_i)$, let

$$\mathcal{G}_p(x) = \frac{1}{d}\sum_{i=1}^d\frac{1}{x-\lambda_i} = \frac{p'(x)}{d\cdot p(x)}$$

and let

$$\mathcal{K}_p(w) = \max\{x : \mathcal{G}_p(x)=w\}$$

be the inverse Cauchy transform. This is similar to the Barrier from the Sparsifier as we have the following.

Theorem 5.9. *For polynomials $p(x)$ and $q(x)$ of degree n and for $w>0$, we have*

$$\mathcal{K}_{p\boxplus_n q}(w) \le \mathcal{K}_p(w)+\mathcal{K}_q(w)-\frac{1}{w}.$$

Combining this, we know that for $p(x)=\frac{1}{x-1}\chi(M)=(x-1)^{n/2-1}(x+1)^{n/2}$, we have

$$\mathcal{G}_p(x) = \frac{1}{n-1}\left(\frac{n/2-1}{x-1}+\frac{n/2}{x+1}\right) \le \frac{x}{x^2-1} = \mathcal{G}_{\chi(M)}(x)$$

where M is the adjacency matrix of a perfect matching. Also, we have

$$\mathcal{K}_p(w) \leq \mathcal{K}_{\chi(M)}(w)$$

for all $w \geq 0$. Then the above theorem yields that

$$\mathcal{K}_{p\boxplus\cdots\boxplus p}(w) \leq d\mathcal{K}_p(w) - \frac{d-1}{w} \leq d\mathcal{K}_{\chi(M)}(w) - \frac{d-1}{w}$$

is an upper bound on the largest root of $p \boxplus \cdots \boxplus p$. Let $w = \frac{\sqrt{d-1}}{d-2}$, then $K_{\chi(M)}(w) = \sqrt{d-1}$. Then we have

$$d\mathcal{K}_p(w) - \frac{d-1}{w} \leq d\sqrt{d-1} - (d-1)\frac{d-2}{\sqrt{d-1}} = d\sqrt{d-1} - (d-2)\sqrt{d-1} = 2\sqrt{d-1}.$$

Note that Theorem 5.7 provides a polynomial-time method for computing the expected characteristic polynomials.

However, this is not enough to obtain a polynomial time algorithm to construct a bipartite Ramanujan graph.

In each time, we can $n!$ choices of the permutation. For each permutation, we can compute how the root of expected characteristic polynomial changes by the choice and choose one permutation, and repeat d times. However, this yields $O(dn^n)$ time algorithm. But, instead, one can choose an edge by edge rather than choosing one perfect matching. One can show that this also yields an interlacing family with the same expectation. As we only have to check n possibilities each time, and repeat dn times, this yields an efficient algorithm. See [2] for more.

References

[1] J. Batson, D. Spielman, and N. Srivastava. Twice-ramanujan sparsifiers. SIAM Review, 56(2):315–334, 2014.

[2] M. Cohen, Ramanujan graphs in polynomial time, 2016 IEEE 57th Annual Symposium on Foundations of Computer Science, arXiv:1604.03544

[3] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bulletin of the Amererican Mathematical Society 43(4), 2006.

[4] K. Liu, the method of interlacing polynomials, https://homes.cs.washington.edu/liukui17/

[5] A. Marcus, D. Spielman, and N. Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees. Ann. of Math., 182-1:307–325, 2015.

[6] A. Marcus, D. Spielman, and N. Srivastava. Interlacing families IV: Bipartite Ramanujan graphs of all sizes, SIAM Journal on Computing 47(6):2488-2509

[7] C. Hall, D. Puder, and W. F Sawin, Ramanujan coverings of graphs, Advances in Mathematics 323, 367-410, 2018.

[8] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," in IEEE Signal Processing Magazine, vol. 30, no. 3, pp. 83-98, May 2013, doi: 10.1109/MSP.2012.2235192.

[9] D. Speilman, Spectral and Algebraic Graph Theory, http://cs-www.cs.yale.edu/homes/spielman/

[10] R. Pemantle, Hyperbolicity and stable polynomials in combinatorics and probability, arXiv:1210.3231