EXPANDER GRAPHS

Contents

1 Overview of the Talk

What are expander graphs and why are they important? The goal of this talk is to obtain intuition about this interesting psuedorandom object, prove a few interesting properties, to demonstrate some examples, and also to see their application to randomized algorithms. At the end, we will briefly cover their relationship to randomness extractors, and say a few words about how point-line incidence theorems are involved in their analysis.

2 Basic Definitions

2.1 Graph Theory

Definition 2.1. A graph.
A graph $G = (V, E)$, where V is the vertex set and E is the edge set. We will denote $|V| = n$, $|E| = m$. Often, we want to represent a graph as a matrix A. Typically, we will use the adjacency matrix formalism: The rows and columns will be both of size n, and each entry $A(i, j) = w_{ij}$, the weight from vertex i to vertex j. If $i = j$, this indicates a self loop. If the graph G is undirected, then the bottom-left half of the matrix A is zero.

Definition 2.2. d-regular graphs.
If every vertex of a graph G has the same degree (number of edges) d, then G is called d-regular.

Definition 2.3. A neighborhood.
A neighborhood of a set S in a graph G is given by $N(S) = \{u | \exists v \in S \text{ s.t. } (u, v) \in E\}$. In words, a neighborhood is every point that has at least one edge to S.

Definition 2.4. A graph cut.
The cut-set of two sets of vertices $S, T \subseteq G$ is given by $e(S, T)$, the set of edges crossing between S and its complement T. Explicitly,

$$e(S, T) = \{(u, v) \in S \times T | (u, v) \in E\}$$

Often, we will take T to be $\bar{S}$, the complement of S (the vertices in the graph not in S).

2.2 Probability

Definition 2.5. Probability distribution.
A probability distribution over a set S defines a function $\phi : S \to [0, 1]$ such that $\int_S \phi(x)dx = 1$ in the continuous case, or $\sum_S \phi(x) = 1$ in the discrete case.

Definition 2.6. Support of a probability distribution.
The support of π is Supp(π) $= \{x : \pi_x > 0\}$.

Definition 2.7. Markov chain.
A Markov chain is a memoryless random process. The next state is only dependent on the current state. For a chain with n states, we can define a transition matrix $P \in \mathcal{M}(n, n)$ such that each column sums to 1 - there is a probability distribution on next states defined for each state. Notably, for d-regular graphs, we can define a transition matrix $P = \frac{1}{d}A$, where A is the adjacency matrix of the graph, since d is the total number of actions for any given node. In a graph, typically a starting vertex must be chosen. The resulting distribution and properties are defined based on the starting vertex. There is also an initial distribution over the vertices, indicating how likely each of them is to be reached in the next step from the starting vertex.

Definition 2.8. Stationary distribution.
Some Markov chains can reach a distribution that is stable. That is, further applications of the transition matrix do not change the distribution on the vertices. A stationary distribution π satisfies $P\pi = \pi$ for transition matrix P - note this means that P must have an eigenvalue of 1 for which the eigenvector is the stationary distribution. Sometimes, the stationary distribution is uniform (as it will be a few times in this talk).

Definition 2.9. Mixing time.
Informally, this refers to the amount of time it takes for a Markov chain to 'mix', or get close enough to a stationary distribution.

Definition 2.10. Statistical distance.
We define the statistical distance between two random variables X, Y with finite and identical range as $\frac{1}{2}|X - Y|_1 = \frac{1}{2}\sum_{w \in \text{range}(X)} |\mathbf{P}\{X = w\} - \mathbf{P}\{Y = w\}|$.

Definition 2.11. Min-entropy.
We define the min-entropy of a random variable X as

$$\sup\{k | \forall x \in \text{range}(X), \mathbf{P}\{X = x\} \leq 2^{-k}\}$$

If X has no randomness, the min-entropy is 0. If X has finite range with n values and is uniform, then the min-entropy is n.

3 Definitions of Expanders

We will write down the three definitions on one board for convenience, and keep it available to look at as we go through the proofs.

Informally, expander graphs are sparse and well-connected. We interpret these properties in the asymptotic sense: that is, we consider an infinite family of graphs $\{G_i\}_{i=1}^{\infty}$. We have that $|V_i|$ is increasing without bound. However, the sparsity condition requires that the degree d_i grows slowly as a function of $|V_{g_i}|$. Typically when we refer to an expander, the degree d is actually constant.

What we are saying is the following: Consider an infinite family of graphs such that for each graph in a sequence, the degree of the graph is increasing very slowly as a function of the increasing number of vertices (sparsity property). Often, we actually just assume that the degree is constant as a function of the number of vertices. Then, for each graph in this sequence, a well-connectedness property holds, and we can define this well-connectedness property in different ways.

The well-connectedness aspect of expander graphs can be defined through various properities of a graph, namely vertex, edge, and spectral properties. Most of these definitions can be shown to be equivalent for appropriate choices of parameters. We will proceed to define the three most common definitions as well as do some work towards showing the equivalence of the vertex-spectral expansion definitions. It is also important to note that the following definitions are valid for directed multigraphs; however, in this talk, we will focus on undirected multigraphs.

3.1 Vertex Expansion

Definition 3.1. Vertex expander.
A graph G is a (k, a)-vertex expander if $\forall$ sets S such that $|S| \leq k$, we have $|N(S)| \geq a \cdot |S|$.

We now state that good expander graphs exist.

Theorem 3.2. *Existence of good expanders.*
For all constants $d \geq 3$, there is a constant $\alpha > 0$ such that for all n, a random d-regular undirected graph on n vertices is a $(\alpha n, d - 1.01)$-vertex expander.

Instead of proving this theorem, we will prove a slightly easier theorem for a special case. First we define a bipartite expander.

Definition 3.3. Bipartite vertex-expander.
A bipartite multigraph G is a (k, a)-vertex expander if for all sets S_{left} of size at most k, we have $|N(S_{left})| \geq a \cdot |S_{left}|$. Here, S_{left} denotes a set of vertices that is a subset of the left part of the bipartite graph.

The first proof we give demonstrates that 'good' bipartite expanders even exist. This proof is not constructive, it merely demonstrates existence. Later, we will talk about explicit constructions of expander graphs, which are useful in application.

First, we will need a short lemma to help with bounding.

Lemma 3.4. $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$

Proof. First note that $k \leq n$.

$$\begin{aligned}\binom{n}{k} \leq \sum_{i=1}^{k} \binom{n}{i} &= \left(\frac{n}{k}\right)^k \sum_{i=1}^{k} \binom{n}{i} \left(\frac{k}{n}\right)^k \\ &\leq \left(\frac{n}{k}\right)^k \sum_{i=1}^{n} \binom{n}{i} \left(\frac{k}{n}\right)^i \\ &\leq \left(\frac{n}{k}\right)^k \left(1+\frac{k}{n}\right)^n \\ &\leq \left(\frac{n}{k}\right)^k \left(e^{\frac{k}{n}}\right)^n = \left(\frac{ne}{k}\right)^k\end{aligned} \tag{1}$$

□

Theorem 3.5. *Existence of good bipartite expanders.*
For any d, there exists $\alpha_d > 0$ such that for all n,

$$\boldsymbol{P}\{G \text{ is a } (\alpha_d n, d-2) - \text{expander}\} \geq \frac{1}{2}$$

where G is chosen uniformly from $\text{Bipartite}_{d,n}$. *Note that this is equivalent to uniformly and independently choosing d neighbors on the right for each left-vertex v.*

Proof. First, we give the main idea of the proof. Interestingly, this proof due to Pinsker in 1973 was one of the first applications of the probabilistic method. The main idea is that we bound the probability that the statement does not hold. If we can bound the probability below $\frac{1}{2}$, we are done. Then, having shown that the probability that we can find a good bipartite expander is bounded below by a non-zero value, we are guaranteed existence. If there were no such bipartite expander, the probability would be necessarily zero.

We now proceed to bound this probability. First we define probability p_k as the probability that there exists a vertex set S of size k on the left portion of the bipartite graph that does not expand by at least $d-2$. We fix a set S such that $|S| = k$, and we have $N(S)$ is a set of $k \cdot d$ random vertices on the right hand side of the graph. We can think of the members of $N(S)$ as an ordered list of $k \cdot d$ vertices, selected randomly with replacement. In order to violate $d-2$ expansion for a specific k, the number of vertices in $N(S)$ must be less than or equal to $k \cdot (d-2) = k \cdot d - 2k$. Therefore, the number of repeats in the list must be at least $2k$. The probability that a given element v_i in the list is a repeat of the previous $i-1$ elements is at most $\frac{i-1}{n} \leq \frac{k \cdot d}{n}$.

We have

$$\begin{aligned}\mathbf{P}\{|N(S)| \leq k \cdot (d-2)\} &\leq \mathbf{P}\{\text{at least } 2k \text{ repeats in the list}\} \\ &\leq \binom{k \cdot d}{2k} \left(\frac{k \cdot d}{n}\right)^{2k}\end{aligned} \tag{2}$$

Therefore,

$$\begin{aligned} p_k &\leq \binom{n}{k}\binom{k\cdot d}{2k}\left(\frac{k\cdot d}{n}\right)^{2k} \\ &\leq \left(\frac{ne}{k}\right)^k \left(\frac{kde}{2k}\right)^{2k}\left(\frac{k\cdot d}{n}\right)^{2k} \\ &= \left(\frac{e^3d^4k}{4n}\right)^k \end{aligned} \tag{3}$$

Recall that we had assumed $k \leq \alpha_d n$, therefore we can choose $\alpha_d = \frac{1}{e^3d^4}$, resulting in $p_k \leq 4^{-k}$. Now we can bound the whole probability that G is not a bipartite $(\alpha_d n, d-2)$-expander.

$$\mathbf{P}_{G\in \text{Bip}_{n,d}}\{G \text{ is not an } (\alpha_d n, d-2) - \text{expander}\} \leq \sum_{k=1}^{\lfloor \alpha_d n \rfloor} 4^{-k} < \frac{1}{2} \tag{4}$$

and we have our result. □

Remark 3.6. We will comment that a general tradeoff between d, α, and a is known in terms of entropy. A random d-regular n vertex graph bipartite multigraph is known to be an $(\alpha n, a)$-vertex expander with high probability for sufficiently large n if

$$d > \frac{H(\alpha) + H(\alpha a)}{H(\alpha) - \alpha a H(\frac{1}{a})}$$

where $H(p) = -\left(p\text{log}(p) + (1-p)\text{log}(1-p)\right)$ is the binary entropy function.

Remark 3.7. Note that this proof only gives existence, and not an explicit construction. One desirable goal is to construct expanders with low degree explicitly. We prefer low degree graphs because larger degree graphs have larger size, which leads to large constants in algorithms of low asymptotic complexity and makes them less practical in practice.

3.2 Edge Expansion

We will just give the definition, and a statement of a theorem of equivalence later on.

Definition 3.8. Edge expander.
A d-regular graph G is a (k, ϵ)-edge expander if for all sets S such that $|S| \leq k$, the cut size $|e(S, \bar{S})| \geq \epsilon \cdot |S| \cdot d$.

That is, essentially ϵ of the edges from S lead outside S. We can view this definition in terms of random walks. If we condition the stationary distribution on being inside S, the probability that the walk leaves S in one step is given by $\frac{|e(S,\bar{S})|}{|S|\cdot d}$.

3.3 Spectral Expansion

The purpose of this definition is to represent the well-connectedness property of an expander graph in terms of random walks – we essentially want to say that random walks on the graph converge quickly to the stationary distribution.

Definition 3.9. 2^{nd} Eigenvalue of a graph.
Let G be a d-regular graph with normed adjacency matrix $M = \frac{1}{d}A$, where A is the adjacency matrix. The largest eigenvalue of M is $\lambda_1 = 1$ with eigenvector $u = \left(\frac{1}{|V|}, \cdots, \frac{1}{|V|}\right)$. Then, the second largest eigenvalue $\lambda_2 = \max_{\|x\|=1, x \perp u} \|Mx\|$. (We can think of the normalized eigenvectors as representing orthogonal components - think of it in terms of SVD).

Now consider a probability distribution on the vertices of the graph G, called π, represented as a vector. Note that we can decompose this vector into the sum of two orthogonal vectors. Letting u be the uniform distribution, we can write $\pi = u + \pi^{\perp}$, where $\pi^{\perp}$ is orthogonal to the uniform vector.

An interesting spectral property follows: We can view M as the normalized adjacency matrix as the transition matrix for a Markov chain starting with distribution π. Then we have

$$\begin{aligned} M\pi - u = M(u + \pi^{\perp}) - u &= Mu - u + M\pi^{\perp} \\ &= M\pi^{\perp} \text{ since the eigenvalue of } u \text{ is } 1 \end{aligned} \tag{5}$$

Then we have

$$\begin{aligned} \|M\pi - u\|^2 = \|M\pi^{\perp}\|^2 &\leq \lambda_2^2 \|\pi^{\perp}\|^2 \\ &= \lambda_2^2 \|\pi - u\|^2 \text{ and we have} \\ \|M\pi - u\| &\leq \lambda_2 \|\pi - u\| \end{aligned} \tag{6}$$

From here, we define what spectral expansion means.
First we define the spectral gap.

Definition 3.10. Spectral gap.
The spectral gap of G is given by $\gamma(G) = 1 - \lambda_2(G)$.

Definition 3.11. Spectral Expansion.
A graph G has spectral expansion γ if $\gamma(G) \geq \gamma$, for $\gamma \in [0, 1]$.

Essentially what this means is that there is a constant λ such that $\lambda_2(G) \leq \lambda$ such that at every step of the Markov chain, distance to uniformity shrinks by at least λ. We write spectral expansion in terms of γ to better fit the intuition that a large γ is good for expansion (a small λ is good for expansion). We want a small λ since then, the Markov chain will mix very quickly.

4 Equivalence of Definitions

4.1 Collision Probability

First, we define an important quantity in probability theory.

Definition 4.1. Collision probability.
For probability distribution π, the collision probability CP is defined to be the probability that two independent samples from π are equal. In equation form, this is simply

$$\text{CP}(\pi) = \sum_x \pi_x^2 = \|\pi\|^2$$

Here are a few properties of the collision probability that we will need.

Lemma 4.2. *Properties of collision probability.*
For every probability distribution $\pi \in [0,1]^n$, we have

1. *CP(π) =* $\|\pi\|^2 = \|\pi - u\|^2 + \frac{1}{n}$
2. *CP(π)* $\geq \frac{1}{|\text{Supp}(\pi)|}$

where u is the uniform distribution, and Supp *denotes the support of the distribution.*

Proof. The proofs are as follows:

1. First, write $\pi = u + \pi^\perp$. Then $\text{CP}(\pi) = \|\pi\|^2 = \|u + \pi^\perp\|^2 = \|u\|^2 + \|\pi - u\|^2 = \frac{1}{n} + \|\pi - u\|^2$.
2. Let $z = |\text{Supp}(\pi)|$. Recall that the support is the set of vertices v such that π_v is non-zero. Then recall the Cauchy-Schwarz Inequality: $\mathbf{x} \cdot \mathbf{y} \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$. We will apply it with $\mathbf{x} = \pi_{\text{Supp}(\pi)}$, and $\mathbf{y} = \mathbf{1}_{\text{Supp}(\pi)}$, where the subscript denotes what the indices run over. We have

$$1 = \sum_v \pi_v = \mathbf{x} \cdot \mathbf{y} \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\| = \sqrt{z \cdot \sum_v \pi_v^2} = \sqrt{z \cdot \text{CP}(\pi)}$$

   Therefore, $\frac{1}{|\text{Supp}(\pi)|} \leq \text{CP}(\pi)$.

□

Now, we use these properties to show that spectral expansion actually implies vertex expansion!

4.2 Spectral Expansion $\implies$ Vertex Expansion

Theorem 4.3. *Spectral $\implies$ Vertex.*
If G is a regular graph with spectral expansion $\gamma = 1 - \lambda$ for some $\lambda \in [0,1]$, then for every $\alpha \in [0,1]$, G is an $(\alpha n, \frac{1}{((1-\alpha)\lambda^2+\alpha)})$-vertex expander.

Proof. We have that $\lambda_2(G) \leq \lambda$. Applying the calculations from before and the lemma, we have for any probability distribution over the vertices π that

$$\begin{aligned} \|M\pi - u\|^2 \leq \lambda_2^2 \|\pi - u\|^2 = \lambda_2^2 \left(\|\pi\|^2 - \frac{1}{n} \right) \\ \text{CP}(M\pi) - \frac{1}{n} \leq \lambda^2 \left(\text{CP}(\pi) - \frac{1}{n} \right) \end{aligned} \tag{7}$$

Letting S be any subset of vertices with $|S| \leq \alpha n$ and π the uniform distribution on S, by the lemma, $\text{CP}(\pi) \geq \frac{1}{|S|}$ and $\text{CP}(M\pi) \geq \frac{1}{|\text{Supp}(M\pi)|} = \frac{1}{|N(S)|}$. Therefore

$$\frac{1}{|N(S)|} - \frac{1}{n} \leq \lambda^2 \left(\frac{1}{|S|} - \frac{1}{n} \right)$$

Now we solve for $|N(S)|$ and use the fact that $n \geq \frac{|S|}{\alpha}$ to get

$$|N(S)| \geq \frac{|S|}{((1-\alpha)\lambda^2 + \alpha)}$$

from which we see that G is a $(\alpha n, \frac{1}{((1-\alpha)\lambda^2+\alpha)})$-vertex expander by definition. □

4.3 Vertex Expansion $\implies$ Spectral Expansion

Here, we state the theorem in the other direction, but do not prove it.

Theorem 4.4. *Vertex $\implies$ Spectral.*
For every $\delta > 0$ and $d > 0$, there exists $\gamma > 0$ such that if G is a d-regular $(\frac{n}{2}, 1+\delta)$ vertex expander, then it also has spectral expansion γ. Specifically, we can take $\gamma = \Omega\left((\frac{\delta}{d})^2\right)$.

Note that we require the maximum size of a vertex subset for which we bound the neighborhood to be half the size of the graph. This is necessary for the following reason: if we allowed αn, then for $\alpha < \frac{1}{2}$, a graph can have good vertex expansion by no spectral expansion. This occurs if the graph is the disjoint union of two good expanders.

4.4 Optimizing the Expansion Constants

We can summarize the exact result in the following corollary:

Corollary 4.5. *Let $\mathcal{G}$ be an infinite family of graphs, all with degree d, constant. Then the following conditions are equivalent:*

1. *There is a constant $\delta > 0$ such that every $G \in \mathcal{G}$ is an $(\frac{n}{2}, 1+\delta)$ vertex expander.*
2. *There is a constant $\gamma > 0$ such that every $G \in \mathcal{G}$ has spectral expansion γ.*

However, the measures are not the same if we remove the restriction on α, and let $k = \alpha n$. For small α, we can get $a \sim d - 1$, just a little bit less than it (i.e. $a = d - 1.01$). It is also true that $d - 1$ is an upper bound on a, $a < d - 1$.

If we start out with a $(1-\lambda)$-spectral expander, then taking $\alpha \to 0$ gives that vertex expander coefficient $a \sim \frac{1}{\lambda^2}$. It therefore follows by combining these two bounds that

$$\lambda > \frac{1}{\sqrt{d}} - o(1)$$

More precisely, we have the following theorem:

Theorem 4.6. *Lower bound on λ.*
For every constant $d \in \mathbb{N}$, any d-regular, n-vertex graph G satisfies $\lambda(G) \geq 2\frac{\sqrt{d-1}}{d} - O(1)$

Later, we will see an explicit construction of a Ramanujan graph which gives

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

4.5 Spectral Expansion $\equiv$ Edge Expansion

Also note that spectral expansion implies edge expansion and vice versa.

Theorem 4.7. *Spectral $\equiv$ Edge.*

1. *If a d-regular, n-vertex graph G has spectral expansion γ, then G is a $(\frac{n}{2}, \frac{\gamma}{2})$-edge expander.*
2. *If a d-regular, n-vertex graph is an $(\frac{n}{2}, \epsilon)$-edge-expander and at least α fraction of the edges leaving each vertex are self-loops for some $\alpha \in [0,1]$, then G has spectral expansion $\alpha \cdot \frac{\epsilon^2}{2}$.*

The first direction is proved by the Expander Mixing Lemma, to which we will now turn.

5 The Expander Mixing Lemma

The Expander Mixing Lemma is interesting in its own right, and shows a fundamental connection between spectral and edge expansion properties of graphs. Intuitively, it shows that expander graphs behave like random graphs (we will see exactly how this is the case soon). From another viewpoint, the lemma says that if you take a random walk on an expander from an arbitrary vertex, you approach the stationary distribution very quickly. Using this perspectiv, we can also view expander graphs as approximating complete graphs, where the first step of a random walk is already uniformly distributed. (Recall that another property of complete graphs that they seek to emulate is being well-connected).

As a note, this theorem was proved by Alon (who has a wealth of publication on the subject of expanders) and Chung in 1988.

Theorem 5.1. *The Expander Mixing Lemma.*
Let graph G be a (n, d, γ)-spectral expander with adjacency matrix A, where $\gamma = 1 - \lambda$, $\lambda \in [0, 1]$. Then for all $S, T \in V$, we have

$$\left| |e(S,T)| - \frac{d|S| \cdot |T|}{n} \right| \leq \hat{\lambda} d \sqrt{|S||T|}$$

where $\hat{\lambda} = \max\{|\lambda_2|, |\lambda_n|\}$ for λ_i an eigenvalue of the normalized adjacency matrix $\frac{1}{d}A$.

Before we get to the proof, let us break down what this theorem is actually stating. $|e(S,T)|$ is the number of edges in an (S,T)-cut, and we will see $\frac{d|S|\cdot|T|}{n}$ is the expected number of edges between S and T in a random d-regular graph with n vertices. It is like we are taking the edges of a graph at random, picking an edge with probability $\frac{d}{n}$. Therefore, we are bounding how close the spectral expander is to a random graph, in terms of the size of the cut set between any two vertex sets. $\hat{\lambda}$ controls the term, so the smaller $\hat{\lambda}$ is (the larger γ is), the smaller the difference. Thus the better the spectral expander, the closer the behavior to a random graph. Also note that our definition of $\hat{\lambda}$ is only not the absolute value of the second largest eigenvalue in the case where G is bipartite: here, $\hat{\lambda} = d$. Now we give the proof of this lemma.

Proof. Let $\mathbf{1}_S, \mathbf{1}_T$ be the characteristic vectors of S and T - that is, 0 if vertex i is not present, 1 if vertex i is present in S or T respectively. We can write these vectors in terms of an orthonormal basis of eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$. Thus we have $\mathbf{1}_S = \sum_i \alpha_i \mathbf{v}_i, \mathbf{1}_T = \sum_i \beta_i \mathbf{v}_i$. Then we have

$$|e(S,T)| = \mathbf{1}_S \cdot A \cdot \mathbf{1}_T = \left(\sum_i \alpha_i \mathbf{v}_i \right) \cdot A \cdot \left(\sum_i \beta_i \mathbf{v}_i \right) = \sum_i \lambda_i \alpha_i \beta_i$$

where the λ_i are the eigenvalues of A. Note that this is basically like SVD, we get the eigenvalue out because we are multiplying by unit orthogonal eigenvectors: $\mathbf{v}_i^T A \mathbf{v}_i = \lambda_i \mathbf{v}_i^T \mathbf{v}_i = \lambda_i$, and $\mathbf{v}_i^T A \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j = 0$ (by orthogonality). Then recall that the first eigenvalue of the adjacency matrix (not normalized) is d, and the corresponding eigenvector is the normalized uniform distribution on the vertices - in other words, $\alpha_1 = \mathbf{1}_S \cdot \sqrt{n}\mathbf{u} = \frac{|S|}{\sqrt{n}}, \beta_1 = \mathbf{1}_T \cdot \sqrt{n}\mathbf{u} = \frac{|T|}{\sqrt{n}}, \lambda_1 = d$. We can therefore re-write as follows:

$$|e(S,T)| = \lambda_1 \alpha_1 \beta_1 + \sum_{i=2}^{n} \alpha_i \beta_i \lambda_i = d \cdot \frac{|S| \cdot |T|}{n} + \sum_{i=2}^{n} \alpha_i \beta_i \lambda_i$$

In other words, the first term is the expected value of the number of edges between S and T for the uniform

distribution. Intuitively, this term spreads the weight evenly over the graph, and thus is the expected value.

$$\begin{aligned}\left||e(S,T)| - d\cdot\frac{|S|\cdot|T|}{n}\right| &= \left|\sum_{i=2}^{n}\alpha_i\beta_i\lambda_i\right| \\ &\leq d\cdot\sum_{i=2}^{n}|\alpha_i\beta_i\lambda_i| \\ &\leq \hat{\lambda}d\cdot\sum_{i=2}^{n}|\alpha_i\beta_i|\end{aligned} \tag{8}$$

Then, by Cauchy-Schwarz and the fact that the norm with the first element zeroed is smaller than the full norm,

$$\left||e(S,T)| - d\cdot\frac{|S|\cdot|T|}{n}\right| \leq \hat{\lambda}d\|\alpha\|_2\|\beta\|_2 = \hat{\lambda}d\|\mathbf{1}_S\|_2\|\mathbf{1}_t\|_2 = \hat{\lambda}d\sqrt{|S||T|}$$

□

Now we can again gain a bit more intuition about what this means by shifting around the statement of the theorem. Let's first do a thought experiment. Consider a d-regular graph G. We could

1. Choose a random vertex $v_1 \in V$ and then pick one of its neighbors v_2.
2. Choose two vertices (v_1, v_2) randomly and independently from $V \times V$.

What the Expander Mixing Lemma effectively states is that $\mathbf{P}\{(v_1, v_2) \in S \times T | S, T \subseteq V\}$ is approximately the same for both scenarios, for graphs with good expansion.

We can see this as follows. The probability for the first case is just $\frac{|e(S,T)|}{nd}$ - we first choose 1 out of n, then 1 out of d. The probability for the second case is $\frac{|S|}{n}\cdot\frac{|T|}{n}$- each time, we choose 1 out of n. Note what happens when we divide by nd throughout the statement of the lemma:

$$\left|\frac{|e(S,T)|}{nd} - \frac{|S||T|}{n^2}\right| \leq \frac{\lambda}{n}\sqrt{|S||T|}$$

Thus we see that the Expander Mixing Lemma shows that these two probabilities are close together.

A converse statement also holds for the Expander Mixing Lemma, which we will state but not prove.

Theorem 5.2. *Converse of Expander Mixing Lemma.*
Let G be a d-regular graph with n vertices and suppose

$$\left||e(S,T)| - d\cdot\frac{|S|\cdot|T|}{n}\right| \leq d\theta\sqrt{|S||T|}$$

holds for every pair of disjoint vertex sets (S, T) for some positive θ. Then

$$\lambda = \mathcal{O}\left(\theta\cdot\left(1+\log\left(\frac{d}{\theta}\right)\right)\right)$$

We can also give a version of the Expander Mixing Lemma for irregular graphs, i.e. graphs that do not have a constant degree d. To do this, we introduce some new definitions.

Definition 5.3. Volume of a vertex set.
We define

$$\text{vol}(S) = \sum_{v\in S} d_v$$

If we define the average degree $\bar{d} = \frac{1}{|S|}\sum_{v\in S} d_v$, then we can succinctly write $\text{vol}(S) = \bar{d}|S|$.

We now state the more general expander mixing lemma.

Theorem 5.4. *General Expander Mixing Lemma.*
For every two disjoint sets of vertices $S, T \subseteq V$, we have

$$\left| |e(S,T)| - \frac{\text{vol}(S) \cdot \text{vol}(T)}{\text{vol}(V)} \right| \leq \lambda \sqrt{\text{vol}(S) \cdot \text{vol}(T)}$$

or

$$\left| |e(S,T)| - \bar{d} \cdot \frac{|S| \cdot |T|}{n} \right| \leq \lambda \cdot \bar{d} \sqrt{|S| \cdot |T|}$$

We can see that it looks roughly the same as the previous statement of the theorem, but instead, we use the average degree. Its proof varies in that we can't simply use the nice $\frac{1}{d}A$ formulation as before. Instead, we have to write $M = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, where D is a diagonal matrix with the degree of vertex i at (i, i). It takes a little more computation, but the result can be achieved by similar means as before (namely, plenty of linear algebra).

6 Constructions of Expanders

So far, we have not seen any explicit constructions of expander graphs, we have only proved things about them - for instance, they exist. However, it is desirable to have explicit constructions - if we can only say 'random graphs are good expanders', then in some situations - where we want to use the pseudorandomness of the expander itself in order to solve a problem - it will be completely useless! If you had the prerequisite randomness necessary in order to find the expander, you would not need the expander itself for your application!

Some other reasons a randomly chosen expander might not be sufficient are as follows.

1. There is some error probability associated with the graph - we may not want to tolerate any error in the case that the chosen graph is not an expander. To deal with this, we could check if the chosen graph is an expander, but computing the expansion is known to be **NP**-hard for most measures (however, spectral expansion can be computed in polynomial time with respect to the size of the graph, as it is just an eigenvalue computation). Spectral expansion however only yields estimates to vertex and edge expansion, and does not get optimal expansion in these other measures.

2. Some applications require exponentially large expander graphs, and thus it is impossible to even write down a randomly chosen expander in some situations.

Now that we see the need for explicit constructions of expanders (this is also an interesting problem combinatorially), let us define two types of constructions of expanders on n vertices.

Definition 6.1. Mildly Explicit Construction. We construct a complete representation of the graph in time $\text{poly}(n)$.

Definition 6.2. Fully Explicit Construction. Given a vertex $v \in [n]$ and $i \in [d]$, where d is the degree of the expander, compute the i^{th} neighbor of v in time $\text{poly}(\log(n))$.

One reason we care about the distinction between mildly explicit and fully explicit constructions is the third case mentioned above. At first, you'd think that a complete representation of the graph should be 'fully explicit' - however, the problem comes when dealing with exponentially large expanders. Suppose we want to perform a random walk - we don't need to store the full graph in memory, if we can just compute the i^{th} neighbor, we are perfectly happy. Performing random walks on expanders can be useful in randomness-efficient error reduction.

There are several approaches to explicitly constructing expander graphs.

6.1 Ramanujan Graphs

First, we briefly discuss a mildly explicit construction of a class of expander graphs called Ramanujan graphs, which are optimal in the spectral-expander sense.

The construction is given as follows:

Definition 6.3. Lubotzsky-Phillips-Sarnak construction of Ramanujan graphs.
$G = (V, E)$ is a graph with $V = \mathbb{F}_q \cup \{\infty\}$, the finite field of prime order q such that $q \equiv 1 \bmod 4$ plus an infinity node. The edges connect each node z with all nodes z' of the form

$$z' = \frac{(a_0 + ia_1)z + (a_2 + ia_3)}{(-a_2 + ia_3)z + (a_0 - ia_1)}$$

for $a_0, a_1, a_2, a_3 \in \mathbb{Z}$ such that $a_0^2 + a_1^2 + a_2^2 + a_3^2 = p$, a_0 is odd and positive, a_1, a_2, a_3 are even, for some fixed prime $p \neq q$ such that $p \equiv 1 \bmod 4$, q is a square modulo p, and $i \in \mathbb{F}_q$ such that $i^2 = -1 \bmod q$.

The degree of this graph is the number of solutions to $a_0^2 + a_1^2 + a_2^2 + a_3^2 = p$, which happens to be $d = p + 1$. It also has $\lambda(G) \leq 2\frac{\sqrt{d-1}}{d}$, so it is an optimal spectral expander, and it is in fact even better than expanders drawn from random graphs. These graphs are mildly explicit due to the need to find the prime q (according to [Vadhan] - I could not find a source saying it has an explicit construction).

The explanation of why this works is rather complicated, and we won't get into it...if you want to know more, you could ask Professor Sarnak, who was one of the three authors of a paper that introduced this construction.

6.2 Iterative Graph Operations

Another approach to constructing expanders is combinatorially. We can define some operations over graphs, and then define an iterative algorithm to build expander graphs. Much success has been had in particular with the zig-zag product, which we will shortly describe.

First we give some definitions.

Definition 6.4. Graph product.
Let $G = (V, E)$ be a graph with n vertices and degree d. G^k denotes the graph whose vertex set is V: Two edges u, v are connected iff they are connected by a path of length k in G. G^k has degree d^k, and if A is the adjacency matrix of G, the adjacency matrix of G^k is A^k.

The spectral expansion is improved by taking the product of a graph with itself, since we are effectively taking eigenvalues with magnitude less than 1 to some power k, which makes them smaller. However, we get this gain at the cost of a larger degree, which blows up exponentially - we want to get good expansion AND keep low degree as the number of vertices increases.

We now define the replacement product.

Definition 6.5. Replacement product. We define $\mathbf{R}(G, H)$ to be the replacement product of G and H, with G having n vertices and degree m, and H having m vertices and degree d. The new graph $\mathbf{R}(G, H)$has $n \cdot d$ vertices and degree $d + 1$ with vertex set $V_G \times V_H$ and a set of edges defined in the following way. We we will think of G as the large graph, and H as the small graph. Basically, we replace every vertex in G with a copy of H. We will refer to the copy of H for vertex v as the cloud of v.

For $v \in V(G), i \in V(H)$, let (v, i) denote the i^{th} vertex in the v^{th} cloud. Note that $i \in [m]$. Let $(u, v) \in E(G)$ be such that v is the i^{th} neighbor of u and u is the j^{th} neighbor of v. Then $((u, i), (v, j)) \in E(\mathbf{R}(G, H))$. Also, if $(i, j) \in E(H)$, then $\forall u \in V(G)$, $((u, i), (u, j)) \in E(\mathbf{R}(G, H))$.

We see a visualization here in Figure 1.

Now we are ready to define the zig-zag product.

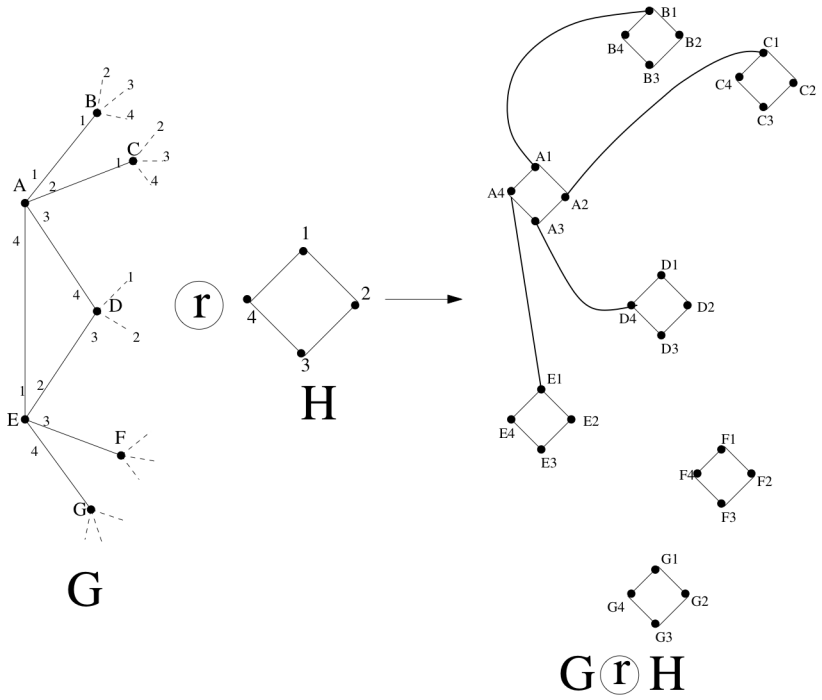

Figure 1: Replacement Product

Definition 6.6. Zig-zag product. We have G, H with $|V_G| = n, |V_H| = m$, degree of G is m, degree of H is d. The zig-zag product ZigZag(G, H)has the same vertex set as in the replacement product. Then we have $((u, i), (v, j)) \in E(\text{ZigZag}(G, H))$ if there exist l, k such that $((u, i), (u, l)), ((u, l), (v, k))$ and $((v, k), (v, j))$ are all in $E(\mathbf{R}(G, H))$. In other words, (u, i) and (v, j) are connected in the zig-zag product if in the replacement product, we can step from (u, i) to (v, j) by first taking a unit step in the u-cloud, taking a step across clouds, and finally taking a unit step in the v-cloud. The zig-zag product is a d^2-regular graph on $n \cdot d$ vertices.

We see a visualization of the zig-zag product in Figure 2.

The zig-zag product is important because it preserves the degree of the graph and enlarges it by a constant factor without making expansion too much worse compared to the original graph.

We introduce some notation for simplicity: (n, d, α) is a graph with n vertices, degree d, and $|\lambda_2| \leq \alpha \cdot d$.

To be able to construct explicit expanders with these graph operations, we need a crucial lemma, which we will use without proof.

Lemma 6.7. *Construction lemma. If G is a (n, m, α)-graph and H is a (m, d, β)-graph, then ZigZag(G, H)is an $(nm, d^2, \alpha + \beta + \beta^2)$-graph.*

Also, it follows easily that H^2 is an (m, d^2, β^2)-graph.

Proof. We only prove the second statement. We have that H is a (m, d, β)-graph, meaning that $|\lambda_2| \leq \beta \cdot d$. Then, $|\lambda_2^2| \leq \beta^2 \cdot d^2$. Since we know that H^2 has degree d^2 and the number of vertices does not change,and we also know that the eigenvalues of H^2 are merely the squares of the eigenvalues of H, it follows that β^2 is the new expansion coefficient. □

From this lemma we can construct a family of expanders. Start with a fixed size base graph H with decent expansion and at each step, take the zig-zag product of a power of the previous graph with H. The degree is kept constant at each step, while the number of vertices grows exponentially and the spectral gap is preserved.

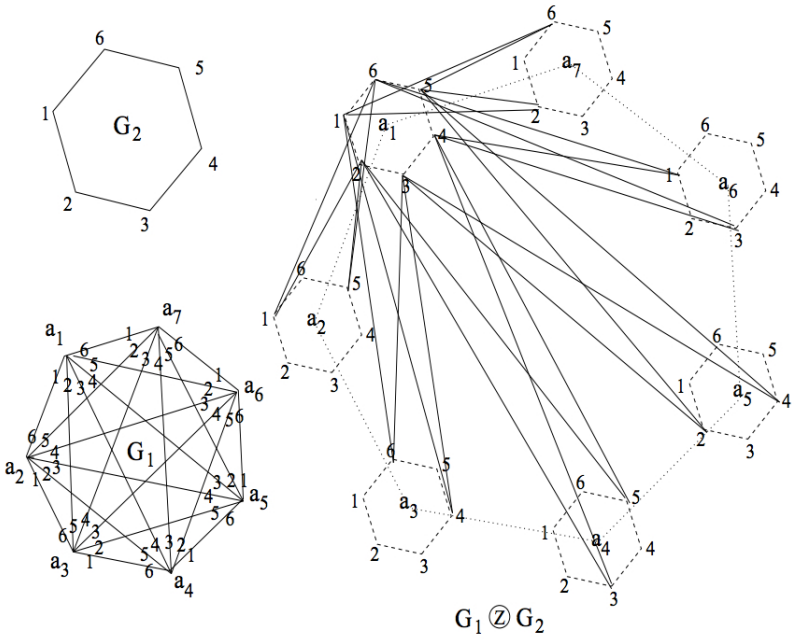

FIGURE 1. Zig-zag product of two graphs.

Figure 2: Zig-zag Product

Theorem 6.8. *Let H be a $(d^4, d, \frac{1}{5})$-graph for some constant d. If we define a family of graphs G_n by*

$$G_1 = H^2, G_{n+1} = \text{ZigZag}(G_n^2, H)$$

then G_n is a $(d^{4n}, d^2, \frac{1}{2})$-graph, and $\{G_i\}_{i=1}^{\infty}$ forms a family of expanders.

Proof. The proof follows by induction. The base case at $n = 1$ is true by the definition of H. Now we have the induction step: suppose that G_n is a $(d^{4n}, d^2, \frac{1}{2})$-graph. Then, G_n^2 is a $(d^{4n}, d^4, \frac{1}{4})$ graph and has the same degree as the number of vertices of H, so the zig-zag product is defined. By the construction lemma, ZigZag(G_n^2, H) is a $(d^{4n+4}, d^2, \frac{1}{2})$-graph since the new α is at most $\frac{1}{4} + \frac{1}{5} + \frac{1}{5^2} \leq \frac{1}{2}$. □

We note that this family of expanders is only mildly-explicit, since the time to compute neighbors is not good enough. This construction can be made fully-explicit with the introduction of another graph operation called tensoring, which we will not go into. Essentially, we change the construction by replacing G_n^2 in the zig-zag product with $G_t \otimes G_t$, which will cause the graph sizes to grow more quickly while preserving degree and expansion from before.

7 Some Applications of Expanders

Now we will see a few reasons expanders are useful.

7.1 Error Reduction in Randomized Algorithms

The setup is as follows. Given a decision problem, we might solve it efficiently with a randomized algorithm A that on input x samples a random bit string $r \in \{0, 1\}^k$ and deterministically computes the answer $A(x, r)$. However, there could be some probability that A is wrong. In this example, for simplicity we will only refer to one-sided error.

Definition 7.1. One-sided error.
If x has the property we are looking for, A outputs 'yes' no matter what. If x does not have the property, A outputs the wrong answer with some probability β.

We would like to reduce the one-sided error using as few bits of randomness as possible.

A naive approach to solving this problem would be to run A t times, each with a new random string, and only output 'no' if at least one of the runs answered 'no'. This drives down the probability of error to β^t; however, we end up using $\mathcal{O}(tk)$ random bits. We can do better using expander graphs.

The reduction is as follows. Let G be a $(2^k, d, \alpha)$-expander with vertex set $V = \{0,1\}^k$ and α such that $\beta + \alpha < 1$. We apply the following algorithm:

1. Pick a starting vertex $v_0 \in V$ uniformly at random.
2. Starting from v_0, perform t steps of a random walk: $(v_0, v_1, \cdots, v_t)$.
3. Output $\bigwedge_{i=0}^{t} A(x, v_i)$.

If we denote $B_x \subseteq \{0,1\}^k$ as the set of strings r for which $A(x,r)$ returns the wrong answer, then A gives the correct answer on x if at least one vertex v_i of the random walk avoids B_x.

We can estimate the probability of this event - it is given by the following theorem.

Theorem 7.2. *Let G be a (n, d, α)-graph and $B \subseteq V$, $|B| = \beta n$ for some $\beta > 0$. The probability that a t-step walk on G starting from a uniformly random vertex v always stays inside B is given by*

$$\boldsymbol{P}\{\forall i, v_i \in B\} \leq (\beta + \alpha)^t$$

We see that $|B| = \beta 2^k$ since β is the original probability of error and applying this theorem gives the event of an error bounded by $(\beta + \alpha)^t$, which rapidly decreases since $\beta + \alpha < 1$. The number of random bits used in this approach is only $\mathcal{O}\left(k + t\log(d)\right)$ random bits: $\log(2^k) = k$ for choosing the initial vertex, and $\log(d)$ for sampling neighbors during each step of the walk.

For this algorithm to be efficient, we must construct the expander efficiently - i.e. we need a fully explicit construction of an expander. This result can be extended to the two-sided error case by changing the AND-operation to a majority vote and applying Chernoff-like bounds to get exponential error decay.

7.2 Connection to Randomness Extractors

As you may recall, a randomness extractor is given by the following definition.

Definition 7.3. Randomness 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 Ext(X, U_d) is $< \epsilon$.

We now show how to use a random walk over an expander to obtain an extractor. As before, we need to define the walk over the graph (what are the vertices?), choose a starting point, and define a way to move from step to step on the walk. 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. Let G be a d-regular expander graph with 2^n vertices, where $\lambda_2 < \frac{1}{2}$. Let s be a sample from X and let z be a uniformly random string of length $t = \left(\frac{n}{2} - \frac{k}{2} + \log(\frac{1}{\epsilon}) + 1\right)\log(d) = \mathcal{O}(n - k + \log(\frac{1}{\epsilon}))$.

Notice that each vertex in our graph can be represented by an n-bit string since $|V| = 2^n$ represents all possibilities for random variable X. Thus, by sampling s we choose a starting point for the walk. Then, the walk itself is defined by z. Like before, we must choose from d choices since the graph is d-regular, this requires $\log(d)$ bits. We do this process $\left(\frac{n}{2} - \frac{k}{2} + \log(\frac{1}{\epsilon}) + 1\right)$ times and use up all the bits of z. Then we output the label (as an n-bit string) of the final vertex of our walk.

Proof. We claim that this process yields a (k, ϵ)-extractor. The main point of the proof is that we need to demonstrate that

$$\left| \left(\frac{1}{d} A \right)^t \pi - U_n \right|_1 < \epsilon$$

namely, that the statistical distance between the result of the random walk and uniform is epsilon small. Here, A is the adjacency matrix of the expander graph and π is the starting distribution, which we know has the property that the min-entropy is at least k (there's an element that occurs with probability $\leq 2^{-k}$).

Let $M = \frac{1}{d}A$. Note that from before, we have $M^t\pi - u = M^t\pi^\perp$, and therefore $\|M^t\pi - u\|_2 = \|M^t\pi^\perp\| \leq \lambda_2^t \|\pi - u\|$. From the min-entropy property, we know that $\|\pi\|_2^2 \leq 2^{-k}$, since we maximize the l_2 norm of a vector where entries sum to 1 by placing weight on as few vectors as possible. Thus we can assume all the weight goes on the first 2^k vectors and by min-entropy each weight is 2^{-k} (they sum to 1), so $\|\pi\|_2^2 \leq \sum_{i=1}^{2^k} 2^{-2k} = 2^{-k}$. By definition $\|u\|_2 = 2^{-\frac{n}{2}}$. Therefore, $\|\pi - u\|_2 \leq \|\pi\|_2 + \|u\|_2 \leq 2^{-\frac{k}{2}} + 2^{-\frac{n}{2}} \leq 2 * 2^{-\frac{k}{2}} = 2^{-\frac{k}{2}+1}$, since $k \leq n$. Recalling that $\lambda_2 \leq \frac{1}{2}$, we have

$$\left| \left(\frac{1}{d} A \right)^t \pi - U_n \right|_1 < \left(\frac{1}{2} \right)^t 2^{-\frac{k}{2}+1} = 2^{-\left(\frac{n}{2} + \log\left(\frac{1}{\epsilon}\right)\right)} = \epsilon 2^{-\frac{n}{2}}$$

Then using the fact that the l_1 norm is bounded by the l_2 norm by $\|v\|_1 \leq \sqrt{j}\|v\|_2$ for $v \in \mathbb{R}^j$, and here $j = 2^n$, we have

$$\left| \left(\frac{1}{d} A \right)^t \pi - U_n \right|_1 < \epsilon$$

as desired. Therefore the function described is a (k, ϵ)-extractor. □

The connection to point-line incidence theorems now becomes clearer. As you may recall, Professor Dvir showed us how to use point-line incidence theorems to analyze extractors. We will not go into that in this talk.

8 An Open Problem

Finally, we will state an open problem about expanders, relating to the desire to give explicit constructions of expander graphs with optimal spectral expansion.

Give an explicit combinatorial construction of constant-degree d expander graphs G with $\lambda G \leq 2\frac{\sqrt{d-1}}{d}$. Note that this is like the Ramanujan constraint, but the construction of those graphs were algebraic - they rely on deep results in number theory, and not combinatoric in the sense that the zig-zag construction was combinatoric.

References

[Dwork] Stanford CS369E Notes

[Kotowski&Kotowski] University of Toronto: Lectures on Expanders

[Sauerwald&Sun] Max Planck Institut: Spectral Graph Theory Notes

[Trevisan11] The Zig-zag Product

[Trevisan14] The Expander Mixing Lemma in Irregular Graphs

[Vadhan] Salil Vadhan's Pseudorandomness Notes

[Williamson2014] Spectral Graph Theory, Expanders, and Ramanujan Graphs