**15-855: Intensive Intro to Complexity Theory**
Spring 2009

Lecture 24: **Expander Graphs**

Recall we always deal with connected, $n$-vertex, $d$-regular undirected graphs with self-loops and multiple edges allowed. Some basics of expanders to recall from last time:

**Definition 0.1.** *We say $G$ is an* $(n, d, \rho)$-combinatorial expander *if for all $S \subseteq V$ with $|S| \leq n/2$,*

$$|E(S, \overline{S})| \geq \rho \cdot d|S|.$$

*where $|E(S, \overline{S})|$ denotes the number of edges between $S$ and its complement $\overline{S} = V \setminus S$.*

**Definition 0.2.** *(The MGG expander.) $V = \mathbb{Z}_m^2$, with $(x, y)$ connected to $(x \pm y, y)$, $(x \pm y + 1, y)$, $(x, y \pm x)$, $(x, y \pm x + 1)$. This is an $(m^2, 8, \rho)$-combinatorial expander for some explicit $\rho > 0$.*

# 1 Random walks and error reduction

Perhaps the most useful fact about expanders is that random walks on them *mix fast*. I.e., if you start at any vertex and take a very short random walk, soon enough your position will be almost uniformly random. For example, we will see the following theorem:

**Theorem 1.1.** *Let $G$ be an $(n, d, .01)$-combinatorial expander. Let $B \subseteq V$ be a set of "bad" vertices with $|B| \leq (1 - \delta)n$. Suppose we pick a random vertex $u_1 \in V$, then take $t - 1$ steps in a standard random walk: to $u_2$, $u_3$, ..., $u_t$. Then it's exponentially unlikely all $u_i$'s fall into $B$:*

$$\mathbf{Pr}[u_i \in B \ \forall i \in [t]] \leq (1 - .00001\delta)^t.$$

Let's see why this is useful in derandomization. Let $L \in \mathsf{RP}$. Perhaps the best algorithm $A$ we know uses $\ell = \ell(n)$ random bits and has one-sided error $1 - \delta$; i.e., if $x \notin L$ then $A$ always says No, if $x \in L$ then $A$ says Yes with probability at least $\delta$. Suppose we would like to get the error down to $1/2$. The straightforward solution would be to repeat $A$ for $O(1/\delta)$ times; this requires using $O(\ell/\delta)$ random bits. If random bits are precious, here is a better solution:

"Take" (implicitly) a strongly explicit $(N, 8, .01)$-combinatorial expander $G$ (like the MGG one), where $N = 2^\ell$.[1] We interpret the vertices $V$ as strings in $\{0, 1\}^\ell$. We spend $\ell$ random bits to pick an initial random vertex $r_1$. We then take a random walk $r_2$, ..., $r_t$ of $t = O(1/\delta)$ steps. Note that this can be done with $3(t - 1)$ random bits and $t \cdot \text{polylog}(N) = t \cdot \text{poly}(\ell)$ time, by strong explicitness; indeed $O(t \cdot \ell) = O(\ell/\delta)$ time for the MGG graphs. We run our algorithm with these random strings $r_1, \ldots, r_t$ and accept if it ever accepts.

Total random bits used: $\ell + O(1/\delta)$, much better than $O(\ell/\delta)$.

[1] This is a square if $\ell$ is even, which we can assume without loss of generality.

But why does this get the error down to 1/2? If $x \notin L$ the algorithm still says No. Otherwise, if $x \in L$, let $B \subseteq \{0,1\}^\ell$ be the random strings which cause $A$ to say No. By assumption, $|B| \leq (1-\delta)2^\ell = (1-\delta)L$. By the Theorem, the probability that *all* $r_1, \ldots, r_t \in B$ and hence the algorithm says No is at most $(1 - .00001\delta)^t = (1 - .00001\delta)^{O(1/\delta)} \leq 1/2$, taking the $O(\cdot)$ constant large enough.

That's nice.

**Exercise 1.2.** *Show how to efficiently find an $n$-bit prime whp using $O(n)$ random bits.*

One can also get the same randomness reduction for BPP algorithms; i.e., two-sided error.

Now that we've shown an application, let's go back and prove some theorems.

## 2 Spectral analysis

We would like to analyze random walks on (regular) graphs. This may be more familiar to you as the analysis of Markov chains. The first step is to form the $n \times n$ *transition matrix* $K$:

$$K[i,j] = \frac{1}{d}A = \mathbf{Pr}[\text{going from vertex } i \text{ to vertex } j] = \begin{cases} 1/d & \text{if } (i,j) \in E \\ 0 & \text{else,} \end{cases}$$

where $A$ is the *adjacency matrix*. Note that $K$ is *symmetric* and *doubly stochastic* (all entries nonnegative, all row and column sums are 1). You can also easily see that

$$K^t[i,j] = \mathbf{Pr}[\text{walk of length } t \text{ starting at } i \text{ ends at } j].$$

Suppose $\pi$ is a probability distribution on vertices; think of it as a row vector with nonnegative entries summing to 1. Suppose $u \sim \pi$ and then we take one random step. What is the resulting probability distribution? It's easily seen to be $\pi K$. In particular, if $\pi = [\frac{1}{n}\frac{1}{n}\cdots\frac{1}{n}]$ represents the uniform distribution, $\pi K = \pi$, since all column-sums in $K$ are 1. We say that the uniform distribution is the *stationary distribution* for the random walk. We also prefer to work with column vectors from now on; since $K$ is symmetric, we've just seen that $K\pi^\top = \pi^\top$.

We now take a detour to linear algebra, a topic we hope you remember. Because $K\pi^\top = 1 \cdot \pi^\top$, the vector $\pi^\top$ is an eigenvector with eigenvalue 1. To study eigenvalues further, define the *Laplacian* matrix

$$L = id - K.$$

The following identity is the key to several observations:

**Exercise 2.1.** *Let $x \in \mathbb{R}^V \cong \mathbb{R}^n$ be any vector, with coordinates indexed by the vertices. Then*

$$x^\top L x = \frac{n}{2} \cdot \mathop{\mathbf{E}}_{(u,v)\ \textit{rand. edge}}[(x[u] - x[v])^2]. \tag{1}$$

Since the RHS is always nonnegative, we conclude that $L$ is a "positive semidefinite (PSD)" matrix. We now recall a fact from linear algebra:

**Fact 2.2.** *Since $L$ is PSD, it has $n$ real, nonnegative eigenvalues, which we sort as*

$$0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n,$$

*and a corresponding orthonormal basis of eigenvectors, $\psi_1, \ldots, \psi_n$.*

Since $K = id - L$, we immediately conclude that $\psi_1, \ldots, \psi_n$ are eigenvectors for $K$ with eigenvalues

$$1 \geq \beta_1 \geq \beta_2 \geq \cdots \geq \beta_n, \qquad \beta_i = 1 - \lambda_i.$$

Actually, we already saw $\beta_1 = 1$, with unit eigenvalue $\sqrt{n} \cdot \pi^\top$; i.e.,

$$\psi_1 = [\tfrac{1}{\sqrt{n}} \tfrac{1}{\sqrt{n}} \cdots \tfrac{1}{\sqrt{n}}]^\top.$$

Since we're assuming $G$ is connected, $\lambda_2$ must be strictly positive. This can be seen from (1): if $x$ is an eigenvector with eigenvalue 0 then one sees that $x$ must be constant on all connected components of $G$; hence $x$ must be parallel to $[11\cdots1]^\top$. Hence:

$$\beta_2 < 1.$$

**Exercise 2.3.** *Show that $\beta_n \geq -1$ with equality iff $G$ is bipartite. (Hint: consider $id + K$ and get a + sign on the RHS of (1).)*

**Exercise 2.4.** *Form $G'$ from $G$ by making it "lazy": adding $d$ self-loops to each vertex. The random walk on $G'$ is like the one on $G$, except at each tick it stays put with probability $1/2$. Show $\beta'_i = \frac{1}{2} + \frac{1}{2}\beta_i$, hence all $\beta'_i$ are nonnegative.*

We now come to an important definition:

**Definition 2.5.** *The "second largest eigenvalue" is*

$$\beta := \max\{\beta_2, |\beta_n|\};$$

*i.e., the maximum of $|\beta_i|$ over all $i \neq 1$. This quantity is strictly less than 1 assuming $G$ is not bipartite. $1 - \beta$ is called the "spectral gap".*

The reason this parameter is key is as follows. From the eigenvector/eigenvalue decomposition of $K$, we have:

**Fact 2.6.** *For any $x \in \mathbb{R}^n$, the vector $Kx$ is equal to $\beta_1 = 1$ times the projection of $x$ on $\psi_1 = [\frac{1}{\sqrt{n}} \cdots \frac{1}{\sqrt{n}}]$, plus $\beta_2$ times the projection of $x$ on $\psi_2$ plus $\cdots$ plus $\beta_n$ times the projection of $x$ on $\psi_n$.*

In particular, *the components of $x$ orthogonal to $[11\cdots1]$ get contracted by a factor of $\beta$ (or smaller).*

# 3 Algebraic expanders

Why have we gone through all this eigenvalue analysis? The idea is that the second largest eigenvalue $\beta$ gives an excellent measure of expansion for $G$.

**Definition 3.1.** *We say $G$ is an $(n, d, \epsilon)$-*algebraic expander *if the second largest eigenvalue $\beta$ is at most $1 - \epsilon$.*

**Proposition 3.2.** *If $G$ is an $(n, d, \epsilon)$-algebraic expander then it is an $(n, d, \epsilon/2)$-combinatorial expander.*

This is extremely useful: The reason is that it's in P to determine the algebraic expansion of a given graph (just compute eigenvalues) whereas it's known to be coNP-hard to certify the combinatorial expansion. For this reason, it's much easier to analyze and use the algebraic expansion of graphs. Indeed, the LPS and MGG expanders are proved via the algebraic definition.

*Proof.* Assume $G$ is an $(n, d, \epsilon)$-algebraic expander and let $S \subseteq V$ have $0 < |S| \leq n/2$ (for $|S| = 0$ there's nothing to prove). Let $x \in \mathbb{R}^V$ be the 0-1 indicator of $S$. From (1) we have

$$x^\top L x = \frac{n}{2} \mathop{\mathbf{Pr}}_{(u,v) \text{ rand. edge}}[(u, v) \text{ cut by } S] = \frac{n}{2} \cdot \frac{|E(S, \overline{S})|}{dn/2} = \frac{|E(S, \overline{S})|}{d}. \tag{2}$$

On the other hand, by the eigenvector/eigenvalue decomposition of $L$, writing $x_i$ for the projection of $x$ onto $\psi_i$, we have

$$x^\top L x = \sum_{i=1}^{n} \lambda_i \|x_i\|_2^2 \geq \lambda_2 \sum_{i=2}^{n} \|x_i\|_2^2 = \lambda_2(\|x\|_2^2 - \|x_1\|_2^2) \tag{3}$$

Here we used $\lambda_1 = 0$ and "the Pythagorean Theorem" (orthonormality of $\psi_i$'s). Clearly

$$\|x\|_2^2 = |S|. \tag{4}$$

And

$$\|x_1\|_2^2 = \langle x, \psi_1 \rangle^2 = \left( \sum_{v \in V} \frac{x[v]}{\sqrt{n}} \right)^2 = \frac{|S|^2}{n}. \tag{5}$$

Combining (2), (3), (4), (5):

$$\frac{|E(S, \overline{S})|}{d} \geq \lambda_2 \left( |S| - \frac{|S|^2}{n} \right) = (1 - \beta_2) \frac{|S|(n - |S|)}{n},$$

hence

$$|E(S, \overline{S})| \geq (1 - \beta_2) d |S| (1 - |S|/n) \geq (1 - \beta) d |S| (1 - |S|/n).$$

The last factor here is at least $1/2$ since $|S| \leq n/2$, completing the proof. □

There is also a reverse to this theorem, which we won't need, which is good because it's a bit trickier to prove:

**Proposition 3.3.** *If $G$ is an $(n, d, \rho)$-combinatorial expander then its $\beta_2$ is at most $1 - \rho^2/2$.*

There is also an extension of Proposition 3.2 called:

**The Expander Mixing Lemma:** *Let $G$ be an $(n, d, \epsilon)$-algebraic expander and let $S, T \subseteq V$. Then*

$$\left| |E(S, T)| - \frac{d}{n} |S||T| \right| \leq (1 - \epsilon) d \sqrt{|S||T|}.$$

*Proof.* Exercise. Mimic the proof of Proposition 3.2 but with $y^\top L x$, where $x$ is the indicator of $S$ and $y$ is the indicator of $T$. □

Thus we see that expanders have nice "pseudorandomness" properties: the number of edges between $S$ and $T$ is close to "what you would expect" for a random $d$-regular graph.

Here's one more simple fact that follows from (1); every graph is at least "slightly expanding":

**Proposition 3.4.** *Assume $G$ is not bipartite. Then $\beta \leq 1 - \frac{1}{dn^2}$.*

*Proof.* The result is not too hard; we leave it as an exercise to prove

$$\beta \leq 1 - \Omega\left(\frac{1}{dn\mathrm{diam}(G)}\right).$$

The hint is to take $x$ to be the second unit eigenvector in (1) and note that, since $x \perp \psi_1$, its coordinates sum to 0. Hence it has at least one component $x[u]$ with $|x[u]| \geq 1/\sqrt{n}$, and at least one other component $x[v]$ with the opposite sign to $x[u]$. Now consider a path from $u$ to $v$... □

# 4 Fast mixing

Here we'll see the first utility of studying eigenvalues:

**Proposition 4.1.** *(Fast mixing.) Let $G$ be an $(n, d, \epsilon)$-algebraic expander. Suppose $p$ is a probability distribution on $V$ (thought of as a column vector); we start a random walk from distribution $p$ and run it for $t$ steps. Then the resulting distribution, $K^t p$, is very close to the uniform distribution $\pi = [\frac{1}{n} \ \cdots \ \frac{1}{n}]^\top$:*

$$\|K^t p - \pi\|_2 \leq (1 - \epsilon)^t.$$

*Proof.* Decompose $p$ in the eigenvector decomposition as $p = \alpha\psi_1 + p'$, where $p' \perp \psi_1$. We have

$$\alpha = \langle p, \psi_1 \rangle = \frac{1}{\sqrt{n}} \sum_v p[v] = \frac{1}{\sqrt{n}};$$

hence $\alpha\psi_1 = \pi$. Note also that $\|p'\|_2 \leq \|p\|_2 \leq \|p\|_1 = 1$. Thus

$$\|K^t p - \pi\|_2 = \|K^t p'\|_2 \leq \beta^t \|p'\|_2 \leq \beta^t = (1 - \epsilon)^t,$$

by definition of $\beta$. □

**Corollary 4.2.** *The* diameter *of $G$ is at most $\lceil \frac{\ln n}{\epsilon} \rceil$.*

I.e., good expanders have diameter $O(\log n)$. This will be key in the proof of $\mathsf{SL} = \mathsf{L}$.

*Proof.* Fix $u, v \in V$. Let $p$ be the distribution with all its probability on $u$. Take $t \geq \frac{\ln n}{\epsilon}$ in the above proposition, so $(1 - \epsilon)^t < \exp(-\epsilon t) = 1/n$. It follows that $K^t p$ cannot have any zero components (else the 2-norm of the difference from $\pi$ would be at least $1/n$); in particular, its component on $v$ is positive. Hence for a random walk of $t$ steps starting at $u$, there is a positive probability of reaching $v$. □

**Exercise 4.3.** *Using Proposition 3.4 and the idea in the proof of Corollary 4.2, show USTCON is in* RL. *(This was discussed in Lecture 9.)*

# 5 Proof of error-reduction by random walks

In this section we prove Theorem 1.1; more specifically, the more detailed version for algebraic expanders.

**Theorem 5.1.** *Let $G$ be an $(n, d, \epsilon)$-algebraic expander. Let $B \subseteq V$ be a set of "bad" vertices with $|B| \leq (1-\delta)n$. Suppose we pick a random vertex $u_1 \in V$, then take $t-1$ steps in a standard random walk: to $u_2$, $u_3$, ..., $u_t$. Then it's exponentially unlikely all $u_i$'s fall into $B$:*

$$\mathbf{Pr}[u_i \in B \ \forall i \in [t]] \leq \sqrt{1-\delta} \cdot (1-\epsilon\delta)^t \leq (1-\epsilon\delta)^t. \tag{6}$$

Note that Theorem 5.1 implies the earlier Theorem 1.1, using Proposition 3.2.

*Proof.* Let $P$ be the projection matrix which "zeroes out" components of a probability vector which are not in $B$: i.e., $P[u,u] = 1$ if $u \in B$, and all other entries are 0. We leave it as an exercise to check the following:

$$\mathbf{Pr}[u_i \in B \ \forall i \in [t]] = \| \overbrace{PK \cdots PKPK}^{t \text{ times}} P\pi \|_1,$$

where $\pi = [\frac{1}{n} \cdots \frac{1}{n}]^\top$ denotes the uniform distribution on vertices. Since $P^2 = P$, we can also write this as

$$\|(PKP)^t P\pi\|_1 \leq \sqrt{n} \cdot \|(PKP)^t P\pi\|_2,$$

using Cauchy-Schwarz. Note that $\|P\pi\|_2 = \sqrt{(1-\delta)/n}$. Hence we can complete the proof by showing that the maximum eigenvalue (spectral norm) of $PKP$ is at most $1-\epsilon\delta$. I.e., we want to show

$$\max_{\|x\|=1} x^\top PKPx \leq 1 - \epsilon\delta. \tag{7}$$

Now given any unit $x \in \mathbb{R}^V$, write $y = Px$ so that the LHS above is $y^\top Ky$. Write $y = y_1 + y'$, where $y_1$ is the projection of $y$ onto $\psi_1$ and $y'$ is the orthogonal component. By eigenvector/eigenvalue decomposition of $K$,

$$y^\top Ky \leq \|y_1\|_2^2 + (1-\epsilon)(\|y\|_2^2 - \|y_1\|_2^2) = \epsilon\|y_1\|_2^2 + (1-\epsilon)\|y\|_2^2. \tag{8}$$

But

$$\|y_1\|_2^2 = \langle y, \psi_1 \rangle^2 = \langle y, P\psi_1 \rangle^2 \leq \|y\|_2^2 \|P\psi_1\|_2^2 = (1-\delta)\|y\|_2^2$$

where the second equality uses that $y = Px$, so in the inner product it doesn't hurt to replace $\psi_1$ by $P\psi_1$, the inequality is Cauchy-Schwarz, and the last equality is the definition of $P$ and $\psi_1 = [\frac{1}{\sqrt{n}} \cdots \frac{1}{\sqrt{n}}]^\top$. Hence (8) is at most

$$\epsilon(1-\delta)\|y\|_2^2 + (1-\epsilon)\|y\|_2^2 = (1-\epsilon\delta)\|y\|_2^2 \leq (1-\epsilon\delta)\|x\|_2^2 = 1 - \epsilon\delta,$$

where the inequality is because $y = Px$ and $P$ is a projection. This completes the proof. □