Lecture 7: Expanders

Lecturer: *Dana Moshkovitz* Scribe: *Andrey Grinshpun and Dana Moshkovitz*

In this lecture we give basic background about expander graphs. Expanders are graphs with strong connectivity properties: every two large subsets of vertices in an expander have many edges connecting them. Surprisingly, one can construct very sparse graphs that are expanders, and this is what makes them so useful.

Expanders have a huge number of applications in theoretical computer science: in construction of fault-tolerant networks, in proofs of complexity-theoretic results like $SL = L$, in design of error correcting codes, etc. We use expanders to amplify the soundness of PCPs, and in subsequent lectures, to obtain PCPs with a constant number of queries.

1 Basic Definitions

When we have a graph $G = (V, E)$ and it is unambiguous to which graph we are referring, we use $n = |V|$. The graphs we consider are all regular, and we denote their degree by D.

Definition 1. *Given $G = (V, E)$ a regular graph of degree D, and given $S, T \subseteq V$, define*

$$E(S,T) = \{(u,v) \in E : u \in S \wedge v \in T\} = E \cap (S \times T)$$

and note that $0 \leq |E(S, V \setminus S)| \leq |S|\, D$

Definition 2. *The* Edge Expansion *or* Cheeger Constant *of G is*

$$\Phi(G) = \min_{S \subseteq V,\, 0 < |S| \leq \frac{n}{2}} \frac{|E(S, V \setminus S)|}{|S|}$$

Intuitively, the above computes the worst bottleneck in the graph.

Definition 3. *A sequence of graphs $\{G_n\}_{n \in \mathbb{N}}$ is a constant degree expander if there exist constants $D \geq 1$, $\varepsilon > 0$, such that:*

- *$\forall n, G_n$ is on n vertices*
- *$\forall n, G_n$ is D-regular*
- *$\forall n, \Phi(G_n) \geq \varepsilon$*

1.1 Examples

For the cycle on n vertices,

$$\Phi(C_n) = \frac{2}{\frac{n}{2}} = \frac{4}{n}$$

which is seen by taking S to be a half-cycle. Thus, $\{C_n\}$ is not a constant degree expander.

For the complete graph on n vertices,

$$\Phi(K_n) = \min_{S \subseteq V, |S| \leq \frac{n}{2}} \frac{|E(S, V \setminus S)|}{|S|}$$

$$= \min_{S \subseteq V, |S| \leq \frac{n}{2}} (\frac{|S|\,|V \setminus S|}{|S|})$$

$$= \min_{S \subseteq V, |S| \leq \frac{n}{2}} (|V \setminus S|) = \lceil \frac{n}{2} \rceil$$

That is, the complete graph has an optimal Cheeger constant $\Theta(D)$. However, K_n has super-constant degree $D = n - 1$, so the family $\{K_n\}$ is not a constant degree expander. That said, constant degree expanders are frequently thought of as sparse approximators of $\{K_n\}$.

2 Constructions

For n natural, a random D-regular graph on n vertices satisfies, with high probability, $\Phi(G) \geq \varepsilon$ for some constant ε which depends on D.

There are explicit constructions of constant degree expanders as well (we do not go into the proofs):

1. (Margulis, '73) We create a graph on vertices $\mathbb{Z}_m \times \mathbb{Z}_m$. The neighbors of (x, y) are

$$(x + y, y), (x - y, y), (x + y + 1, y), (x - y + 1, y)$$

$$(x, y + x), (x, y - x), (x, x + y + 1), (x, y - x + 1)$$

2. (Lubotzky, Phillips, Sarnak, '86) Use vertices $\mathbb{Z}_p$ for p prime, $p = 1 (\mod 4)$. The neighbors of x are $x + 1, x - 1, x^{-1}$. This construction creates Ramanujan Expanders (defined later).

3. (Reingold, Vadhan, Wigderson, '00) They show an iterative combinatorial construction: repeatedly, square the adjacency matrix of the graph and reduce its degree (by an operation they call "Zig Zag Product").

3 Linear Algebra View

The *adjacency matrix* of a graph $G = ([n], E)$ is the matrix A defined by

$$A_{i,j} = \chi_{(i,j) \in E}$$

That is, $A_{i,j}$ is 1 if $(i, j) \in E$ and 0 otherwise.

We write $\vec{1}$ for $(1, 1, \ldots, 1)$. If G is D-regular with adjacency matrix A, then

$$A\vec{1} = D\vec{1}$$

so $\vec{1}$ is an eigenvector with associated eigenvalue D. This eigenvector corresponds to a uniform distribution over the vertices of the graph.

Take

$$D = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{n-1} \geq -D$$

to be the eigenvalues of A. It may be checked that $\lambda_0 = \lambda_1$ iff the graph is not connected, and that $\lambda_{n-1} = -D$ iff the graph is bipartite. Recall that since A is symmetric, its eigenvectors may be normalized to form an orthonormal basis.

Define the *second eigenvalue* λ of A by

$$\lambda = \max_{0<i\leq n-1} |\lambda_i| = \max(\lambda_1, -\lambda_{n-1})$$

The following claim, which we often use in the sequel, says that the 2-norm of a vector parallel to $\vec{1}$ shrinks by at least λ after being multiplied by A:

Claim 3.1. *For $v \in \mathbb{R}^n$ with $\langle v, \vec{1}\rangle = 0$,*

$$\|Av\| \leq \lambda \|v\|.$$

Proof. We choose $\{w_i\}$ orthonormal eigenvectors of A such that the eigenvalue associated to w_i is λ_i, and we may write.

$$v = \sum_i \alpha_i w_i$$

and note in the above sum, since $\langle v, \vec{1}\rangle = 0$, we have that $\alpha_0 = 0$. We also have that $\sum_i \alpha_i^2 = \|v\|^2$.

$$\|Av\|^2 = \langle Av, Av\rangle = \langle \sum_i \alpha_i Aw_i, \sum_i \alpha_i Aw_i\rangle = \sum_{i,j} \alpha_i \alpha_j \langle Aw_i, Aw_j\rangle$$

$$= \sum_{i,j} \alpha_i \alpha_j \lambda_i \lambda_j \langle w_i, w_j\rangle = \sum_i \alpha_i^2 \lambda_i^2 \leq \sum_i \alpha_i^2 \lambda^2 = \lambda^2 \sum_i \alpha_i^2 = \lambda^2 \|v\|^2$$

□

Define the *spectral gap* of A by $D - \lambda$. The following theorem, presented here without a proof, connects the expansion of the graph with its spectral gap:

Theorem 4 (Cheeger)**.**

$$\frac{\Phi(G)^2}{2D} \leq D - \lambda \leq 2\Phi(G)$$

The following theorem, also without a proof, bounds the possible size of the spectral gap.

Theorem 5.

$$\lambda \geq (1 - o(1))\sqrt{D}$$

Graphs that achieve the bound of Theorem 5 are called *Ramanujan Expanders*.

The following useful lemma shows that small second eigenvalue guarantees that between any two subsets of vertices in the graph, the number of edges resembles the number of edges in a random graph with the same number of edges:

Lemma 3.2 (Expander Mixing Lemma)**.**

$$\forall S, T \subseteq V \quad \left| \frac{|E(S,T)|}{Dn} - \frac{|S|}{n}\frac{|T|}{n} \right| \leq \frac{\lambda}{D}\sqrt{\frac{|S|}{n}\frac{|T|}{n}}$$

Proof. (Below whether a vector is interpreted as a column vector or as a row vector should be clear from context.)

Take f to be the indicator vector on S; i.e. $f_i = 1 \Leftrightarrow i \in S$ and take g to be the indicator vector on T. Then we have

$$|E(S,T)| = fAg$$

Choose $f_\parallel, f_\perp$ (similarly, $g_\parallel, g_\perp$) such that the following three conditions hold:

$$f = f_\parallel + f_\perp$$

$$\exists c \in \mathbb{R} \text{ s.t. } f_\parallel = c\vec{1}$$

$$\langle f_\parallel, f_\perp \rangle = 0$$

i.e., we have $f_\parallel = \langle f, \frac{\vec{1}}{\sqrt{n}} \rangle \frac{\vec{1}}{\sqrt{n}} = \frac{|S|}{n}\vec{1}$. Then,

$$fAg = (f_\parallel + f_\perp)A(g_\parallel + g_\perp) = f_\parallel Ag_\parallel + f_\parallel Ag_\perp + f_\perp Ag_\parallel + f_\perp Ag_\perp$$

Next we analyze these four terms. The first term corresponds to the uniform distributions on S and T, and yields the main term:

$$f_\parallel Ag_\parallel = \frac{|S|\,|T|}{n}\frac{\vec{1}}{\sqrt{n}}A\frac{\vec{1}}{\sqrt{n}} = \frac{|S|\,|T|\,D}{n}$$

We wish to show the remaining 3 summands are small.

$$f_\perp Ag_\parallel = f_\perp \frac{|T|}{n}\vec{1} = \frac{|T|}{n}\langle f_\perp, \vec{1}\rangle = 0$$

Similarly, we get $f_\parallel Ag_\perp = 0$.

Finally, the error term comes from the following:

$$\begin{aligned} |f_\perp Ag_\perp| &\leq \|f_\perp\|_2 \|Ag_\perp\|_2 \quad \text{(Cauchy-Schwarz)} \\ &\leq \lambda \|f_\perp\|_2 \|g_\perp\|_2 \quad \text{(Claim 3.1)} \\ &\leq \lambda \|f\|_2 \|g\|_2 \\ &= \lambda\sqrt{|S|\,|T|} \end{aligned}$$

□

4 Random Walks on Expanders

Throughout this section, we will use $\alpha = \frac{\lambda}{D}$, and we will take u to be the uniform distribution over the vertices, i.e. $u = \frac{\vec{1}}{n}$.

Define a *random walk* on a graph $G = (V, E)$ w.r.t. p a distribution over V to be a sequence X_i of random variables with X_0 distributed according to p and X_{i+1} a neighbor of X_i chosen uniformly at random.

The random walk defines a Markov process over the vertices of the graph with transition matrix $\hat{A} = \frac{A}{D}$.

Intuitively, the following theorem says that a random walk on an expander quickly forgets where it started.

Theorem 6. *For any distribution p over the vertices and for any $t \geq 0$,*

$$\|\hat{A}^t p - u\|_2 \leq \alpha^t \|p - u\|_2 \leq \alpha^t$$

Proof. We proceed by induction on t. The case for $t = 0$ is clear. Note that $\hat{A}^t p$ is a distribution, so $\langle \hat{A}^t p - u, \vec{1} \rangle = \langle \hat{A}^t p, \vec{1} \rangle - \langle u, \vec{1} \rangle = 1 - 1 = 0$. Hence,

$$\begin{aligned} \|\hat{A}^{t+1} p - u\|_2 &= \|\hat{A}(\hat{A}^t p - u)\|_2 \quad (\hat{A}u = u) \\ &\leq \alpha \|\hat{A}^t p - u\|_2 \quad \text{(Claim 3.1)} \\ &\leq \alpha^{t+1} \|p - u\|_2 \quad \text{(Induction hypothesis)} \end{aligned}$$

□

We now prove that a random walk on an expander has an exponentially small probability to stay in any small subset of the vertices:

Theorem 7. *For $B \subseteq V, |B| = \beta n$, the probability that a random walk of length t starting from the uniform distribution over the vertices stays in B is at most $(\beta + \alpha)^t$.*

Note that the uniform distribution is a stationary distribution: $\hat{A}u = u$.

Define $P = P_B$ to be the projection onto B, that is $(P_B)_{i,j}$ is 1 if $i = j \in B$ and 0 otherwise. Note $P^2 = P$ and that for any vector v, $\|Pv\|_2 \leq \|v\|_2$. Note also that the probability that the random walk stays in B is $\left|(P\hat{A})^t P u\right|$.

Claim 4.1. *For any vector v such that $Pv = v$,*

$$\|P\hat{A}v\|_2 \leq (\beta + \alpha)\|v\|_2$$

Proof. This is trivial if $v = \vec{0}$. Otherwise, without loss of generality, take $\sum_i v_i = 1$. Then we may write $v = u + z$ with $\langle u, z \rangle = 0$. Then, by a triangle inequality,

$$\|P\hat{A}v\|_2 = \|P\hat{A}u + P\hat{A}z\|_2 \leq \|P\hat{A}u\|_2 + \|P\hat{A}z\|_2.$$

We have that $P\hat{A}u = Pu$, and so $\|P\hat{A}u\|_2 = \sqrt{\frac{\beta}{n}}$. By Cauchy-Schwarz, using $Pv = v$,

$$1 = \sum_{i \in B} v_i \leq \sqrt{\sum v_i^2}\sqrt{\beta n} = \|v\|_2 \sqrt{\beta n}.$$

Thus, $\|v\|_2 \geq \frac{1}{\sqrt{\beta n}}$, and

$$\|P\hat{A}u\|_2 \leq \beta \|v\|_2.$$

By assumption $\langle z, u \rangle = 0$ and so $\langle z, \vec{1} \rangle = 0$ and thus, by Claim 3.1, we get

$$\|P\hat{A}z\|_2 \leq \|\hat{A}z\|_2 \leq \alpha \|z\|_2 \leq \alpha \|v\|_2$$

Combining the bounds on $\|P\hat{A}u\|_2$ and $\|P\hat{A}z\|_2$, the claim follows. □

We now prove Thoerem 7:

Proof. From Jensen's inequality,

$$\left(\frac{1}{n}(P\hat{A})^t Pu\right)^2 \leq \frac{1}{n}\|(P\hat{A})^t Pu\|_2^2,$$

and so $\left|(P\hat{A})^t Pu\right| \leq \sqrt{n}\|(P\hat{A})^t Pu\|_2$.

We prove by induction that for every $t \geq 0$, for $v_t = (P\hat{A})^t Pu$, it holds that $Pv_t = v_t$ and $\|v_t\|_2 \leq (\beta+\alpha)^t\|Pu\|_2$: The case $t = 0$ is clear. Assume the claim holds for t, and let us prove it for $t+1$:

$$\begin{aligned}
\|(P\hat{A})^{t+1}Pu\|_2 &= \|(P\hat{A})v_t\|_2 \\
&\leq (\beta+\alpha)\|v_t\|_2 \quad \text{(Claim 4.1, using } Pv_t = v_t\text{)} \\
&\leq (\beta+\alpha)^{t+1}\|Pu\|_2 \quad \text{(Induction hypothesis)}
\end{aligned}$$

Overall,

$$\left|(P\hat{A})^t Pu\right| \leq \sqrt{n}(\beta+\alpha)^t\|Pu\|_2 = \sqrt{\beta}(\beta+\alpha)^t \leq (\beta+\alpha)^t.$$

□

5 Randomness-Efficient Amplification for PCPs

Expanders can be used to amplify the success probability of randomized algorithms in a randonmess-efficient way. We next show a similar application of expanders to amplification of soundness of PCPs.

Amplification of PCP soundness refers to taking a result like:

$$NP \subseteq PCP_{1,s}[r,q]$$

And obtaining from it a result with a much smaller soundness error (though, perhaps, slightly worse other parameters):

$$NP \subseteq PCP_{1,\approx s^k}[r',q']$$

Amplification can be done by running the initial PCP verifier k times independently. Unfortunately, this causes the randomness to become $r' = rk$. We achieve much smaller randomness $r' = r + k\log D$ using degree-D expanders.

We use expander graphs $\{G_{2^r}\}$. As usual, we use D to denote their degree, λ to denote their second eigenvalue, and α for λ/D.

Let V be a verifier with completeness 1, soundness s, randomness r and q queries. The new verifier V' uses the same proof V does, and proceeds as follows:

1. Perform a random walk $r_0, \ldots, r_k$ on G_{2^r} starting from a uniformly distributed vertex r_0.
2. Run V with randomness r_i for all $0 \leq i \leq k$.
3. If V rejects in any of its applications, reject. Otherwise, accept.

The verifier V' uses $r' = r + k\log D$ bits of randomness, and makes $q' = qk$ queries to the proof. It has perfect completeness, and as to its soundness – let $B \subseteq \{0,1\}^r$ be the set of random strings that cause V to accept. We just saw that the probability that $r_i \in B$ for all $0 \leq i \leq k$ is bounded by $(s+\alpha)^k$, so this is our new soundness.