# Notes 15: Expanders

## 1. Expanders

We will consider regular unweighted graphs unless otherwise specified. Self-loops are allowed.

There are a few different definitions of expanders (i.e. well-connected graphs) in the literature:

- (Edge expanders) For every $S \subseteq V$ with $|S| \leqslant |V|/2$, many edges go across $S$ and $\overline{S}$:
$$\varphi(S) = \frac{w(S, \overline{S})}{d(S)} \geqslant \gamma$$
- (Spectral expanders) Largest two eigenvalues $\lambda_1$ and $\lambda_2$ of $\mathcal{A}$ has a large gap:
$$\lambda_1 - \lambda_2 \geqslant \gamma$$
- (Vertex expanders) For every $S \subseteq V$ with $|S| \leqslant |V|/2$, many vertices outside of $S$ is adjacent to $S$:
$$\frac{|N(S) \setminus S|}{|S|} \geqslant \gamma \,,$$
where $N(S) = \{i \in V \mid (i,j) \in E \text{ for some } j \in S\}$ is the set of neighbors of $S$

These definitions are related to each other. Cheeger–Alon–Milman shows that the first two definitions are equivalent (with possibly different $\gamma$). We will later see that spectral expanders satisfying an additional condition are vertex expanders (again with different $\gamma$).

The complete graph $K_n$ with self-loops on $n$ vertices satisfies all of the above definitions with $\gamma$ bounded away from 0 (and independent of $n$). Here $K_n$ is the graph that contains an edge $(i,j)$ for every unordered pair of vertices $i$ and $j$, including when $i = j$. Its adjacency matrix $A$ is the all-one matrix $J$.

$K_n$ is dense and has $\Theta(n^2)$ edges. Interestingly, there are sparse expanders with only $O(n)$ edges.

## 2. Spectral radius

In fact, we will look at graphs with small spectral radius. Given a $d$-regular graph with normalized adjacency matrix $\mathcal{A}$ ($= A/d$), its spectral radius $\lambda$ is $\max\{|\lambda_2|, |\lambda_n|\}$. In other words, the maximum magnitude of all non-trivial eigenvalues of $\mathcal{A}$. (Recall that the largest eigenvalue $\lambda_1$ of $A$ is trivially $d$, with eigenvector $\mathbb{1}$. All other eigenvalues $\lambda_2 \geqslant \ldots \geqslant \lambda_n$ are non-trivial.) In particular, a graph with spectral radius $1 - \gamma$ is a spectral expander.

Graphs with small spectral radius are good approximators of $K_n$. Indeed, a graph has small spectral radius if and only if its Laplacian is close that of $K_n$, properly scaled.

**Claim 2.1.** *Let $H$ be the weighted graph with adjacency matrix $\frac{d}{n}J$. A $d$-regular graph $G$ on $n$ vertices has spectral radius at most $\lambda$ if and only if $\|L_G - L_H\| \leqslant \lambda d$.*

Again the above norm is the operator norm. For a real symmetric matrix, its operator norm equals the largest eigenvalue (in absolute value).

*Proof.* $G$ has spectral radius at most $\lambda$ if and only if all non-trivial eigenvalues of its adjacency matrix $A_G$ are in the range $[-\lambda d, \lambda d]$. This holds if and only if all non-trivial eigenvalues of $L_G = D_G - A_G = dI - A_G$ are in the range $[d - \lambda d, d + \lambda d]$. All these eigenvalues correspond to the eigenspace orthogonal to the eigenvector $\mathbb{1}$. That is, for every $x \perp \mathbb{1}$, Courant–Fischer tells us
$$(d - \lambda d) x^\top x \leqslant x^\top L_G x \leqslant (d + \lambda d) x^\top x \,.$$

For $H$, all non-trivial eigenvalues of its adjacency matrix $A_H$ are 0, and the largest eigenvalue is $\frac{d}{n}$. All non-trivial eigenvalues of $L_H = dI - A_H$ are $d$. Again these eigenvalues correspond to the eigenspace orthogonal to the eigenvector $\mathbb{1}$. That is, for every $x \perp \mathbb{1}$,
$$x^\top L_H x = d x^\top x \,.$$

Therefore for every $x \perp \mathbb{1}$,
$$|x^\top (L_G - L_H) x| \leqslant \lambda d x^\top x \,.$$
Since $L_G - L_H$ is a real symmetric matrix, this is equivalent to $\|L_G - L_H\| \leqslant \lambda d$.

We have only considered $x \perp \mathbb{1}$, because $\mathbb{1}$ is a common eigenvector of $L_G$ and $L_H$ of eigenvalue 0. □

It is well-known that a random $d$-regular graph on $n$ vertices has small spectral radius (bounded away from 1) with high probability. If you are curious, you may read Chapter 8 of *"Expander Graphs and Their Applications"* by Hoory, Linial, and Wigderson. We will see that graphs with small spectral radius in turn behave like random graphs.

## 3. Pseudorandom properties from spectral radius

Given a graph $G = (V, E)$ and subsets $S \subseteq V$ and $T \subseteq V$, let $\overleftrightarrow{E}(S,T) = \{(u,v) \mid u \in S, v \in T, (u,v) \in E\}$ be the set of directed edges with one endpoint in $S$ and another endpoint in $T$. We are counting the number of directed edges (ordered pairs), so an undirected edge with both endpoints in $S \cap T$ is counted twice.

If a graph is generated as a random $d$-regular graph, we expect each edge to appear with probability $\frac{d}{n}$, so in expectation there are $\frac{d}{n}|S||T|$ edges in $\overleftrightarrow{E}(S,T)$. This behavior holds approximately in graphs with small spectral radius.

**Lemma 3.1** (Expander Mixing Lemma)**.** *Let $G$ be a $d$-regular graph with spectral radius $\lambda$. For every $S \subseteq V$ and $T \subseteq V$,*

$$\left| \left|\overleftrightarrow{E}(S,T)\right| - \frac{d}{n}|S||T| \right| \leqslant \lambda d\sqrt{|S||T|}\,.$$

*Proof.* Note that

$$\mathbb{1}_S^\top L_G \mathbb{1}_T = \mathbb{1}_S^\top (dI - A_G)\mathbb{1}_T = d|S \cap T| - \left|\overleftrightarrow{E}(S,T)\right|\,.$$

Since $G$ is a good approximator of $H$ (defined in Claim 2.1),

$$\mathbb{1}_S^\top L_H \mathbb{1}_T = \mathbb{1}_S^\top \left(dI - \frac{d}{n}J\right)\mathbb{1}_T = d|S \cap T| - \frac{d}{n}|S||T|\,.$$

So

$$\left|\overleftrightarrow{E}(S,T)\right| - \frac{d|S||T|}{n} = \mathbb{1}_S^\top L_H \mathbb{1}_T - \mathbb{1}_S^\top L_G \mathbb{1}_T\,.$$

Since $G$ has spectral radius $\lambda$,

$$\mathbb{1}_S^\top (L_G - L_H)\mathbb{1}_T \leqslant \|\mathbb{1}_S\|\|(L_G - L_H)\mathbb{1}_T\| \leqslant \|\mathbb{1}_S\|\|L_G - L_H\|\|\mathbb{1}_T\| = \lambda d\sqrt{|S||T|}\,.$$ □

One can slightly strengthen the above bound. Let $\alpha = |S|/n$ and $\beta = |T|/n$ be the fractional size of $S$ and $T$. We can orthogonalize $\mathbb{1}_S$ and $\mathbb{1}_T$ with respect to $\mathbb{1}$ and get

$$\mathbb{1}_S^\top (L_G - L_H)\mathbb{1}_T^\top = (\mathbb{1}_S - \alpha\mathbb{1})^\top (L_G - L_H)(\mathbb{1}_T - \beta\mathbb{1})\,.$$

Also

$$\|\mathbb{1}_S - \alpha\mathbb{1}\|\|\mathbb{1}_T - \beta\mathbb{1}\| = n\sqrt{(\alpha - \alpha^2)(\beta - \beta^2)}\,.$$

This yields the improved bound

$$\left| \left|\overleftrightarrow{E}(S,T)\right| - \alpha\beta dn \right| \leqslant \lambda dn\sqrt{(\alpha - \alpha^2)(\beta - \beta^2)}\,.$$

## 4. Vertex expansion

**Theorem 4.1** (Tanner)**.** *Let $G$ be a $d$-regular graph with spectral radius $\lambda$. Then for every $S \subseteq V$ of fractional size $\alpha = |S|/n$,*

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

This gives a strong bound for subsets that are very small, i.e. $\alpha$ much smaller than $\lambda^2$. For this $\alpha$, the lowerbound is approximately $|S|/\lambda^2$, so the neighborhood has much more vertices than $|S|$.

*Proof.* We apply (strengthened) Lemma 3.1 with $T = V \setminus N(S)$. Let $\delta = |N(S)|/n$, so $\delta = 1 - \beta$. There are no edges between $S$ and $T$, and it must hold that

$$\alpha\beta dn \leqslant \lambda dn\sqrt{(\alpha - \alpha^2)(\beta - \beta^2)}\ .$$

We get the lowerbound by rearranging terms.

$$\alpha^2\beta^2 \leqslant \lambda^2(\alpha - \alpha^2)(\beta - \beta^2) \qquad \Longleftrightarrow \qquad \alpha\beta \leqslant \lambda^2(1-\alpha)(1-\beta)\ ,$$

which is the same as

$$\frac{\beta}{1-\beta} \leqslant \frac{\lambda^2(1-\alpha)}{\alpha} \qquad \Longleftrightarrow \qquad \frac{1-\delta}{\delta} \leqslant \frac{\lambda^2(1-\alpha)}{\alpha}\ ,$$

and is equivalent to

$$\frac{1}{\delta} \leqslant \frac{\lambda^2(1-\alpha)+\alpha}{\alpha} \qquad \Longleftrightarrow \qquad \delta \geqslant \frac{\alpha}{\lambda^2(1-\alpha)+\alpha}\ . \qquad \square$$

## 5. Random walk on expanders

Consider $(t-1)$-step random walk in a regular graph $G$ starting from a uniformly distributed initial vertex $v_0$. If $G$ is the complete graph $K_n$ with self-loops, the random walk $(v_0, \ldots, v_{t-1})$ is simply a sequence of $t$ independent vertices, uniformly distributed. Now if $G$ is a graph with small spectral radius, the vertices $v_0, \ldots, v_t$ in the random walk are no longer independent of each other. Remarkably, random walk on $G$ still behaves similarly to random walk on $K_n$.

For example, let $S \subseteq V$ be any subset of vertices. If we look at the number of times the random walk hits $S$, this number will be very close to its expectation $t|S|/n$. More generally, let $f : V \to [0,1]$ be any bounded function (such as $f = \mathbb{1}_S$). Let $\lambda$ be the spectral radius of the regular graph $G$.

**Theorem 5.1** (Chernoff Bound for expanders)**.**

$$\mathbb{P}\left[\frac{1}{t}\sum_{0\leqslant i<t} f(v_i) > \mathop{\mathbb{E}}_{v\in V} f(v) + \varepsilon + \lambda\right] \leqslant \exp(-\Omega(\varepsilon^2 t))\ .$$

Note that when $\lambda = 0$, we recover the usual Chernoff bound (up to the hidden constant in the big-$\Omega$).

We will not prove this theorem.

The theorem implies that the success rate of a randomized algorithm can be amplified using a random walk on expanders. Suppose you have an algorithm that, given any input, outputs the correct yes/no answer 70% of the time, where the probability is over the random bits used by the algorithm. You can amplify the success probability by repeating the algorithm and taking the majority vote. But repeating the algorithm many times requires a lot of random bits. One way to save random bits is to perform random walk on $d$-regular graphs with $d$ not too big (say $d = 100$) and small spectral radius $\lambda$.

We associate the vertex set $V$ of the graph $G$ with the random bitstings used by a single run of the randomized algorithm. The subset $S \subseteq V$ corresponds to those random bits where the algorithm yields incorrect outputs. Taking a $(t-1)$-step random walk, the chance that more than half of the bitstrings cause the algorithm to err is exponentially small in $t$. How many random bits do we need to generate the bitstrings? $\log_2 n$ for $v_0$, plus $\log_2 d$ per step of the walk (to choose a random neighbor), for a total of $\log_2 n + (t-1)(\log_2 d)$ random bits. This is much fewer than $t \log_2 n$ bits required to generate $t$ independent bitstrings.