Topics in Theoretical Computer Science March 2, 2015

Lecture 3

Lecturer: Ola Svensson *Scribes: Jérome Amiguet, Romain Edelmann, Florian Tramèr*

1 Introduction

In this lecture, we come back to the *probabilistic method* and introduce Expander Graphs. We will focus primarily on the bipartite case and show an application in the field of error correcting codes and randomness reduction. Before giving a formal definition of such expander graphs, we state the intuition behind them: a graph is a "good" expander graph if every subset of its vertices which is not"too large", has "many connections" to the outside. Different definitions give rise to so called edge-expanders and node-expanders. In the first case we are interested in the number of edges leaving a given subset. In the second, we are interested in the number of vertices which share an edge with some vertex in a subset. We will formalize these in section 4.

Expander Graphs were introduced in the 1970's and have found important applications in communication theory, in particular in the field of error correction codes we mentioned, as well as in computational complexity theory (PCP theorem, hardness of approximation results), pseudorandom number generation or network design. In this lecture, we focus on the combinatorial definition of expander graphs. The next lecture will focus on the algebraic definition.

2 Bipartite Expander Graphs

We begin with some useful definitions. Note that all graphs considered in this lecture are undirected.

Definition 1 *A bipartite graph $G = (L \cup R, E)$ is a graph consisting of two disjoint sets of vertices L and R such that every edge from $E \subseteq L \times R$ connects one vertex of L and one vertex of R (L and R are thus independent sets).*

Definition 2 *A D-regular graph is a graph where every vertex has degree exactly D. In the context of a bipartite graph $G = (L \cup R, E)$, we say that G is D-left-regular iff all vertices in L have degree D. The analogue definition of a D-right-regular graph will not be used.*

Definition 3 *Given a graph $G = (V, E)$ and some subset $S \subseteq V$, we define the neighborhood of S, denoted $\mathrm{N}(S)$, as the set of all vertices in V which are adjacent to some vertex in S. Formally,*

$$\mathrm{N}(S) = \{v \in V \mid \exists u \in S,\ (u, v) \in E\}.$$

We are now ready to give the (combinatorial) definition of a Bipartite Expander Graph:

Definition 4 *A $(n, m, D, \gamma, \alpha)$ bipartite expander graph is a D-left-regular bipartite graph $G = (L \cup R, E)$ where $|L| = n$, $|R| = m$ (with $m \leq n$) and $\forall S \subseteq L$ such that $|S| \leq \gamma \cdot n$, the neighborhood $\mathrm{N}(S)$ of S satisfies $|\mathrm{N}(S)| \geq \alpha \cdot |S|$.*

We can note some trivial bounds on the expander parameters α and γ: since the graph is D-left-regular (every vertex in L has degree D), we must have $\alpha \leq D$ since for any $S \subseteq L$, we must have $\mathrm{N}(S) \leq D \cdot |S|$. Similarly, if we let $\alpha = D$, then we must have $\gamma \leq \frac{1}{D}$ (a set $S \subseteq L$ with $|S| > \frac{n}{D}$ can't possibly have a neighborhood satisfying $\mathrm{N}(S) \geq \alpha \cdot |S| > n > m$). So we can already see that there must be some inherent dependency between the parameters α and γ.

A question we can ask now is whether such expander graphs actually exist. The answer is clearly yes, since a complete bipartite graph, for instance, is a bipartite expander graph. However, for a such a complete graph we have $D = m$. What we are thus striving for is an expander where D is constant, while retaining high expansion parameters α and γ.

We now present a probabilistic argument for the existence of such bipartite expander graphs which also yields a randomized algorithm for constructing an expander graph. Interestingly, while this probabilistic method has been known for a long time, it was only in 2002 that a first deterministic construction of expander graphs was given [1].

Theorem 5 *For any $\epsilon > 0$, $m \leq n$, there exist $\gamma > 0$ and $D \geq 1$, such that a $(n, m, D, \gamma, (1-\epsilon)D)$ bipartite expander graph exists.*
Additionally, let

$$\gamma \cdot n = \frac{m}{D \cdot e^{1/\epsilon}}$$

and

$$D = f(\epsilon) \cdot \log(\frac{n}{m})$$

for some function f.

Proof Let L and R be two set of vertices such that $|L| = n, |R| = m$ and $m \leq n$, then we can construct a D-left-regular bipartie graph G by choosing D neighbours uniformly at random for each $v \in L$.

As G is a D-left-regular bipartite graph, we are interested in when G is not a $(n, m, D, \gamma, \alpha)$ bipartite expander graph. That occurs if there is a subset $S \subseteq L$ and a subset $M \subseteq R$ such that

(i). $|S| \leq \gamma \cdot n$;

(ii). $\mathrm{N}(S) \subset M$ and $|M| = \alpha \cdot |S|$.

The neighborhood of S must be a strict subset of M. We could consider M to be smaller and always extend it by enough vertices $v \in R/(\mathrm{N}(S) \cup M)$, so that $|M| = \alpha \cdot |S|$. Let $\mathrm{P}(S, M)$ be the indicator variable for this event, we have that

$$\mathrm{P}(S, M) = \begin{cases} 1 & \text{if } \mathrm{N}(S) \subset M \\ 0 & \text{otherwise} \end{cases}.$$

Now if we look at the probability of that event, we can see that it is bounded as follow

$$\Pr[\mathrm{P}(S, M)] \leq \left(\frac{|M|}{m}\right)^{D \cdot |S|}.$$

This inequality is derived from the fact that the vertices from S can connect to at most $D \cdot |S|$ vertices in R. The probability that an edge from S to R falls into M is exactly $\frac{|M|}{m}$. If all edges were independent (which is clearly not the case since we can't have two edges between the same vertex pair), the probability that each edge from S to R falls into M would be exactly $\left(\frac{|M|}{m}\right)^{D \cdot |S|}$. Here we are interested in an upper bound only which is provided by this independent scenario.

Negating the statement of definition 4, we can say that the graph G which we sampled is not a bipartite expander graph *iff* $\exists S \subseteq L$, $|S| \leq \gamma n$, $\exists M \subseteq R$, $|M| = \alpha |S|$, such that $\mathrm{N}(S) \subset M$, where $\alpha = (1-\epsilon)D$. Informally, we can find some subset S which doesn't have a large enough neighborhood. We will now compute

the probability that our graph G is not an expander:

$$\begin{aligned}
\Pr[G \textit{ is not an expander}] &\leq \sum_{1\leq s\leq \gamma\cdot n}\ \sum_{\substack{S\subseteq L\\ |S|=s}}\ \sum_{\substack{M\subseteq R\\ |M|=\alpha|S|}} \Pr[\mathrm{P}(S,M)] && \text{follows from a union bound}\\
&\leq \sum_{s=1}^{\gamma n}\binom{n}{s}\binom{m}{\alpha s}\left(\frac{\alpha s}{m}\right)^{Ds}\\
&\leq \sum_{s=1}^{\gamma n}\left(\frac{ne}{s}\right)^{s}\left(\frac{me}{\alpha s}\right)^{\alpha s}\left(\frac{\alpha s}{m}\right)^{Ds} && \text{standard upper bound for combinations}\\
&= \sum_{s=1}^{\gamma n}\left[\left(\frac{ne}{s}\right)\left(\frac{me}{\alpha s}\right)^{\alpha}\left(\frac{\alpha s}{m}\right)^{D}\right]^{s}\\
&\leq \sum_{s=1}^{\infty} x^{s} = \frac{x}{1-x} \qquad \text{if} \quad |x|<1 \qquad \text{where } x=\left(\frac{ne}{s}\right)\left(\frac{me}{\alpha s}\right)^{\alpha}\left(\frac{\alpha s}{m}\right)^{D}
\end{aligned}$$

We now bound the term x by a constant:

$$\begin{aligned}
x &= \left(\frac{ne}{s}\right)\left(\frac{(1-\epsilon)D\cdot s\cdot e^{1/\epsilon}}{m}\right)^{\epsilon D} && \text{since } \alpha=(1-\epsilon)D\\
&\leq \left(\frac{e}{\gamma}\right)\left(\frac{(1-\epsilon)D\cdot \gamma\cdot n\cdot e^{1/\epsilon}}{m}\right)^{\epsilon D} && \text{since } s\leq \gamma n\\
&\leq \left(\frac{n\cdot D\cdot e^{1+1/\epsilon}}{m}\right)(1-\epsilon)^{\epsilon D} && \text{since } \gamma=\frac{m}{n\cdot D\cdot e^{1/\epsilon}}
\end{aligned}$$

Setting $D=\dfrac{\log_{\frac{1}{(1-\epsilon)}}(10e^{1/\epsilon+1}Dn/m)}{\epsilon}$, we get

$$x\leq\frac{1}{10},\quad \Pr[G \textit{ is not an expander}]\leq\frac{1}{8}.$$

Note that we give a definition of D in terms of D. In it's closed form, D is related to the *Lambert-W* Function and grows as $f(\epsilon)\cdot log(n/m)$ which is what the theorem definition states.

Therefore, we have proved that with a positive probability, G is an expander graph and thus expander graphs satisfying the conditions stated in our theorem must exist! ■

3 Applications

3.1 Error Correcting Codes

Consider the case where *Alice* must send a message to *Bob* over a noisy channel that can flip a fraction of the k bits of the original message. Knowing that the channel is noisy, *Alice* adds redundancy by encoding the message into $n > k$ bits hoping that *Bob* will be able to decode it correctly even in the presence of flipped bits. We will soon see how to use expander graphs to build such a code.

Definition 6 *The* rate *of a code* $C\subseteq\{0,1\}^n$ *is* $\dfrac{\log_2|C|}{n}$.

Definition 7 *The* distance δ *of a code* $C\subseteq\{0,1\}^n$ *is* $\min\limits_{c_1\neq c_2}\dfrac{d_H(c_1,c_2)}{n}$, *where* d_H *is the Hamming distance.*

Note that there is a tradeoff to be made between rate and distance. When the rate is high, the distance tends to be small, and *vice versa*. A code is said to be asymptotically good if it has constant rate and distance. We can obtain a asymptotically good code using bipartite expanders.

Given a bipartite D-left-regular graph $G = (L \cup R, E)$ with $L = \{a_1, \ldots, a_n\}$, $R = \{b_1, \ldots, b_m\}$ and $m = \frac{n}{2}$, we define C as follows:

$$(x_1, \ldots, x_n) \in C \iff \forall j \in \{1, \ldots, m\} \sum_{i:(a_i,b_j)\in E} x_i \equiv 0 \mod 2$$

Using this, we get m constraints on the elements of the code. Each constraint can divide by at most 2 the number of elements which belong to C, thus we have a rate of at least $\frac{\log_2(2^n \cdot 2^{-m})}{n} = 1 - \frac{m}{n} = \frac{1}{2}$.

Lemma 8 *If G is a $(n, m, D, \gamma, (1-\epsilon)D)$ bipartite D-left-regular expander, with $\epsilon < \frac{1}{2}$, then the distance δ of the code is equal to γ.*

Proof For the sake of contradiction, suppose $\exists x, y \in C : d_H(x, y) < \gamma \cdot n$. Let $S = \{a_i \mid x_i \neq y_i\}$ be the elements of L where the corresponding bits of x and y disagree.

As both x and y are codewords, we have that:

$$\forall a_i \in S, j \in \{1, \ldots, m\} \sum_{(a_i,b_j)\in E} (x_i + y_i) \equiv 0 \mod 2$$

Thus, since x and y disagree on the bits indexed by S:

$$\forall a_i \in S, j \in \{1, \ldots, m\} \sum_{(a_i,b_j)\in E} 1 \equiv 0 \mod 2$$

Let $\mathrm{N}(S) \subseteq R$ be the set of neighbours of vertices in S. For every $b_j \in \mathrm{N}(S)$, the above sum is by definition not empty. Thus, every $b_j \in \mathrm{N}(S)$ must have at least 2 corresponding elements in S. This implies that there are at least $2|\mathrm{N}(S)|$ edges leaving S. As $2|\mathrm{N}(S)| \geq 2\alpha|S| > D|S|$, this leads to a contradiction, as S has $D|S|$ outgoing edges. ■

Remark There exists a linear decoding algorithm, described in [2]. The idea being to flip an x_i as long as there is a majority of neighbours of a_i whose corresponding constraints are not satisfied.

3.2 Saving Randomness

Suppose we have a probabilistic algorithm which when given an input x and a random string $r \in \{0,1\}^k$, outputs $f(x, r) = 1$ if x is a `YES`-instance and outputs $f(x, r) = 1$ with probability less than $\frac{1}{12}$ if x is a `NO`-instance. We want to be able to lower the rate of false positives without using additional randomness and so by only increasing the running time by a constant factor.

Let $n = 2^k$ and consider G, a $(n, n, D, \gamma, \alpha)$ bipartite expander, with $\alpha = \frac{D}{2}$ and $\gamma = \frac{1}{De^{1/2}} \geq \frac{1}{2D}$. For every $S \subseteq L$ such that $|S| \geq \frac{n}{6D}$, we have that $\mathrm{N}(S) \geq \alpha \cdot |S| \geq n/12$.

Assume x to be a `NO`-instance. Let us define $B = \{r \in \{0,1\}^k \mid f(x, r) = 1\}$ to be the set of random strings that incorrectly lead the algorithm to accept x. We have that $|B| \leq \frac{n}{12}$ since the probability of a false positive is bounded by $\frac{1}{12}$. We assign each element r of $\{0,1\}^k$ to a unique distinct vertex of L and to a unique distinct vertex of R. For every random string r, we define $r_1, \ldots, r_D$ to be the D neighbours of the vertex in L that corresponds to r.

Our modified algorithm, on input $(x; r)$, will now run the original algorithm on the D random strings associated to r and output

$$\begin{cases} 1 & \text{if} \quad f(x, r_1) = f(x, r_2) = \cdots = f(x, r_D) = 1 \\ 0 & \text{otherwise} \end{cases}$$

This algorithm will produce a false positive only if all $r_1, \ldots, r_D \in B$, that is if $\mathrm{N}(S) \subseteq B$. Let $S \subseteq L$ be the set of vertices whose neighbours are all in B. As $|B| \leq \frac{1}{12}$, we can upper bound $|S|$ by $\frac{n}{6D}$. Thus, the algorithm will produce false positives with probability at most $\frac{1}{6D}$. It will do so using only k random bits and will have a running time increased by a factor D.

4 Exercises

Note that to remain consistent with our discussions during the exercise session, in the following we will define the neighborhood $\mathrm{N}(S)$ of a set of vertices S in a slightly different manner than that of definition 3. In the exercise session, we used a different formulation (say $\mathrm{N}'(S)$) which was defined as $\mathrm{N}'(S) = \mathrm{N}(S) \cup S$ (the neighborhood of a set S is the union of the set S itself and all vertices adjacent to at least one vertex in S.

Definition 9 *A n vertex D-regular graph $G = (V, E)$ is an ϵ-node expander if for all $S \subset V$ such that $|S| \leq \frac{n}{2}$, we have $\mathrm{N}'(S) \geq (1+\epsilon)|S|$.*

Definition 10 *A n vertex D-regular graph $G = (V, E)$ is an ϵ-edge expander if for all $S \subset V$ such that $|S| \leq \frac{n}{2}$, we have $\delta(S) = |E(S, \overline{S})| \geq \epsilon \cdot D \cdot |S|$. Here, $\delta(S)$ denotes the number of edges in the cut between S and $\overline{S}$.*

Definition 11 *Given a Graph $G = (V, E)$, we define the diameter of G to be the "longest shortest path" between any two distinct vertices $u, v \in V$. More formally, if we denote by $d(u, v)$ the length of the shortest path from u to v in G, then*

$$\text{diameter}(G) = \max_{\substack{u,v \in V \\ u \neq v}} d(u, v).$$

If there is no path from u to v in G, we define $d(u, v) = \infty$. Furthermore, for any disconnected graph G, $\text{diameter}(G) = \infty$.

Problem 1 Show that a ϵ-node expander has diameter $O(\log n)$.

Solution Pick any $u, v \in V$. Set $S_0 = \{u\}$ and $T_0 = \{v\}$. From the fact that G is a ϵ-node expander, we know that $\mathrm{N}'(S_0) \geq (1+\epsilon)|S_0|$ and that $\mathrm{N}'(T_0) \geq (1+\epsilon)|T_0|$. So the set of vertices that u (respectively v) can reach in a single step is at least $(1+\epsilon)$. Applying this idea further, we get

$$\mathrm{N}'(S_1) \geq (1+\epsilon)|S_1| \geq (1+\epsilon)^2$$
$$\dots$$
$$\mathrm{N}'(S_k) \geq (1+\epsilon)^k$$

as long as $|S_k| \leq \frac{n}{2}$. Let k_u denote the value for k such that $|S_k| \leq \frac{n}{2}$ and $\mathrm{N}'(S_k) > \frac{n}{2}$. Then we know that more than half of all the vertices are reachable from u in at most k_u steps. Similarly, we denote by k_v the value for k such that $|T_k| \leq \frac{n}{2}$ and $\mathrm{N}'(T_k) > \frac{n}{2}$. Obviously, we must have $\mathrm{N}'(S_{k_u}) \cap N'(T_{k_v}) \neq \emptyset$. Thus, the path from u to v is of length at most $k_u + k_v$. We compute

$$(1+\epsilon)^{k_u} \leq \frac{n}{2} \leq \mathrm{N}'(S_{k_u})$$
$$k_u \leq \frac{1}{\log(1+\epsilon)} \log(n/2) = O\left(\frac{\log(n)}{log(1+\epsilon)}\right) = O(\log(n))$$

and the same bound holds for k_v. Thus, the path from u to v has length $O(\log n)$. Since u and v are arbitrary vertices from our graph, we can conclude that $\text{diameter}(G) = O(\log n)$.

Problem 2 Relate ϵ-node expanders to ϵ-edge expanders, and *vice versa*.

Solution Assume $G = (V, E)$ to be an ϵ-node expander and let $S \subset V$ be such that $|S| \leq |V|/2$. By definition, we have that $|\mathrm{N}'(S)| \geq (1+\epsilon)|S|$. Thus we must have at least $\epsilon|S|$ edges going out of S (thus $|E(S, \overline{S})| \geq \epsilon|S|$), making G a (ϵ/D)-edge expander.

Let's now assume $G = (V, E)$ to be an ϵ-edge expander and let $S \subset V$ be such that $|S| \leq |V|/2$. By definition, we have at least $\epsilon D|S|$ edges going out of S. As G is D-regular, those edges must at least reach $\epsilon D|S|/D = \epsilon|S|$ nodes outside of S, thereby making G a ϵ-node expander.

References

[1] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing*, STOC '02, pages 659–668, New York, NY, USA, 2002. ACM.

[2] M. Sipser and D. A. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42:1710–1722, 1996.