

Lecture 20

In which we prove properties of expander graphs.

1 Quasirandomness of Expander Graphs

Recall that if G is a d -regular graph, and A is its adjacency matrix, then, if we call $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ the eigenvalues of A with repetitions, we are interested in the parameter $\sigma_2(G) := \max_{i=2,\dots,n} \{|\lambda_i|\}$, and we have

$$\sigma_2(G) = \left\| A - \frac{d}{n}J \right\|$$

where J is the matrix with a one in each entry, and $\|\cdot\|$ is the matrix norm $\|M\| := \max_{x, \|x\|=1} \|Mx\|$.

Our first result today is to show that, when $\sigma_2(G)$ is small, the graph G has the following *quasirandomness* property: for every two disjoint sets S, T , the number of edges between S and T is close to what we would expect in a random graph of average degree d , that is, approximately $\frac{d}{|V|}|S||T|$.

For two (possibly overlapping) sets of vertices S, T , we define $edges_G(S, T)$ to be the number of edges with one endpoint in S and one endpoint in T , with edges having both endpoints in $S \cap T$, if any, counted twice.

Lemma 1 (Expander Mixing Lemma) *Let $G = (V, E)$ be a d -regular graph, and let S and T be two disjoint subsets of vertices. Then*

$$\left| edges_G(S, T) - \frac{d}{|V|} \cdot |S| \cdot |T| \right| \leq \sigma_2(G) \cdot \sqrt{|S| \cdot |T|}$$

PROOF: We have

$$edges_G(S, T) = \mathbf{1}_S^\top A \mathbf{1}_T$$

and

$$|S||T| = \mathbf{1}_S^\top J \mathbf{1}_T$$

so

$$\begin{aligned} & \left| \text{edges}_G(S, T) - \frac{d}{|V|} \cdot |S| \cdot |T| \right| \\ &= \cdot \left| \mathbf{1}_S^\top A \mathbf{1}_T - \frac{d}{|V|} \mathbf{1}_S^\top J \mathbf{1}_T \right| \\ &= \cdot \left| \mathbf{1}_S^\top \left(A - \frac{d}{|V|} J \right) \mathbf{1}_T \right| \\ &\leq \|\mathbf{1}_S\| \cdot \left\| A - \frac{d}{|V|} J \right\| \cdot \|\mathbf{1}_T\| \\ &= \sqrt{|S|} \cdot \sigma_2(G) \cdot \sqrt{|T|} \end{aligned}$$

□

Note that, for every disjoint S, T , we have $\sqrt{|S| \cdot |T|} \leq |V|/2$, and so the right-hand side in the expander mixing lemma is at most $\frac{\sigma_2(G)}{d} \cdot |E|$, which is a small fraction of the total number of edges if σ_2 is small compared to d .

2 Random Walks in Expanders

A *t-step random walk* is the probabilistic process in which we start at a vertex, then we pick uniformly at random one of the edges incident on the vertices and we move to the other endpoint of the edge, and then repeat this process t times.

If $P := \frac{1}{d}A$ is the normalized adjacency matrix of an undirected regular graph G , then $P(u, v)$ is the probability that, in one step, a random walk started at u reaches v . This is why the normalized adjacency matrix of a regular graph is also called its *transition matrix*.

Suppose that we start a random walk at a vertex chosen according to a probability distribution \mathbf{p} , which we think of as a vector $\mathbf{p} \in \mathbb{R}^V$ such that $\mathbf{p}(u) \geq 0$ for every u and $\sum_u \mathbf{p}(u) = 1$. After taking one step, the probability of being at vertex v is $\sum_u \mathbf{p}(u)P(u, v)$, which means that the probability distribution after one step is described by the vector $\mathbf{p}^\top \cdot P$, and because of the symmetric of P , this is the same as $P\mathbf{p}$.

Iterating the above reasoning, we see that, after a t -step random walk whose initial vertex is chosen according to distribution \mathbf{p} , the last vertex reached by the walk is distributed according to $P^t\mathbf{p}$.

The parameter σ_2 of P^t is equal to $(\sigma_2(G)/d)^t$, and so if G has a parameter σ_2 bounded away from d , and if t is large enough, we have that the parameter σ_2 of P^t is very small, and so M^t is close to $\frac{1}{n}J$ in matrix norm. If P^t was actually equal to $\frac{1}{n}J$, then $P^t \cdot \mathbf{p}$ would be equal to the uniform distribution, for every distribution \mathbf{p} . We would thus expect $P^t \cdot \mathbf{p}$ to be close to the uniform distribution for large enough t .

Before formalizing the above intuition, we need to fix a good measure of distance for distributions. If we think of distributions as vectors, then a possible notion of distance between two distributions is the Euclidean distance between the corresponding vectors. This definition, however, has various shortcomings and, in particular, can assign small distance to distributions that are intuitively very different. For example, suppose that \mathbf{p} and \mathbf{q} are distributions that are uniform over a set S , and over the complement of S , respectively, where S is a set of size $|V|/2$. Then all the entries of $\mathbf{p} - \mathbf{q}$ are $\pm 2/n$ and so $\|\mathbf{p} - \mathbf{q}\| = 2/\sqrt{n}$, which is vanishingly small even though distributions over disjoint supports should be considered as maximally different distributions.

A very good measure is the *total variation distance*, defined as

$$\max_{S \subseteq V} \left| \sum_{v \in S} \mathbf{p}(v) - \sum_{v \in S} \mathbf{q}(v) \right|$$

that is, as the maximum over all events of the difference between the probability of the event happening with respect to one distribution and the probability of it happening with respect to the other distribution. This measure is usually called *statistical distance* in computer science. It is easy to check that the total variation distance between \mathbf{p} and \mathbf{q} is precisely $\frac{1}{2} \cdot \|\mathbf{p} - \mathbf{q}\|_1$. Distributions with disjoint support have total variation distance 1, which is largest possible.

Lemma 2 (Mixing Time of Random Walks in Expanders) *Let G be a regular graph, and P be its normalized adjacency matrix. Then for every distribution \mathbf{p} over the vertices and every t , we have*

$$\|\mathbf{u} - P^t \mathbf{p}\|_1 \leq \sqrt{|V|} \cdot (\sigma_2(G)/d)^t$$

where \mathbf{u} is the uniform distribution.

In particular, if $t > \frac{dc}{d - \sigma_2(G)} \cdot \ln \frac{|V|}{\epsilon}$, then $\|\mathbf{u} - P^t \mathbf{p}\|_1 \leq \epsilon$, where c is an absolute constant.

PROOF: Let $\bar{J} = J/|V|$ be the normalized adjacency matrix of a clique with self-loops. Then, for every distribution \mathbf{p} , we have $\bar{J}\mathbf{p} = \mathbf{u}$. Recall also that $\sigma_2(G) = \|P - \bar{J}\|$.

We have

$$\begin{aligned}
& \| \mathbf{u} - M^t \mathbf{p} \|_1 \\
& \leq \sqrt{|V|} \cdot \| \mathbf{u} - P^t \mathbf{p} \| \\
& \leq \sqrt{|V|} \cdot \| \bar{J} \mathbf{p} - P^t \mathbf{p} \| \\
& \leq \sqrt{|V|} \cdot \| \bar{J} - P^t \| \cdot \| \mathbf{p} \| \\
& \leq \sqrt{|V|} \cdot (\sigma_2(G)/d)^t
\end{aligned}$$

□

The last result that we discussed today is one more instantiation of the general phenomenon that “if $\sigma_2(G)$ is small then a result that is true for the clique is true, within some approximation, for G .”

Suppose that we take a $(t - 1)$ -step random walk in a regular graph G starting from a uniformly distributed initial vertex. If G is a clique with self-loops, then the sequence of t vertices encountered in the random walk is a sequence of t independent, uniformly distributed, vertices. In particular, if $f : V \rightarrow [0,1]$ is a bounded function, the Chernoff-Hoeffding bounds tell us that the empirical average of $f()$ over the t points of the random walk is very close to the true average of $f()$, except with very small probability, that is, if we denote by v_1, \dots, v_t the set of vertices encountered in the random walk, we have

$$\mathbb{P} \left[\frac{1}{t} \sum_i f(v_i) \geq \mathbb{E} f + \epsilon \right] \leq e^{-2\epsilon^2 t}$$

where $n := |V|$. A corresponding Chernoff-Hoeffding bound can be proved for the case in which the random walk is taken over a regular graph such that $\sigma_2(G)$ is small.

Lemma 3 (Chernoff-Hoeffding Bound for Random Walks in Expanders) *Let $G = (V, E)$ be a regular graph, and (v_1, \dots, v_t) the distribution of t -tuples constructed by sampling v_1 independently, and then performing a $(t - 1)$ -step random walk starting at v_1 . Let $f : V \rightarrow [0, 1]$ be any bounded function. Then*

$$\mathbb{P} \left[\frac{1}{t} \sum_i f(v_i) \geq \mathbb{E} f + \epsilon + \frac{\sigma_2(G)}{d} \right] \leq e^{-\Omega(\epsilon^2 t)}$$

We will not prove the above result, but we briefly discuss one of its many applications.

Suppose that we have a polynomial-time probabilistic algorithm A that, on inputs of length n , uses $r(n)$ random bits and then outputs the correct answer with probability, say, at least $2/3$. One standard way to reduce the error probability is to run the

algorithm t times, using independent randomness each time, and then take the answer that comes out a majority of the times. (This is for problems in which we want to compute a function exactly; in combinatorial optimization we would run the algorithm t times and take the best solutions, and in an application in which the algorithm performs an approximate function evaluation we would run the algorithm t times and take the median. The reasoning that follows for the case of exact function computation can be applied to the other settings as well.)

On average, the number of iterations of the algorithms that give a correct answer is $\geq 2t/3$, and the cases in which the majority is erroneous correspond to cases in which the number of iterations giving a correct answer is $\leq t/2$. This means that the case in which the modified algorithm makes a mistake correspond to the case in which the empirical average of t independent 0/1 random variables deviates from its expectation by more than $2/3 - 1/2 = 1/6$, which can happen with probability at most $e^{-t/18}$, which becomes vanishingly small for large t .

This approach uses $t \cdot r(n)$ random bits. Suppose, instead, that we consider the following algorithm: pick t random strings for the algorithm by performing a t -step random walk in an expander graph of degree $O(1)$ with $2^{r(n)}$ vertices and such that $\sigma_2(G) \leq d/12$, and then take the majority answer. A calculation using the Chernoff bound for expander graphs show that the error probability is $e^{-\Omega(t)}$, and it is achieved using only $r(n) + O(t)$ random bits instead of $t \cdot r(n)$.