CS174 Spring 99 Lecture 20 Summary John Canny

Expander Graphs

A very important class of sparse graphs are **expander** graphs. Among other things, they are the model for a good network. Basically, an expander has the property that every subset of its vertices has a large set of neighbors. That implies that any pair of nodes are connected by a short path. Furthermore, removing random edges in the graph (simulating connection failures) does not reduce this property by much, so a network which is an expander is fault-tolerant. Expanders often have the following property:

Defn A graph G is d-regular if every vertex in G has degree d.

Defn A d-regular graph $G_d(V, E)$ is a c-expander if every subset $S \subseteq V$ of $\leq |V|/2$ vertices satisfies $|N(S) - S| \geq c|S|$, where $N(S)$ is the set of vertices which are neighbors of vertices in S.

To see that an expander makes a good network, suppose that you want to route a message from a vertex a to another vertex b in G. Every vertex has d neighbors. By the properties of an expander, there are $(1 + d)(1 + c)$ vertices at distance ≤ 2 from a. Working out from there, there will be $\geq (1+d)(1+c)^k$ vertices at distance $\leq k+1$ from a. You can continue expanding from a until the reachable set of vertices V_a has $> |V|/2$ vertices. The vertex b may not be among them. But if you expand from b in the same way, you eventually obtain a set V_b of $> |V|/2$ vertices reachable from b. The sets V_a and V_b both contain $> |V|/2$ vertices so they must overlap. The overlap contains vertices on a path from a to b.

In this way, we have shown that for any pair of vertices a and b, there is a path of length $\leq 2(k + 1)$ from a to b, where $k = \log_{(1+c)} |V|/2d$. The larger the value of c, the shorter the path between any two vertices.

Random d-regular graphs typically make good expanders. But one problem with generating graphs randomly is that there is no direct way to compute c, so you cant check directly whether the random graph is a good expander (and generate another it if it isnt). Fortunately, there is another property of graphs which can be easily computed, and which gives a lower bound on c. It is based on the eigenvalues of a matrix derived from G.

Eigenvalues and Expanders

Defn The **adjacency matrix** of a graph $G = (V, E)$ with $|V| = n$ is an $n \times n$ symmetric matrix A, where

$$A_{ij} = \begin{cases} 1 & \text{if edge } \{i, j\} \in E \\ 0 & \text{otherwise} \end{cases}$$

Recall the definition of an eigenvalue of a matrix:

Defn If A is an $n \times n$ matrix, then λ is an **eigenvalue** of A if

$$Av = \lambda v$$

for some n-vector v, called the **eigenvector** associated with λ.

In future, instead of saying "eigenvalue of the adjacency matrix of G" we will just say "eigenvalue of G".

For a d-regular graph, the largest eigenvalue is $\lambda = d$, corresponding to an eigenvector which is all ones. This follows because every row of A has exactly d ones, so multiplying A by a vector of all ones gives a vector of all d's. For G to be an expander, the critical quantity turns out to be the **second largest eigenvalue** of (the adjacency matrix of) G. Let $\lambda = \lambda(G)$ henceforth denote the second-largest eigenvalue of G.

Theorem If λ is the second-largest eigenvalue of a d-regular graph $G = (V, E)$, then G is a c-expander with

$$c = (d - \lambda)/2d$$

Since the second-largest eigenvalue is positive, the upper bound for the expansion factor is $1/2$. Typical expander graphs will approach that factor quite closely. Before proving the theorem we need to establish a lemma. Let X and Y be two subsets of V. Let $E(X, Y)$ denote the number of edges in E between a vertex in X and a vertex in Y. Then

lemma For every partition of V into disjoint subsets X and Y,

$$E(X, Y) \geq \frac{(d - \lambda)|X||Y|}{|V|}$$

Proof Let $n = |V|$, $x = |X|$, $y = |Y| = n - x$. In what follows, we will associate a vector with each vertex in V. Assuming the vertices are numbered $1, \ldots, n$, the vector associated with v is $(0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 is the number of the vertex. We will use v for both the vertex and its vector. Now let $D = dI$ be a diagonal matrix with d along its diagonal, and let w be any n-vector.

$$\begin{aligned} w^T(D - A)w &= \textstyle\sum_{u \in V} d(u^T w)^2 - 2\sum_{\{u,v\} \in E}(u^T w)(v^T w) \\ &= \textstyle\sum_{\{u,v\} \in E}(u^T w - v^T w)^2 \end{aligned}$$

Now consider the w defined as

$$u^T w \quad \begin{cases} -y & \text{if} \quad u \in X \\ x & \text{if} \quad u \in Y \end{cases}$$

If v is an eigenvector of A, $Av = \lambda v$, and so $(D - A)v = (d - \lambda)v$. In other words A and $D - A$ have the same eigenvectors. The eigenvalues of $(D - A)$ are of the form $d - \lambda$. Since the largest eigenvalue of A is d, the smallest eigenvalue of $(D - A)$ is 0. The second-largest eigenvalue of A corresponds to the second-smallest eigenvalue of $(D - A)$. All other eigenvalues of $(D - A)$ are positive.

The vector w that we chose above is orthogonal to the eigenvector $(1, \ldots, 1)$ which has eigenvalue 0 for $(D - A)$. That means w can be expressed as a linear combination of the other eigenvectors of $(D - A)$, which are orthogonal, because $(D - A)$ is a real, symmetric matrix:

$$w = \sum_{i=2,\ldots,n} \alpha_i a_i$$

where the a_i are unit eigenvectors, α_i are scalars, and a_1 is the eigenvector of all ones. Let $\lambda_1 < \lambda_2 \cdots$ be the eigenvalues of $(D - A)$, then

$$\begin{aligned} w^T(D-A)w &= \textstyle\sum_{i=2,\ldots,n} \lambda_i \alpha_i^2 a_i^T a_i \\ &\geq \lambda_2 \textstyle\sum_{i=2,\ldots,n} \alpha_i^2 \\ &= (d-\lambda) w^T w \\ &= (d-\lambda)(xy^2 + yx^2) \\ &= (d-\lambda)xyn \end{aligned}$$

Earlier in the proof, we showed that

$$w^T(D-A)w = \sum_{\{u,v\}\in E} (u^T w - v^T w)^2$$

which is $= E(X,Y)(x+y)^2 = E(X,Y)n^2$. So

$$w^T(D-A)w = E(X,Y)n^2 \geq (d-\lambda)xyn$$

which after a final rearrangement, gives

$$E(X,Y) \geq (d-\lambda)xy/n \qquad\qquad QED$$

Now we can complete the proof of the main theorem:

Proof (theorem) Let $S \subset V$ be any subset of $s \leq n/2$ vertices of G. By the lemma above, taking $X = S$, $Y = V - S$, there are at least

$$[(d-\lambda)s(n-s)]/n \geq [(d-\lambda)s]/2$$

edges from S to its complement. Since no vertex in the complement is adjacent to more than d of these,

$$|N(S)| \geq [(d-\lambda)s]/2d \qquad QED$$

It is very helpful to be able to verify that a random graph is a good expander. Since it is cheap to generate another random graph, it is much better to do that than to work with a graph that is a poor expander. We cant check directly that a graph is a good expander efficiently. But now we know that if the graph has a small second eigenvalue, then it is a good expander. The eigenvalues of an n-vertex graph can be computed in $O(n^3)$ time, and even less if the graph is sparse. So a simple scheme for computing expanders is to generate a random graph, compute its eigenvalues, and accept if the second-largest eigenvalue is small enough, otherwise, to generate another random graph.

We have assumed so far that expanders are d-regular graphs. A random d-regular graph (i.e. a sample chosen with equal probability from the space of all d-regular graphs) will have good probability of having a small second eigenvalue, as shown by the next theorem, which we will not prove. Let λ denote the second-largest eigenvalue of G.

Theorem (Broder and Shamir 1987) For a random d-regular graph,

$$Pr[\lambda \leq 3d^{\frac{3}{4}}] \quad \geq \quad 1 - O(e^{-\frac{1}{2}d^{\frac{1}{2}}})$$

From the earlier theorem, we know that such a graph has expansion factor c given by

$$c \geq (d - \lambda)/2d$$

And since λ grows as $O(d^{\frac{3}{4}})$, the expansion factor is good for most random graphs.

It remains for us to come up with a good scheme for generating d-regular random graphs. We will assume for simplicity that d is even. It is tricky to generate perfectly random d-regular graphs, but a good approximation is achieved by thinking of a d-regular graph as a union of $d/2$ graphs which are 2-regular. The 2-regular graphs are defined via permutations. Let $\pi : \{1, \ldots, n\} \rightarrow \{1, \ldots, n\}$ be a random permutation of n vertices. We add the edge $\{i, \pi(i)\}$ to G for every i. Then every vertex has degree two, because it is the source and destination for some argument of the permutation. There is one exception, in that if $\pi(i) = i$, we have a self-loop. These do not affect the bounds on λ, so its OK for the permutation to have self-loops. Also, when we take the union of $d/2$ 2-regular graphs, we may encounter an edge $\{i, j\}$ more than once. This again doesnt affect the bounds. So we allow a d-regular random graph to have both self-loops and repeated edges.

Algorithm for generating random d-regular graphs Assume d is even.

- Let G be an n-vertex graph with no edges.
- Generate $d/2$ random permutations π_i of $\{1, \ldots, n\}$
- For each permutation π_i, add all edges of the form $\{j, \pi(j)\}$ for $j = 1, \ldots, n$ to G.

Recall that the expected number of self-loops in a random permutation is 1, independent of n. So for the procedure above, the expected number of self-loops is $d/2$.

For edge duplicates, notice that we are adding $nd/2$ edges out of a possible $n(n-1)/2$ edges. Letting $N = n^2/2$, that means placing $d\sqrt{N}/\sqrt{2}$ balls into N boxes, so the expected number of collisions (birthday problem) is about $d^2/4$

So overall, the number of self-loops or multiple edges in a graph generated this way are fixed functions of d. As n grows, the fraction of such edges $\rightarrow 0$.