18.409 An Algorithmist's Toolkit September 29, 2009

Lecture 6

Lecturer: Jonathan Kelner *Scribe: Anthony Kim(2009)*

Topics

- Diameters and their relationship to λ_2
- Expanders
- Butterfly networks

1 Diameters and Eigenvalues

So far, every time we've dealt with eigenvalues, it's had something to do with connectivity. For example, the spectral gap can be used to approximate the quality of cuts; it also describes how well a graph can mix under a random walk. They both are saying similar things: the eigenvalue is saying how connected a graph is. A walk will mix quickly if there's a lot connected to everything else. The min-cut will likewise be large if there's a lot of connectivity.

For almost every reasonable property about a graph, there's something you can write down regarding its relation to the second eigenvalue of the Laplacian. Today, we're going to show the relation between λ_2 and the diameter of a graph.

Definition 1 *The diameter, δ, is the longest, shortest path between any two vertices of a graph. In other words, we can define the distance between two vertices u and v in G as the shortest path connecting the two. The diameter of G is the largest distance between any two vertices in G.*

It's not immediately clear why the diameter should be related to λ_2; the following provides intuition[1]:

1. Well-connected graphs have big λ_2
2. Well-connected graphs have a small δ
3. So, graphs with big λ_2 should have small δ

Before proceeding, we'll be making the following assumption:

Assumption 2 *G is a d-regular graph (this is for simplicity and not really a limiting assumption).*

We'll also be utilizing lazy random walks in our investigation. As a reminder,

Definition 3 *A lazy random walk is simply a random walk along a graph with self loops added in:*

$$M = \underbrace{\frac{A}{2d}}_{Random\ Walk} + \underbrace{\frac{I}{2}}_{Self\text{-}loops} \tag{1}$$

where A is the graph's adjacency matrix and I is the identity matrix.

Since we're using adjacency matrices, the interesting eigenvalues will be close to 1. So, let μ_2 be the second largest eigenvalue of M and λ be the gap (i.e., $\lambda = 1 - \mu_2$).

[1]Note, however, that this is not a proper syllogism

1.1 A First Bound

Claim 4 *For any G,*

$$\delta \leq \frac{\ln(n)}{\lambda} \tag{2}$$

where δ is the diameter of G, n is the number of vertices in the graph, and $\lambda = 1 - \mu_2$, where μ_2 is the second largest eigenvalue of the M associated with G.

Claim 4 means that λ, up to a logarithmic factor, really does provide a direct bound on the diameter of the graph. For most graphs, this isn't very tight, but it's a good place to start. So, why is it true?

Proof We will use random walks to prove claim 4. Let u and v be vertices that are *as far as possible from one another*. Start a random walk at u and let $p_t(v)$ be the probability of a walk being at vertex v at time t. If $p_t(v) > 0$, then $\delta \leq t$. Intuitively, this means that if we start at u and, after t time steps, there is some probability of ending up at v (the farthest vertex from u), then there must be a path of length t between the two. If there wasn't, $p_t(v)$ would be 0.

Recalling that the stationary distribution is $\pi(v) = 1/n$ for regular graphs, we can equivalently state that if $|p_t(v) - \pi(v)| < \frac{1}{n}$, then $\delta \leq t$. Why? Because if $|p_t(v) - \pi(v)| < \frac{1}{n}$, then $p_t(v) > 0$, implying $\delta \leq t$.

Recall from our earlier lecture on random walks that

$$|p_t(v) - \pi(v)| < (1-\lambda)^t \sqrt{\frac{d(v)}{\min_y d(y)}} = (1-\lambda)^t$$

Since G is regular, $d(v) = d$ for all vertices (allowing the last equality above). We'll now look at what happens when we set $t = \frac{\ln n}{\lambda}$:

$$(1-\lambda)^t = (1-\lambda)^{\frac{\ln n}{\lambda}} < \left(\frac{1}{e}\right)^{\ln n} = \frac{1}{n}$$

With the inequality coming from the fact that $(1-\lambda)^{1/\lambda} < \frac{1}{e}$ for all $\lambda > 0$. Thus, for $t = \frac{\ln n}{\lambda}$, we have that $|p_t(v) - \pi(v)| < \frac{1}{n}$, and therefore, $\delta \leq t = \frac{\ln n}{\lambda}$. ■

1.2 A better bound

As stated earlier, bounding δ by $\frac{1}{n}$ is not that great. So, can we do better? Yes. And we do so by using an important trick that frequently comes up. First, note that if the $(u,v)^{\text{th}}$ entry of A^k is non-zero, then there's a path of at most length k from u to v. Replacing A with M doesn't change this (it just makes the non-zero entries smaller). If e_u and e_v are basis vectors, then

$$|p_k(v) - \pi(v)| = |e_v^T M^k e_u - \frac{1}{n}| < \frac{1}{n}$$

which would imply $\delta \leq k$. Let's then let $p(x)$ be a polynomial of degree k:

$$p(x) = \sum_{j=1}^{k} c_j x^j$$

Note that we can also interpret M as a variable and apply p to it as follows:

$$p(M) = \sum_{j=1}^{k} c_j M^j$$

If $p(M)$ has no zero entries, then $\delta \leq k$. Why? Because all non-zero elements of M indicate all vertices that can be reached in one step, M^2 is all vertices that can be reached in two steps, and so on. Thus,

for any non-zero entry (u, v) in $p(M)$, there must have been a non-zero element in M^j for $0 < j \leq k$, implying the existence of a path from u to v of at most length k. If this is true for all entries in $p(M)$, then a path of at most length k exists from any vertex to any other vertex, which means the diameter is at most k.[2]

Claim 5 *Suppose p has degree k, $p(1) = 1$, and $|p(\mu_i)| < \frac{1}{n}$ for all $i \geq 2$, then*

$$\delta \leq k \tag{3}$$

Proof For this proof, it is sufficient to show that every entry in $p(M)$ is non-zero. First, recall that we can write down any matrix, M, as the following:

$$M = \sum_i \mu_i v_i v_i^T$$

where μ_i and v_i are the i-th eigenvalue and eigenvector, respectively. Since

$$M^k = \sum_i \mu_i^k v_i v_i^T$$

we can write

$$p(M) = \sum_{j=0}^{k} c_j M^j = \sum_{j=0}^{k} c_j \sum_i \mu_i^j v_i v_i^T = \sum_i \left(\sum_{j=0}^{k} c_j \mu_i^j \right) v_i v_i^T = \sum_i p(\mu_i) v_i v_i^T$$

Therefore, we can write the $(a, b)^{\text{th}}$ entry of $p(M)$ as follows

$$\begin{aligned}
e_a^T p(M) e_b &= e_a^T \left(\sum_i p(\mu_i) v_i v_i^T \right) e_b \\
&= \sum_i p(\mu_i)(e_a^T v_i)(v_i^T e_b) \\
&= \sum_i p(\mu_i) v_i(a) v_i(b) \\
&= \frac{1}{n} + \sum_{i=2}^{n} p(\mu_i) v_i(a) v_i(b) \\
&\geq \frac{1}{n} - \left| \sum_{i=2}^{n} p(\mu_i) v_i(a) v_i(b) \right| \\
&\geq \frac{1}{n} - \sum_{i=2}^{n} |p(\mu_i)| \, |v_i(a)| \, |v_i(b)| \\
&\geq \frac{1}{n} - \max_{i \geq 2} |p(\mu_i)| \sum_{i=2}^{n} |v_i(a)| \, |v_i(b)| \\
&\geq \frac{1}{n} - \max_{i \geq 2} |p(\mu_i)| \\
&> 0
\end{aligned}$$

Where the penultimate step follows from $\sum_i |v_i(a)| \, |v_i(b)| \leq 1$. Let V be the matrix where rows are the eigenvectors v_i's. Then $V \cdot V^T = I$ by the orthonormal condition. It follows that $V^T \cdot V = I$ and the columns

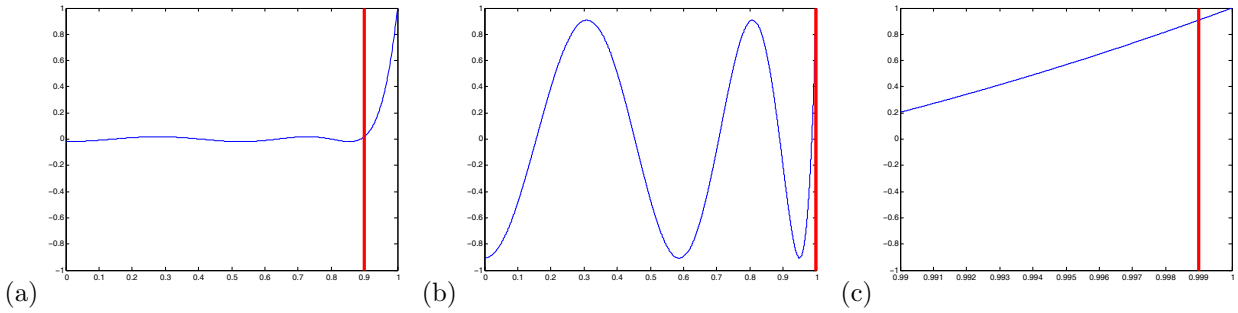
[2]Note that we're not saying anything about δ if $p(M)$ has zero entries. Since there are no restrictions on c_j, it's possible that the summation produces a zero entry for $p(M)$ where for all positive c_j a non-zero entry would have existed.

of V form an orthonormal basis. Hence $\left(\sum_{i=2}^{n}|v_i(a)||v_i(b)|\right)^2 \leq \left(\sum_{i=2}^{n}|v_i(a)|^2\right)\left(\sum_{i=2}^{n}|v_i(b)|^2\right) \leq 1$. The ultimate step follows from our assumption that $|p(\mu_i)| < \frac{1}{n}$ for all $i \geq 2$. Thus, if $p(1) = 1$ and $|p(\mu_i)| < \frac{1}{n}$ for all $i \geq 2$, we have that every entry in $p(M)$ is non-zero, implying that $\delta \leq k$. ■

Claim 6 *For any $t \in (0,1)$, I assert the existence of a magic polynomial, $p_k^{(t)}$, with the following properties:*

1. *$p_k^{(t)}$ is of degree k*
2. *$p_k^{(t)}(1) = 1$*
3. *$\left|p_k^{(t)}(x)\right| \leq 2\left(1+\sqrt{2t}\right)^{-k}$ for any $x \in [0, 1-t]$*

We will provide no proof for this claim here, but the polynomials are derived from Chebyshev polynomials, and we'll use them again later. To provide some intuition, figure 1 shows graphs of these polynomials for $k = 10$ with varying t. Notice that to keep the same bound for smaller values of t, a larger k is required due to the "oscillations" that the polynomial must take on in order to achieve $p(1) = 1$ while keeping $p(x)$ small for $x \in [0, 1-t]$.



(a) (b) (c)

Figure 1: (a) $t = 0.1$ (b) $t = 0.001$ (c) $t = 0.001$ zoomed in for x from 0.99 to 1

If we set $t = \lambda$, we get a degree k polynomial, p, such that

1. $p(1) = 1$
2. $|p(x)| \leq 2\left(1+\sqrt{2\lambda}\right)^{-k}$ for any $x \in [0, \mu_2]$.

Additionally, if we set $k = \left(1+\frac{1}{\sqrt{2\lambda}}\right)\ln(2n)$, then it is possible to show that $p(x) < \frac{1}{n}$ for all $x \in [0, \mu_2]$, which gives the following bound:

$$\delta \leq \left(1+\frac{1}{\sqrt{2\lambda}}\right)\ln(2n) \tag{4}$$

This is much better than our previous bound of $\delta \leq \frac{\ln(n)}{\lambda}$. So, strangely, by putting in a particular polynomial, we get a bound that grows with $1/\sqrt{\lambda}$ as opposed to just $1/\lambda$. This foreshadows our next unit on iterative linear algebra.

1.3 Example Application

Suppose that you have a symmetric matrix M and want the eigenvector associated with the largest eigenvalue. For purposes of this example, let them be normalized such that the largest eigenvalue is 1. Then an easy way to get an approximate answer for the eigenvector is to compute $M^k x$ for a large k and a random x.

Why does this give the eigenvector associated with $\mu_1 = 1$? If all other eigenvalues are less than 1, then for a large enough k, they will diminish in importance, until all that is left is v_1. This is a very intuitive and natural algorithm that takes about $1/\lambda$ steps to get close.

But we just found a much faster algorithm! Assuming that we know some good bound on λ (if we don't, we could easily search for it), we can compute $p_k^{(\lambda)}(M)x$ instead of $M^k x$ to get the dominant eigenvector. This method converges much faster, and we'll get into this more in a few lectures.

2 Expanders

If you had to know one set of graphs in your life, these are the ones to know. They often are a counterexample to many long-standing conjectures. Also, they turn up literally everywhere. If you didn't know any better, you would think that they don't exist from the described properties. But they're almost every single graph. Specifically, we'll be looking at families of d-regular graphs $(G_n)_n$ as n goes to infinity:

Definition 7 *$(G_n)_n$ is an expander family if $\lambda_2(G_n) \geq c$ for some constant c and for all n.*

Most of the graphs we've looked at are not expanders. For example, path graphs have $\lambda_2 \leq O(1/n^2)$ and binary trees have $\lambda_2 \leq O(1/n)$. This means that λ_2 very quickly goes to zero as $n \to \infty$ for both cases. Expanders don't have this property. Even as $n \to \infty$, λ_2 stays above a constant. Given this, it's not clear that they should exist.

Note: We should think of d as a constant. In other words, we'll pick a d and study expanders in that family.

2.1 Relating Expanders to Cuts

The first thing we'll look at is Cheeger's inequality for expanders. Recall that

$$\frac{\lambda_2}{2} \leq \phi(G)$$

For expanders, this implies

$$\frac{c}{2} \leq \phi(G)$$

What does this mean? It means that any set, S, of vertices with $|S| \leq n/2$ has at least $(c/2)|S|$ edges leaving it. This is a strong property: for expanders, there are no small cuts that can be made in the graph. Every cut that balances the sizes of the sets of vertices cuts a constant fraction of the edges in the graph.

The other side of Cheeger's inequality says

$$\Theta(1)\frac{\phi(G)^2}{d} \leq \lambda_2$$

Again, for expanders, this can be rewritten.

$$\phi(G) \leq \sqrt{\frac{cd}{2\Theta(1)}}$$

Since d is a constant, this says that the isoperimetry, ϕ, is also bounded above by a constant. Normally, there's a large gap between the upper-bound and lower-bounds in Cheeger's inequality. Here we've sandwiched ϕ between two constants. Therefore, an equivalent definition of expanders is as follows:

Definition 8 *$(G_n)_n$ is expander family if $\phi(G) \geq c'$ for some constant c' and all n.*

2.2 Do Expanders Exist?

A natural question is if expanders exist, what are the required parameters of the graphs? It turns out that random graphs are expanders, and so, almost all graphs are expanders. But, there's a limit as to how "good" of an expander you can have:

Claim 9 *For any G,*

$$\lambda_2 \leq d - 2\sqrt{d-1} + o(1) \tag{5}$$

In other words, even though λ_2 is always larger than a constant, there's a limit as to how well-connected it can be. We won't prove this. But this is not that strong a bound. We already know that $\lambda_2 \leq d$. This is just $O(\sqrt{d})$ smaller. Furthermore, this bound is tight since Ramanujan graphs meet it. So that's the limit of what an expander can be.

2.3 Expanders and Randomness

Expanders are all over the study of randomness, but we'll just study one interesting property. We'll use $\mu_2 = d - \lambda_2$ to simplify formulas, where now, μ_2 is the second largest eigenvalue of the adjacency matrix. Suppose you make a graph by randomly including each edge with probability d/n. In other words, construct a graph such that each vertex has an expected number of d edges leaving it. Since the total number of possible edges is $|S||T|$ and there's a d/n probability of having each edge, the expected number of edges between any two sets S and T will be $\frac{d|S||T|}{n}$.

Claim 10 *Expander Mixing Lemma: If you choose any two vertex sets, S and T, the difference in the total number of edges between the two and the expected number for a random graph is bounded. Formally,*

$$\begin{aligned}\left|e(S,T) - \frac{d|S||T|}{n}\right| &\leq \frac{\mu_2}{n}\sqrt{|S|\left|\bar{S}\right||T|\left|\bar{T}\right|} \\ &\leq \mu_2\sqrt{\min(|S|,\left|\bar{S}\right|)\cdot\min(|T|,\left|\bar{T}\right|)}\end{aligned}$$

This is surprising because there is no randomness here. This is just a property associated with expanders, but it behaves similarly to random graphs.

Proof Let α and β be the fraction of total vertices that are in the sets S and T:

$$|S| = \alpha n \qquad |T| = \beta n.$$

Let x and y be the characteristic vectors of S and T, respectively.

Definition 11 *A characteristic vector, x, of a set S is a vector of length n that has $x_i = 1$ if $v_i \in S$, and $x_i = 0$ otherwise.*

We can now write the number of edges between sets S and T as $e(S,T) = x^T A y$. Now, as you've probably noticed, it's beneficial to use vectors that are perpendicular to the all-ones vector, $\mathbf{1}$. So, we'll rewrite x and y as

$$v = x - \alpha\mathbf{1} \qquad w = y - \beta\mathbf{1}.$$

Clearly, $v \cdot \mathbf{1} = w \cdot \mathbf{1} = 0$, implying orthogonality. Rewriting the number of edges between sets S and T, we get

$$\begin{aligned}e(S,T) &= x^T A y \\ &= (v + \alpha\mathbf{1})^T A (w + \beta\mathbf{1}) \\ &= v^T A w + v^T A \beta\mathbf{1} + \alpha\mathbf{1}^T A w + \alpha\mathbf{1}^T A \beta\mathbf{1}.\end{aligned}$$

Using the following identities,

$$A\mathbf{1} = d\mathbf{1} \qquad \mathbf{1}^T A = d\mathbf{1}^T,$$

we get

$$\begin{aligned} e(S,T) &= v^T Aw + \beta v^T A\mathbf{1} + \alpha \mathbf{1}^T Aw + \alpha\beta \mathbf{1}^T A\mathbf{1} \\ &= v^T Aw + \beta v^T d\mathbf{1} + \alpha d\mathbf{1}^T w + \alpha\beta d\mathbf{1}^T\mathbf{1} \\ &= v^T Aw + \alpha\beta dn, \end{aligned}$$

where we have used the orthogonality of v and w with $\mathbf{1}$ (i.e., $v^T\mathbf{1} = 0$) to cancel out the middle two terms. We now have the following bound:

$$\begin{aligned} |e(S,T) - \alpha\beta dn| &= \left|v^T Aw\right| \\ &\leq |v||Aw| \\ &\leq |v|\mu_2|w| \\ &= \frac{\mu_2}{n}\sqrt{(\alpha n)\left((1-\alpha)n\right)(\beta n)\left((1-\beta)n\right)} \\ &= \frac{\mu_2}{n}\sqrt{|S|\left|\bar{S}\right||T|\left|\bar{T}\right|}, \end{aligned}$$

where the third line follows from the fact that w is orthogonal to $\mathbf{1}$, and thus, μ_2 is the largest eigenvalue that can affect w, and the fourth line follows from the fact that $|v| = \sqrt{n\alpha(1-\alpha)}$. To see this, note that

$$\begin{aligned} |v| &= |x - \alpha\mathbf{1}| \\ &= \sqrt{|S|(1-\alpha)^2 + \left|\bar{S}\right|(-\alpha)^2} \\ &= \sqrt{\alpha n(1-\alpha)^2 + (1-\alpha)n\alpha^2} \\ &= \sqrt{\alpha n(1-\alpha)(1-\alpha+\alpha)} \\ &= \sqrt{\alpha n(1-\alpha)}, \end{aligned}$$

and the same steps show that $|w| = \sqrt{n\beta(1-\beta)}$. Thus, we have shown that $|e(S,T) - \alpha\beta dn| \leq \frac{\mu_2}{n}\sqrt{|S|\left|\bar{S}\right||T|\left|\bar{T}\right|}$.■

2.4 Some Properties We Now Know

- Random walks on expanders mix in a logarithmic number of steps
- Expanders have logarithmic diameter
- Expanders have a constant isoperimetric number

2.5 Vertex Expansion

We've discussed cutting a graph and looking at the number of edges cut. An equivalent way of thinking about a cut is to select a set of vertices and then count the number of edges with one vertex in the set and one out. Another useful metric can be obtained by counting the number of vertices that are neighbors of a set. In other words, for a set of vertices, X, let $N(X)$, be the set of vertices, Y, such that $(x,y) \in E$ such that $x \in X$ and $y \in \bar{X}$.

Claim 12

$$N(X) \geq \frac{d^2|X|}{\mu^2 + (d^2 - \mu^2)|X|/n}$$

We don't prove this, but the high level idea is the following:

- Select a set of vertices from G. Call this set X.
- Let Y be the set of vertices that are neither in X nor in $N(X)$. In other words, $Y = V \backslash (N(X) \cup X)$.
- Now, by construction, we have that $e(X, Y) = 0$.

Algebra gets a little messy, but you can just plug the above into the expander mixing lemma to show this bound. It turns out also that for X/n small and $\mu = 2\sqrt{d-1}$, we can achieve

$$N(X) \geq \frac{d}{4}|X|.$$

Why is this interesting? What this is saying is that for any set X, there are at least $d/4$ neighbors not in X. Since each vertex has d neighbors total, this bound is quite strong. It turns out that this is about as good as you can get with spectral graph theory. To see this, we will generalize the vertex expansion as follows.

We want to show bounds of the form $|N(S)| \geq \gamma|S|$. In other words, we want to say that the vertex expansion of G is greater than or equal to γ for any S. Sometimes we'll only care about expansions of smaller sets (e.g., for $|S| \leq 0.01n$).

Definition 13 *G is an (α, β)-expander if for $\alpha\beta < 1$ and all sets S with $|S| \leq \alpha n$ have $|N(S)| \geq \beta|S|$.*

We showed that Ramanujan graphs are $(\alpha, d/4)$ expanders for some constant α. Some applications need expansion greater than $d/2$ but with small (constant) α. These exist, but we can't prove better than $d/2$ with spectral techniques.[3]

2.6 Bipartite Expanders

Many of the applications of expanders use bipartite expanders. These are just expanders that are bipartite graphs. It is easier to show that these exists (it will be a homework problem!).

Definition 14 *A d-regular bipartite graph is an (α, β)-expander if every set S on the left with $|S| \leq \alpha n$ has $N(S) \geq \beta|S|$.*

Whenever $\alpha\beta < 1$, there exists some d such that almost all d-regular graphs on n nodes (for n sufficiently large), are (α, β)-expanders.

[3] It turns out that random graphs work here. In 2002, Capalbo, Reingold, Vadhan, and Wigderson gave an explicit construction technique with expansion $d - o(1)$.

18.409 Topics in Theoretical Computer Science: An Algorithmist's Toolkit
Fall 2009