Notes for Lecture 11

1 Eigenvalues, Expansion, and Random Walks

As usual by now, let $G = (V, E)$ be an undirected d-regular graph with n vertices, M be its transition matrix, $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of M, and $x_1, \ldots, x_n$ be a system of orthonormal eigenvectors.

Let p be a probability distribution over vertices V, and consider the following process: pick at random a vertex v according to p, then perform a t-step random walk in G starting from v. We would like to address the following questions:

1. Is there a clean formula that specifies the distribution of the final vertex of the walk in terms of p, M and t?
2. For large t, does the distribution of the final vertex converge to a fixed distribution independent of p?
3. If so, what is this distribution?
4. And how fast is the convergence?

As we will see, the answers are

1. Yes, if we write distributions as row vectors, then it's pM^t.
2. Yes (provided the graph is connected and not bipartite, both necessary conditions).
3. The uniform distribution.
4. It depends on $\max\{\lambda_2, -\lambda_n\}$, or just on λ_2 if we take a "lazy" random walk.

Let us begin with the first question. If $p \in \mathbb{R}^V$ is a row vector that represents a probability distribution over vertices (hence, $p(v) \geq 0$ for all v, and $\sum_v p(v) = 1$), then consider the vector pM. Its v-th entry is

$$(pM)(v) = \sum_u p(u)M(u, v)$$

which clearly represents the probability of reaching v by first picking a vertex u according to distribution p, and then moving to a random neighbor of u. In particular, pM is itself a probability distribution.

Reasoning inductively, we see that, for every $t \geq 1$, pM^t is a probability distribution, and it represents the distribution of the final vertex in a t-step random walk in G that starts at a vertex selected accoding to p.

It follows that, in order to understand random walks in G, we need to understand the action of the matrix M^t.

Fortunately, we already know the eigenvalues and the eigenvectors of M^t: the matrix M^t has eigenvalues $\lambda_1^t, \ldots, \lambda_n^t$, and eigenvectors $x_1, \ldots, x_n$.

If p is a probability distribution, let us write $p = \alpha_1 x_1 + \alpha_n x_n$. Then

$$pM^t = \alpha_1 x_1 + \alpha_2 \lambda_2^t x_2 + \ldots + \alpha_n \lambda_n^t x_n$$

Where $x_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)$ and $\alpha_1 = p \cdot x_1^T = \frac{1}{\sqrt{n}} \sum_v p(v) = \frac{1}{\sqrt{n}}$, so that $\alpha_1 x_1 = (\frac{1}{n}, \ldots, \frac{1}{n})$ is the uniform distribution, which we shall denote by p_U from now on.

Now we would like to argue that pM^t converges to p_U for large t. Indeed consider the vector $pM^t - p_U$, which measures the "non-uniformity" of pM^t; its length is

$$\begin{aligned} ||pM^t - p_U|| &= ||\alpha_2 \lambda_2^t x_2 + \ldots + \alpha_n \lambda_n^t x_n|| \\ &= \sqrt{\alpha_2^2 \lambda_2^{2t} + \cdots \alpha_n^2 \lambda_n^{2t}} \\ &\leq \max_{i=2,\ldots,n} |\lambda_i|^t \cdot \sqrt{\alpha_2^2 + \cdots + \alpha_n^2} \\ &\leq \max_{i=2,\ldots,n} |\lambda_i|^t \cdot ||p|| \\ &\leq \max_{i=2,\ldots,n} |\lambda_i|^t \end{aligned}$$

Where we use the fact that if p is a probability distribution then

$$||p|| = \sqrt{\sum_v p^2(v)} \leq \sqrt{\sum_v p(v)} = 1$$

Let $\bar{\lambda}_2 := \max_{i=2,\ldots,n} |\lambda_i| = \max\{\lambda_2, -\lambda_n\}$ be the second largest eigenvalue in absolute value. Our calculation shows that

$$||pM^t - p_U|| \leq \bar{\lambda}_2^t$$

and so if choose $t = O(\frac{1}{1-\bar{\lambda}_2} \log n)$ then we can have, say

$$||pM^t - p_U|| \leq \frac{1}{100n}$$

and, in particular, $\left|pM^t(v) - \frac{1}{n}\right| \leq \frac{1}{100n}$ for every v. This means that it only takes $t = O(\frac{1}{1-\bar{\lambda}_2} \log n)$ steps for a random walk to converge to the uniform distribution, and that the diameter of G is upper bounded by $O(\frac{1}{1-\bar{\lambda}_2} \log n)$ for a stronger reason.

It is instructive to see that any bound on the convergence of random walks must depend both on $1 - \lambda_2$ and on $1 - |\lambda_n|$.

Suppose, for starters, that G has two connected components S, $V - S$ and that p is uniform on S. Then $pM = p$, $pM^t = p$, and a random walk will never converge to the uniform distribution. (Algebraically, we have $\lambda_2 = 1$, $\alpha_2 \neq 0$.) This demonstrates the necessity of a dependence on $1 - \lambda_2$.

Suppose know that $G = (V, E)$ is a complete bipartite graph with bipartition $A, V - A$, such that $n/2$ vertices are on each side. (Hence G is regular of degree $n/2$.) Let us construct the eigenvectors and eigenvalues of G. We now that $x_1 := \frac{1}{\sqrt{n}}(1, \ldots, 1)$ is an eigenvector for the eigenvalue 1. Consider now the vector x_n such that $x_n(v) = \frac{1}{\sqrt{n}}$ if $v \in A$ and $x_n(v) = -\frac{1}{\sqrt{n}}$ if $v \notin A$. Then we can verify that $x_1 \perp x_n$, and that x_n is an eigenvector of eigenvalue -1. Finally, we note that a vector x is orthogonal to both x_1 and x_n if and only if $\sum_{v \in A} x(v) = \sum_{v \notin A} x(v) = 0$, and that any such vector is an eigenvector of eigenvalue 0. Choose $x_2, \ldots, x_{n-1}$ to be an orthonormal basis for the above-described space of dimension $n - 2$, and so we have constructed an orthonormal set of eigenvectors $x_1, \ldots, x_n$ for the eigenvalues $\lambda_1 = 1$, $\lambda_2 = \cdots \lambda_{n-1} = 0$, $\lambda_n = -1$.

Consider the distribution p that is uniform on A. Then pM is uniform on $V - A$, pM^2 is uniform on A, pM^3 is uniform on $V - A$, and so on, and the random walk does not converge to the uniform distribution. To see what happens algebraically, $p = \frac{1}{\sqrt{n}} x_1 + \frac{1}{\sqrt{n}} x_n$, so $pM^t - p_U = \lambda_n^t \frac{1}{\sqrt{n}} x_n = (-1)^t \frac{1}{\sqrt{n}} x_n$.

The theory that we have developed so far connects edge expansion with $1 - \lambda_2$, but not with $1 - |\lambda_n|$. There is, however, a simple trick that allows to relate edge expansion to the behavior of random walks.

For a d-regular graph G with transition matrix M, define the *lazy random walk* on G as the random walk of transition matrix $M_L := \frac{1}{2} I + \frac{1}{2} M$; equivalently, we can think of it as a standard random walk on the $2d$-regular graph G_L that is identical to G except that every vertex has d self-loops.[1]

If M has eigenvalues $\lambda_1 = 1, \ldots, \lambda_n$ and eigenvectors $x_1, \ldots, x_n$, then it is easy to see that $x_1, \ldots, x_n$ are also eigenvectors for M_L, and that the eigenvalues of M_L are $\frac{1}{2} + \frac{1}{2}\lambda_1 = 1, \frac{1}{2} + \frac{1}{2}\lambda_2, \ldots, \frac{1}{2} + \frac{1}{2}\lambda_n$. In particular, all the eigenvalues of M_L are non-negative, and so $\bar{\lambda}_2(M_L) = \lambda_2(M_L) = \frac{1}{2} + \frac{1}{2}\lambda_2(M)$.

It follows that if G has normalized edge expansion h, then $\bar{\lambda}_2(M_L) \leq 1 - \frac{h^2}{4}$.

If G is connected, then every cut is crossed by at least one edge, and so the normalized edge expansion is at least inverse-polynomial.

$$h(G) = \min_{S \subseteq V, |S| \leq \frac{|V|}{2}} \frac{edges(S, V - S)}{d|S|} \geq \frac{1}{d \cdot \frac{n}{2}} = \frac{2}{nd}$$

From Cheeger's inequality, we have that the gap between largest and second-largest eigenvalue is also inverse-polynomial

$$1 - \lambda_2 \geq \frac{h^2}{2} \geq \frac{2}{n^2 d^2}$$

And so, in M_L, we have

$$1 - \bar{\lambda}_2(M_L) = 1 - \lambda_2(M_L) = \frac{1}{2} - \frac{1}{2}\lambda_2(M) \leq 1 - \frac{1}{n^2 d^2}$$

This means that, in a connected graph, a lazy random walk of length $O(n^2 d^2 \log n)$ reaches a nearly uniformly distributed vertex.

[1] The name "lazy" random walk refers to the fact that the random walk M_L behaves essentially like the random walk M, except that at every step there is a probability $\frac{1}{2}$ of doing nothing.

Another conclusion that we can reach is that the diameter of G_L is at most $O\left(\frac{2}{1-\lambda_2}\log n\right)$, and the same bound applies to G, since G and G_L have the same diameter. If G has normalized edge expansion h, then G has diameter at most $O(h^{-2}\log n)$.

2 The Expander Mixing Lemma

So far we have studied properties (and characterizations) of graphs in which λ_2 and $\bar{\lambda}_2$ are bounded away from 1. Graphs satisfying the stronger requirement that $\bar{\lambda}_2$ is close to zero enjoy a number of additional, sometimes surprising, properties. Such graphs are also useful in a number of applications, and the question of explicitly constructing such graphs having bounded degree is well studied (and it will be the subject of the next two lectures). For now we note that once we have constructed a family of arbitrary large d-regular graphs satisfying $\lambda_2 \leq 1-\epsilon$ for some fixed $\epsilon > 0$, then we also immediately get a family of $2^k d^k$-regular graphs satisfying $\bar{\lambda}_2 \leq \left(1-\frac{\epsilon}{2}\right)^k$, because given a d-regular graph G satisfying $\lambda_2(G) \leq 1-\epsilon$, we can first convert it into the $2d$-regular graph G_L satisfying $\bar{\lambda}_2(G_L) \leq 1-\frac{\epsilon}{2}$ and then take the k-th power of G_L. (The k-th power of a graph G with transion matrix M is the graph G^k whose transition matrix is M^k: G^k has one edge for every length-k path in G.)

One of the main results about graphs with small $\bar{\lambda}_2$ is the *Expander Mixing Lemma*.

Lemma 1 (Expander Mixing Lemma) *Let G be a d-regular graph with n vertices and second largest eigenvalue in absolute value $\bar{\lambda}_2$, let A, B be two disjoint sets of vertices. Then*

$$\left|edges(A,B) - d\frac{|A||B|}{n}\right| \leq \bar{\lambda}_2 \cdot d \cdot \sqrt{|A| \cdot |B|} \leq \bar{\lambda}_2 \cdot d \cdot n \tag{1}$$

Thus, in a very good expander, the number of edges between any two sufficiently large sets of vertices is approximately what it would be in a random d-regular graph.

Although there is a short direct proof, it is instructive to use the proof of the lemma as an opportunity to introduce the following notion.

Definition 1 (Matrix Norm) *Let M be an $n \times m$ matrix, then we define its norm as*

$$||M|| := \max_{x \in \mathbb{R}^n} \frac{||xM||}{||x||} = \max_{x \in \mathbb{R}^n,\ ||x||=1} ||xM|| \tag{2}$$

Note that if M is the transition matrix of an undirected regular graph, then $||M|| = 1$. Define J_n to be the $n \times n$ matrix that has a 1 in each entry. We drop the subscript n when it's clear from the context.

Claim 2 *Let M be the transition matrix of an undirected graph G, and let $\bar{\lambda}_2$ be the second largest eigenvalue of M in absolute value. Then*

$$\bar{\lambda}_2 = ||M - \frac{1}{n}J|| \tag{3}$$

The point of the claim is that when $\bar{\lambda}_2$ is small then M is "close" to $\frac{1}{n}J$, which is the transition matrix of the complete graph.

Note that a random walk in $\frac{1}{n}J$ reaches, already at the first step, a random vertex, so this is consistent with the intuition that a random walk in a graph with small $\bar{\lambda}_2$ converges rapidly to the uniform distribution.

Let us now turn to the proof of the Expander Mixing Lemma. Let $1_A \in \mathbb{R}^n$ be the vector such that $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ otherwise; define 1_B similarly. We have

$$edges(A, B) = 1_A d M 1_B^T \tag{4}$$

and

$$1_A \frac{1}{n} J 1_B^T = \frac{1}{n} \cdot |A| \cdot |B| \tag{5}$$

so that

$$\left| edges(A, B) - d\frac{|A||B|}{n} \right| = d \left| 1_A \left(M - \frac{1}{n} J \right) 1_B^T \right|$$

and, applying Caucy-Schwarz and the definition of norm, we have

$$\left| 1_A \left(M - \frac{1}{n} J \right) 1_B^T \right| \leq ||1_A \left(M - \frac{1}{n} J \right) || \cdot ||1_B|| \leq ||1_A|| \cdot ||M - \frac{1}{n} J|| \cdot ||1_B||$$

And now we see that

$$||1_A|| = \sqrt{|A|}$$
$$||1_B|| = \sqrt{|B|}$$
$$||M - \frac{1}{n} J|| = \bar{\lambda}_2$$

3 Exercises

1. Prove Claim 2

   [Hint: the maximizing vector is the eigenvector of $\bar{\lambda}_2$]