CS 830: Randomness and Complexity	3/8/05

Lecture 15: Spectral Analysis of Random Walks

instructor: Jin-Yi Cai — Scribe: Bess Berg, Michael Kowalczyk

In this lecture we consider a random walk on a regular graph. We show the rate of convergence to the uniform distribution depends on the second largest eigenvalue of the adjacency matrix of the graph.

1 A useful lemma

Ultimately we would like to understand the properties of a random walk on a regular graph, but first we will prove a useful lemma. Let $G = (V, E)$ be a d-regular graph with $|V| = n$, and let M be the adjacency matrix for G. For any $A, B \subseteq V$ let $E(A, B)$ be the set of all edges between A and B; that is,

$$E(A,B) = \{(a,b) |\ a \in A, b \in B, (a,b) \in E\}.$$

Recall from last lecture that for randomly, independently chosen A and B of fixed sizes (say $|A| = k$ and $|B| = \ell$) we have that the expected value of $|E(A, B)|$ is $(d/n)|A||B|$. Now we give an upper bound for the difference between this expected value and the actual value of $|E(A, B)|$.

Lemma 1.1 $\big||E(A,B)| - \frac{d}{n}|A||B|\big| \leq \lambda\sqrt{|A||B|}$, *where λ is the second largest eigenvalue of M.*

Corollary 1.2 $\big||E(A,B)| - \frac{d}{n}|A||B|\big| \leq \lambda n$, *where λ is the second largest eigenvalue of M.*

Proof of Corollary:

$$\begin{aligned}
\big||E(A,B)| - \frac{d}{n}|A||B|\big| &\leq \lambda\sqrt{|A||B|} \qquad \text{(by the lemma)}\\
&\leq \lambda\frac{|A|+|B|}{2} \qquad \text{(geometric mean is at most arithmetic mean)}\\
&\leq \lambda n \qquad \text{(both } |A| \text{ and } |B| \text{ are at most } n).
\end{aligned}$$

♣

Proof of Lemma: Let $\chi_A, \chi_B \in \{0,1\}^n$ be the characteristic vectors of sets A and B respectively (so the i^{th} component of χ_A is 1 if and only if vertex v_i is in A, etc.). We claim that $|E(A,B)| = \chi_A^T M \chi_B$. To see this note that $\chi_A^T M$ is a vector whose i^{th} component is the number of edges from v_i to A. Then multiplying $\chi_A^T M$ with χ_B simply picks out the components i such that $v_i \in B$ and sums over them, for a total that is exactly the number of edges between A and B. We note that expressing a graph-related value such as $|E(A, B)|$ in terms of the graph matrix M is a standard trick.

We now employ another standard trick, which is to take advantage of the spectral analysis of M. In particular, M is a symmetric matrix and therefore has a full set of real eigenvalues,

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, with corresponding eigenvectors $v_1, v_2, \ldots, v_n$ that make up an orthonormal basis for $\mathbb{R}^n$. Because the graph is d-regular, the largest eigenvalue is $\lambda_1 = d$ with corresponding eigenvector $v_1 = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)^T$. We express χ_A and χ_B with respect to this basis,

$$\chi_A = \sum_{i=1}^{n} \langle \chi_A, v_i \rangle v_i, \qquad \chi_B = \sum_{i=1}^{n} \langle \chi_B, v_i \rangle v_i,$$

noting that $\langle \chi_A, v_1 \rangle = \frac{1}{\sqrt{n}}|A|$ and $\langle \chi_B, v_1 \rangle = \frac{1}{\sqrt{n}}|B|$. We obtain:

$$\begin{aligned}
|E(A,B)| &= \chi_A^T M \chi_B \\
&= \Big(\sum_{i=1}^{n} \langle \chi_A, v_i \rangle v_i^T\Big) M \Big(\sum_{i=1}^{n} \langle \chi_B, v_i \rangle v_i\Big) \\
&= \Big(\sum_{i=1}^{n} \langle \chi_A, v_i \rangle v_i^T\Big) \Big(\sum_{i=1}^{n} \langle \chi_B, v_i \rangle \lambda_i v_i\Big).
\end{aligned}$$

Because the eigenvectors v_i are orthogonal, we can multiply these two sums and obtain only n terms; that is, $v_i^T v_j = 1$ if and only if $i = j$ and otherwise $v_i^T v_j = 0$:

$$\begin{aligned}
&= \sum_{i=1}^{n} \lambda_i \langle \chi_A, v_i \rangle \langle \chi_B, v_i \rangle \\
&= d\frac{|A||B|}{n} + \sum_{i=2}^{n} \lambda_i \langle \chi_A, v_i \rangle \langle \chi_B, v_i \rangle.
\end{aligned}$$

Note we have used another standard trick: we isolated the main term, the term for which $i = 1$. Letting $\lambda = \max_{i \geq 2}\{|\lambda_i|\}$ we complete the proof:

$$\begin{aligned}
\Big||E(A,B)| - \frac{d}{n}|A||B|\Big| &= \Big|\sum_{i=2}^{n} \lambda_i \langle \chi_A, v_i \rangle \langle \chi_B, v_i \rangle\Big| \\
&\leq \sum_{i=2}^{n} \big|\lambda_i \langle \chi_A, v_i \rangle \langle \chi_B, v_i \rangle\big| \\
&\leq \lambda \sum_{i=2}^{n} \big|\langle \chi_A, v_i \rangle \langle \chi_B, v_i \rangle\big| \\
&\leq \lambda \sqrt{\sum_{i=1}^{n} \langle \chi_A, v_i \rangle^2} \sqrt{\sum_{i=1}^{n} \langle \chi_B, v_i \rangle^2} \quad \text{(by the Cauchy-Schwartz inequality)} \\
&= \lambda ||\chi_A||_2 ||\chi_B||_2 \\
&= \lambda \sqrt{|A||B|}.
\end{aligned}$$

♣

Note that we can formulate this result so that it resembles the Hash Mixing Lemma from previous lectures. Let $\Gamma_i(x)$ denote the i^{th} neighbor of x. Then the three equations below are equivalent, where the last in particular is analagous to the Hash Mixing Lemma.

$$
\begin{aligned}
\left| |E(A,B)| - d\frac{|A||B|}{n} \right| &\leq \lambda n & (1)\\
\left| \frac{|E(A,B)|}{nd} - \frac{|A|}{n}\frac{|B|}{n} \right| &\leq \frac{\lambda}{d} & (2)\\
\left| \Pr_{\substack{x\in V\\ 1\leq i\leq d}}[x \in A, \Gamma_i(x) \in B] - \Pr_{x,y\in V}[x\in A, y\in B] \right| &\leq \frac{\lambda}{d} & (3)
\end{aligned}
$$

2 Karp-Pippenger-Sipser Generator

Recall that our ultimate goal is efficient amplification of the success probability of some algorithm deciding an RP language. On input x, such an algorithm takes some randomness, say a string z, and runs a polytime predicate $P(x,z)$ to decide whether x is in the language or not. Since the language is in RP, we are promised that $P(x,z) = 1$ with probability at least $1/2$ if x is in the language, and otherwise $P(x,z) = 0$ with probability 1 (the probabilities are taken over the random strings z). The idea of the Karp-Pippenger-Sisper generator is to let the strings correspond to vertices in a d-regular graph, and on any random seed z let $G(z) = \{z_i \in \Gamma(z)\}$; that is, $G(z)$ is the set of neighbors of z. Fix any input x in the language, and let $\overline{W_x}$ be the set of random strings z such that $P(x,z) = 0$. That is, $\overline{W_x}$ is the set of *bad* witnesses for x. We would like to show that, taken over all z, the probability that $G(z) \subseteq \overline{W_x}$ is small.

Define the set A:

$$
\begin{aligned}
A &= \{v \in V |\ \text{all neighbors of } v \text{ are in } \overline{W_x}\}\\
&= \{v |\ G(v) \subseteq \overline{W_x}\}
\end{aligned}
$$

Let $B = W_x$, and note that by the properties of the RP algorithm we have $|B| \geq n/2$. Also note that we have $E(A,B) = \varnothing$, so that $|E(A,B)| = 0$. We use the previous lemma to start off the following sequence of equivalent equations.

$$
\begin{aligned}
\left|0 - d|A||B|/n\right| &\leq \lambda\sqrt{|A||B|} & (4)\\
d|A||B|/n &\leq \lambda\sqrt{|A||B|} & (5)\\
\frac{d}{n}\sqrt{|A||B|} &\leq \lambda & (6)\\
\frac{d^2}{n^2}|A||B| &\leq \lambda^2 & (7)\\
|A|/n &\leq \left(\frac{\lambda}{d}\right)^2 \cdot \frac{n}{|B|} & (8)\\
|A|/n &\leq 2\left(\frac{\lambda}{d}\right)^2 & (9)
\end{aligned}
$$

(The last equation follows because $|B| \geq n/2$.) Then indeed, if λ is small, using neighbors of a random seed z is a good low-cost (in terms of randomness) way to obtain more trials of the algorithm.

3 Random Walks on Expanders

Let $\hat{M} = \frac{1}{d}M$, so that $\hat{M}$ is the stochastic matrix that gives the transition probabilites of a random walk on the graph G. We will use spectral analysis on $\hat{M}$ to show that no matter what initial distribution over the vectors we start with, after not too many applications of the random walk it will become very close to the uniform distribution.

Let $w = (w_1, \ldots, w_n)^T$ be a weight distribution on V, so that $\sum_i w_i = 1$ and $w_i \geq 0$ for all i. $\hat{M}_{ij}$ is the probability of the random walk moving to v_i given that it starts at v_j, and therefore with initial distribution w, the probability of ending up at vertex v_i is $\sum_j \hat{M}_{ij} w_j$. Then the vector $\hat{M}w$ gives the distribution after one step of the walk, starting from distribution w. In fact, it is not hard to see that starting in distribution w the distribution after k steps of the walk is $\hat{M}^k w$.

We can analyze the dynamics of the walk more closely by expressing w in terms of the eigenvectors of $\hat{M}$. (Note that the eigenvectors for $\hat{M}$ are the same as for M, just with their corresponding eigenvalues scaled by $1/d$.) Let $w = \sum \alpha_i v_i$, where $\alpha_i = \langle w, v_i \rangle$. Note that $\alpha_1 = \frac{1}{\sqrt{n}}$ and also that, since we have scaled by $\frac{1}{d}$, the corresponding eigenvalue for v_1 is $\lambda_1 = 1$. Once again we isolate the main term and write

$$\begin{aligned} \hat{M}^k w &= \alpha_1 \lambda_1^k v_1 + \sum_{i=2}^{n} \alpha_i \lambda_i^k v_i \\ &= \alpha_1 v_1 + \sum_{i=2}^{n} \alpha_i \lambda_i^k v_i. \end{aligned}$$

Notice that $\alpha_1 v_1 = \frac{1}{n}(1, 1, \ldots, 1)^T$, where $u \doteq \frac{1}{n}(1, 1, \ldots, 1)^T$ is the uniform distribution. So we have

$$\hat{M}w - u = \sum_{i=2}^{n} \lambda_i^k \alpha_i v_i.$$

Once again letting $\lambda = \max_{i \geq 2} |\lambda_i|$, we obtain:

$$\begin{aligned} ||\hat{M}^k w - u||_2 &= ||\sum_{i=2}^{n} \lambda_i^k \alpha_i v_i||_2 && (10) \\ &\leq \lambda^k ||\sum_{i=2}^{n} \alpha_i v_i||_2 && (11) \\ &\leq \lambda^k ||\sum_{i=1}^{n} \alpha_i v_i||_2 && (12) \\ &= \lambda^k ||w||_2 && (13) \\ &\leq \lambda^k. && (14) \end{aligned}$$

Line (11) follows from the following:

$$\begin{aligned}
||\sum_{i=2}^{n} \lambda_i^k \alpha_i v_i||_2^2 &= \sum_{i=2}^{n} (\lambda_i \alpha_i)^2 \\
&\leq (\lambda^k)^2 \sum_{i=2}^{n} \alpha_i^2 \\
&= (\lambda^k)^2 ||\sum_{i=2}^{n} \alpha_i v_i||_2^2;
\end{aligned}$$

and line (13) follows from the following:

$$\begin{aligned}
||w||_2 &= \sqrt{\sum_{i=1}^{n} w_i^2} \\
&\leq \sqrt{\sum_{i=1}^{n} w_i} \\
&= 1
\end{aligned}$$

The point is that after k applications of the random walk, the 2-norm distance of the resulting distribution from the uniform distribution is at most λ^k; and this is true for any initial distribution w! So if λ is small, say, λ is some constant bounded away from 1, then with $k = O(\log n)$ applications of the random walk we can get within $\lambda^k = 1/poly(n) \ll 1/n$ of the uniform distribution. This means that we have the desireable property that for every i we will have some reasonable probability of landing on v_i.