CS 683: Advanced Algorithms	**Random Walks on Graphs**
March 13, 2001	**Scribe: Kate Jenkins, Russ Woodroofe**

1 Introduction to Random Walks

Last time we finished discussing applications of eigenvalues on planar graphs. This leads us naturally into our next topic: random walks on graphs, and their applications.

It will be useful to consider random walks on large graphs to study actions on other objects:

Eg:

1) We will model card shuffling as a random walk on the $n!$ permutations of n objects.

2) We will look at a 2 dimensional lattice of particles (which will represent the states of some system)

Two representative questions we might ask (and which we will discuss over the next 2 lectures) are:

1) Is a graph G connected? Of course, we can check this in polynomial time. If G is large, however, we would also like to check this in small space. Deterministically, it is known how to do this in $O(\log^2 n)$ (or even a little better), although these algorithms are not polynomial in time. We will use random walk techniques to give a probabilistic algorithm which takes $O(\log n)$ space, and expected polynomial time.

2) (From card shuffling) Given a large set, we often would like to pick an element approximately uniformly at random. How quickly can we do this? The techniques we will use in this problem will also be useful in approximately counting the size of the set, if this is not known.

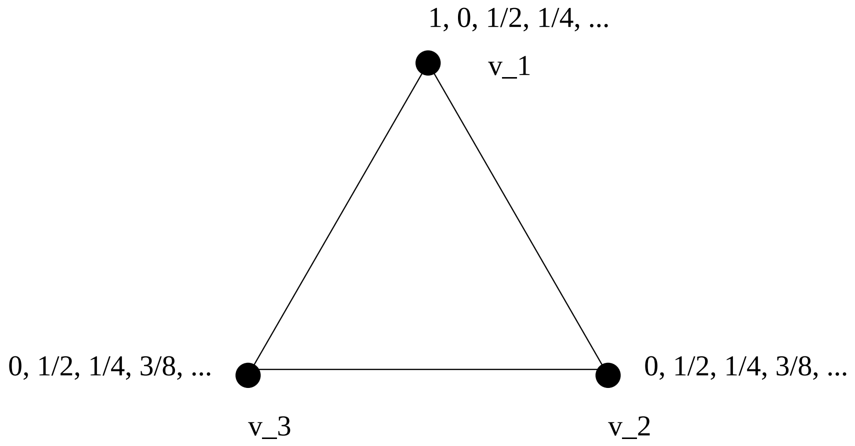
2 Basics

Suppose G is an undirected d-regular graph on n nodes. Then a *random walk* starts at some node v, chooses a neighbor w of v uniformly at random, moves to w, and repeats. Then after k steps, we have a probability distribution of which vertex one might be at. This corresponds with a vector $v^{(k)}$ with one coordinate for each node (representing the probability we are at that node) satisfying $\sum_{i \in V} v_i^{(k)} = 1$.

(More generally, we can consider starting our random walk with a probability distribution $v^{(0)}$. Then starting at node i corresponds to the distribution $v_i^{(k)} = 1$, $v_j^{(k)} = 0$ for $j \neq i$)

Example

Consider K_3, the triangle. We start at some vertex v_1 with probability 1. After 1 step, we have equal probability of walking towards each of the other vertices, so we are at v_1 with probability 0, and at v_2, v_3 with probability $\frac{1}{2}$ each. And so forth (pictured below, with the probability of being at a given vertex in steps 0 through 3 labelled)



Recall that the adjacency matrix A of a graph $G = (V, E)$ is an $n \times n$ matrix (where $n = |V|$) with $A_{ij} = 1$ if $(i, j) \in E$, 0 otherwise. The Laplacian $L = I - A$. We define $M = \frac{1}{d}A$ to be the *transition matrix* of G (so M_{ij} is the probability that we move from i to j in a step starting at i). So in the example above,

$$M = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

A *stationary distribution* of the random walk is a vector (probability distribution) σ which is unchanged by one step: ie, σ such that $M\sigma = \sigma$. Equivalently, σ is an eigenvector for M with eigenvalue 1 (and since it's a probability distribution, $\sum_{i \in V} \sigma_i = 1$, $\sigma_i \geq 0$). Then note the following:

Lemma 2.1 *A, M have the same eigenvectors, with eigenvalues scaled by $\frac{1}{d}$*

Proof. We have that λ an eigenvalue of A iff there is some x with $Ax = \lambda x$. But this occurs iff $Mx = \frac{1}{d}Ax = \frac{\lambda}{d}x$, ie, iff $\frac{\lambda}{d}$ is an eigenvalue of M (with the same eigenvalue x). ■

3 A Question

Recall from previous lectures that $\sigma = \frac{\overline{1}}{n}$ is a stationary distribution (since $\overline{1}$ an eigenvector for A). Also, σ is unique iff G is connected. This lets us ask the following question:

Question: If we start from an arbitrary initial distribution v, and iterate the random walk, will we converge to σ? (assuming G connected, so $\sigma = \frac{\overline{1}}{n}$ unique.)

Answer: As stated, no. For example, if $G = X \cup Y$ is bipartite, then we know whether we're in X or Y after k steps by the parity of k (so no convergence). (It turns out that this is the only thing that can go wrong, but as many of the graphs we will be interested in turn out to be bipartite, this is a serious drawback!)

The answer is yes, however, if we change our random walk to allow ourselves to "stall" (stay at the same vertex v) at any v with probability $\frac{1}{2}$. This new idea of random walk has transition matrix $M'_{ij} = \frac{1}{2d}$if $i, j \in E$, 0 if $i \neq j$ not in E, and $\frac{1}{2}$ if $i = j$. Thus, $M' = \frac{1}{2}I + \frac{1}{2d}A = \frac{1}{2}I + \frac{1}{2}M$. Furthermore, $Mx = \lambda x$ iff $(\frac{1}{2}I + \frac{1}{2}M)x = (\frac{1}{2} + \frac{\lambda}{2})x$, and thus M, M' have the same eigenvectors, and λ an eigenvalue of M corresponds to $\frac{1}{2} + \frac{1}{2}\lambda$ an eigenvalue of M'.

4 Bounding Eigenvalues

Recall our notation from last lecture: we let $\lambda_i = \lambda_i(L)$ be the ith eigenvalue of the Laplacian (arranged in increasing order $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \ldots \leq \lambda_n$). Then $\lambda_i(A) = d - \lambda_i$ (arranged in decreasing order), and hence $\lambda_i(M) = \frac{1}{d}(d - \lambda_i) = 1 - \frac{1}{d}\lambda_i$ and $\lambda_i(M') = 1 - \frac{\lambda_i}{2d}$. Denote this last $\lambda'_i = \lambda_i(M')$.

How large can λ_n be?

Claim 4.1 *The Laplacian matrix L of a d-regular graph G has $\lambda_n \leq 2d$ (and $\lambda_n = 2d \iff G$ is bipartite)*

Proof.

Let x be the n^{th} eigenvector, with corresponding eigenvalue λ_n.

$$Lx = \lambda_n x$$

What happens to the i^{th} coordinate when $x \to Lx$?

$$x_i \to \sum_{(i,j)\in E} (x_i - x_j)$$

Suppose x_i has the maximum absolute value. We can assume without loss of generality that $x_i > 0$. Then $\forall i, j, x_i - x_j \leq 2x_i$, so:

$$\lambda_n x_i = (Lx)_i = \sum_{(i,j)\in E} (x_i - x_j) \leq 2dx_i$$

(Note that the only way to have this be an equality is if all coordinates x_i have the same magnitude and edges only connect pairs of nodes with opposite sign. This means we have a bipartite graph.) ■

Since $\lambda'_i = 1 - \lambda_i/2d$, the claim implies that in M' we have:

$$1 = \lambda'_1 > \lambda'_2 \geq \lambda'_3 \geq \ldots \geq \lambda'_n \geq 0$$

(Note that $\lambda'_n \geq 0$ holds because $\lambda'_i = 1 - \lambda_i/2d$, which holds because we set the probability of "stalling" in the random walk to be 1/2. Had we chosen a smaller probability, λ'_n could be negative).

5 Convergence of Random Walks

5.1 Proof of eventual convergence

Now we are going to prove that regardless of the initial probability distribution, a random walk on a graph (with stalling) always converges to the stationary distribution σ. The stationary distribution σ is defined as before to be the eigenvector of M (and M') with eigenvalue 1. (Note that an eigenvector with eigenvalue 1 in M also has eigenvalue 1 in M').

Theorem 5.1 *A random walk on a d-regular graph G (with self-loops as in M') converges to σ from any initial distribution v.*

Proof. The eigenvectors of M' form a basis $\omega_1, \ldots, \omega_n$, so we can write v in terms of this basis: $v = \sum_i \alpha_i \omega_i$. Hence,

$$M'v = \sum_i \alpha_i M'\omega_i = \sum_i \alpha \lambda'_i \omega_i$$

Iterating for k steps,

$$v \to (M')^k v = \sum \alpha_i (\lambda'_i)^k \omega_i$$

For $i \neq 1$, we have $\lambda'_i < 1$ and hence $(\lambda'_i)^k \to 0$ as $k \to \infty$. These converge to 0 exponentially quickly, and the rate is determined by λ'_2 (the second largest eigenvalue).

$\lambda'_1 = 1$, so $\lim_{k\to\infty}(M')^k v = \alpha_1 \omega_1 = \sigma$.

■

5.2 Rate of convergence

How quickly does v converge to the stationary distribution?

We will use the following fact for $x \in \mathbf{R}^d$:

$$\|x\|_1 \leq \sqrt{d}\|x\|_2 \leq \sqrt{d}\|x\|_1$$

Let $q = \alpha_1\omega_1$ (= first term of above sum).

Then,

$$\begin{aligned}\left\|q - (M')^k v\right\|_1 &\leq \sqrt{n}\left\|q - (M')^k v\right\|_2 \\ &= \sqrt{n}\left\|\sum_{i=2}^{n} \alpha_i (\lambda'_i)^k \omega_i\right\|_2 \\ &= \sqrt{n}\Big(\sum_{i=2}^{n} \alpha_i^2 (\lambda'_i)^{2k}\Big)^{1/2}\end{aligned}$$

$$\leq \sqrt{n}\Big(\sum_{i=2}^{n} \alpha_i^2 (\lambda_2')^{2k}\Big)^{1/2}$$

$$= \sqrt{n}(\lambda_2')^k \Big(\sum_{i=2}^{n} \alpha_i^2\Big)^{1/2}$$

$$\leq \sqrt{n}(\lambda_2')^k$$

The last inequality follows since $(\sum_{i=2}^{n} \alpha_i^2)^{1/2} \leq \|v\|_2 \leq \|v\|_1 = 1$.

5.3 Maximum Relative Error

The maximum relative error per node is

$$\max_{i \in v} \frac{\Big|q_i - \Big[(M')^k v\Big]_i\Big|}{\Big|q_i\Big|}$$

q is the first eigenvector, so $q_i = 1/n$. Hence the maximum relative error is:

$$= n \max_{i \in v} \Big|q_i - \Big[(M')^k v\Big]_i\Big|$$

$$= n \Big\|q - (M')^k v\Big\|_\infty$$

$$\leq n \Big\|q - (M')^k v\Big\|_1$$

$$\leq n\sqrt{n}(\lambda_2')^k$$

$$= n^{1.5}(1 - \frac{\lambda_2}{2d})^k$$

How many steps (k) must we take to be within max relative error δ?

We need $n^{1.5}(1 - \frac{\lambda_2}{2d})^k \leq \delta$, so take

$$k = \frac{2d}{\lambda_2}\Big((3/2)\ln n + \ln(1/\delta)\Big) \leq \frac{3d}{\lambda_2}\Big(\ln n + \ln(1/\delta)\Big)$$

To bound k in terms of the graph expansion α, we use the fact that $\lambda_2 \geq \frac{\alpha^2}{2d}$(which was proved at great expense in a previous lecture). Thus we can take

$$k = \frac{3d}{\frac{\alpha^2}{2d}}\Big(\ln n + \ln(1/\delta)\Big) = \frac{6d^2}{\alpha^2}\Big(\ln n + \ln(1/\delta)\Big)$$

6 Connection to Markov Chains

We now present some of the more general phrasing and results used by mathematicians when talking about random walks.

A Markov Chain C has a state set V (assume $|V| = n$, finite) and transition matrix M, where M_{ij} = Probability of going from state i to j.

Let G_C be the "state graph" of C. The nodes of G_C are V. G_C contains a directed edge (i, j) if $M_{ij} > 0$. (Note that the matrix equivalent to iterating C is $v \rightarrow v^T M$).

We present two definitions.

Definition 6.1 *Markov Chain C is "irreducible" if the state graph G_C is strongly connected.*

Definition 6.2 *Markov Chain C is "aperiodic" if the* $\gcd($*all cycle lengths in* $G_C) = 1$.

Here is a basic result about Markov Chains, which we will give without proof.

Theorem 6.3 *Let C be a finite, irreducible, aperiodic Markov Chain. Then:*
(1) $\exists$ unique stationary distribution σ (i.e. $\sigma^T M = \sigma^T$)
(2) $\sigma_i > 0$, $\forall i \in V$
(3) The expected time to return to state i starting from state i is $1/\sigma_i$.
(4) If $N(i,t)$=number of visits to i in first t steps, then $\frac{N(i,t)}{t} \rightarrow \sigma_i$ almost surely.