College of Computer & Information Science — Spring 2016
Northeastern University — 11 January 2016
CS 7880: Network Algorithms — Scribe: Rajmohan Rajaraman

- Administrivia
- Introduction: random walks
- Brief introduction to random walk analysis using spectral methods
- Bounds on hitting time and cover time

1 Introduction to random walks

Let G be a d-regular directed graph (which is any directed graph in which the in-degree and out-degree of every vertex is d). Let A denote its adjacency matrix, with $A_{ij} = 1$ whenever there is an edge from i to j. A random walk on G is a process that starts at time 0 at an arbitrary vertex and proceeds as follows: if the walk is at vertex i at time t, then it is at vertex j at time $t + 1$ where j is chosen uniformly at random from all the out-neighbors of i. We can capture the random walk by the matrix $M = A/d$, which we call the *random walk matrix*.

A vector x is referred to as a probability vector is $\sum_i x(i) = 1$. The location of the walk at time t can be given by the probability vector $p_t = p_0 M^t$, where p_0 is the unit vector with $p_0(i) = 1$ if the walk starts at i and 0 otherwise. We say that x is a stationary distribution if $xM = x$. Some immediate questions are:

- Do stationary distributions always exist? It is immediate that $u = (1/n, \ldots, 1/n)$ is a stationary distribution of a random walk for every d-regular graph. More generally, finite Markov chain has a stationary distribution.
- Is the stationary distribution unique? In general, any finite irreducible ergodic Markov chain has a unique stationary distribution.
- Does a random walk always converge to a stationary distribution? If it does, how long does it take?

We now study the convergence of a random walk on G to the stationary distribution u. Let $\lambda(G)$ be defined as follows:

$$\lambda(G) = \max_{x \perp u} \frac{\|xM\|}{\|x\|}.$$

Lemma 1. *For any initial probability distribution π, we have*

$$\|\pi M^k - u\| \leq \lambda(G)^k.$$

Proof: Since $\|xM\| \leq \lambda(G)\|x\|$ for any $x \perp u$, it follows that $\|(\pi - u)M\| \leq \lambda(G)\|\pi - u\|$. Since $(\pi - u)M^t \perp u$ for all t, it follows that $\|(\pi - u)M^t\| \leq \lambda(G)^t \|\pi - u\|$. But $uM^t = u$ and $\|\pi - u\| \leq 1$, yielding the desired claim. $\square$

From the above lemma, it is clear that a random walk will converge to the stationary distribution if $\lambda(G) < 1$; smaller the value of $\lambda(G)$, the faster it will converge. In particular, in $\ln(n/\varepsilon)/(1-\lambda(G))$ steps, every entry of πM^k will be at least $(1-\varepsilon)/n$. Is there a good upper bound on $\lambda(G)$ be? We will see that for undirected nonbipartite graphs $\lambda(G) \leq 1 - \Omega(1/n^3)$, which implies that a random walk on nonbipartite graphs converges to the uniform distribution in a polynomial number of steps.

We will later show that $\lambda(G)$ is the same as the second largest eigenvalue of M (in magnitude), which will then allow us to prove the polynomial convergence bound stated above.

2 Bounds on the hitting time and cover time of random walks

In addition to the convergence rate of random walks, measures of interest include *hitting time*, the expected time that it takes for the random walk to visit a particular vertex and *cover time*, the expected time it takes to visit every vertex in the graph. In this section, we establish worst-case bounds on the hitting time and cover time of random walks. Our analysis approach will be combinatorial and different than the algebraic method presented in the previous section.

Let $H(u, v)$ denote the expected time it takes for a random walk starting from u to first hit v. So $H(u, u)$ denotes the expected time it takes for a random walk starting from u to return to u for the first time. Suppose X_i denote the number of steps between the ith visit and the $(i+1)$st visit to u of the walk. Then, $H(u, u)$ is simply the limit, as s tends to infinity of $(X_1 + X_2 + \ldots + X_s)/s$. (Note that both the mean and variance of X_i can be shown to be finite.) But as s tends to infinity, $s/(X_1 + X_2 + \ldots + X_s)$ is simply the fraction of time the walk is at s, which equals the visit probability in the stationary distribution. So we have:

$$\pi(v) = \frac{d(v)}{2m} = \frac{1}{H(v, v)}.$$

We thus know that $H(v, v) = 2m/d(v)$.

Hitting time. Define the *hitting time* of a graph to be the maximum, over all vertices u and v in G, of $H(u, v)$. Consider the time taken for a random walk starting from vertex u to hit a vertex v. To obtain a bound on the expectation of this time, $H(u, v)$, we consider a shortest path from u to v given by $u = v_0, v_1, v_2, \ldots, v_k = v$, for some k. We place an upper bound on $H(v_i, v_{i+1})$ as follows.

$$H(v_i, v_{i+1}) = 1/d(v_i) + (1 - 1/d(v_i))\left(\widetilde{H}(v_i, v_i) + H(v_i, v_{i+1})\right),$$

where $\widetilde{H}(v_i, v_i)$ is the expected time it takes for the walk starting from v_i to return to v_i to the first time under the condition that the first step is not to v_{i+1}. The above equation can be simplified to

$$H(v_i, v_{i+1}) = 1 + (d(v_i) - 1)\widetilde{H}(v_i, v_i).$$

We have a bound for $H(v_i, v_i) = 2m/d(v_i)$. What about $\widetilde{H}(v_i, v_i)$? We have:

$$H(v_i, v_i) \geq (1 - 1/d(v_i))\widetilde{H}(v_i, v_i).$$

We thus obtain

$$H(v_i, v_{i+1}) \leq 1 + d(v_i)H(v_i, v_i)$$

Adding over all i, we obtain

$$H(u,v) \le k + 2mk \le (2m+1)n.$$

We can get a better upper bound for regular graphs, for which we have $H(v_i, v_i) = n$. So plugging this in Equation 1, and adding over all i in the shortest path, we obtain

$$H(u,v) \le k + n \sum_{i=0}^{k-1} d(v_i).$$

The following lemma, left as an exercise, is now handy.

Lemma 2. *Let u and v be two vertices in any undirected graph G. Then, the sum of the degrees of the vertices along any shortest path between u and v in G is at most $3n$.*

This yields us $H(u,v) \le 3n^2 + n$.

Theorem 1. *The hitting time of any n-vertex, m-edge undirected graph is at most $(2m+1)n$. The hitting time of n-vertex regular graph is at most $3n^2 + n$.*

The above bound for hitting time for nonregular graphs is a bit weak. In fact, we can establish a slightly better bound for covering the entire graph.

Cover time. The cover time of a graph is the maximum, over every start vertex u, of the expected time taken for a random walk starting from u to visit every vertex of the graph. Above, we showed using the stationary distribution π that $H(v,v) = 2m/d(v)$. Note that $\pi(v)/d(v) = 1/2m$. Therefore, the probability that the walk is going from vertex u to v in stationary distribution is $1/(2m)$ for any edge (u,v). From this, we can calculate the expected number of steps for the walk to traverse the edge from v to u, if it is presently traversing the edge from u to v. It is simply $H(v,v)d(v) = 2m$.

We now place an upper bound on the cover time by simply adding the above traversal time bound over the Eulerian walk edges of a spanning tree to obtain a cover time bound of at most $2m(n-1)$.

Theorem 2. *The cover time for any undirected graph with n vertices and m edges is at most $2m(n-1)$.*