18.409 An Algorithmist's Toolkit September 22, 2009

Lecture 4

Lecturer: Jonathan Kelner *Scribe: Dan Iancu (2009)*

1 Random walks

Let $G = (V, E)$ be an undirected graph. Consider the random process that starts from some vertex $v \in V(G)$, and repeatedly moves to a neighbor of the current vertex chosen uniformly at random.

For $t \geq 0$, and for $u \in V(G)$, let $p_t(u)$ denote the probability that you are at vertex u at time t. We can think of p_t as a vector in $\mathbb{R}^n$. Clearly,

$$\sum_{u \in V(G)} p_t(u) = 1$$

Observe that you are at a vertex u at time t, then at time $t+1$ you are at each neighbor v of u with probability $1/d(u)$, where $d(u)$ denotes the degree of u. So,

$$\begin{aligned} p_{t+1}(v) &= \sum_{(u,v) \in E(G)} \mathbf{Pr}[\text{at } u \text{ at time } t] \cdot \mathbf{Pr}[\text{'go to } v \text{ at time } t+1\text{' given 'at } u \text{ at time } t\text{'}] \\ &= \sum_{(u,v) \in E(G)} p_t(u) \cdot \frac{1}{d(u)} \end{aligned}$$

We can write this using matrix notation as follows. Define the matrix $W = W_G$:

$$[W_G]_{i,j} = \begin{cases} \frac{1}{d(j)} & \text{if } (i,j) \in E(G) \\ 0 & \text{otherwise} \end{cases}$$

Note that $[W_G]_{i,j}$ is the probability of going from j to i. We have

$$W_G = A \cdot D^{-1},$$

where A is the adjacency matrix of G, and D is the diagonal matrix with $[D]_{i,i}$ the degree of the i-th vertex of G.

2 Stationary distribution

We define a probability vector π which corresponds to the *stationary distribution* of the random walk. Let

$$\pi(u) = \frac{d(u)}{\sum_{v \in V(G)} d(v)}.$$

Claim 1 *π is a probability distribution.*

Proof We have

$$\sum_{u \in V(G)} \pi(i) = \sum_{u \in V(G)} \frac{d(u)}{\sum_{v \in V(G)} d(v)} = \frac{\sum_{u \in V(G)} d(u)}{\sum_{v \in V(G)} d(v)} = 1.$$

■

We next show that, if the random walk follows the distribution π at time t, then it has the same distribution at time $t+1$. This is expressed using matrix notation in the following claim.

Claim 2 $W \cdot \pi = \pi$.

Proof Let $k \in V(G)$. We have

$$[W \cdot \pi]_k = \sum_{i=1}^{n} W_{k,i}\pi_i = \frac{1}{\sum_{v \in V(G)} d(v)} \sum_{(i,k) \in E(G)} \frac{1}{d(k)} \cdot d(k) = \frac{1}{\sum_{v \in V(G)} d(v)} \cdot d(k) = \pi(k).$$

■

This statement is equivalent to the matrix W having eigenvalue 1, with corresponding eigenvector π (note that, since π is a multiple of the vector of node degrees, $D \cdot \mathbf{1}$, we could also take the latter as the eigenvector).

The natural next step at this point would be to claim that the random walk of a graph G always converges to the stationary distribution π. This however turns out to be false. It is easy to see that for a bipartite graph G. Consider for example the case $G = C_6$, the cycle on 6 vertices, and let the vertex set of G be $V(G) = \{1, 2, \ldots, 6\}$. Assume without loss of generality that the random walk starts at time $t_0 = 1$ at vertex 6. Then, at time t, the current vertex is odd if and only if t is odd. Therefore, the walk does not converge to any distribution.

3 Lazy Random Walks

There is an easy way to fix the above periodicity problem. We introduce a modified version of the original walk, which we call *lazy random walk*. In a lazy random walk at time t:

- we take a step of the original random walk with probability 1/2,
- we stay at the current vertex with probability 1/2.

We can show that the above modification breaks the periodicity of the random walk. The transition probabilities are encoded in the following matrix:

$$W' = (W + I)/2 = (I + A \cdot D^{-1})/2,$$

where I denotes the identity matrix.

The fact that W and W' are not symmetric matrices makes their analysis complicated. We will thus define new matrices. The *normalized walk matrix* is defined as

$$N = D^{-1/2} \cdot W \cdot D^{1/2} = D^{-1/2} \cdot A \cdot D^{-1/2}.$$

The *normalized lazy walk matrix* is defined as

$$N' = D^{-1/2} \cdot W' \cdot D^{1/2} = (I + D^{-1/2} \cdot A \cdot D^{-1/2})/2.$$

Claim 3 *The matrices N and W have the same eigenvalues and related eigenvectors.*

Proof Suppose that v is an eigenvector of N, with eigenvalue λ. Let $q = D^{1/2} \cdot v$. Then,

$$N \cdot v = \lambda \cdot v = D^{-1/2} \cdot W \cdot D^{1/2} \cdot v = D^{-1/2} \cdot W \cdot q.$$

Multiplying by $D^{1/2}$ on the left we obtain

$$W \cdot q = \lambda \cdot D^{1/2} \cdot v = \lambda \cdot q.$$

Therefore, q is an eigenvector of W with eigenvalue λ. ■

Observe that, by Claim 2, W has eigenvector $D \cdot \mathbf{1}$, with eigenvalue 1. Therefore, by Claim 3, the normalized walk matrix N has eigenvector $D^{1/2} \cdot \mathbf{1}$, with eigenvalue 1.

4 Connections to Laplacians

We've used the Laplacian L. The normalized Laplacian $\mathcal{L}$ is defined as

$$\mathcal{L} = D^{-1/2} \cdot L \cdot D^{-1/2}.$$

Claim 4 *$N = I - \mathcal{L}$.*

Therefore, the eigenvalues of N are given by $1-$(eigenvalues of $\mathcal{L}$). So, it makes sense to order them in the opposite way

$$1 = \mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$$

We can now translate our theorems about the eigenvalues of Laplacians to theorems about μ_is. We have

- For each i, $\mu_i \in [-1, 1]$.
- If G is connected, then $\mu_2 < 1$.
- The -1 eigenvalues occur only for bipartite graphs.

Let μ_i' be the eigenvalues of N'. Then

- For each i, $\mu_i' \in [0, 1]$.
- If G is connected, then $\mu_2' < 1$.

5 ℓ_2 Convergence

Define the *spectral gap* to be

$$\lambda := 1 - \mu_2'.$$

For probability distributions p, q, we define their ℓ_2 distance to be

$$\|p - q\|_2 = \sqrt{\sum_i (p(i) - q(i))^2}.$$

The following theorem gives a bound on the rate of convergence of the lazy random walk to the stationary distribution π.

Theorem 5 *Let p_0 be an arbitrary initial distribution, and p_t be the distribution after t steps of the lazy random walk. Then,*

$$\|p_t - \pi\|_2 \leq (1-\lambda)^t \cdot \sqrt{\frac{\max_x d(x)}{\min_y d(y)}}.$$

Proof [Proof for regular graphs] Observe that for a matrix $M = Q^{-1} \cdot \Lambda \cdot Q$, we have $M^k = Q^{-1} \cdot \Lambda^k \cdot Q$. Thus, for an eigenvector v of M, $M^k \cdot v = \lambda^k \cdot v$.

Recall that $N' = (I + D^{-1/2} \cdot A \cdot D^{-1/2})/2$. Since G is regular, $D = d \cdot I$, for some integer $d > 0$. Thus,

$$N' = I + \frac{1}{d} A$$

and the stationary distribution is simply the uniform distribution on $V(G)$

$$\pi = \frac{1}{n} \cdot \mathbf{1}.$$

Let $c_i = v_i^T p_0$, where v_i denotes the eigenvector corresponding to the i-th eigenvalue. We have

$$N'^k \cdot p_0 = \sum_{i=1}^{n} c_i \cdot \mu_i^k \cdot v_i = c_1 \cdot v_1 + \sum_{i=2}^{n} c_i \cdot \mu_i^k \cdot v_i$$

Since $c_1 = v_1^T p_0 = 1/n$, it follows that

$$\begin{aligned} \|p_k - \pi\|_2 &= \|\sum_{i=2}^{n} c_i \cdot \mu_i^k \cdot v_i\|_2 = \sqrt{\sum_{i=2}^{n} c_i^2 \cdot \mu_i^{2k}} \le \mu_2^k \sqrt{\sum_{i=2}^{n} c_i^2} \\ &\le \mu_2^k \sum_{i=1}^{n} (v_i^T p_0)^2 \le \mu_2^k = (1-\lambda)^k. \end{aligned}$$

■

Using a similar argument, we can also show an analogous bound for ℓ_∞ convergence.

Theorem 6 *For any vertex $v \in V(G)$,*

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

6 Conductance

Cheeger's inequality carries over too, by replacing the isoperimetric number by a new parameter, which we call *conductance* Φ.

Definition 7 (Conductance) *For $S \subseteq V(G)$, let*

$$\Phi(S) = \frac{e(S)}{\min\left(\sum_{v \in S} d(v), \sum_{v \in \bar{S}} d(v)\right)}.$$

We define the conductance to be

$$\Phi(G) = \min_{S \subseteq V} \Phi(S).$$

Using the above definition, Cheeger's inequality now becomes:

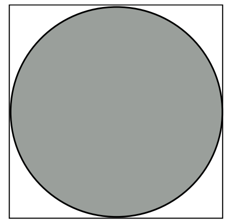
$$\Theta(1) \cdot \Phi^2(G) \le 1 - \mu_2' \le \Theta(1) \cdot \Phi(G).$$

The parameter $\Phi(G)$ is related to the rate of convergence to the stationary distribution. In particular, bounds on $\Phi(G)$ let us prove that a walk mixes quickly.

The intuitive interpretation of the connection between conductance and the rate of convergence is as follows. If a graph has high conductance, it is well-connected. Therefore, a large amount of probability mass can very quickly move from one part of the graph to another.

7 Introduction to Monte Carlo methods

Assume that we want to estimate $\pi = 3.1415\ldots$ by throwing darts in the following dartboard:



Assume that the square corresponds to $[-1,1] \times [-1,1]$. If you pick a point in the square uniformly at random, the probability that you pick one inside the circle is equal to $\pi/4$. Suppose that you pick n points in $[-1,1] \times [-1,1]$, uniformly at random. Then,

$$\mathbf{E}[\text{number of points inside circle}] = n \cdot \pi/4$$

So, you can return the estimate

$$\hat{\pi} = (\text{number of points inside circle}) \cdot 4/n.$$

A natural question is how close this estimate would be to the right answer.

In order to answer the above question, we will introduce the Chernoff bound. Suppose we have a random variable $r \in \{0,1\}$, such that $\mathbf{Pr}[r=1] = p$, and $\mathbf{Pr}[r=0] = 1-p$. Assume that we draw n independent samples $r_1, \ldots, r_n$, and let $R = \sum_i r_i$. By the linearity of expectation, we have

$$\mathbf{E}[R] = \mathbf{E}[\sum_i r_i] = \sum_i \mathbf{E}[r_i] = n \cdot p$$

We will say that R *ϵ-approximates* $\mathbf{E}[R]$ if

$$(1-\epsilon)\mathbf{E}[R] \leq R \leq (1+\epsilon)\mathbf{E}[R]$$

This is a multiplicative error measure.

Theorem 8 (One version of the Chernoff bound) *The probability that R fails to ϵ-approximate $\mathbf{E}[R]$ is*

$$\mathbf{Pr}\left[|R - \mathbf{E}[R]| \geq \epsilon \mathbf{E}[R]\right] \leq 2e^{-np\epsilon^2/12} = 2e^{-\mathbf{E}[R]\epsilon^2/12}.$$

Some notes on the above bound:

- The bound is near tight.
- It is necessary for the trials to be independent, in order for the bound to hold.
- It provides a multiplicative, but not an additive error guarantee.
- For fixed ϵ, it falls off exponentially in n. So, if we have failure probability $1/2$, we can improve it to $1/2^k$ by performing $m = n \cdot k$ trails.
- Therefore, smaller n requires more trials.
- If we want ϵ-approximation with probability $1-\delta$, then we need

$$N \geq \Theta\left(\frac{\log(1/\delta)}{p\epsilon^2}\right).$$

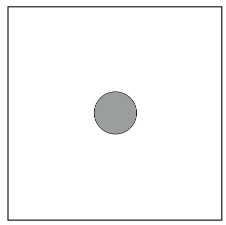
That is, we need enough trials to get $\Theta(\log(1/\delta)/\epsilon^2)$ successes.

Back to the dartboard example, if we want to estimate π within, say, 5%, with probability at least 0.99, then we have $\epsilon = 0.05$, $\delta = 1/100$. Therefore, we need

$$N \geq \Theta\left(\frac{\log(100)}{(\pi/4)(0.05)^2}\right)$$

Observe that it is easy to make δ smaller, but it is harder to make ϵ smaller.

If we are bad darts, then we run into trouble. This happens if we have a big dartboard, and a small circle.



In particular, if p is exponentially small, then we need exponentially many trials to expect a constant number of successes.

We can also run into trouble if it is hard to throw darts at all. That is, if it is hard to draw samples uniformly at random from the ambient space. We will develop some techniques for fixing the above problems in certain scenarios.

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