Lecture 8: Random Walks in Undirected Graphs

Aditya Bhaskara[*]

CS 5968/6968: Techniques in Algorithms and Approximation

February 4th, 2016

Abstract

We introduce the notion of a random walk in an undirected graph. What happens after we take T random steps in the walk? We show how to analyze walks and their convergence properties.

1 Walks in graphs

Suppose we have an undirected graph G. For the next couple of lectures, we will consider graphs that aren't necessarily regular. A random walk in G is the following random procedure:

A particle starts at some vertex $u^{(0)}$ in the graph (this could either be fixed, or itself sampled from some given *start distribution* $p^{(0)}$). At time t, the particle moves from $u^{(t-1)}$ to a uniformly-random neighbor of $u^{(t-1)}$.

A simple example is a walk on an $n \times n$ grid (chessboard). Say the particle starts at the bottom left corner, and at every step, moves in one of the four possible directions with equal probability. This is sometimes referred to as *Brownian motion*, used in physics to model the movement of a particle in a fluid.

We are interested in how the walk *evolves*. In particular, we are interested in the likelihood that the particle is at some vertex v at time t. How can we compute such quantities? The key is to do it recursively: for the particle to be at v at time t, it must have been at one of the neighbors of v at time $(t-1)$, and must have chosen the edge to v. Thus if we denote the probability of interest by $p^{(t)}(v)$, we have

$$p^{(t)}(v) = \sum_{u \in \mathrm{nbr}(v)} p^{(t-1)}(u) \cdot \frac{1}{\deg(u)}. \tag{1}$$

This gives a clean way to compute $p^{(t)}(v)$ for all v, by computing $p^{(t-1)}(v)$ for all v, which in turn can be done by computing $p^{(t-2)}(\cdot)$, and so on.

Recall that we know the start distribution $p^{(0)}(v)$

2 Transition matrix

Let define $p^{(t)}$ as a vector in $\mathbb{R}^n$ whose ith entry is $p^{(t)}(i)$. We can use Eq. (1) to write down $p^{(t)}$ in terms of $p^{(t-1)}$. The *transition matrix* M will be defined so that we have

$$p^{(t)} = M p^{(t-1)}.$$

[*]bhaskara@cs.utah.edu

(1) tells us what the entries of M should be. Considering the vth row of the output, we get

$$p^{(t)}(v) = \sum_u M_{uv} p^{(t-1)}(u).$$

Comparing with (1), we must set $M_{uv} = 0$ if there is no edge uv, and $M_{uv} = \frac{1}{\deg(u)}$ when uv is an edge. If all the degrees are equal to d, this is precisely the adjacency matrix scaled by a factor d. Otherwise, a bit of thought shows that M is the adjacency matrix with its ith column scaled by $1/\deg(i)$.

Let us write A to denote the adjacency matrix, and D to denote the diagonal matrix of degrees Then the discussion above implies that

I.e., $D(i,i) = deg(i)$ and $D(i,j) = 0$ for $i \neq j$.

$$M = AD^{-1}.$$

We can thus write $p^{(t)}$ in a simple way:

$$p^{(t)} = Mp^{(t-1)} = M^2 p^{(t-2)} = \cdots = M^t p^{(0)}.$$

What if we repeat this process forever? Does the distribution $p^{(t)}$ coverge to some vector? Does it cycle between a small number of vectors? If it does converge, at what rate?

We now develop machinery to answer questions of this kind. The key observation is that the vector $p^{(t)}$ evolves exactly like the *power method* we saw in the last lecture. This suggests the connection to eigenvalues of M. Further, if the process does converge, it must be to some vector p that satisifes $p = Mp$, or equivalently, p must be an eigenvector of M, with eigenvalue 1.

3 Eigenvalues of M

The first thing to notice about M is that it is not symmetric. Thus it is not immediately clear that it even has real eigenvalues. Luckily, our matrix has the following special form:

Note that $B(i,j) = 1/\sqrt{deg(i)deg(j)} = B(j,i)$. Also, $D^{1/2}$ and $D^{-1/2}$ are the diagonal matrices whose (i,i)th entries are $\sqrt{deg(i)}$ and $1/\sqrt{deg(i)}$ respectively.

$$M = PBP^{-1}, \qquad \text{where } B = D^{-1/2}AD^{-1/2} \text{ and } P = D^{1/2}.$$

The matrix B is simply the normalized adjacency matrix we encountered earlier, which is a symmetric matrix. The equation above implies that M and B are *similar*, i.e., they have the same eigenvalues. In fact, if v is an eigenvector of B satisfying $Bv = \lambda v$, then Pv is an eigenvector M,

The notion of similar matrices is a standard one in linear algebra.

$$M(Pv) = PBP^{-1}Pv = \lambda Pv.$$

Now, the eigenvalues of B are something we understand! As we saw earlier, they all lie between -1 and 1, and the multiplicity of 1 is precisely the number of connected components in G. We also saw that the smallest eigenvalue is -1 if and only if G is bipartite.

Suppose G is connected. This means that B has a unique eigenvector $\mathbf{u}$ of eigenvalue 1 (up to scaling), and we also know its entries – the ith entry is $\sqrt{\deg(i)}$. Thus the corresponding eigenvector of M is $D^{1/2}\mathbf{u}$, whose ith entry is $\deg(i)$. Let us scale it so that the entries sum to 1, and denote this vector by

π. The ith entry of π is thus

$$\pi(i) = \frac{\deg(i)}{\sum_j \deg(j)} = \frac{\deg(i)}{2m},$$

where m is the number of edges in G.

4 Stationary distribution

The stationary distribution for a random walk is a distribution that remains unchanged upon taking a step of the walk. I.e., a vector $\mathbf{u}$ with nonnegative entries that sum up to 1, and which satisfies $M\mathbf{u} = \mathbf{u}$. Observe that this is exactly what we saw with π above. Thus π is one stationary distribution for the random walk M.

Is it the only stationary distribution? If G is connected, the uniqueness of the top eigenvector implies that this is indeed the case. Furthermore, the key property of the stationary distribution is that for *any* start distribution $p^{(0)}$, taking sufficiently many steps of the random walk converges to the stationary distribution, as long as the graph is not bipartite.

To see this, suppose we start with $p^{(0)}$ and repeatedly apply M. Thus we have $p^{(t)} = M^t p^{(0)}$. From the analysis of the power iteration we saw last class, we have that as long as $\langle p^{(0)}, \pi \rangle > 0$ and the gap between the largest and the second largest (in magnitude) eigenvalues is > 0, the power method converges. The first condition is true because for any distribution $p^{(0)}$, we have $\langle p^{(0)}, \pi \rangle \geq \min_i \pi(i) > 0$ (because degrees are all > 0). The second condition follows because in any graph that is not bipartite, we have a non-zero gap $1 - \max_{i \neq n} |\lambda_i(M)|$.

5 Convergence to stationarity

We can also quantify the convergence analysis above.

5.1 definition. The time T for which $\|p^{(T)} - \pi\|_1 < 1/4$ is called the *mixing time* for the random walk M.

Suppose G is connected. Then we know that $1 - |\lambda_{n-1}| \geq 1/n^3$, as a consequence of one of the problems on HW2. Further, if we have a graph that is not bipartite, we can show that $1 - |\lambda_1| \geq 1/n^3$.

The exercise was for d-regular graphs, and had a bound $1/n^2$ for the Laplacian. When we consider the normalized Laplacian, we get a bound $1/n^3$, since the max degree is $\leq n$.

Finally, since $\min_i \pi(i) \geq 1/m > 1/n^2$, we have that for $t = O(\log n/n^3)$, we will have $\|\pi - p^{(t)}\|_2 < 1/n$. A mild technicality here is that in the last class, we measured closeness in the ℓ_2 norm, while in the definition of the mixing time, we care about ℓ_1 norm. To go from one bound to the other, we use the inequality that for any n dimensional vector x,

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2.$$

Thus the above reasoning shows that the mixing time in any connected, n-vertex non-bipartite graph is $O(n^3 \log n)$.

In the next class, we will see how to get rid of the non-bipartite assumption, by looking at a simple modification of the random walk.