Notes for Recitation 25

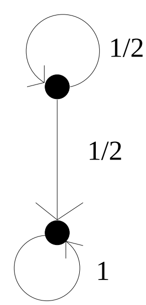
1 Random walks on graphs

In lecture yesterday, we saw what happens when you take a random walk on the roulette wheel. In today's recitation, we will study a more general paradigm that allows us to model the typical movement pattern of a 6.042 student right after the final exam.

Let directed graph G have vertices V and edges E. The 6.042 student comes out of the final exam located on a particular node of the graph, corresponding to the exam room. What happens next is unpredictable, as the student is in a total haze. At each step of the walk, if the 6.042 student is at node u at the end of the previous step, he/she picks one of the edges (u, v) uniformly at random from the set of all edges directed out of u, and walks to the node v.

If $|V| = n$, let the vector $P^{(j)} = (p_1^{(j)}, \ldots, p_n^{(j)})$ be such that $p_i^{(j)}$ is the probability of being at node i after j steps.

a. We will start by looking at a simple graph. If the student starts at node 1 (the top node) in the following graph, what is $P^{(0)}, P^{(1)}, P^{(2)}$? Give a nice expression for $P^{(n)}$.



Solution. $P^{(0)} = (1, 0), P^{(1)} = (1/2, 1/2), P^{(2)} = (1/4, 3/4),\ P^{(n)} = (1/2^n, 1 - 1/2^n)$. ■

b. Given an arbitrary graph, show how to write an expression for $p_i^{(j)}$ in terms of the $p_k^{(j-1)}$'s.

Solution. We have

$$p_i^{(j)} = \sum_{k \ | \ (k,i) \in E} \frac{1}{\text{degree}(k)} p_k^{(j-1)}.$$

■

c. Does your answer to the last part look like any other system of equations you've seen in this course?

Solution. It should – these are similar to the equations that we got from PageRank! ■

d. Let the *limiting distribution* vector π be

$$\lim_{k\to\infty} \frac{\sum_{i=1}^{k} P^{(i)}}{k}.$$

What is the limiting distribution of the graph from part a? Would it change if the start distribution were $P^{(0)} = (1/2, 1/2)$ or $P^{(0)} = (1/3, 2/3)$?

Solution. Say we start with the distribution $(x, 1-x)$. The distribution after i steps is

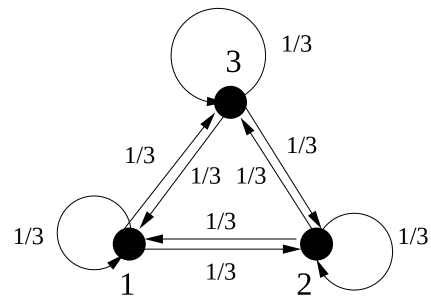
$$P^{(i)} = (x/2^i, 1 - x/2^i)$$

Plugging this into the formula for the limiting distribution, we have

$$\begin{aligned}
\pi &= \lim_{k\to\infty} \frac{\sum_{i=1}^{k} P^{(i)}}{k} \\
&= \lim_{k\to\infty} \frac{\sum_{i=1}^{k} (x/2^i, 1 - x/2^i)}{k} \\
&= \lim_{k\to\infty} \left(\frac{\sum_{i=1}^{k} x/2^i}{k}, \frac{\sum_{i=1}^{k} 1 - x/2^i}{k} \right) \\
&= \lim_{k\to\infty} \left(\frac{x - x/2^k}{k}, \frac{k - (x - x/2^k))}{k} \right) \\
&= (0, 1)
\end{aligned}$$

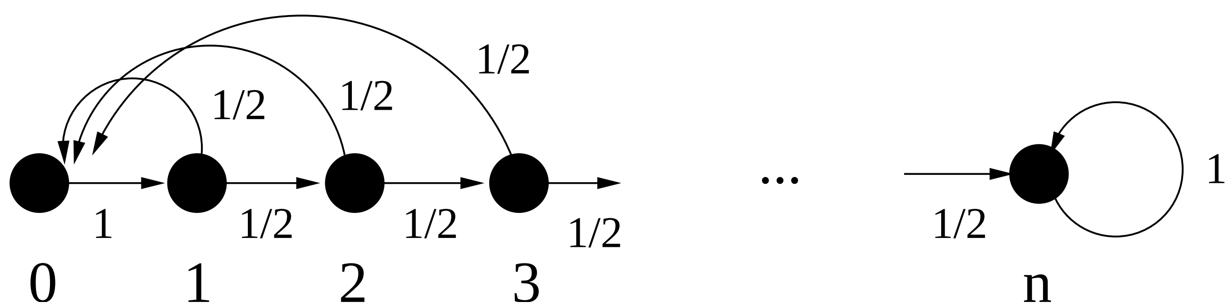
■

e. Let's consider another directed graph. If the student starts at node 1 with probability 1/2 and node 2 with probability 1/2, what is $P^{(0)}, P^{(1)}, P^{(2)}$ in the following graph? What is the limiting distribution?



Solution. $P^{(0)} = (1/2, 1/2, 0)$, $P^{(1)} = (1/3, 1/3, 1/3)$, $P^{(2)} = (1/3, 1/3, 1/3)$, and in general, the limiting distribution is $(1/3, 1/3, 1/3)$. ■

f. Now we are ready for the real problem. In order to make it home, the poor 6.042 student is faced with n doors along a long hall way. Unbeknownst to him, the door that goes outside to paradise (that is, freedom from 6.042 and more importantly, vacation!) is at the *very end*. At each step along the way, he passes by a door which he opens up and goes through with probability 1/2. Every time he does this, he gets teleported back to the 6.042 exam room. Let's figure out how long it will take the poor guy to escape from 6.042. What is $P^{(0)}, P^{(1)}, P^{(2)}$? What is the limiting distribution?



Solution. $P^{(0)}$ just has $p_0^{(0)} = 1$ and all other probabilities 0, since the student is at the exam room and has just started. $P^{(1)}$ just has $p_1^{(1)} = 1$ since the student made the only move possible, and all other probabilities 0. $P^{(2)}$ has $p_0^{(2)} = 1/2$, since the student opens the first door with probability 1/2, and $p_2^{(2)} = 1/2$, since the student passes the first door with probability 1/2.

Eventually, we reach node n, at which point we stay there forever. Thus, the limiting distribution has $p_i = 0$ for $i = 0, 1, \ldots, n-1$, and $p_n = 1$. ■

g. Show that the expected number of teleportations $T(n)$ you make back to the exam room before you escape to the outside world is $2^{n-1} - 1$.

Solution. The probability that you manage to reach the last door to the outside without getting teleported back to the 6.042 exam room is $\frac{1}{2^{n-1}}$. Thus, by the mean time to failure formula, $T(n)$, the expected number of times you get teleported back to the 6.042 exam room is $2^{n-1} - 1$ (on the last try you don't get teleported back, but rather, you succeed.) ■

2 More random walks

Consider an undirected connected graph $G = (V, E)$. It turns out that such graphs have a unique limiting distribution, independent of the initial distribution. For node i, let $\deg(i)$ be its degree. Let $m = \sum_i \deg(i) = 2|E|$, and $n = |V|$. Consider the vector of probabilities

$$\pi^* = \left(\frac{\deg(1)}{m}, \frac{\deg(2)}{m}, \ldots, \frac{\deg(n)}{m} \right).$$

We will show that π^* is a limiting distribution of G.

Note that in general, such a clean description of a limiting distribution does not exist for directed graphs, such as for the web graph that PageRank uses. Intuitively this makes sense, as otherwise one could create a lot of dummy links that point to your web site to increase its degree, and therefore artificially increase its rank.

a. In order to show that π^* is a limiting distribution, we need to pick a starting distribution. Let's choose as our starting distribution $P^{(0)} = \pi^*$. Using the formula from part 1b, prove by induction that $P^{(n)} = P^{(0)}$ for all n.

Solution. Let $P(n)$ be the proposition that $P^{(n)} = P^{(0)}$.

Base case: Clearly, $P^{(0)} = P^{(0)}$.

Inductive step: We know from problem 1b that $p_i^{(n+1)} = \sum_{k \mid (k,i) \in E} \frac{1}{\text{degree}(k)} p_k^{(n)}$. Using this fact, we have

$$\begin{aligned}
p_i^{(n+1)} &= \sum_{k \mid \{i,k\} \in E} \frac{1}{\deg(k)} \cdot p_k^{(n)} \\
&= \sum_{k \mid \{i,k\} \in E} \frac{1}{\deg(k)} \cdot p_k^{(0)} && \text{(Inductive Hypothesis)} \\
&= \sum_{k \mid \{i,k\} \in E} \frac{1}{\deg(k)} \cdot \frac{\deg(k)}{m} && \text{(π^* is the starting distribution)} \\
&= \sum_{k \mid \{i,k\} \in E} \frac{1}{m} && \text{(simplifications)} \\
&= \frac{\deg(i)}{m} \text{(definition of degree)} \\
&= p_i^{(0)} && \text{(π^* is the starting distribution),}
\end{aligned}$$

as desired. ∎

b. Use the definition of a limiting distribution along with the results of part a to show that π^* is a limiting distribution of G.

Solution. Using the definition of a limiting distribution, we have

$$\begin{aligned}
\pi &= \lim_{k\to\infty} \frac{\sum_{j=1}^{k} P^{(j)}}{k} \\
&= \lim_{k\to\infty} \frac{\sum_{j=1}^{k} P^{(0)}}{k} && \text{(from part 2a)} \\
&= \lim_{k\to\infty} \frac{kP^{(0)}}{k} && \text{(simplification)} \\
&= P^{(0)} && \text{(evaluation of the limit)} \\
&= \pi^* && \text{(π^* is the starting distribution)}
\end{aligned}$$

■