15-251: Great Theoretical Ideas In Computer Science

Recitation 10: Markov Chains and Random Walks

Definitions and Theorems

Review on Markov chains:

- A **Markov chain** is a directed graph where each directed edge is labeled with a non-zero probability, and the sum of the probabilities over outgoing edges from any vertex is 1. We call the vertices in a Markov chain **states** and the edges **transitions**.
- A Markov chain with n states can be described by an n by n **transition matrix** K, where each entry K$[i, j]$ corresponds to the probability for transition (i, j). The rows of K sum to 1.
- Given a Markov chain and a row vector π_0 for the initial probability distribution over the states,

  $$\pi_t = \pi_0 K^t$$

  where π_t represents the probability distribution after t steps. If the chain is strongly connected, then there exists a unique distribution π, called the **invariant distribution**, for which $\pi = \pi K$.
- We define random variable T_{ij} as the number of steps it takes to reach state j from state i.
- Suppose a given Markov chain is strongly connected. Then, for all i,

  $$\mathbf{E}[T_{ii}] = \frac{1}{\pi[i]}.$$

A new definition:

- We can turn any connected undirected graph into a Markov chain as follows. First, let the vertices represent the states. Then, from any vertex i, choose a transition by selecting a neighbor of i randomly. Traversing a graph in this way is known as a **random walk**.

A Cake-Walk

(a) Describe the transition matrix K for a random walk on a connected undirected graph.

(b) Show that

$$\pi = \left[\frac{d_1}{2m}, \frac{d_2}{2m}, \ldots, \frac{d_n}{2m}\right]$$

is the invariant distribution for the random walk, where d_i is the degree of vertex i.

(c) What is $\mathbf{E}[T_{ii}]$, the expected number of steps it takes to return to vertex i starting from i?

So Close, Yet So Far

Let G = (V, E) be a connected undirected graph.

(a) Let $(i, j) \in$ E be an edge in G. Show that for a random walk on G, $\mathbf{E}[T_{ij}] \leq 2m$.
(Hint: Use the law of total expectation on $\mathbf{E}[T_{jj}]$.)

(b) Use the above result to show that for any $s, t \in$ V, $\mathbf{E}[T_{st}] \leq n^3$.
(Hint: Use linearity of expectation.)

RFS

The undirected connectivity decision problem takes as input an undirected graph G = (V, E) and two vertices $i, j \in$ V. The output is "yes" if there is a path from i to j, and "no" otherwise.

You are trying to solve an instance of the undirected connectivity problem. Unfortunately, there is a worldwide shortage of memory, so you can only afford $O(\log n)$ bits of space. This means that your trusty BFS and DFS algorithms won't work, since you can't even keep track of your visited nodes! Show that there is nevertheless a polynomial-time randomized Monte Carlo algorithm which will solve the problem with probability of error less than $\frac{1}{1000}$.