CS 683: Advanced Design & Analysis of Algorithms 3/10/2008

Lecture 22

Lecturer: John Hopcroft *Scribes: Maxime Brugidou (mb727) and Pierre-Alexandre Meyer (pm326)*

1 PageRank

Given an aperiodic strongly connected component, a random walk on it converges to a stationary problem. We define

$$\text{PageRank} = \frac{1}{\text{ReturnTime}}$$

The World Wide Web can be considered as strongly connected by jumping once in a while to a node chosen uniformly at random.

1.1 Examples

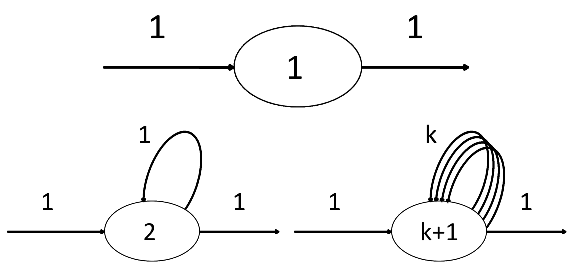

Figure 1: PageRank without the jump

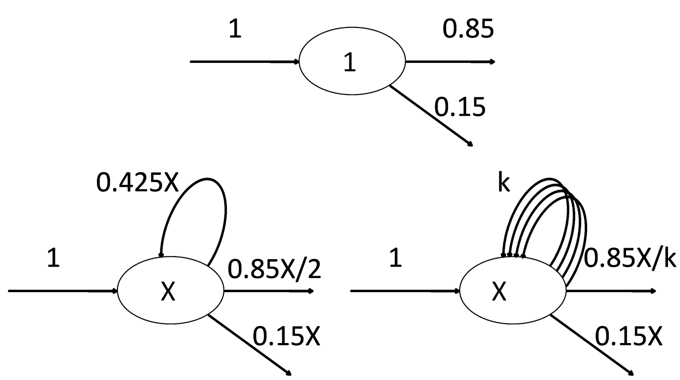

Figure 2: PageRank with the jump

In the end the maximum X for $k \to \infty$ if $\frac{1}{0.15} = 6.667$.

1.2 PageRank in detail

As we can see on the graphic

$$\text{return time} = \text{hitting time} + \text{expected time to restart}$$

PageRank is defined as a tradeoff between hitting time (which capture the restart) and the short cycles that capture random walks

$$(1-\alpha)A^t P + \alpha \frac{1}{n}(1) = P$$

where α is the restart factor ($\alpha = 0.15$ in general).

The restart can be done by choosing only trusted nodes (trusted websites are chosen with higher probability).

2 Hidden Markov Models (HMM)

Given a statistical model (directed graph) composed of states, output symbols and transitions, we will try to solve the following questions:

- How probable is a given output sequence?
- What is most likely the sequence of states?
- Given an upper bound on the number of states and an output sequence, what is the most likely HMM?

2.1 Example

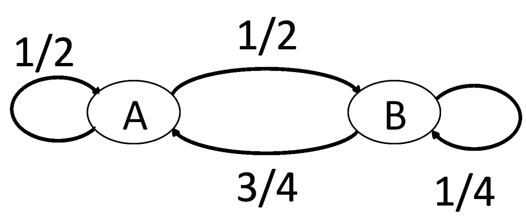

Figure 3: Two states A and B

Let's consider a model with two states A and B, two symbols h and t and the following probabilities:

- $p(A \to B) = \frac{1}{2}$
- $p(B \to A) = \frac{3}{4}$
- $p(A \to A) = \frac{1}{2}$
- $p(B \to B) = \frac{1}{4}$
- $p(h|A) = \frac{1}{2}$
- $p(t|A) = \frac{1}{2}$
- $p(h|B) = \frac{2}{3}$
- $p(t|B) = \frac{1}{3}$

Initially, the probability of being in state A is 1.

What is the probability of sequence `hhht`?

$t = 2$	$\frac{1}{8}\frac{1}{2}\frac{1}{2} + \frac{1}{6}\frac{3}{4}\frac{1}{2}$	$\frac{1}{8}\frac{1}{2}\frac{2}{3} + \frac{1}{6}\frac{1}{4}\frac{2}{3}$	0	0
$t = 1$	$\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$	$\frac{1}{2}\frac{1}{2}\frac{2}{3} = \frac{1}{6}$	0	0
$t = 0$	$\frac{1}{2}$	0	0	0
	$A \to A$	$A \to B$	$B \to A$	$B \to B$

In general, given n states and a sequence of T symbols, the complexity to fill in the previous table is $O(n^3T)$ in time and $O(n^2T)$ in space.