Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

FS 2019 Prof. R. Wattenhofer

Computer Engineering II

Exercise Sheet Chapter 11

We categorize questions into four different categories:

Quiz Short questions which we will solve rather interactively at the start of the exercise sessions.

Basic Improve the basic understanding of the lecture material.

Advanced Test your ability to work with the lecture content. This is the typical style of questions which appear in the exam.

Questions marked with $^{(g)}$ may need some research on Google.

Quiz

1 Quiz Questions

a) Can a sequence of random variables, where each random variable depends on the 2 previous random variables, be described by a Markov chain?

b) Is the following true: If a finite Markov chain is irreducible and there is a state which has period 1, then the Markov chain is ergodic?

c) If, in a Markov chain, there is a state with period 2 and a state with period 3, then the Markov chain is always which of the following? Irreducible, not irreducible, ergodic, aperiodic.

d) Which of the 6 possibilities to create or remove a link between the websites u and v or from v to w can improve the PageRank of u?

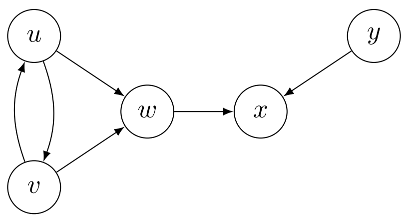


Figure 1: An example of a web graph with 5 websites. Website x does not link to any other website, i.e., x is a *sink*.

e) Is the following true: By performing a simple random walk for a long enough time, the probability of being in a specific state gets arbitrarily close to the probability for that state given by the unique stationary distribution, independent of the chosen starting node?

2 Soccer Betting

The *FC Basel* soccer club is a particularly moody team. Upon winning a game, they tend to win subsequent games. After losing a game, however, they often end up losing the next game as well. A group of international scientists, consisting of soccer experts, mathematicians, and psychologists, has recently conducted a thorough analysis of this behavior. In particular, they have discovered that upon winning a game, the FCB wins the next game with a probability of 0.6 as well. With probabilities 0.2 each, the next game will be a tie or a loss. After a loss, the FCB will win/tie/lose its next game with probability 0.1/0.2/0.7, respectively. Finally, after a tie, the next game being a win or a loss is equally probable. The probability that the next game also ends up being a tie is 0.4.

a) Model the FCB's moodiness using a Markov chain.

b) In *two games* from now (they will play one game in between), the FCB will play against the FC Zurich. The Swiss TOTO offers you the following odds:

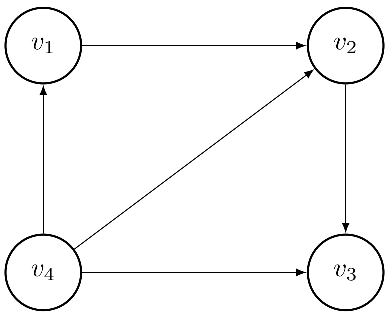
Win: 3.5 Tie: 4.0 Loss: 1.5

Given that the FCB won the game three games ago, but lost the last two games, what would be your bet? Why?

c) More recent studies have shown that the FCB is even moodier than expected. In fact, after losing *two games in a row*, the probability of winning its next game reduces to 0.05, that of getting a tie to 0.1. Change your Markov chain model to incorporate the new findings. How does the change influence your bet?

3 PageRank

Usually, one would compute the PageRank of a network not by hand, but by running an algorithm on a computer – since the relevant instances are far too tedious to compute by hand. Therefore we only consider a very small example with four nodes for this exercise:



a) The naïve suggestion is to simply use the nodes' (in-)degree as a rank, by computing $(1, 1, 1, 1) \cdot A$, where A is the adjacency matrix of the graph. What are the naïve PageRanks, and why might that not be a good idea?

b) An improved-naïve version would normalize each non-zero row in the adjacency matrix by dividing each row by the sum of its entries. Denote the obtained matrix by A'. What are the improved-naïve PageRanks $(1, 1, 1, 1) \cdot A'$? Is the problem from a) solved now?

c) Since websites with higher rank should have a larger impact on the rank of the nodes they link to, you decide to iterate the process from b). Denoting $(1, 1, 1, 1)$ with q_0, calculate the iterated version of this "PageRank" by computing

$$\begin{aligned} q_1 &= q_0 \cdot A' \\ q_2 &= q_1 \cdot A' \\ q_3 &= q_2 \cdot A' \\ &\vdots \end{aligned}$$

(You should only need to iterate very few times.) When does the rank converge, and what is the issue? Would the same happen when using the Google matrix?

4 Queues

Router R, working for its favorite network, is very reliable: In each time step, it forwards exactly one packet from its queue. It likes the network so much because in each time step there arrive 0, 1 or 2 packets and each possibility occurs with the same probability. If, at time t, R's queue is empty, but at least one packet arrives in the time step between t and $t + 1$, it forwards one of the arriving packets immediately. The queue can accommodate up to 10 packets. If, at time t, the queue is full, R forwards one of the contained packets as usual, but it also fills the queue to its limit again, if a new packet should arrive between t and $t + 1$.

a) Model the load of the queue by using a Markov chain. What is the stationary distribution?

Hint: Try to rewrite the definition of stationary distribution (Definition 12.10 in the script) as a system of linear equations. Make use of the fact that most entries are 0.

b) A colleague wants to play a joke on R and tells it that, due to a change in the network, the probability that 2 packets arrive is now k times as high as the probability that 0 packets arrive. Can R now calculate the new stationary distribution (for any $k > 0$) or does it need to know the exact probabilities? What can R conclude about the relation between the probabilities (in the stationary distribution) of two subsequent nodes in the Markov chain (i.e. about the relation between the probabilities of two loads of the queue which differ by 1)?

c) Due to a magic upgrade, R's queue now can accommodate all packets that ever arrive. What is the stationary distribution now (for the initial probabilities for the number of arriving packets)? Can you see a problem with that? What if the probabilities for the arrival of 0, 1 and 2 packets were 2/5, 2/5 and 1/5, respectively?

d) R has become old. The magic has disappeared and its maximum queue size is back to 10 packets. Moreover, it is annoyed by colleagues sending packets when they know that its queue is full. Thus, if its queue is full at time t and 2 packets arrive between t and $t + 1$ it simply flushes its queue. How does the stationary distribution change?

e) R has become very old. It considers retiring to some island, but since it still likes the work it does, it lets chance decide about the time of its retirement. Some of its friends are still young and it has a lot of good advice for its younger colleagues, so, in order to not retire too soon, it chooses to throw a very biased coin each time its queue is full and 2 packets arrive: If the coin shows head (which happens with probability 0.000001) it retires immediately, otherwise it continues its work as it did when it was young (no flushing this time). How about the stationary distribution now?

5 Chess Tours

Assume we have an empty chess board.

a) If a Knight (Horse/Springer/Pferd) is starting from one of the 4 middle squares of the board and moves by picking the next square uniformly at random from its set of legal moves for that square, how many moves on average will it take for it to return back to the same square?

$Hint_1$: The "degree sum theorem" says that the sum of degrees of the nodes of an undirected graph is equal to twice the number of edges in the graph.
$Hint_2$: If a Knight's tour is represented as a undirected graph, the 4 corner nodes have degrees of 2, and so on.

b) If a Knight starts at one of the corners, does the answer change? If so, to what?

c) If a Rook (Elephant/Castle/Turm) is placed on the bottom left corner a 2X2 chess board, and moves by picking the next square uniformly at random from its set of legal moves for that square, how many moves on average will it take for it to reach the top right corner of the board?