Rolf Wanka
Matthias Kergaßner

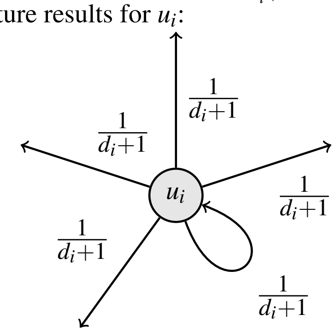
Erlangen, 24 May 2022

Exercises to
Swarm Intelligence
Summer 2022
Sheet 4

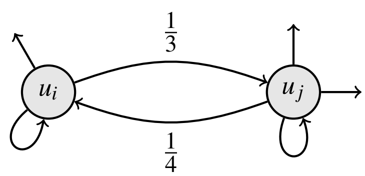
Problem 12:

In the following, we describe random walk on a graph.

Let $G = (V,E)$ be an undirected graph with $n = |V|$ nodes. We replace every undirected edge $\{u_i, u_j\}$ with two directed edges (u_i, u_j) and (u_j, u_i). The **degree** of a node is the number of edges connecting this node to other nodes. For each node u_i with degree d_i, the probabilities on edges leaving u_i is $\frac{1}{d_i+1}$, as well as the probability to stay at u_i is $\frac{1}{d_i+1}$ depicted as a self-loop. The following picture results for u_i:



Thus, for each original undirected edge, there are two values on the directed replacements, one for the direction leaving the node and one for the direction entering the node. The following picture shows the probabilities for an original edge $\{u_i, uj\}$ with u_i and u_j having different degrees:



(a) Model the probability to be at node u_j after s steps if started in node u_i, for all nodes $u_i, u_j \in V$. Specify a matrix M so that the probabilities can be calculated by multiplication with M (*Hint*: Modified adjacency matrix).

(b) Show: For $\vec{\pi} = \left(\frac{d_1+1}{2|E|+|V|}, \frac{d_2+1}{2|E|+|V|}, \ldots, \frac{d_n+1}{2|E|+|V|}\right)$: $\vec{\pi} \cdot M = \vec{\pi}$

(c) Now consider a *regular* graph whose transition probabilities on the edges are determined as described above. A graph is **regular** if all nodes have the same degree d. Thus, each node is connected to exactly d other nodes by an edge. The probability p for all edges and each direction is always the same:

$$p = \frac{1}{d+1}$$

Let some of the nodes of G be colored *red*, all other nodes are colored *blue*.

The walk starts at the first node u_1 (the index is just denoting the order in which nodes are visited). In every transition step, use an edge to enter node u_{i+1} or to stay at the current node u_i, with probability p. During the walk, one counts how often one has seen a next "red" node, and how often one has seen a next "blue" node so far.

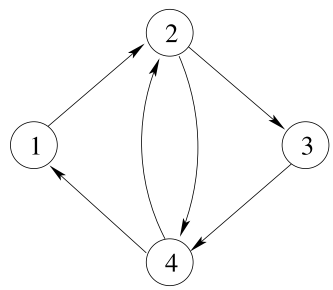
Can you give (estimate, if the walk will be stopped after finite many steps) the actual number of red nodes in G based on these numbers? If yes, how?

(d) Recall the first sheet of problems. What is the connection of the random walk in (c) to eigenvalues?

Problem 13:
The goal of the *PageRank algorithm* is to evaluate the relevance of web pages with respect to a given query. Similar to the previous problem, the algorithm simulates a user surfing aimlessly through the World Wide Web, who follows a link placed on each page with equal probability. The more often the user comes across a particular page, the more important it is considered to be, and it is given a higher "PageRank."

In this problem, we will learn the basics of the PageRank algorithm.

(a) Consider the following web graph consisting of web pages 1, 2, 3 and 4, and the depicted links:

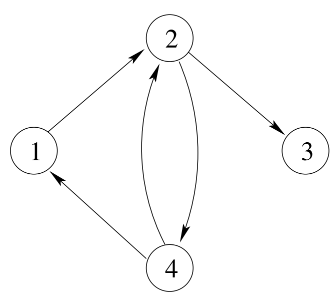


Let the vector $\vec{\pi}^{(t)} = (\pi_1^{(t)}, \pi_2^{(t)}, \pi_3^{(t)}, \pi_4^{(t)})$ denote the probabilities that the user is on the respective page at time t. Suppose the user starts on page 1, i.e., $\vec{\pi}^{(0)} = (1,0,0,0)$. Where will she/he be with what probability after two steps?

(b) Again, model the system using a matrix A such that $\vec{\pi}^{(t+1)} = \vec{\pi}^{(t)} \cdot A$.

(c) To which vector $\vec{\pi}$ does the system converge? So what is the ranking of the web pages?

(d) The following situation may occur in the web graph:



What is the problem? How could this problem be solved? (Think about what to do if you don't find a link at all, or an interesting link on a page). Using your method, determine a ranking for pages 1, 2, 3, and 4.

Final remark: Since for large graphs, such as the World Wide Web graph, the vector $\vec{\pi}$ cannot be computed exactly the PageRank is determined iteratively, using $\vec{\pi}^{(t+1)} = \vec{\pi}^{(t)} \cdot A$, and an extension with a strategy that avoids the problem from part (d).