Rolf Wanka
Matthias Kergaßner

Erlangen, 2 June 2022

Exercises to
Swarm Intelligence
Summer 2022
Sheet 5

These problems are for the meetings on June 10th and on June 14th/17th. No exercise meeting on June 7th. Please use the meeting on June 10th instead. Please start implementing your solutions at home before visiting the meetings. The room will always be *0.157-115 – CIP Pool EEI*, the times will be the regular exercise times.

Problem 14:
In this problem the **PageRank algorithm** should be implemented. The programming language may be chosen at will (Java, C++ or MATLAB are recommended).

The *PageRank algorithm* can calculate not only the relevance of web pages, but also the relevance of the chapters of a book, where there are cross-references between chapters. As an example we consider the *Taschenbuch der Algorithmen (English version of this book appeard as: Algorithms Unplugged)*, which was published in 2008 by Springer Verlag. (`doi:10.1007/978-3-540-76394-9`)

The book graph ("which chapter refers to which?") is given as XML file *graph.xml* (available on the StudOn page for the course in the directory "Exercises/Sheet 05." A picture of the book graph can also be found there in the file *graph.pdf*

Use the *PageRank algorithm* to determine a ranking for the individual book chapters.

Hints and remarks:

- The PageRank Algorithm:
  1: Initialize the relevance of each chapter to $\frac{1}{\#\text{ chapters}}$
  2: Choose d, e. g., $d \leftarrow \frac{4}{5}$
  3: **while** termination criterion not met **do**
  4: **for all** Chapter S **do**
  5: Set new relevance of S to $\frac{1-d}{\#\text{ chapters}} + d \cdot \sum_{\forall R:(R,S)\in E} \frac{\text{relevance of R}}{\text{outdegree of R}}$
  6: **end for**
  7: **end while**
- Simple XML parsers that you can use to read in the input graph are given on the StudOn page for the course as:
  - *DumbXMLParser.java*,
  - *DumbXMLParser2.m*
  - *DumbXMLParser3.cpp*
- At first, design which data structures you need (nodes, graph, ...).

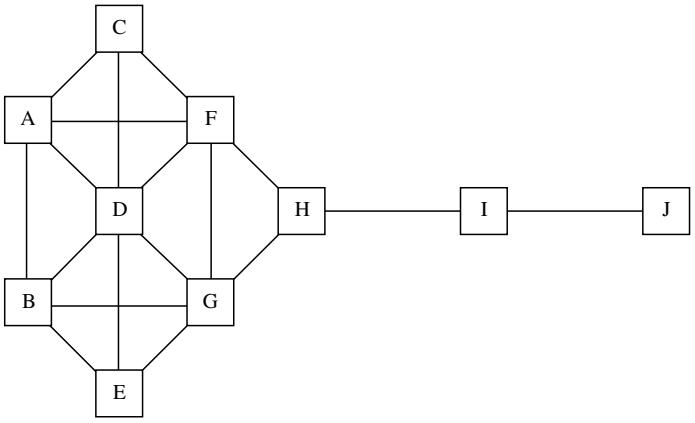
(a) What ranking do you get for $d = 4/5$?

(b) How does the choice of d affect your result?

(c) How does the initialization of the relevances (line 1) influence your result?

Problem 15:
In this problem, the analysis of social networks is considered.

The following graph, which is well known in the scientific community, describes by means of an edge which two users regularly exchange emails with each other. Because of its shape, the graph is called *Kite graph*.



The XML file *kite.xml* of the Kite graph can be found on the StudOn page of the course in the directory "Exercises/Sheet 05."

We now want to look at how "central" a user is in a network and how this can be measured in the context of our question about the measurability of emergence. Three measures are widely used, they are called **indicators of centrality**:

(1) The *Degree Centrality* $C_{\mathrm{D}}(e)$ of user e is:

$$C_{\mathrm{D}}(e) = \frac{\deg(e)}{n-1} \ ,$$

where $\deg(e)$ denotes the degree of node e, i. e., how many neighbors user e has.

(2) The *Betweeness Centrality* $C_{\mathrm{B}}(e)$ of user e is:

$$C_{\mathrm{B}}(e) = \frac{2}{(n-1)(n-2)} \cdot \sum_{s \neq e \neq t,\, t \neq s} \frac{\sigma_{st}(e)}{\sigma_{st}} \ ,$$

where σ_{st} denotes the number of different shortest paths between s and t, and $\sigma_{st}(e)$ denotes the number of those of them using e.

(3) The *Closeness Centrality* $C_{\mathrm{C}}(e)$ of user e is:

$$C_{\mathrm{C}}(e) = \frac{\sum_{t \neq e} \mathrm{dist}(e,t)}{n-1} \ ,$$

where $\mathrm{dist}(e,t)$ denotes the number of edges on a shortest path between e and t.

(a) Compute for the Kite graph the three indicators of centrality.

	A	B	C	D	E	F	G	H	I	J
C_{D}										
C_{B}										
C_{C}										

(b) Discuss the *meaning* (what is the idea behind the formulas?) and *informativeness* (do the computed numbers make sense?) of the three indicators of centrality.