Social Network Analysis for Computer Scientists
Extra Assignment

`http://liacs.leidenuniv.nl/~takesfw/SNACS`

Deadline: December 19, 2017, 13:00

This document contains two exercises that consist of various numbered questions that together form the extra assignment of the Social Network Analysis for Computer Scientists course taught at Leiden University. Please do not be late with handing in your work. You have to hand in the solutions to these exercises individually. Discussing the harder questions with fellow students is allowed, but writing down identical solutions is not. Hand in your solutions via e-mail (`f.w.takes@liacs.leidenuniv.nl`) in a PDF-file generated using LaTeX.

For each question, clearly describe how you obtained your answer. Write down any nontrivial assumptions that you make. You can use any programming language, scripting language or toolkit. All practical exercises can be done on the student workstations. Most of the practical questions can be answered using Gephi (`http://gephi.org`). In any case, always clearly describe which toolkit or programming language you used and how you obtained your answer using these tools.

Questions or remarks? Contact the lecturer or assistant, walk by room `157b` or ask your questions during one of the weekly lectures.

Exercise 1: Diameter Computation (30p)

Apply the BoundingDiameters algorithm on paper to find the exact diameter (maximum distance, length of a longest shortest path) of the undirected graph in Figure 1. The algorithm is discussed during the lectures and explained in:

F.W. Takes and W.A. Kosters, Determining the Diameter of Small World Networks, in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 1191-1196, 2011.
doi: `http://dx.doi.org/10.1145/2063576.2063748` or see
`http://liacs.leidenuniv.nl/~takesfw/pdf/diameter.pdf`.

You do not have to do the prepruning step discussed in the paper. Explain your steps in detail, and mention any nontrivial assumptions. As a selection strategy, alternate between choosing the node with the largest upper bound value and the node with the smallest lower bound value, breaking ties by taking the node with the highest degree. How many iterations did it take you to compute the diameter? Compare this value to the naive method for computing the diameter.

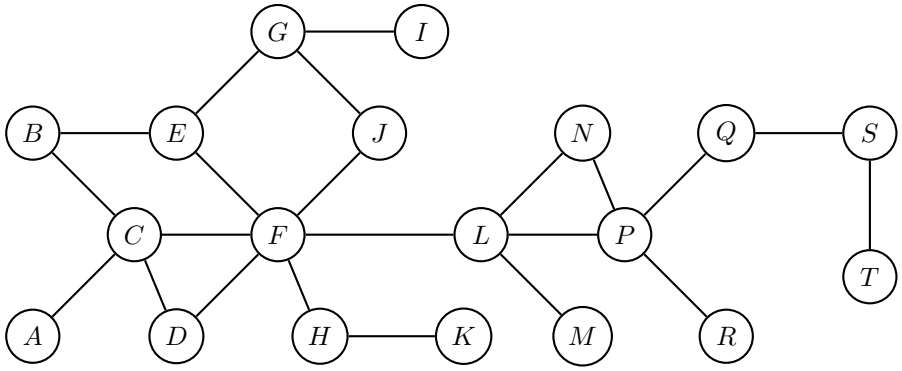

Figure 1: An undirected graph with 19 nodes and 22 edges.

Exercise 2: An Online Social Network (70p)

This is a practical exercise, for which you can use any toolkit or programming language. A dataset of an online social network is given in the shared UNIX file `/vol/share/groups/liacs/scratch/SNACS/2014/network.in`. For convenience and speed, you can copy the file located in that folder to the local hard disk (usually, that is `/scratch` on a student workstation).

The file contains a list of directed friendships of an online social network. Each line looks like `userA[whitespace]userB[newline]` and represents one directed link from a person identified by `userA` to a person identified by `userB`. You may assume that these identifiers are integers that fit in a 4-byte `signed int` in C. For each of the following questions, remember to write down how you obtained your answer and include any relevant source code.

(2×2p) Question 2.1 How many directed links does this network have?

(2×3p) Question 2.2 How many users (nodes) does this social network have? Hint: a node counts as a node if it is a source or a target of a link.

(2×6p) Question 2.3 Give the indegree and outdegree distribution of this graph (so, for each degree value the number of times that it occurs). Present the results in a representative diagram, for example generated using a simple tool such as `gnuplot` or `matplotlib`.

(2×7p) Question 2.4 How many weakly connected components and how many strongly connected components does this network have? How many nodes and links are in the largest strongly connected component of this graph?

(2×7p) Question 2.5 Give the exact or approximated distance distribution of the largest strongly connected component of this graph as a diagram.

(20p) Question 2.6 Visualize the social network in `larger.in` as a vector graphic so that it can be printed on A4 paper. Give the size and optionally the color of a node a sensible meaning based on node centrality, and explain your choices. Which visualization algorithm did you use, what parameters of the algorithm did you choose? Include your network as a full-page A4 vector graphic PDF in your report, or include it as a separate file.