Assignment 1 Social Network Analysis Deadline: 7 February

The individual learning objective of this assignment is to be able to get a first glimpse of a large network data set you are unfamiliar with. It is the essence of this exercise to face something new and unknown. For this reason, you are requested to choose one network data from the Stanford Large Network Dataset Collection:

http://snap.stanford.edu/index.html

A warning in advance: working with large network data might be problematic for multiple reasons. Be creative and make compromises if you encounter difficulties. Try to make sense also of what the data is about before you start working with it.

Exercises

1. Write an R-script that reads your data. Transform your data to a data frame. Argue about the format you have chosen to analyze large network data.
2. *Make sure that before or while doing any calculations or plots that you handle non-response / missing data meaningfully if there is any.*
3. Justify what you do with isolates and multiple components if there are any.
4. Analyze the density of your network. Create a table that contains further descriptive network statistics for your network. Please include average degree (in-degree and out-degree), standard deviation of degree (in-degrees and out-degrees), reciprocity, and transitivity.
5. Create a graph of the degree (out-degree and in-degree) *distributions* in your network. What do you observe? What do you imply from the observed degree distributions?
6. A, If there are node-level variables in the data: Create a formatted table that contains descriptive information for node level variables. B, If there are no node-level variables: Create a formatted table that contains distribution of node-level centrality values.
7. Considering the large network size, select a form of visualization that could be meaningful. Visualize your network and color the nodes according to a selected actor-variable or according to a selected measure of centrality.
8. Check assortativity in the network by centrality measures. What do you observe? Provide an interpretation of your findings. What kind of theoretical arguments could possibly explain your results?
9. And finally, a thought exercise. Assume that the robustness or vulnerability of the network is examined. Try to come up with a measurement of robustness / vulnerability and argue for the usefulness of your measure. Speculate about some implications for the concrete network.

Assignment 1 Social Network Analysis Deadline: 7 February

Please explain your answers verbally for each exercise. Send in a PDF (doc) file with your answers that clearly indicates which data you have used and a COMMENTED R script until the strictly enforced deadline of 7 February midnight (23:59 CET) to karoly.takacs@liu.se.

So, two files to be submitted:

1. **a text file with explanations (not simply an R output)**
2. **and a commented R script.**