Assignment 2 Social Network Analysis, Deadline: 22 February

For this assignment, you need to use the Correlates of War dataset, which is available freely at: https://correlatesofwar.org/data-sets
The data set contains several ties (e.g., militarized interstate conflicts, formal alliances). You have your own freedom to develop your own hypotheses using any part of this data. You could also use parts of datasets (e.g., limiting the investigation to a decade, to a continent, or to a certain tie only).

The purpose of this assignment is to get some practice in practical selection of data, describing cohesion and community structure in networks:

- by calculating basic network properties,
- by performing visualizations of networks and degree distributions,
- by performing community detection algorithms,
- and by comparing and interpreting results of different algorithms.

Exercises

Write an R-script that reads your network data, makes elementary checks, and calculates everything that is included in this assignment. You will be asked to write a textual response to the following questions:

1) Formulate theoretical arguments and expectations about interstate relations. Think in network terms. Select the appropriate part of the data for analysis.
2) On the selected data, make a table that contains descriptive network statistics for the network and for key node-level variables. Please mention the historical period covered, the nature of tie(s), the number of nodes and ties, average degree, density, the standard deviation of degrees (or outdegrees and of indegrees if network is directed), and average path length. Identify the number of isolates if there are any. Please check if the network is a single component. Which node has the highest degree, and which has the highest betweenness centrality? What do these measures represent in terms of the specific type of relations and data you have selected?
3) Visualize the network. Vary degree color and size according to key node attributes. Try to pick a node attribute that relates to an important characteristic (e.g., power, GDP per capita).
4) Implement and compare the results (community compositions, their size) of TWO different community detection algorithms to find communities (e.g., Girvan-Newman, Louvain, walktrap, cohesive blocks, Leiden). Please take into account that most of these methods can use undirected graphs only, others (such as Girvan-Newman) might work with weighted graphs. Before carrying out this task, simplify the data if necessary. Create a plot where nodes are either colored based on detected community membership, or you circulate the groups with some color in the background.
5) Make meaningful comparisons between the results of community detection algorithms. Summarize the differences and similarities and try to explain these. Please refer to the relevant literature describing the logic of these algorithms if necessary.

WRITE YOUR NAME IN BOTH the R SCRIPT (AT THE TOP) AND IN THE PDF DOCUMENT THAT YOU submit. Please send your work that contains clear written descriptions of your answers to these questions and tasks in a PDF/doc file that does not contain code, just text, visualization, and tables. In addition, please attach a commented(!) R-script or R Markdown file, and the data you used until 21 February midnight (23:59 CET) to karoly.takacs@liu.se.