

# Network Analysis Project

Social Data Management

January 16th, 2018

The objective of this *individual* network analysis project is to evaluate the capacity to analyze a network dataset. This project constitutes 50% of your final grade for the *Social and Uncertain Data Management*.

## 1 Requirements

You should download a social network dataset of your choice (except the dataset used in the practical lab) and analyse its properties.

You can take your dataset from wherever you wish. Some popular dataset repositories are the Stanford Large Network Collection (<http://snap.stanford.edu/data/index.html>) and the Koblenz Network Collection (<http://konect.uni-koblenz.de/>).

The *minimal requirements* for a passing grade are to:

- Show the number of nodes and edges in the graph.
- Draw the graph if small enough; for large graphs this may be unfeasible.
- Draw the histogram of degrees. Compare the distribution with the distribution for a random graph having the same average degree. Discuss the results.
- Draw the histogram of clustering coefficient, and the average clustering coefficient. Compare it with the one of a random graph and discuss the results.
- Draw the histogram of distances in the graphs, the diameter and the average distance. Compare with random graphs and discuss the results.

The *extra requirements* are to go beyond basic analysis, and discuss other relevant measures. Below are some suggestions, but you can add your own:

- Detect the communities in the graph, and discuss the results.
- Count the number the triangles in the graph, and compare to a random graph.
- Compute and discuss other centrality measures: betweenness, PageRank, etc.
- Do a comparative analysis of your social dataset and a non-social one (e.g., transport, Web).
- etc.

## 2 Submission & Evaluation

Send your submissions by email to Silviu Maniu ([silviu.maniu@lri.fr](mailto:silviu.maniu@lri.fr)) by **Friday, February 9th 2018, 11:59pm**, for full credit. Submissions sent by Saturday, February 10th 2018, 11:59pm will incur 5 points of penalty out of 20. Submissions received after this date will receive no credit.

Your submission should contain at least a report in PDF format, containing your analysis. In case you implemented extra code which aided you in the analysis, please include it in the submission, along with a discussion of its implementation; this will be taken into account in the grade.

The submission will be evaluated based on the clarity of the report, the correctness of its analysis (especially in the comparison with random networks) and whether it has fulfilled at least the minimal requirements above.