Network Analysis Project

Social and Graph Data Management

December 2024

The objective of this *collective* network analysis project is to evaluate the capacity to analyze a real-world graph that is larger than the one used in the practical labs. You are expected to complete the project in groups of 2 or 3 students (not 1 nor 4).

1 Requirements

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

You can source your dataset from wherever you wish. Some popular dataset repositories are the Stanford Large Network Dataset Collection (`http://snap.stanford.edu/data/index.html`) and the Network Data Repository (`https://networkrepository.com/`), others can be found in ML competition websites.

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.
- Analyze the degree correlations of the graph.
- Detect the communities in the graph, 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:

- Do a comparative analysis of your social dataset and a non-social one (e.g., transport, Web).
- Other comparisons or analysis that you may find interesting.

2 Submission & Evaluation

Upload your submissions on eCampus by **Saturday, January 4th 2025, 13:00**, for full credit. Submissions sent by Sunday, January 5th 2025, 12:00, 21:00 (Paris time) will incur up to 5 points of penalty out of 20. Submissions received after this date will receive no credit.

There should be a single submission per group (so please coordinate your submission), and the file name should be of the following form (NB1: in case of compound names just omit the internal "-" or other non-ascii characters inside the compound names, NB2: the brackets indicate you may adapt depending on nb of authors, and obviously should not be part of the file name):

`NameStudent1-NameStudent2-hw2.zip` and if there are 3 students in your group:

`NameStudent1-NameStudent2-NameStudent3-hw2.zip`

Your submission should consist of

- jupyter notebook for your code
- any additional file you may consider of interest (keep the overall size below a few MBs please)

In the report (which may be the notebook, or a separate file, as you see fit), you are expected to discuss how you used generative AI (chatGPT, copilot and the likes) to help you complete the project: how many prompts you used, how much time you spent interacting with the tool, and you are asked to provide a link to (or, if impractical, a screenshot of) the most helpful conversations. In case you chose not to use any generative AI (but we do not encourage you particularly to avoid those tools, so this would be rather surprising), instead of discussing how you used it, just add a signed statement (oath) that you did not use any generative AI. Any (limited) use of external sources, e.g., discussion with fellow students from another team, must imperatively be mentioned, with a clear explanation of how it contributed to the result.

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. You may possibly be asked (and again, you may not) to discuss the report by visio or in person, at short notice, and possibly individually. This means you (obviously) have to master the content of the code or report you submit.