Social and Economic Network Science 2024

Final project

In small groups of 2-3 persons, you should carry out a small analysis project using the techniques learned in the course (50% of the final grade).

In practice:

Choose a network dataset.

> You are welcome to use one of the datasets already presented in class. Alternatively, you can choose another suitable dataset: for example, several general-population surveys include personal network data (see slides of 28 February), and some other possible sources are listed below (not an exhaustive list). If you choose a dataset not used in class, please:
>
> - Check availability of information about the data: for example, sources, collection methods, possible problems (such as outliers, missing data, etc.).
> - Talk to the tutors in advance to make sure your choice is suitable.

Think of a substantive question that your data can help answer.

> The question should be rooted in your main social scientific discipline (economics, sociology, demography, geography, history etc.), and should be amenable to examination using your data. Don't be too ambitious: it is better to choose a 'small' manageable question than a 'big' unfeasible one; for the same reason, it might be better to aspire to provide correlational rather than full-fledged causal results.
>
> Identify the measures, models, statistics and/or network visualisations you have learned in the course that, when applied to your data, can help you answer your question.
>
> Perform your analyses in Python, using networkx. You may choose to use Gephi for visualisations.

Write a short report in which you present your analyses.

This will be done in two steps.

Preliminary proposal

First, prepare a small preliminary outline of the project you might carry out. It should not take more than one slide and consist of a very brief explanation (3 lines or possibly a figure/diagram) of the research question, the chosen data set and the analyses envisaged. At this stage, it will not be necessary to have explored the literature, nor to have performed the analyses. What is expected is only the identification, at least approximate and provisional, of a way forward.

You will briefly present your idea to the class at the time of our last session on **3 April 2024**.

Actual mini-project

You must then develop the idea you initially sketched out, taking into account any comments from tutors and other students. You will write a report, which should not exceed **four pages** (including graphs, tables, bibliography etc.). It will follow the standard rules for presenting a project:

- introduction, stating the question
- main body of the text, presenting the data, highlighting the chosen approach, and explaining the results
- conclusion, on the contributions and limits of the analysis
- list of references.

It is advisable to divide the main body of your text into numbered sections, each with a title. Explain all steps. Graphs, tables, and other materials should be numbered and accompanied by captions to aid interpretation; they can be incorporated into the text or placed in appendices.

Put together your work, which will consist of A) main report, preferably in pdf format; B) Python notebook, in .ipynb or, preferably, .html format; C) dataset[1]; D) other appendices (if needed).

Name your files Name1_Name2_Name3_A, Name1_Name2_Name3_B, etc.

Send your final file to paola.tubaro@ensae.fr and floriana.gargiulo@cnrs.fr by **30 April 2024**.

Marking criteria

The final project is marked according to 5 criteria, with a maximum of 4 points for each: 1/Originality of topic, 2/Appropriateness of interpretation, 3/Quality of analysis, 4/Methods, and 5/Writing.

Some data sources (non-exhaustive list)

Awesome Network Analysis, by François Briatte: https://github.com/briatte/awesome-network-analysis

UCI Network Data Repository, University of California at Irvine : http://networkdata.ics.uci.edu/

Stanford Large Network Dataset Collection: http://snap.stanford.edu/data/

[1] If you re-use one of the datasets already used in class, it is not necessary to include it. If you use an alternative dataset which is publicly available online, a link to its URL location is sufficient. If you use a dataset which cannot be shared because its access requires specific authorizations (for example, through the Progedo/Quetelet system), you should just explain the procedure through which you requested (and obtained) it.