

Assignment № 2: Analysing the Structure of a Social Network

Introduction to Computational Social Science
Central European University

Mark Wittek, Timur Naushirvanov

Fall 2024/25

Deadline: 11 November 2024

Necessary Background

In this assignment, you will analyse the [LastFM Asia Social Network dataset](#), which includes user interactions and preferences related to music. The network consists of nodes representing users and edges representing mutual connections. Some nodes also have features related to users' musical tastes.

The key goal is to examine the structure of the social network and identify patterns such as community structures or correlations between network centrality measures and musical preferences using network analysis techniques.

What I am supposed to do?

In this assignment, you have **two main tasks**:

1. Conduct a **network analysis** of the dataset (in Python or R), including visualisation and interpretation.
2. Prepare a **two-page research report** (about 850 words) summarising your analysis and its results (and submit it in PDF format). If you include visualisations, the report may exceed two pages, but the word count should remain around the suggested amount.

Task 1: Network Analysis

Before working with the data, it is essential to define a specific research question that you aim to address. Below are some examples to inspire you:

- What is the overall structure of the LastFM network? Can it be captured using block modelling or community detection algorithms?
- How do user centralities (e.g., degree, betweenness) relate to music preferences? Is there a correlation?
- What does the degree distribution reveal about the network (e.g., scale-free or random network)?

Data Processing Steps:

- Read the dataset into Python or R.
- Visualise the network, highlight major communities, and interpret the network structure.
- Compute the degree distribution and create a histogram.
- Apply a stochastic block model or compute correlation between node properties (e.g., centrality and music taste) or calculate a segregation measure based on music taste.

You can consult various network analysis libraries (e.g., `'networkx'`, `'graph-tool'`, `'igraph'`) for the network creation and visualisations.

NB: You are not required to perform all of the suggested analyses. Creating a visualization and conducting one additional analysis (e.g., correlating centrality scores with musical preferences or running a block model) is sufficient.

Task 2: Research Report

Once you have completed your analysis, incorporate your findings into the text! In your research report, you should summarise your work, including the data and the key results of your analysis. The suggested structure for your report is as follows (but remember, you are always free to be *creative!*):

- **15% - Introduction:** Provide background information, define the research question, and explain the motivation behind the study.
- **20% - Data Preprocessing:** Detail the data collection process and explain the methods used to clean and preprocess the data.
- **35% - Data Analysis:** Describe the analytical techniques and tools employed to extract insights from the data.
- **25% - Results and Discussion:** Present key findings, relate them to the research question, discuss limitations, and highlight any surprising results. **Briefly discuss the importance of this or similar research for Computational Social Science.**

NB: In research, it is normal to encounter results that challenge expectations or seem unclear. Your goal in this course is to explore social science issues using computational methods — not to produce groundbreaking results. Focus on engaging with the process and applying the tools appropriately.

Can I use ChatGPT or other AI tools?

Since this course is not focused on programming, you **are welcome** to use AI tools if you find yourself struggling to come up with the right **coding** solutions. However, please be transparent about this and acknowledge your use of AI in the relevant parts of your code. This will not affect your grade, but it will demonstrate your thoughtfulness and respect for the course instructors.

Grading

For grading, you submit your **Python or R code** (well-commented!) and a **two-page research report** (about 850 words). Your final grade will be based on a total of **100 points**, distributed across the following categories:

- **Code Submission and Report (50 points):**
 - You have submitted both a meaningful code implementation and a comprehensive research report.
- **Code Quality (15 points):**
 - **5 points** - Basic functionality is implemented, but the code may lack structure or proper documentation.

- **10 points** - Code is mostly well-structured, with clear functions; comments and documentation are present but may be sparse.
- **15 points** - Code is clean, well-organised, fully functional, and well-documented. The use of AI is not being abused.
- **Research Report Quality (35 points):**
 - **10 points** - The research question is not very meaningful, the research strategy is questionable, and the results are poorly discussed.
 - **20 points** - The research question is clear, but the strategy or discussion of results could be improved.
 - **25 points** - The research question is clear, and the research strategy is reasonable. Results are discussed but lack depth.
 - **30 points** - The research question is well-defined, with a solid research strategy. Results are discussed in a meaningful way, although there may be room for improvement.
 - **35 points** - The research question is well-defined, with an excellent research strategy. Results are thoroughly discussed, with insightful conclusions, and the relevance of Computational Social Science is elaborated.

In special cases, points may be awarded in increments that fall between the established steps to better reflect the quality of your work.