Mini Project 3 – Art Analysis

Note:
The mini projects are individual assignments. You could and it is encouraged to brainstorm the assignments with your classmates, but you need to have individual codes and reports. All written reports and codes are going to be checked for plagiarism.

Task Description

The cultural heritage sector is undergoing a significant transformation. Museums, galleries, libraries, and archives are increasingly digitizing their collections, reshaping how these works are accessed. This digitization democratizes cultural access, expanding the ways in which artifacts are shared and represented. However, transitioning from physical to virtual spaces demands careful consideration, as merely replicating physical collections online is not the most effective approach for engaging with them. Therefore, developing innovative methods for retrieving and interacting with this information is essential to fully realizing its potential.

For this assignment, you'll work with data from WikiArt, an extensive online art encyclopedia. This dataset provides an opportunity to explore and analyze various aspects of the art world, including artists, movements, institutions, nationalties, as well as examine their interrelationships and associations.

Data:
The data consists of 4 CSV files, generated by scraping data from WikiArt:

- **artists.csv**: *URL of the artist at WikiArt, id, image URL, nation, name, total of art work, interval of active years*.
- **relationships.csv**: *URL of the artist at WikiArt, list of friends, list of artist that they were in influenced by, list of artist that they influenced, list of art institutions that the artist studied, list of schools that was part of, type (artist or collection)*.
- **institution.csv**: *City, country, name, URL of the institution at WikiArt*.
- **schools.csv**: *name, school URL at WikiArt*.

The goal is to answer the following questions:

1. Which were the most influential artists?
2. Which were the most influential movements?
3. Which were the most influential institutions?
4. Which nationalities concentrate the majority of artists?
5. Which are the biggest communities in the network?

Step1 – Preprocessing and EDA

- Checking for missing values.
- Data distribution (movements, nationalities, artworks, etc).

Step 2 – Networkx

- Create a network based on the given data: consider the nodes, the edges, the attributes, etc.

- Visualize the network (or parts of the network) given visualization tools in python, if needed use Cytoscape.
    - **BONUS points:** interactive visualization (2 points) – you can get maxim 3 during the entirety of the course.

Step 3 – Network Analysis

- Based on differenet network analysis techniques (centrality measures, community analysis, clustering algorithms), answer the questions (Goals: 1-5) above.

Reporting

Write a scientific report including:

- **Introduction**: Description of the problem: what are you trying to solve? Elaborate on the goals.

- **Methodology**:
  Preprocessing and EDA:
  - To include: preliminary data analysis - dataset loading, data processing, EDA, relevant statistics, etc).

  Network creation and analysis:
  - To include:
    - Describe how you chose to represent the network.
    - Describe the methods used and how they are applied to answer your research questions.

- Results (based on applying the above methodology, how do you answer the questions 1-5 in the Goals section):
    - To include: description of the output of the analyses, central findings, visualizations, tables, brief discussion commenting on results etc.

- Conclusion: (what was easy, what was difficult, do your results support your initial expected outcomes, future work, etc).

You will be peer-reviewed based on the following criteria: problem statement, methodology used (feasibility and application of the methods), code (is it working or not), results, visualization, network analysis, report text quality (clarity and style – scientific or not).

You will also be graded on the overall quality of your report. To give you an idea of a proper project report, a sample is uploaded in Moodle.
You need to use Google Colab for your code. Remember to include the installation steps in the code if you use any specific package.

Upload the code and the written report into Moodle!