COMP30850 Assignment 1

Deadline: Sunday 21st February 2021

Overview:

The goal of this assignment is to construct and characterise network representations of two movie-related datasets. Your networks should model the *co-starring* relations between actors in these two dataset - i.e. the collaboration network of actors who appear together in the same movies.

The assignment should be implemented as a single Jupyter Notebook (not a script). The notebook should be clearly documented, using comments and Markdown cells to explain your code and analysis. Use tables, charts, and visualisations where appropriate to support your results.

Note: The final task in the assignment requires using the Gephi visualisation tool. The rest of the analysis should be performed in your notebook.

Tasks:

Complete the following 4 tasks:

1. **Network construction**
    - Download your 2 personal datasets, which are in a single ZIP file at:

      http://mlg.ucd.ie/modules/COMP30850/datasets/<STUDENT_NUMBER>.zip

      For example, if your student number is 145023491, your ZIP file is at:

      http://mlg.ucd.ie/modules/COMP30850/datasets/145023491.zip
    - Each dataset is represented as a JSON file, containing information about the actors and the movies in which they starred. For each dataset:
        a) Parse the JSON data and create an appropriate *co-starring* network using NetworkX, where nodes represent individual actors. You should justify your choice of network representation.
        b) Identify and remove any isolated nodes from the network.
2. **Network characterisation**
    - For each of the networks created in Task 1:
        a) Apply a range of different methods to characterise the structure and connectivity of the network.
        b) Apply different centrality measures to identify important nodes in the network. Clearly explain and interpret the results which are produced.
    - Based on the results from the above, discuss how the two networks are different (or similar) to one another.

3. Ego-centric analysis

- For each of the networks created in Task 1:
    - Select one of the important nodes in the network as identified in Task 2b. Generate an *ego network* for this node and characterise the structure of the new smaller network.

4. Network visualisation

- For each of the networks created in Task 1:
    a) Export the network as a GEXF file.
    b) Load the GEXF file in Gephi. Use the layout functionality to produce a useful visualisation of the network. Save a final image of your network visualisation as a PNG file.

Guidelines:

- The assignment should be completed individually. Any evidence of plagiarism will result in a 0 grade.
- Submit your assignment via the COMP30850 Brightspace page. Your submission should be in the form of a single ZIP file containing:
    1. Your Jupyter notebook (i.e. IPYNB file). In the notebook please clearly state your full name and your student number.
    2. The exported GEXF files for your two networks.
    3. The final PNG images for your two networks, produced with Gephi.
- Hard deadline: Submit by the end of 21st February 2021
    - 1-5 days late: 10% deduction from overall mark
    - 6-10 days late: 20% deduction from overall mark
    - Assignments will not be accepted after 10 days without Extenuating Circumstances formally approved by UCD.