Homework 4: Network analysis and visualization

Logistics

- This is a group assignment. Submit one zip file per project group.
- Due date: 12:30pm Tuesday, November 30, 2021
- Method: Upload a zip file to eTL. The zip file name should be [group number].zip such as [10.zip] without brackets.
- Zip file content (file names are case sensitive):
  - report.pdf (document that contains your story and visualizations)
  - *.py or *.ipynb (any Python code or Jupyter Notebook file you created for this homework)

Background

We briefly touched on bipartite graph in lecture. We are going to learn more about the bipartite graph and its projections in this homework.

The student-kaggle pair dataset that you used in the previous homework is a classic example of bipartite graph. You have two different types of nodes: Person and Dataset. Edges are defined as whether a person liked a dataset.

To do list

- Download the following csv file: sp21_student_kaggle.csv.
- The following items will be checked for the grade.
  - Create an undirected bipartite graph by referring to the following documentation. (0.5 pt)
    - https://networkx.org/documentation/stable/reference/algorithms/bipartite.html
  - Create a node-link diagram differentiating the two types of nodes. For example, you can use different colors to denote Person and Dataset. Also, encode edge weights into the visualization. For example, you can use alpha or width to encode edge weights. (1 pt)
  - Answer this question, "How many components are there in this graph?" (0.5 pt)
  - Extract the largest component by referring to the following documentation. This largest component is called the "main component" of the graph. (0.5 pt)
    - https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.components.connected_components.html
  - List the names of people who do not belong to the main component. (0.5 pt)
  - Convert the main component (which is still a bipartite graph) into two unipartite graphs by referring to the following documentation. (0.5 pt)
    - https://networkx.org/documentation/stable/reference/algorithms/bipartite.html
  - Visualize both persons graph and datasets graph using node-link diagrams. (1 pt)
  - Compute closeness centralities for all nodes in the persons graph. (0.5 pt)
  - Compute betweenness centralities for all nodes in the persons graph. (0.5 pt)
  - Compute the correlation coefficient between the two centrality measures. (0.5 pt)