THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY
Department of Computer Science and Engineering
COMP4641: Social Information Network Analysis and Engineering
Spring 2023 Assignment 3
Due time and date: 11:59pm, May 7 (Sun), 2023.

IMPORTANT NOTES

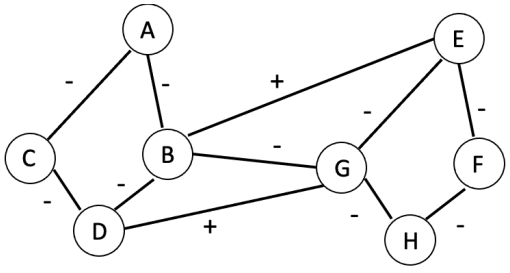
- **Your grade will be based on the correctness and clarity.**
- **Late submission: 25 marks will be deducted for every 24 hours after the deadline.**
- **ZERO-Tolerance on Plagiarism: All involved parties will get zero mark.**

Q1. Centrality

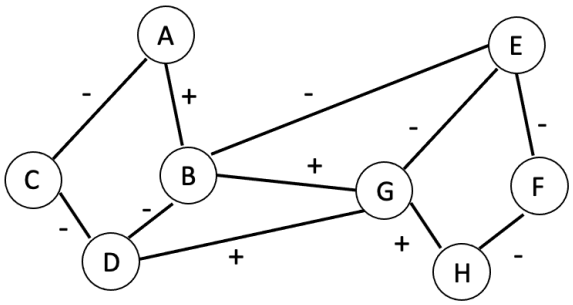
In this question, we compute various centrality measure on the Football dataset[1] using NetworkX. For a particular centrality measure, we consider the top-10 values as high, and the bottom-10 values as low.

1. Follow the steps from NetworkX official website[2] to import the dataset and visualize the graph.
2. Find nodes with high value in one centrality measure but low value in another centrality measure. Note that it is possible that such nodes do not exist.
   (a) Find nodes with high degree centrality but low closeness centrality.
   (b) Find nodes with high degree centrality but low betweeness centrality.
   (c) Find nodes with high degree centrality but low eigenvector centrality.
   (d) Find nodes with high closeness centrality but low degree centrality.
   (e) Find nodes with high closeness centrality but low betweeness centrality.
   (f) Find nodes with high closeness centrality but low eigenvector centrality.
   (g) Find nodes with high betweeness centrality but low degree centrality.
   (h) Find nodes with high betweeness centrality but low closeness centrality.
   (i) Find nodes with high betweeness centrality but low eigenvector centrality.
   (j) Find nodes with high eigenvector centrality but low degree centrality.
   (k) Find nodes with high eigenvector centrality but low closeness centrality.
   (l) Find nodes with high eigenvector centrality but low betweeness centrality.

Q2. Signed Network You are given the following two signed networks. By following the procedure in the lecture notes, show whether they are balanced or not.



[1] http://www-personal.umich.edu/~mejn/netdata/football.zip
[2] https://networkx.org/documentation/stable/auto_examples/graph/plot_football.html



Q3. Power Law

In this question, you have to estimate the α of the power-law distribution of the Facebook Government network. It contains 7,057 nodes and 89,455 (undirected) edges. The dataset can be downloaded from `https://snap.stanford.edu/data/gemsec-Facebook.html`. The detailed steps are as follows.

1. Rename the dataset to government_edges.txt and load it.
2. Show the degree distribution using a log-log plot.
3. Use linear regression on the degree distribution (starting from degree=20). Show the regression fit, and estimate α of the power-law distribution.
4. Show the complementary cumulative denstiy function (CCDF) using a log-log plot.
5. Use linear regression on the CCDF (starting from degree=20). Show the regression fit, and estimate α of the power-law distribution.

We have prepared a code template at `https://colab.research.google.com/drive/16lwh6QobQfNZ_l96b897YtU9Z_diCcQY?usp=sharing`. Steps 1 and 2 have already been provided for you. Please implement steps 3-5.

Submission Guidelines

Please submit a Python notebook A.ipynb for your code, and a report (report.pdf) for your results and answers. Zip all the files into A_awangab_12345678 (replace awangab with your ust account and 12345678 your student id). Please submit the assignment by uploading the compressed file to Canvas. Note that the assignment should be clearly legible, otherwise you may lose some points if the assignment is difficult to read. Plagiarism will lead to zero point on this assignment.