

THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY
Department of Computer Science and Engineering
COMP4641: Social Information Network Analysis and Engineering
Spring 2023 Assignment 1
Due time and date: 11:59pm, Mar 12 (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.**

NetworkX

In this question, you are required to use NetworkX to do basic data analysis on a Facebook Government network dataset. It contains 7,057 nodes and 89,455 (undirected) edges. The dataset can be downloaded from <https://snap.stanford.edu/data/gemsec-Facebook.html>.

1. Use the function `nx.read_edgelist(filepath, delimiter=',')` to load the dataset `government_edges.txt` as G .
 - Rename `government_edges.csv` as `government_edges.txt` before loading.
 - Output (i) the number of nodes and edges of G ; and (ii) the average degree of G .
 - Show visualization of the graph using the online tool: <https://cosmograph.app/>.
2. Use `nx.gnp_random_graph()` to generate a undirected graph G_{np} such that (i) the number of nodes is the same as that of G ; and (ii) the expected mean degree is the same as the average degree as G .
 - Output (i) the number of nodes and edges of G_{np} ; and (ii) the average degree of G_{np} .
 - Write G_{np} to file by `nx.write_edgelist(Gnp, "random.csv", data=False)`. Visualize it using the online tool: <https://cosmograph.app/>.
3. Plot the degree histograms (as on **p11** of `Lecture_4_SmallWorld.pdf`) of G and G_{np} , using both linear-linear and log-log scales on the axes.
 - Compare the histograms, what conclusions can be drawn?
4. Plot the number of paths versus the (shortest) path distance (as in **p19** of `Lecture_4_SmallWorld.pdf`) for G and G_{np} .
 - From the plots obtained, do G and G_{np} have similar average path lengths?
5. Plot the average clustering coefficient versus degree (as in **p26** of `Lecture_4_SmallWorld.pdf`) for G and G_{np} .
 - Both axes should be in log scale.
6. Plot the number of connected components with size (as in **p30** of `Lecture_4_SmallWorld.pdf`).
 - From the plots obtained, what do G and G_{np} have in common?

Submission Guidelines

Please submit a Python notebook `A.ipynb` for your code, and a report (`report.pdf`) for your results and answers. The submitted folder should be 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.