ELL 880 - Social Network Analysis – Fall 2024
Assignment 1
(Due August 16, 2024)
Mark: 8

Properties of a real Social Network utilizing tools and libraries of your choice

Part A – Dataset

1. Download a Social Network Data from the links given below.
   a. It should be sufficiently big (nodes> 1000)
   b. Do no choose a very large dataset – you may have problems with your assignments.
   c. Once you select a dataset make an entry here:
      https://docs.google.com/spreadsheets/d/1MJFTk6RcoFL4gddyWsS5rOo7IcVM1tma5abOnhn5ROc/edit?usp=sharing
   d. **Make sure the dataset you use is unique to you (marks will be deducted if students use same Data set)**
   e. Note that you will use the same dataset for Assignment 2 (GNN).

Part B – Tools [3 marks]

2. Calculate the Degree Distribution
   i. Assign sizes to vertices based on their total degree.

3. Filter the network by degree such that only the:
   i. Bottom 10% of nodes and the connection among them are visible.
   ii. Top 5% of nodes and the connections among them are visible.

4. Find
   i. All the connected components of the network
   ii. The size of the giant component of the network

Output: Paste screenshots in a document and submit in Moodle

You can use any tool that you want – but Gephi may be easiest for you.

Part C – Libraries [5 marks]

5. Now try to do the same exercise using a library of your choice.

The program should be written in such a way that it reads in the input network by itself (preferably from a hard-coded URL of the dataset) and does not as for any input.

Output: Submit in Moodle the link to the Jupyter notebook/Google Co-lab.

You can use any library that you want – but SNAP or NetworkX may be easiest for you. You can use more than one library.

List of Dataset Sites

1. https://www.kaggle.com/
2. http://networkrepository.com/
3. https://snap.stanford.edu/data/
4. http://www.sociopatterns.org/datasets/
5. https://networkdata.ics.uci.edu/
6. https://graphchallenge.mit.edu/data-sets
7. https://www.pik-potsdam.de/members/donges/network-datasets