Social Computing (CS60017)
Autumn 2021
Assignment 1: Measuring network structure

Assignment given: 16th September, 2021
Deadline: 26th September, 2021
Full marks: 40

[IMPORTANT] General instructions:

- This assignment teach you the actual network structures in the real world networks
- We strongly suggest using the snap library for graph analysis: https://snap.stanford.edu/snap/
- Please add one readme.txt file with instructions for running your code for each question. If we cannot run your code, there might be severe penalty.
- Please add one answers.txt file with answers to each of the questions as output by your code. The answers in the file should be generated by your code at runtime only.
- Please submit your code files, readme.txt and log.txt in a single zip file named <ROLL_NO>_Assgn1.zip
- Please follow the nomenclature of the files strictly as instructed above.
- If we detect plagiarism, there will be severe penalties. So please don't copy.
- Please use CSE Moodle to submit your zip file.
- It's preferable that you use c++ or python for your code.

Q1. Download the dataset and write a program to generate the following structural metrics given below.
http://snap.stanford.edu/data/ego-Facebook.html

A. Structure of the network- [1+1.5+1+1.5+ 2 + 3 = 10]
    1) Calculate the number of nodes present in the network.
    2) Remove all nodes with IDs divisible by 3 and calculate the number of nodes present in the network.
    3) Calculate the number of edges present in the network.
    4) Calculate the number of edges present in the network after removal of all nodes with IDs divisible by 3.
    5) Find out the number of nodes with the highest degree. Your code should print the node IDs with the highest degree in the network.

6) Plot of the distribution of the shortest path lengths in the network. Your code should create the plotted image in the “plots” directory in a file named, “shortest_path_<network_name>.png”.

B. Components of the network- [2+2+2+3+1 = 10]

1) Find out the fraction of nodes in the largest connected component of the network.
2) Number of articulation points: A node is an articulation point if, when removed, increases the number of connected components. Your code should print the number of articulation points in the network.
3) Calculate the number of connected components in the network.
4) Plot of the distribution of sizes of connected components. Your code should create the plotted image in the “plots” directory in a file named: “connected_comp_<network_name>.png”.
5) Find out the diameter of the largest connected component of the network.

Q2. Download the SNAP “Email-EuAll” dataset http://snap.stanford.edu/data/email-Eu-core.html.

A. Size of the network [1.5 x 2=3]

(a) Your code should print the following line to stdout: Number of nodes
(b) Your code should print the following line to stdout: Number of edges:

B. Degree of nodes in the network [1.5 x 2=3]

(a) Number of nodes which have degree = 4 Your code should print the following line to stdout: Number of nodes with degree=4
(b) Plot of the Degree distribution Your code should create the plotted image in the “plots” directory in a file named: “deg-dist.jpg”

C. Calculate the following- [3 x 2=6]:

(a) Find out the number of nodes and edges in largest weakly connected component (WCC)
(b) Find out the number of nodes and edges in largest strongly connected component (SCC)

D. Calculate the average Clustering Coefficient [2]

E. Calculate number of triangles and rectangles. [4]

F. Calculate Number of edge bridges: An edge is a bridge if, when removed, increases the number of connected components. [2].

Your code should print the following line to stdout: “Number of edge bridges”