**CSE 5245 (SPR’18)Introduction to Network Science – Lab 1**

This assignment will allow you to explore network node measures of interest (as we just discussed in class) along the axes of effectiveness, efficiency to compute and redundancy (many of these measures are correlated). All aspects will need to be executed on stdlinux and we will be using Python (a language/environment worth learning about in its own right!). I recommend the following tutorial for those less familiar with python: http://web.cse.ohio-state.edu/~champion.17/3461/python\_tutorial.pdf

 There are three components to this assignment:

1. Implementation of various measures using the Python NetworkX library (which is located in /usr/local/python-2.7.10/lib/python2.7/site-packages/)
2. For each node in each graph the value of each of the measures.
3. Meta-level analysis of the correlation of measures (within a graph) and across graphs.

**Preliminaries:**

You will need to set your environment variables in stdlinux to subscribe to Python (use the subscribe command on your Unix shell). Execute permissions need to be set to the script (chmod +x examples.py). **Please take care of the preliminaries soon and see the TA to check if your environment is not working – do not wait till the last minute!**

You will focus on the following four datasets which you will need to download:

<https://snap.stanford.edu/data/wiki-Vote.html>

<https://snap.stanford.edu/data/ca-GrQc.html>

<https://snap.stanford.edu/data/p2p-Gnutella08.html>

<https://snap.stanford.edu/data/egonets-Facebook.html>

**Assignment Details:**

You will compute/implement the following node measures, several of which are directly available on Networkx (some you may need to implement):

1. Degree Centrality (normalized)
2. Closeness Centrality (normalized)
3. Betweenness Centrality
4. Eigenvector Centrality
5. Pagerank Centrality (with d=0.85)
6. Clustering Coefficient
7. For each node in each graph your program will output the value of each of the above 6 measures.
8. Perform a meta-level analysis of correlation of various measures within a graph and perhaps attempt to characterize different graphs based on this meta-level analysis (do you find anything interesting? Think about time to compute and look at specific nodes for which some measures are high while others are low – look at overall correlation across measures. This is your chance to explore and possibly come up with new/novel observations.)

The command to use once you are ready to submit is: **submit c5245aa 01 <files to submit>.** The submitted set of files should include all code; a readme file describing how to execute/run code and your report. Your report should be precise and defend both your basic design choices and your meta-level analysis. We will discuss possible conclusions/analysis choices in class.

This is a team project – teams of two. It is recommended that at least one team member have some priori experience with Python. Graduate students may elect to do this assignment on their own. Please ensure that both team members contribute equally to all aspects – implementation, documentation and report writing. Each report must have a small section describing the contribution of each team member.

**Feedback:** That said we welcome feedback on any difficulties that may arise at any time. Please do not hesitate to contact either of us with any questions.

**Due Date:** Sunday February 4th 2018 by midnight.