***Data620 - Week 4 Assignment (Spring 2019)***

**Member:** Joseph Garcia

**Week 4 Assignment**

Centrality measures can be used to predict (positive or negative) outcomes for a node.

Your task in this week’s assignment is to identify an interesting set of network data that is available on the web (either through web scraping or web APIs) that could be used for analyzing and comparing centrality measures across nodes. As an additional constraint, there should be at least one categorical variable available for each node (such as “Male” or “Female”; “Republican”, “Democrat,” or “Undecided”, etc.)

In addition to identifying your data source, you should create a high level plan that describes how you would load the data for analysis, and describe a hypothetical outcome that could be predicted from comparing degree centrality across categorical groups.

For this week’s assignment, you are *not required* to actually load or analyze the data. Please see also Project 1 below.

You may work in a small group on the assignment. You should post your document to GitHub.

***Solution***

***Background***

The data that I am using is "Bitcoin OTC trust weighted signed network" by Jure Leskovec from the Stanford Network Analysis Project (SNAP**.** Since Bitcoin users are anonymous, there is a need to maintain a record of users' reputation to prevent transactions with fraudulent and risky users.

***Data Source***

Nodes: 5,881

Edges: 35,592

Range of Edge weight -10 to +10

Percentage of positive edges 89%

***High Level Plan***

The data would be loaded into a Jupyter notebook with Python networkx package to do the analysis.

***Hypothetical Outcome***

A hypothetical outcome would be to verify via network analysis that will have positive trust and negative trust. Using the categorical variable that is well trusted will likely have many connections, and consequentially, a higher degree centrality, than others.