Data 620 – Week 4 Assignment

Nicholas Schettini and Amanda Arce

**Requirements:**

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.

**Data Set:**

The open flights data base <https://openflights.org/>

According to the openflights website, “OpenFlights is a tool that lets you map your flights around the world, search and filter them in all sorts of interesting ways, calculate statistics automatically, and share your flights and trips with friends and the entire world (if you wish). It's also the name of the open-source project to build the tool.”

Also according to the website, the database has over 10,000 airports and other travel methods.

**Plan:**

We will import various packages in Python, such as Pandas, numpy, networks, and more. We will attempt to analyze US and Canada airports and if possible, domestic and international flights.