# DATA 620 Assignment Week 4 Homework

Web Analytics, Spring 2022

Prof: Alain Ledon

Members: Joshua Hummell, Scott Reed

## 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.

# Data

The data source that I have identified for the assignment and for the project one is the IMDB Movie database. The data set contains a plethora of information, but we will be looking at the connection between actor 1 and actor 2. The dataset contains useful categorical data such as Genre, Rating, Director etc. The data is in a text file and was created by Katie Truong, which can be found at <https://github.com/katie-truong/Jupyter>. This dataset has been pre-cleaned by the author, but we will have to remove several columns as well as match up and create another column. We want to add a column that shows the gender of the actors/actresses.

# High Level Plan

We will download the .csv file from the source page and then load the .csv file into a Jupyter notebook. From there we will use the *networkx* package to do the analysis.

We believe that those Actor One nodes that are female will have higher degree centrality on average as compared with those Actor Two nodes that have negative trust. It seems reasonable that actors and actresses will have roles with each other.