*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 by end of day on Sunday.*

I plan to use Zachary's Karate Club dataset for this assignment. This dataset represents a microcosm of social interactions within a karate club. The dataset consists of nodes that represent individual club members and edges that signify friendships or interactions between them. Analyzing this dataset provides a unique opportunity to understand the relationships and social dynamics within the club.

We plan to download the dataset online (from http://konect.cc/networks/ucidata-zachary/) and save it to my github repository. We plan to use the NetworkX library to load the graph from the dataset.

Degree centrality measures the number of connections or links that a node (individual) has in a network. By comparing degree centrality across categorical groups, such as gender, age, or occupation, we can assess the extent to which individuals within different groups are socially connected or influential within the network.

For example, let's consider a social network analysis conducted on a professional networking platform. We divide the individuals into two categorical groups: "Senior Professionals" and "Junior Professionals." By calculating the degree centrality for each group, we can compare the average degree centrality values and draw conclusions about the social influence or popularity of individuals in these groups.

A hypothetical outcome could be that the "Senior Professionals" group exhibits higher average degree centrality compared to the "Junior Professionals" group. This would suggest that senior professionals, who have more experience and established connections, tend to have a larger number of professional connections or interactions within the network. Higher degree centrality in this context could indicate greater social influence, reputation, or access to resources within the professional community.