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.

The dataset selected for this assignment is based on the collaboration among Spotify artists. It tracks the different musicians that are on weekly charts (tend to be more popular) that worked with other artists to publish music. The categorical variable in this instance is genre of music, which is tracked on a one to many basis in a list format. It might be easier to select the first item from a list of genres for presentational purposes to simplify the information on the nodes. The dataset is available for download on Kaggle; however, it is separated into two input files based on edges and nodes although it should be fairly easy to join the two together given there is a unique identifier present in both files. We would download the files and save to Github, which would make it easy to reference the files in our Jupyter notebook and utilize Networkx pandas method to import into the workbook.

By calculating centrality measures based on a categorical genre we might attempt to predict which music genres are more likely to be utilized by musical artists when releasing songs. This might potentially signify which areas of music might encourage more collaboration or of groups of artists that may be expanding the traditional boundaries of their genre. It may also signify whether artists are willing/able to help others get their first break with a hit song as music discovery has now shifted substantially from studios to artists allowing many lesser known musicians to be found/discovered on platforms like Spotify. If you were an aspiring producer perhaps identifying new talent via the centrality metrics/network analysis might help you partner with them to create more music.