DATA 620 - Assignment 3

For this project, I am going to extend the work that I did on Assignment 2 and work with the dataset for movie ratings. However, this time, I want to focus more on which movies are watched and rated by the same user.

While I did not incorporate a categorical variable in the data set I used in assignment 2; for this assignment I will use the tags data set associated with movie titles to generate a “genre” categorical variable for each movie title.

My plan of action would be to create a data set that is based on the individual movies watched by each user and affix the “genre” categorical variable for each movie. The data network will represent the relationship between movies that were watched by the same user. Thus an edge represents the connection between two movies that were watched and rated by the same user. Based on this network, one of the hypothetical outcomes that I believe could be predicted from comparing degree centrality across categorical variables, is what is the most universally watched movie and genre. In other words, are there movies that appear to be universally popular? I also imagine that there will likely be individuals who mostly watch films of a specific genre; so it will be interesting to see if there are movies that are so common/popular that they are able to cross over such that they are watched even when there are certain genres that predominate.