Written Report Final Project DS210

This project takes a data set that involves social circles from Twitter. The dataset includes node features (profiles), circles, and ego networks. I first read in the txt file with the data set on it and created a graph. I implemented page rank to gain insights on connectivity in the graph, all in which outputted low rank scores such as 0.0000024, which tells us that there is some degree of connectivity in the graph, but not a strong connection. I then implemented friends of friends to further identify nodes (a user on twitter in my case) to single out who may be connected. Six degrees of separation was used to find on average how separated users are, which most users separated by a degree of 3 or 4. Breadth first search was used to find the shortest path to the last node of the graph, once again to strengthen analysis on connectivity in the graph. All in all, I sought to understand how connected social users on twitter are. I chose this focus because Twitter is a large social platform with a variety of users. However, because of the nature of twitter being social media, it is implied that there are mutual connections with different users, but there are many social circles. This analysis allowed me to gain insight on social circles in my data sample. In future analysis, I would like to visualize the social circles, identify the user with the highest connectivity, and overall understand the connections of the social circles and why they are created.