DS210 Final

I chose to do my project on six degrees of separation. I am comparing the degrees of separation between github users and facebook users and seeing which has the average shortest path.

For my data set, I used Stanford's SNAP database. The links are below:

https://snap.stanford.edu/data/ego-Facebook.html (Facebook)

https://snap.stanford.edu/data/github-social.html (Github)

The facebook data file was a txt file and did not require any cleaning or preprocess as the nodes and edges were separated nicely. For the Github csv file, I needed to convert this file into a txt file. I used python for this step labeled "preprocess". I converted the file into a txt file and removed all the commas.

The data file contained two columns. Each column has nodes of users and the edges are created between these nodes.

I created two functions to run my code. I first had to create a function to construct a graph with nodes and edges. I used hashmaps in order to read the data file more efficiently.

The second function is calculating the average shortest distance. I used breadth first search and hashmaps to get the distance in efficiently.

I have two testing codes. The first test is testing the graph construction and if it creates the right amount of nodes and edges. The second test is testing if the average shortest distance calculation is approximately accurate.

Finally, the main code calls the two data sets and runs it through the two functions. On average Facebook has around 3.66 degrees of separation and Github has around 3.22 degrees of

separation. (Note these numbers are approximations) To conclude Github users are more connected than facebook users as they have a shorter path in terms of degrees of separation.

Overall, I found it interesting that both Github and Facebook users have fairly low degrees of separation. At the same time, under 4 degrees of separation makes sense. If we look at Linkedin, they label it as 1st, 2nd, 3rd+ degrees of separation. It seems like beyond 4 degrees of separation is very unlikely on social networking programs. Also people do not need to "know" each other online to be friends. I think this also contributes to why the degrees of separation are low online.