Jaywoo Jo
Holly Lee

Hemank Kohli

CS225 Project Goals

For our project, as the assignment describes, we will be loading a data file as a graph and use that graph to run algorithms of interest to the data set. As the AMA was just a few days ago, we have not yet had the time to speak with our mentor or go over detailed aspects of running the specific data set we're interested in. As a result, we want to preface with the fact that the following might be susceptible to change in the following days.

We are interested in the Social Circles: Facebook data set (http://snap.stanford.edu/data/ego-Facebook.html) using Stanford's SNAP by Jure Leskovec. Due to the simple format of these data sets, this seems a rather straitforward option for this project. The dataset consists of “circles” or a ”friends list” from Facebook, consisting of roughly 4039 Nodes and 88234 Edges. It states that the features of the dataset have been anonymized to protect users. Social circles help to organize personal social networks, and are also implemented in Google+ and Twitter. Each circle is a subset of a user's friends, and may be disjoint, overlapped, etc. While we are not too sure on how circles and ego networks work, we will be investigating the graphs during the project.

Our goal is to run one of the Traversals covered in class and two other options from the complex or uncovered options we were provided with, and output this data. We are currently interested in running a BFS Traversal, perhaps the Shortest Path using Dijkstra's Algorithm and a Graph Coloring or PageRank.

The challenge in the next few days will be finding a way to output the data, and building code and test cases for our algorithms, for which we will be speaking with our mentor soon to figure out the details.