DS 210 Final Project Report

In the early days of social media and social networking, Google put out their own version called “Google Plus”, which was rather popular among google users because no extra account was necessary past a google email account. With this being said, the social network collected data on the users to put into a data set.

Thus, the data set I chose to use for the final project is the Google Plus data, which details the “circles” from the Google+ social networking site. The graph is directed, meaning that the edges that connect nodes are directed from one node to another (rather than being bidirectional). I imported this data set from the Stanford University data based “SNAP” as a text file of large integers, which represents users of the social network.

In my code, I used functions to first read the text file and convert to a vector of tuples, where the tuples are the two connected users. From there, I made my graph in the form of a hashmap where the keys were the starting nodes and the values were the nodes that these starting nodes were connected to. From there, I iterated through my graph to determine which node has the most connections, which I found was the key “111091089527727420853” which had 49041 connections. Wow, someone had a big circle! This means that the user associated with the key “111091089527727420853” had connections with 49041 other users.

What I saw on the database website that I thought was interesting to note is that the data was only gathered by those who chose to share their “circles”. With that being said, there were still so many data entries of users.

Reflecting on my results, I also find it interesting that someone had that many connections! I am wondering if this is some sort of celebrity, but I don’t remember that many people using Google Plus, although it was slightly popular among my friends and I. This makes me interested to see what similar data among larger networking sites such as Twitter, Instagram, or Facebook would look like now.