

Kyle Yu

Professor Kontothanassis

DS 210

13 December 2023

Final Project Report

Goal + Dataset + Motivations:

My project aimed to analyze the principle of six degrees of separation, answering the question: What is the usual distance between pairs of vertices in a graph? For the dataset, I examined the Stanford Network Analysis Project's Facebook data set on social circles. The file, "facebook.txt" contains 4039 nodes representing various Facebook accounts and 88234 edges representing these accounts' friends on the social media app. I was intrigued by this theory because during my time here at BU, I have met so many people who are friends with my friends not only from BU but friends from California or other places. Every time this would happen, I would always be so amazed. A recent example is a friend that I met through a club here at BU and how she is friends with a bunch of people that I am friends with at her high school in California. The vast, intertwining network of people fascinated me, especially because professional networking is so important these days.

Approach + Methodology:

I read the "facebook.txt" data from the file into the memory and transformed it into a graph format with each node representing a Facebook user and each edge representing a friendship between two users, using a struct and implemented function. These two functions were in separate modules, which are named "graph.rs" and "read.rs." I employed the Breadth-First Search algorithm from the "functions.rs" module to determine the distances between each node and all other nodes in the graph. Using the count_degree function from the "functions.rs" module, I was able to categorize the nodes based on their average distance to other users.

[illegible]

[illegible]

2. a vector of vectors containing the distance between nodes (1, 2, 3, 4, 5, 6, 7) and total number of nodes within those distances:

$$[[1, 675], [2, 1157], [3, 857], [4, 350], [5, 207], [6, 103], [7, 31]]$$

3. the total average distance between nodes:

2.187175

Conclusion:

Based on result 2 for an average distance of 1, 2, 3, 4, 5, 6, and 7 average distance to other users, there were 675, 1157, 857, 350, 207, 103, and 31 nodes. The results illustrate that the majority of people are within 6 degrees of separation from each other. As you can see in the pie chart, the shades of green are within the 6 degrees of separation, while the tiny sliver of red represents over 6 degrees of separation. To further prove my point, I calculated the final, total average distance between all nodes, which came out to be 2.187175. We can clearly see, through my exploration of Facebook data, that the principle of 6 degrees of separation is true.

Average Distance	Number of Nodes
1	675
2	1157
3	857
4	350
5	207
6	103
Over 6	31

