

Sumin Hong

Professor

DS 210

14 December 2023

Final Project Report

DS210 Final Project Using BFS Algorithm

For my project, I used “Social circles: Facebook” dataset from Stanford Network Analysis Project(SNAP) <https://snap.stanford.edu/data/ego-Facebook.html>. Part of this dataset is “facebook_combined.txt” which consists of ‘circles’ (or ‘friends lists’) from Facebook. This data was collected from survey participants using the Facebook app. There are 4039 nodes and 88234 edges in which nodes represent the undirected user on Facebook and edges represent their connections such as following or followed list. I selected this data because I was interested in the data set of social media how the follow recommendations work and how the following list is related to each other of users. I also found the six degrees of separation interesting, which is a concept of social network where two people who do not know each other would be linked together within six degrees of relationship. By analyzing the distance between pairs of vertices in the graph, I thought that I could figure out if the six degrees of separation would actually apply to actual case in social media.

In order to figure out the correctness of question, I used a method of BFS (Breadth-First Search) algorithm. This algorithm examines adjacent node first instead of digging into one branch and then starting another branch in the nodes. This method is frequently used when looking for a shortest distance between the nodes since it enables the algorithm not to go through all possible edges. Before I perform BFS to calculate the distances between each node, I first

turned the text file into a graph format. Then, using BFS algorithm, calculate the distances, and use the average distance from every vertex to other vertices as a conclusion to show if the six degree of separation applies.

Getting into more specifics, in the `readfile.rs` modules, I implemented `read_file` function in order to read the text file, `'facebook_combined.txt'` where each line is unwrapped in order to be formatted for graph. It is then used in main function in `main.rs` module and graph were created with `create_undirected` function from `bfs.rs` from the list of edges. It also prints adjacent list of each node which is also printed when tested run. From this, we can see how each node is connected to other and which one is connected to each other. By using `compute_distances_bfs` from `bfs.rs`, it is possible to calculate the distance between nodes which is in graph format from previous code.

As a result, I have observed that the average distance between vertices in facebook is 3.69, which is a small number compared to what I predicted it to be six. From the 4039 users and 88234 connectivity of them on Facebook, it is found to be 3.69 connections apart on average between each user, which means if you find someone and go on the following lists, it is about four friends of connectivity apart from you. From this analysis, it is concluded that the concept of six degrees of separation is applied since the average distance is within six and this finding how social networking can occur between relationship.

How to Run:

1. Download the file called "Final_Project."
2. On terminal, navigate to the directory of the file downloaded.
3. Run the code using `<cargo run --release>`

References:

<https://snap.stanford.edu/data/ego-Facebook.html>
<https://stackoverflow.com/questions/71189961/bfs-algorithm-tracking-the-path>
<https://payasr.github.io/Are%20Graphs%20hard%20in%20Rust.pdf>
<https://www.sotr.blog/articles/breadth-first-search>