Jason Huang

Professor Kontothanassis

DS210

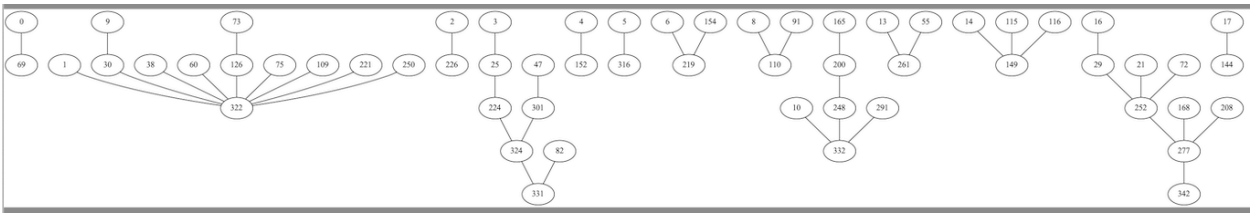
4 May, 2022

Facebook Social Circles

The data set I picked is from Stanford Large Network Dataset Collection. The dataset can be found at this URL: https://snap.stanford.edu/data/ego-Facebook.html. The data set is named ego-Facebook, representing anonymized social circles from Facebook as an undirected graph with 4,039 nodes and 88,234 edges. I found this to be an interesting dataset as it represents interactions between people. These graphs are powerful tools for understanding the complex social relationships throughout society, and how individuals influence each other. These social network graphs can be useful in many fields including, sociology, marketing, and politics.

An interesting part I deduced from the dataset is that the average distance between pairs of vertices is 3.6925. Meaning people within this dataset (representing the US), are connected to each other by an average of 3.6925 degrees. Consistently when doing random walks over the data with walk lengths of 6, there are particular vertices that are highly occurring. I choose 6 based on six degrees of separation. Those vertices are 3437, 107, 1684, and 0. These vertices are likely more “connected” than other vertices, and that’s why they are consistently ranked higher. Attached to the output is a dot and a png file to represent the graph. Attached below is a portion of the png file. This png file is a minimal representation of the actual graph as it only considers

one edge from every node. An actual representation of the graph using petgraph would be much more complex and reduce readability and output efficiency.



The code is organized into several modules. The graph module creates the graph. The print_graph module generates a visualization of the graph. The exploring module contains the algorithms being run on the graph. Lastly, the tests module contains a few tests for this project.

To run the file, enter the directory from the terminal/command, and run the package with cargo run. Attached below is the output from my directory. The run time with cargo run –release is less than 1 minute with the dataset provided by Stanford.

```
(base) jhuang@crc-dot1x-nat-10-239-245-168 DS210_Final_Project % cargo run --release
    Finished release [optimized] target(s) in 0.06s
     Running `target/release/rust_facebook_project`
Generated PNG image: graph.png
Top 5 vertices:
3437: 0.002237822154724936
107: 0.002073339075639776
1684: 0.0018982251739692183
0: 0.001875875512840855
1912: 0.0011683251354353196
Average distance between pairs of vertices: 3.6925068496963913
(base) jhuang@crc-dot1x-nat-10-239-245-168 DS210_Final_Project %
```