Jenny Tang
Professor Kontothanassis
DS210
5 May 2024

DS210 Final Write-up

Data

The data was acquired from SNAP:
https://snap.stanford.edu/data/ego-Twitter.html

The data had 81306 nodes with 1768149 edges, so it was adequate for the project at hand. Furthermore, Stanford's write-up showed that it was useful in analyzing social circles and connectivity on Twitter, now known as X. This data was suitable for me because I wanted to analyze the interconnectedness of uses on Twitter (X).

Modules and Results

My graph.rs module utilized the parse_graph function to turn graph data from a string format into a HashMap representation. It also uses print_degree_centrality to print out the degree centrality values for the nodes to be put in the graph.

The bfs.rs module stands for breadth-first search, which is an algorithm I attempted to use to return the distance of nodes from the starting node in a HashMap. The algorithm was meant to find the shortest path from a source node on a graph in relation to other nodes.

The degree_centrality.rs module computes the degree centrality of nodes within Twitter (X), or the number of edges (connections, in this case), a node has. In the case of Twitter, the nodes are important based on the number of connections they have. The print_degree_centrality function returns the values to the terminal.

main.rs ensures the other modules were connected and ran together smoothly. I performed some operations on it relating to graph.rs and degree centrality. The results returned were as followed:

```
    Finished release [optimized] target(s) in 0.63s
     Running `target\release\tang_final.exe`
Node 7860742: Degree Centrality 5655
Node 22462180: Degree Centrality 8324
Node 40981798: Degree Centrality 10336
Node 34428380: Degree Centrality 8251
Node 43003845: Degree Centrality 9368
Node 3359851: Degree Centrality 7359
BFS result: {1: 0}
```

Results Analysis
The BFS result shows that node 1 is reachable from itself with a distance of 0.

My degree centrality values were returned as such:

Node 43003845: Degree Centrality 9368
Node 3359851: Degree Centrality 7359
Node 22462180: Degree Centrality 8324
Node 34428380: Degree Centrality 8251
Node 7860742: Degree Centrality 5655
Node 40981798: Degree Centrality 10336

The degree centrality of the nodes are high—in the thousands—showing that Twitter is highly interconnected as a social media platform. Thus, this proves that Twitter's interconnectedness creates a highly dynamic and influential network and online community.

The higher degree centrality mean that a user is connected to a more significant number of other users in the network. For example, though node 42002845 has a high degree centrality of 9368, user 40981798's centrality is higher, meaning that there is a higher level of influence with the latter, as there are larger connections involved. Overall, Twitter has a highly connected network.