Yaochen Zhang Final Project Report

Data name and resource:
Gnutella peer-to-peer network, August 8 2002
https://snap.stanford.edu/data/p2p-Gnutella08.html

Data Explanation:
Gnutella is a peer-to-peer network protocol. Founded in 2000, it was the first decentralized peer-to-peer network of its kind, leading to other, later networks adopting the model. The dataset is a sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002. There are total of 9 snapshots of Gnutella network collected in August 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts.

Problem and Solution:
The peer-to-peer network is a directed graph. In a directed graph, the edges all have directions. This means there is an edge existing between every node leading from one to another. One important problem I would like to discuss about in this project is about connectivity and six degrees of separation. When solving the problem like this, Breadth First Search and Depth First Search are the most useful tools. The DFS and BFS are two common graph traversal algorithms used to visit nodes in a graph in a specific order. DFS starts at a root node and visits its neighbors as far as possible along each branch before backtracking. It explores a branch of the tree as far as possible before backtracking to explore other branches.

In my program, I mainly use BFS to perform most of computations. Comparing with DFS, Breadth First Search uses a queue, it starts from the starting vertex then explores its neighbors and its neighbor's neighbor, and so on. In my coding, I use BFS to compute all the distance starting from the vertex and computes and average distance between each node. In the test part, I set up a small directed graph to test my function.

Conclusion:
The over all coding was really a touch work, and I tried to follow the goal of my project as much as possible, but there are always some tiny errors occur during the coding. The peer-to-peer network data set contains 6301 nodes with 20777 edges, has the average clustering coefficient of 0.0109, 2383 triangles and the shortest path is 9.