

Kiko Xu

## Final Project Report

Collaborator: Philip Hoang

This project is about finding the distance in terms of numbers of edges between each node. The data set regards the road networks in California. The nodes in the data set represent the intersections of roads while the edges represent the roads. This data set is unweighted and undirected.

For this project, I took a random sampling of 30,000 data points because the data set is too large so it takes too long to run the code. The code of this project focuses mainly on finding the shortest distance from node 0 to other nodes. Using breadth first search, I found the shortest distance because the data is unweighted and undirected. Another focus is the distribution average shortest path for each node. I also looked at different statistics like the average distance, maximum distance, median distance, and standard deviation of the data.

It outputs the average distance, maximum distance, median distance, standard deviation, distribution of average shortest path lengths for each node, and the shortest distance from node 0 to the other nodes. From the output, we could see that the nodes(intersections) are mostly just connected to each other one way because the shortest path is mostly just the number of the nodes. We could also see that the distribution average shortest path for each node range has a wide range from around 7,000 to 15,000.