Introduction: For this project, I analyzed the social circles dataset from Facebook, which was collected from survey participants using the Facebook app. The dataset includes over 4000 nodes and 88,000+ undirected edges between them. The nodes indicate a user's profile and the undirected edges indicate the friendship between two users.

In order to analyze these graphs, I employed two algorithms: Dijkstra's shortest path algorithm and the Breadth First Search algorithm. Further explanation of the implementation of the algorithms is contained down below.

Interesting findings: Upon implementing the functions for the algorithms and an evaluative function for six degrees of separation, I found that any node can traverse to another node within >6 trips. In the context of Facebook, this means that any user is connected to another user within 6 shared friendships (undirected edges). As a result, the notion of six degrees of separation is true for this dataset. Now, this is an older dataset, so we can't claim the same would always hold true in Facebook's current climate. But it's still neat to see!

How to use:

I separated the algorithms into separate modules. The main file already imports them in for usage. I already created the functions to translate the raw data and set them within the main file. All you need to do is switch the file path to your own. The dataset is provided in the repository, so if you clone the project, you should have it downloaded already.

The main function within the main file contains two commented-out blocks: one for the Dijkstra algorithm and one for the breadth-first algorithm. Both algorithms accept a starting node and a graph represented as a vector of vectors as input, and return a HashMap where the keys are the nodes and the values are the distances from the starting node to each node in the graph. This is essentially the pathing from one node to every other node in the dataset. The default is set to the 0th node, but you can change it to your custom. Additionally, both blocks of code contain the function for testing for six degrees of separation and will output true on whether or not the dataset upholds the theorem.