Short project description:
My project explored the Six Degrees of Separation Theory that any person on the planet can be connected to any other person through a chain of acquaintances that has no more than five intermediaries. My program computes the usual distance between pairs of vertices in two graphs using the **Breadth-First Search algorithm** and outputs the difference between these two distances.

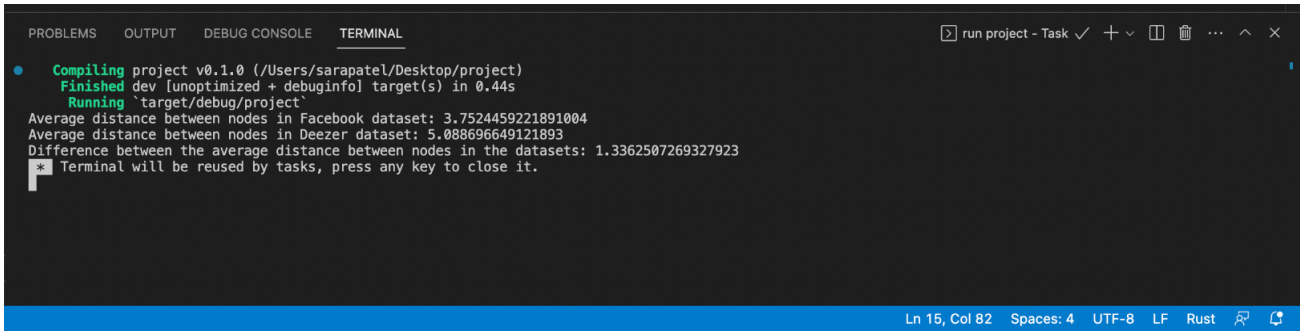
Datasets:

- fb_data.txt: This dataset consists of 'circles' (or 'friends lists') from Facebook. Facebook data was collected from survey participants using this Facebook app. The dataset includes node features (profiles), circles, and ego networks.
    - http://snap.stanford.edu/data/ego-Facebook.html
- deezer.csv: This data was collected from the music streaming service Deezer (November 2017). These datasets represent friendship networks of users from 3 European countries. I chose to specifically use data representing friendships among Romanian users.
    - https://www.kaggle.com/datasets/andreagarritano/deezer-social-networks?resource=download

Components of project:

- **read_txt** and **read_csv**: read edge files and organize them into hashmaps that include the key node followed by all the nodes it is connected to. These two functions are the exact same, but read_txt splits by whitespace while read_csv splits by commas.
- **bfs**: implements Breadth-First Search algorithm by visiting all the nodes in a graph in a breadth-first order. Starts at the given root node and visits all the nodes at the current level before moving onto the next.
- **average_distance**: for each pair of vertices, adds shortest distance between two given nodes to total_distance and adds 1 to num_pairs. Lastly, divides total_distance by num_pairs to compute the average distance between nodes for the dataset.
- **main**: prints the average distance between nodes for the Facebook dataset and the Deezer dataset. Lastly, it returns the absolute difference between the average distance between nodes for these two datasets.
- **tests**:
    - **test_read_txt_empty_file**: tests read_txt on an empty dataset to make sure length of graph output is 0.
    - **test_read_csv**: tests read_csv on a small dataset with only 2 edges. Checks length of graph as well as the friends of each node.
    - **test_bfs_shortest_path**: checks the bfs function's output on a a small dataset with only 2 edges.
    - **test_bfs_no_path**: checks the bfs function's output when given a nonexistent edge.

Output:



Conclusion:

The output did not reflect what I expected the average distance between nodes would be. Since these datasets are pretty small and fewer relationships are represented, I expected the average distance to be larger. Before this project, I was very unconvinced of the Six Degrees of Separation Theory, but upon looking at these outputs it seems very possible. With such widespread communication and travel in the modern world, it is probable that the theory is more factual than I previously thought.