

Final Project Report

Yunho Choi (U53197861), ynhchoi@bu.edu

- Includes:
 - General Description of the Project
 - What the project does
 - How to run it
 - What the output looks like
 - Analysis of the output
 - Conclusion

Six Degrees of Separation

Definition: The theory that any person on the planet can be connected to any other person on the planet through a chain of acquaintances that has no more than five intermediaries.

Definition implemented from <https://www.techtarget.com/whatis/definition/six-degrees-of-separation#:~:text=Six%20degrees%20of%20separation%20is,no%20more%20than%20five>

General Description of the Project

- This final project involves implementing the Breadth First Search (BFS) algorithm on a dataset called "gplus_combined.txt" to find the six degrees of separation between two people on Google+. The dataset contains the connections between different users on Google+.
 - Access to the dataset: <https://snap.stanford.edu/data/ego-Gplus.html>
- The **goal** of the project:
 - To implement the BFS algorithm to find the "six degrees of separation" between any two users in the dataset called "gplus_combined.txt."
 - To analyze the distribution of these degrees of separation across a range of pairs of users.
 - To explore the social network of users in the dataset and gain insights into the degree of separation between any two users using graph algorithms
- Note:
 - Degree 1 means that the two users are directly connected, while Degree 5 means that there are five intermediaries between the two users

What the Project Does

- The dataset("gplus_combined.txt") is parsed to generate a graph structure where each user is represented as a node, and the connections between users are represented as edges. The graph is then used as input to the BFS algorithm, which starts with a user (source node) and searches the graph for the destination user by exploring all of the users who are directly connected to the source user. If the destination user is not found, the algorithm moves on to explore the next level of users who are connected to the source user through mutual friends. This process continues until either the destination user is found or the maximum depth (6 degrees of separation) is reached.
- The output of the project is the shortest path (if exists) between the source and destination users. This project also provides the number of hops required to reach the destination user from the source user. If there is no path between the source and destination user within 6 degrees of separation or if they are non-reachable, it returns `None`, indicating that there is no path.

How to Run

- The number of pairs that can be chosen from the "gplus_combined.txt" can be changed in the main.rs module. In line 14 in the main.rs module, there is this a line of code: `let num_pairs = 100;`. The `100` can be modified to natural numbers (ranging from 1 to 15,000,000 in this case) when selecting the number of pairs to find six degrees of separation.
- After that, simply type `cargo run` or `cargo run --release` in the terminal to print out the value.
 - This might take up a lot of time depending on the given value in `num_pairs` variable.

What the output looks like

Output (2000 pairs)

```
Degrees of Separation:  
Degree 3: 1316 pairs  
Degree 5: 6 pairs  
Degree 1: 2 pairs  
Degree 4: 197 pairs  
Degree 2: 479 pairs
```

Output (1000 pairs)

```
Degrees of Separation:  
Degree 3: 677 pairs  
Degree 2: 210 pairs  
Degree 1: 5 pairs  
Degree 4: 105 pairs  
Degree 5: 3 pairs
```

Output (500 pairs)

```
Degrees of Separation:  
Degree 5: 1 pairs  
Degree 4: 43 pairs  
Degree 3: 341 pairs  
Degree 1: 2 pairs  
Degree 2: 113 pairs
```

Output (100 pairs)

```
Degrees of Separation:  
Degree 2: 21 pairs  
Degree 3: 71 pairs  
Degree 4: 8 pairs
```

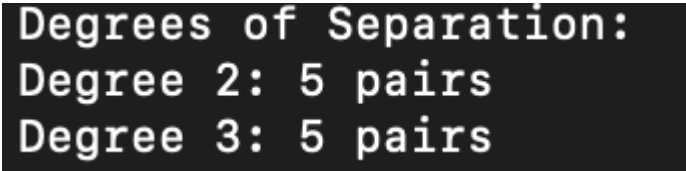
Output (50 pairs)

```
Degrees of Separation:  
Degree 2: 11 pairs  
Degree 4: 6 pairs  
Degree 3: 33 pairs
```

Output (30 pairs)

```
Degrees of Separation:  
Degree 3: 23 pairs  
Degree 2: 6 pairs  
Degree 4: 1 pairs
```

Output (10 pairs)



```
Degrees of Separation:  
Degree 2: 5 pairs  
Degree 3: 5 pairs
```

Analysis of the output

- Result of different number of pairs: 2000, 1000, 500, 100, 50, 30, and 10 (in the same order as the image)
- Looking at the results, it appears that the majority of the pairs, in all cases, have a degree of separation of 2 or 3, which is consistent with the popular notion of "six degrees of separation". The number of pairs with higher degrees of separation, 4 or 5, decreases as the number of pairs selected decreases.
 - It's interesting to observe that for some of the smaller numbers of pairs analyzed, for example 10 or 30, there were significantly less pairs with a degree of separation higher than 3. For example, from the output above, when 30 pairs are selected, there is only one pair that has degree of 4, and when 10 pairs are selected, there are no pairs that have higher degree than 3. This suggests that in a smaller sample size, the likelihood of finding individuals connected by more than three intermediaries is much lower.

Conclusion

- For all the pairs of users that were analyzed, there exists a path between them within six degrees of separation, which confirms the popular belief that any two people in the world are connected by at most six degrees of separation.
- The number of pairs with shorter degrees of separation is significantly higher than those with longer degrees of separation.
- For smaller samples (10 or 30 pairs), there seems to be a higher concentration of pairs with a lower degree of separation, possibly due to chance or the particular structure of the social network.
- For larger samples (1000 or 2000 pairs), the number of pairs with a higher degree of separation is more significant, indicating that there may be some users who are more isolated or have fewer connections to the rest of the network.
- Overall, this analysis provides evidence supporting the theory of six degrees of separation, using graph algorithms to study social networks.