

Benjamin Frechette
DS210 - Algorithms and Programming for Data Science
May 7, 2023

For this project, I decided to analyze the degrees of separation between users on Facebook. While I intended to work with a similar but larger dataset of Twitter users, configuring the algorithms to work with such a large dataset proved to be too difficult for skill level. I chose this dataset because I felt that mapping out human social circles was more interesting than analyzing any artificially generated data. What I found is that there is an average of 3 degrees of separation and a median of 4 degrees of separation between any two users, which is notably smaller than the six degrees commonly used as a benchmark. While I can only speculate, I feel this is likely because the interconnectedness of communities is significantly higher in online spaces than in real life, so navigating through social circles is much easier.

For my project, I utilized the breadth-first search algorithm discussed in lecture to determine the distance between every user. I modified the code provided in class to store every calculated distance into a single vector, which would then be run through two functions I created to calculate the average and median distance. Every function except for the `read_file` function are stored between two modules: `calcs` and `graphs`. The `graphs` module includes some altered functions provided in lecture that build the graph models. The `calcs` module contains the functions that carry out the breadth-first algorithm and calculate the median and average degrees of separation. The test functions included in my project are meant to test the functions of my own creation, as well as test the altered breadth-first algorithm to ensure correctness.

To run the project, type `cargo run facebook_combined.txt` into the terminal, which will output two lines telling you the average and median degrees of separation after about ten seconds of processing.