

CS 225 Project Goals

We have chosen to work with the College Message Temporal Network (from UC Irvine) dataset from the Stanford website. We chose this dataset because it would be very interesting to get insights on what a private social network may look like; how far one person's reach can go would be fascinating to see using our graph traversal algorithms. Since this is a very large dataset (.txt file), we will only commit a small portion of it to Git. This dataset can be found at the following link: <http://snap.stanford.edu/data/CollegeMsg.html>.

For our first graph traversal, we will implement a **Breadth First Search**. We will implement this in C++ using a queue as we learned in class; this should then output the data in a text file according to the traversal order.

We plan to implement our graph data structure as an adjacency matrix as described in the lecture videos; then, on this implementation of the data structure, we will go forward with the three graph traversal algorithms that are mentioned. We will save the adjacency matrix in a file, to ensure that we don't have to parse the data every time.

The other algorithms we plan to implement are **Dijkstra's Algorithm** and the **Strongly Connected Component algorithms**.

Dijkstra's algorithm would give us the shortest path between two nodes in the graph. In this case, this can be interpreted as the minimum number of intermediate messages to connect two users on the network.

The Strongly Connected Component algorithm will allow us to determine the number of strongly connected components in the graph. This can be interpreted as the number of groups of people who only communicate amongst themselves in this UCI social network (i.e. the number of "cliques").

Additionally, we will write unit tests in Catch. We will practice test-driven development, writing unit tests incrementally as we go to make debugging easier.