CS 225: Final Project Goals

aliceh4, joonhee4, rohans5, sumedhv2

- Dataset: CollegeMsg temporal network (http://snap.stanford.edu/data/CollegeMsg.html)
  - "This dataset is comprised of private messages sent on an online social network at the University of California, Irvine. Users could search the network for others and then initiate conversation based on profile information. An edge (u, v, t) means that user u sent a private message to user v at time t. The dataset here is derived from the one hosted by Tore Opsahl, but we have parsed it so that it can be loaded directly into SNAP as a temporal network."
  - We will use this dataset to look at how students at UCI interact with each other through this messaging service, implementing algorithms that can tell us unique information about the students.

- Traversal:
  - We will use a depth-first search (DFS) traversal to traverse our graph.
  - Will implement this recursively.

- Algorithms:
  - Kruskal's Algorithm (Covered)
    - We will use Kruskal's Algorithm to detect vertex relationships and build minimum spanning trees.
    - Kruskal's algorithm uses an undirected edge-weight graph. We can apply this to find the minimum number of friends that two users must interact with in order to become friends themselves.
  - PageRank (Complex/Uncovered)
    - We can use the PageRank algorithm to find the most "popular" people in this specific social network.
    - PageRank works by counting the number and quality of connections a webpage has to estimate how important the website is. We can apply this algorithm to our dataset/graph implementation to estimate the importance rankings of specific people.

- Additional Resources:
  - We will additionally be using the linear algebra C++ library Armadillo for computing the PageRank algorithm and ranking popularity of people on the social network.