Traversals: DFS

“DFS has many more applications. One natural application is **topological sorting** (i.e., resolving dependencies): the order in which a recursive DFS *finishes processing* the vertices of a (directed acyclic) graph of dependencies corresponds to a valid topological order.” - https://www.quora.com/What-are-the-advantages-of-using-BFS-over-DFS-or-using-DFS-over-BFS-What-are-the-applications-and-downsides-of-each

For our purpose, since we need to process a big quantity of dataset during our graph search, we lean towards using DFS traversal for its better ability to handle space complexity. We believe that since in a Memetracker search, we are getting a lot of children on our nodes, DFS search will work in our favor in tracing our data.

Algorithm:

1. We are leaning towards using Prim’s algorithm, for the sake of its ability to build the Minimum Spanning Tree from any vertex, and it runs faster in dense graphs, which is what we need while searching through our multi-edgy graphs in the Memetracker dataset. If we use Fibonacci heaps we can improve our time complexity as well (O(E + log V)).

https://www.geeksforgeeks.org/difference-between-prims-and-kruskals-algorithm-for-mst/

2. We are also looking at implementing the Iterative deepening depth-first search. We think that its characteristics, which is “optimal like breadth-first search, but uses much less memory; at each iteration, it visits the nodes in the search tree in the same order as depth-first search, but the cumulative order in which nodes are first visited is effectively breadth-first.” This will help us avoid memory errors and save runtime. Worst case O(b^d), similar to bfs traversal, where b is branching factor, and d is the depth of the shallowest solution
https://en.wikipedia.org/wiki/Iterative_deepening_depth-first_search

Dataset: From Stanford Large Network Dataset Collection

- Memetracker: http://snap.stanford.edu/data/memetracker9.html

“96 million memes from the Memetracker. Memetracker tracks the quotes and phrases that appear most frequently over time across this entire online news spectrum. This makes it possible to see how different stories compete for news and blog coverage each day, and how certain stories persist while others fade quickly.

Overall Memetracker tracks more than 17 million different phrases and about 54% of the total phrase/quote mentions appear on blos and 46% in news media.”