Gray Houston

ghousto

0026483532

Note: Throughout this analysis, the variable *n* will always refer to the number of nodes/vertices in the graph and the variable *m* will always refer to the number of edges in the graph.

Project 5 Analysis

1. The asymptotic performance of my implementation of the topological sort algorithm is O(*n* + *m*), where *n* is the number of nodes in the graph and *m* is the number of edges. In my implementation, the source nodes (i.e. nodes with in-degree of 0) are identified when the graph is read in from the file. Therefore, determining the source nodes will not contribute to the algorithm’s performance/running time.

Assuming that the source vertices have previously been identified and stored in a list, my implementation performs a BFS-esque traversal of the graph. When sources nodes are removed from the graph and added to the topological ordering, only that source node’s children (i.e. the vertices pointed toward by the source) are checked if they have become “relative” source vertices.

By only checking a source node’s children for new sources and only processing these new sources, each node is accessed and processed exactly once. This traversal of every node exactly once while creating the topological ordering produces an O(*n*) asymptotic performance.

In addition, the number of nodes traversed by looking at all of a source vertex’s children for all relative source nodes is equal to *m*. Proving that every edge is traversed exactly once is slightly non-trivial:

All source nodes only contain out-bound edges

When a source is removed from the graph, each out-edge to a child is removed

Right before a source is deleted, each out-edge is followed to its children

Therefore, all the out-edges of a deleted source node are traversed exactly once

All of nodes of the graph are eventually removed as “relative” source nodes

Therefore, the out-edges of all of the nodes are traversed exactly once

The total of all out-edges in the graph is *m*

Therefore, the total number of node access that occur by checking a source’s children for new sources vertices runs asymptotically in O(*m*) time.

Furthermore, because the O(*n*) traversal to create the topological ordering is not directly related to the O(*m*) traversal used to check a source vertex’s children, the performance of these two processes is summed instead of multiplied. Therefore, the overall asymptotic performance of the algorithm is O(*n* + *m*).

1. The asymptotic performance of my implementation of the longest path algorithm runs in O(*n* + *m*) time. My algorithm uses a topological ordering and the following analysis will assume that one has already been computed. If the ordering hasn’t been created, my algorithm makes one by calling the function described above in question 1. The creating of the topological ordering would incur an additional O(*n* + *m*) amount of computation, but does not alter the overall asymptotic performance of the algorithm.

My implementation works by traversing the nodes in topological order to relax the distances between nodes. This traversal of all *n* vertices runs in O(*n*) time asymptotically and, as I will prove below, the relaxation of all edges runs in O(*m*) time.

The key to the function of this longest path algorithm is the relaxation of the distances that measure how long it takes to reach an arbitrary node *v*. Because the edges are not weighted, these distances measure the maximum number of nodes traversed to reach node *v* from any of the source nodes. These distances are stored in an array where an array index corresponds to the node’s id.

When an arbitrary node *v* is reached in the topological order, each child of *v* is accessed to determine if the path to *v* is longer than any other path that has reached each child vertex so far. Thereby, all of the outbound edges of node *v* are traversed. In the same manner, all of the out-edges of each node in the digraph is traversed. Because all of the nodes only relax their out-edges once, O(*m*) node traversals and comparisons occur. Therefore, the relaxation of node distances asymptotically runs in O(*m*) time.

Because these two processes are computed independently, the overall performance is the sum the performance of the two tasks. Therefore, the asymptotic running time for the whole algorithm is O(*n* + *m*).