A close up of a logo

Description automatically generated

Analysing Properties of a Wikipedia Page Graph

By Clayton Herbst (22245091)

Graphs are data structures that consist of a finite set of vertices and a set of ordered or unordered edges that form the link between these vertices (Wikipedia 2019). A single Wikipedia page (the node) can be considered as a member of a set of Wikipedia pages (the graph) interconnected through hyperlinked URL’s (the directed edges) from which various characteristics and properties can be determined.

## Shortest Path Solution

In order to find the minimum number of URL links a user must follow to navigate from one Wikipedia page to another, a Breadth First Search (BFS) approach can be implemented starting at the specified starting vertex and ending at the destination vertex. The BFS algorithm starts at a vertex and construct a spanning tree for the given graph called the breadth-first tree (CITS2200 Topic 12) using a first-in-first-out queue abstract data structure. Modifying this common graph search technique allows the algorithm to keep track of the distance of each node within the breadth-first spanning tree from its root. Considering a Wikipedia graph, G, the pseudo code to find the shortest path between vertex *v* and vertex *y* is as follows:

*Procedure BFS(v, y) //v is the starting node, y is the destination node*

*Push v on to the tail of Q*

*dist[v] 🡨 0*

*while Q is not empty*

*Pop vertex w from the head of Q for each vertex x adjacent to w do*

*if colour[x] is white then dist[x] ← dist[w]+1*

*if x equals y then return dist[x]*

*end if*

*colour[x] ← grey*

*Push x on to the tail of Q*

*end if*

*end for*

*colour[w] ← black*

*end while*

The time complexity of this shortest path algorithm implementation is O(V + E) in the worst case where V is the number of vertices in the graph and E is the number of edges in the graph. This has been derived by summing the time complexity of enqueuing all vertices and examining all edges which are O(V) and O(E) respectively (CITS2200 Topic 12). The BFS provides optimal time complexity across an unweighted graph (Wikipedia 2019).

## Hamiltonian Path

## Strongly Connected Components

## Graph Centres

## References

Graph definition: <https://en.wikipedia.org/wiki/Graph_(abstract_data_type)>

Shortest Path: <https://en.wikipedia.org/wiki/Shortest_path_problem#Unweighted_graphs>