DC : Red Black Trees

Implementing RB Tree on large graphs

Input : California Road networks (given pair of nodes that connects each other)

California road network

Dataset information

A road network of California. Intersections and endpoints are represented by nodes and the roads connecting these intersections or road endpoints are represented by undirected edges.

Dataset statistics	
Nodes	1965206
Edges	2766607
Nodes in largest WCC	1957027 (0.996)
Edges in largest WCC	2760388 (0.998)
Nodes in largest SCC	1957027 (0.996)
Edges in largest SCC	2760388 (0.998)
Average clustering coefficient	0.0464
Number of triangles	120676
Fraction of closed triangles	0.02097
Diameter (longest shortest path)	849
90-percentile effective diameter	5e+02

Space Used :

1) RB node : key(string), rightchild, leftchild, parent, color, index(int)
2) Adjacency list (space : nodes + edges)
3) Vertices : It is an array which stores the string we input corresponding to its index.(so that access can be in constant time)

Time Taken : (averaged)

1) Single insert after 1900000 insertion = 0.0067 ms
2) Single search atter 1900000 insertion = 0.0047 ms
3) Total time taken (insertion + Searching for all nodes) = 8.132 s

Procedure :

1) We are taking input and instantiating a RB node for every unique vertex in the graph.
2) As we insert new nodes we are assigning each node with an index value (which is initial 0 and incremented as we take input) and storing its key to an array of vertices at this assigned index.
3) Also simultaneously we are making an adjacency list of data given.

How it works : Let's say we have a starting point ABC and ending point XYZ.

1) We will search ABC and XYZ in the RB tree and get its index values.
2) After we get 2 indexes we have an adjacency list by which we can find paths by DFS or BFS.
3) If we want to print a path we can get it from the vertices array as it stores string corresponding to index.

Benefits :

1) While inserting strings checking will be faster if it is already inserted or not.
2) Searching for Starting and ending points will be faster.