**Project report**

Group members: Umar Mansoor 19i-0594 Shoaib Asghar 19i-0406 Saad Sagheer 19i-04966

**Introduction:**

In the CA-GrQc.txt dataset there are 5242 nodes and 28980 edges. By reading file, we noted total number of nodes and edges. Firstly, we have to take graph as undirected as given in the dataset, e.g., node 10310 is directing towards 3466 and 3466 is also pointing towards 103100. After that, we have drawn directed graph. The reason for creating directed graph is that we have to find in degree, out degree, source node, isolated node and sink node; and the reason for making undirected graph is that we have to find articulation node and bridge edge etc. We read file only once throughout the project. We also make graph for CA-AstroPh.txt dataset for bonus purpose.

**Working:**

We used adjacency list to make graph. We read two nodes from file and make an edge by making source node as first node as head of link list and store destination in its linked list. We wrote indegree and outdegree algorithms, from which we found out whether the node is isolated, source, sink and we can also find out indegree and out degree distribution. To find bridge edge and articulation point we use undirected graph. For bridge edge, we removed an edge from both nodes and apply DFS algorithm on the graph to check whether it is a bridge edge or not. For articulation node, we removed a node and removed all its edges and then apply DFS algorithm on the graph to check whether it is an articulation node or not. We also make SSC algorithm.

**Problems Faced:**

* Since the data provided was very large, so it was difficult for us to understand the given data as it was very time consuming to process data for making graph.
* It was difficult to understand if the graph we have to make is undirected or directed, but after analyzing graph and file we concluded that it is undirected graph.
* Since number of nodes are very large, we cannot print them on console.
* For shortest path length distribution and diameter, we faced difficulties to implement them as graph was not weighted.
* For dataset CA-AstroPh.txt file, it took more than 10 minutes on our devices only to make a graph and print total number of nodes and edges, so it is very hard to test the given algorithms in that file.