National University
Of Computer and Emerging Sciences

**CS2001-Data Structures**

# Project Report: Report for dataset of ca-GrQc.txt

**Submitted to: Sir Hassan Raza**


**Submitted by:**

**Talha Nadeem (i21-0313)**

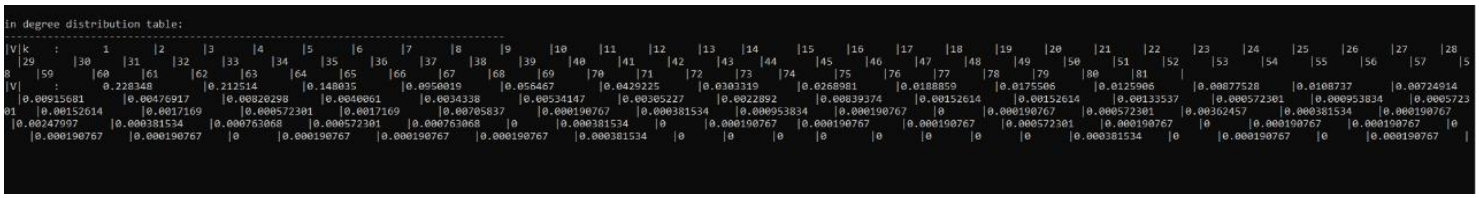
**Hamzah Ahmed (i21-0254)**


**Semester: 3rd**

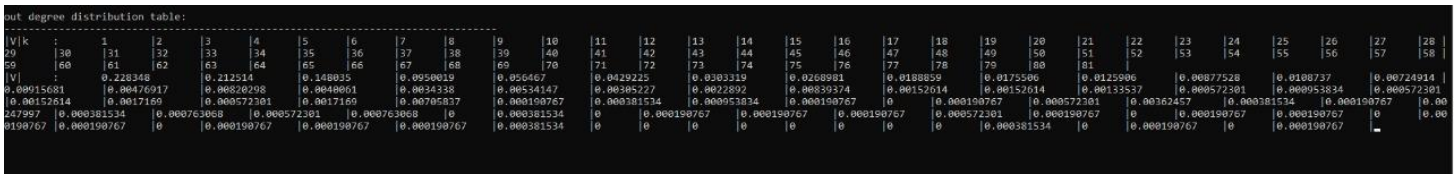

**Department: Department of DS & AI.**

# Functions

- The number of nodes in the data set are 5242
- The total number of edges in the graph are 28980
- The total no. of source nodes are 0 since the graph contains an edge forming from every node in the graph
- The total no. of sink nodes are 0 since the graph contains an edge ending at every node
- The total number of isolated nodes are 0 since there is no node with 0 in and out degree
- The total number of bridge edges is 1197. Which means there are 1197 points which if broken, the subgraph would be disconnected.
- The total number of articulation nodes could not be found out since the graph given to us was a directed one and articulation nodes exist in undirected graph only.
- The shortest path length distribution could not be reported as the function takes too much time to execute, however, it can be displayed in the demo
- The diameter of the graph could not be reported as the function takes too much time to execute, however it can be displayed in the demo
- Following is the in-degree distribution of the dataset:




- Following is the out degree distribution of the graph




- The SCC could also not be displayed since the function takes way too much time to execute

We have a total of 5242 authors in the dataset since there are 0 since nodes and there are 5242 unique IDs on the left side of our dataset(source side). It was observed that there were loops in the dataset which means that if author A co-authored a book with author B then author B also co-authored a book with author A. Each publication can have multiple authors. There is an author which is pointing to many nodes which means that there are books with one author and many coauthors. Each index of the array of linked lists contains more than 1 node in it which means there are many co-authors to a single book. The connection of an author to its multiple co-authors also means that there are multiple subgraphs in our dataset. It is to be noted that since there are 0 sink nodes. Then all of the coauthors are also authors of their own book.