

INFORMATION RETRIEVAL

ASSIGNMENT 5

Submitted by –

Vaibhav Varshney(MT17065)

Dataset

- The graph network taken is “Collaboration network of Arxiv General Relativity”.
- It is undirected graph originally consisting of 5242 nodes and 14, 496 edges.

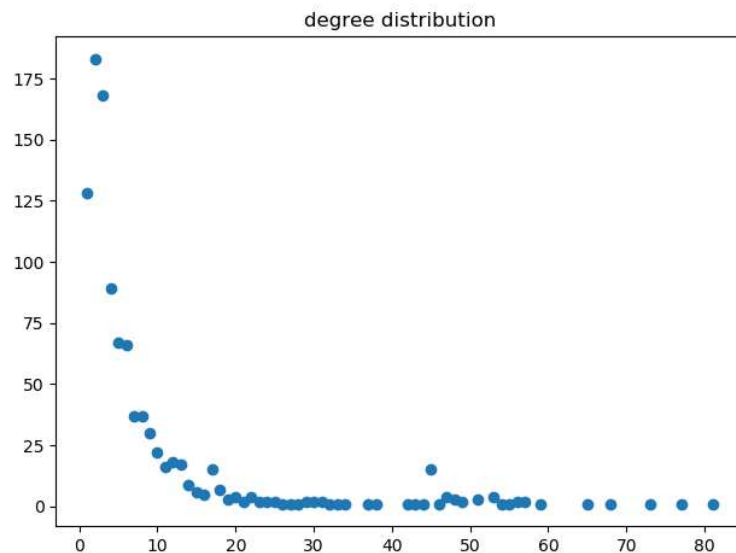
Methodology

- The graph is converted into an adjacency matrix with 1000 nodes only.

Analysis

- **Degree Distribution-** In the study of graphs and networks, the degree of a node in a network is the number of connections it has to other nodes and the degree distribution is the probability distribution of these degrees over the whole network.

The degree distribution obtained in this graph is as shown below:



- **Cluster Coefficient-** The local clustering coefficient of a vertex (node) in a graph quantifies how close its neighbours are to being a clique (complete graph).

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}.$$

Average clustering coefficient in this graph is: 0.529

- **Closeness centrality-** Closeness Centrality indicates how close a node is to all other nodes in the network. Calculated as the average of the shortest path length from the node to every other node in the network.
- **Betweenness centrality-**

Observations and Inferences

The following facts have been observed by seeing the results of the experiment:

- It has been observed that the average degree is around 30.
- Average clustering coefficient is 0.539 over the 1000 nodes.
- The closeness centrality is around 19 for most of the nodes but for some exceptional cases it is around 980.
- Betweenness centrality is around 0.01 average.
