ASSIGNMENT 1 : Web and Social Computing [IT752]

By

Prajwal M P

192508IT015

Purpose

To analyse a given graph dataset and find its properties.
Nodes : The number of vertices in the graph. We analyze different size graphs based on number of nodes ranging from few hundred to 200k+

Edges: The number of connections between the edges representing some relationship between them.

Average degree : It gives the mean number of edges per node in a graph. A higher average degree could mean better connectivity.

Density : This gives how well connected the nodes. A high density means the number of edges are closer to maximum possible number of edges. A low score could mean the graph is sparse.

Average Clustering: It gives the mean average clustering coefficient. A higher value means the graph is well connected and has clusters with high number of connections.

Number of connected components: A connected component is one which there is a path between every pair of nodes.

Diameter: It is the longest shortest path in a graph. It gives the longest shortest distance between any two nodes in the graph.

DATASET #1

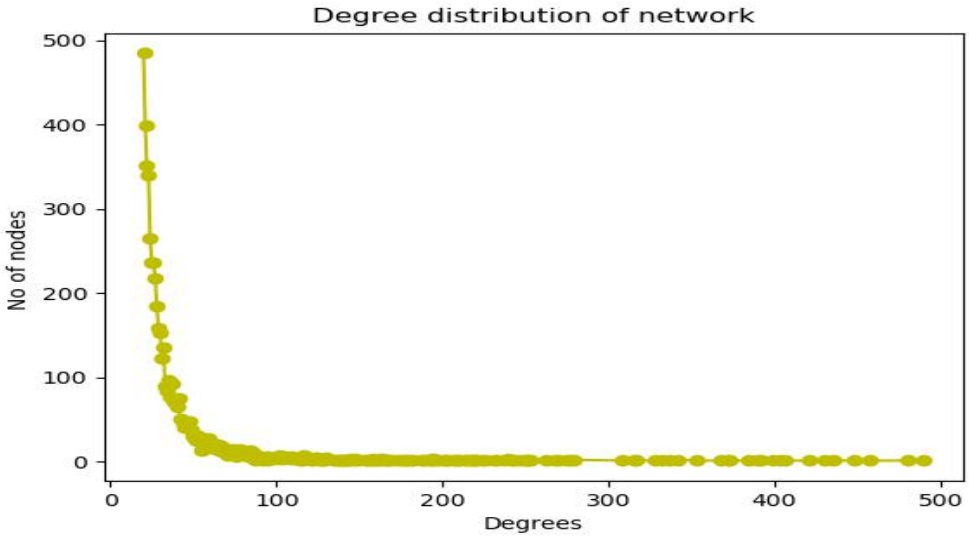
Procedure

Download the “General Relativity and Quantum Cosmology collaboration network” from https://snap.stanford.edu/data/ca-GrQc.html
Use networkx in python to analyse the dataset

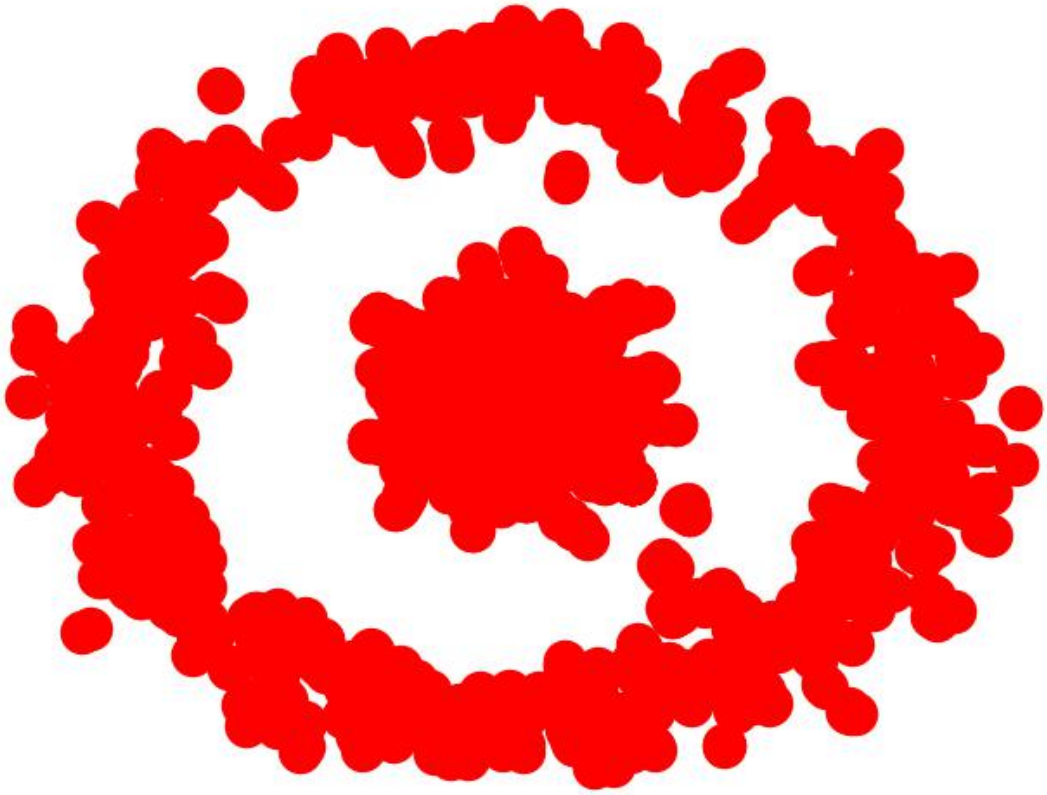
Findings

The following are the properties of the CA-Grq dataset
Nodes : 5242
Edges : 14496
Average degree : 5.5307
Density : 0.00105
Average clustering : 0.5296
Number of connected components : 355
Diameter : 17

Graph Degree Distribution



Graph plot



If ER model with same parameters were created then

Average Shortest path : 3.82068247881
Average Clustering Co-efficient : 0.00217847756537

If WS model with same parameters were created then
Average Shortest path : 4.07752997617
Average Clustering Co-efficient : 0.0431872224837

If AB model with same parameters were created then
Average Shortest path : 2.62709409514
Average Clustering Co-efficient : 0.0285012164725

DATASET #2

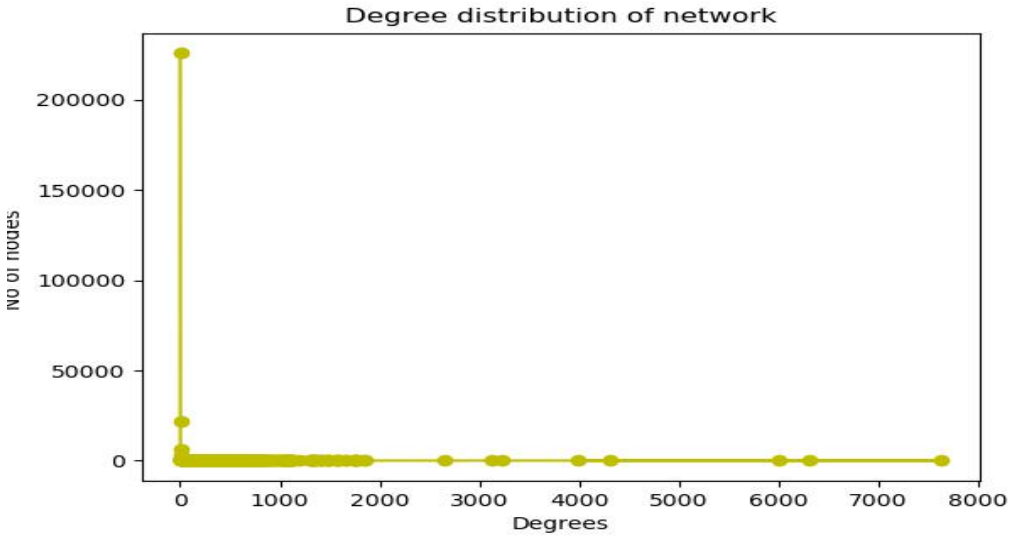
Procedure

Download the “EU email communication network” from https://snap.stanford.edu/data/email-EuAll.html
Use networkx in python to analyse the dataset

Findings

The following are the properties of the dataset
Nodes : 265214
Edges : 420045
Average degree : 2.7568
Density : 0.000001039
Average clustering : 0.06708495948
Number of connected components : 15836
Diameter : 14

Graph Degree Distribution



DATASET #3

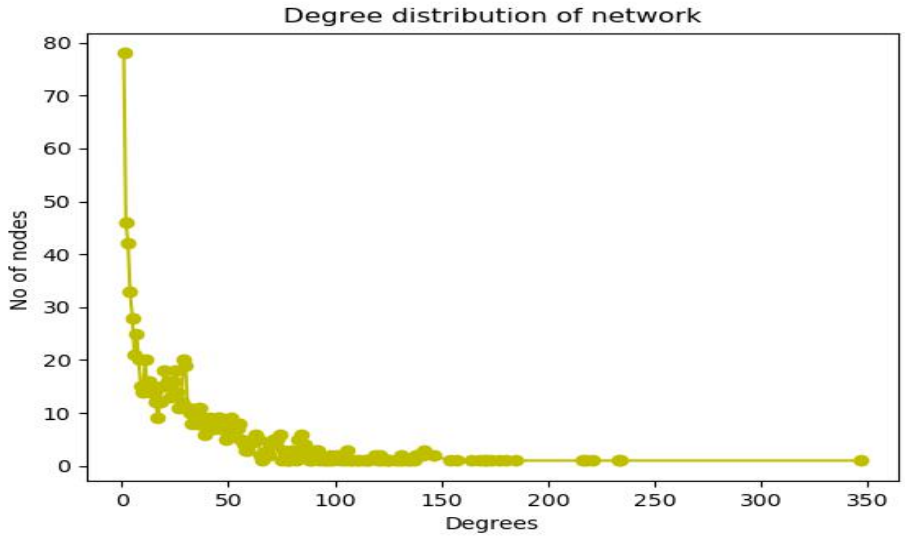
Procedure

Download the “email-Eu-core network” from
https://snap.stanford.edu/data/email-Eu-core.html
Use networkx in python to analyse the dataset

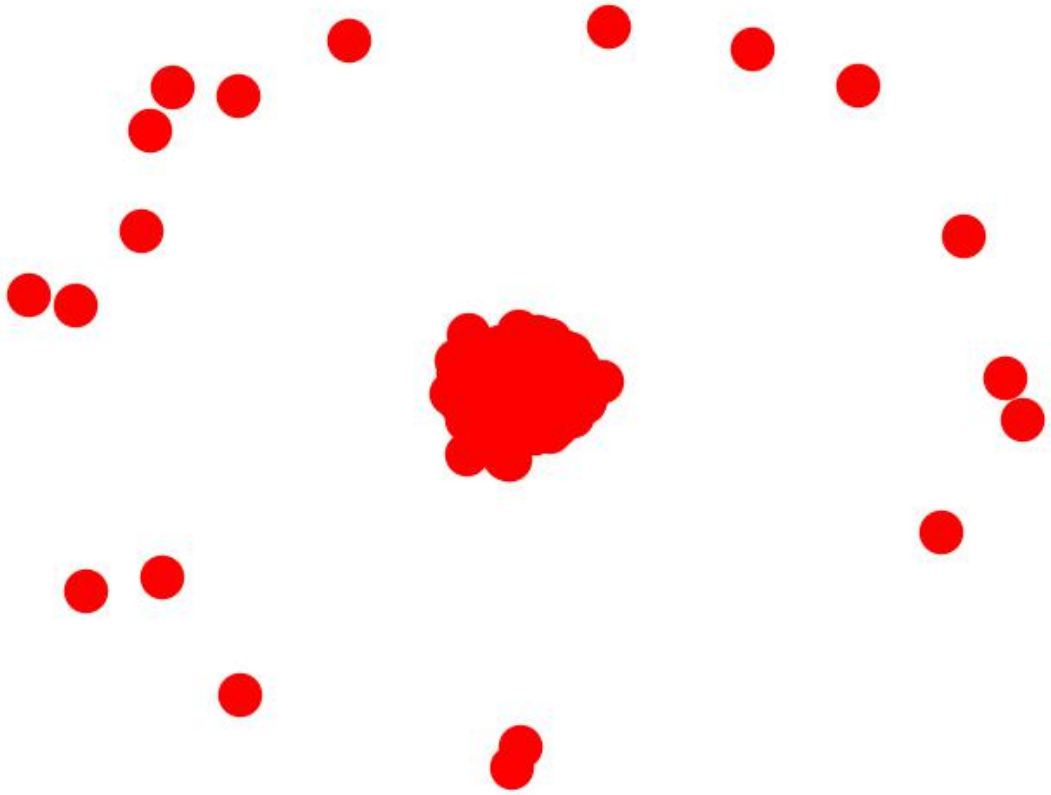
Findings

The following are the properties of the dataset
Nodes : 1005
Edges : 25571
Average degree : 33.2458
Density : 0.03311331
Average clustering : 0.3994
Number of connected components : 20
Diameter : 7

Graph Degree Distribution



Graph plot



If ER model with same parameters were created then

Average Shortest path : 2.16604031635
Average Clustering Co-efficient : 0.096767068534

If WS model with same parameters were created then
Average Shortest path : 2.28928465244
Average Clustering Co-efficient : 0.0331287533442

If AB model with same parameters were created then
Average Shortest path : 3.32316505124
Average Clustering Co-efficient : 0.0468789653425

DATASET #4

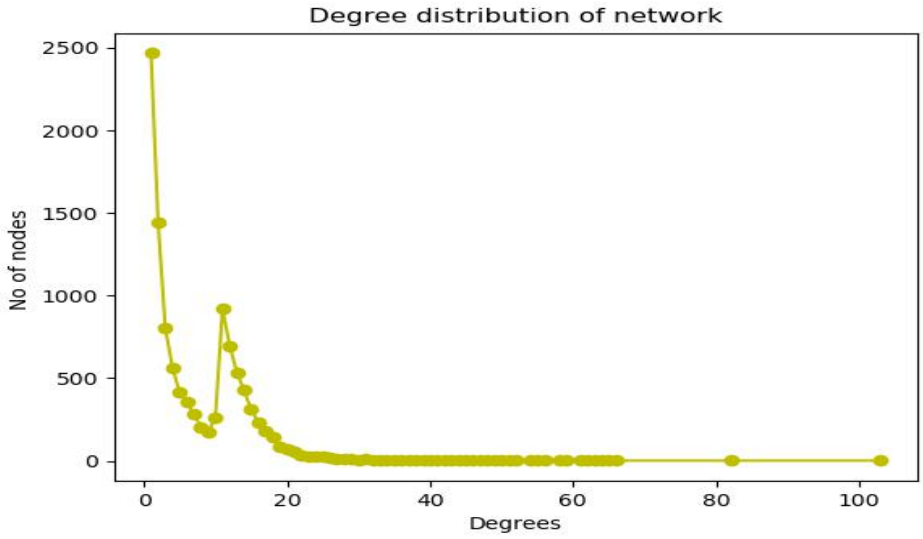
Procedure

Download the “Gnutella peer-to-peer network” from https://snap.stanford.edu/data/p2p-Gnutella04.html
Use networkx in python to analyse the dataset

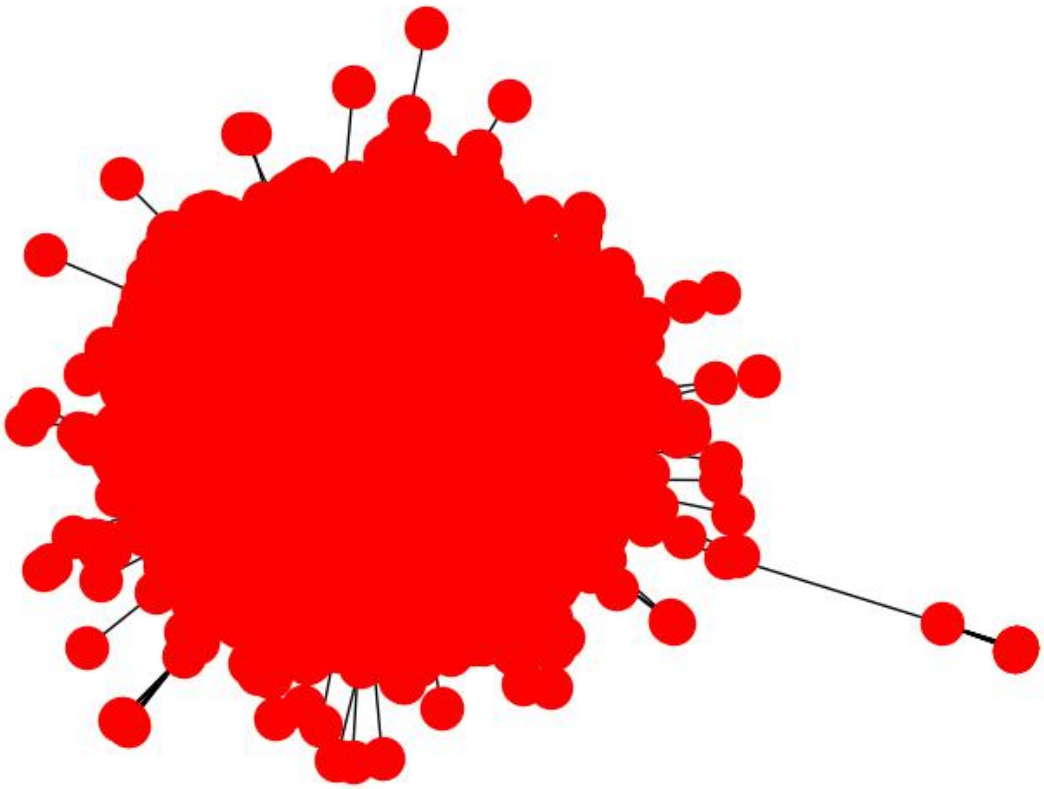
Findings

The following are the properties of the dataset
Nodes : 10876
Edges : 39994
Average degree : 33.2458
Density : 0.03311331
Average clustering : 0.0062
Number of connected components : 20
Diameter : 9

Graph Degree Distribution



Graph plot



If ER model with same parameters were created then

Average Shortest path : 1.504135
Average Clustering Co-efficient : 0.0068534

If WS model with same parameters were created then
Average Shortest path : 2.026544
Average Clustering Co-efficient : 0.012342

If AB model with same parameters were created then
Average Shortest path : 1.230124
Average Clustering Co-efficient : 0.008345

Conclusion:

Average Degree is low in 1-3 dataset but dataset 4 has a high average degree.
Average clustering is higher is real dataset than generated ones.
Dataset 3 only has 1 component. Other datasets have multiple components.
Diameter is very very low compared to Number of nodes in all datasets.