Web and Social Computing (IT752)
Lab Assignment 1

Submitted to : Dr. Sowmya Kamath
Submitted by : Shaikh Sahil Ahmed
Roll number: 192IT022

Dataset 1: wiki-Vote
Dataset 2: p2p-Gnutella04
Dataset 3: p2p-Gnutella06

The above 3 datasets were used and I concluded various factors like:

1. Degree Distribution
2. Diameter
3. Geodesic path length
4. Clustering Coefficient
5. Strongly Connected Components
6. Sparseness
7. k-connectedness
8. SCC properties

1. Degree Distribution

Finding degree distribution for all the 3 datasets are plotted as shown:

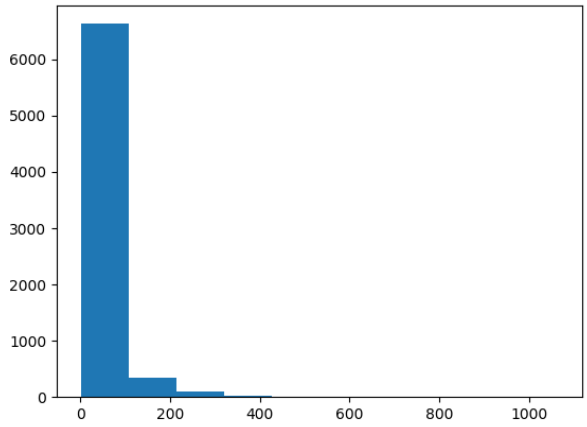


Fig 1. Histogram plot for dataset 1

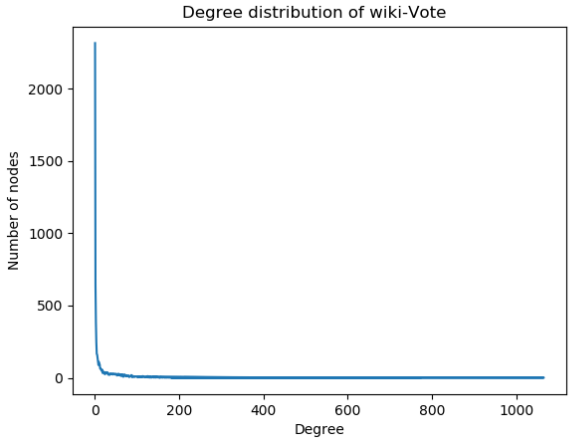
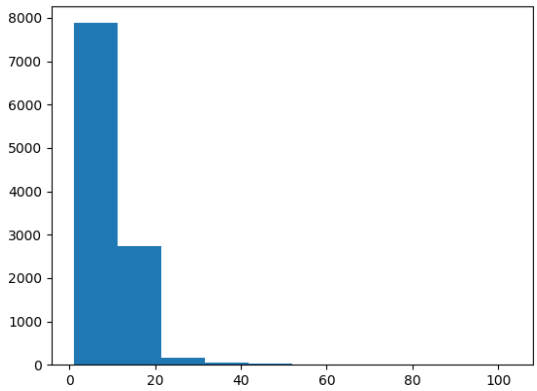
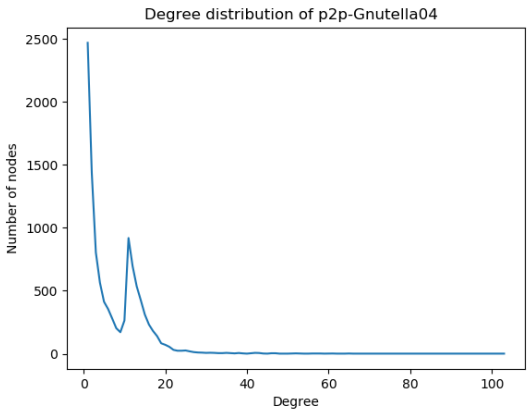


Fig 2. Degree distribution plot for dataset 1





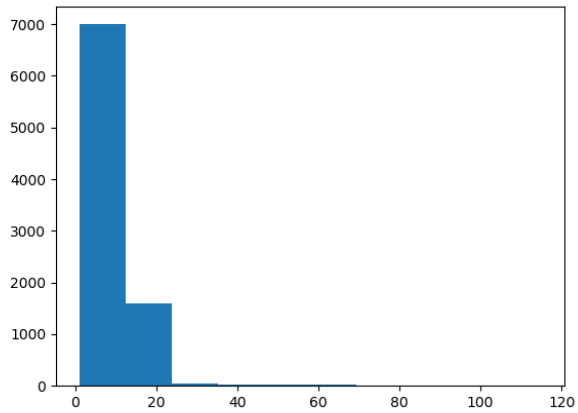


Fig 3. Histogram plot for dataset 1

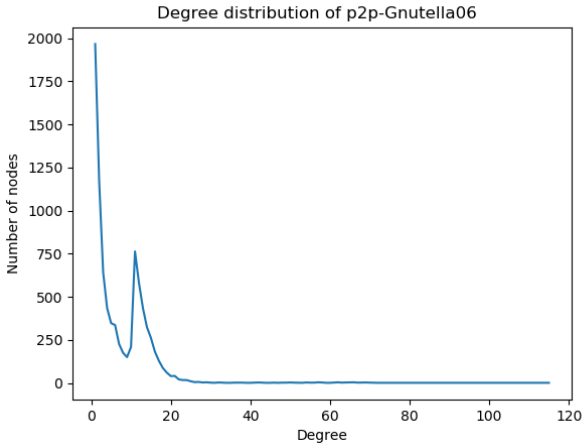


Fig 4. Degree distribution plot for dataset 1

Fig 5. Histogram plot for dataset 2

Fig 6. Degree distribution plot for dataset 2

As compared to dataset 1 and dataset 2, we can see from the graph that dataset 3 is lighter than others as it is having very fewer nodes with a very high degree and most of the are having a very little degree.

2. Diameter

The maximum eccentricity from all the vertices is considered as the ***diameter*** of the ***Graph*** G or it can be defined as the maximum shortest path between any pair of nodes. For dataset 1 and dataset2 , I got an infinite diameter, and for dataset 3 I got 10. As the path length between some nodes of dataset 1 and 2 may be infinite, that is there is no path from some vertex to another.

3. Geodesic path length

It is defined as the number of edges along the shortest path connecting a given pair of nodes. As there are various edges between different nodes, geodesic path length can be calculated for different vertices. I have randomly taken 2 nodes to find the geodesic path length. For dataset 1, source node id is 7092 and destination node id is 3 and geodesic path length between them is 3 and the average geodesic path length for the whole graph is 2.17391. For dataset 2, source node id is 20 and destination node id is 796 and geodesic path length between them is 4 and the average geodesic path length for the whole graph is 2.28615. For dataset 3, source node id is 7092 and destination node id is 3 and geodesic path length between them is 2 and the average geodesic path length for the whole graph is 2.301994.

4. Clustering Coefficient

It is a measure of how connected node i's neighbors are connected to each other. Clustering coefficient is found for various node. I have taken few outputs for various nodes.

<u>For dataset 1:</u>
('30', 0.15079365079365079)
('1412', 0.04926108374384237)
('3352', 0.11341014393178656)
('5254', 0.08517596871667259)
('5543', 0.14312251473021217)

<u>For dataset 2:</u>
('789', 0.032679738562091505)
('790', 0.025)
('794', 0.001282051282051282)
('1271', 0)
('1272', 0.027777777777777776)

<u>For dataset 3:</u>

('246', 0.012800819252432157)
('563', 0.010047593865679535)
('567', 0.0024489795918367346)
('636', 0.0024198427102238356)
('638', 0.0008163265306122449)
Average clustering coefficient for dataset 1 is 0.14089, for dataset 2 is 0.006216 and for dataset 3 is 0.006676.

5. Strongly connected components

A graph is said to be strongly connected or diconnected if every vertex is reachable from every other vertex. For dataset 1, number of strongly connected components are 5816, for dataset 2 is 6560 and for dataset 3 is 5492.

6. Sparseness

A graph with only a few edges is a sparse graph. By finding the edge density of the graph we can how much the graph is dense. Edge density lies between 0 and 1, going more towards 0 is sparse and going more towards 1 is a dense graph. Edge density of dataset 1 graph is 0.0039814, for dataset 2 is 0.0067604 and dataset 3 is 0.00082985 Here, dataset 3 graph is more sparse as compared to dataset 1 and dataset 2.

7. Finding K for K-Connectedness graph

A K-component is a maximal subgraph of a graph G that has, at least, node connectivity K, we need to remove at least K nodes to break it into more components. For dataset 1 value of K is 24 and for dataset 2 and dataset 3 is 1. As dataset is having only 1 component, the diameter was 10 and for other dataset 1 and 2, the graph is diconnected due to which the value of diameter is showing infinite.

8. SCC properties

A directed graph is strongly connected if there is a path between all pairs of vertices. A strongly connected component (**SCC**) of a directed graph is a maximal strongly connected subgraph.

There are more strongly connected components in dataset 2 as compared to dataset 1 and 3.