Web and Social Computing
ASSIGNMENT 3

Shrinidhi M
192IT024

There are 3 datasets used in this assignment to check the properties with different sizes of networks. They are:

1. **p2p-Gnutella08**

   ```
   Type: DiGraph
   Number of nodes: 6301
   Number of edges: 20777
   Average in degree:   3.2974
   Average out degree:   3.2974
   ```

2. **p2p-Gnutella09**

   ```
   Type: DiGraph
   Number of nodes: 8114
   Number of edges: 26013
   Average in degree:   3.2059
   Average out degree:   3.2059
   ```

3. **p2p-Gnutella04**

   ```
   Type: DiGraph
   Number of nodes: 10876
   Number of edges: 39994
   Average in degree:   3.6773
   Average out degree:   3.6773
   ```

DataSet1: p2p-Gnutella08

```
import networkx as nx
import matplotlib.pyplot as plt
G1 = nx.read_edgelist("/content/p2p-Gnutella08.txt",create_using=nx.DiGraph(), nodetype = int)
print(nx.info(G1))
```

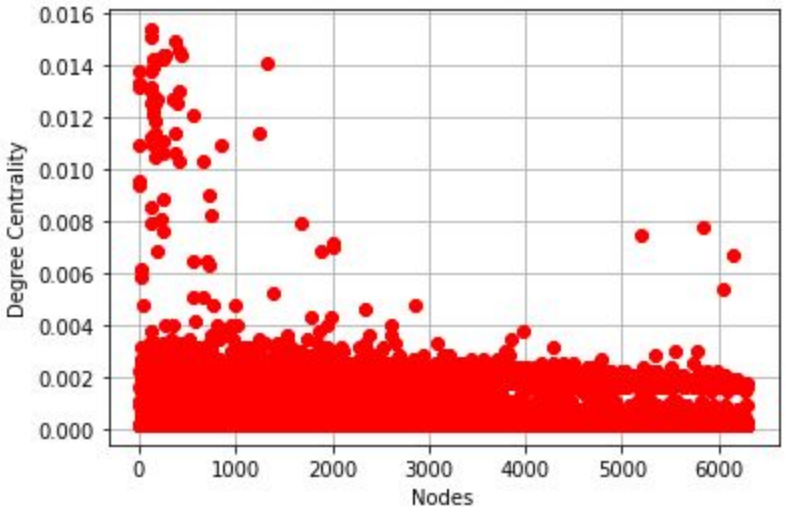
```
Name:
Type: DiGraph
Number of nodes: 6301
Number of edges: 20777
Average in degree:   3.2974
Average out degree:   3.2974
```

1.Degree Centrality

The degree can be interpreted in terms of the immediate risk of a node for catching whatever is flowing through the network (such as a virus, or some information). In the case of a directed network (where ties have direction), we usually define two separate measures of degree centrality, namely indegree and outdegree. Accordingly, indegree is a count of the number of ties directed to the node and outdegree is the number of ties that the node directs to others. When ties are associated to some positive aspects such as friendship or collaboration, indegree is often interpreted as a form of popularity, and outdegree as gregariousness.

Minimum degree centrality node and value 1 3.771165667307765e-05
Maximum degree centrality node and value 68 0.013387638118942564
Nodes in Decreasing order of Degree Centrality(top 5 nodes):

[(68, 0.013387638118942564),
(8537, 0.003130067503865445),
(13901, 0.0027152392804615905),
(6508, 0.0025266809970962024),
(5523, 0.0019610061470000375)]

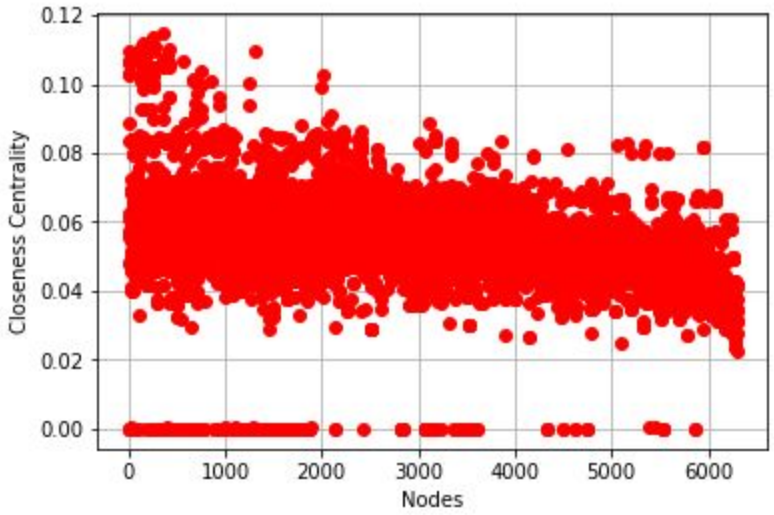


2. Closeness Centrality

In a connected graph,closeness centrality (or closeness) of a node is a measure of centrality in a network, calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus the more central a node is, the closer it is to all other nodes.

Minimum Closeness centrality node and value 41 0.0
Maximum Closeness centrality node and value 68 0.07772797401843072
Nodes in Decreasing order of Closeness Centrality(top 5 nodes):

[(68, 0.07772797401843072),
(280, 0.04921174442232865),
(642, 0.048466210542942985),
(735, 0.04743061775824379),
(1654, 0.047314767685390335)]

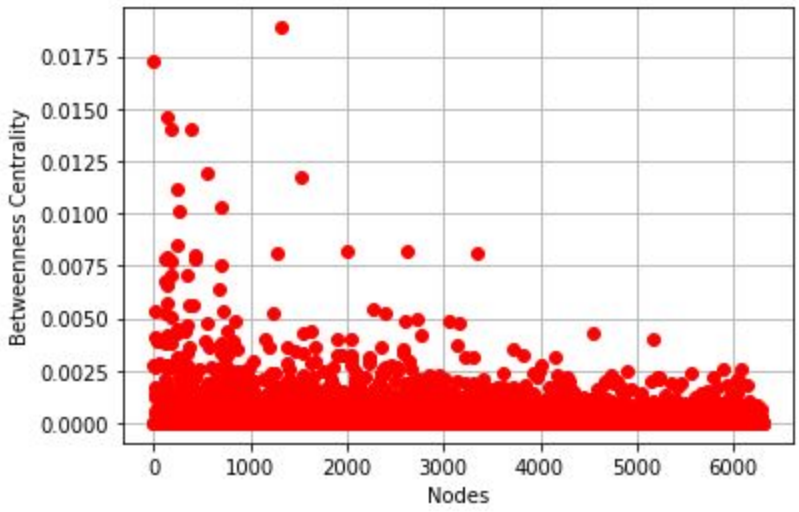


3. Betweenness Centrality

In graph theory, betweenness centrality is a measure of centrality in a graph based on shortest paths. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex. Betweenness centrality finds wide application in network theory: it represents the degree of which nodes stand between each other. For example, in a telecommunications network, a node with higher betweenness centrality would have more control over the network, because more information will pass through that node. Betweenness centrality was devised as a general measure of centrality: it applies to a wide range of problems in network theory, including problems related to social networks, biology, transport and scientific cooperation.

Minimum Betweenness centrality node and value 0 0.0
Maximum Betweenness centrality node and value 1317 0.0188687055523367
Nodes in Decreasing order of Betweenness Centrality(top 5 nodes):

[(1317, 0.0188687055523367),
(3, 0.017235200252786236),
(146, 0.0146110184536896),
(390, 0.014068027594482927),
(175, 0.01400021903697455)]



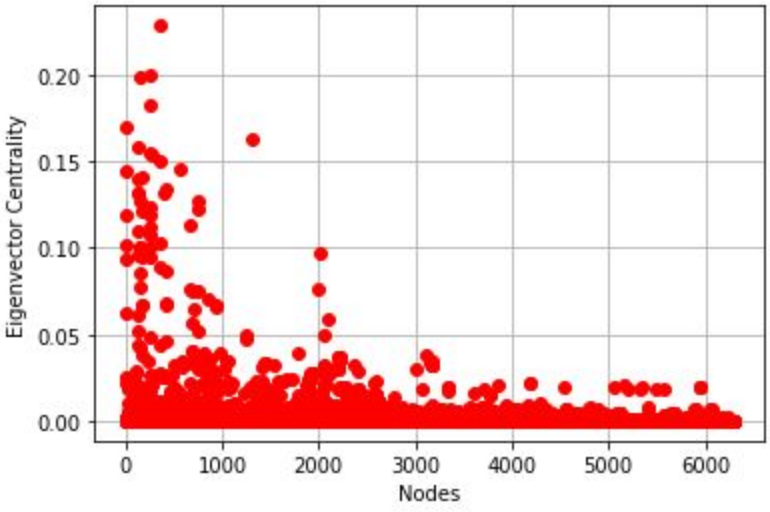
4. Eigen Vector Centrality

In graph theory, eigenvector centrality (also called eigencentrality) is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

Minimum degree centrality node and value 0 5.316146072083429e-14
Maximum degree centrality node and value 367 0.22871171476594773
Nodes in Decreasing order of Eigenvector Centrality(top 5 nodes):

[(367, 0.22871171476594773),
(249, 0.19997338445402307),
(145, 0.19842978241076356),

(264, 0.18287133350231022),
(5, 0.1700107854403149)]



2. Implementation of HITS & Page Rank Algorithm:

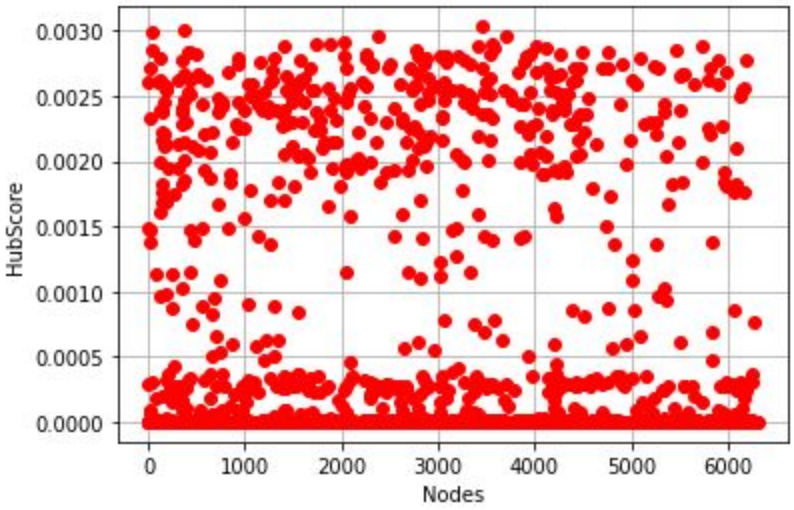
a. HITS algorithm:

Algorithm is a Link Analysis Algorithm that rates webpages, developed by Jon Kleinberg. This algorithm is used to the web link-structures to discover and rank the webpages relevant for a particular search.
HITS uses hubs and authorities to define a recursive relationship between webpages. Before understanding the HITS Algorithm, we first need to know about Hubs and Authorities.

- Given a query to a Search Engine, the set of highly relevant web pages are called roots. They are potential authorities..
- Pages which are not very relevant but point to pages in the Root are called hubs. Thus, an Authority is a page that many hubs link to whereas a Hub is a page that links to many authorities.

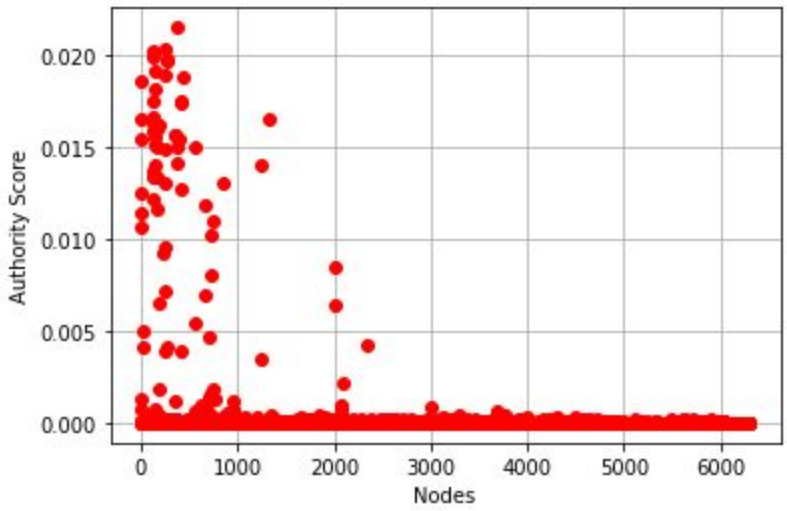
Maximum hub score webpage and value 3459 0.003032395968743096
Minimum hub score webpage and value 1 0.0
Nodes in Decreasing order of hub score (top 5 nodes):

[(3459, 0.003032395968743096), (366, 0.0030048537286680 94), (36, 0.002992944003658231), (2374, 0.0029659158718999402), (3693, 0.002954220175645511)]



Maximum authority score webpage and value 367 0.02146884302565421
Minimum authority score webpage and value 0 0.0
Nodes in Decreasing order of autority score (top 5 nodes):

[(367, 0.02146884302565421),
(249, 0.020290000606315882),
(123, 0.020245125623950845),
(127, 0.019874039870167143),
(266, 0.019718056113887456)]

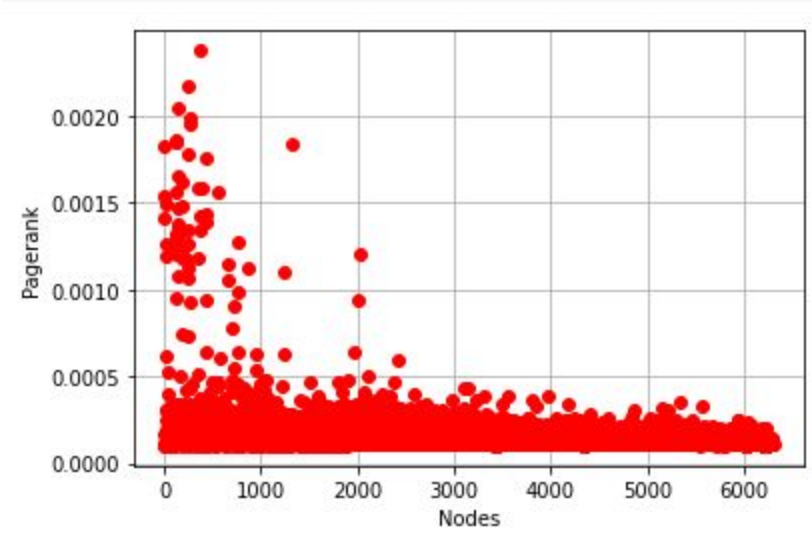


b. Page Rank Algorithm:

PageRank (PR) is an algorithm used by Google Search to rank websites in their search engine results. PageRank was named after Larry Page, one of the founders of Google. PageRank is a way of measuring the importance of website pages.

Maximum page rank webpage 367 0.002378560316719018
Minimum page rank webpage 0 0.00010061505853980176
Nodes in Decreasing order of Page Rank (top 5 nodes):

[(367, 0.002378560316719018),
(249, 0.002177444615709691),
(145, 0.0020465276122979643),
(264, 0.001990378022436873),
(266, 0.001957552521809092)]


0.0020
0.0015
0.0010
0.0005
0.0000
Pagerank
0
1000
2000
3000
4000
5000
6000
Nodes

DataSet2: `p2p-Gnutella09`

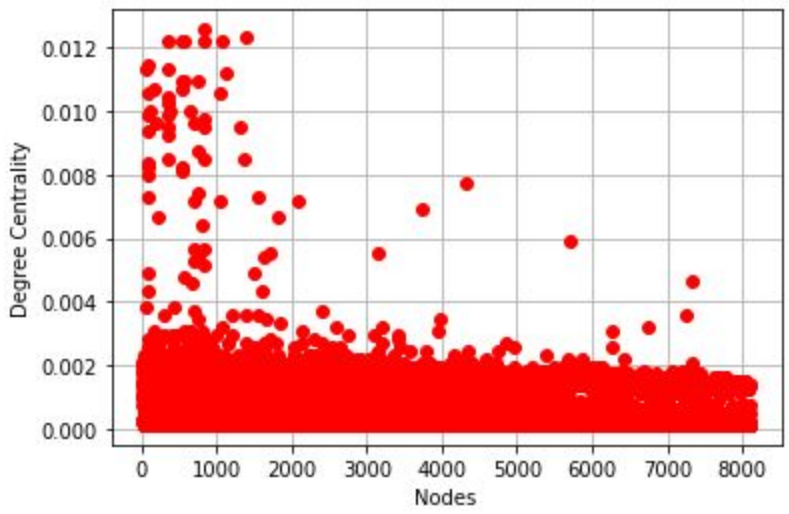
```
import networkx as nx
import matplotlib.pyplot as plt
G1 = nx.read_edgelist("/content/p2p-Gnutella09.txt",create_using=nx.DiGraph(), nodetype = int)
print(nx.info(G1))
```

```
Name:
Type: DiGraph
Number of nodes: 8114
Number of edges: 26013
Average in degree:   3.2059
Average out degree:   3.2059
```

1.Degree Centrality

Minimum degree centrality node and value 18 0.0001232589670898558
Maximum degree centrality node and value 822 0.01257241464316529
Nodes in Decreasing order of Degree Centrality(top 5 nodes):

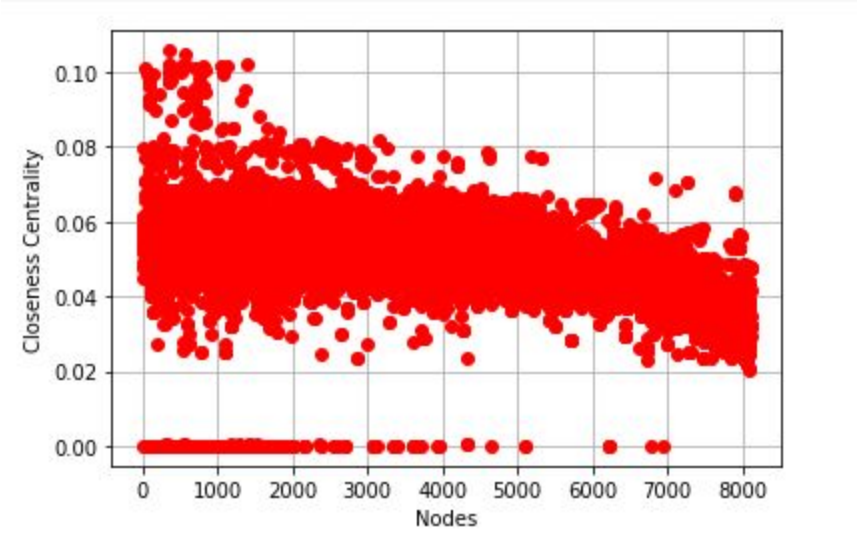
[(822, 0.01257241464316529),
(1389, 0.01232589670898558),
(530, 0.012202637741895724),
(356, 0.012202637741895724),
(825, 0.012202637741895724)]



2. Closeness Centrality

Minimum Closeness centrality node and value 64 0.0
Maximum Closeness centrality node and value 351 0.10568044890887182
Modes in Decreasing order of Closeness Centrality(top 5 nodes):

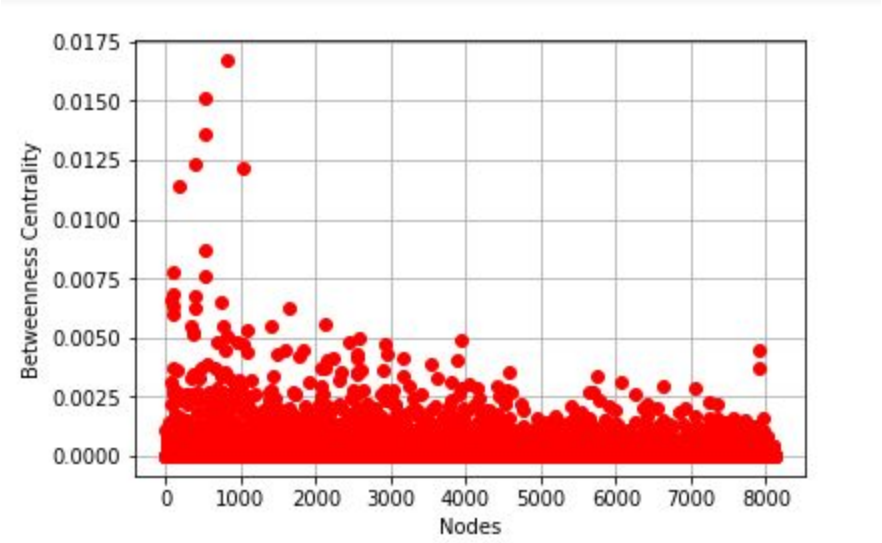
[(351, 0.10568044890887182),
(563, 0.1048886069285748),
(352, 0.10214511283176766),
(1389, 0.10198979238779443),
(565, 0.1019554831954684)]



3. Betweeness Centrality

Minimum Betweenness centrality node and value 1 0.0
Maximum Betweenness centrality node and value 822 0.016700241990001318
Nodes in Decreasing order of Betweenness Centrality(top 5 nodes):

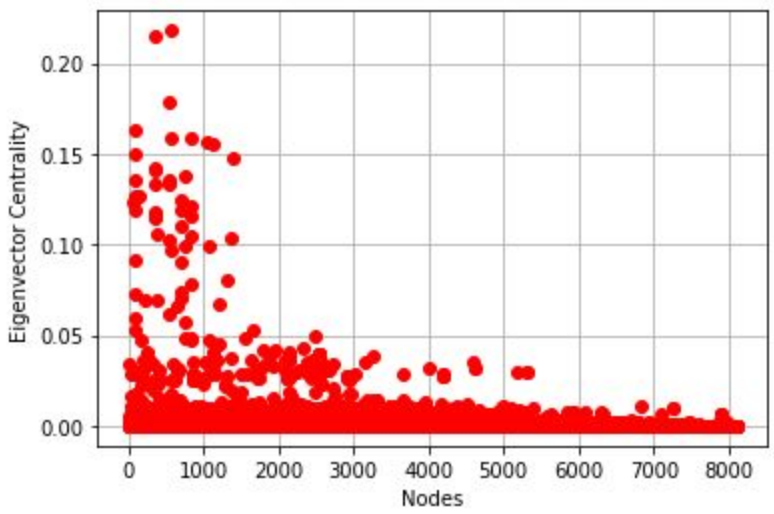
[(822, 0.016700241990001318),
(531, 0.01511095588624004),
(530, 0.013577129309631569),
(387, 0.012336390890243382),
(1041, 0.012171428963987313)]



4. Eigen Vector Centrality

Minimum degree centrality node and value 64 5.528405829828154e-15
Maximum degree centrality node and value 563 0.21803452971513723
Nodes in Decreasing order of Eigenvector Centrality(top 5 nodes):

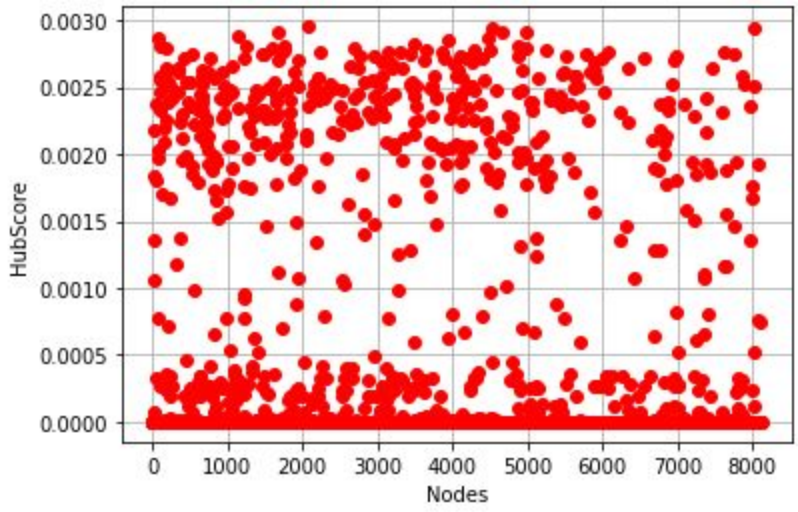
[(563, 0.21803452971513723),
(351, 0.21500951781680866),
(534, 0.17867464605959194),
(88, 0.16314940818894527),
(836, 0.15905019237713142)]



2. Implementation of HITS & Page Rank Algorithm:

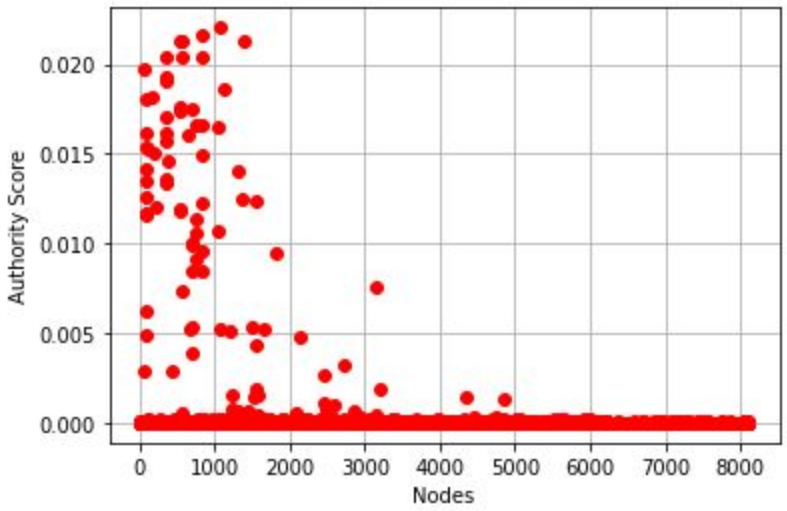
a.HITS algorithm:

Maximum hub score webpage and value 2070 0.0029530804907342205
Minimum hub score webpage and value 1 0.0
Nodes in Decreasing order of hub score (top 5 nodes):
[(2070, 0.0029530804907342205), (8024, 0.0029434588003124243), (4515, 0.0029368301681521112), (4987, 0.002914633014596839), (1664, 0.0029116898541687464)]



Maximum authority score webpage and value 1074 0.022047866014719305
Minimum authority score webpage and value 64 0.0
Nodes in Decreasing order of autority score (top 5 nodes):

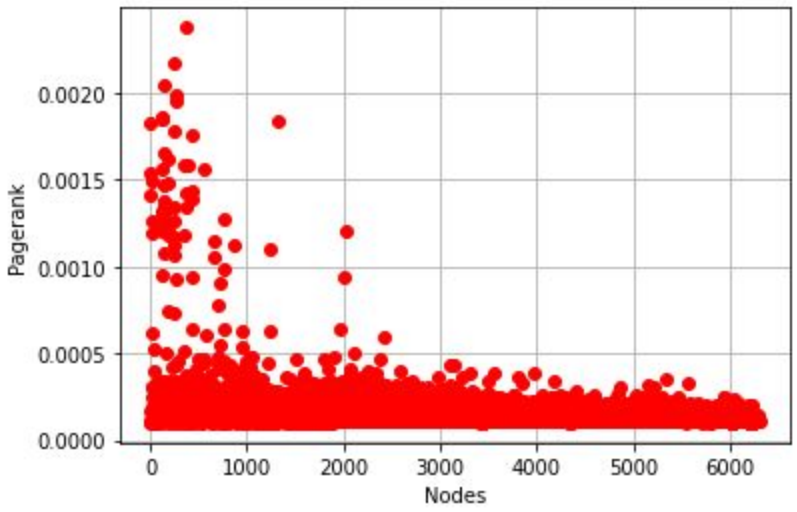
[(1074, 0.022047866014719305),
(822, 0.021585438545584965),
(1389, 0.021336664235616764),
(530, 0.021276874200390546),
(565, 0.021266873357450262)]



b.Page Rank Algorithm:

Maximum page rank webpage 351 0.0015871762721460221
Minimum page rank webpage 64 8.078747851616502e-05
Nodes in Decreasing order of Page Rank (top 5 nodes):

[(351, 0.0015871762721460221),
(563, 0.001546015088697242),
(822, 0.0014157202073277666),
(534, 0.0013874770361764072),
(565, 0.0013770841876928335)]


0.0020
0.0015
0.0010
0.0005
0.0000
Pagerank
0
1000
2000
3000
4000
5000
6000
Nodes

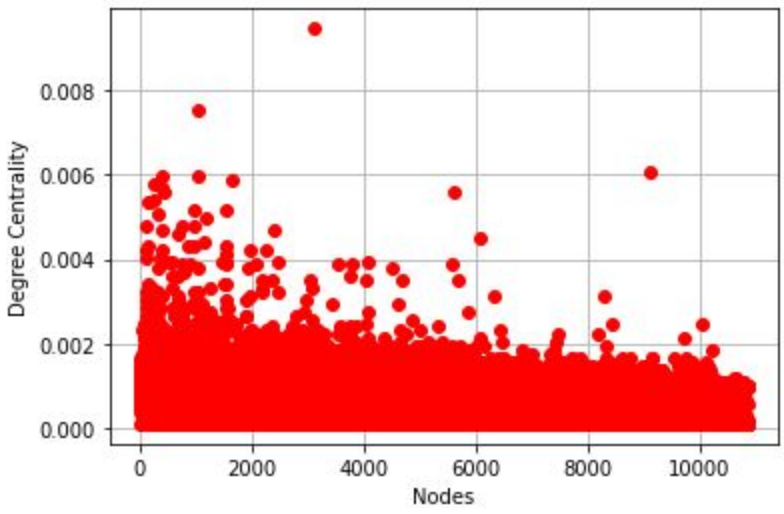
DataSet3: `p2p-Gnutella04`

```
import networkx as nx
import matplotlib.pyplot as plt
G1 = nx.read_edgelist("/content/p2p-Gnutella04.txt",create_using=nx.DiGraph(), nodetype = int)
print(nx.info(G1))
```

```
Name:
Type: DiGraph
Number of nodes: 10876
Number of edges: 39994
Average in degree:   3.6773
Average out degree:   3.6773
```

1.Degree Centrality

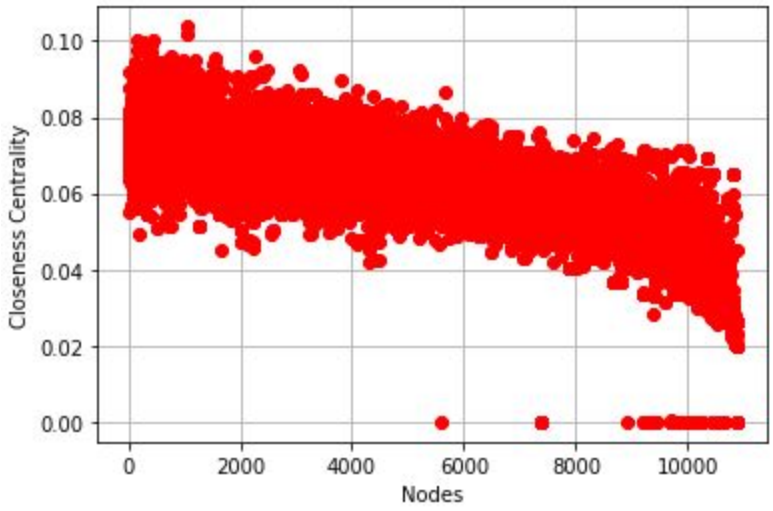
Minimum degree centrality node and value 24 9.195402298850575e-05
Maximum degree centrality node and value 3109 0.009471264367816092
Nodes in Decreasing order of Degree Centrality(top 5 nodes):
[(3109, 0.009471264367816092),
 (1054, 0.007540229885057472),
 (9134, 0.006068965517241379),
 (407, 0.005977011494252874),
 (1056, 0.005977011494252874)]



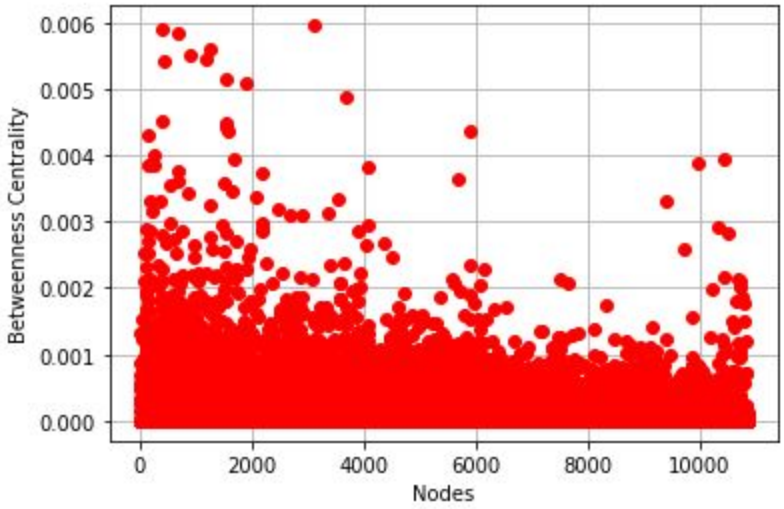
2. Closeness Centrality

Minimum Closeness centrality node and value 5586 0.0
Maximum Closeness centrality node and value 1056 0.10380262056360066
Nodes in Decreasing order of Closeness Centrality(top 5 nodes):

[(1056, 0.10380262056360066),
 (1054, 0.10198606198101959),
 (453, 0.10036900899181406),
 (171, 0.10025340313027084),
 (263, 0.0996736382452919)]



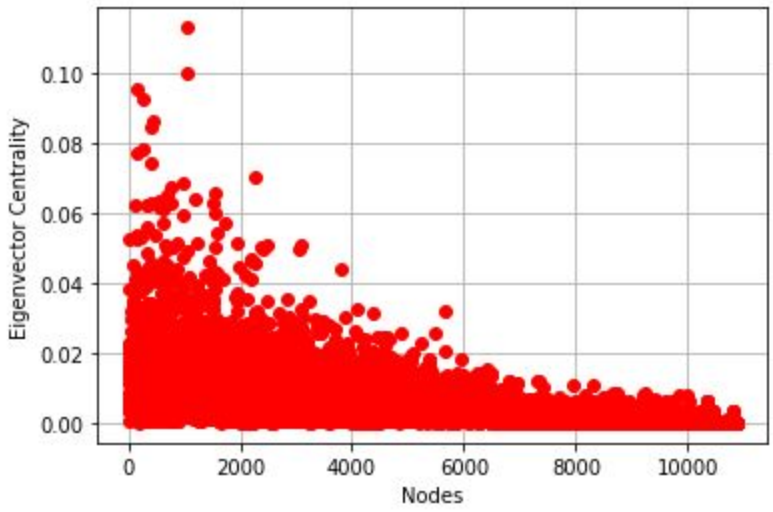
3. Betweeness Centrality



4. Eigen Vector Centrality

Minimum degree centrality node and value 5586 1.0765174705404389e-14
Maximum degree centrality node and value 1056 0.11290585493190843
Nodes in Decreasing order of Eigenvector Centrality(top 5 nodes):

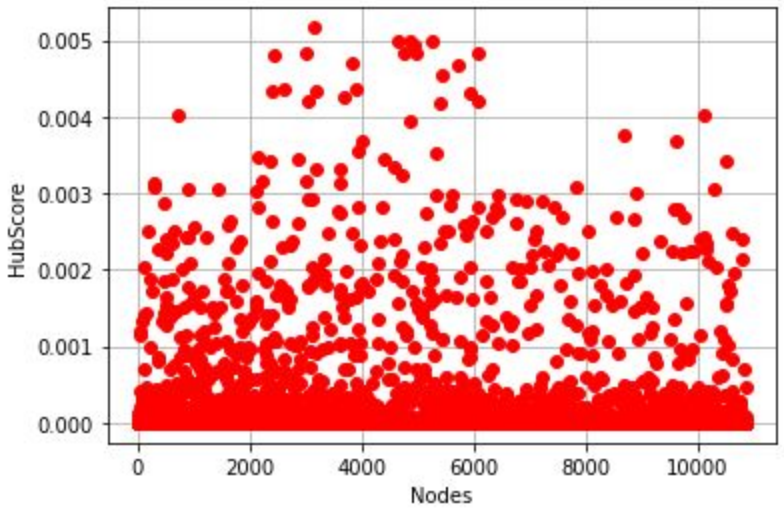
[(1056, 0.11290585493190843),
(1054, 0.099946897044748),
(171, 0.09509198621223416),
(263, 0.0924503773764383),
(453, 0.08599512470188617)]



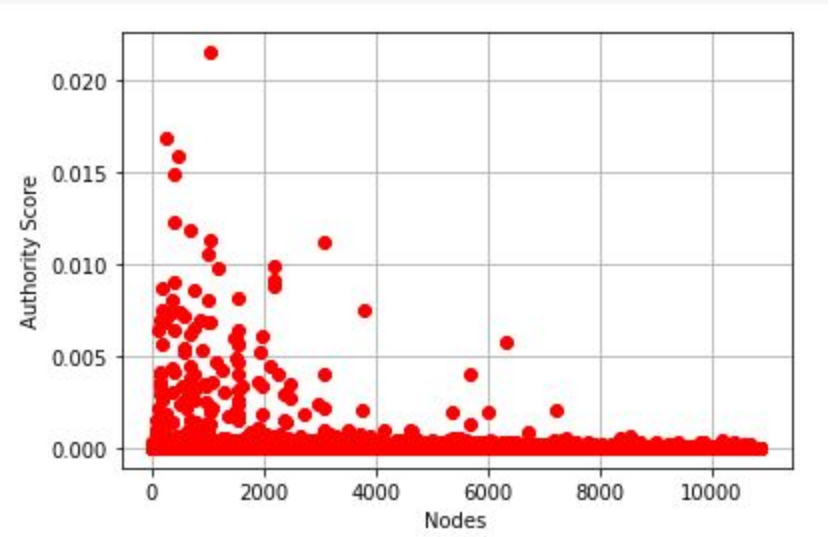
2. Implementation of HITS & Page Rank Algorithm:

a.HITS algorithm:

Maximum hub score webpage and value 3154 0.005167046979475104
Minimum hub score webpage and value 2 0.0
Nodes in Decreasing order of hub score (top 5 nodes):
[(3154, 0.005167046979475104), (4645, 0.0049902914760886615), (4866, 0.0049902914760886615), (5256, 0.0049902914760886615), (4942, 0.004944090430236253)]



Maximum authority score webpage and value 1054 0.021553778629464077
Minimum authority score webpage and value 5586 0.0
Nodes in Decreasing order of autority score (top 5 nodes):



b.Page Rank Algorithm:

Maximum page rank webpage 1056 0.0006711727183638689
Minimum page rank webpage 5586 5.499573860784478e-05

Nodes in Decreasing order of Page Rank (top 5 nodes):