INFORMATION RETRIEVAL
ASSIGNMENT 3 - REPORT

GROUP MEMBERS:-

ANURAG GAUTAM MT22015

MUSKAAN GUPTA MT22113

SALONI GARG MT22063

Ques1 Briefly describe the dataset chosen and report the following:

Ans1 a) Dataset Name :- **email-Eu-core network**

Network representation in the form of adjacency matrix.

For an adjacency matrix, if there is an edge 30->142 then adj_matrix[30[142]=1 otherwise it will be 0.

Adjacency matrix is shown below and first considers all the nodes from 0 to the maximum element present in the network.

Adjacency Matrix for the dataset

```
for row in adj_matrix:
    print(row)
```

```
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0
[0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0
[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0
[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1
[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0
[0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0
```

Edge list for the network

Edge List

```
[ ] print(edge_list)

    [[0, 1], [2, 3], [2, 4], [5, 6], [5, 7], [8, 9], [10, 11], [12, 13], [12, 14], [15, 16], [17, 18], [12, 19], [20, 21], [20, 22], [23, 24],
```

Number of Nodes in the network

```
[ ] #number of nodes
    for i in list1:
      if i not in list2:
        list2.append(i)
    n=max(list2)
    node=len(list2)
    print("Number of node",node)

    Number of node 1005
```

Number of edges

Number of edges

```
# number of edges
print("number of edges=",len(edge_list))

number of edges= 25571
```

Average In degree

```
print("Avg In -degree", avg_In_degree)

Avg In -degree 25.443781094527363
```

Average Out Degree

```
print("Avg Out-degree", avg_out_degree)

Avg Out-degree 25.443781094527363
```

The average in degree and out degree for the network is the same as if the network is directed and the nodes with indegree will get balanced with nodes with outdegree.

Node with maximum In degree

```
print("Node with Max in-degree", max_in_degree_node)

Node with Max in-degree 160
```

Node with maximum Out degree

```
print("Node with Max Out-degree", max_out_degree_node)
```

```
Node with Max Out-degree 160
```

Density of the network

```
print("The density of the network", density)
```

```
The density of the network 0.025342411448732432
```

Network density tell that density=0 , graph has no edges

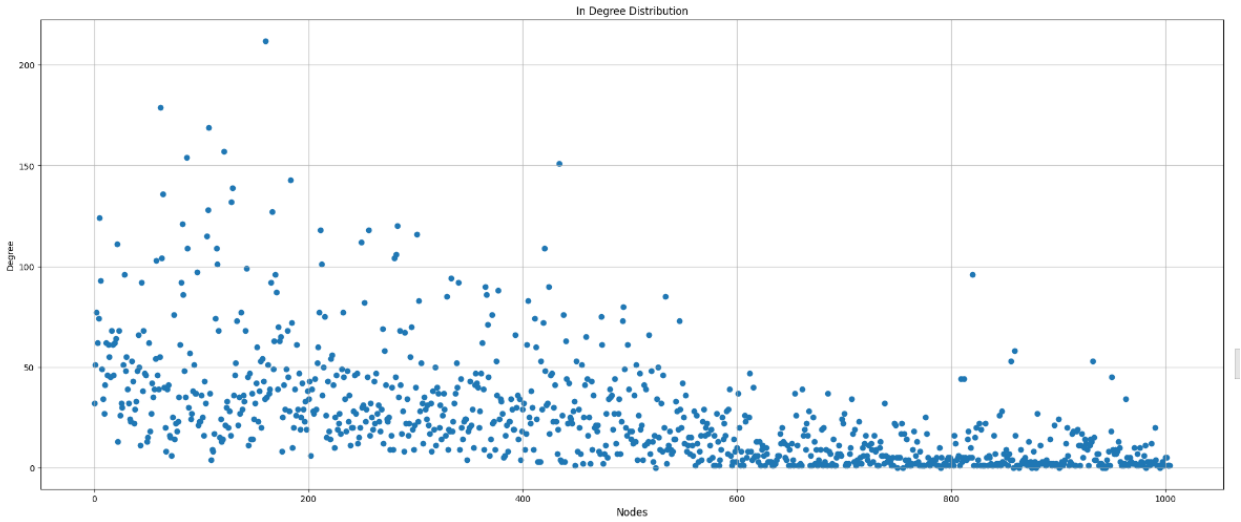
density=1, graph is complete

Network density is calculated as:- Total no. of edges / Total no. of edges that can be present in the graph.

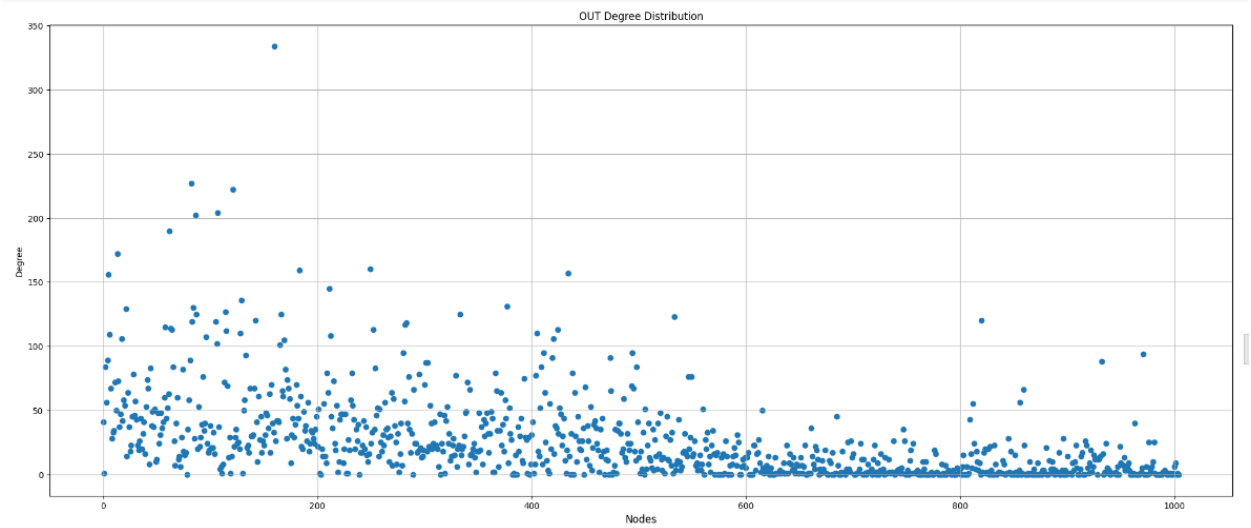
Total no. of edges that can be present in the graph= n*(n-1)

Ques 1b) Plot degree distribution of the network (in case of a directed graph, plot in-degree and out-degree separately).

Ans1 b) Plot for indegree distribution



Plot for outdegree distribution



Ques 1c) Calculate the local clustering coefficient of each node and plot the clustering coefficient distribution (lcc vs frequency of lcc) of the network.

Ans1 c) Local clustering coefficients for each node

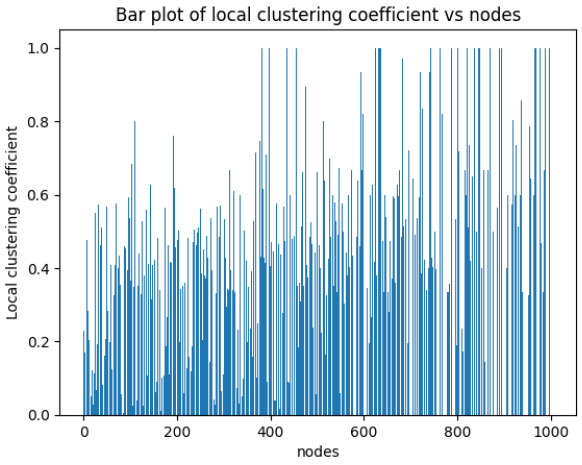
For calculating the clustering coefficients for the network we have used the formula Clustering coefficients= N/n(n-1) where n= total no. of node neighbors.

N= total no of edges among n neighbors of that node in the network

```
for i in range(len(clustering_coefficients)):
  print("Clustering_coefficient for ",i,"=",clustering_coefficients[i])
```

```
Clustering_coefficient for  0 = 0.31007751937984496
Clustering_coefficient for  1 = 0.23058823529411765
Clustering_coefficient for  2 = 0.06840539922214596
Clustering_coefficient for  3 = 0.17060041407867496
Clustering_coefficient for  4 = 0.06741321388577827
Clustering_coefficient for  5 = 0.008353637312913331
Clustering_coefficient for  6 = 0.026238161776121722
Clustering_coefficient for  7 = 0.12635814889336017
Clustering_coefficient for  8 = 0.4777777777777778
Clustering_coefficient for  9 = 0.45479082321187586
Clustering_coefficient for  10 = 0.2839962997224792
Clustering_coefficient for  11 = 0.09477477477477478
Clustering_coefficient for  12 = 0.20488721804511278
Clustering_coefficient for  13 = 0.008411273617475363
Clustering_coefficient for  14 = 0.07789678675754626
Clustering_coefficient for  15 = 0.2988235294117647
Clustering_coefficient for  16 = 0.11106997408367271
Clustering_coefficient for  17 = 0.05087572977481234
Clustering_coefficient for  18 = 0.21548821548821548
Clustering_coefficient for  19 = 0.12152917505030181
Clustering_coefficient for  20 = 0.11796042617960426
Clustering_coefficient for  21 = 0.02898239587805925
Clustering_coefficient for  22 = 0.4725274725274725
Clustering_coefficient for  23 = 0.1129746835443038
Clustering_coefficient for  24 = 0.41700404858299595
Clustering_coefficient for  25 = 0.48295454545454547
Clustering_coefficient for  26 = 0.5490196078431373
```

Plot for clustering coefficient



Ques 2 PageRank, Hubs and Authority

Ans 2a) PageRank for each node:

Pagerank is a method that evaluates the importance of a website by considering the number and quality of links pointing to it, also known as its InDegree nodes. It ranks websites based on their perceived significance, with higher-ranked sites considered more important than lower-ranked ones.

```python
# Sort the pages based on their PageRank score from high to low
sorted_pages = sorted(pagerank_scores.items(), key=lambda x: x[1], reverse=True)

# Print the PageRank score for each node
for node, score in sorted_pages:
  print(f"Node {node} has a PageRank score of {score:.4f}")
```

```
Node 1 has a PageRank score of 0.0094
Node 130 has a PageRank score of 0.0069
Node 160 has a PageRank score of 0.0068
Node 62 has a PageRank score of 0.0053
Node 86 has a PageRank score of 0.0051
Node 107 has a PageRank score of 0.0050
Node 365 has a PageRank score of 0.0048
Node 121 has a PageRank score of 0.0047
Node 5 has a PageRank score of 0.0045
Node 129 has a PageRank score of 0.0045
Node 183 has a PageRank score of 0.0043
Node 64 has a PageRank score of 0.0042
Node 434 has a PageRank score of 0.0042
Node 532 has a PageRank score of 0.0041
Node 128 has a PageRank score of 0.0041
Node 106 has a PageRank score of 0.0040
Node 21 has a PageRank score of 0.0038
Node 166 has a PageRank score of 0.0037
Node 301 has a PageRank score of 0.0036
Node 82 has a PageRank score of 0.0035
Node 227 has a PageRank score of 0.0035
Node 333 has a PageRank score of 0.0035
Node 211 has a PageRank score of 0.0035
Node 377 has a PageRank score of 0.0034
Node 105 has a PageRank score of 0.0034
Node 256 has a PageRank score of 0.0034
Node 283 has a PageRank score of 0.0033
Node 420 has a PageRank score of 0.0032
Node 44 has a PageRank score of 0.0031
Node 87 has a PageRank score of 0.0031
Node 820 has a PageRank score of 0.0031
Node 96 has a PageRank score of 0.0031
Node 114 has a PageRank score of 0.0031
Node 280 has a PageRank score of 0.0030
Node 249 has a PageRank score of 0.0030
Node 28 has a PageRank score of 0.0030
Node 282 has a PageRank score of 0.0029
Node 6 has a PageRank score of 0.0029
Node 81 has a PageRank score of 0.0029
Node 165 has a PageRank score of 0.0029
Node 63 has a PageRank score of 0.0028
Node 58 has a PageRank score of 0.0028
```

Ans 2b) Authority and Hub score for each node

The Authority score of a node refers to the number of incoming edges from other nodes that point towards it, while the Hub score of a node represents the number of outgoing edges from it to other nodes. These scores are calculated based on the interconnectivity of nodes in a network using the Hits algorithm built-in in the Networkx library.

```
[ ]  1 # Compute the HITS algorithm for the graph
     2 hub_scores, authority_scores= nx.hits(Directed_g, max_iter= 100, normalized = True)
```

Hub score for each node

```
[ ]  1 # Sort the hub score high to low
     2 sorted_hubs = dict(sorted(hub_scores.items(), key=lambda item: item[1], reverse=True))
     3 # Print the hub score for each node
     4 for node, score in sorted_hubs.items():
     5     print(f"Node {node} has a Hub score of {score:.6f}")
```

```
Node 169 has a Hub score of 0.005562
Node 115 has a Hub score of 0.005468
Node 63 has a Hub score of 0.005423
Node 58 has a Hub score of 0.005384
Node 424 has a Hub score of 0.005380
Node 166 has a Hub score of 0.005360
Node 128 has a Hub score of 0.005352
Node 13 has a Hub score of 0.005345
```

Authority score for each node

+ Code + Text

```
1 # Sort the authority score high to low
2 sorted_authority = dict(sorted(authority_scores.items(), key=lambda item: item[1], reverse=True))
3 # Print the authority score for each node
4 for node, score in sorted_authority.items():
5     print(f"Node {node} has a authority score of {score:.6f}")
```

```
Node 160 has a authority score of 0.007220
Node 107 has a authority score of 0.006898
Node 62 has a authority score of 0.006696
Node 434 has a authority score of 0.006485
Node 121 has a authority score of 0.006472
Node 183 has a authority score of 0.006041
Node 128 has a authority score of 0.005948
Node 249 has a authority score of 0.005729
Node 256 has a authority score of 0.005704
Node 129 has a authority score of 0.005678
Node 283 has a authority score of 0.005624
Node 82 has a authority score of 0.005478
Node 106 has a authority score of 0.005459
Node 114 has a authority score of 0.005443
Node 87 has a authority score of 0.005425
Node 166 has a authority score of 0.005395
Node 211 has a authority score of 0.005305
Node 105 has a authority score of 0.005188
```

Compare the results obtained from both the algorithms in parts 1 and 2 based on the node scores

```
Node 160 has a Hub score of 0.010629
Node 82 has a Hub score of 0.009617
Node 121 has a Hub score of 0.009530
Node 107 has a Hub score of 0.008788
Node 62 has a Hub score of 0.008233
Node 249 has a Hub score of 0.008018
Node 434 has a Hub score of 0.007541
Node 183 has a Hub score of 0.007200
Node 86 has a Hub score of 0.007003
Node 114 has a Hub score of 0.006398
Node 105 has a Hub score of 0.006320
Node 211 has a Hub score of 0.006317
Node 129 has a Hub score of 0.006263
Node 21 has a Hub score of 0.006185
Node 87 has a Hub score of 0.006105
Node 142 has a Hub score of 0.005976
Node 283 has a Hub score of 0.005945
Node 333 has a Hub score of 0.005935
Node 212 has a Hub score of 0.005905
Node 83 has a Hub score of 0.005796
Node 282 has a Hub score of 0.005794
Node 405 has a Hub score of 0.005693
Node 169 has a Hub score of 0.005562
Node 115 has a Hub score of 0.005468
Node 63 has a Hub score of 0.005423
Node 58 has a Hub score of 0.005384
Node 424 has a Hub score of 0.005380
Node 166 has a Hub score of 0.005360
Node 128 has a Hub score of 0.005352
Node 13 has a Hub score of 0.005345
Node 820 has a Hub score of 0.005321
Node 5 has a Hub score of 0.005048
Node 106 has a Hub score of 0.004967
Node 81 has a Hub score of 0.004888
Node 252 has a Hub score of 0.004816
Node 420 has a Hub score of 0.004766
Node 303 has a Hub score of 0.004668
Node 254 has a Hub score of 0.004633
```

.continued

```
Node 160 has a authority score of 0.007220
Node 107 has a authority score of 0.006898
Node 62 has a authority score of 0.006696
Node 434 has a authority score of 0.006485
Node 121 has a authority score of 0.006472
Node 183 has a authority score of 0.006041
Node 128 has a authority score of 0.005948
Node 249 has a authority score of 0.005729
Node 256 has a authority score of 0.005704
Node 129 has a authority score of 0.005678
Node 283 has a authority score of 0.005624
Node 82 has a authority score of 0.005478
Node 106 has a authority score of 0.005459
Node 114 has a authority score of 0.005443
Node 87 has a authority score of 0.005425
Node 166 has a authority score of 0.005395
Node 211 has a authority score of 0.005305
Node 105 has a authority score of 0.005188
Node 212 has a authority score of 0.005182
Node 115 has a authority score of 0.005129
Node 169 has a authority score of 0.004991
Node 280 has a authority score of 0.004844
Node 81 has a authority score of 0.004800
Node 86 has a authority score of 0.004799
Node 28 has a authority score of 0.004781
Node 63 has a authority score of 0.004781
Node 282 has a authority score of 0.004745
Node 21 has a authority score of 0.004735
Node 64 has a authority score of 0.004569
Node 58 has a authority score of 0.004558
Node 820 has a authority score of 0.004498
Node 142 has a authority score of 0.004425
Node 329 has a authority score of 0.004371
Node 303 has a authority score of 0.004369
Node 83 has a authority score of 0.004364
Node 340 has a authority score of 0.004334
Node 424 has a authority score of 0.004227
Node 420 has a authority score of 0.004225
```

..conti.

From the results above, the pagerank and authorities work upon incoming nodes but still shows different nodes as important because pagerank takes nodes with maximum in-degree i.e. for node 1 whereas, authority score is calculated using maximum incoming nodes from the hubs, so that doesn't includes total incoming nodes. In Hubs, the Outgoing nodes are referred and thus, 160 has maximum OutDegree nodes Other than that we can observe that node 160 has maximum hub score and authority score which shows that it is a highly connected and influential node in the network. This suggests that it is both central and authoritative, and plays a key role in the network.