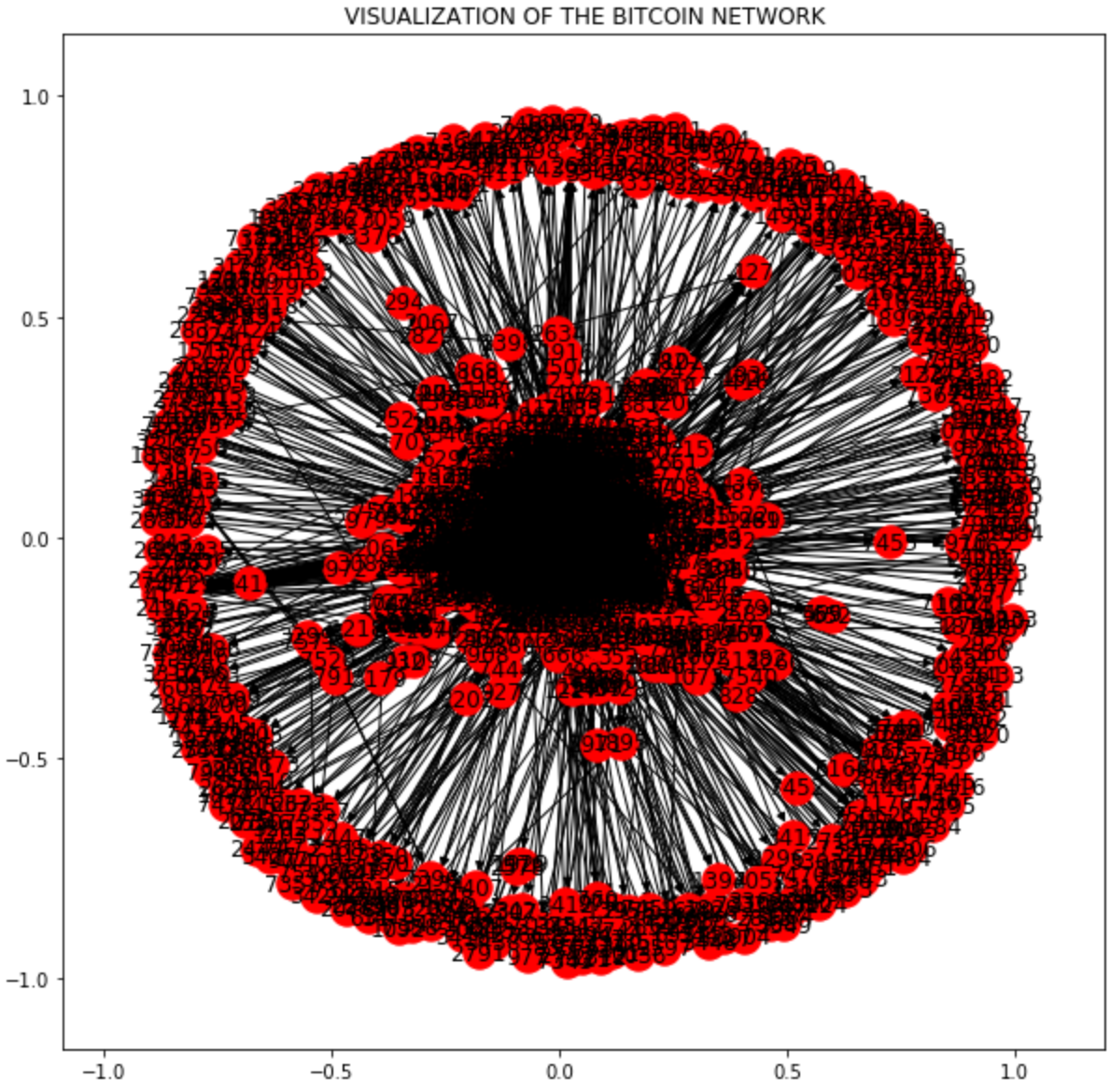
ASSIGNMENT-6
REPORT & ANALYSIS

- Firstly I had chosen a network dataset named as **"BITCOIN ALPHA WEB OF TRUST NETWORK"**. This network describes trust between two bitcoin users. If an edge exists from node A to node B then A trusts on B. The dataset comprises columns such as source node, destination node, edge weight and time stamp.



- The network is directed graph (has directed edges), signed edge weights (edges may have negative weights too).

	source	target	weight	time stamp
0	7188	1	10	1407470400
1	430	1	10	1376539200
2	3134	1	10	1369713600
3	3026	1	10	1350014400
4	3010	1	10	1347854400

Fig1 . First 5 rows of the data

- Statistics of the Network are as follows,

PARAMETER	VALUE
TOTAL NUMBER OF NODES IN THE NETWORK	3783
TOTAL NUMBER OF EDGES IN THE NETWORK	24186
AVERAGE IN-DEGREE IN THE NETWORK	6.39 (rounded off to two decimal places)
AVERAGE OUT-DEGREE IN THE NETWORK	6.39 (rounded off to two decimal places)
NODE WITH MAXIMUM IN-DEGREE IN THE NETWORK	1
NODE WITH MAXIMUM OUT-DEGREE IN THE NETWORK	1
DENSITY OF THE NETWORK	0.002 (rounded off to two decimal places)

Fig 2. Network Statistics

- **The Average indegree and out degree** remain the same because, every outdegree constituted by a single node will be counter-balanced by indegrees of the nodes having incoming edges.
- **The density of the network** tells us how dense the network is, if the density of the network is 1, then the graph is a complete graph, if it is 0 then the network has no edges. **Density of network = (total number of edges in the network) / (n)*(n-1) ; n = total count of nodes in network (in case of directed graph)**

- **In Edge list** If we represent all the sets of edges present in the network in a list then it is called edge list representation. Fraction of the output image of how the edge list looks like can be seen in the below picture.

EDGE LIST REPRESENTATION OF THE NETWORK

```
[(7188, 1), (430, 1), (3134, 1), (3026, 1), (3010, 1), (804, 1), (160, 1), (95, 1), (377, 1), (888, 1), (89, 1), (1901, 1),
(161, 1), (256, 1), (351, 1), (3329, 1), (3341, 1), (649, 1), (1583, 1), (87, 1), (37, 1), (309, 1), (821, 1), (1496, 1), (63
7, 1), (964, 1), (594, 1), (2249, 1), (554, 1), (20, 1), (2227, 1), (1315, 1), (519, 1), (1316, 1), (2149, 1), (1724, 1), (1
8, 1), (57, 1), (118, 1), (3254, 1), (1177, 1), (112, 1), (11, 1), (586, 1), (35, 1), (15, 1), (1445, 1), (152, 1), (2, 1),
(113, 1), (44, 1), (2401, 1), (10, 1), (2378, 1), (126, 1), (3245, 1), (783, 1), (493, 1), (1358, 1), (1180, 1), (529, 1), (3
33, 1), (1538, 1), (2282, 1), (1519, 1), (2966, 1), (474, 1), (330, 1), (958, 1), (17, 1), (1295, 1), (38, 1), (1952, 1), (22
3, 1), (625, 1), (1392, 1), (3355, 1), (1881, 1), (58, 1), (96, 1), (1580, 1), (196, 1), (146, 1), (416, 1), (1198, 1), (331
9, 1), (1867, 1), (896, 1), (617, 1), (3300, 1), (1877, 1), (462, 1), (3279, 1), (454, 1), (1860, 1), (121, 1), (151, 1), (15
70, 1), (1573, 1), (1063, 1), (1353, 1), (459, 1), (2334, 1), (1267, 1), (1060, 1), (1061, 1), (7431, 1), (1355, 1), (71, 1),
(3070, 1), (2113, 1), (3001, 1), (396, 1), (2260, 1), (142, 1), (2238, 1), (123, 1), (2942, 1), (1509, 1), (7410, 1), (1760,
1), (2876, 1), (259, 1), (1493, 1), (2845, 1), (370, 1), (2844, 1), (2167, 1), (156, 1), (2808, 1), (255, 1), (736, 1), (760
3, 1), (346, 1), (9, 1), (75, 1), (175, 1), (2754, 1), (22, 1), (155, 1), (1261, 1), (2552, 1), (2586, 1), (710, 1), (578,
1), (472, 1), (2472, 1), (1606, 1), (4, 1), (563, 1), (744, 1), (3422, 1), (250, 1), (249, 1), (2427, 1), (3414, 1), (1590,
1), (2305, 1), (1900, 1), (3392, 1), (1065, 1), (1147, 1), (1072, 1), (3375, 1), (1024, 1), (709, 1), (1382, 1), (3332, 1),
(1885, 1), (3330, 1), (1197, 1), (3316, 1), (1886, 1), (154, 1), (1522, 1), (2391, 1), (3298, 1), (432, 1), (3292, 1), (3290,
1), (7597, 1), (3274, 1), (1875, 1), (291, 1), (158, 1), (1579, 1), (1846, 1), (1066, 1), (290, 1), (3233, 1), (891, 1), (24
```

Fig 3. Edge List representation of network

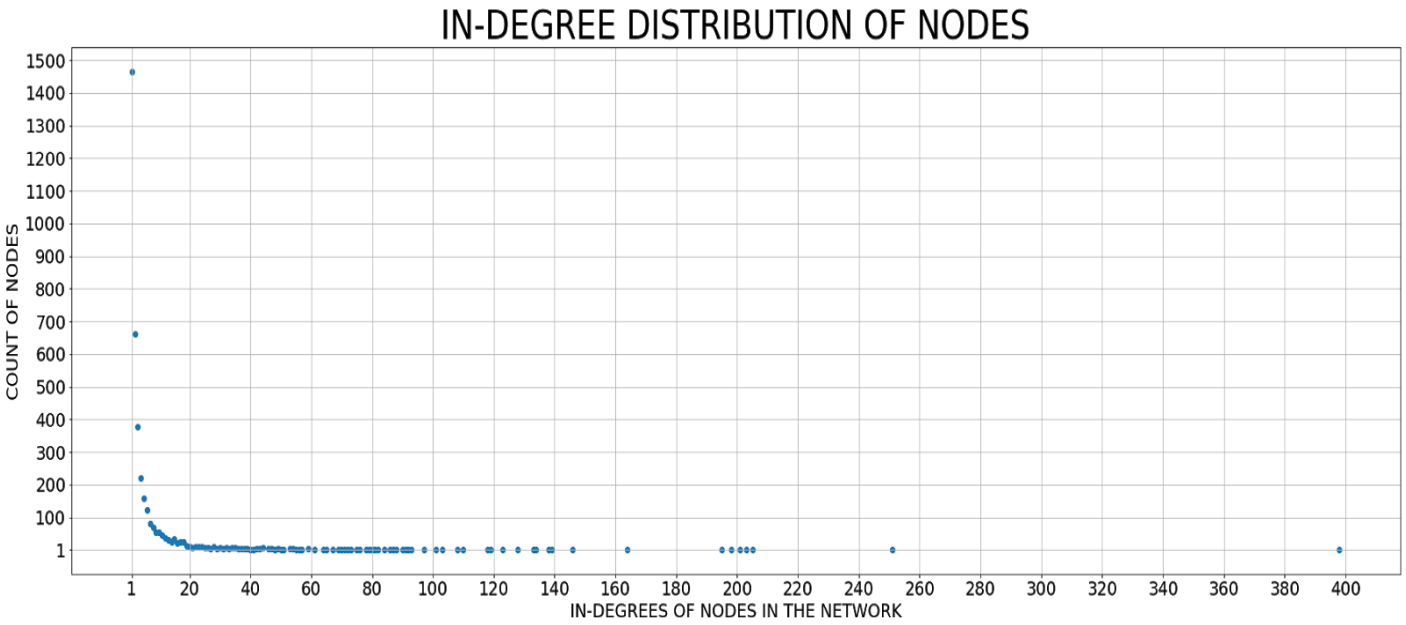
- **In the adjacency matrix** representation of the network, if an edge goes from node A to node B then the entry of adj_mtx[A][B] will be made 1. If in case of undirected graphs or networks both adj_mtx[A][B] and adj_mtx[B][A] will be made 1. **Cropped image** of entire adjacency matrix is shown below,

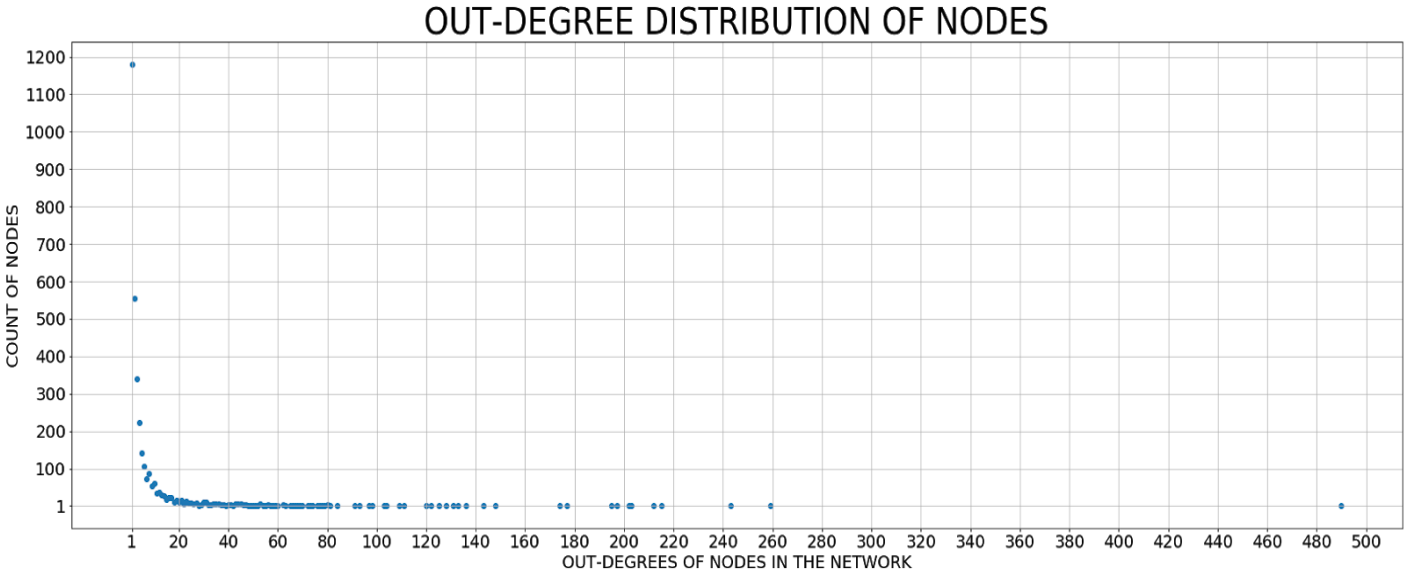
ADJACENCY MATRIX

	1	2	3	4	5	6	7	8	9	10	...	7595	7596	7597	7598	7599	7600	7601	7602	7603	7604
1	0	1	0	1	0	0	0	0	1	1	...	0	0	1	0	0	0	0	0	1	0
2	1	0	0	1	1	0	1	1	1	1	...	0	0	0	0	0	0	0	0	1	0
3	0	1	0	0	1	1	1	1	0	1	...	1	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	1	1	...	0	1	0	0	0	0	0	0	1	0
5	0	1	1	0	0	1	0	1	0	0	...	0	0	0	0	0	0	0	0	0	0
6	0	0	1	0	1	0	1	1	0	0	...	0	0	0	0	0	0	0	0	0	0
7	0	1	1	0	0	1	0	1	0	0	...	1	0	0	1	1	1	1	1	0	1
8	0	1	1	0	1	1	1	0	0	1	...	1	0	0	0	0	0	0	0	0	1
9	1	1	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	1	0
10	1	1	1	1	0	0	0	1	0	0	...	0	0	0	0	0	0	0	0	0	0
11	1	1	1	1	1	1	1	1	0	1	...	0	0	1	0	0	0	0	0	1	0
12	0	0	1	0	1	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
13	0	1	1	0	1	0	1	1	0	0	...	0	0	0	0	0	0	0	0	0	0
14	0	1	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	1	0
15	1	1	0	0	0	0	0	0	1	1	...	0	0	0	0	0	0	0	0	0	0
16	0	1	0	1	0	1	0	0	0	0	...	0	1	0	0	0	0	0	0	0	0
17	0	1	0	1	0	0	0	0	1	1	...	0	0	0	0	0	0	0	0	1	0

Fig 4. Adjacency Matrix representation of network

- **In-degree & out-degree distribution** of nodes describe the range of indegrees & Out-degrees that the nodes in the given network are possessing respectively.





- **<u>Clustering coefficient</u>** of a node can be defined as the measure of certainty of a node being a part of a large cluster. Clustering coefficient lies between the range of 0 to 1. Higher the coefficient higher is the certainty.(cropped output)
    - NODES WITH CLUSTERING CO-EFFICIENT 0 -> 2317
    - NODES WITH CLUSTERING CO-EFFICIENT 1 -> 173

NODE	CLUSTERING-COEFFICIENT
261	1.0
352	1.0
399	1.0
418	1.0
524	1.0
538	1.0
556	1.0
600	1.0
736	1.0
746	1.0
755	1.0
758	1.0
771	1.0
802	1.0
818	1.0
902	1.0
910	1.0

- CLUSTERING COEFFICIENT OF THE NETWORK:: 0.12635262321014096
- **Clustering coefficient of the node = $(N_v)/(N)(N-1)$; (in case of directed graph)**
- N= Total count of neighbours for the node; N_v = Total count of edges among N neighbours of that node in the network.

- **<u>Degree Centrality</u>** tells the value of indegree and out degree of every node in the network. (Cropped Output)

NODE	IN-DEGREE
1	398
3	251
2	205
11	203
4	201
177	198
7	195
10	164
5	146
6	139
26	138
8	134
9	133
12	128
15	123

NODE	OUT-DEGREE
1	490
8	259
3	243
4	215
7	212
11	203
177	202
15	197
2	195
10	177
5	174
22	148
6	148
9	143
33	136

- **Pagerank** is an algorithm developed by Larry Page which assigns the score based on how important and relevant the website/node is. Higher the score the more its importance. Pagerank assumes no self edges or multi-edges between the nodes and the edges are authentic and true. If a node has more number of incoming edges it generally tends to have higher page rank. It infers us that it is the most authentic page that can be trusted upon. The pagerank scores for the given network node wise can be seen below. (cropped output image)

PAGE-RANK SCORES

==========

{1: 0.016993099228405292, 3: 0.008960320960927014, 4: 0.008023439583332513, 2: 0.0066231832832233375, 177: 0.0066078111083345
09, 7: 0.006543114032848893, 11: 0.006192044052141409, 10: 0.005598872299087015, 13: 0.005287958636203423, 6: 0.0047801213904
85108, 16: 0.004707642392717352, 5: 0.004622857702795663, 7564: 0.004203501017394427, 9: 0.004200846400007224, 8: 0.004107053
817032611, 12: 0.004083127744453426, 26: 0.0039019566135912753, 33: 0.003606231563084478, 14: 0.0035589459835260726, 798: 0.0
03507194907540784, 17: 0.003495447122947391, 18: 0.0034018120194040016, 25: 0.003366926463452764, 22: 0.0033578456422887156,
15: 0.003286117909070405, 79: 0.0032814003133206982, 7603: 0.0032054173740647123, 95: 0.002948042055839976, 69: 0.00292804319
02535906, 40: 0.0028759663345321314, 21: 0.0028670672239916864, 23: 0.0027621295735930852, 30: 0.0027393055183080562, 27: 0.0
0268384322969599, 57: 0.002619249813194172, 19: 0.0026160666008117834, 58: 0.00260278770672031, 145: 0.0025221379014626934, 2
8: 0.0025003778728845292, 42: 0.0024818084985976097, 24: 0.002433165613813747, 50: 0.00243186601227601, 29: 0.002388321694051
232, 45: 0.0023269786253837044, 51: 0.0022682057226275825, 48: 0.0022326280993909273, 43: 0.002201340579827938, 36: 0.0021867
875014486638, 35: 0.0021851765196186654, 38: 0.00218216892336012, 117: 0.0021601214502419024, 39: 0.002142403722776401, 46:
0.0021289749273413396, 61: 0.0021253184242606203, 84: 0.0021249415098918705, 32: 0.0020847247103996346, 7604: 0.0020625547795
917886, 31: 0.002015112440032661, 85: 0.0019991539714934985, 34: 0.0019911500490423246, 52: 0.0019448225223099005, 65: 0.0019
410068184605067, 125: 0.0019251337191082418, 78: 0.0019108845637578145, 41: 0.0018898556957635932, 130: 0.0018510811760545493,
89: 0.0018278969974936 35, 67: 0.0018159565462051342, 44: 0.0017402701916294 44, 49: 0.0017123600231220607, 239: 0.001710774247

- **Authority and hub-scores** are other two metrics used to measure the importance of web pages. highly related web-pages to the given query are called as root nodes.pages that are not very relevant but point to authorities/roots are called hubs. A good authority or a root node has many hubs pointing to it. The authority and hub scores for the given network node wise can be seen below. (cropped output image)

HUB SCORES

==========

{11: 0.008537684146378718, 177: 0.0069610043633879385, 3: 0.0068841928868758435, 2: 0.006829069854496472, 7: 0.00670052410224
19755, 8: 0.006528863504334652, 1: 0.0064088516365644315, 22: 0.006132935795890347, 10: 0.006024201182498987, 26: 0.005889592
272392635, 24: 0.005540072414254607, 58: 0.005469946837299 06, 95: 0.00544198036583848, 5: 0.005399418719920633, 30: 0.0053505
99951392969, 6: 0.005194347787695366, 19: 0.005059119020232108, 33: 0.005025729714269618, 85: 0.004953059533326474, 15: 0.004
870427357673242, 9: 0.004860602173090094, 25: 0.004712132582187772, 4: 0.004691121478640904, 17: 0.004455626994995476, 43: 0.
004345421028049623, 12: 0.004263498693882872, 21: 0.004179224971228584, 42: 0.004172923596950638, 32: 0.004034106748731866, 3
6: 0.0038080363896429545, 51: 0.003763388313850775, 29: 0.003485373462676724, 88: 0.0034612985649328334, 40: 0.00335257685440
31124, 154: 0.0033159002481623653, 47: 0.0033078008055692195, 31: 0.0032788345216668837, 70: 0.003237748705468323, 67: 0.0031
973532468442868, 49: 0.0031683785575469057, 34: 0.0031148885550329047, 27: 0.0030299677813691 76, 75: 0.003027177357114127, 7
3: 0.0030059439763306726, 145: 0.003004655095224043, 14: 0.002964090277389432, 20: 0.0029636943928530155, 90: 0.0029143330509
118696, 124: 0.0028537218447468454, 190: 0.002851810880619545, 125: 0.0028069565647828127, 52: 0.0027401317166933672, 491: 0.
0027332309544775752, 7603: 0.0026777507358885346, 16: 0.0026644015998716 92, 142: 0.0026605221853644675, 83: 0.002644225357217
9593, 62: 0.0026232863670861 46, 2336: 0.0026194792090452 5, 173: 0.0026058951401279 4, 156: 0.0025951530476664 22, 174: 0.00258
04557248153455, 166: 0.0025722717450812945, 97: 0.002558323662683099, 123: 0.0024902887522977703, 103: 0.0024875660204487803,
197: 0.0024384243220122396, 65: 0.0023906422136575 9, 87: 0.0023801112496524067, 136: 0.0023764083543394935, 116: 0.0023650485

AUTHORITY SCORES

==========

{11: 0.0077489839735902805, 3: 0.0069533608643944835, 2: 0.00681199455141867, 177: 0.006191924885749605, 7: 0.006059056891417
5415, 1: 0.005881193763593962, 26: 0.005754122352972832, 10: 0.005389497289307632, 5: 0.005046574210713 64, 9: 0.0049807011979
27936, 24: 0.0049521885112465665, 8: 0.004879300311058208, 6: 0.0048171309043961, 95: 0.004673457088070035, 29: 0.00456671225
5379265, 19: 0.004532679241755887, 31: 0.00419539259947677, 42: 0.004170545806443817, 22: 0.0041292334157049205, 33: 0.004124
227775817787, 4: 0.004089896516654834, 43: 0.004067548464244481, 25: 0.004009489947653647, 32: 0.0039776405340792 96, 21: 0.00
3934323394202373, 17: 0.003928190812986274, 15: 0.003882770209023 3293, 145: 0.0038084418325084377, 12: 0.003807720156968241,
85: 0.0037939565053383235, 36: 0.003621013467926249, 51: 0.0034736044551958497, 47: 0.0033746106781864242, 30: 0.003354302105
8363956, 41: 0.0033377473653706845, 14: 0.0031940616583407736, 7604: 0.003158914149452438, 27: 0.0031570745138219 99, 40: 0.00
30567818087404493, 34: 0.003054934617233669, 7603: 0.0029732494473495357, 13: 0.002972490476815529, 67: 0.002967291156581427
5, 70: 0.0027819194576384045, 90: 0.0027799697578562666, 75: 0.002778674774309781, 73: 0.0027122365976914773, 125: 0.00270416
5574800716, 154: 0.0026719011851675 27, 124: 0.0026689816059925105, 20: 0.0026101916075421 01, 49: 0.0025767830774362155, 103:
0.0025299336530964682, 16: 0.0025047054503686 17, 142: 0.0024789801035171475, 83: 0.0024357094330081 95, 123: 0.002381864688048
2716, 92: 0.002351750557467288, 798: 0.0023009149609117 93, 113: 0.0022969195339818875, 174: 0.0022653875257253858, 52: 0.0022
646493006235465, 116: 0.0022635087630090 25, 58: 0.0022583022657176 57, 62: 0.0022379368675558 25, 156: 0.0022339128857587 64, 8
8: 0.0022334523987690 4, 133: 0.0022279707540133835, 166: 0.0022221460713322933, 35: 0.0022163423410008 76, 197: 0.002215285893

- ## INFERENCES

1. As the density of the network is 0.002 (rounded off to two decimal places) it seems that the network (graph) is too shallow or sparse.(Maximum possible edges are 1,43,07,306 but the existing edges in the network are 24,186)

2. Observing the indegree distribution of the nodes it seems that around 40% of the nodes in the network have an indegree of 1 and most of the nodes have an in-degree lying between 1 to 90 itself.

3. Observing the out-degree distribution of the nodes it seems that around 32% of the nodes in the network have an out-degree of 1 and most of the nodes have an out-degree lying between 1 to 80 itself.

4. Observing the clustering coefficients of the nodes it seems that around 62% of the total nodes in the network are not part of any clusters and also only 4% of the total nodes are part of huge clusters.

5. Nodes 1,2,3,11,4 are the top-5 nodes having the highest number of incoming edges into them, and the nodes 1,8,3,4,7 are the top-5 nodes having the highest number of outgoing edges from them.

6. As the nodes with most incoming edges have higher page rank , when seen in the degree centrality measure, nodes 1,3,2,11,4,177,7,10,5,6 are top-10 nodes to have higher in-degree value and also the top-10 nodes to have higher pagerank scores. These nodes are the ten most valuable/important nodes in the network as per page rank algorithm.

7. The top-10 authority nodes are **11**,**3**,**2**,0,**7**,**1**,**26**,**10**,5,9. It infers us that these nodes have a good count of indegree from hubs. Which is almost similar to the result of pagerank. The top-10 hubs in the network are **11**,177,**3**,**2**,**7**,8,**1**,22,**10**,**26**. It seems that the nodes 1,2,3,7,10,11,26 are the most important hubs as well as authorities. This infers that there are some nodes in the network whose in-degree is as high as their out-degree thus they are acting as best hubs as well as best authorities.

References:

[1]https://www.geeksforgeeks.org/hyperlink-induced-topic-search-hits-algorithm-using-networxx-module-python/
[2] https://www.geeksforgeeks.org/page-rank-algorithm-implementation/