IR Assignment 3

Report

Parul Jain - MT20164

Sristi Sharma - MT21098

Question 1:

Dataset Description :

We used the Wiki-Vote dataset. Wiki-Vote is a collection of Wikipedia votes that are used to determine who is qualified to be a Wikipedia administrator. Wikipedia is a platform for a vast collection of information edited by various administrators and contributors. To become an admin of Wikipedia, one needs to issue a Request for Adminship (RfA), for which votes are counted based on votes, admin is being chosen.

1. **Number of Nodes :**

   Minimum no. of Nodes - 3

   Maximum no. of Nodes - 8297

```
Minimum Node No.:  3
Maximum Node No.:  8297
```

   Unique Nodes :

```
Number of Unique Nodes:  7115
```

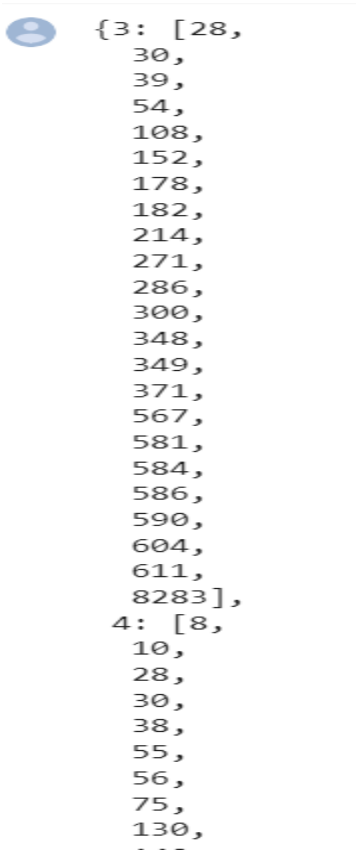
2. **Number. of Edges :**

```
Number of Unique Edges:  103689
```

3. **Adjacency Matrix :**

```
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

Adjacency List :



4. **Avg In-degree :**

   Formula : Average In-degree = Sum(In-degrees) / No. of In-degrees

```
Average In-Degree:        14.573295853829936
```

5. **Avg Out-degree :**

   Formula : Average Out-degree = Sum(Out-degrees) / No. of Out-degrees

```
Average Out-Degree:      14.573295853829936
```

6. **Node with Max In-degree :**

```
Node with Max In-degree:         4037
Max In-degree:   457.0
```

7. **Node with Max Out-degree :**

```
Node with Max out-degree:        2565
Max Out-degree:  893.0
```

8. **The density of the network :** The density of the network is defined as the fraction of edges present over all possible edges.

Formula: $\eta = |E| / |V|(|V|-1)$

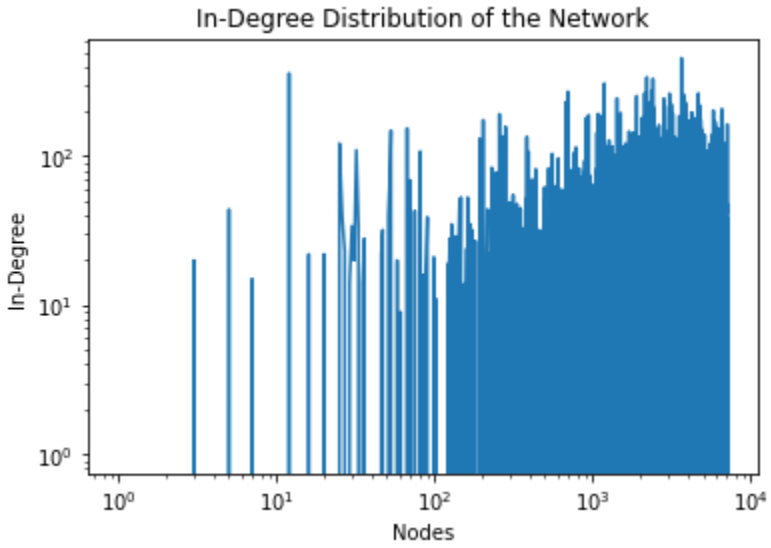
$|E|$ = **Number of edges**
$|V|$ = **Number of vertices**

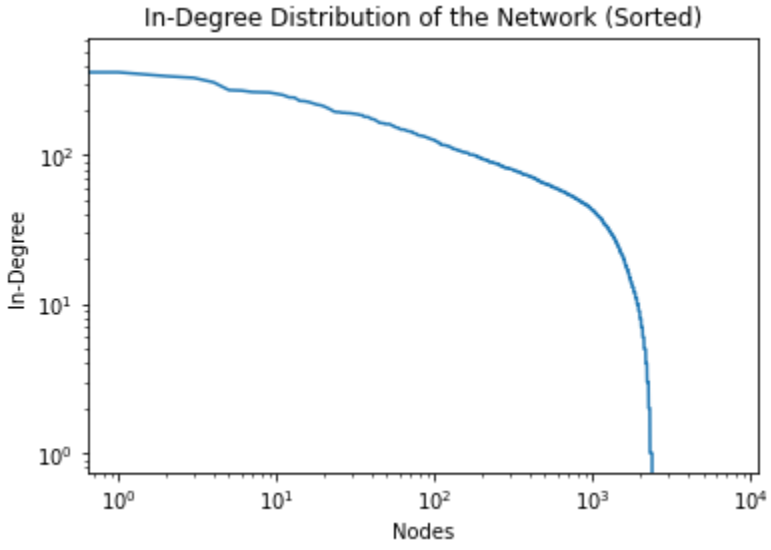
```
Density of the network:  0.0020485375110809584
```

```
Density of the network:  0.0020485375110809584
```

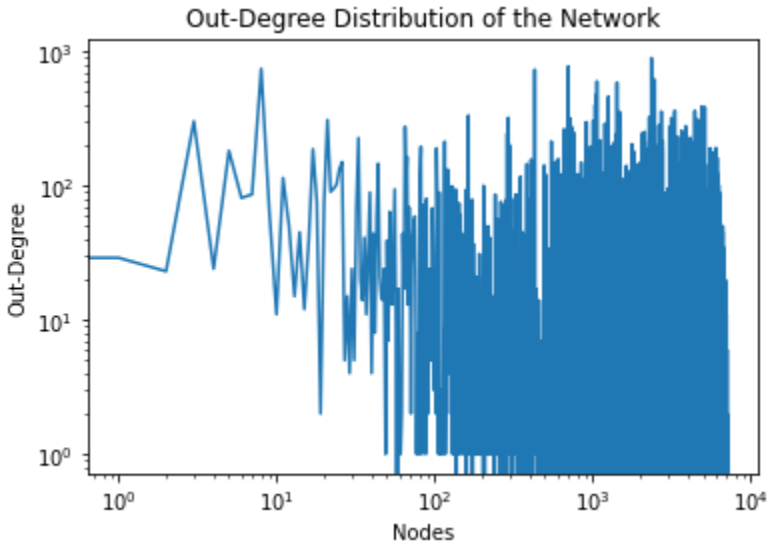
9. **Plot degree distribution of the network (in case of a directed graph, plot in-degree and out-degree separately).**

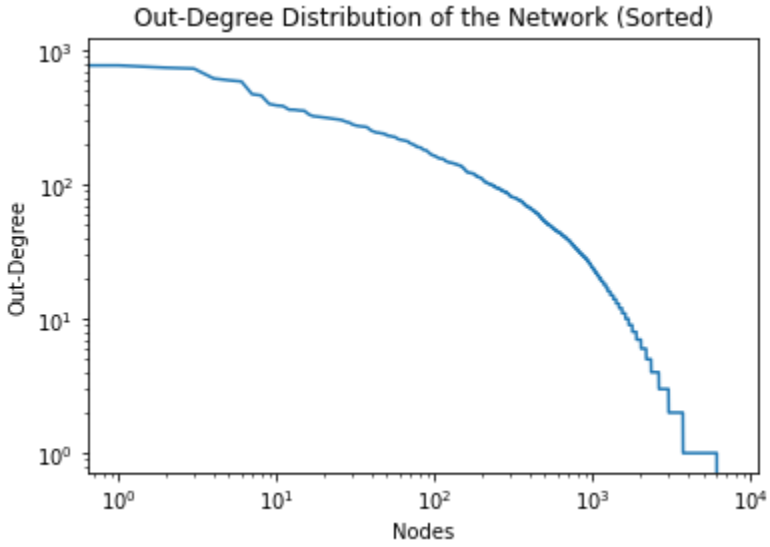
   a. **In-degree distribution :**





b. Out-degree distribution :

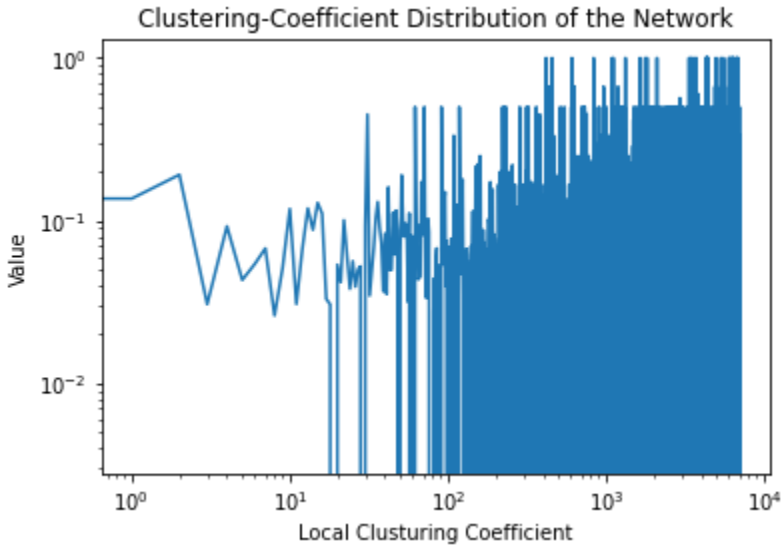


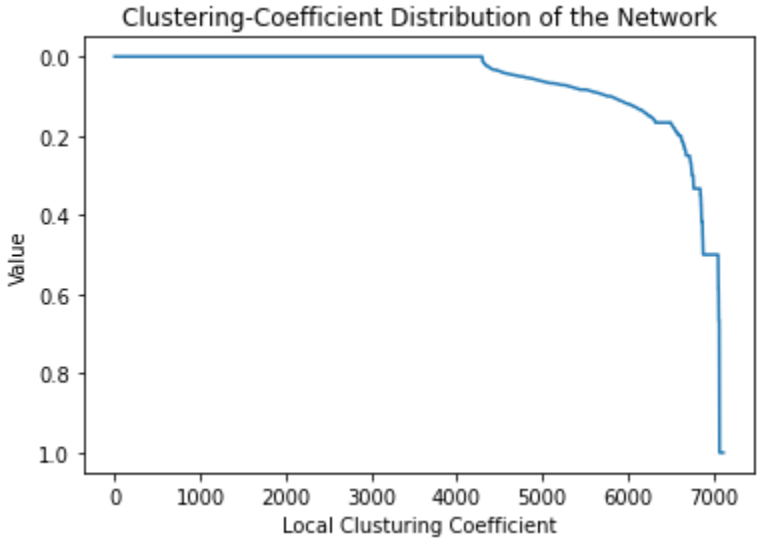


10. Calculate the local clustering coefficient of each node and plot the clustering-coefficient distribution of the network.

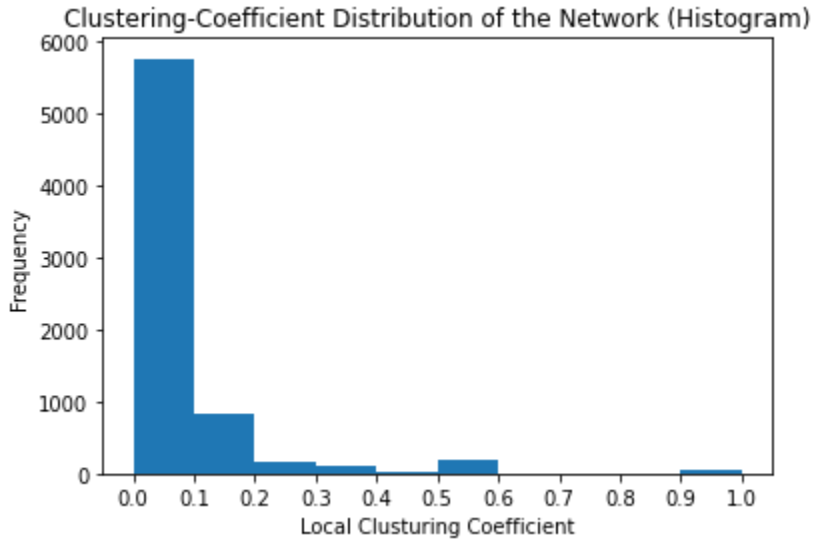
```
{3: 0.07509881422924901,
 4: 0.13669950738916256,
 5: 0.191699604743083,
 6: 0.030659391432531737,
 7: 0.09239130434782608,
 8: 0.04319713435735535,
 9: 0.05416666666666667,
 10: 0.06757865937072503,
 11: 0.02612160941473519,
 12: 0.051201923076923075,
 13: 0.11818181818181818,
 14: 0.030740568234746156,
 15: 0.06693877551020408,
 16: 0.11904761904761904,
 17: 0.08787878787878788,
 18: 0.12878787878787878,
 19: 0.1106612685560054,
 20: 0.03306307860387557,
 21: 0.030526315789473683,
 22: 0.0,
 23: 0.05368382080710848,
 24: 0.04190308073803219,
 25: 0.10074906367041199,
 26: 0.060582306830907054,
 27: 0.03805086390992041,
 28: 0.05639097744360902,
 29: 0.03923937360178971,
 30: 0.05,
 31: 0.05238095238095238,
 32: 0.0,
 33: 0.09239130434782608,
 34: 0.45,
 35: 0.03461538461538462,
 36: 0.049471755487115514,
 37: 0.06842105263157895,
```

Clustering Coefficient Distribution Graph :





Clustering-Coefficient Distribution of the Network (Histogram):


Clustering-Coefficient Distribution of the Network (Histogram)
Frequency
Local Clusturing Coefficient
0
1000
2000
3000
4000
5000
6000
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Question_2)

1) **Calculate Pagerank**

```
pr
```

```
'5818': 0.0006044957008480287,
'6229': 0.0006076925581037548,
'6262': 0.0003462320080l054115,
'6780': 0.0006121431328794814,
'7726': 0.0002339955931300 3906,
'8288': 0.0001725520263834 6065,
'21': 5.0487823458630175e-05,
'794': 0.0001944303630l971945,
'2814': 0.0003992975717399 2565,
'22': 5.0487823458630175e-05,
'601': 9.0856470141385 88e-05,
'670': 0.000115379822101417,
'1029': 0.0001722208560783937,
'1039': 0.0001039763489530 5467,
'1593': 0.0003834003522033 2726,
'24': 5.0487823458630175e-05,
'311': 0.0002208477193511 7008,
'633': 0.000274781526749 0076,
'742': 0.0001766932678097 9478,
'769': 0.000482916622454 4633,
'856': 0.000326654847878437,
'1100': 0.0001461362510391 8607,
'1267': 0.000480202343391 6766,
'1315': 0.0003053429179076 2003,
'1357': 0.000499611383713 4112,
'1416': 0.0003576030248531 6934,
'1419': 0.000177490194087 9044,
'1473': 0.000549703096557 2309
```

✓ 0s

2)

2) Find Hub_Score and Authority Score

```
] hubs,authority=nx.hits(wiki_graph,max_iter=100,normalized='True')
  hubs,authority
```

```
    '1012': 0.0002919109637567062,
    '1018': 0.0004316058146715862,
    '1031': 0.0005201037194784301,
    '1034': 0.0003575655715398478,
    '1061': 0.0004234813592740686,
    '1075': 0.0001925529003534098,
    '1080': 0.0002628649124648275 7,
    '1097': 0.0004507578161443358,
    '1111': 0.0005221980132109163,
    '1124': 0.0001539415543686372 3,
    '1127': 0.0002657837258784777,
    '1128': 0.0001217171281209634 5,
    '1137': 0.0002619640158775 77,
    '1151': 0.0006177669374588397,
    '1154': 0.0003262020273838909,
    '1164': 0.0006890649949626053,
    '1165': 0.0004434433614292344,
    '1185': 0.0007664839264511934,
    '1186': 0.0010000980405660817,
    '1191': 0.0004454076933200136,
    '1196': 0.0001312684149771958 4,
    '1199': 0.0005465937360779417,
    '1200': 0.0004841111120312053,
    '1201': 0.0007067257587802613,
    '1203': 0.0002260489726937507 3,
    '1211': 0.0013434246809247402,
    '1230': 0.0005414997952412405,
    '1239': 0.0008544476339116682,
```

3) Sorting Authority Score

```
authority_sorted
```

```
'6570': 0.0004433676932568525,
'1859': 0.00044300150246841545,
'7882': 0.0004429918632685488,
'6243': 0.0004416349637031831,
'7788': 0.00044133451213086456,
'6780': 0.0004405161840874033,
'2724': 0.000440192754327615,
'789': 0.00043926955629563233,
'4977': 0.0004392359446259285,
'960': 0.00043889849335563137,
'5305': 0.0004384838153655973,
'7301': 0.000437941505499588,
'3310': 0.00043737457139706874,
'2329': 0.00043695223282713957,
'2922': 0.00043669611924342707,
'86': 0.0004360778453974603,
'859': 0.0004353023585300519,
'764': 0.00043484036220027377,
'7054': 0.0004340797675222387,
'4500': 0.00043398097979674 81,
'7924': 0.00043394092569221354,
'6554': 0.00043392339716656636,
'7443': 0.00043368600772457176,
```

4) Sorting Hubs_Score

```
[21] hubs_sorted
```

```
'3321': 0.0002466010921240168,
'1925': 0.00024637071204017236,
'580': 0.0002461926479840 6094,
'325': 0.00024593380506693314,
'4455': 0.0002457041979985899,
'3196': 0.0002453128507793918,
'1837': 0.0002449184348955648,
'4442': 0.00024427337991353914,
'7132': 0.0002442422672742842,
'5485': 0.0002442372834529296,
'7682': 0.0002441703934768462,
'4661': 0.00024388069206027148,
'7131': 0.00024325970063001886,
'4056': 0.0002429534098217357,
'2706': 0.00024255214304318048,
'3326': 0.00024193251017878029,
'6024': 0.0002413771677929 4897,
'2692': 0.00024074286726494403,
```

5) Sorting PageRank_Score

```
pagerank_sorted
```

```
'7910': 0.0005134497390115 82,
'4299': 0.0005133166022698446,
'3631': 0.0005131168971589021,
'8287': 0.0005117251710434227,
'5886': 0.000509609182711119,
'1319': 0.0005086508582796598,
'4795': 0.0005085704830244999,
'1679': 0.0005071313832162692,
'8290': 0.0005069324864598211,
'3529': 0.0005065996333424629,
'4587': 0.0005045819688432686,
'5144': 0.0005040281354984446,
'5872': 0.0005037077081840194,
'863': 0.0005031261013833063,
'1622': 0.0005028131225905317,
'3910': 0.0005028055176424482,
'2727': 0.0005026266433502654,
'5925': 0.0005023689272194698,
'6417': 0.0005019710353047945,
'5335': 0.0005008885956372853,
'6347': 0.0005001054605639593,
'5262': 0.0004999326512138794,
'1357': 0.0004996113837134112,
```

6) Top 5 values of Hubs_Sorted_Score

```
for k,v in list(hubs_sorted.items())[:5]:
  print(str(k)+":"+str(v))
```

```
2565:0.007940492708143138
766:0.007574335297501244
2688:0.006440248991029861
457:0.006416870490261075
1166:0.006010567902411204
```

7) Top 5 values of Authority_Sorted_Score

```
for k,v in list(authority_sorted.items())[:5]:
  print(str(k)+":"+str(v))
```

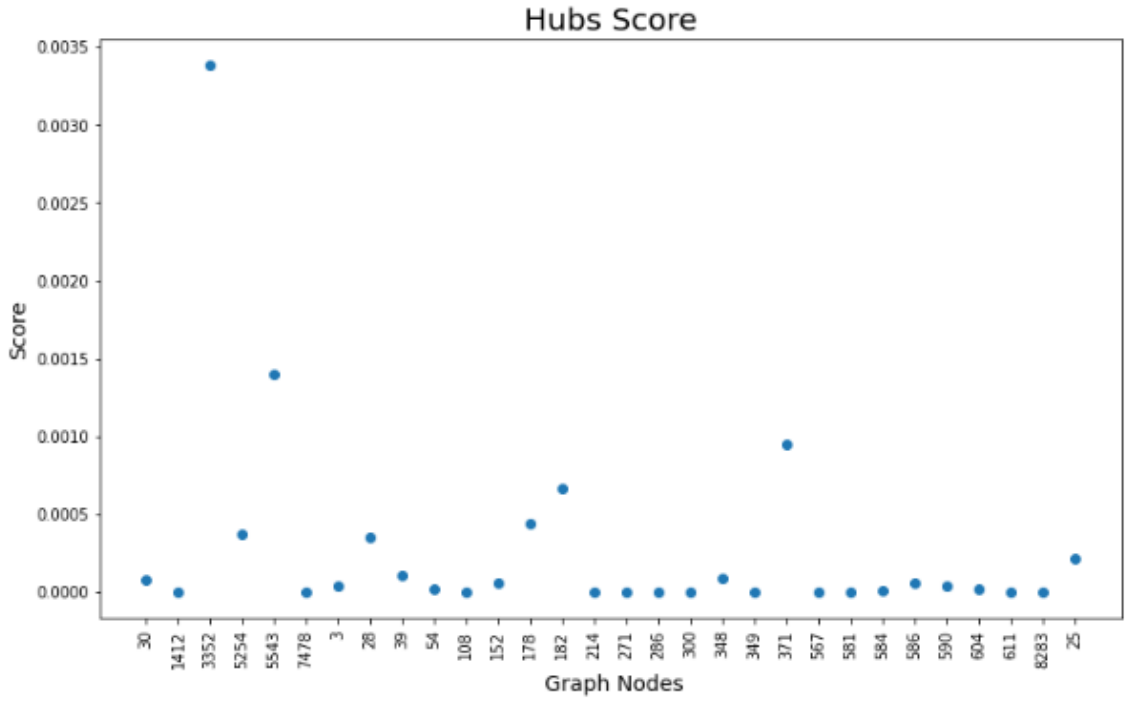
```
2398:0.002580147178008874
4037:0.002573241124229791
3352:0.0023284150914976813
1549:0.002303731480457178
762:0.002255874856287139
```

8) Top 5 values of PageRank_Sorted_Score

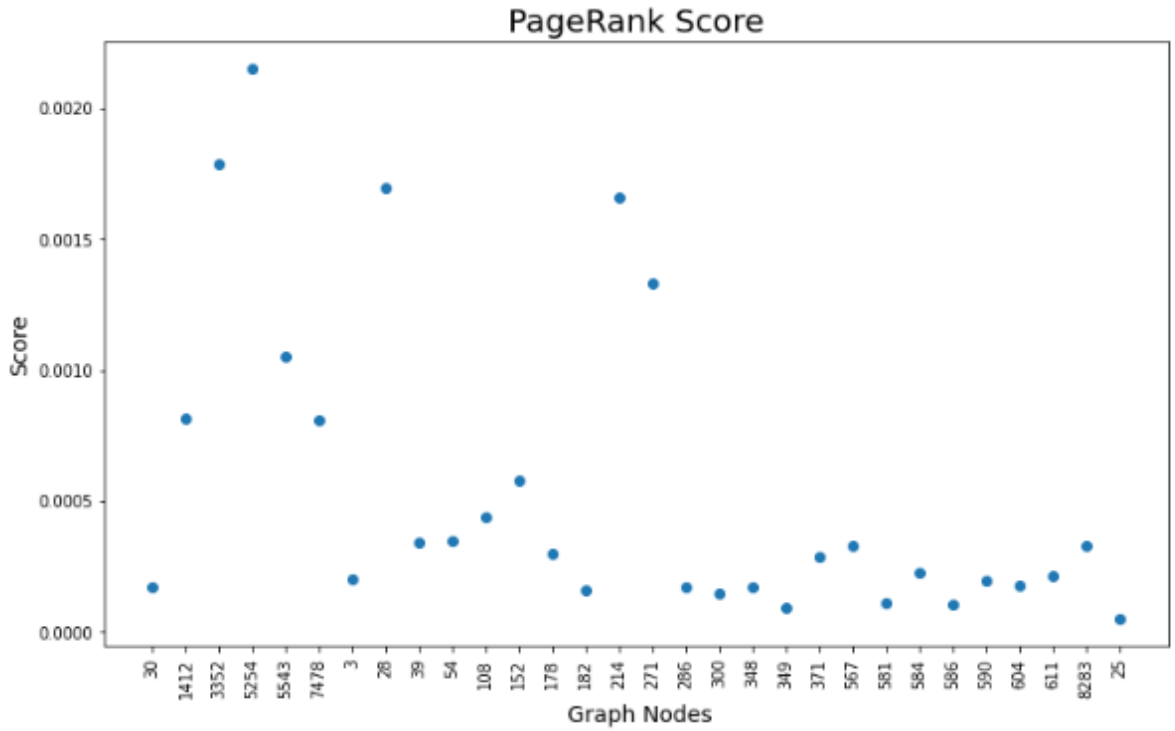
```
for k,v in list(pagerank_sorted.items())[:5]:
  print(str(k)+":"+str(v))
```

```
4037:0.004612715891167545
15:0.0036812207295292714
6634:0.003524813657640258
2625:0.0032863743692308997
2398:0.002605333171725021
```

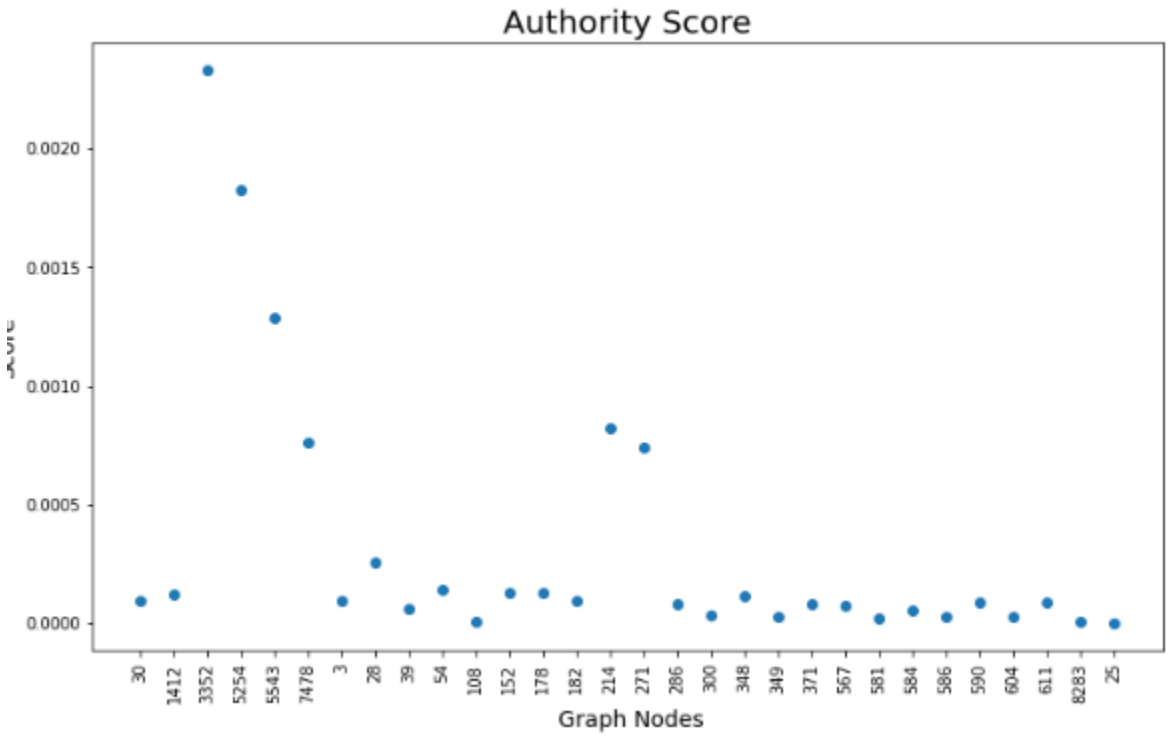
9) Plot Top 30 Values of Hub_Score on Graph



10) Plot Top 30 Values of PageRank_Score on Graph



11) Plot Top 30 Values of Authority_Score on Graph



12) Plot Top 20 Values of on Bar Graph Comparison of PageRank Score, Authority Score, Hubs Score:

We can see the difference between the scores of PageRank,Authority,Hubs scores corresponding nodes.