

## REPORT

Ipsita R (2020379)

Hrishav Basu (2020067)

Aayush Singh (2020009)

### Question 1.

Dataset picked- wiki-Vote

#### 1. Number of Nodes

```
[31] print('no.of nodes:', len(node_set))
```

```
no.of edges: 7115
```

#### 2. Number of Edges

```
print('no.of edges:', len(edges))
```

```
→ no.of edges: 103689
```

#### 3. Avg In-degree

```
7] in_dg = sum(in_)/len(node_set)  
print("Avg. In-Degree: ", in_dg)
```

```
Avg. In-Degree: 14.573295853829936
```

#### 4. Avg. Out-Degree

```
5] out_dg = sum(out_)/len(node_set)  
print("Avg. Out-Degree: ", out_dg)
```

```
Avg. Out-Degree: 14.573295853829936
```

#### 5. Node with Max In-degree

```
3] print('Node with Max In-degree',max_in, 'Node: ',in_idx)
```

```
Node with Max In-degree 457 Node: 4037
```

#### 6. Node with Max out-degree

```
7] print('Node with Max out-degree',max_out,'Node: ',out_idx)
```

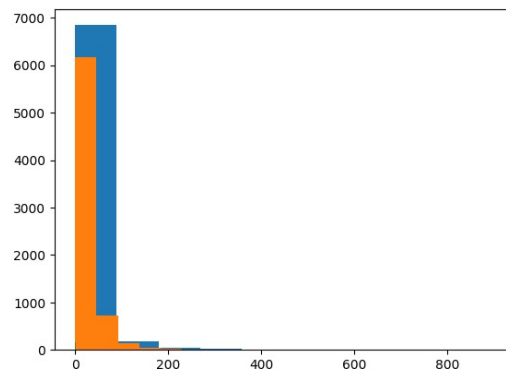
```
Node with Max out-degree 893 Node: 2565
```

## 7. The density of the network

```
8] print('density',density)
```

```
density 0.0020485375110809584
```

## Plot degree distribution of the network



## 2. Calculate the local clustering coefficient of each node and plot the clustering-coefficient distribution (lcc vs frequency of lcc) of the network.

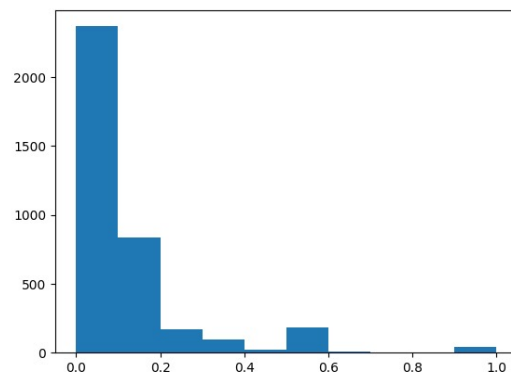
$$C_n = \frac{2T_n}{d_n(d_n - 1)}$$

We used this formula to calculate the local clustering, where,

$C_n$  is the local clustering coefficient of node  $n$ .

$T_n$  is the number of triangles a particular node is a part of.

$d_n$  is the total degree of that particular node.



## Question 2-

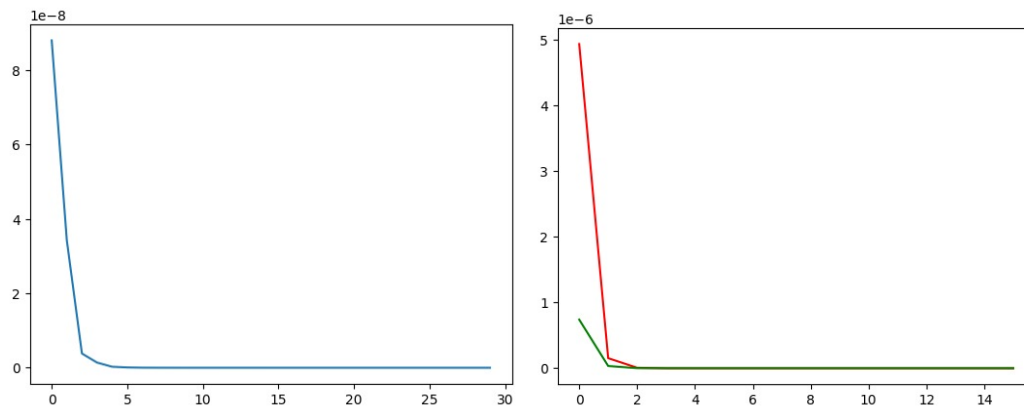
PageRank, Authority, hub score for each node:

```
top 10 rank scores :
2565    0.0043372949187308815
11      0.003017206269367328
766     0.002968177479349323
457     0.002963411320667381
4037    0.002878218886740526
1549    0.0028581648714845506
1166    0.002669208905008099
2688    0.0023843472728713416
15      0.002163159726354969
1374    0.002131987766043142
```

```
top 10 authority scores :
2565    0.15769611748358103
766     0.13015243025685455
1549    0.12938941353080033
1166    0.11950594168986171
2688    0.11008403659853248
457     0.10999186611635883
3352    0.09179709631226124
11      0.08956574261869124
1151    0.08717924518500951
1374    0.08692950770481205
```

```
top 10 hub scores :
2565    0.157696117537377
766     0.13015243029945367
1549    0.12938941344572305
1166    0.11950594165584667
2688    0.11008403661789759
457     0.10999186615700852
3352    0.09179709627666102
11      0.08956574247014454
1151    0.08717924513642718
1374    0.08692950771109112
```

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



PageRank is a centrality measure based on the quality and quantity of incoming links to a node, while HITS computes hub and authority scores based on outgoing and incoming links. In the Wikipedia voting network, node 2565 has the highest PageRank score, node 2565 has the highest hub score, and node 2398 has the highest authority score. The importance of nodes is determined by the structure of incoming and outgoing links, but the relationship between centrality measures and link structure can be complex in real-world networks.