Assignment 4

Shrinidhi M 192IT024
V Venkatesh kumar 192IT025

Dataset

The dataset taken from https://api.covid19india.org/. The report used the data of 27891 patients from the mentioned link. The following analysis made for the mentioned data. In the analysis each patient is considered as separate node.

1.Network

The Total number of nodes/patients are 27891 but for some patients link from which person he affected by the disease is obtained and for some other patients the link is unknown. The node which has a link is considered as node of a Digraph. So average Indegree and Outdegree calculated.

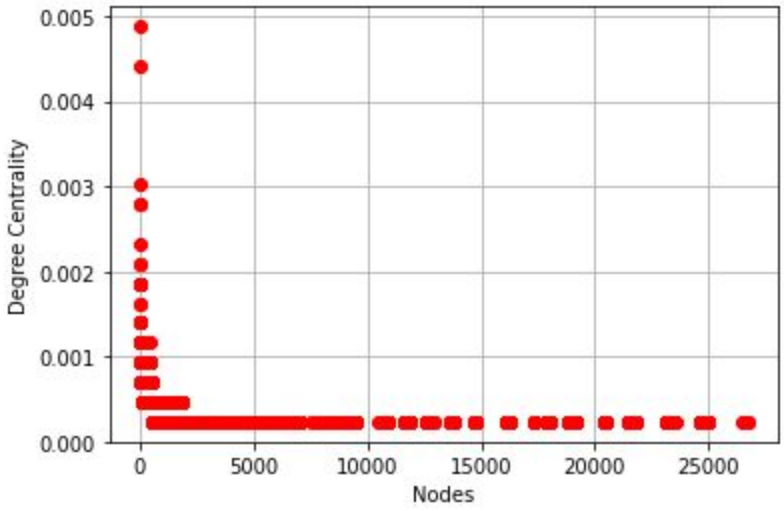
```
Name:
Type: DiGraph
Number of nodes: 4308
Number of edges: 2997
Average in degree:   0.6957
Average out degree:   0.6957
```

2.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 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 with 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 1947 0.00023218017181332715
maximum degree centrality node and value 1 0.00487578360807987
nodes in Decreasing order of Degree Centrality(top 5 nodes):

[(1, 0.00487578360807987),
 (2, 0.0044114232644533216),
 (3, 0.0030183422335732 53),
 (4, 0.0027861620617599257),
 (5, 0.0027861620617599257)]
```

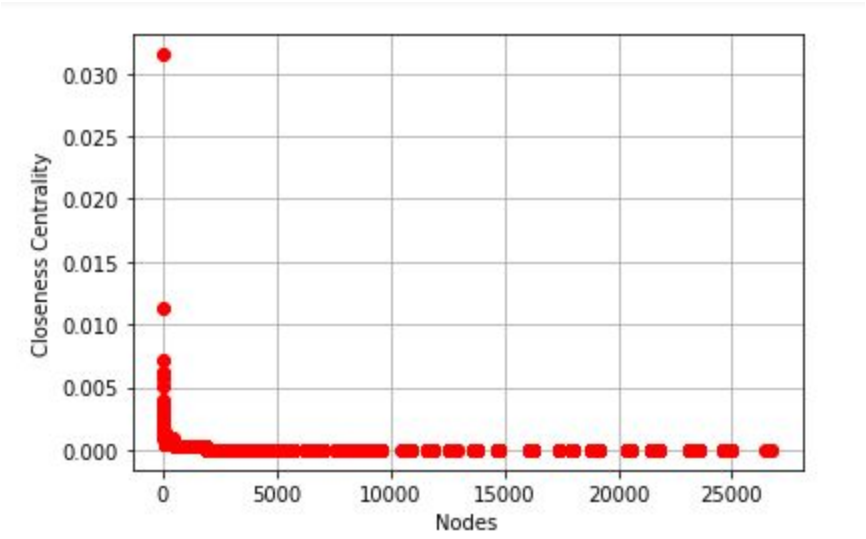


3.Closeness Centrality

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 1947 0.0
maximum Closeness centrality node and value 1 0.03151788258382824
nodes in Decreasing order of Closeness Centrality(top 5 nodes):

[(1, 0.03151788258382824),
 (2, 0.011378779512733814),
 (3, 0.007122372031118683),
 (4, 0.006211352119336349),
 (5, 0.005747057654884521)]
```

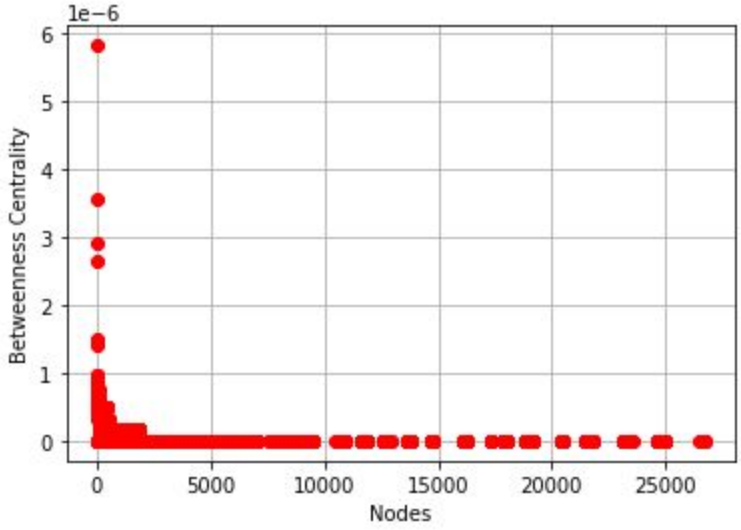


4.Betweenness centrality

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 1 0.0
maximum Betweenness centrality node and value 2 5.823376348313825e-06
nodes in Decreasing order of Betweenness Centrality(top 5 nodes):

[(2, 5.823376348313825e-06),
 (3, 3.5587299906362264e-06),
 (4, 2.9116881741569125e-06),
 (5, 2.6420874172905315e-06),
 (29, 1.5097642384517324e-06)]
```

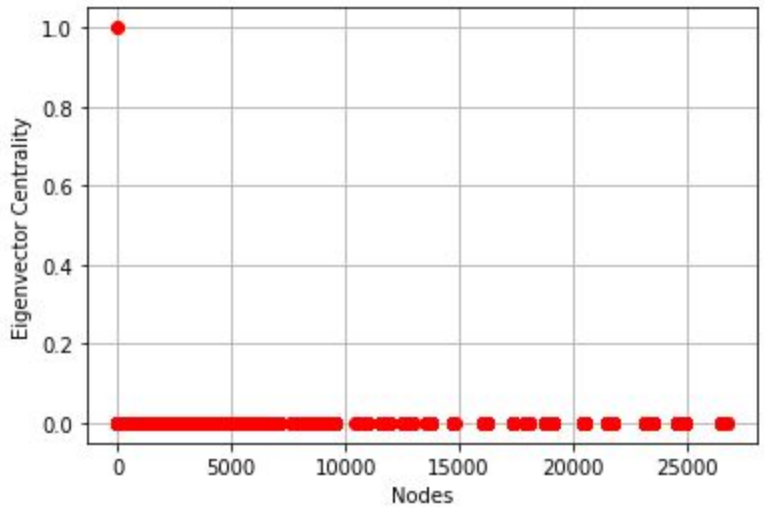


5.EigenVector Centrality

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 1947 1.9651461524775854e-08
maximum degree centrality node and value 1 0.9999995796382716
nodes in Decreasing order of Eigenvector Centrality(top 5 nodes):

[(1, 0.9999995796382716),
 (2, 0.000674182690530485),
 (3, 0.0003835375745790499),
 (4, 0.0002469009625972835),
 (5, 0.00021215717862147984)]
```

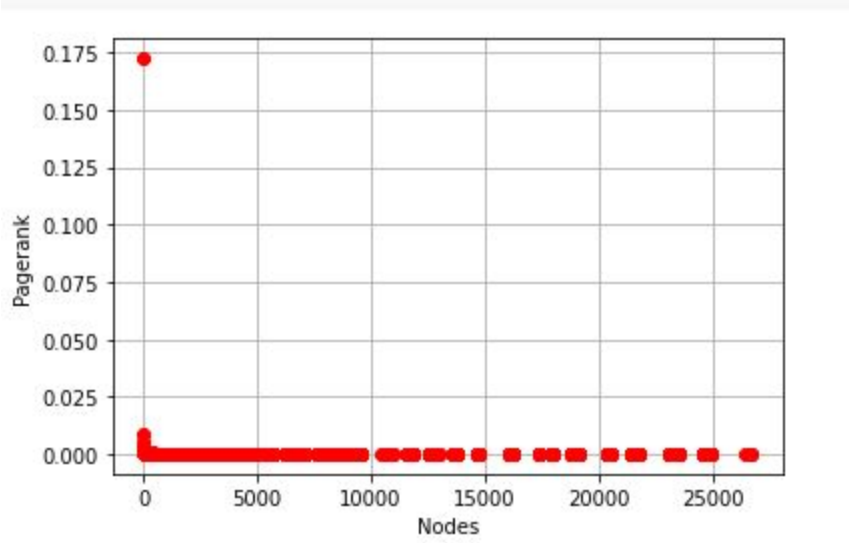


6.Page Rank Algorithm

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.

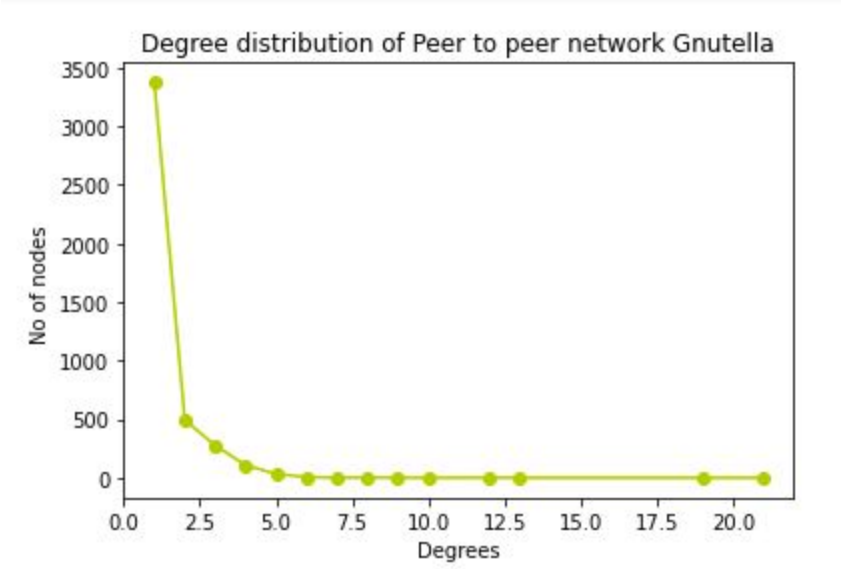
```
maximum page rank webpage 1 0.17241724832870342
minimum page rank webpage 1947 0.00011032272655598561
nodes in Decreasing order of Page Rank (top 5 nodes):

[(1, 0.17241724832870342),
 (2, 0.008534903029093573),
 (3, 0.005301429699335947),
 (4, 0.004444840172267846),
 (5, 0.00406814439042213)]
```



7. Degree Distribution

In the study of graphs and networks, the degree of a node in a network is the number of connections it has to other nodes and the degree distribution is the probability distribution of these degrees over the whole network



8.Diameter

It is the shortest distance between the 2 most distant nodes within the network. In other words, once the shortest path length from every node to all or any other nodes is calculated, the diameter is that the longest of all the calculated path lengths.

```
NetworkXError: Found infinite path length because the digraph is not strongly connected
```

9.Geodesic Path length

A shortest path, or geodesic path, between two nodes in a very graph could be a path with the minimum number of edges. The length of a geodesic path is termed geodesic distance or shortest distance. Geodesic paths aren't necessarily unique, but the geodesic distance is well defined since all geodesic paths have the identical length.

```
Example shortest path:
2
Avg. shortest path length:
1.75
```

10.Clustering coefficient & average clustering coefficient

A clustering coefficient could be a measure of the degree to which nodes during a graph tend to cluster together. Evidence suggests that in most real-world networks, and specifically social networks, nodes tend to make integrated groups characterized by a comparatively high density of ties; this likelihood tends to be greater than the typical probability of a tie randomly established between two nodes.

```
Avg clustering coefficient:
0.0
```

11.Summary

Each patient's status of 27891 is considered and classified as Deceased, Recovered with respect to status. Figure 1 shows the number of deceased people in different states. Figure 2 shows the number of people who recover from Covid with respect to number of different states.

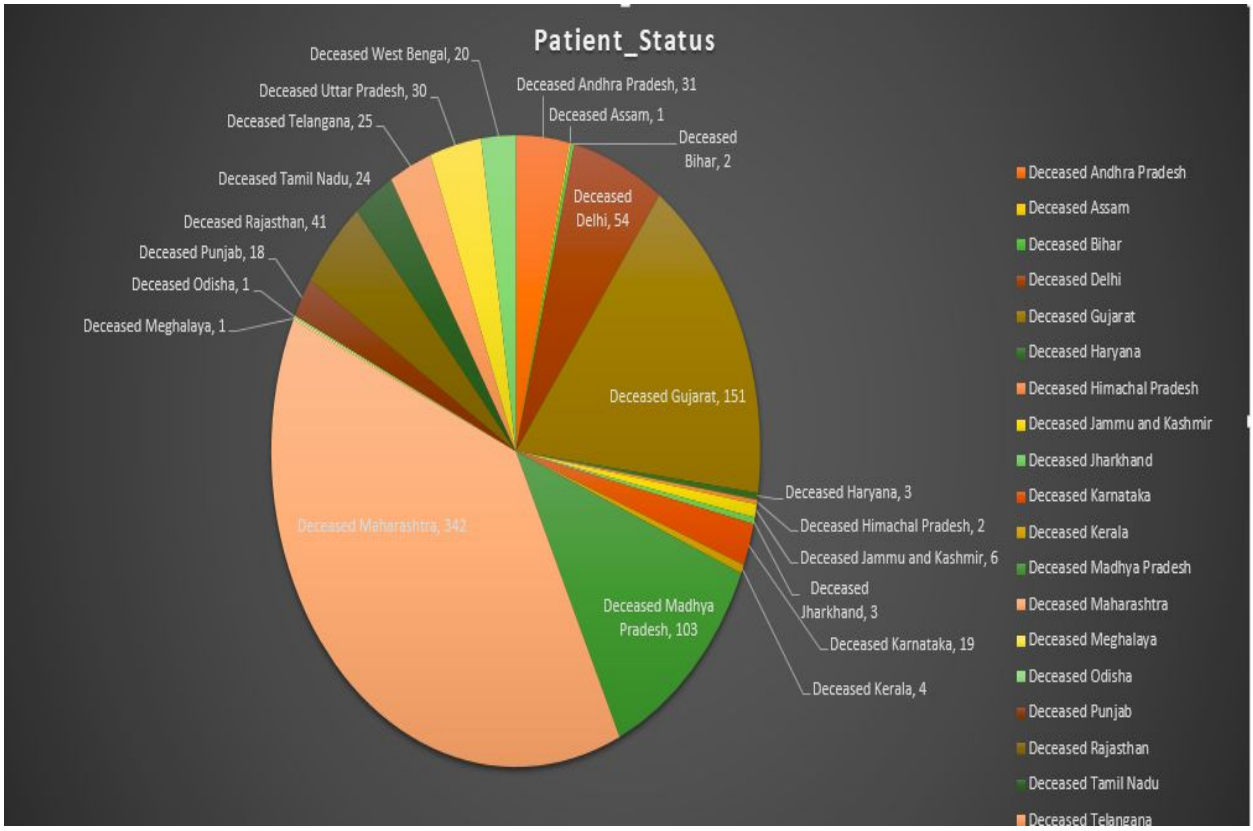


Figure1: State wise deceased candidates

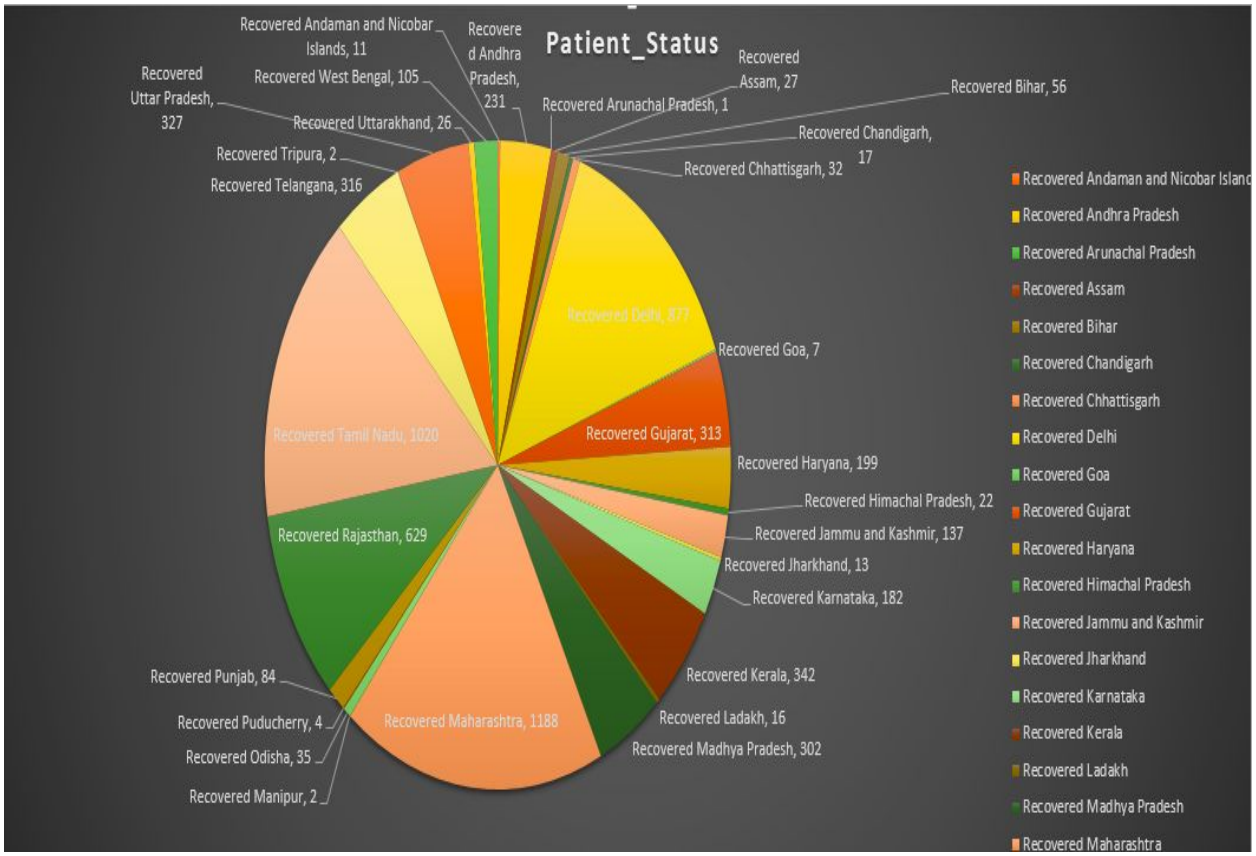


Figure1: State wise recovered candidates