NIRANJAN (192IT014)
SHIVANSH (192IT023)

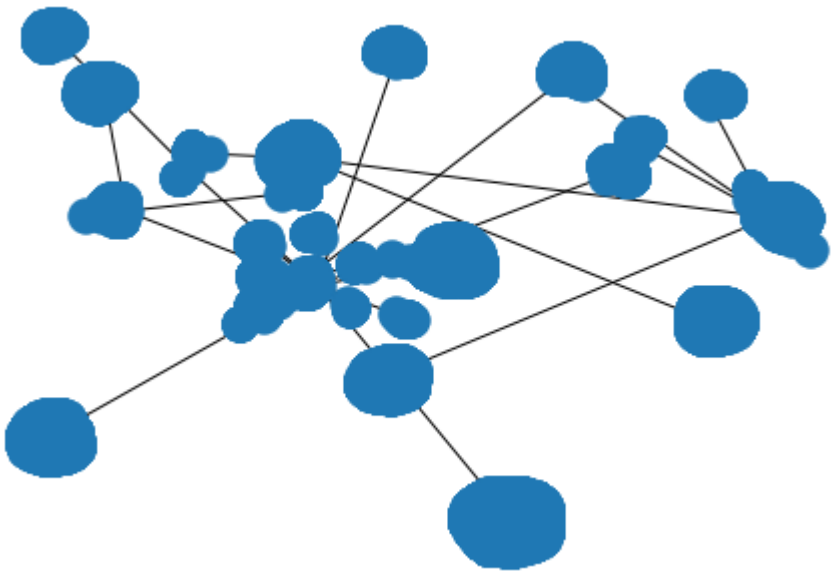
LAB ASSIGNMENT- 4

This assignment is to demonstrate the usefulness of all the concepts studied by applying it to large scale COVID19 graph data of India.

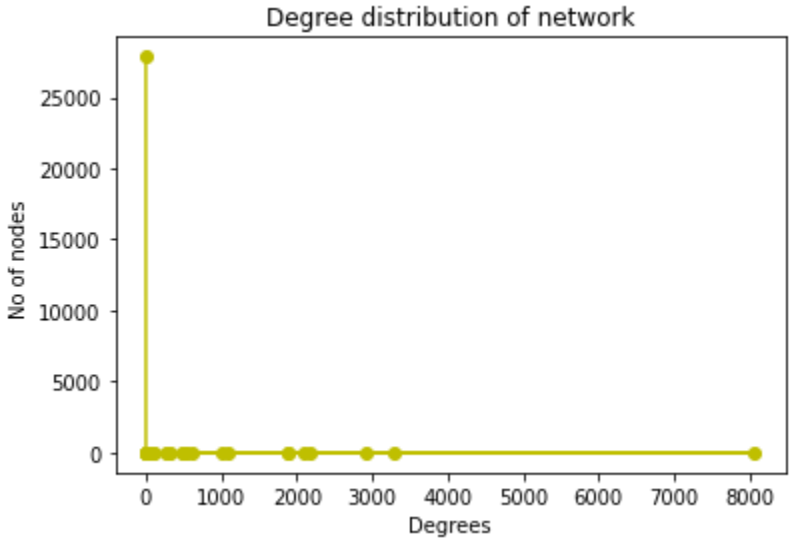
DATASET:

The dataset contains two columns: indian states and patients. The dataset taken is from 30th Jan-2020 (first case of India) to 26th April-2020.
The graph contains 27890 nodes and 27890 edges.

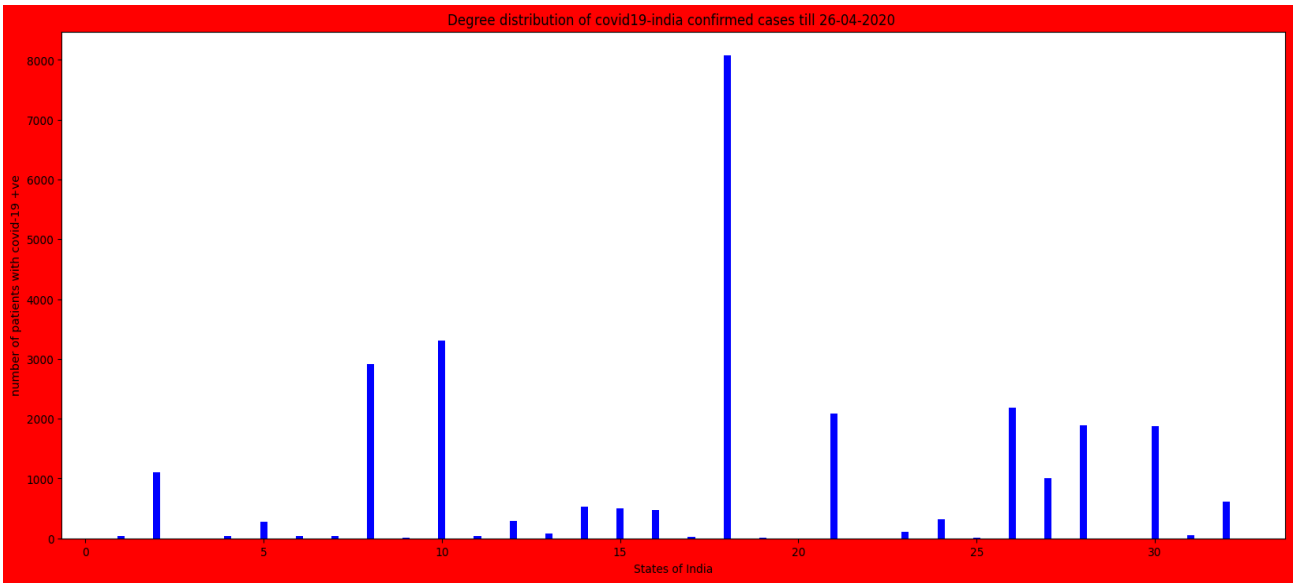
average degree: 2.0



Following is the plot of the **degree distribution** of the network



The degree distribution clearly captures only a small amount of information about a network. But that information still gives important clues into structure of a network. The network follows power law degree distribution. Most nodes have relatively small degree and very few nodes have high degree.



we can visualize in the above graph that only in 3 states of India the patients of Covid19 are above 2000 and in only one state the cases are above 3000.

***Average Clustering Coefficient**: 0.0*

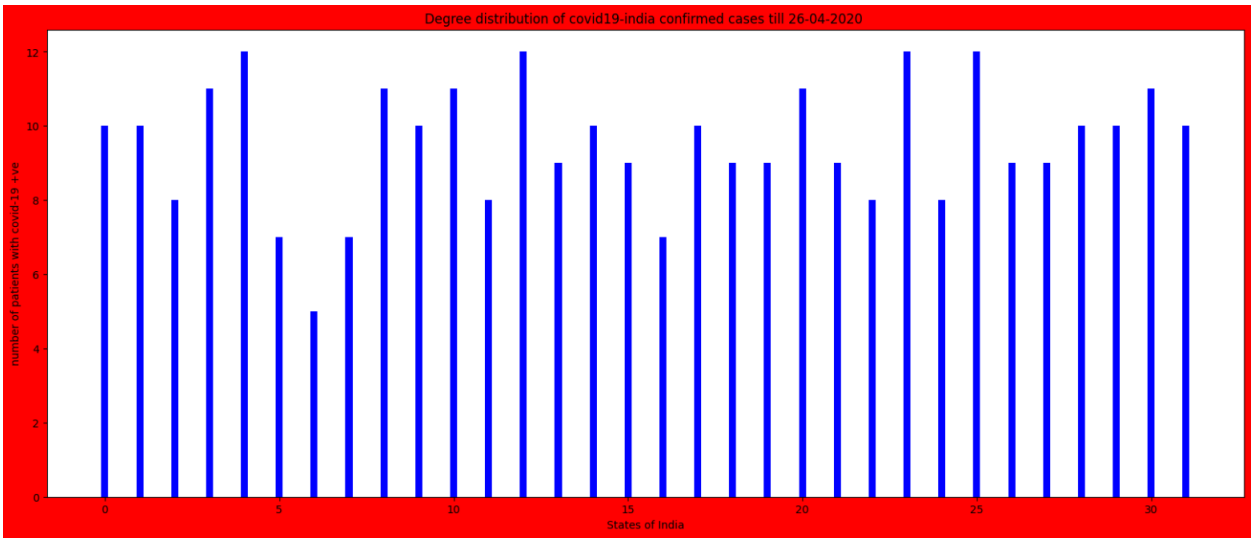
A measure of how connected node's neighbors are to each other. This means that the triangle is not formed between the nodes or the nodes are not well connected.

Is the network strongly connected: False
Number of **strongly connected components**: 27890
Largest strongly connected component: {12}

Density: $3.58564308508731e^{-05}$
Sparseness (taken as 1/density): 27889.0

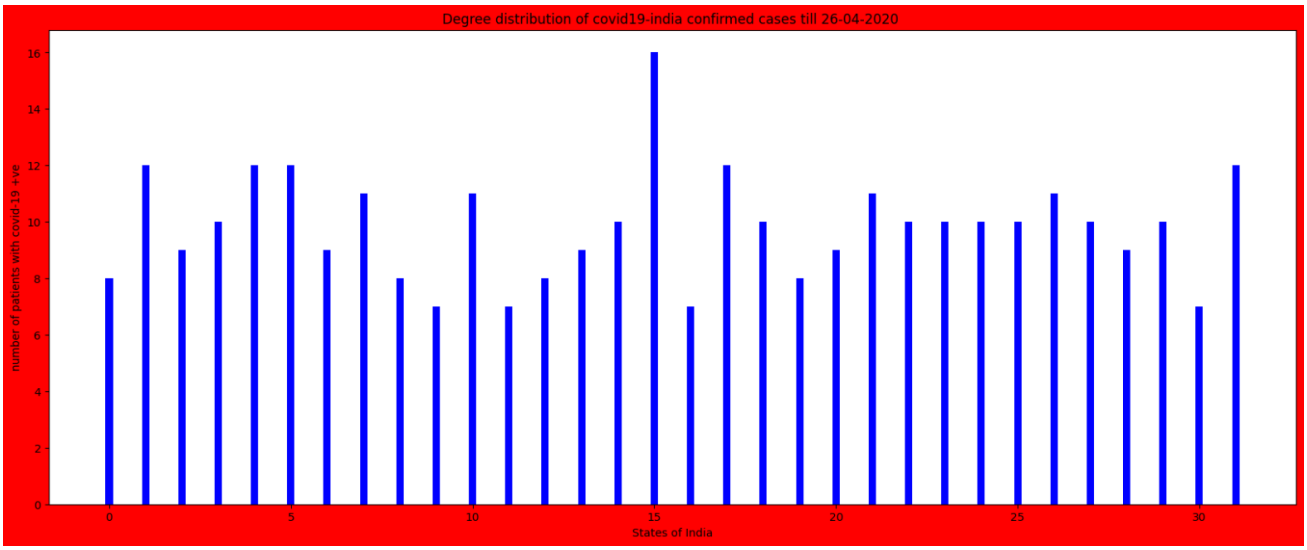
Network Density is the measure of network health and effectiveness. The network density $3.58564308508731e^{-05}$ signifies that the network is not dense.
Connected components are: 2

Erdos –Renyl Model



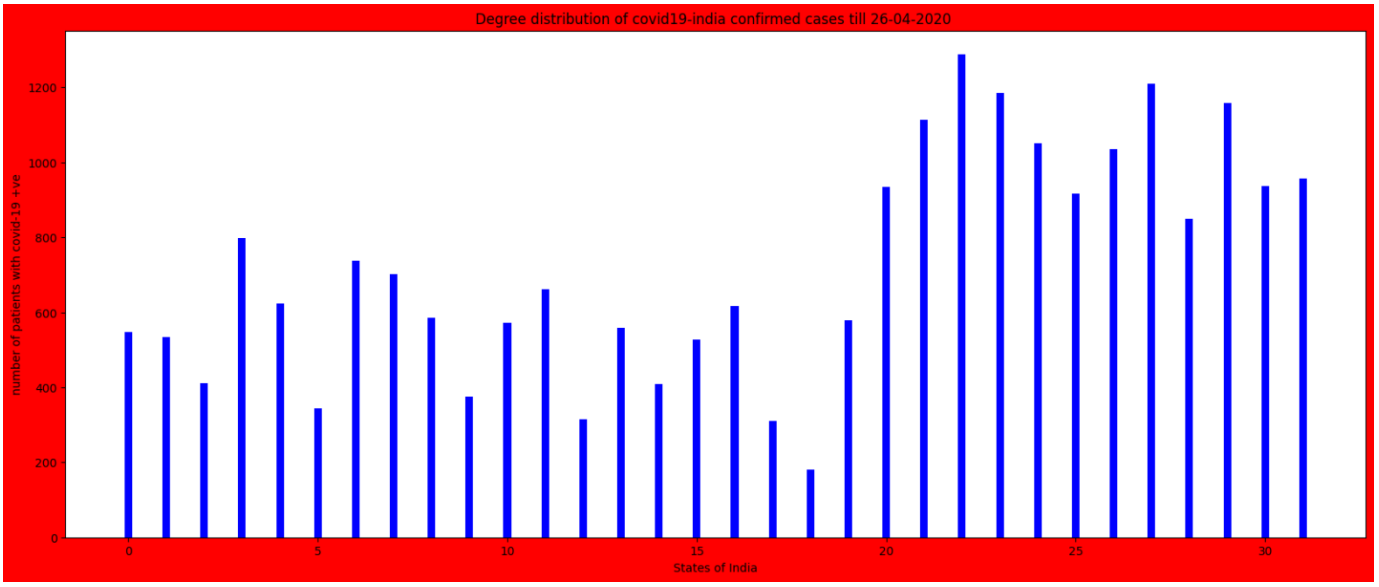
average clustering coefficient: $4.780686028445082e^{-05}$

Watts - Strogatz Model



average clustering coefficient: 0.017950920115204216

Albert – Barabasi Model

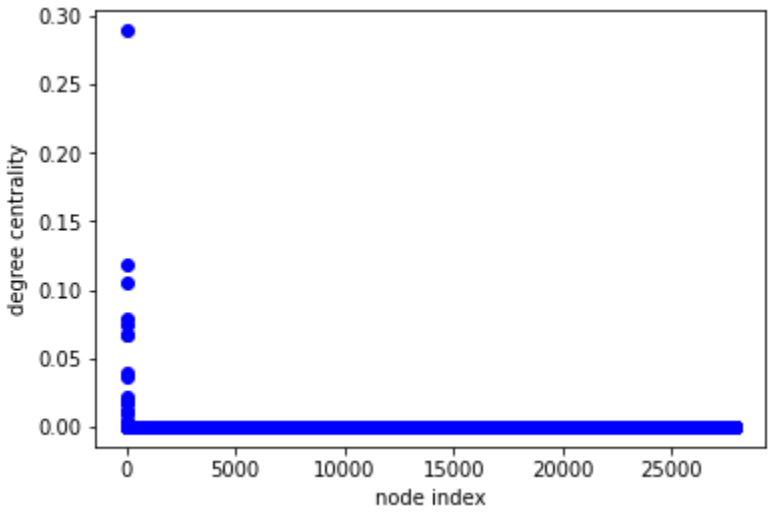


average clustering coefficient: 0.00819838510911233

The Albert- Barabasi Model gives us the similar degree distribution as followed by our network and also approaximately same average clustering coefficient. This is because our network follows pareto distribution (power-law degree distribution).

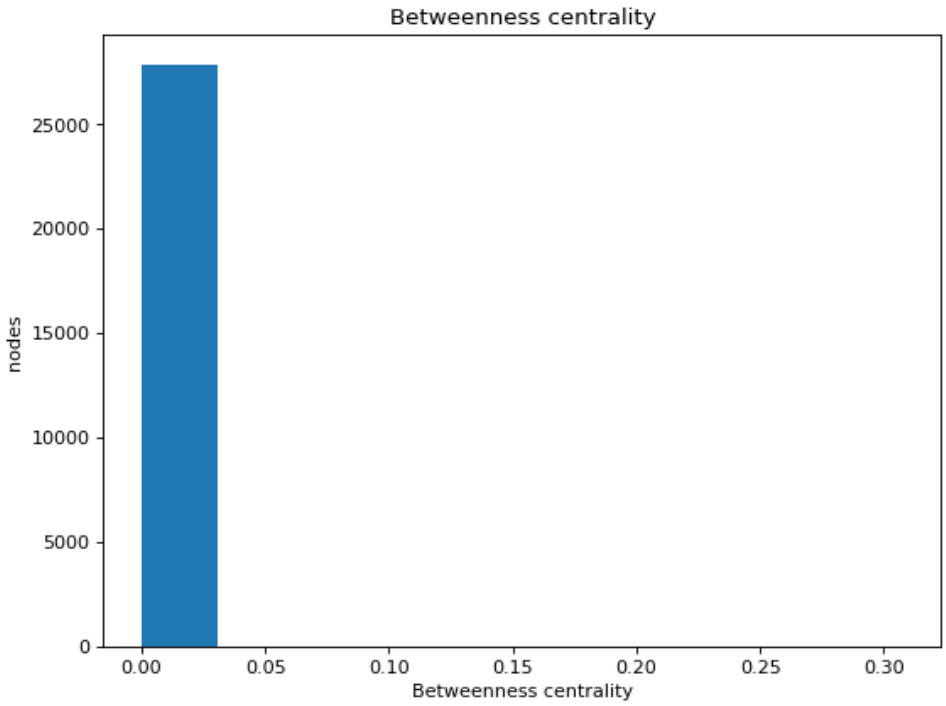
Degree Centrality:
It makes the assumption that important nodes have many connections. Degree centrality is based on number of neighbours. The node with higher degree centrality has more number of neighbours or friends. It assigns a score based on number of links held by each node.

index of max degree centrality [18]
index of min degree centrality [681]



As we can see, few nodes have higher degree centrality and majority of the nodes have lower centrality. i.e. only few nodes are central in the network means having large connections in the network.

Betweeness Centrality:
It makes the assumption that important nodes are those who connect other nodes. Betweenness centrality measures the extent to which a vertex lies on paths between other vertices. Vertices with high betweenness may have considerable influence within a network by virtue of their control over information passing between others.
node - 12 has maximum betweenness centrality of : 0.30772382903810114

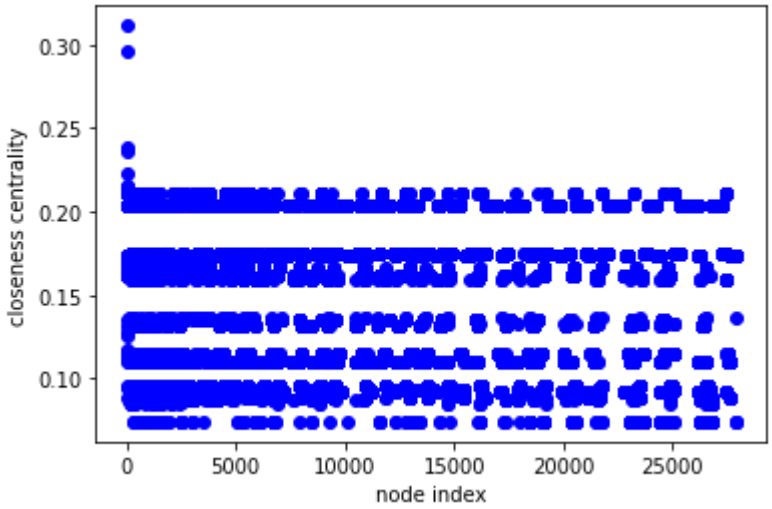


As we can observe that maximum nodes have less betweenness centrality i.e. maximum nodes does not occur in shortest path. Only one or two nodes will be there that occurs in all the shortest paths and there betweenness centrality is maximum.

Closeness Centrality:

Closeness centrality is a way of detecting nodes that are able to spread information very efficiently through a graph. The closeness centrality of a node measures its average farness (inverse distance) to all other nodes.

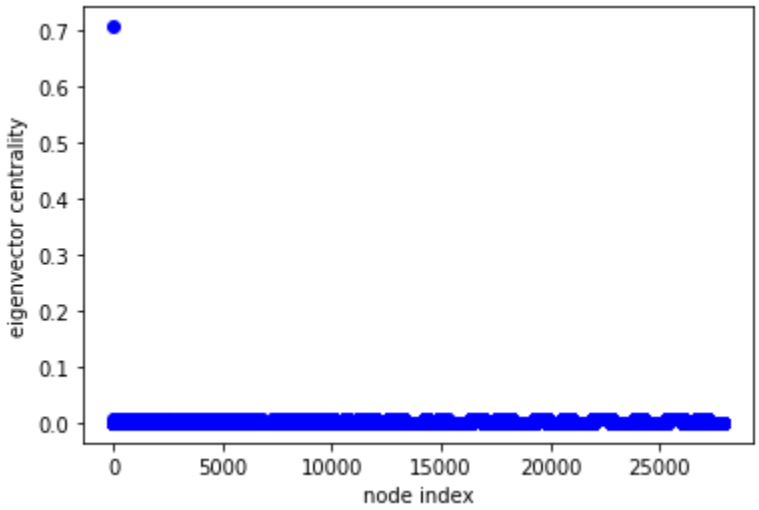
index of max closeness centrality [12]
index of min closeness centrality [348]



We can observe that very few nodes have higher closeness centrality i.e. very few nodes are far to all other nodes in the network (or the nodes or patients are close to each other).

Eigen Vector Centrality:
Eigen centrality is important as it can identify nodes with influence on the whole network not just the directly connected nodes.

index of max eigenvector centrality [18]
index of min eigenvector centrality [348]

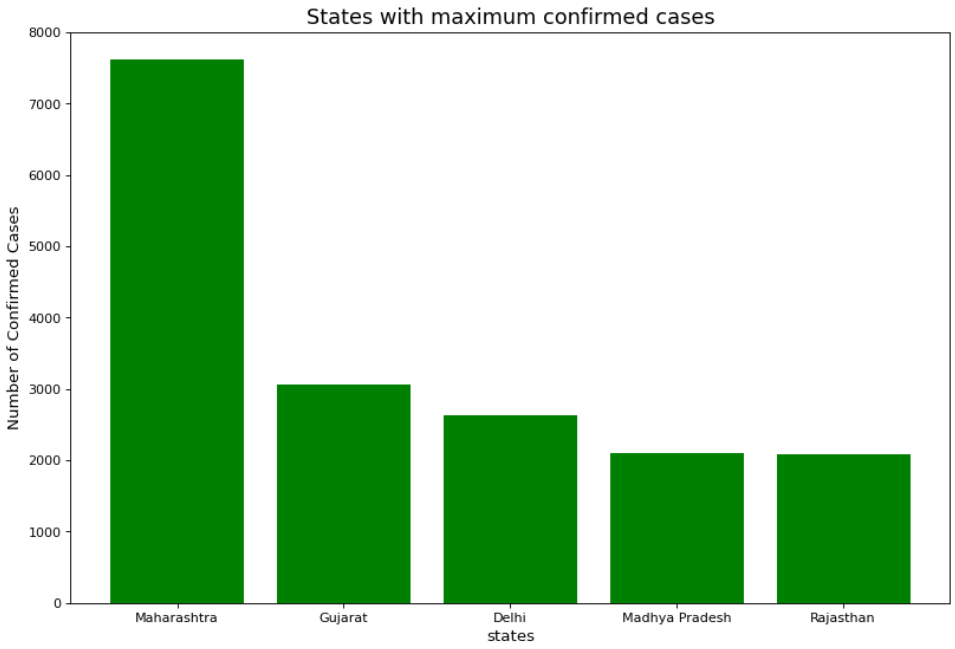


As we can observe that most of the nodes have less value of eigen centrality. These nodes might have higher number of connections but not well connected among themselves. Hence have less value of eigen centrality. As we can see in the graph that only one node have high eigen centrality i.e. the connection of this node is well connected. The node having higher eigen centrality does not mean that the node has high number of connections but it has connections with vertices having more benefit. This means there is no transmission from one state to another till 26th April-2020.

ADDITIONAL OBSERVATIONS

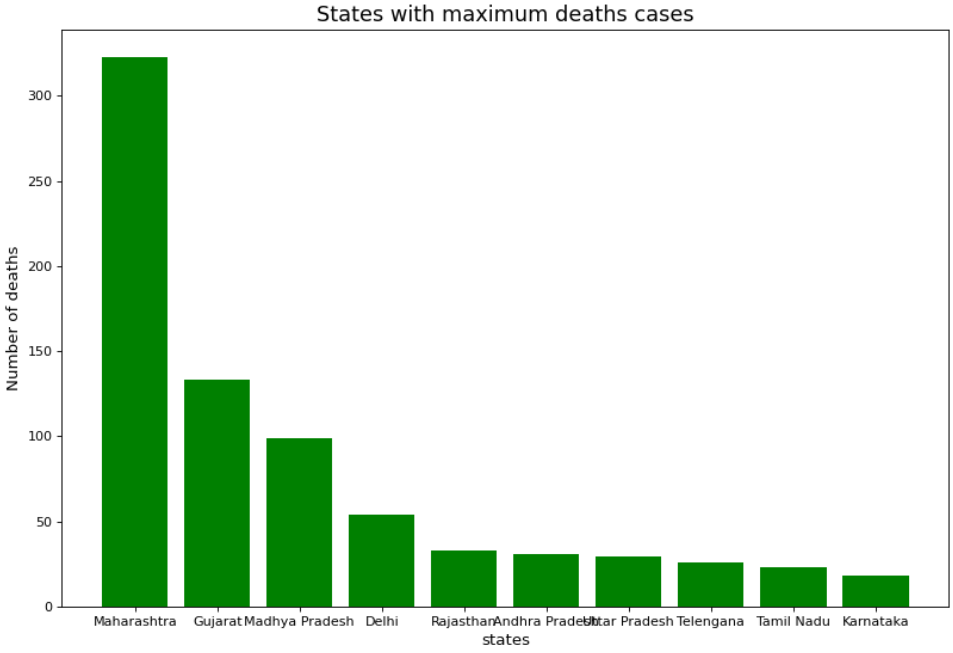
Total Confirmed Cases: 26605

Following is a plot of states with maximum no. of confirmed cases:



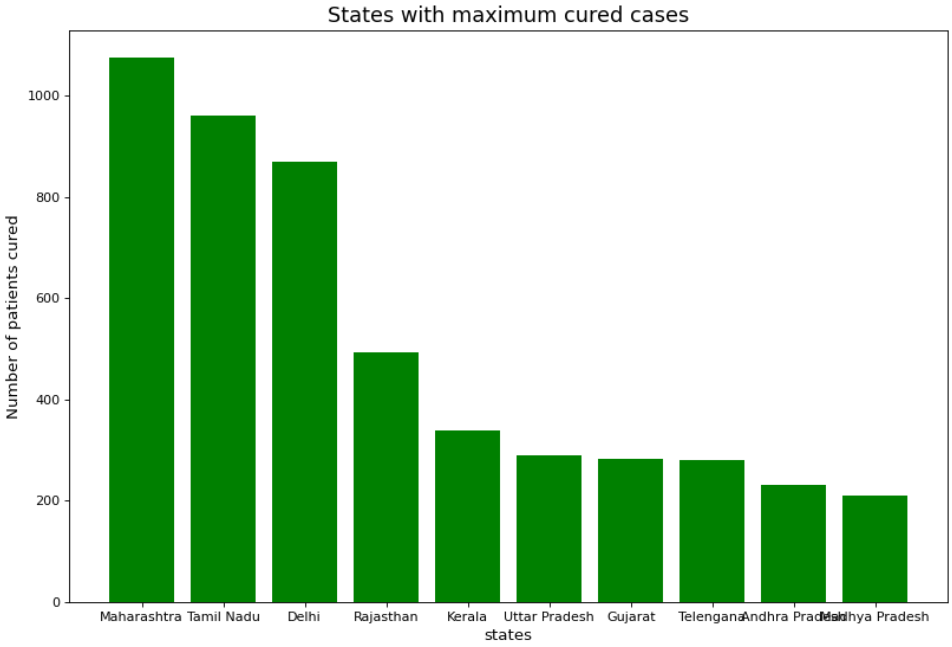
Total Deaths: 826

Following is the plot of States with maximum death cases:



Total Cured: 5914

Following is the plot of States with maximum cured cases:



Observation: The above graph tells that Maharashtra is the worst affected state because of this pandemic and Tamil Nadu has fastly cured covid patients as compared to other states.