REPORT

On

ASSIGNMENT 4

Under the guidance of
Dr. Sowmya Kamath

Submitted By :

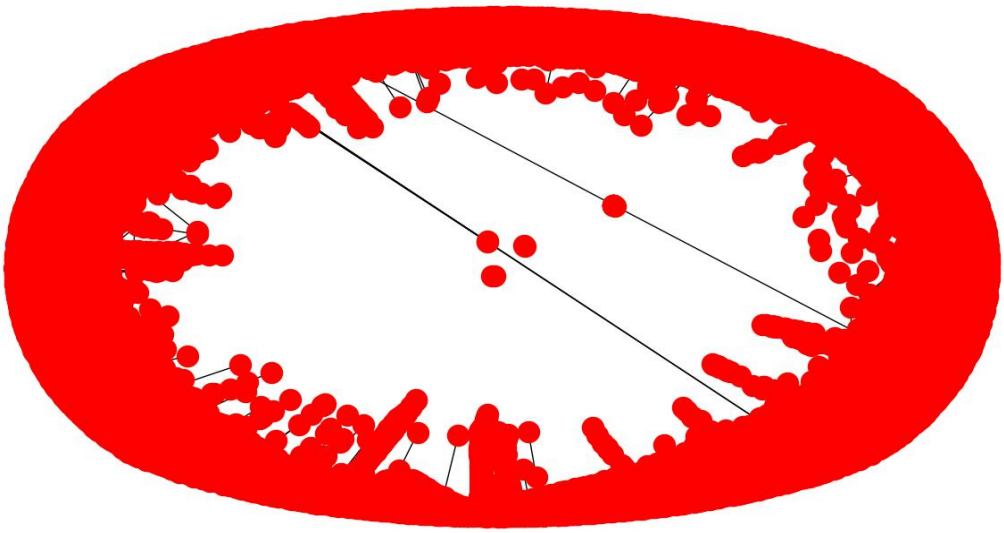
PRAJWAL M P(192IT015)

BAVYA BALAKRISHNAN (192IT003)

The graph is constructed based on the data collected from COVID19-India API, A volunteer-driven, crowdsourced database for COVID-19 stats & patient tracing in India. We used the daily data from January 30, 2020 to June 04, 2020.

1. If patient level information is available we have taken each patient as a node
2. If only districtwise information is available, one node is taken for each district
3. If only statewise information is available, one node is created for each state
4. Each vertex has the attributes such as age,gender,State,Current status,Type of Transmission and weight.
5. Weight of each vertex indicate the number of patients.
6. In the Covid graph a directed edge between patient x and patient y indicates that patient x contracted disease from patient y.
7. An *outbreak* begins when one or more nodes are infected from outside the population. These are called *imported infections*. IMP node represent such transmission.

Covid Graph



Properties of Covid graph are given below taking both directed and undirected graph into consideration

Properties of Graph(Directed)

E0 is the node representing outside countries

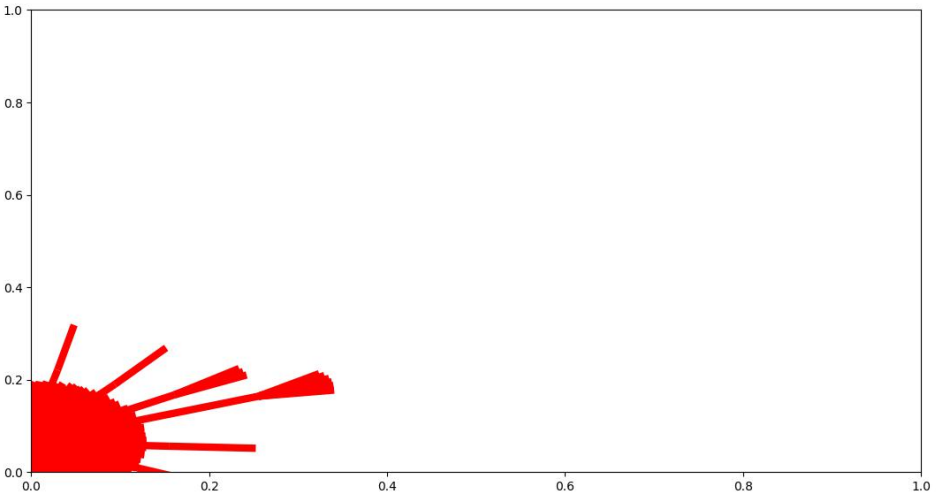
```
Name:
Type: DiGraph
Number of nodes: 30128
Number of edges: 2458
Average in degree:   0.0816
Average out degree:   0.0816
('Is graph strongly connected : ', False)
('Number of strongly connected components : ', 30125)
('Patient with max degree:', u'E0')
('No of people affected:', 962)
```

```
('Density : ', 2.7080438253066443e-06)
('Average clustering coefficient : ', 0.0002897405745254920 4)
```

Properties of Graph(UnDirected)

```
Name:
Type: Graph
Number of nodes: 30129
Number of edges: 30189
Average degree:   2.0040
('Density : ', 6.651562910393433e-05)
('Average clustering coefficient : ', 0.0005761428544328831)
('Number of connected components : ', 17)
('Is graph connected : ', False)
```

A giant component is a connected component of a network that contains a significant proportion of the entire nodes in the network. Typically as the network expands the giant component will continue to have a significant fraction of the nodes.
Giant component present in the undirected Covid graph is given below.

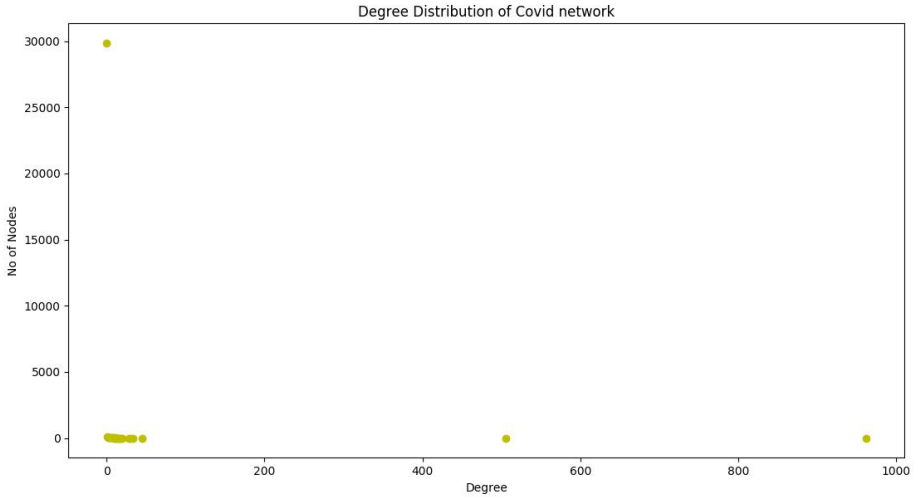


The properties of giant component found in Covid graph are as given below

```
Name:
Type: Graph
Number of nodes: 1122
Number of edges: 1128
Average degree:   2.0107
('Density : ', 0.0017936620759730379)
('Average clustering coefficient : ', 0.0)
('Diameter: ', 11)
('k value in k-connectedness : ', 1)
('average of shortest paths : ', 2.6184079340924593)
```

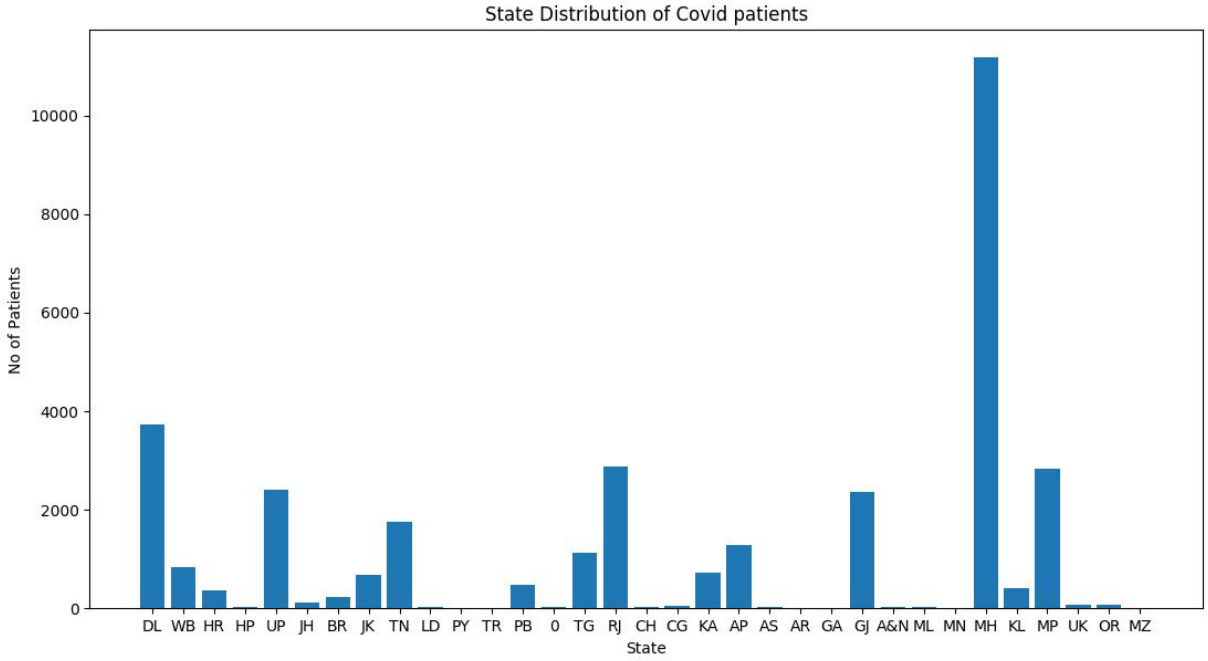
It indicates how the spread of disease among patients are interlinked. The nodes outside the giant component are mostly the ones with no information on disease contraction. Regardleess of how small the average node degree is, giant component can result in heavy spread of the disease. If this is the case, then interventions targeted at the highly connected individuals alone could stop a disease spreading.
We found 1122 nodes part of this giant component and diameter 11 indicating that there are no 2 patients more than 11 hops away.

The patient highest indegree is expected to have spread the disease the most. By analysing the indegree distribution of Covid graph we can estimate the level of community spread. As per the plot below, around 30000 nodes didnt cause any spread of disease. According to the sample data collected, around 962 patients are infected from outside India. 27734 nodes are isolated as we couldn't trace the sourec and this can be a due to 2 reasons. Either because of Community spread or because of inadequacy of data.

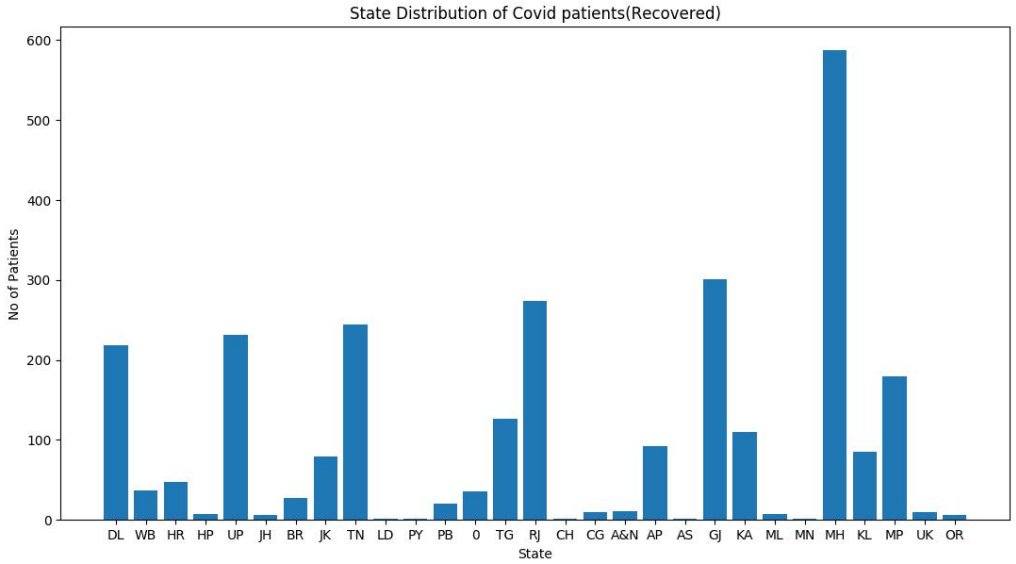


We can also observe that the number of nodes with very high degree are very less and network of Covid 19 seem to follow a power law, whereby there is steadily decreasing curve of connectedness between entities. We can call it a small world network. We observe the spread of disease from a small set of initiators to a much larger population. It also suggest the importance of treating the hubs of Corona first to prevent the community spread.

Statewise distribution of Corona patients

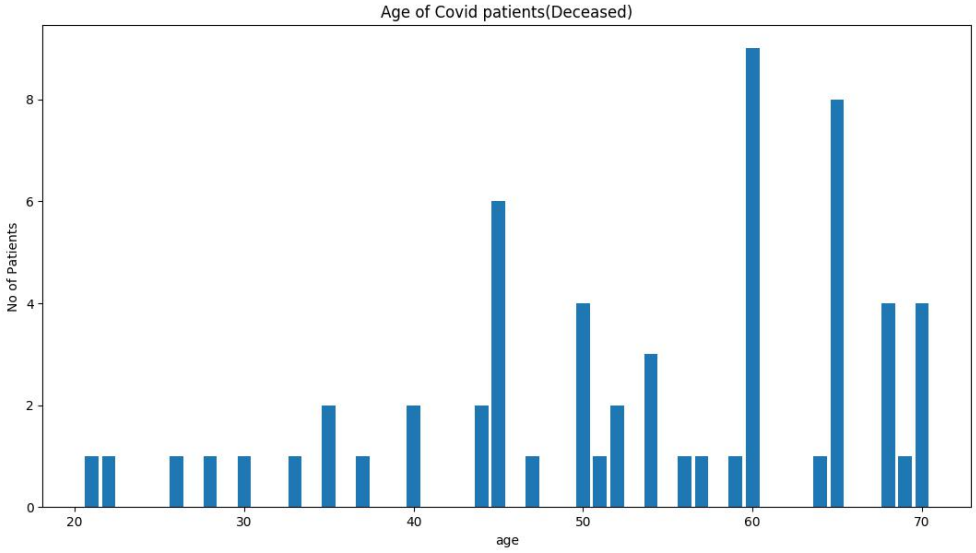


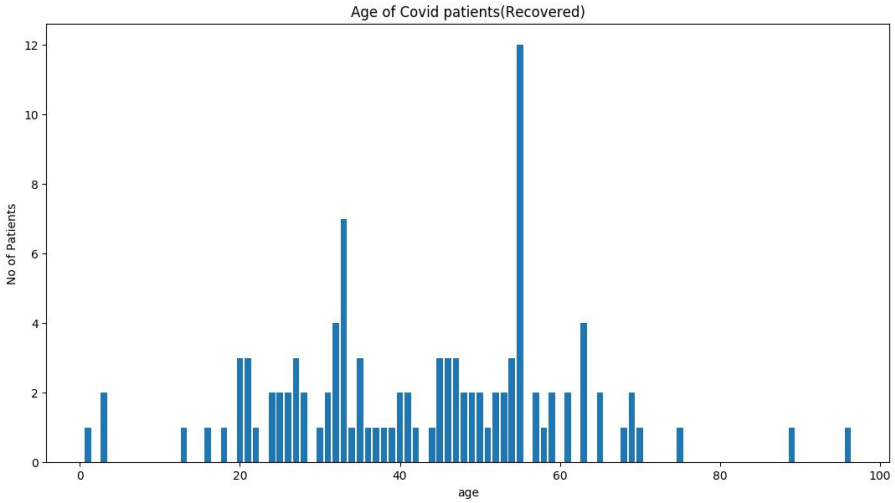
Maharastra reported the most number of cases till June 4, followed by Delhi. In the number of recovered cases also MH is at the top.



Age

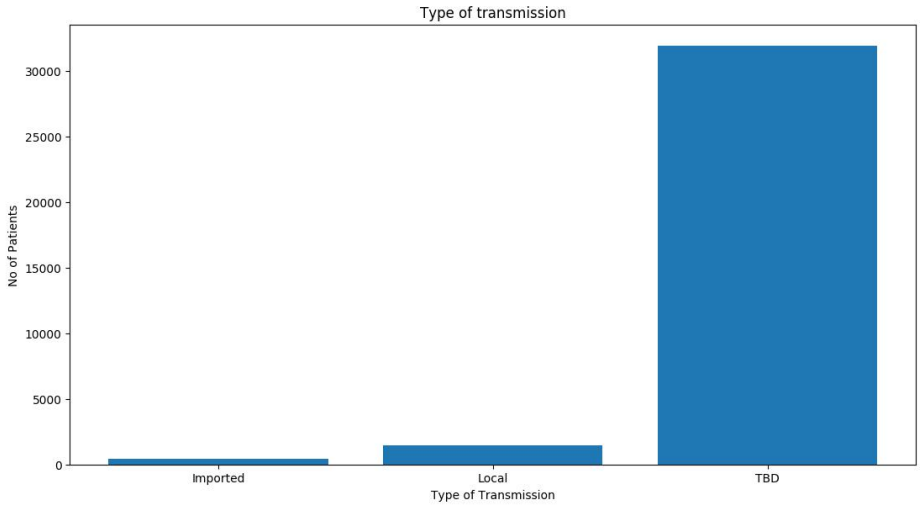
When we consider the age of patients highest mortality rate is observed among patients above age 60. Recovery rate is high among patients below age 60





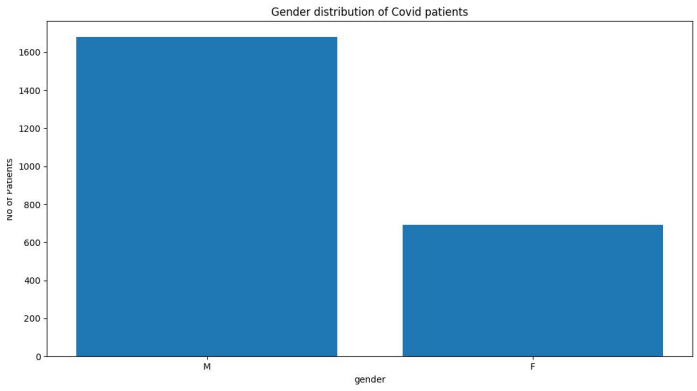
Type of transmission

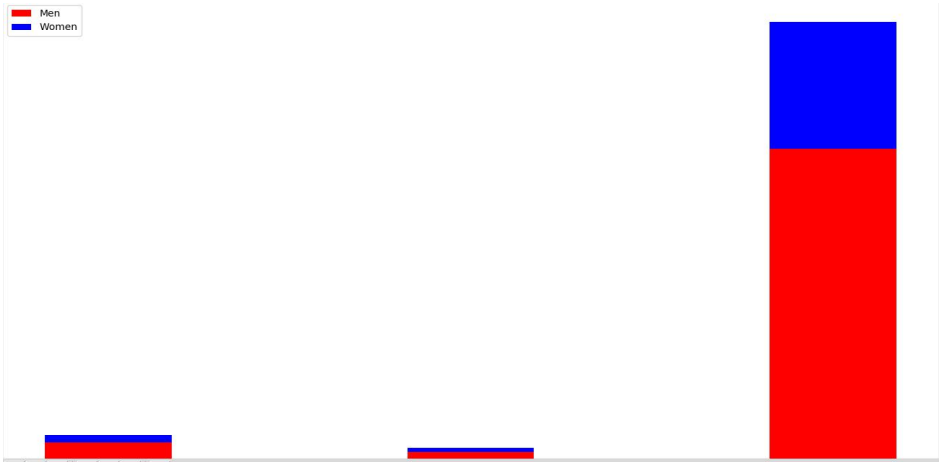
Sample data provided from the API do not have enough information on the type of transmission



Gender

We observed that Men with Covid 19 are more at risk for worse outcomes and death, independent of age





Centrality of Nodes

Measuring centrality help us to identify Super-spreaders of disease

```
('Degree centrality: Maximum value :', 0.03193149002555847)
('Degree centrality: Node with Maximum value :', u'E0')
The top 10 Node centrality pairs are  {u'P8460': 0.0009293988780827829, u'P6': 0.0005642778902645467, u'P182': 0.0006970491585620872, u'P20410
': 0.0011285557805290935, u'IMP': 0.016762372622564477, u'P689': 0.0014936767683473298, u'KA-C1': 0.0005642778902645467, u'P28995': 0.00099578
45122315532, u'P32857': 0.0006306635244133l7, u'E0': 0.03193149002555847}
('Eigenvector centrality: Maximum value :', 0.4082064740484362)
('Eigenvector centrality: Node with Maximum value :', u'P194')
The top 10 Node centrality pairs are  {u'P8460': 0.010043892321095546, u'P13641': 0.002827834197078254, u'P683': 0.4082064740484362, u'P684':
0.4082064740484362, u'P145': 0.4082064740484362, u'P146': 0.4082064740484362, u'P16136': 0.00900444150700087, u'E0': 0.0025924256711322533, u'
P194': 0.4082064740484362, u'P195': 0.4082064740484362}
('Closeness centrality: Maximum value :', 0.03072800289345486)
('Closeness centrality: Node with Maximum value :', u'E0')
The top 10 Node centrality pairs are  {u'P8460': 0.0009459952866199754, u'P182': 0.0006653651059001741, u'P20410': 0.0010527759528687049, u'IM
P': 0.017749893944047558, u'P689': 0.0016869761148393372, u'KA-C1': 0.0005642778902645467, u'P28995': 0.0009968217877651276, u'P32857': 0.0005
99130348192651, u'E0': 0.03072800289345486, u'P2892': 0.0007604172638859131}
('Betweenness centrality: Maximum value :', 8.373677546482334e-08)
('Betweenness centrality: Node with Maximum value :', u'P32857')
The top 10 Node centrality pairs are  {u'P13641': 2.203599354337456e-08, u'P30071': 2.313779322054329e-08, u'P182': 2.313779322054329e-08, u'P
20416': 5.619178353560513 6e-08, u'P30609': 3.085039096072439e-08, u'P29405': 7.602417772464224e-08, u'P21463': 2.974859128355566e-08, u'P32857
': 8.373677546482334e-08, u'P2892': 4.627558644108658e-08, u'P26467': 5.288638450409895e-08}
```

Nodes with high centralities

E0: Its among the top 10 nodes in all 4 centrality measures. Node representing an Indonesian national; The patient had flown into Delhi on March 9 and after a few days took a train to Karimnagar, about 170 km from Hyderabad, for a religious meeting with a group from Indonesia and Malaysia. This patient spread the disease to many people

P8460

P194: Showing highest eigen vector centrality.Eigen vector centrality computes the centrality of a node as a function of the centralities of its neighbours

P32857: With highest betweenness centrality reflects the amount of control that this node exerts over the interactions of other nodes in the network.

Centrality Distribution

It shows that majority of nodes have very low centralities. Though few nodes exhibit high centrality. Those are the most important nodes.