

# WSC Assignment 4

Pritish Naik  
MTech 1<sup>st</sup> Year 2<sup>nd</sup> sem  
(Reg.No: 192422IT010)

## Content of Report

1. COVID19 Graph plot built using who infected whom
  - a. Directed graph
  - b. Undirected graph
2. Giant Component plot and its analysis
3. Centrality calculation
  - a. Degree Centrality
  - b. Closeness Centrality
  - c. Betweenness Centrality
4. Gephi visualization of COVID19 Graph
5. Similarity between completely random graph and COVID19 graph

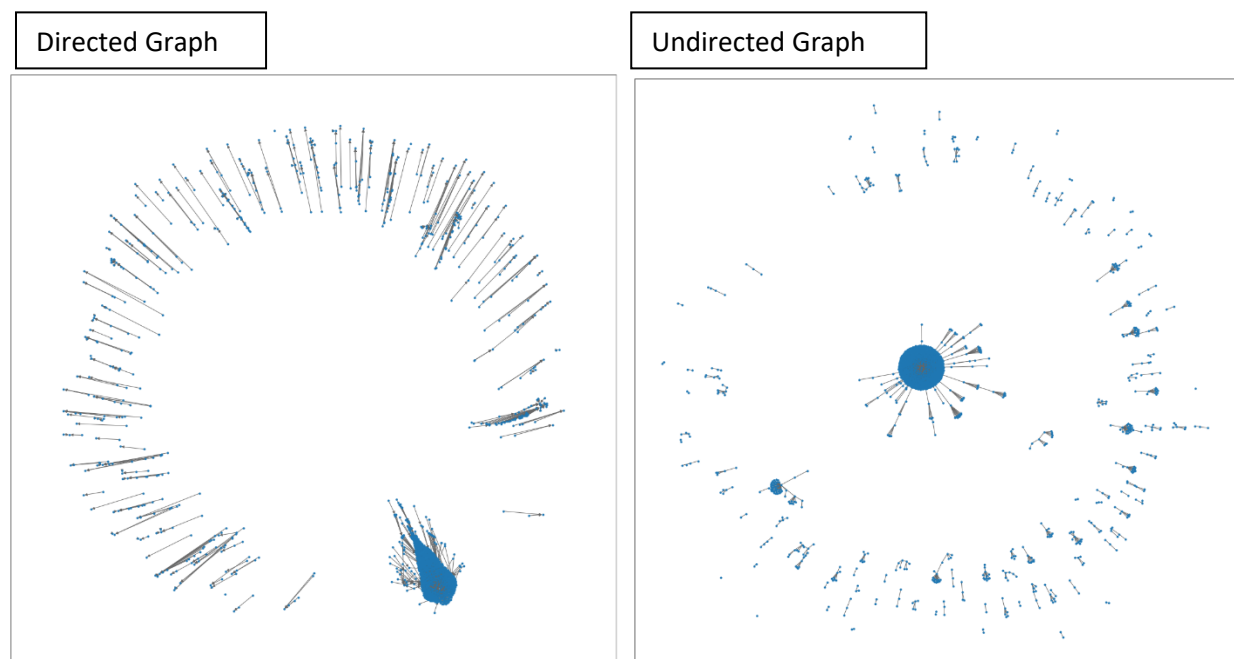
# 1. Who infected Whom?

## Dataset:

[api.covid19india.org/raw\\_data1.json](https://api.covid19india.org/raw_data1.json) (Data till EoD Apr 19th)

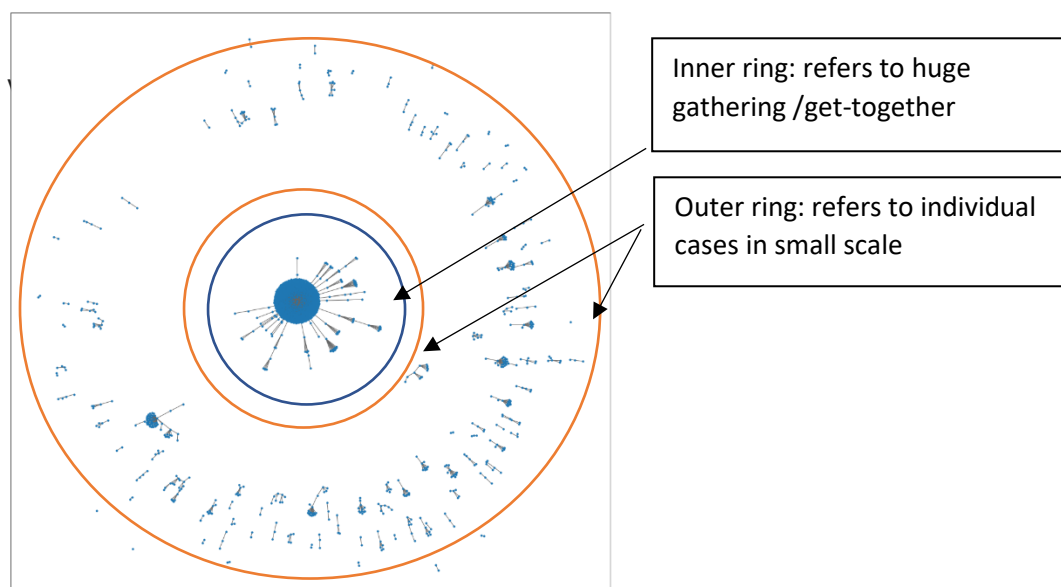
This dataset was imported into excel then converted into CSV format Imported in the colab python code

## Plot of dataset COVID19 data:

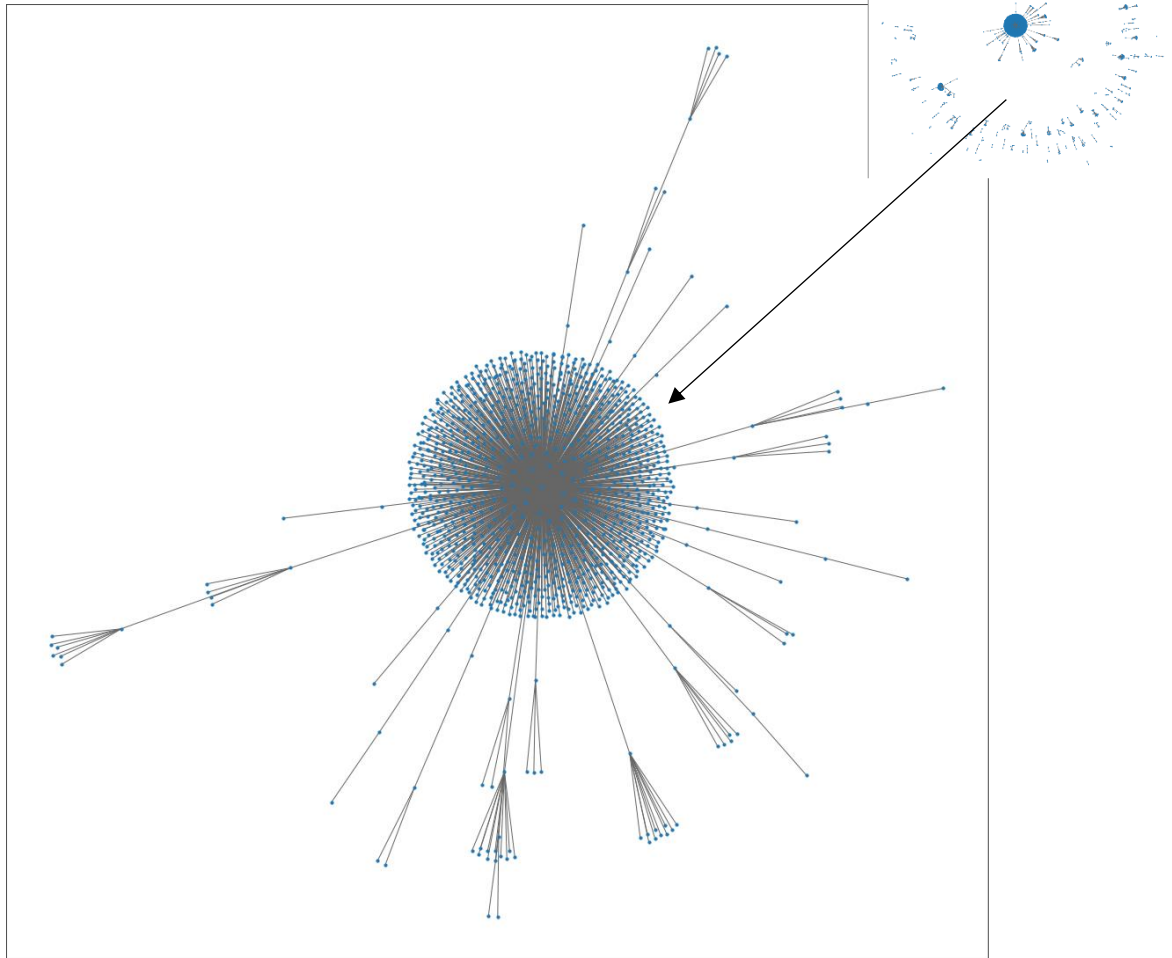


## Observation:

The COVID graph is very sparse. There is a subgraph in the centre of the graph (Giant component). Let's find out number of patients in the giant component



## 2. Giant Tablighi component:



### Analysis of giant component:

#### **Centre**

Event 90000 is the Tablighi Zamaat Event that took place in Nizamuddin is the super spreader

#### **Radius**

Radius = 3 virus can be infected in 3 successions i.e if A gets infected then A -> B -> C -> D

#### **Periphery**

Periphery patients are the ones who infected from Tablighi Zamaat event and farthest for source. Periphery can tell us the patients who got infected indirectly from Tablighi Event didnt infect others. So they broke the chain of infection.

Another interesting insight which is debatable is that there are 10 Female patients P6832, P6835, P9262, P9263, P12565, P12568, P12569, P12570, P12571, P12572

who were successful in breaking the chain of infection whereas there are 7 male patients (P12566,P12567, P12581, P12582, P12583, P12584, P14699, P14701) who were successful in breaking chain of infection. So based on data we can say that women may be better in containing coronavirus in themselves as compared to men

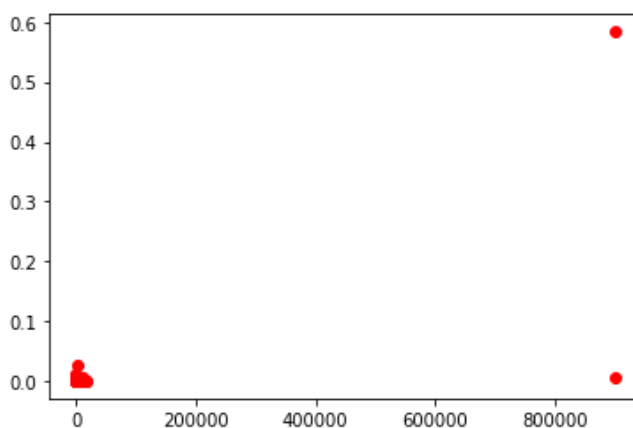
### 3. Centrality calculations of COVID19:

#### 3.1 Degree centrality

##### Degree centrality of COVID19:

Here the assumption is that important(interesting) nodes have many connections. This is most basic type of centrality measure where we count the neighbours that are directly connected to the node. This analysis will tell us about the COVID patients who were able to infect patients directly. (Directed spreaders)

Mainly used in calculation of  $R_0$  – Reproduction rate of virus



According to degree centrality, following are the rankings of super spreader of corona virus:

Rank 1: P90000 – (Pls note: P90000 number is given by me, in dataset, the Name of this event is E0)

This refers to Tablighi Zamaat event happened in Nazamuddin. As we don't know the exact patient, the event is measured as one entity

Rank 2: P689

P689 was 70 years old Female from Chikkaballapura Hospitalized on 26-03-2020. The case was registered in Karnataka India. She Travelled from Mecca, Saudi Arabia, details awaited. Infected 42 others

Rank 3: Patient no 182

P182 was 70 years old Male from Punjab Hospitalized on 19-03-2020. The case was registered in Punjab India. He travelled from Germany via Italy. Infected 20 others

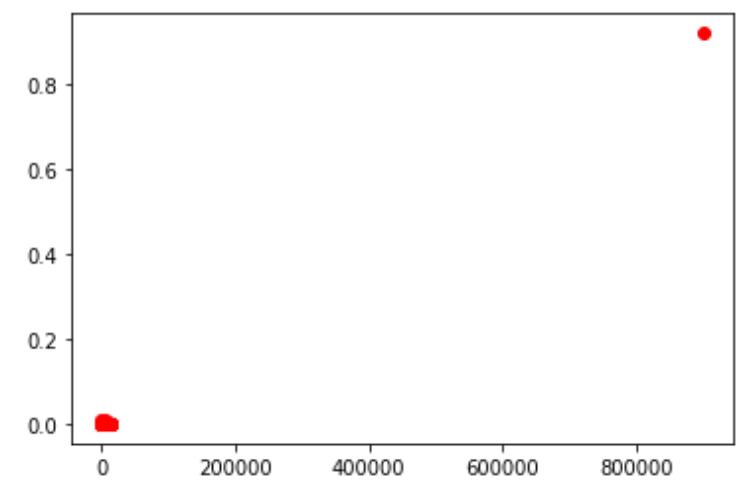
Rank 4: Patient no 6

Patient no 6 is 69 years old Italian tourist who was infected in Rajasthan. Infected 16 others who were of 55 age.

Rank 5: Patient no 301

Patient no 301 is 69 years old who travelled from Thailand. Infected 14 others.

### Degree centrality of Giant Component:



According to degree centrality, following are the ranks of node as per their importance:

Rank 1: The Tablighi Zamaat event

Rank 2: Patient no. 4862

This 50-year-old patient was in direct contact with Tablighi zamaat event and infected 11 others

Rank 3: Patient no. 531

This patient is from MP was in direct contact with Tablighi zamaat event and infected 10 others

Rank 4: Patient no. 8463

This 55-year-old patient was in direct contact with Tablighi zamaat event from Belagavi infected 6 others

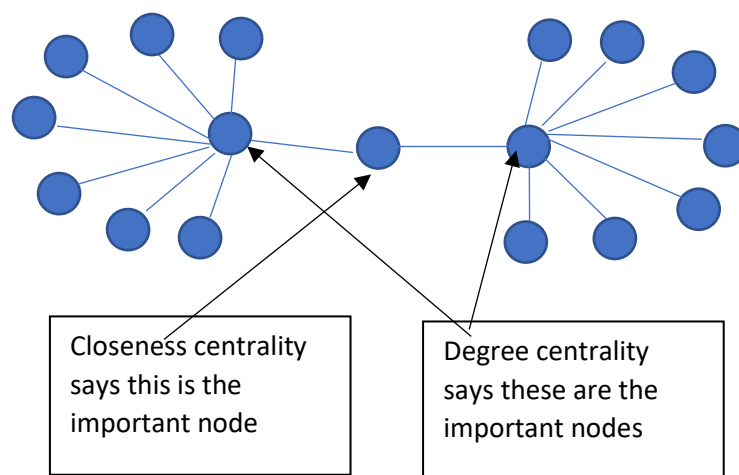
Rank 5: Patient no. 2891

This 20-year-old patient was in direct contact with Tablighi zamaat event from Belagavi infected 5 others

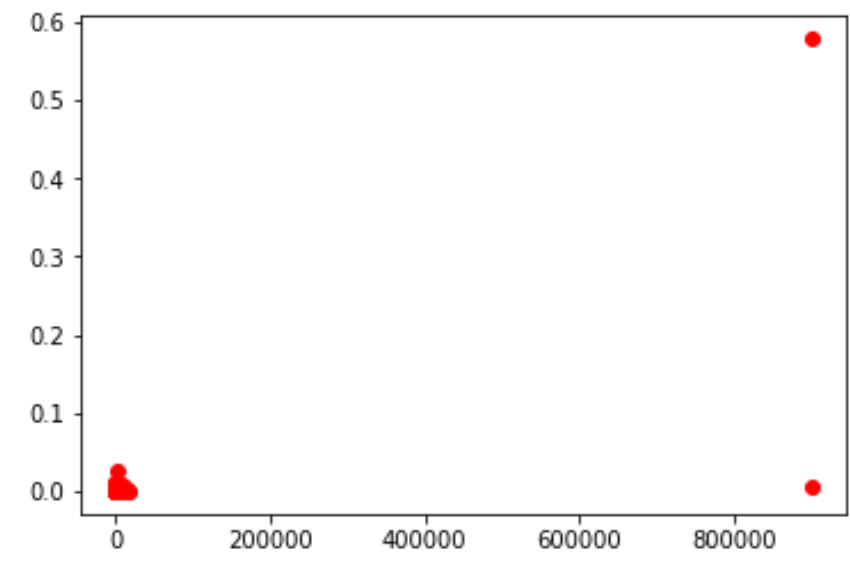
### 3.2 Closeness Centrality:

Here, we assume important nodes are close to other nodes in the Network. In other words, closeness centrality says that nodes who are central are a short distance away from all the other nodes in the network. In our COVID19 context, the closeness centrality talks about indirect spreaders

Following diagram will illustrate why do we need Closeness centrality when we have degree centrality. So the definition of important node is different



Closeness centrality plot:



According to Closeness centrality, following are the ranks of node as per their importance

Rank 1: Tablighi Zamaat event

Rank 2: P689

P689 was **70 years old** Female from Chikkaballapura Hospitalized on 26-03-2020. The case was registered in Karnataka India. She Travelled from Mecca, Saudi Arabia, details awaited. Infected 42 others

Rank 3: P182

P182 was **70 years old** Male from Punjab Hospitalized on 19-03-2020. The case was registered in Punjab India. He travelled from Germany via Italy. Infected 20 others

Rank 4: P6

Patient no 6 is **69 years old** Italian tourist who was infected in Rajasthan. Infected 16 others who were of 55 age.

Rank 5: P301

Patient no 301 is **69 years old** who travelled from Thailand. Infected 14 others.

**Interesting insight from Closeness centrality:**

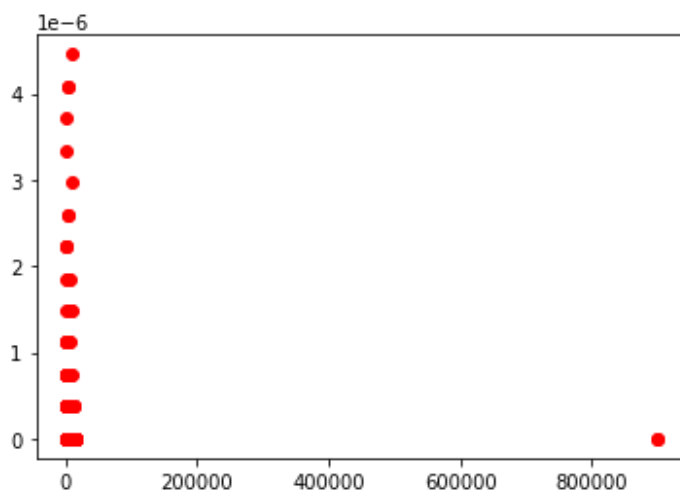
Excluding Tablighi Zamaat event, All the patients who are ranked high in closeness centrality have age 69/70. This might mean old age patients are infecting people who are close to them or few hops away from them.



### Betweenness Centrality:

Here the assumption is that important nodes are the ones who connect other nodes in the network. The metric used to calculate betweenness centrality is distance. Distance between two nodes is the shortest path between them. This measure Talks about how corona virus infected people from one city to another city. Betweenness centrality will help the government to study how COVID19 managed to spread across country even though strict lockdown was imposed.

Betweenness centrality also talks whether the officials were successful in quarantining people who were in contact with COVID19 positive patient.



According to Betweenness centrality, following are the ranks of node as per their importance.

Rank 1: P8463

This is 55-year-old patient from Belagavi. This patient got infected from the patient below P2862 (Rank 2) who is from Orissa.

Rank 2: P2892

This patient is from Orissa was infected people directly/indirectly two states which are far from each other .i.e in **Belagavi** in Karnataka and **Korda** district in Orissa.

Rank 3: P4862

This patient had attended Tablighi Zamaat and infected 11 others in **Bengaluru** city.

Rank 4: P531

This patient is from **Natipora** district Srinagar J&K . Responsible to infect 8 people in **Bandipora** and 2 in **Rajouri** district. Interesting point about this patient is, he has high

degree and Betweenness centrality but low closeness centrality. So, we can say this patient has spread virus indirectly. i.e quarantining of people who were COVID19 positive

Rank 5: P1215

This patient is from Meerut Uttar Pradesh. He has spread COVID in his city and infected 8 others in Mysuru Karnataka district

#### Interesting insight:

(Rank 1) P8463 hails from Hirebagewadi has been in contact with (Rank 2) P2862 from Belagavi who was COVID19 positive. P8463 went back to his city and infected people in his city. That's the reason why they have high betweenness centrality. So betweenness centrality takes about how COVID spreads inter-cities.

Summary of ranks and insights provided by centrality measures is given below:

Ranks	Centralities		
	Degree	Closeness	Betweenness
1	E90000	E90000	P8463
2	P689	P689	P2892
3	P182	P182	P4862
4	P6	P6	P531
5	P301	P301	P1215

Degree Centrality talks about Super direct COVID spreaders i.e who infected others by **direct** contact

Closeness Centrality talks about Super indirect COVID spreaders i.e who infected others **indirectly**

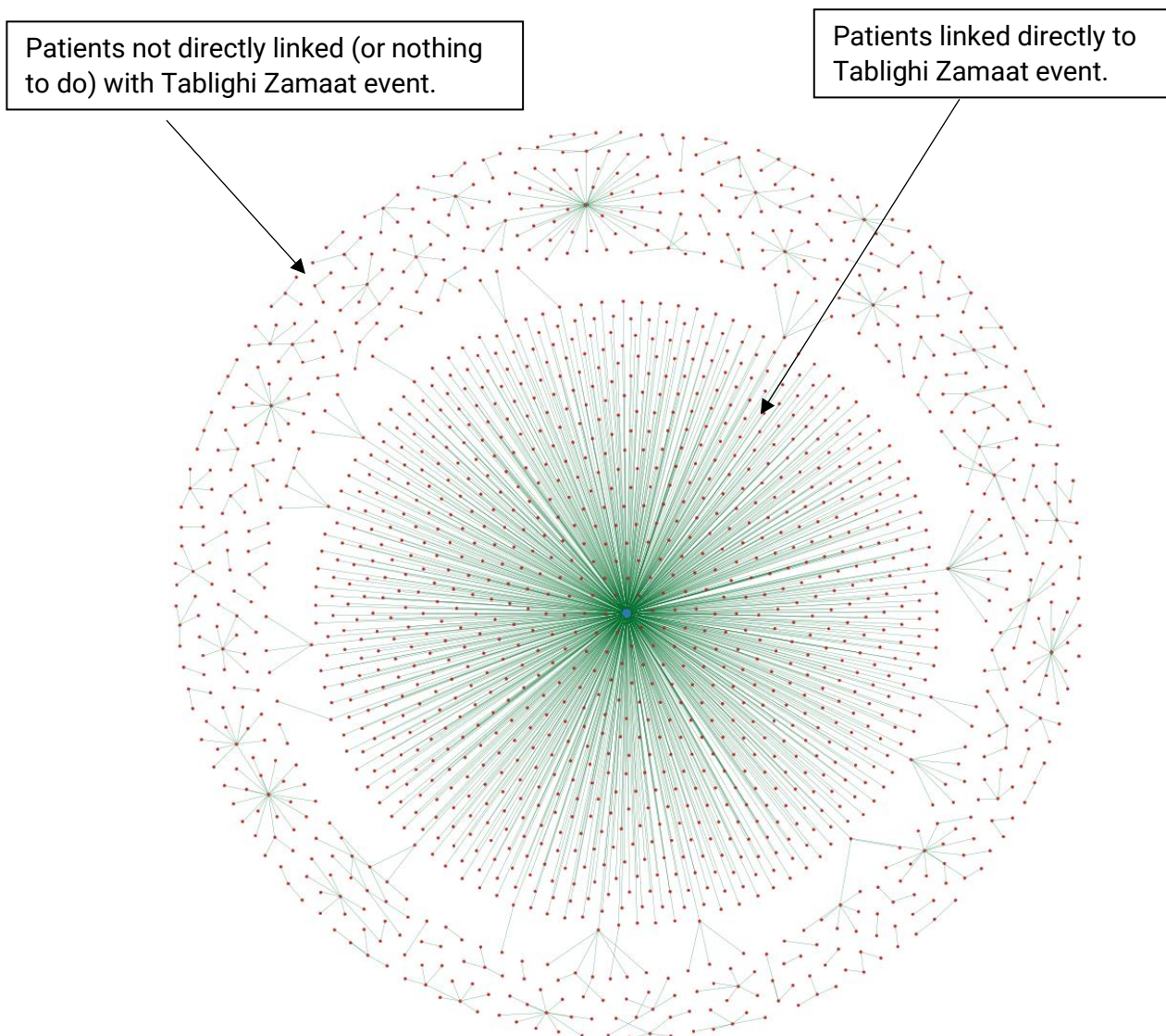
Betweenness Centrality talks about Super direct/indirect COVID **spreaders in different cities/localities**. Like connection link

## 4. Gephi Visualization of COVID19:

### How was the visualisation created?

The COVID 19 dataset used is from 10<sup>th</sup> March to 19<sup>th</sup> April 2020. The visualization shown below was accidentally created while I was trying out with the layouts. When the data is loaded, it is visualized in the form of lattice box. To get this visualization, first apply ForceAtlas 2 where the strongly connected nodes are clustered together, and we can also see several clusters that are disconnected from the rest of the graph. After nodes stop moving, apply Fructerman Reingold Layout which will make the nodes evenly spaced.

In the graph, patients are the nodes and the directed edges show how spread of infection happened from already infected patient. The node size is proportional to number of outgoing infections(edges) from the patient.



## Results by Gephi:

#Node: 1627

#Edges: 1523

Average degree: 0.936

Network Diameter: 5

Graph density: 0.001

Modularity: 0.59

Number of Weakly Connected Components: 118

Number of Strongly Connected Components: 1627

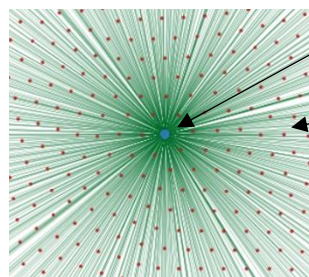
Average Clustering Coefficient: 0.002

Average Path length: 1.1952729319077096

## Analysis:

This graph shows two rings:

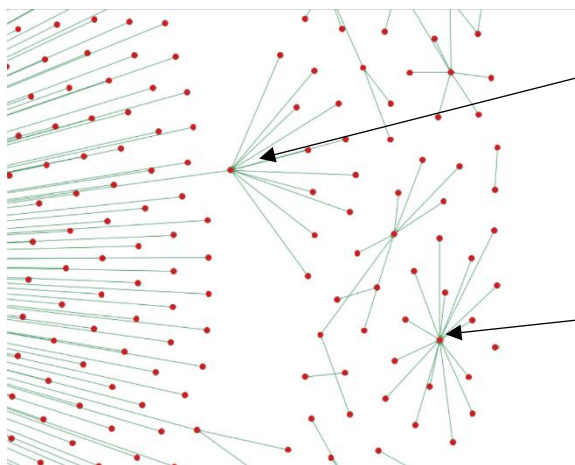
1. Inner ring: The blue node in the middle of the graph is the Tablighi Zamaat event happened in Delhi. Nodes (Patients) in the inner ring are directly connected to the patients from Tablighi Zamaat.



Tablighi Zamaat event

Patient Infected after the event

Outer ring: Here there are nodes where the spread of corona virus cannot be directly linked to Tablighi Zamaat event.



Patient directly linked to Tablighi Zamaat event, He/she infected 11 other people.

Patient with no link to Tablighi Zamaat event. He/she infected 16 other people.

It's interesting to note that when you look at this degree distribution, the Tablighi Zamaat event has accumulated a degree that's very large, almost 1500. Whereas, most patients have a very small degree. And that's typical of **power law degree distributions**.

Another interesting observation is the Completely random graph (ER Model) generated in Gephi with similar number of nodes and edges was able to predict Degree distribution, Network Diameter and average path length close to the Covid19 model. So, it can be deduced that probability that a person will infect another person is 0.00116 considering that corona virus is infecting randomly.

Nodes: 1627  
Edges: 1534  
Wiring Probability: 0.00116  
Average Degree: 0.943  
Network Diameter: 7  
Average Path length: 1.71

Completely random graph  
#Nodes: 1627  
#Edges: 1534  
Avg.degree: 0.943

COVID19 graph  
#Node: 1627  
#Edges: 1523  
Avg.degree: 0.936

