Epidemic Spread Analysis

line 1: Mohamad Jad AL Khalil   
line 2: *dept. of* *Electronic Engineering and Computer Science*  
line 3: *Queen Mary, University of London*line 4: London, England  
line 5: ec20794@qmul.ac.uk

line 1: 3rd Given Name Surname  
line 2: *dept. name of organization*  
*(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: 2nd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization*line 4: City, Country  
line 5: email address

line 1: 4th Given Name Surname  
line 2: *dept. name of organization*  
*(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address

*Abstract*

This document will be discussing how a new epidemic can start from a single patient and see how the epidemic can spread through a community. By using a friendship network and colocation data of people within the United Kingdom, we will see how different changes to a network can change the rate of spread of an epidemic. This paper considers data found online based on the COVID-19 pandemic and uses similar values for the models to simulate the same conditions.

# Introduction

For the past 18 months, the world has quickly been ravaged by the spread of COVID-19 at a rate no one could have predicted. World Economies have been left struggling with many left jobless and struggling to make ends meet. The spread of an epidemic has been proven to cause mass distress and cause issues to any community in its path. Network analysis methods have been around for a long time with many uses such as forming marketing strategies by seeing how communities are connected and which nodes are most central within a community. These methods can also be used to form strategies against epidemic spread but since COVID-19 was a new epidemic and little information was known about the disease, many governments did not know how to tackle the issue before it became such a large-scale problem.

In the time that COVID-19 has been around, there have been very few cases of reinfection after being infected and recovering from the disease, thus the model that will be used to analyse the spread will be the Gillespie SIR model from the Epidemics on Networks (EoN)[1] package for Python. This is a Markovian model which means the spread is based on the parameters given to the model but is also random throughout the network.

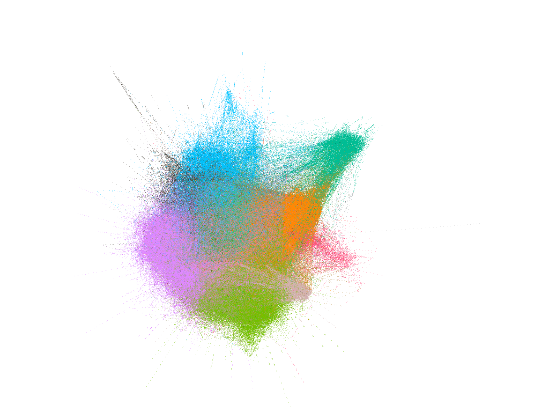
# Literature Review

# Dataset and Network

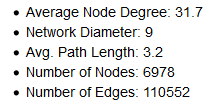
The dataset used for this network analysis is from Gowalla[2] which is a social networking website where people can share their location by checking in. The network is an undirected friendship network combined with a dataset of location data from check ins by users on specific dates and times. By editing the checkins dataset, it has been used to create a list of nodes and their location on a specific day. By doing this and limiting the location of people to within the United Kingdom (UK), it was possible to look for people who were within 500 meters of each other on a specific day. Any users within the specified range of 500 meters has an edge created between them thus creating a new edge list of interactions between people. By combining this edge list with the friendship dataset, we create a full edge list of all possible connections between people in the area. The initial network of this dataset has been plotted out using Gephi [3]and split by communities to produce the graph show in “Fig. 1”.

As we can see from the Network diagram, there are many communities within the region specified for the edge list. There are eight major communities found within the dataset and many smaller communities.

The entire formatted dataset ranges data collected over 400 days, but it is being used as if all the connections between nodes that are made over those 400 days are there from the beginning to see how the spread of such an epidemic would occur over a reasonably sized network and what changes to nodes and edges could reduce the effect of the epidemic on society.



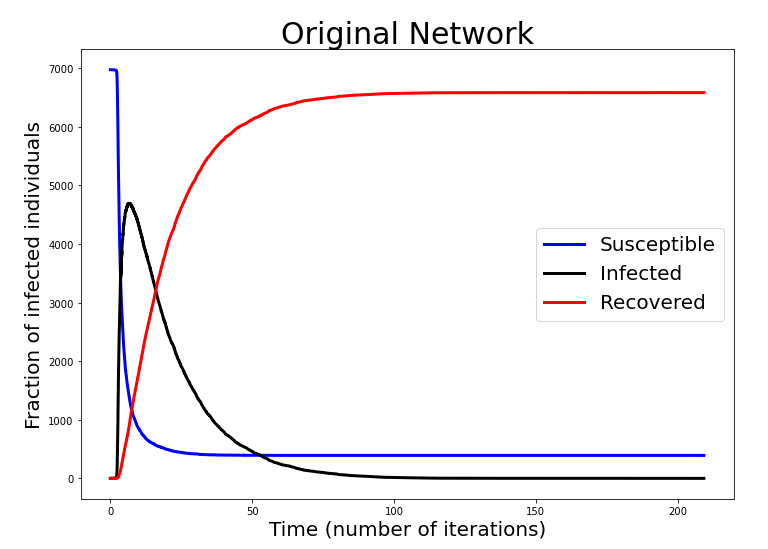
1. Initial Network



1. Initial Network Statistics

# Network Analysis Methodology

## Initial Network (RACHEL TRY FILL IN YOUR WORK HERE)



1. Initial SIR Model

The first thing to decide was how to set the right parameters of the chosen Susceptible, Infected, Recovered (SIR) model. As the name suggests, once someone has been infected, set by infection rate (β), they are removed after the recovery period, which is set by recovery rate (γ), and are either recovered or dead but in both cases are removed since they cannot infect anyone else anymore. The parameters we needed were already calculated by researchers in an article [4] on the parameters of the SIR model for COVID-19.

With the initial parameters chosen to be 0.09 for β and 0.06 for γ, I ran the initial SIR model on the network, and it produced the curves shown in “Fig. 3”. It shows that there is a massive spike in the number of infections within the first ten days of the first person being infected. There is a spike where about 65% of the population are all infected at once which would cause major issues for medical services especially since doctors and nurses would be included within the number of infected people. This requires measures to be taken to reduce the peak of the infection and try stall the rate of spread as much as possible to flatten the curve.

## Centrality Measures

The first task to be carried out on the network is checking the effects of the different centrality measures on the spread of an epidemic. Since networks rely on these central nodes for connection to other parts of the network, severing these links should have positive effects on the way an epidemic would spread within the network. The idea would be that the most central nodes (people) would have to take extra precautions to avoid human interaction for the good of the rest of the community.

We decided to try two different values for x (percentage of nodes removed) to see how effective the method would be and how it could scale. The two values chosen were 1%, to see if the most central nodes had a large effect on the spread of a disease, and 5%, to see if there would be a massive decrease in the rate of infection compared to 1%.

The centrality measures chosen for the task are:

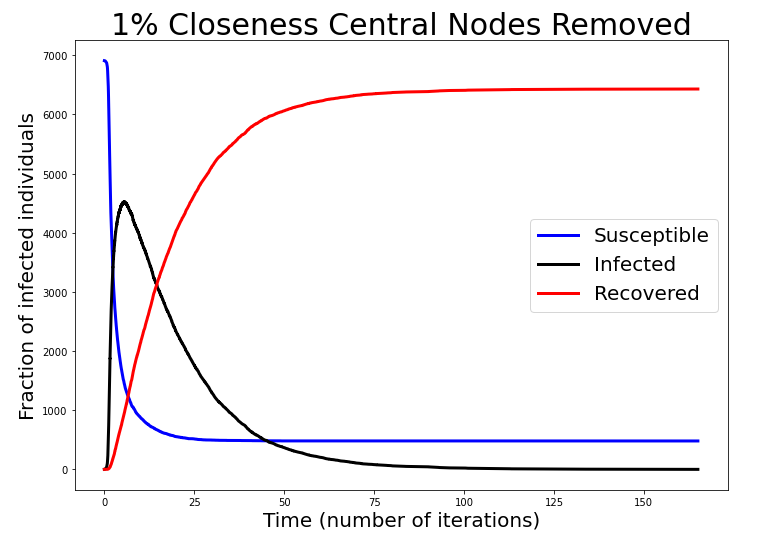
* **Closeness Centrality** – The average shortest path between a node and all other nodes in a network. This is useful because if a node has a high closeness centrality, it would be somewhere in the middle of the network so a lot of interactions would occur between this person and people who are interacting with others whereas a node with low centrality might only be able to communicate with one person and thus has a longer path to reach others.
* **Degree Centrality –** The total number of edges that are connected to a node. A node with the most edges connected to it would be considered the most central node in the network using this measure. This would refer to popular or key people in a network. Since they have so many edges, removing them from a network would cause a lot of connections to be lost as all their edges would be removed. This could in turn sever communities from each other thus making certain communities free from infection from the epidemic since it could not reach them.
* **Betweenness Centrality –** The likelihood that a path goes through a node to reach another node. Since they are the nodes that create the shortest paths between people, removing them would increase the time needed for an epidemic to spread thus lowering the maximum infected number.
* **Eigenvector Centrality –** The rank of a node in connecting other nodes together. The nodes with the highest eigenvector centrality are the nodes that most likely connect communities together and without them connections between communities could be lost. Even though there could be many nodes connecting communities, removing the most central ones could still have a major effect on the spread of an epidemic.

### Closeness Centrality

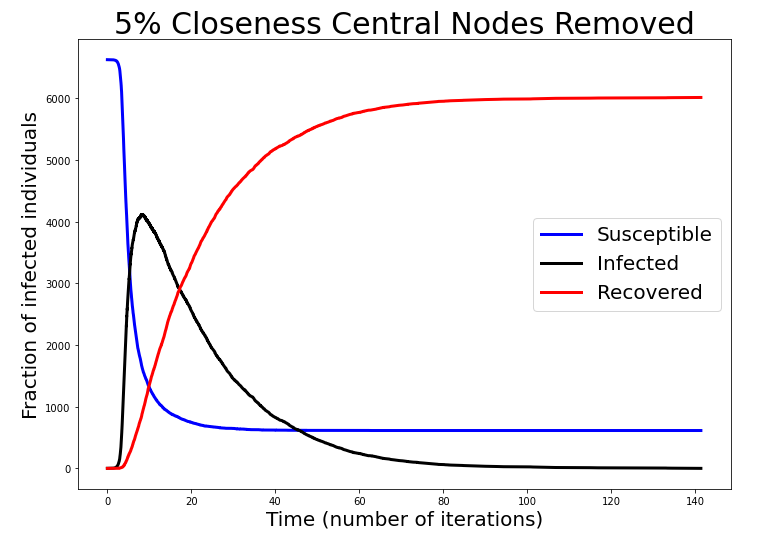
First, we started by removing 1% of the most central nodes in the network using NetworkX [5] and its feature to calculate the centrality for each node in the network.

By calculating the closeness centrality for each node in the network, a list was then created for each node and its centrality. The list was then sorted in descending order and then the top 1% of nodes from the list were removed from the network.

This gave us the result shown in “Fig. 4” and the results for removing 5% of the central nodes in “Fig. 5”.



1. 1% closeness centrality SIR

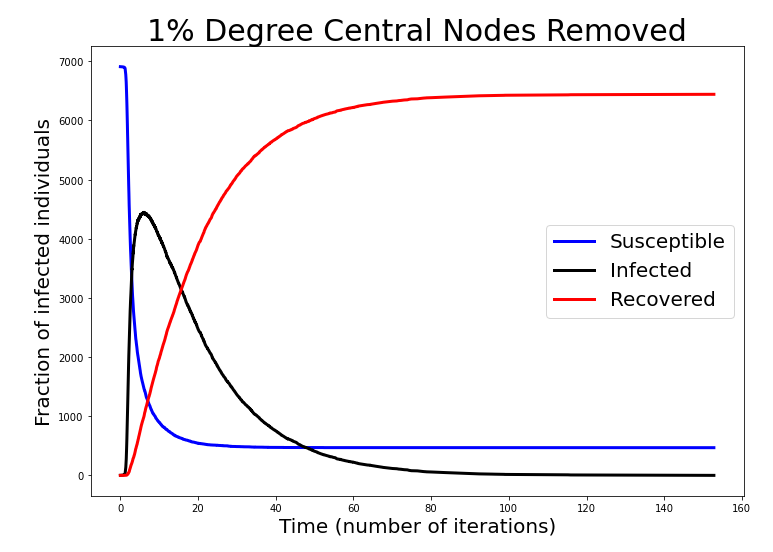


1. 5% closeness centrality SIR

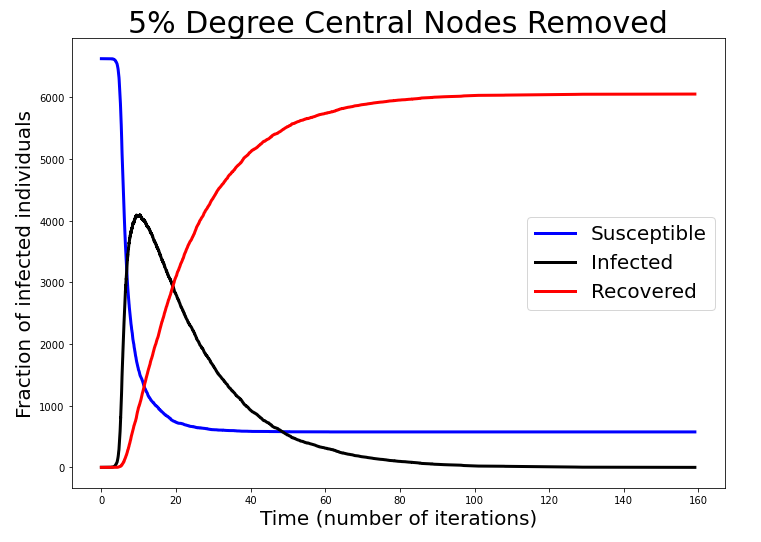
As can be seen from the graphs above, removing central nodes from the network has a positive effect on the number of infections in the network. Removing 5% of the nodes with the highest closeness centrality results in an 8% reduction in the maximum number of people infected on a given day from 65% to 57%. Removing the most central 1% of nodes had little effect however and only reduced the number of people infected by 1%.

### Degree Centrality

The same method was applied to the rest of the centrality measure but using the respective centrality calculations from NetworkX.

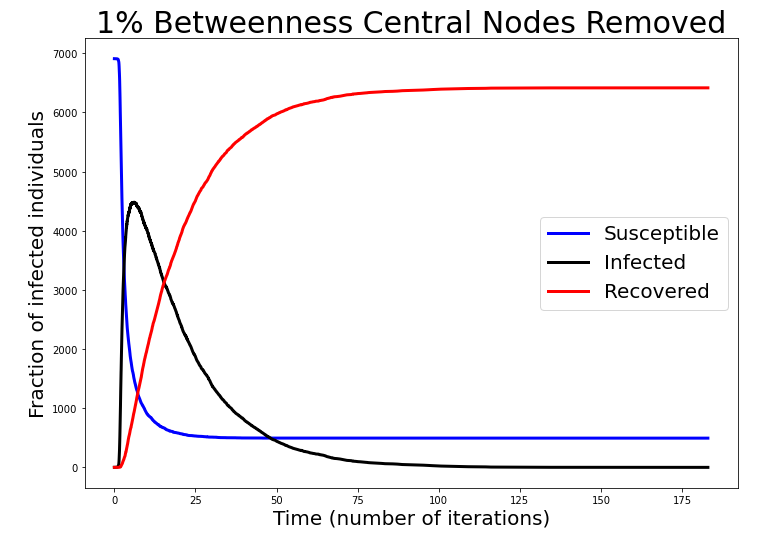


1. 1% degree centrality SIR

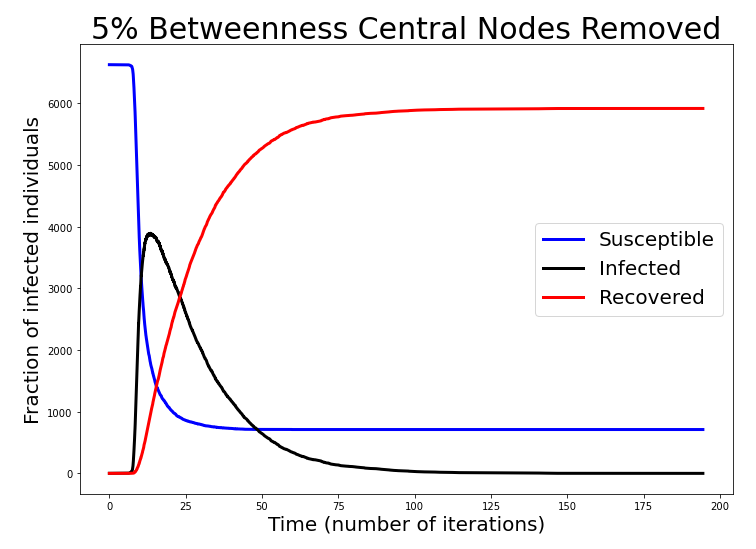


1. 5% degree centrality SIR

### Betweenness Centrality

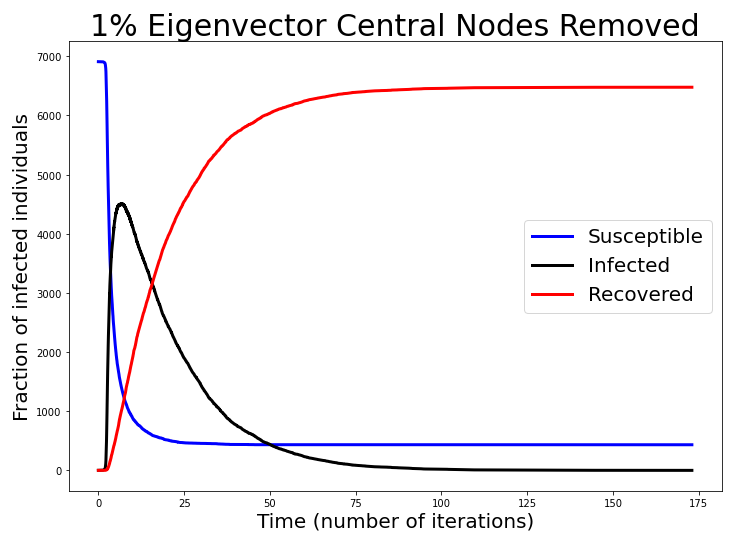


1. 1% betweenness centrality SIR

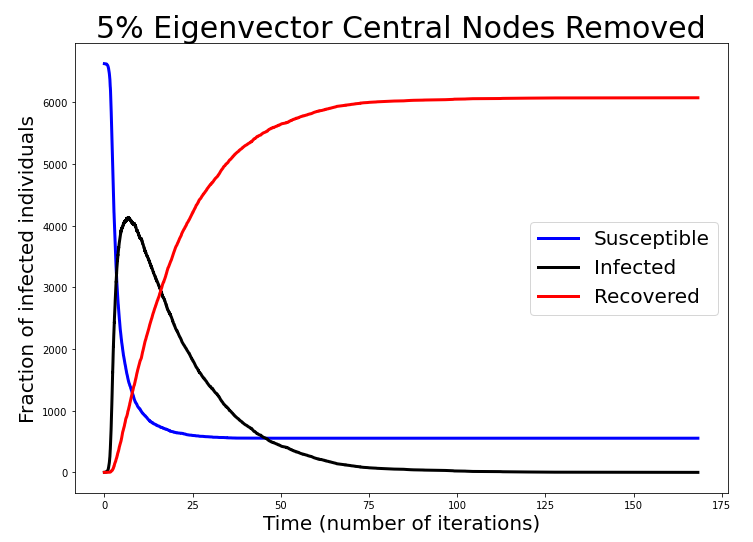


1. 5% betweenness centrality SIR

### Eigenvector Centrality



1. 1% eigenvector centrality SIR

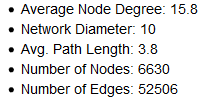


1. 5% eigenvector centrality SIR

### Result discussion

As can be seen from “Fig. 4” to “Fig. 11”, there is not much difference between the different centrality measures and their effect on the rate of spread in this network. Although betweenness centrality does slow the rate of spread more than any of the other centrality measures.

This is possibly due to the increase in the average path length in the network. “Fig. 12” shows the network statistics for the network resulting from removing the 5% top betweenness centrality nodes.



1. 5% betweenness centrality statistics

By removing just 5% of the most central nodes, the average node degree halves as well as the number of edges and the network diameter increases thus successfully spreading out the total population.

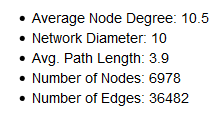
## Imposing Restrictions on movement

### Partial Lockdown

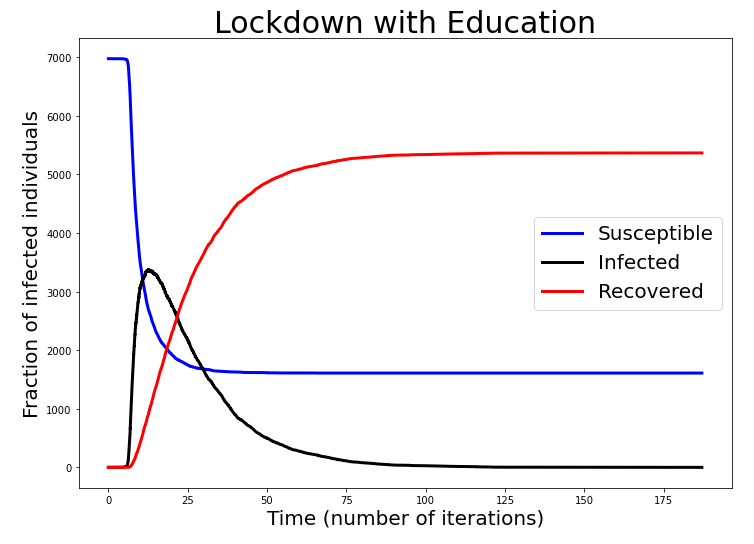
The first restriction attempted to be imposed on the network was a lockdown scenario where education systems would still run in person so students and teachers would still interact with each other. To do this I had to find the percentage of the population which worked in key roles in society. Using the data for the UK’s work force [6] I found that approximately 33% of workers in the UK are classified as key workers.

To simulate 67% of the population being put under lockdown, I split the edge list using a train test split from the SciKitLearn package for python and took 67% of edges and removed them from the network. This left the graph looking quite disconnected as shown in the stats from “Fig. 13”.

With this type of lockdown most of the nodes have no contact with outside nodes other than the necessary shopping although in this simulation that is not considered. “Fig. 14” shows the SIR model on this disconnected network and it can be seen that the number of infections drops significantly compared to the original network.



1. Partial Lockdown statistics

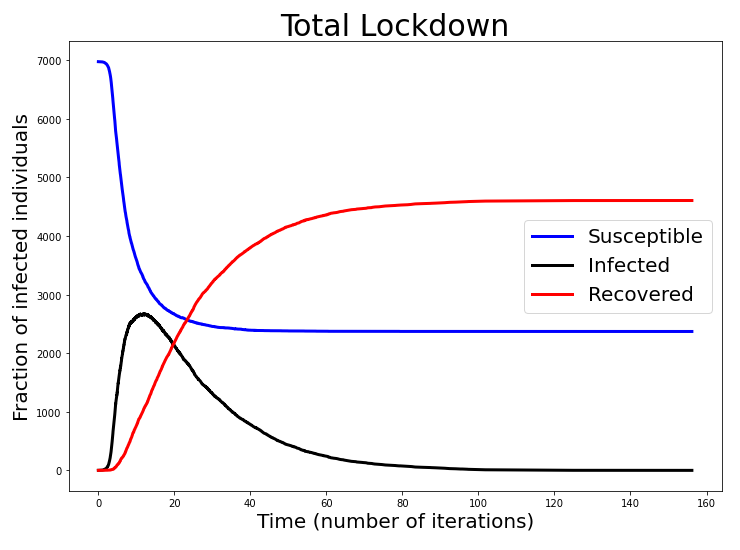


1. Partial Lockdown SIR curve

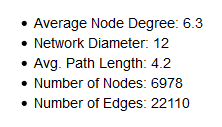
### Total Lockdown

By implementing a total lockdown, we would be trying to limit human interactions to the bare necessities so education would all have to be online and all shopping would be heavily controlled. I have calculated the percentage of the population that would be put on lockdown would be 80% from the data in [6].

Having this total lockdown drastically reduces the number of people who get infected with a maximum number of infections at any time being around 2800 people as shown by “Fig. 15”. This will help healthcare systems cope much more with the epidemic outbreak that if people had freedom of movement but it would not reduce the time for the epidemic to end by much.



1. Total Lockdown SIR



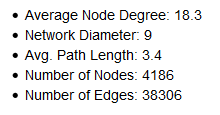
1. Total Lockdown Statistics

### Node Removal Analysis

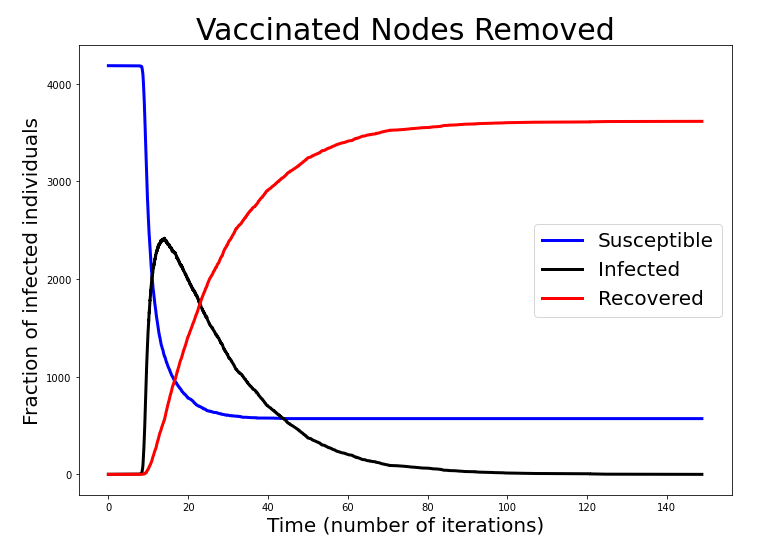
We already attempted to remove central nodes from a network to see the effect of it on the spread of an epidemic but in this node removal analysis, we will be simulating the effect of having a portion of the population vaccinated from the epidemic.

First, 40% of the population (nodes) are chosen at random and removed from the network since we consider vaccinated nodes as immune and therefore regardless of if they are in the network or not, nothing can pass through them.

When the nodes are removed from the network we are given the statistics shown in “Fig. 17” and the SIR curve in “Fig. 18”.



1. Vaccinated network statistics



1. Vaccinated Network SIR

We can see that having a large portion of the population vaccinated from the disease reduces the maximum number of infections by over 50%.

In reality, it is not possible to have the population vaccinated so early in an epidemic though so different measures must be taken until vaccinations can be given out to the population.

# Conclusions

# References

[1]        J. C. Miller and T. Ting, “EoN (Epidemics on Networks): A fast, exible Python package for simulation, analytic approximation, and analysis of epidemics on networks,” *arXiv*, pp. 1–28, 2020, doi: 10.21105/joss.01731.

[2]        “loc-Gowalla @ snap.stanford.edu.” [Online]. Available: https://snap.stanford.edu/data/loc-Gowalla.html.

[3]        “Gephi - The Open Graph Viz Platform.” https://gephi.org/ (accessed Apr. 19, 2021).

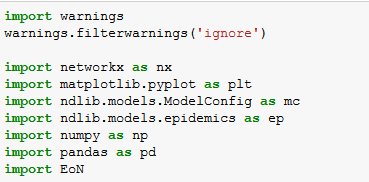
[4]        D. Kumar Bagal, A. Rath, A. Barua, and D. Patnaik, “Estimatingtheparametersofsusceptible-infected-recoveredmodelofCOVID-19casesinIndiaduringlockdownperiods,” no. January, 2020.

[5]        “NetworkX — NetworkX documentation.” https://networkx.org/ (accessed Apr. 19, 2021).

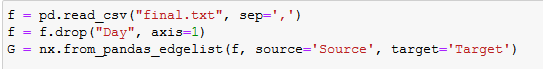
[6]        R. Mcsweeney, “Coronavirus and key workers in the UK,” *Office for National Statistics*, no. May, pp. 1–7, 2020.

# Appendix

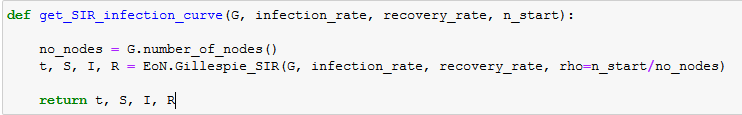
## Modules Imported



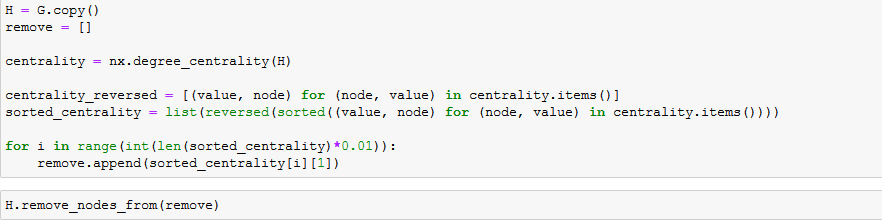
## Load Dataset



## Model



## Code for removing nodes/edges



## Remove Random Nodes

