

Epidemic Spreading Analysis with Social Networks

Danish Mehra
School of Electronic
Engineering and Computer
Science
Queen Mary University of
London
ec20716@qmul.ac.uk

Neishka Srivastava
School of Electronic
Engineering and Computer
Science
Queen Mary University of
London
n.srivastava@se20.qmul.ac.uk

Ropan Bhattacharya
School of Electronic
Engineering and Computer
Science
Queen Mary University of
London
r.bhattacharya@se20.qmul.ac.uk

Sujith Aleshwaram
School of Electronic
Engineering and Computer
Science
Queen Mary University of
London
s.aleshwaram@hss19.qmul.ac.uk

Abstract—Assessment of epidemic spread, and its control has become a very pertinent topic over the past couple of years with the global spread of covid19. In this paper we have provided a comparative study of the major methodologies used for lowering the peak of the epidemic curve. Here we have taken into consideration the hospital capacity catering to the concerned population. For epidemic spread modeling we have primarily used SIR model. Overall, as an outcome of this study, firstly, we have presented a high-level comparison of the vaccination strategies and identified based on the infection curve how certain strategies are more effective. We have used network methodologies like community detection, or other network centrality metrics to define these strategies. Secondly, we have also presented an effective targeted lockdown strategy which is based on identifying the key hotspots for lockdown using network centrality metrics. Here, we have identified the extent of lockdown needed to lower the peak sufficiently below the healthcare capacity.

Keywords—Epidemic, SIR, SIS, Lockdown, Targeted Vaccination

I. INTRODUCTION

Epidemics have currently become one of the most researched topics around the world with the spread of Covid19. Epidemics spread through complex heterogeneous network of people by various modes of transmission (for example physical contact between a susceptible and an infected person). In epidemic analysis these complex networks are studied using Social Network Analysis, while different infectious disease models are used to simulate an epidemic spread within these networks. One such common model used is Susceptible Infectious Recovered (SIR) epidemic model [1]. In SIR model, people are susceptible initially; once getting infected they become disease spreader and finally they recover and become disease resistant.

Understanding epidemic spread within a population usually involves large scale analyses which can include looking at movements of people between cities or countries. Nevertheless, understanding how spreading works at a small scale (i.e. within cities or localities) can provide quick inciteful information to devise strategies to combat an epidemic. To do this it is crucial to take into account the structure of the social network and use derived information like communities within the network, network centrality measures, to model localized changes in the network and simulate and epidemic spread to identify the effects of these changes. These models can then be extended to form strategies around vaccination and lockdown.

In this report we have presented a comparison of the strategies devised based on network related metrics that could be used for vaccination and lockdown. The primary

network data that has been used to simulate the epidemic spread is a user location dataset [2] that is publicly available. Using this dataset, we have created a hypothetical epidemic spread between friends who meet at the same place within a given time window (colocation user graph - undirected) and analysed how the network structure influence this spread.

- To model various vaccination strategies, we have performed node removal analysis based on network centrality measures such as eigenvector and betweenness centrality. We have also overlayed this network with a separate friendship network dataset (friendship user graph - undirected) to identify communities and community-based centrality measures. These community-based measures have also been used as part of node removal analysis to model a vaccine strategy based on targeted vaccination to people who have high degree within a community.
- To model tactical interjections like lockdown, we have first identified location hotspots based on indegree centrality measures in a directed graph of user and their check-in location. We have subsequently used this as a basis for third node removal analysis to simulate the effect on epidemic spread of targeted lockdown.

Finally, we have closed this paper by discussing the outcome of the above simulations performed to model various vaccination and lockdown strategies. Here we have also presented the reasoning behind the higher effectiveness of certain measures, also discussing the challenges in implementing them.

II. RELATED WORKS

Multiple research works have been done around epidemic spreading, especially over the last decade. In this section we discuss related research efforts and explain how centrality measures help in understanding disease spreads in Social networks.

Study of epidemic spreading has drawn parallels from information/ idea diffusion in networks. However, one of the key differences with information cascade is the fact that in epidemic spreading there is no decision making and as a result the process of contagion can be quite complex and very difficult to model. One simple model used quite commonly to simulate the epidemic branching is based on idea of waves. Essentially in first wave, a person carrying a disease enters the population and transmits to k individuals with probability p . In second wave, these k individuals go on to meet k different people each as a result having a second wave of $k*k=k^2$ people. The process continues in subsequent waves. This is similar to Reed-Frost model in its discrete time version, where

infections happen sequentially in generations [6][7]. On the contrary a Markovian model used for modelling epidemic spread assumes that infectious individuals infect each of their susceptible neighbour independently at rate β and recovers at rate γ [4]. One of the seminal works in modelling epidemic spread has been provided by Tom Britton [4] where he has performed stochastic modelling and statistical analysis on infectious disease outbreaks taking place in communities structured as social networks. In his paper he has analysed disease spreading using both Markovian and Reed-Frost methodologies. He has also modelled various prevention strategies including vaccination. Overall, in this paper he has presented what can be estimated and how. [4]

Another work by Dekker [4] discussed few centrality measures such as Closeness centrality, Betweenness centrality, Eigenvector centrality which are best in identifying disease spreaders.

- Closeness centrality is the most popularly used measure to find out how far the nodes are from each other.
- Betweenness centrality helps to find out how a single node is involved in shortest paths in the network.
- Eigenvector centrality is used to find out how inter community links are connected in a huge network.

From Dekker's findings, we can say eigenvector centrality is sometimes partial to densely connected community network instead of giving appropriate measure on whole network of communities [1].

Similarly, Christley discovered how centrality measure helps in identifying a root person where the disease spreading originated. He also states that identifying the related parameter which predicts the vulnerability of people to infection will help in analysing and controlling the further spread of disease. Also, node degree is helpful in estimating the importance of this parameter [2].

Ajelli's study on epidemic spread prediction says both metapopulation model and agent-based model are good at predicting the spread of epidemic. An agent-based model is a stochastic and spatially-explicit simulation model where the agents represent individuals. On the other hand, Ajelli used Global Epidemic and Mobility (GLEaM) model for his analysis on a metapopulation approach. In this model the world is divided into geographical regions thus defining a sub-population network where connections among subpopulations represent the individual fluxes due to the transportation and mobility infrastructure [8,9]. Infection spread occurs inside each urban area and is described by compartmental schemes based on discrete stochastic dynamics of the individuals and the disease. Ajelli concluded that although the metapopulation model prediction results are less accurate, it did provided accurate spreading patterns of epidemics among the connections. At same time agent-based model can give better predictions results than the other model using high confidence datasets [3].

Quantifying the nodal spreading abilities and identifying the potential influential spreaders has been one of the most engaging topics recently, which is essential and beneficial to facilitate information flow and ensure the stabilization

operations of social networks. In order to accurately identify the potential super-spreaders, the CumulativeRank algorithm has been proposed by Zhang, D et al. This algorithm combined the local and global performances of nodes for measuring the nodal spreading abilities. They carried out extensive experiments with the Susceptible-Infected-Recovered (SIR) model on real-world social networks to demonstrate the accuracy and stability of the proposed algorithm.

Key Terms

Following key terms were used in the work below.

- **Basic Reproductive Number:** $R_0 = p \cdot k$, where p = infection probability, k = branching number. It determines if the disease will spread or die out. In the branching process model, if $R_0 < 1$ the disease will die out after a finite number of waves. If $R_0 > 1$, with probability > 0 , the disease will persist by infecting at least one person in each wave.
- **SIR model:** Here individuals are susceptible then they get infected and finally they recover at a rate γ . A node stays infected for t time.
- **SIS model:** Here individuals are susceptible then they get infected. However even after recovering they stay susceptible. Here if recovery rate is greater than infection rate, the epidemic dies out.

III. DATASET AND NETWORK PRESENTATION

Dataset Description and Pre-Processing

A. Checkins Dataset:

Checkins Dataset is collected from users sharing their locations by checking-in. This is accessed using their public API and consists of a total of 6,442,890 check-ins of these users over the period of Feb. 2009 - Oct. 2010.

In our simulations, we have used the checkins dataset for the month of July 2010. This is based on the fact that there is considerably higher number of movements in this month as it's the onset of the summer break and most people opt to travel at this time.

To preprocess the data, we have divided the check-in time into 6 different time slots like shown below.

00:00 - 4:00 --> Group A
 4:00 - 8:00 --> Group B
 8:00 - 12:00 --> Group C
 12:00 - 16:00 --> Group D
 16:00 - 20:00 --> Group E
 20:00 - 00:00 --> Group F

We have identified co-located users by grouping the checkins dataset over location id and time slots.

Finally we have created a graph with colocated users based on the grouping done above.

B. Friendship Network:

The friendship network is a dataset that represents an undirected network. The data is a collection of tuples where each element is a user and the tuple represents their friendship. This was collected using Gowalla's public API, and consists of 196,591 nodes (users) and 950,327 edges (friendships). In our analysis we have used the entire friendship network dataset for community detection.

Graphs Generated

For the epidemic analysis we have generated 3 graphs. Here we describe them in detail.

A. Colocation Graph

Colocation graph is an undirected graph which consists of users as nodes and edges representing the fact that they are colocated. This graph was generated from checkins dataset for the month of July 2010. As mentioned in the previous section we identified colocated users by grouping them over location id and time slot.

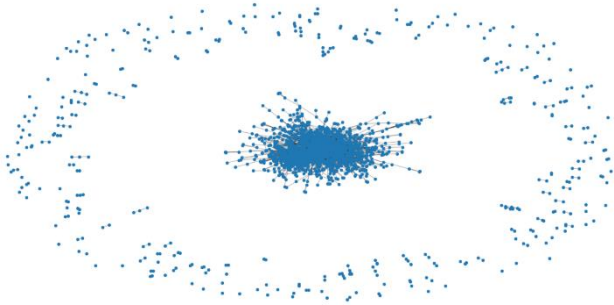


Figure 3.1 Colocation Graph

Graph specifics

- Number of Nodes: 2068
- Number of Edges: 5918
- Average degree: 2.86
- Maximum degree in the Graph: 83

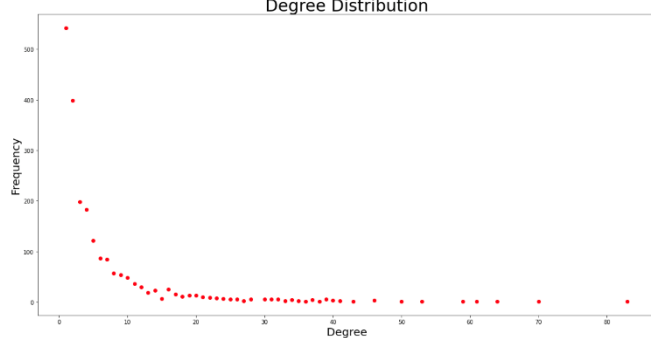


Figure 3.2 Degree Distribution of Colocation Graph

As can be seen from the above figure, about 75% of the nodes have degree less than 5. However, it can also be seen that there are some nodes which have a really high degree (above 50). These essentially represent individuals who frequently visit social hotspots.

B. Friendship Network Graph

Friendship network is an undirected graph which consists of users as nodes and edges representing the friendship between them. This graph was used to identify communities and overlay the community driven metrics over the colocation dataset.

Graph specifics

- Number of Nodes in the Community Graph: 1812
- Number of Edges in the Community Graph: 5041
- Average degree: 2.78
- The maximum degree in the Graph: 79

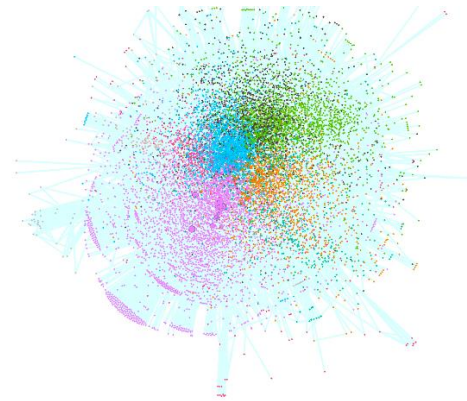


Figure 3.3 Undirected Community Graph

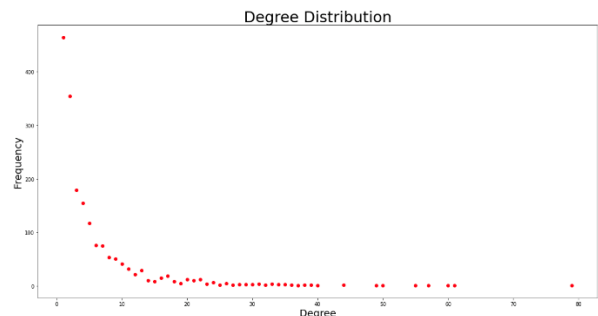


Figure 3.4 Degree Distribution of Community graph

We have identified in total 24 communities using greedy modularity community detection algorithm of NetworkX. As can be seen from the above figure, about 70% of the nodes have degree less than 5. However, it can also be seen that there are some nodes which have a really high degree (above 50). These essentially represent individuals who are social and have a lot of friends. These individuals can end up being super spreaders.

C. Hotspot Graph

Hotspot graph generated is a directed graph where we have user nodes pointing to location nodes. The edges here represents the fact that the user has checked at that location. The hotspot graph has been used to identify crowded locations which can act as hotspot for epidemic spread.

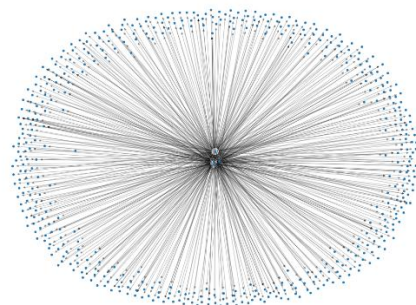


Figure 3.5 Hotspot Graph

Graph specifics

- Number of Nodes in the Hotspot Graph: 98135
- Number of Edges in the Hotspot Graph: 122621
- Average degree: 1.25
- The maximum Indegree in the Graph: 185

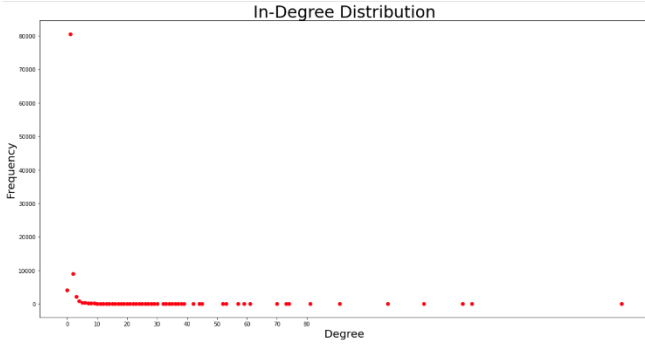


Figure 3.6 Degree Distribution of Hotspot Graph

In figure 3.5, we can see quite a varied degree distribution. About 90% of the nodes have degree of less than 5. A handful of nodes have a really high degree (above 50). In fact, we have nodes with degree above 100 as well. These essentially represents hotspots where users usually come in contact with each other. We have analyzed as part of our simulation a lockdown strategy which targets these hotspots.

IV. NETWORK ANALYSIS METHODOLOGY

Mentioned below are the steps for the conducted experiments to visualize the infection curve and to devise certain measures to control the spread of epidemic.

Infection Curve

- Using the check-ins dataset, based on users being collocated at the same point of time, we have generated a graph of users with edges defined as users being collocated.
- After generating the graph, we have run simulations for epidemic spread for transmission probability and reproduction rate.
- We have also plotted the epidemic spread using both SIR model and SIS model.

Node Removal Experiments

We have performed node removal to model epidemic spread keeping in mind both vaccinations and tactical interjections like lockdown. For vaccinations, we have performed the following experiments.

1. Eigenvector Centrality
 - We have removed nodes based on highest Eigenvector Centrality.
 - We have performed this experiment with 0%-30% of the nodes removed through 5 different simulations.
 - For each simulation of the infection curve, we have performed 500 iterations to model epidemic spread.
2. Betweenness Centrality
 - We have removed nodes based on highest Betweenness Centrality.
 - We have performed this experiment with 0%-50% of the nodes removed through 5 different simulations.

Community Detection and Node Removal

Further on vaccinations, we have also performed experiments with community detection.

- We have identified communities within the friendship network dataset.

- We have then identified and removed nodes with highest degree within a community.
- We have performed this experiment with 0%-60% of the nodes removed through 7 different simulations.

Hotspot Detection and Node Removal

To analyze the tactical measures like lockdown on the effectiveness of controlling epidemic spread, we have also performed node removal experiments with hotspot detection.

- Identifying the hotspots based on the highest indegree centrality for a directed graph between users and their check-in location with edge being the count of times the user visited that location.
- We have then identified and removed hotspots with highest indegree and then generated a graph of users based on co-location dataset.
- We have then performed this experiment with removed nodes.
- Overall, we have performed 7 different simulations with different percentage of hotspots being removed.

V. RESULTS AND DISCUSSION

In the following section, we have discussed in detail the outcomes of the individual simulations that we have run considering different epidemic models.

A. Epidemic Spreading Simulation

To generate an epidemic spreading simulation in the friendship network. By taking the co-location data, students have to propose a model on how a hypothetical infection would spread in the population, as friends co-locate and are susceptible to infect each other.

As we discussed above, our first simulation starts by looking at the spread of epidemic in a friendship network using co-location data.

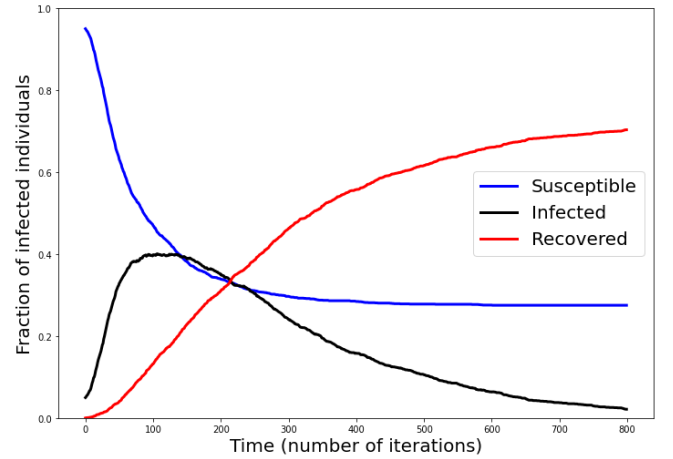


Figure 4.1 Infection Curve (SIR)

Through time, we can see that the fraction of individuals getting infected increases until a peak is achieved (nearly 120 iterations). At peak, about 40% of the population is infected. The peak then decreases as the rate of recovery increases through time. Fraction of individuals susceptible starts falling from 100% and stabilizes between 30% to 40%. At this point, the infection rate has waned and the recovery rate is quite high.

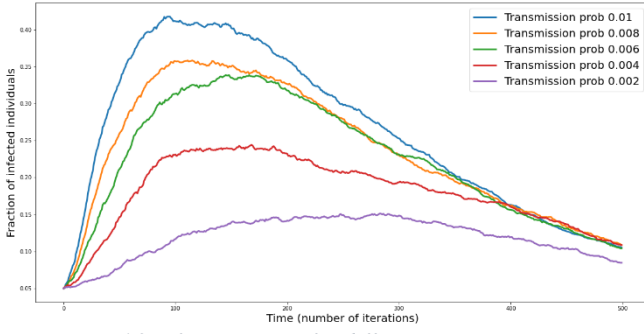


Figure 4.2 Infection Curve for different transmission rate (SIR)

As can be seen, the peak of the fraction of infected individuals increases as we increase the transmission rate. We have performed the simulations below with transmission probability of 0.01.

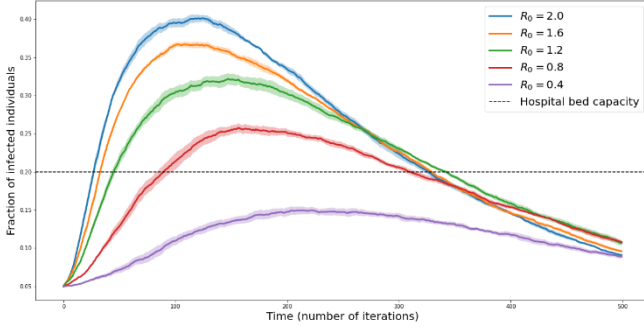


Figure 4.3 Hospital Bed Capacity Curve (Infection Curve)

As can be seen, decreasing R_0 value, decreases the peak of the fraction of infected individuals. The line representing the hospital bed capacity tells us the maximum number of beds available at a particular point in time. For getting proper healthcare, it is advised that the peak of the infection curve stays below the hospital bed capacity line at all times.

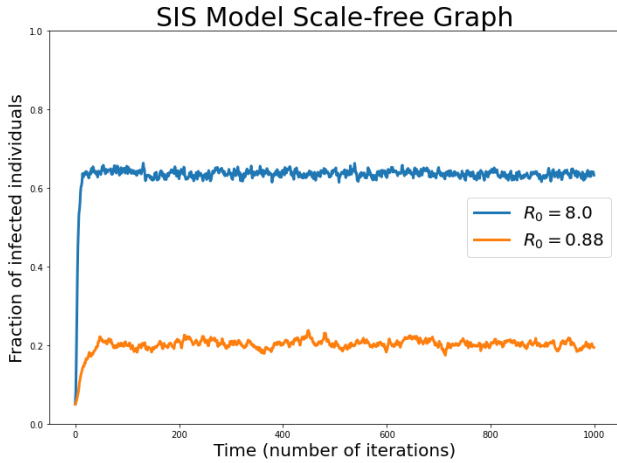


Figure 4.4 SIS Model Scale Free Graph

This model suggests firstly the peak is achieved within a few iterations after which the infection curve plateaus. This is because individuals continue to stay susceptible after having recovered.

B. Node Removal Experiments

In the following section, we provide the outcomes of removing the nodes having highest eigenvector centrality degree. We have taken different percentage values and

conducted an experiment showcasing the spread of epidemic through number of iterations.

1. Eigenvector Centrality

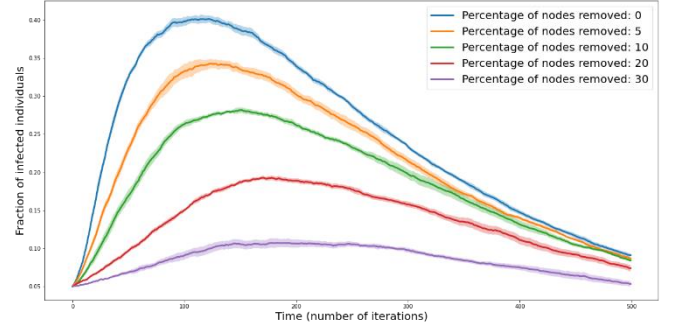


Figure 4.5 Eigenvector Centrality and Node Removal

As observed from the graph, removing the top 30% of the nodes having high eigenvector centrality drastically flattens the curve hence controlling the epidemic. If we draw a parallel with vaccination, using targeted vaccination based on highest eigenvector degree, we would need to vaccinate only about 20% of the population to push the infection rate down below the hospital capacity of 20% at peak.

2. Betweenness Centrality

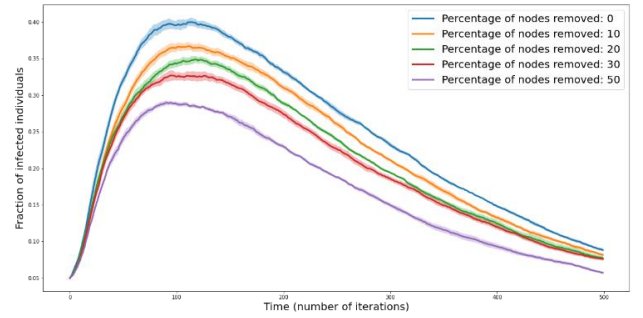


Figure 4.6 Betweenness centrality and Node Removal

As observed from the graph, even after removing the top 50% of the nodes having high betweenness centrality flattens the curve although the peak still remains above 20% hospital bed capacity. Thus, with regards to targeted vaccination based on highest betweenness centrality, we would need to vaccinate more than 50% of the population to push the infection rate down below the hospital capacity at peak.

C. Community Detection and Node Removal

We detected 24 communities in friendship network dataset using NetworkX greedy modularity algorithm. Here we have presented the outcome of node removal experiments done after Community Detection. As mentioned previously nodes are removed as a fraction of each community based on highest degrees within a community.

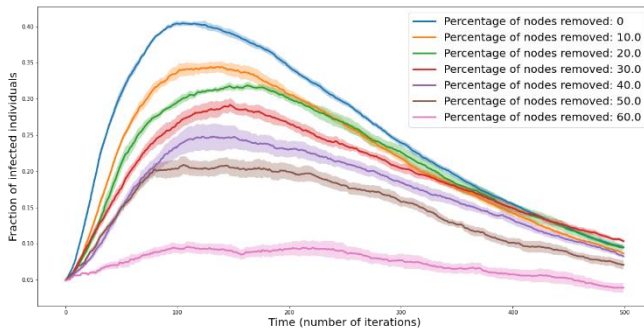


Figure 4.7 Community Based Node Removal

From the plot, we can observe that after removing the top 50% of the nodes based on highest degree in each community, the curve flattens with the peak below 20% hospital bed capacity. Thus, with regards to targeted vaccination based on community-based criteria, we would need to vaccinate at least 50% of the population to push the infection rate down below the hospital capacity at peak.

D. Hotspot Detection and Node Removal

Here we have discussed the outcomes of the experiments to demonstrate the effectiveness of tactical measures like lockdown on epidemic spread. As mentioned in the previous section, we have identified the hotspots based on the highest indegree centrality for a directed graph between users and their check-in location with edge being the count of times the user visited that location.

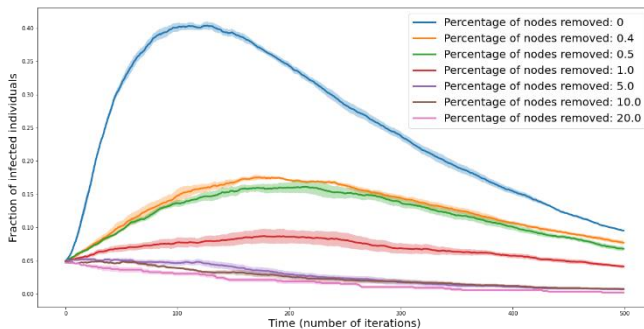


Figure 4.8 Hotspot Detection Based Node Removal

Here we can see even after removing the locations with 0.4% highest indegree locations we are able to stabilize the peak below the hospital capacity of 20%. In fact, by removing 1% of the locations, we are able to achieve considerable flattening of the curve. Extending this to imposing lockdown, it is clear that shutting down key hotspots helps control the epidemic spread.

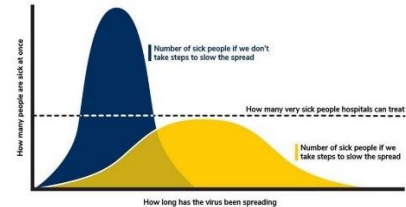
VI. CONCLUSIONS AND PERSPECTIVES

From the observed simulations we can conclude that to manage proper healthcare, we have to devise strategies that can control the peak of the infection curve so that it stays below the hospital bed capacity line at all times. One strategy can be tactical measures like closing down major hotspots, stopping major community interactions and restricting movement of high degree nodes which represent people could contribute to flattening the curve and hence reducing the epidemic spread. Second approach can be more strategic and be devised around vaccination and full herd immunity. An

efficient way of vaccinating can be targeted immunization. As we have presented in our report approaches like using eigenvector centrality to determine critical nodes of a social network and vaccinating them on priority yields comparatively lower infection spread. However, it must be added that determining the eigenvector centrality of all the nodes in a population can be a challenging task. Nevertheless, we can draw an inference that targeted vaccination to individuals who are important and larger social circle can yield faster benefits.

Steps to take to flatten the curve:

- Wash your hands
- Avoid travelling to areas with the virus
- Stay home if sick
- Avoid hosting large gatherings



Challenges and Future Directions

Due to computational issues, we worked on a smaller sample size. Working with larger samples will give us a much deeper understanding on the measures that can help us efficiently control the epidemic spread. A small part of the network is generally observed, hence could be misleading. We focused our colocation study on a micro level which slightly but not completely imitates the real-world scenario. As a thought experiment for further studies, we can consider looking further inside the community structure. Here with targeted community-based measures, we can try to stop the epidemic from spreading outside the community. One of the key areas would be to identify the bridging nodes which link communities together.

VII. REFERENCES

- [1] Dekker, Anthony. (2013). Network Centrality and Super-Spreaders in Infectious Disease Epidemiology.
- [2] R. M. Christley, G. L. Pinchbeck, R. G. Bowers, D. Clancy, N. P. French, R. Bennett, J. Turner, Infection in Social Networks: Using Network Analysis to Identify High-Risk Individuals, American Journal of Epidemiology, Volume 162, Issue 10, 15 November 2005, Pages 1024–1031, <https://doi.org/10.1093/aje/kwi308>
- [3] Ajelli, M., Gonçalves, B., Balcan, D. et al. Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. BMC Infect Dis 10, 190 (2010). <https://doi.org/10.1186/1471-2334-10-190>
- [4] Britton, T. "Epidemic models on social networks—With inference," Statistica Neerlandica, Netherlands Society for Statistics and Operations Research, vol. 74(3), pages 222–241, August 2020
- [5] Zhang, D., Wang, Y. & Zhang, Z. Identifying and quantifying potential super-spreaders in social networks. Sci Rep 9, 14811 (2019). <https://doi.org/10.1038/s41598-019-51153-5>
- [6] Ball, F. G., Mollison, D., & Scalia - Tomba, G. (1997). Epidemics with two levels of mixing. The Annals of Applied Probability, 7, 46– 89.
- [7] Barabási, A. - L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286, 509– 512.
- [8] Balcan D, Hu H, Goncalves B, Bajardi P, Poletto C, Ramasco JJ, Paolotti D, Perra N, Tizzoni M, Van den Broeck W, Colizza V, Vespignani A: Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. BMC Medicine. 2009, 7: 45-10.1186/1741-7015-7-45.
- [9] Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A: Multiscale mobility networks and the large scale spreading of infectious diseases. Proc Natl Acad Sci USA. 2009, 106: 21484-21489. 10.1073/pnas.0906910106.

