Analyzing the Spread of Infectious Disease Using a Probabilistic Model

Shrikant Saxena*, Shivang Khare†, Sujata Pal‡ and Vidushi Agarwal§
Department of Computer Science and Engineering
Indian Institute of Technology Ropar, Ropar, India
* 2017csb1111@iitrpr.ac.in , † 2017csb1109@iitrpr.ac.in
‡ sujata@iitrpr.ac.in , § vidushi.19csz0010@iitrpr.ac.in

Abstract—Infectious diseases are those that can be transmitted from person to person upon some form of contact. In this regard, airborne infectious diseases can wreak quite a havoc as they have a high degree of infectiousness and can easily infect a healthy person who comes in proximity of an infected person for a specific interval of time. The situation can take the form of an epidemic in no time if the outbreak of a disease is not checked at an earlier stage. In this paper, we simulate the spread of airborne infectious disease in the city population. Disease transmission from an infected person to a healthy person is modeled based on proximity and contact time. We analyze how population density affects the spread of disease. Moreover, we also analyze how practices like wearing a mask and hotspot lockdowns might slow down the spread of infection. Finally, we analyze how an epidemic mitigates when a certain fraction of the population becomes immune to the disease. Observations and inferences drawn from the simulation results can help make policies to tackle the spread of airborne infectious disease in a city community.

Keywords: Epidemic, infectious diseases, hotspot lockdown, wearing masks.

I. INTRODUCTION

Urban areas are characterized by high population density. When combined with extensive population mobility, the spread of airborne disease becomes rapid. The Covid-19 epidemic is a real example of such a disease spread. It is an airborne contagious disease caused by the novel coronavirus(SARS COV-2). The disease has affected about 96 million people and is responsible for the death of about 2 million people worldwide as of January 2021 [1].

Several measures are taken care of to slow down the spread of this disease. Pal *et al.* [2] in their research propose a three phased system in which thermal imaging is utilized to identify and quarantine individuals who are suspected to have infection based on their elevated body temperatures. In our paper, we consider two preventive measures employed to mitigate the spread of disease. They are locking down the hotspot areas and wearing a mask. Hotspot area refers to the regions with a high number of infected people. Masks act as a barrier between the susceptible person and the infection source and help prevent infection spread.

A healthy person could catch the infection if he inhaled a certain quantity of disease-causing pathogens from the infected person. The quantity of pathogens inhaled by the person depends upon the proximity and time duration of contact with the infection source [3]. Therefore, we analyze disease transmission through a probabilistic distribution model by using proximity between persons and the time of contact as its parameters. Wearing a mask can significantly reduce the number of pathogens exhaled or inhaled. We have also taken this parameter into account in our simulation model.

The rest of the paper is organized as follows: Section II summarizes some of the previous related works in this area. Section III discusses the epidemic model used in the simulation. Section IV discusses the actual design specifications of the simulation. Section V discusses the results obtained and their analysis followed by the conclusion in section VI.

II. RELATED WORKS

Mittal *et al.* [3] presents a mathematical framework to estimate the risk of airborne transmission of COVID-19. EpiSimS [4] presents a discrete-event-driven stochastic model for the spread of an infectious agent in an urban population. However, EpiSimS does not consider the traveling pattern of people and their encounters with one another. These factors directly affect how an epidemic develops in a population. We have addressed this issue in our research. The epidemic model formulated by us in this paper involves a movement model wherein the people in the city walk along predefined paths to reach their destination. The speed of movement has also been kept within a predetermined range mimicking the actual traveling characteristics of people. Upon reaching the destination, people wait at the destination for a random interval of time chosen from a predetermined range. Moreover, people plan their journey such that their journey path does not pass through an infection hotspot. Mei *et al.* [5] in their research have also implemented such a model using the Geographic Information System (GIS) to model a city's infrastructure prone to an epidemic. However, they have not considered the impact of various preventive measures employed for an infection outbreak. We address this drawback in our paper. We include two epidemic mitigation methods in the simulation, locking-down hotspots and wearing masks to prevent disease transfer. We analyze how these methods are effective in slowing down the spread of epidemic disease in the population.

The specific contributions of this paper are summarised as follows:

1) We simulate the spread of an infectious disease in a city environment based on a probabilistic disease transfer model from person to person.
2) We analyze how the infection spreads in the population and how the population experiences multiple waves of infection events.
3) We analyze how the spread of epidemic varies with population density.
4) We incorporate two epidemic mitigation methods in our simulation, wearing masks by people and locking-down hotspots. We analyze how effective these two methods are in mitigating an epidemic.

III. Epidemic Model

A. Infection transmission model

The airborne transmission of respiratory infections involves the following sequence of events:

1) Viruses containing droplets are generated, released, and aerosolized through the nose and mouth of the infected person.
2) Aerosolized droplets are transported by air currents to a susceptible person.
3) The droplets are inhaled and deposited in the respiratory mucosa of the susceptible person, causing infection.

The infection may get transmitted from an infected person i to a healthy person j upon an event of contact with the probability P_{ij}. The probability that a susceptible person catches an infection while in a particular location depends on: the number of infected people occupying the area, the time duration of visit, activity performed at the location, and the infectiousness of the infected persons (Eubank *et al.* [6]). According to Valle *et al.* [7], σ is the average frequency of disease transmission events per hour of contact between fully infectious and fully susceptible people. For events that occur randomly in time, the number of occurrences of disease transmission in a period of time t obeys a *Poisson* probability law with parameter σt [7], as proposed by Valle *et al.* [7] in their research. Given a random variable X denoting the occurrences of disease transmission in the time interval t,

$$P(X = k) = \frac{e^{-\sigma t}(\sigma t)^k}{k!}, k = 0, 1, 2, \ldots, N \quad (1)$$

As illustrated in Figure 1, a susceptible person is within the minimum threshold distance from n infected persons for a time period t. The random variable X_i represents the infection transmission event from the infected person i to the susceptible person. The probability that there is at least one occurrence of disease transmission event from an infected person to susceptible person can be derived as follows:

$$P\left(\sum_{i=1}^{n} X_i = 1\right) = 1 - \prod_{i=1}^{n} P(X_i = 0)$$

$$\because P(X_i = 0) = e^{-\sigma t}$$

$$P\left(\sum_{i=1}^{n} X_i = 1\right) = 1 - e^{-n\sigma t} \quad (2)$$

We define Δ as the minimum distance of approach in which the disease can be transmitted from an infected person to a susceptible person. If n infected individuals and a susceptible individual stay within a distance threshold Δ (varies with the disease) for a time interval T, the susceptible individual can catch an infection with a probability of $P = 1 - e^{-n\sigma T}$. The proximity of contact, δ is defined as the distance of separation between two individuals. In case $\delta \leq \Delta$, the duration of the contact is recorded and the transmission of infection occurs according to Equation 2. In case $\delta \geq \Delta$, the probability of infection transmission is taken to be zero. σ can be estimated based on the history of the disease.
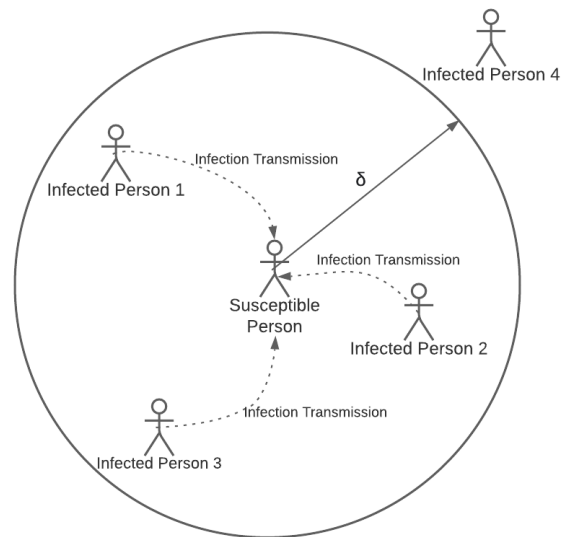

Fig. 1: Disease transmission

B. Effect of wearing a face mask

A face mask prevents the exhalation and inhalation of infected droplets in the form of an aerosol depending upon the quality of the product. A well-fitting N-95 stops about 95% of the particles passing through it. Wearing a proper mask by either the infected person or the healthy person reduces the transmission risk by a factor of about 20 [3]. Our simulation models the city population such that the fraction F_{mask} of the total individuals are wearing a mask. We define a risk mitigation factor R_i associated with each person i. R_i is modeled as follows:

$$R_i \sim \mathcal{N}(\mu_{mit}, \sigma_{mit}), i \in \mathbb{M}$$
$$R_i = 1, i \notin \mathbb{M}, \quad (3)$$

where μ_{mit} and σ_{mit} represent the mean and variance value for the factor by which infection transmission probability is mitigated upon wearing a mask. $\mathbb{M}$ is the set of individuals wearing a mask. When an infected person i and a healthy person j come into contact within the minimum threshold

distance Δ (defined in section III-A), the revised effective probability of disease transmission is given by the following expression:

$$P = \frac{1 - e^{-\sigma t}}{R_i R_j} \tag{4}$$

Various possible scenarios are shown in Figure 2. Extending the case where n infected individuals $i_1, i_2, \ldots, i_n$ come into contact with a susceptible person j within the minimum threshold distance, the effective probability of disease transmission will be:

$$P = \frac{1 - e^{-n\sigma t}}{\sum_{y=1}^{n} R_{i_y} R_j} \tag{5}$$

Where R_i is the infection transmission risk mitigation factor if person i is wearing a mask, as defined in equation 3.
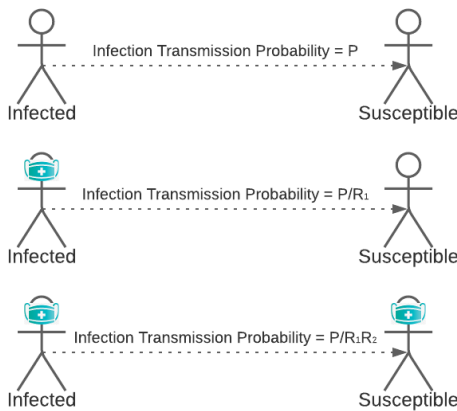

Fig. 2: Effect of wearing a face mask on infection transmission probability

C. Disease Progression

According to Mei et al. [8], once a person is infected, the disease spreads through the following phases:

1) Incubation period: During this period, the person is infected but the symptoms of the disease have not surfaced and hence the person feels no need to take recursive measures.
2) Symptomatic period: The symptoms of the disease surface and the person seeks medical health and other safety measures like self-isolation to prevent the further spread of disease.
3) Recovery: In case a person recovers from the disease, the person remains immune from it for some time interval due to the presence of antibodies.

To simulate the phenomenon, we have taken 3 parameters in our model, τ_{inc}, τ_{sym}, and τ_{imm}. These parameters come from Normal prior distributions.

$$\begin{aligned}
\tau_{inc} &\sim \mathcal{N}(\mu_{inc}, \sigma_{inc}) \\
\tau_{sym} &\sim \mathcal{N}(\mu_{sym}, \sigma_{sym}) \\
\tau_{imm} &\sim \mathcal{N}(\mu_{imm}, \sigma_{imm})
\end{aligned} \tag{6}$$

The parameters $\tau_{incubation}$, $\tau_{symptomatic}$, and $\tau_{immunity}$ vary from disease to disease. The progression of a person from various disease stages are illustrated in Figure 3.
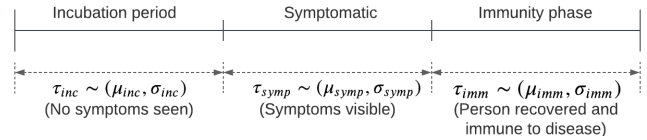

Fig. 3: Stages of disease progression

D. Hotspot identification

The simulated city area is bounded in a $P \times Q$ grid. The area is further divided into sub-regions, the dimension of each region being $S \times T$. Let N and N_G denote the number of detected disease transmission cases in the city and in the grid-square G. Then the grid-square G is declared a hotspot if the following conditions are satisfied:

$$\frac{N_G}{N} \geq \phi, \tag{7}$$

where ϕ is the minimum threshold fraction. Once the grid-square G is declared a hotspot, individuals are not allowed to pass through this sub-region during their commute (hotspot lockdown phase).

E. Movement model

The individuals follow the following movement model in our simulation design:

1) An individual first selects a random destination.
2) The individual then finds the shortest valid path to the destination node taking care that the path does not pass through any hotspot.
3) The individual then follows the path with speeds that vary randomly between thresholds ν_{high} and ν_{low}.
4) Upon reaching the destination, the individual waits for a random interval of time between thresholds T_{min} and T_{max}. The map is composed of multiple map nodes. Each map node has its specific location and is connected to neighbor nodes via predefined paths. The map nodes which are lying within containment zones are temporarily removed from the map. The agent then selects a random map node as its destination and finds the shortest path using Djikstra's shortest path algorithm to reach the destination. Algorithm 1 gives the pseudocode for the random destination selection and path-finding algorithm.

IV. SIMULATION DESIGN

For evaluation of the proposed model, ONE simulator is used. The movement of people is simulated in the form of node movement following the $RandomMapBased$ movement model with restrictions in place so that the nodes do not enter the hotspots. Infection transmission from person to person is simulated through message passing between the

Algorithm 1 Find Movement Path

Input: Map Node list $\mathcal{N}$, hotspot list $\mathcal{H}$, starting point s
Output: Shortest path P to a random destination

1: **procedure** FIND PATH($\mathcal{N},\mathcal{H},s$)
2:     **for each** $n \in \mathcal{N}$ **do**
3:         **for each** $h \in \mathcal{H}$ **do**
4:             $l \leftarrow n \cdot location$
5:             **if** l *lies within* $h \cdot area$ **then** $\mathcal{N} \cdot remove(n)$
6:             **end if**
7:         **end for**
8:     **end for**
9:     $d \leftarrow RandomNode(\mathcal{N})$
10:    $P \leftarrow Dijkstras(s, d, \mathcal{N})$
11: **return** P
12: **end procedure**

TABLE I: Simulation Parameters

Parameter	Value
Number of nodes	[400-850]
Duration	24 hours
Transmission range	6 feet

TABLE II: Values of constant for epidemic model

Model	Parameters	Values
Disease transmission	σ	5
	Δ	6 ft.
Face Mask model	μ_{mit}, σ_{mit}	20, 1
Disease progression	μ_{inc}, σ_{inc}	30 min., 6 min.
	μ_{sym}, σ_{sym}	110 min., 22 min.
	μ_{imm}, σ_{imm}	7 min., 1.5 min.
Movement	ν_{low}, ν_{high}	8 Km/hr, 50 Km/hr
	T_{min}, T_{Max}	5 min., 1 hr.

Fig. 4: Total cases versus time for varying population densities

nodes which come in contact according to Equation 2. Values of various simulation parameters are given in Table I.

The experiment consists of two parts. In the first part, we vary the population density of the city. One of the individuals is randomly infected with the disease and the spread of infection is observed. We repeat the process for different population densities and analyze the outcomes. The second part of the experiment consists of simulating 500 individuals under 4 scenarios as described below:

1) Scenario 1: No individual is wearing a mask and there is no provision for hotspots to be closed.
2) Scenario 2: Half of the individuals are wearing masks. There is no provision for hotspots to be closed.
3) Scenario 3: No individual is wearing a mask. Provision of hotspot identification is in place.
4) Scenario 4: Half of the individuals are wearing masks. Provision of hotspot closing is in place.

The values for various infection model parameters are given in Table II.

In each scenario, one of the 500 individuals was randomly infected with the disease and the outcomes were recorded.

V. RESULTS AND ANALYSIS

A. Increase in total number of cases and variation in active cases

The graph for the total number of cases as shown in Figure 4 features multiple ripples. Each ripple consists of an initial exponential growth period followed by a slowdown period where the rate of epidemic spread reduces. The reason for the exponential growth is the increase in the number of infection sources while the slowdown period is due to the decrease in the number of susceptible people owing to many people already being infected or going through the immunity

phase. These people act as a barrier to infection spread in the population. After these people again become infection susceptible, the graph features another ripple. The epidemic spread rate at the beginning of the simulation is quite slow since the number of infection sources is very less.

The graph for the number of active cases as shown in Figure 5 features a very low rate of epidemic spread at the beginning followed by exponential growth. The reason for this change
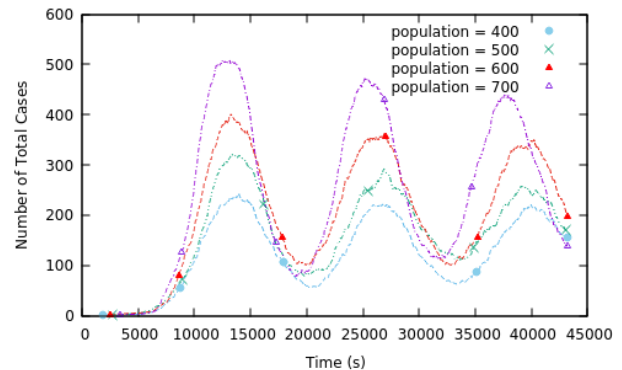

Fig. 5: Active cases versus time for varying population densities

is the increase in the number of infection sources. The graph features maximas and minimas similar to a sine curve. The maximas decrease in magnitude over time. The inferences that can be drawn from these observations are as follows:

1) The rate of infection spread is low in the beginning when the number of infection sources is less.
2) The population faces multiple infection waves over time.

3) The rate of infection transmission increases when the proportion of susceptible persons in the population increases.
4) The rate of infection transmission decreases when the proportion of already infected people and people going through the immunity phase in the population increases.
5) The subsequent waves of infections become less and less severe, shown by the decreasing magnitudes of maximas.

B. Variation in infection spread with change in population density

To study the impact of the increase in population density, we performed the simulation with varying amounts of people in the city environment. As expected, the total number of cases and the average rate of infection spread increased with an increase in population density of the environment. The number of active cases at any time for different city populations is shown in Figure 5. The total number of infection transmissions at the end of simulation under different population densities is shown through the bar graph in Figure 6. The Pearson
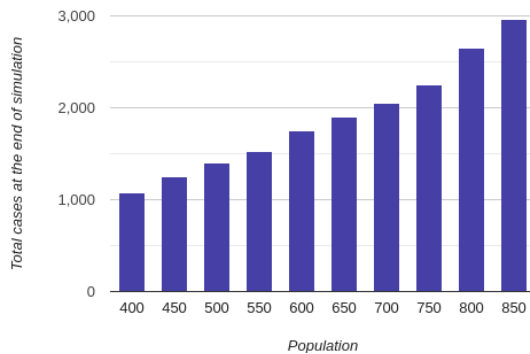

Fig. 6: Total cases at the end of simulation for different population densities

correlation coefficient for the total number of cases versus city population is 0.98. This shows that there is a strong direct relationship between the total number of cases and the population density.

C. Comparison of different epidemic mitigation methods

As previously stated in section VIII, the impact of local lockdowns and wearing a mask in the epidemic is studied using four scenarios listed below:
1) Scenario 1: No mitigation method employed.
2) Scenario 2: Half the city population is wearing a mask.
3) Scenario 3: Local lockdowns are imposed in hotspot areas.
4) Scenario 4: Half the city population is wearing a mask and the provision of local lockdowns is also imposed.

The total number of cases and the active number of cases are depicted in graphs in Figure 7 and Figure 8 respectively. Wearing a mask proves out to be the most effective provision to mitigate the disease spread. The imposition of local lockdowns mitigates the spread of infection to some extent but the

impact is not much pronounced. The mitigation of the spread of an epidemic in various scenarios can be gauged through the bar graph as shown in Figure 9. To find the effectiveness with which the infection spread is mitigated in a scenario, we use the following expression:

$$Effectiveness_i = \frac{N_1}{N_i} \qquad (8)$$

Where N_1 and N_i denote the total number of cases at the end of simulation in Scenario 1 and Scenario i respectively. Values of the total number of cases and the effectiveness of methods in alleviating the disease spread are presented in Table III .
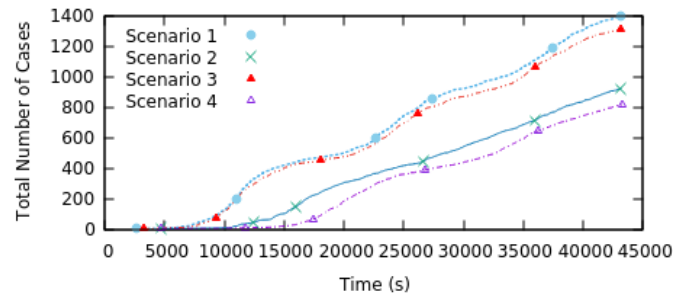

Fig. 7: Total cases versus time for different epidemic mitigation methods
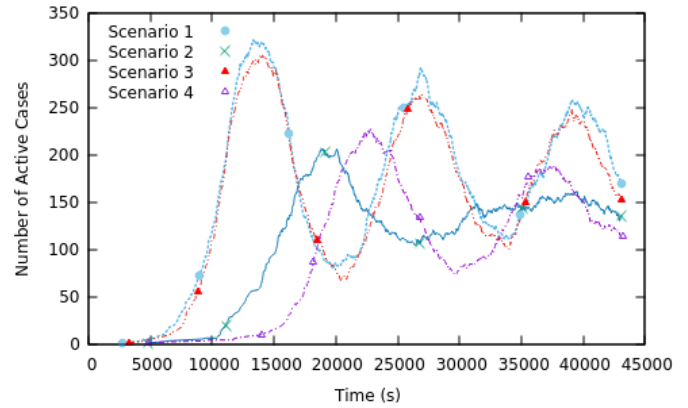

Fig. 8: Active cases versus time for different epidemic mitigation methods

TABLE III: Impact of epidemic mitigation methods in different scenarios

S. No	Mitigation methods	Total cases	Effectiveness
1	None	1394	1.00
2	Wearing mask	922	1.51
3	Hotspot lockdown	1309	1.07
4	Wearing mask, hotspot lockdown	815	1.71

D. More contribution of unmasked people in the spread of an epidemic as compared to masked people

We distribute the infection transmission cases under the following 4 categories:
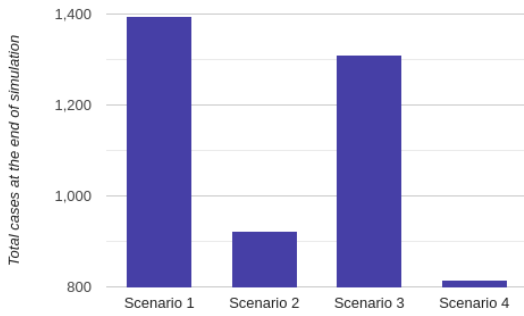
Fig. 9: Total cases at the end of simulation for different epidemic mitigation methods

1) Category 1: Infection transmission from a masked person to a masked person.
2) Category 2: Infection transmission from a masked person to an unmasked person.
3) Category 3: Infection transmission from an unmasked person to a masked person.
4) Category 4: Infection transmission from an unmasked person to an unmasked person.

The number of infection cases under each category is depicted in Table IV and the bar graph in Figure 10. Although the probability of infection spreading from a masked person to an unmasked person and vice versa follows identical probability distribution, the number of cases under category 3 is almost three times the number of cases in category 2.

Let us define the sets $\mathbb{W}, \mathbb{M}, \mathbb{I}, \mathbb{U}$ as follows:

1) Set $\mathbb{W}$ is the set of people in the city population not wearing a mask.
2) Set $\mathbb{M}$ is the set of people in the city population wearing a mask.
3) $\mathbb{I}$ is the set of infected people in the city population.
4) $\mathbb{U}$ is the set of uninfected people in the city population.

Upon analysis, it is found that the number of persons in the set $\mathbb{W} \cap \mathbb{I}$ grows more rapidly than that in the set $\mathbb{M} \cap \mathbb{I}$. This is because the infection susceptibility of an unmasked person is more in comparison to a masked person. Non-mask-wearing people are majorly infected by other non-mask-wearing infected people as compared to mask-wearing infected people since the former outnumbers the latter.
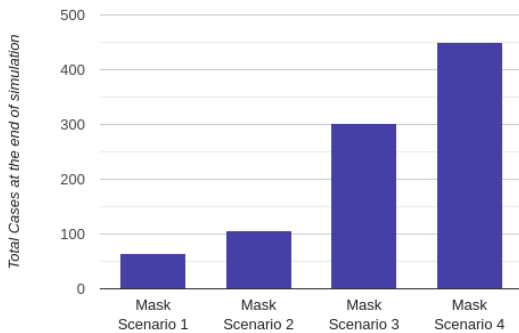

Fig. 10: Total cases at the end of simulation for different mask categories

TABLE IV: Number of infection cases under various categories

Category	Number of cases
Masked to masked	64
Masked to unmasked	106
Unmasked to masked	302
Unmasked to Unmasked	450

VI. CONCLUSION

In this paper, we studied how population density and various preventive measures minimize the spread of an epidemic in a population. There was a strong positive correlation present between the population density and the total number of infected cases. We compared the total number of infection cases with and without any epidemic mitigation methods in some places. Wearing masks came out to be the most effective epidemic mitigation measure. Unmasked people contributed excessively to the spread of an epidemic, not only because of more probability to transmit the disease to an uninfected person but also due to the high probability of getting infected themselves and becoming an infection source. Placing infection hotspots under lockdown also contributed to the mitigation of the epidemic to some extent. Combining both of these methods reduced the total number of infection cases by a factor of 1.7.

Despite our best efforts, the simulation has its limitations. We have not addressed the effect that quarantining the infected persons would have on the spread of infection in the city population. Further, we have also not captured how certain locations in the city (like a mall or cinema hall) tend to have a higher footfall of people throughout the day. We would like to address these limitations in our future works.

REFERENCES

[1] "COVID-19 coronavirus pandemic," Accessed: 20th January, 2021 https://www.worldometers.info/coronavirus, 2008.
[2] S. Pal, P. Sinha, and S. Kumar, "Crowd-sourced centralized thermal imaging for isolation and quarantine," in *Computational Modeling and Data Analysis in COVID-19 Research*, C. R. Panigrahi, B. Pati, M. Rath, and R. Buyya, Eds. CRC Press, ch. 8, p. 20.
[3] R. Mittal, C. Meneveau, and W. Wu, "A mathematical framework for estimating risk of airborne transmission of COVID-19 with application to face mask use and social distancing," *Physics of Fluids*, vol. 32, no. 10, p. 101903, 2020.
[4] S. M. Mniszewski, S. Y. Del Valle, P. D. Stroud, J. M. Riese, and S. J. Sydoriak, "Episims simulation of a multi-component strategy for pandemic influenza," in *Proceedings of the 2008 Spring simulation multiconference*, pp. 556-563, 2008.
[5] S. Mei, B. Chen, Y. Zhu, M. Lees, A. Boukhanovsky, and P. Sloot, "Simulating city-level airborne infectious diseases," *Computers, Environment and Urban Systems*, vol. 51, pp. 97 − 105, 2015.
[6] S. Eubank, H. Guclu, S. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modeling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, pp. 180–4, 2004.
[7] S. Del Valle, J. Hyman, H. Hethcote, and S. Eubank, "Mixing patterns between age groups in social networks," *Social Networks*, vol. 29, no. 4, pp. 539 − 554, 2007.
[8] S. Mei, P. Sloot, R. Quax, Y. Zhu, and W. Wang, "Complex agent networks explaining the hiv epidemic among homosexual men in amsterdam," *Mathematics and Computers in Simulation*, vol. 80, no. 5, pp. 1018 − 1030, 2010.