HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 4 (2023)

Modeling the Spread of Infectious Diseases Within a Social Environment

Andy Liu

San Marino High School

ABSTRACT

The rapid spread and devastating effect of the COVID-19 pandemic made infectious disease mitigation a concern of the whole world. Many countries implemented different policies in order to mitigate the spread of the disease, with varying degrees of success. Each policy/measure was an attempt to find a balance between feasibility and effectiveness. I created a simulation to model a viral outbreak in a population with moving and interacting individuals. Using this model, I tested the impact of various disease prevention or mitigation policies at incremental degrees of strictness. I found that each method displayed an exponential relationship between the thoroughness of its implementation and its effectiveness. The model shows that any approach to mitigate infectious disease spread in densely populated regions should be enforced fully to significantly deter disease spread.

Introduction

The rise of the COVID-19 pandemic has fostered intense interest from both the academic and public communities in strategies to mitigate the spread of disease. Research in this area is not only applicable to the latest COVID-19 outbreak, but also relevant to any highly infectious disease. This can include other worldwide pandemics, such as SARS, or even less severe infections, such as the seasonal flu. Infectious disease mitigation will remain a relevant area of study even as institutions worldwide declare that the pandemic phase of COVID-19 has passed.

One of the important aspects of infectious diseases is the cumulative effect of many individual interactions on the spread of the disease. A single key transmission can sometimes lead to many others through these interactions, so modeling each interaction is crucial to gain an understanding of the basis of disease spread. Many simulations incorporated these interactions in disease modeling using either a number of 'interaction factors' derived from case studies, or other mathematical descriptions. One such mathematical description is the SEIR model, which divides the population into susceptible, exposed, infective, and recovered groups. (Zhang et al., 2022).

An alternative way to model the individual interactions is to explicitly simulate movements of people, as done in(Stevens, 2020). His simulation consisted of a group of people moving in straight lines and bouncing off of the borders of the area and other people. One person was in the infected state at the beginning of the simulation, and the infection spread upon contact. After some time, infected people would recover from the disease. Harry Stevens pointed out that the spread of the infection was exponential and examined the effects of quarantining and two levels of social distancing. Crucially, he noted that it was nearly impossible to fully enforce either strategy, which reduced their effectiveness and was the main barrier to disease mitigation.

In this study, I built on Harry Steven's work by increasing the scale and complexity of the simulation. I attached multiple variables to character movement, devised a more accurate mathematical description of the process of disease transmission, increased the number of characters and simulation area, and validated the model against real data collected on COVID-19. I also explicitly incorporated "hotspots" of interactions in the model to reflect social interactions in the endemic phase of COVID-19, or in annual flu seasons. Using this improved model, I tested the effectiveness of multiple methods of disease mitigation when enforced at different levels of completeness.

Methods

Simulation Setup

A 1000 characters are generated in an area of 960 x 540 pixels with a social gathering node of radius R_N (150) at the center of the area (Figure 1A). A grouping constant G with value 0.7 determines the probability that a character is assigned to one of the social gathering nodes at the beginning of the simulation. Characters that are assigned to a node will be placed inside the node radius. Characters not assigned to a node are placed randomly within the simulation area. During every frame of the simulation, each character has a set "migration probability", in the range of 0.1% – 0.01%, to be reassigned with probability G to a node or no node.

For simplicity, most simulations used a model with one node. In addition, a model with 2 symmetrically placed social gathering nodes was also tested. In this model, each node was placed at ⅕ of the simulation area width away from the vertical borders. Characters that are (re)assigned to a node are distributed randomly between the two nodes.

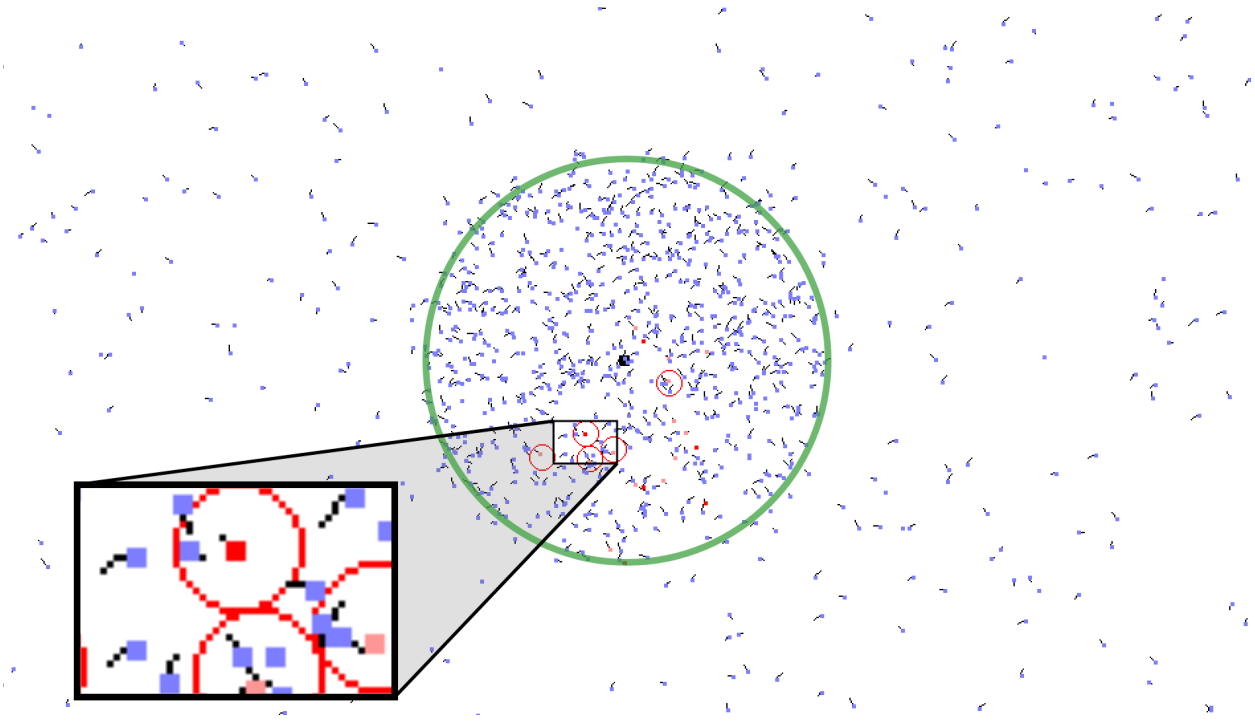

Figure 1. Visual representation of the model setup. The green circle depicts a social gathering "node" , blue and red dots depict healthy and infectious (incubating or diseased state) characters , black lines depict character movement vectors , and red circles depict the infection radius. The simulation area is 960 px by 540 px, and the node radius is 150 px.

Characters

People in the simulation are represented by characters, which are visually depicted as small red or blue squares depending on their condition (Figure 1A). Characters can either be healthy, incubating, diseased, or immune. Healthy is the default state. Upon infection, they will enter the incubating state in which they can infect healthy characters but otherwise do not display any properties of the diseased state. After a set amount of time, incubating characters

will enter the diseased state. Diseased characters can still infect healthy characters but have reduced movement speed and have a defined chance to die and be removed from the simulation every tick. If a diseased character is quarantined, its movement will be halted, and it will lose its ability to infect other characters for the duration of the quarantine. If a diseased character survives the disease for a set amount of time, they will recover and enter the immune state. They are no longer infectious and are now the same as a healthy character except with a dramatically reduced probability to be reinfected.

A tick occurs at regular intervals 10 times every second and determines when incubating and diseased characters attempt to infect healthy characters. During every tick, healthy characters within a defined radius of an incubating or diseased character, termed the infection radius, will have a defined probability (P) to be infected. P is defined by

Equation 1:

$$P = R * I^N * (1 - M_S) * (1 - M_0) \quad (1)$$

Table 1. Definition of variables in Equation 1.

Variable	Meaning
P	Probability of infection
R	Infection rate (parameter)
I	Immunity multiplier (parameter)
N	Number of times the character has previously been infected
M_s	The effectiveness of the mask the target is wearing as a decimal from 0 to 1. Equal to 0 if the target is not wearing a mask.
M_o	The effectiveness of the mask the transmitter is wearing as a decimal from 0 to 1. Equal to 0 if the transmitter is not wearing a mask.

Character Movement

A movement vector, defined in Equation 2, is generated randomly for every character. For characters assigned to the node, a node attraction vector is added to their movement. The attraction vector is a normalized vector pointing towards the node multiplied by the node attraction constant, which is a parameter of the simulation. For healthy and immune characters, an avoidance vector is added pointing away from diseased (not incubating) characters when they are within the infection radius of the diseased character.

Equation 2:

$$\vec{v} = (\langle a, b\rangle * S + \frac{\langle x_n - x, y_n - y\rangle}{||\langle x_n - x, y_n - y\rangle||} * N - \frac{\langle x_c - x, y_c - y\rangle}{||\langle x_c - x, y_c - y\rangle||} * A) * W \quad (2)$$

Table 2. Definition of variables in Equation 2.

Variable	Meaning
$\vec{v}$	Movement vector of the character for that frame
a, b	Random variables from -1 to 1, inclusive. Represent base x and y random movement increments
S	Character speed constant (parameter)
x_n	X position of the node to which the character is assigned.
x	X position of the character
y_n	Y position of the node to which the character is assigned.
y	Y position of the character
N	Node attraction constant (parameter)
x_c	X position of the first diseased character whose infection radius the character came within.
y_c	Y position of the first diseased character whose infection radius the character came within.
A	Avoidance constant (parameter)
W	Infected character slowdown constant (parameter)

Simulation Runs

At the beginning of each simulation, one character is given the diseased state, and all other characters are in the healthy state. The simulations ran as described above for 5000 frames. In each frame, the number of characters in each state, the number of characters that have died, the average number of times each character has been infected, and the R_0 (average number of times each infected character has transmitted the disease) of the disease are recorded in a 2d array. The simulation was repeated 5 times in most cases for every set of testing conditions, and the average of the results from the replicate runs are reported.

Simulation Parameters

Table 3 summarizes the value of parameters used in the simulation. The relative incubation time and recovery time were obtained from another study (Byrne et al., 2020). All movement speed and weighting values were randomly chosen values that gave reasonable simulation results. Two parameters, the infection probability and node radius, were obtained by iterative simulations and adjustments until the infection curve best matched the data of COVID-19 spread in Los Angeles (CSSEGIS, 2020/2023). Simulations were run with a starting guess for the target parameters, and the value of the parameter was incrementally adjusted to reduce the difference between simulated results and real data. When the simulation output was within a certain range of accuracy with the real data, the increment value was decreased. This process was repeated until the target parameters gave simulation results that reproduced the Los Angeles data.

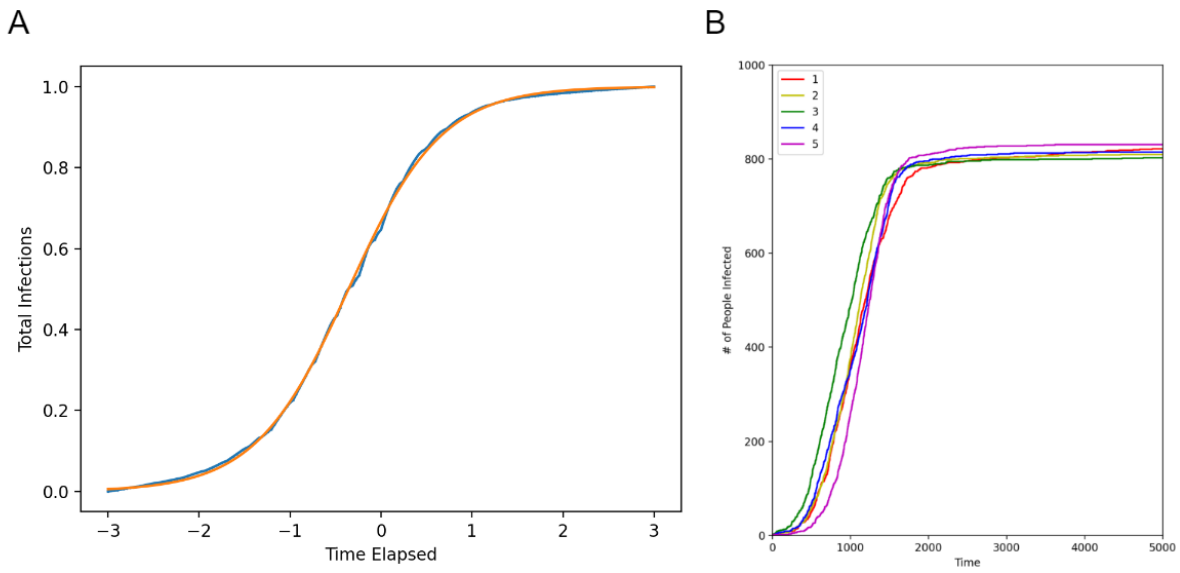

Figure 2. Model validation. (**A**) Comparison of normalized cumulative infection curves for COVID-19 from Los Angeles (blue) and from the simulation (orange). Both time and number of infections were normalized for comparison purposes and in order to optimize the value of some of the parameters in the simulation (see METHODs). The time scale between real life and the simulation is approximately 0.03 days/frame, or 2 days/sec. (**B**) Simulated cumulated infection curves from 5 five replicate runs using baseline settings.

Results

A dynamic model for infectious disease spread

Figure 1 displays the base setup for the simulation. The area in the center with a radius R_N (green circle, center shown as black square) is termed "node" and represents a social gathering center. The characters, denoted by the smaller blue and red squares, move in a randomly determined path that is weighted towards the node (see METHODS). The characters' movement is shown by the protruding black lines in Figure 1. Characters can either be healthy, incubating, diseased, and immune, as defined in the METHODS. Red depicts incubating and diseased characters who are infectious; these individuals have a probability to infect nearby blue healthy characters within an infection radius, shown as the red circle in the inset. This probability is altered based on whether each character is wearing a mask. Infectious characters also have reduced movement, as shown by the movement lines in Figure 1b. The parameters of the simulation were either taken from public data on COVID-19, or obtained from fits of the simulation curves to COVID-19 data in Los Angeles (see Methods).

The main outputs of the simulation are cumulative infection curves, which graph the number of people that have been infected over time since the outbreak of the disease (example in Figure 2B). These curves are sigmoidally shaped, with an initial lag, an exponential rise phase, and reach a plateau at the end of the simulation, at which point all characters are either healthy or immune. The simulation model was validated by testing the reproducibility of multiple runs and by comparison to real life data. In Figure 2A, the infection curve from COVID-19 data in Los Angeles is compared to the simulation results from the baseline model. The time axis is normalized using the points at which the total number of infections reaches 50% and 90% of the total as calibration points. The good agreement between the two curves suggests that my model gives a reasonable approximation to situations during an infectious disease outbreak. Simulation runs were repeated to assess the robustness of the model. Figure 2B compares the out-

come of 5 simulation runs using the same baseline settings, which show that the variation between replicate runs is <1.8% from the average.

Table 3. Values, definition, and additional information on the parameters used in the simulation.

Parameter	Value	Definition or Defined in	Source
Number of Characters	1000		Arbitrary
S	300 pixels / sec	Equation 2	Arbitrary
Character Avoidance	100 pixels / sec	Equation 2	Arbitrary
G, grouping probability	0.7	The proportion of characters assigned to a node.	Arbitrary
Migration Chance	0.005 / sec		Arbitrary
R	150 pixels	Radius of the node	Parameterized based on the data from Los Angeles (CSSEGIS, 2020/2023)
Node Attraction	40 pixels / sec	Equation 2	Arbitrary
Number of Nodes	1 or 2		Arbitrary
Incubation Time	2 sec		(Byrne et al., 2020)
Recovery Time	10 sec	.	(Byrne et al., 2020)
Infection Radius	10 pixels		Arbitrary
Infection Rate	0.078		Parameterized based on the data from Los Angeles (CSSEGIS, 2020/2023)
Infection Cooldown	0.1 sec	This regulates how often the infection chance is applied	Arbitrary
Immunity Multiplier	0.1	This decreases the chance of reinfection upon recovery	Arbitrary
Death Chance	0.003	This chance is applied at the same time as infection rate is checked.	Arbitrary
Infected Slowdown	0.5	Equation 2	Arbitrary

This model allows me to further extract information on how many people were infected overall, as reflected by the total number of infections at the end of the simulation, as well as how long it took the disease to spread, as reflected by the time it took to reach 90% of maximal infection. These graphs make it easier to see the correlation between the tested variable and the effectiveness of disease mitigation.

Simulation Results: the effect of disease mitigation policies

In order to evaluate the effectiveness of public health policies, I tested the impact of masking, quarantining, and travel reduction on the spread of disease.

Masking

Figure 3 compares the infection curves for populations with increasing proportions of mask-wearing individuals. Increasing the percentage of people wearing masks decreases the number of people in the population who are infected (Figure 3B). Masking also increases the time it takes for the infection to spread (Figure 3C). However, the effect of masking does not correlate linearly with the percentage of people wearing masks. Mask wearing for only 20% and 40% of the population does not significantly delay disease spread and only reduces the total number of infections by 7.6% and 19.0%, respectively. On the other hand, the difference in total number of infected characters between 60% and 80% masking is 3 times that between 0% and 20% masking.

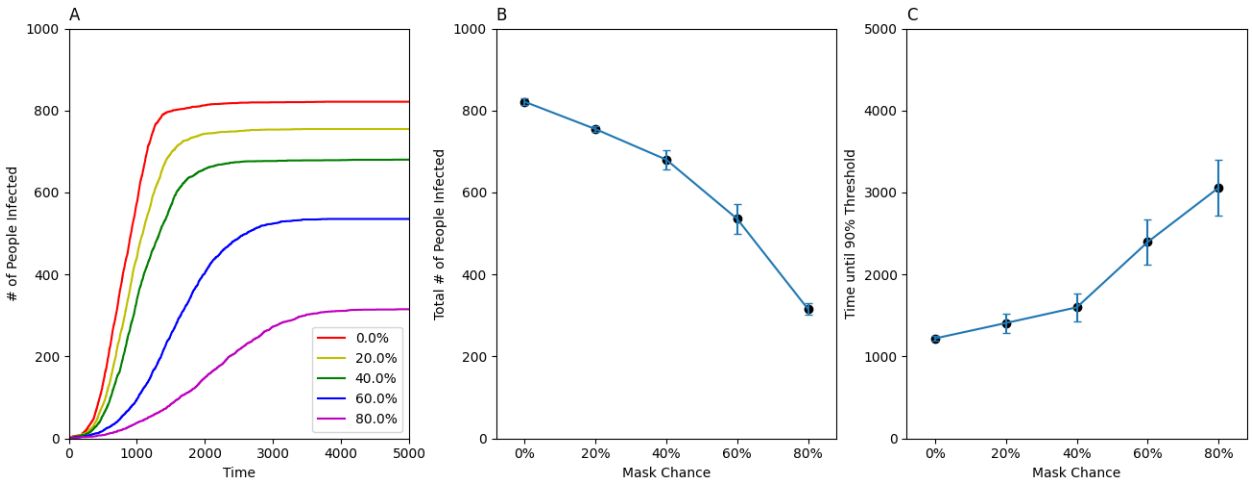

Figure 3. (**A**) Cumulative infection curves for populations in which 0%, 20%, 40%, 60%, and 80% of people wear masks are shown in red, yellow, green, blue, and purple, respectively. Mask effectiveness was set at 90%. (**B**) The total number of people infected at plateau in populations in which 0%, 20%, 40%, 60%, and 80% of people wore masks. (**C**) Time to reach 90% of the maximum value. Values are shown as mean ± S.D. From 5 simulations.

A similar, but less pronounced pattern is observable when varying the effectiveness of the masks themselves (Figure 4). Masks with at least 75% effectiveness are required to substantially delay disease spread, and masks with a minimum of 90% effectiveness are required to significantly reduce the total number of infections. The effect of masks is significantly larger for mask effectiveness ratings of 90% and 95%.

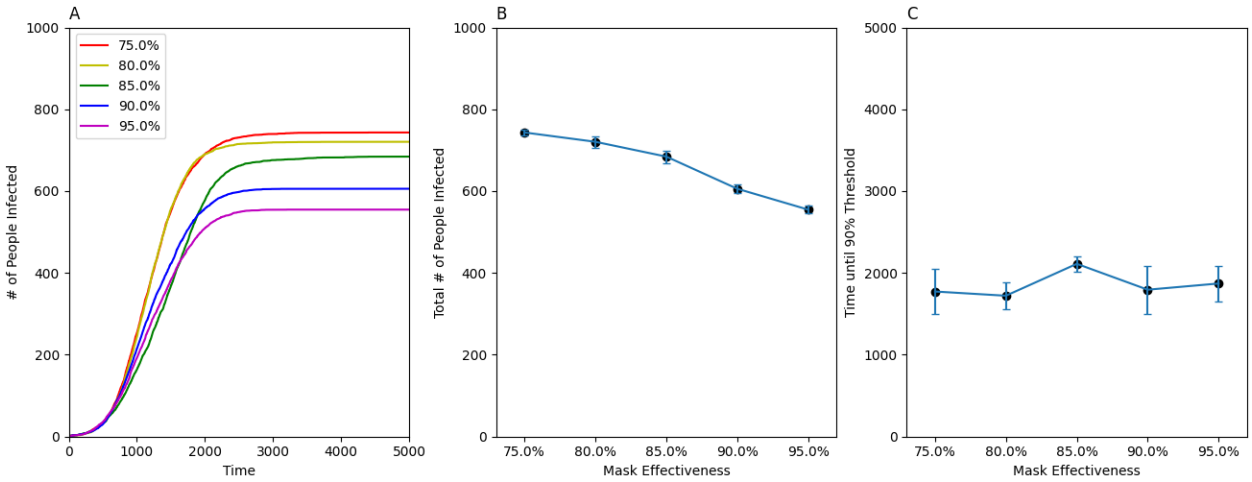

Figure 4. (**A**) Cumulative infection curves for populations in which 50% of people wore masks that block 75%, 80%, 85%, 90%, and 95% of transmissions are shown in red, yellow, green, blue, and purple, respectively. (**B**) The total number of people infected in populations at the plateau as a function of the effectiveness of masks. (**C**) Time to reach 90% of the maximum value. Values are shown as mean ± S.D. from 5 simulations.

Quarantining

Another prevention policy investigated was quarantined. Upon infection, individuals would have a set chance to "quarantine", preventing them from infecting other characters until they recover from the disease. Figure 5 compares the infection curves for populations with a 0%, 20%, 40%, 60%, and 80% probability to quarantine upon infection. Quarantining did not have a major effect until at least approximately 60% of the infected individuals are quarantined. Quarantining also delays the spread of infection more substantially than reducing the total number of infections. For example, when 80% of infected individuals were quarantined, the time to infect 90% of the characters increased from ~1100 to ~2800 frames, whereas the total number of infected people was reduced by 19.6%.

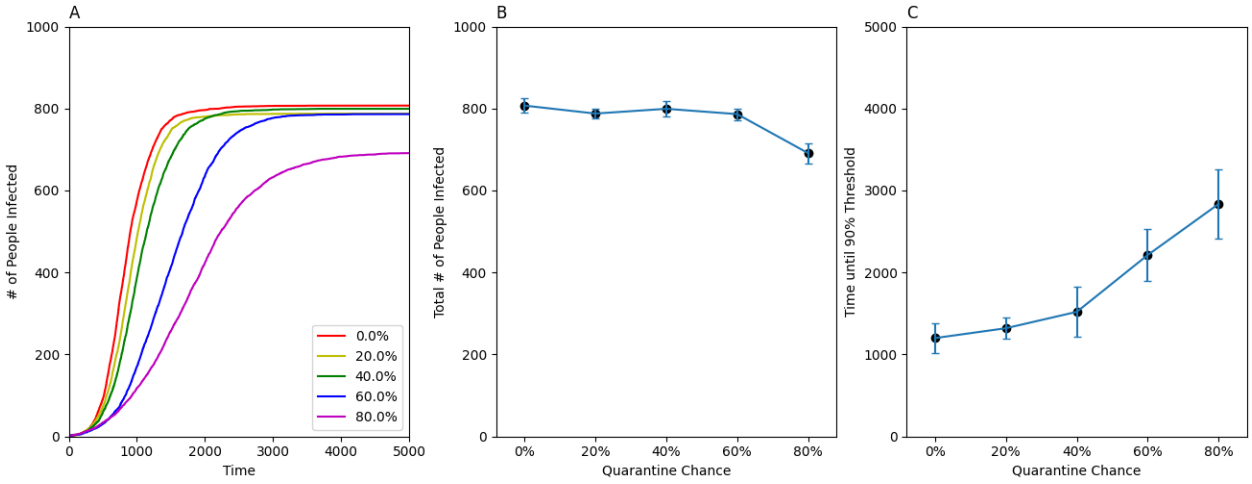

Figure 5. (**A**) Representative cumulative infection curves for populations in which 0%, 20%, 40%, 60%, and 80% of people quarantine upon infection are shown in red, orange, green, blue and purple, respectively. (**B**) Total number of people infected in populations in which 0%, 20%, 40%, 60%, and 80% of infected individuals quarantine. (**C**) Time to reach 90% of the maximum value. Values shown are mean ± S.D. from 5 replicate runs.

Travel Reduction

Finally, I investigated the effects of having multiple social gathering nodes and altering the amount of travel between them. Figure 6 is a visual representation of a simulation setup with multiple nodes. Characters are either assigned to a specific node or no node at any given time. They have a chance to be randomly reassigned every frame (termed migration probability), which causes them to migrate from one node to another and simulates traveling between social gathering hotspots such as major metropolitan cities.

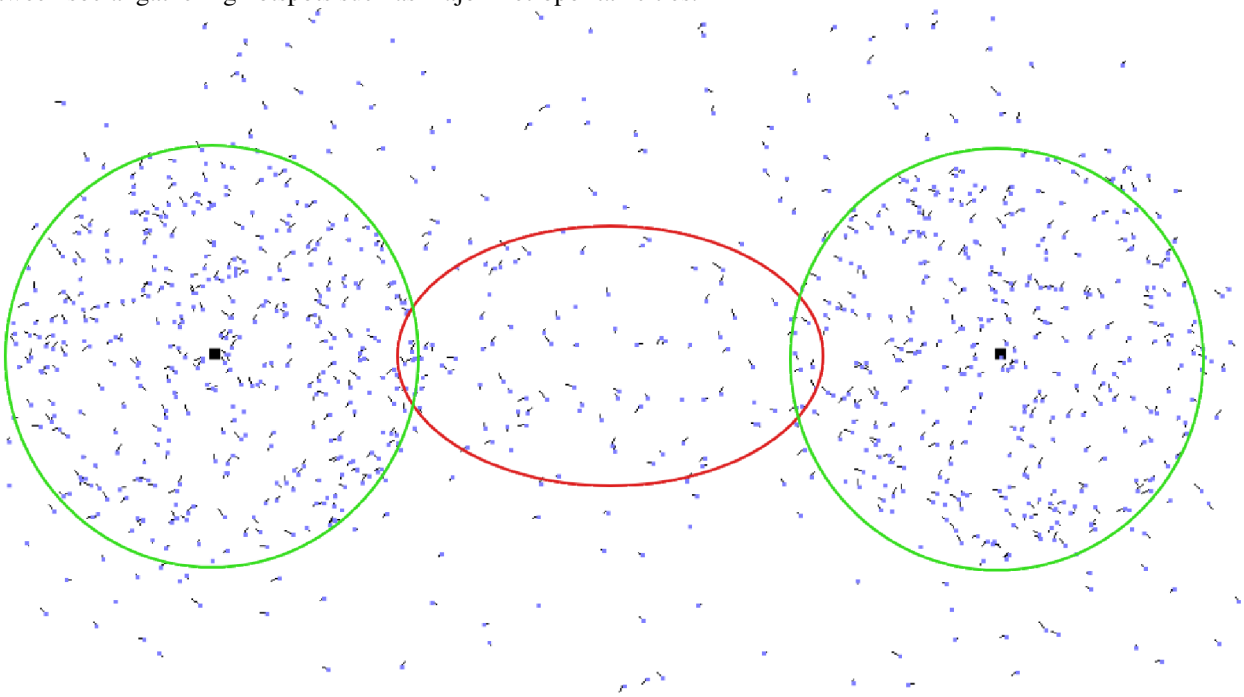

Figure 6. Visual representation of the model with two social gathering nodes. Each node is shown as a green circle. Migrating characters are circled in red.

Figure 7 shows the infection curves for a population that could migrate between two nodes with probabilities of 0.1%, 0.05%, 0.03%, 0.02%, and 0.01% per second. In real life, this would equate to limiting the amount of direct travel between two cities. At lower migration probabilities, the infection curves are separated into two sigmoidal phases. The first phase reflects disease spread in the node in which the outbreak initiated, and the second phase reflects infection spreading to a node nearby due to migration from the initiating node. Decreasing the amount of travel between nodes decreased the speed at which the disease spreads in the non-initiating node as well as the total number of infections in that node. However, significant impact was only observed when migration probability is lowered to 0.02% and below.

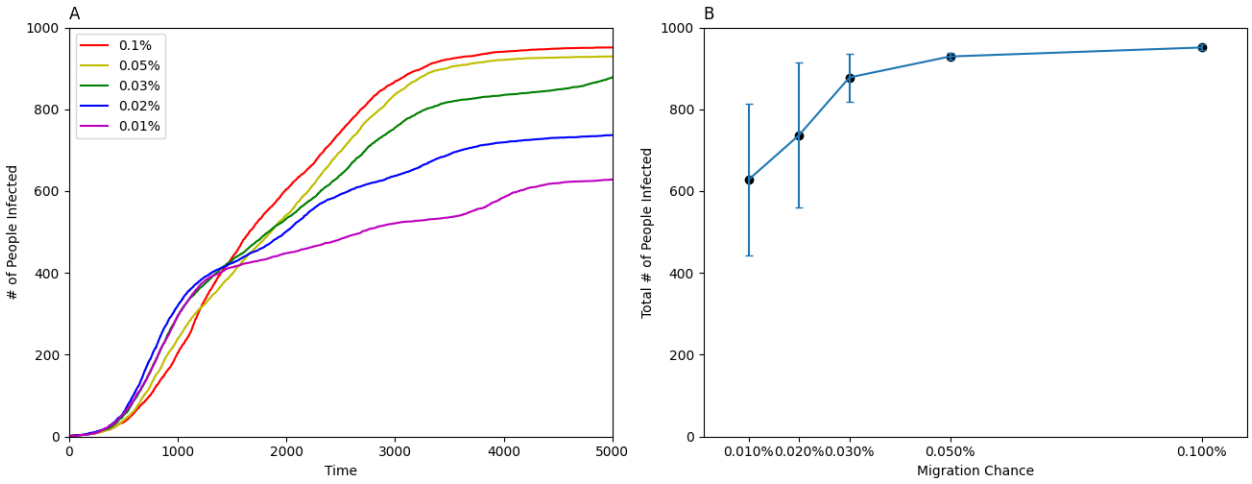

Figure 7 (**A**) Representative cumulative infection curves for populations that had a 0.1%, 0.05%, 0.03%, 0.02%, and 0.01% migration chance per second are shown in red, orange, green, blue, and purple, respectively. (**B**) Total number of infected individuals at the end of the simulation for populations with a 0.1%, 0.05%, 0.03%, 0.02%, and 0.01% migration chance per second. Values shown are mean ± S.D. from 5 replicate runs.

Discussion

From the data in Figure 3, it can be concluded that the number of people wearing masks greatly affects the spread of the disease. Furthermore, masking appears to be increasingly effective as the usage rate approaches 100%. This is likely due to the fact that, as more people wear masks, the chances that both characters in a potential transmission interaction are wearing masks increases, in which case transmission is nearly impossible because the effects of masks compound in that scenario. This level of protection may be required to significantly slow the spread of disease in this model due to the extremely high concentration of characters in the area of the node.

The same reason may underlie why masks that block less than 75% of transmissions had negligible effects on the infection curve. In a population where 50% of people wear masks, the total number of infections approaches half that of populations with no masking at all as the effectiveness of masks approaches 100%. It is probable that the masks are mainly preventing the mask-wearing individuals from being infected, as opposed to preventing them from spreading it to others. This is again due to the high density of the population in the social gathering node, as every character within the node radius is exposed to infectious individuals many times. In the absence of quarantine, individuals without masks will eventually be infected upon encounters with infectious characters.

Quarantining diseased individuals is one of the most feasible mitigation strategies. My simulation showed that this strategy can be effective in delaying the spread of disease, but not as effective in reducing the total number of cases. However, the effectiveness of quarantining diseased individuals was only significant at levels of 80% or above. Quarantining does not prevent healthy characters from being infected; it prevents infected characters from spreading the disease. This means that in order to be effective, it must be enforced thoroughly enough to prevent contact with any diseased characters. In an extremely dense population as simulated in the model, most of the population must quarantine upon infection in order for the policy to be effective.

While reducing the amount of travel between two social gathering centers significantly slowed the spread of the disease in the non-initiating node. However, the only way to reduce the total amount of infections would be to impose strict restrictions on travel between the two nodes, or at least decrease travel sufficiently such that no diseased characters migrated before recovering or infecting healthy characters.

Overall, the simulation results suggest that any individual health policy that aims to mitigate the spread of an infectious disease in a densely populated area can work, but these policies must be enforced to near completion. Methods used to stop disease's spread do not display a linear relationship with their effectiveness, but rather, grow exponentially more effective as their implementation nears 100%.

References

Byrne, A. W., McEvoy, D., Collins, A. B., Hunt, K., Casey, M., Barber, A., Butler, F., Griffin, J., Lane, E. A., McAloon, C., O'Brien, K., Wall, P., Walsh, K. A., & More, S. J. (2020). Inferred duration of infectious period of SARS-CoV-2: Rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open*, *10*(8), e039856. https://doi.org/10.1136/bmjopen-2020-039856

CSSEGIS. (2023). *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*. https://github.com/CSSEGIS/COVID-19 (Original work published 2020)

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

Pete Shinners (2011). PyGame - Python Game Development. Retrieved from http://www.pygame.org

Stevens, H., *Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve"—Washington Post*. (n.d.). Retrieved June 5, 2023, from https://www.washingtonpost.com/graphics/2020/world/corona-simulator/

Zhang, Y., Tao, Y., Shyu, M.-L., Perry, L. K., Warde, P. R., Messinger, D. S., & Song, C. (2022). Simulating COVID19 transmission from observed movement. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-07043-4

Appendix

The code files for the simulation and graph are available at the following link:
https://github.com/AndyLiu0/COVIDsimulation