An Agent-Based Model to Analyze COVID-19 Outbreak and Policies at Stanford University

Kimia Fereydooni, Kyle Depaz Feliciano, Seraj Desai, Thibaut Badoual

Abstract

The COVID-19 pandemic has had a profound impact on education at all levels. Elementary schools, high schools, and universities alike are grappling with the challenges of educating students safely. Universities have taken various approaches to reducing infection within their campuses. As Stanford University approaches its first in-person quarter for undergraduates in the winter of 2021, several policies on social distancing, testing, and virtual classes have been released. The goal of this paper is to model the spread of COVID-19 on campus using an agent-based model to investigate how the number of susceptible, infected and recovered individuals are influenced by policies surrounding which student spaces are safest to re-open first, how testing frequency can minimize further infection, and how differences in mask and social distancing protocol depending on the space can affect infection rates. This model changes recovery rate, quarantine rate, contact rate, and infectious period. Our results have shown that opening up different campus zones is feasible AS LONG AS there is a corresponding increase in testing frequency.

Introduction

Objective: The aim of this paper is to develop an agent-based model that can help Stanford University leaders test the effectiveness of multiple COVID-19 intervention protocols while tracking SIR (Susceptible - Infected - Recovered) statistics. This model allows for a detailed analysis of individual agents or students as they move from compartment (room) to compartment. This is summed into population-wide statistics. Several locations on campus will be modeled while the mask and social distancing protocols in each are modulated. This can help provide recommendations on which public spaces may be successfully reopened, and under what conditions.

Background and Justification: Compartmental models have been used in many fields, including biology, chemistry, and ecology [2]. Most of these models use ordinary differential equations (ODEs) to formulate and solve the system numerically. Although they are effective when analyzing and predicting system behavior, they retain several disadvantages. Compartmental models are computationally expensive due to the fine discretization needed to achieve desired accuracy. Additionally, in order to obtain differential equations, the underlying set of assumptions may be unrealistic, physically meaningless, or difficult to justify. One should also consider the information lost inherent in the process of first homogenizing a discrete system to solve the differential equations, then re-discretizing the results into nodal values [1, 9]. Furthermore, while PDE models enable the analysis of spatial developments via diffusion terms, this type of mixing doesn't physically represent the mixing of infected populations. This is because diffusion will always cause mixing from high-to-low concentrations, whereas true mixing of infected people can go from low-to-high concentrations due to the chaotic and stochastic nature of mobility.

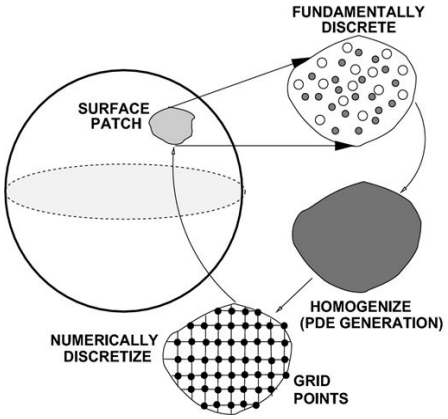


Figure 1. Process of developing a continuum model from a fundamentally discrete system. Reproduced from Zohdi [9].

On the other hand, Agent-based models (ABM), which are computationally similar to particle interaction models, are able to track individual agents based on rules of engagement, and are easy to implement. ABM brings significant benefits when applied to human systems and it is best to use them when the interaction between the agents are nonlinear, complex and discontinuous as is the case with the contagious disease spread. Most importantly, they organically capture the stochastic nature of mixing, and is well-suited for smaller populations, such as for Stanford University.

Some ABM models applied to Covid-19 have already proven themselves. One of the widespread methods [3] is to spatially model each individual in a grid. Individuals can move randomly around them. If they come into contact with an infected person, they have a certain probability of getting sick. This contact can take two forms. Either all individuals are small beads that when they collide with each other can transmit the virus. Or each individual moves independently of the others and can be infected if in its close perimeter (defined by a radius) an infected person is present. This type of ABM, although very visual and realistic at first glance, was not chosen for this project for the simple reason that it loses all its meaning on a small scale. Representing the movement of students in a gym or dining hall in this way is far from ideal and realistic.

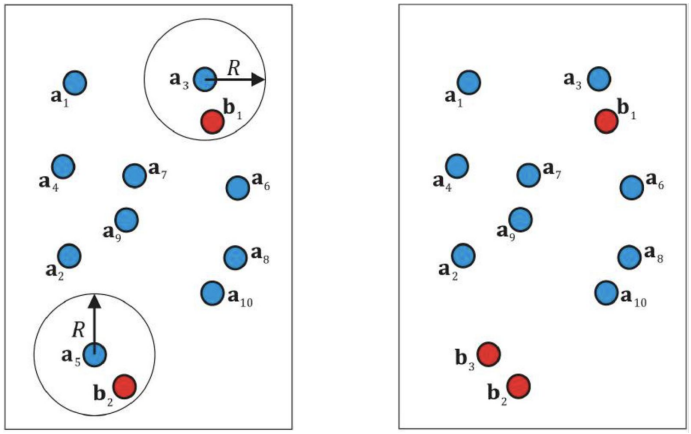


Figure 2. Example of an ABM [3]. Each individual is represented by a sphere. In red, the individual is infected. R is the radius of contamination.

Guidance:

The goal of this paper is to answer the following questions:

- Can Stanford reopen public spaces to students, and to what extent?
- How quarantining would change the total number of infected people in time?
- How much more testing should Stanford do by opening certain spots?

In addition, the results will be compared to the compartmental SIR models and the “equivalent contact rate” for each location will be calculated based on the plots.

Conclusion:

By analyzing each agent individually based on the rules of interaction, we were able to find that reopening of higher-infection zones (i.e. gym) MUST correspond with increased testing frequency. The reopening of higher-infection zones does not pose a significant danger as long as infected students are immediately cognizant of their infectiousness and follow quarantine protocols. This has the effect of reducing the “equivalent contact rate”.

Methods

Epidemiology model and Data:

The model begins with 3346 students of the distribution shown in figure 3, 228 of which are already infected with COVID-19. The percentage of infected individuals seeded into this initial population is dependent on data retrieved from John Hopkins' Coronavirus Resource Center [4], The Covid Tracking Project [5], and Mayo Clinic [7, 8]. 7-day positivity rate averages for each state were used to predict the number of infected individuals coming back to Stanford University in the winter quarter. We assumed that the population of individuals from each state would have the same positivity rate as the positivity rate in their respective state of origin. Data on the student population was obtained from [6].

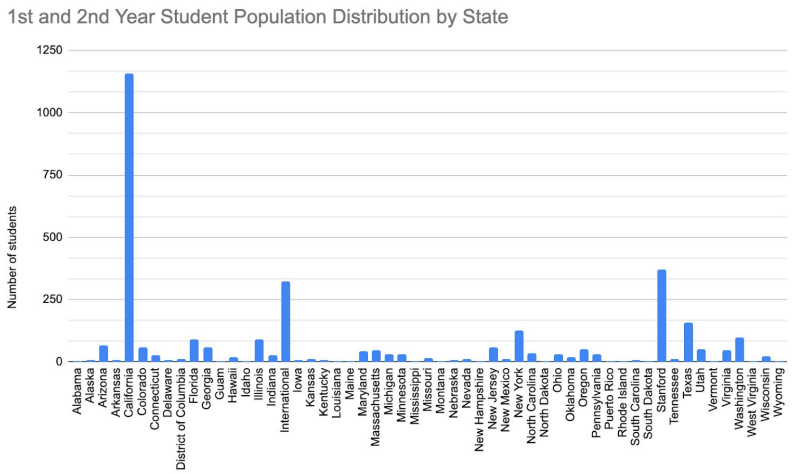


Figure 3. Distribution of 1st and 2nd year Stanford University students by state

The model contains four ideal types of cells, including a residential cell, an outdoors cell called "The Oval," an indoors active cell called "Gym," and a dining hall cell, called "Dining Hall". Each location is assigned a protocol based on the environment and social policies relevant to that type of location, which is mathematically represented by the Poisson distribution parameter, λ. This parameter is used later in the model to sample the contact period (B) of each agent from the Poisson distribution.

For example, let's assume the λ value for the gym and the oval is 1.15 and 0.6 respectively. Figure 4 shows the Poisson distribution for these λ values. As seen from the plots, there's a higher chance for someone at the gym to infect more people compared to someone at the Oval which is because of the fact that the λ value at the gym is more than that of the Oval. As a result, by just adjusting the value of λ, we can control the number of people that can get infected in each location.

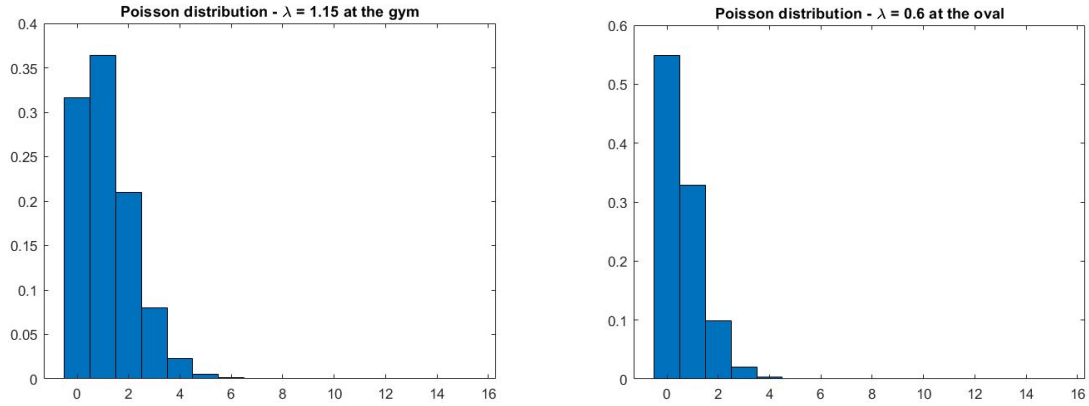


Figure 4. Poisson distribution for two values of λ

Note that to sample the contact period of each agent at a specific location, other distribution functions like Gaussian or Lognormal could also be used; however, we decided to use Poisson distribution because of two reasons: 1. Since Poisson distribution is defined by a single parameter, λ, we only had to introduce one variable to each location, whereas if we wanted to use Gaussian distribution, we had to introduce two variables (mean and standard deviation). 2. Poisson distribution outputs only integers which makes the most sense as the number of people that can get infected by one agent at each time step has to be an integer.

Location	Environment Indoors/outdoors	Social policy mask/no mask	λ
Gym	Outdoors	No mask	β
Student housing	Indoors	No mask	$1.5 * \beta$
Dining hall	Indoors	Mask	β
The oval	Outdoors	Mask	$0.8 * \beta$

Table 1: Infectiousness Parameters

At each time step of the simulation (which represents one full day), each agent interacts with n other agents, n sampled from Poisson distribution based on the location. For example, since λ has a higher value in the dining hall compared to the oval, an infected person in the dining hall will likely infect more people compared to the same person in the oval. After each time step, agents get assigned to new locations (based on their transition probabilities) and the simulation continues the same way for a total number of days. Note that the new location assignment is weighted differently for each location based on its popularity. For example, the gym is more popular than the oval, so students end up there 30 percent of the time whereas this value is 10% for the oval which has less popularity.

Model's assumptions regarding virus transmission:

1. Agents can only infect each other when they are in the same location. This is because COVID-19 gets transmitted from person to person if they are within a certain physical distance from each other.
2. No infections can occur while agents are moving from one location to the other. This is based on the assumption that the students don't have many direct interactions with each other when they are on their way from one place to another compared to when they are already at a location with a certain number of people. This assumption was made to make the computations easier as we don't need to worry about the transition process between relocations.

To account for the recovery of agents, each agent is assigned an infectious period which is sampled from a Gaussian distribution with mean and standard deviation of $1/\gamma$ and 2 days respectively. In the model, once the total number of the agents' sick days pass their infectious period, they get recovered and never get infected for the rest of the simulation, meaning that they become immune to the virus.

Finally, to take into account the effect of a quarantine imposed on infected persons, a last location has been added. In this location, no new infections can take place, patients are all isolated from each other and they are only allowed to leave once they are fully recovered. The quarantine procedure is as follows. At each time step, every infected person has a certain probability of testing positive. To do so, he or she must take a test and have a positive result. Since Stanford's testing policy is to encourage students to be tested every week, a "day sick" variable is created, between 3 and 6, representing the number of days on average that an individual is infectious without testing and going into quarantine. In order to take into account the students who do not comply with Stanford's testing guidelines, a random variable is added, assuming that 20% of the students do not get tested regularly. Thus, any infected person who complies with the guidelines and whose number of days since infection is greater than "day sick" is sent to quarantine.

Mobility model:

At each timestep of the simulation, the code loops over all the locations and all the agents in that specific location. As an example, let's assume the first location is the gym which has a total number of 60 students in it on a specific day. The model loops over each student individually and samples the number of interactions they have with other students based on the λ of that location. For example, student A in the gym is sampled to have 3 interactions with a susceptible, infected and recovered student. Out of these interactions, only the one with the susceptible student results in a new infection because an student who already is infected can't get infected twice and the recovered student is immune to the virus (based on the model assumptions). This is the process in which each student gets infected at each location and time step. Regarding the recovery modeling, at each time step, the number of infected days of an infected student is stored and if that number is more than the recovered days assigned to that person, then the person is assumed to be recovered. And finally, each student is assigned a value which represents how many days it takes for them to realize they have been infected. Once their number of infected days reaches that value, they are moved to the forth location, quarantine, and are kept there until they get recovered. This number is pretty practical as it can represent how often students should get tested, or how accurate the test results are.

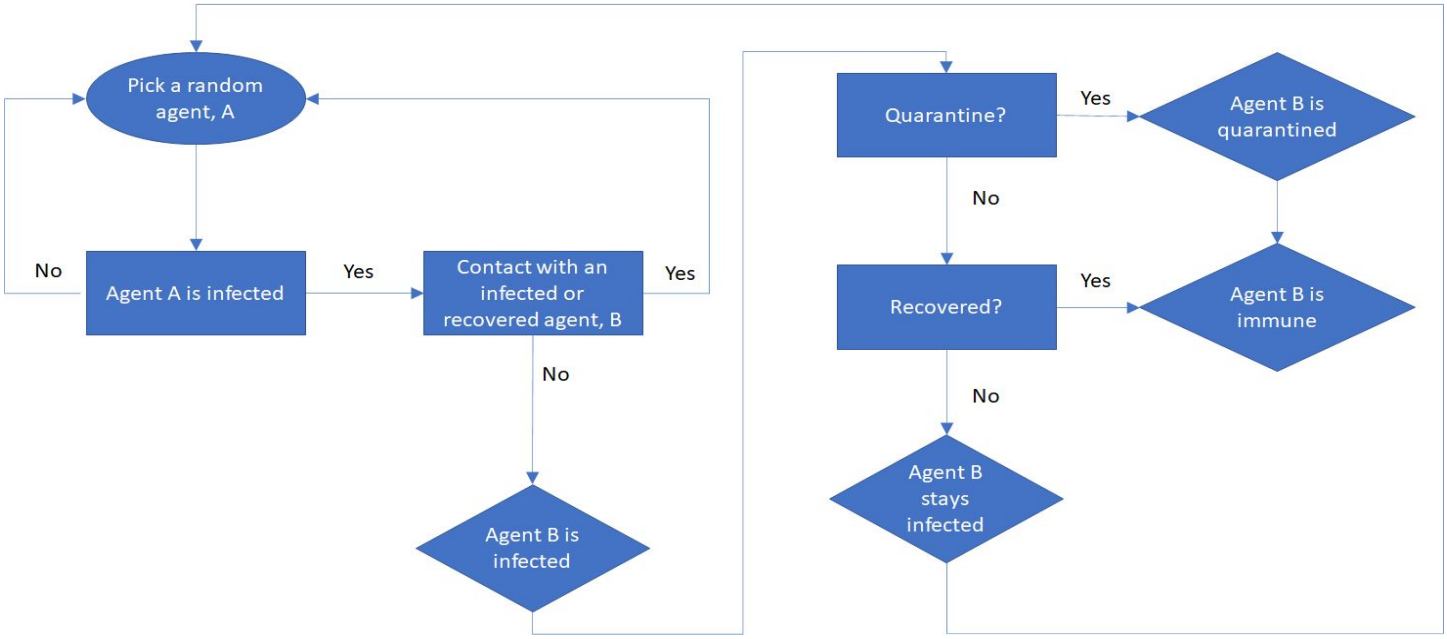


Figure 5. Flow chart of the proposed model

Validation:

Validation of the Agent-Based-Model (ABM) was performed by comparing the results against a comparable SIR model. This was done by implementing a binary search algorithm that searches for an R_0 that allows the SIR model to closely follow the ABM curves. For simplicity, the goal was to closely match the infectious (I) curve and neglect the S and R trajectories. Results are shown below.

Results and Discussion

Three different studies were conducted in order to validate the model and justify the conclusions:

1) How does the agent-based model compare to the conventional SIR model (validation)?
2) How does mobility affect the rate of infection (justification)?
3) How does testing frequency affect the rate of infection (justification?)

Agent-Based-Model Reasonably Matches SIR Model

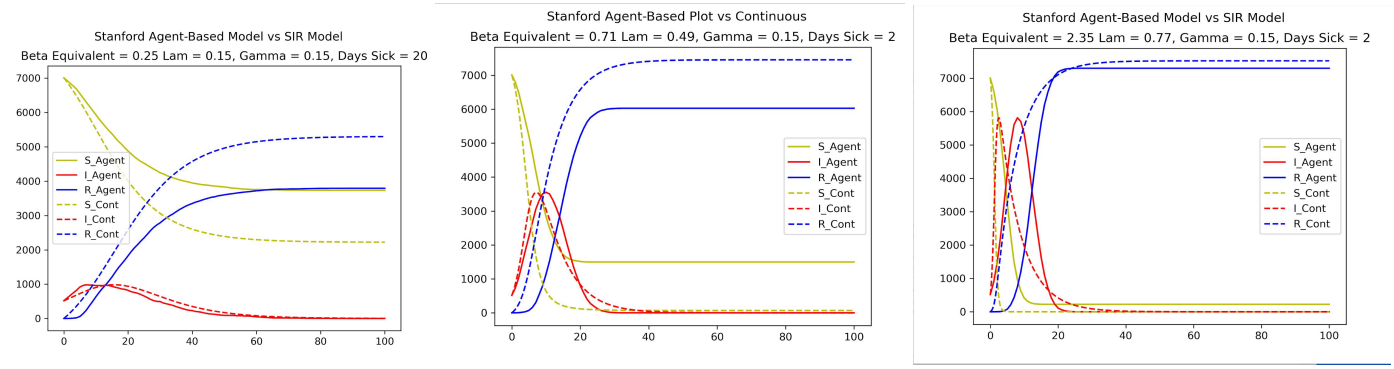


Figure 6: Comparison of Agent-Based Model vs. SIR Model

The above plots illustrate that an "equivalent" contact rate can be found such that the SIR model parallels the infectious curve, I, of the Agent-Based Model. It should be noted that for smaller λ values, the ABM predicts less total infections. At higher values of λ, the ABM predicts a similar value for total infections, but is slightly delayed.

The ABM underprediction total recovery for lower λ and time delay for higher λ likely stems from the fact that in the SIR model, people immediately begin recovering. However in the ABM, the first recovery happens after ~6 days and thus, the ABM is shifted to the right by ~6 days. The effect of this lagging recovery curve has different effects depending on λ (how quickly the infections spread). For very a slow-spreading infection (λ = small, leftmost plot), the infection curves (I) parallel each other without lag. This indicates that both models predict the same time to attain steady state for infections. However, the 6-day delay for the ABM recovery (R) curve means that it has 6 days fewer to grow before the number of infections go to zero, and hence, it will be lower than the SIR R curve. Conversely, the R curve for fast-spreading infections (λ = large, rightmost plot) is less sensitive to time delays because the entire population will become infected quickly--the 6-day time delay means that the ABM recovery has only 100-6 = 94 days to grow. But for fast infections, 94 days poses no challenge for the infection to infect everyone. The time delay for high λ on the infectious (I) curve is more conspicuous simply because the curve is thinner, so any left/right shift is readily observable.

The ABM model reasonably parallels the SIR model. Significant differences in the results between both models were primarily attributed to the recovery curve, which is a minor bookkeeping difference between the SIR and the ABM.

Test Rate and Quarantine Determines the Percent of Susceptible Population That Gets Sick

Agents were restricted to the residential cell by setting P_transition equal to [100% 0% 0% 0%]. Beta and lambda were 1.54 and 0.77, respectively. The "days sick" variable was changed, effectively manipulating the number of days on average that an individual is infectious without testing and going into quarantine. If days sick is set to 3 or 6, which roughly approximates to testing every week and testing every other week, respectively, nearly 100% of the population eventually gets sick and enters the recovered category. However, when testing is maintained at multiple times per week, where days sick is 0 and 1 and students are able to almost immediately test and quarantine after infection, we notice that the infection peak is substantially reduced and a large percentage of the population, about 10% of individuals in the model where days sick is set to 1, never get sick. When testing is increased further, such that individuals can identify infection on the day that they become infectious, we see for the first time that the number of susceptible individuals is maintained above the number of recovered individuals.

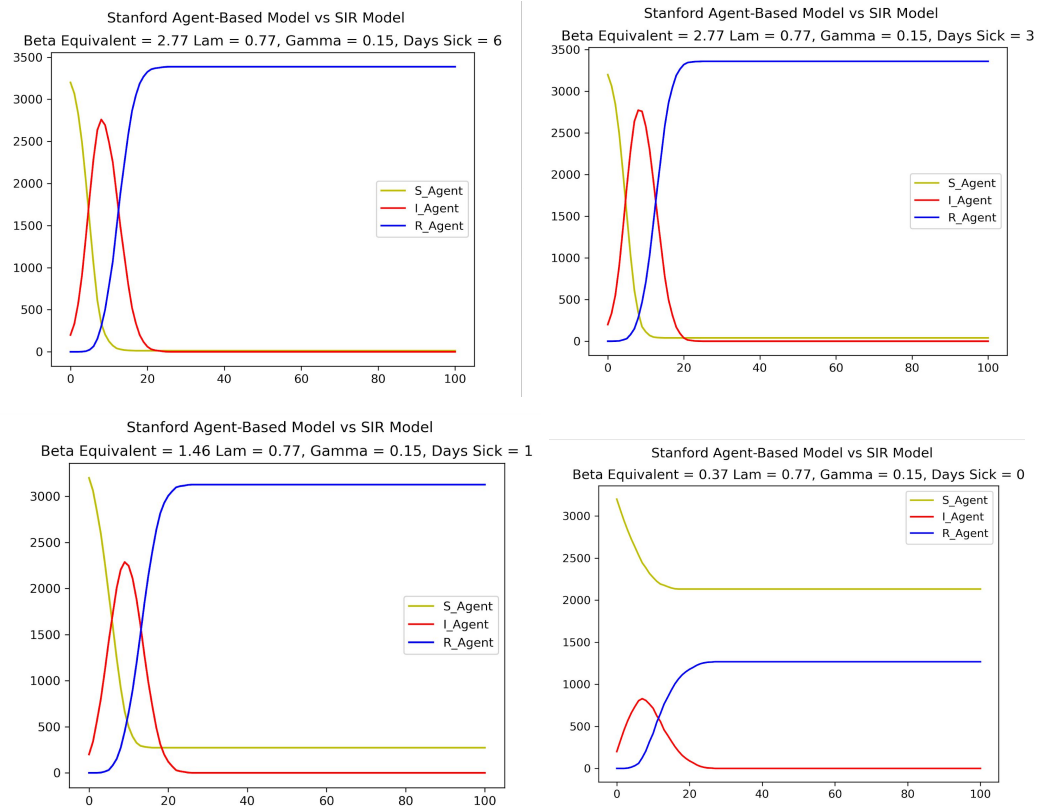


Figure 7: Comparison of Testing Rates for **High** λ

Furthermore, if λ is reduced, biweekly testing (days sick = 3) offers benefits over weekly testing (days sick = 6).

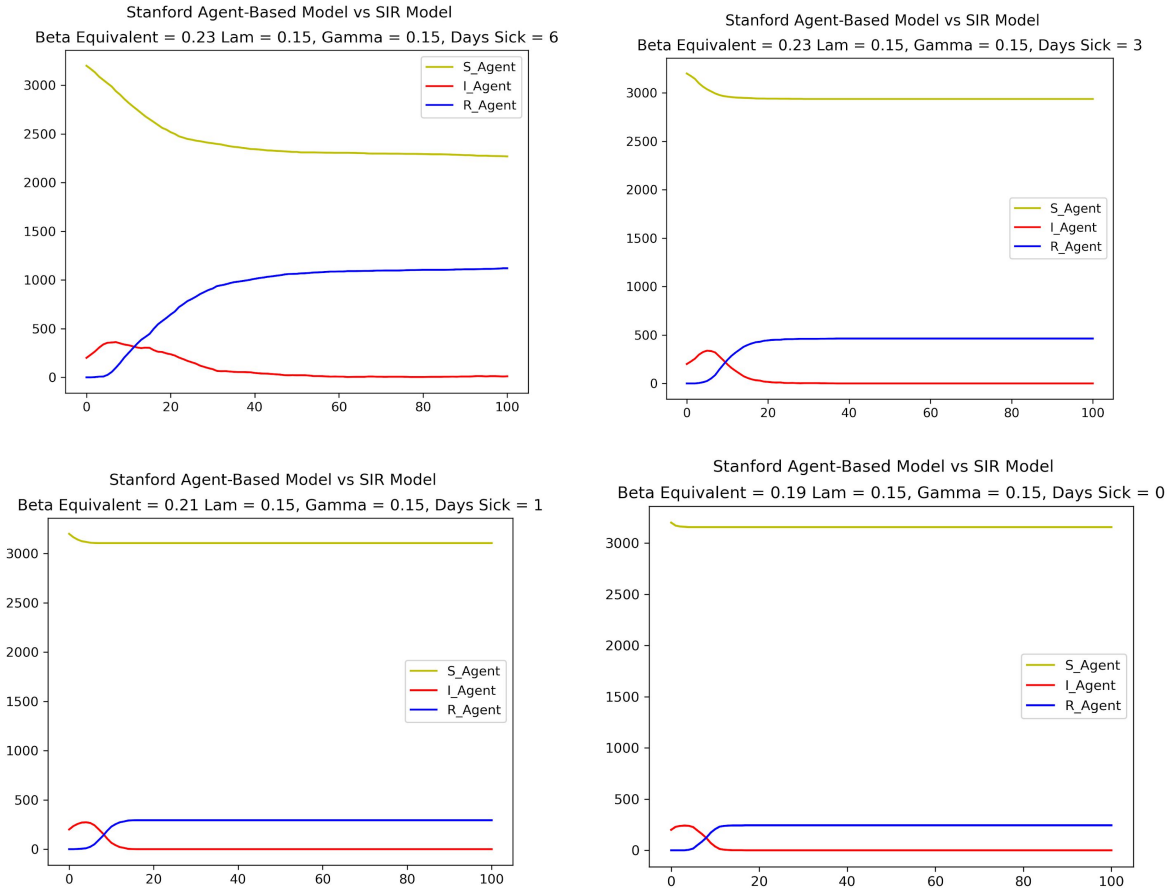
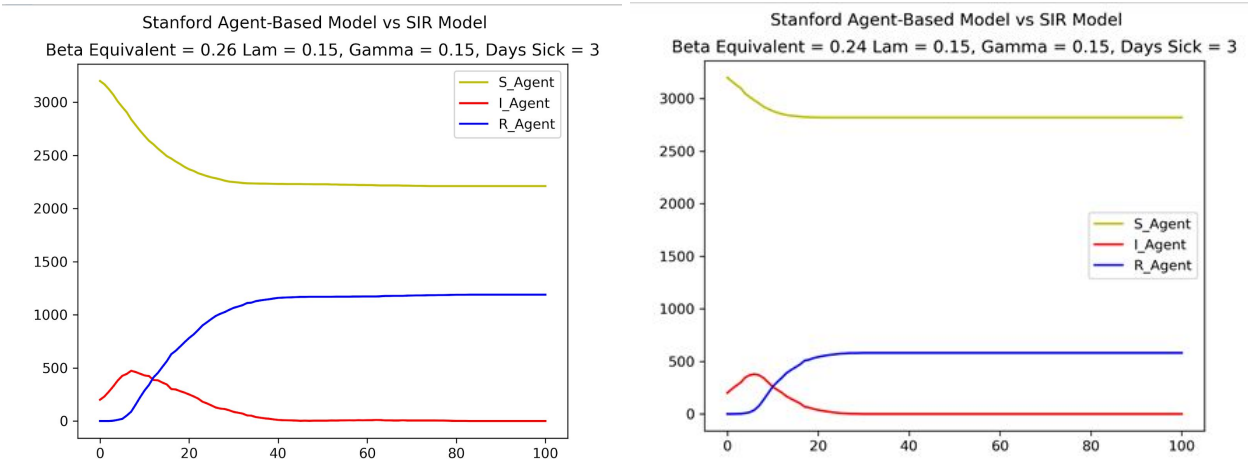


Figure 8: Comparison of Testing for **Low** λ

As shown, testing twice a week (days sick = 3) has notable benefits over weekly testing (days sick = 6). However, testing twice a week provides limited benefit if λ is higher, as discussed above.

Effect of Changing Agent Movement Parameters Depends on λ and Testing

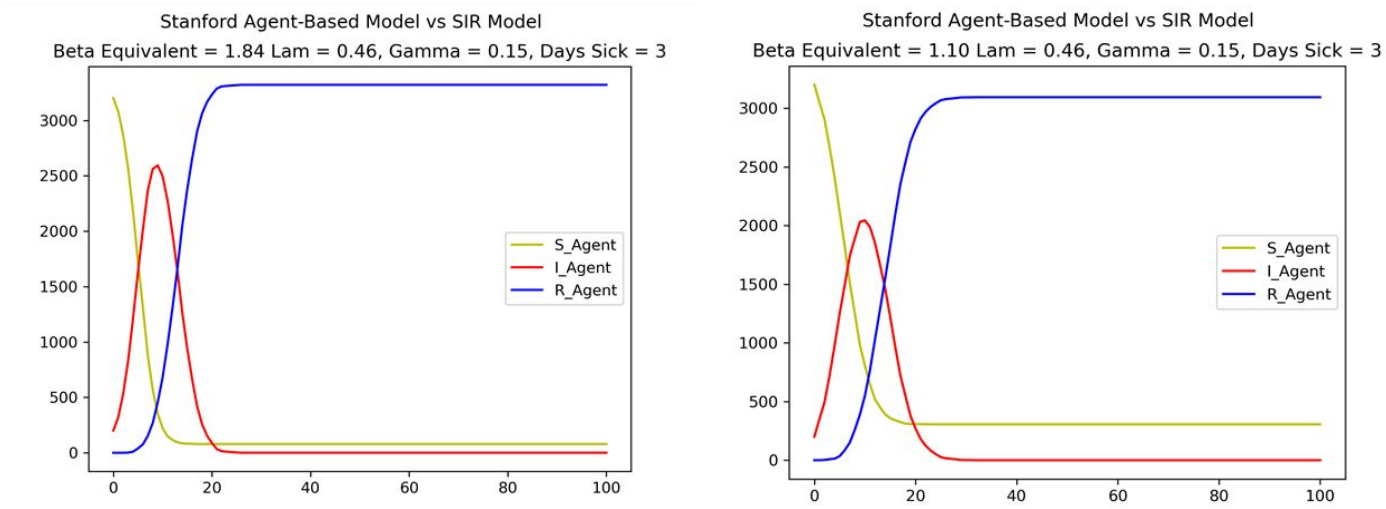
The effect of agent movement depends on other factors. The figures below illustrate:



Low λ: Stationary Low λ: Move between 4 locations

Figure 9: Effect of Agent Movement For **Low** λ

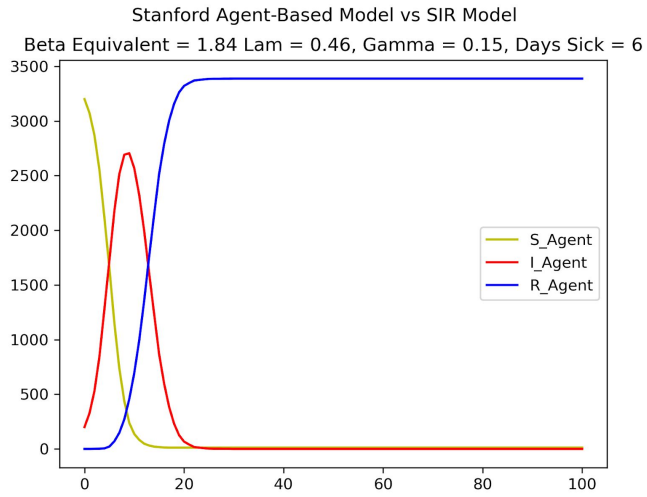
The above plots suggest that increased mobility can actually reduce the infection rate--this is because the stationary case assumed that all individuals are confined to the residence halls, which will have a higher infection rate due to being indoors. Conversely, increased movement (right plot) enables students to visit other locations with less infection rates, as shown in Table 1.



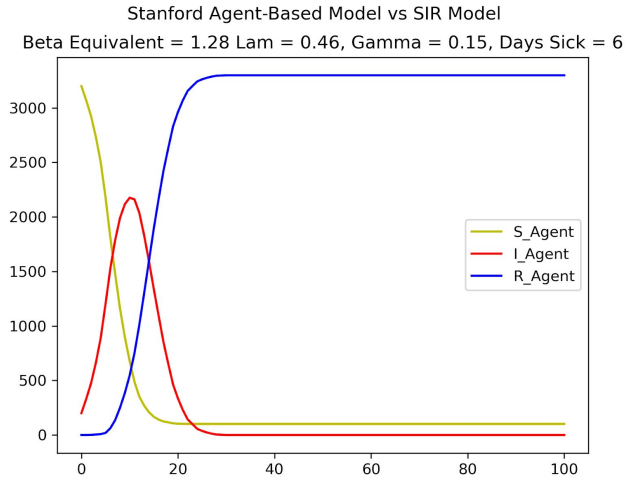
High λ: Stationary High λ: Move between 4 locations

Figure 10: Effect of Agent Movement For **High** λ

Similar trends are observed for the high λ case where increased mobility reduced infections. This is again because of the assumption that stationary students are kept inside residence halls which have higher infection rates. Finally, the effect of agent movement can be compared for different testing rates:

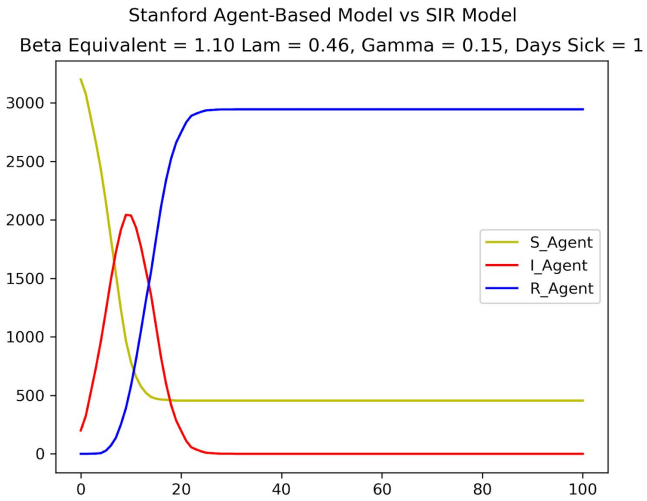


Low Testing Frequency: Stationary

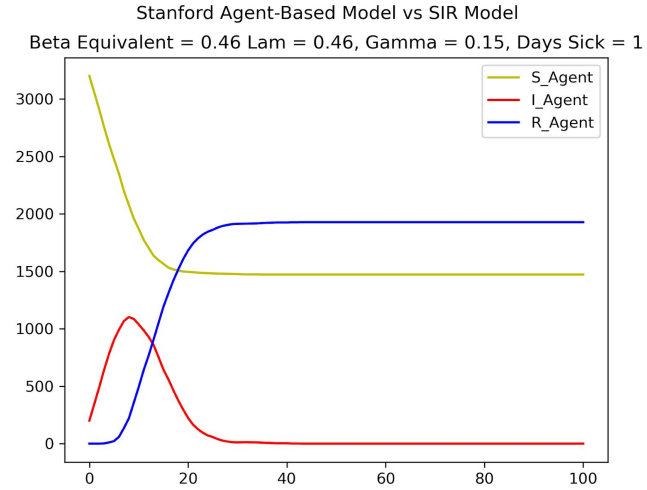


Low Testing Frequency: Move between 4 locations

Figure 11: Effect of Agent Movement For **Low Testing Rates**



High Testing Frequency: Stationary



High Testing Frequency: Movement

Figure 12: Effect of Agent Movement For **High Testing Rates**

Again, increased mobility decreases the infection rates because of the high contagiousness assumed in the stationary location of the residence halls. But interestingly, the infectiousness of increased movement is largely reduced with increased testing (leftmost plot of Figure 11 vs leftmost plot of Figure 12). Additionally, any benefit of mobility is increased with higher testing rates as suggested by Figures 11 and 12.

Conclusion

The above plots suggest that Stanford can begin to reopen common areas and while keeping infections at a reasonable level only if there is a corresponding increase in testing frequency. This matches intuition because high mobility suggests higher λ values, which maps to a faster spread of infections. Only increased testing frequency can acutely pinpoint these infections and trigger quarantine. The specific testing frequency also depends on the acceptable level of peak infections. For example, a maximum allowable threshold of 1000 students that can freely move about requires "every-other-day"/tri-weekly testing. The effects of testing and mobility are highly coupled and warrant a combined treatment.

References

[1] Viguerie, A., Veneziani, A., Lorenzo, G. *et al.* (2020). Diffusion–reaction compartmental models formulated in a continuum mechanics framework: application to COVID-19, mathematical analysis, and numerical study. *Comput Mech* 66, 1131–1152. https://doi.org/10.1007/s00466-020-01888-0

[2] Jones, T.J., Watts, K., & Whytock, R.C. (2018). Using fluid dynamic concepts to estimate species movement rates in terrestrial landscapes. *Ecological Indicators, 93,* 344-350. https://doi.org/10.1016/j.ecolind.2018.05.005

[3] Erik Cuevas, (2020). An agent-based model to evaluate the COVID-19 transmission risks in facilities *Computers in Biology and Medicine*, Volume 121, https://doi.org/10.1016/j.compbiomed.2020.103827

[4] John Hopkins University and Medicine (2020). *Daily State-by-State Testing Trends* [Data set]. John Hopkins University. https://coronavirus.jhu.edu/testing/individual-states/usa

[5] The COVID Tracking Project (2020). *Data by State* [Data set]. The Atlantic: The COVID Tracking Project. https://covidtracking.com/data

[6] SIRIS Student Analytics (2020). *ME 233 Autumn 2020 Student Locations* [Data set]. Stanford University SIRIS Student Analytics. http://siris.su.domains/ (note: data retrieved by ME 233 teaching team and made available through Canvas website, canvas.stanford.edu)

[7] Mayo Clinic (2020). *Guam coronavirus map: What do the trends mean for you?* [Data set]. Mayo Clinic. https://www.mayoclinic.org/coronavirus-covid-19/map/guam/

[8] Mayo Clinic (2020). *Puerto Rico coronavirus map: What do the trends mean for you?* [Data set]. Mayo Clinic. https://www.mayoclinic.org/coronavirus-covid-19/map/puerto-rico

[9] Zohdi, T.I. An agent-based computational framework for simulation of global pandemic and social response on planet X (2020). Comput Mech 66, 1195–1209. https://doi.org/10.1007/s00466-020-01886-2