GEOGRAPHICALLY ASSISTED AGENT-BASED MODEL FOR COVID-19 TRANSMISSION (GEOACT)

Kaushik Ganapathy
Halicioglu Data Science Institute
University of California San Diego
La Jolla, CA 92093
krganapa@ucsd.edu

Jiaxi Lei
Halicioglu Data Science Institute
University of California San Diego
La Jolla, CA 92093
jil1119@ucsd.edu

Evan Price
Halicioglu Data Science Institute
University of California San Diego
La Jolla, CA 92093
evprice@ucsd.edu

Akshay Bhide
Halicioglu Data Science Institute
University of California San Diego
La Jolla, CA 92093
abhide@ucsd.edu

March 8, 2021

ABSTRACT

As schools attempt to reopen amid the COVID-19 pandemic, there is an increasing need to detect and quantify potentially risky activities in schools to help schools test out their individual reopening plans in order to prevent an outbreak. In this paper, we describe the development of a spatially explicit agent-based model to help detect risky activities and assess reopening plans for individual schools by incorporating elements such as behavioral factors, environmental factors, and effects from pharmaceutical and non-pharmaceutical interventions. Following this we describe the development of a gateway infrastructure powered by Apache Airavata to allow general-purpose users to run model simulations with user-defined parameters. Finally, we use the aforementioned model to estimate COVID-19 case counts and the effectiveness of proposed interventions over a two week period for a real school to demonstrate model usability.

1 Introduction

The COVID-19 pandemic has drastically changed the face of the globe since its advent, with the virus continuing to spread everyday. As a consequence, a series of legislative directives have been implemented, resulting in the decline of global travel and business activities. A particularly adverse consequence of these directives is the closure of schools forcing the nation's youth to study and learn from home. The prospect of reopening schools amid the pandemic has sparked a vociferous debate among parents, school teachers, administrators, policy makers and public health officials alike. This issue is particularly relevant as parents of school-going children slowly get back to work, in-turn necessitating schools for supervising and educating these kids during the day. At the same time, this need must also be viewed with caution as voluntarily exposing children to risk-potential activities can result in super spreader events, thus jeopardizing the lives of children, their families and immediate communities. To help mitigate such risks states across the country have proposed legislation to mandate schools to file reopening-plans with state-level public health agencies in order to facilitate an informed, phased, reopening of schools. However, there still exists no uniform system to assess the effectiveness of reopening plans as COVID-19 transmission is highly dependent on school-specific parameters such as floor plans, school resources, the student population, school's neighboring communities, and class schedules. In this paper, we describe the development of a comprehensive spatially-explicit agent-based model (ABM) to predict COVID-19 case-loads and assess the effectiveness of non-pharmaceutical interventions (NPIs) for specific schools. The model takes school-specific data such as floor-plans and class schedules to run simulations on a variety of user-controlled parameters. The system is designed to enable school administrators enable school administrators

and policy makers to assess the effectiveness of reopening plans, and potentially uncover any high- risk activities beforehand.

2 Methods and Implementation

There are several aspects to consider while developing the model to assess COVID-19 transmission. In this section, we shall broadly describe the mechanisms that power infection-transmission in our model, implementation of pharmaceutical and non-pharmaceutical interventions (NPIs), and an overview of how these aspects are incorporated into a comprehensive Airavata gateway to enable general purpose users to carry out simulations.

Our model is an agent-based take to a classic SEIR model. In this model agents are broadly divided into four categories.

1. Human Agents: This class of agents represents the various human beings who comprise a school. These agents are further sub-divided into students and teachers each of which have distinct behavioral properties.
2. Room Agents: This class of agents represents the myriad of room types that are present in a school such as classrooms, lunch-cafeterias, libraries etc. Room agents have distinct properties which vary over time, and are dependent on the human agents that are contained within them
3. Bus Agents: Bus agents represent the school buses which ply on different routes for a specific school. Like room agents, the behavioral dynamics of bus agents are innately connected to the human agents which are present within them.
4. School Agents: School Agents represents the entire school as an single entity. Like other agents described above the state of school agents vary temporally and are government by the combined states of Human, Room, and Bus Agents.

Each human agent having one of three states namely: healthy (susceptible), exposed (exposed), infectious (infected and capable of spreading). Our model does not take into account the recovered state, since we are interested in measuring transmission rates. Particularly, it must be noted that our model factors in effects from asymptomatic transmission as well as pharmaceutical interventions such as vaccination and testing individuals which makes it less likely for an individual to be capable of transmission.

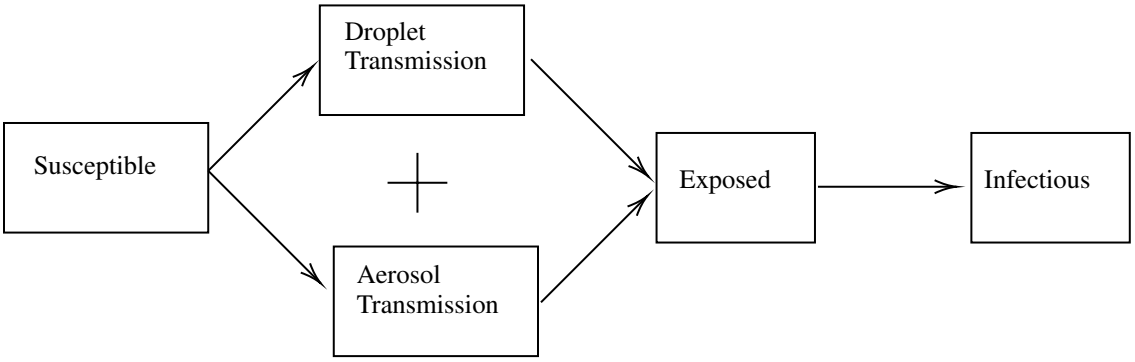

Figure 1: Flowchart depicting an overall summary of the various states that a human agent takes over the course of simulations. Note that the methods of exposure via aerosol and droplet transmission are implemented via two independent models, and that suspetible individuals in the model do not include those vaccinated.

2.1 Disease Transmission

A key step in the generation of model results would be a human agent's transition between susceptible (healthy) and exposed states. Upon competing a review of infectious disease epidemiology, it can be inferred that COVID-19 has two primary routes of infecting people namely: Surface (droplet) transmission[1] and aerosol transmission[2] .

2.1.1 Surface Transmission

Surface transmission is transmission which occurs through contact with SARS-COV-2 on surfaces. Findings have shown that a sizable proportion of COVID-19 transmission occurs via surface transmission caused by inadvertent

exposure to SARS-COV-2 virions on common areas, with SARS-COV-2 virions having been found for over 3 days post contact on fomites.[3] Surface transmission occurs as a result of exposure to large droplets of saliva and mucous containing a sizable amount of viral load. Since our model uses an agent-based approach, a key relationship to identify is one that links distances between pairs of individuals to likelihoods of infection. In the model this relationship is encoded by utilizing the findings of Chu et al.[4] Chu consolidated findings from 172 COVID studies, 38 of them being directly focused on infection rate's relationship to distance.

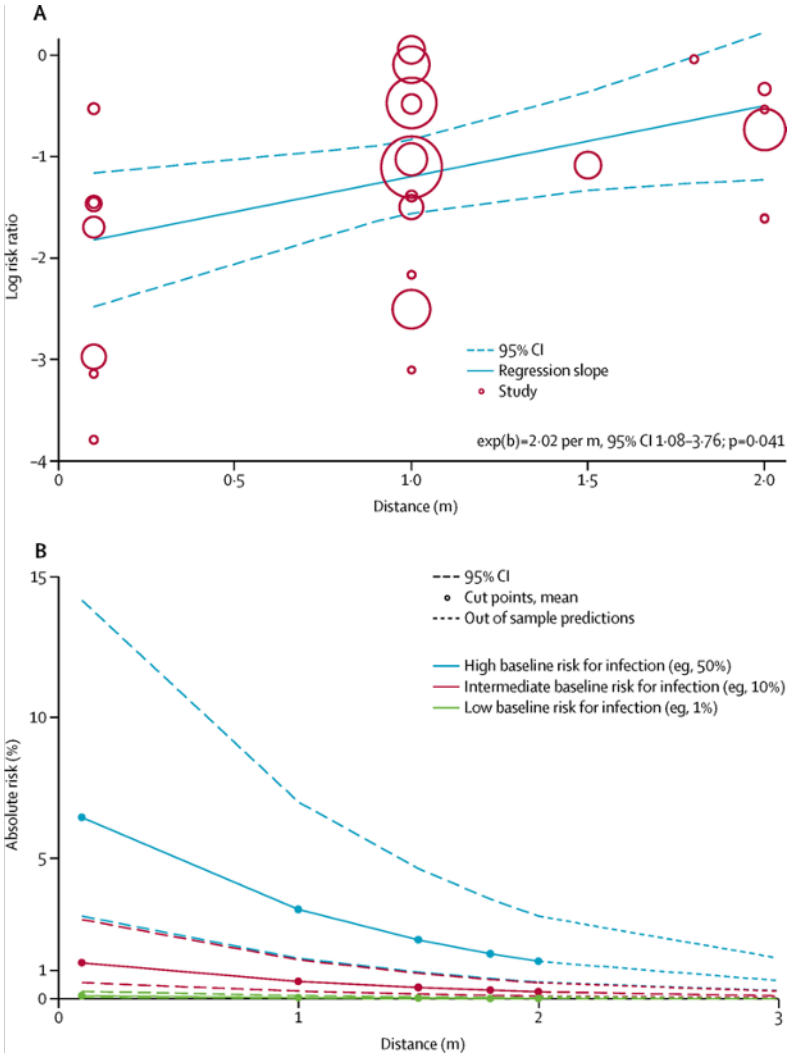

Figure 2: Findings by Chu et al. depicting the relationship between distance between human agents and risks of exposure via droplet transmission. Note that the rate of decay of the curve is approximately 2.02 m^{-1}

.

As seen from Figure 2, there is an apparent exponentially decreasing between the distance a human agent is from an infectious agent and their risk of contracting COVID via droplet transmission. The agent-based model replicates the above findings to infect other human agent's by using SciPy's curve fitting function to fit an exponential-decay curve to the empirical data provided above.

2.1.2 Aerosol Transmission

In addition to surface transmission, there is growing evidence that indoor airborne transmission associated with relatively small, micron-scale aerosol droplets plays a dominant role in the spread of COVID-19, especially in small micro spaces[2,5] . These microscopic aerosol drops are emitted into the environment when infectious human agents cough, sneeze, sing, speak or breathe, thereby expelling an array of liquid droplets. The range of the exhaled pathogens is determined by the radii of the carrier droplets, which typically lie in the range of 0.5 μm - 1 mm. Recent experimental

evidence concluded that particles with r > 2μm are less infectious,[6,7] making the incorporation of aerosol transmission significantly important especially in the agent-based model.

While, the larger virus-laden droplets settle to the floor, aerosol transmisson occurs from microscopic particles which are suspended by the ambient airflow and are picked up by ventilation systems or inhaled, which in-turn causes a secondary route for infection. The parameters required for aerosol-based disease transmission parameters are derived from several earlier studies[8–10] , and models for aerosol transmission are generally based on the principle of a well-mixed room[7] . Models operating under such assumptions are in-turn based upon the seminal work of Wells and Riley[11,12] which relate the transmission rate (β_α) and room ventilation outflow rate (Q) as follows:

$$\beta_\alpha \propto \frac{1}{Q}$$

In addition, the effects of viral deactivation, sedimentation dynamics and the polydispersity of the suspended droplets were considered by Nicas et al, Stilianakis & Drossinos[7,13,14] .

In the agent-based model, we incorporate the effects of aerosol transmission which takes both these effects into account by replicating the COVID-19 aerosol transmission model pioneered by Bazant and Bush,[7] which derives the mean aerosol transmission rate as follows:

$$\bar{\beta_\alpha} = p_m^2 f_d \lambda_q \quad [\frac{\text{Transmissions}}{\text{Unit Time}}] \tag{1}$$

$$= p_m^2 \frac{Q_b}{\lambda_c(\bar{r})V} Q_b C_q \tag{2}$$

$$= (p_m Q_b)^2 \frac{C_q}{\lambda_c(\bar{r})V} \tag{3}$$

$$= (p_m Q_b)^2 \frac{C_q}{(\lambda_a + \lambda_f + \lambda_s + \lambda_v)(\bar{r})V} \tag{4}$$

$$= p_m^2 \frac{\lambda_q Q_b}{(\lambda_a + \lambda_f + \lambda_s + \lambda_v)(\bar{r})V} \tag{5}$$

where p_m is the mask penetration factor, λ_q is the breathing flow rate (Q_b) multiplied by the Infectiousness of Air (C_q), λ_a is the air changes per hour(ACH), λ_f is the air filtration rate, λ_v is the deactivation rate, λ_s is the size-dependent sedimentation rate, and V is the volume of the room.

All of the aforementioned parameters barring the room volume have been derived from the indoor-covid-safety web application published by Bazant and Bush. The room volume has been estimated by multiplying floor area times a room height of 12 ft. from floorplans obtained from the San Diego County Institutional Facilities and Planning Office.

The transmission rate obtained is defined as the mean number of transmissions between a single pair of healthy and infectious individuals. Hence, this number is multiplied by the number of infectious individuals, as well as the time of exposure to estimate the mean number of transmissions occurring. The mean number of transmissions are then used as proxies for infection-probabilities (with a cut-off at 1.0) at the end of each day to determine whether a susceptible human agent should be infected or not.

A key assumption to note here is the granularity of the model we considering exposure to SARS-COV-2. In the agent-based model, infection via aerosol transmission is factored in by considering the mean breathing rate, mean viral loads exhaled, and the mean mask permeability factor, of all human agents in a given room and then uniformly infect each agent accordingly. In other words, the aerosol infections are done on a room agent level rather than a hum,an agent level. The reason why such an assumption is possible is because our aerosol model considers a well-mixed room, which in-turn means that the likelihood of getting infected via aerosols is agnostic to the spatial location of individuals. While this assumption may not hold true in the real world, the absence of room specific air-flow patterns hinders the development of a more realistic model.

2.1.3 Transmission Dynamics

Apart from having two distinct routes of infection another factor our agent-based model considers is the temporal dynamics of COVID-19 transmission. Specifically our model models the incubation time post-exposure, as well as the an individual's varying level of infectiousness during the course of the disease. Both these insights are modeled

using results from He et al.[15,16] , which in-turn is based on temporal patterns of viral shedding from 94 patients with laboratory-confirmed COVID-19 with infectiousness profiles obtained from a separate sample of 77 infector–infectee transmission pairs.

As seen in Figure 3(A), the temporal curve governing infectiousness of individuals peaks at roughly -1 days (1 day prior to symptom onset). Additionally, it can be also noted that the above curve can be approximated with the help of a gamma distribution with appropriate shape, scale and location parameters. Similarly, observing Figure 3(B), it can be seen that the temporal distribution of symptom-onset times peaks at roughly 3 days post exposure, and follows an approximate lognormal distribution.

These observations are in correspondence with that of He et al,[15] Hence, these distributions are incorporated into the model by fitting an appropriate curve using SciPy's curve fit function. Each human being is then assigned a random symptom-onset time post exposure using values from distribution 3(B), This value is then used to calculate the number of days prior to symptom onset for each individual using distribution 3(A). Once the corresponding temporal point on distribution 3(A) is obtained, the value for infectiousness of an individual at a specific point of time is obtained, which is in conjunction with the distribution described in Figure 2(B), to estimate the risk for a new susceptible human to be infected by a given human agent.

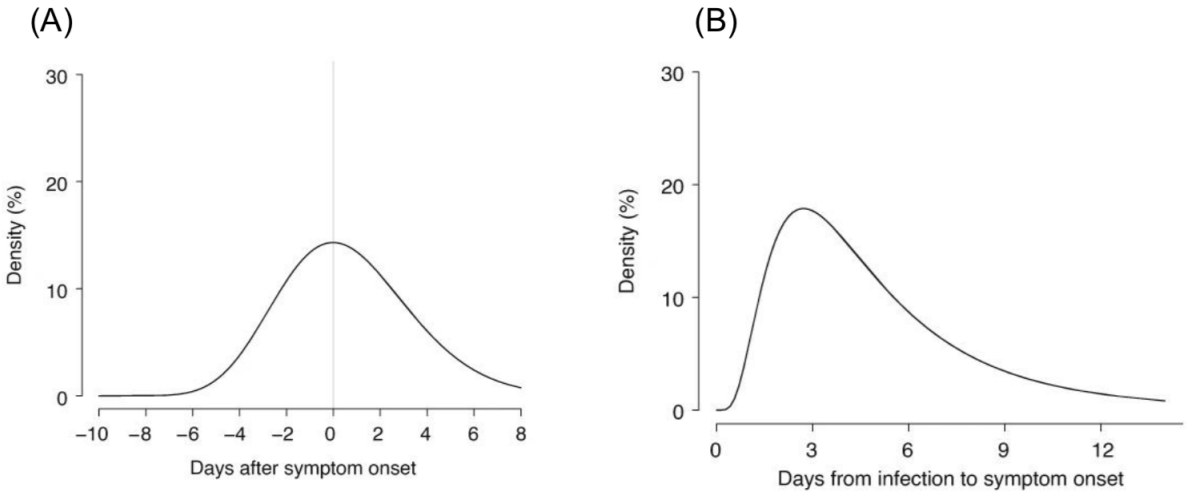

Figure 3: Findings by He et al. depcting the incubation curve and infectiousness curves for COVID-19. Figure 3(A) shows the temporal distribution of the infectiousness of COVID-19 patients post symptom onset, peaking at approximately -1 days. Figure 3(B) shows the temporal distribution that governs symptom onset post exposure, which peaks at approximately 3 days.

2.2 Agent Behavior

Having described the mechanisms that govern a human agent's exposure to SARS-COV-2, we now describe the encoded behavioral dynamics which directly influence infection risks for different types of agents.

2.2.1 Student Agents

In our spatially explicit model we simulate the actual locations of student agents over time, as they move through their regular class day. For instance, in our model, students are either seated in their assigned desks (emulating class-time instruction) or engage in group activities with nearby classmates during class hours. The changes in the spatial distribution of students over time thus contributes to the length of exposure to potentially infected agents, and thus affects the rates of transmission. Similarly, the specific micro-space occupied by students also varies over time which in-turn can vary the rates of transmission. For instance, in our model school agents (from grades 1-5) have recess in a outdoor yard where they interact with a cohort of students (friends). Such a situation causes a higher likelihood of transmission among cohorts. On the contrary, the risks from aerosol transmission is removed since these students are

outdoors. Another example would be when students agents from grades 1-5 have lunch in the cafeteria. This activity can in-turn increase the likelihoods of inter-class exposure.

An important aspect to note is the variance in the physical activity of human agents. For instance in the model, student agents may alter their rate of speaking and breathing depending on the current activity, which in-turn has an effect on the extent of COVID-19 transmission. The exact descriptions on how transmissions vary has been described at length in the disease transmission section.

2.2.2 Teacher Agents

Similar to student agent behavior described above, the behaviors of teachers can also influence model outcomes. While students have assigned seating during classes, teachers move from one area of the class to another, thus emulating interactions between students and teachers. As a result of this, transmission between different student cohorts in a classroom can occur, which results in greater potential for extensive transmission within the same classroom. After class hours, teachers are confined within a faculty lounge which can increase the likelihood for transmission between teachers from different classes. As described in the previous section, teacher agents also experience variation in the physical activity over time. However, a key difference from student agents is the differing likelihoods to engage in these specific activities over time. For instance, a teacher has increased tendencies to talk loudly or shout during class instruction which directly increases the rate of aerosol transmission from them, while student agents experience this only during group activities.

2.2.3 Room Agents

The behavior or state of room agents may influence the outcome from model runs as well. At the moment, the state of room agents primarily influence the extent of aerosol transmission. For instance, in our model, the type of ventilation in a given room (which determines the number of air changes per hour [λ_a] or ACH) at a given time of the day can significantly alter the extent of aerosol transmission in a room. Additionally, room agents may also have specific seating layouts or a maximum threshold on the number of occupants can also alter the extents of transmission.

2.2.4 Bus Agents

Finally, the behavior of bus agents over time will also have a direct impact in determining the extent of transmission. For instance, transmission rates can greatly vary as bus agents pass through and pick student agents up from different areas with varying infection rates. This occurs as a direct consequence from different student agents spend different amounts of time on the bus, together with students from other cohorts. Furthermore, akin to a room agent, the buses may also have altered state of airflow (ventilation) during various sections of the trip which can greatly affect aerosol transmission. Finally, the innate design of bus trips which limits the ability of bus drivers to enforce stipulated standard operating procedures on student agents can adversely affect transmission during the bus ride.

2.3 Intervention Methods

Like agent behaviors, planned intervention strategies may also directly influence rates of COVID-19 transmission. Our model allows users to estimate and understand the effectiveness of specific interventions or even a combination of several techniques planned at the same time. Broadly, intervention methods fall under one of two categories:

1. Non-pharmaceutical Interventions (NPIs)
2. Pharmaceutical Interventions

2.3.1 Non-pharmaceutical Interventions (NPIs)

Non-pharmaceutical interventions are those strategies apart from vaccination, testing and therapeutic drugs that reduce the extent of transmission of COVID-19. In our model, we incorporate the following NPIs.

2.3.1.1 Mask Wearing

Mask wearing has been used extensively to control outbreaks of previous airborne infections such as SARS, MERS, and Influenza. Several studies have described the effectiveness of proper mask-wearing in controlling COVID-19 transmission[17, 18] . In our model, the effects from mask wearing are encoded as a mask penetration factor p_m which varies between 1.0 (no mask), 0.30 (for a multilayer cotton mask), 0.5 (for a regular cotton mask), 0.1

(for a surgical mask) and 0.05 (for a N-95 respirator). As noted earlier these values have been obtained from the indoor-covid-safety web application[7] .

A key simplification in our model occurs when two human agents are wearing different mask types. Consider two human-agents having mask penetration factors p_{m1} and p_{m2} respectively. When this occurs, our agent-based model calculates the likelihood of infection using the mean of the mask penetration factors (i.e. $\frac{1}{2}(p_{m1} + p_{m2})$).

2.3.1.2 Ventilation

Ventilation of closed spaces allows for the dilution of pathogen concentration in rooms, thus resulting in decreasing amounts of transmission. The effects of ventilation as an effective strategy to reduce COVID-19 spread has been well studied[19,20] and agreed upon. In the model the effects from ventilation changes are directly encoded in the form of an air changes per hour (ACH or λ_a) parameter which in-turn controls the aerosol transmission rate as discussed earlier. The vallues of this parameter generally range between very low values (≈ 0.3) in closed poorly ventilated places to over 20 in environments such as hospitals or moving school buses with open windows. Once again, these values are directly sourced from the indoor-covid-safety web application.

2.3.1.3 Cohorts and Class Sizes

Limiting classroom occupancy and reducing interactions among students from different classes have long been considered effective NPIs. Each of these measures would equivalently result in reduced COVID-19 transmission[21,22] . Our model allows users to input a attendance_prop parameter which reduces the total school strength by the specified proportion. On a similar vein, cohorts groups have also been implemented to ensure that student-agents have an increased likelihood of coming into close contact with specified groups of students at specific times of the day. For instance, student agents may be configured to spend recess with specific agents who may have an increased likelihood to be from the same class. In this case, the usage of cohorts would thus reduce the extent of transmission across different classrooms, thus reducing the potential for a super-spreader event. Similar to attendance, the parameters controlling the formation of cohorts can be also controlled by users.

2.3.1.4 Microspatial Layouts

As a consequence of being spatially explicit, our model is able to assess the reduction in COVID-19 cases from proposed micro-spatial changes. In classroom spaces, we have encoded the ability to have two distinct seating layout styles:

1. Circular Desks: Each desk contains a predetermined number of students (considering the maximum capacity of the room) each of whose assigned seats are arranged in the form of a circle.
2. Individual Seats: Each student is assigned their own individual desk.

Similarly for bus agents, we have encoded the ability to have two distinct seating styles:

1. Regular Seating: Students are arranged in a 2+2 or 2+1 seating pattern which in-turn is determined by the model of the bus used.
2. Zig-Zag Seating: Each student has their own individual seat with no two students seated directly behind each other.

Our model will thus allow school administrators and policymakers to reliably assess if the planned changes in micro-space layout can effectively reduce transmission of COVID-19.

2.3.2 Pharmaceutical Interventions

As previously outlined, pharmaceutical interventions encompass vaccinations, therapeutic drugs, and testing each of which has the potential to reduce transmission by a sizable degree. The incorporation of pharmaceutical interventions into our model has been made relevant with the advent of the COVID-19 vaccine[23] . We incorporate the effects of the following pharmaceutical interventions:

2.3.2.1 Vaccines

The development of a plethora of COVID-19 vaccines has been viewed with hope to bring about an epidemiological end to the ongoing pandemic. As greater amounts of people get access to the vaccine, it is important to estimate when it can be safe to open common spaces. In our model, we incorporate the effects from vaccination by removing a human-agent's ability to transmit or be infected by COVID-19. In addition, we have user controlled parameters which allows users to adjust the proportion of students and teachers who are vaccinated and observe changes in transmission.

2.3.2.2 Testing

Another important pharmaceutical intervention is testing for COVID-19. With greater extents and frequency of testing, we can successfully isolate infectious individuals faster, and thereby slow down disease transmission. In our model, we model the effects of testing by detecting and removing infectious agents at regular intervals of time. We also provide user controlled parameters to adjust this frequency for both students and teachers independently.

2.4 Data Procurement and Processing

The spatially-explicit agent-based model uses a plethora of input sources in order to carry out simulations for specific schools. In this section, we detail the process of procuring and processing the data in order to run simulations.

2.4.1 School Floorplans

A vital requirement of the model is spatially accurate school floor-plans. These were provided to us by the instructional facilities planning department of San Diego Unified School District (SDUSD) for a selected set of public schools and school principals for private schools.

The school floor-plans so obtained were either in PDF or CAD (.DWG) file formats. Therefore, methods were developed to convert (digitize) the floor-plans into ESRI shapefiles which could be then used our model. Once floor-plans were obtained, they are loaded into ArcGIS Pro. Layouts in CAD file formats could be directly imported into ArcGIS Pro. However, in case of PDF layouts, a second step of manual digitization followed. The process of digitization used the on-screen digitization tool where users manually overlay geometries such as polygons and lines over the requisite PDF. Additionally, to add external features, and more context to the digitization, comparisons were performed using satellite imagery of relevant school campuses as obtained from Google Maps or ESRI's ArcGIS Pro ecosystem. Once a preliminary digitized output was obtained, relevant features (such as buildings, entry points etc.) were annotated over the outline. A uniform annotation scheme has been established for digitization to enable standardized digitization for additional school layouts.

While layouts derived from CAD files has information regarding the scaling and projection systems used by layouts, those digitized from PDF layouts do not. Estimates for scaling the layouts were produced by the affine transform function in the shapely library in Python in correspondence with the seating capacity of classrooms for elementary schools estimated from findings by Tanner et al[24] . A maximum capacity of 20 students per classroom was assumed to estimate the true scaling factor, and a uniform height of 12 feet was assumed. For our model, all floor-plan digitization efforts were carried out with support from the Department of Urban Studies and Planning at the University of California, San Diego.

Number of Students plus 1 Teacher	Elementary School [Square Feet (Meters)]	Secondary School [Square Feet (Meters)]
10	539 (50.13)	704 (65.47)
11	564 (52.45)	768 (71.42)
12	637 (59.24)	832 (77.38)
13	686 (63.80)	896 (83.33)
14	735 (68.36)	960 (89.28)
15	784 (72.91)	1024 (95.23)
16	833 (77.47)	1088 (101.18)
17	882 (82.03)	1152 (107.14)
18	931 (86.58)	1216 (113.09)
19	980 (91.14)	1280 (119.04)
20	1029 (95.70)	1344 (124.99)

Figure 4: Guidelines for minimum standards of elementary school class sizes by Tanner et al.

Additional data on schools such as the size of the school population were obtained from https://www.niche.com and conversations with officials from SDUSD or school principals.

2.4.2 Bus Route Data

Another key aspect of our agent-based model is bus routes. Data on the exact specification of buses used and seating arrangements proposed by SDUSD was provided during an onsite field trip to a school bus yard.

Additionally, school bus route data was provided for every elementary, middle and high school in the county. The data was provided in the form of an excel file which was first cleaned and parsed. Locations of stops were provided as abstract street addresses or Thomas Brothers Grids. Data cleaning pipelines using the Pandas data-processing library were used to convert the stops into those that could be geocoded. Following this, the ArcGIS online geocoding service (mediated via the Python API) was used to geocode stop representations into locations. Context-based (manual) geocoding was used in the case of errors, which was in-turn implemented by cross referencing representations of locations with google maps. The stops were then compiled in the form of a ESRI shapefile.

Finally, methods were written in Python to retrieve current information on case counts within the San Diego County by zip-codes which can be used in conjuction with the aforementioned shapefile to simulate community transmission with students in specific spending different amounts of time in the bus, with other children from different areas, each of which may have different rates of COVID-19 occurrence.

2.5 Codebase Development

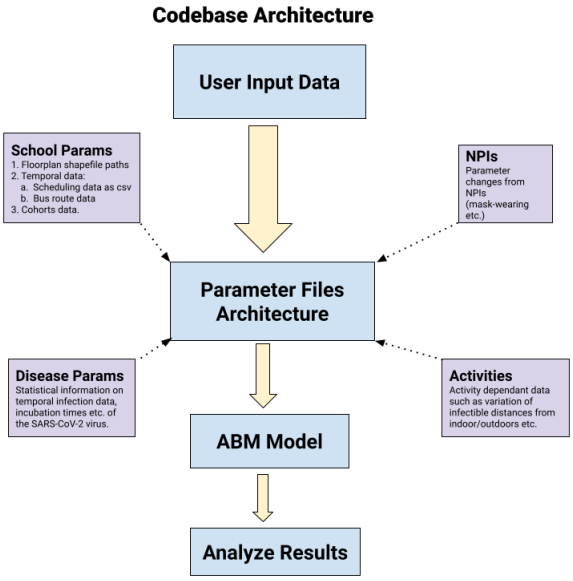

Figure 5: An overview of the computational infrastructure to produce model results.

The system used to simulate COVID-19 transmission in schools enabled through an interconnected system of open source libraries and APIs. Specifically use the mesa_geo library in Python to enable us to run simulations. Additionally, we have engineered efficient parallelization schemes to scale this library and parallelize our model runs on HPC systems. As described in the earlier sections, the model uses heterogeneous data from a plethora of sources to run realistic GIS-enabled agent-based modeling simulations.

Once input parameters are obtained they are classified into fixed and variable model parameters depending on the values that are to be modified in independent simulation runs. School-specific fixed model parameters are written in a schoolpararms.ini file.This file contains references to the locations of annotated shapefiles, class schedules, grade-wise population distribution, and temporal information such as the time duration of the total and each step of the simulation. The next set of fixed parameters provides information on schedule-specific events in a simulation such as class, ventilation, and sanitation schedules. These parameters are organized as CSV file and references to these contained in the schoolpararms.ini. In addition to temporal data, there are constant model parameters which play a crucial role in simulating the transmission dynamics during the course of simulations. These are appropriately contained in the diseaseparams.ini file. Sections of this file correspond to various transmission parameters such as distributions

to simulate incubation times, and infectiousness. The values of these parameters have been obtained with techniques described in the methods section. The parameters discussed until this point are specific to schools or transmission of COVID-19. However, since our model involves simulating transmission in schools, it became essential to incorporate fixed model parameters that vary by the type of activity that a student or teacher performs during the course of the day. Specifically, the activityparams.ini file contains information on the distances each human agent can move in a 5 minute interval at varying points of the day. Additionally, the transmission_rate.py file contains information that corresponds the physiological activities (breathing, talking, running etc.) to appropriate numeric representations using schemes described in the methods section. Similarly, the NPIs.ini file contains information on fixed model parameter changes from the incorporation of interventions into the simulation. For instance, this file contains the various mask penetration factors which change the degrees of transmission. Finally, the last set of parameters are the variable parameters whose combinations reflect the exhaustive set of parameter combinations to be simulated in a model run. For instance, we could attempt changing changing the location of the lunch between the cafeteria and in class and change attendance rates between 75% and 100%. This corresponds to a set of 4 unique combinations. A job creator has been implemented which takes in the set of variable parameters and packages them into individual jobs, which are essentially folders containing files representing a specific combination of fixed and variable model parameters. The data contained in this folder allows for estimating the state of an agent (A) at a specific time (T) performing an activity (P) in a given space at the school (S). This data is then used by the core simulation framework on local or HPC systems to produce files that contain data collected at each time step of the simulation.

Once the simulation is complete, the system generates a folder of CSV files containing the values of the number of COVID-19 positive individuals at each time step of the simulation for the specific combination of fixed and variable parameters used for the job. This data is then graphed and analyzed using the Pandas, Matplotlib, Seaborn, and SciPy libraries to provide transmission curves. Furthermore, options have been provided to generate images depicting aspects of the agent state (such as position, and health status) at each time step to analyze and identify risky activities that contribute to higher transmission. Finally the model also has provisions to produce heat-maps of containing the exact locations where agents first got exposed to COVID-19. Thus, the model workflow and parameters schema developed allows one to reliably simulate and analyze COVID-19 transmission dynamics and patterns in schools.

2.6 Gateway Development and Architecture

Since our model will be primarily used by general purpose users such as school administrators and policy makers, we attempt to incorporate the entire codebase described above into an Apache Airavata gateway to help reliably set-up and run the model workflows on HPC systems.

A preliminary gateway was set up with the help of extensive collaborations with the Science Gateways as a Platform (SciGap) team hosted at geoact.sdsc.edu. Once a template gateway was set up, set up a new experiment titled GeoACT was configured on the gateway which act as the ultimate conduit for general purpose users to run simulations. In order to enable our code-base to be used by the gateway, the entire code-base was hosted on a public GitHub repository located at https://github.com/covABM/SchoolABM. Following this, schemes for various input parameters was configured with the help of the Airavata web GUI. These input parameters ranged from simple numeric entries, to even radio buttons for choices. The current set of user controlled parameters is detailed below:

Parameter Name	**Input Type**
Number of Gradeschool Students	Numeric Entry
Number of Preschool Students	Numeric Entry
Number of Preschool Students	Numeric Entry
Number of Faculty	Numeric Entry
Proportion of Students traveling via School Bus	Numeric Entry
School Zipcode	Numeric Entry
Number of Days	Numeric Entry
Mask Type	Radio Button: Choices of Surgical, Cloth, N95, Mutlilayer, None
Ventilation	Radio Button: Choices of Closed Windows, Open Windows, Open Windows + Fans, Mechanical Ventilation, None
Proportion of Teachers Vaccinated	Numeric Entry
Testing Policy for Teachers	Radio Button: Weekly, Biweekly, None
Testing Policy for Students	Radio Button: Weekly, Biweekly, None

Table 1: User defined parameters and their configured input types on the Airavata Gateway.

Each of the above parameters was provisioned with a command line argument which would in-turn dictate how these parameters get passed along for computations. Having set up the user inputs, the next step was to configure the computational script. In this regard a shell script was written which parsed through the relevant command line arguments by going through the exhaustive list of user defined options as detailed in the command line arguments for each input parameter. In this regard, the getopts unix function was utilized. Once the values for each of the parameters was obtained they were stored into unix shell variables. Subsequently, the GitHub repository set up previously repository

was cloned, and the user input parameters were written into relevant parameter files as described earlier with the help of the crudini package. Once the configuration files for relevant model inputs have been successfully written, model script was executed and the Airavata gateway sends these computations configured as a slurm job to the compute queue on Comet HPC system at the San Diego Supercomputer Center. Access to the computational resources has been mediated with the help of a community login associated with a startup grant from XSEDE. Following this step, provisions to collect generated model output files was also implemented using the Airavata web GUI. Specifically, provisions were given to enter the location of the desired output, along with the file type. At the moment, we allow users to retrieve transmission curves (as PNGs) as well as locations of first exposure to COVID-19 as heat maps from the model.

3 Results and Discussion

In this section we shall attempt to use the model to simulate COVID-19 transmission using a given school layout by varying a non-exhaustive set of user defined parameters. Additionally, we shall also analyze the results obtained from the model using quantitative and context driven qualitative techniques.

3.1 System Setup

In this section, we describe the system setup we have used to generate the model results. As detailed in previous sections, the model uses school floor-plans obtained from the Instructional Facilities Planning Department (IFPD) from SDUSD. As seen in the figure below, the floor-plan consists of classrooms which are in-turn classified as those used for instructing grades 1-5, pre-school, kindergarten (KG), and special education students. In addition, the school consists of two (main and small) recess yards. Finally, the school also has a cafeteria where all grade-school students (1-5) have meals. It must be noted that children belonging to kindergarten, pre-school and special education always have meals in class, whereas grade school children may have lunch in classrooms with the help of the adjustable inclass_lunch parameter. Furthermore, NPIs such as masks are assumed to be not present during lunch hours, and aerosol transmission was considered to be absent for students who were outdoors during recess.

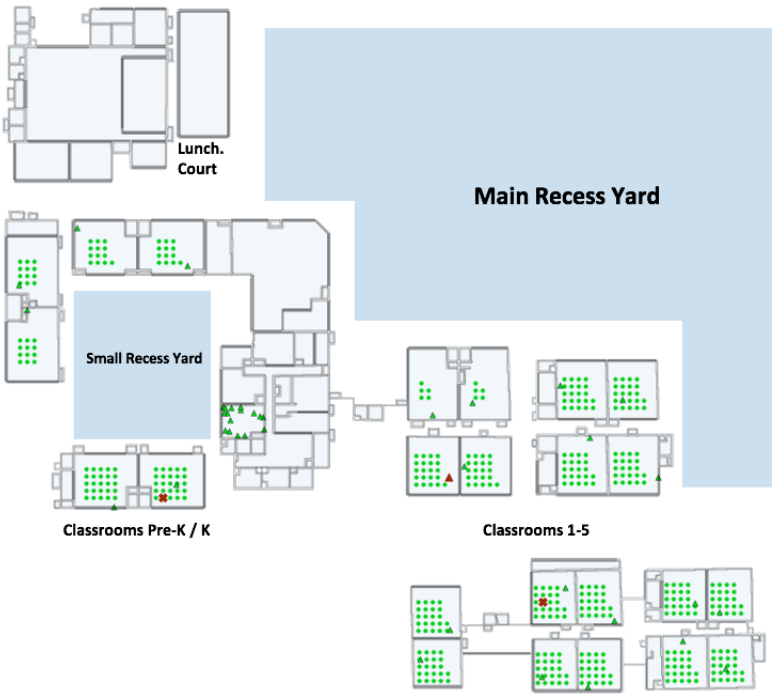

Figure 6: Schematic representing the layout and floor-plan of the school for model runs.

The simulation uses a total of 500 agents consisting of 460 students and 40 teaching faculty and staff. Specifically, there are 350 grade-school students with roughly 21 students per classroom on average. In addition, there are also 50 kindergarten, and 50 preschool students with 25 students in a classroom. Finally, there are 10 special- education students with 5 students per classroom. Each classroom has a teacher, and the remainder reside in faculty-lounge for

the duration of the simulation. If students wear masks, they are assumed to wear surgical masks whereas teachers are assumed to be completely vaccinated. Similarly, during class hours students were assumed to be breathing heavily 75% of the time, whispering 20% of the time, and talking loudly 5% of the time at each five minute interval. Similarly during group activities, students were assumed to whisper 5% of the time, breath heavily 15% of the time, and talk normally 20% of the time and talk loudly 60% of the time during each five minute interval. Finally, it is assumed that all classrooms in the school have open windows during periods of instruction.

Students a follow a routine day-to-day schedule consisting of a typical school day from 08:00 AM to 03:30 PM with the following template class schedule. However, the model incorporates flexibility on class-schedules with independent classes within the school population being able to operate on different schedules. In this model run, it is assumed that each of the classes have the same schedule.

Time	**Activity**
08:00 AM - 09:30 AM	Class
09:30 AM - 10:00 AM	Recess
12:00 Noon - 12:45 PM	Class
12:45 PM - 1:30 PM	Lunch
1:30 PM - 03:30 PM	Class

Table 2: Class schedule used by student agents during the course of the 2 week simulation.

The following model run simulates the following schedule over a two week period with steps at every 5 minutes, and the results describe the evolution of COVID-19 cases through each step of the simulation. Furthermore, we alter the following parameters to investigate changes in the rate of transmission.

1. mask_prob [0, 0.5, 1.0]: Represents the proportion of the student population properly wearing masks. Masks are assumed to be taken off during meals.
2. attend_rate [0.75, 1.0]: Represents the proportion of the total pre-COVID student population currently attending class in-person.
3. inclass_lunch [True, False]: Represents whether the agents have lunch in class or in the cafeteria. Kindergarten, Pre-K, and Special Education students are assumed to always have lunch in class.
4. student_testing_freq [7, 14]: Represents the testing frequency for students in days. It is to be noted that we assume no latency between testing and obtaining test results while removing student agents.

3.2 Results

In this section, we describe the results obtained from the simulations. Each parameter combination generated a folder consisting of model_val.csv and agent_val.csv files which were then parsed using the Pandas library. The associated parameter combinations and output results can be viewed in the Table 3 outlined in the following page. A FacetGrid plot was generated using the seaborn library showing line-plots of the proportion of the population infected with COVID-19 at each 5 minute interval in a 14 day period which is available to download at here. Upon computing a histogram of the observed results, we observe that the 0.75 percentile of the proportion of cases is 28% corresponding to an actual case count of 140.

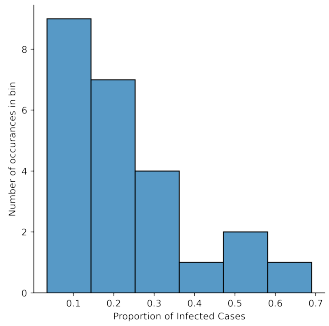

Figure 7: Histogram of observed proportion of population infected by COVID-19 across parameter combinations

combination_number	student_testing_freq	iteration	mask_prob	inclass_lunch	attend_rate	init_patient	seat_dist	output
1	7	0	0.0	True	0.75	3	5	0.2746
2	14	0	0.0	True	0.75	3	5	0.2
3	7	0	0.5	True	0.75	3	5	0.1653
4	14	0	0.5	True	0.75	3	5	0.264
5	7	0	1.0	True	0.75	3	5	0.0346
6	14	0	1.0	True	0.75	3	5	0.16
7	7	0	0.0	False	0.75	3	5	0.2186
8	14	0	0.0	False	0.75	3	5	0.488
9	7	0	0.5	False	0.75	3	5	0.216
10	14	0	0.5	False	0.75	3	5	0.0986
11	7	0	1.0	False	0.75	3	5	0.0693
12	14	0	1.0	False	0.75	3	5	0.1226
13	7	0	0.0	True	1.0	3	5	0.278
14	14	0	0.0	True	1.0	3	5	0.564
15	7	0	0.5	True	1.0	3	5	0.074
16	14	0	0.5	True	1.0	3	5	0.292
17	7	0	1.0	True	1.0	3	5	0.052
18	14	0	1.0	True	1.0	3	5	0.042
19	7	0	0.0	False	1.0	3	5	0.23
20	14	0	0.0	False	1.0	3	5	0.69
21	7	0	0.5	False	1.0	3	5	0.2
22	14	0	0.5	False	1.0	3	5	0.422
23	7	0	1.0	False	1.0	3	5	0.046
24	14	0	1.0	False	1.0	3	5	0.126

Table 3: Parameter combination and model simulation results.

However upon observing the Table 3, we can see that there are several instances (parameter combinations 10, 11, 17, 18, 19 etc.) where the observed case counts are less than 10% (50 Students). To detect if there specific parameters which have particular effectiveness in reducing case counts, we investigate further by building a Random Forest Regressor. The parameter combinations would be the input vector X, and the proportion of COVID-19 positive cases after the 14 day simulation period as the output vector y. This is done not for predictive purposes, but rather to assess if any single parameter plays a critical role in determining the proportion of COVID-19 positive cases in a 14 day period. The random forest regressor was built using scikit-learn using 50 randomly initiated decision trees. The feature importance's were extracted and are displayed below.

attend_rate	inclass_lunch	student_testing_freq	mask_prob
0.1444	0.1055	0.2127	0.5371

Table 4: Table showing the various feature importances from the fitted Random Forest Regressor

As we can see clearly, the importance of mask_prob is 53.71% which is followed by student_testing_freq with 21.27%, attend_rate by 0.144 % and and finally inclass_lunch with 0.1055%. Thus, we see that mask_prob has an overwhelming influence on determining model results. The nature of this factor is substantiated by the results in Table 3 which show that out of all the parameter combinations having a net model result of 10% total proportion or less, over 71% of them arise from having a mask adoption of 100%. Furthermore even the remaining cases arise from having mask adoption of 50%. This suggests that wearing masks is incredibly useful for containing the spread of COVID-19. Simialrly, we can see that increasing the testing frequency also almost always decreases the total number of COVID-19 cases. As seen by the results in Table 3, increasing testing frequency has resulted in caseload reduction in $\frac{12}{14}$ possible cases. The role that testing plays can be also observed from downloading the FacetGrid plot which clearly shows the trajectory of transmission curves changing (For instance in parameter combinations 9, 13, 19) as a consequence of testing. Having understood that mask wearing plays a crucial role in curtailing transmission of COVID-19, we try to quantify this effect statistically. Given that we have one run per simulation, a comparison of means via ANOVA is infeasible. Thus we use the Kolmogorov- Smirnov (ks_2sample) test to quantify the differences in distribution of each pair of parameter combinations (amounting to 276 combinations). We ran the ks_2sample test using SciPy. Furthermore to aid simplicity in comparisons, we took the mean proportion of people infected each day thus creating distributions of 14 elements each, thereby corresponding to a day-by-day granularity. We observed statistically significant differences in 116 out of the 276 combinations. Furthermore we observed that 14 out of the 116 parameter combinations varied in all the three parameters, and that only 13 of them solely deviated on the basis on in mask probability. The results can be seen on the following table:

Thus, we can see that mask probability is one of the most important factors in determining reduction in COVID-19 transmission. Furthermore, we can see that mask probability plays a role in determining (either solely or in combination) significant differences in transmission in over 87.6% of the cases ($\frac{102}{116}$ times). While the evidence as to the importance of mask-wearing is encouraging, we believe that running sensitivity analysis using multiple samples for each parameter combination across a multitude of different parameters will help quantify results better.

Factors	Significant Frequency
attend_rate-mask_prob	14
attend_rate-inclass_lunch-mask_prob-student_testing_freq	14
inclass_lunch-mask_prob-student_testing_freq	14
mask_prob	13
attend_rate-inclass_lunch-mask_prob	12
attend_rate-mask_prob-student_testing_freq	12
mask_prob-student_testing_freq	12
inclass_lunch-mask_prob	11
attend_rate	4
inclass_lunch-student_testing_freq	2
attend_rate-inclass_lunch	2
student_testing_freq	2
inclass_lunch	2
attend_rate-student_testing_freq	1
attend_rate-inclass_lunch-student_testing_freq	1

Table 5: Table showing the various feature combinations which cause significant changes in COVID-19 Transmission

4 Conclusions

We can see that our prototype model and workflow can be used to detect risky activities in school, and assess the effectiveness of intervention strategies in schools. We hope that the Airavata gateway we set up provides value to schools by allowing them to simulate their reopening plans, thus allowing parents, teachers and school administrators to understand how COVID-19 might impact their specific school considering the intervention strategies they had planned. In the future, we anticipate users to be allowed to control several different parameters which can greatly influence model outcomes. Additionally, we will attempt to run our model on a multitude of different parameter combinations to strongly quantify the effects of various factors on COVID-19 transmission in school environments. We hope that such efforts will allow for the development of similar systems for other diseases such as Influenza or Measles which may be more relevant post the pandemic.

References

[1] World Health Organization et al. Modes of transmission of virus causing covid-19: implications for ipc precaution recommendations: scientific brief, 27 march 2020. Technical report, World Health Organization, 2020.

[2] Song Tang, Yixin Mao, Rachael M Jones, Qiyue Tan, John S Ji, Na Li, Jin Shen, Yuebin Lv, Lijun Pan, Pei Ding, et al. Aerosol transmission of sars-cov-2? evidence, prevention and control. *Environment international*, 144:106039, 2020.

[3] Kenichi Azuma, U Yanagi, Naoki Kagi, Hoon Kim, Masayuki Ogata, and Motoya Hayashi. Environmental factors involved in sars-cov-2 transmission: effect and role of indoor environmental quality in the strategy for covid-19 infection control. *Environmental health and preventive medicine*, 25(1):1–16, 2020.

[4] Derek K Chu, Elie A Akl, Stephanie Duda, Karla Solo, Sally Yaacoub, Holger J Schünemann, Amena El-harakeh, Antonio Bognanni, Tamara Lotfi, Mark Loeb, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of sars-cov-2 and covid-19: a systematic review and meta-analysis. *The Lancet*, 395(10242):1973–1987, 2020.

[5] Michael A Kohanski, L James Lo, and Michael S Waring. Review of indoor aerosol generation, transport, and control in the context of covid-19. In *International forum of allergy & rhinology*, volume 10, pages 1173–1179. Wiley Online Library, 2020.

[6] Mahesh Jayaweera, Hasini Perera, Buddhika Gunawardana, and Jagath Manatunge. Transmission of covid-19 virus by droplets and aerosols: A critical review on the unresolved dichotomy. *Environmental research*, page 109819, 2020.

[7] Martin Z Bazant and John WM Bush. Beyond six feet: A guideline to limit indoor airborne transmission of covid-19. *medRxiv*, 2020.

[8] Chih-Chieh Chen and Klaus Willeke. Aerosol penetration through surgical masks. *American journal of infection control*, 20(4):177–184, 1992.

[9] A Synnefa, E Polichronaki, E Papagiannopoulou, M Santamouris, G Mihalakakou, P Doukas, PA Siskos, E Bakeas, A Dremetsika, A Geranios, et al. An experimental investigation of the indoor air quality in fifteen school buildings in athens, greece. *International Journal of Ventilation*, 2(3):185–201, 2003.

[10] David L Johnson, Robert A Lynch, Evan L Floyd, Jun Wang, and Jacob N Bartels. Indoor air quality in classrooms: Environmental measures and effective ventilation rate modeling in urban elementary schools. *Building and Environment*, 136:185–197, 2018.

[11] William Firth Wells et al. Airborne contagion and air hygiene. an ecological study of droplet infections. *Airborne Contagion and Air Hygiene. An Ecological Study of Droplet Infections.*, 1955.

[12] EC Riley, G Murphy, and RL Riley. Airborne spread of measles in a suburban elementary school. *American journal of epidemiology*, 107(5):421–432, 1978.

[13] Mark Nicas, William W Nazaroff, and Alan Hubbard. Toward understanding the risk of secondary airborne infection: emission of respirable pathogens. *Journal of occupational and environmental hygiene*, 2(3):143–154, 2005.

[14] Nikolaos I Stilianakis and Yannis Drossinos. Dynamics of infectious disease transmission by inhalable respiratory droplets. *Journal of the Royal Society Interface*, 7(50):1355–1366, 2010.

[15] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.

[16] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Author correction: Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(9):1491–1493, 2020.

[17] Zhiguo Zhou, Dongsheng Yue, Chenlu Mu, and Lei Zhang. Mask is the possible key for self-isolation in covid-19 pandemic. *Journal of medical virology*, 92(10):1745–1746, 2020.

[18] John Zhai. Facial mask: a necessity to beat covid-19. *Building and environment*, 2020.

[19] Chanjuan Sun and Zhiqiang Zhai. The efficacy of social distance and ventilation effectiveness in preventing covid-19 transmission. *Sustainable cities and society*, 62:102390, 2020.

[20] Rajesh K Bhagat, MS Davies Wykes, Stuart B Dalziel, and PF Linden. Effects of ventilation on the indoor spread of covid-19. *Journal of Fluid Mechanics*, 903, 2020.

[21] Laura Di Domenico, Giulia Pullano, Chiara E Sabbatini, Pierre-Yves Boëlle, and Vittoria Colizza. Modelling safe protocols for reopening schools during the covid-19 pandemic in france. *Nature communications*, 12(1):1–10, 2021.

[22] Laura Di Domenico, Giulia Pullano, Chiara E Sabbatini, Pierre-Yves Boëlle, and Vittoria Colizza. Can we safely reopen schools during covid-19 epidemic? *medRxiv*, 2020.

[23] Simran Preet Kaur and Vandana Gupta. Covid-19 vaccine: A comprehensive status report. *Virus research*, page 198114, 2020.

[24] CK Tanner. Minimum classroom size and number of students per classroom. 2000.