Delineo: Data-Driven Simulation of COVID-19 Spread for Community-Level Decisions and Resilience

M. Insley, S. Bliss, S. Chan, Z. Kim, S. Li, H. Noren, E. Sullivan, R. Abichandani, M. Agarwal, E. Asare, C. Bakhit, J. Bakhit, A. Bao, T. Berkery, S. Cabrejos, S. Cao, Y. Chen, A. Dehade, L. DiBernardo, S. Earla, A. Gandhi, J. Huang,
M. Y. Jia, K. Ko, A. Li, I. Lian, Da. Lu, B. Lee, M. Lee, R. Muratore, N. Najjar, C. Nwosu, Y. Pan, J. Park, S. Li, S. Rakshit, E. Song, O. Wang, A. Wei, O. Wei, P. Xu, D. Yang, J. Yu, R. Yu, J. Zhang, Y. Zhang, A. Dahbura

Johns Hopkins University | Whiting School of Engineering | Baltimore, MD
Design Day 2021

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Introduction

Since the first emergence of SARS-CoV-2 in December 2019, COVID-19 has become a public health crisis and challenge all over the world. However, questions remain on its modes of transmission, treatments, and vaccines even after a year of the initial outbreak. Although many projection models have been launched to estimate the number of cases and deaths, these models failed to account for different types of communities and environments people live in. In that, significant insight into the mechanisms for disease spread at the community level is necessary and significant, especially across an extremely diverse array of communities, each with its own structure and idiosyncrasies.

Therefore, to address this issue, the Delineo proposes a combination of agent-based and compartment models that brings variability and adaptability to each individual, community, and country. By using geolocation and demographic data to simulate interactions, Delineo predicts the spread of the virus in an individualized, complex setting and under targeted environments. Our goal is for Delineo to fill this critical gap by providing a community-level simulator to public health experts and decision-makers so that they have tools they need to navigate the maze of uncertainty created by these invisible threats.

Objectives

The Delineo community-level COVID-19 simulator has the goal of piercing the veil of pandemic spread by providing:

1) insights into how COVID-19 and future transmissible diseases spread at the local level for specific communities, and
2) an environment for users to explore which interventions tailored to the local level are the most effective at containing the disease, leading to the novel concept of *precision public health*.

Using community-level simulation will allow local-level decision-makers to implement *precision* interventions that are less one-size-fits-all and more tailored to combat spread in a specific community, in all likelihood with less negative impact on the economic and psychological well-being of the community.

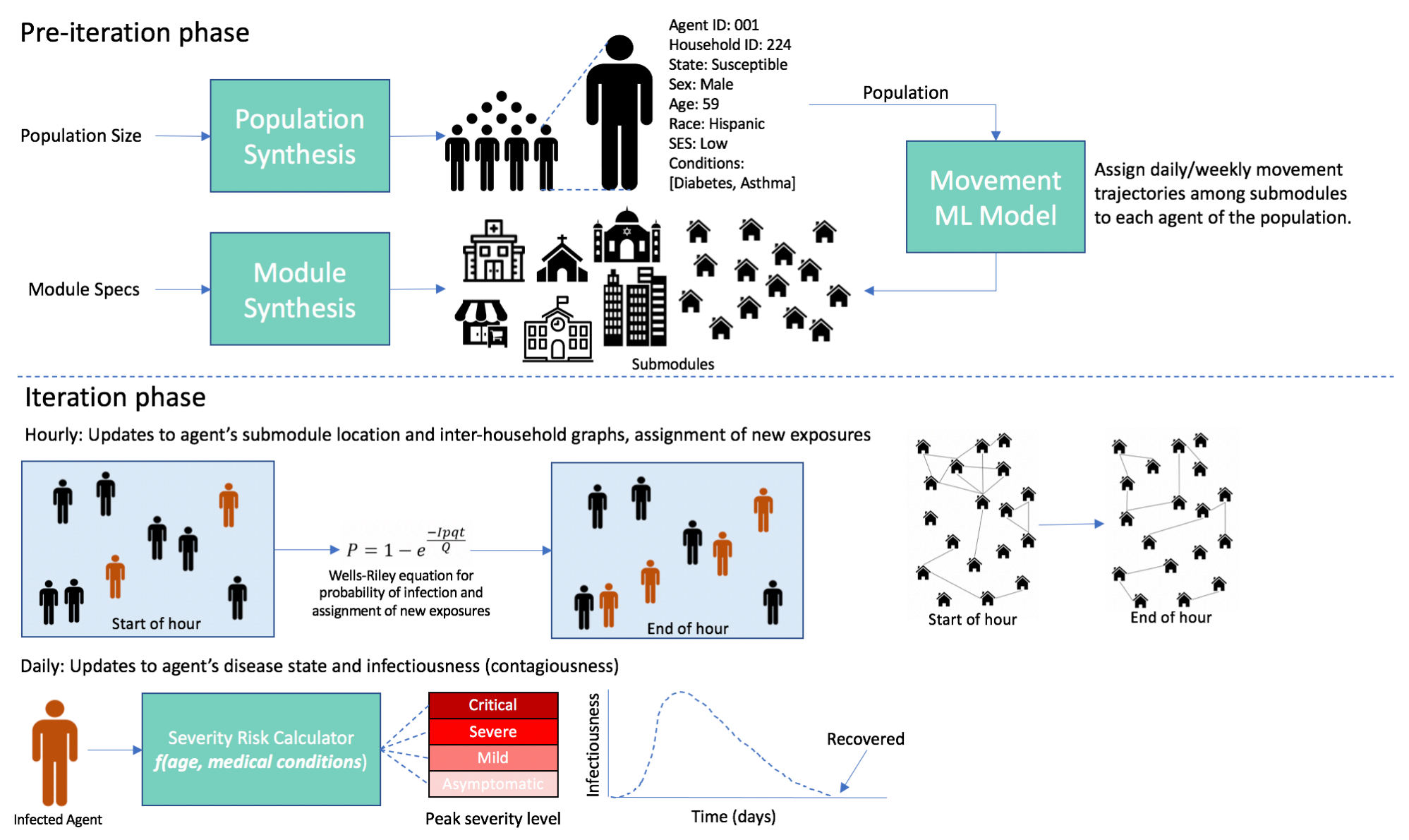
Materials and Methods

We consulted experts in the field to determine the best methods in creating our novel disease model. Their mentorship led us to the utilization of the Iterative Proportional Fitting Procedure (IPFP) and the Wells-Riley equation as described in the following sections. Additionally, research was done to generate informed conclusions about the general disease characteristics of COVID-19 as well as trends in contagiousness, transmissibility, and disease trajectory.

We also sourced and applied various datasets that informed our simulation. Firstly, the demographic information of agents within our synthetic population was generated from United States Census Bureau data, where we were able to compile age, sex, race, and socioeconomic status (SES) information per geographic location. Additionally, our team employed geolocation data from Safegraph, a data company, to inform the mobility networks and agent trajectories within our model.

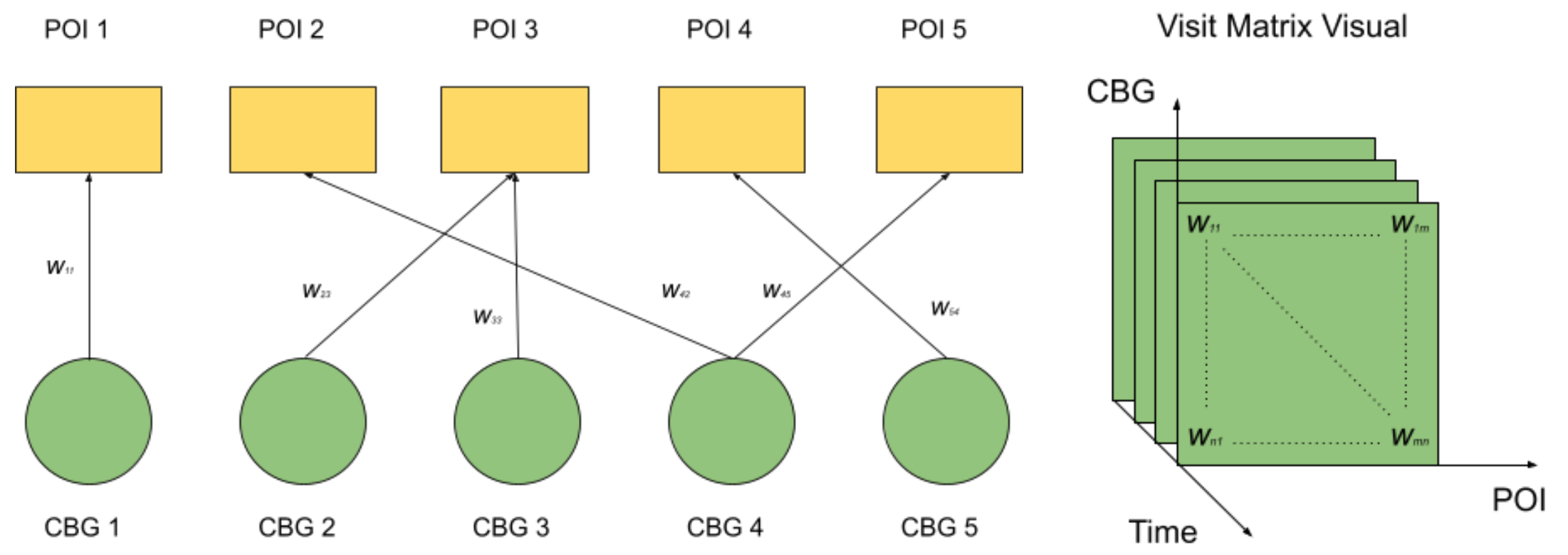
Results

Simulation Model & Synthesis

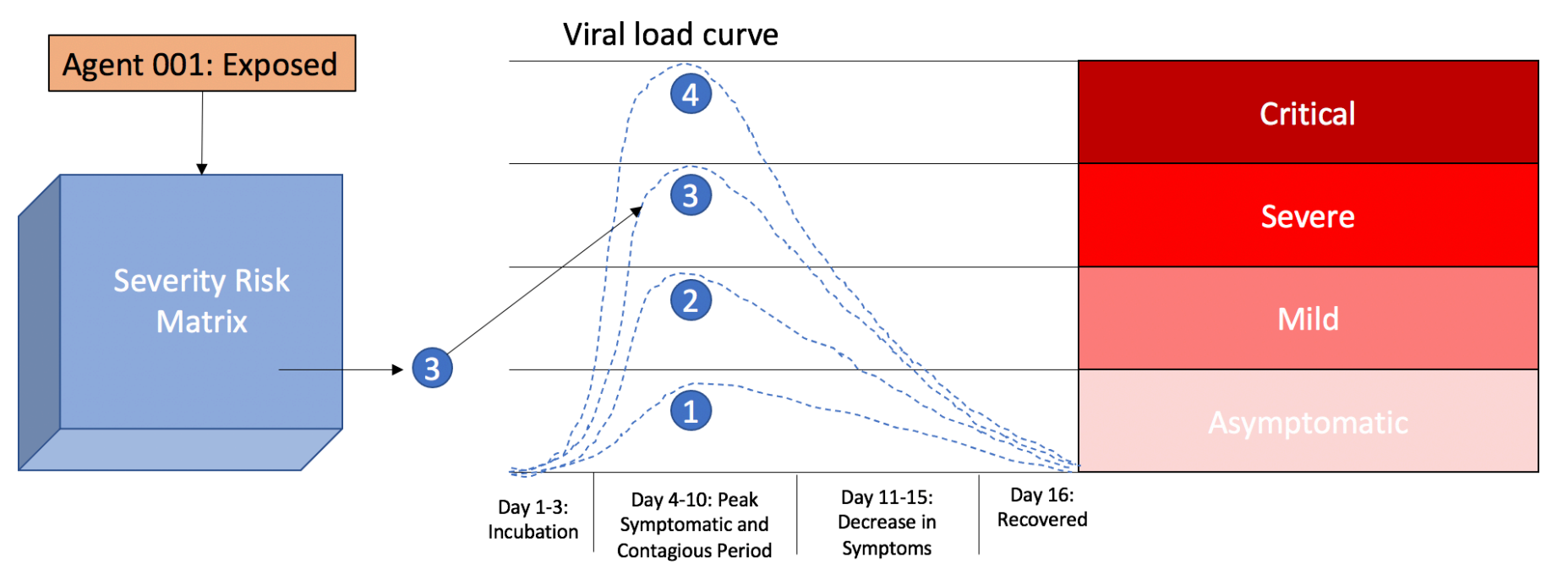


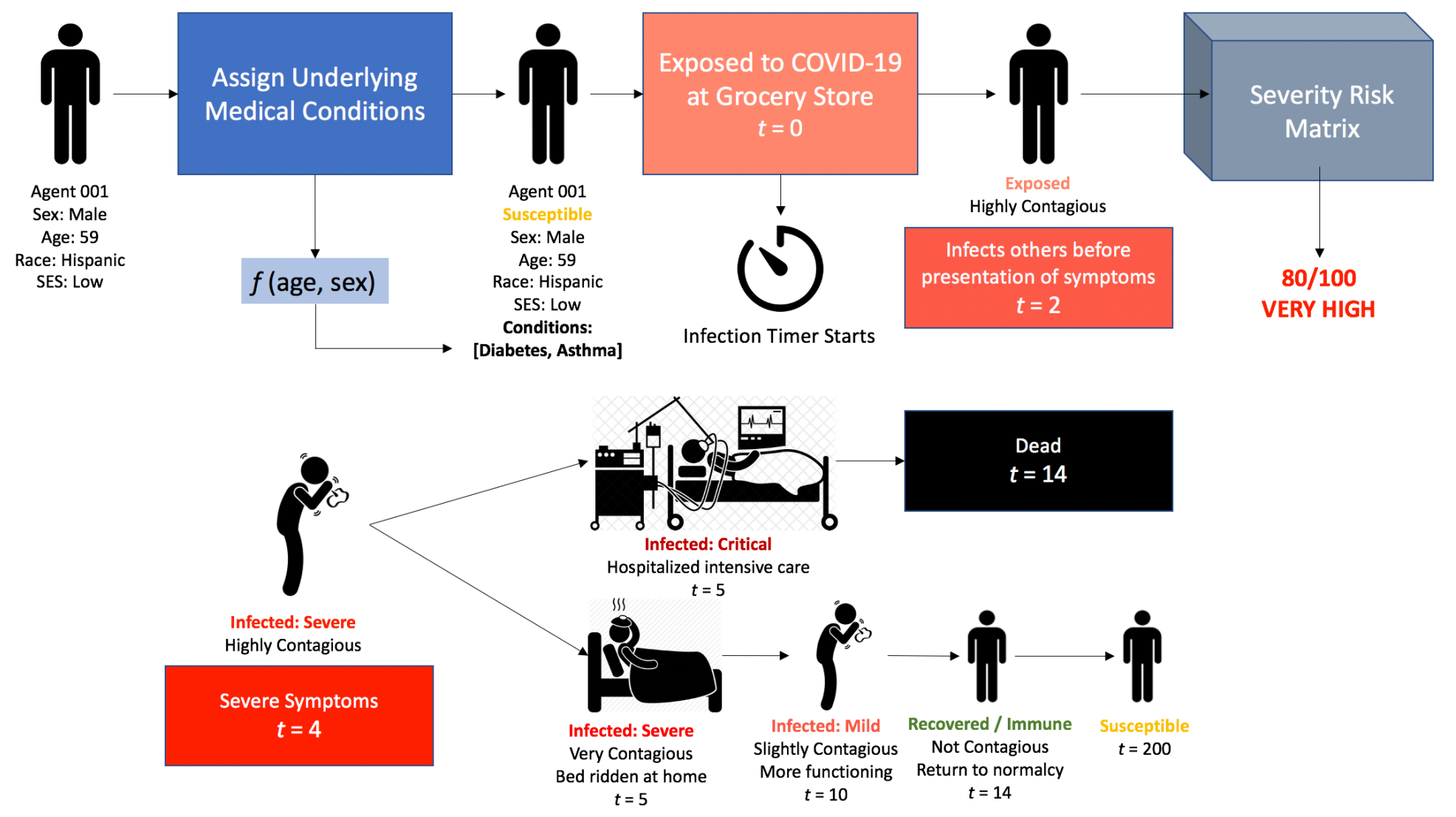
There are two phases built into our simulation. The **pre-iteration phase** involves the creation of a synthetic population using an open-source project called SynthPop, a reimplementation of PopGen using the modern scientific Python stack. After incorporating additional demographic information such as gender, age, race, and socioeconomic status into the synthetic population through existing census data, the underlying medical conditions are distributed among the population based on known demographic predispositions. Afterwards, SynthPop assigns agents into households given their specific demographic data as well as daily and weekly movement trajectories among the submodules, or facilities, within the simulation.
The **iteration phase** initiates the simulation. Hourly updates change each agent's submodule location and assigns new exposures of the disease if needed. The determination of infection is informed by the Well-Riley equation which provides a probability of infection informed by multiple facility-based variables. Additionally, household-based interactions will shift between each hour. As for daily updates, the model will take in new and existing exposures to revise each agent's disease state and contagiousness.

Agent Mobility Patterns: Iterative Proportional Fitting Procedure



Our model uses the **Iterative Proportional Fitting Procedure** (IPFP), a classical algorithm in computer science that performs data scaling based on current datasets, to create a mobility network containing information about the number of individuals (agents) traveling from each Census Block Group (CBG) to Point of Interest (POI). A Census Block Group is the smallest geographical unit for which the Census bureau publishes sample data and has a population of around 600 to 3000 people. We define a POI as any location (facility) that an agent can visit such as a coffee shop, school, workplace, etc.
Safegraph provides our model with accurate datasets outlining how many people from each CBG arrive at a particular POI every hour. This dataset is utilized to construct a CBG Marginal and a POI Marginal that are linked to the Visit Matrix which records the number of people travelling from a Census Block Group to a Point of Interest. Since all POI's have varying durations of visit, two assumptions were made. Firstly, that each visitor to a POI spends a median length of time calculated in that POI and secondly, that each new visitor in a POI is equally likely to arrive again any time from that time to the next hour. Extreme outliers are also truncated. The Dwell Time Correction Factor is then used to adjust the CBG and POI values, which are utilized in IPFP to calculate the adjusted visitor matrix which serves as a probabilistic model for the general mobility network.
From this, trajectories for individual agents are modelled. After extrapolating that data to represent 20% of the US population, it is scaled up by a predetermined factor in the IPFP process to accurately delineate the complete picture. Our method also tries to reduce information entropy and limits the amount of error.
Once a mobility network is assembled using our Machine Learning model, it is integrated with the main simulation. Agents of our synthetic population are then assigned to each 'role' (POI) depending on their CBG such that the simulation accurately tracks real-world behavior, trends, demographics and restrictions such as stay-at-home orders.



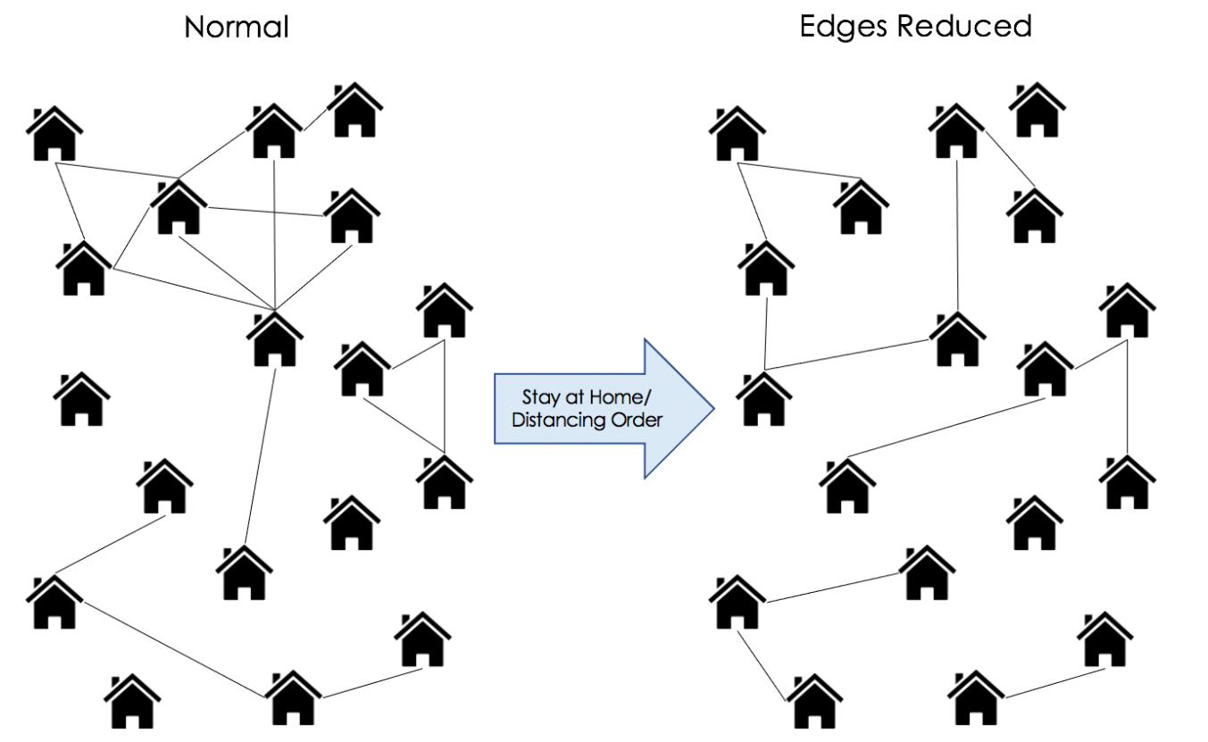


Severity-Based Disease Model

Our disease model splits the severity of infection into four states, asymptomatic, mild, severe, and critical, depending on the typical symptoms displayed by an individual.
An agent's possibility of contracting disease depends on their **severity risk (SR) score** and the probability of infection within a facility. The SR score, which ranges from 0-100, takes into account sex, age, and underlying conditions. This score will determine the probability that an agent is assigned one of the four peak disease states, which determines how quickly an agent recovers and/or is removed from the population. These disease states will also dictate an agent's level of contagiousness, along with time spent in the "infected" state. If an agent is established to be infected, they will move through all the lower disease states until they reach their assigned peak state, then recover. From our research, we have determined that a portion of those critical will die from the disease and thus be removed from the simulation's population.

$$P = 1 - e^{\frac{-Ipqt}{Q}}$$

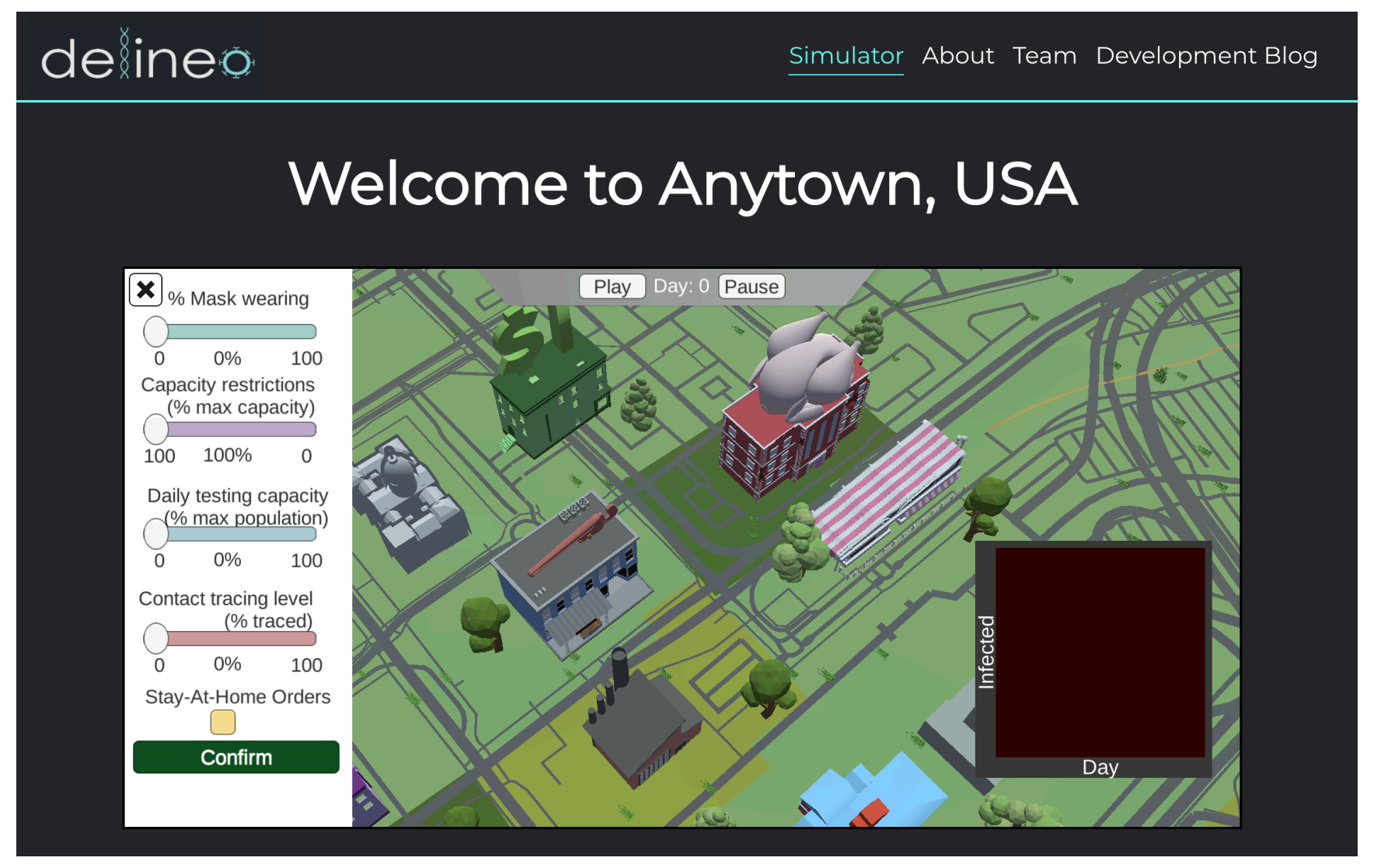
P = *Probability of infection*
I = *Number of infector individuals in the space*
p = *Average breathing rate of individuals in the space*
q = *Quanta generation rate*
t = *Exposure time*
Q = *Air flow rate from HVAC system*



Facility Submodule Spread: Wells–Riley & the Household Graph Approach

A submodule is any location/place/building that individuals in the population can be found at any point in time. A submodule is defined by its overall type/label (retail, school, home, etc), its size, and its location (detailed by its latitude & longitude). Based on the overall type/label of the submodule, the **Wells-Riley equation** is utilized to determine a probability of infection (*P*). This equation has typically been used to assess ventilation strategies and their associations to airborne pathogen infectivity. In our model, the facility's properties will inform the variables *Q* and *p* within the equation, whereas the agents and their disease states will inform variables *I* and *p*. Every hour, the simulation takes in a facility's factors to determine *P* and new exposures (infections) are assigned.
However, a household tends to spend more time together in proximity. Additionally, household interactions are unique to how agents interact within a facility. Thus, we have utilized **a graph approach to model inter-household interactions**. With this approach, each household becomes a node in a graph and its edges represent interactions between households. Within the simulation, there will be *m* groups of *n* households interacting. In conjunction to other graph algorithms, these groups of households would be represented on the graph as SCCs (Strongly Connected Components). All nodes within an SCC are reachable from one another. This allows us to quantify what kind of interaction occurs. Edge weights represent the level of contact/distance between households, with closer interactions having larger edge weights. When assigning infection, we ignore edges between two non-infected, susceptible households (households where all agents have not been infected). Therefore, we can iterate through infected households and use the edge weights to assign infections.
Our utilization of the Wells-Riley equation and the graph approach **allows flexibility to insert different interventions** such as mask wearing, capacity restrictions, testing, contact tracing, quarantine, and stay-at-home orders into our model. For interventions that affect agent behavior and properties such as mask wearing, we modify the Wells-Riley variables *q* and *p* to indicate their effects on contagiousness. Additionally, interventions that decrease agent movement into facilities such as contract tracing, quarantine, stay-at-home orders, and capacity restrictions will decrease the number of individuals within each facility (variable *I*) as well as the number of edges within the overall household graph. These modifications impact the accumulation of overall number of infected individuals within our simulation.

Delineo Website & Anytown USA



The Delineo website provides an intimate user experience that allows for the general population to gain a better understanding of the various factors and public health measures that impact disease spread. In addition to the general information provided about our project, we've built a simplified web-based version of our model, **Anytown USA**. Unlike Delineo which can be customized to any area within the United States, Anytown USA is a disease model of a typical small town in rural America, providing the user with an educational experience of how different variables and interventions affect the spread of COVID-19.

Conclusion

Community-based simulation is a critical component for understanding the mechanisms underlying the spread of disease. Delineo, and its simplified version Anytown USA, provides a tool for public health experts, policymakers, researchers, teachers and students to lift the veil of uncertainty on infectious respiratory diseases such as COVID-19 to design safe and effective preventative measures and intervention strategies. Please visit www.covidweb.isi.jhu.edu to use Anytown USA and for additional information.