

# Agent-Based Stochastic SIR Model with Variable Distributions for Infection and Recovery Probabilities, Including Hospitalization Dynamics

Arushi Srivastava (210100027)

**Abstract**—This report presents my project's findings on modelling of a contagion on a population and effective hospital placement to reduce the effect of the outbreak. It is vital to model disease prevention strategies and find the most optimised combination of resources to mitigate risks in real-life scenarios. For example, during COVID-19, understanding how to efficiently allocate hospital beds, implement targeted lockdowns, and roll out vaccination programs became essential to managing the spread. Modelling these strategies allows us to simulate different scenarios and assess the impact of each intervention, helping to determine the best course of action. By anticipating needs and outcomes, we can create well-informed protocols that can save lives, reduce strain on healthcare systems, and minimise socioeconomic disruption in the event of an outbreak.

## I. BACKGROUND & MOTIVATION

I was inspired to take this up as my topic when I researched about a game I had played quite a lot when I was young, called *Plague Inc.* *Plague Inc.* is a game where you create and evolve a disease to try and infect everyone on Earth. The challenge is to spread it as widely as possible without getting cured. You have to use strategy to optimise your limited resources: deciding whether to focus on how it spreads (like through air or water) or on symptoms that make it more deadly. You also need to adapt to different environments, like hot or cold countries, to keep infecting people effectively. It's all about making smart choices to keep your disease spreading and avoid detection. I was always interested on how to best the game, and had searched many different techniques back then on how to best the game, and win based on strategy and not on luck.

I knew I wanted to create an optimization model with similar parameters to this nostalgic game, where a disease is spreading hard and fast in its attempt to wipe out humanity. But with a twist. Instead of designing a more powerful contagion, I wanted to develop a model where the focus is on saving as many lives as possible in a fast-evolving environment. In my version, while the disease advances with

new traits, the population responds with dynamic strategies for containment, treatment, and recovery. The goal shifts to building resilience and optimising resources to counteract the outbreak, modelling how society can work strategically to protect and heal against the spread.



Fig. 1. Growth in Startup Ecosystem in India

## II. AGENT TYPES, SCHEDULES, AND CONNECTIVITY

In my simulation, I use an agent-based approach with three types of agents: **children**, **adults**, and **young adults**. Each agent follows a predefined schedule that determines where they are located and when, influencing their potential interactions and exposure to infection.

**Children** follow this schedule:

- **Home:** 0:00 AM – 9:00 AM, 6:00 PM – 12:00 AM
- **School:** 9:00 AM – 3:00 PM
- **Park:** 3:00 PM – 5:00 PM

**Adults** follow this schedule:

- **Home:** 0:00 AM – 8:00 AM, 8:00 PM – 12:00 AM
- **Work:** 8:00 AM – 8:00 PM

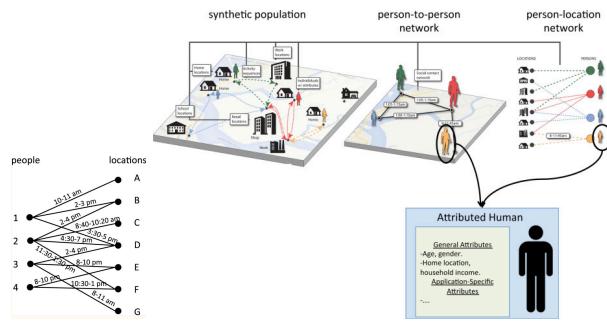
**Young Adults** follow this schedule:

- **College:** 9:00 AM – 6:00 PM
  - **Rest at Hostel:** 6:00 PM – 9:00 AM
- Every **60 days**, young adults go home for **24 hours**.

## Project Report

The agents are connected based on shared locations, with connections forming whenever agents are at the same place during overlapping time intervals. This results in a person-to-person contact network, which is key to modeling the infection spread.

I use a **location-based connectivity** structure that follows a bipartite graph format. One set of vertices represents the agents (people), and the other set represents locations (such as home, school, work, and park). If two agents are co-located at the same location during the same time, an edge is formed between them. This connectivity is critical to simulate how infections spread through contact between agents.



**Fig. 2.** Agent- Co Location Based Interactions

This model draws inspiration from the "**Agent-Based Computational Epidemiological Model**" paper, which employs a similar approach using an implicit dependency graph. The interactions between agents are based on their co-location, with edges forming whenever they meet at a shared location during the same time. The time each agent spends with another agent (contact duration), is found using this co-location system. These interactions play a central role in the infection transmission process, which is modeled stochastically, incorporating factors like infectivity, susceptibility, and recovery probabilities. In the next section, I will discuss further on how the probability of infection is calculated.

### III. INFECTION PROBABILITY CALCULATION FOR EACH AGENT

A. In my model, the probability of infection between two agents is calculated when a susceptible individual and an infectious individual are co-located. The formula used to calculate the probability of the infected individual (i) infecting the susceptible individual (j) is:

$$p_{i \rightarrow j} = 1 - (1 - r_i s_j \rho)^{\tau}$$

Where:

- $p_{i \rightarrow j}$  is the probability that the infectious agent (i) will infect the susceptible agent (j).
- $\tau$  is the duration of their exposure.

- $r_i$  is the infectivity of agent (i), which determines how likely they are to transmit the infection.
- $s_j$  is the susceptibility of agent (j), indicating how likely they are to become infected.
- $\rho$  the transmissibility, a constant property of the disease that represents the probability of a completely susceptible person becoming infected by a completely infectious person during one hour (my simulation period) of exposure. This value is calibrated to produce a desired attack rate, which is the fraction of the total population that becomes infected

### B. Values of infectivity, susceptibility and transmissibility

For the **infectivity** of each agent, I modeled it based on an exponential decay, similar to the approach used in the paper "*A Stochastic Agent-Based Model of the SARS-CoV-2 Epidemic in France*". In that paper, the risk of contamination was highest at the onset of symptoms and decreased over time. To account for the risk of transmission before symptoms develop, they assumed that infected individuals became contagious one day after infection, with their contagiousness decreasing exponentially the further they were from symptom onset. I applied a similar method, using an exponential decay function for infectivity:

$$I(t) = A \cdot e^{-(bt)} \text{ Where:}$$

- $A$  is the initial infectivity value, which I set as a uniform distribution between two values (x and y) for each type of agent.
- $b$  is the decay rate, determining how quickly the infectivity decreases over time.
- $t$  represents time since infection

### ADD GRAPH

The values he range for initial infectivity (A) and the decay rate for each type of agent were determined using the **DTMC** model I employed to simulate parameter values. This exponential decay fits well with the dynamics of viral replication, as it captures the high transmission risk early in the infection and the gradual decrease as time progresses.

For **susceptibility**, I used a **gamma distribution**, similar to the approach outlined in the paper "*Individual Variation in Susceptibility or Exposure to SARS-CoV-2 Lowers the Herd Immunity Threshold*". Each agent type (children, adults, young adults) had different shape and scale values for the gamma distribution. The values for the range of shape and scale parameters used for each type of agent were estimated using the **DTMC** model. This method allows for variability in susceptibility between individuals and accounts for how some people may be more or less susceptible to infection based on individual characteristics

## Project Report

### C. Use of Discrete-Time Markov Chain (DTMC) for analysis of values of infectivity, susceptibility, transmissibility

For my analysis, I used a Discrete-Time Markov Chain (DTMC) model to study the SIR epidemic process, focusing on how the values of susceptibility, infectivity, and transmissibility (i.e., the attack rate) affect the spread of the disease. I referenced the paper “*An Introduction to Stochastic Epidemic Models*” for finding these values

#### Deterministic Approach to SIR Modelling

4 Linda J. S. Allen

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta}{N}SI + b(I+R) \\ \frac{dI}{dt} &= \frac{\beta}{N}SI - (b+\gamma)I \\ \frac{dR}{dt} &= \gamma I - bR,\end{aligned}\quad (3)$$

where  $\beta > 0$ ,  $\gamma > 0$ ,  $b \geq 0$ , and the total population size satisfies  $N = S(t) + I(t) + R(t)$ . The initial conditions satisfy  $S(0) > 0$ ,  $I(0) > 0$ ,  $R(0) \geq 0$ , and  $S(0) + I(0) + R(0) = N$ . We assume that the birth rate equals the death rate so that the total population size is constant,  $dN/dt = 0$ .

The basic reproduction number (2) and the birth rate  $b$  determine the dynamics of model (3). The dynamics are summarized in the following theorem.

**Theorem 2.** Let  $S(t)$ ,  $I(t)$ , and  $R(t)$  be a solution to model (3).

i) If  $\mathcal{R}_0 \leq 1$ , then  $\lim_{t \rightarrow \infty} I(t) = 0$  (disease-free equilibrium).

ii) If  $\mathcal{R}_0 > 1$ , then

$$\lim_{t \rightarrow \infty} (S(t), I(t), R(t)) = \left( \frac{N}{\mathcal{R}_0}, \frac{bN}{b+\gamma} \left( 1 - \frac{1}{\mathcal{R}_0} \right), \frac{\gamma N}{b+\gamma} \left( 1 - \frac{1}{\mathcal{R}_0} \right) \right)$$

(endemic equilibrium).

iii) Assume  $b = 0$ . If  $\mathcal{R}_0 \frac{S(0)}{N} > 1$ , then there is an initial increase in the number of infected cases  $I(t)$  (epidemic), but if  $\mathcal{R}_0 \frac{S(0)}{N} \leq 1$ , then  $I(t)$  decreases monotonically to zero (disease-free equilibrium).

#### DTMC approach to SIR modelling

##### 3.3 SIR Epidemic Model

Let  $S(t)$ ,  $I(t)$ , and  $R(t)$  denote discrete random variables for the number of susceptible, infected, and immune individuals at time  $t$ , respectively. The DTMC SIR epidemic model is a bivariate process because there are two independent random variables,  $S(t)$  and  $I(t)$ . The random variable  $R(t) = N - S(t) - I(t)$ . The bivariate process  $\{(S(t), I(t))\}_{t=0}^{\infty}$  has a joint probability function given by

$$p_{(s,i)}(t) = \text{Prob}\{S(t) = s, I(t) = i\}.$$

This bivariate process has the Markov property and is time-homogeneous.

Transition probabilities can be defined based on the assumptions in the SIR deterministic formulation. First, assume that  $\Delta t$  can be chosen sufficiently small such that at most one change in state occurs during the time interval  $\Delta t$ . In particular, there can be either a new infection, a birth, a death, or a recovery. The transition probabilities are denoted as follows:

$$p_{(s+k,i+j),(s,i)}(\Delta t) = \text{Prob}\{(\Delta S, \Delta I) = (k, j) | (S(t), I(t)) = (s, i)\},$$

The **transition matrix** captures the probabilities of moving between different infected states during the time interval  $\Delta t$ :

An Introduction to Stochastic Epidemic Models 7

$$\begin{pmatrix} 1 & d(1)\Delta t & 0 & \cdots & 0 & 0 \\ 0 & 1 - (b+d)(1)\Delta t & d(2)\Delta t & \cdots & 0 & 0 \\ 0 & b(1)\Delta t & 1 - (b+d)(2)\Delta t & \cdots & 0 & 0 \\ 0 & 0 & b(2)\Delta t & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & d(N-1)\Delta t & 0 \\ 0 & 0 & 0 & \cdots & 1 - (b+d)(N-1)\Delta t & d(N)\Delta t \\ 0 & 0 & 0 & \cdots & b(N-1)\Delta t & 1 - d(N)\Delta t \end{pmatrix},$$

Fig. 3. Transition matrix for an SIS Model

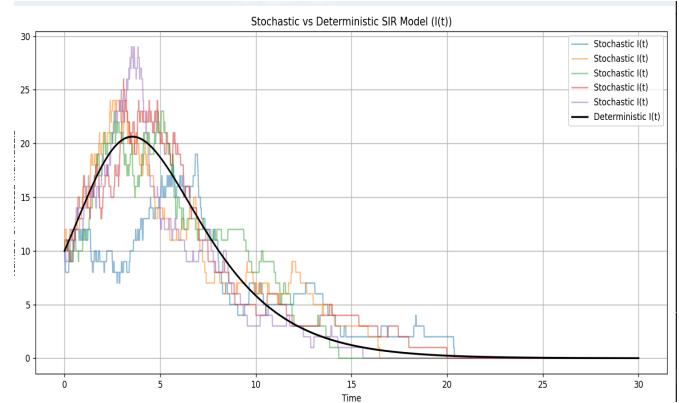


Fig. 4. Result of Deterministic vs. DTMC Approach for  $N = 100$ ,  $\beta = 1$ ,  $\gamma = 0.5$ ,  $\Delta t = 0.01$ ,  $T = 30$ , steps =  $T / \Delta t$ , initial state =  $(90, 10)$ , no. of paths = 5

To calculate the ranges for susceptibility, infectivity, and transmissibility for each agent based on their age, I made the following changes:

1. I defined the probability value as

$$p_{i \rightarrow j} = 1 - (1 - r_i s_j \rho)^{\tau}$$

Here, **contact duration** is set to 0.25, representing an average of 6 hours (a quarter of a day) that agents are in contact with others.

2. The **effective beta** is calculated by multiplying the probability of infection (from the formula above) with the contact rate. This gives us the probability of an infection occurring when two agents come into contact:

effective beta=probability×contact rate

3. The **transmissibility** value was the attack rate, or the fraction of infections desired over the total population

4. By adjusting the initial values of infectivity, susceptibility, and transmissibility, we can test whether an epidemic will occur. For us, an epidemic is defined as affecting 60% or more of the population. The model, based on the **DTMC** approach and reproductive numbers, helps us understand this spread.

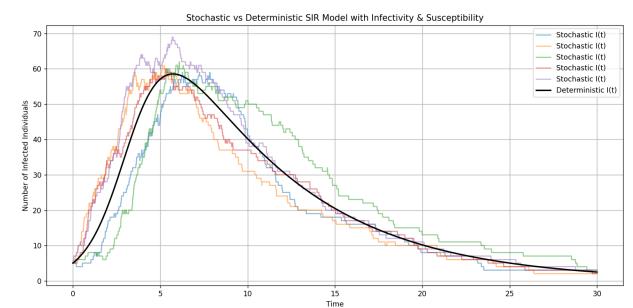


Fig. 5. Result of Deterministic Vs Modified DTMC

## Project Report

approach with  $N = 100$ , transmissibility = 0.8, infectivity = 1, decay rate (for infectivity) = 0.04, susceptibility = 1, contact rate = 3, contact duration = 0.25,  $\gamma = 0.1428$ ,  $\Delta t = 0.01$ ,  $T = 30$ , steps =  $T / \Delta t$ , initial state = (95, 5), no. of paths = 5

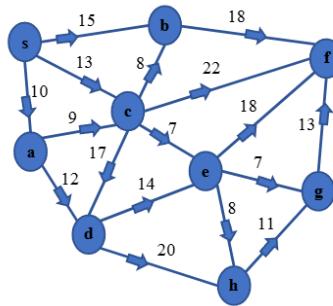
### IV. RECOVERY RATE CALCULATION

The recovery rate  $\gamma$  is defined as 1 divided by the infectious period of the agent. For each agent type, recovery happens if and only if and when they are hospitalized (details on hospitalization will be explained later). The time to recovery follows a log-normal distribution, with different mean and variation values assigned for each agent type, this was done in reference to the paper “*Delayed epidemic peak caused by infection and recovery rate fluctuations*”

### V. HOSPITAL ALLOCATION & PATIENT ISOLATION STRATEGY

For hospital allocation, the "Greedy Algorithm based Health Care Resources Management System in the Times of a Pandemic" paper was used. Dijkstra's algorithm was applied to select the nearest hospital where a patient could be admitted. Each hospital was mapped as a node, and the system was treated as a weighted directed graph based on distance information between hospitals. Hospital capacity was also checked before admitting any patient.

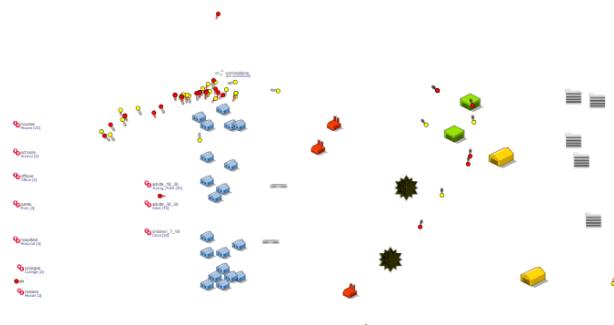
A patient was only sent to the hospital if their symptom severity (a variable) crossed 60%. If not, they remained infected in the system until they eventually recovered. Hospitalization served the purpose of isolating highly symptomatic individuals from susceptibles, while mildly symptomatic patients stayed in the community, continuing to spread the infection, and recovered after a while on their own.



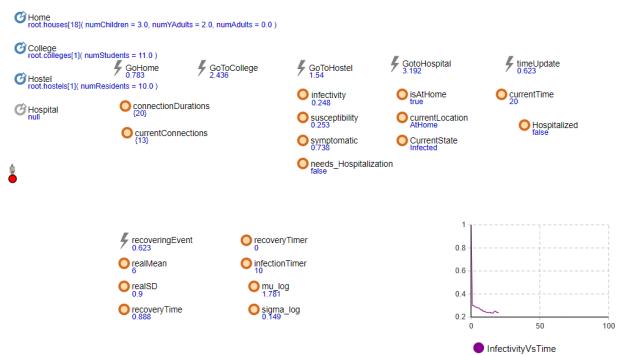
**Fig. 6.** Weighted Dijkstras Algorithm for Hospital Allocation

### Hypothesis for Participants:

### VI. MODEL AND RESULTS



**Fig. 7.** Population and Environment Model

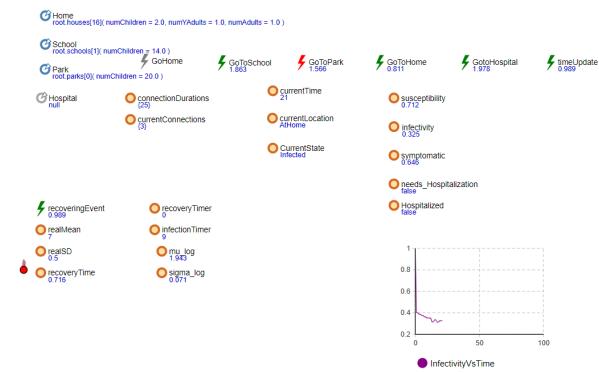


**Fig. 7.** Data for one Young Adult



**Fig. 7.** Data for one Adult

## Project Report



**Fig. 7.** Data for one Child



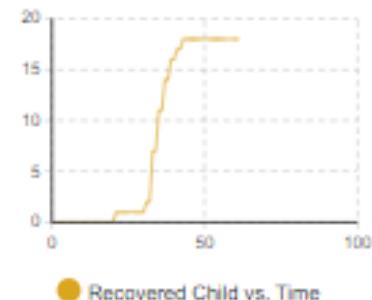
● Current Infections Child vs. Time



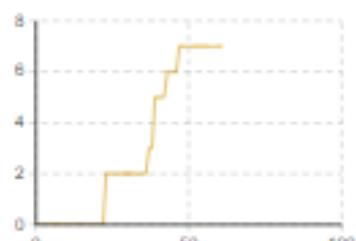
● Current Infections Adult vs. Time



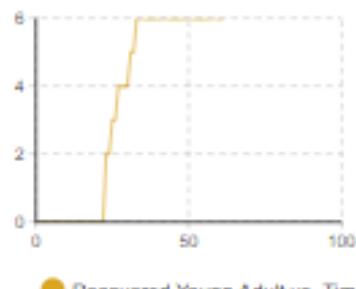
**Fig. 8.** Current Infections of Each Kind of Agent



● Recovered Child vs. Time

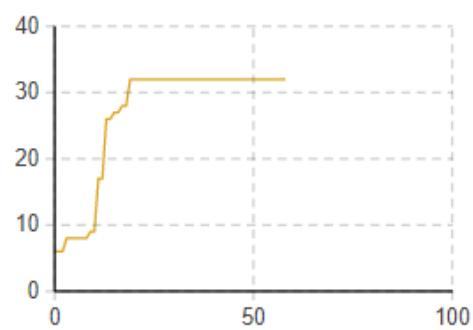


● Recovered Adult vs. Time



● Recovered Young Adult vs. Time

**Fig. 9.** No. Of Recoveries for Each Kind of Agent



● Total Infections vs. Time

## Project Report

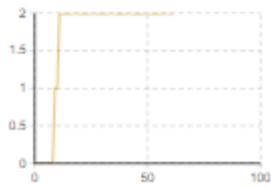
**Fig. 10.** Total Infections Overall



● Total Infections Child vs. Time



● Total Infections Adult vs. Time



● Total Infections Young Adult vs. Time

**Fig. 11.** Total Infections Overall for Each Kind Of Agent

## VII. Acknowledgments

I would like to express our sincere gratitude to Prof. Avinash Bhardwaj and Prof. Siuli Mukhopadhyay for guiding me throughout the project with their inputs and recommendations, as well as giving us a fascinating project to work on.

### Project Files & References:

<https://drive.google.com/drive/folders/1hOtKfEsMPa7QxND4d9CVIjqcC0EL-wW0?usp=sharing>