# Agent based simulators for Covid-19: Simulating larger models using smaller ones

Daksh Mittal, Sandeep Juneja, Shubhada Agrawal
TIFR Mumbai

August 29, 2022

**Abstract**

Agent-based simulators (ABS) are a popular epidemiological modelling tool to study the impact of various non-pharmaceutical interventions in managing an epidemic in a city (or a region). They provide the flexibility to accurately model a heterogeneous population with time and location varying, person-specific interactions. Government policies such as partial and localised lock downs, case isolation, home quarantine, school closures, partially opened workplaces, etc. and important pandemic developments over time including presence of variants as well as vaccines, are easily incorporated in an ABS. Typically, for accuracy, each person is modelled separately. This however may make computational time prohibitive when the city population and the simulated time is large. In this paper, we primarily focus on the COVID-19 pandemic and dig deeper into the underlying probabilistic structure of a generic ABS to arrive at modifications that allow smaller models to give accurate statistics for larger ones. We observe that simply considering a smaller aggregate model and scaling up the output leads to inaccuracies. We exploit the observation that in the initial Covid spread phase, the starting infections create a family tree of infected individuals more-or-less independent of the other trees and are modelled well as a multi-type super-critical branching process. Further, although this branching process grows exponentially, the relative proportions amongst the population types stabilise quickly. Soon after, for large city population, once enough people have been infected, the future evolution of the pandemic is closely approximated by its mean field limit with a random starting state. We build upon these insights to develop a shifted, scaled and restart based algorithm that accurately evaluates the ABS's performance using a much smaller model while carefully reducing the bias that may otherwise arise. We theoretically support the proposed algorithm through an asymptotic analysis where the population size increases to infinity.

## 1 Introduction

COVID-19 is an ongoing pandemic caused by the SARS-COV 2 virus that began in December 2019 in Wuhan and through more infectious and dangerous variants continues to rage into its third year. The reported cases worldwide are around 496 million and reported fatalities 6.17 million as of April 08, 2022. Early on, countries across the world implemented a variety of non-pharmaceutical interventions such as testing, tracing, tracking, isolation, quarantining, enforcement of mask wearing, social distancing, shutting public transportation, lockdowns, etc. to reduce the disease spread. Later, since December 2020, vaccines became a key weapon in fight against the virus. Predicting the consequences of presence of different variants and the vaccines, as well as of non medical interventions is important in selecting the preferred course of action. Epidemiologists often resort to sophisticated mathematical models that predict the related statistics of health indicators including the total number of infected, hospitalisations, critical cases and fatalities as a function of time.

The classical 'aggregate' models for disease spread capture the mean field dynamics of the disease within the population through simple ordinary differential equations (ODEs). These models typically involve a few parameters that need to be estimated from data, are easily solved using a numerical ODE solver and are good at highlighting macroscopic trends (see, [18] for seminal work. A few recent references amongst thousands are [5], [8], [15], [3]). However, such aggregate models often are inadequate in accurately capturing the initial uncertainty inherent in the disease spread and its impact on future disease evolution, impact of more detailed, localised and complex interventions, emergence of new variants, and

their co-evolution with the existing ones, and other important features such as detailed implementation of a vaccination strategy.

**Agent-based simulators and their advantages:** In this paper we focus on another popular class of models, namely the agent-based simulators (ABS) (See [16], [7]). In such models, a synthetic copy of the region is constructed on a computer that attempts to capture the population interaction spaces and detailed disease spread as well as disease spreading interactions as they evolve in time. Typically, each individual in the region is modelled as an agent, so that total number of agents equals the total region population. The constructed individuals reside in homes, children may go to schools, adults may go to work. Individuals also engage with each other in community spaces (to capture interactions in marketplaces, restaurants, public transport, and other public places). Homes, workplaces, schools sizes and locations and individuals associated with them, their gender and age, are created to match the region census data and are distributed to match its geography. Government policies, such as partial, location specific lock downs for small periods of time, case isolation of the infected and home quarantine of their close contacts, closure of schools and colleges, partial openings of workplaces, etc. are easily modelled. Similarly, variable compliance behaviour in different segments of the population that further changes with time, is easily captured in an ABS. Further, it is easy to introduce new variants as they emerge, the individual vaccination status, as well as the protection offered by the vaccines against different variants as a function of the evolving state of the epidemic and individual characteristics, such as age, density of individual's interactions, etc (see [17]). Thus, this microscopic level modelling flexibility allows ABS to become an effective strategic and operational tool to manage and control the the disease spread. See [7] for an ABS used for UK and USA related studies specific to COVID-19, [9] for a COVID-19 study on Sweden, [1], [12] for a study on Bangalore and Mumbai in India. See, e.g., [10] and [16] for an overview of different agent-based models.

**Key drawback:** However, when a reasonably large population is simulated, especially over a long time horizon, an ABS can take huge computational time and this is its key drawback. This becomes particularly prohibitive when multiple runs are needed using different parameters. For instance, any predictive analysis involves simulating a large number of scenarios to provide a comprehensive view of potential future sample paths. In model calibration, the key transmission parameters are fit to the infection data during the early stages of the disease, requiring many computationally demanding simulations. Many other important parameters such as infectiousness of new variants are similarly selected to match the observed data over relevant time horizons, again requiring massive computational effort.

**Key contributions:** The key contribution of this paper is to develop a **shift-scale-restart** algorithm (described later) that carefully exploits the closeness of the underlying infection process (process of number infected of each type at each time), first to a multi-type super critical branching process, and then, suitably normalised, to the mean field limit of the infection process, so that the output from the smaller model accurately matches the output from the larger one. Essentially, both the smaller and the larger model, with identical initial conditions of small number of infections evolve similarly in the early days of the infection growth. Interestingly, in a super-critical branching process, while the number of infections grows exponentially, the proportions of the different types infected quickly stabilises, and this allows us to shift a scaled path from a smaller model to a later time with negligible change in the underlying distribution. Therefore, once there are enough infections in the system, output from the smaller model, when scaled, matches that from the larger model at a later 'shifted' time. This shifting and scaling of the paths from the smaller model does good job of representing output from the larger model when there are no interventions to the system. However, realistically, government intervenes and population mobility behaviour changes with increasing infections. To get the timings of these interventions right, we restart the smaller model and synchronise the timings of interventions in the shifted and scaled path to the actual timings in the original path of the larger model.

We present numerical results where our strategy is implemented on a model for Mumbai with 12.8 million population that realistically captures interactions at home, school, workplace and community as well as mobility restrictions through interventions such as lockdown, home quarantine, case isolation, schools closed, limited attendance at workplaces, etc. Using our approach, we more or less exactly replicate 12.8 million population model for Mumbai with all its complexities, using only 1 million people model, providing an almost 12.8 times speed-up.

Further, we provide theoretical support for the proposed approach through an asymptotic analysis where the population size $N$ increases to infinity. We show that early on till time $\log(N/\log N)/\log(\rho)$, where $\rho$ is the exponential epidemic growth rate in early stages, the epidemic process is well approximated by an associated branching process. Then, after time $\log(\epsilon N)/\log(\rho)$ for any $\epsilon > 0$, the epidemic process is closely approximated by its mean field limit that can be seen to follow a large dimensional discretized ODE. While, our simulation model allows each person's state to lie in an uncountable state space, our theoretical analysis is conducted under somewhat simpler finite state assumptions, that subsumes general compartmental models. As is empirically evident, the results hold more generally. However, the analysis required is more intricate and requires further research. We also identify the behaviour of simple compartmental models such as susceptible-infected and recovered (SIR) and some generalisations in the early branching process phase.

**The remaining paper is structured as follows:** To illustrate the key ideas practically, we first show the implementation of our algorithm on a realistic model of Mumbai in Section 2. Then in Section 3, we briefly summarise our agent based simulator. In Section 4 we spell out the shift-scale-restart algorithm. We provide theoretical asymptotic analysis supporting the efficacy of the proposed algorithm in Section 5. The technical proofs are presented in the Section 6, where for completeness we also spell out the input data used in our simulation model for Mumbai. [1]

## 2 Speeding up ABS: The big picture

A naive approach to speed up the ABS maybe to use a representative smaller population model and scale up the results. Thus, for instance, while a realistic model for Mumbai city may have 12.8 million agents (see [17]), we may construct a sparser Mumbai city having, say, a million agents, that matches the bigger model in essential features, so that, roughly speaking, in the two models each infectious person contributes the same total infection rate to all susceptibles at each time. The output numbers from the smaller model may be scaled by a factor of 12.8 to estimate the output from the larger model. We observe, somewhat remarkably, that this naive approach is actually accurate if the initial seed infections in the smaller model (and hence also the larger model) are large, say, of the order of thousands, and are identically distributed in both the smaller and the larger model (see Figure 1 where the number exposed are plotted under a counter-factual no interventions scenario. The comparative statements hold equally well for other statistics such as the number infected, hospitalised, in ICUs and deceased). The rationale is that in this setting both the smaller and the larger model have sufficient infections so that the proportion of the infected population in both the models well-approximate their identical mean field limits.

However, modelling initial randomness in the disease spread is important for reasons including ascertaining the distribution of when and where an outbreak may be initiated, the probability that some of the initial infection clusters die-down, getting an accurate distribution of geographical spread of infection over time, capturing the intensity of any sample path (the random variable $W$ in the associated branching process, described in Theorem 5.2), etc. These are typically captured by setting the initial infections to a small number, say, around a hundred, and the model is initiated at a well chosen time (see [1]). In such settings, we observe that the scaled output from the smaller model (with proportionately lesser initial infections) is noisy and biased so that the simple scaling fix no longer works (see Figure 2. Later in Section 6.2, we explain in a simple setting why the scaled smaller model is biased and reports lower number of infections compared to the larger model in the early infection spread phase). In fact, we observe that in the early days of the infection, the smaller and the larger model with the **same number of initial infections**, similarly clustered, behave more or less identically (see Figure 3), so that the smaller model with the unscaled number of initial seed infections provides an accurate approximation to the larger one. Here again in the early phase in the two models, each infectious person contributes roughly the same total infection rate to susceptibles at home, workplace and community. Probabilistically this is true because early on, both the models closely approximate an identical multi-type branching process. Shift-scale-restart algorithm outlined in Section 4 exploits these observations to speed up the simulator. We briefly describe it below.

**Fixing ideas:** Suppose that for Mumbai with an estimated population of 12.8 million, a 12.8 million agent model is seeded with 100 randomly distributed infections on day zero. To get the statistics of

[1] A three page extended abstract of this paper appeared in [21].

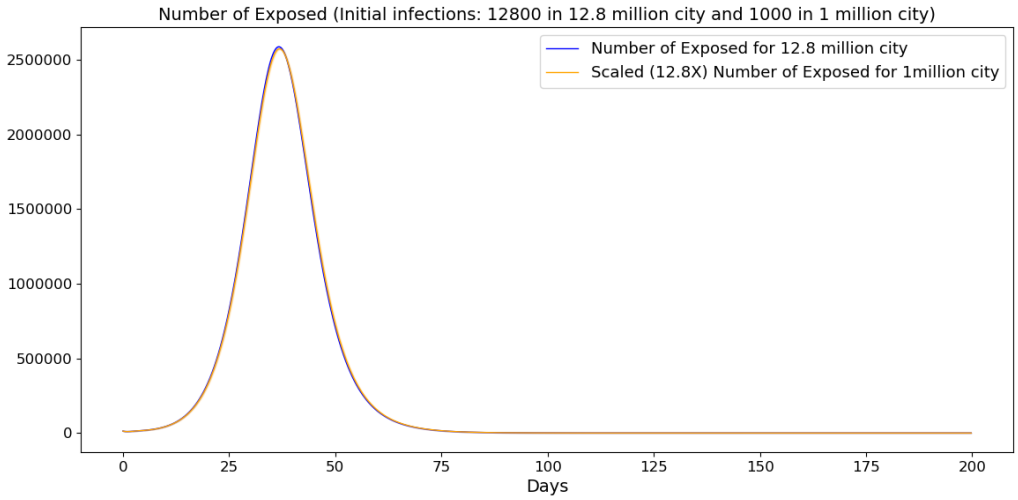


Figure 1: Scaled no. exposed in the smaller model match the larger model when we start with large, 12800, no. of infections.

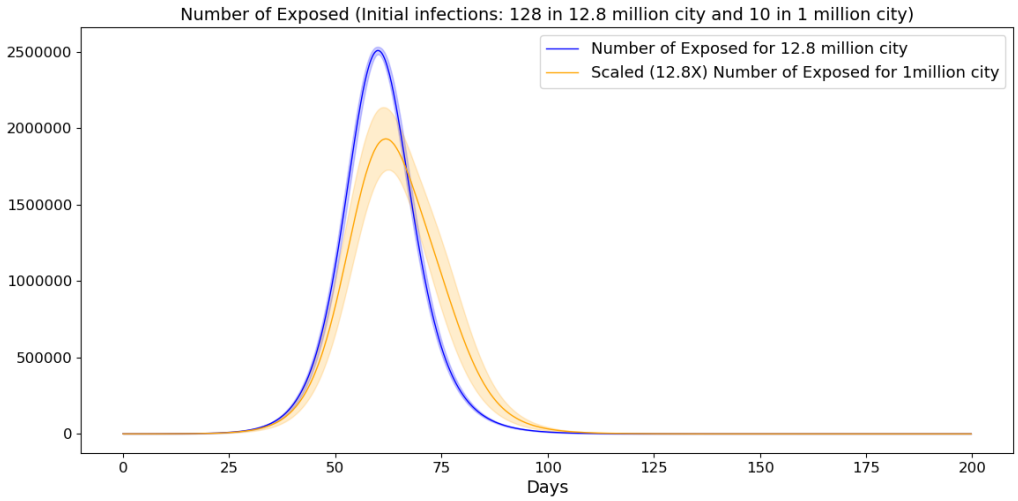


Figure 2: Scaled no. exposed in smaller model do not match the larger model when we start with few, 128, no. of infections.

interest such as expected hospitalisations and fatalities over time, instead of running the 12.8 million agent model, we start a 1 million agent model seeded with 100 similarly randomly distributed infections at day zero and generate a complete path for the requisite duration. To get the statistics for the larger model, we first observe that under the no-intervention scenario, the output of the smaller model more or less exactly matches that of the larger model for around first 35 days (Figure 4). As our analysis in Section 5 suggests, the two models closely approximate the associated branching process till time $\frac{\log(N/(i\log N))}{\log \rho}$ where $N$ denotes the population of the smaller model and equals 1 million, $i$ denotes the number exposed at time zero and equals 100, and $\rho$ denotes the exponential growth rate in the early fatalities and is estimated from fatality data to equal 1.21. The quantity $\frac{\log(N/i\log N)}{\log \rho}$ is estimated to equal 34.5 suggesting that both are close to the underlying branching process around day 35. After this initial period of around 35 days, the city has an average of $x$ thousand infections. We then determine the day when the city had $\frac{x}{12.8}$ thousand infections. This turns out to be day 21.5 in our example. We take the path from day 21.5 onwards, scale it by a factor of 12.8, and concatenate it to the original path starting at day 35. Theoretical justification for this time shift comes from the branching process theory, where while a super-critical multi-type branching process can be seen to grow exponentially with a sample path dependent intensity, the relative proportions amongst types along each sample path stabilise fairly quickly and become more-or-less stationary (see Theorem 5.1). This shifted and scaled output after 35 days matches that of the larger model remarkably well. See Figure 4 where the generated infections from the larger 12.8 million model and the shifted and scaled smaller 1 million model are compared. The choice of day 35 is not critical above. Similar results would be achieved if we used lesser, as low as 25 days in the original model.

In a realistic setting, administration may intervene once the reported cases begin to grow. Suppose in the above example, an intervention happens on day 40. (This is not unreasonable, as in modelling Mumbai, our calibration exercise had set the day zero to February 13, 2020 (see, [12]), the resulting infections and reported cases reached worrying levels around the second week of March. Restrictions in the city were imposed around March 20, 2020.) In that case, the shifted and scaled path from the day 21.5 would need to have the restrictions imposed on day 26.5 (so that it approximates day 40 for the larger model). We achieve this by using the first generated a path till day 35, computing the appropriate scaled path time (21.5 days, in this case), and then, using common random numbers, restarting an identical path from time zero that has restrictions imposed from day 26.5. This path is scaled from day 21.5 onwards and concatenated to the original path at day 35. Later in Section 5.5 we note that the restarted path need not use common random numbers from the original path. Even if it is generated independently, we get very similar output.

To summarise, our shift-scale approach works because, early on when the small model is close to the branching process, shifting across time is valid as the proportions across types do not distort significantly. Thereafter, the scaling property holds (smaller model, suitably scaled, well approximates the larger model) even when the number infected increase to become of the same order as the susceptible, since both the smaller and the larger model closely approximate their common mean field limit evolution process (see Theorem 5.4 for a theoretical justification).

Figures 4, 5 and 6 compare the number exposed, the number hospitalised and the number of cumulative fatalities in a 12.8 million Mumbai city simulation (in no intervention scenario) with the estimates from the shift-scale-restart algorithm applied to the smaller 1 million city. Figures 1 to 6 are under no intervention scenario for Mumbai as described in Section 6.3. Figures 7, 8 and 9 compare the number exposed, the number hospitalised, and the number of cumulative fatalities in a 12.8 million Mumbai city simulation (intervention: home quarantine from day 40) with estimates from the shift-scale-restart algorithm applied to the smaller 1 million city. Figure 10 compares the exposed population process for the 12.8 million population Mumbai model with the smaller 1 million one as per our algorithm under realistic interventions (lockdown, case isolation, home quarantine, masking etc.) introduced at realistic times, as implemented in [13] using similar parameters, for 250 days. We see that the smaller model faithfully replicates the larger one with negligible error.

# 3 Agent Based Simulator

In this section, we informally describe the main drivers of the dynamics in our infection spread model. A more detailed discussion can be seen in [1].

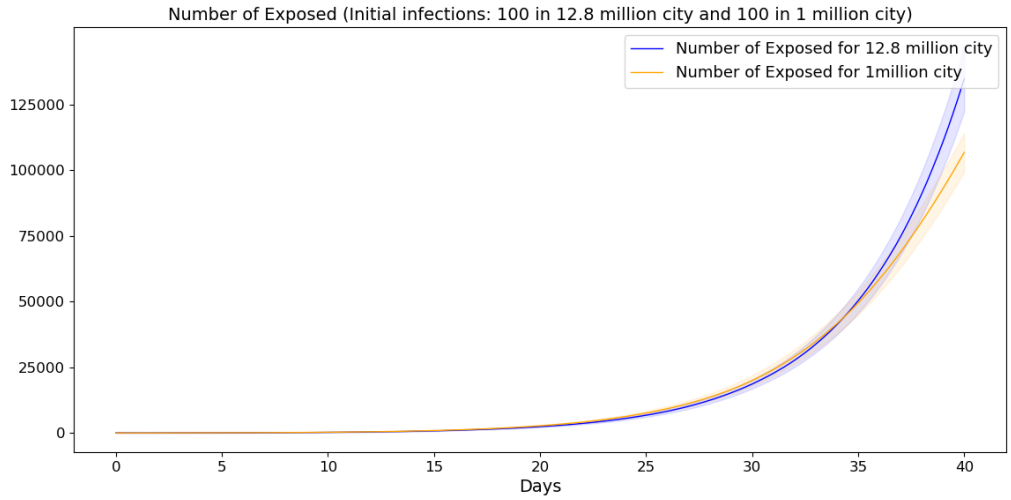


Figure 3: Smaller and larger model are essentially identical initially when we start with same number of few, 100, no. of infections

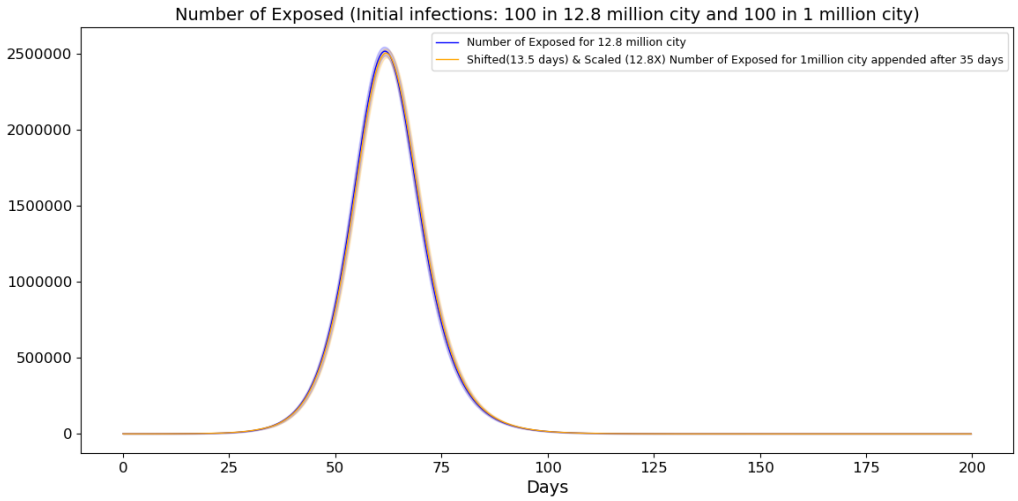


Figure 4: Shift and scale smaller model (no. of exposed) matches the larger model under no intervention scenario.

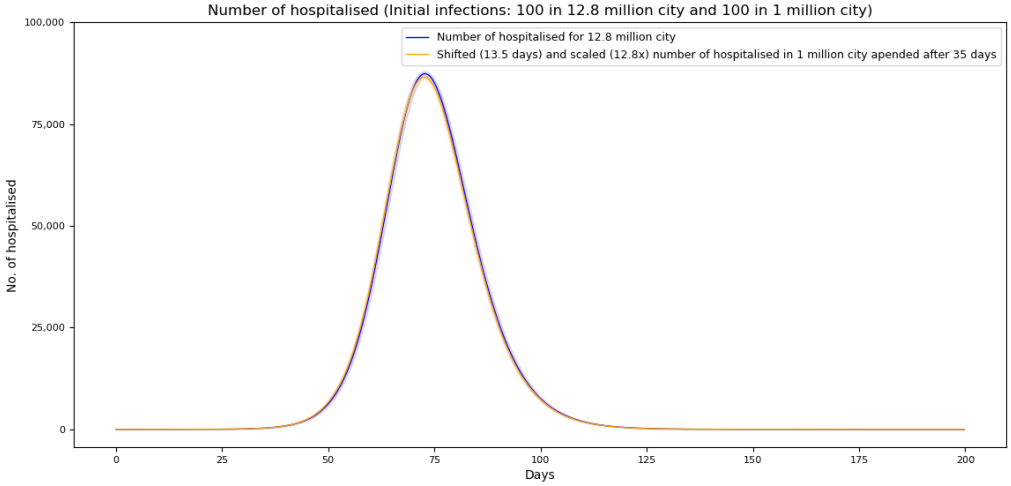


Figure 5: Shift and scale smaller model (no. of hospitalised) matches the larger model under no intervention scenario.

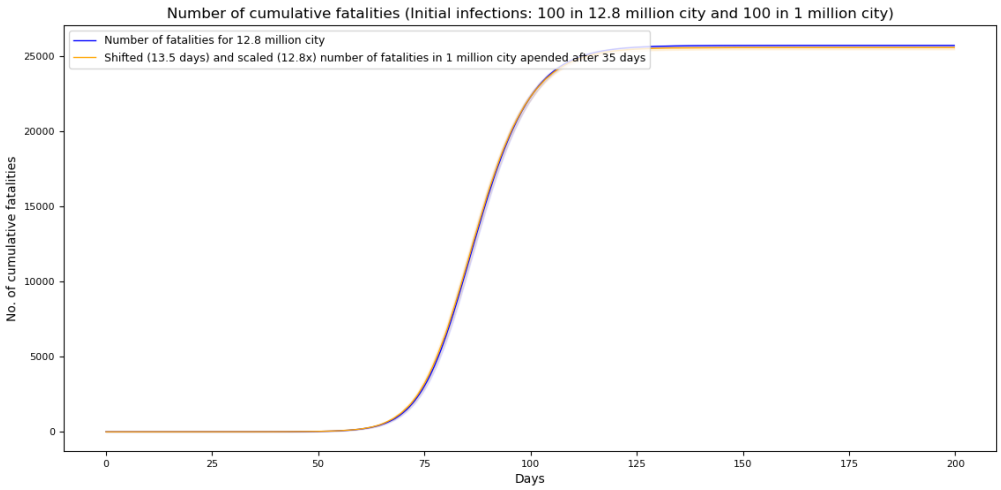


Figure 6: Shift and scale smaller model (no. of cumulative fatalities) matches the larger model under no intervention scenario.

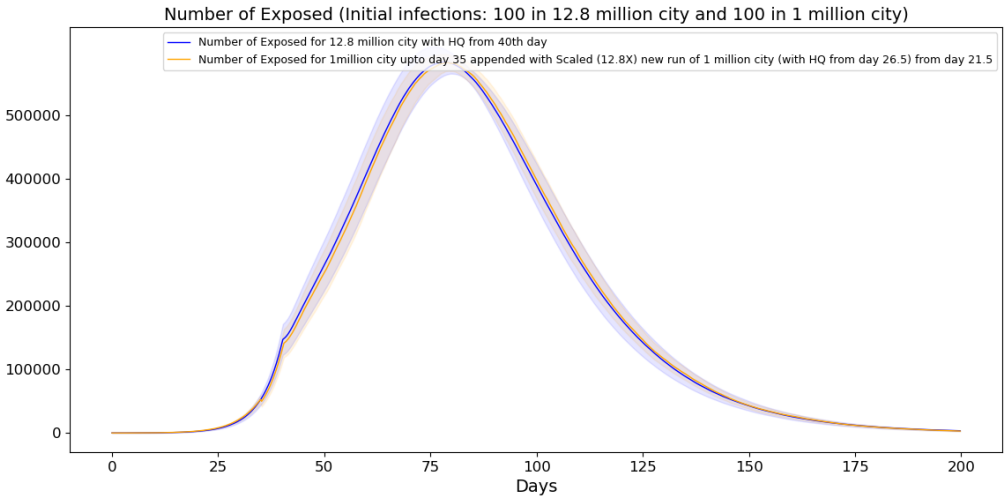


Figure 7: Shift-scale-restart smaller model (no. of exposed) matches the larger model (intervention: home quarantine from day 40).

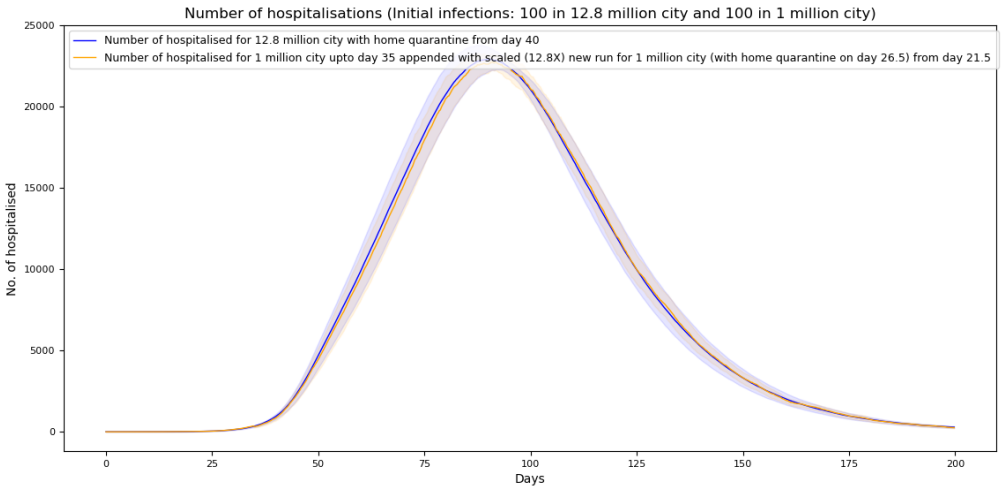


Figure 8: Shift-scale-restart smaller model (no. of hospitalised) matches the larger model (intervention: home quarantine from day 40).

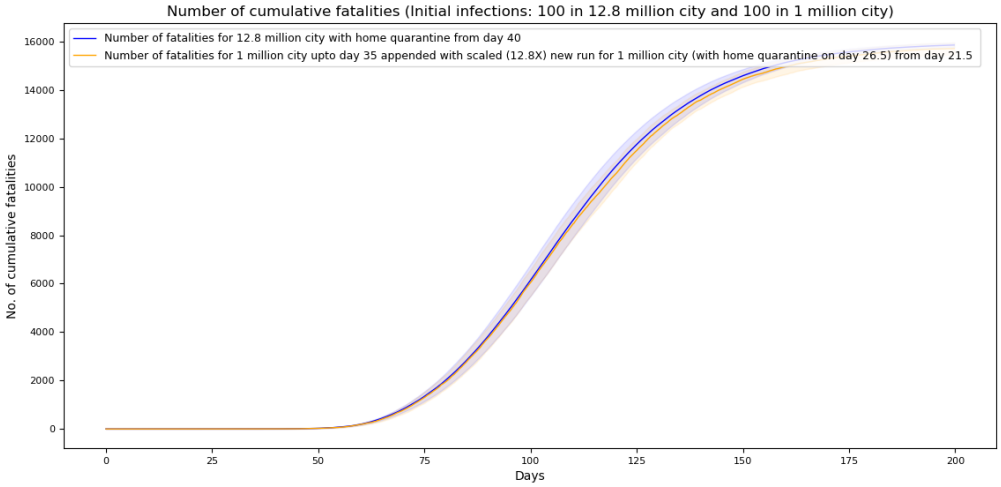


Figure 9: Shift-scale-restart smaller model (no. of cumulative fatalities) matches the larger model (intervention: home quarantine from day 40).

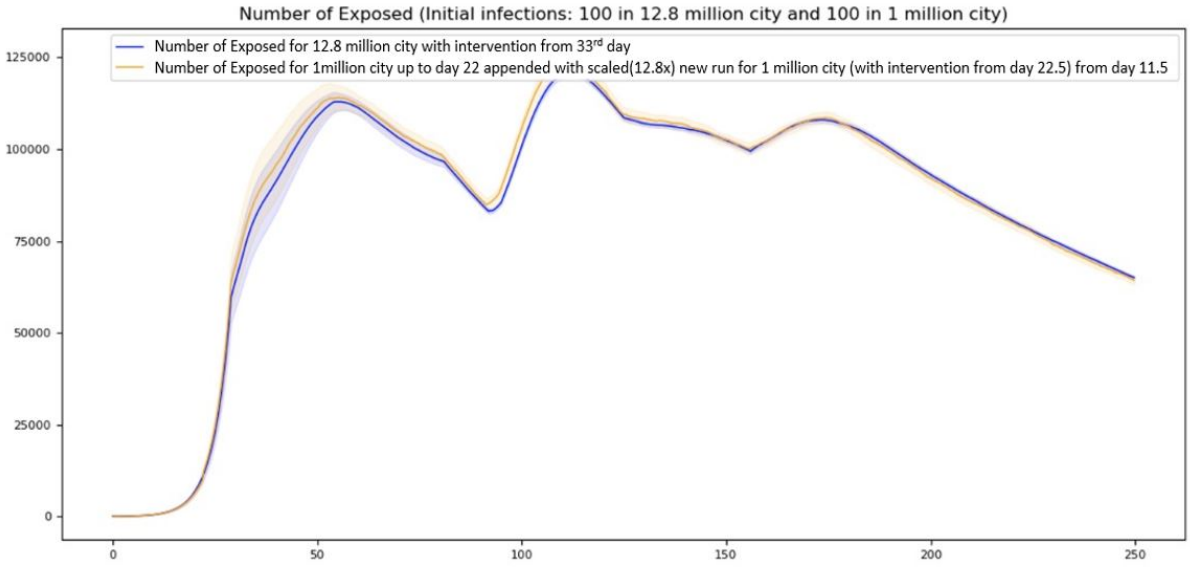


Figure 10: Shift-scale-restart smaller model match the larger one under real world interventions over 250 days.

The model consists of individuals and various interaction spaces such as households, schools, workplaces and community spaces. Infected individuals interact with susceptible individuals in these interaction spaces. The number of individuals living in a household, their age, whether they go to school or work or neither, schools and workplaces size and composition all have distributions that may be set to match the available data. The model proceeds in discrete time steps of constant width $\Delta t$ (six hours in our set-up). At a well chosen time zero, a small number of individuals can be set to either exposed, asymptomatic, or symptomatic states, to seed the infection. At each time $t$, an infection rate $\lambda_n(t)$ is computed for each susceptible individual $n$ based on its interactions with other infected individuals in different interaction spaces (households, schools, workplaces and community). In the next $\Delta t$ time, each susceptible individual moves to the exposed state with probability $1 - \exp\{-\lambda_n(t) \cdot \Delta t\}$, independently of all other events. Further, disease may progress independently in the interval $\Delta t$ for the population already afflicted by the virus. The computation of $\lambda_n(t)$ and the probabilistic dynamics of disease progression are briefly summarized below under simplified assumptions (details available in [1], also see [7]). Simulation time is then incremented to $t + \Delta t$, and the state of each individual is updated to reflect the new exposures, changes to infectiousness, hospitalisations, recoveries, quarantines, etc., during the period $t$ to $t + \Delta t$. The overall process repeats incrementally until the end of the simulation time.

**Computing $\lambda_n(t)$ :** A susceptible individual $n$ at any time $t$ receives a total infection rate $\lambda_n(t)$ which is sum of the infection rates $\lambda_n^h(t)$ (from home), $\lambda_n^s(t)$ (from school), $\lambda_n^w(t)$ (from workplace) and $\lambda_n^c(t)$ (from community) coming in from infected individuals in respective interaction spaces of individual $n$.

$$\lambda_n(t) = \lambda_n^h(t) + \lambda_n^s(t) + \lambda_n^w(t) + \lambda_n^c(t).$$

Briefly, the transmission rate ($\beta$) of virus by an infected individual in each interaction space is the expected number of eventful (infection spreading) contact opportunities with all the individuals in that interaction space. It accounts for the combined effect of frequency of meetings and the probability of infection spread during each meeting. An infected individual can transmit the virus in the infective (pre-symptomatic or asymptomatic stage) or in the symptomatic stage. Each individual has two other parameters: a severity variable (individual attendance in school and workplace depends on the severity of disease) and a relative infectiousness variable, virus transmission is related linearly to this. Both bring in heterogeneity to the model.

Specifically, let $e_{n'}(t) = 1$ when an individual $n'$ can transmit the virus at time $t$, $e_{n'}(t) = 0$ otherwise. To keep notation simple, we avoid describing the details of individual infectiousness, severity, age dependent mobility and community density factor (see [1] for details). Let $\beta_h$, $\beta_s$, $\beta_w$, and $\beta_c$ denote the transmission coefficients at home, school, workplace and community spaces, respectively. A susceptible individual $n$ who belongs to home $h(n)$, school $s(n)$, workplace $w(n)$ and community space $c(n)$ sees the following infection rates at time $t$ in different interaction spaces:

$\lambda_n^h(t)$ is the average transmission rate coming in to individual $n$ from each infected individual in his home.

$$\lambda_n^h(t) = \beta_h \frac{\sum_{n':h(n')=h(n)} e_{n'}(t)}{n_{h(n)}}.$$

$\lambda_n^s(t)$ and $\lambda_n^w(t)$ are also computed similarly based on infected individuals in respective interaction spaces of the individual $n$.

Different wards in the city constitute different communities. For Mumbai, since slums form more than half the population and are extremely dense, we further divide each ward into slum and non-slum community. To compute infection rate to an individual from the community, we first compute the infection rate $h_{c'}(t)$ for each community $c'$ at a given time $t$. This is set to the sum of transmission rate from all the infected individuals of the community assigned a weight that is proportional to the strictly decreasing function of the distance between the individual and the community centre. Specifically,

$$h_{c'}(t) = \beta_c \left( \frac{\sum_{n':c(n')=c'} f(d_{n',c'}) \cdot e_{n'}(t)}{\sum_{n'} f(d_{n',c'})} \right) \tag{1}$$

where $f(d)$ is a distance kernel that is strictly decreasing in $d$. As in [6], one may select $f(d) = 1/(1 + (d/a)^b)$ and $d \ll a$ for appropriately chosen non-negative constants $a$ and $b$. Note that $h_{c'}(t)$ is dominated by $\beta_c$ and it equals $\beta_c$ only when everyone in the community $c'$ is infected.

The infection rate $\lambda_n^c(t)$ seen by an individual $n$ in community $c$ is set to sum of the community infection rates from different communities of the city multiplied with weights that are again proportional to the strictly decreasing function of the distance between the two communities. This quantity is further adjusted for the distance between individual $n$ and its community centre. Specifically,

$$\lambda_n^c(t) = \frac{f(d_{n,c(n)})}{\sum_{c'} f(d_{c(n),c'})} \sum_{c'} f(d_{c(n),c'}) h_{c'}(t). \tag{2}$$

Again observe that, if $f(d_{n,c(n)}) < 1$, then $\lambda_n^c(t) < \beta_c$. Further, an age dependent factor determining the mobility of the individuals in the community may also be accounted in $\lambda_n^c(t)$ (see [1]). Containment zones were an important government intervention where as the cases in a region rise, they are contained through localised movement restrictions. We omit the associated modelling details (see [14]), although they are captured in our experiments. In [1] an implementation of contact tracing methodology is discussed. However, this has minimal effect on disease spread, and is not considered in the code we use for experiments in this paper.

**Disease progression :** Our model of COVID-19 progression is based on descriptions in [22] and [7]. An individual may have one of the following states: susceptible, exposed, infective (pre-symptomatic or asymptomatic), recovered, symptomatic, hospitalised, critical, or deceased. Briefly, an individual after getting exposed to the virus at some time observes an incubation period which is random with a Gamma distribution. Individuals are infectious for an exponentially distributed period. This covers both presymptomatic transmission and possible asymptomatic transmission. We assume that a third of the patients recover, these are the asymptomatic patients; the remaining develop symptoms. Individuals either recover or move to the hospital after a random duration that is exponentially distributed (see Section 6.3 for model input data details). The probability that an individual recovers depends on the individual's age. While hospitalised individuals may continue to be infectious, they are assumed to be sufficiently isolated, and hence do not further contribute to the spread of the infection. Further progression of hospitalised individuals to critical care and further to fatality is also age dependent.

**Public health safety measures (PHSMs):** We introduce methodologies to model different PHSMs (or Interventions) such as lockdowns, home quarantine, case isolation, social distancing of elderly population, mobility restrictions, masks etc. See [1] for details on how these interventions are modelled in the simulator. Table 6 in Section 6.3 summarizes some of the key interventions implemented. The PHSM's mentioned above when implemented put some restrictions on the individual's mobility. However, it is often the case that when several restrictions are in place, only a fraction of the population comply with these restrictions. Therefore, we restrict the movement of only the individuals in the compliant fraction of households in the city.

# 4 Shift-scale-restart algorithm

Let $\mu_0(N)$ denote the initial distribution of the infected population at time zero in our simulation model with population $N$ and let the simulation run for a total of $T$ time units. E.g., for Mumbai at a suitably chosen time 0, we select 100 people at random from the non-slum population and mark them as exposed. (Since initially the infection came from international travellers flying into Mumbai, it is reasonable to assume that most of them were residing in non-slums). The simulation dynamics may be summarised in Algorithm 1.

For $k, N \in \mathbb{N}$, $k > 1$, let $kN$ be the number of individuals in the larger city, and $N$ in the smaller city. Roughly speaking, the larger city has $k$ times more homes, schools and workplaces compared to the smaller city. The joint distribution of people in homes, schools or workplaces is unchanged, and transmission rates $\beta_h$, $\beta_s$, $\beta_w$ and $\beta_c$ are unchanged.

When we initiate both the larger as well as the smaller city with a same few and well spread infections, the disease spread similarly in homes, schools and workplaces. To understand the disease spread through communities, for simplicity assume that there is a single community and $f$ equals 1 in (1) and (2). It is easy to see that each susceptible person sees approximately $1/k$ times the community infection rate in the larger city compared to the smaller city. On the other hand, the larger city has $k$ times more

**Algorithm 1** Simulation Dynamics

1: At $t = 0$, start the simulation with $I_0$ infections distributed as per $\mu_0(N)$.
2: **while** $t < T$ **do**
3: For each susceptible individual $n$, calculate $\lambda_n(t)$. Its status then changes to exposed with probability $1 - \exp(-\lambda_n(t))$.
4: All individuals in some state other than susceptible, independently transition to another state as per the disease progression dynamics.
5: $t \leftarrow t + 1$.
6: The above simulation is independently repeated many times and average of performance measures such as number exposed, number infected, number hospitalised, and number of fatalities as a function of time are reported.

susceptible population. This is true even when there are more wards and for a general distance function $f$. Therefore, early on in the simulation, the total number of people getting infected through communities is also essentially identical between the larger and the smaller city, and the infection process in the two cities evolves very similarly.

Let $t_S$ denote the time till the two cities evolve essentially identically (as seen empirically and suggested by theoretical analysis, this is close to $\log_\rho N^*$ for $N^* = N/\log N$. Here $\log_\rho m = \log m/\log \rho$ for any $m \in \Re^+$ and $\rho$ denotes the initial infection exponential growth rate.)

First consider the contrafactual no intervention scenario where the population intermingling behaviour does not change even as the disease spreads through the population. In this setting we propose that a shift and scale Algorithm 2, that builds upon Algorithm 1, be applied to the smaller city to generate output that resembles the larger city. Algorithm 2 is graphically illustrated in Figure 11.

**Algorithm 2** Shift and scale algorithm in no-intervention setting

1: At $t = 0$, start the simulation with $I_0$ infections distributed as per $\mu_0(N)$. Generate the simulation sample path $[y_1, y_2, ..., y_T]$ where $y_t$ denotes the statistics of the affected population (e.g., number exposed, number hospitalised, number of fatalities) at time $t$.

2: Suppose there are $x$ infections at $t_S$, determine an earlier time $t_{x/k}$ in the simulation when there where approximately $\frac{x}{k}$ infections in the city.

3: The statistics of affected for the larger city is then obtained as

$$[y_1, y_2, ..., y_{t_S}, k \times y_{t_{x/k}+1}, ..., k \times y_{T-(t_S - t_{x/k})}].$$

4: The above simulation is independently repeated many times and average of performance measures such as number exposed, number infected, number hospitalised, and number of fatalities as a function of time are reported.

In a realistic scenario, as the infection spreads, the administration will intervene and impose mobility restrictions. Thus, our simulation adjustments to the small city need to account for the timings of these interventions accurately. Let $t_I$ denote the first intervention time (e.g., lockdown; typically after $\beta \log_\rho N$ time for small $\beta \in (0, 1)$). Let $t_{min} \approx \min\{t_I, t_S\}$. We need to restart our simulation to ensure that the shift and scaled path incorporates the intervention at the correct time. The Algorithm 3 achieves this and is graphically illustrated in Figure 12.

As we see empirically, and as is suggested by Proposition 5.1, in Algorithms 2 and 3, evolution after time $t_{x/k}$, the infection process is more or less deterministic.

**Algorithm 3** Shift, scale and restart algorithm

1: At $t = 0$, start the simulation with $I_0$ infections distributed as per $\mu_0(N)$. Generate the simulation sample path $[y_1, y_2, ..., y_{t_{min}}]$ where $y_t$ denotes the statistics of the affected population (e.g., number exposed, number hospitalised, number of fatalities) at time $t$.

2: Suppose there are $x$ infections at $t_{min}$, determine an earlier time $t_{x/k}$ in the simulation when there where approximately $\frac{x}{k}$ infections in the city.

3: Restart a new simulation of the city using common random numbers, but with the intervention introduced at time $t_{\frac{x}{k}} + t_I - t_{min}$. Simulate it upto time $T - (t_{min} - t_{x/k})$. Denote the time series of statistics of the affected population in the restart simulation by

$$z_1, z_2, ..., z_{t_{x/k}}, ..., z_{T-(t_{min}-t_{x/k})}.$$

4: The approximate statistics of the affected population for the larger city is then obtained as

$$[y_1, y_2, ..., y_{t_{min}}, k \times z_{t_{x/k}+1}, ..., k \times z_{T+t_{x/k}-t_{min}}].$$

5: The above simulation is independently repeated many times and average of performance measures such as number exposed, number infected, number hospitalised, and number of fatalities as a function of time are reported.

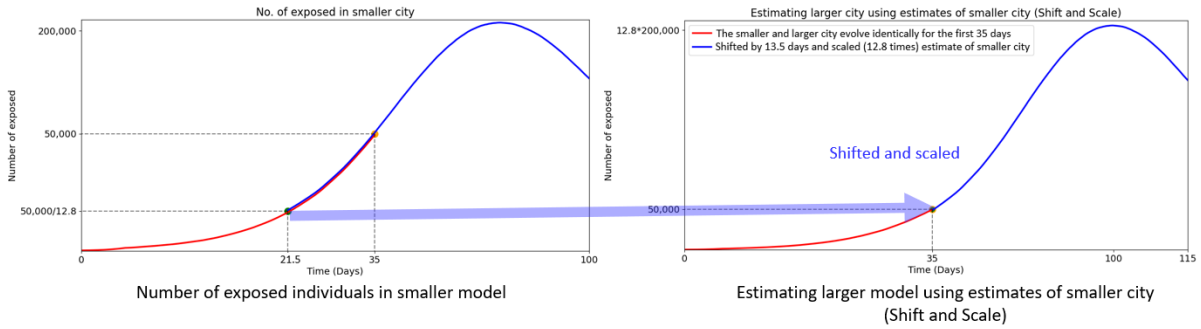


Figure 11: Shift and scale under no intervention scenario

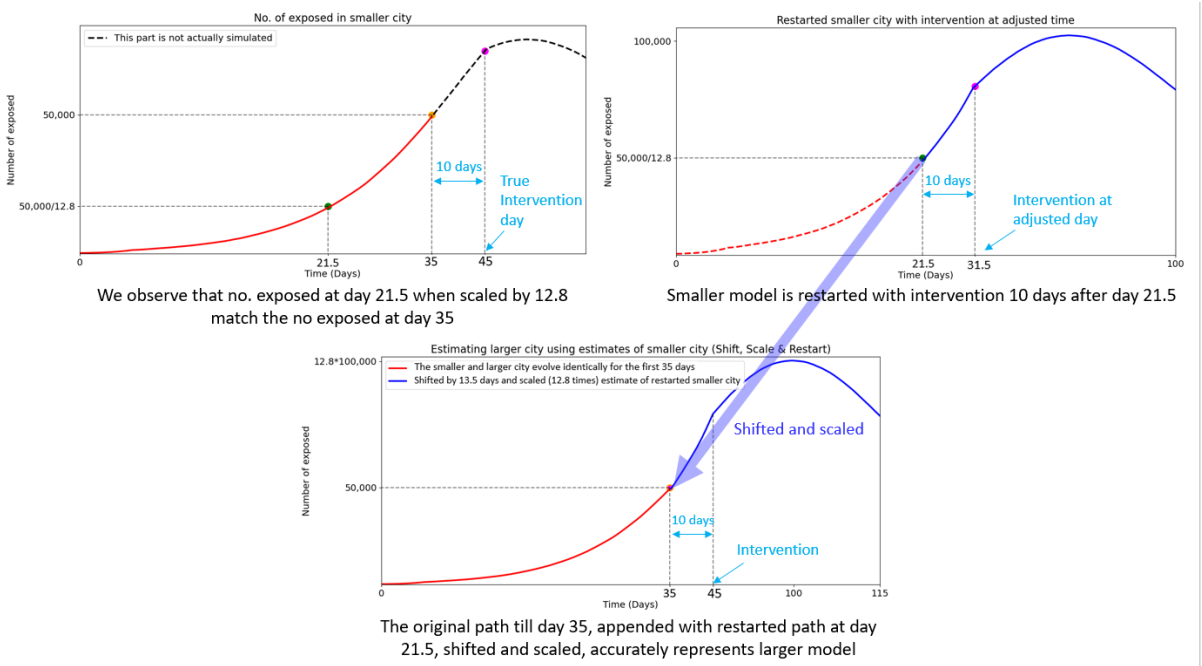


Figure 12: Shift, scale and restart algorithm

# 5 Asymptotic Analysis of Epidemic Process

In the SSR algorithm, theoretical justification is needed for the fact that early on in the small city simulation, we could take a path at one time period, scale it, and stitch it to the path at another appropriately chosen time period to accurately generate a path for the larger city. We provide this through analyzing our city in an asymptotic regime as the city population $N$ increases to infinity. To bring out the key observations simply, we consider a simpler model where the interaction spaces of homes, workplaces and schools are ignored and only a single community interaction space is retained.

In the following analysis, we first review standard multi-type branching process in Section 5.1. We then define an epidemic process in Section 5.2. Further, we define a multi-type super-critical branching process tailored to the epidemic process in Section 5.3. We describe the coupling between this and the epidemic process in Section 5.4. We then state the results demonstrating the closeness of the epidemic and the branching process in the early disease spread phase in Section 5.5. The results demonstrating the closeness of the epidemic process to its mean field limit once the epidemic process has grown are given in in Section 5.6. The technical proofs are given separately in the Appendix.

## 5.1 Supercritical multi-type branching process review

In this section, we first define a multi-type branching process and state a key result associated with a super critical multi-type branching process and an assumption outlining a set of sufficient conditions for it to hold. See [2] for details.

Let $\tilde{\boldsymbol{B}}_t$ be a $\tilde{\eta}$ dimension vector denoting the multi-type branching process at time $t$, where $\tilde{\eta} < \infty$. Component $i$ of $\tilde{\boldsymbol{B}}_t$ denotes the number of individuals of type $i$ at time $t$. As is well known, in a multi-type branching process, at the end of each time period an individual may give birth to children of different types and itself dies (it may reincarnate as a child of same or different type). The number of children each individual of type $i$ gives birth to is independent and identically distributed. Therefore, multi-type branching process is a Markov chain $(\tilde{\boldsymbol{B}}_t \in \mathbb{R}^{+\tilde{\eta}} : t \geq 0)$. Suppose $\tilde{\boldsymbol{B}}_t = (b_1, \ldots, b_{\tilde{\eta}})$, then $\tilde{\boldsymbol{B}}_{t+1}$ is sum of independent offsprings of $b_1$ type 1 parents, independent offsprings $b_2$ type 2 parents, and so on. Thus, $\tilde{\boldsymbol{B}}_{t+1}$ is sum of $b_1 + b_2 + ... + b_{\tilde{\eta}}$ independent random vectors in $\mathbb{R}^{+\tilde{\eta}}$.

Consider a matrix $\tilde{K} \in \Re^{+\tilde{\eta}\times\tilde{\eta}}$ such that $\tilde{K}(i,j) = \mathbb{E}\left(\tilde{B}_1(j)|\tilde{\boldsymbol{B}}_0 = \boldsymbol{e_i}\right)$, that is expected number of

type $j$ offsprings of a single type $i$ individual in one time period. Then,

$$\mathbb{E}\left(\tilde{\boldsymbol{B}}_{t+1}\right) = \tilde{K}^T \mathbb{E}\left(\tilde{\boldsymbol{B}}_t\right) = (\tilde{K}^T)^{t+1} \mathbb{E}\left(\tilde{\boldsymbol{B}}_0\right).$$

Recall that the Perron Frobenius eigenvalue of a non-negative irreducible matrix (a non-negative square matrix $A$ is irreducible if there exists an integer $m > 0$ such that all entries of $A^m$ are strictly positive) is its largest eigenvalue in absolute value, and can be seen to be positive. The following assumption is standard in multi-type branching process theory (see [2]).

**Assumption 1.** *$\tilde{K}$ is irreducible and its Perron Frobenius eigenvalue $\tilde{\rho} > 1$. Furthermore,*

$$\mathbb{E}\left(\tilde{B}_1(j) \log \tilde{B}_1(j) | \tilde{\boldsymbol{B}}_0 = \boldsymbol{e_i}\right) < \infty,$$

*for all $1 \leq i, j \leq \tilde{\eta}$.*

By $X_n \xrightarrow{P} X$ we denote that the sequence of random variables $\{X_n\}$ converges to $X$ in probability as $n \to \infty$. Theorem 5.1 below is well known (see [2]).

Suppose that Assumption 1 holds. Then, corresponding to eigenvalue $\tilde{\rho}$, there exist strictly positive right and left eigenvectors of $\tilde{K}$, $\tilde{\boldsymbol{u}}$ and $\tilde{\boldsymbol{v}}$ such that $\tilde{\boldsymbol{u}}^T \tilde{\boldsymbol{v}} = 1$ and $\sum_{i=1}^{\tilde{\eta}} \tilde{u}(i) = 1$.

**Theorem 5.1.** *Under Assumption 1,*

$$\lim_{t \to \infty} \frac{\tilde{K}^t}{\tilde{\rho}^t} = \tilde{\boldsymbol{u}} \tilde{\boldsymbol{v}}^T.$$

*Furthermore,*

$$\frac{\tilde{\boldsymbol{B}}_t}{\tilde{\rho}^t} \xrightarrow{P} W \tilde{\boldsymbol{v}} \quad \text{as } t \to \infty, \tag{3}$$

*where $W$ is a non-negative random variable such that $P\{W > 0\} > 0$ and $\mathbb{E}\left(W | \tilde{\boldsymbol{B}}_0 = \boldsymbol{e_i}\right) = \tilde{u}(i)$ for all $i = 1, ..., \eta$. Let $A = \{\omega : \tilde{B}_t(\omega) \to \infty \text{ as } t \to \infty\}$. Then, for any $\epsilon > 0$ and for all $j \in [1, \tilde{\eta}]$*

$$\lim_{t \to \infty} P\{\omega : \omega \in A, \left| \frac{\tilde{B}_t(j)}{\sum_{i=1}^{\tilde{\eta}} \tilde{B}_t(i)} - \frac{\tilde{v}(j)}{\sum_{i=1}^{\tilde{\eta}} \tilde{v}(i)} \right| > \epsilon\} = 0. \tag{4}$$

## 5.2 Epidemic process dynamics

Some notation is needed to help specify the dynamics of the epidemic process.

- The city comprises of $N$ individuals and our interest is in analyzing the city asymptotically as $N \to \infty$.
- An individual at any time can be in one of the following disease states: Susceptible, exposed, infective, symptomatic, hospitalised, critical, dead and recovered. Individuals are infectious only in infective or symptomatic states. Denote all the disease states by $\mathcal{D}$. To keep the discussion simple, we ignore the possibility of an individual getting reinfected, although incorporating them in a realistic way would not alter our conclusions.
- Each individual has some characteristics that are assumed to remain unchanged throughout the epidemic regardless of individual's disease state. These include individual's age group, disease progression profile (e.g., some may be more infectious than others), community transmission rates (e.g., individuals living in congested slums may be modelled to have higher transmission rates), mobility in the community (e.g., elder population may travel less to the community compared to the working age population). We assume that set of all possible individual characteristics are finite, and denote them by $\mathcal{A}$. Let $N_a$, for $a \in \mathcal{A}$, denote the total number of individuals with characteristic $a$ in system $N$, and set $\pi_a = \frac{N_a}{N}$ for all $a \in \mathcal{A}$. We assume that $\pi_a$ is independent of $N$ as $N \to \infty$.

- Hence, each individual at any time may be classified by a type $\boldsymbol{s} = (a, d)$, where $a$ denotes the individual characteristic and $d$ the disease state. Let $\mathcal{S} = \mathcal{A} \times \mathcal{D}$ denote the set of all types.
- Let $\mathcal{U} \subset \mathcal{S}$ denote all the *types* with susceptible disease state. Hence, $\mathcal{S} \setminus \mathcal{U}$ denote the types where individuals are already affected (that is, they have been exposed to the disease at some point in the past). Let $\eta = |\mathcal{S} \setminus \mathcal{U}|$.
- Denote the number of individuals of type $\boldsymbol{s}$ at time $t$ by $X_t^N(\boldsymbol{s})$ and set $\boldsymbol{X}_t^N = (X_t^N(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U})$. Then, $\boldsymbol{X}_t^N \in \mathbb{Z}^{+\eta}$.
- Let $A_t^N = \sum_{\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}} X_t^N(\boldsymbol{s})$ denote the total number of affected individuals in the system $N$ at or before time $t$.

**Dynamics:** At time zero, for each $N$, $\boldsymbol{X}_0^N$ is initialised by setting a suitably selected small and fixed number of people randomly from some distribution $\mu_0(N)$ and assigning them to the exposed state. All others are set as susceptible. The distribution $\mu_0(N)$ is assumed to be independent of $N$ so we can set $\boldsymbol{X}_0^N = \boldsymbol{X}_0$ for all N.

Given $\boldsymbol{X}_t^N$, $\boldsymbol{X}_{t+\Delta t}^N$ is arrived at through two mechanisms. For the ease of notation we will set $\Delta t = 1$. 1) Infectious individuals at time $t$ who make Poisson distributed infectious contacts with the rest of the population, moving the contacted susceptible population to exposed state, and 2) through population already affected moving further along in their disease state. Specifically,

- We assume that an infectious individual with characteristic $a \in \mathcal{A}$ spreads the disease in the community with transmission rate $\beta_a$. Thus, the total number of infectious contacts it makes with all the individuals (both susceptible and affected) in one time step is Poisson distributed with rate $\beta_a$. The individuals contacted are selected randomly and an individual with characteristic $\tilde{a} \in \mathcal{A}$ is selected with probability proportional to $\psi_{a,\tilde{a}}$ (independent of $N$). $\psi_{a,\tilde{a}}$ helps model biases such as an individual living in a dense region is more likely to infect another individual living in the same dense region. As a normalisation, set $\sum_{\tilde{a}} \pi_{\tilde{a}} \psi_{a,\tilde{a}} = 1$.
- Once the number of infectious contacts for a particular characteristic $\tilde{a} \in \mathcal{A}$ have been generated, each contact is made with an individual selected uniformly at random from all the individuals with characteristic $\tilde{a}$.
- If an already affected individual has one or more contact with an infectious individual, its type remains unchanged. On the other hand, if a susceptible individual has at least one contact from any infectious individual, it will get exposed at time $t + 1$. Each susceptible individual with characteristic $\tilde{a} \in \mathcal{A}$ who gets exposed at time $t$, increments $\boldsymbol{X}_{t+1}^N(\tilde{a}, \textit{exposed})$ by 1. A susceptible individual that has no contact with an infectious individual remains susceptible in the next time period.
- Once an individual gets exposed, its disease progression is independent of all the other individuals and depends only on its characteristics, that is, the disease progression profile of the characteristic class the individual belongs to. The waiting time in each state (except susceptible, dead and recovered) is assumed to be geometrically distributed. Transition to symptomatic, hospitalised, critical, dead or recovered state happens with respective characteristic (disease progression profile) dependent transition probabilities. Thus, an individual of type $\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}$, some disease state other than susceptible, at time $t$ transitions to some other state $\boldsymbol{q}$ at time $t + 1$ with the transition probability $P(\boldsymbol{s}, \boldsymbol{q})$ in one time step. The probability $P(\boldsymbol{s}, \boldsymbol{q})$ is independent of time $t$ and $N$. If the transition happens, $X_{t+1}^N(\boldsymbol{s})$ is decreased by 1 and $X_{t+1}^N(\boldsymbol{q})$ is increased by 1. Observe that, if characteristic of types $\boldsymbol{s}$ and $\boldsymbol{q}$ is different, then $P(\boldsymbol{s}, \boldsymbol{q}) = 0$.

## 5.3 Associated branching process dynamics

For each $t$, let $\boldsymbol{B}_t \in \mathbb{Z}^{+\eta}$, $\boldsymbol{B}_t = (B_t(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U})$ where $B_t(\boldsymbol{s})$ denote the number of individuals of $\boldsymbol{s}$ at time $t$ in the branching process.

**Dynamics:** At time zero $\boldsymbol{B}_0 = \boldsymbol{X}_0$. Given $\boldsymbol{B}_t$, we arrive at $\boldsymbol{B}_{t+1}$ as follows:

- At time $t$, every infectious individual of type $a$, for all $a$, gives birth to independent Poisson distributed offspring of type $(\tilde{a}, exposed)$ at time $t+1$ with rate $\pi_{\tilde{a}}\psi_{a,\tilde{a}}\beta_a$ for each $\tilde{a}$. $B_{t+1}(\tilde{a}, exposed)$ is increased accordingly.
- Once an individual gets exposed, disease progression of the individual has same transition probabilities as in each epidemic process. An individual of type $\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}$, that is, in disease state other than susceptible at time $t$, transitions to some other disease state $\boldsymbol{q}$ at time $t+1$ with probability $P(\boldsymbol{s}, \boldsymbol{q})$. If the transition happens, $B_{t+1}(\boldsymbol{s})$ is decreased by 1 and $B_{t+1}(\boldsymbol{q})$ is increased by 1.

Let $A_t^B = \sum_{\boldsymbol{s}\in\mathcal{S}\setminus\mathcal{U}} B_t(\boldsymbol{s})$ denote the total number of offsprings generated by time $t$ in the branching process.

**Specifying the expected offsprings matrix $K \in \Re^{+\eta\times\eta}$ for $\{\boldsymbol{B}_t\}$:** Let each entry $K(\boldsymbol{s}, \boldsymbol{q})$ of $K$ denote the expected number of type $\boldsymbol{q}$ offspring of a single type $\boldsymbol{s}$ individual in one time step. Let $\mathcal{H} \subset \mathcal{S}$ denote all the types with the disease states that are infectious or may become infectious in subsequent time steps (that is, types with disease state either exposed, infective or symptomatic). Let $\mathcal{H}^c$ denote its complement. Individuals in $\mathcal{H}^c$ do not contribute to community infection. Let $\tilde{\mathcal{H}} \subset \mathcal{S}$ denote all the types with the disease states that are infectious (that is infective or symptomatic state). Let $\mathcal{E}$ denote $\mathcal{H} \setminus \tilde{\mathcal{H}}$, that is, the set of all the types corresponding to exposed individuals. Let $\hat{\eta} = |\mathcal{H}|$.

As described above, an individual of type $\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}$, may give birth to other exposed individuals if it is infectious, and/or may itself transition to some other type in one time step. Then, $K$ can be written as,

$$\begin{aligned} &K(\boldsymbol{s}, \boldsymbol{q}) = P(\boldsymbol{s}, \boldsymbol{q}) + \pi_{\tilde{a}}\psi_{a,\tilde{a}}\beta_a \text{ for all } \boldsymbol{s} = (a, d) \in \tilde{\mathcal{H}} \text{ and } \boldsymbol{q} = (\tilde{a}, \tilde{d}) \in \mathcal{E} \\ &K(\boldsymbol{s}, \boldsymbol{q}) = P(\boldsymbol{s}, \boldsymbol{q}) \text{ otherwise.} \end{aligned} \tag{5}$$

It follows that

$$\begin{aligned} \mathbb{E}\left(\boldsymbol{B}_{t+1}\right) &= K^T \mathbb{E}\left(\boldsymbol{B}_t\right) \\ &= (K^T)^{t+1} \mathbb{E}\left(\boldsymbol{B}_0\right). \end{aligned} \tag{6}$$

Recall that Theorem 5.1 holds for a standard multi-type branching process under Assumption 1. In particular, $\tilde{K}$ is assumed to be irreducible. However, $K$ defined above is not irreducible. Lemma 5.1 below sheds further light on $K$. Theorem 5.2 observes that conclusions of Theorem 5.1 continue to hold for the branching process $\{\boldsymbol{B}_t\}$ associated with each epidemic process $\{\boldsymbol{X}_t^N\}$. For any matrix $M$, let $\rho(M)$ denote its spectral radius, that is, the maximum of the absolute values of all its eigenvalues.

Henceforth, we assume that $\rho = \rho(K) > 1$ in all subsequent analysis.

**Lemma 5.1.** *There exist matrices $K_1 \in \Re^{+\hat{\eta}\times\hat{\eta}}$, $C \in \Re^{+(\eta-\hat{\eta})\times(\eta-\hat{\eta})}$ and $M \in \Re^{+(\eta)\times(\eta-\hat{\eta})}$ such that*

$$K = \left(\begin{smallmatrix} K_1 & M \\ 0 & C \end{smallmatrix}\right),$$

*where $K_1 \in \Re^{+\hat{\eta}\times\hat{\eta}}$ is irreducible. Further, $\rho(C) \leq \rho(K_1)$.*

As $K_1$ defined in Lemma 5.1 is irreducible, its Perron Frobenius eigenvalue is equal to the spectral radius $\rho(K_1)$.

**Theorem 5.2.** *Spectral radius of $K$ is equal to the spectral radius of $K_1$, that is, $\rho(K) = \rho(K_1)$. Furthermore, $\rho = \rho(K)$ is a unique eigenvalue of $K$ with maximum absolute value and*

$$\lim_{t\to\infty} \frac{K^t}{\rho^t} = \boldsymbol{u}\boldsymbol{v}^T,$$

*where $\boldsymbol{u}$ and $\boldsymbol{v}$ are the strictly positive right and left eigenvectors of $K$ corresponding to eigenvalue $\rho = \rho(K)$ such that $\boldsymbol{u}^T\boldsymbol{v} = 1$ and $\sum_{i=1}^{\eta} u(i) = 1$. In addition,*

$$\frac{\boldsymbol{B}_t}{\rho^t} \xrightarrow{P} W\boldsymbol{v} \quad \text{as } t \to \infty,$$

*where $W$ is a non-negative random variable such that $P\{W > 0\} > 0$ iff $B_0(\boldsymbol{s}) \neq 0$ for some $\boldsymbol{s} \in \mathcal{H}$ and $\mathbb{E}\left(W|\boldsymbol{B}_0 = \boldsymbol{e_i}\right) = u(i)$ for all $i = 1, ..., \eta$. Also, let $A = \{\omega : B_t(\omega) \to \infty\}$ as $t \to \infty$. Then, for any $\epsilon > 0$ and for all $j \in [1, \eta]$,*

$$\lim_{t\to\infty} P\{\omega : \omega \in A, \left| \frac{B_t(j)}{\sum_{i=1}^{\eta} B_t(i)} - \frac{v(j)}{\sum_{i=1}^{\eta} v(i)} \right| > \epsilon\} = 0.$$

Observe that $\mathbb{E}\left(B_1(j) \log B_1(j)|\boldsymbol{B}_0 = \boldsymbol{e_i}\right) < \infty$ for all $1 \leq i, j \leq \eta$ holds because the offsprings for our branching process have a Poisson distribution.

## 5.4 Coupling the epidemic processes and associated branching process

- Recall that at time zero, $\boldsymbol{B}_0 = \boldsymbol{X}_0$. We couple each exposed individual of each type in the epidemic process to an exposed individual of the same type in the branching process at time zero.
- The coupled individuals in each of these processes follow the same disease progression (using same randomness) and stay coupled throughout the simulation.
- Further, when infectious, they generate identical Poisson number of contacts (in epidemic process) and offsprings (in branching process). Specifically, when a coupled individual with characteristic $a \in \mathcal{A}$ is infectious, the number of contacts it makes in a time step with all the individuals with characteristic $\tilde{a} \in \mathcal{A}$ in epidemic process is Poisson distributed with rate $\pi_{\tilde{a}}\psi_{a,\tilde{a}}\beta_a$. This equals the offsprings generated by the corresponding coupled individual in the branching process where the offsprings are with characteristic $\tilde{a} \in \mathcal{A}$ and are of type $(\tilde{a}, \textit{exposed})$.
- In epidemic process, each contact is made with an individual randomly selected from the population with the same characteristic. If a contact is with a susceptible person, then that person is marked exposed in the next time period and is coupled with the corresponding person in the branching process.
- On the other hand, if in epidemic process, a new contact is with an already affected individual, then this does not result in any person getting exposed so that the corresponding offspring in the branching process is uncoupled. The descendants of uncoupled individuals in the branching process are also uncoupled. In our analysis, we will show that such uncoupled individuals in the branching process vis-a-vis epidemic process are a negligible fraction of the coupled individuals till time $\log_\rho N^*$ for large $N$, where recall that $N^* = N/\log N$ and $\log_\rho N^* = \log N^*/\log \rho$.

Let $I_t^N$ denote the number of coupled individuals between epidemic process and branching process at time $t$. Since all the affected individuals in epidemic process are coupled with some individual in branching process of the same type, $I_t^N = A_t^N$ and we have,

$$X_t^N(\boldsymbol{s}) \leq B_t(\boldsymbol{s}) \quad \forall \boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}.$$

Further,

$$A_t^N \leq A_t^B. \tag{7}$$

The new uncoupled individuals born from the coupled individuals in branching process at each time $t$ are referred to as "ghost individuals" and are denoted by $G_t^N$. As mentioned earlier, these and their descendants remain uncoupled to epidemic system $N$. Let $D_{i,j}^N$ denote all the descendants of $G_i^N$ (ghost individuals born at time $i$) after $j$ time steps. Let $H_t^N$ denote the total number of uncoupled individuals in the branching process at time $t$. Then,

$$H_t^N = \sum_{i=1}^{t} D_{i,t-i}^N. \tag{8}$$

## 5.5 Analysis: The initial branching process phase

Following result shows that epidemic process is close to the multi-type branching process till time $\log_\rho N^*$ (recall that $N^* = N/\log N$ and $\log_\rho N^* = \log N^*/\log \rho$) as $N \to \infty$, where $\rho$ denotes the exponential growth rate of the branching process.

**Theorem 5.3.** *As $N \to \infty$, for all $\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}$,*

$$\sup_{t \in [0, \log_\rho (N^*/\sqrt{N})]} \left| X_t^N(\boldsymbol{s}) - B_t(\boldsymbol{s}) \right| \xrightarrow{P} 0. \tag{9}$$

$$\sup_{t \in \left[0, \log_\rho N^*\right]} \left| \frac{X_t^N(\boldsymbol{s})}{\rho^t} - \frac{B_t(\boldsymbol{s})}{\rho^t} \right| \xrightarrow{P} 0. \tag{10}$$

Result (9) was earlier shown for SIR setting in [4]. We extend this result to general models and also prove the result (10) in this more general setting.

Following lemma is used in proving Theorem 5.3.

**Lemma 5.2.** *For all $\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}$,*

$$\mathbb{E}\left(\left|B_t(\boldsymbol{s}) - X_t^N(\boldsymbol{s})\right|\right) \leq \tilde{e} \frac{\rho^{2t}}{N},$$

*for some constant $\tilde{e} > 0$.*

Recall from Theorem 5.2 that,

$$\frac{\boldsymbol{B}_t}{\rho^t} \xrightarrow{P} W \boldsymbol{v} \quad \text{as } t \to \infty,$$

where $W$ is a non-negative random variable representing the intensity of branching process and $\boldsymbol{v} \in \Re^{+\eta}$ is the left eigenvector corresponding to eigenvalue $\rho$ of matrix $K$. Therefore, initially (till time $\log_\rho N^*$) epidemic process grows exponentially at rate $\rho$, with sample path dependent intensity being determined by $W$.

The following proposition follows directly from Theorem 5.2 and Theorem 5.3 and justifies the fact that the proportions across different types stabilize quickly in the epidemic process, and thus early on, till time $\log_\rho N^*$, paths can be patched from one time period to the other with negligible error due to change in the proportions.

**Proposition 5.1.** *For $t_N \to \infty$ as $N \to \infty$ and $\limsup_{N\to\infty} \frac{t_N}{\log_\rho N^*} < \infty$, then for all $\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}$:*

$$\left| \frac{X_{t_N}^N(\boldsymbol{s})}{\sum_{\boldsymbol{q} \in \mathcal{S}\setminus\mathcal{U}} X_{t_N}^N(\boldsymbol{q})} - \frac{v(\boldsymbol{s})}{\sum_{\boldsymbol{q} \in \mathcal{S}\setminus\mathcal{U}} v(\boldsymbol{q})} \right| \xrightarrow{P} 0, \quad \text{as } N \to \infty.$$

**Remark 1.** In Algorithm 3 we had suggested that the restarted simulations should use common random numbers as in the original simulation paths so as to identically reproduce them. However, in our experiments we observe that even if the restarted paths are generated using independent samples, that leads to a negligible anomaly. To understand this, observe that to replicate an original path, we essentially need to replicate $W$ along that path, as after small initial period, the associated branching process is well specified once $W$ is known. (Note that this $W$ is implicitly generated in a simulation, it is not explicitly computed). Common random numbers achieve this. However, for approximating the statistics of the larger city after $t_{\min}$ (Day 22 in Figure 10), independent restarted simulations provide equally valid sample paths as the ones using the common random numbers. The only difference is that the $W$ associated with each independently generated path may not match the $W$ associated with the original path, so patching them together at time $t_{\min}$ may result in a mismatch. However, since we are reporting statistics associated with the average of generated paths, these statistics depend linearly on the average of the $W$'s associated with generated sample paths. Thus, if restarted paths are independently generated, the corresponding average of associated $W$'s may not match the average of $W$'s associated with the original simulations. Empirically, we observe negligible mismatch as the average of $W$'s appear to have small variance.

Compartmental models are widely used to model epidemics. Usually, in these models, we start with infected population which is a positive fraction of the overall population. However, little is known about the dynamics if we start with a small, constant number of infections. Below we describe the results for some popular compartmental models using Theorem 5.3 and 5.1.

**Example 1** (SIR Model)**.** Susceptible-Infectious-Recovered (SIR) models are the simplest aggregate compartmental models used in epidemiology. All the individuals in the city are assumed to be identical, i.e., they have same community transmission rate and disease progression transition probabilities. Individuals can be in three disease states (susceptible, infected, recovered). Let $S(t), I(t), R(t)$ denote the number of susceptible, infected and recovered at any time $t$. Let community transmission rate of an infected individual be $\beta$ and an infected individual recovers at rate $r$. SIR dynamics can be expressed by the following system of differential equations:

$$\begin{aligned}\frac{dS(t)}{dt} &= -\frac{\beta}{N}S(t)I(t),\\ \frac{dI(t)}{dt} &= \frac{\beta}{N}S(t)I(t) - rI(t),\\ \frac{dR(t)}{dt} &= rI(t).\end{aligned}$$

A discrete version of this model fits our setting. Again, for the ease of notation, time step is set to 1. Let $\beta$ denote community transmission rate of an infected individual, i.e., number of contacts made by an individual with other individuals in one time step is Poisson distributed with rate $\beta$, and an infected individual recovers at rate $r$. Again, till time $t = \log_\rho N^*$, the epidemic process is close to the branching process. For branching process, let the infected population be denoted as type 1, and the recovered population as type 2. Then, the $K$ matrix of the corresponding branching process is:

$$K = \left(\begin{smallmatrix} 1+\beta-r & r \\ 0 & 1 \end{smallmatrix}\right).$$

Observe that, matrix $K_1 = \big(1+\beta-r\big)$, $C = \big(1\big)$, $M = \big(r\big)$ and $\rho(K) = \rho(K_1) = 1+\beta-r$. Finally,

$$v = \left(\begin{smallmatrix} \frac{1}{r} \\ \frac{\beta-r}{r} \end{smallmatrix}\right).$$

Therefore, initially (till time $\log_\rho N^*$ for large $N$) epidemic process grows exponentially at rate $\rho = 1+\beta-r$, and the ratio of infected to recovered individuals in this phase quickly stabilizes to $\frac{\beta-r}{r}$.

**Example 2** (SEIR model)**.** Recall that SEIR model corresponds to susceptible-exposed-infected-recovered model. As the name suggests, SEIR model considers an additional disease state of exposed as a refinement to SIR model. Again, all the individuals are identical. Community transmission rate of an infected individual is $\beta$ (i.e., number of contacts made by an individual with other individuals is Poisson distributed with rate $\beta$). An exposed individual transitions to infected state with rate $p$ and an infected individual recovers at rate $r$.

Let the exposed population be denoted as type 1, infected population as type 2, and recovered population as type 3. Then, the $K$ matrix of the branching process is:

$$K = \begin{pmatrix} 1-p & p & 0 \\ \beta & 1-r & r \\ 0 & 0 & 1 \end{pmatrix}.$$

Observe that, $C = \big(1\big)$, $M = \left(\begin{smallmatrix} 0 \\ r \end{smallmatrix}\right)$, $K_1 = \left(\begin{smallmatrix} 1-p & p \\ \beta & 1-r \end{smallmatrix}\right)$ and $\rho(K) = \rho(K_1) = 1 + \frac{1}{2}\big(((p+r)^2 + 4p(\beta-r))^{1/2} - (p+r)\big)$. Finally,

$$v = \begin{pmatrix} \beta \\ p+\rho-1 \\ r+\frac{rp}{\rho-1} \end{pmatrix}.$$

Therefore, initially (till time $t = \log_\rho N^*$, for large $N$) epidemic process grows exponentially at rate $\rho = 1 + \frac{1}{2}\big(((p+r)^2 + 4p(\beta-r))^{1/2} - (p+r)\big)$, and the proportions of exposed, infected and recovered quickly stabilize to $\big(\beta, (p+\rho-1), (r+\frac{rp}{\rho-1})\big)$ normalised by their sum.

**Example 3** (SIR model with two age groups)**.** Usually different age groups have different disease progression profiles. To account for this, we consider SIR model with two age groups. We limit ourselves to two groups so that the growth rate of epidemic process can be expressed in closed form (solution to a quadratic equation. In general, the degree of the equation is same as number of age groups). Suppose that the fraction of individuals in each age group is $\pi_1$ and $\pi_2 = 1 - \pi_1$. There are only 3 disease states (susceptible, infected and recovered). Community transmission rate of an infected individual is $\beta$ (i.e., number of contacts made by an individual with other individuals is Poisson distributed with rate $\beta$). The infected individual make contacts with people belonging to age group 1 and age group 2 at rate $\pi_1\beta$ and $\pi_2\beta$, respectively. Further, suppose that an infected individual from first and second age group recover at rates $r_1$ and $r_2$, respectively.

We then have four types of population - First and second type denote the number infected in the first and second age group. Third and fourth type correspond to number recovered in first and second type age group. Then the $K$ matrix of this branching process is:

$$K = \begin{pmatrix} \pi_1\beta + 1 - r_1 & \pi_2\beta & r_1 & 0 \\ \pi_1\beta & \pi_2\beta + 1 - r_2 & 0 & r_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here, as per the notation in Lemma 3.1, $C = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $M = \left(\begin{smallmatrix} r_1 & 0 \\ 0 & r_2 \end{smallmatrix}\right)$, $K_1 = \begin{pmatrix} \pi_1\beta + 1 - r_1 & \pi_2\beta \\ \pi_1\beta & \pi_2\beta + 1 - r_2 \end{pmatrix}$ and

$$\rho(K) = \rho(K_1) = 1 + \frac{1}{2}\left(\left((\beta - r_1 - r_2)^2 + 4\pi_1 r_2(\beta - r_1) + 4\pi_2 r_1(\beta - r_2)\right)^{1/2} + (\beta - r_1 - r_2)\right)$$

Finally,

$$v = \begin{pmatrix} \pi_1\beta \\ r_1 + \rho - 1 - \pi_1\beta \\ \frac{r_1}{\rho - 1}\pi_1\beta \\ \frac{r_2}{\rho - 1}(r_1 + \rho - 1 - \pi_1\beta) \end{pmatrix}.$$

Therefore, initially (for time $\leq \log_\rho N^*$ and large $N$) epidemic process grows exponentially at rate

$$\rho = 1 + \frac{1}{2}\Big(\left((\beta - r_1 - r_2)^2 + 4\pi_1 r_2(\beta - r_1) + 4\pi_2 r_1(\beta - r_2)\right)^{1/2} + (\beta - r_1 - r_2)\Big),$$

and the population of the four types (infected in two age groups and recovered in two age groups) quickly stabilizes to be proportional to

$$\left(\pi_1\beta, (r_1 + \rho - 1 - \pi_1\beta), (\frac{r_1}{\rho - 1}\pi_1\beta), (\frac{r_2}{\rho - 1}(r_1 + \rho - 1 - \pi_1\beta))\right).$$

## 5.6 Deterministic phase

As Theorem 5.3 and Proposition 5.1 note, early in the infection growth, till time $\log_\rho N^*$ for large $N$, while the number affected grows exponentially, the proportion of individuals across different types stabilizes. Here, the types corresponding to the susceptible population are not considered because at this stage, the affected are a negligible fraction of the total susceptible population. The growth in the affected population in this phase is sample path dependent and depends upon non-negative random variable $W$.

However, at time $\log_\rho(\epsilon N)$ for any $\epsilon > 0$ and large $N$, this changes as the affected population equals $\Theta(N)$. Hereafter, the population growth closely approximates its mean field limit whose initial state depends on random $W$ and where the proportions across types may change as the time progresses. Our key result in this setting is Theorem 5.4. To this end we need Assumption 2.

Let $t_N = log_\rho(\epsilon N)$. Let $\boldsymbol{\mu}_t^N$ denote the empirical distribution across types at time $t_N + t$. This corresponds to augmenting the vector $\boldsymbol{X}_{t_N+t}^N$ with the types associated with the susceptible population at time $t_N + t$ and scaling the resultant vector with factor $N^{-1}$.

**Assumption 2.** *There exists a random distribution* $\bar{\boldsymbol{\mu}}_0(W) \in \mathbb{R}^{+(\eta+1)}$ *that is independent of* $N$ *such that* $\boldsymbol{\mu}_0^N \xrightarrow{P} \bar{\boldsymbol{\mu}}_0(W)$ *as* $N \to \infty$.

Observe that $\bar{\boldsymbol{\mu}}_0(W)$ above is path dependent in that it depends on the random variable $W$. The above assumption is seen to hold empirically. While, we do not have a proof for it (this appears to be a difficult and open problem), Corollary 5.1 below supports this assumption. This corollary follows from Lemma 5.2.

**Corollary 5.1.** *For* $\epsilon \in (0,1)$, $t = \log_\rho(\epsilon N)$ *and for all* $\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}$,

$$\mathbb{E}\left(\left|\frac{X_t^N(\boldsymbol{s})}{\rho^t} - \frac{B_t(\boldsymbol{s})}{\rho^t}\right|\right) \leq \epsilon\tilde{e},$$

*for some constant* $\tilde{e} > 0$.

Let $c_t^N(a)$ denote the total incoming-infection rate from the community as seen by an individual with characteristic $a \in \mathcal{A}$ at time $t$. It is determined from the disease state of all the individuals at time $t$ and is $\mathcal{F}_t$ measurable. In our setup, $c_t^N(a)$ equals

$$\sum_{\boldsymbol{q}=(\tilde{a},\tilde{d})\in\mathcal{S}\setminus\mathcal{U}} \mu_t^N(\boldsymbol{q})\mathbb{1}(\text{type } \boldsymbol{q} \text{ is infectious})\psi_{\tilde{a},a}\beta_{\tilde{a}}, \tag{11}$$

where $\mathbb{1}$ denotes the indicator function.

For each individual $n \leq N$, let $S_n^t$ denote its type at time $t$. Then, for $\boldsymbol{s} = (a,d) \in \mathcal{S}$

$$\mathbb{P}\left(S_n^{t+1} = \boldsymbol{s} | \mathcal{F}_t\right) = h(S_n^t, \boldsymbol{s}, c_t^N(a)),$$

for a continuous function $h$. In particular, the transition probability only depends on the disease-state of the individual at the previous time, the disease-state to which it is transitioning, and the infection rate incoming to the individual at that time.

Recall that from Assumption 2 we have defined $\bar{\boldsymbol{\mu}}_0(W) \in \mathbb{R}^{+(\eta+1)}$ such that $\boldsymbol{\mu}_0^N \xrightarrow{P} \bar{\boldsymbol{\mu}}_0(W)$ as $N \to \infty$. Define $\bar{\boldsymbol{\mu}}_t(W) \in \mathbb{R}^{+(\eta+1)}$ such that for all $t \in \mathbb{N}$, $\boldsymbol{s} = (a,d) \in \mathcal{S}$,

$$\bar{\mu}_t(\boldsymbol{s}, W) := \sum_{\boldsymbol{s'}\in\mathcal{S}} \bar{\mu}_{t-1}(\boldsymbol{s'}, W) h(\boldsymbol{s'}\boldsymbol{s}, \bar{c}_{t-1}(a, W)), \tag{12}$$

where

$$\bar{c}_{t-1}(a, W) = \sum_{\boldsymbol{q}=(\tilde{a},\tilde{d})\in\mathcal{S}\setminus\mathcal{U}} \bar{\mu}_{t-1}(\boldsymbol{q}, W)\mathbb{1}(\text{state } \boldsymbol{q} \text{ is infectious})\psi_{\tilde{a},a}\beta_{\tilde{a}}.$$

**Theorem 5.4.** *Under Assumption 2 and for* $t \in \mathbb{N}$,

$$\boldsymbol{\mu}_t^N \xrightarrow{P} \bar{\boldsymbol{\mu}}_t(W) \text{ as } N \to \infty.$$

In particular, if $\bar{\boldsymbol{\mu}}_t$ denotes the mean field limit of the normalised process at time $t + \log_\rho(\epsilon N)$, then, the number of infections observed in a smaller model with population $N$ is approximately $N * \bar{\boldsymbol{\mu}}_t$ and that of a larger model is approximately $kN * \bar{\boldsymbol{\mu}}_t$. Thus, the larger model infection process can be approximated by the smaller model infection process by scaling it by $k$.

**Remark 2.** While the deterministic equations (conditioned on $W$) in Theorem 5.4 can be easily solved when the number of types is small and initial state is established via simulation, these become much harder in a realistic model with all its complexity, where the number of types is extremely large and may be uncountable if non-memoryless probability distributions are involved. One may thus view the key role of the simulator as a tool that identifies the random initial state of these deterministic mean field equations and then solves them somewhat efficiently using stochastic methods.

## Acknowledgments

We acknowledge the support of A.T.E. Chandra Foundation for this research. We further acknowledge the support of the Department of Atomic Energy, Government of India, to TIFR under project no. 12-R&D-TFR-5.01-0500.

## References

[1] S. Agrawal, S. Bhandari, A. Bhattacharjee, A. Deo, N.M. Dixit, P. Harsha, S. Juneja, P. Kesarwani, A.K. Swamy, P. Patil, N. Rathod, R. Saptharishi, S. Shriram, P. Srivastava, R. Sundaresan, N. K. Vaidhiyan, and S. Yasodharan. City-scale agent-based simulators for the study of non-pharmaceutical interventions in the context of the covid-19 epidemic. *Journal of the Indian Institute of Science*, 100(4):809–847, 2020.

[2] K.B. Athreya and P.E. Ney. *Branching processes*. 1972.

[3] O. N. Bjørnstad, Katriona Shea, Martin Krzywinski, and Naomi Altman. The SEIRS model for infectious disease dynamics. *Nature*, 2020.

[4] T. Britton and E. Pardoux. *Stochastic Epidemic Models with Inference*. 2019.

[5] Mohammadali Dashtbali and Mehdi Mirzaie. A compartmental model that predicts the effect of social distancing and vaccination on controlling COVID-19. *Nature*, 2021.

[6] Neil M Ferguson, Derek AT Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S Burke. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209–214, 2005.

[7] Neil M Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Technical Report 9, MRC Centre for Global Infectious Disease Analysis, Imperial College London, U.K., 2020. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/.

[8] J. Fernández-Villaverdea and Charles Jones. Estimating and simulating a SIRD Model of COVID-19 for many countries, states, and cities. *Journal of Economic Dynamics and Control*, 2022.

[9] Jasmine M Gardner, Lander Willem, Wouter van der Wijngaart, Shina Caroline Lynn Kamerlin, Nele Brusselaers, and Peter Kasson. Intervention strategies against covid-19 and their estimated impact on swedish healthcare capacity. *medRxiv*, 2020.

[10] M Elizabeth Halloran, Neil M Ferguson, Stephen Eubank, Ira M Longini, Derek AT Cummings, Bryan Lewis, Shufu Xu, Christophe Fraser, Anil Vullikanti, Timothy C Germann, et al. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644, 2008.

[11] T.E. Harris. *The Theory of Branching Processes*. 1963.

[12] P. Harsha, S. Juneja, D. Mittal, and R. Saptharishi. Covid-19 epidemic in Mumbai: Projections, full economic opening, and containment zones versus contact tracing and testing: An update. Technical report, TIFR Mumbai, India, 2020. https://arxiv.org/abs/2011.02032.

[13] P. Harsha, S. Juneja, P. Patil, N. Rathod, R. Saptharishi, A.Y. Sarath, S. Sriram, P. Srivastava, R. Sundaresan, and N.K. Vaidhiyan. COVID-19 Epidemic Study II: Phased emergence from the lockdown in Mumbai. Technical report, India, 2020. https://arxiv.org/abs/2006.03375.

[14] P. Harsha, S. Juneja, and R. Saptharishi. Covid-19 epidemic in mumbai: Long term projections, full economic opening, and containment zones versus contact tracing and testing. Technical report, TIFR Mumbai, India, 2020. http://www.tcs.tifr.res.in/~sandeepj/avail_papers/Mumbai_September_Report.pdf.

[15] H.W. Hethcote. The mathematics of infectious diseases. *SIAM*, 2000.

[16] Elizabeth Hunter, Brian Mac Namee, and John D Kelleher. A taxonomy for agent-based models in human infectious disease epidemiology. *Journal of Artificial Societies and Social Simulation*, 2017.

[17] S. Juneja and D. Mittal. Modelling the Second Covid-19 Wave in Mumbai. Technical report, TIFR Mumbai, India, 2021. https://arxiv.org/pdf/2105.02144.pdf.

[18] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[19] A. Malani, D. Shah, G. Kang, G.N. Lobo, J. Shastri, M. Mohanan, R. Jain, S.T. Agrawal, S. Juneja, S. Imad, and U. Kolthur-Seetharam. Seroprevalence of SARS-CoV-2 in slums versus non-slums in Mumbai, India. *The Lancet Global Health*, 9(2):E110–E111, 2020.

[20] M.W. Meckes. *Lecture notes on matrix analysis*. 2019.

[21] D. Mittal, S. Juneja, and S. Agrawal. Shift, scale and restart smaller models to estimate larger ones: Agent based simulators in epidemiology. *To appear in ACM SIGMETRICS - Performance Evaluation Review*, 2022.

[22] Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick GT Walker, Han Fu, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 2020.

# 6 Appendix

In this section we first give proofs (Section 6.1) of the results stated in Section 5. In Section 6.2, we provide a heuristic argument for the mismatch between larger model and smaller model when we start with small number of initial infections. Finally, we give details of the parameters and the city statistics used in numerical experiments in Section 6.3.

## 6.1 Proofs

### 6.1.1 Proof of Lemma 5.1

- Observe that types $\boldsymbol{s} \in \mathcal{H}^c \backslash \mathcal{U}$ correspond to hospitalised, critical, dead or recovered states and types $\boldsymbol{q} \in \mathcal{H}$ correspond to exposed infective or symptomatic state. As any individual in hospitalised, critical, dead or recovered disease states neither transitions nor gives birth to offspring in exposed, infective or symptomatic disease states, therefore $K(\boldsymbol{s}, \boldsymbol{q}) = 0$ for all $\boldsymbol{s} \in \mathcal{H}^c \setminus \mathcal{U}$ and $\boldsymbol{q} \in \mathcal{H}$.

- Recall that $K_1 \in \Re^{+\hat{\eta} \times \hat{\eta}}$ is such that $K_1(\boldsymbol{s}, \boldsymbol{q}) = K(\boldsymbol{s}, \boldsymbol{q})$ for all $\boldsymbol{s} \in \mathcal{H}$ and $\boldsymbol{q} \in \mathcal{H}$. Observe that $K_1$ correspond to the "types" with the disease states that are infectious or may become infectious in subsequent time steps. As every infectious individual gives birth to exposed individuals of each characteristic therefore $K_1$ is irreducible.

- Recall that matrix $C \in \Re^{+(\eta-\hat{\eta}) \times (\eta-\hat{\eta})}$ is such that $C(\boldsymbol{s}, \boldsymbol{q}) = K(\boldsymbol{s}, \boldsymbol{q})$ for all $\boldsymbol{s} \in \mathcal{H}^c \setminus \mathcal{U}$ and $\boldsymbol{q} \in \mathcal{H}^c \setminus \mathcal{U}$. The individuals of the type $\boldsymbol{s} \in \mathcal{H}^c \setminus \mathcal{U}$ do not contribute to any infections in the city. In addition, this individual either transitions to the next disease state or remains in the same disease state with positive probability. Therefore matrix $C$ is an upper triangular matrix with diagonal entries equalling the probability of an individual in some disease state remaining in the same disease state in one time step (which is less than 1). Hence, all the eigenvalues of matrix C are positive and less than or equal to 1.

- Recall that matrix $M \in \Re^{+\hat{\eta} \times (\eta-\hat{\eta})}$ is such that $M(\boldsymbol{s}, \boldsymbol{q}) = K(\boldsymbol{s}, \boldsymbol{q})$ for all $\boldsymbol{s} \in \mathcal{H}$ and $\boldsymbol{q} \in \mathcal{H}^c \setminus \mathcal{U}$.

Therefore, there exists $K_1 \in \Re^{+\hat{\eta}\times\hat{\eta}}$, $C \in \Re^{+(\eta-\hat{\eta})\times(\eta-\hat{\eta})}$ and $M \in \Re^{+(\eta)\times(\eta-\hat{\eta})}$ such that

$$K = \begin{pmatrix} K_1 & M \\ 0 & C \end{pmatrix},$$

where $K_1 \in \Re^{+\hat{\eta}\times\hat{\eta}}$ is irreducible, and $\rho(C) < \rho(K) = \rho(K_1)$.

### 6.1.2 Proof of Theorem 5.2

Recall from Lemma 5.1,

$$K = \begin{pmatrix} K_1 & M \\ 0 & C \end{pmatrix}.$$

Observe that, set of eigenvalues of $K$ is same as combined set of eigenvalues of $K_1$ and $C$. Furthermore,

- $\rho(C) \leq 1$ from proof of Lemma 5.1, therefore $\rho(K) = \rho(K_1)$.
- $K_1$ is irreducible from Lemma 5.1, therefore its spectral radius $\rho(K_1)$ is Perron-Frobenius eigenvalue. In addition, spectral radius $\rho(K) = \rho(K_1)$ of $K$ is also a unique eigenvalue of $K$ with highest absolute value.
- $\lim_{t\to\infty} \frac{K^t}{\rho^t} = \boldsymbol{u}\boldsymbol{v}^T$ then follows from [20] (Lemma 8.16, pg. 69).

Results (3) and (4) in Theorem 5.1 for supercritical multi-type branching process (Section 5.1) were proved in [2], Chap. V. Critical step in proving these results was the fact that corresponding matrix $\tilde{K}$ had following property $\lim_{t\to\infty} \frac{\tilde{K}^t}{\tilde{\rho}^t} = \tilde{\boldsymbol{u}}\tilde{\boldsymbol{v}}^T$ (see Lemma 1 - Chap. V, pg. 194, in [2]).

As shown above, matrix $K$ of the associated branching process also satisfies this property ($\lim_{t\to\infty} \frac{K^t}{\rho^t} = \boldsymbol{u}\boldsymbol{v}^T$), therefore results (3) and (4) in Theorem 5.1 hold for our branching process as well.

### 6.1.3 Proof of Lemma 5.2

Recall that total number of uncoupled individuals in branching process at time $t$ is $H_t^N$. Therefore, for all $\boldsymbol{s} \in \mathcal{S} \setminus \mathcal{U}$

$$\left|B_t(\boldsymbol{s}) - X_t^N(\boldsymbol{s})\right| \leq H_t^N. \tag{13}$$

From (8) and (13) we have

$$\left|B_t(\boldsymbol{s}) - X_t^N(\boldsymbol{s})\right| \leq \sum_{i=1}^{t} D_{i,t-i}^N. \tag{14}$$

Before proceeding with further analysis, we assume two results (15) and (16), which we will show later.

For some constant $\hat{\beta} > 0$,

$$\sum_{i=1}^{t} \mathbb{E}\left(D_{i,t-i}^N\right) \leq \frac{\hat{\beta}}{N} \mathbf{1}^T \sum_{i=0}^{t} K^{t-i} \mathbb{E}\left((A_{i-1}^B)^2\right) \mathbf{1}, \tag{15}$$

where $A_i^B$ is total number of affected individuals in branching process at time $i$ and $\mathbf{1} \in \Re^{+\eta}$ is a vector with each entry equal to 1.

For some constant $\tilde{c}$,

$$\mathbb{E}\left(\left(A_i^B\right)^2\right) \leq \tilde{c}\rho^{2i}. \tag{16}$$

Assuming (15) to be true, from (14) and (15) we have

$$\mathbb{E}\left(\left|B_t(\boldsymbol{s}) - X_t^N(\boldsymbol{s})\right|\right) \leq \frac{\hat{\beta}}{N} \mathbf{1}^T \sum_{i=0}^{t} K^{t-i} \mathbb{E}\left((A_{i-1}^B)^2\right) \mathbf{1}. \tag{17}$$

As $\lim_{t\to\infty} \frac{K^t}{\rho^t} = \boldsymbol{u}\boldsymbol{v}^T$, therefore there exists $K_3 \in \Re^{+\eta\times\eta}$ such that for all $t \geq 0$

$$K^t \leq \rho^t K_3. \tag{18}$$

Assuming (16) to be true, from (16), (17) and (18) we have,

$$\begin{aligned}
\mathbb{E}\left(\left|B_t(\boldsymbol{s}) - X_t^N(\boldsymbol{s})\right|\right) &\leq \frac{\hat{\beta}}{N}\mathbf{1}^T \sum_{i=0}^{t} \rho^{t-i} K_3 \tilde{c} \rho^{2i-2}\mathbf{1} \\
&\leq \frac{\hat{\beta}}{N}\mathbf{1}^T K_3 \tilde{c} \frac{\rho^{2t-1}}{(\rho-1)}\mathbf{1} \\
&= \tilde{e}\frac{\rho^{2t}}{N},
\end{aligned}$$

where $\tilde{e} = \left(\frac{\hat{\beta}}{\rho}\mathbf{1}^T \frac{K_3\tilde{c}}{(\rho-1)}\mathbf{1}\right)$.

**Proof of (15):** Recall that ghost individuals $G_t^N$ is the number of uncoupled exposed individuals born from the coupled individuals in branching process when compared to epidemic process at time $t$. These extra individuals in branching process are born whenever a coupled infectious individual in epidemic process makes contact with already affected individual. Therefore, $G_t^N$ is equal to the number of contacts made to already affected individuals in epidemic process in one time step between time $t-1$ and $t$. Let $Y_t^N$ be the total number of community contacts in one time step between time $t-1$ and $t$. For a given contact, probability of hitting an already affected individual is less then $\psi_{max}\frac{A_{t-1}^N + Y_t^N}{N}$, where $\psi_{max} = \max_{a\in\mathcal{A},\tilde{a}\in\mathcal{A}}\{\psi_{a,\tilde{a}}\}$. Therefore, we have,

$$\begin{aligned}
\mathbb{E}\left(G_t^N\right) &= \mathbb{E}\left(\mathbb{E}\left(G_t^N|\mathcal{F}_{t-1}\right)\right) \\
&\leq \mathbb{E}\left(\mathbb{E}\left(\psi_{max}\frac{(A_{t-1}^N + Y_t^N)}{N}Y_t^N|\mathcal{F}_{t-1}\right)\right) \\
&\leq \mathbb{E}\left(\mathbb{E}\left(\psi_{max}\frac{A_{t-1}^N Y_t^N}{N}|\mathcal{F}_{t-1}\right)\right) + \mathbb{E}\left(\mathbb{E}\left(\psi_{max}\frac{(Y_t^N)^2}{N}|\mathcal{F}_{t-1}\right)\right).
\end{aligned}$$

To upper bound $Y_t^N$, we assume that all the affected individuals in epidemic process at time $t-1$ are infectious and that community hits by each infectious individual is Poisson distributed with $\beta_{max} = \max_{a\in\mathcal{A}}\{\beta_a\}$. Then,

$$\mathbb{E}\left(G_t^N\right) \leq \psi_{max}\mathbb{E}\left(\beta_{max}\frac{(A_{t-1}^N)^2}{N}\right) + \psi_{max}\mathbb{E}\left(\frac{\beta_{max}A_{t-1}^N + \beta_{max}^2(A_{t-1}^N)^2}{N}\right).$$

Since $A_{t-1}^N \geq 1$, and setting $\hat{\beta} = \psi_{max}(\beta_{max}^2 + 2\beta_{max})$, we have

$$\mathbb{E}\left(G_t^N\right) \leq \hat{\beta}\mathbb{E}\left(\frac{(A_{t-1}^N)^2}{N}\right).$$

To upper bound the $\mathbb{E}\left(D_{i,t-i}^N\right)$, we assume that each state $\boldsymbol{s} \in \mathcal{S}\setminus\mathcal{U}$ has $G_i^N$ ghost individuals born at time $i$. As ghost individuals $G_i^N$ born at time $i$ have descendants according to the same branching process dynamics,

$$\mathbb{E}\left(D_{i,t-i}^N\right) \leq \frac{\hat{\beta}}{N}\mathbf{1}^T K^{t-i}\mathbb{E}\left((A_{i-1}^N)^2\right)\mathbf{1}.$$

Taking summation on both sides we have,

$$\sum_{i=1}^{t}\mathbb{E}\left(D_{i,t-i}^N\right) \leq \frac{\hat{\beta}}{N}\mathbf{1}^T \sum_{i=1}^{t} K^{t-i}\mathbb{E}\left((A_{i-1}^N)^2\right)\mathbf{1}. \tag{19}$$

From (7) and (19), we have

$$\sum_{i=1}^{t}\mathbb{E}\left(D_{i,t-i}^N\right) \leq \frac{\hat{\beta}}{N}\mathbf{1}^T \sum_{i=1}^{t} K^{t-i}\mathbb{E}\left((A_{i-1}^B)^2\right)\mathbf{1}.$$

**Proof of (16):** Recall that $A_i^B = \sum_{\boldsymbol{s}\in\mathcal{S}\setminus\mathcal{U}} B_i(\boldsymbol{s})$. Then, $A_i^B = \mathbf{1}^T\boldsymbol{B}_i = \boldsymbol{B}_i^T\mathbf{1}$. Hence,

$$\begin{aligned}\mathbb{E}\left(\left(A_i^B\right)^2\right) &\leq \mathbb{E}\left((\mathbf{1}^T B_i)^2\right)\\ &= \mathbf{1}^T\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{B}_i^T\right)\mathbf{1}.\end{aligned} \tag{20}$$

Let $Var(\boldsymbol{B}_1) \in \Re^{+\eta\times\eta}$, be a matrix with $(\boldsymbol{s},\boldsymbol{q})$ entry equalling $\mathbb{E}\left(B_1(\boldsymbol{s})B_1(\boldsymbol{q})\right) - \mathbb{E}\left(B_1(\boldsymbol{s})\right)\mathbb{E}\left(B_1(\boldsymbol{q})\right)$, $V_{\boldsymbol{s}}$ denote $Var(\boldsymbol{B}_1|\boldsymbol{B}_0 = \boldsymbol{e}_{\boldsymbol{s}})$ and $C_0$ denote the matrix with $(\boldsymbol{s},\boldsymbol{q})$ entry set to $\mathbb{E}\left(B_0(\boldsymbol{s})B_0(\boldsymbol{q})\right)$. From (Chap. 2, pg 37, [11]) we have,

$$\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{B}_i^T\right) = (K^i)^T C_0 K^i + \sum_{j=1}^{i}(K^{i-j})^T\Big(\sum_{\boldsymbol{s}\in\mathcal{S}\setminus\mathcal{U}} V_{\boldsymbol{s}}\mathbb{E}\left(\boldsymbol{B}_{j-1}(\boldsymbol{s})\right)\Big)K^{i-j}. \tag{21}$$

Furthermore, for our branching process, there exists a constant $v > 0$ such that $V_{\boldsymbol{s}} \leq vK$ for all $\boldsymbol{s}\in\mathcal{S}\setminus\mathcal{U}$. Using this in (21) we observe that

$$\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{B}_i^T\right) = (K^i)^T C_0 K^i + v\sum_{j=1}^{i}(K^{i-j})^T\Big(\sum_{\boldsymbol{s}\in\mathcal{S}\setminus\mathcal{U}} \mathbb{E}\left(B_{j-1}(\boldsymbol{s})\right)\Big)K^{i-j+1}. \tag{22}$$

From (6) and (18) we have

$$\boldsymbol{B}_i \leq \rho^i (K_3)^T\boldsymbol{B}_0. \tag{23}$$

.

From (22) and (23) we have

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{B}_i^T\right) &\leq \rho^{2i}(K_3)^T C_0 K_3 + v\tilde{d}\sum_{j=1}^{i}\rho^{2i-j}(K_3)^T K_3\\ &\leq \rho^{2i}\left((K_3)^T C_0 K_3 + \frac{\rho v\tilde{d}}{\rho-1}(K_3)^T K_3\right)\\ &= \rho^{2i}\tilde{K}_3,\end{aligned}$$

where, $\tilde{d} := \sum_{\boldsymbol{s}\in\mathcal{S}\setminus\mathcal{U}}\left(\mathbb{E}\left(B_0^T\right)K_3\right)(\boldsymbol{s})$ and $\tilde{K}_3 := \left((K_3)^T C_0 K_3 + \frac{\rho v\tilde{d}}{\rho-1}(K_3)^T K_3\right)$.

Using above bound in (20) and setting $\tilde{c} = \mathbf{1}^T\tilde{K}_3\mathbf{1}$, we get

$$\mathbb{E}\left(\left(A_i^B\right)^2\right) \leq \tilde{c}\rho^{2i}.$$

### 6.1.4 Proof for Theorem 5.3

For $\zeta > 0$, we have

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0,\log_\rho(N^*/\sqrt{N})]}\left|X_t^N(\boldsymbol{s}) - B_t(\boldsymbol{s})\right| \geq \zeta\right) &\leq \frac{1}{\zeta}\mathbb{E}\left(\sup_{t\in[0,\log_\rho(N^*/\sqrt{N})]}\left|X_t^N(\boldsymbol{s}) - B_t(\boldsymbol{s})\right|\right)\\ &\leq \frac{1}{\zeta}\sum_{t=0}^{\log_\rho(N^*/\sqrt{N})}\mathbb{E}\left(\left|X_t^N(\boldsymbol{s}) - B_t(\boldsymbol{s})\right|\right),\end{aligned}$$

where the first inequality follows from the Markov's inequality, and the second follows from observing that maximum of non-negative random variables is bounded by their sum. Similarly,

$$\mathbb{P}\left(\sup_{t\in[0,\log_\rho N^*]}\left|\frac{X_t^N(\boldsymbol{s})}{\rho^t}-\frac{B_t(\boldsymbol{s})}{\rho^t}\right|\geq\zeta\right)\leq\frac{1}{\zeta}\mathbb{E}\left(\sup_{t\in[0,\log_\rho N^*]}\left|\frac{X_t^N(\boldsymbol{s})}{\rho^t}-\frac{B_t(\boldsymbol{s})}{\rho^t}\right|\right)$$
$$\leq\frac{1}{\zeta}\sum_{t=0}^{\log_\rho N^*}\mathbb{E}\left(\left|\frac{X_t^N(\boldsymbol{s})}{\rho^t}-\frac{B_t(\boldsymbol{s})}{\rho^t}\right|\right).$$

Thus, it is sufficient to show that

$$\lim_{N\to\infty}\left\{\sum_{t=0}^{\log_\rho(N^*/\sqrt{N})}\mathbb{E}\left(\left|X_t^N(\boldsymbol{s})-B_t(\boldsymbol{s})\right|\right)\right\}=0,$$

and

$$\lim_{N\to\infty}\left\{\sum_{t=0}^{\log_\rho N^*}\mathbb{E}\left(\left|\frac{X_t^N(\boldsymbol{s})}{\rho^t}-\frac{B_t(\boldsymbol{s})}{\rho^t}\right|\right)\right\}=0.$$

From Lemma 5.2, we have,

$$\mathbb{E}\left(\left|X_t^N(\boldsymbol{s})-B_t(\boldsymbol{s})\right|\right)\leq\tilde{e}\frac{\rho^{2t}}{N},\tag{24}$$

Taking sum from $t=0$ to $t=\log_\rho(N^*/\sqrt{N})$ above, we observe that

$$\sum_{i=0}^{\log_\rho(N^*/\sqrt{N})}\mathbb{E}\left(\left|X_i^N(\boldsymbol{s})-B_i(\boldsymbol{s})\right|\right)\leq\frac{\rho^{2\log_\rho(N^*/\sqrt{N})}}{N}\frac{\tilde{e}}{\rho^2-1},$$

recall that $N^*=N/\log N$. Therefore,

$$\lim_{N\to\infty}\mathbb{E}\left(\sup_{t\in[0,\log_\rho(N^*/\sqrt{N})]}\left|X_t^N(\boldsymbol{s})-B_t(\boldsymbol{s})\right|\right)=0.$$

Using (24) to bound $\sum_{i=0}^t\frac{1}{\rho^i}\left(\mathbb{E}(\left|X_i^N(\boldsymbol{s})-B_i(\boldsymbol{s})\right|)\right)$ we have,

$$\sum_{i=0}^t\frac{1}{\rho^i}\left(\mathbb{E}(\left|X_i^N(\boldsymbol{s})-B_i(\boldsymbol{s})\right|)\right)\leq\frac{\rho^t}{N}\frac{\tilde{e}}{\rho-1}.$$

Setting $t=\log_\rho N^*$, we observe that

$$\sum_{i=0}^t\mathbb{E}\left(\left|\frac{X_i^N(\boldsymbol{s})}{\rho^i}-\frac{B_i(\boldsymbol{s})}{\rho^i}\right|\right)\leq\frac{1}{\log N}\frac{\tilde{e}}{\rho-1}.$$

Thus,

$$\lim_{N\to\infty}\mathbb{E}\left(\sup_{t\in[0,\log_\rho N^*]}\left|\frac{X_t^N(\boldsymbol{s})}{\rho^t}-\frac{B_t(\boldsymbol{s})}{\rho^t}\right|\right)=0.$$

### 6.1.5 Proof of Theorem 5.4

We will prove Theorem 5.4 through induction. For the ease of notation we hide $W$ in representing the mean field process $\bar{\boldsymbol{\mu}}_t(W)$ and $\bar{c}_t(W,a)$ in the proof. Assumption 2 forms the base case for induction. Assume that $\boldsymbol{\mu}_t^N\xrightarrow{\mathrm{P}}\bar{\boldsymbol{\mu}}_t$. From the definition we know that for all $\boldsymbol{s}=(a,d)\in\mathcal{S}$,

$$\mu_{t+1}^N(\boldsymbol{s}) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\left(S_n^{t+1} = \boldsymbol{s}\right).$$

Recall that

$$\mathbb{P}\left(S_n^{t+1} = \boldsymbol{s}|\mathcal{F}_t\right) = h(S_n^t, \boldsymbol{s}, c_t^N(a)),$$

Define $\mathcal{M}_{t+1}^N(\boldsymbol{s})$ to be

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\left(S_n^{t+1} = \boldsymbol{s}\right) - h(S_n^t, \boldsymbol{s}, c_t^N(a)).$$

Clearly,

$$\mu_{t+1}^N(\boldsymbol{s}) = \sum_{\boldsymbol{s'}\in\mathcal{S}} \left[h(\boldsymbol{s'}, \boldsymbol{s}, c_t^N(a))\mu_t^N(\boldsymbol{s'}) \; + \; \mathcal{M}_{t+1}^N(\boldsymbol{s})\right]. \tag{25}$$

Observe that $\mathcal{M}_{t+1}^N(\boldsymbol{s})$ is a sequence of 0-mean random variables whose variance converges to 0 as $N \to \infty$. Therefore

$$\mathcal{M}_{t+1}^N(\boldsymbol{s}) \xrightarrow{P} 0.$$

Assuming that as $N \to \infty$,

$$h(\boldsymbol{s'}, \boldsymbol{s}, c_t^N(a))\mu_t^N(\boldsymbol{s'}) \xrightarrow{P} h(\boldsymbol{s'}, \boldsymbol{s}, \bar{c}_t(a))\bar{\mu}_t(\boldsymbol{s'}), \tag{26}$$

we get

$$\mu_{t+1}^N(\boldsymbol{s}) \xrightarrow{P} \sum_{\boldsymbol{s'}\in\mathcal{S}} h(\boldsymbol{s'}, \boldsymbol{s}, \bar{c}_t(a))\bar{\mu}_t(\boldsymbol{s'}) \; =: \; \bar{\mu}_{t+1}(\boldsymbol{s}).$$

To see (26), observe that $h(\boldsymbol{s'}, \boldsymbol{s}, c_{t-1}^N(a))\mu_t^N(\boldsymbol{s'})$, can be re-written as

$$\begin{aligned} &h(\boldsymbol{s'}, \boldsymbol{s}, \bar{c}_t(a))\mu_t^N(\boldsymbol{s'}) + \\ &(h(\boldsymbol{s'}, \boldsymbol{s}, c_t^N(a)) - h(\boldsymbol{s'}, \boldsymbol{s}, \bar{c}_t(a)))\mu_t^N(\boldsymbol{s'}). \end{aligned}$$

Taking limit as $N \to \infty$ for the first term above, we get

$$h(\boldsymbol{s'}, \boldsymbol{s}, \bar{c}_t(a))\mu_t^N(\boldsymbol{s'}) \xrightarrow{P} h(\boldsymbol{s}, \boldsymbol{s'}, \bar{c}_t(a))\bar{\mu}_t(\boldsymbol{s'}),$$

since $\mu_t^N(\boldsymbol{s})$ converges to $\bar{\mu}_t$ in probability, by induction hypothesis. Moreover, the second term equals 0. To see this, observe that the second term above can be bounded by

$$\left|h(\boldsymbol{s'}, \boldsymbol{s}, c_t^N(a)) - h(\boldsymbol{s'}, \boldsymbol{s}, \bar{c}_t(a))\right|. \tag{27}$$

From (11), we can see that $c_t^N(a)$ is a continuous function of $\mu_t^N$ and $\mu_t^N \xrightarrow{P} \bar{\mu}_t$ by induction hypothesis. Therefore

$$c_t^N(a) \; \xrightarrow{P} \; \sum_{\boldsymbol{q}\in\mathcal{S}} \bar{\mu}_t(\boldsymbol{q})\mathbb{1}(\text{state } \boldsymbol{q} \text{ is infectious})\psi_{\tilde{a},a}\beta_{\tilde{a}} = \; \bar{c}_t(a).$$

Observe that $c_t^N$ and $\bar{c}_t$ are bounded (by definition). Moreover, $h$ is uniformly continuous in its arguments since it is a continuous function defined on a compact set. Thus, taking limit as $N \to \infty$, (27) converges to zero since $h(\boldsymbol{s'}, \boldsymbol{s}, c_t^N(a)) \xrightarrow{P} h(\boldsymbol{s'}, \boldsymbol{s}, \bar{c}_t(a))$ uniformly as $N \to \infty$.

## 6.2 Bias in the smaller model: A heuristic justification

In this section we compare infections in a large city with the appropriately scaled version of the smaller city and show that the smaller city underestimates the reported number of infections compared to the larger city. This bias is more pronounced when the initial number of infections are small, and becomes much less significant when this initial number infected becomes large. As our analysis suggests, it is the initial variability in the infection process that causes this bias. For simplicity of presentation, we use deterministic SIR model as used in Section 5.5.

First consider a city with total population $N$. Assume that there are $i_0$ infected in the city at time 0. Let $i_N(t)$ and $r_N(t)$ be the number infected and recovered at time $t$. Denote by $i^B(t)$ the number of infections in the associated branching process at time $t$ when starting from 1 infection at time 0.

We know that for large $N$ till time $t_N = \log_\rho \frac{N^*}{i_0}$ (recall that $N^* = N/\log N$) the city evolves approximately as a branching process. Therefore, each starting infection gives rise to an approximately independent tree of descendant infections till this time, and

$$i_N(t_N) = \sum_{j=1}^{i_0} i_j^B(t_N) + o(\rho^{t_N}),$$

where $i_j^B(t_N)$ for each $j \in [1, i_0]$ is distributed as an independent copy of $i^B(t)$. Recall from Example 3.1 that the proportion amongst the infected and the recovered stabilises in the branching process as time increases. Hence,

$$r_N(t_N) = \frac{r}{\beta - r} \sum_{j=1}^{i_0} i_j^B(t_N) + o(\rho^{t_N}),$$

and taking expectations,

$$\mathbb{E}\left(i_N(t_N)\right) = i_0 \mathbb{E}\left(i^B(t_N)\right) + o(\rho^{t_N}).$$

$\mathbb{E}\left(i_N(t_N+1)\right)$ can be estimated as follows,

$$\begin{aligned}
\mathbb{E}\left(i_N(t_N+1)\right) &= \mathbb{E}\left(\beta \sum_{j=1}^{i_0} i_j^B(t_N)\left(1 - \frac{1}{N}\Big(\sum_{j=1}^{i_0} i_j^B(t_N) + \sum_{j=1}^{i_0} r_j^B(t_N)\Big)\right)\right) + \mathbb{E}\left((1-r)\sum_{j=1}^{i_0} i_j^B(t_N)\right) + o(\rho^{t_N}) \\
&= \mathbb{E}\left(\beta \sum_{j=1}^{i_0} i_j^B(t_N)\left(1 - \frac{1}{N}\Big(\sum_{j=1}^{i_0} i_j^B(t_N) + \frac{r}{\beta - r}\sum_{j=1}^{i_0} i_j^B(t_N)\Big)\right)\right) + \frac{1}{N}(1-r) i_0 \mathbb{E}\left(i^B(t_N)\right) + o(\rho^{t_N}) \\
&= \mathbb{E}\left(\beta \sum_{j=1}^{i_0} i_j^B(t_N)\left(1 - \frac{\beta}{N(\beta - r)}\sum_{j=1}^{i_0} i_j^B(t_N)\right)\right) + \frac{1}{N}(1-r) i_0 \mathbb{E}\left(i^B(t_N)\right) + o(\rho^{t_N}) \\
&= (1 + \beta - r) i_0 \mathbb{E}\left(i^B(t_N)\right) - \frac{\beta^2 i_0}{N(\beta - r)} \mathbb{E}\left([i^B(t_N)]^2\right) - \frac{\beta^2 i_0(i_0-1)}{N(\beta - r)}\Big[\mathbb{E}\left(i^B(t_N)\right)\Big]^2 + o(\rho^{t_N}) \\
&= (1 + \beta - r) i_0 \mathbb{E}\left(i^B(t_N)\right) - \frac{\beta^2 i_0}{N(\beta - r)} \mathrm{Var}(i^B(t_N)) - \frac{\beta^2 i_0^2}{N(\beta - r)}\Big[\mathbb{E}\left(i^B(t_N)\right)\Big]^2 + o(\rho^{t_N}).
\end{aligned} \tag{28}$$

Now consider a larger city with total population $kN$ for $k > 1$. At time 0, we proportionately increase the initial infections to $ki_0$. Let $i_{kN}(t)$ and $r_{kN}(t)$ be the number of infections and recovered at time $t$ in this larger city.

Again, till time $t_{kN} = \log_\rho \frac{kN^*}{ki_0} = \log_\rho \frac{N^*}{i_0} = t_N$, the city evolves as a branching process. Therefore, each starting infection has an approximately independent infection tree till this time, and

$$i_{kN}(t_N) = \sum_{j=1}^{ki_0} i_j^B(t_N) + o(\rho^{t_N}),$$

where again each $i_j^B(t_N)$ is distributed as an independent copy of $i^B(t)$. Again,

$$r_{kN}(t_N) = \frac{r}{\beta - r} \sum_{j=1}^{ki_0} i_j^B(t_N) + o(\rho^{t_N}).$$

Following similar steps as earlier to estimate $\mathbb{E}\left(i_{kN}(t_N+1)\right)$, we have

$$\mathbb{E}\left(i_{kN}(t_N+1)\right) = k(1+\beta-r)i_0\mathbb{E}\left(i^B(t_N)\right) - \frac{\beta^2 i_0}{N(\beta-r)}\text{Var}(i^B(t_N)) - k\frac{\beta^2 i_0^2}{N(\beta-r)}\left[\mathbb{E}\left(i^B(t_N)\right)\right]^2 + o(\rho^{t_N}). \tag{29}$$

Scaling $k$ times the estimate of $i_N(t_N+1)$ in (28), we have

$$k\mathbb{E}\left(i_N(t_N+1)\right) = k(1+\beta-r)i_0\mathbb{E}\left(i^B(t_N)\right) - k\frac{\beta^2 i_0}{N(\beta-r)}\text{Var}(i^B(t_N)) - k\frac{\beta^2 i_0^2}{N(\beta-r)}\left[\mathbb{E}\left(i^B(t_N)\right)\right]^2 + o(\rho^{t_N}). \tag{30}$$

Observe that the second term $[\frac{\beta^2 i_0}{N(\beta-r)}\text{Var}(i^B(t_N))]$ on right-hand side of (29) is smaller than second term $[k\frac{\beta^2 i_0}{N(\beta-r)}\text{Var}(i^B(t_N))]$ on right-hand side of (30). Other terms are equal in (29) and (30). This suggests that $\mathbb{E}\left(i_{kN}(t_N+1)\right) > k\mathbb{E}\left(i_N(t_N+1)\right)$. Thus a large city with $kN$ population should have more infections at time $t_N+1$ as compared to the scaled smaller city with population $N$. Observe that the variance term in the output (30) of smaller model induces an error when we estimate the larger model by scaling the output of smaller model. Inductively, this error between the larger model and the scaled smaller model can be seen to hold till $\text{O}(log_\rho N)$ time. This is also evident in Figure 2.

Also observe that in the estimate of $k\mathbb{E}\left(i_N(t_N+1)\right)$ for smaller city, the ratio of the third term to the second term for smaller city is $i_0\frac{\left[\mathbb{E}\left(i^B(t_N)\right)\right]^2}{\text{Var}(i^B(t_N))}$, that is the third term becomes dominant over the second term as $i_0$ increases. As the third term is the same for both the larger city and the scaled smaller city, therefore as $i_0$ (initial number of infections) increases, the difference between the number of infections in the larger city and the scaled smaller city becomes smaller, suggesting that scaling the smaller city provides a good approximation to the larger city when $i_0$ is large. This is also evident from Figure 1.

## 6.3 Input data for numerical experiments

In this section for completeness, we summarise the city statistics and disease related parameters used in our numerical experiments. This data is similar to that reported in [1] where it is more fully justified.

Recall that for the numerical experiments, we first create a synthetic city that closely models the actual Mumbai. A synthetic model is set to match the numbers employed, numbers in schools, commute distances, etc in Mumbai. Tables 3 and 4 show the household size distribution and school size distribution in the model. Fraction of working population is set to 40.33%. Workplace size distribution can be seen in [1].

For Figures 1 to 9, we consider **one community space** and the whole population is considered to be living in non-slums. For Figure 10, we consider **48 community spaces**, to model the 24 administrative wards in Mumbai further divided into slums and non-slums. Mumbai slums are densely crowded. This leads to difficulties in maintaining social distancing and increases the transmission rate between the infected and the uninfected. We account for this by selecting a higher community transmission rate in the slums (2 times the non-slums). This leads to slum and non-slum prevalence that closely matches the seroprevalence data observed in July 2020 (see [19]). Table 1 summarises the non-slum population age distribution and Table 2 summarises the non-slum population age distribution.

Transmission rates are set as follows : $\beta_h = 1.227$, $\beta_w = 0.919$, $\beta_s = 1.82$, $\beta_c = 0.233$. See, Table 5 and Figure 13 for details of the disease progression. Symptomatic patients are assumed to be more infectious during the symptomatic period than during the pre-symptomatic infective stage (1.5 times more infectious in our model).

Table 6 summarizes the details of interventions modelled in the simulator. In Figures 1 to 6, no interventions were implemented. In Figures 7 to 9, intervention was home quarantine from day 40.

Table 1: Non-slum population age distribution

| Age (yrs) | Fraction of population |
|---|---|
| 0-4 | 0.0757 |
| 5-9 | 0.0825 |
| 10-14 | 0.0608 |
| 15-19 | 0.0669 |
| 20-24 | 0.0705 |
| 25-29 | 0.0692 |
| 30-34 | 0.0777 |
| 35-39 | 0.0716 |
| 40-44 | 0.0762 |
| 45-49 | 0.0664 |
| 50-54 | 0.0795 |
| 55-59 | 0.0632 |
| 60-64 | 0.0560 |
| 65-69 | 0.0380 |
| 70-74 | 0.0227 |
| 75-79 | 0.0136 |
| 80+ | 0.0094 |

Table 2: Slum population age distribution

| Age (yrs) | Fraction of population |
|---|---|
| 0-4 | 0.0757 |
| 5-9 | 0.0825 |
| 10-14 | 0.0875 |
| 15-19 | 0.0963 |
| 20-24 | 0.0934 |
| 25-29 | 0.0917 |
| 30-34 | 0.0921 |
| 35-39 | 0.0849 |
| 40-44 | 0.0606 |
| 45-49 | 0.0529 |
| 50-54 | 0.0632 |
| 55-59 | 0.0503 |
| 60-64 | 0.0327 |
| 65-69 | 0.0221 |
| 70-74 | 0.0073 |
| 75-79 | 0.0044 |
| 80+ | 0.0021 |

For Figure 10, interventions were introduced based on the actual interventions as happened in the city. Specifically, no intervention for first 33 days, i.e. till 20 March, 2020 (Simulation starts on 13th Feb, 2020). Lockdown from March 20 to May 17. Masks are active from 9 April, 2020 (from day 53). Rules for higher social distancing of elderly are enforced from May 1, 2020 (from day 75). Schools are closed throughout the pandemic period except in the earlier no intervention period. Other interventions such as home quarantine and case isolation are also implemented post the lockdown. Attendance schedules at workplaces after May 17 are set as follows : 5% in May 18-31, 20% in all of June , 33% in all of July, and 50% thereafter. These numbers were selected keeping in mind the prevalent restrictions and observing the transportation and the Google mobility data. Compliance levels are set at 60% in non-slums and 40% in slums. These appeared reasonable and these matched the fatality data early on in the pandemic (till June 2020).

Table 3: Household size distribution

| Household size | Fraction of households |
|---|---|
| 1 | 0.0485 |
| 2 | 0.1030 |
| 3 | 0.1715 |
| 4 | 0.2589 |
| 5 | 0.1819 |
| 6 | 0.1035 |
| 7-10 | 0.1165 |
| 11-14 | 0.0126 |
| 15+ | 0.0035 |

Table 4: School size distribution

| School size | Fraction of schools | School size | Fraction of schools |
|---|---|---|---|
| 0-100 | 0.0185 | 500-600 | 0.0889 |
| 100-200 | 0.1284 | 600-700 | 0.063 |
| 200-300 | 0.2315 | 700-800 | 0.0481 |
| 300-400 | 0.2315 | 800-900 | 0.0408 |
| 400-500 | 0.1574 | 900+ | 0 |

Table 5: Disease progression parameters

| Parameter description | Values |
|---|---|
| Incubation Period | Gamma distributed with shape 2 and scale 2.29 |
| Asymptomatic Period | Exponentially distributed with mean duration 0.5 of a day |
| Symptomatic Period | Exponentially distributed with mean duration of 5 days |
| Hospitalisation Period | 8 days |
| Critical Period | 8 days |

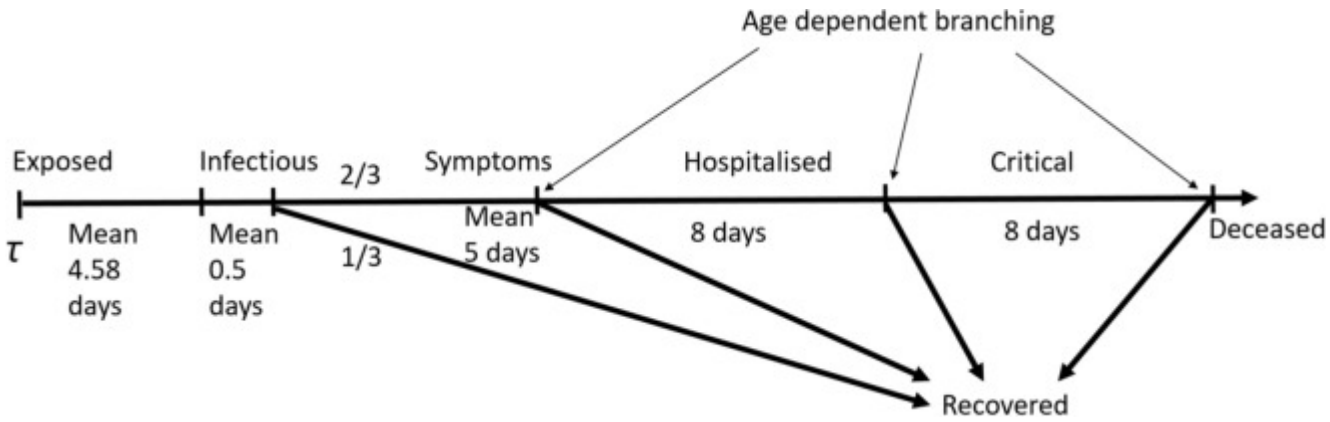


Figure 13: A simplified model of COVID-19 progression.

Table 6: Interventions as implemented in the simulator

| Intervention | Description |
|---|---|
| No intervention | Business as usual |
| Lockdown | For compliant households, household rates are doubled, no workplace interactions except for 25% leakage (for essential services), community interactions reduce by 75%. For non-compliant households, workplace interactions only have a leakage of 25%, community interactions are unchanged, and household interactions increase by 50% |
| Case Isolation | Compliant symptomatic individuals stay at home for 7 days, reducing non-household contacts by 75%. Household contacts remain unchanged. |
| Home Quarantine | Following identification of a symptomatic case in a compliant household, all household members remain at home for 14 days. Household contact rates double, contacts in the community reduce by 75% |
| Social distancing of the elderly | All compliant individuals over 65 years of age reduce their community interactions by 75% |
| Schools and colleges closed | Self explanatory |
| Masks | Reduce community transmission by 20% |