Agent based simulators for epidemic modelling: Simulating larger models using smaller ones

Daksh Mittal
Columbia Business School
New York, USA
dmittal27@gsb.columbia.edu

Sandeep Juneja
TIFR
Mumbai, India
juneja@tifr.res.in

Shubhada Agrawal
Georgia Institute of Technology
Atlanta, USA
shubhadaiitd@gmail.com

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 as well as detailed governmental mobility restrictions. 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 dig deeper into the underlying probabilistic structure of a generic, locally detailed ABS for epidemiology to arrive at modifications that allow smaller models (models with less number of agents) to give accurate statistics for larger ones, thus substantially speeding up the simulation. 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 disease 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. Once enough people have been infected, the future evolution of the epidemic 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 apply our algorithm for Covid-19 epidemic in a city and theoretically support the proposed algorithm through an asymptotic analysis where the population size increases to infinity. We develop nuanced coupling based arguments to show that the epidemic process is close to the branching process early on in the simulation.

KEYWORDS

Agent-based simulators, Epidemiological modelling, Multi-type branching process, Mean field limit, Asymptotic analysis

1 INTRODUCTION

Agent-based simulators (ABS) are a popular tool in epidemiology (See [11], [5]). In this paper we focus on ABS used to model epidemics such as Covid-19 in cities and we illustrate our methodological contributions through the use of ABS for modelling Covid evolution in the city of Mumbai. As is apparent from the paper, the underlying ideas are valid more generally for any epidemic in a large city that may have an initial exponential growth phase that tapers down when sufficient population is infected (See [4], [6]).

As is well known, in an ABS in epidemic modelling of a city, a synthetic copy of it is constructed on a computer that captures the population interaction spaces and detailed disease spread as well as disease spreading interactions as they evolve in time. Typically, each individual in the city is modelled as an agent, so that total number of agents equal the total city 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 city census data and are distributed to match its geography. Government policies, such as partial, location specific lockdowns 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. lead to mobility reduction and 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 [12]). 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 [5] for an ABS used for UK and USA related studies specific to COVID-19, [7] for a COVID-19 study on Sweden, [1], [12] for a study on Bangalore and Mumbai in India. See, e.g., [8] and [11] 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 parameters such as transmission rates, infectiousness of new variants are fit to the observed infection data over relevant time horizons, requiring many computationally demanding simulations.

Key contributions: We 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 a 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. We numerically observe similar results with a smaller half-million population model, although further reduction leads to increase in errors as the branching process phase is very small and mean field approximations break down. Mumbai is a densely populated city. To check for validity of our approach to less dense cities, we reduce the parameters that capture interactions and numerically observe that the proposed approach works well even in substantially sparser cities.

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 $\frac{\log(N/i\log N)}{\log(\rho)}$, where ρ is the exponential epidemic growth rate in early stages and i denotes the number exposed at time zero, the epidemic process is well approximated by an associated branching process. Our analysis relies on developing a nuanced coupling between the epidemic and the branching process during this period. Then, after time $\log(\epsilon N)/\log(\rho)$ for any $\epsilon > 0$, and large N, we show that 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.

Structure of remaining paper: 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. We first conduct our analysis in a simpler set-up where the coupling between the epidemic and the branching process and the related analysis is easier. A more general analysis, and a more intricate coupling is analysed in [13], where we also provide detailed proofs for the results of Section 5. In Section 6 we demonstrate the performance of our algorithm for less dense cities and smaller half million and one lakh models. In section 7 we conclude.

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 [10]), 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. In [13], 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).

Additionally, 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.

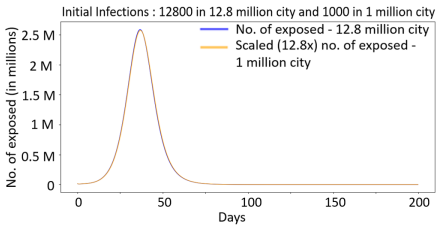

Figure 1: Scaled no. exposed in the smaller model match the larger model when we start with large, 12800 infections.

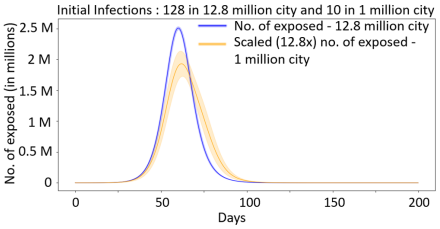

Figure 2: Scaled no. exposed in smaller model do not match larger model when we start with few, 128 infections.

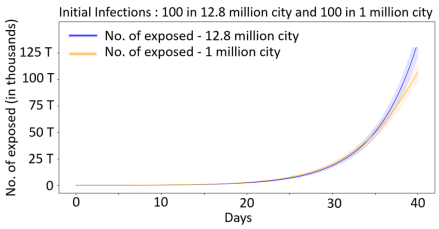

Figure 3: Smaller and larger model are essentially identical initially when we start with same no. of few, 100 infections.

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 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 ρ 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 5.2). 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 reasonable, as in modelling Mumbai, our calibration exercise had set the day zero to February 13, 2020 (see, [10]), 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

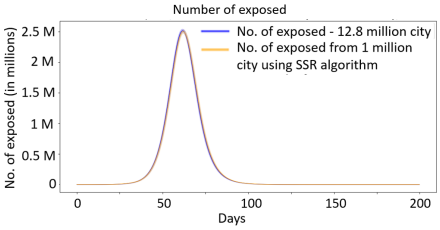

Figure 4: Shift and scale smaller model (no. of exposed) matches the larger model under no intervention scenario.

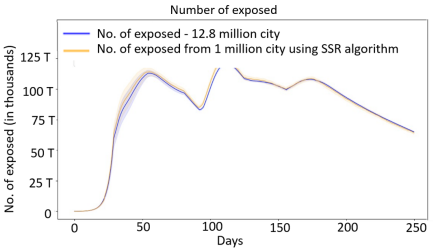

Figure 5: Shift-scale-restart smaller model match the larger one under real world interventions over 250 days.

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. In [13] we note that the restarted path need not use common random numbers. Even if it is generated independently, we get very similar output.

Figure 4 compares the number exposed 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. Numerical parameters for the no-intervention scenario for Mumbai are described in [13]. Figure 5 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 [1] using similar parameters, for 250 days. In this experiment, smaller and larger city evolves similarly till day 22. Intervention is at day 33, therefore smaller city is restarted with intervention at day 22.5 and scaled estimates from day 11.5 of the restarted simulation are appended at day 22 of the initial run (i.e. shifted by 10.5 days). We see that the smaller model faithfully replicates the larger one with negligible error. Similar results hold for other health statistics such as the number hospitalised and the number of cumulative fatalities, and for larger models that include variants (see [13]).

3 AGENT BASED SIMULATOR

We informally and briefly describe the main drivers of the dynamics in our model. Details are given in [13], [1].

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 model proceeds in discrete time steps of constant width Δt (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. $\lambda_n(t)$ is a 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. Componentwise calculations of these rates are detailed in the [13] (Section 3.1).

In the next Δt 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 Δt for the population already afflicted by the virus. The probabilistic dynamics of disease progression as well as implementation of public health safety measures are briefly summarized in [13] (Section 3) under simplified assumptions (underlying model is similar to that in [5] and [1]). 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.

4 SHIFT-SCALE-RESTART ALGORITHM

Let $\mu_0(N)$ denote the initial distribution of the infected population at time zero in our 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 $I_0 = 100$ 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). Algorithm 1 summarises the simulation dynamics.

Scaling the model: 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 β_h, β_s, β_w and β_c are unchanged (recall that each transmission rate of an infected individual in each interaction space is the expected number of 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).

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.

When we initiate both the larger as well as the smaller city with a same few and well spread infections, the disease spreads similarly in homes, schools and workplaces. To understand the disease spread through communities, for simplicity assume that there is a single community. 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 susceptible population. This is true even when there are more communities and for a general distance function f (see [13] Section 3 for role of f in community infection rates). Therefore, early on in the simulation, the total number 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 $N^* = N/i \log N$. Here $\log_\rho m = \log m / \log \rho$ for any $m \in \mathfrak{R}^+$ and ρ denotes the initial infection exponential growth rate.)

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 2 achieves this and is graphically illustrated in Figure 6. In the contrafactual no intervention scenario where the population intermingling behaviour does not change even as the disease spreads through it, we can just use one simulation and need not restart the simulation.

As we see empirically, and as is suggested by Proposition 5.4, in Algorithm 2, evolution of the infection process after time $t_{x/k}$, given the state at that time, is more or less deterministic (even though the process may be close to the associated branching process at this time).

Algorithm 2 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: As is Step 6, Algorithm 1.

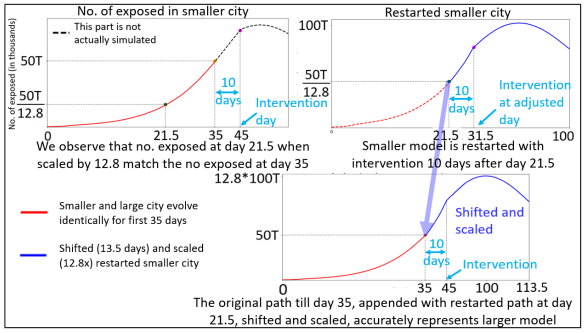

Figure 6: Shift, scale and restart algorithm

5 ASYMPTOTIC ANALYSIS

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. A more intricate analysis involving interactions at homes, schools and workplaces is given in [13]

We briefly review the standard multi-type branching process in [13] (See [2], [9] for detailed review). In the following analysis, we define an epidemic process in Section 5.1. Further, we define a multi-type super-critical branching process tailored to the epidemic process in Section 5.2. We describe the coupling between the two processes in Section 5.3. We then state the results demonstrating the closeness of the epidemic and the branching process in the early disease spread phase in Section 5.4. The results demonstrating the closeness of the epidemic process to its mean field limit once the epidemic process has grown are given in Section 5.5.

5.1 Epidemic process dynamics

Notation: 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 $\mathcal{D}$. For simplicity, we ignore possible reinfections, although incorporating them would not alter our conclusions.

• Each individual has some characteristics that are assumed to remain unchanged throughout the epidemic regardless of it'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 π_a is independent of N as $N \to \infty$.

• Hence, each individual at any time may be classified by a type $\mathbf{s} = (a, d)$, 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, $\mathcal{S} \backslash \mathcal{U}$ 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 $\mathbf{s}$ at time t by $X_t^N(\mathbf{s})$ and set $\mathbf{X}_t^N = (X_t^N(\mathbf{s}) : \mathbf{s} \in \mathcal{S} \setminus \mathcal{U})$. Then, $\mathbf{X}_t^N \in \mathbb{Z}^{+\eta}$.

• Let $A_t^N = \sum_{\mathbf{s} \in \mathcal{S} \backslash \mathcal{U}} X_t^N(\mathbf{s})$ denote the total number of affected individuals in the system N at or before time t.

Dynamics: At time zero, for each N, $\mathbf{X}_0^N$ 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 $\mathbf{X}_0^N = \mathbf{X}_0$ for all N.

Given $\mathbf{X}_t^N$, $\mathbf{X}_{t+\Delta t}^N$ is arrived at through two mechanisms. (For the ease of notation we will set $\Delta t = 1$.) i) 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 ii) 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 β_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 β_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 individuals living in a dense

region are more likely to infect other individuals 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 gets exposed at time $t+1$. Each susceptible individual with characteristic $\tilde{a} \in \mathcal{A}$ who gets exposed at time t, increments $\mathbf{X}^N_{t+1}(\tilde{a}, exposed)$ 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 $\mathbf{s} \in \mathcal{S} \setminus \mathcal{U}$, some disease state other than susceptible, at time t transitions to some other state $\mathbf{q}$ at time $t+1$ with the transition probability $P(\mathbf{s}, \mathbf{q})$ in one time step. The probability $P(\mathbf{s}, \mathbf{q})$ is independent of time t and N. If the transition happens, $X^N_{t+1}(\mathbf{s})$ is decreased by 1 and $X^N_{t+1}(\mathbf{q})$ is increased by 1.

5.2 Associated branching process dynamics

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

Dynamics: At time zero $\mathbf{B}_0 = \mathbf{X}_0$. 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 $\mathbf{s} \in \mathcal{S} \setminus \mathcal{U}$, that is, in disease state other than susceptible at time t, transitions to some other disease state $\mathbf{q}$ at time $t+1$ with probability $P(\mathbf{s}, \mathbf{q})$. If the transition happens, $B_{t+1}(\mathbf{s})$ is decreased by 1 and $B_{t+1}(\mathbf{q})$ is increased by 1.

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

The expected offsprings matrix $K \in \mathfrak{R}^{+\eta \times \eta}$ for $\{\mathbf{B}_t\}$: Let each entry $K(\mathbf{s}, \mathbf{q})$ of K denote the expected number of type $\mathbf{q}$ offspring of a single type $\mathbf{s}$ 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 $\mathbf{s} \in \mathcal{S} \setminus \mathcal{U}$, 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,

$$\begin{aligned} &K(\mathbf{s}, \mathbf{q}) = P(\mathbf{s}, \mathbf{q}) + \pi_{\tilde{a}} \psi_{a,\tilde{a}} \beta_a \ \forall\ \mathbf{s} = (a, d) \in \tilde{\mathcal{H}}, \mathbf{q} = (\tilde{a}, \tilde{d}) \in \mathcal{E} \\ &K(\mathbf{s}, \mathbf{q}) = P(\mathbf{s}, \mathbf{q}) \text{ otherwise.} \end{aligned} \tag{1}$$

It follows that

$$\mathbb{E}(\mathbf{B}_{t+1}) = K^T \mathbb{E}(\mathbf{B}_t) = (K^T)^{t+1} \mathbb{E}(\mathbf{B}_0). \tag{2}$$

Theorem 5.2 for multi-type branching processes holds under a standard assumption that the expected offspring matrix K is irreducible (for more details, see [13]). However, K as defined in (1) is not irreducible. Lemma 5.1 below sheds further light on the structure of K. Theorem 5.2 observes that standard conclusions continue to hold for the branching process $\{\mathbf{B}_t\}$ associated with epidemic processes $\{\mathbf{X}^N_t\}$.

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 $K_1 \in \mathfrak{R}^{+\hat{\eta} \times \hat{\eta}}, C \in \mathfrak{R}^{+(\eta-\hat{\eta}) \times (\eta-\hat{\eta})}$ and $M \in \mathfrak{R}^{+(\eta) \times (\eta-\hat{\eta})}$ such that*

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

where $K_1 \in \mathfrak{R}^{+\hat{\eta} \times \hat{\eta}}$ 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 its 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 $\lim_{t\to\infty} K^t/\rho^t = \mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ and $\mathbf{v}$ are the strictly positive right and left eigenvectors of K corresponding to eigenvalue $\rho = \rho(K)$ such that $\mathbf{u}^T\mathbf{v} = 1$ and $\sum_{i=1}^{\eta} u(i) = 1$.*

In addition, $\mathbf{B}_t/\rho^t \xrightarrow{P} W\mathbf{v}$ as $t \to \infty$, where W is a non-negative random variable such that $P\{W > 0\} > 0$ iff $B_0(\mathbf{s}) \neq 0$ for some $\mathbf{s} \in \mathcal{H}$ and $\mathbb{E}(W|\mathbf{B}_0 = \mathbf{e_i}) = u(i)$ 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 $\mathbb{E}(B_1(j) \log B_1(j) | \mathbf{B}_0 = \mathbf{e_i}) < \infty$ for all $1 \leq i, j \leq \eta$ holds because the offsprings for our branching process have a Poisson distribution.

5.3 Coupling epidemic and branching process

• Recall that at time zero, $\mathbf{B}_0 = \mathbf{X}_0$. 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 $(\tilde{a}, exposed)$.

• 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 $\log_\rho N^* = \log N^*/\log \rho$ and ρ denotes the exponential growth rate of the branching process.

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

5.4 The initial branching process phase

Following result shows that epidemic process is close to the multi-type branching process till time $\log_\rho N^*$.

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

$$\sup_{t \in [0, \log_\rho(N^*/\sqrt{N})]} \left| X_t^N(\mathbf{s}) - B_t(\mathbf{s}) \right| \xrightarrow{P} 0. \tag{3}$$

$$\sup_{t \in \left[0, \log_\rho N^*\right]} \left| \frac{X_t^N(\mathbf{s})}{\rho^t} - \frac{B_t(\mathbf{s})}{\rho^t} \right| \xrightarrow{P} 0. \tag{4}$$

Result (3) was earlier shown for SIR setting in [3]. We extend this result to general models and also prove the result (4) in this more general setting.

Recall from Theorem 5.2 that, $\mathbf{B}_t/\rho^t \xrightarrow{P} W\mathbf{v}$ as $t \to \infty$, where W is a non-negative random variable representing the intensity of branching process and $\mathbf{v} \in \mathfrak{R}^{+\eta}$ is the left eigenvector corresponding to eigenvalue ρ of matrix K. Therefore, initially (till time $\log_\rho N^*$) epidemic process grows exponentially at rate ρ, 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 5.4. *For $t_N \to \infty$ as $N \to \infty$ and $\limsup_{N\to\infty} \frac{t_N}{\log_\rho N^*} < \infty$, then for all $\mathbf{s} \in \mathcal{S} \setminus \mathcal{U}$:*

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

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. In [13] we describe the results for some popular compartmental models using Theorem 5.3 and 5.4.

5.5 Deterministic phase

As Theorem 5.3 and Proposition 5.4 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 5.5. To this end we need Assumption 1.

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

Assumption 1. *There exists a random distribution $\bar{\mu}_0(W) \in \mathfrak{R}^{+(\eta+1)}$ that is independent of N such that $\mu_0^N \xrightarrow{P} \bar{\mu}_0(W)$ as $N \to \infty$.*

Observe that $\bar{\mu}_0(W)$ above is path dependent in that it depends on the random variable W. The above assumption is seen to hold empirically.

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. In our setup, $c_t^N(a)$ equals

$$\sum_{\mathbf{q}=(\tilde{a},\tilde{d})\in\mathcal{S}\setminus\mathcal{U}} \mu_t^N(\mathbf{q}) \mathbb{1}(\text{type } \mathbf{q} \text{ is infectious}) \psi_{\tilde{a},a}\beta_{\tilde{a}}, \tag{5}$$

where $\mathbb{1}$ denotes the indicator function. For each individual $n \le N$, let S_n^t denote its type at time t. Then, for $\mathbf{s} = (a, d) \in \mathcal{S}$

$$\mathbb{P}\left(S_n^{t+1} = \mathbf{s} | \mu_t^N\right) = h(S_n^t, \mathbf{s}, c_t^N(a)),$$

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.

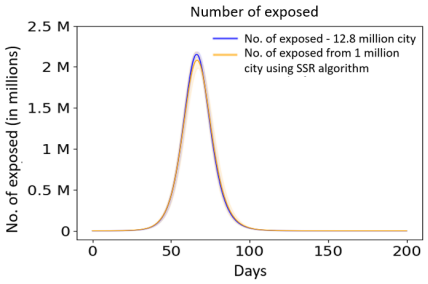

Figure 7: Shift and scale smaller model (no. of exposed) matches the larger model under no intervention scenario ($\beta_c^{new} = \beta_c/10$).

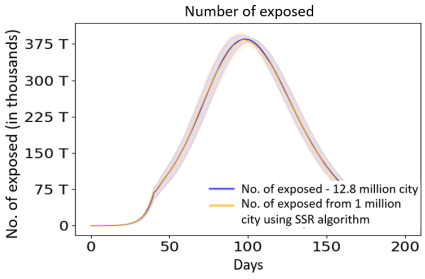

Figure 8: Shift-scale-restart smaller model (no. of exposed) matches the larger model under intervention ($\beta_c^{new} = \beta_c/10$).

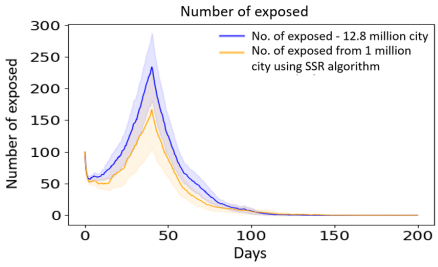

Figure 9: Shift-scale smaller model is within 20% of the larger model under intervention. $\beta_c^{new} = \beta_c/20$, $(\beta_h^{new}, \beta_w^{new}, \beta_s^{new}) = (\beta_h, \beta_w, \beta_s)/4$.

Recall that from Assumption 1 we have defined $\bar{\mu}_0(W) \in \mathfrak{R}^{+(\eta+1)}$ such that $\mu_0^N \xrightarrow{P} \bar{\mu}_0(W)$ as $N \to \infty$. Define $\bar{\mu}_t(W) \in \mathfrak{R}^{+(\eta+1)}$ such that for all $t \in \mathbb{N}$, $\mathbf{s} = (a, d) \in \mathcal{S}$,

$$\bar{\mu}_t(\mathbf{s}, W) := \sum_{\mathbf{s}' \in \mathcal{S}} \bar{\mu}_{t-1}(\mathbf{s}', W) h(\mathbf{s}'\mathbf{s}, \bar{c}_{t-1}(a, W)), \tag{6}$$

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

Theorem 5.5. *Under Assumption 1 and for $t \in \mathbb{N}$, $\mu_t^N \xrightarrow{P} \bar{\mu}_t(W)$ as $N \to \infty$.*

In particular, if $\bar{\mu}_t$ 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 $N * \bar{\mu}_t$ and that of a larger model is approximately $kN * \bar{\mu}_t$. Thus, the larger model infection process can be approximated by the smaller model infection process by scaling it by k.

6 FURTHER EXPERIMENTS

A. SSR on sparser cities: While our experiments are designed for Mumbai, a very dense city, it is reasonable to wonder if the results would hold for less dense cities or regions. In this section we report the experimental results for the sparser cities that are created by reducing the contact and hence the transmission rates between individual in different interaction spaces. Specifically, we consider two reasonable scenarios:

• Scenario 1: The community beta is kept at one tenth the level of the community beta for Mumbai considered in our earlier experiments in Section 2. Figure 7, 8 shows that the result using SSR algorithm matches the larger model in the no-intervention scenario and the scenario with an intervention of home quarantine after 40 days. Although we did not conduct elaborate experimentation, similar results should be true under more general and extensive interventions.

• Scenario 2: Beta community is reduced by a factor of 20, beta home, school and workplace by a factor 4. Under no intervention scenario the result using SSR algorithm matched the larger model (See [13]). However, it was observed that even in the no-intervention scenario, the number of infections were very low after 200 days. Under home quarantine after 40 days, the exposed cases from the smaller model are within 20% of the larger model although both are more-or-less negligible numbers, the highest less than 250.

B. How small can the smaller model be: In the earlier experiments in Section 2, we had compared the results of a 12.8 million city to the results from using SSR on a smaller 1 million city. A natural question to ask is how small the city can be for SSR to continue to be accurate. We found that results using SSR algorithm on 500,000 city were accurate while 100,000 city poorly matched the larger model. Experiment details are given in [13].

7 CONCLUSION

In this paper we considered large ABS models used to model epidemic spread in a city or a region. These models are of great use in capturing details of population types, their behaviour, time varying regulatory instructions, etc. A key drawback of ABS models is that computational time can quickly become prohibitive as the city size and the simulated time increases. In this paper we proposed a shift-scale-restart algorithm that exploits the underlying probabilistic structure and allows smaller cities to provide extremely accurate approximations to larger ones using much less computational time. We supported our experiments and the algorithm through asymptotic analysis where we showed that initial part of the epidemic process is close to a corresponding multi-type branching process and very quickly the evolution of the epidemic process is well approximated by its mean field limit. The fact that in multi-type branching process the distribution of population across types stabilizes quickly helped us shift sample paths across time without loss of accuracy. To show closeness of the epidemic process to a branching process we developed careful coupling arguments.

REFERENCES

[1] Shubhada Agrawal, Siddharth Bhandari, Anirban Bhattacharjee, Anand Deo, Narendra M. Dixit, Prahladh Harsha, Sandeep Juneja, Poonam Kesarwani, Aditya Swamy, Krishna Preetam Patil, Nihesh Rathod, Ramprasad Saptharishi, Sharad Shriram, Piyush Srivastava, Rajesh Sundaresan, Nidhin Koshy Vaidhiyan, and Sarath Yasodharan. 2020. 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) (2020), 809–847.

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

[3] Tom Britton and Etienne Pardoux. 2019. *Stochastic Epidemic Models with Inference.* Springer, Switzerland.

[4] Jean-Paul Chretien, Steven Riley, and Dylan B George. 2015. Mathematical modeling of the West Africa Ebola epidemic. *eLife* 4 (2015), e09186.

[5] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, Han Fu, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Lucy C Okell, Sabine Van Elsland, Hayley Thompson, Robert Verity, Erik Volz, Haowei Wang, Yuanrong Wang, Patrick GT Walker, Caroline Walters, Peter Winskill, Charles Whittaker, Christl A Donnelly, Steven Riley, and Azra C Ghani. 2020. *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.

[6] Enrique Frias-Martinez, Graham Williamson, and Vanessa Frías-Martínez. 2011. An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. 57–64.

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

[8] 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. 2008. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences* 105, 12 (2008), 4639–4644.

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

[10] Prahladh Harsha, Sandeep Juneja, Daksh Mittal, and Ramprasad Saptharishi. 2020. *COVID-19 Epidemic in Mumbai: Projections, full economic opening, and containment zones versus contact tracing and testing: An Update.* Technical Report. TIFR Mumbai, India.

[11] Elizabeth Hunter, Brian Mac Namee, and John D Kelleher. 2017. A taxonomy for agent-based models in human infectious disease epidemiology. *Journal of Artificial Societies and Social Simulation* 20(3) (2017), 1–2.

[12] Sandeep Juneja and Daksh Mittal. 2021. *Modelling the Second Covid-19 Wave in Mumbai.* Technical Report. TIFR Mumbai, India.

[13] Daksh Mittal, Sandeep Juneja, and Shubhada Agrawal. 2022. Agent based simulators for Covid-19: Simulating larger models using smaller ones. https://arxiv.org/abs/2209.02887