arXiv:2012.12839v2 [cs.MA] 24 Dec 2020

Cohorting to isolate asymptomatic spreaders: An agent-based simulation study on the Mumbai Suburban Railway

Alok Talekar[1], Sharad Shriram[2], Nidhin Vaidhiyan[2], Gaurav Aggarwal[1], Jiangzhuo Chen[3], Srini Venkatramanan[3], Lijing Wang[3], Aniruddha Adiga[3], Adam Sadilek[1], Ashish Tendulkar[1], Madhav Marathe[3], Rajesh Sundaresan[2,4] and Milind Tambe[1]

[1] Google Inc.
[2] Indian Institute of Science, Bangalore
[3] University of Virginia
[4] Strand Life Sciences

ABSTRACT

The Mumbai Suburban Railways, *locals*, are a key transit infrastructure of the city and is crucial for resuming normal economic activity. Due to high density during transit, the potential risk of disease transmission is high, and the government has taken a wait and see approach to resume normal operations. To reduce disease transmission, policymakers can enforce reduced crowding and mandate wearing of masks. *Cohorting* – forming groups of travelers that always travel together, is an additional policy to reduce disease transmission on *locals* without severe restrictions. Cohorting allows us to: (*i*) form traveler bubbles, thereby decreasing the number of distinct interactions over time; (*ii*) potentially quarantine an entire cohort if a single case is detected, making contact tracing more efficient, and (*iii*) target cohorts for testing and early detection of symptomatic as well as asymptomatic cases. Studying impact of cohorts using compartmental models is challenging because of the ensuing representational complexity. Agent-based models provide a natural way to represent cohorts along with the representation of the cohort members with the larger social network. This paper describes a novel multi-scale agent-based model to study the impact of cohorting strategies on COVID-19 dynamics in Mumbai. We achieve this by modeling the Mumbai urban region using a detailed agent-based model comprising of 12.4 million agents. Individual cohorts and their inter-cohort interactions as they travel on locals are modeled using local mean field approximations. The resulting multi-scale model in conjunction with a detailed disease transmission and intervention simulator is used to assess various cohorting strategies. The results provide a quantitative trade-off between cohort size and its impact on disease dynamics and well being. The results show that cohorts can provide significant benefit in terms of reduced transmission without significantly impacting ridership and or economic & social activity.

KEYWORDS

Covid-19, Public Transportation, Cohorts, Social Simulation

1 INTRODUCTION

COVID-19 is the worst public health disaster in the 21st century, second only to the 1918 pandemic in the last 100 years. An estimated 36 million confirmed cases and over a million fatalities all over the world have been reported as of 8th October 2020. The response to the pandemic has varied across countries, but given that no pharmaceutical interventions were available, almost all countries instituted significant social distancing measures to control the spread. As an extreme measure, countries had enforced lockdowns to reduce mobility in varying capacity. Data clearly shows that urban regions were impacted far earlier than rural regions. Population density and the resulting social interactions are important driving factors.

The pandemic has had significant impact on India thus far. The Government of India acted early and instituted a nation-wide lockdown. The lockdown had a significant effect on urban and regional mobility patterns[1]. Like other parts of the world, Indian cities, many of which have some of the highest population densities, have been affected significantly by the pandemic. Mumbai, one of the largest metropolises in the world with 12.4 million inhabitants in the Brihanmumbai area[2] is the economic hub of the country. Mumbai has the highest population density in the country (over $32,300\ people/km^2$), and has been significantly impacted by the pandemic. *Locals* had a daily ridership of over 8.2 million passengers before the lockdown – this represents 20-40% of city's total population[3]. *Locals*, the Mumbai suburban trains, are a lifeline of the city and play a central role in the vibrant social and economic activities of the metropolis, and resumption of service is critical for the resumption of normal economic and social activities in Mumbai. However, this can potentially contribute to a surge in disease spread [18] since the locals and the stations are densely packed during normal working hours. This creates a dense social network, ideal for airborne disease transmission, with long distance edges that can quickly spread the disease across the metropolis. The source code for the simulator is open-sourced and available at https://github.com/cni-iisc/epidemic-simulator/tree/mumbai_local

1.1 Our contributions and novelty

This is the first paper to explore and quantify the public health benefit of *cohorting* strategies while facilitating increased mobility and economic activity. Policy makers are looking for ways to resume the service of *locals* in Mumbai and this work enables just that. Cohorts [4] are motivated by two central hypotheses. (*i*) Two social interaction networks with the same interaction density can have varying impacts on disease dynamics; specifically, for small social networks, locally dense but globally weakly connected networks might be better at controlling the spread. (*ii*) It is easy to do contact

[1] See https://ourworldindata.org/covid-mobility-trends.. Our results estimate that there was over 70% drop in mobility in India [2]
[2] This is the area under the jurisdiction of the Brihanmumbai Municipal Corporation.
[3] The estimate accounts for multiple rides by the same individual
[4] An illustrative examples is shown on this video

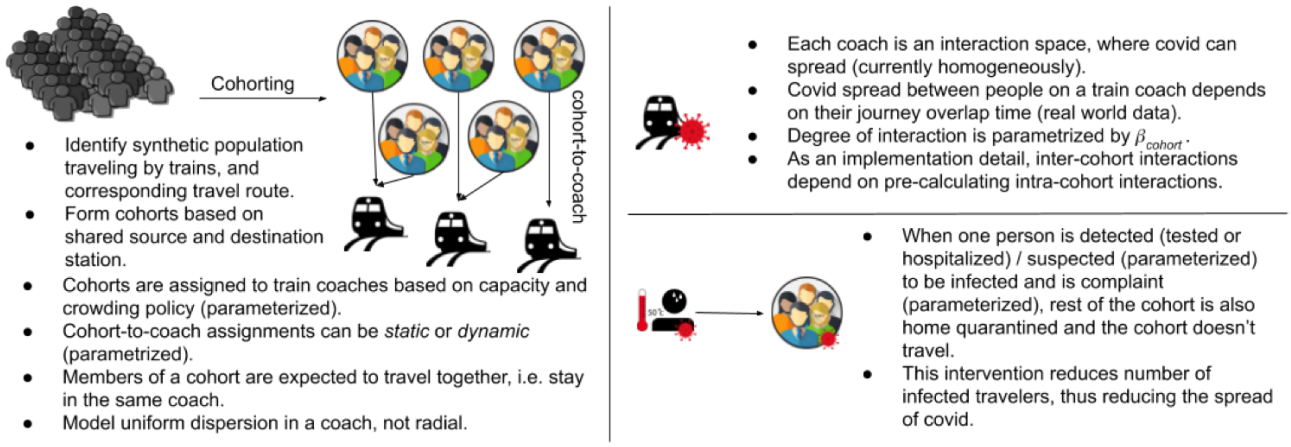


Figure 1: Schematic representation of our agent-based modeling of local trains.

tracing when travelers form cohorts; early and efficient contact tracing followed by quarantining and monitoring of individuals who came in close contact with the infected individuals can help mitigate the disease spread. The first hypothesis is the rationale for *social bubbles* [24] in the context of restricted social interactions. From this perspective, cohorts can be thought as *traveler bubbles*. They change the dynamic social network structure by creating locally dense sub-graphs but reducing interactions between these dense sub-graphs. The second hypothesis is that cohorting helps us implement the *test-isolate* [25] strategy with greater efficiency. Together, cohorting can potentially help reduce the overall impact of the epidemic while supporting social and economic activities — the precise trade-offs of these two components is the focus of the paper.

Novelty. The work has a number of innovative components. This includes the following. (*i*) *Multi-scale agent simulations:* representing cohorts within a social network is challenging and calls for a multi-scale approach – agents interacting through interaction spaces instead of interacting directly with each other – in order to reduce computational complexity without compromising the quality of the solution. Another multi-scale notion is that of hierarchical interaction spaces that stratify interactions within a larger interaction space like in a workplace or a school or a neighbourhood. This is essential for simulating contact tracing and the test-isolate strategy effectively. (*ii*) *Multi-theory models:* that incorporate multiple social and epidemic theories. This includes theories of disease transmission and social choice theories (route choices to workplaces). (*iii*) *Calibration to diverse data sets.* Agent models should be calibrated to complex measured data. We incorporate a large variety of data to calibrate and validate the models. (*iv*) *Evaluation of realistic policies:* Study of policies that are directly derived from needs on the ground. The public health outcomes we derive from such a data driven model provides innovative insights that can inform policy.

2 RELATED WORK

The literature on agent-based models (ABMs) is vast, see for example [5, 6, 11] (ABM frameworks), [13] (modeling of small pox spread), [17] (benefit of fine-grained model), [1, 9] (fine-grained modeling of slums in Delhi and influenza spread), [32] (benefit of using auxiliary data sources for modeling Ebola spread), [29] (use of survey data to model behaviour), [3, 31, 34] (time-varying graph networks), [28] (vaccination), [30] (ABMs for computational epidemiology), [21, 26] (review of epidemic spread in complex networks) and [35] (impact of mobility on COVID-19 spread).

We shall focus only on those closely related to our work in this paper. Mei et al. [22] developed a city-scale ABM for Beijing (6M agents) using detailed demographic, mobility, socio-economic data. The data they needed are quite detailed, the interventions modeled have limited flexibility, and the test-isolate strategy is not modeled. Cooley et al. [10] developed a city-scale ABM (7.5M agents) based on survey data to study H1N1 epidemic spread in the New York City. They calibrated the parameters based on the 1957-8 influenza spread data. The interventions are limited to wearing masks, vaccinations and social distancing. They conclude that interventions targeting the subway system alone are not sufficiently effective in mitigating the infection spread. The papers [14, 16, 18] use a density model developed in [15] for narrow and enclosed areas to study the correlation between mobility and infection spread in crowded spaces. The validation of these models are limited to statistical data analysis [18] and small-scale simulations (200 – 100,000 agents) [15, 16]. Social bubbles have been studied by [20, 24, 27] using ABMs. However, these ABMs are modeled with limited interaction spaces [20], or do not consider local demographic variations [27], or limit the number of agents to about 100,000 [7].

In this work, we overcome several of the limitations highlighted above. We model the interactions more realistically by considering multiple and hierarchical interactions spaces like homes/neighbourhoods/communities, schools, and workplaces, in addition to interactions during commute. Our simulator works with 12.4M agents. The contact rate parameters are calibrated to the COVID-19 India data. The simulator enables time-varying interventions that model on-ground policies (containment zones, lockdown fatigue, compliance, contact tracing policies, and quarantine duration). Additionally, the simulator implements the test-isolate strategy.

3 METHODOLOGY

We implement cohorting strategies on top of a city-scale agent-based epidemic simulator. The agent-based simulator models interactions in households, workplaces, schools, neighbourhoods, and communities. In this paper, we focus on the modeling and the implementation of cohorts as an additional interaction space related to transportation.

The agent-based simulator consists of two components: (*i*) a synthetic city generator and (*ii*) a disease spread simulator that simulates the spread of the infection on the generated synthetic city. The synthetic city is generated taking into account the demographics of the city such as: age distribution, household size distribution, unemployment ratio, commute distance distributions, school size distributions etc. The synthetic city has as many agents as the population in the city of interest.

An attribute particularly important to the current study is the workplace location of an individual. We first assign a workplace ward (Mumbai has 24 wards) to an individual in accordance with the inter zone travel pattern given in Table 4 in [4]. The individual is then randomly assigned a workplace in the workplace ward (workplace location is uniformly distributed across a ward), with the intention to match the commute distance distributions for Mumbai.

All interactions between agents are modeled via interaction spaces to enable scalable computation at city-scale. Examples of interaction spaces include train coaches, workplaces, households, neighbourhoods. An agent can have membership to multiple interaction spaces, and disease spreads between agents via shared interaction spaces. Additionally, we bring hierarchical interactions spaces (project-team : workplace, neighborhood : wider-community, class-rooms : schools) to enable modeling of test-isolate strategies. The infection spreading model for an interaction space can depend on the type of the interaction space.

The model for disease progression in an infected individual accounts for disease states such as susceptible, exposed, pre-symptomatic and asymptomatic (to model asymptomatic transmission), symptomatic, hospitalised, critical, deceased and recovered, with age-dependent state transition probabilities. The age-dependent state transition probabilities are based on estimates in [33].

With all these attributes, agents constitute the nodes of a geospatial social contact network with mapped mobility. The city-scale epidemic simulator takes (*i*) the instantiated network (synthetic city), (*ii*) the disease progression model, (*iii*) the intervention, testing and contact tracing protocols, and then simulates the spread of infection in the network. We expand on the modeling of the transport interaction space and cohorts in the next subsections.

3.1 Mumbai *locals* Dataset

We digitized the Mumbai Rail Map in [19], and captured train line, train station, transit time between consecutive stations along a line, along with station latitude, longitude information. We restrict ourselves to the census city limits of Mumbai and Mumbai suburbs, to be consistent with agents being generated, and have 52 stations in the network. This data is used to pre-compute shortest travel route and corresponding time between any two stations across any line. In determining total travel time, if a journey has multiple legs, each transfer interval is assumed to be 7 minutes based on frequency of *locals* at stations under pre-COVID schedule. A leg represents transit along a single train line that does not require any transfer. During instantiation, the pre-computed travel times are used in determining shortest commute time route and corresponding source-destination stations for individuals. In the simulator, the pre-computed travel times are used to model the time spent in a journey.

3.2 Cohorts

To model cohorts, in addition to the attributes mentioned above, we first determine if an office-going agent would take the train to go to work. For travel between home and workplace we optimize for travel time across various modes of travel, accounting for frequency and cost differential of road commute options compared to using *locals*, the expected speed of travel via road 21.6 km/h [23] along with geodesic to road detour index [8] (value considered for Mumbai = 1.7), and multiple possible train stations an agent would consider as the primary commute stations both for home and workplace location. This allows us to know which agents take trains, and what route they follow. Our generated data shows that about 30% of the population takes train for daily workplace commute, which is inline with the known ridership numbers of *locals*. Each cohort can have up to 3 legs in the journey (based on *locals* routes), where each leg represents travel along a single line of *locals* in a single coach.

All individuals in a cohort are assumed to travel together while commuting. Hence they are assigned based on shared origin and destination stations, that is every member of the cohort would assemble at the common origin station, and then travel together to the destination station together with other members of the cohort and remain in the same train coach for each leg of the journey as every other member of the cohort. In our simulation, individuals are picked randomly to form a cohort as long as they share origin and destination stations for commute. Cohort size is parameterized, and the case of cohort size = 1 represents non-cohorting (business as usual) scenario. Cohorts are formed at the start of the simulation and stay same throughout the simulation.

3.2.1 Cohort-to-coach matching. Each day cohorts get assigned to train coaches for their morning and evening commute. For every train line (four in our case) and direction of travel, we initialise an empty coach with a given seating capacity and a *crowding factor*. For any given section of travel, the number of individuals in a coach cannot exceed the occupancy limit defined as the product of seating capacity and the crowding factor. In other words, the crowding factor captures the fact that the number of travelers in a single coach almost exceeds the total number of available seats (by a significant factor during normal operations). Our experiments use crowding factor as one of the parameters that can be controlled when formulating a policy. For each coach, we also keep count of cohorts who could not be accommodated due to unavailability of space. Next, we pick cohorts at random. For the chosen cohort, we check if the existing coaches on the journey legs for the cohort (depending on their source destination pair, cohorts can travel across multiple lines and hence could have more than one journey leg) can accommodate the cohort. If the cohort can be accommodated on all its journey legs, the cohort is assigned to the corresponding coaches, and the capacity of the coaches along the cohort's journey

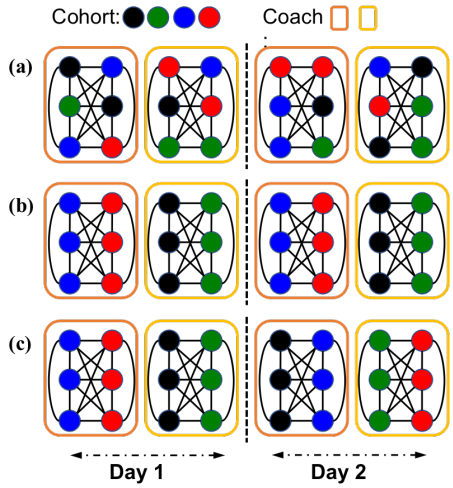


Figure 2: An example of different cohort-to-coach strategies and their impact on network structure. There are four cohorts (colored by black, green, blue, and red) and two coaches (colored by orange and yellow). Two days are shown in the figure. (a) No cohort-to-coach strategy. Passengers from all cohorts are randomly mixing between coaches on different days. (b) Static cohort-to-coach strategy. Members of a cohort travel together and coach assigned to the cohort is fixed across days. (c) Dynamic cohort-to-coach strategy. Members of a cohort travel together and coach assigned to the cohort is arbitrary across days.

legs is correspondingly reduced. If the cohort cannot be accommodated in at least one journey leg, the cohort is put back into bin of un-allocated cohorts, and the counter on coaches that could not accommodate the cohort is incremented. If the counter on a coach exceeds a threshold (five in our simulations), the current coach is pushed into an array of occupied coaches, and a new coach is instantiated for that train line. We continue this till all cohorts are assigned to coaches.

Assigning cohort-to-coach. We study two cohort-to-coach assignment strategies – static and dynamic. Static assignment strategy is used to model the case where the coach assigned to a cohort is fixed for each commute via a ticketing policy in practice. Specifically, each cohort takes the *same coach each day* in each direction, and this coach is shared with the *same cohorts* each day as well. Dynamic assignment strategy is used to model the case where the ticketing policy is more relaxed, and each cohort can potentially choose any train coach arbitrarily as long as they meet the basic criteria of cohorting where every member of the cohort travels together in the same coach. The dynamic strategy is easier to implement and requires minimal enforcement in practice compared to static assignment. The impact of the strategies is evaluated and discussed in the paper later. Figure 2 illustrates the cohorting strategies and their impact on network structure.

3.2.2 One-Off Travel. We model one-off travel for scenarios where individuals may not have a fixed daily commute destination, by separating travelers into 2 pools–those traveling in cohorts and those traveling individually. Each of these pools travels in separate coaches, and in practice could travel in different trains at different times to avoid infection spread at stations. We implement one-off travel by earmarking a fraction of the office going individuals as one-off travelers. The one-off travelers can be effectively viewed cohorts of size one, while the remaining group is used to form cohorts of required size. We assign separate coaches for one-off travelers and normal cohorts, thereby restricting all interactions between normal cohorts and one-off travelers. Even though we don't model the transition of an individual from one-off traveler to cohort traveler, we expect this can be achieved via testing or quarantining before joining a cohort.

4 MODELING DETAILS

Once the cohort assignments are done, the simulator proceeds in time steps of 6hrs. The simulator is seeded with 100 infected individuals at the start of the simulation. The start of simulation, with reference to actual timelines, is obtained as part of the simulator calibration step; we discuss in more detail in Section 4.1. At time step t, a susceptible individual n is exposed to a daily disease transmission rate $\lambda_n(t)$. The computation of $\lambda_n(t)$ takes into account the individual's interactions in all the interaction spaces they are associated with. The strength of interactions between an individual and the various interaction spaces are determined by the intervention policy active at the given time step. The intervention policies are chosen based on the announced policies by the government. Some parameters (e.g., compliance parameter, contact tracing parameters) are tuned so that the simulator output is in reasonable agreement with the actual data from the city.

We model the contribution of an interaction space to an individual's infection rate via a mean-field approximation of the infectivity prevalent in the interaction space. Such an approximation helps us to, (i) avoid simulating pair wise interactions between individuals, and (ii) one common computation of the infectivity in the interaction space suffices for all individuals in the interaction space. Individual variations are modeled separately, but is easily implemented via the use of individual specific scaling factors. We follow the same methodology for cohorts too. We describe the mean-field approximations used in modeling cohorts in (1), (2) and (3).

Given the disease transmission rate seen by an individual, a susceptible individual will be infected at the current timestep t with probability $(1 - exp(-\lambda_n(t) \times \Delta_t))$, where Δ_t is the duration of the simulations timestep in days (1/4 in our simulations). Thus, computation of $\lambda_n(t)$ for every individual at every timestep is a key component of the simulator.

In this study, we introduce cohorts and model the contribution from trains to an individual's transmission rate. Contribution from trains is modeled in two parts: 1) contribution from the same cohort, which we call the intra-cohort interaction, and from other cohorts that share a train coach with the cohort under consideration, which we call the inter-cohort interactions. For cohort i, the intra-cohort transmission rate is modeled as

$$\lambda_{intra_cohort}(i,t) = \beta_{coach}\tau(i) \sum_{n':Co(n')=i} I_{n'}(t)\rho_{n'}(t)\kappa_c(n',t), \tag{1}$$

where $\tau(i)$ denotes the travel time for cohort i across all journey legs, the summation is across all individuals n' who belong to cohort i. $Co(n')$ denotes the individual to cohort mapping. $I_{n'}(t)$ denotes whether individual n' is infective at time t, $\rho_{n'}(t)$ is the infectiousness factor for individual n', and $\kappa_c(n',t)$ is a modulation factor for cohort related interactions for individual n', and is used to model the effect of various intervention strategies. For example, when an individual n''s cohort is quarantined, we set $\kappa_c(n',t)$ to zero, to model the individual's lack of interaction via cohorts.

β_{coach} denotes the transmission rate parameter, that can be used to calibrate the simulator behavior with actual observations.

The inter-cohort interaction seen by cohort i is modeled as

$$\lambda_{inter_cohort}(i,t) = \sum_{j:j\neq i} \lambda_{intra_cohort}(j,t)\frac{OT(i,j,t)}{\tau(j)}, \tag{2}$$

where j denotes all other cohorts and $OT(i,j,t)$ denotes the overlap time in the journeys of cohorts i and j, where two cohorts are considered to overlap only if they share a coach.

Once the intra-cohort and inter-cohort transmission rates are computed, the contribution of trains to a susceptible individual n's transmission rate is modeled as

$$\lambda_n^{trains}(t) = (\lambda_{intra_cohort}(i,t) + \lambda_{inter_cohort}(i,t))\kappa_c(n,t), \tag{3}$$

where i denotes individual n's cohort, and $\kappa_c(n,t)$ is the workplace modulation factor for individual n.

Unlike other interaction spaces where the interaction space is considered active for all simulation timeslots, for trains we consider them active twice daily in two 6hr simulation timeslots. The first timeslot corresponds to morning travel and the second timeslot corresponds to evening travel.

In (3) we have assumed that, inside a coach, the interactions are homogeneous among all individuals in the coach. This assumption can be translated to uniform mixing of individuals in a coach. Even though we can expect cohort members to stay together and hence restrict interactions with other cohorts, the above assumption models the worst case scenario. The assumption is further reasonable as it will be difficult to implement physical distancing between cohorts and also to model the level of interactions between cohorts.

For cohorts, unlike the modeling of infection spread in other interaction spaces where we assume that the number of interactions of an individual remains constant for a given time interval, we assume that the number of interactions increases in proportion to the number of individuals traveling in a coach. We assume a linear increase in interactions as the number of individuals increase, though one could consider other alternatives, such as a monotonically increasing concave function.

4.1 Calibration of parameters

Like β_{coach} there are other tunable parameters in the simulator and it is important to tune the parameters to appropriate values to obtain reasonable outputs and insights from the simulator. We perform calibration in two steps. We seed a fixed number of individuals (100 in our case) in exposed state at the start of the simulation, and simulate a *no-intervention* (no mitigation strategies are enabled) scenario. We then tune the transmission parameters for home, workplace and community ($\beta_H, \beta_W, \beta_C$) to minimise the difference in slopes of the log of cumulative fatalities seen in the simulator to that of the actual cumulative fatalities observed in India between 26 March 2020 and 10 April 2020 (from 10 fatalities to 199 fatalities). We consider the *no-intervention* policy for calibration as the fatalities before April 10 2020 can be assumed to have got infected prior to imposition of any restrictions in India. A linear fit for the log fatalities curve is based on the assumption that cumulative fatalities grow exponentially with time, which was observed to be consistent with actual data. We observe a good match in the slopes of the fatalities curves for India and Mumbai, suggesting similar disease transmission rates in Mumbai and in the whole of India in the initial days of the pandemic.

Transmission parameters in interaction spaces are tied to one of the above three parameters. For example, we assume that the transmission parameter for a project team is nine times that of the transmission parameter for the larger workplace. This is based on the assumption that an individual spends 90% of the time in office with their team members and only the remaining 10% with the larger workplace group. We further try to equalise the contribution of household, workplace and community towards disease spread. We use *stochastic approximation* methods to arrive at the appropriate parameters values. Once we match the growth rate of fatalities, we calibrate the start of simulation date with the actual fatalities timeline, so that the time series of fatalities of the simulator matches in expectation with the actual data.

4.1.1 Calibration of β_{coach}. Due to sparsity of data on the impact of trains on disease transmission we have not been able to calibrate β_{coach} independently. We use the following heuristic argument to compute a nominal β_{coach} from β_H, where β_H is the transmission rate parameter associated with households.

The infection transmission rate seen by an individual n from their household at time t is modeled as

$$\lambda_n^H(t) = \beta_H \frac{1}{n_H^\alpha} \sum_{n'=1}^{n_H} I_{n'}(t)\rho_{n'}(t)\kappa_H(n',t), \tag{4}$$

where the summation is across all individuals in the household, $(1-\alpha)$ denotes the crowding factor for households, $I_{n'}(t)$ denotes whether individual n' is infective at time t, $\rho_{n'}(t)$ denotes individual n''s infectiousness factor and $\kappa_H(n',t)$ denotes individual n''s household based modulation factor.

Thus, β_H can be interpreted as the household transmission rate per day for an individual. Let r_H denote the number of typical contacts for an individual at home for a day. Then, the probability of transmission from a contact can modeled as $p_c = \frac{\beta_H}{r_H}$.

4.1.2 Calibration of β_{coach}. Due to sparsity of data on the impact of trains on disease transmission we have not been able to calibrate β_{coach} independently. We use the following heuristic argument to compute a nominal β_{coach} from β_H, where β_H is the transmission rate parameter associated with households.

The infection transmission rate seen by an individual n from their household at time t is modeled as

$$\lambda_n^H(t) = \beta_H \frac{1}{n_H^\alpha} \sum_{n'=1}^{n_H} I_{n'}(t)\rho_{n'}(t)\kappa_H(n', t), \tag{5}$$

where the summation is across all individuals in the household, $(1-\alpha)$ denotes the crowding factor for households, $I_{n'}(t)$ denotes whether individual n' is infective at time t, $\rho_{n'}(t)$ denotes individual n''s infectiousness factor and $\kappa_H(n', t)$ denotes individual n''s household based modulation factor.

Thus, β_H can be interpreted as the household transmission rate per day for an individual. Let r_H denote the number of typical contacts for an individual at home for a day. Then, the probability of transmission from a contact can modeled as $p_c = \frac{\beta_H}{r_H}$.

Let the number of typical contacts per minute per individual in a train coach be r_T. Then, the effective transmission rate per minute per individual in coach, $\hat{\beta}_{coach}$, can be expressed in terms of r_T as

$$\hat{\beta}_{coach} = p_c r_T = \frac{\beta_H r_T}{r_H}. \tag{6}$$

If we assume that the number of close contacts per day in a household $r_H = 50$ contacts per day, and close contacts per minute per individual in a coach $r_T = 1/100$, $\hat{\beta}_{coach}$ can be computed in terms of β_H as:

$$\hat{\beta}_{coach} = \beta_H \frac{r_T}{r_H} = \beta_H \frac{1}{100}\frac{1}{50} = \beta_H 0.0002. \tag{7}$$

In the simulator, for every simulation timestep Δ_t, the transmission rate is further multiplied by Δ_t to obtain the mean disease transmitting contacts per simulation timestep. For trains, since we already account for commute time, and since we assume that one journey is restricted to one simulation time step, we need to discount the further multiplication by Δ_t. Thus, a nominal value for the β_{coach} parameter we use in simulator, in terms of β_H can be obtained as

$$\beta_{coach} = \frac{\hat{\beta}_{coach}}{\Delta_t} = \hat{\beta}_{coach} 4 = \beta_H 0.0008, \tag{8}$$

where we have used $\Delta_t = 1/4$ days, to account for a simulation timestep duration of 6hrs.

4.2 Intervention Modeling

A key feature of the simulator in [12] is its ability to simulate various time varying intervention strategies. Interventions are modeled by modulating an individual's edge weights with various interaction space. For example, when an individual is self-isolated at home, contact rates with their household is reduced by 25%, contact rates with their workplace is reduced to zero and contact rates with the community is reduced to 10%. The simulator also supports testing and contact tracing protocols. Contact tracing in the close network of an individual can be initiated for each of the following events: (i) an individual is hospitalised, (ii) an individual tests positive, (iii) an individual reports symptoms. The fraction of such events that trigger contact tracing and the fraction of individuals who would be contact traced are all configurable.

Table 1: Model parameters

Parameter	*Symbol*	*Mumbai*
Transmission coefficient at home	β_h	0.7928 (calibrated)
Transmission coefficient at school	β_s	0.2834 (calibrated)
Transmission coefficient at workplace	β_w	0.1417 (calibrated)
Transmission coefficient at community	β_c	0.0149 (calibrated)
Subnetwork upscale factor	$\tilde{\beta}$	9
Transmission coefficient at transport space	β_{coach}	0.0005
Household crowding	$1-\alpha$	0.2
Community crowding	r_c	2
Distance kernel $f(d) = 1/(1+(d/a)^b)$	(a, b)	$(2.709, 1.279)$
Infectiousness shape (Gamma distributed)	(shape,scale)	$(0.25, 4)$
Severity probability	$\Pr\{C_n = 1\}$	0.5
Age stratification	$M_{n,n'}$	Not used
Project subnetwork size range	$n_{\mathscr{W}(n)}$	$3-10$
Family friends' subnetwork range	no symbol	2-5 families

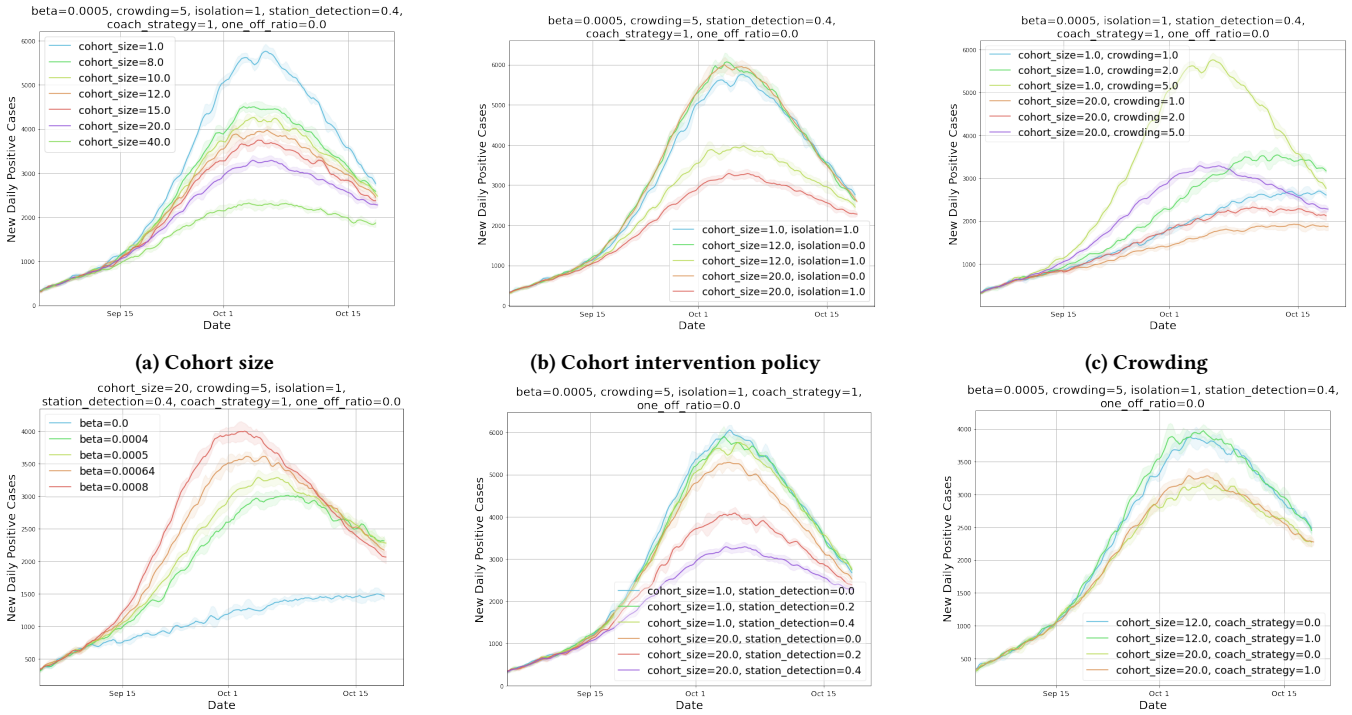
4.3 Intervention Modeling - Cohorts

In the specific case of cohorts, we study the impact of isolating an entire cohort when an individual is hospitalised or tested positive or is sufficiently symptomatic. When an individual is hospitalised or tested positive, then all their cohort members are placed under self isolation. A symptomatic individual may self-declare or be detected at a station and that can also trigger isolation of the other cohort members.

5 RESULTS AND KEY FINDINGS

Based on the methodology described in Section 3.2, 3.74M individuals take trains for daily commute in our simulations. In terms of interventions, we consider a pre-lockdown period starting from March 16 and lockdown extending till May 18 in Mumbai. From May 18, the following interventions are active: (i) 14 day home quarantining of symptomatic individuals and their household members, for compliant households, (ii) social distancing by compliant elderly (age > 65), (iii) schools are closed. Further, post May 18, we assume a phased re-opening of offices, with 5% offices operating till May 31, followed by 15% attendance till June 30, followed by 25% attendance till July 31, followed by 33% attendance till Aug 31 and full attendance thereafter. Further, we assume masks were enforced from April 09. Compliance probabilities to the regulations (including mask wearing) are 60% in high density areas and 40% in other areas. Testing and contact tracing is active from March 16. On hospitalisation or identification of a positive case, contact tracing, isolation and testing is triggered, based on a set of specified probabilities. This models the *test-isolate* intervention. Trains are assumed to start from September 07. Progressive containment of wards based on number of hospitalised cases is active till the end of August.

We run the simulation from mid February, but exclude presenting results for the duration prior to the anticipated restart date for the *locals*. As an implementation detail, we built in our model the ability to store and load state at any timestep, and were able to store state for state prior to the anticipated *locals* restart date. We only stored the infection state of individuals. Randomization seed, and other parameters related to cohorting were not stored. This ability to store once and load multiple times helped reduce



(a) Cohort size **(b) Cohort intervention policy** **(c) Crowding**

(d) β_{coach} **(e) Symptomatic detection at station** **(f) Cohort-to-coach mapping strategy**

Figure 3: Y-axis represents daily new detected positive cases. Parameters which vary are labeled in the legend, where as parameters which don't are noted above the plot. The shaded region represents 1 standard deviation away from the mean, estimated over 5 simulation runs with the selected configuration. Beta represents β_{coach}, the transmission rate parameter of the interaction in train coaches. Crowding represents the crowding factor of train coaches, which impacts occupancy limit of a train coach. Isolation value of 0 or 1 represents lack or presence of cohort isolation policy respectively. Coach_strategy of 0 or 1 represents static or dynamic coach assignment respectively. One_off_ratio represents proportion of travelers that travel one off, and the remainder $(1.0 - one_off_ratio)$ travel in cohorts of the selected cohort_size. Station detection represents the proportion of symptomatic infected commuters detected at the station, via thermal screening or other testing mechanisms.

computation time significantly. In a few scenarios for long term projections and estimating network saturation, we run the simulator for significantly longer duration.

In Figure 3, we plot new daily positive cases, which represent the new cases detected on the specified date. This value is much lower than the actual underlying new infections on that day. All plots in the group peak and then trend downwards. To explain this, we plot Figure 4 which shows that the cumulative detected cases to that date side by side with new daily positive cases, and we observe that the network saturates by Dec, with inflection point of new daily positive cases in mid Oct, which matches with the peak of new daily cases. Also interestingly we see that network saturation level for larger cohort sizes is much lower, which shows that cohorting helps reduce total cases significantly.

It is important to note that the daily positive cases take into account interaction in other interaction spaces (household, office, neighbourhood, community) in addition to the transport interaction space. The plots thus show the impact of cohorting strategies on the progression of the *overall disease burden in the city*. Another feature of the simulator is that it models spatial variance in population density and the associated increased contacts in slum areas, and therefore increased spread in these areas. A third important feature worth highlighting is that we model contact tracing, which help contain the spread, this helps to not only match actual observed cases, but also model the benefit of contact tracing enabled case isolation in a realistic way.

5.1 Details

1. Impact of cohort size. Figures 3a and 3b show the change in infection spread dynamics for various cohort sizes. We observe:

- Without isolation, all cohorts have similar dynamics.
- With isolation, we observe reduction in disease spread, as potentially asymptomatic cases are removed from the mix.
- Significant reduction in disease spread is observed as cohort sizes increases.

2. Impact of crowding factor. Figure 3c plots daily positive cases for different crowding factors.

- Disease spread is extremely sensitive to crowding.

- Crowding factor of 1 without isolation has lower daily cases than crowding factor 2 with isolation, suggesting that reducing the crowding in trains by half has more impact than enforcing isolation of cohorts.

3. Impact of contact rate. In Figure 3d, we plot daily positive cases for different β_{coach}, the contact rate parameter for cohorts. Since we do not have a calibrated β value, it is important to study the robustness of the results to various β values. We observe:

- Higher β_{coach} values cause higher disease transmission.
- Higher β_{coach} values reach the peak earlier.
- We observe similar trend in the disease progression curves for different β values, suggesting that the qualitative results are robust to variations in β values.

4. Impact of detection probability. In Figure 3e, we study the impact of varying the detection probability of symptomatic individuals at stations. A detection triggers the isolation of the entire cohort. Detection of positive symptomatic individuals is critical for cohorting to succeed and could potentially be achieved using thermal scanners at stations, and/or by employing random testing of individuals at stations. As detection increases, spread of diseases falls, and without detection at cohorting has only marginal improvement compared to no-cohorting scenarios.

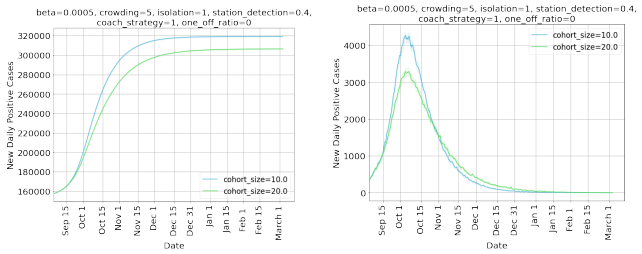
5. Impact of inter-day cohort-to-coach assignment strategies. In Figure 3f, we study the impact of static coach assignment, where restricting cohorts to travel with the same set of cohorts on a daily basis against a much relaxed policy of allowing cohorts choose their train of choice and time of travel. The only restriction is that the cohorts should travel together. We observe no significant benefits from static coach assignment, suggesting strict coach assignment may not be warranted.

6. Impact on count of quarantined individuals 5a shows that the number of people quarantined due to cohorting increases with increased cohort sizes. Interestingly, the increase in total number of quarantined individuals in city due to cohorting is small (Figure 5b). This can be attributed to the observation that cohorting with reasonable cohort size can reduce the daily positive cases substantially, and stem disease progression (Figure 3a). Cohorting is able to quarantine individuals even before testing, and contact tracing mechanisms kick in, and has an ability to smartly identify & isolate.

7. Impact of one-off travel. Figure 6 shows the impact of allowing one-off travel along side cohorts. We assume one-off travelers use separate coaches and thus avoid interacting with those in cohorts. Figure 6 (right) shows new daily positive cases detected with increasing cohort sizes. We observe that even with significant proportion of one-off travelers (40%), cohorting strategy reduces disease transmission significantly.

5.2 Ideal Cohort Size

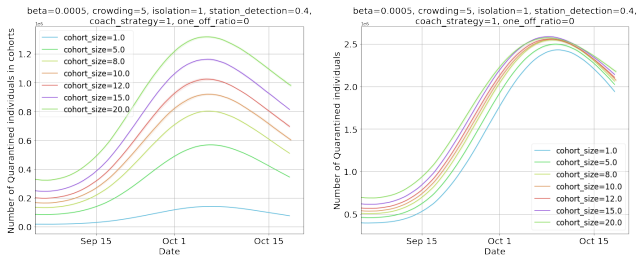
Even though our study finds larger cohort sizes attenuate disease transmission more severely, there are practical considerations like coach capacity utilization, and enforceability of isolation policies. In practice a cohort size of 12 to 20 might be more appealing.



(a) Cumulative positive cases **(b) Daily positive cases**

Figure 4: Effect of cohort size on peak daily case load and total case load at saturation.

.



(a) # of individuals quarantined due to cohorting in locals **(b) # of total individuals quarantined in the city**

Figure 5: As cohort size increases, there is a greater contribution of quarantined cases due to cohorting, but the increase in total quarantined cases in the city is marginal.

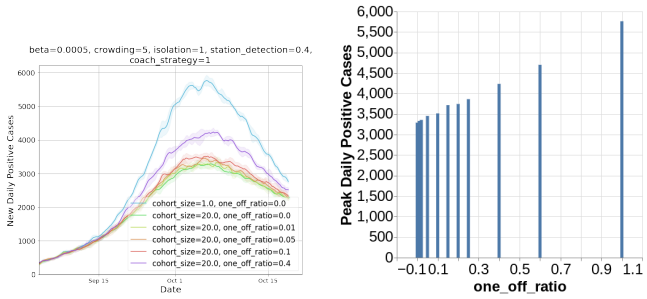


Figure 6: Effect of one-off travelers: As fraction of one-off travelers increase, the daily positive cases increase but partial cohorting does far better than business as usual (cohort size of 1).

6 KEY FINDINGS AND POLICY IMPLICATIONS

- Cohorting can significantly reduce disease transmission. Larger cohort sizes are more effective at reducing disease transmission.
- Effectiveness of cohorting depends on effectiveness of timely case detection (via thermal screening for symptomatic cases

at stations) and isolation (enforcing quarantine for all cohort members when a single member is detected to be positive).
- Static coach assignment does not have significant impact on reducing disease transmission as compared to dynamic coach assignment.
- Disease transmission is most sensitive to crowding in trains, and crowding needs to be effectively managed. While reducing crowding on train coaches, care must be taken to not crowd the stations. This aspect needs to be studied further.
- One-off travel up to 10% will only show marginal increase in disease transmission, and even at 40% one-off travel, disease transmission can be significantly reduced by cohorting. There is benefit in incremental implementation of cohorting.
- While the model has focused on Mumbai metro rail system, we believe these findings are generally applicable to other public transit systems (like metros, buses) in India as well other large metropolitan areas.

ACKNOWLEDGMENTS

This work was supported by a Google Gift Grant. The SahasraT supercomputing cluster at the Supercomputer Education and Research Centre (SERC), Indian Institute of Science, was used for most of the simulations. NKV and SS's work was done at the IISc-Cisco Centre for Networked Intelligence; the centre's support is gratefully acknowledged. RS's work was done while on sabbatical leave at Strand Life Sciences; their support is gratefully acknowledged. NKV, SS and RS are grateful for the discussions, comments, suggestions from Sandeep Juneja, Ramprasad Saptharishi, Piyush Srivastava and Prahladh Harsha from the School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai.

We (JC, SV, LW, AA, MV) thank members of the Biocomplexity Institute and Initiative, University of Virginia for useful discussion and suggestions. This work was partially supported by National Institutes of Health (NIH) Grant R01GM109718, NSF BIG DATA Grant IIS-1633028, NSF DIBBS Grant OAC-1443054, NSF Grant No.: OAC-1916805, NSF Expeditions in Computing Grant CCF-1918656, CCF-1917819 & NSF RAPID CNS-2028004, NSF RAPID OAC-2027541. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] Abhijin Adiga, Shuyu Chu, Stephen Eubank, Christopher J Kuhlman, Bryan Lewis, Achla Marathe, Madhav Marathe, Eric K Nordberg, Samarth Swarup, Anil Vullikanti, et al. 2018. Disparities in spread and control of influenza in slums of Delhi: findings from an agent-based modelling study. *BMJ open* 8, 1 (2018).

[2] Aniruddha Adiga, Lijing Wang, Adam Sadilek, Ashish Tendulkar, Srinivasan Venkatramanan, Anil Vullikanti, Gaurav Aggarwal, Alok Talekar, Xue Ben, Jiangzhuo Chen, et al. 2020. Interplay of global multi-scale human mobility, social distancing, government interventions, and COVID-19 dynamics. *medRxiv* (2020).

[3] Alessia Antelmi, Gennaro Cordasco, Carmine Spagnuolo, and Vittorio Scarano. 2020. A Design-Methodology for Epidemic Dynamics via Time-Varying Hypergraphs. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS' 20)*. 61–69.

[4] Judy Baker, Rakhi Basu, Maureen Cropper, Somik Lall, and Akie Takeuchi. 2005. *Urban poverty and transport: the case of Mumbai.* The World Bank.

[5] Michael Balmer, Konrad Meister, Marcel Rieser, Kai Nagel, and Kay W Axhausen. 2008. Agent-based simulation of travel demand: Structure and computational performance of MATSim-T. *Arbeitsberichte Verkehrs-und Raumplanung* 504 (2008).

[6] Christopher L Barrett, Keith R Bisset, Stephen G Eubank, Xizhou Feng, and Madhav V Marathe. 2008. EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *SC'08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*. IEEE, 1–12.

[7] Per Block, Marion Hoffman, Isabel J Raabe, Jennifer Beam Dowd, Charles Rahal, Ridhi Kashyap, and Melinda C Mills. 2020. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour* (2020), 1–9.

[8] Francis P. Boscoe, Kevin A. Henry, and Michael S. Zdeb. 2012. A Nationwide Comparison of Driving Distance Versus Straight-Line Distance to Hospitals. *The Professional Geographer* 64, 2 (2012), 188–196. https://doi.org/10.1080/00330124.2011.583586 arXiv:https://doi.org/10.1080/00330124.2011.583586

[9] Jiangzhuo Chen, Shuyu Chu, Youngyun Chungbaek, Maleq Khan, Christopher Kuhlman, Achla Marathe, Henning Mortveit, Anil Vullikanti, and Dawen Xie. 2016. Effect of modelling slum populations on influenza spread in Delhi. *BMJ open* 6, 9 (2016).

[10] Philip Cooley, Shawn Brown, James Cajka, Bernadette Chasteen, Laxminarayana Ganapathi, John Grefenstette, Craig R Hollingsworth, Bruce Y Lee, Burton Levine, William D Wheaton, et al. 2011. The role of subway travel in an influenza epidemic: a New York City simulation. *Journal of urban health* 88, 5 (2011), 982.

[11] Sara Y Del Valle, Phillip D Stroud, James P Smith, Susan M Mniszewski, Jane M Riese, Stephen J Sydoriak, and Deborah A Kubicek. 2006. EpiSimS: epidemic simulation system. *Los Alamos, NM: Los Alamos National Laboratory* (2006).

[12] Shubhada Agrawal et.al. 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.

[13] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 6988 (2004), 180–184.

[14] Edward L Glaeser, Caitlin S Gorback, and Stephen J Redding. 2020. *How much does covid-19 increase with mobility? evidence from new york and four other us cities.* Technical Report. National Bureau of Economic Research.

[15] Lara Goscé, David AW Barton, and Anders Johansson. 2014. Analytical modelling of the spread of disease in confined and crowded spaces. *Scientific reports* 4 (2014), 4856.

[16] Lara Goscé and Anders Johansson. 2018. Analysing the link between public transport use and airborne transmission: mobility and contagion in the London underground. *Environmental Health* 17, 1 (2018), 84.

[17] Jürgen Hackl and Thibaut Dubernet. 2019. Epidemic spreading in urban areas using agent-based transportation models. *Future Internet* 11, 4 (2019), 92.

[18] Jeffrey E Harris. 2020. The subways seeded the massive coronavirus epidemic in new york city. *NBER Working Paper* w27021 (2020).

[19] Mandar Rane Jaikishan Patel, Snehal Patil. 2013. Mumbai Rail Map. https://commons.wikimedia.org/wiki/File:Mumbai_Rail_Map_-_English.jpg

[20] Trystan Leng, Connor Whie, Joe Hilton, Adam J Kucharski, Lorenzo J Pellis, Helena Stage, Nicholas G Davies, Matt J Keeling, and Stefan Flasche. 2020. The effectiveness of social bubbles as part of a Covid-19 lockdown exit strategy, a modelling study. *medRxiv* (2020).

[21] Madhav Marathe and Anil Kumar S Vullikanti. 2013. Computational epidemiology. *Commun. ACM* 56, 7 (2013), 88–96.

[22] Shan Mei, Bin Chen, Yifan Zhu, Michael Harold Lees, AV Boukhanovsky, and Peter MA Sloot. 2015. Simulating city-level airborne infectious diseases. *Computers, Environment and Urban Systems* 51 (2015), 97–105.

[23] Shantonil Nag. 2017. Traffic speeds: India's fastest & slowest cities. https://www.cartoq.com/traffic-speeds-indias-fastest-slowest-cities/

[24] Bjarke Frost Nielsen and Kim Sneppen. 2020. COVID-19 superspreading suggests mitigation by social network modulation. *medRxiv* (2020).

[25] Wold Health Organization. 2020. WHO Director-General's opening remarks at the media briefing on COVID-19. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---16-march-2020

[26] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. 2015. Epidemic processes in complex networks. *Reviews of modern physics* 87, 3 (2015), 925.

[27] Jose Paulo Guedes Pinto, Patricia Camargo Magalhaes, Gerusa Maria Figueiredo, Domingos Alves, and Diana Maritza Segura-Angel. 2020. Local protection bubbles: an interpretation of the decrease in the velocity of coronavirus's spread in the city of Sao Paulo. *medRxiv* (2020).

[28] Prathyush Sambaturu, Bijaya Adhikari, B Aditya Prakash, Srinivasan Venkatramanan, and Anil Vullikanti. 2020. Designing Effective and Practical Interventions to Contain Epidemics. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS' 20)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1187–1195.

[29] Meghendra Singh, Achla Marathe, Madhav V Marathe, and Samarth Swarup. 2018. Behavior model calibration for epidemic simulations. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS' 18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1640–1648.

[30] Samarth Swarup, Stephen G Eubank, and Madhav V Marathe. 2014. Computational epidemiology as a challenge domain for multiagent systems. In *Proceedings*

of the 2014 international conference on Autonomous agents and MultiAgent systems (AAMAS' 14). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1173–1176.

[31] Michele Tizzani, Simone Lenti, Enrico Ubaldi, Alessandro Vezzani, Claudio Castellano, and Raffaella Burioni. 2018. Epidemic spreading and aging in temporal networks with memory. *Physical Review E* 98, 6 (2018), 062315.

[32] Srinivasan Venkatramanan, Bryan Lewis, Jiangzhuo Chen, Dave Higdon, Anil Vullikanti, and Madhav Marathe. 2018. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics* 22 (2018), 43–49.

[33] 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. 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases* 20, 6 (2020), 669 – 677.

[34] Christian L Vestergaard and Mathieu Génois. 2015. Temporal gillespie algorithm: Fast simulation of contagion processes on time-varying networks. *PLoS Comput Biol* 11, 10 (2015), e1004579.

[35] Ding Wang, Brian Yueshuai He, Jingqin Gao, Joseph YJ Chow, Kaan Ozbay, and Shri Iyer. 2020. Impact of COVID-19 Behavioral Inertia on Reopening Strategies for New York City Transit. arXiv:2006.13368