Urban Mobility Model Reduces Data Demands for Estimating COVID-19 Growth and Informs Reopening Policy

Zhenyu Han[1†]**, Fengli Xu**[2†]**, Yong Li**[1*]**, Tao Jiang**[3]**, and James Evans**[2,4*]

[1]Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing, P. R. China
[2]Knowledge Lab Department of Sociology, University of Chicago, Chicago, IL, USA
[3]School of Electronics Information and Communications, Huazhong University of Science and Technology, Wuhan, P. R. China
[4]Santa Fe Institute, Santa Fe, NM, USA
[†]Zhenyu Han contributed equally to this work with Fengli Xu.
[*]Yong Li and James Evans are corresponding authors. E-mail: liyong07tsinghua.edu.cn, jevansuchicago.edu

ABSTRACT

The SARS-CoV-2 virus is primarily transmitted through in-person interaction, and so its growth in urban space is a complex function of human mobility that cannot be adequately explained with standard epidemiological models. Recent studies leveraged fine-grained urban mobility data to accurately model viral spread, but such data pose privacy concerns and are often difficult to collect, especially in developing regions. To reduce the data barrier, here we propose a novel metapopulation SEIR model coupled with an urban mobility model. Experiments on real-world data collected from cities in the United States, India and Brazil show that our model can estimate complex, distinctive COVID-19 growth curves with high accuracy. Estimated mobility changes are consistent with empirical observations. Our model also reproduces urban "superspreading", where a few neighborhoods account for most new infections and it can inform location-aware reopening policies to achieve a better balance of social cost and disease prevention.

Introduction

The mechanism of SARS-CoV-2 spread in urban space is inherently complex, embedded in heterogeneous urban human mobility behavior. This has led to distinctive growth patterns of confirmed cases in cities worldwide, ranging from exponential to sublinear growth[1–3]. Nevertheless, most classic epidemiological frameworks, such as the Susceptible-Exposed-Infectious-Recovered (SEIR) model[4], adopt the assumption of homogeneous population mixing[5], which can only explain exponential growth curves and is at odds with empirical observations. In response, recent research has introduced the mechanism of self-containment to extend the expressive power of the SEIR model to capture subexponential growth[6]. Self-containment policies are simulated by gradually removing susceptible persons from the overall population, but this strategy continues to model urban population as a group of homogeneous individuals, which cannot explain widely observed urban "superspreading events" and the spatial heterogeneity of COVID risk[7,8]. Recent studies have begun to leverage fine-grained empirical mobility data to augment models of coronavirus spread[9–11], but these raise significant privacy concerns[12] and create a high demand for mobility data collection[13], which may not be possible to satisfy in the developing world. These concerns and limitations pose a pressing need for an expressive, general epidemiological model that can capture the complex mechanism of coronavirus spread and estimate varying trajectories of infection without mobility data. This is especially important to inform the design of flexible and adaptive containment and reopening policies.

Here, we present a novel metapopulation SEIR model that natively incorporates fundamental patterns of urban mobility behavior. We model urban space as a set of spatially distributed neighborhoods, and capture human mobility behavior as population flows among neighborhoods with a underlying urban mobility model. Inspired by Stouffer's law of opportunity-driven urban movement[14] and Tobler's first law of geography[15], we use a general gravity model that predicts mobility flows as proportional to neighborhood population and inversely proportional to travel distance. Atop simulated mobility flows, we overlay a metapopulation SEIR model to characterize the infection within each neighborhood and track inter-neighborhood disease spread arising from urban mobility. This model allows us to jointly consider spatially heterogeneous urban mobility behavior and urban population distribution without costly mobility data.

Experiments on real-world data collected from 30 of the most infectious counties and cities in the United States, India and Brazil show that our model can accurately estimate the complex and distinctive growth curves tracing COVID-19 confirmed cases ($R^2 > 0.980$). Estimated mobility changes are consistent with real-world observations derived from Apple Mobility Trends Reports[16] (Pearson's R = 0.872), suggesting the effectiveness of our model in substituting for human mobility data. Our model can also characterize urban "superspreading events" by tracking the spatially heterogeneous coronavirus spread through urban neighborhoods, where a small portion (20%) of neighborhoods account for a large portion (68.3%) of new infections. Moreover, experiments show that urban "superspreading events" are a joint result of uneven urban population distribution and heightened population flows associated with populous neighborhoods. We demonstrate that by focusing on those regions predicted to have the highest infection risk, the proposed model can facilitate location-aware reopening policies that result in significantly more effective epidemic control at a similar level of mobility restriction without the need for costly mobility monitoring[17]. Moreover, our proposed model serves as a simulation framework for the cost-effective evaluation of reopening policy under different scenarios, *e.g.*, different levels of mobility activity and different infection rates from new variants. In this way, our model reduces the demand for mobility data in estimating COVID-19 growth and informing the design of flexible and targeted reopening policies, which can guide countries through the transition to a post-pandemic world, especially in developing regions with limited data and access to therapies and vaccines.

Results

Urban mobility rules driven metapopulation SEIR model

We adopt a metapopulation scheme to model the fine-grained spread of disease in urban space. This breaks down urban space into numerous spatially distributed neighborhoods that each contains a sub-population, and maintains a separate SEIR model with its own susceptible (S), exposed (E), infectious (I), and recovered (R) states (see Supplementary Figure 1 in SI Appendix). We use a gravity model to characterize mobility flows between neighborhoods, accounting for two fundamental rules in human mobility behaviour: social interaction and travel cost (see Methods M1). The rule of social interaction emphasizes the attraction of social interaction opportunity, which characterizes Stouffer's law of opportunity-driven human movement[14]) and can be approximated by the population size of destination neighborhoods. Besides, the rule of travel cost captures the locality of urban movements[18]—Tobler's first law of geography[15]. These rules can be principally integrated into a general gravity model that predicts mobility flows proportional to neighborhood population and inversely proportional to travel distance. Therefore, our proposed model simulates occurrences of new infections within each neighborhood with separate SEIR models, and their dispersal to other neighborhoods according to predicted mobility flows. There are three free parameters in the proposed model: (1) a learnable infection rate β that accounts for infectiousness of coronavirus, *e.g.*, the effect of social distancing policies[19,20] and the emergence of new variants[21]; (2) a learnable quarantine rate κ that accounts for the capacity of testing and quarantine[22,23]; (3) and a learnable mobility level M that accounts for changes in mobility behavior, *e.g.*, the effect of stay-at-home orders and suspensions of public transportation[24]. Other epidemiological parameters are set according to recent COVID-19 studies (see Supplementary Table.S1 in SI Appendix). Urban population distributions are derived from the open source WorldPop database[25].

We evaluate our model with the task of estimating growth curves tracing confirmed cases of COVID-19 in 30 of the most infectious counties and cities in the United States, Brazil and India. To characterize the potential for changes in mobility behavior, we segment the simulation period based on the implementations of nationwide intervention policies in the United States, Brazil and India, and then fit separate parameters for each segmented period (Supplementary Figure 2 in SI Appendix). Ground truth data for confirmed cases are collected from official statistics released by each country[26–28] (see Supplementary Table.S2, S3, S4 in SI Appendix). Selected cities exhibit complex and distinctive growth curves (Fig. 1A, red dotted lines), which can be classified into four categories: linear (Hudson, King, and most Brazilian cities), concave (Bergen, Miami-Dade, Nassau, etc.), convex (Cook, Davidson, Los Angeles, etc.), and S-shaped (Harris, New Orleans, Will, and most Indian cities). All of these distinctive growth trajectories can be accurately estimated by our model, with R^2 scores above 0.980 for all cities (Fig. 1A, blue lines). By contrast, the standard SEIR model fails to reproduce these empirical growth patterns, only predicting exponential or no growth with $R_0 > 1$ and $R_0 < 1$, respectively (Fig. 1A, green lines)[6]. Moreover, we theoretically prove that our model is expressive enough to capture the complex forms of COVID-19 growth (see Methods M2), substantially extending the expressive power of classic epidemiological models. We find that different levels of urban mobility have a significant impact on the shape of growth curves (Supplementary Figure 3 in SI Appendix). These results justify the need to incorporate rules of urban mobility to characterize the complex growth of COVID-19 cases. We further evaluate our model by predicting future confirmed cases in all 20 U.S. counties within two weeks, *i.e.*, May 1-14 (Fig. 1B). Results demonstrate that our model significantly outperforms the standard SEIR model by reducing the normalized root mean square error (NRMSE) from 7.222 to 0.294. Experiments on mobility behavior change manifest a high correlation (Pearson's R = 0.872) between the mobility reduction estimated by our model and real-world observations derived from Apple Mobility Trends Reports[16] (Fig. 1C), which demonstrates that our model can accurately estimate empirical mobility behavior changes without mobility data. These results

suggest that our model can be robustly generalized to cities and regions in the developing world where mobility data is lagging or limited.

Reproducing and rationalizing superspreading events in urban space

Superspreading events have been widely observed in epidemics like SARS, Measles and Smallpox[29], as well as COVID-19, where a small portion of infected people and locations are responsible for a disproportionate number of new infections. Researchers have identified accumulating superspreading events of COVID-19 through case study[7,8], phylogenetic analysis[22,30] and statistical evaluation[31,32]. Superspreading cannot be adequately explained by the standard SEIR model reliant on the fundamental assumption of homogeneous population mixing. As a result, researchers have tried to characterize superspreading events with dispersion parameter k, which measures how the transmissive power of each individual deviates from the general population[8,29] with a negative binomial model. This statistical approach, however, neglects the behavioral mechanism underlying the superspreading phenomenon, *e.g.*, spatially heterogeneous population mixing. Moreover, recent research has found evidence that superspreading events can be better characterized with fat-tailed power law distributions[32], which further suggest their link to human behavior patterns. Because our model is tuned to characterize the spatial heterogeneity of disease spreading in urban space, here we examine the mechanism of superspreading through the analytical framework provided by our model.

We use the infectee-infector ratio to measure spatially heterogeneous infection risk among different neighborhoods, which we define as the average number of new infections per previously infected person. We rank neighborhoods based on the infectee-infector ratio, and calculate the cumulative distribution function of the infected population and new infections occurrences (Fig. 2A). If the infection risk is spatially homogeneous, we expect the cumulative distribution function to grow linearly with the number of neighborhoods following the black dashed line. Our model reproduces a highly skewed distribution where the most infectious 20% of neighborhoods are responsible for 68.3% of new infections. We use the Gini index[33] to quantify unevenness in the spatial distribution of the infected population and new infection occurrences, which measure 0.630 and 0.663, respectively, indicating a high degree of spatial heterogeneity. We further validate simulated spatially heterogeneous infection risk with fine-grained confirmed cases in New York City[34] (Fig. 2B). We observe a high correlation between predicted and real-world infections in each neighborhood (Spearman'R = 0.583). These results indicate our model can effectively extend the expressive power of the standard epidemiological model to explain superspreading events and prediction is consistent with real-world observation.

To investigate the mechanism behind superspreading events, we look at the effect of urban population distribution and urban mobility behaviors. First, recent research found empirical evidence that the uneven urban population distribution might contribute to the spatially heterogeneous infection risk in urban space[35], and the empirically observed urban population distributions follow a highly uneven exponential distribution (Supplementary Figure 4 in SI Appendix). Specifically, if urban population distribute evenly and urban movements are complete random, our model will degenerate into a standard SEIR model with homogeneous population mixing, which predicts the infection risk to be spatially homogeneous. To examine the effect of uneven urban population distribution, we simulate a urban system with real-world urban population distribution and random urban movement (Fig. 2C). We observe that the spatial unevenness of distribution of infected population and new infections is largely reduced, with the Gini index decreases to 0.198 and 0.090 respectively. It indicates the uneven urban population distribution alone cannot fully explain the superspreading events in urban space. Second, the urban movement is considered as another potential cause for the superspreading events in urban space[36], which is characterized by the rules of travel cost and social interaction in our model. To examine the effect of each rule, we evaluate two variants of urban mobility model that only consider one rule, respectively. When the urban mobility model only considers the rule of travel cost (see Fig. 2D), our model reproduces a similar level of spatial heterogeneity in infection risk as random movement (Fig. 2C). However, if the urban mobility model only considers the rule of social interaction, we reproduce a strong superspreading effect with Gini index of 0.632 and 0.539 for new infections and infected populations (see Fig. 2E), which are similar to the results of the complete model (Fig. 2A). We also theoretically analyze the superspreading effect in Supplementary Method 1 in SI Appendix. These results suggest that superspreading events are a joint result of an uneven urban population distribution and heightened population flows naturally associated with populous neighborhoods.

Informing location-aware reopening policies

As a result of superspreading events, highly skewed spatial distributions of COVID-19 cases suggest the potential to design targeted, location-aware intervention policies that more effectively curb coronavirus spread while minimizing social cost. Specifically, due to the emergence of new coronavirus variants and the shortage of vaccine supply, prolonged citywide lockdowns in many countries pose a pressing need for the design of reopening policies that can facilitate gradual but consistent recovery across communities[37]. Here, we leverage our proposed model to inform location-aware reopening policies that only restrict the mobility of neighborhoods with the highest predicted infection risks. We set a budget of social cost equal to the

percentage of city population remaining under mobility restriction, then evaluate the policy's efficacy in curbing disease spread as the percentage of confirmed cases that would be prevented compared to a complete reopening.

Populous neighborhoods tend to be more vulnerable to coronavirus transmission[35] (Spearman correlation = 0.74, see Supplementary Figure 5 in SI Appendix). A naive reopening policy might involve maintaining mobility restrictions for the most populated neighborhoods and first reopening other neighborhoods, but this strategy imposes high social costs and as shown in Fig. 2. Given a social cost budge of lifting the mobility restriction on 95% city population, if we reopen the least populated neighborhoods, then 28.49% of overall confirmed cases can be avoided in counties in U.S. compared to complete reopening, which outperforms the baseline of randomly selecting neighborhoods for reopening (Fig. 3A). Nevertheless, infection risk involves factors beyond neighborhood population and with the same social cost budget, if we reopen the neighborhoods with the least predicted infection risks, the overall confirmed cases can be reduced by 60.13% (Fig. 3A). Similar observations are made for Indian and Brazilian cities (Fig. 3C, E), where reopening policies informed by our model consistently outperform baselines of reopening least populated or random neighborhoods (see Supplementary Table.S5, S6, S7 in SI Appendix). These results suggest that our model can inform the design of substantially more cost-effective, location-aware reopening policies.

To examine sources of differential efficacy, we visualize neighborhoods selected for remaining under mobility restriction in three cities (Fig. 3B, D and F). The case of Los Angeles county (L.A.) illustrates how the baseline policy focused only on population density would miss the high infection risk that occurs in less populous neighborhoods around central L.A. because it cannot model these neighborhoods' mobility flows with nearby, densely populated neighborhoods (Fig. 3B). For cities in developing countries such as Delhi and São Paulo, we find that reopening policies informed by our model can better capture infections in multiple-center cities by considering both the population size and distance from nearby city centers (Fig. 3D, F). For example, in Fig. 3D, we show how our model selects neighborhoods in central areas that are likely the hubs of urban mobility, but that this has only 8.7% overlap with the top populated neighborhoods. More detailed analysis suggests the location-aware reopening policy informed by our model can also effectively alleviate super-spreading events in cities (Supplementary Figure 6 in SI Appendix).

Simulating the epidemic development under different scenarios

As efforts to reduce the spread of COVID-19 continue, the dynamic of coronavirus spread shifts as new variants emerge[21], urban mobility behavior changes[24] and new social distancing policies are implemented[19]. As a general framework that incorporates urban mobility with virus spread, our model can simulate epidemic development under different scenarios with distinct parameter settings. Therefore, we can evaluate the scenarios of varying mobility, infection rates, quarantine rates and intervention times in the epidemic development in 20 U.S. counties (Fig. 4A-D), respectively. First, we investigate how mobility affects disease transmission by assuming a different level of mobility in our model (see (Supplementary Method 2 in SI Appendix for details in tuning mobility behavior). The simulation shows that restricting mobility behavior is effective in reducing coronavirus transmission, where 10% higher real-world mobility activity will lead to nearly threefold the number of overall infections (2.93 times; IQR $1.57 \sim 7.24$), and 10% less mobility will halve the number of cases (0.525 times; IQR $0.237 \sim 0.783$). As a point of comparison, the average observed mobility drop in U.S. counties during the pandemic is approximately 30.3% (Supplementary Figure 7 in SI Appendix, with IQR $0.227 \sim 0.392$), which suggests the mobility behavior change during pandemic was effective in curbing the spread of coronavirus. Controlling the infection rate is also an effective method to fight the disease, where a 10% decrease in infection rate nearly halves infections (0.524 times; IQR $0.231 \sim 0.661$). This suggests the great importance of reducing the infection rate of coronavirus, which can be achieved by improving the distribution and administration of effective vaccinations[38,39]. Quarantine is another widely adopted approach to contain the spreading of coronavirus[40], but with a more complicated asymmetric effect. Our model predicts that the number of confirmed cases will be 29% higher given a 10% decrease in quarantine rate (1.29 times; IQR $1.06 \sim 2.38$), while a 10% increase in quarantine rate will prevent approximately 78% of citywide infections (0.221 times; IQR $0.0739 \sim 0.525$). This asymmetric effect of quarantine rates suggests the necessity of ensuring that testing capacity and hospital resources are sufficient. These capacities can greatly reduce overall infections once they pass a critical threshold. Finally, timely responses to COVID-19 spread are considered critical for combating the virus[41]. To explore policy response with our model, we shifted the time point of mobility behavior change to simulate the potential effect of different response times. We find that a delay of 10 days causes 3.26 times higher overall infections (IQR $1.58 \sim 9.53$), while adopting policies 10 days earlier reduces the infections by 57.5% (0.425 times; IQR $0.12 \sim 0.71$). These analyses reveal how timely interventions such as mobility reduction, social distancing and quarantine can play important roles in curbing viral spread and flattening growth curves.

Discussion

Our model aims to capture the fine-grained transmission process of coronavirus in cities with a minimal data burden. We show that by observing only the population distribution and number of confirmed cases, our model can accurately reproduce the complex growth curves of COVID-19 cases and forecast future trends in cities from both developed and developing countries.

Our model provides a theoretical framework to explain the distinctive growth curves in different cities from the perspective of urban mobility. Furthermore, our model can reproduce and explain superspreading events in cities resulting from the uneven distribution of urban population and spatially heterogeneous mobility. In these ways, our model markedly augments the expressive power of current epidemiological models. We demonstrate that targeted, location-aware reopening policies can be designed based on the prediction of our model, which can achieve better balance between epidemic control and social cost. As a general simulation framework, our model can evaluate likely outcomes under different natural and policy scenarios. Finally, the minimal data requirement ensures our model can be easily generalized to cities and regions without fine-grained mobility data, which will be particularly beneficial to developing countries.

Our study has two clear limitations. First, our parsimonious model only incorporates three learnable parameters, which account for infection rate, quarantine rate and mobility level. We do not consider all of the rich, contextual features that might influence the spread of coronavirus, such as weather[42], changing attitudes towards containment policies[43], and population demographics like age[23] and socioeconomic status[9]. Second, our model adopts the same set of parameters for all neighborhoods, which might limit model expressiveness in capturing neighborhood heterogeneity. Nevertheless, analysis demonstrates that our model can accurately trace citywide COVID-19 cases, suggesting sufficient expressiveness for characterizing urban viral transmission dynamics. Moreover, the minimum model setting improves its robustness and generalizability.

Our model and analysis hold several implications for understanding the COVID-19 epidemic that could improve the design of reopening policies. First, our analysis suggests that widely observed superspreading events are not solely due to the randomness of rare events, but also systematically linked with the underlying heterogeneity in human mobility. We further demonstrate the need to jointly consider urban environment and human mobility behavior to effectively curb superspreading. Second, our model leverages rule-based mobility to inform reopening policies without fine-grained mobility data, demonstrating how better epidemic control can co-exist with better privacy protection. Third, experiments show that our estimated mobility changes are consistent with real-world observations from Apple Mobility Trends Reports. This affirms that our model can offer insights into fine-grained urban transmission dynamics and be robustly transferred to cities without mobility data. Finally, our model can be used to comparatively and inexpensively evaluate possible epidemic scenarios in underdeveloped countries, such as divergent vaccination rates and quarantine policies. Our model proposes dynamically customized intervention policies to facilitate decision making in underdeveloped countries that lack fine-grained mobility data.

Methods

M1 Metapopulation SEIR model

Standard SEIR models use ordinary differential equations (ODE) to trace epidemic development, which divides the population into four statuses: susceptible (S), exposed (E), infected (I) and recovered (R). A fundamental assumption underlying the standard SEIR model is homogeneous population mixing such that each susceptible individual will have similar infection risk. As a result, SEIR can neither reproduce the complex growth curves of COVID-19 confirmed cases[6] nor explain the mechanism of superspreading events[8].

Here, we aim to extend the power of the standard SEIR model by introducing a metapopulation framework that incorporates fine-grained urban mobility behavior, as shown in (Supplementary Figure 1 in SI Appendix). Specifically, we divide urban space into numerous neighborhoods, and maintain a separate SEIR process for the sub-population in each one. The overall city population is divided into neighborhoods based on real-world population distributions[25]. In each simulation epoch, our model has two stages to trace epidemic development and population mixing, respectively. In the epidemic development stage, we compute changes in population status according to the following equations for each neighborhood:

$$\frac{\mathrm{d}s^n}{\mathrm{d}t} = -\beta s^n e^n - \beta i^n e^n + \mathrm{Input}_S(n,t)$$

$$\frac{\mathrm{d}e^n}{\mathrm{d}t} = (1-\omega_i)\beta s^n e^n + (1-\omega_i)\beta s^n i^n - \frac{1}{\tau}e^n + \mathrm{Input}_E(n,t)$$

$$\frac{\mathrm{d}i^n}{\mathrm{d}t} = \omega_i \beta s^n e^n + \omega_i \beta s^n i^n + \frac{1}{\tau}e^n - \kappa i^n - \gamma i^n + \mathrm{Input}_i(n,t)$$

$$\frac{\mathrm{d}r^n}{\mathrm{d}t} = \kappa i^n + \gamma i^n + \mathrm{Input}_R(n,t),$$

where s^n, e^n, i^n, r^n are the susceptible, exposed, infected and recovered people in neighborhood n. β is the infection rate, ω_i is the ratio that new infections immediately show symptoms and transition to infected, τ is average incubation time of exposed persons, and κ denotes the quarantine rate at which infected persons are removed from the population.

For the infection states for the whole city, we have

$$S = \sum_n s^n$$
$$E = \sum_n e^n$$
$$I = \sum_n i^n$$
$$R = \sum_n r^n.$$

Here we formally introduce the input terms of our formulations. Inputs are determined based on incoming and outgoing population due to urban mobility. Specifically, in the stage of population mixing, we simulate the urban mobility behavior based on the following gravity model:

$$m_{ij} = M \frac{N_i^\rho N_j^\theta}{\exp(d_{ij}/r)},$$

where M is the mobility level depicting intensity of urban mobility, N_i and N_j are the population size of the original and destination neighborhoods, and d_{ij} is the Manhattan distance between them. ρ, θ, r are empirical coefficients we set based on recent urban mobility research[44].

For each input terms, we have

$$\text{Input}_S(n,t) = M \sum_m \frac{N_m^\rho N_n^\theta}{\exp(d_{mn}/r)} \times \frac{s^m}{N^m}$$

$$\text{Input}_E(n,t) = M \sum_m \frac{N_m^\rho N_n^\theta}{\exp(d_{mn}/r)} \times \frac{e^m}{N^m}$$

$$\text{Input}_I(n,t) = M \sum_m \frac{N_m^\rho N_n^\theta}{\exp(d_{mn}/r)} \times \frac{i^m}{N^m}$$

$$\text{Input}_R(n,t) = M \sum_m \frac{N_m^\rho N_n^\theta}{\exp(d_{mn}/r)} \times \frac{r^m}{N^m},$$

where the summation is for each neighborhood. The second terms of the above equations depicts the percentage of susceptible, exposed, infected and recovered people in the source neighborhood. We assume each person has an equal possibility of traveling, regardless of their status. This gravity model assumes that mobility flows between neighborhoods are negatively correlated with travel distance and positively correlated with population size, which simultaneously captures rules of social interaction and travel distance. We set the infection rate β, quarantine rate κ and mobility level M as learnable parameters to estimate epidemic dynamics. Besides, we set other epidemiological parameters based on recent COVID-19 research (see Supplementary Table. S1 in SI Appendix).

M2 Theoretical analysis of complex growth curves

To evaluate our model's capacity for capturing complex growth curves of COVID-19 confirmed cases, we consider two extreme scenarios. First, we consider a complete lockdown scenario ($M \to 0$) in which coronavirus is contained to several initial neighborhoods. In this, the effective population is the combined population of these neighborhoods, which will be significantly smaller than the whole population ($N_{eff} \ll N$). It is equivalent to an exponential depletion of susceptible people resembling the self-containment mechanism in the recent SIR-X model, which has been proved to reproduce sub-exponential growth curves[6]. Moreover, our model provides a micro foundation to explain the sub-exponential growth of COVID-19 cases from the perspective of urban mobility behavior.

The second extreme case is the circumstance of highly efficient city-wide population mixing ($M \to \infty$). In this scenario, population mixing in each simulation epoch makes persons of different status distribute proportionately across neighborhoods based on the size of each subpopulation. Therefore, the epidemic development stage of each neighborhood runs a similar SEIR

model, but with different population sizes:

$$\frac{\mathrm{d}s^n}{\mathrm{d}t} = -\beta s^n e^n - \beta i^n e^n$$

$$\frac{\mathrm{d}e^n}{\mathrm{d}t} = (1-\omega_i)\beta s^n e^n + (1-\omega_i)\beta s^n i^n - \frac{1}{\tau}e^n$$

$$\frac{\mathrm{d}i^n}{\mathrm{d}t} = \omega_i\beta s^n e^n + \omega_i\beta s^n i^n + \frac{1}{\tau}e^n - \kappa i^n - \gamma i^n +$$

$$\frac{\mathrm{d}r^n}{\mathrm{d}t} = \kappa i^n + \gamma i^n,$$

where s^n, e^n, i^n, r^n have similar proportion composition in each neighborhood. For example, the number of susceptible people in the neighborhood can be computed as

$$s^n = \frac{s^n + e^n + i^n + r^n}{S + E + I + R}S$$
$$= p_n S,$$

where p_n is the proportion of population in neighborhood n.

Therefore, city-wide epidemic development is a linear sum of this set of homogeneous SEIR models. For example, coronavirus transmission (the decrease in susceptible population) occurs in each simulation epoch and can be computed as follows:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{\mathrm{d}s^1}{\mathrm{d}t} + \frac{\mathrm{d}s^2}{\mathrm{d}t} + \cdots + \frac{\mathrm{d}s^m}{\mathrm{d}t}$$
$$= -(\beta s^1 e^1 + \beta s^1 i^1) - (\beta s^2 e^2 + \beta s^2 i^2) - \cdots - (\beta s^m e^m + \beta s^m i^m)$$
$$= -(\beta SE + \beta SI)p_1^2 - (\beta SE + \beta SI)p_2^2 - \cdots - (\beta SE + \beta SI)p_m^2$$
$$= \left(\left(\sum_{i=1}^{m} p_i^2 \beta\right)SE + \left(\sum_{i=1}^{m} p_i^2 \beta\right)SE \right).$$

Similar equations can be derived for exposed, infected, and recovered populations. Therefore, we see that city-wide epidemic development is equivalent to an SEIR model with different parameters, where the equivalent infection rate is $\beta_{equ} = \left(\sum_{i=1}^{m} p_i^2 \beta\right)$. As a result, it can reproduce exponential growth curves from the standard SEIR model.

Our proposed metapopulation model lies between these two extreme cases, with mobility level parameter ranges from complete lockdown ($M \to 0$) to highly efficient population mixing ($M \to \infty$). Therefore, by changing the mobility level, our model can reproduce complex growth curves ranging from sub-linear to exponential (see Supplementary Figure 3 in SI Appendix). As a result, our model is provably expressive to capture the complex growth curves of COVID-19 confirmed cases.

Data Availability

The empirical data sets from cities in US, India and Brazil that support this study are available in GitHub, https://github.com/0oshowero0/COVID urban-mobility-model.

Code Availability

The source code for numeric simulation is available online: https://github.com/0oshowero0/COVID19-urban-mobility-model.

Acknowledgements

Author contributions

Fengli Xu, Yong Li, and James Evans jointly launched this research. Zhenyu Han performed the experiments and prepared the figures. Fengli Xu, Yong Li, Tao Jiang and James Evans provided the research outline, research design and critical revisions. All authors jointly analyzed the results and wrote the paper.

Additional information

Supplementary information
The online version contains supplementary material available at.

Competing interests
The authors claim no competing interests.

References

1. Komarova, N. L., Schang, L. M. & Wodarz, D. Patterns of the COVID-19 pandemic spread around the world: exponential versus power laws. *J. R. Soc. Interface* **17**, 20200518 (2020).

2. Singer, H. M. The COVID-19 pandemic: growth patterns, power law scaling, and saturation. *Phys. Biol.* **17**, 055001 (2020).

3. Hoseinpour Dehkordi, A., Alizadeh, M., Derakhshan, P., Babazadeh, P. & Jahandideh, A. Understanding epidemic data and statistics: A case study of COVID-19. *J. Med. Virol.* **17**, 868–882 (2020).

4. Hethcote, H. W. The mathematics of infectious diseases. *SIAM review* **42**, 599–653 (2000).

5. Brauer, F., Castillo-Chavez, C. & Castillo-Chavez, C. *Mathematical models in population biology and epidemiology*, vol. 2 (Springer, 2012).

6. Maier, B. F. & Brockmann, D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in china. *Science* **368**, 742–746 (2020).

7. Majra, D., Benson, J., Pitts, J. & Stebbing, J. SARS-CoV-2 (COVID-19) superspreader events. *J. Infect.* (2020).

8. Adam, D. *et al.* Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (sars-cov-2) infections in hong kong. *Res. Sq.* (2020).

9. Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* (2020).

10. Hou, X. *et al.* Intracounty modeling of covid-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proc. Natl. Acad. Sci.* **118** (2021).

11. Chiu, W. A., Fischer, R. & Ndeffo-Mbah, M. L. State-level needs for social distancing and contact tracing to contain covid-19 in the united states. *Nat. human behaviour* **4**, 1080–1090 (2020).

12. Ahmad, N. & Chauhan, P. State of data privacy during covid-19. *Computer* **53**, 119–122 (2020).

13. Anwari, N., Ahmed, M. T., Islam, M. R., Hadiuzzaman, M. & Amin, S. Exploring the travel behavior changes caused by the covid-19 crisis: A case study for a developing country. *Transp. Res. Interdiscip. Perspectives* **9**, 100334 (2021).

14. Stouffer, S. A. Intervening opportunities: A theory relating mobility and distance. *Am. Sociol. Rev.* **5**, 845–867 (1940).

15. Tobler, W. R. A computer movie simulating urban growth in the detroit region. *Econ. geography* **46**, 234–240 (1970).

16. Apple mobility trends reports. https://www.apple.com/covid19/mobility (2020).

17. Jay, J. *et al.* Neighbourhood income and physical distancing during the covid-19 pandemic in the united states. *Nat. human behaviour* **4**, 1294–1302 (2020).

18. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nat. physics* **6**, 818–823 (2010).

19. Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain covid-19 in china. *Nature* **585**, 410–413 (2020).

20. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in europe. *Nature* **584**, 257–261 (2020).

21. Koyama, T., Platt, D. & Parida, L. Variant analysis of sars-cov-2 genomes. *Bull. World Heal. Organ.* **98**, 495 (2020).

22. Moreno, G. K. *et al.* Revealing fine-scale spatiotemporal differences in SARS-CoV-2 introduction and spread. *Nat. Commun.* **11**, 1–13 (2020).

23. Zhang, J. *et al.* Changes in contact patterns shape the dynamics of the COVID-19 outbreak in china. *Science* **368**, 1481–1486 (2020).

24. The State Council of the People's Republic of China. The announcement from wuhan's headquarter on the novel coronavirus prevention and control [in chinese] (2020).

25. WorldPop, global high resolution population denominators project. https://www.worldpop.org/ (2020).

26. Johns Hopkins University Center for Systems Science and Engineering. 2019 novel coronavirus COVID-19 (2019-nCoV) data repository by johns hopkins CSSE (2020). https://github.com/CSSEGISandData/COVID-19 (2020).

27. COVID19India. Covid19india. https://www.covid19india.org/ (2021).

28. Ministério da Saúde. Coronavírus brasil (in Portuguese). https://covid.saude.gov.br/ (2021).

29. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).

30. Wang, L. *et al.* Inference of person-to-person transmission of COVID-19 reveals hidden super-spreading events during the early outbreak phase. *Nat. Commun.* **11**, 5006 (2020).

31. Cirillo, P. & Taleb, N. N. Tail risk of contagious diseases. *Nat. Phys.* **16**, 606–613 (2020).

32. Wong, F. & Collins, J. J. Evidence that coronavirus superspreading is fat-tailed. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 29416–29418 (2020).

33. Gastwirth, J. L. The estimation of the lorenz curve and gini index. *Rev. Econ. Stat.* **54**, 306–316 (1972).

34. NYC Department of Health and Mental Hygiene. NYC Health Data. https://www1.nyc.gov/site/doh/covid/covid-19-data.page (2021).

35. Frieden, T. R. & Lee, C. T. Identifying and interrupting superspreading events—implications for control of severe acute respiratory syndrome coronavirus 2. *Emerg. infectious diseases* **26**, 1059 (2020).

36. Bouffanais, R. & Lim, S. S. Cities — try to predict superspreading hotspots for COVID-19 (2020).

37. Spelta, A. & Pagnottoni, P. Mobility-based real-time economic monitoring amid the covid-19 pandemic. *Sci. Reports* **11**, 1–15 (2021).

38. Solís Arce, J. S. *et al.* Covid-19 vaccine acceptance and hesitancy in low-and middle-income countries. *Nat. Medicine* **27** (2021).

39. Machingaidze, S. & Wiysonge, C. S. Understanding covid-19 vaccine hesitancy. *Nat. Medicine* 1–2 (2021).

40. Volpert, V., Banerjee, M. & Petrovskii, S. On a quarantine model of coronavirus infection and data analysis. *Math. Model. Nat. Phenom.* **15**, 24 (2020).

41. World Health Organization. Public health surveillance for COVID-19: interim guidance. *World Heal. Organ.* (2020).

42. Prata, D. N., Rodrigues, W. & Bermejo, P. H. Temperature significantly changes COVID-19 transmission in (sub)tropical cities of brazil. *Sci. Total. Environ.* **729**, 138862 (2020).

43. Sabat, I. *et al.* United but divided: Policy responses and people's perceptions in the EU during the COVID-19 outbreak. *Heal. Policy* **124**, 909–918 (2020).

44. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21484–21489 (2009).
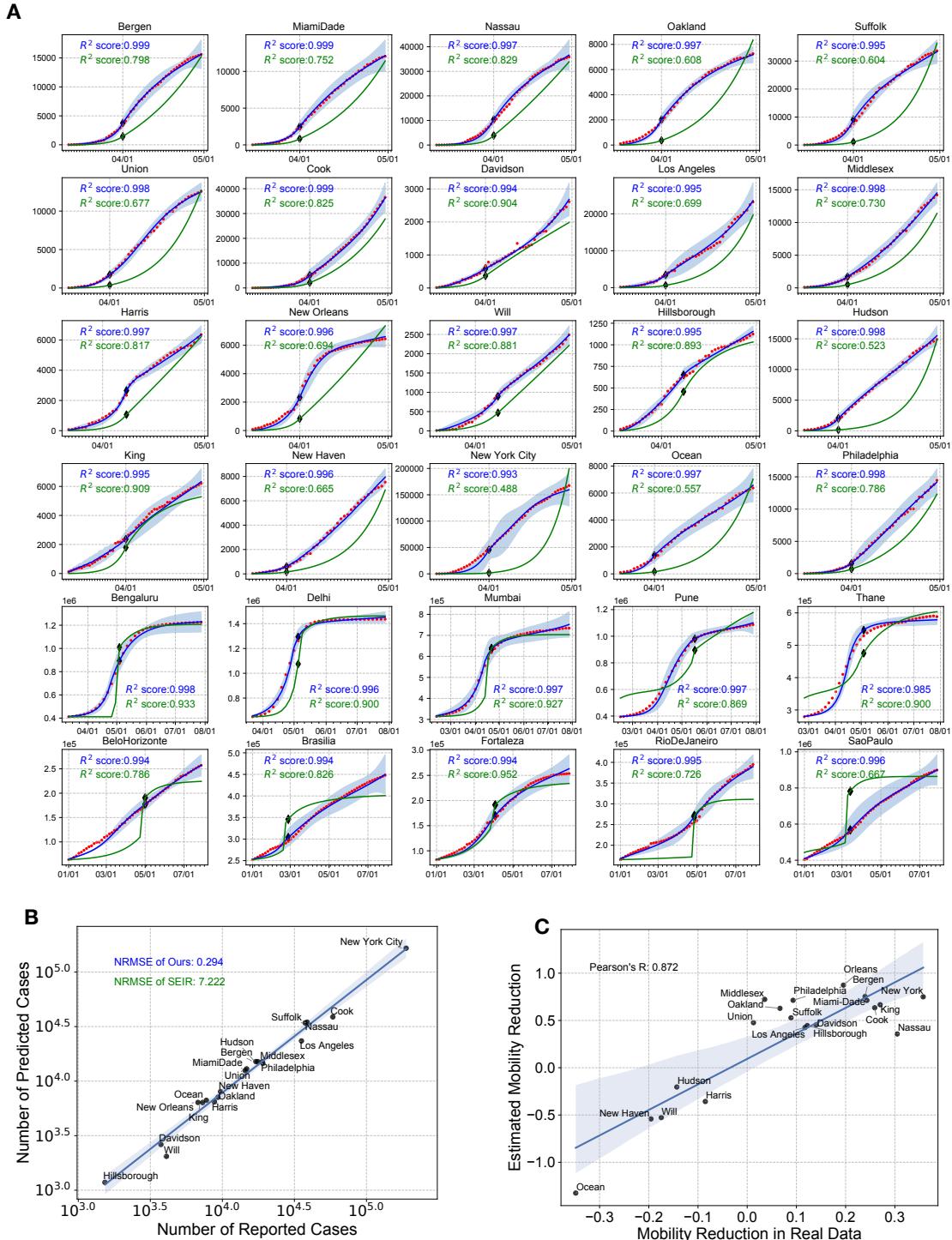
Figure 1. Predicting COVID-19 growth in the 30 most infectious counties and cities in the U.S., India and Brazil. (A) Reproducing growth curves for COVID-19 confirmed cases. Red dots are empirical observations of confirmed cases, blue lines are growth curves estimated by our model, and green lines are growth curves estimated by the standard SEIR model. The shaded area represents the 99% confidence interval. Diamond markers denote the time of policy interventions. Our model can better reproduce empirically observed COVID-19 growth curves, resulting in significantly higher accuracy. **(B)** Evaluating our model with the task of predicting future confirmed cases within 14 days for all 20 U.S. counties. The normalized root mean square error (NRMSE) of our model is 0.294, compared to 7.222 for the standard SEIR model. **(C)** Correlation between empirical mobility reduction and model estimation before and after announcement of nationwide emergency status around March 25 for U.S. counties. Our model can accurately estimate mobility behavior change with a Pearson correlation coefficient of 0.872.
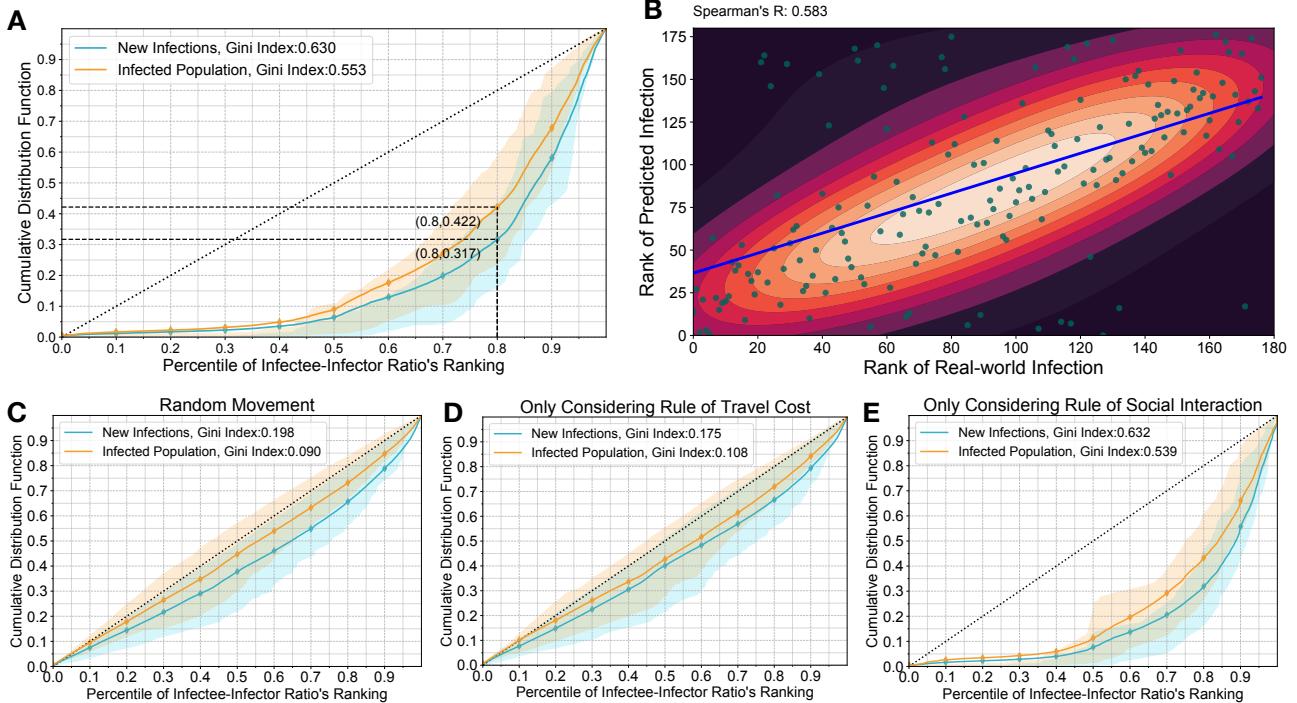
Figure 2. Reproducing and rationalizing superspreading events in cities. (**A**) The distribution of infected population and occurrences of new infections across neighborhoods ranked by infectee-infector ratio. If the infected populations and new infections are distributed evenly, they are expected to follow the black dashed line. Our model reproduces a significant uneven distribution with the bottom 80% neighborhoods only accounting for 42.2% infected persons (orange line) and 31.7% new infections (blue line). Shaded areas represent the interquartile range for all urban counties. The Gini index for the distributions of new infections and infected populations are 0.630 and 0.553 respectively, which suggests the infection risk is spatially heterogeneous in cities. (**B**) The correlation between predicted infection risk and real-world observation across neighborhoods in New York city. We find a high correlation between the rank of predicted confirmed cases and real-world observed confirmed cases across New York neighborhoods (Spearman's R = 0.583), which suggests the spatially heterogeneous infection risk predicted by our model is consistent with empirical observations. (**C**) When substituting the mobility model with random movement, the spatial unevenness of the distributions of infected population and new infections is largely reduced. (**D**) When only considering the rule of travel cost, we reproduce a similar level of spatial unevenness as the random movement. (**E**) When only considering the rule of social interaction, we reproduce similar level of spatial unevenness as the complete model in (A).
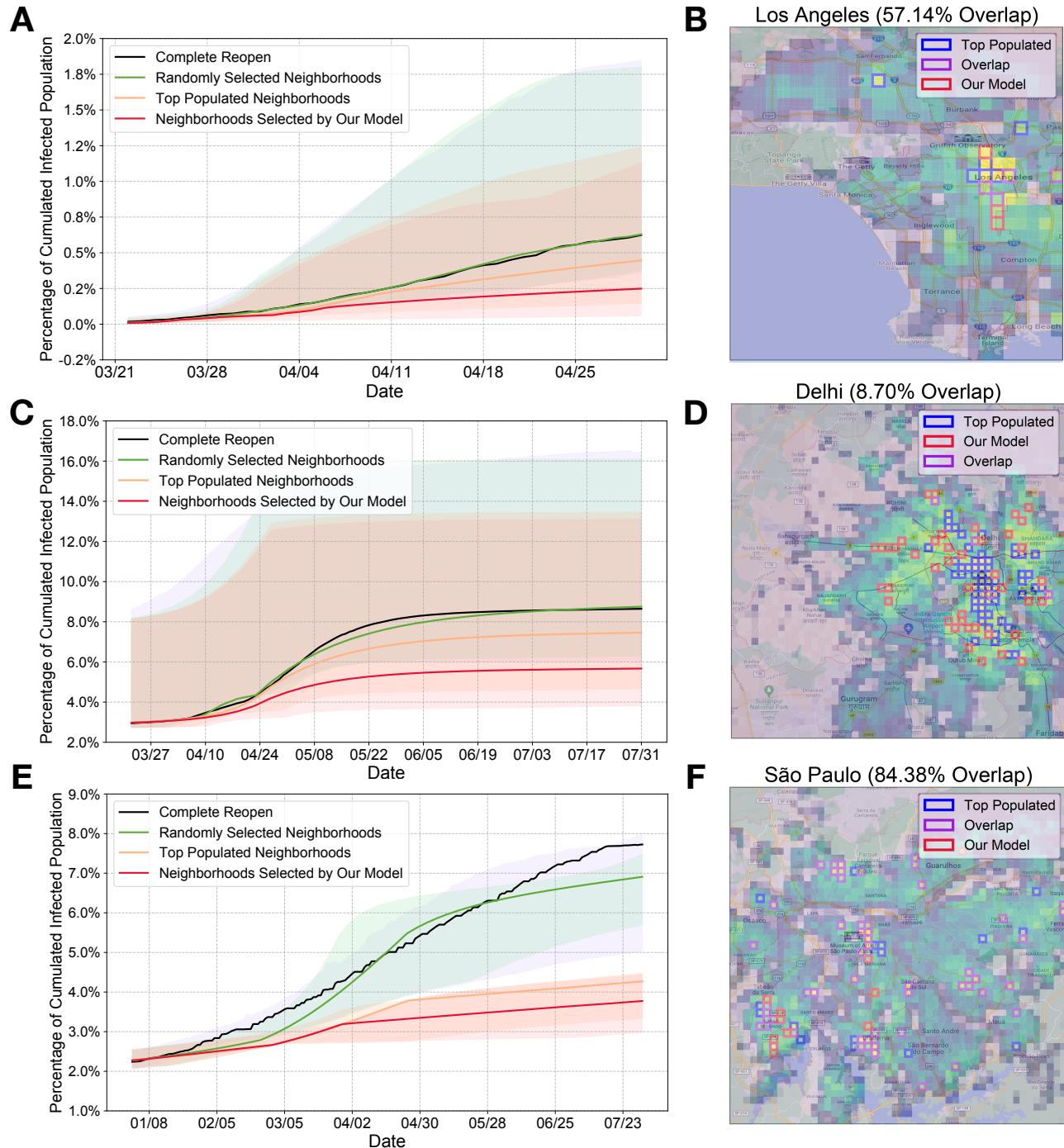
Figure 3. The effectiveness of different location-aware reopening policies. **(A, C, E)** Performance of different reopening policies in U.S. counties, Indian and Brazilian cities. We calculate the percentage of infected population in cities of each country. The reopening policy informed by our model leads to a 60.13% decrease in cumulative infections in U.S. counties, which significantly outperforms the strategy of not reopening top populated neighborhoods (28.49%). Besides, the reopening policy informed by our model leads to a 34.41% and 51.14% decrease in cumulative infections in India and Brazil, respectively. (Solid lines denote the median value of all counties and cities, and shaded areas represent interquartile ranges). **(B, D, F)** Visualization of neighborhoods selected for remaining under mobility control in different reopening policies in Los Angeles, Delhi and São Paulo. The neighborhoods selected by our model differ significantly with the top populated strategies. Our model jointly considers population distribution and urban mobility behavior, which better capture the most infectious locations for both single-center (LA) and multi-center (Delhi and São Paulo) cities.
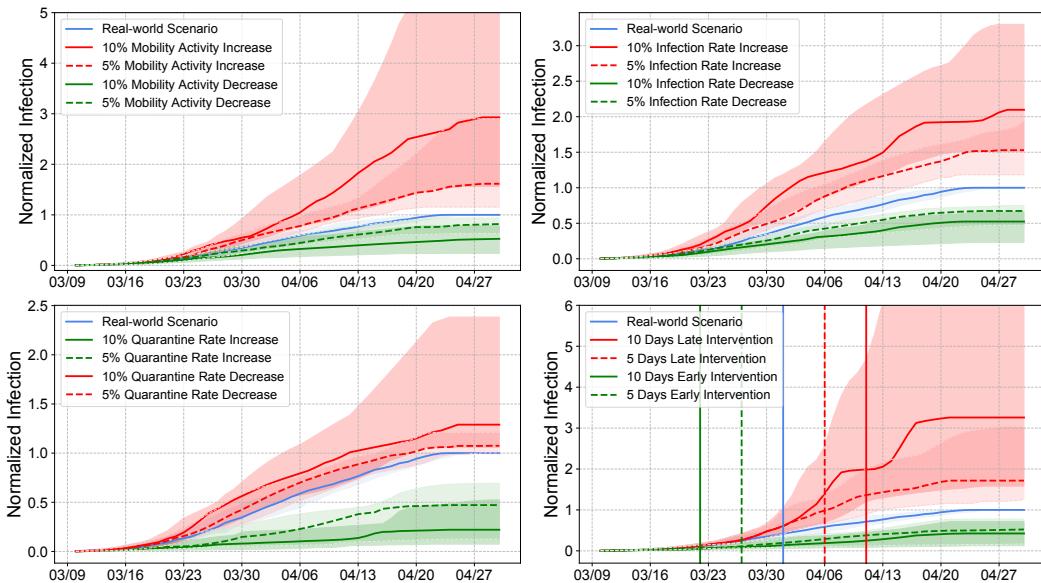
Figure 4. Estimated growth curves for COVID-19 infections in U.S. counties under different scenarios. All Figures show the median value of confirmed cases across 20 U.S. counties under different scenarios normalized by the overall real-world confirmed cases, where the shaded areas denote interquartile range. (**A**) Estimated growth curves with different urban mobility activities. (**B**) Estimated growth curves under different quarantine rates. (**C**) Estimated growth curves under different infection rates. (**D**) Estimated growth curves with different intervention dates. The vertical lines denote the dates of intervention policy.