scientific reports

OPEN

Model predicted human mobility explains COVID-19 transmission in urban space without behavioral data

Zhenyu Han[1,5], **Fengli Xu**[1,2,5], **Yong Li**[1]✉, **Tao Jiang**[3] & **James Evans**[2,4]✉

The SARS-CoV-2 virus is primarily transmitted through in-person interactions, and so its growth in urban space is a complex function of human mobility behaviors that cannot be adequately explained by standard epidemiological models. Recent studies leveraged fine-grained urban mobility data to accurately model the viral spread, but such data pose privacy concerns and are often difficult to collect, especially in low- and middle-income countries (LMICs). Here, we show that the metapopulation epidemiological model incorporated with a simple gravity mobility model can be sufficient to capture most of the complex epidemic dynamics in urban space, largely reducing the need for empirical mobility data. Extensive experiments on 30 cities in the United States, India and Brazil show that our model consistently reproduces complex, distinctive COVID-19 growth curves with high accuracy. It also provides a theoretical explanation of the emergence of urban "superspreading", where a few high-risk neighborhoods account for most subsequent infections. Furthermore, with the aid of the proposed framework, we can inform mobility-aware travel restrictions to achieve a better balance between social cost and disease prevention, which facilitates sustainable epidemic control and supports the gradual transition to a post-pandemic world.

As a respiratory infectious disease, COVID-19 predominantly spreads through person-to-person contact and interaction. Consequently, its transmission mechanism in urban space is inherently complex. This complexity involves both macro- and micro-scopic properties. These range from exponential to sub-linear increases in confirmed cases[1–3]to the "superspreading phenomena"[4, 5]. Conventional homogeneous mixing models, such as the Susceptible-Exposed-Infectious-Recovered (SEIR) model[6], can only explain exponential growth curves in the absence of mobility data, which contradicts empirical observation. Going beyond homogeneous models, researchers introduced the metapopulation framework, which breaks the homogeneous mixing assumption by dividing the target region into several metapopulation patches[7, 8]. It demonstrates superior performance in large-scale epidemic modeling, where population flows between patches are typically modeled by a random diffusion process[9, 10]. Nevertheless, when delving into the intra-city dynamics, such a straightforward mobility rule fails due to high heterogeneity in fine-grained human mobility. To solve this problem, researchers have begun to leverage individual-level cell phone trajectories[11, 12], survey-based contact matrices[13, 14], or smart card check-in records[15–17]to depict intra-city human mobility. Albeit significant improvements in epidemic modeling, these methods pose high demands on mobility behavioral data[18], which raises significant privacy concerns[19].

To overcome the above limitations, we provide new evidence that combining a metapopulation epidemiological model with a parsimonious gravity mobility model, rather than empirical mobility data, can be sufficient to capture most of the complex epidemic dynamics. This finding allows us to provide a simple yet novel explanation for the complex epidemic dynamics with minimum data demand. Specifically, we adopt a metapopulation framework that models the urban space as a set of spatially distributed neighborhoods and run a separate SIR variant model to capture the local transmissions in each neighborhood. To capture the person-to-person interactions, we principally integrate the metapopulation epidemiological framework with a parsimonious gravity mobility model, which predicts mobility flows with the attractions of social interaction opportunity[20]and the cost of travel distance[21]. The simplicity and deep fusion of the proposed model allow us

[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. [5]Zhenyu Han and Fengli Xu contributed equally. ✉email: liyong07@tsinghua.edu.cn; jevans@uchicago.edu

to develop a unified model calibration process that simultaneously fits mobility parameters and epidemiological parameters with the COVID-19 incidence records. It largely reduces the data demand in previous works that either fit complicated standalone mobility models with behavioral data[22]or adopt a completely data-driven approach[23], which is especially important to ease data privacy concerns and applications in LMICs.

Experiments on real-world data collected from 30 of the most infectious counties and cities in the United States, India and Brazil show that the proposed 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[24] (Pearson's $R = 0.872$), suggesting the effectiveness of our model in capturing empirical mobility trends. 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 subsequent 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. Therefore, by focusing on those regions predicted to have the highest infection risk, we can facilitate mobility-aware travel restrictions that result in a more desirable balance of social cost and disease prevention[25]. Furthermore, we perform a cost-effective evaluation of intervention policies under different scenarios, e.g., different levels of mobility activity, and different infection rates from new variants. We demonstrate that complex epidemic growth patterns can be adequately captured based on the predicted urban mobility, which allows us to provide practical and sustainable epidemic control policy recommendations to LMICs even without behavior data, paving the way for a safe transition to a post-pandemic world.

Results

Urban mobility driven metapopulation epidemiological model

To incorporate urban mobility with epidemic transmission, we adopt a metapopulation scheme that divides urban space into several neighborhoods (1 km × 1 km grid cells, see Methods and Supplementary Fig. 1 in SI Appendix), providing a basis for modeling spatial heterogeneity within cities. Each neighborhood contains a sub-population, and maintains a separate SIR variant with its own susceptible (S), asymptomatic (A), infectious (I), and recovered (R) states (see Supplementary Fig. 1 in SI Appendix). We adopt the implicit assumption in SIR models for homogeneous intra-neighborhood mobility, where each individual has an equal probability of contacting other individuals in the same neighborhood. Moreover, our theoretical analysis shows a parsimonious gravity model is sufficient to capture the inter-neighborhood urban mobility that explains complex COVID-19 growth curves (see Methods for details). Therefore, we propose to principally integrate the gravity mobility model into the metapopulation epidemiological model. The simplicity of the model design allows us to jointly fit mobility parameters and epidemiological parameters with the COVID-19 growth curves, which largely reduces the dependence on fine-grained empirical mobility data. Specifically, the designed gravity model principally integrates the fundamental rules of social interaction and travel cost to predict urban mobility flows. The rule of social interaction emphasizes the attraction of social interaction opportunities of Stouffer's law[20], which can be approximated by the destination neighborhood's population. Besides, the rule of travel cost captures the locality of urban movements[26]- Tobler's first law of geography[21], where people tend to visit nearby places. The parameters of the mobility and epidemiological models are calibrated in a coupled way, which captures the complex interplay between urban mobility and epidemic transmission. Leveraging this framework, we successfully introduce the heterogeneous population mixing by the mobility model, further extending the understanding of disease transmission patterns in regions without empirical mobility data.

We evaluate the proposed model by 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 changing mobility behavior due to non-pharmaceutical interventions such as stay-at-home orders, we segment the simulation period based on the implementations of nationwide intervention policies of these countries, and then fit separate parameters for each segmented period (Supplementary Fig. 2 in SI Appendix). Ground truth data of confirmed cases are collected from official statistics released by each country[27–29] (see Supplementary Table.S1, S2, S3 in SI Appendix). The 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). Through principally capturing the heterogeneous intra-city interactions, our model can accurately predict distinctive growth curves with R^2 scores consistently above 0.980 for all cities (Fig. 1A, blue lines). By contrast, the homogeneous mixing SIR variant 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). It follows the classical assumption that people across the entire city homogeneously mix with a unified contact rate. The much lower accuracy of the baseline model validates the conclusion that metapopulation models with heterogeneous contact patterns generally perform better than homogeneous mixing models[7,8]. Our model can also achieve high accuracy in predicting daily incidences with R^2scores up to 0.936, representing a minimum 12.3% improvement over the baseline model across all cities (see Supplementary Fig. 3 in SI Appendix). Besides, we evaluate the robustness of our framework by replacing the gravity mobility model with the radiation mobility model[30]. The results are shown in Supplementary Fig. 10, where we observe negligible performance differences compared with the gravity mobility model in Fig. 1. We also investigate how neighborhood scales affect the model performance. In Supplementary Fig. 11, we observe that using larger neighborhood scales leads to slightly lower accuracy, where the R^2drops less than 5% in 10 km x 10 km setting. These results combined to demonstrate the robustness of our findings. Moreover, we theoretically demonstrate that the proposed model is able to capture complex forms of transmission patterns, where the SIR-X model that predicts sub-exponential growth[31] is a special case (see Methods for details). We find that different levels of urban

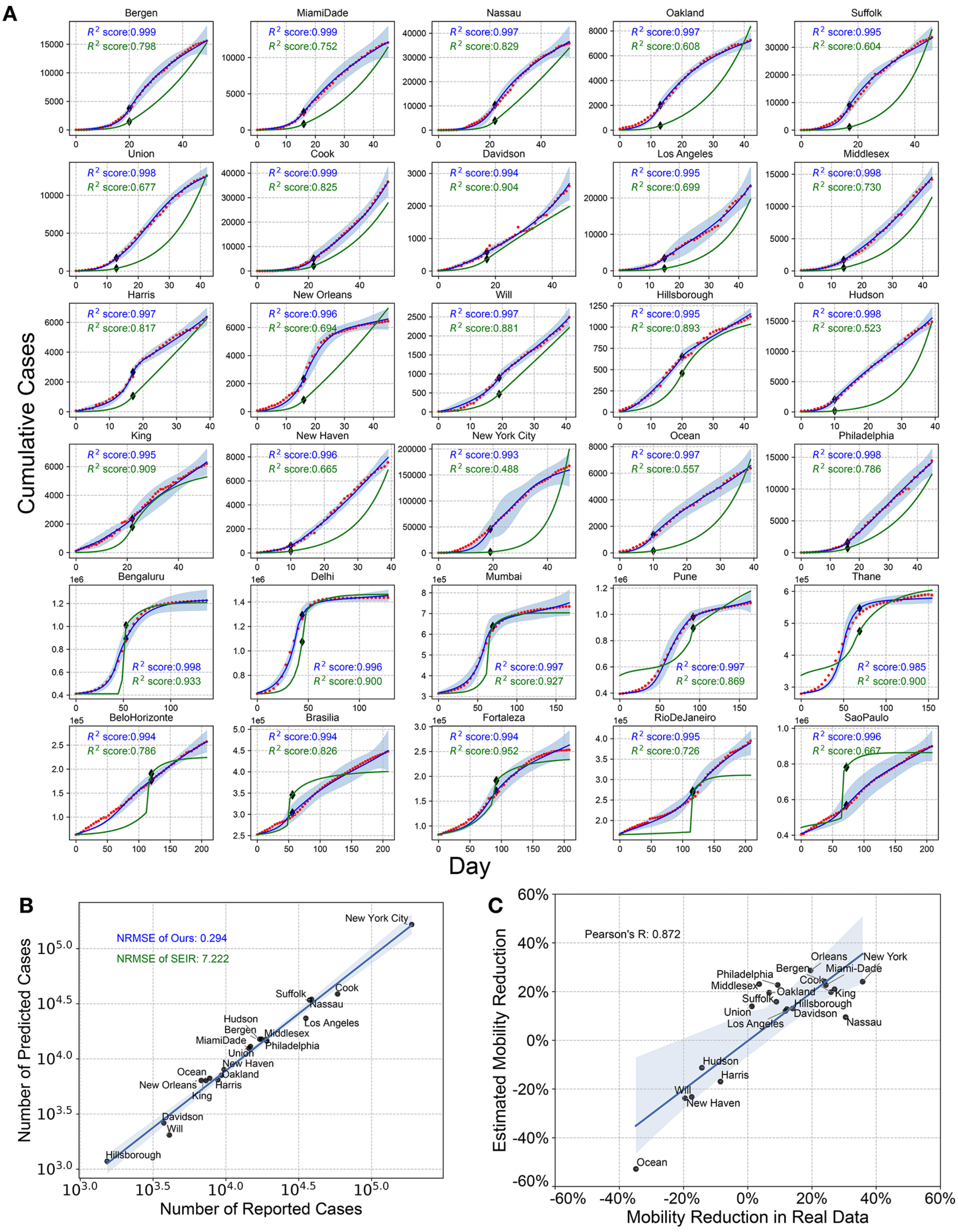

Fig. 1. Predicting COVID-19 growth in the 30 most infectious counties and cities in the U.S., India and Brazil. **(A)** Reproducing growth curves for COVID-19 confirmed cases. Red dots are empirical observations of confirmed cases, blue lines are growth curves estimated by our model, and green lines are growth curves estimated by a homogeneous mixing SIR variant model. Diamond markers denote the time of policy interventions. It demonstrates that the proposed model can effectively reproduce empirically observed COVID-19 growth curves, resulting in significantly higher accuracy. **(B)** Evaluating the proposed model with the task of predicting future confirmed cases within 14 days for all 20 U.S. counties. **(C)** Correlation between empirical mobility reduction and model estimation before and after the behaviour change around the first week in April for U.S. counties. Our model can accurately estimate mobility behavior change with a Pearson correlation coefficient of 0.872. Shaded areas in all the figures represent the 99% confidence interval.

mobility have a significant impact on the shape of growth curves (Supplementary Fig. 4 in SI Appendix), which further justifies the need to incorporate human mobility to characterize the complex growth of COVID-19 cases. We further evaluate the model by predicting future confirmed cases in all 20 U.S. counties in Fig. 1B, where the model demonstrates great performance of a low normalized root mean square error (NRMSE) of 0.294. Moreover, Fig. 1C shows mobility behavior change manifests 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[24], which demonstrates that the proposed model can accurately estimate empirical mobility behavior changes without mobility data (see Methods for details). These results suggest that our model can effectively alleviate the privacy concerns in data-intensive epidemiological modeling and ease the data collection burden in LMICs.

Reproducing and rationalizing superspreading events in urban space

Superspreading events have been widely observed in epidemics like SARS, Measles, and Smallpox[4], as well as COVID-19, where a small portion of infected people and locations are responsible for a disproportionately high number of secondary infections. Researchers have identified accumulating superspreading events of COVID-19 through phylogenetic analysis[32, 33], statistical evaluation[34, 35]and contact tracing based on mobility behavioral data[5]. Superspreading can neither be adequately explained by the standard SEIR model nor the recently proposed SIR-X model[31], which both rely 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 secondary infections generated by each infected individual deviate from the general population[4, 5] with a negative binomial model. These works neglect the behavioral mechanism underlying the superspreading phenomenon, e.g., spatially heterogeneous population mixing. By incorporating the human mobility model, our proposed model can serve as an analytical framework to further explain the underlying mechanisms behind the complex transmission dynamics.

We use the infectee-infector ratio (IIR) to measure the severity of superspreading in different neighborhoods, defined as the number of secondary infections per previously infected person at each timestamp. A higher IIR in the neighborhood demonstrates that each infected person can cause more infections at this timestamp, representing a stronger superspreading phenomenon. By ranking all neighborhoods in a city based on their average IIR throughout the epidemic period, we can clearly observe the superspreading phenomenon from the cumulative distribution function of the infected population and secondary infections in 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. However, our model reproduces a highly skewed distribution where the most infectious 20% of neighborhoods are responsible for 68.3% of secondary infections. We use the Gini index[36]to further quantify the unevenness in the spatial distribution of the infected population and secondary infection occurrences, where a higher value represents stronger superspreading in the range of 0–1. We observe high Gini indexes of 0.553 and 0.630 for the infected population and secondary infections respectively, indicating a high degree of superspreading phenomenon. To align the intra-city prediction with real-world data, we further validate model-informed spatially heterogeneous infection risk with fine-grained confirmed cases in New York City[37] (Fig. 2B). We observe a high correlation between predicted and real-world infections in each neighborhood (Spearman's $R = 0.583$). These results indicate our model can effectively extend the standard epidemiological model to reproduce superspreading events, and the prediction is consistent with real-world observation without the support of empirical mobility data.

After successfully reproducing the superspreading of disease, we further investigate the underlying behavior mechanism behind such phenomenon leveraging the urban mobility model. First, recent research found empirical evidence that the uneven urban population distribution might contribute to the spatially heterogeneous infection risk in urban spaces[38], and the empirically observed urban population distributions follow a highly uneven exponential distribution (Supplementary Fig. 5 in SI Appendix). Specifically, if the urban population is distributed evenly and urban movements are completely random, our model will degenerate into an SIR model variant with homogeneous population mixing, which predicts the infection risk to be spatially homogeneous. To examine the effect of uneven urban population distribution, we substitute the gravity model by random urban movement in Fig. 2C. We observe largely reduced spatial unevenness with Gini index decreases to 0.090 and 0.198 for infected population and secondary infections. It indicates the uneven urban population distribution alone cannot fully explain the superspreading events in urban space. Second, the urban movement is considered another potential cause for the superspreading events in urban spaces[39]. To examine the effect of social interaction and travel cost, we evaluate two variants of the 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 secondary infections and infected populations (see Fig. 2E), which are similar to the results of the complete model (Fig. 2A). These observations suggest that superspreading events are a joint result of uneven urban population distribution and heightened population flows naturally associated with populous neighborhoods, which reveals the key urban dynamics behind the widely observed superspreading phenomenon.

Informing mobility-aware travel restriction policies without mobility behavioral data

As a result of superspreading phenomena, highly skewed spatial distributions of COVID-19 cases suggest the potential to design targeted, mobility-aware travel restrictions that can more effectively curb coronavirus spread while minimizing social cost. Specifically, due to the emergence of new coronavirus variants and the inequity distribution of vaccine supplies, strict non-pharmaceutical interventions might happen again in many

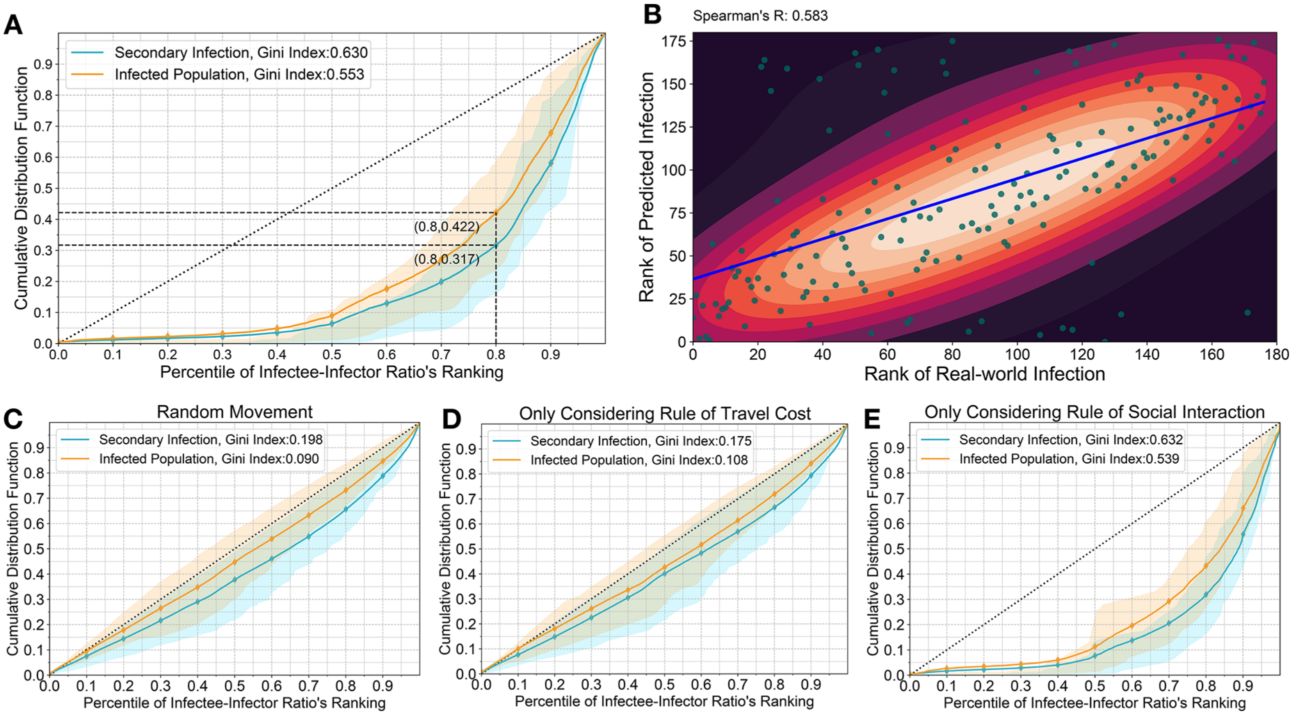

Fig. 2. Reproducing and rationalizing superspreading events in cities. **(A)** The distribution of infected population and occurrences of secondary infections across neighborhoods ranked by the average infectee-infector ratio. If the infected populations and secondary infections are distributed evenly, they are expected to follow the black dashed line. Our model reproduces a significantly uneven distribution with the bottom 80% neighborhoods only accounting for 42.2% infected persons (orange line) and 31.7% secondary infections (blue line). Shaded areas represent the interquartile range for all urban counties. The Gini index for the distributions of secondary 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. The color code represents 10 levels of kernel density estimates, where brighter color represents higher density. **(C)** When substituting the mobility model with random movement, the spatial unevenness of the distributions of infected population and secondary 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 a similar level of spatial unevenness as the complete model in (A).

LMICs, posing a pressing need for the design of better travel restriction policies that can facilitate gradual and sustainable epidemic control across communities[40]. Previous works leverage large-scale empirical mobility data to quantify high-risk locations[11, 23], while we argue that model-predicted human mobility flow could be a good substitution to inform such policies when mobility behavioral data is not available. Here, we leverage our model to inform mobility-aware travel restrictions 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 the city population under travel restriction (5% of the city population for U.S. counties), then evaluate the policy's efficacy under the same social cost budget in curbing disease spread as the percentage of confirmed cases that would be prevented compared to the no intervention scenario.

Populous neighborhoods tend to be more vulnerable to coronavirus transmission[38] (Spearman's $R=0.74$, see Supplementary Fig. 6 in SI Appendix). A widely adopted policy, referenced as *Top Populated* policy, prioritizes the introduction of travel restrictions in the most populated neighborhoods to make the most of the social cost budget. Under the *Top Populated* policy, 28.49% of overall confirmed cases can be avoided compared to no intervention, which outperforms the baseline of the *Randomly Selected* policy, which randomly chooses neighborhoods for mobility control under the same social cost budget (Fig. 3A). Nevertheless, infection risk involves factors beyond neighborhood population. If we control mobility in neighborhoods with the largest predicted infection risks by our model, overall confirmed cases can be reduced by 60.13%. We highlight how considering the population factor alone is not enough to efficiently control epidemic transmission given the same social cost budget because it ignores the diffusion effect of disease transmission, where the disease is more likely to spread to nearby neighborhoods even in a relatively small population. As an intuitive example, *Top Populated* policy will restrict faraway population centers, although the disease outbreak might not influence these areas.

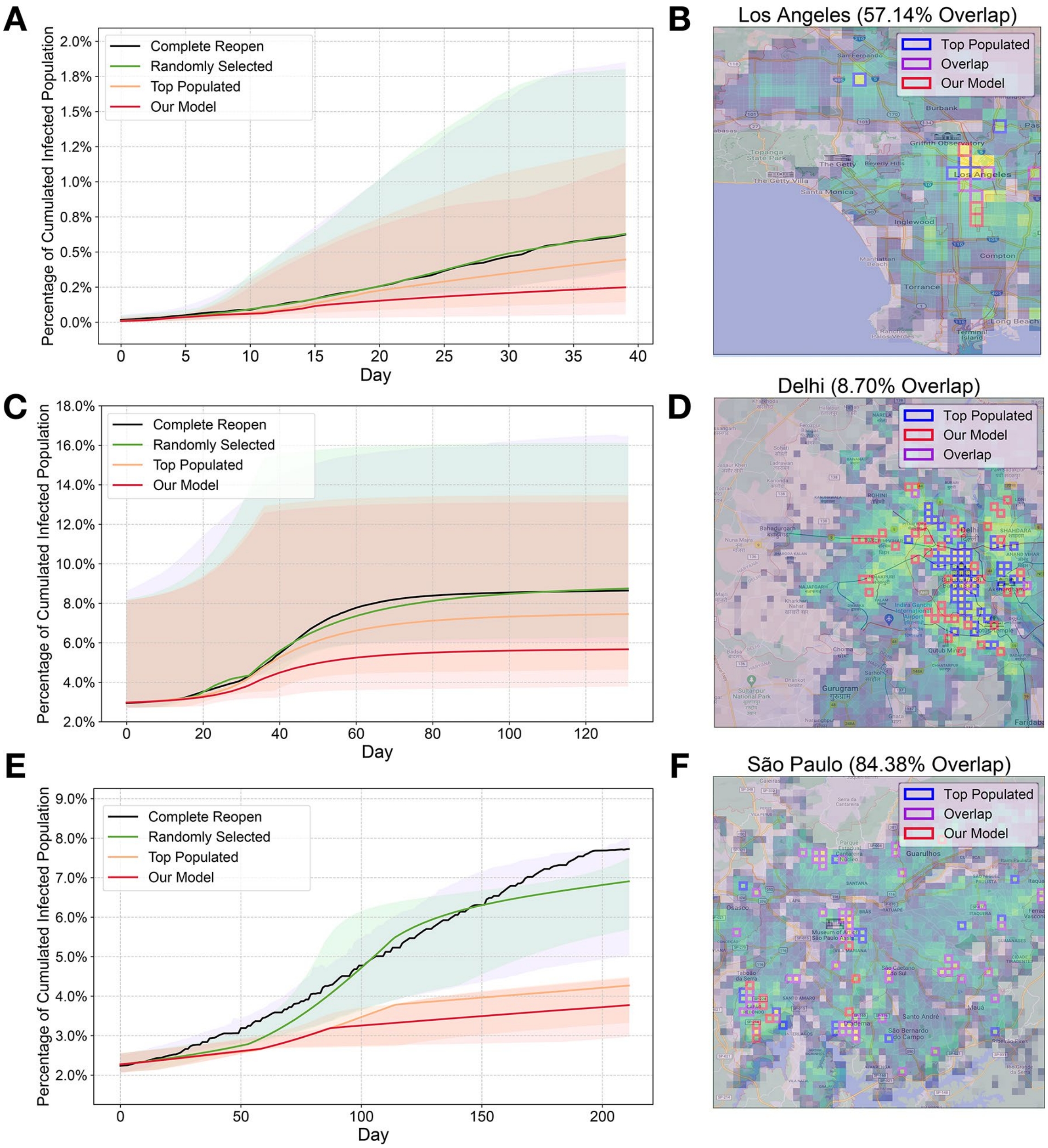

Fig. 3. The effectiveness of different mobility-aware travel restrictions. **(A, C, E)** Performance of different policies in U.S. counties, Indian and Brazilian cities. We calculate the percentage of infected population in cities of each country. The policy informed by our model leads to a 60.13% decrease in cumulative infections in U.S. counties, which significantly outperforms the *Top Populated* policy (28.49%). Besides, the 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 under mobility control in different restriction policies in Los Angeles, Delhi and São Paulo. The neighborhoods selected by our model differ significantly with *Top Populated* policy. Our model jointly considers population distribution and human mobility behavior in urban spaces, which better capture the most infectious locations for both single-center (LA) and multi-center (Delhi and São Paulo) cities.

Similar observations are made for Indian and Brazilian cities (Fig. 3C, E), where policies informed by our model consistently outperform the baselines of *Top Populated* and *Randomly Selected* policies (see Supplementary Table S4, S5, S6 in SI Appendix). These results suggest that our model can inform the design of substantially more cost-effective, mobility-aware travel restrictions.

To examine sources of differential efficacy, we visualize the neighborhoods under travel restriction in three cities (Fig. 3B, D and F). The case of Los Angeles county (L.A., in Fig. 3B) illustrates how the baseline policy focused only on the population would miss the high infection risk that occurs in less populous neighborhoods

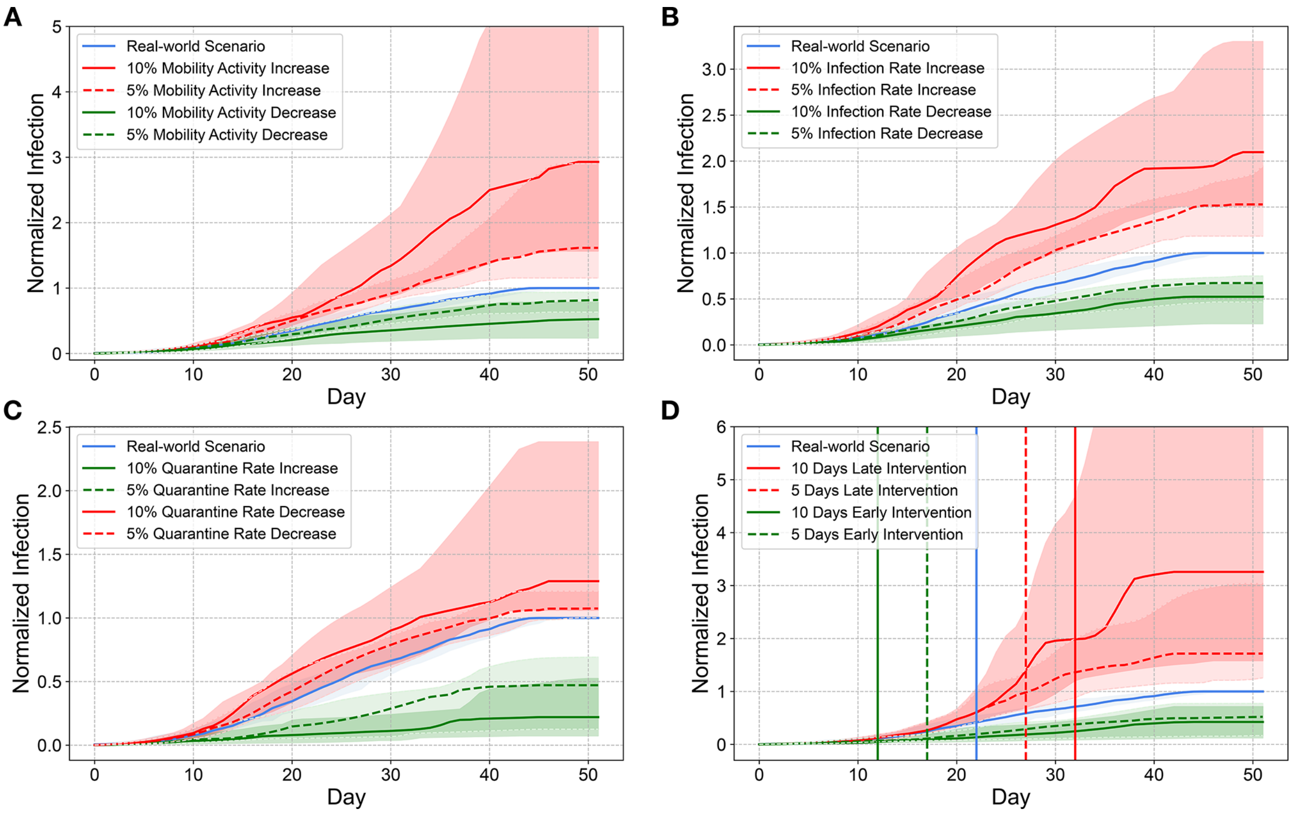

Fig. 4. Estimated cumulative growth curves for COVID-19 infections in U.S. counties under different scenarios. The figures show the median value of confirmed cases across 20 U.S. counties under different human mobility activities **(A)**, infection rates **(B)**, quarantine rates **(C)**, and intervention dates **(D)**. We normalize the infection in each county according to the real-world number of confirmed cases at the end of the studied period. Shaded areas represent interquartile ranges for all the counties, and the vertical lines denote the dates of intervention policy.

around central L.A. because it cannot model these neighborhoods' mobility flows with nearby, densely populated neighborhoods. For cities in LMICs such as Delhi and São Paulo (Fig. 3D, F), we find that policies informed by our model still outperform baseline policies, but with different overlaps. Delhi has a compact city center with a much denser population compared with the peripheral area, which hinders the *Top Populated* policy to capture possible outbreaks in the peripheral area. Hence, the overlap between selected locations of *Top Populated* policy and our model is small (8.70%). On the contrary, the population of São Paulo is distributed more evenly across urban space, where population density peaks are far away from each other, making the mobility connection between them relatively small. Therefore, the location's importance is mainly determined by population size. As a result, our model generates largely similar location choices as the *Top Populated* policy (84.38%). Departing from a static population distribution, previous research pointed out that urban mobility structure-encompassing mobility flows that interconnect various neighborhoods-greatly shapes epidemic dynamics and the effectiveness of NPIs. Specifically, cities with a centralized mobility distribution are more vulnerable to the epidemic, but the benefit from travel restrictions is greater[41]. This phenomenon, combined with changing mobility patterns due to the epidemic[42], firmly supports our argument that dynamic, mobility-aware guidance is critical for epidemic control. Our results further demonstrate that integrating the metapopulation epidemiological model with a parsimonious gravity mobility model is adequate to inform the design of practical mobility-aware travel restrictions in the absence of fine-grained mobility behavioral data, which is especially important to improve epidemic control in LMICs. More detailed analysis suggests the travel restrictions informed by our model can also effectively alleviate the superspreading events in cities (Supplementary Fig. 7 in SI Appendix).

Simulating the epidemic development under different scenarios

In light of the development trajectory of the COVID-19 pandemic, the viral spread dynamics shift with the emergence of new variants[43], changes in urban mobility behavior[44], and the implementation of different social distancing policies[45]. As a general framework that incorporates human mobility with virus spread, our model can simulate epidemic development under different scenarios with distinct parameter settings. We evaluate the scenarios of varying mobility, infection rates, quarantine rates and intervention times respectively (Fig. 4A-D). First, we investigate how mobility affects disease transmission by assuming different levels of mobility in our model. 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 ~ 7.24), and 10% less mobility will halve the number of cases (0.525 times; IQR 0.237 ~ 0.783). As a point of comparison, the average observed mobility drop in U.S. counties during the pandemic is approximately 30.3% (Supplementary Fig. 8 in SI Appendix, with IQR 0.227 ~ 0.392), which suggests the mobility behavior change during the pandemic was effective in curbing the spread of coronavirus. Reducing the infection rate is also a promising direction to control COVID-19 pandemic, where a 10% decrease in infection rate nearly halves infections (0.524 times; IQR 0.231 ~ 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[46, 47]or implementing a compulsory mask-wearing policy in public places. Quarantine is another widely adopted approach to contain the spreading of coronavirus[48], 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 ~ 2.38), while a 10% increase in quarantine rate will prevent approximately 78% of citywide infections (0.221 times; IQR 0.0739 ~ 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 pandemic are considered critical for combating the virus[49]. 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 ~ 9.53), while adopting policies 10 days earlier reduces the infections by 57.5% (0.425 times; IQR 0.12 ~ 0.71). These analyses reveal how timely interventions such as mobility reduction, social distancing and quarantine can play important roles in curbing the spread of SARS-CoV-2 virus.

Discussion

The complex transmission process of contagion diseases in urban space is a long-standing research problem that has drawn an increasing amount of research attention in recent years. Previous works found the shape of epidemics largely correlated with urban population distribution[50, 51]. Specifically, the peakedness of influenza season, i.e., the degree to which the influenza epidemics are compressed into a short period, is positively correlated with urban population size[50], while the cases of COVID-19 are observed to be more evenly distributed over time in crowded cities[51]. However, the population distribution does not capture the full picture of person-to-person interaction in urban space[52]. Therefore, some recent works proposed to use fine-grained human mobility behavioral data to model the complex transmission process[22, 23], which however poses huge burdens for data collection and privacy preservation. Different from the previous literature, we aim to capture the complex dynamics of urban epidemics with minimal data requirements. Specifically, we use model-predicted urban mobility to facilitate contagion disease modeling, which jointly considers population distribution and spatial proximity to accurately reproduce the complex transmission dynamics. It provides a theoretical framework to explain and reproduce the distinctive growth curves in different cities from the perspective of urban mobility, without the need for large-scale data collection. Our model can explain the superspreading events in cities resulting from the uneven distribution of urban population and spatially heterogeneous mobility. We demonstrate that targeted, mobility-aware travel restrictions can be designed based on our model's prediction, which can achieve a better balance between epidemic control and social cost to facilitate sustainable epidemic control for the gradual transition to a post-pandemic world. We also evaluate the likely outcomes under different natural and policy scenarios, where a faster policy response could greatly reduce the epidemic transmission. Therefore, our results provide experimental evidence that the metapopulation epidemiological model incorporated with a parsimonious urban mobility model can be sufficient to capture complex urban epidemic dynamics, which will be particularly beneficial to LMICs without fine-grained mobility data.

Our study has limitations. First, our model only incorporates basic assumptions regarding epidemic transmission. We do not consider all of the rich, contextual features that might influence the spread of the virus, such as weather[53]and changing attitudes towards containment policies[54]. More complex mechanisms in epidemic dynamics can be explored by explicitly considering these factors given high-quality fine-grained data. Second, although we aim to provide a widely available solution for LMICs where mobility data is hard to obtain, a better balance between data demand and the simplicity of the model can be achieved. For instance, given daily commuting data, one can further calibrate the gravity model to capture typical travel distance within cities, which might vary across countries with different economic and social contexts[55]. With the help of fine-grained demographic data in cities, we can arm the model with age-stratified infection risks to better reflect the heterogeneity of infection towards different age groups, which is helpful to inform more equitable interventions for elder people. Besides, providing vaccination data, we can evaluate the effectiveness of more targeted vaccination processes that may improve the efficiency of limited vaccine supplies. Nevertheless, analyses demonstrate that we can accurately trace citywide COVID-19 patterns with our model, suggesting sufficient expressiveness for characterizing urban viral transmission dynamics. Moreover, the minimum model setting improves its robustness and generalizability.

Our model and analyses hold several implications for better understanding and controlling the epidemic spreading. First, we demonstrate that the metapopulation epidemiological model incorporated with an urban mobility model is adequate to reproduce the complex epidemic dynamics. Experiments show that our estimated mobility changes are consistent with real-world observations, which affirms that our model can offer insights into fine-grained urban transmission dynamics and can be robustly transferred to other cities, facilitating decision-making in LMICs that lack fine-grained mobility data. Second, our analyses suggest that the 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 the urban environment and human mobility behavior to effectively curb superspreading. Third, we observe that adopting a fine-grained mobility model (i.e., with smaller neighborhood scales), is beneficial to better capturing epidemic dynamics. This finding is in line with recent public health research[56–58], suggesting the necessity of fine-grained mobility control during an epidemic. Fourth, during the prolonged pandemic, the coronavirus has been experiencing frequent antigenic mutation that heightens its spreading ability and even reduces vaccine protection[59,60], which might pose the need for more targeted, cost-efficient mobility control policies that can balance the disease transmission and social cost.

Methods

Model design

To model the complex epidemic dynamics, recently proposed SIR model[31]introduces self-quarantine mechanism for susceptible population and successfully reproduces sub-exponential growth curves. However, such a mechanism is idealistic, and the fundamental homogeneous population mixing setting prevents it from fully explaining the microscopic complexity of urban epidemic transmission, i.e., superspreading phenomena. To overcome this problem, some research incorporates mobility models with epidemiological models to better capture the complex transmission in urban spaces[11, 23, 61–67], where the mobility parameters are calibrated with intensive mobility data. It limits their application in LMICs or finer-grained inner-city scenarios since behavioral data is hard to acquire in these places. In that case, these works focus on town-town level[67], county-county level[66], state-state or province-province level[61–63] or even country-country level[64, 65]. After successfully reproducing the superspreading of disease, we further investigate the underlying behavior mechanism behind such phenomenon leveraging the urban mobility model.

Here, we provide both theoretical and empirical evidence that metapopulation epidemiological model incorporated with parsimonious urban mobility model can be sufficient to capture the complex epidemic dynamics in urban spaces without empirical urban mobility data. Specifically, we model urban space with numerous 1 km × 1 km grid cells, which we reference as "neighborhoods" in this study (see Supplementary Fig. 1 in SI Appendix). We set the population in each neighborhood according to the population distributions provided by WorldPop dataset[68]. For the contagion process, we adopt an SIR variant model that uses ordinary differential equations (ODE) to trace epidemic development, which contains four states: susceptible (S), asymptomatic (A), infected (I) and recovered (R). The adopted model can reflect the difficulty of detecting large asymptomatic population by separating them from infected population and learning different parameters. Each neighborhood maintains a separate SIR variant model for its sub-population, while epidemiological parameters are shared across all neighborhoods. In each simulation epoch, our model has two stages that trace the contagion process and mobility process, respectively. In the contagion stage, we calculate the number of susceptible, asymptomatic, infected and recovered people in each neighborhood according to the following equations:

$$
\begin{aligned}
\frac{\mathrm{d}s_n}{\mathrm{d}t} &= -\beta\, s_n a_n - \beta\, s_n i_n,\\
\frac{\mathrm{d}a_n}{\mathrm{d}t} &= (1-\omega)\beta\, s_n a_n + (1-\omega)\beta\, s_n i_n - \frac{1}{\tau} a_n - \kappa\, a_n,\\
\frac{\mathrm{d}i_n}{\mathrm{d}t} &= \omega\,\beta\, s_n a_n + \omega\,\beta\, s_n i_n + \frac{1}{\tau} a_n - \kappa\, i_n - \gamma\, i_n,\\
\frac{\mathrm{d}r_n}{\mathrm{d}t} &= \kappa\, a_n + \kappa\, i_n + \gamma\, i_n,
\end{aligned}
$$

where s_n, a_n, i_n, r_n are the susceptible, asymptomatic, infected and recovered people in neighborhood n. β is the infection rate, γ is the recovery rate, ω is the ratio that new infections immediately show symptoms and transit to infected, and τ is the average time of transition from asymptomatic persons to symptomatic ones. Note that both asymptomatic and symptomatic people can lead to secondary infections immediately after she/he gets infected, which depicts the most severe scenario of epidemic spreading.} Besides, κ denotes the testing and quarantine rate at which asymptomatic and infected persons are removed from the population. Note that we follow the assumption of SIR that the interactions between susceptible and asymptomatic/infected people scale with the population, as demonstrated in the $\beta\, s_n a_n$ and $\beta\, s_n i_n$ terms accordingly.

For the infection states of the whole city, we have

$$
S = \sum\nolimits_n s_n, A = \sum\nolimits_n a_n, I = \sum\nolimits_n i_n, R = \sum\nolimits_n r_n.
$$

In the mobility stage, we leverage the gravity equation to reproduce the urban mobility flow between neighborhoods as follows:

$$
flow_{mn} = M\frac{N_m^{\rho} N_n^{\theta}}{\exp\left(d_{mn}/r\right)},
$$

where M is the mobility level representing the intensity of the mobility flow among neighborhoods, N_m and N_n are the population size of the original and destination neighborhoods, and d_{mn} is the Manhattan distance between them. ρ , θ , r are predefined coefficients to represent the strength of the above two rules. The gravity model assumes that mobility flows between neighborhoods are negatively correlated with travel distance and positively correlated with population size.

Therefore, we can obtain the input flow and output flow of neighborhood n as follows:

$$\text{InputFlow}_n = \sum{}_{m, m \neq n} \text{flow}_{mn} = \sum{}_{m, m \neq n} M \frac{N_m^\rho N_n^\theta}{\exp\left(d_{mn}/r\right)},$$

$$\text{OutputFlow}_n = \sum{}_{m, m \neq n} \text{flow}_{nm} = \sum{}_{m, m \neq n} M \frac{N_n^\rho N_m^\theta}{\exp\left(d_{nm}/r\right)}.$$

Assuming the population flows have the same population composition (S, A, I and R) as the origin neighborhood, we have

$$\text{InputFlow}_{S_n}(t) = \sum{}_{m, m \neq n} M \frac{N_m^\rho N_n^\theta}{\exp\left(d_{mn}/r\right)} \times \frac{s_m}{N_m},$$

$$\text{OutputFlow}_{S_n}(t) = \sum{}_{m, m \neq n} M \frac{N_n^\rho N_m^\theta}{\exp\left(d_{nm}/r\right)} \times \frac{s_n}{N_n},$$

$$\text{InputFlow}_{A_n}(t) = \sum{}_{m, m \neq n} M \frac{N_m^\rho N_n^\theta}{\exp\left(d_{mn}/r\right)} \times \frac{a_m}{N_m},$$

$$\text{OutputFlow}_{A_n}(t) = \sum{}_{m, m \neq n} M \frac{N_n^\rho N_m^\theta}{\exp\left(d_{nm}/r\right)} \times \frac{a_n}{N_n},$$

$$\text{InputFlow}_{I_n}(t) = \sum{}_{m, m \neq n} M \frac{N_m^\rho N_n^\theta}{\exp\left(d_{mn}/r\right)} \times \frac{i_m}{N_m},$$

$$\text{OutputFlow}_{I_n}(t) = \sum{}_{m, m \neq n} M \frac{N_n^\rho N_m^\theta}{\exp\left(d_{nm}/r\right)} \times \frac{i_n}{N_n},$$

$$\text{InputFlow}_{R_n}(t) = \sum{}_{m, m \neq n} M \frac{N_m^\rho N_n^\theta}{\exp\left(d_{mn}/r\right)} \times \frac{r_m}{N_m},$$

$$\text{OutputFlow}_{R_n}(t) = \sum{}_{m, m \neq n} M \frac{N_n^\rho N_m^\theta}{\exp\left(d_{nm}/r\right)} \times \frac{r_n}{N_n},$$

where the summation is for each neighborhood.

The term $\frac{\mathrm{s_m}}{\mathrm{N_m}}, \frac{\mathrm{a_m}}{\mathrm{N_m}}, \frac{\mathrm{i_m}}{\mathrm{N_m}}, \frac{\mathrm{r_m}}{\mathrm{N_m}}$ represents the percentage of susceptible, asymptomatic, infected and recovered people in neighborhood m. The mobility stage will update the s, a, i, r for each neighborhood with the same temporal resolution (daily in our study) as the contagion stage, which is demonstrated as follows: $s_n = s_n + InputFlow_S(n,t) - OutputFlow_S(n,t) = s_n + M \sum_{m, m \neq n} \frac{N_m^p N_n^\theta}{exp(d_{mn}/r)} \times \frac{s_m}{N_m} - M \sum_{m, m \neq n} \frac{N_n^p N_m^\theta}{exp(d_{nm}/r)} \times \frac{s_n}{N_n}$,
$a_n = a_n + InputFlow_A(n,t) - OutputFlow_A(n,t) = a_n + M \sum_{m, m \neq n} \frac{N_m^p N_n^\theta}{exp(d_{mn}/r)} \times \frac{a_m}{N_m} - M \sum_{m, m \neq n} \frac{N_n^p N_m^\theta}{exp(d_{nm}/r)} \times \frac{a_n}{N_n}$,
$i_n = i_n + InputFlow_I(n,t) - OutputFlow_I(n,t) = i_n + M \sum_{m, m \neq n} \frac{N_m^p N_n^\theta}{exp(d_{mn}/r)} \times \frac{i_m}{N_m} - M \sum_{m, m \neq n} \frac{N_n^p N_m^\theta}{exp(d_{nm}/r)} \times \frac{i_n}{N_n}$,
$r_n = r_n + InputFlow_R(n,t) - OutputFlow_R(n,t) = r_n + M \sum_{m, m \neq n} \frac{N_m^p N_n^\theta}{exp(d_{mn}/r)} \times \frac{r_m}{N_m} - M \sum_{m, m \neq n} \frac{N_n^p N_m^\theta}{exp(d_{nm}/r)} \times \frac{r_n}{N_n}$.

There are three learnable parameters in the above model: (1) a learnable infection rate β that accounts for infectiousness of coronavirus, e.g., the effect of social distancing policies[45, 69]and the emergence of new variants[43]; (2) a learnable quarantine rate κ that accounts for the capacity of testing and quarantine for infected people[14, 32]; (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[44]. In each subpopulation, these parameters are shared across the city considering the limited variance in spatial transmissibility and testing ability. We set the other parameters according to recent epidemiological research on COVID-19 (see Supplementary Table.S7 in SI Appendix).

Supported by our proof that complex urban epidemics can be explained by coupling metapopulation epidemic model with urban mobility model, we design the above concise model that principally fuses epidemic and mobility processes and only has three learnable parameters. It allows us to develop a unified calibration process that simultaneously fits mobility parameters and epidemiological parameters from COVID-19 infection records. Specifically, we adopt a Gaussian prior for the learnable parameters (β , κ and M) and use Bayesian optimization to optimize them. 10 best models out of 40 runs are adopted for all the experiments. Statistics about the fitted parameters are provided in Supplementary Table 8. To validate the robustness of the results under different calibration methods, we also calibrate our model on daily cases in Supplementary Fig. 9. Compared with calibrating on cumulative cases in Supplementary Fig. 3, the performance differences are negligible.

Through the above calibration process, we can estimate empirical mobility changes without mobility data. The parameter M depicts the overall mobility level in the city, which is calibrated solely based on the epidemic curves. By comparing the changes of M during different stages, one can estimate the relative change in human mobility due to the epidemic. Specifically, we fit the epidemic curves before and after the first week of April in the U.S. for each county, which provides two sets of parameters that depict the mobility changes. By comparing the relative change of the mobility level M with empirical mobility changes from Apple Mobility Trends Records, we observe a high Pearson correlation of 0.872 (as demonstrated in Fig. 1C). It reflects that our model along with the underlying calibration process is able to capture real-world behavior changes. On average, our model estimates that the mobility level (M) decreases from 1.09×10^{-4} to 7.26×10^{-5} for all the counties, which corresponds to a 33.4% decrease.

We should notice that the neighborhood scale may influence the model accuracy. We perform extensive experiments on different neighborhood scale settings in Supplementary Fig. 11, ranging from 1 km $\times$ 1 km to 10 km $\times$ 10 km. The results demonstrate that using larger neighborhood scales can lead to lower R^2, while overall accuracy is still preserved. The worst-case scenario occurs when the epidemic dynamic follows a linear pattern, where R^2 drops less than 5% for linear growth patterns. However, even in these cases, our model still achieves high R^2 scores, demonstrating the robustness of the methodology.

Previous epidemiological research that leverages mobility model, such as the famous GLEaM model[22], adopts a separate fitting process for their mobility model, which is decoupled from the epidemic simulation. This assumption limits their application since urban mobility patterns are tightly coupled with epidemic transmission due to the spontaneous behavior change and imposed non-pharmaceutical interventions[70, 71]. Different from these works, we jointly calibrate the urban mobility parameter and epidemic parameter in our work. This design allows us to uncover the human mobility patterns that explain the specific epidemic dynamic and largely reduces the need of real-world mobility data.

Implementation details of the baseline SIR model

We implement an SIR variant model as a representative homogeneous mixing baseline in Fig. 1 (green lines). Specifically, the baseline model considers each city as a general unit, where each person uniformly contacts all the other people who not quarantined according to the following equations:

$$\frac{\mathrm{d}S}{\mathrm{d}t}=-\beta\,SA-\beta\,SI,$$

$$\frac{\mathrm{d}A}{\mathrm{d}t}=(1-\omega\,)\beta\,SA+(1-\omega\,)\beta\,SI-\kappa\,A-\frac{1}{\tau}A,$$

$$\frac{\mathrm{d}I}{\mathrm{d}t}=\omega\,\beta\,SA+\omega\,\beta\,SI+\frac{1}{\tau}A-\kappa\,I-\gamma\,I,$$

$$\frac{\mathrm{d}r_n}{\mathrm{d}t}=\kappa\,I+\gamma\,I,$$

where S, A, I, R are the susceptible, asymptomatic, infected, recovered people in the whole city. This baseline model shares the same ODE form with our proposed epidemiological model within each metapopulation patch, which provides a homogeneous counterpart without the aid of the mobility model.

Theoretical analysis of complex growth curves

To demonstrate our model's capacity for capturing the complex growth curves of COVID-19 cases, we consider two extreme cases of urban mobility. First, when $M\to 0$, there is no mobility flow between neighborhoods. Therefore, under this scenario, the virus can only infect a limited number of people who locate in neighborhoods that have initial cases, resulting in a significantly smaller effective population (people who can get infected) than the city population ($N_{eff}\ll N$). It is equivalent to the self-containment mechanism in recently proposed SIR-X model that assumes susceptible population is exponentially depleting due to travel restriction[31], which has proven effective in reproducing sub-exponential growth curves. Rigorous proofs can be found in (https://www.science.org/doi/suppl/10.1126/science.abb4557/suppl_file/abb4557_maier_sm.pdf.). Therefore, SIR-X model can be regarded as a special case of our model with $M\to 0$. Moreover, our model provides a more realistic micro behavioural foundation that explains the complex growth curves of COVID-19 cases with mobility behavior changes.

Second, with the limit of strong intra-city mobility, i.e., $M\to\infty$, the population flow between any neighborhood pair $flow_{mn}$ is converging to infinity. Under this scenario, the individuals in each neighborhood are totally shuffled during every simulation epoch, making S, A, I and R distributed proportionally in each neighborhood. Therefore, the composition portions of S, A, I and R are the same for each neighborhood. Besides, we define p_n as the ratio between the sub-population in neighborhood n and the total population in the city:

$$p_n=\frac{s_n+a_n+i_n+r_n}{S+A+I+R},$$

where $p_1=p_2=\cdots=p_n=p$. Therefore, the city-wide epidemic is a linear sum of SIR models with different sub-population sizes but homogeneous population composition, which has the following equations:

$$\begin{aligned}\frac{dS}{dt}=\frac{ds_1}{dt}+\frac{ds_2}{dt}+\cdots+\frac{ds_n}{dt}&=\Big(-(\beta s_1a_1+\beta s_1i_1)\Big)+\Big(-(\beta s_2a_2+\beta s_2i_2)\Big)+\cdots+\Big(-(\beta s_na_n+\beta s_ni_n)\Big)\\&=\Big(-(\beta SA+\beta SI)p_1^2\Big)+\Big(-(\beta SA+\beta SI)p_2^2\Big)+\cdots+\Big(-(\beta SA+\beta SI)p_n^2\Big)\\&=-n(\beta SA+\beta SI)p^2;\end{aligned}$$

$$\frac{dA}{dt} = \frac{da_1}{dt} + \frac{da_2}{dt} + \cdots + \frac{da_n}{dt} = \left((1-\omega)(\beta s_1 a_1 + \beta s_1 i_1) - \frac{1}{\tau}a_1\right) + \left((1-\omega)(\beta s_2 a_2 + \beta s_2 i_2) - \frac{1}{\tau}a_2\right) + \cdots +$$

$$\left((1-\omega)(\beta s_n a_n + \beta s_n i_n) - \frac{1}{\tau}a_n\right)$$

$$= \left((1-\omega)(\beta SA + \beta SI)p_1^2 - \tfrac{1}{\tau}Ap_1\right) + \left((1-\omega)(\beta SA + \beta SI)p_2^2 - \tfrac{1}{\tau}Ap_2\right) + \cdots +$$

$$\left((1-\omega)(\beta SA + \beta SI)p_n^2 - \frac{1}{\tau}Ap_n\right)$$

$$= n\left((1-\omega)(\beta SA + \beta SI)p^2 - \frac{1}{\tau}Ap\right);$$

$$\frac{dI}{dt} = \frac{di_1}{dt} + \frac{di_2}{dt} + \cdots + \frac{di_n}{dt} = \left(\omega(\beta s_1 a_1 + \beta s_1 i_1) + \frac{1}{\tau}a_1 - \kappa i_1 - \gamma i_1\right) + \left(\omega(\beta s_2 a_2 + \beta s_2 i_2) + \frac{1}{\tau}a_2 - \kappa i_2 - \gamma i_2\right) + \cdots +$$

$$\left(\omega(\beta s_n a_n + \beta s_n i_n) + \frac{1}{\tau}a_n - \kappa i_n - \gamma i_n\right)$$

$$= \left(\omega(\beta SA + \beta SI)p_1^2 + \frac{1}{\tau}Ap_1 - \kappa Ip_1 - \gamma Ip_1\right) + \left(\omega(\beta SA + \beta SI)p_2^2 + \frac{1}{\tau}Ap_2 - \kappa Ip_2 - \gamma Ip_2\right) + \cdots +$$

$$+\left(\omega(\beta SA + \beta SI)p_n^2 + \frac{1}{\tau}Ap_n - \kappa Ip_n - \gamma Ip_n\right) = n\left(\omega(\beta SA + \beta SI)p^2 + \frac{1}{\tau}Ap - \kappa Ip - \gamma Ip\right);$$

$$\frac{dR}{dt} = \frac{dr_1}{dt} + \frac{dr_2}{dt} + \cdots + \frac{dr_n}{dt} = \left(\kappa i_1 + \gamma i_1\right) + \left(\kappa i_2 + \gamma i_2\right) + \cdots + \left(\kappa i_n + \gamma i_n\right)$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}r_1}{\mathrm{d}t} + \frac{\mathrm{d}r_2}{\mathrm{d}t} + \cdots + \frac{\mathrm{d}r_n}{\mathrm{d}t} = (\kappa\, a_1 + \kappa\, i_1 + \gamma\, i_1) + (\kappa\, a_2 + \kappa\, i_2 + \gamma\, i_2) + \cdots + (\kappa\, a_n + \kappa\, i_n + \gamma\, i_n)$$

$$= \left(\kappa Ip_1 + \gamma Ip_1\right) + \left(\kappa Ip_2 + \gamma Ip_2\right) + \cdots + \left(\kappa Ip_n + \gamma Ip_n\right) = n(\kappa I + \gamma I)p.$$

Therefore, our model is degenerated into a homogeneous SIR model variant, whose equivalent infection rate and recovery rate are $\beta_{equ} = np^2\beta$, $\gamma_{equ} = np\gamma$. As a result, our model is able to reproduce the exponential growth curves as the standard SIR model. Since the mobility parameter M ranges from 0 to ∞ , our proposed model can reproduce a wide spectrum of growth curves between these two extreme cases (see Supplementary Fig. 4 in SI Appendix). Therefore, our model is provably expressive to capture the complex growth curves of urban epidemics.

Data availability

All of the data used in our work can be freely accessed. The empirical epidemic data and vaccination data are available in the GitHub repository maintained by Johns Hopkins University in https://github.com/CSSEGISandData/COVID-19. The global population data is provided by WorldPop in https://www.worldpop.org/. And the mobility data we use to validate our model prediction is obtained from Apple. Because Apple no longer provides the data, please contact Zhenyu Han (hanzy19@tsinghua.org.cn) to acquire it upon reasonable request.

Code availability

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

Received: 15 June 2024; Accepted: 17 January 2025

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. 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).
5. Adam, D. C. et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**, 1714–1719 (2020).
6. Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2000).
7. Grenfell, B. & Harwood, J. Meta) population dynamics of infectious diseases. *Trends Ecol. Evol.* **12**, 395–399 (1997).

8. Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11157–11162 (2005).
9. Colizza, V. & Vespignani, A. Invasion threshold in heterogeneous Metapopulation Networks. *Phys. Rev. Lett.* **99**, 148701 (2007).
10. Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3**, 276–282 (2007).
11. Zhou, Y. et al. Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data. *Lancet Digit. Health.* **2**, e417–e424 (2020).
12. Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971 (2020).
13. Mistry, D. et al. Inferring high-resolution human mixing patterns for disease modeling. *Nat. Commun.* **12**, 323 (2021).
14. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, eabb8001 (2020).
15. Sun, L., Axhausen, K. W., Lee, D. H. & Huang, X. Understanding metropolitan patterns of daily encounters. *Proc. Natl. Acad. Sci. U S A.* **110**, 13774–13779 (2013).
16. Ku, D. et al. Safe traveling in public transport amid COVID-19. *Sci. Adv.* **7**, eabg3691 (2021).
17. Mo, B. et al. Modeling epidemic spreading through public transit using time-varying encounter network. *Transp. Res. Part. C Emerg. Technol.* **122**, 102893 (2021).
18. 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. Interdisciplinary Perspect.* **9**, 100334 (2021).
19. Ahmad, N. & Chauhan, P. State of data privacy during COVID-19. *Computer* **53**, 119–122 (2020).
20. Stouffer, S. A. & Intervening Opportunities A theory relating mobility and Distance. *Am. Sociol. Rev.* **5**, 845–867 (1940).
21. Tobler, W. R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46**, 234–240 (1970).
22. 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).
23. Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
24. Apple. Apple Mobility Trends Reports. (2020). https://web.archive.org/web/20220331161154/http://covid19.apple.com/mobility
25. Jay, J. et al. Neighbourhood income and physical distancing during the COVID-19 pandemic in the United States. *Nat. Hum. Behav.* **4**, 1294–1302 (2020).
26. Song, C., Koren, T., Wang, P. & Barabási A.-L. Modelling the scaling properties of human mobility. *Nat. Phys.* **6**, 818–823 (2010).
27. Johns Hopkins University Center for Systems, S. & Engineering. (2020).
28. Covid19India. COVID19India. (2021). https://www.covid19india.org/
29. Ministério da, S. Coronavírus Brasil (in Portuguese). (2021). https://covid.saude.gov.br/
30. Simini, F., González, M. C., Maritan, A. & Barabási A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
31. Maier, B. F. & Brockmann, D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, eabb4557 (2020).
32. Moreno, G. K. et al. Revealing fine-scale spatiotemporal differences in SARS-CoV-2 introduction and spread. *Nat. Commun.* **11**, 5558 (2020).
33. 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).
34. Cirillo, P. & Taleb, N. N. Tail risk of contagious diseases. *Nat. Phys.* **16**, 606–613 (2020).
35. Wong, F. & Collins, J. J. Evidence that coronavirus superspreading is fat-tailed. *Proc. Natl. Acad. Sci. U. S. A.* 117, 29416–29418 (2020).
36. Gastwirth, J. L. The estimation of the Lorenz curve and Gini Index. *Rev. Econ. Stat.* **54**, 306–316 (1972).
37. NYC Government. COVID-19: Data. (2021). https://www1.nyc.gov/site/doh/covid/covid-19-data.page
38. Frieden, T. R. & Lee, C. T. Identifying and interrupting superspreading events—implications for control of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**, 1059 (2020).
39. Bouffanais, R. & Lim, S. S. Cities — try to predict superspreading hotspots for COVID-19. *Nature* **583**, 352–355 (2020).
40. Spelta, A. & Pagnottoni, P. Mobility-based real-time economic monitoring amid the COVID-19 pandemic. *Sci. Rep.* **11**, 1–15 (2021).
41. Aguilar, J. et al. Impact of urban structure on infectious disease spreading. *Sci. Rep.* **12**, 3816 (2022).
42. Xu, Y. et al. Urban dynamics through the lens of human mobility. *Nat. Comput. Sci.* **3**, 611–620 (2023).
43. Koyama, T., Platt, D. & Parida, L. Variant analysis of SARS-CoV-2 genomes. *Bull. World Health Organ.* **98**, 495 (2020).
44. The State Council of the People's Republic of, C. (2020).
45. Lai, S. et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **585**, 410–413 (2020).
46. Solís Arce, J. S. et al. COVID-19 vaccine acceptance and hesitancy in low-and middle-income countries. *Nat. Med.* 27 (2021).
47. Machingaidze, S. & Wiysonge, C. S. Understanding COVID-19 vaccine hesitancy. *Nat. Med.* **27**, 1338–1339 (2021).
48. Volpert, V., Banerjee, M. & Petrovskii, S. On a quarantine model of coronavirus infection and data analysis. *Math. Model. Nat. Phenom.* **15**, 24 (2020).
49. WHO. Public health surveillance for COVID-19: interim guidance. (2020). https://www.who.int/publications/i/item/WHO-2019-nCoV-SurveillanceGuidance-2022.2
50. Dalziel, B. D. et al. Urbanization and humidity shape the intensity of influenza epidemics in US cities. *Science* **362**, 75–79 (2018).
51. Rader, B. et al. Crowding and the shape of COVID-19 epidemics. *Nat. Med.* **26**, 1829–1834 (2020).
52. Han, Z. et al. Who will survive and revive undergoing the epidemic: analyses about POI visit behavior in Wuhan via check-in records. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **5**, 1–20 (2021).
53. 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).
54. Sabat, I. et al. United but divided: policy responses and people's perceptions in the EU during the COVID-19 outbreak. *Health Policy*. **124**, 909–918 (2020).
55. Mazzoli, M. et al. Field theory for recurrent mobility. *Nat. Commun.* **10**, 3895 (2019).
56. Espinoza, B., Castillo-Chavez, C. & Perrings, C. Mobility restrictions for the control of epidemics: when do they work? *Plos One*. **15**, e0235731 (2020).
57. Poletto, C., Tizzoni, M. & Colizza, V. Human mobility and time spent at destination: impact on spatial epidemic spreading. *J. Theor. Biol.* **338**, 41–58 (2013).
58. Moreno, V. et al. The role of mobility and health disparities on the transmission dynamics of tuberculosis. *Theoretical Biology Med. Modelling*. **14**, 1–17 (2017).
59. Telenti, A. et al. After the pandemic: perspectives on the future trajectory of COVID-19. *Nature* **596**, 495–504 (2021).
60. Callaway, E. Fast-spreading COVID variant can elude immune responses. *Nature*, 500–501 (2021).
61. Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
62. Tizzoni, M. et al. On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* **10**, e1003716 (2014).
63. Bengtsson, L. et al. Using mobile phone data to predict the spatial spread of cholera. *Sci. Rep.* **5**, 1–5 (2015).
64. D'Silva, J. P. & Eisenberg, M. C. Modeling spatial invasion of Ebola in West Africa. *J. Theor. Biol.* **428**, 65–75 (2017).
65. Sorichetta, A. et al. Mapping internal connectivity through human migration in malaria endemic countries. *Sci. Data* **3** (2016).

66. Truscott, J. & Ferguson, N. M. Evaluating the adequacy of gravity models as a description of human mobility for Epidemic Modelling. *PLoS Comput. Biol.* 8 (2012).
67. Bharti, N., Xia, Y., Bjornstad, O. N. & Grenfell, B. T. Measles on the Edge: Coastal heterogeneities and infection dynamics. *PLoS ONE* 3 (2008).
68. Tatem, A. J. WorldPop, open data for spatial demography. *Sci. Data*. **4**, 170004 (2017).
69. Flaxman, S. et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020).
70. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *Lancet Digit. Health*. **2**, e638–e649 (2020).
71. Gibbs, H. et al. Changing travel patterns in China during the early stages of the COVID-19 pandemic. *Nat. Commun.* **11**, 5012 (2020).

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under U21B2036.

Author contributions

F.X., Y.L., and J.E. jointly launched this research. Z.H. performed the experiments and prepared the figures. F.X., Y.L., T.J., and J.E. provided the research outline, research design and critical revisions. All authors jointly analyzed the results and wrote the paper.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-87363-3.

Correspondence and requests for materials should be addressed to Y.L. or J.E.