

Supplementary Information of model predicted human mobility explains COVID-19 transmission in urban space without behavioral data

Zhenyu Han, Fengli Xu, Yong Li, Tao Jiang, James Evans

Yong Li and James Evans are corresponding authors. E-mail: liyong07tsinghua.edu.cn, jevansuchicago.edu

Datasets

In this work, we use five datasets: population distribution data and COVID-19 confirmed cases data are for model fitting. Besides, policy announcement data, real-world mobility data and neighborhood level infection data for validation.

A) Population distribution data: we derive the population distribution data from the WorldPop project¹. It contains fine-grained population distribution at 100m resolution for cities around the globe. We choose the 2020 dataset of the United States, India and Brazil in our research, and aggregate the population into neighborhoods for each urban county. We determine the border of each county or city based on Google Maps.

B) COVID-19 confirmed cases data: for U.S. counties, we adopt the COVID-19 confirmed cases time-series data released by Johns Hopkins University². According to the severity of infection, we select 20 counties in U.S. to validate our model. Most of them are from the top 20 infected counties, and we also include three countries that have less infected people to validate model effectiveness on mild scenarios, which are Davidson, Will and Hillsborough. Statistics of the selected counties are available in Table 1. For India and Brazil cities, we adopt the official release of confirmed cases data^{3,4} to fit the model. We select the 5 most infectious cities in India and Brazil accordingly, which are shown in Table 2, Table 3.

C) Policy announcement data: we refer to the COVID-19 State and County Policy Orders dataset⁵ at the state and county level to determine the date of behavior change in the U.S. (Fig. 1). We calculate the density of COVID-19 related policies and adopt this index as indicative of political attitude towards the disease. The selected time point for behavior change in early periods is April 1st, when the number of policies reached a peak (and a few days earlier). Specific policies have good correspondence with that date. For example on March 21, New Jersey Governor announced a “stay-at-home” order and on March 22nd, non-essential businesses were ordered to close or end all in-person functions in New York, New Jersey and Connecticut⁶.

D) Real-world mobility data: we use Apple Mobility Trends Reports⁷ as the real-world mobility data to validate our model-estimated mobility behavior changes.

E) Neighborhood level infection data: we use fine-grained infection data from New York City to validate the location-aware reopening policy design proposed by our model. These infection data are available at MODZCTA level, which divides New York City into 177 neighborhoods. These data are provided by the Health Department of New York City⁸, which are hosted in the NYC Coronavirus Disease 2019 (COVID-19) Data Repository⁹.

Supplementary Methods

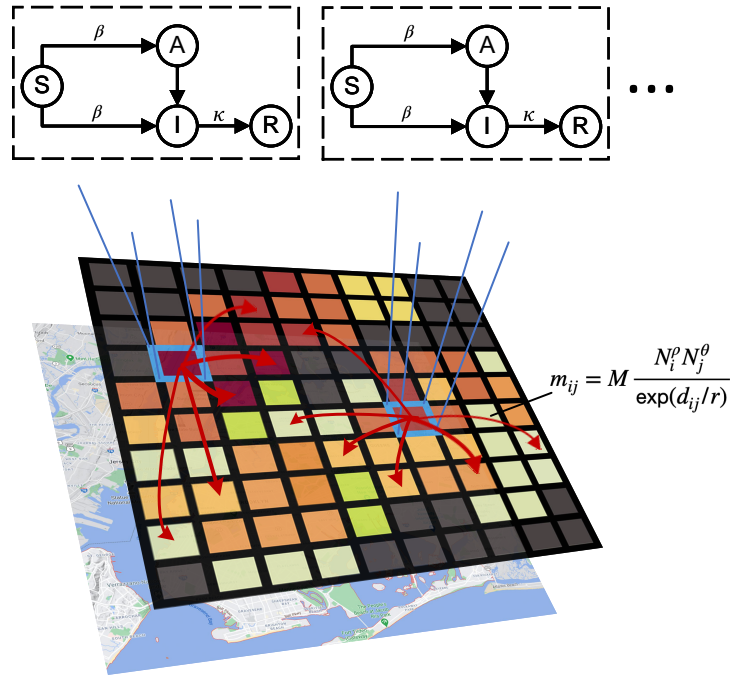
SM1. Calibration of mobility activity

After the model calibration process on epidemic curves, the calibrated mobility level M can be used to estimate the mobility change without empirical mobility data. To evaluate the accuracy of this estimation, we collect Apple Mobility Trends Reports as an external data source. However, as described in the Methods section, parameter M is a dimensionless scaling factor, which forbids direct comparisons. Here, we rescale the changes of parameter ΔM according to a linear mapping function as follows:

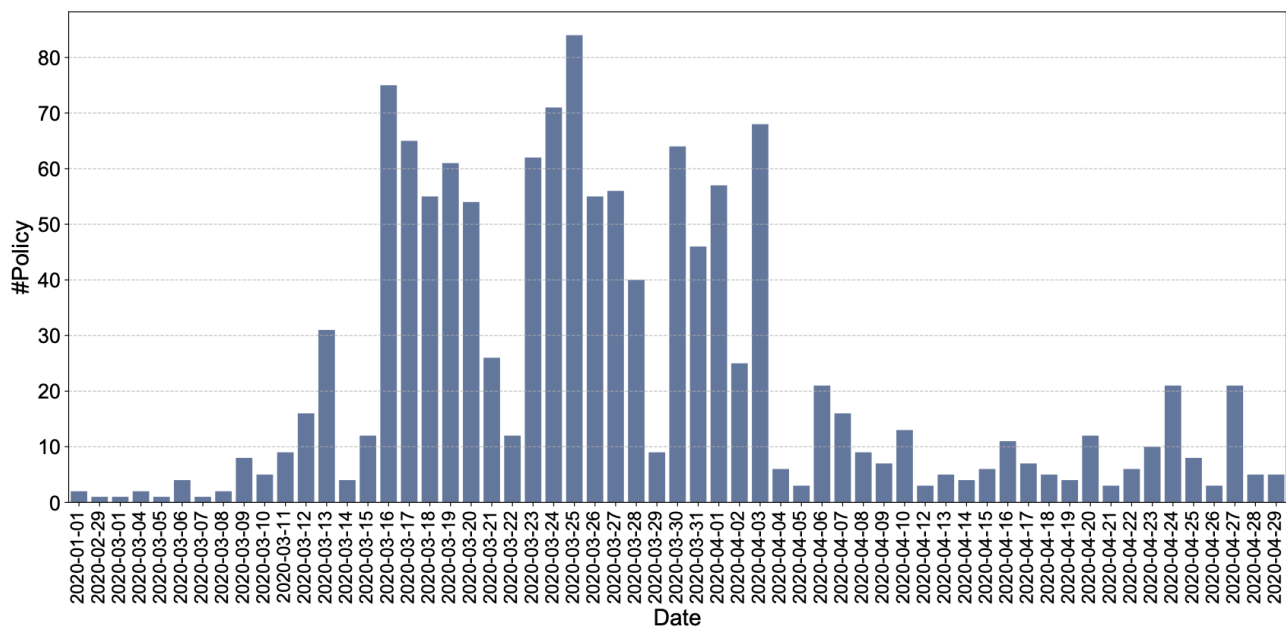
$$\Delta \bar{M} = k \times \Delta M + b,$$

where $\Delta \bar{M}$ and ΔM are the rescaled and original model-informed mobility changes accordingly. In Figure 1C, we investigate the correlation between model-informed mobility change ($\Delta \bar{M}$) with the empirical mobility change (Apple data), where the high Pearson correlation of 0.872 validates that the proposed model can accurately estimate empirical mobility behavior changes without mobility data.

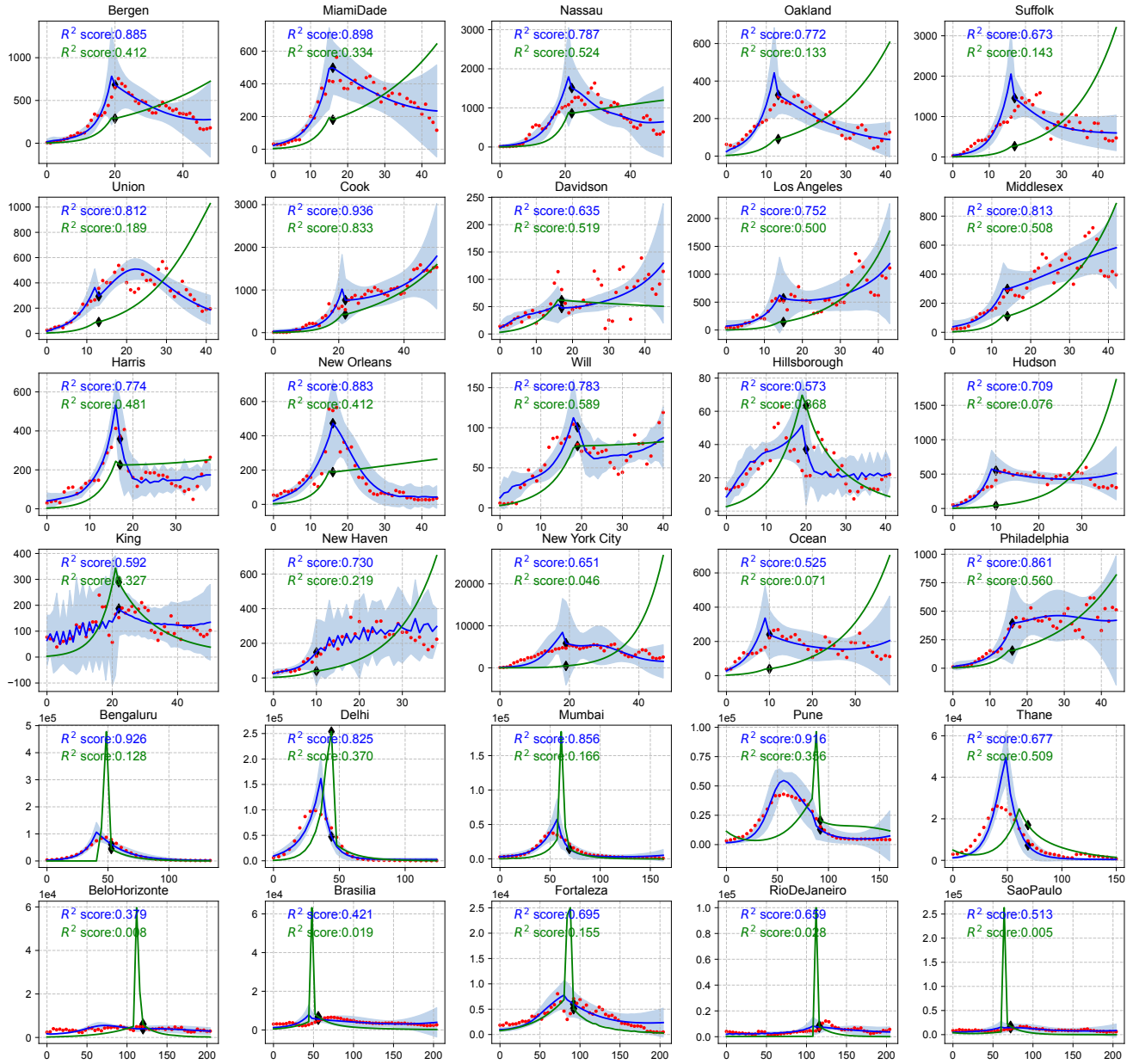
Supplementary Figures



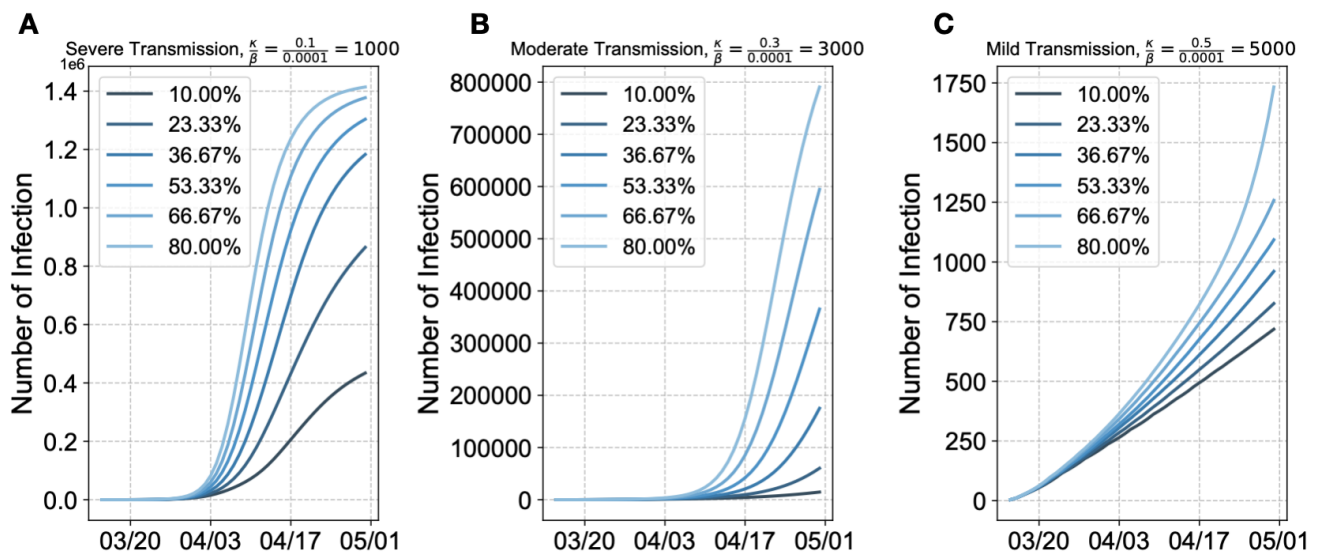
Supplementary Figure 1. Illustration of the proposed metapopulation model. We segment urban space into numerous neighborhoods, and divide the whole city population into subpopulations based on them. Separate SIR variant are run on the subpopulation within each neighborhood, and the susceptible, asymptomatic, infected, and recovered individuals are mixed across neighborhoods according to their mobility predicted by the gravity model.



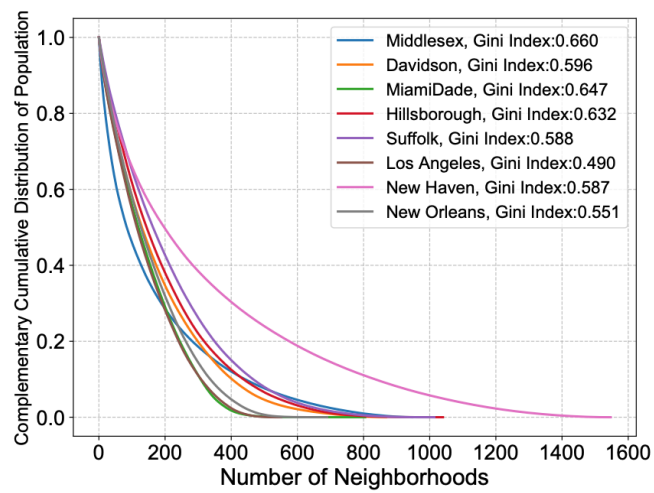
Supplementary Figure 2. The number of COVID-19 intervention policies announced in the United States. Most policies are released at the end of March after the Declaration of National Emergency. Thus, we choose April 1st as the time point of behavior change in the United States.



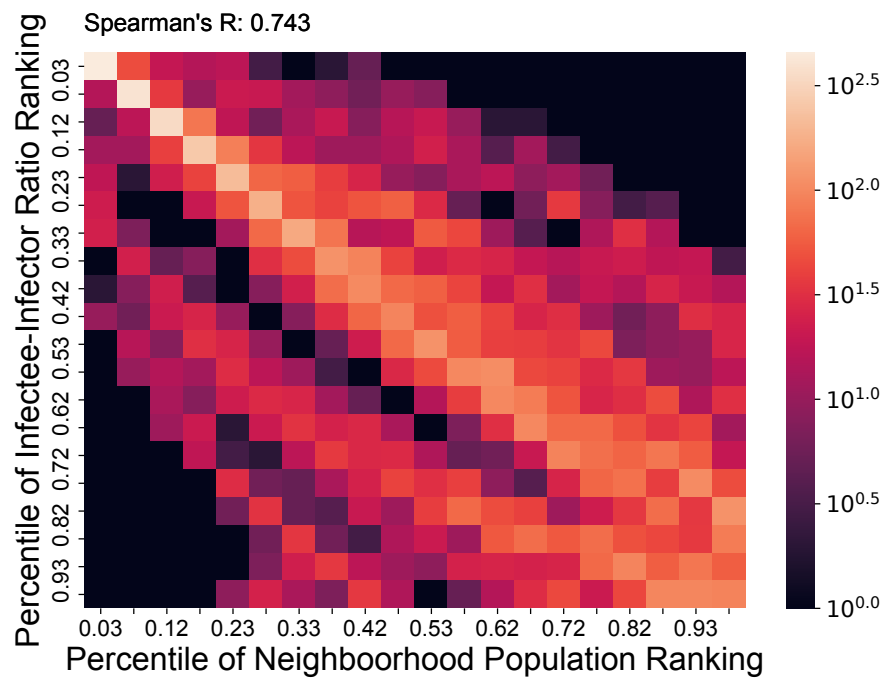
Supplementary Figure 3. Daily case counts of our model and traditional SEIR, calibrated on cumulative cases. In the main text, we demonstrate the cumulative cases curves to better demonstrate the ability of our model to capture different epidemic dynamics. Here we provide the corresponding daily case counts to better demonstrate the robustness of our model. The daily cases is processed by a 3-point moving average.



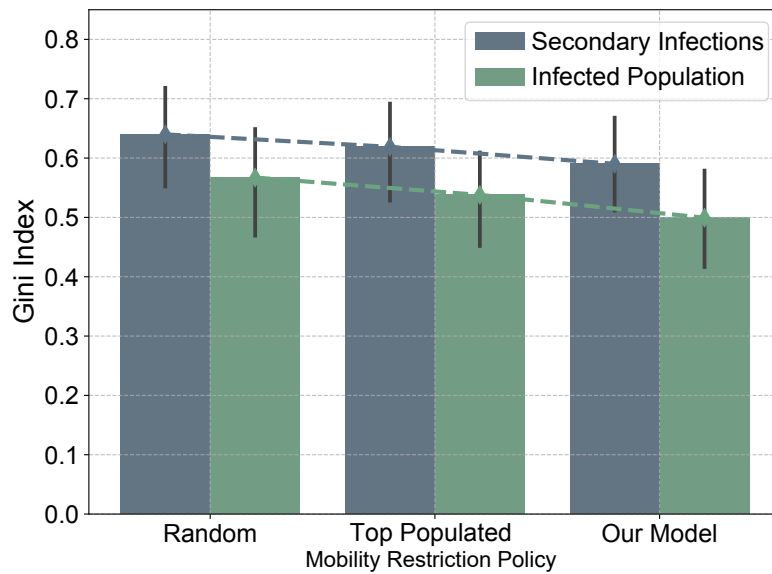
Supplementary Figure 4. Estimated growth curves of COVID-19 cases with different mobility levels and transmission rates. Color of lines represent different mobility levels compared with maximum mobility, where all people will move to other neighborhoods. Different mobility levels will greatly change the shape of growth curves, where low mobility levels are more likely to result in sub-exponential growth curves.



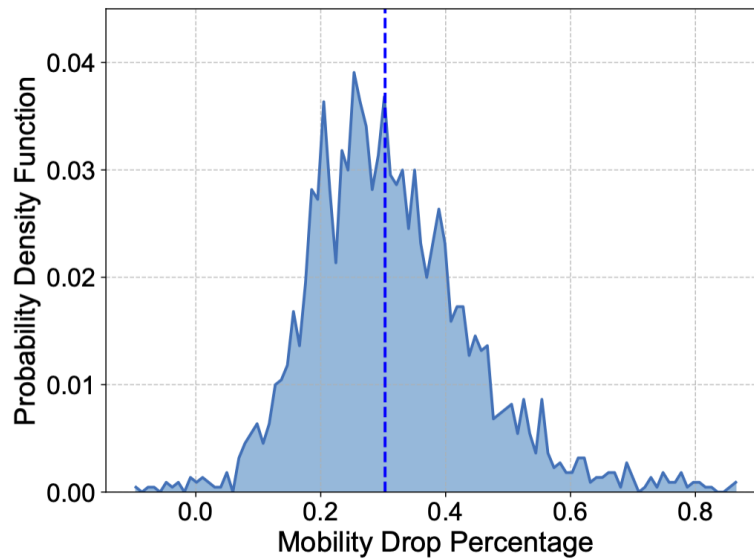
Supplementary Figure 5. Heterogeneity of population distribution in cities. Population distributions of selected U.S. counties demonstrate strong heterogeneity where top populated neighborhoods account for considerable population.



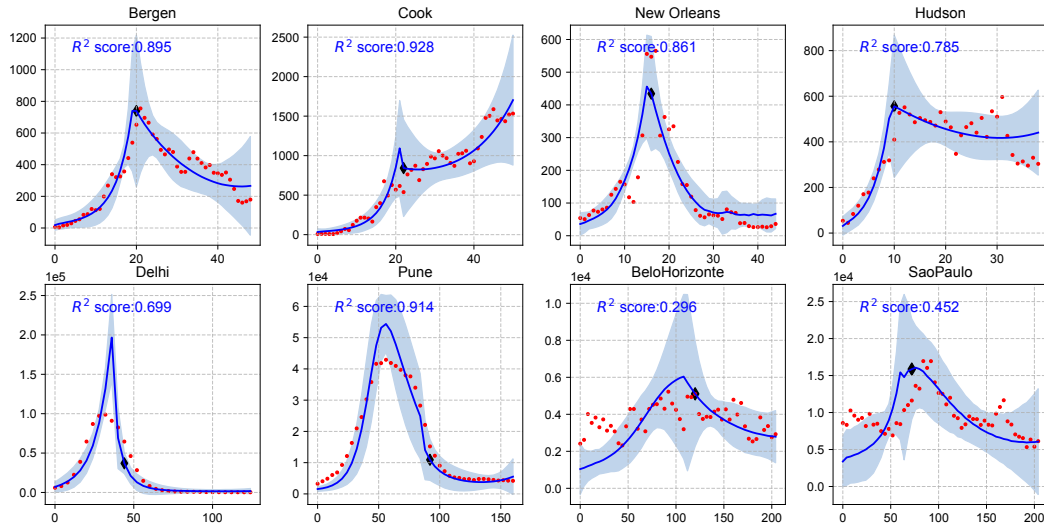
Supplementary Figure 6. Correlation between disease transmission and population. The heatmap of neighborhood's ranking on population size and infectee-infectior ratio across 20 United States counties, which manifests a relative high Spearman correlation of 0.743, but also a high variance in infectee-infectior ratio that cannot be explained by population size.



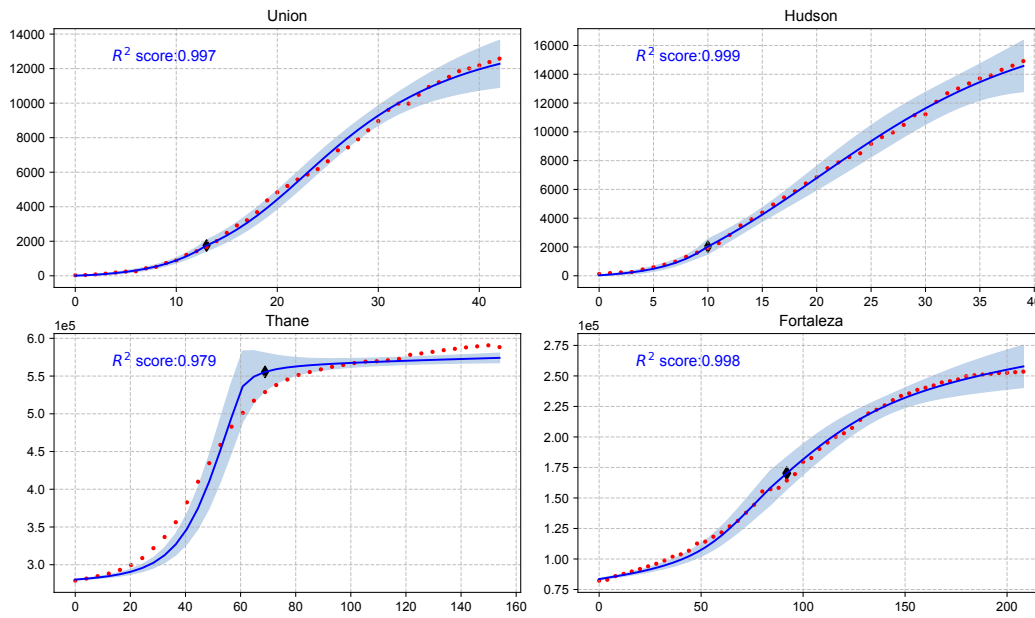
Supplementary Figure 7. The effect of superspreading events under different mobility reduction policies in U.S. urban counties. It is measured by the Gini indices of the spatial distribution of secondary infections and the infected population. The policy informed by our model results in smaller spatial unevenness of infection risk, which indicates less superspreading.



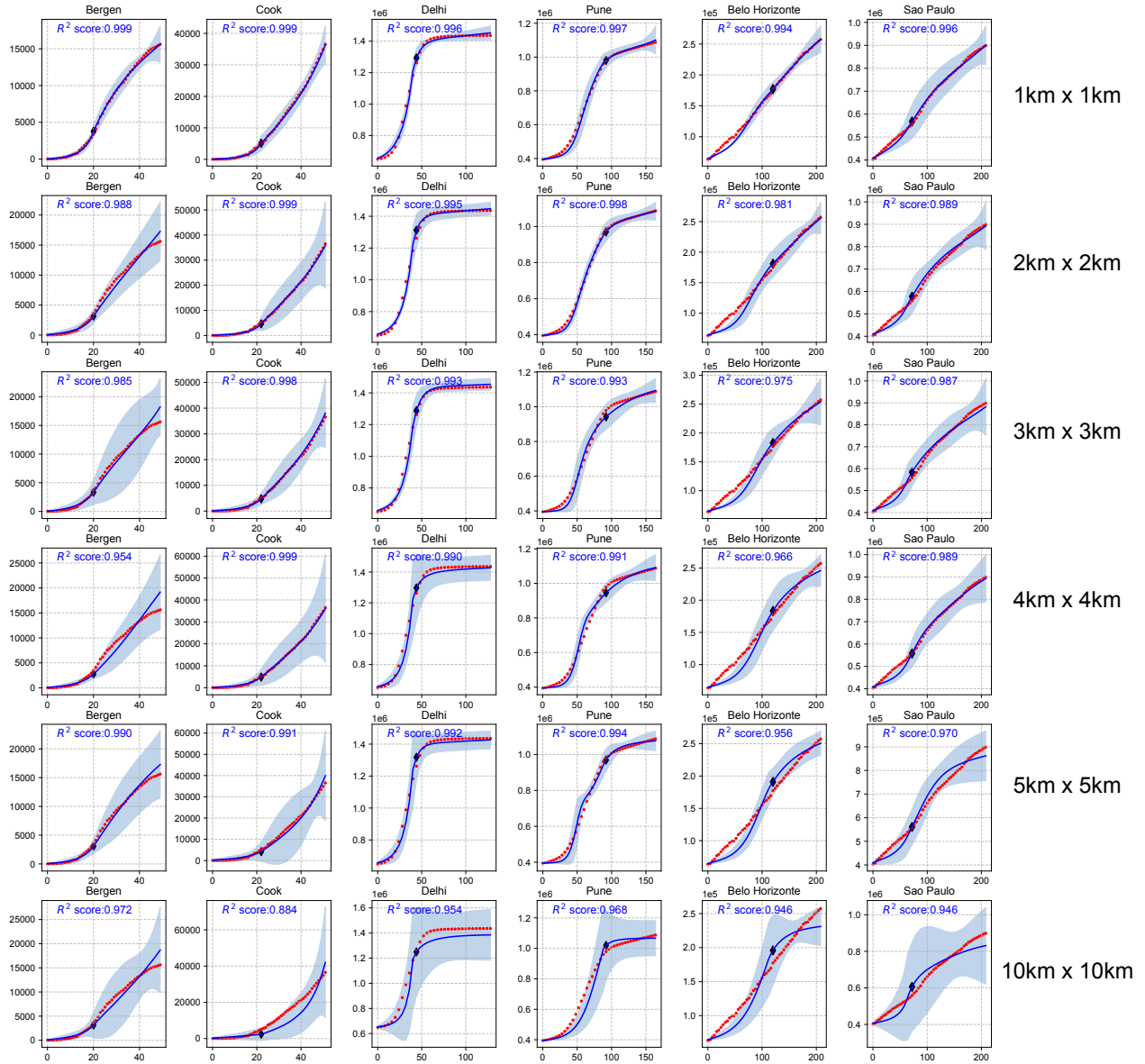
Supplementary Figure 8. Mobility drops in selected U.S. urban counties before and after March 25. We evaluate the real-world mobility drops from Apple Mobility Trends Reports⁷. The mobility drop has a median value of 30.3% (shown in the vertical dashed line), with IQR $0.227 \sim 0.392$.



Supplementary Figure 9. Daily case counts of our model, calibrated on daily cases. We choose representative counties and cities that cover different epidemic dynamics and population densities to present the differences between calibrated on cumulative cases and daily cases. The daily cases is processed by a 3-point moving average.



Supplementary Figure 10. Performance of using radiation model in reproducing COVID-19 growth curves. We replace the gravity model by the radiation model in Union, Hudson, Thane and Fortaleza to validate the robustness of our model. Compared with the results using the gravity model (see Fig.1a), the performance differences are negligible. The gravity based model achieved high R^2 of 0.998 in Union and Hudson, 0.985 in Thane, and 0.994 in Fortaleza. While the radiation based model achieved 0.997 in Union, 0.999 in Hudson, 0.979 in Thane, and 0.998 in Fortaleza.



Supplementary Figure 11. Performance under different neighborhood scales in reproducing COVID-19 growth curves. We adopt different neighborhood scale settings, ranging from 1 km \times 1 km to 10 km \times 10 km, in six representative cities with varying epidemic dynamics. We observe a trend 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 by 4.8% and 5.0% for Belo Horizonte and Sao Paulo, respectively. However, even in these cases, our model still achieves a high R^2 of 0.946, demonstrating the robustness of the methodology.

Supplementary Tables

Supplementary Table 1. Statistics for selected urban U.S. counties

County	State	Time Period	Behaviour Change	Confirmed Cases	Land Area (km^2)	Population
New York	New York	2020-03-13 ~2020-04-30	2020-04-02	167478	783.8	8645904
Cook	Illinois	2020-03-10 ~2020-04-30	2020-04-02	36513	4235	6880976
Nassau	New York	2020-03-10 ~2020-04-30	2020-04-02	35854	1173	1627850
Suffolk	New York	2020-03-15 ~2020-04-30	2020-04-02	33664	6146	1691197
Los Angeles	California	2020-03-17 ~2020-04-30	2020-04-02	23220	1302	5941621
Bergen	New Jersey	2020-03-12 ~2020-04-30	2020-04-02	15610	639	2199635
Hudson	New Jersey	2020-03-22 ~2020-04-30	2020-04-02	14916	162	707938
Philadelphia	Pennsylvania	2020-03-16 ~2020-04-30	2020-04-02	14468	400	2158260
Middlesex	Massachusetts	2020-03-18 ~2020-04-30	2020-04-02	14208	2195	3249380
Union	New Jersey	2020-03-19 ~2020-04-30	2020-04-02	12578	940	116520
Miami-Dade	Florida	2020-03-16 ~2020-04-30	2020-04-02	12063	6297	2788473
New Haven	Connecticut	2020-03-22 ~2020-04-30	2020-04-02	7536	2233	1160721
Oakland	Michigan	2020-03-19 ~2020-04-30	2020-04-02	7267	2349	1212662
New Orleans	Louisiana	2020-03-16 ~2020-04-30	2020-04-02	6452	906	354576
Ocean	New Jersey	2020-03-22 ~2020-04-30	2020-04-02	6375	2370	584915
Harris	Texas	2020-03-22 ~2020-04-30	2020-04-09	6356	4602	6099809
King	Washington	2020-03-10 ~2020-04-30	2020-04-02	6207	5975	2362968
Davidson	Tennessee	2020-03-15 ~2020-04-30	2020-04-02	2612	1363	931460
Will	Illinois	2020-03-20 ~2020-04-30	2020-04-09	2492	2200	1927804
Hillsborough	Florida	2020-03-19 ~2020-04-30	2020-04-09	1124	3279	1628413

Supplementary Table 2. Statistics for selected cities in India.

County	Time Period	Behaviour Change	Confirmed Cases	Land Area (km^2)	Population
Delhi	2021-03-22 ~2021-07-31	2021-05-04	1436265	1484	23764826
Bengaluru Urban	2021-03-12 ~2021-07-31	2021-05-01	1227339	2196	14196185
Pune	2021-02-14 ~2021-07-31	2021-05-14	1089386	484.6	4857275
Mumbai	2021-02-11 ~2021-07-31	2021-04-17	734779	603.4	14744188
Thane	2021-02-24 ~2021-07-31	2021-04-21	589018	147	3581270

Supplementary Table 3. Statistics for selected cities in Brazil.

City	Time Period	Behaviour Change	Confirmed Cases	Land Area (km^2)	Population
São Paulo	2021-01-01 ~2021-07-31	2021-03-10	904083	1521	17974819
Brasilia	2021-01-01 ~2021-07-31	2021-02-23	450166	5802	3611998
Rio de Janeiro	2021-01-01 ~2021-07-31	2021-04-26	398836	1255	8015617
Belo Horizonte	2021-01-01 ~2021-07-31	2021-04-27	259535	330.9	3244199
Fortaleza	2021-01-01 ~2021-07-31	2021-04-01	253811	313.8	3285519

Supplementary Table 4. Population of restricted regions in U.S. urban counties.

City	Total Population	Algorithm	Controlled Population	Percentage
New York	8645904	Random	415867	4.81%
		Top Populated	422266	4.88%
		Most Infectious	416457	4.82%
Cook	6880976	Random	332365	4.83%
		Top Populated	325095	4.72%
		Most Infectious	321007	4.67%
Nassau	1627850	Random	80892	5.00%
		Top Populated	80794	4.96%
		Most Infectious	78782	4.84%
Suffolk	726471	Random	84953	5.02%
		Top Populated	87807	5.19%
		Most Infectious	84430	4.99%
Los Angeles	5941621	Random	305118	5.14%
		Top Populated	309545	5.21%
		Most Infectious	309550	5.21%
Bergen	2199635	Random	97258	4.42%
		Top Populated	97958	4.45%
		Most Infectious	97878	4.45%
Hudson	707938	Random	44248	6.25%
		Top Populated	41062	5.80%
		Most Infectious	41064	5.80%
Philadelphia	2158260	Random	105545	4.89%
		Top Populated	104777	4.85%
		Most Infectious	105076	4.87%
Middlesex	3249380	Random	139986	4.31%
		Top Populated	141161	4.34%
		Most Infectious	141165	4.34%
Union	116520	Random	5728	4.92%
		Top Populated	7974	6.84%
		Most Infectious	7977	6.85%
Miami-Dade	2788473	Random	137034	4.91%
		Top Populated	143188	5.13%
		Most Infectious	143190	5.14%
New Haven	1160721	Random	53958	4.65%
		Top Populated	55476	4.78%
		Most Infectious	53062	4.57%
Oakland	1212662	Random	59464	4.90%
		Top Populated	59046	4.87%
		Most Infectious	59809	4.93%
New Orleans	354576	Random	18170	5.12%
		Top Populated	18820	5.31%
		Most Infectious	17823	5.03%
Ocean	584915	Random	28652	4.90%
		Top Populated	29652	5.07%
		Most Infectious	28349	4.85%
Harris	6099809	Random	298287	4.89%
		Top Populated	304437	4.99%
		Most Infectious	309631	5.08%
King	2362968	Random	118185	5.00%
		Top Populated	115710	4.90%
		Most Infectious	115713	4.90%
Davidson	931460	Random	47858	5.15%
		Top Populated	47964	5.15%
		Most Infectious	46693	5.01%
Will	1927804	Random	94054	4.88 %
		Top Populated	97218	5.04 %
		Most Infectious	95510	4.95%
Hillsborough	1628413	Random	80746	4.96%
		Top Populated	80131	4.92%
		Most Infectious	84324	5.18%

Supplementary Table 5. Population of restricted regions in Indian cities.

City	Total Population	Algorithm	Controlled Population	Percentage
Delhi	23764826	Random	2394590	10.08%
		Top Populated	2388527	10.05%
		Most Infectious	2388592	10.05%
Bengaluru	14196185	Random	1272813	8.97%
		Top Populated	1296836	9.14%
		Most Infectious	1296837	9.14%
Pune	4857275	Random	489703	10.08%
		Top Populated	492292	10.14%
		Most Infectious	492311	10.14%
Mumbai	14744188	Random	1475510	10.00%
		Top Populated	1485880	10.08%
		Most Infectious	1485927	10.08%
Thane	3581270	Random	405313	11.32%
		Top Populated	358563	10.01%
		Most Infectious	358569	10.01 %

Supplementary Table 6. Population of restricted regions in Brazilian cities.

City	Total Population	Algorithm	Controlled Population	Percentage
São Paulo	17974819	Random	1813335	10.09%
		Top Populated	1813363	10.09%
		Most Infectious	1813429	10.09%
Brasília	3611998	Random	352588	9.76%
		Top Populated	359142	9.94%
		Most Infectious	359151	9.94%
Rio de Janeiro	8015617	Random	795824	9.93%
		Top Populated	804380	10.04%
		Most Infectious	804406	10.04%
Belo Horizonte	3244199	Random	327975	10.11%
		Top Populated	327954	10.11 %
		Most Infectious	327986	10.11%
Fortaleza	3285519	Random	335223	10.20%
		Top Populated	332139	10.11%
		Most Infectious	332154	10.11%

Supplementary Table 7. Parameters in our model.

Name	Description	Value	Source
Infection rate (β)	Probability of getting infected when a susceptible person collocates with an exposed or infected person. We assume the transmission probability is equal for exposed and infected people.	Learnable parameter	-
Quarantine rate (κ)	The quarantine rate for the infected population.	Learnable parameter	-
Incubation period (τ)	The average incubation period of exposed individuals.	5.2	Reference to 10–12
Recovery rate (γ)	The rate with which infected individuals recover or die.	1/14	According to 13
Probability of symptomatic infection (ω)	The probability that a new infected person becomes symptomatic (I) Otherwise (s)he will become a asymptomatic patient (A), and the corresponding probability is $1 - \omega$.	0.7	Reference to 12
Gravity model parameter M, ρ, θ, r	Scaling factor that represents the overall mobility; the attraction force of source neighborhood population and destination neighborhood population; the distance parameter for modeling the locality of mobility	Learnable parameter, 0.46, 0.64, 82 km	Reference to 14,15

Supplementary Table 8. Average value of calibrated parameters in our model.

City	β		κ		M	
	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
Bergen	2.86×10^{-4}	1.94×10^{-4}	5.32×10^{-1}	5.70×10^{-1}	8.87×10^{-5}	2.18×10^{-5}
Cook	1.86×10^{-4}	9.70×10^{-5}	6.46×10^{-1}	5.12×10^{-1}	8.18×10^{-5}	2.97×10^{-5}
Davidson	2.32×10^{-4}	2.64×10^{-4}	4.75×10^{-1}	4.72×10^{-1}	1.48×10^{-4}	8.08×10^{-5}
Harris	3.11×10^{-4}	3.26×10^{-4}	7.80×10^{-1}	8.88×10^{-1}	1.54×10^{-4}	2.09×10^{-4}
Hillsborough	1.12×10^{-4}	2.28×10^{-4}	4.13×10^{-1}	6.98×10^{-1}	1.30×10^{-4}	7.21×10^{-5}
Hudson	2.13×10^{-4}	2.52×10^{-4}	5.35×10^{-1}	6.73×10^{-1}	1.57×10^{-4}	1.98×10^{-4}
King	1.34×10^{-3}	7.25×10^{-4}	9.65×10^{-1}	9.13×10^{-1}	1.86×10^{-4}	6.22×10^{-5}
Los Angeles	2.69×10^{-4}	3.89×10^{-4}	8.26×10^{-1}	8.39×10^{-1}	1.20×10^{-4}	6.93×10^{-5}
Miami-Dade	1.52×10^{-4}	2.39×10^{-4}	5.65×10^{-1}	7.07×10^{-1}	1.25×10^{-4}	3.64×10^{-5}
Middlesex	2.62×10^{-4}	2.55×10^{-4}	7.30×10^{-1}	7.54×10^{-1}	8.31×10^{-5}	2.32×10^{-5}
New York City	5.16×10^{-4}	1.46×10^{-4}	8.73×10^{-1}	4.21×10^{-1}	4.02×10^{-5}	1.00×10^{-5}
Nassau	3.98×10^{-4}	2.90×10^{-4}	5.20×10^{-1}	4.81×10^{-1}	4.20×10^{-5}	2.71×10^{-5}
New Haven	6.57×10^{-4}	5.28×10^{-3}	6.40×10^{-1}	9.88×10^{-1}	1.57×10^{-4}	2.42×10^{-4}
New Orleans	1.03×10^{-3}	2.14×10^{-3}	4.93×10^{-1}	7.33×10^{-1}	8.70×10^{-5}	1.14×10^{-5}
Oakland	6.25×10^{-4}	3.73×10^{-4}	5.61×10^{-1}	5.00×10^{-1}	7.76×10^{-5}	2.94×10^{-5}
Ocean	8.90×10^{-4}	5.08×10^{-4}	5.84×10^{-1}	5.08×10^{-1}	4.32×10^{-5}	1.00×10^{-4}
Philadelphia	2.00×10^{-4}	4.04×10^{-4}	4.10×10^{-1}	6.60×10^{-1}	1.29×10^{-4}	3.71×10^{-5}
Suffolk	4.63×10^{-4}	1.90×10^{-4}	6.44×10^{-1}	5.18×10^{-1}	9.08×10^{-5}	4.25×10^{-5}
Union	4.91×10^{-4}	2.48×10^{-4}	4.17×10^{-1}	2.60×10^{-1}	1.91×10^{-5}	9.97×10^{-5}
Will	2.38×10^{-4}	3.23×10^{-4}	5.45×10^{-1}	7.52×10^{-1}	1.48×10^{-4}	2.26×10^{-4}
Bengaluru	2.90×10^{-4}	4.57×10^{-4}	9.02×10^{-1}	9.63×10^{-1}	1.13×10^{-4}	2.64×10^{-5}
Delhi	4.86×10^{-4}	2.81×10^{-4}	9.36×10^{-1}	9.30×10^{-1}	2.21×10^{-4}	3.41×10^{-5}
Mumbai	2.95×10^{-4}	3.49×10^{-4}	8.23×10^{-1}	8.92×10^{-1}	8.13×10^{-5}	1.33×10^{-4}
Pune	4.47×10^{-4}	2.48×10^{-4}	8.69×10^{-1}	7.11×10^{-1}	2.81×10^{-5}	1.11×10^{-4}
Thane	5.34×10^{-4}	3.18×10^{-4}	9.52×10^{-1}	8.83×10^{-1}	1.15×10^{-4}	8.00×10^{-5}
Belo Horizonte	5.58×10^{-4}	4.22×10^{-4}	8.21×10^{-1}	7.36×10^{-1}	8.55×10^{-6}	4.72×10^{-6}
Brasilia	5.13×10^{-4}	1.68×10^{-4}	9.20×10^{-1}	7.65×10^{-1}	1.45×10^{-5}	8.85×10^{-5}
Fortaleza	4.20×10^{-4}	4.22×10^{-4}	8.68×10^{-1}	8.36×10^{-1}	2.30×10^{-5}	5.34×10^{-5}
Rio de Janeiro	5.48×10^{-4}	4.04×10^{-4}	9.46×10^{-1}	9.19×10^{-1}	1.24×10^{-4}	1.70×10^{-5}
São Paulo	5.35×10^{-4}	3.45×10^{-4}	9.48×10^{-1}	8.94×10^{-1}	1.23×10^{-4}	1.13×10^{-5}

References

1. WorldPop, global high resolution population denominators project. <https://www.worldpop.org/> (2020).
2. 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).
3. COVID19India. Covid19india. <https://www.covid19india.org/> (2021).
4. Ministério da Saúde. Coronavírus brasil (in Portuguese). <https://covid.saude.gov.br/> (2021).
5. Department of Health & Human Services. Covid-19 state and county policy orders. <https://healthdata.gov/dataset/covid-19-state-and-county-policy-orders> (2020).
6. Bakker, M., Berke, A., Groh, M., Pentland, A. S. & Moro, E. Effect of social distancing measures in the new york city metropolitan area. *N. Engl. J. Public Policy* (2020).
7. Apple mobility trends reports. <https://www.apple.com/covid19/mobility> (2020).
8. NYC Department of Health and Mental Hygiene. NYC Health Data. <https://www1.nyc.gov/site/doh/covid/covid-19-data.page> (2021).
9. NYC Department of Health and Mental Hygiene. NYC Coronavirus Disease 2019 (COVID-19) Data. <https://github.com/nychealth/coronavirus-data> (2020).
10. Kraemer, M. U. *et al.* The effect of human mobility and control measures on the covid-19 epidemic in china. *Science* **368**, 493–497 (2020).
11. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet* **395**, 689–697 (2020).
12. Ferretti, L. *et al.* Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science* **368** (2020).
13. Bruce Aylward, W. L. Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>.
14. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
15. 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).