Supplementary Materials of: "Estimating the establishment of local transmission and the cryptic phase of the COVID-19 pandemic in the USA"

Jessica T. Davis[1*], Matteo Chinazzi [1*], Nicola Perra [2*,1], Kunpeng Mu[1], Ana Pastore y Piontti[1], Marco Ajelli[3,4], Natalie E. Dean[5], Corrado Gioannini[6], Maria Litvinova[6], Stefano Merler[3], Luca Rossi[6], Kaiyuan Sun[7], Xinyue Xiong[1], M. Elizabeth Halloran[8,9], Ira M. Longini Jr.[5], Cécile Viboud[7], Alessandro Vespignani[1,6,†]

[1]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA USA
[2]Networks and Urban Systems Centre, University of Greenwich, London, UK
[3]Bruno Kessler Foundation, Trento Italy
[4]Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA
[5]Department of Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, USA
[6]ISI Foundation, Turin, Italy
[7]Fogarty International Center, NIH, USA
[8]Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[9]Department of Biostatistics, University of Washington, Seattle, WA. USA
[*]These authors contribute equally to this work
[†]To whom correspondence should be addressed; E-mail: a.vespignani@northeastern.edu.

Materials and Methods

Here, we provide the details about the model calibration, present the sensitivity analysis of some key parameters, and describe the details of the importation sources estimation. We also include an analysis of the empirical data, several indicators (air traffic, population, density), and the data obtained via the model.

Global Epidemic and Mobility Model

We adopt the Global Epidemic and Mobility model (GLEAM), a stochastic, spatial, epidemic model based on a metapopulation approach that has been used and published previously (*7,*

8). In the model, the world is divided into over 3,200 geographic subpopulations constructed using a Voronoi tessellation of the Earth's surface. Subpopulations, centered around major transportation hubs (e.g. airports), consist of cells with a resolution of 15 x 15 arc minutes (approximately 25 x 25 kilometers). High resolution data are used to define the population of each cell (*50*). Other attributes of individual subpopulations, such as age specific contact patterns, health infrastructure, etc., are added according to available data (*21*).

GLEAM integrates a human mobility layer, represented as a network, using both short-range (i.e. commuting) and long-range (i.e. flights) mobility data from the Offices of Statistics for 30 countries on 5 continents as well as the Official Aviation Guide (OAG) and IATA databases (updated in 2019) (*13, 14*). The air travel network consists of the daily passenger flows between airport pairs (origin and destination) worldwide mapped to the corresponding subpopulations. We define a worldwide homogeneous standard for GLEAM to overcome differences in the spatial resolution of the commuting data across different countries. Where information is not available, the short-range mobility layer is generated synthetically by relying on the "gravity law" or the more recent "radiation law" both calibrated using the real data available (*51*). These approaches assume more frequent travel to nearby or closer subpopulations and less frequent travel to distant locations. In Fig. 1 we show a representation of the geographical resolution of the model and the mobility network for the contiguous United States (US).

Initial conditions are set specifying the number and location of individuals capable of transmitting the infection. GLEAM is then able to track over time the proportion of the population in each disease compartment for all subpopulations. At the start of each simulated day, travelers move to their destinations via the flight network. The probability of air travel changes from day to day, varies by age group, and can consider the effects of location specific airline traffic reductions. Short-range mobility (i.e. commuting) varies between workdays and weekends, by age group, and by disease status. Each full day is simulated using 12 distinct time steps, and this process is repeated for every simulated day. Individuals and their traveling patterns are tracked as shown in the flow diagram for the GLEAM algorithm (Fig. 2).

The combined population structure and mobility network create a synthetic world that is used to simulate the unfolding dynamics of the epidemic. The infection dynamics occur within each subpopulation. We adopt a classic $SLIR$ model in which individuals can be classified into four compartments: susceptible, latent, infectious, or removed. Susceptible individuals become latent through interactions with infectious individuals. During both the latent and infectious stages we assume that individuals are able to travel. Following the infectious period, individuals then progress into the removed compartment where they are no longer able to infect others, meaning they have either recovered, been hospitalized, isolated or have died. Individuals transition between compartments using stochastic binomial chain processes assuming parameter values from available literature that define the natural history of disease. In Table. 1 we report the parameter estimates used in the model.

Interventions Timeline

In order to realistically depict the evolution of the epidemic, a comprehensive set of policy in-

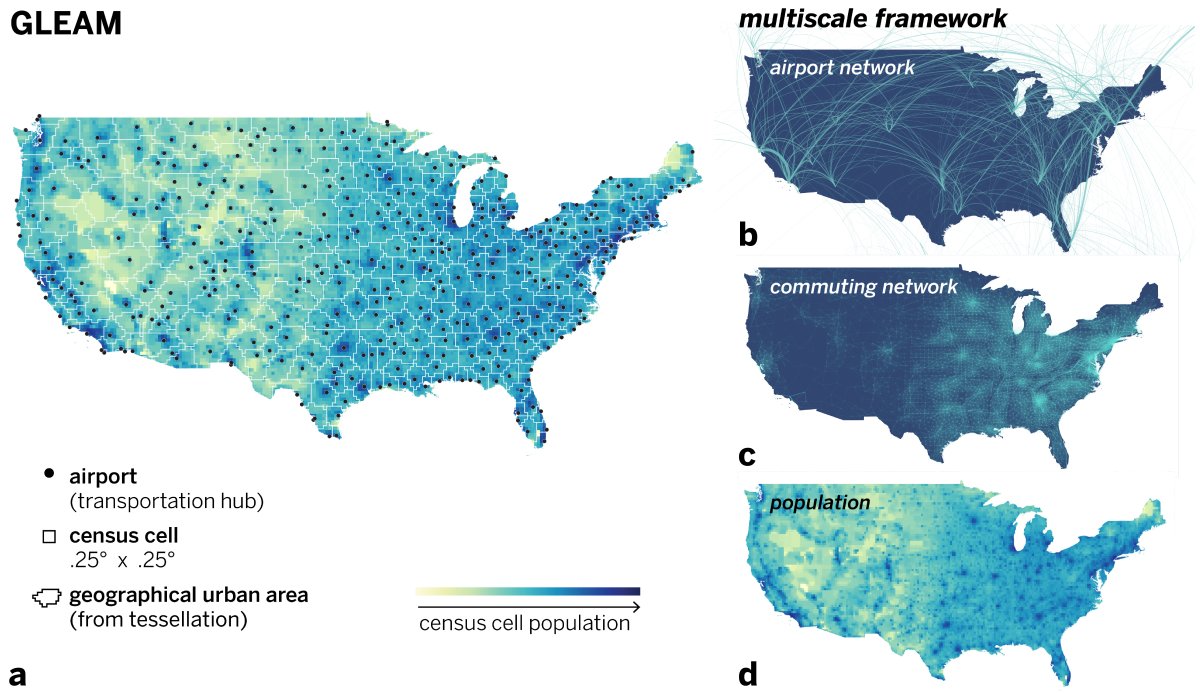

Fig. S1: Schematic representation of GLEAM. (A) The subpopulation structure of the US. Subpopulations are geographic regions, formed from the Voronoi tessellation, that are centered around airports. They are comprised of census cells that are approximately 25km x 25km. (B) Diagram of the origin destination airport network (long range mobility network) in the US. (C) Diagram of the commuting network (short-range mobility network) in the US. (C) The population layer of GLEAM showing the population size of census cells.

Parameters	**Range**	**Ref.**
Latent period (mean)	[4, 7] days	(*52*)
Infectious period (mean)	[2, 4] days	(*16*)
Days until recovery	[10, 14] days	(*16, 22*)
Generation time	[6, 8] days	
Reproductive number	[1.6, 3.3] in steps of 0.01	
Starting date	[2019-11-15, 2019-12-01]	(*20, 23–26*)

Table S1: Summary of parameter ranges explored in the sensitivity analysis of the model. Reproductive number and starting date are uniform prior for the model's calibration. Reference parameters are reported in the main text.

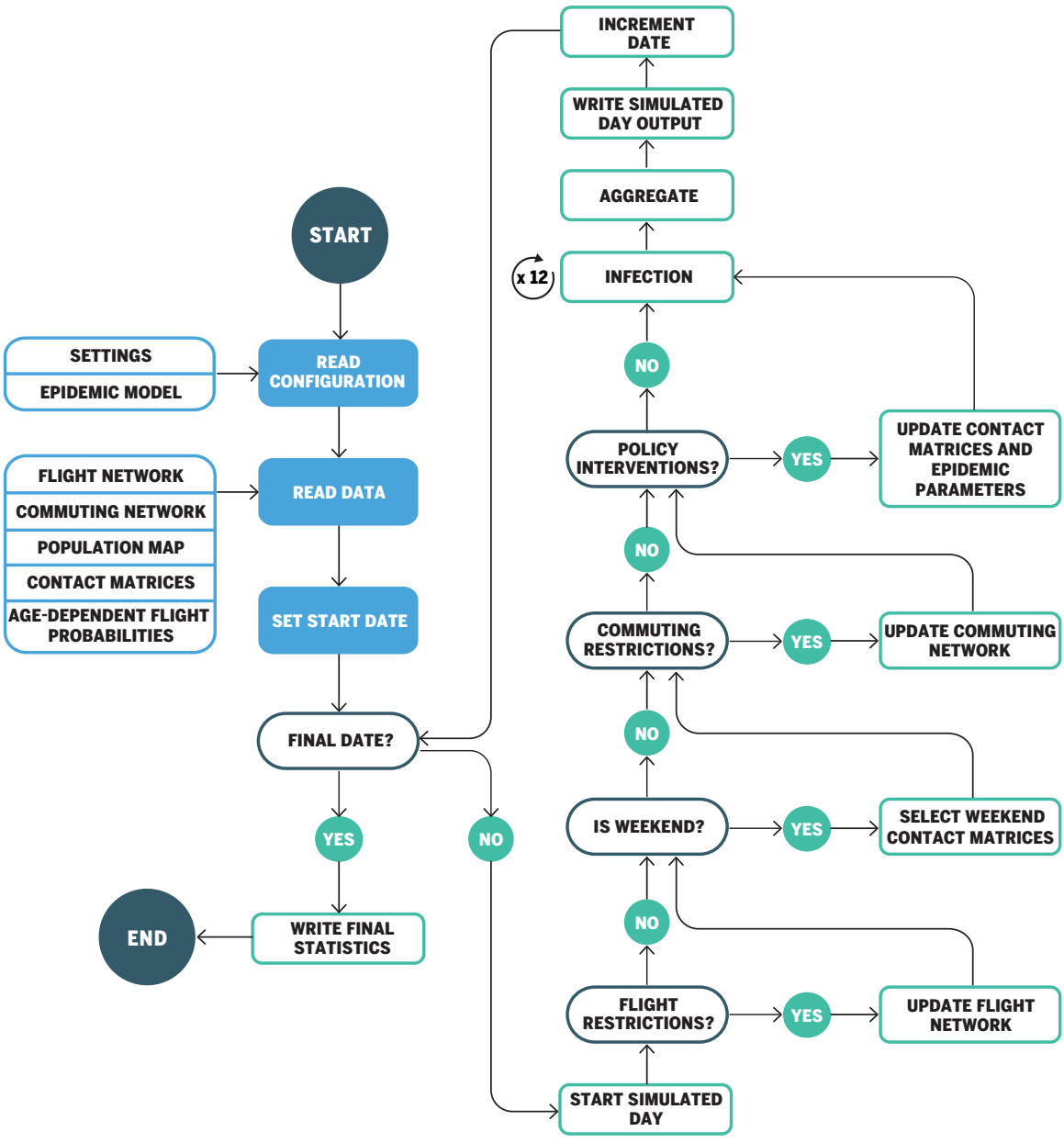

Fig. S2: Flow diagram of GLEAM's algorithm

terventions is applied to modify disease transmissibility and population mobility. On January 15, partial international travel reductions (from 10% to 40%) are applied for individuals traveling to/from China. Between January 23 and 28, flight and commuting reductions are applied to Wuhan and other subpopulations in the Hubei province to enforce government-mandated quarantines.

In addition, on January 25, commuting reductions are applied also to all other subpopulations in mainland China. To do so, we collected daily travel data starting January 1, 2020 until February 25, 2020 from the Baidu Qianxi platform (*53*), which provides three mobility indices (inflow index, outflow index, and intra-city index). The indices are proxies for the number of travelers moving in, out of, and inside a city, respectively. We extracted the mobility outflow index of 27 provinces and 4 municipalities for the current year 2020 and the previous year (with the same lunar date), and then mapped all provinces and municipalities to the metapopulation structure of the model to estimate the travel flow changes during the epidemic where the travel reduction can be estimated as $1 - \frac{I_{cur}}{I_{pre}}$, where I_{cur} and I_{pre} are the mobility outflow index of current year 2020 and previous year on the same lunar date, respectively.

On February 1, due to the increasing amount of restrictions implemented by various countries and airlines (*54–59*), stronger travel reductions are applied between mainland China and the rest of the world. We use actual worldwide (international and domestic) origin-destination traffic data from the OAG database to quantify travel reductions. We also apply case detection based on travel history and additional travel bans across pairs of countries according to the Oxford COVID-19 Government Response Tracker (OxCGRT) (*60*). We account as well for the intra-country mobility and contacts reduction in workplaces and social settings (*46*) using the COVID-19 Community Mobility reports obtained from Google (*61*).

Starting in mid-March all around the world, countries started to close schools as a means to slow the spread of COVID-19. We use the timeline of school closures provided by OxCGRT (*60*). As our model considers contact matrices for different settings, namely households, schools, workplaces and community contacts (*21*), we are able to quantify the decrease in contacts that individuals have in each of these environments. To implement the school closure in the United States we follow (*62*) where authors study the effects of school closure in the context of seasonal influenza epidemics. According to the date when schools were closed in the different states we consider a reduction of contacts in all individuals attending an educational institution (*60, 63*). In the United States, this intervention was applied at state level. Following the school closure, most states issued a "stay at home order". In this case, we consider that only contacts in the household and only essential workplaces remained open. Using the COVID-19 Community Mobility reports (*61*) we are able to compute the relative reduction on the number of contacts in workplaces, and community interaction as well the relative reduction in the intra-country mobility. This data is available at the state level for United States, Italy and Spain among other countries, starting on February 15, 2020. These reports are updated regularly and the last date for the data used in this manuscript ends on July 31, 2020. For countries where we do not have mobility reports available we assume that on the date that schools are closed there is a reduction in mobility of 50%, and 80% reduction when there is 'stay at home order'. When the interventions are relaxed the mobility reduction is relaxed accordingly.

Global Model Calibration

The model described is stochastic and outputs an ensemble of possible epidemic outcomes for each set of initial conditions. We seed the epidemic in Wuhan, China assuming a starting date between 11/15/2019 and 12/1/2019 with 20 initial infections (*26*). Given the doubling time of the epidemic this might corresponds to the epidemic starting from mid October to late November, 2019. We simulate epidemic scenarios sampling reproductive numbers (R_0) from a uniform prior in the range 1.6 to 3.3. We use an Approximate Bayesian Computation (ABC) Rejection Algorithm. The ABC rejection algorithm samples a set of parameter points θ (for instance R_0) according to a prior distribution and simulates through the model the dataset E'. A distance measure $s(E', E)$ determines the difference between E' and the evidence E based on a given metric. If the generated E' is outside a tolerance from the evidence E (i.e. $s(E', E) > \epsilon$) the sampled parameter value is discarded. The sampled parameters that are accepted provide an estimate of the likelihood with respect to the evidence E and allows us to calculate the posterior distribution $P(\theta, E)$. As evidence, E, we considered the cumulative number of SARS-CoV-2 cases internationally imported from China during the time window January 12 to January 21, 2020. The distance measure is at each date the difference between the SARS-CoV-2 cumulative imported cases generated by the model and the evidence with a tolerance provided by the under-detection interval estimated in Ref. (*29*). We also account for a non detectable 40% rate of asymptomatic individuals (sensitivity analysis ranging from 35% to 50%) (*64, 65*). The rejection algorithm accepts only configurations that satisfy the distance measure each day. This approach allows us to calibrate the model by incorporating both the growth rate of importations and their magnitude, scaled according to the under-detection estimates. The detailed list of importation events used is provided in Table S1 of the supplementary materials of Ref. (*12*). Using the ABC calibration and the age-stratified contact matrices for the US, the obtained posterior distribution $P(R_0 = x|E)$ for the basic reproductive number R_0 in the US has a median 2.7 [95% CI 2.4-3.1] (Fig. 3). The median values for state level reproductive numbers range from 2.6-2.8, with doubling times T_d in the range of $3.3 - 3.6$ days.

Sensitivity Analysis

In the main text we report the results of the model calibrated only on importation data from mainland China before the implementation of any travel restrictions. Here we show the results using the same compartmental model and parameters, however we calibrate the model using both the international importation events as well as an additional constraint on the weekly incidence of deaths reported in the US from March 21 to March 28, 2020. The deaths recorded in this period are most likely from infections that occurred before local interventions were implemented. The additional calibration further selects the simulation results that more closely match the evolution of the outbreak in the US. The average reproductive number R_0 for the individual states ranges $2.6 - 2.8$ with an average doubling time of 3.7 [90%CI 3.3-4.2] days. Both calibrations produce similar results and do not alter the conclusions of the main analysis. In Fig. 4 we show the distribution of the first dates when each continental state observed at least 10 locally generated transmission events per day. Model differences between the median date estimates between this calibration and the one in the main text are at most 1 day.

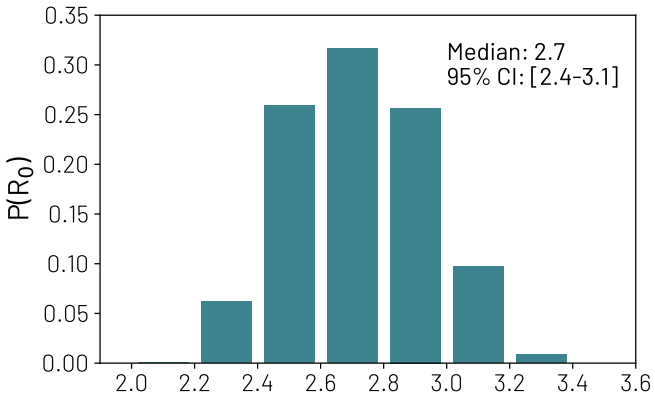

Fig. S3: The posterior distribution of the reproductive number for the US. The median value and the 95% reference range are shown.

Epidemic surveillance data

The data of officially confirmed cases and deaths for each state is taken from John Hopkins University Coronavirus resource centre (*36*). The data used in the analysis has been downloaded on May 23 and contains data up to that date.

Importation networks

The importation networks are obtained as follows. As a first step, we track the importations by air transportation (considering both individuals in the latent and infectious compartments) in any census areas of the US in all the runs selected using the ABC calibration. We then compute the day, in each run, in which the number of daily transitions from S to L is at least 10 in each state. In order words, we evaluate the date, in each run, when the state experienced the first local outbreak. We then track, in each run, the arrivals of latent and infectious individuals before or at the time of the local outbreak. In doing so, we aggregate the observations from census areas to states. Then in each build, for each state, we construct a directed and weighted network in which importation sources link the target states. The width of the link is the number of infections imported. By considering the observations in all the runs selected during the calibration phase, we build the chord diagram shown in the manuscript. To make them comparable, as the number of importations is a function of the source-target pair, we normalized the incoming flows for each target such that the sum of all incoming links is one. Furthermore, to help the readability of the plots, we aggregated sources considering macro areas such as Europe and Asia. We keep US (to isolate the national importations) and mainland China (as the epicenter of the pandemic) separate. All the other sources are grouped together and labeled "Others". More specifically, source countries of importations in the USA are grouped as:

- **Asia**: Afghanistan, Armenia, Azerbaijan, Bahrain, Bangladesh, Brunei, Cambodia, Cyprus, India, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, Korea, Kuwait, Kyrgyzstan, Lao PDR, Lebanon, Malaysia, Maldives, Mongolia, Myanmar, Nepal, Oman, Pakistan, Philippines, Qatar, Saudi Arabia, Singapore, Sri Lanka, Taiwan, Tajikistan, Thai-

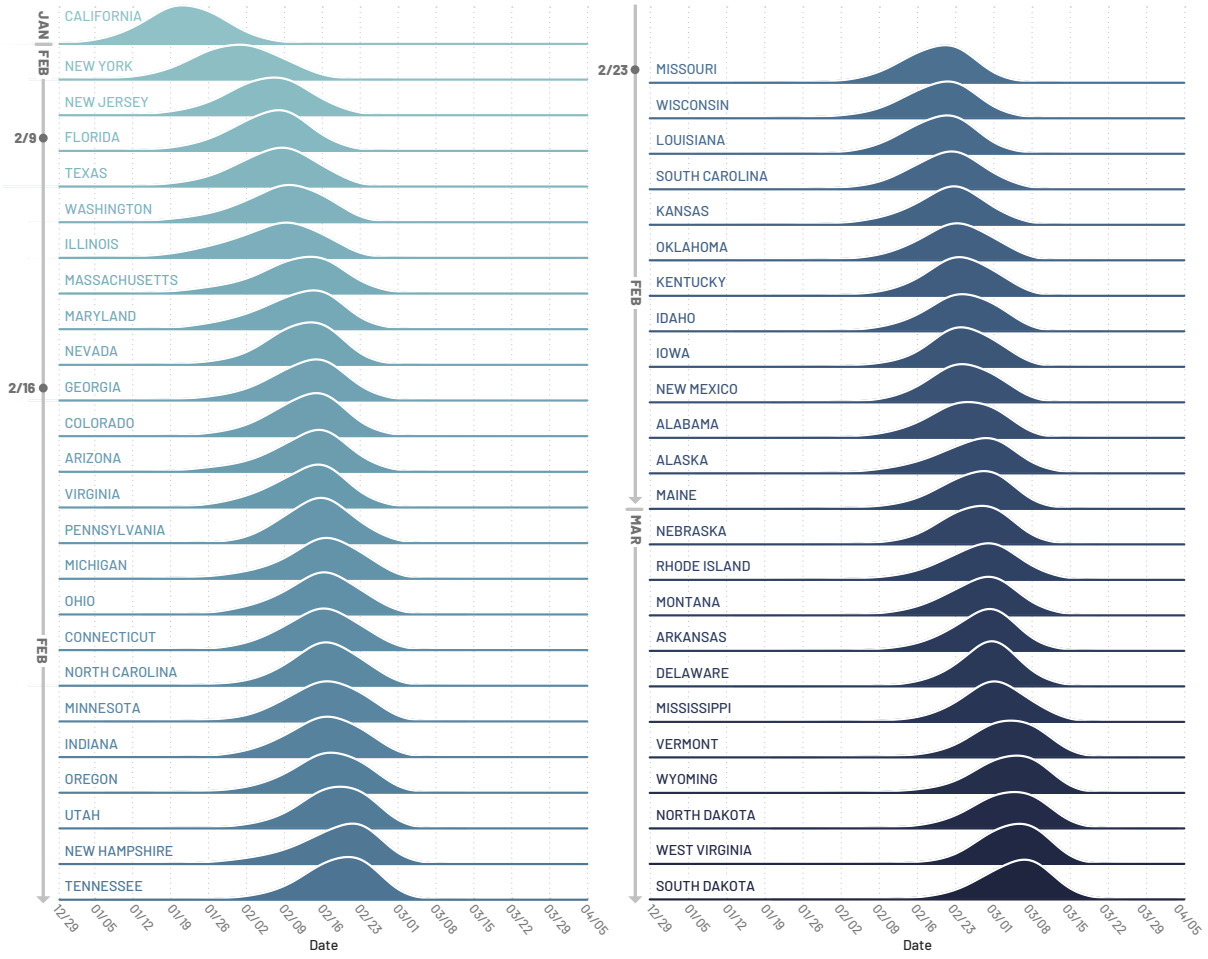

Fig. S4: Timing of the onset of local transmission. Distributions of the probability that each state had reached at least 10 locally generated COVID-19 transmission events per day by any given week since December 1, 2019.

land, Turkey, United Arab Emirates, Uzbekistan, Vietnam, Yemen

- **China**: mainland China
- **Europe**: Albania, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Gibraltar, Greece, Hungary, Iceland, Ireland, Isle of Man, Italy, Jersey, Kosovo, Latvia, Lithuania, Luxembourg, Macedonia, Malta, Moldova, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom
- **Others**: Algeria, Angola, Antigua and Barbuda, Argentina, Aruba, Australia, Bahamas, Barbados, Belize, Bermuda, Bolivia, Botswana, Brazil, British Virgin Islands, Burundi, Cameroon, Canada, Cape Verde, Caribbean Netherlands, Cayman Islands, Chile, Colombia, Congo, Cook Islands, Costa Rica, Cuba, Curaçao, Côte d'Ivoire, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Fiji, French Guiana, French Polynesia, Gambia, Ghana, Greenland, Grenada, Guadeloupe, Guatemala, Guinea, Guyana, Haiti, Honduras, Jamaica, Kenya, Liberia, Madagascar, Martinique, Mauritius, Mexico, Morocco, Mozambique, Namibia, New Zealand, Nicaragua, Nigeria, Palau, Panama, Paraguay, Peru, Rwanda, Samoa, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, St-Barthélemy, St. Kitts and Nevis, St. Lucia, St. Maarten, St. Vincent and Grenadines, Sudan, Suriname, Tonga, Trinidad and Tobago, Tunisia, Turks and Caicos Islands, Uganda, Uruguay, Vanuatu, Venezuela, Zambia, Zanzibar, Zimbabwe
- **USA**: all the US states plus the US territories (American Samoa, Guam, Northern Mariana Islands, Puerto Rico, U.S. Virgin Islands)

Following this procedure, in Table 2 we report the seeding share for the first 25 states that experienced a local outbreak (same states as in the left column of Fig. 2 in the manuscript).

In Table 3 we report the seeding shares for the remaining states that experienced a local outbreak (same states as in the right column of Fig. 2 in the manuscript).

In Table 4 we report the the share of introduction of SARS-Cov-2 infections for all the first 25 states that experienced a local outbreak (same states as in the left column of Fig. 2 in the manuscript), considering, however, all the infections imported up to March 1, 2020. In other words, we don't constrain on the time of the start of the local outbreak (i.e. first daily 10 local transmissions), but we consider a general cutoff date for all states. The flows are radically different, especially for the first states, where the critical role of China before the travel restrictions of January 23, is replaced by a much larger internal flow. In fact, for all states the share of domestic importations is above 50%.

In Table 5 we report the share of introduction of SARS-Cov-2 infections through March 1, 2020, for the remaining states that experienced a local outbreak (same states in the right column

State	USA	China	Asia	Europe	Others
California	0.13	0.60	0.17	0.02	0.06
New York	0.27	0.24	0.25	0.14	0.10
New Jersey	0.33	0.12	0.26	0.16	0.13
Florida	0.59	0.02	0.11	0.13	0.15
Texas	0.60	0.04	0.19	0.07	0.11
Washington	0.60	0.04	0.22	0.05	0.10
Illinois	0.52	0.06	0.22	0.08	0.11
Massachusetts	0.53	0.04	0.16	0.15	0.12
Maryland	0.51	0.03	0.24	0.12	0.10
Nevada	0.63	0.01	0.14	0.07	0.16
Georgia	0.64	0.02	0.18	0.08	0.09
Arizona	0.71	0.01	0.09	0.04	0.15
Virginia	0.57	0.02	0.20	0.11	0.10
Colorado	0.73	$<$0.01	0.08	0.06	0.13
Pennsylvania	0.70	$<$0.01	0.09	0.09	0.12
Ohio	0.71	0.01	0.11	0.07	0.10
Michigan	0.65	0.02	0.14	0.08	0.11
Connecticut	0.49	0.02	0.19	0.16	0.14
North Carolina	0.70	0.01	0.11	0.08	0.10
Minnesota	0.67	$<$0.01	0.13	0.06	0.13
Indiana	0.67	0.01	0.13	0.08	0.11
Oregon	0.72	0.01	0.14	0.03	0.10
Utah	0.77	$<$0.01	0.08	0.04	0.10
New Hampshire	0.55	0.02	0.14	0.16	0.13
Tennessee	0.73	$<$0.01	0.09	0.06	0.12

Table S2: Importation of seeding events. Sources are listed from the second column on. Targets are the states listed in the first column. This is the list of the first 25 states that experienced an outbreak. Numbers are rounded to the second digit.

State	USA	China	Asia	Europe	Others
Missouri	0.76	<0.01	0.08	0.05	0.10
Wisconsin	0.82	<0.01	0.03	0.04	0.11
Louisiana	0.76	<0.01	0.06	0.06	0.11
South Carolina	0.75	<0.01	0.09	0.07	0.08
Kansas	0.78	<0.01	0.07	0.05	0.09
Oklahoma	0.77	<0.01	0.08	0.06	0.10
Kentucky	0.75	<0.01	0.08	0.07	0.10
Idaho	0.83	<0.01	0.06	0.03	0.08
Iowa	0.77	<0.01	0.06	0.05	0.12
New Mexico	0.84	<0.01	0.04	0.04	0.07
Alabama	0.74	<0.01	0.12	0.06	0.08
Alaska	0.75	<0.01	0.11	0.03	0.11
Maine	0.77	<0.01	0.06	0.08	0.09
Nebraska	0.79	<0.01	0.05	0.04	0.12
Rhode Island	0.85	<0.01	0.01	0.08	0.05
Montana	0.88	<0.01	0.01	0.03	0.08
Arkansas	0.81	<0.01	0.05	0.06	0.08
Delaware	0.70	<0.01	0.04	0.12	0.14
Mississippi	0.81	<0.01	0.04	0.07	0.08
Vermont	0.87	<0.01	0.04	0.04	0.05
Wyoming	0.81	<0.01	0.07	0.03	0.09
North Dakota	0.90	0.	0.01	0.01	0.08
West Virginia	0.84	<0.01	0.02	0.06	0.08
South Dakota	0.87	0.	<0.01	0.03	0.09

Table S3: Importation of seeding events. Sources are listed from the second column on. Targets are the states listed in the first column. This is the list of the last 24 states that experienced an outbreak. Numbers are rounded to the second digit.

State	USA	China	Asia	Europe	Others
California	0.69	<0.01	0.13	0.05	0.13
New York	0.56	<0.01	0.13	0.17	0.14
New Jersey	0.57	<0.01	0.12	0.16	0.14
Florida	0.74	<0.01	0.03	0.08	0.15
Texas	0.82	<0.01	0.06	0.04	0.08
Washington	0.82	<0.01	0.07	0.03	0.07
Illinois	0.75	<0.01	0.08	0.06	0.11
Massachusetts	0.71	<0.01	0.06	0.11	0.11
Maryland	0.72	<0.01	0.10	0.09	0.09
Nevada	0.79	<0.01	0.04	0.04	0.13
Georgia	0.82	<0.01	0.06	0.05	0.07
Arizona	0.83	<0.01	0.02	0.02	0.13
Virginia	0.77	<0.01	0.08	0.07	0.08
Colorado	0.86	<0.01	0.03	0.03	0.08
Pennsylvania	0.83	<0.01	0.03	0.05	0.09
Ohio	0.85	<0.01	0.03	0.04	0.07
Michigan	0.82	<0.01	0.05	0.05	0.08
Connecticut	0.67	<0.01	0.09	0.12	0.13
North Carolina	0.84	<0.01	0.04	0.05	0.07
Minnesota	0.81	<0.01	0.05	0.03	0.10
Indiana	0.82	<0.01	0.05	0.04	0.08
Oregon	0.87	<0.01	0.04	0.02	0.07
Utah	0.89	<0.01	0.03	0.02	0.06
New Hampshire	0.70	<0.01	0.07	0.12	0.11
Tennessee	0.85	<0.01	0.03	0.03	0.08

Table S4: Introduction of SARS-Cov-2 infections thorugh March 1. Sources are listed from the second column on. Targets are the states listed in the first column. This is the list of the first 25 states that experienced an outbreak. Numbers are rounded to the second digit.

State	USA	China	Asia	Europe	Others
Missouri	0.87	<0.01	0.03	0.03	0.08
Wisconsin	0.89	<0.01	0.01	0.02	0.08
Louisiana	0.87	<0.01	0.03	0.03	0.07
South Carolina	0.86	<0.01	0.04	0.04	0.06
Kansas	0.88	<0.01	0.03	0.03	0.07
Oklahoma	0.87	<0.01	0.03	0.03	0.07
Kentucky	0.86	<0.01	0.04	0.03	0.07
Idaho	0.91	<0.01	0.03	0.01	0.05
Iowa	0.86	<0.01	0.03	0.02	0.09
New Mexico	0.92	<0.01	0.02	0.02	0.04
Alabama	0.85	<0.01	0.07	0.03	0.05
Alaska	0.90	<0.01	0.05	0.01	0.05
Maine	0.85	<0.01	0.03	0.04	0.07
Nebraska	0.87	<0.01	0.02	0.02	0.09
Rhode Island	0.89	<0.01	0.01	0.06	0.04
Montana	0.94	<0.01	<0.01	0.01	0.05
Arkansas	0.89	<0.01	0.02	0.03	0.05
Delaware	0.80	<0.01	0.03	0.06	0.10
Mississippi	0.89	<0.01	0.03	0.03	0.05
Vermont	0.91	<0.01	0.03	0.02	0.04
Wyoming	0.86	<0.01	0.05	0.02	0.08
North Dakota	0.92	0.	0.01	0.01	0.06
West Virginia	0.90	<0.01	0.01	0.03	0.05
South Dakota	0.91	0.	<0.01	0.01	0.08

Table S5: Importation of infection flows up to March 1. Sources are listed from the second column on. Targets are the states listed in the first column. This is the list of the last 24 states that experienced an outbreak. Numbers are rounded to the second digit.

of Fig. 2 in the manuscript).

In Figure 5 we provide a visual representation, via a chord plot, of the data shown in Tables 4 and 5.

Correlation analysis

As mentioned and shown in the main text, during the early phases of the spreading, mobility plays a crucial role. In order to highlight this aspect, here we report the full correlation analysis between the real data and the mobility indicators. In particular, we compute the order in which states reached 100 cases/deaths in the real surveillance data and compare it with the order of states according to their air traffic (considering both national and international travels). Note how the correlation plot in the main text considered as a mobility indicator the sum of the two

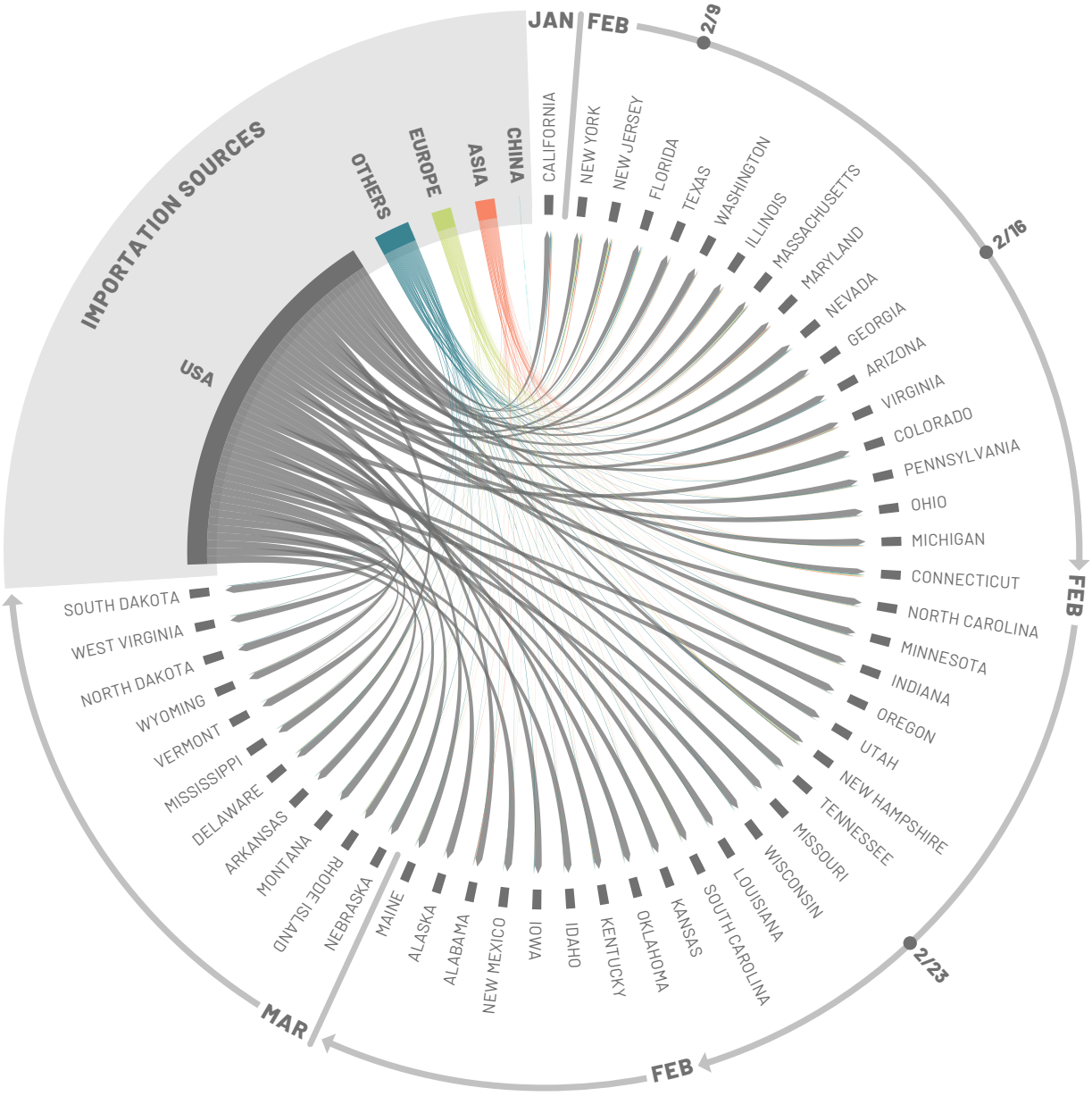

Fig. S5: Share of importations of infections in all continental states from US, China, Europe, Asia and all other countries up to March 1, 2020. US states are ordered, clockwise, according to the start of the local outbreak.

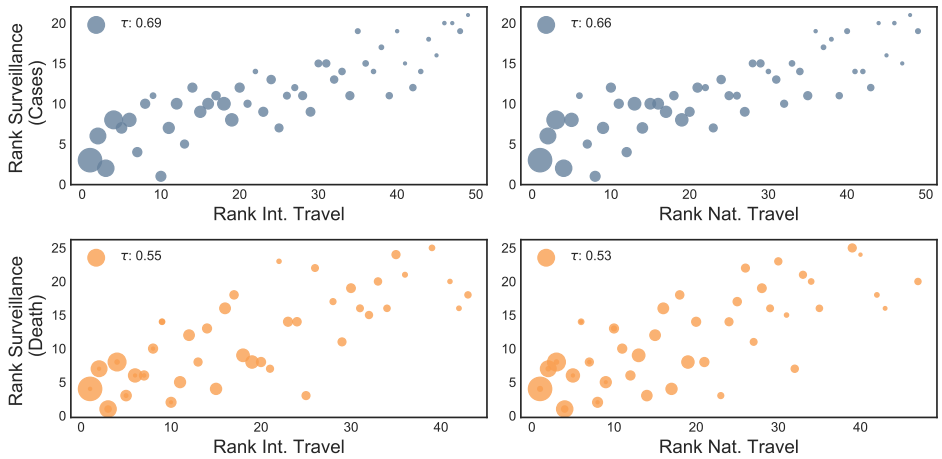

Fig. S6: Correlation between the order in which states reached 100 confirmed cases (top row) or deaths (bottom row) and their International (left) or National (right) air traffic. The size of each state is assigned proportional to the population size

types of traffic. In Fig. 6 we show the result reporting also the value of the Kendall's tau. Overall the correlations, especially for the number of cases, are comparable with those shown in the main text.

The states that were the first to experience the outbreak, besides being hubs in the air transportation network, are also very populous. It is then natural to wonder how rankings based on population compare with respect to those based on air traffic. In Figure 7 we show the comparison. In particular, we order states according to their population (left column) and density (right column) and to the epidemic indicators from surveillance (cases and deaths). Not surprisingly, we find high correlation levels across the board. However, it is interesting to note how those reported considering air travels are comparable (0.67 when considering both national and international traffic). The correlation for the number of deaths (bottom row) is lower with respect to the number of cases (top row). Furthermore, it is interesting to notice how the correlations are even smaller when considering the population density (right column). In fact, they go down to 0.38 when comparing them with the order in which a state surpassed the threshold of 100 deaths.

In Figure 8 we repeat the same analysis considering the model's projections. The correlations are comparable to the previous. Also in the model, population density is less correlated. It is important to observe that air travel traffic, population and population densities are not independent indicators. Figure 9 highlights this observation. Particularly high is the correlation between air traffic and population. Also, population and population densities are well correlated while air traffic and density are not. This is due, in part, to the many states that, due to their location, see lots of traffic but are not very dense.

The correlations between rankings reported above have been computed by using the Kendall's tau (*66*) as implemented by the *scipy.stats* library (*67*). The metric is designed to compare the rankings obtained ordering items, states in our case, according to pairs of different quantities. The Kendall's tau is defined only in the case that the ranks have the same size. In case the two

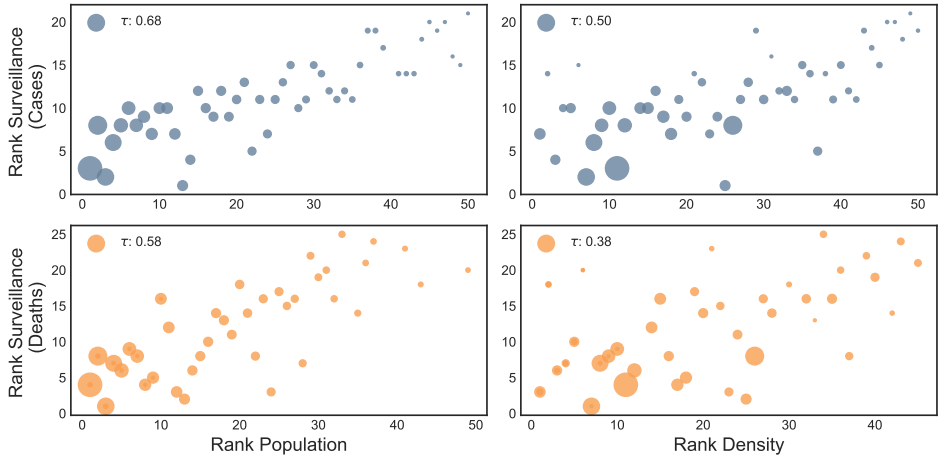

Fig. S7: Correlation between the order in which states reached 100 confirmed cases (top row) or deaths (bottom row) and their population (left) or population density (right). The size of each state is assigned proportional to the population size

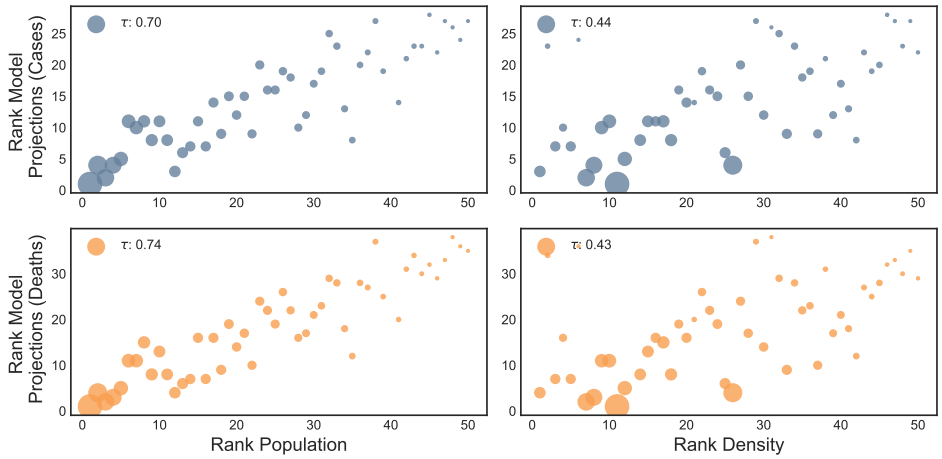

Fig. S8: Correlation between the order in which states reached 100 confirmed cases (top row) or deaths (bottom row) according to the model and their population (left) or population density (right). The size of each state is assigned proportional to the population size

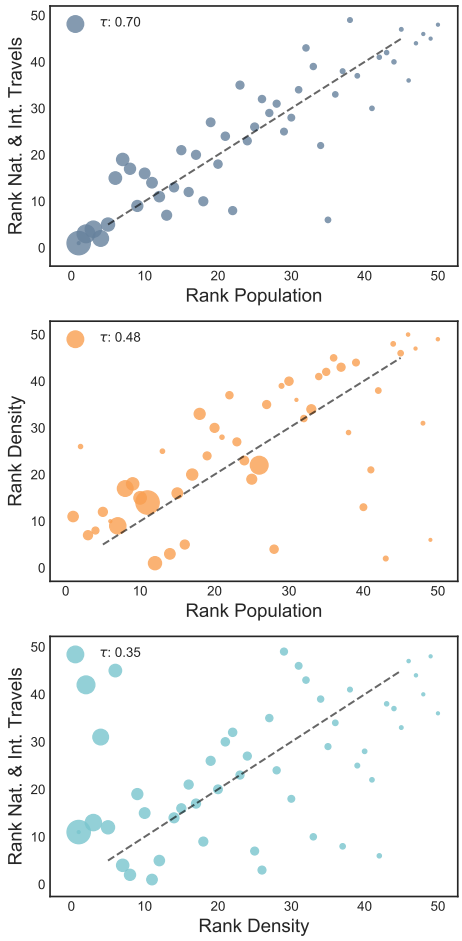

Fig. S9: Correlation between travel, population and population density for the continental US states.

ranks have different size (i.e. some states did not yet go above a given threshold) the metric is applied to the common subset of the two.

References

1. Centers for Disease Control and Prevention (CDC), "First Travel-related Case of 2019 Novel Coronavirus Detected in United States" (2020); https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html.

2. Business Insider, "2nd case of the Wuhan coronavirus in the US was just confirmed in Chicago" (2020); https://www.businessinsider.fr/us/wuhan-coronavirus-second-confirmed-case-in-the-us-chicago-2020-1.

3. Arizona Department of Health Services, "Public Health Agencies Confirm 2019 Novel Coronavirus Case in Arizona" (2020); https://www.azdhs.gov/director/public-information-office/index.php#news-release-012620.

4. Los Angeles County Public Health, "Public Health Confirms First Case of 2019 Novel Coronavirus in Los Angeles County" (2020); http://publichealth.lacounty.gov/phcommon/public/media/mediapubhpdetail.cfm?prid=2227.

5. Illinois Department of Public Health, "Second Illinois 2019 Novel Coronavirus Case Identified" (2020); http://www.dph.illinois.gov/news/second-illinois-2019-novel-coronavirus-case-identified.

6. CDC, "Updated Guidance on Evaluating and Testing Persons for Coronavirus Disease 2019 (COVID-19)"; https://emergency.cdc.gov/han/2020/han00429.asp.

7. D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, A. Vespignani, Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*. **106**, 21484-21489 (2009).

8. D. Balcan, B. Gonçalves, H. Hu, J.J. Ramasco, V. Colizza, A. Vespignani, Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *Journal of Computational Science* **1**, 132-145 (2010).

9. D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J.J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, A. Vespignani, Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC medicine* **7**, 45 (2009).

10. Q. Zhang, K. Sun, M. Chinazzi, A. Pastore y Piontti, N.E. Dean, D.P. Rojas, S. Merler, D. Mistry, P. Poletti, L. Rossi, M. Bray, M.E. Halloran, I.M. Longini, A. Vespignani, Spread of Zika virus in the Americas. *Proceedings of the National Academy of Sciences* **114**, E4334 (2017).

11. A. P. y Piontti, N. Perra, L. Rossi, N. Samay, A. Vespignani, *Charting the Next Pandemic: Modeling Infectious Disease Spreading in the Data Science Age* (Springer, 2018).

12. M. Chinazzi, J.T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A.P. y Piontti, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, E.M. Halloran, I.M. Longini, A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).

13. International Air Transportation Association. https://www.iata.org/.

14. Official Aviation Guide. https://www.oag.com/.

15. M. Gatto,E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences* **117**, 10484 (2020).

16. S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch, Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period *Science* **368**, 860 (2020).

17. R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Jeffrey, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489 (2020).

18. J. T. Wu, K. Leung, G. M. Leung, 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).

19. S. Lai, N. W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J. R. Floyd, A. Wesolowski, M. Santillana, C. Zhang, X. Du, H. Yu, and A. J. Tatem, Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* (2020). doi.org/10.1038/s41586-020-2293-x

20. N. Imai, A. Cori, I. Dorigatti, M. Baguelin, C. A. Donnelly, S. Riley, N. M. Ferguson, "Report 3: Transmissibility of 2019-nCoV" (Imperial College London, 2020) www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-3-transmissibility-of-covid-19/.

21. D. Mistry, M. Litvinova, M. Chinazzi, L. Fumanelli, M. F. Gomes, S. A. Haque, Q.-H. Liu, K. Mu, X. Xiong, M. E. Halloran, I.M. Longini Jr., S. Merler, M. Ajelli, A. Vespignani. Inferring high-resolution human mixing patterns for disease modeling. *arXiv* [Preprint]. 25 February 2020. https://arxiv.org/abs/2003.01214

22. R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases* (2020). https://doi.org/10.1016/S1473-3099(20)30243-7

23. A. Rambaut, "Preliminary phylogenetic analysis of 11 nCoV2019 genomes, 2020-01-19" (2020); http://virological.org/t/preliminary-phylogenetic-analysis-of-11-ncov2019-genomes-2020-01-19/329.

24. K. Anderson, "Estimates of the clock and TMRCA for 2019-nCoV based on 27 genomes" (2020); http://virological.org/t/clock-and-tmrca-based-on-27-genomes/347

25. T. Bedford, R. Neher, J. Hadfield, E.Hodcroft, M. Ilcisin, N. Müller, "Genomic analysis of nCoV spread. Situation report 2020-01-23" (2020); https://nextstrain.org/narratives/ncov/sit-rep/2020-01-23

26. L. van Dorp, M. Acmana, D. Richard, L.P. Shawd, C.E. Ford, L. Ormond, C.J. Owen, J. Pang, C.C.S. Tan, F.A.T. Boshier, A.T. Ortiz, F. Balloux, Emergence of genomic diversity and recurrent mutations in SARS-CoV-2 *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* **83** 104351 (2020).

27. M. Sunnåker, A.G. Busetto, E. Numminen, J. Corander, M. Foll, C. Dessimoz, Approximate Bayesian Computation, *PLoS Comput Biol*. **9**, e1002803 (2013).

28. P.M. Salazar, R. Niehus, A.R. Taylor, C.O. Buckee, M. Lipsitch. Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. *medRxiv* 2020.02.04.20020495 [Preprint]. 11 February 2020. doi:10.1101/2020.02.04.20020495

29. R. Niehus, P. M. De Salazar, A. Taylor, M. Lipsitch, Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China. *The Lancet Infectious Diseases* (2020). https://doi.org/10.1016/S1473-3099(20)30229-2.

30. Global security index https://www.ghsindex.org/.

31. D. Cereda, M. Tirani, F. Rovida, V. Demicheli, M. Ajelli, P. Poletti, F. Trentini, G. Guzzetta, V. Marziano, A. Barone, M. Magoni, S. Deandrea, G. Diurno, M. Lombardo, M. Faccini, A. Pan, R. Bruno, E. Pariani, G. Grasselli, A. Piatti, M. Gramegna, F. Baldanti, A. Melegaro,S. Merler The early phase of the COVID-19 outbreak in Lombardy, Italy, *arXiv* [Preprint]. https://arxiv.org/abs/2003.09320 (2020)

32. O. Pybus, A. Rambaut, L. du Plessis, A.E. Zarebski, M. U. G. Kraemer, J. Raghwani, B. Gutiérrez, V. Hill, J. McCrone, R. Colquhoun, B. Jackson, Á. O'Toole, J. Ashworth Preliminary analysis of SARS-CoV-2 importation establishment of UK transmission lineages, https://www.oxfordmartin.ox.ac.uk/publications/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/ (2020).

33. Santa Clara County Public Health, "County of Santa Clara Identifies Three Additional Early COVID-19 Deaths", https://www.sccgov.org/sites/covid19/Pages/press-release-04-21-20-early.aspx (2020).

34. CDC Cases, Data Surveillance https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/commercial-lab-surveys.html.

35. F.P. Havers, C. Reed, T. Lim, J.M. Montgomery, J.D. Klena, A.J. Hall, A.M. Fry, D.L. Cannon, C. Chiang, A. Gibbons, I. Krapiunaya, M. Morales-Betoulle, K. Roguski, M.A.U. Rasheed, B. Freeman, S. Lester, L. Mills, D.S. Carroll, S.M. Owen, J.A. Johnson, V. Semenova, C. Blackmore, D. Blog, S.J. Chai, A. Dunn, J. Hand, S. Jain, S. Lindquist, R. Lynfield, S. Pritchard, T. Sokol, L. Sosa, G. Turabelidze, S.M. Watkins, J. Wiesman, R.W. Williams, S. Yendell, J. Schiffer, N.J. Thornburg, Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. *JAMA Intern Med.* (2020). doi:10.1001/jamainternmed.2020.4130.

36. John Hopkins University Coronavirus Resource Centre https://coronavirus.jhu.edu/.

37. M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, Open COVID-19 Data Working Group, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, S. V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, 493-497 (2020).

38. D. Weinberger, T. Cohen, F. Crawford, F. Mostashari, D. Olson, V. E. Pitzer, N. G. Reich, M. Russi, L. Simonsen, A. Watkins, C. Viboud. Estimation of Excess Deaths Associated With the COVID-19 Pandemic in the United States, March to May 2020. *JAMA Intern Med.* 1 July 2020 doi:10.1001/jamainternmed.2020.3391.

39. "Nextstrain: Real-time tracking of pathogen evolution", https://nextstrain.org/ (2020).

40. M. Worobey, J. Pekar, B. B. Larsen, M. I. Nelson, V. Hill, J. B. Joy, A. Rambaut, M. A. Suchard, J. O. Wertheim, P. Lemey. *bioRxiv* 10.1101/2020.05.21.109322 [Preprint]. 23 March 2020. doi:10.1101/2020.05.21.109322

41. A. S. Gonzalez-Reiche, M. M. Hernandez, M. Sullivan, B. Ciferri, H. Alshammary, A. Obla, S. Fabre, G. Kleiner, J. Polanco, Z. Khan, B. Alburquerque, A. van de Guchte, J. Dutta, N. Francoeur, B. S. Melo, I. Oussenko, G. Deikus, J. Soto, S. H. Sridhar, Y.-C. Wang, K. Twyman, A. Kasarskis, D. R. Altman, M. Smith, R. Sebra, J. Aberg, F. Krammer, A. Garcia-Sarstre, M. Luksza, G. Patel, A. Paniz-Mondolfi, M. Gitman, E. M. Sordillo, V. Simon, H. van Bakel. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* (2020). doi:10.1126/science.abc1917.

42. A. Adiga, S. Venkatramanan, J. Schlitt, A. Peddireddy, A. Dickerman, A. Bura, A. Warren, B. D. Klahn, C. Mao, D. Xie, D. Machi, E. Raymond, F. Meng, G. Barrow, H. Mortveit, J. Chen, J. Walke, J. Goldstein, M. L. Wilson, M. Orr, P. Porebski, P. A. Telionis, R. Beckman, S. Hoops, S. Eubank, Y. Y. Baek, B. Lewis, M. Marathe, C. Barrett. *medRxiv* 10.1101/2020.02.20.20025882 [Preprint]. 2 March 2020. doi:10.1101/2020.02.20.20025882

43. C. R. Wells, P. Sah, S. M. Moghadas, A. Pandey, A. Shoukat, Y. Wang, Z. Wang, L. A. Meyers, B. H. Singer, and A. P. Galvani. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proceedings of the National Academy of Sciences* **117**, 7504 (2020).

44. L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, V. Colizza. Expected impact of lockdown in Île-de-France and possible exit strategies. *medRxiv* 10.1101/2020.04.13.20063933 [Preprint]. 17 April 2020. doi:10.1101/2020.04.13.20063933

45. A. Aleta, D. Martin-Corral, A. Pastore y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini, S. Merler, A. Pentland, A. Vespignani, E. Moro, and Y. Moreno. Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. *medArxiv* 10.1101/2020.05.06.20092841 [Preprint]. 18 May 2020. doi:10.1101/2020.05.06.20092841.

46. J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, and H. Yu. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* (2020). doi:10.1126/science.abb8001.

47. N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, CMMID COV ID-19 working group, R. M. Eggo, Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine* (2020). https://doi.org/10.1038/s41591-020-0962-9

48. Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, T. Feng, Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases* (2020). https://doi.org/10.1016/S1473-3099(20)30287-5.

49. J. Raude, M. Debin, C. Souty, C. Guerrisi, C. Turbelin, A. Falchi, I. Bonmarin, D. Paolotti, Y. Moreno, C. Obi, J. Duggan, A. Wisniak, A. Flahault, T. Blanchon, V. Colizza, Are people excessively pessimistic about the risk of coronavirus infection? *medArxiv*. 10.31234/osf.io/364qj [Preprint]. 5 March 2020. doi:10.31234/osf.io/364qj

50. Socioeconomic Data and Applications Center (SEDAC), Columbia University. http://sedac.ciesin.columbia.edu/gpw.

51. F. Simini, M. C. González, A. Maritan, A.-L. Barabási, A universal model for mobility and migration patterns, *Nature* **484**, 96-100 (2012).

52. J. A. Backer, D. Klinkenberg, J. Wallinga, Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020.*Eurosurveillance* **25**(5), 2000062 (2020). https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062

53. Baidu Qianxi, http://qianxi.baidu.com/ (2020).

54. New York Times, "North Korea Bans Foreign Tourists Over Coronavirus, Tour Operator Says", https://www.nytimes.com/2020/01/21/world/asia/coronavirus-china-north-korea-tourism-ban.html (2020).

55. CNA, "Scoot cancels flights to China's Wuhan over virus outbreak", https://www.channelnewsasia.com/news/singapore/wuhan-virus-scoot-cancels-flights-mtr-train-12309076 (2020).

56. Toui tre News, "Vietnam aviation authority ceases all flights to and from coronavirus-stricken Wuhan", https://tuoitrenews.vn/news/business/20200124/vietnam-aviation-authority-ceases-all-flights-to-and-from-coronavirusstricken-wuhan/52707.html (2020).

57. Reuters, "Russia ramps up controls, shuts China border crossings over virus fears", https://www.reuters.com/article/us-china-health-russia-border/russian-regions-in-far-east-close-border-with-china-amid-coronavirus-fears-tass-idUSKBN1ZR0TU (2020).

58. Center for Disease Control, "Novel Coronavirus in China", https://wwwnc.cdc.gov/travel/notices/warning/novel-coronavirus-china (2020).

59. The Australian, "Travelers from China to be denied entry to Australia", https://www.theaustralian.com.au/subscribe/news/1/?sourceCode=TAWEB_WRE170_a&dest=https%3A%2F%2Fwww.theaustralian.com.au%2Fnation%2Ftravellers-from-china-to-be-denied-entry-into-australia%2Fnews-story%2F7b7619d44af78dd7395a934e22b52997&memtype=anonymous&mode=premium (2020).

60. T. Hale, S. Webster, A. Petherick, T. Phillips, B. Kira *Oxford COVID-19 Government Response Tracker*, Blavatnik School of Government (2020).

61. Google LLC "Google COVID-19 Community Mobility Reports", https://www.google.com/covid19/mobility/ Accessed: August 4,2020.

62. G. De Luca,K. Van Kerckhove, P. Coletti, C. Poletto, N. Bossuyt, N. Hens, V. Colizza, The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium. *BMC infectious diseases* **18**, 1-16 (2018).

63. The New York Times, "See which states and cities have told their residents to stay at home." https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html

64. D. Oran, E. Topol, Prevalence of asymptomatic SARS-CoV-2 infection. *Ann. Internal Med* (2020,). https://doi.org/10.7326/M20-3012.

65. CDC,"COVID-19 Pandemic Planning Scenarios." https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html

66. M. G. Kendall, A new measure of rank correlation. *Biometrika* **30**, 81 (1938).

67. Scipy.org: Kendall Tau. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kendalltau.html.