



OPEN

A high-resolution flux-matrix model describes the spread of diseases in a spatial network and the effect of mitigation strategies

Guillaume Le Treut^{1✉}, Greg Huber^{1✉}, Mason Kamb¹, Kyle Kawagoe², Aaron McGeever¹, Jonathan Miller³, Reuven Pnini³, Boris Veytsman^{4,5} & David Yllanes^{1,6}

Propagation of an epidemic across a spatial network of communities is described by a variant of the SIR model accompanied by an intercommunity infectivity matrix. This matrix is estimated from fluxes between communities, obtained from cell-phone tracking data recorded in the USA between March 2020 and February 2021. We apply this model to the SARS-CoV-2 pandemic by fitting just one global parameter representing the frequency of interaction between individuals. We find that the predicted infections agree reasonably well with the reported cases. We clearly see the effect of “shelter-in-place” policies introduced at the onset of the pandemic. Interestingly, a model with uniform transmission rates produces similar results, suggesting that the epidemic transmission was deeply influenced by air travel. We then study the effect of alternative mitigation policies, in particular restricting long-range travel. We find that this policy is successful in decreasing the epidemic size and slowing down the spread, but less effective than the shelter-in-place policy. This policy can result in a pulled wave of infections. We express its velocity and characterize the shape of the traveling front as a function of the epidemiological parameters. Finally, we discuss a policy of selectively constraining travel based on an edge-betweenness criterion.

When plague hit Florence in August 1630, the Florentine authorities made a number of high-stakes decisions which proved highly effective¹. One of the reasons the Florence *Sanità* could organize this response was the ample time they had, forewarned as they were by the Milanese authorities in November 1629. Today’s public-health authorities work under much more compressed timescales, as evidenced by the SARS-CoV-2 pandemic. Long-distance travel radically changes the dynamics of spreading, which raises a number of questions about the spatial dynamics of transmission in modern times. Epidemic outbreaks in the last two decades have provided the scientific community with a wealth of material to study these questions, going beyond the classic Susceptible-Infected-Recovered (SIR) theory with perfect mixing^{2–6}. Several studies have shown how the total epidemic size can be affected by factors such as inhomogeneity in transmission rates^{7–14} or in the mode of transmission^{15,16}. Classically, motion of individuals was taken into account by introducing diffusion terms in the standard SIR equations, allowing the emergence of spatio-temporal patterns^{17–20}. Recently, in the context of the SARS-CoV-2 pandemic, such approaches have been especially valuable in order to study the effect of containment policies such as lockdown and quarantine^{21,22}. These models are, however, limited in that they do not, in principle, take long-distance air travel into account. Several works have, therefore, considered disease spread in a network, typically constructed from air-traffic data^{19,23,24}, where edges can connect locations separated by large geographical distances. This approach can lead to very accurate predictions at the country scale²⁵ but predictions at finer scales remain challenging. Another study considering human mobility emphasized how spatial variation in

¹Chan Zuckerberg Biohub, 499 Illinois Street, San Francisco, CA 94158, USA. ²Department of Physics, Kadanoff Center for Theoretical Physics, University of Chicago, Chicago, IL 60637, USA. ³Okinawa Institute of Science and Technology, Onna-son, Okinawa 904-0495, Japan. ⁴Chan Zuckerberg Initiative, Redwood City, CA 94063, USA. ⁵School of Systems Biology, George Mason University, Fairfax, VA 22030, USA. ⁶Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), 50018 Zaragoza, Spain. ✉email: guillaume.letreut@czbiohub.org, greg.huber@czbiohub.org

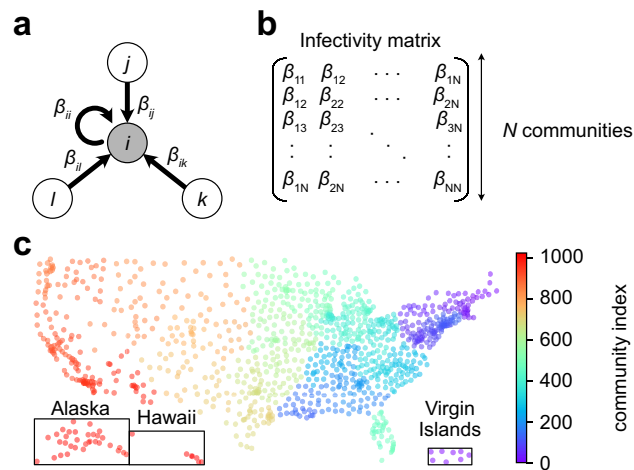


Figure 1. Model for the spreading of SARS-CoV-2 in a network of communities in the USA. (a) One community i interacting with three other communities j , k and l , with the transmission rates β_{ij} , β_{ik} and β_{il} respectively. (b) Infectivity matrix. (c) $N = 2^{10}$ communities in the USA. Each community aggregates a number of Census Block Groups (CBGs).

public-health infrastructure reflected on epidemiological parameters can affect the dynamics of spread to different countries²⁶. Data-based studies of epidemic spread and the impact of social distancing through the analysis of social-network structure have also been very informative^{16,27–29}.

A recent paper by Chang et al.³⁰ obtained a model for the spatio-temporal spread of a disease at a high spatial resolution by using extensive mobile tracking information to identify physical interactions between individuals. Chang et al. showed that the actual spread of the SARS-CoV-2 epidemic can be well explained from the mobility data of individuals. The model relied on the simulation of interactions among individuals on a bipartite graph where nodes, representing locations at a very fine spatial resolution, are divided in two sets: Census Block Groups (residential areas) and Points Of Interest (non-residential), each of them having its own transmission rate. The model was fitted to reproduce known reported cases of COVID-19 in 10 metropolitan areas, and could then be used to make short-term predictions about the spreading or study the effect of different mitigation strategies.

Here we take an approach similar to that of Chang et al., using mobility data to calibrate a model for disease spread. However, we investigate this propagation at the scale of a large country, the USA, rather than metropolitan areas. Specifically, we introduce a spatial SIR epidemiological model in which effective transmission rates between $N = 2^{10}$ communities are computed from mobility data of individuals belonging to these communities. We show that this model captures well the spread of the SARS-CoV-2 epidemic. Remarkably, we find that a simple model consisting of an interaction frequency dropping under the effect of lockdown, and of a single flux matrix encoding the travel of individuals, faithfully reproduces the reported cases of COVID-19 both globally and locally in each community. Strikingly, the SARS-CoV-2 epidemic spreads in a delocalized fashion, infecting distant communities very quickly. Moreover, an even simpler model with uniform transmission rates between the communities gives results very close to the model based on mobility data, emphasizing the prevalent role air travel played in the spread of the SARS-CoV-2 epidemic. We then study how interventions that change travel patterns can localize epidemics. In particular, we investigate the hypothetical effect of a policy preventing long-distance travel. In addition to “flattening” the curve, spreading through nearest-neighbor interactions creates traveling waves, which we characterize both analytically and numerically. These results allow us to discuss which interventions are more effective, limiting short-range contacts (a lockdown), or limiting long-range trips (a quarantine). We also propose an alternative mitigation strategy based on an edge-betweenness criterion.

Model

We consider a metapopulation model of N communities numbered $1, 2, \dots, N$. Denoting by S_a , I_a and R_a the numbers of susceptible, infected and recovered individuals in community a , the standard SIR equations read:

$$\frac{dS_a}{dt} = -S_a \sum_b \beta_{ab} \frac{I_b}{M_b}, \quad \frac{dR_a}{dt} = \gamma I_a, \quad \frac{dI_a}{dt} = -\frac{dS_a}{dt} - \frac{dR_a}{dt}, \quad (1)$$

where β_{ab} is the transmission rate from infected individuals in community b to susceptible individuals in community a and γ is the recovery rate, assumed to be the same for all communities. Diagonal elements of the infectivity matrix $[\beta_{ab}]$ describe intracommunity infections, while off-diagonal elements describe inter-community infections (Fig. 1a,b). We also introduce the local epidemic sizes $T_a = I_a + R_a$. The total population in each community M_a is constant through time,

$$S_a(t) + I_a(t) + R_a(t) = M_a. \quad (2)$$

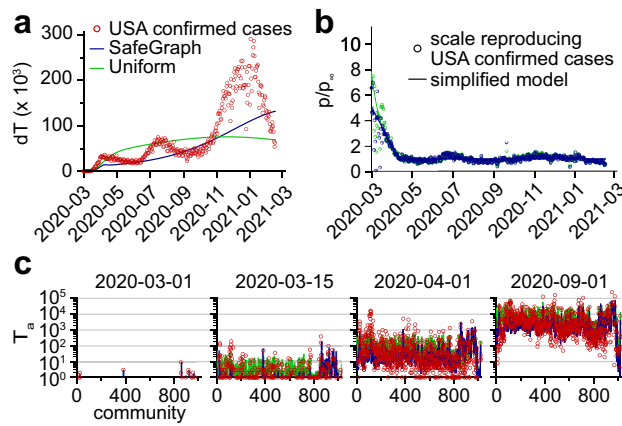


Figure 2. Model based on SafeGraph mobility data or a uniform infectivity matrix. **(a)** The model is fitted to COVID-19 confirmed cases in the USA. There is one fitting parameter per day. **(b)** Fitting parameters obtained. The shape suggests a simplified model with two limiting values before and after lockdown. **(c)** The simplified models obtained reproduce the spread of the SARS-CoV-2 epidemic in the community network. A direct comparison between the local epidemic sizes predicted by the models and the reported values can be found in Fig. 5. See also Movies S1 and S2.

The model in Eq. (1) has been extensively studied^{31–35}. We show in the Supplementary Information (SI) that the dynamics can be reduced to an ODE of just one N -vector variable, and the endemic equilibrium can be obtained by solving a transcendental equation involving the infectivity matrix $[\beta_{ab}]$.

In order to estimate the infectivity matrix, we used mobility data compiled by SafeGraph³⁶, tracking the location of about 20 million USA cell phones between March 2020 and February 2021. The locations consist of more than 200,000 Census Block Groups (CBGs). Each cell phone is assigned for physical residence the location where it spent the most time, and daily visits to other locations are recorded. For computational purposes, we coarse-grained the physical locations into $N = 2^{10}$ communities, which are shown in Fig. 1c. Let f_{ab} be the number of individuals from community a visiting community b per unit of time. We will assume that

$$f_{ab} \ll M_a, \quad f_{ab} \ll M_b, \quad \text{for all } a \text{ and } b. \tag{3}$$

The variation in susceptible individuals in community a due to new infections during the time interval Δt has the form:

$$S_a(t + \Delta t) - S_a(t) = - S_a(t) \times \text{Pr}(\text{meeting an infected individual}) \times \beta \Delta t, \tag{4}$$

where β is the disease-specific transmission rate when a susceptible individual has contact with an infected individual. There are three kinds of infection to consider: (i) the infected person and the infector belong to the same community, (ii) an infected person visits a neighboring community and infects a resident of this community, and (iii) a susceptible person visits a neighboring community and gets infected by one of its residents. We will neglect the rarer “tourist to tourist” infections, when an infected person visits a neighboring community and infects there a visitor from yet another community. The term $\text{Pr}(\text{meeting an infected individual})$ can therefore be evaluated as a function of the pseudo-flux matrix $[f_{ab}]$ for each of the three aforementioned cases (SI). After summation of the three contributions, we obtain:

$$\beta_{aa} = \beta p, \quad \beta_{ab} = \beta p \frac{f_{ab} + f_{ba}}{M_a}. \tag{5}$$

where p is the frequency with which an individual is having contact with another individual of its community, and is assumed to be the same for all communities. Equation (5) determines the infectivity matrix $[\beta_{ab}]$ from Eq. (1) up to a proportionality constant, namely βp .

Results

The model reproduces the spatial dynamics. In order to assess the validity of the infectivity matrix based on mobility data, Eq. (5), we confronted the model’s predictions against COVID-19 case numbers reported in the USA by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University³⁷. Specifically, we fitted the daily $\beta p(t)$ values so as to minimize the sum of squared errors between values predicted by the model and values reported by the CSSE (Material and methods). The fitted $\beta p(t)$ values show a steep decay during the month of March 2020, followed by a plateau lasting until February 2021 (Fig. 2b).

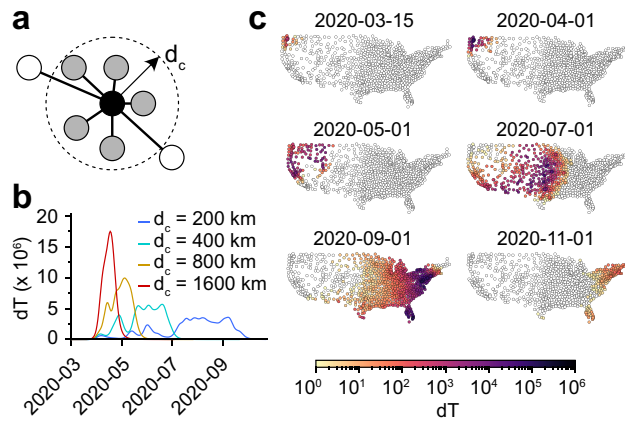


Figure 3. Limiting long-distance travels without local lockdown. (a) Only transmission rates for communities separated by a distance $d_{ij} < d_c$ are retained. (b) Daily new infections for increasing values of the cutoff distance. (c) Spatial visualization of the daily new infections using a cutoff $d_c = 200$ km. See also Movie S3.

A two-phase simplified model. The time variations of $\beta p(t)$ shown in Fig. 2b imply a drastic decrease in the interaction frequency among individuals across all communities. This result reflects the effect of the shelter-in-place policies that were implemented at the beginning of the SARS-CoV-2 pandemic in many USA states. We can use this observation to estimate the effect of these policies quantitatively: the interaction frequency among individuals, namely p , is about 5 times smaller in the plateau following shelter-in-place policies than it was at the onset of the SARS-CoV-2 pandemic. Following this observation, we defined a simplified model, in which the fit with reported cases was carried out while enforcing a softplus shape for $\beta p(t)$ (Fig. 2b and [Material and methods](#)). This simplified model reproduced the average progression of the epidemic (Fig. 2a), but it didn't capture the three oscillations visible in the number of new cases. We conclude that these oscillations in the number of new cases were mostly driven by a similar oscillatory pattern in the interaction frequency as seen in Fig. 2b. In Fig. 2c, we show for a few dates an overlay of the number of new COVID-19 cases in each community according to the official reports and as predicted by our simplified model. Despite non-negligible variance, we find that the model reproduces to a satisfying degree the real dynamics. This can also be seen through a direct comparison of the model predictions with the real reports of new cases, for all dates combined, as shown in Fig. 5a. This result suggests that the infectivity matrix constructed from mobility data (Eq. 5) is a reasonable approximation of the “true” infectivity matrix.

Turning down long-range interactions. The previous results suggest that the decrease in new COVID-19 cases was mostly driven by a country-wide reduction in the interaction frequency among individuals. Here we investigate the hypothetical effect of an alternative policy, namely a travel restriction while keeping unchanged the interaction frequency among individuals. Specifically, we modified the infectivity matrix so that communities separated by a physical distance larger than a prescribed cutoff do not interact: $\beta_{ij} = 0$ if $d_{ij} > d_c$ (Fig. 3a). We seeded the infection in a community belonging to the state of Washington and simulated the spread of the epidemics using a fixed interaction frequency ($\beta p(0)$ from the simplified model). As expected, we observed a reduction in the number of daily new cases dT (we define the local variables $dT_a(t) = T_a(t) - T_a(t-1)$) when the cutoff distance d_c decreased, illustrating the “curve-flattening” effect that was targeted by travel restriction policies (Fig. 3b). In this idealized scenario with a single seed for the infection, the epidemic propagates as a traveling wave from the west coast to the east coast (Fig. 3c). However, as can be seen by comparing Fig. 2a to Fig. 3b, travel restriction policies are not as efficient as lockdown policies to decrease the spread of an epidemic.

Properties of infectivity matrices. The daily infectivity matrices constructed from the SafeGraph mobility data can be viewed as elements of a random-matrix ensemble. Remarkably, matrix elements seem to be distributed according to the law $\beta_{ij} = \langle \beta_{ij} \rangle e^{\xi}$, where $\langle \beta_{ij} \rangle$ is the mean infectivity matrix and $\xi \equiv N(0, 1)$ is a centered reduced Gaussian variable (Fig. S1). The probability density of the eigenvalues is also shown in Fig. S1. A connection can be made with random matrix theory (RMT)^{38–40}, initially introduced by E. Wigner to model the spectra of the nuclei of heavy atoms, where the interactions between many nucleons are assumed to be drawn from a random ensemble. In RMT, random matrices are classified according to their symmetry, or according to their corresponding level statistics (namely, the probability density of the spacing between consecutive eigenvalues) that exhibit different degrees of level repulsion³⁸. In particular, the Wigner-Dyson (WD) statistics is typical of the Gaussian Orthogonal Ensemble (in which eigenvalues “interact”), while the Poisson statistics is typical of a matrix with independent eigenvalues. The level-spacing statistics of the infectivity matrices is shown in Fig. 4a. We found that they were time-independent. Yet interestingly, the level statistics interpolates between the WD statistics and the Poisson statistics⁴¹. As shown earlier, removing links between communities according to their geographical distance results in a slower spread and a smaller epidemic size. Concomitantly, the level spacing distribution converges toward a Poisson statistics (Fig. 4c). The crossover from WD (entropy $S = 0.7169$) to Poisson ($S = 1$) distribution as links between communities are successively removed suggests an isolation policy

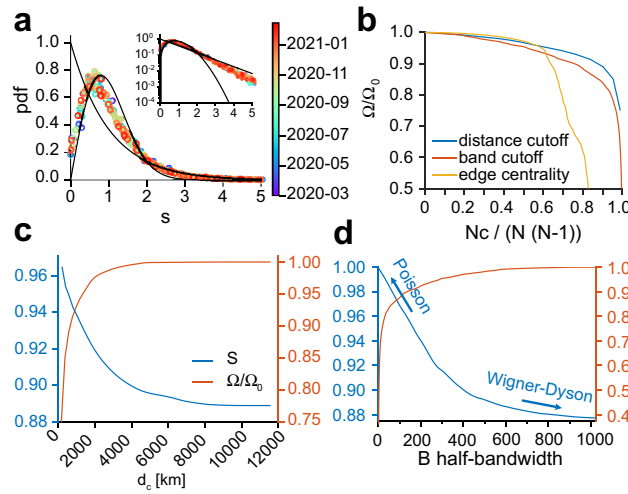


Figure 4. Random-matrix properties and mitigation strategies. **(a)** Level spacing distribution (unfolded^{38,45–47}) of infectivity matrices between 2020-03-01 and 2021-02-15 (pooled by 7-day windows). The solid lines denote the Wigner-Dyson and the Poisson statistics. The level distribution is shown in Fig. S1. **(b)** The epidemic size Ω/Ω_0 versus the relative number of cuts $N_c/N/(N-1)$ for the 3 mitigation methods: distance cutoff, bandwidth, and edge-betweenness centrality. Ω_0 is the epidemic size for $N_c = 0$. Level-spacing entropy S and epidemic size as a function of **(c)** the cutoff distance d_c in the range 170–12,000 km, and **(d)** the half-bandwidth $1 \leq B < 1023$.

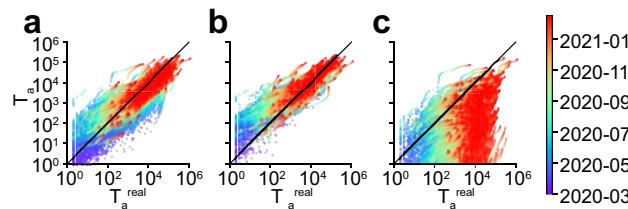


Figure 5. Comparison of the (fitted) model predictions with reported values using **(a)** the infectivity matrix derived from SafeGraph fluxes (same model as Fig. 2), **(b)** an infectivity matrix with uniform transmission rates, and **(c)** the infectivity matrix derived from SafeGraph fluxes truncated to geographical distances $d_c < 400$ km. Local epidemic sizes T_a are shown. One symbol is associated to one given community and one given day.

that can lead to an effective reduction of the epidemic size (SI). As alternative mitigation strategies, we choose to induce the transition toward a Poisson distribution by decimating links between communities according to their “nominal” distance or the “edge-betweenness” centrality^{42–44} (Fig. 4b). Figure 4d shows how the level spacing distribution converges toward a Poisson statistics when the nominal distance threshold is lowered. We find the edge-betweenness centrality to be more efficient in decreasing the epidemic size. This is because decimation according to edge-betweenness first targets links with the largest transmission rates. One could also consider a moderate policy: instead of eliminating links completely, one could impose constraints on the flux of individuals that are allowed to commute via central pre-determined links.

Classes of infectivity matrices. Although SafeGraph mobility data³⁶ provide a realistic picture of people movement between communities, one might ask what is the sensitivity of the results to the infectivity matrix derived from the mobility fluxes. We thus considered a model with uniform transmission rates among communities, namely $\beta_{ab} = \beta$, which suppresses spatial effects. In particular, this model leads to the natural variables v_a (see SI) to be uniform: $v_a(t) = v^{\text{UN}}(t)$. Surprisingly, carrying out the same fit to the reported cases (Fig. 1a,b) was only marginally inferior to the fit carried out with the SafeGraph mobility data (see Fig. 5b). By contrast, carrying out the same fit with the infectivity matrix derived from SafeGraph mobility data but with long-range interactions turned down resulted in a significantly different dynamics (see Fig. 5c). There are several implications of this result. (1) Although changes in the structure of the infectivity matrix can lead to drastically different dynamics (Figs. 3, 4 and 5c), it appears that the infectivity matrix derived from SafeGraph fluxes falls in the same “universality class” as the uniform model. This suggests that long-range movements (e.g. air traffic) played a prevalent role in the spread of SARS-CoV-2. (2) Our choice to reduce the complexity of the model to only one fitting parameter might not be adequate to discriminate between models falling into the same universality

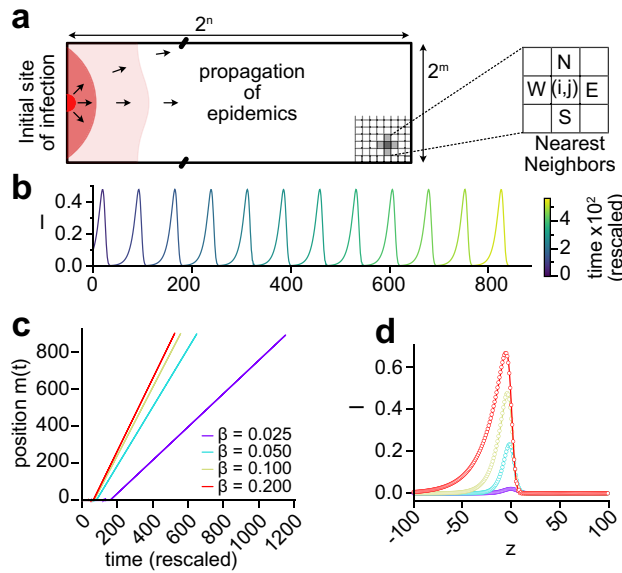


Figure 6. Existence of a wave with nearest-neighbors-only interactions. **(a)** Equation (6) is solved on a square lattice of $2^n \times 2^m$ sites. **(b)** Traveling front of infected individuals moving along the x direction (left to right). We took $\alpha = \beta = \gamma = 0.1$, so that the rescaled recovery rate is $\tilde{\gamma} = \gamma/(a\beta) = 0.2$. **(c)** The dynamics simulated for different values of β . The position $m(t)$ ($S(t, m(t)) = (1 + S_\infty)/2$) of the wave is asymptotically linear in time, hence a constant velocity $\tilde{v} = \sqrt{1 - \tilde{\gamma}}$. **(d)** The simulated wave profiles (symbols) are in agreement with the profiles predicted by Eq. (7) (solid lines). The corresponding S,I,R profiles are shown in Fig. S3. A simulation with $\tilde{\gamma} = 0.2$ is shown in Movie S4.

class. Instead of fitting only one scalar $p(t)$ at each time point, we also investigated the possibility of fitting the N^2 transmission rates $\beta_{ab}(t)$ minimizing the errors with reported cases (see **Material and methods** and Fig. S2). Although this latter approach is clearly prone to overfitting, it shows that there exists a parametrization of the model which reproduces very closely reported cases. We anticipate that the proper number of fitting parameters lies in between those two extreme scenarios.

Analysis of traveling waves. The results from the previous sections suggest the apparition of a wave-front when transmissions are short range. To investigate this phenomenon, we consider a simplified model of communities lying on a two-dimensional square lattice. Each community index is replaced by coordinates $(i, j) \in \llbracket 1, 2^n \rrbracket \times \llbracket 1, 2^m \rrbracket$. In particular, we consider that only individuals from neighboring communities interact together. We rewrite Eq. (1):

$$\frac{dS_{ij}}{dt} = -S_{ij}[\alpha I_{ij} + \beta(I_{i+1,j} + I_{i-1,j} + I_{i,j+1} + I_{i,j-1})], \quad \frac{dI_{ij}}{dt} = -\frac{dS_{ij}}{dt} - \gamma I_{ij}, \quad \frac{dR_{ij}}{dt} = \gamma I_{ij}, \quad (6)$$

where α (respectively β) is the intra-community (resp. inter-community) transmission rate. After rescaling the time and space variables, and defining the rescaled recovery rate $\tilde{\gamma} = \gamma/(a\beta)$ where $a = 4 + \alpha/\beta$, we look for wave solutions in the continuum limit by introducing the shape functions $S(x, y, t) = g(x - \tilde{v}t)$ and $I(x, y, t) = h(x - \tilde{v}t)$, where $\tilde{v} = v/\sqrt{a\beta}$ represents the velocity of the wave in rescaled time and space (see SI). The shape functions satisfy the ODE:

$$\begin{aligned} h'' &= -\frac{\tilde{v}}{g}h' + \left(\frac{\tilde{\gamma}}{g} - 1\right)h, \\ g' &= -f + \frac{\tilde{\gamma}}{\tilde{v}}h. \end{aligned} \quad (7)$$

We find that the velocity of the traveling wave is bounded from below (SI):

$$\tilde{v} \geq \tilde{v}_c = 2\sqrt{1 - \tilde{\gamma}}, \quad (8)$$

which is in agreement with previous reports^{18,48} and with results from marginal-stability analysis^{21,49,50}. Interestingly, by an independant argument, we also established (see SI) that $\tilde{v} \leq \tilde{v}_c$. Therefore the traveling wave must move at velocity $\tilde{v} = \tilde{v}_c$. This indicates that the SIR dynamics in Eq. (6) falls into the Fisher-Kolmogorov-Petrovsky-Piscunov universality class, resulting in pulled waves⁵¹⁻⁵⁴. Although we have taken the continuum limit of a nearest-neighbor model, this analysis is also valid for any finite-range infection matrix with the appropriate rescaling of variables.

We performed simulations of the dynamics given by Eq. (6) on a square lattice with a varying aspect ratio (Fig. 6a) and seeding the infection at one site on the west boundary. As expected, there is a front of new infections, moving from west to east as time progresses. A timelapse of a traveling front of infected individuals with $\alpha = \beta = \gamma = 0.1$ is shown in Fig. 6b. The wave position increases asymptotically linearly with time, but the velocity \tilde{v} of the wave varies with $\tilde{\gamma}$ (Fig. 6c). The profiles obtained are in agreement with the shape functions obtained by solving Eq. (7), as shown in Fig. 6d.

Discussion

Biological systems are inherently complex and it can be a challenge to characterize them by a small number of parameters. In the case of epidemics, the huge number of parameters (for example, intercommunity spread involves N^2 transmission rates for N communities) can obscure the salient mechanisms and make it difficult for policy makers to find efficient interventions. Reducing the dimension of a model, when possible, is therefore of great value. In this article, we have proposed a spatial SIR model with an infectivity matrix based on the local travel patterns.

Despite its simplicity (there is only one global parameter to be adjusted), we find that our model is able to capture the spatial spread of the SARS-CoV-2 epidemic, with a delocalized multicenter spreading caused by long-distance travelers bringing infection into far regions, then becoming secondary centers of infection. The time evolution of the interaction frequency reflects the shelter-in-place policies that were implemented by various USA states in the early stages of the SARS-CoV-2 pandemic. In fact, the rich diversity of human responses could be summarized by a simplified model for the interaction frequency, whose asymptotes represent the values “before lockdown” and “after lockdown”, with few assumptions about the infection process.

If a model with relatively few parameters describes the observations, one can assume that it also describes the situation when these parameters are changed by an intervention. Therefore, we suggest such a model could be used as the basis for efficient policies—or at least reliably to estimate the consequences of adopting policies. As an example, we have shown the hypothetical effect of an alternative to the shelter-in-place policy in the case of the SARS-CoV-2 pandemic. Specifically, we have investigated a travel restriction policy in which individuals can only move within an area of fixed radius centered around their residence. By contrast to the lockdown policy, we find that infections spread through a well-defined wave front, traveling with a certain velocity. This scenario might be preferable since it gives time to communities and public health infrastructures to prepare for the onset of the epidemics, while still “flattening the curve” of new infections (Fig. 3b). We have also provided both an analytical and numerical analysis elucidating the mechanism of formation of a wavefront. In particular, we give the velocity and the shape of the wavefront for an epidemic spreading through nearest neighbors interactions.

During the course of our research, an epidemiological study bearing similarities with our approach was published⁵⁵. In that study, the authors developed a county-resolved metapopulation model describing the spread of SARS-CoV-2 in the USA, informed with mobility data from SafeGraph. Instead of calibrating the model by fitting the country-wide number of reported cases as in the present study, the authors fitted their model to reproduce reported cases of COVID-19 on a per-county basis. Furthermore, the model required fitting many time-dependent parameters on a per-county basis, including per-county transmission rates, making the fitting procedure very high dimensional. Finally, the dynamics of the disease spread introduced differs from Eq. (1) since S , I and R compartments were introduced for each commute channel $i \rightarrow j$ rather than for each community. Altogether, the complexity of that model makes it less amenable to analytical study.

In conclusion, we used a simple model of intercommunity spread of an infectious disease to show the transition between different regimes of epidemic progression. Because of the complexity of the infection process (e.g., variations in individuals’ responses, mutations, etc.) and of human behavior, we are still far from a global forecasting system able to predict the spread of different infectious agents throughout the world. As with weather forecasting, observables must be measured in real time in order to inform complex models to yield short-term forecasts. One salient feature in our approach is showing how such widespread measurements (namely, the mobility data) can be integrated in a model describing the spread of an infectious agent, and showing what types of predictions can be obtained. It is a step toward more predictive epidemiology models, grounded in measurable quantities.

Material and methods

Mobility data. The mobility data was obtained from SafeGraph, a company that aggregates anonymized location data from numerous applications in order to provide insights about physical places, via the SafeGraph Community. To enhance privacy, SafeGraph excludes census block group (CBG) information if fewer than two devices visited an establishment in a month from a given census block group. In this manuscript, we use data extracted from the “Social Distancing Metrics” dataset³⁶, with dates between February 2020 and February 2021. SafeGraph has stopped sharing mobility data under this format, however the same data can be accessed under a different format through the Neighborhood Patterns dataset⁵⁶. However the matrices of fluxes between our coarse-grained communities can be found in the ‘data’ folder in our github repository. We obtained the coarse-grained communities by running a K -means clustering algorithm to group the 220,333 CBGs into 2^{10} communities. We used the implementation from Scikit Learn. We then ran a hierarchical clustering algorithm to re-index communities so that communities close in space had close indices, as shown in Fig. 1. We used the linkage function from SciPy. We then computed for every day the “flux matrix” where each entry f_{ab} represents the number of cell phone whose residence CBG belongs to community a which visited a CBG belonging to community b . The average flux matrix was constructed by averaging all the daily flux matrices. We used population counts for CBGs in agreement with the United States Census Bureau’s and available in the SafeGraph Open Census Data⁵⁷ (file “cbg_b01.csv”, column “B01001e1”). We checked that the population counts from the United States Census

Bureau were approximately proportional to the residential mobile-phone counts, therefore validating mobile tracking as a proxy for actual population.

Reports on SARS-CoV-2 infections. In order to fit our model, we used USA cases of COVID-19 reported by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University³⁷ which can be accessed at the GitHub repository github.com/CSSEGISandData/COVID-19. New cases of COVID-19 reported by the CSSE were mapped to the closest mobility-data-derived community based on their latitude and longitude. We therefore obtained the time evolution of SARS-CoV-2 infections through the mobility-data derived communities. Integer absolute new cases were converted into relative population fractions using the community population counts obtained from the United States Census Bureau.

Model fit. *Daily scales.* Here we consider the SIR dynamics (see Supplementary Information):

$$\begin{aligned}\frac{ds_a}{dt}(t) &= -p(t)s_a(t) \sum_b \beta_{ab}j_b(t), \\ \frac{dj_a}{dt}(t) &= p(t)s_a(t) \sum_b \beta_{ab}j_b(t) - \gamma j_a(t).\end{aligned}\quad (9)$$

We also define the chi-square at each time t :

$$\chi^2(t) = \sum_{a=1}^N M_a^2 (s_a(t+1) - \hat{s}_a(t+1))^2, \quad (10)$$

where M_a is the population at location a and $(1 - \hat{s}_a)M_a$ is the number of reported cases at location a . Assuming that $p(t-1), p(t-2), \dots, p(0)$ have been previously evaluated, the dynamics (s_a, j_a) is determined up to time t , and we have:

$$\begin{aligned}s_a(t+1) &= s_a(t) - p(t) \int_t^{t+1} du s_a(u) \sum_b \beta_{ab}j_b(u), \\ j_a(t+1) &= j_a(t) + p(t) \int_t^{t+1} du \left(s_a(u) \sum_b \beta_{ab}j_b(u) - \gamma j_a(u) \right).\end{aligned}\quad (11)$$

To obtain the scale $p(t)$, we solve for:

$$p^{\text{fit}}(t) = \text{argmin}(\chi(t)^2). \quad (12)$$

Simplified model. We look for a simplified model in which the scales have a functional form close to a ramp function:

$$p_\theta(t) = \theta_3 \ln(1 + e^{-\theta_1(t-\theta_2)}) + \theta_4. \quad (13)$$

We first set θ by fitting this function to the daily scales:

$$\theta^* = \text{argmin} \left(\sum_t (p_\theta(t) - p^{\text{fit}}(t))^2 \right) \quad (14)$$

Then we adjust the slope of the ramp in order to minimize the error with reported cases. In particular, we set $\theta = (\theta_1^*, \theta_2^*, \psi, \theta_4^*)$, and we solve:

$$\psi^* = \text{argmin} \left(\sum_t \chi(t)^2 \right). \quad (15)$$

We thus define $\theta^{\text{simplified}} = (\theta_1^*, \theta_2^*, \psi^*, \theta_4^*)$, and the model for simplified scales is:

$$p^{\text{simplified}}(t) = p_{\theta^{\text{simplified}}}(t). \quad (16)$$

Daily infectivity matrices. We also considered another approach, consisting of fitting the N^2 transmission rates $\beta_{ab}(t)$ at every time t . In particular, we solve:

$$\left\{ \beta_{ab}^{\text{opt}}(t) \right\} = \text{argmin}(\chi(t)^2), \quad (17)$$

subject to the constraints $\beta_{ab} \geq 0$. We show the results of this fit in Fig. S2, however this approach is prone to overfitting since there are N^2 fitting parameters and only N data point at each time t .

Simulations with nearest-neighbors-only interactions. The curves shown in Fig. 6b,c and the symbols shown in Fig. 6d were obtained by integrating Eq. (6). We used the function `solve_ivp` from SciPy with the “DOP853” integration method. At $t = 0$, we considered $S = 1$ everywhere except at the sites of coordinates $(0, 2^{m-1} - 1)$ and $(0, 2^{m-1})$ (see Fig. 6a), where we set $S = 0$ and $I = 1$. The Laplacian was computed using the 9-point stencil $\Delta_{\text{discrete}} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{pmatrix} / 4$. We considered periodic boundary conditions along the y direction and Dirichlet boundary conditions along the x direction.

Data availability

Data and scripts used in this study are available at the GitHub repository github.com/czbiohub/epidemiology_flux_model. Except the raw mobility data, which belongs to SafeGraph.

Received: 23 January 2022; Accepted: 6 September 2022

Published online: 24 September 2022

References

- Henderson, J. *Florence Under Siege: Surviving Plague in an Early Modern City* (Yale University Press, 2019).
- Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Charact.* **115**, 700–721 (1927).
- Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653. <https://doi.org/10.1137/S0036144500371907> (2020).
- Harko, T., Lobo, F. S. & Mak, M. Exact analytical solutions of the susceptible-infected-recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Appl. Math. Comput.* **236**, 184–194 (2014).
- 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. *Lancet* **395**, 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9) (2020).
- Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: A mathematical modelling study. *Lancet Infect. Dis.* **20**, 535–558. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4) (2020).
- Allard, A., Moore, C., Scarpino, S. V., Althouse, B. M. & Hébert-Dufresne, L. The role of directionality, heterogeneity and correlations in epidemic risk and spread (2020). arXiv preprint [arXiv:2005.11283](https://arxiv.org/abs/2005.11283).
- Aleta, A. *et al.* Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. *medRxiv* <https://doi.org/10.1101/2020.12.15.20248273> (2020).
- Hébert-Dufresne, L., Althouse, B. M., Scarpino, S. V. & Allard, A. Beyond R_0 : Heterogeneity in secondary infections and probabilistic epidemic forecasting. *J. R. Soc. Interface* **17**, 20200393 (2020).
- Britton, T., Ball, F. & Trapman, P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**, 846–849 (2020).
- Neipel, J., Bauermann, J., Bo, S., Harmon, T. & Jülicher, F. Power-law population heterogeneity governs epidemic waves. *PLoS ONE* **15**, e0239678 (2020).
- Sun, K. *et al.* Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* <https://doi.org/10.1126/science.abe2424> (2021).
- Kawagoe, K. *et al.* Epidemic dynamics in inhomogeneous populations and the role of superspreaders. *Phys. Rev. Res.* **3**, 033283. <https://doi.org/10.1103/PhysRevResearch.3.033283> (2021).
- Pozderac, C. & Skinner, B. Superspreading of SARS-CoV-2 in the USA. *PLoS ONE* **16**, 1–10. <https://doi.org/10.1371/journal.pone.0248808> (2021).
- Huber, G. *et al.* A minimal model for household effects in epidemics. *Phys. Biol.* **17**, 065010 (2020).
- 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. <https://doi.org/10.1038/s41562-020-0931-9> (2020).
- Murray, J. *Mathematical Biology II. Spatial Models and Biological Applications* (Springer, 2003).
- Postnikov, E. B. & Sokolov, I. M. Continuum description of a contact infection spread in a SIR model. *Math. Biosci.* **208**, 205–215 (2007).
- Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
- Chu, A., Huber, G., McGeever, A., Veytsman, B. & Yllanes, D. A random-walk-based epidemiological model. *Sci. Rep.* **11**, 19308. <https://doi.org/10.1038/s41598-021-98211-5> (2021).
- Te Vrugt, M., Bickmann, J. & Wittkowski, R. Effects of social distancing and isolation on epidemic spreading modeled via dynamical density functional theory. *Nat. Commun.* **11**, 1–11 (2020).
- Tsori, Y. & Granek, R. Epidemiological model for the inhomogeneous spatial spreading of COVID-19 and other diseases. *PLoS ONE* **16**, e0246056 (2021).
- Tizzoni, M. *et al.* Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Med.* **10**, 1–31 (2012).
- Linka, K., Peirlinck, M., Sahli Costabal, F. & Kuhl, E. Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. *Comput. Methods Biomech. Biomed. Eng.* **23**, 710–717 (2020).
- Ivorra, B., Ngom, D. & Ramos, A. M. Be-CoDiS: A mathematical model to predict the risk of human diseases spread between countries. validation and application to the 2014–15 Ebola virus disease epidemic. arXiv preprint [arXiv:1410.6153](https://arxiv.org/abs/1410.6153) (2014).
- Hsu, S. & Zee, A. Global spread of infectious diseases. *J. Biol. Syst.* **12**, 289–300 (2004).
- Nande, A., Adlam, B., Sheen, J., Levy, M. Z. & Hill, A. L. Dynamics of COVID-19 under social distancing measures are driven by transmission network structure. *PLoS Comput. Biol.* **17**, e1008684 <https://doi.org/10.1371/journal.pcbi.1008684> (2021).
- Ventura, P. C., Aleta, A., Rodrigues, F. A. & Moreno, Y. Modeling the effects of social distancing on the large-scale spreading of diseases. [arXiv:2105.09697](https://arxiv.org/abs/2105.09697) (2021).
- Mayberry, J., Nattestad, M. & Tuttle, A. The structure of an outbreak on a college campus. *Math. Mag.* **94**, 83–98 (2021).
- Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
- Hethcote, H. W. An immunization model for a heterogeneous population. *Theor. Popul. Biol.* **14**, 338–349 (1978).

32. Post, W., DeAngelis, D. & Travis, C. Endemic disease in environments with spatially heterogeneous host populations. *Math. Biosci.* **63**, 289–302 (1983).
33. May, R. M. & Anderson, R. M. Spatial heterogeneity and the design of immunization programs. *Math. Biosci.* **72**, 83–111 (1984).
34. Hethcote, H. W. & Van Ark, J. W. Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs. *Math. Biosci.* **84**, 85–118 (1987).
35. Lloyd, A. L. & May, R. M. Spatial heterogeneity in epidemic models. *J. Theor. Biol.* **179**, 1–11 (1996).
36. SafeGraph. SafeGraph Social Distancing Metrics. <https://docs.safegraph.com/docs/social-distancing-metrics> (2021).
37. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
38. Mehta, M. *Random Matrices*, 3rd ed, Sec. 1.5 (Elsevier, 2004).
39. Akemann, G., Baik, J. & Di Francesco, P. *The Oxford Handbook of Random Matrix Theory* (Oxford University Press, 2011).
40. Livan, G., Novaes, M. & Vivo, P. *Introduction to Random Matrices: Theory and Practice* (Springer, 2020).
41. Rosenzweig, N. & Porter, C. Repulsion of energy levels in complex atomic spectra. *Phys. Rev.* **120**, 1698–1714 (1960).
42. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* **99**, 7821–7826 (2002).
43. Luo, F., Zhong, J., Yang, Y., Scheuermann, R. H. & Zhou, J. Application of random matrix theory to biological networks. *Phys. Lett. A* **357**, 420–423 (2006).
44. DeMeo, P., Ferrara, E., Fiumara, G. & Ricciardello, A. A novel measure of edge centrality in social networks. *J. Knowl.-Based Syst.* **30**, 136–150 (2012).
45. Brody, T. A. *et al.* Random-matrix physics: Spectrum and strength fluctuations. *Rev. Mod. Phys.* **53**(385–479), 391 (1981).
46. Guhr, T., Müller-Groeling, A. & Weidenmüller, H. A. Random matrix theories in quantum physics: Common concepts. *Phys. Rep.* **299**(189–425), 228 (1998).
47. Hakke, F. *Quantum Signatures of Chaos* 3rd edn, Sec 4.5 (Springer, 2010).
48. Naether, U., Postnikov, E. & Sokolov, I. Infection fronts in contact disease spread. *Eur. Phys. J. B* **65**, 353–359 (2008).
49. Dee, G. & Langer, J. Propagating pattern selection. *Phys. Rev. Lett.* **50**, 383 (1983).
50. Ben-Jacob, E., Brand, H., Dee, G., Kramer, L. & Langer, J. Pattern propagation in nonlinear dissipative systems. *Phys. D: Nonlinear Phenomena* **14**, 348–364 (1985).
51. Bramson, M. *Convergence of Solutions of the Kolmogorov Equation to Travelling Waves*, Vol. 285 (American Mathematical Soc., 1983).
52. Paquette, G., Chen, L.-Y., Goldenfeld, N. & Oono, Y. Structural stability and renormalization group for propagating fronts. *Phys. Rev. Lett.* **72**, 76 (1994).
53. Ebert, U. & van Saarloos, W. Front propagation into unstable states: Universal algebraic convergence towards uniformly translating pulled fronts. *Phys. D Nonlinear Phenomena* **146**, 1–99 (2000).
54. Birzu, G., Hallatschek, O. & Korolev, K. S. Fluctuations uncover a distinct class of traveling waves. *Proc. Natl. Acad. Sci.* **115**, E3645–E3654 (2018).
55. Sen, P., Yamana, T. K., Kandula, S., Galanti, M. & Shaman, J. Burden and characteristics of COVID-19 in the united states during 2020. *Nature* **598**, 338–341 (2021).
56. SafeGraph. *SafeGraph Neighborhood Patterns*. <https://docs.safegraph.com/docs/neighborhood-patterns> (2021).
57. SafeGraph. *SafeGraph Open Census Data*. <https://docs.safegraph.com/docs/open-census-data> (2021).

Acknowledgements

The authors acknowledge the generous support of the Chan Zuckerberg Initiative and Chan Zuckerberg Biohub. The authors also thank Mark Rychnovsky for useful conversations on random matrix theory in the early stages of this project, and Naomi Rankin for her help with a preliminary version of the lattice simulations. D.Y. acknowledges support by MINECO (Spain) through Grant No. PGC2018-094684-B-C21, partially funded by the European Regional Development Fund (FEDER). We also are grateful to the anonymous reviewers of *Scientific reports* who asked deep questions, that helped us to better understand our results.

Author contributions

The concept for this research grew out of discussions among all authors; G.L.T. performed research; G.H., A.M., M.K. analyzed data; K.K., J.M., R.P., performed analytical/numerical computations; G.L.T., B.V. and D.Y. wrote the paper, with contributions from all authors.

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/s41598-022-19931-w>.

Correspondence and requests for materials should be addressed to G.L.T. or G.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022