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

Guillaume Le Treut[a,1], Greg Huber[a,1], Mason Kamb[a], Kyle Kawagoe[b], Aaron McGeever[a], Jonathan Miller[c], Reuven Pnini[c], Boris Veytsman[d,e], and David Yllanes[a,f]

[a]*Chan Zuckerberg Biohub, 499 Illinois Street, San Francisco, CA 94158, USA*
[b]*Department of Physics, Kadanoff Center for Theoretical Physics, University of Chicago, Chicago, Illinois 60637, USA*
[c]*Okinawa Institute of Science and Technology, Onna-son, Okinawa 904-0495, Japan*
[d]*Chan Zuckerberg Initiative, Redwood City, CA 94063, USA*
[e]*School of Systems Biology, George Mason University, Fairfax, VA 22030, USA*
[f]*Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), 50018 Zaragoza, Spain*
[1]To whom correspondence should be addressed. E-mails: guillaume.letreut@czbiohub.org, greg.huber@czbiohub.org.

This manuscript was compiled on December 22, 2021

Abstract

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 have applied this model to the 2020 dynamics of the SARS-CoV-2 pandemic. We find that the numbers of susceptible and infected individuals predicted by the model agree with the reported cases in each community by fitting just one global parameter representing the frequency of interaction between individuals. The effect of "shelter-in-place" policies introduced across the USA at the onset of the pandemic is clearly seen in our results. We then consider the effect that alternative policies would have had, namely restricting long-range travel. We find that this policy is successful in decreasing the epidemic size and slowing down the spread, at the expense of a substantial restriction on mobility as a function of distance. When long-distance mobility is suppressed, this policy results in a traveling wave of infections, which we characterize analytically. In particular, we show the dependence of the wave velocity and profile on the transmission parameters. Finally, we discuss a policy of selectively constraining travel based on an edge-betweenness criterion.

Keywords: epidemiology | SIR model | COVID-19 | mobility data

When plague hit Florence in August 1630, the Florentine authorities made a number of high-stakes decisions which proved highly effective [1], ranging from the establishment of isolation hospitals and cemeteries, to the introduction of virtual religious rites, with priests performing liturgies at street corners and their fold listening in their houses. 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 COVID-19 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 scale [25] but predictions at finer scales remain challenging. Another study considering human mobility emphasized how spatial variation in public-health infrastructure reflected on epidemiological parameters can affect the dynamics of spread to different countries [26]. 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 COVID-19 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 very 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 COVID-19 epidemic spreads in a delocalized fashion, infecting distant communities very quickly. We therefore study how interventions that change travel patterns can localize epidemics.

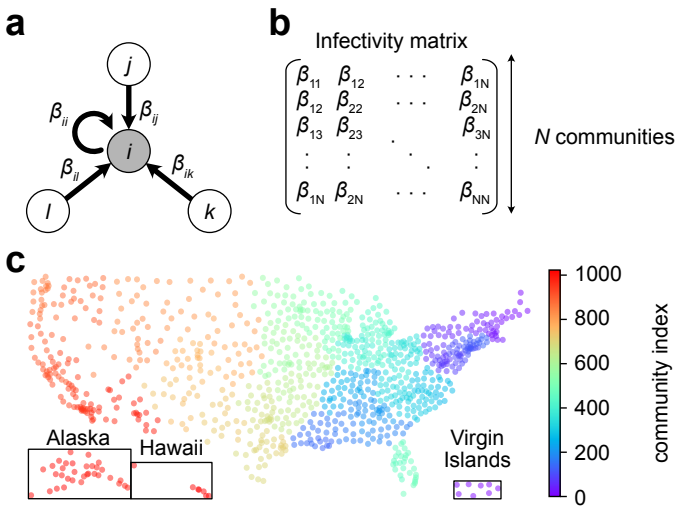

Figure 1 – Model for the spreading of COVID-19 in a network of communities in the USA. **(a)**. One community i interacting with three other communities j, k and l, with the infectivity coefficients β_{ij}, β_{ik} and β_{il} respectively. **(b)** Infectivity matrix. **(c)** $N = 2^{10}$ communities in the USA. Each community aggregates a number of Census Block Groups (CBGs).

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 travelling 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 then propose an alternative mitigation strategy based on an edge-betweenness criterion.

Model

We consider a metapopulation model of N communities numbered 1, 2,. . . , 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:

$$\frac{\mathrm{d}S_a}{\mathrm{d}t} = -S_a \sum_b \beta_{ab} \frac{I_b}{M_b}, \qquad \frac{\mathrm{d}R_a}{\mathrm{d}t} = \gamma I_a, \qquad \frac{\mathrm{d}I_a}{\mathrm{d}t} = -\frac{\mathrm{d}S_a}{\mathrm{d}t} - \frac{\mathrm{d}R_a}{\mathrm{d}t}, \tag{1}$$

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 (Figure 1**a-b**). We also introduce local epidemic sizes $T_a = I_a + R_a$. The total population in each community M_a is constant through time,

$$S_a(t) + I_a(t) + R_a(t) = M_a. \tag{2}$$

The model in equation (1), known as the "multi-patch" model [31], has been extensively studied [32–35]. 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 $[\beta_{ab}]$, as shown in the Supplementary Information (SI).

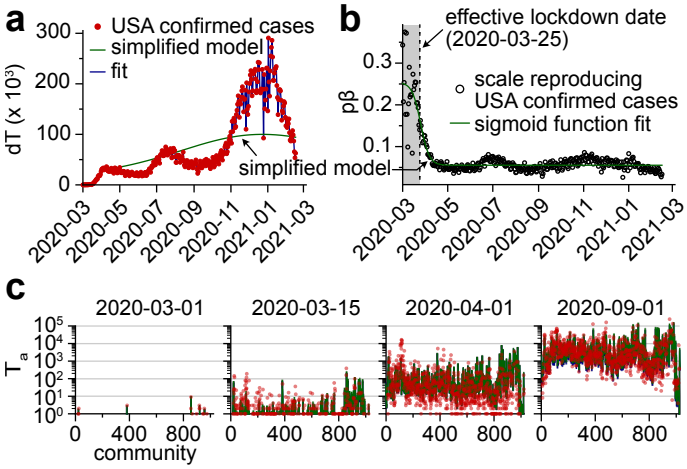

Figure 2 – Model based on mobility data. **(a)** The model is fitted to COVID-19 confirmed cases in the USA. There is one fitting parameter per day. **(b)** Fitting parameters obtained. The shape suggests a sigmoid function, defining a simplified model with two limiting values before and after lockdown. **(c)** The model obtained reproduces the spread of the COVID-19 epidemic in the community network. A direct comparison between the local epidemic sizes predicted by the model and the reported values can be found in the SI.

In order to estimate the infectivity matrix, we used mobility data compiled by SafeGraph [36], 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 Figure 1**c**. 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:

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

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 Pr (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:

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

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

Results

The model reproduces the spatial dynamics

In order to assess the validity of the infectivity matrix based on mobility data, equation (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 University [37]. Specifically, we fitted the daily $\beta p(t)$ values so that the total epidemic size predicted by the model, $\Omega = \sum_a T_a$, reproduced the values reported by the CSSE (Figure 2**a**). The fitted $\beta p(t)$ values show a steep decay during the month of March 2020, followed by a plateau lasting until February 2021 (Figure 2**b**). Although we fitted the model to the real SARS-CoV-2 confirmed cases using the global quantity Ω, local epidemic sizes T_a predicted by the model agreed very well to the real values. In Figure 2**c**, we see that the number of new SARS-CoV-2 cases in each community predicted by the model follows closely the empirical values. This result suggests that the infectivity matrix constructed from mobility data (equation (5)) is a good approximation of the "true" infectivity matrix.

A two-phase simplified model

The time variations of $\beta p(t)$ shown in Figure 2**b** 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. In order to define a simplified model, we fitted the time variations of $\beta p(t)$ to a sigmoid function. This simplified model reproduced the average progression of the epidemic (Figure 2**a**), 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 Figure 2**b**.

Turning down long-range interactions

The previous results suggest that the decrease in new SARS-CoV-2 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 $d_{ij} > d_c$ (Figure 3**a**). 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. As expected, we observed a reduction in the number of daily cases dT when the cutoff distance d_c decreased, illustrating the "curve-flattening" effect that was targeted by travel restriction policies (Figure 3**b**). 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 (Figure 3**c**). However, as can be seen by comparing Figure 2**b** to Figure 3**b**, travel restriction policies are not as efficient as lockdown policies to decrease the spread of an epidemic.

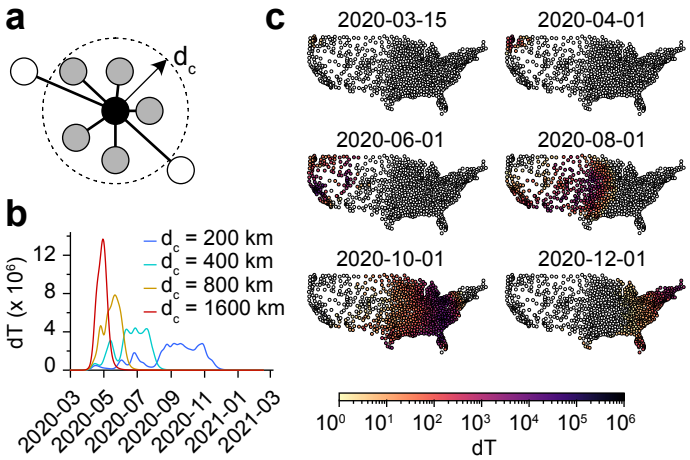

Figure 3 – Limiting long-distance travels without local lockdown. **(a)** Only infectivity coefficients for communities separated by a distance $d_{ij} < d_c$ are retained. **(b)** Daily new infections for increasing values of the cutoff distance. **(c)** Spatial visualization of the daily new infections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

Agreement with 1D analysis of a wave

The results from the previous section suggest the apparition of a wavefront 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 a is replaced by coordinates $(i, j) \in [\![1, 2^n]\!] \times [\![1, 2^m]\!]$. In particular, we consider that only individuals from neighboring communities interact together. We rewrite equation (1):

$$\begin{aligned} \frac{\mathrm{d}S_{i,j}}{\mathrm{d}t} &= -S_{i,j}\left[\alpha I_{i,j} + \beta(I_{i+1,j} + I_{i-1,j} + I_{i,j+1} + I_{i,j-1})\right], \\ \frac{\mathrm{d}I_{i,j}}{\mathrm{d}t} &= -\frac{\mathrm{d}S_{i,j}}{\mathrm{d}t} - \gamma I_{i,j}, \quad \frac{\mathrm{d}R_{i,j}}{\mathrm{d}t} = \gamma I_{i,j}, \end{aligned} \tag{6}$$

where α (respectively β) is the intra-community (resp. inter-community) infectivity rate. We look for wave solutions in the continuum limit and introduce the Ansatz: $S(x, y, t) = F(x - x_0 - vt)$ and $R(x, y, t) = G(x - x_0 - vt)$, where v represents the velocity of the wave. We furthermore introduce the functions $\phi(z)$ and $\psi(z)$ such that $F(z) = (1 + \phi(kz))/2$ and $G(z) = (1 - \psi(kz))/2$, with $k = (2v)/\beta$. Working in units where the lattice spacing is equal to 1, we obtain that ϕ and ψ satisfy the following ODE (SI):

$$\begin{aligned} \phi'' &= -\left(\frac{1}{1+\phi} + CD\right)\phi' + C(1 + CD^2)(\psi - \phi), \\ \psi' &= CD(\psi - \phi), \end{aligned} \tag{7}$$

where $C = a\beta^2/(4v^2)$, $D = 2\gamma/(a\beta)$, and $a = 4 + \alpha/\beta$. We show solutions to equation (7) for different values of C in Figure 4**b**, with $\alpha = \beta = \gamma = 0.1$. Interestingly, we find that C is bounded from above, which imposes that the velocity of the travelling wave is bounded from below (SI):

$$v \geq v_{\mathrm{vmin}} = 2\beta\sqrt{a - \frac{\gamma}{\beta}}, \tag{8}$$

which is in agreement with previous estimates [18, 38] and with results from marginal-stability analysis [21, 39, 40]. 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.

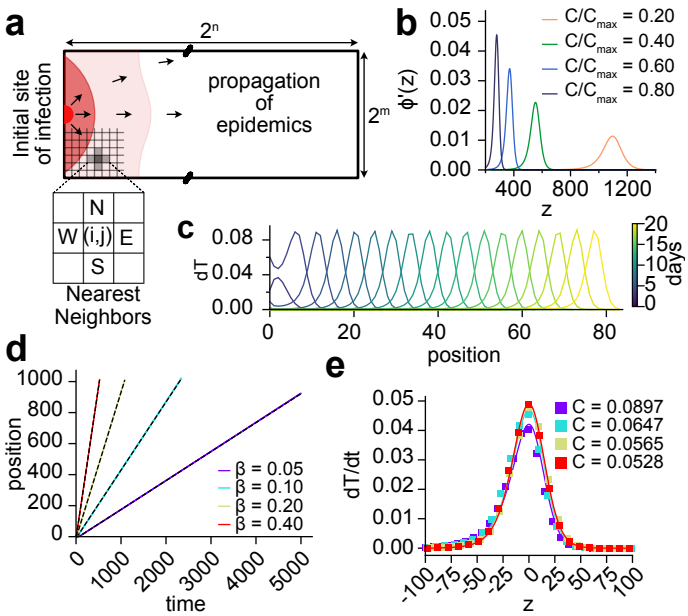

Figure 4 – Existence of a wave with nearest-neighbors-only interactions. **(a)** Equation (6) is solved on a square lattice of $2^n \times 2^m$ sites. **(b)** $\phi'(z)$ is represented, where ϕ is the solution to equation (7). We took $\alpha = \beta = \gamma = 0.1$. **(c)** Wave profiles moving from west to east on the lattice shown for different days (color scale). We used $\beta = 0.1$. **(d-e)** The dynamics with different values of β. The position of the peak of the wave is asymptotically linear in time, hence a constant velocity v of the wave **(d)**. The wave profiles obtained are in agreement with the profiles predicted by equation (7) **(e)**.

We performed simulations of the dynamics given by equation (6) on a square lattice with a varying aspect ratio (Figure 4**a**). In Figure 4**c**, we show the traveling wave obtained for different values of C while keeping $\beta = 0.1$. There is a front of new infections, moving from west to east as time progresses. The velocity v of the wave is constant, resulting in the wave position to increase linearly with time. The value of the velocity v changes with β as shown in Figure 4**d**. The profiles obtained are in agreement with the theory, since they agree with the prediction $\mathrm{d}T/\mathrm{d}t(x, y, t) = vk\phi'(k(x - x_0 - vt))/2$ (Figure 4**e**).

Properties of infectivity matrices

There is an interesting connection between our model and the theory of random matrices [41, 42], 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. Similarly, the infectivity matrices constructed daily 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:

$$\beta_{ij} = \langle \beta_{ij} \rangle e^{\xi}, \tag{9}$$

where $\langle \beta_{ij} \rangle$ is the mean infectivity matrix and $\xi \equiv N(0, 1)$ is a centered reduced Gaussian variable (SI). Despite absolute changes in the flux matrices reflecting the lockdown policies enforced at the onset of the SARS-CoV-2 pandemic, the eigenvalue distribution (SI) and the level-spacing statistics (Figure 5**a**) of the infectivity matrices remain the same through time. Interestingly, the level statistics interpolates between the Wigner-Dyson (WD) statistics [41], typical of the Gaussian Orthogonal Ensemble, and the Poisson statistics of independent levels. The crossover from WD (entropy $S = 0.7169$) to Poisson ($S = 1$) distribution as links between communities are successively removed suggests an isolation policy that can lead to an effective reduction of the epidemic size (SI). We have already shown that removing links between

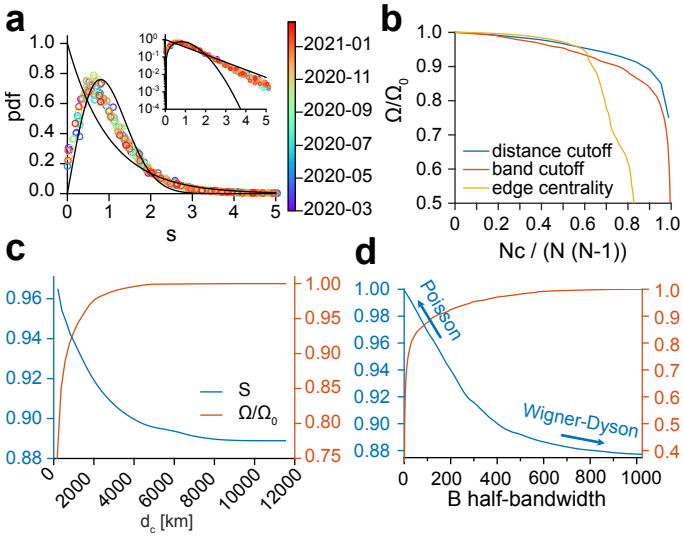

Figure 5 – Random-matrix properties and mitigation strategies. **(a)** Level spacing distribution (unfolded [41, 46–48]) of infectivity matrices between 2020-03-01 and 2021-02-15 (pooled by 7-day windows). The solid lines denote the Wigner-Dyson and the Poisson statistics. **(b)** The epidemic size Ω/Ω_0 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 170-12000 km, and **(d)** the half-bandwidth $1 \leq B < 1023$

communities according to their geographical distance results in a slower spread and a smaller epidemic size. Alternatively, we choose to make the transition toward a Poisson distribution by decimating links between communities according to their "nominal" distance or the "edge-betweenness" centrality [43–45] (Figure 5**b**). We find the edge-betweenness centrality to be more efficient in decreasing the epidemic size. Intuitively, this is because decimation according to edge-betweenness first targets links with the largest transmission rates. In all of these cases, one can also consider a moderate policy: instead of eliminating links completely, one can impose constraints on the flux of people that are allowed to commute via central pre-determined links.

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 infection rates for N communities) can obscure the salient mechanisms and make it difficult for policymakers 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 constructed from mobility data of individuals in the USA. We find that this model accurately reproduces local epidemic spread and requires only one global parameter to be adjusted: the interaction frequency between individuals. 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 minimal model using a sigmoid function for the interaction frequency, whose asymptotes represent the values "before lockdown" and "after lockdown", with few assumptions about the infection process. The individual variances between people and communities turn out to self-average, giving a clear picture of the spread.

Despite its simplicity, 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. 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 infection spread through a well-defined wave front, travelling 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 (Figure 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 published [49]. In that study, the authors developed a county-resolved metapopulation model describing the spread of COVID-19 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 equation (1) since S, I and R compartments are introduced for each commute channel $i \to j$ 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

Data availability

Data and scripts used in this study are available at the GitHub repository github.com/czbiohub/epidemiology_flux_model.

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" dataset [36], with dates between February 2020 and February 2021. This dataset tracks about 20 million cell phones in the USA with a daily resolution. Every day, the residence for each cell phone is set to the CBG where most time was spent. In addition, whenever a cell phone location is detected in a CBG different from its residence CBG during that same day, it is considered as a visit. In the raw data, there are more than 200 000 CBGs. We therefore first coarse grained the communities by running a K-means clustering algorithm to group the CBGs in 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 Figure 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 Data (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 University [37] which can be accessed at the GitHub repository 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

The fit presented in Figure 2 was performed as follows. We can rewrite equation (1) as:

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = f(p; X(t)), \tag{10}$$

where $X = (S, I)$ is a $2N$-vector and p is the global scalar factor from equation (5). We used a constant flux matrix to compute β_{ab}: the average flux matrix from March 1, 2020 to February 15, 2021. We defined the epidemic size $\omega = (\sum_a (I_a + R_a))/(\sum_a M_a) = G(X)$. Knowing the state $X_n = X(t_n)$ at the time t_n, we computed the state and the epidemic size at time t_{n+1} as:

$$\begin{aligned} X_{n+1} = F(p; X_n) &= X_n + \int_{t_n}^{t_{n+1}} \mathrm{d}u\, f(p; X(u)), \\ \omega_{n+1} &= G \circ F(p; X_n). \end{aligned} \tag{11}$$

Let $\omega^{\mathrm{real}}(t)$ be the "real" epidemic sizes computed from the cases reported in the CSSE dataset. We define the discrepancy function J as the difference between the computed and "real" epidemic sizes,

$$J(p, X_n) = \omega_{n+1} - \omega_{n+1}^{\mathrm{real}}. \tag{12}$$

Starting from t_0 = 2020-03-01, and initializing X_0 from the CSSE dataset, we recursively calculate p_n (Figure 2**b**) as

$$p_n = \underset{p}{\text{argroot}} J(p, X_n), \qquad X_{n+1} = F(p_n; X_n). \tag{13}$$

We used for the calculations a constant recovery rate $\gamma = 0.1\,\text{day}^{-1}$.

Simulations with nearest-neighbors-only interactions

The curves shown in Figure 4**c-d** and the symbols shown in Figure 4**e** were obtained by integrating equation (6).We used the function solve_ivp from SciPy with the "Radau" 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 Figure 4**a**), so that the total fraction of infected individuals was 10^{-8}. The Laplacian was computed using the 9-point stencil $\Delta_{\text{discrete}} = \left(\begin{smallmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{smallmatrix}\right)/12$. We considered periodic boundary conditions along the y direction and Dirichlet boundary conditions along the x direction.

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).

References

1. Henderson, J. *Florence Under Siege: Surviving Plague in an Early Modern City* xviii+363 (Yale University Press, New Haven and London, 2019).
2. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* **115,** 700–721 (1927).
3. Hethcote, H. W. The Mathematics of Infectious Diseases. *SIAM Rev.* **42,** 599–653 (2020).
4. 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. *Applied Mathematics and Computation* **236,** 184–194 (2014).
5. 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 (2020).
6. Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S. & Eggo, R. M. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **20,** 535–558 (2020).
7. 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: 2005.11283 [physics.soc-ph].

8. Aleta, A., Martín-Corral, D., Bakker, M. A., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini, I. M., Pentland, A., Vespignani, A., Moreno, Y. & Moro, E. Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. *medRxiv* (2020).

9. Hébert-Dufresne, L., Althouse, B. M., Scarpino, S. V. & Allard, A. Beyond R_0: heterogeneity in secondary infections and probabilistic epidemic forecasting. *Journal of the Royal Society Interface* **17,** 20200393 (2020).

10. 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).

11. Neipel, J., Bauermann, J., Bo, S., Harmon, T. & Jülicher, F. Power-law population heterogeneity governs epidemic waves. *PloS one* **15,** e0239678 (2020).

12. Sun, K., Wang, W., Gao, L., Wang, Y., Luo, K., Ren, L., Zhan, Z., Chen, X., Zhao, S., Huang, Y., Sun, Q., Liu, Z., Litvinova, M., Vespignani, A., Ajelli, M., Viboud, C. & Yu, H. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* **371** (2021).

13. Kawagoe, K., Rychnovsky, M., Chang, S. Y., Huber, G., Li, L. M., Miller, J., Pnini, R., Veytsman, B. & Yllanes, D. Epidemic dynamics in inhomogeneous populations and the role of superspreaders. *Phys. Rev. Research* **3,** 033283 (Sept. 2021).

14. Pozderac, C. & Skinner, B. Superspreading of SARS-CoV-2 in the USA. *PLOS ONE* **16,** 1–10 (Mar. 2021).

15. Huber, G., Kamb, M., Kawagoe, K., Li, L. M., Veytsman, B., Yllanes, D. & Zigmond, D. A minimal model for household effects in epidemics. *Physical Biology* **17,** 065010 (2020).

16. Aleta, A., Martín-Corral, D., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini Jr., I. M., Merler, S., Pentland, A., Vespignani, A., Moro, E. & Moreno, Y. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4,** 964–971 (2020).

17. Murray, J. Mathematical biology II. Spatial models and biological applications. *Springer-Verlag, New York* (2003).

18. Postnikov, E. B. & Sokolov, I. M. Continuum description of a contact infection spread in a SIR model. *Mathematical biosciences* **208,** 205–215 (2007).

19. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *science* **342,** 1337–1342 (2013).

20. Chu, A., Huber, G., McGeever, A., Veytsman, B. & Yllanes, D. A random-walk-based epidemiological model. *Sci. Rep.* **11,** 19308 (2021).

21. Te Vrugt, M., Bickmann, J. & Wittkowski, R. Effects of social distancing and isolation on epidemic spreading modeled via dynamical density functional theory. *Nature communications* **11,** 1–11 (2020).

22. Tsori, Y. & Granek, R. Epidemiological model for the inhomogeneous spatial spreading of COVID-19 and other diseases. *PloS one* **16,** e0246056 (2021).

23. Tizzoni, M., Bajardi, P., Poletto, C., Ramasco, J. J., Balcan, D., Gonçalves, B., Perra, N., Colizza, V. & Vespignani, A. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC medicine* **10,** 1–31 (2012).

24. Linka, K., Peirlinck, M., Sahli Costabal, F. & Kuhl, E. Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. *Computer Methods in Biomechanics and Biomedical Engineering* **23,** 710–717 (2020).

25. 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* (2014).

26. Hsu, S. & Zee, A. Global spread of infectious diseases. *Journal of Biological Systems* **12,** 289–300 (2004).

27. 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 (2021).

28. 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* (2021).

29. Mayberry, J., Nattestad, M. & Tuttle, A. The Structure of an Outbreak on a College Campus. *Mathematics Magazine* **94,** 83–98 (2021).

30. Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D. & Leskovec, J. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589,** 82–87 (2021).

31. Lloyd, A. L. & May, R. M. Spatial heterogeneity in epidemic models. *Journal of theoretical biology* **179,** 1–11 (1996).

32. Hethcote, H. W. An immunization model for a heterogeneous population. *Theoretical population biology* **14,** 338–349 (1978).

33. Post, W., DeAngelis, D. & Travis, C. Endemic disease in environments with spatially heterogeneous host populations. *Mathematical Biosciences* **63,** 289–302 (1983).

34. May, R. M. & Anderson, R. M. Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences* **72,** 83–111 (1984).

35. Hethcote, H. W. & Van Ark, J. W. Epidemiological models for heterogeneous populations: proportionate mixing, parameter estimation, and immunization programs. *Mathematical Biosciences* **84,** 85–118 (1987).

36. SafeGraph. *SafeGraph Social Distancing Metrics* https://docs.safegraph.com/docs/social-distancing-metrics, year = 2021. 2021.

37. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* **20,** 533–534 (2020).

38. Naether, U., Postnikov, E. & Sokolov, I. Infection fronts in contact disease spread. *The European Physical Journal B* **65,** 353–359 (2008).

39. Dee, G. & Langer, J. Propagating pattern selection. *Physical Review Letters* **50,** 383 (1983).

40. Ben-Jacob, E., Brand, H., Dee, G., Kramer, L. & Langer, J. Pattern propagation in nonlinear dissipative systems. *Physica D: Nonlinear Phenomena* **14,** 348–364 (1985).

41. Mehta, M. *Random Matrices* 3rd. Sec. 1.5 (Elsevier, 2004).

42. Livan, G., Novaes, M. & Vivo, P. *Introduction to Random Matrices: Theory and Practice* (Springer, 2020).

43. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proceedings of the national academy of sciences* **99,** 7821–7826 (2002).

44. Luo, F., Zhong, J., Yang, Y., Scheuermann, R. H. & Zhou, J. Application of random matrix theory to biological networks. *Physics Letters A* **357,** 420–423 (2006).

45. DeMeo, P., Ferrara, E., Fiumara, G. & Ricciardello, A. A novel measure of edge centrality in social networks. *Jour. Knowledge-Based Systems* **30,** 136–150 (2012).

46. Brody, T. A., Flores, J., French, J. B., Mello, P., Pandey, A. & Wong, S. S. Random-matrix physics: spectrum and strength fluctuations. *Reviews of Modern Physics* **53.** p. 391, 385–479 (1981).

47. Guhr, T., Müeller-Groeling, A. & Weidenmüller, H. A. Random matrix theories in quantum physics: common concepts. *Phys. Rep.* **299.** p. 228, 189–425 (1998).

48. Hakke, F. *Quantum signatures of Chaos* 3rd. Sec. 4.5 (Springer, 2010).

49. 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).

Supplementary Information

Contents

1 Supplementary figures **17**

2 Formal solution to the spatial SIR **19**

3 Initial stage of an epidemic **21**

4 Construction of the infectivity matrix from mobility data **21**

5 Spreading with a wave of infection **23**

6 Level statistics and connectedness **29**

Supplementary References **32**

1 Supplementary figures

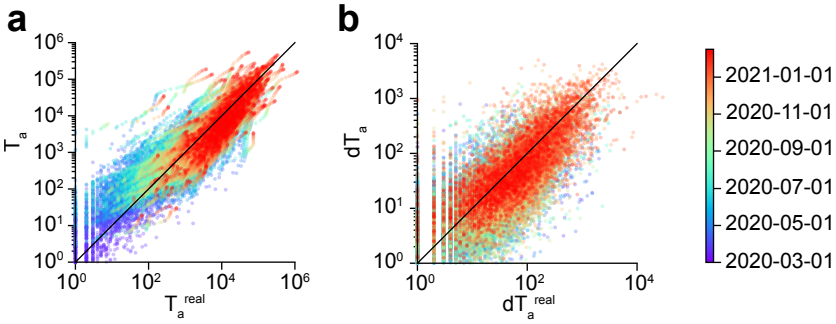

Figure S1 – Supplementary figure to Figure 2 of the main text. Comparison of the model predictions with reported values for **(a)** local epidemic sizes and **(b)** local daily new cases. One symbol is associated to one given community and one given day.

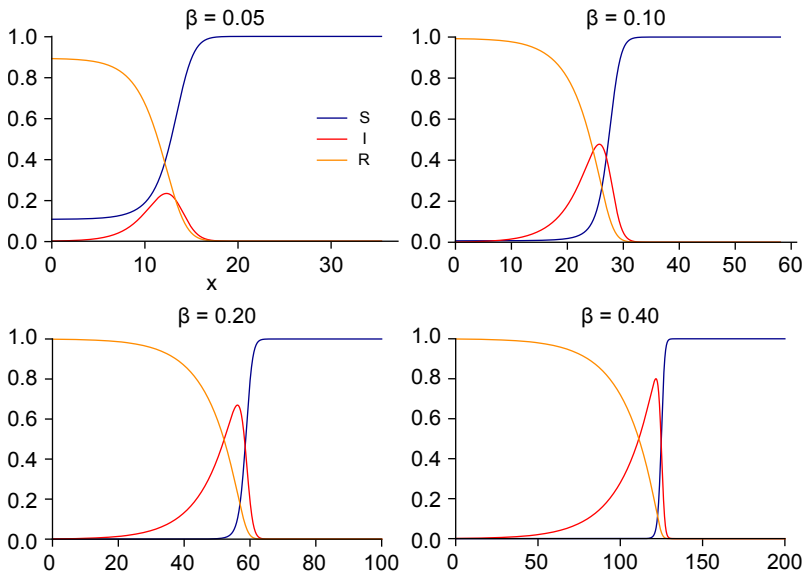

Figure S2 – Supplementary figure to Figure 4 of the main text. Wave profiles for the values of β shown in Figure 4**d-e**.

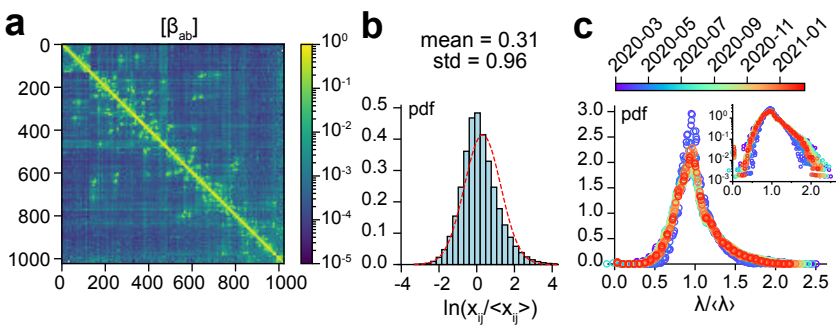

Figure S3 – Supplementary figure to Figure 5 of the main text. **(a)** Mean infectivity matrix (between 2020-03-01 and 2021-02-15). Entries were pooled in squares of size 8×8. The maximum is shown. **(b)** The noise distribution suggests a log-normal distribution of each entry around its mean. **(c)** Eigenvalue distribution of infectivity matrices as a function of time (pooled by 7-day windows).

2 Formal solution to the spatial SIR

2.1 Matrix SIR

We consider N communities affected by a growing epidemic. The spreading of the epidemic is described by the multi-compartment SIR model. Every community a is characterized by the number of susceptible individuals S_a, the infected individuals I_a and the number of recovered individuals R_a. The dynamics for S_a, I_a and R_a with $a \in [\![1, N]\!]$ is described by the equations

$$\begin{aligned} \frac{\mathrm{d}S_a}{\mathrm{d}t} &= -S_a \sum_b \beta_{ab} \frac{I_b}{M_b}, \\ \frac{\mathrm{d}I_a}{\mathrm{d}t} &= -\frac{\mathrm{d}S_a}{\mathrm{d}t} - \gamma I_a, \\ \frac{\mathrm{d}R_a}{\mathrm{d}t} &= \gamma I_a. \end{aligned} \tag{14}$$

We assume that the population in each community is fixed, i.e.

$$S_a(t) + I_a(t) + R_a(t) = M_a = \text{const.} \tag{15}$$

The boundary conditions for equation (14) are:

$$\begin{aligned} S_a(0) &= (1 - \epsilon_a) M_a, \\ I_a(0) &= \epsilon_a M_a, \\ R_a(0) &= 0, \end{aligned} \tag{16}$$

where the parameter ϵ_a describes the initial infection (the initial number of infections per per person in the given community).

We introduce the rescaled variables: $s_a = S_a/M_a$, $j_a = I_a/M_a$ and $r_a = R_a/M_a$. Then equations (14) and (16) become

$$\begin{aligned} \frac{\mathrm{d}s_a}{\mathrm{d}t} &= -s_a \sum_b \beta_{ab} j_b, \\ \frac{\mathrm{d}j_a}{\mathrm{d}t} &= -\frac{\mathrm{d}s_a}{\mathrm{d}t} - \gamma j_a, \\ \frac{\mathrm{d}r_a}{\mathrm{d}t} &= \gamma j_a, \end{aligned} \tag{17}$$

with

$$\begin{aligned} s_a(0) &= (1 - \epsilon_a), \\ j_a(0) &= \epsilon_a, \\ r_a(0) &= 0. \end{aligned} \tag{18}$$

2.2 Solution of matrix SIR equations: the canonical basis

In order to solve equations (17) and (18), we define:

$$\xi_a(t) = \sum_b \beta_{ab} j_b(t), \quad \nu_a(t) = \int_0^t \mathrm{d}u\, \xi_a(u). \tag{19}$$

Substituting equation (19) in equation (17), we solve for s_a and obtain:

$$s_a(t) = (1 - \epsilon_a)e^{-\nu_a(t)}. \tag{20}$$

We also note that $r_a(t)$ can be written as:

$$r_a(t) = \gamma \sum_b \beta^{-1}_{ab} \nu_a(t), \tag{21}$$

where $[\beta^{-1}_{ab}]$ denotes the inverse matrix such that $\sum_c \beta_{ac}\beta^{-1}_{cb} = \delta_{ab}$. We therefore obtain the parametrization:

$$\begin{aligned} s_a(t) &= (1 - \epsilon_a)e^{-\nu_a(t)}, \\ j_a(t) &= 1 - (1 - \epsilon_a)e^{-\nu_a(t)} - \gamma \sum_b \beta^{-1}_{ab} \nu_a(t), \\ r_a(t) &= \gamma \sum_b \beta^{-1}_{ab} \nu_a(t), \end{aligned} \tag{22}$$

as a function of $\nu_a(t)$. Note that when the infectivity matrix is diagonal, namely $\beta_{ab} = \beta\delta_{ab}$, we recover the parametrization from reference [4]:

$$\begin{aligned} s(t) &= s(0)u, \\ j(t) &= \frac{\gamma}{\beta} \ln u - s(0)u + 1, \\ r(t) &= -\frac{\gamma}{\beta} \ln u, \end{aligned} \tag{23}$$

with $u(t) = e^{-\nu(t)}$.

The dynamics is determined by the functions $\nu_a(t)$. Starting from equation (17), we multiply the equations by β_{ab} and sum, obtaining

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_b \beta_{ab} j_b\right) + \gamma \sum_b \beta_{ab} j_b &= \sum_b \beta_{ab}(1 - \epsilon_b)\frac{\mathrm{d}\nu_b}{\mathrm{d}t} e^{-\nu_b} \Leftrightarrow \\ \frac{\mathrm{d}\xi_a}{\mathrm{d}t} + \gamma \xi_a &= \sum_b \beta_{ab}(1 - \epsilon_b)\frac{\mathrm{d}}{\mathrm{d}t}\left(-e^{-\nu_b}\right) \Leftrightarrow \\ \xi_a(t) - \xi_a(0) + \gamma \nu_a(t) &= -\sum_b (1 - \epsilon_b)\left(e^{-\nu_b} - 1\right) \Leftrightarrow \\ \frac{\mathrm{d}\nu_a}{\mathrm{d}t} + \gamma \nu_a &= \sum_b \beta_{ab}\left(1 - (1 - \epsilon_b)e^{-\nu_b}\right). \end{aligned} \tag{24}$$

Equation (24) can be solved numerically. In the $t \to \infty$ limit, we obtain:

$$\gamma \nu_a(\infty) = \sum_b \beta_{ab}\left(1 - (1 - \epsilon_b)e^{-\nu_b(\infty)}\right), \tag{25}$$

which is the same as setting $j_a = 0$ in equation (22). Equation (25) can be solved iteratively. The size of the epidemic is given by:

$$\begin{aligned} \Omega(t) &= \sum_a \left(I_a(t) + R_a(t)\right), \\ &= \sum_a M_a\left(1 - (1 - \epsilon_a)e^{-\nu_a(t)}\right), \\ &\xrightarrow[t\to\infty]{} \sum_a M_a\left(1 - (1 - \epsilon_a)e^{-\nu_a(\infty)}\right). \end{aligned} \tag{26}$$

2.3 Solution of matrix SIR equations: SVD basis

A solution can also be written using the singular value decomposition of the infectivity matrix β:

$$\beta_{ab} = \sum_{\alpha} \lambda_\alpha v_{\alpha a} u_{\alpha b}. \tag{27}$$

For a given vector $\mathbf{x}$, we use the following notation: $x_\alpha = \sum_a v_{\alpha a} x_a$. From equation (24) we thus obtain:

$$\frac{\mathrm{d}v_\alpha}{\mathrm{d}t} + \gamma v_\alpha = \lambda_\alpha \sum_a u_{\alpha a} \left(1 - (1-\epsilon_a) e^{-\sum_\tau v_{\tau a} v_a}\right). \tag{28}$$

3 Initial stage of an epidemic

We again start from equation (17), which we linearize around the initial state $s_a^{(0)} = 1$, $j_a^{(0)} = 0$, and $r_a^{(0)} = 0$. To first order $\mathbf{j}$ and $\mathbf{r}$ satisfy the ODE:

$$\begin{aligned} \frac{\mathrm{d}j_a}{\mathrm{d}t} &= \sum_b \beta_{ab} j_b - \gamma j_a, \\ \frac{\mathrm{d}r_a}{\mathrm{d}t} &= \gamma j_a. \end{aligned} \tag{29}$$

$[\beta_{ab}]$ is a real square positive matrix. According to the Perron-Frobenius theorem, there exists the maximal eigenvalue $\lambda_\omega > 0$, such that any other eigenvalue $\lambda_\alpha < \lambda_\omega$, and the associated eigenvector $\mathbf{v}_\omega$ is positive. At large times, we have:

$$\mathbf{j} \sim (\mathbf{j}(0) \cdot \mathbf{v}_\omega)\, e^{(\lambda_\omega - \gamma)t} \mathbf{v}_\omega, \tag{30}$$

where $(\mathbf{A} \cdot \mathbf{B})$ denotes a scalar product. The epidemic grows only if the basic reproduction number $\mathcal{R}_0 = \frac{\lambda_\omega}{\gamma} > 1$. Furthermore, if the infectivity matrix can be factorized as $\beta_{ab} = f_a g_b$, then β is of rank 1 and λ_ω is the only non-zero eigenvalue. We then have $\lambda_\omega = \sum_a f_a g_a$ and $\mathbf{v}_\omega = \mathbf{f}$.

4 Construction of the infectivity matrix from mobility data

The mobility data obtained from SafeGraph allows us to construct a pseudo-flux matrix in which each entry f_{ab} represents the number of individuals from community a visiting community b per day. Let us now consider one community a, having S_a susceptible individuals and I_a infected individuals. The variation in susceptible individuals due to new infections during the time interval Δt has the form:

$$S_a(t+\Delta t) - S_a(t) = -S_a \times \Pr\,(\text{meeting an infected individual}) \times \beta \Delta t, \tag{31}$$

where $\beta\Delta t$ represents the probability to get infected when meeting an infected individual. We now list the different contributions.

4.1 Intra-community contributions

The contributions coming from infected individuals in the same community are:

$$\begin{aligned} & S_a \times p_a \frac{I_a}{\sum_b f_{ba} + M_a} \times \beta \Delta t, \\ & = p_a \beta \Delta t S_a I_a \frac{1}{M_a} \left(1 - \frac{f_{*a}}{M_a}\right) + \mathrm{o}\left(\frac{f_{*a}}{M_a}\right), \end{aligned} \tag{32}$$

where we have introduced $f_{*a} = \sum_b f_{ab}$, and the parameter p_a representing the frequency with which an individual is interacting with other individuals in community a. Here $I_a/(f_{*a} + M_a)$ is the probability to meet an infected individual when interacting with an individual in the community.

4.2 Inter-community contributions: incoming visitors

The contributions to the infections in community a from infected individuals visiting from another community b are:

$$\begin{aligned} & S_a \times p_a \frac{f_{ba}}{f_{*a} + M_a} \frac{I_b}{M_b} \times \beta \Delta t, \\ & = p_a \beta \Delta t S_a I_b \frac{f_{ba}}{M_a M_b} \left(1 - \frac{f_{*a}}{M_a}\right) + \mathrm{o}\left(\frac{f_{*a}}{M_a}\right), \end{aligned} \tag{33}$$

where we have assumed that $f_{ba} I_b / M_b$ is the number of infected visitors from b.

4.3 Inter-community contributions: returning natives

The contributions coming from individuals from community a infected while visiting another community b are:

$$\begin{aligned} & f_{ab} \frac{S_a}{M_a} \times p_b \frac{I_b}{f_{*b} + M_b} \times \beta \Delta t, \\ & = p_b \beta \Delta t S_a I_b \frac{f_{ab}}{M_a M_b} \left(1 - \frac{f_{*b}}{M_a}\right) + \mathrm{o}\left(\frac{f_{*b}}{M_a}\right), \end{aligned} \tag{34}$$

where we have assumed that $f_{ab} S_a / M_a$ is the number of susceptible visitors from a visiting community b.

Neglecting the f_{*a}/M_a terms, we obtain after adding all contributions:

$$\frac{\mathrm{d}S_a}{\mathrm{d}t} = -S_a \sum_b \beta_{ab} \frac{I_b}{M_b}, \tag{35}$$

with:

$$\begin{aligned} \forall a \in [\![1, N]\!], \quad & \beta_{aa} = \beta p_a, \\ \forall a < b, \quad & \beta_{ab} = \beta \frac{f_{ab} p_b + f_{ba} p_a}{M_a}. \end{aligned} \tag{36}$$

Note that in general, the matrix $[\beta_{ab}]$ is not symmetric. In the manuscript, we make the assumption that the interaction frequency is the same in all communities, namely $p_a = p$.

5 Spreading with a wave of infection

5.1 ODE for the wave profile

We consider the SIR model on a 2d lattice, with infections limited to nearest-neighbors, and with uniform population $M_a = M$. For the sake of simplicity, we will consider the rescaled S, I, R variables with $S + I + R = M = 1$. We write:

$$\begin{aligned}
\frac{\mathrm{d}S_{i,j}}{\mathrm{d}t} &= -S_{i,j}\left(\alpha I_{i,j} + \beta\left(I_{i-1,j} + I_{i+1,j} + I_{i,j-1} + I_{i,j+1}\right)\right), \\
\frac{\mathrm{d}I_{i,j}}{\mathrm{d}t} &= -\frac{\mathrm{d}S_{i,j}}{\mathrm{d}t} - \gamma I_{i,j}, \\
\frac{\mathrm{d}R_{i,j}}{\mathrm{d}t} &= \gamma I_{i,j},
\end{aligned} \tag{37}$$

where i and j denote the indices along the first and second dimensions. We introduce the discrete laplacian:

$$\Delta I_{i,j} = \frac{\left(I_{i-1,j} + I_{i+1,j} - 2I_{i,j}\right) + \left(I_{i,j-1} + I_{i,j+1} - 2I_{i,j}\right)}{\delta x^2},$$

where δx is the lattice spacing. Equation (37) becomes:

$$\begin{aligned}
\frac{\mathrm{d}S_{i,j}}{\mathrm{d}t} &= -\beta S_{i,j}\left(aI_{i,j} + (\delta x)^2 \Delta I_{i,j}\right), \\
\frac{\mathrm{d}I_{i,j}}{\mathrm{d}t} &= -\frac{\mathrm{d}S_{i,j}}{\mathrm{d}t} - \gamma I_{i,j}, \\
\frac{\mathrm{d}R_{i,j}}{\mathrm{d}t} &= \gamma I_{i,j},
\end{aligned} \tag{38}$$

where $a = 4 + \alpha/\beta$. In our typical application, we will consider $\alpha = \beta$, so $a = 5$. From now on, we will set $\delta x = 1$. In applications, allowing δx to vary gives us the scaling of our parameters with respect to distance. Additionally, if we had chosen a non-nearest-neighbor model, allowing δx to vary can help us to understand how far one should restrict travel in order to stop an epidemic. In the continuum we have:

$$\begin{aligned}
\frac{\partial S}{\partial t}(x,y,t) &= -\beta S(x,y,t)\left(aI(x,y,t) + \Delta I(x,y,t)\right), \\
\frac{\partial I}{\partial t}(x,y,t) &= -\frac{\partial S}{\partial t}(x,y,t) - \gamma I(x,y,t), \\
\frac{\partial R}{\partial t}(x,y,t) &= \gamma I(x,y,t),
\end{aligned} \tag{39}$$

We look for solutions which can be written as a wave of the form:

$$\begin{aligned}
S(x,y,t) &= F\left(x - x_0 - vt\right), \\
R(x,y,t) &= G\left(x - x_0 - vt\right),
\end{aligned} \tag{40}$$

where v is the velocity of the wave and x_0 is the position of the wave at $t = 0$. We introduce the functions ϕ and ψ:

$$\begin{aligned}
F(x) &= \frac{1 + \phi\left(\frac{2v}{\beta}x\right)}{2}, \\
G(x) &= \frac{1 - \psi\left(\frac{2v}{\beta}x\right)}{2}.
\end{aligned} \tag{41}$$

By equation (39) they satisfy the differential equation:

$$\begin{aligned} \phi'' - \psi'' &= -\frac{\phi'}{1+\phi} + C(\psi - \phi), \\ \psi' &= CD(\psi - \phi), \end{aligned} \tag{42}$$

where

$$C = \frac{a\beta^2}{4v^2}, \qquad D = \frac{2\gamma}{a\beta}. \tag{43}$$

Equation (42) can be rewritten as a first-order ODE:

$$\begin{aligned} \theta' &= -\left(\frac{1}{1+\phi} + CD\right)\theta + C(1+CD^2)(\psi - \phi), \\ \phi' &= \theta, \\ \psi' &= CD(\psi - \phi), \end{aligned} \tag{44}$$

5.2 Solution for the special case $\gamma = 0$

We start with a simpler but instructive case, namely for a vanishing recovery rate $\gamma = 0 \Rightarrow D = 0$. Any infected individual remains infected forever, thus we have $R(x,t) = 0 \Rightarrow G(x) = 0 \Rightarrow \psi(z) = 1$. In that case, equation (44) becomes:

$$\begin{aligned} \theta' &= -\frac{\theta}{1+\phi} + C(1-\phi), \\ \phi' &= \theta, \end{aligned} \Leftrightarrow \mathbf{z}' = f(\mathbf{z}), \qquad \text{with} \quad \mathbf{z}^T = (\theta, \phi) \tag{45}$$

To the left of the wave, the profiles of S and R must be flat. Thus by setting $\phi''(-\infty) = 0$, we have the initial condition:

$$\theta(0) = C(1 - \phi(0)^2). \tag{46}$$

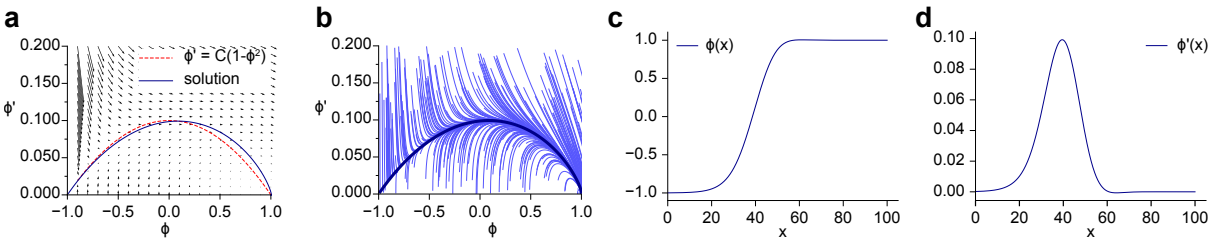

Figure S4 – **(a)** Phase space for equation (45). The black arrows represent the field $f(\mathbf{z})$. The solution to equation (45) with the initial condition from equation (46) and $\phi(0) = -0.99$ is shown as a solid blue line. **(b)** Solutions to equation (45) with initial conditions drawn randomly in the phase plane converge toward the stable solution. **(c-d)** Profiles of the stable solution from **(a)**. We used $C = 0.1$.

In Figure S4**a** we represent the phase space associated with equation (45) for a value $C = 0.1$. In particular, we can visualize the vector field $f(\theta, \phi)$ in the (θ, ϕ)-plane. We obtain the stable solution using the initial condition equation (46) and $\phi(0) = -0.99$. In Figure S4**b**, we show that solutions initialized from random initial conditions in the phase plane converge toward the stable solution. The profiles $(\theta(x), \phi(x))$ associated with the stable solution are shown in Figure S4**c-d**.

5.3 Solution for the general case $\gamma > 0$

We now come back to the general case $\gamma > 0$. We formally rewrite the system equation (44) as:

$$\mathbf{z}' = f(\mathbf{z}), \tag{47}$$

where $\mathbf{z}^T = (\theta, \phi, \psi)$. Any fixed point must satisfy the condition:

$$f(\mathbf{z}) = 0 \Leftrightarrow \mathbf{z} = \begin{pmatrix} 0 \\ \mu \\ \mu \end{pmatrix}, \quad \mu \in]-1, 1]. \tag{48}$$

Note that for such fixed points, $\phi = \psi \Rightarrow S + R = 1 \Rightarrow I = 0$. Therefore, these fixed points correspond to states with no infected individuals, namely $I = 0$. The jacobian of f is:

$$J(\theta, \phi, \psi) = \begin{pmatrix} -\left(\frac{1}{1+\phi} + CD\right) & \frac{\theta}{(1+\phi)^2} - C(1 + CD^2) & C(1 + CD^2) \\ 1 & 0 & 0 \\ 0 & -CD & CD \end{pmatrix}. \tag{49}$$

Looking for the right eigenvector $\mathbf{v}$ associated with the eigenvalue λ, namely $J\mathbf{v} = \lambda\mathbf{v}$, we obtain:

$$\mathbf{v} \propto \begin{pmatrix} \frac{C(1+CD^2)\lambda - (\lambda - CD)\frac{x}{(1+\phi^2))}}{\lambda + \frac{1}{1+\phi} + CD} \\ -(\lambda - CD) \\ CD \end{pmatrix}. \tag{50}$$

The eigenvalue λ is the solution of the equation:

$$-\lambda\left(\frac{1}{1+\phi} + CD + \lambda\right)(CD - \lambda) + \frac{\theta}{(1+\phi)^2}(CD - \lambda) + C(1 + CD^2)\lambda = 0. \tag{51}$$

Therefore, on the line of fixed points, where $\theta = 0$, we have $\lambda = 0$ or:

$$\lambda^2 + \frac{1}{1+\phi}\lambda - CD\left(\frac{1}{1+\phi} + CD\right) + C(1 + CD^2) = 0, \tag{52}$$

with the discriminant:

$$\Delta = \frac{1}{(1+\phi)^2} - 4C\left(1 - \frac{D}{1+\phi}\right). \tag{53}$$

If $\Delta < 0$, then solutions will exhibit an oscillatory behavior in the vicinity of this fixed point. However, by definition of the model in equation (37), $S(t)$ decreases monotonically, therefore oscillations are not permitted. In particular, a spatial oscillation in ϕ would impart an oscillation in time to S. Thus we are looking for fixed points with two real eigenvalues λ_+ and λ_-, such that $\Delta \geq 0$:

$$\begin{aligned} \Delta \geq 0 &\Leftrightarrow (1+\phi)^2 - D(1+\phi) - \frac{1}{4C} \leq 0 \\ &\Leftrightarrow \frac{D}{2}\left(1 - \sqrt{1 + \frac{1}{CD^2}}\right) \leq 1 + \phi \leq \frac{D}{2}\left(1 + \sqrt{1 + \frac{1}{CD^2}}\right). \end{aligned} \tag{54}$$

In the case where $\gamma = 0$, we can prove that this condition is sufficient to guarantee that ϕ is monotonic. Since by definition $1 + \phi > 0$, only the upper bound condition needs to hold. Therefore, the Jacobian has real eigenvalues only and only if:

$$1 + \phi \leq \frac{D}{2}\left(1 + \sqrt{1 + \frac{1}{CD^2}}\right), \tag{55}$$

and the eigenvalues are:

$$\lambda_{\pm} = -\frac{1}{2(1+\phi)}\left(1 \pm \sqrt{1 - 4C\left(1 - \frac{D}{1+\phi}\right)(1+\phi)^2}\right). \tag{56}$$

Clearly $\lambda_+ < 0$. As for λ_-:

$$\lambda_- < 0 \Leftrightarrow \phi > D - 1. \tag{57}$$

Therefore, the fixed point $\mathbf{z}^T = (0, \mu, \mu)$ is stable as long as:

$$D < 1 + \mu \leq \frac{D}{2}\left(1 + \sqrt{1 + \frac{1}{CD^2}}\right). \tag{58}$$

Note that the fixed point $\mathbf{z}^T = (0, 1, 1)$, reached when $x \to \infty$ must be stable (it corresponds to $S = 1$, $I = R = 0$). Using equation (55), this yields the condition:

$$C \leq \frac{1}{16(1 - \frac{D}{2})} \Leftrightarrow v \geq v_{\text{vmin}} = 2\beta\sqrt{a - \frac{\gamma}{\beta}}. \tag{59}$$

Equation (59) gives a strictly positive lower bound for the wave velocity when $\beta > \gamma/a$. Typical propagation velocities are close to v_{min} [18, 21, 38].

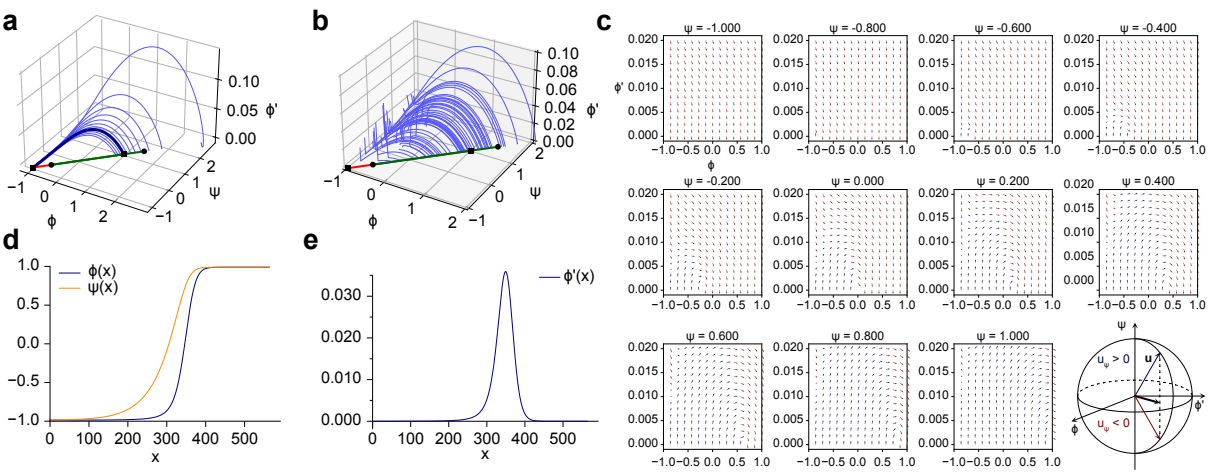

Figure S5 – **(a)** Solutions to equation (44) shown in phase space, with initial conditions chosen according to equations (60) and (61), with $-1 < \phi(0) < D - 1$. Several trajectories "overshoot" the fixed point $(0, 1, 1)$ and some other trajectories "undershoot" it. The physical solution must converge to $(0, 1, 1)$ and is indicated with a thick blue line. Black squares indicate $z^T = (0, -1, -1)$ and $z^T = (0, 1, 1)$ while black circles indicate $z^T = (0, D-1, D-1)$ and $z^T = (0, D(1 + \sqrt{1 + 1/(CD^2)})/2 - 1, D(1 + \sqrt{1 + 1/(CD^2)})/2 - 1)$. **(b)** Solutions to equation (44) shown in phase space, with initial condition chosen randomly in the phase space ($\phi(0) < \psi(0)$). **(c)** Visualization of the unit vector $\mathbf{u} \propto f(\mathbf{z})$ in phase space. $\mathbf{u}$ is shown in blue when $u_\psi > 0$ and in red otherwise. **(d-e)** Profiles of the admissible solution shown in **(a)**. We took $\alpha = \beta = \gamma = 0.1$ and $C = 0.05$

As for the initial condition, we will consider typically that $\lim_{t\to+\infty} I(t) = 0$. In practice, we will consider $I(-\infty) = \epsilon$, and set the initial condition:

$$\psi(0) = \phi(0) + 2\epsilon, \tag{60}$$

where we used $S + I + R = 1$. Left to the wave we expect a flat profile, with S and R converging toward their respective asymptotical values, so $\phi''(-\infty) = \psi''(-\infty) = 0$. Using equation (42), we thus set the initial condition:

$$\begin{aligned} \theta(0) &= \phi'(0), \\ &= C(1 + \phi(0))(\psi(0) - \phi(0)), \\ &= 2C\epsilon(1 + \phi(0)) \end{aligned} \tag{61}$$

In Figure S5**a**, we show solutions to equation (44), for different values of $\phi(0)$, and using the initial conditions from equations (60) and (61). We see that depending on the initial condition $\phi(0)$, trajectories might either "overshoot" or "undershoot" the fixed point $(0, 1, 1)$. We can conjecture that there is only one trajectory landing exactly on this fixed point. Since this fixed point must belong to the trajectory ($S \approx 1$ and $R \approx 0$ at $t = 0$), and making the conjecture that it belongs to only one trajectory, it is the admissible solution to the problem defined by equations (39) and (40). Numerically, we find the root of the function $\phi(0) \to \phi(x_\infty) - 1$, where x_∞ is a sufficiently large number. In Figure S5**a**, we can also see the oscillatory behavior near fixed points $z^T = (0, \mu, \mu)$ such that $\mu > D\left(1 + \sqrt{1 + 1/(CD^2)}\right)/2 - 1$. In Figure S5**b**, we show trajectories with initial conditions randomly drawn in phase space, while preserving the condition $\phi(0) < \psi(0)$, converging toward the line of fixed points. In Figure S5**c**, we represent the values in phase space of the unit vector $\mathbf{u} \propto f(\mathbf{z})$. Finally, the profiles $(\theta(x), \phi(x), \psi(x))$ associated with the admissible solution are shown in Figure S5**d-e**. The difference $\psi - \phi > 0$ represents the fraction of infected individuals.

5.4 Monotonicity and uniqueness of ϕ when $\gamma = 0$

In the case where $\gamma = 0$, we can prove that the condition in Equation (59) is also sufficient to prove that ϕ is monotonic in space, and thus S is monotonic in space and time, the latter of which is a requirement of our system. Therefore, the ansatz ϕ is consistent with the steady state solution to the PDE. Additionally, such a solution is unique given a value of $0 < C < 1/16$. We restart from equation (45).

For a given $-1 < \phi < 1$, it is clear that $\theta' > 0$ for small enough $\theta > 0$. This tells us that so long as the trajectory does not start with $\theta \leq 0$, we will not have to consider $\theta \leq 0$ if we can prove that $-1 < \phi < 1$. It is simple to see that $-1 < \phi$ by observing the vector field in the viscinity of $(\phi, \theta) = (-1 + \epsilon, \theta)$ for $\theta > 0$ and some small $\epsilon > 0$, so we will focus on the other bound. Importantly, $\theta > 0$ is equivalent to the monotonicity condition on ϕ and on $S(x, t)$ in the steady state.

The proof will require a few facts about the trajectory which we will prove:

- There is exactly one x for which $\theta' = 0$.
- $0 < \theta < \frac{1}{4}(1 - \phi)$.
- The neighborhood around $\phi = -1, \theta = 0$ behaves like a saddle point.

The first fact will help prove the second. The second point will ultimately show that our solution is bounded from above by $\phi < 1$. The third point proves the uniqueness of the solution with the proper boundary conditions. We now prove the first of our points. In phase space, one particularly informative curve is $\theta'(\phi, \theta) = 0$. This is given by

$$\theta = C(1 - \phi^2) \tag{62}$$

Any trajectory that is above this curve ($\theta > C(1 - \phi^2)$) at some point $-1 < \phi < 0$, must diverge to infinity near $\phi = -1$. This can be seen in Figure S4**b**. Since the initial condition is $(\phi, \theta) = (-1, 0)$ for $x \to -\infty$, we only care about solutions for which all $-1 < \phi < 0$ values have with $\theta \leq C(1-\phi^2)$. Of course, if there is any point for which $\theta = C(1-\phi^2)$ and $\phi < 0$, then for a slightly smaller ϕ, we know that $\theta > C(1-\phi^2)$. Therefore, we are only interested in solutions where $\theta < C(1 - \phi^2)$ for $-1 < \phi < 0$. However, we can see from this same argument that any trajectory emanating from the point $(\phi, \theta) = (-1, 0)$ must point below the curve $\theta = C(1-\phi^2)$. Therefore, we know that the solution for ϕ automatically meets the aforementioned condition for $x \to -\infty$. We are now left to understand what happens to trajectories when they start with $0 < \theta < C(1 - \phi^2)$.

Since the curve $\theta(\phi) = C(1 - \phi^2)$ is concave down, is symmetric around $\phi = 0$, and $\theta' = 0$ on the curve, we know that the trajectory can only cross the above the curve for $\phi > 0$. This actually tells us that any curve $(\phi(x), \theta(x))$ that has obeys $\theta(x) < C(1 - \phi(x)^2)$ for $x \to -\infty$ will also obey this condition for any x with $-1 < \phi(x) < 0$. The fact that θ and ϕ must be increasing with x before crossing this curve means that, indeed, the trajectory must cross the curve for some x. After crossing, we can say that the trajectory never crosses the curve again via similar reasoning to the last part of the argument that the trajectory doesn't cross $\theta = C(1-\phi^2)$ for $-1 < \phi < 0$. Therefore, the trajectory must cross the curve $\theta(\phi) = C(1 - \phi^2)$ exactly once.

Now we move on to the second point, that $0 < \theta < \frac{1}{4}(1 - \phi)$. To show this, we can just calculate $\dfrac{d\theta}{d\phi} = \dfrac{\theta'}{\phi'}$ on this line and show that it is less than $-1/4$. We will use the fact that $C < 1/16$. Thus,

$$\begin{aligned}
\frac{d\theta}{d\phi} = \frac{\theta'}{\phi'} &= \frac{-\dfrac{\theta}{1+\phi} + C(1-\phi)}{\theta} \\
&= -\frac{1}{1+\phi} + 4C \\
&< -\frac{1}{2} + 4\frac{1}{16} \\
&< -\frac{1}{4}
\end{aligned} \tag{63}$$

Therefore, $0 < \theta < \frac{1}{4}(1 - \phi)$. In particular, after the trajectory crosses $\theta = C(1 - \phi^2)$, it is subject to the constraint $C(1 - \phi^2) < \theta < \frac{1}{4}(1 - \phi)$. This creates a wedge shape boundary for the trajectory that restricts $\phi < 1$. This proves that the ansatz $\phi(x)$ is compatible with the monotonicity condition for $S(x, t)$.

Finally, we see that the neighborhood around $\phi = -1, \theta = 0$ looks like a saddle point when analyzing the spectrum of the Jacobian in equation (49) or looking at Figure S4**b**. Reversing the arrow of time, we see that as the trajectory follows the ridge to $\phi = -1, \theta = 0$, all of the neighboring solutions will fall away. Thus, this solution is unique.

This completes the proof that there is only one unique solution for each $0 < C < 1/16$ and that this ansatz for ϕ is consistant with the monotonicity conditions for $S(x, t)$.

6 Level statistics and connectedness

6.1 The level spacing distribution of the flux matrix

We consider the level statistics of the symmetric matrix $F_{ab} = (f_{ab} + f_{ba})/(M_a M_b)$, where f_{ab} is the number of people residing in community a visiting community b during one day and M_a is the total population of site a, see equation (15). The symmetric flux is related to the infectivity matrix through $\beta_{ab} = p\beta F_{ab} M_b$, where the factor $p\beta$ is estimated by a fit to the SafeGraph mobility data over $K = 353$ days from March 1st 2020 to February 16th 2021 and is given by $p\beta = 0.05 \pm 0.02$, as is shown in Figure 2**b** of the main text. Following common practices in the treatment of random matrices [41], the diagonal elements $F_{aa}^{(k)}$ of the K matrices of size $N \times N$ were drawn from a generic gamma distribution with mean $\mu_a^{(k)} = (N-1)^{-1} \sum_{b \neq a} F_{ab}^{(k)}$ and variance $\mu_a^{(k)}/2$ $\{k = 1, 2, \ldots, K\}$. To obtain the unfolded spectra [41], the eigenvalues E_{ak} are re-scaled as

$$\epsilon_{ak} = N\bar{G}(E_{ak}), \qquad a = 1, 2, \ldots, N, \; k = 1, 2, \ldots, K \tag{64}$$

where $G(E)$ is the empirical level staircase function $G(E) \equiv (NK)^{-1} \sum_{ak} \Theta(E - E_{ak})$ (*i.e.* the cumulative eigenvalue probability function) and $\bar{G}$ is its smoothed interpolation [46–48]. The nearest level spacings $s_{ak} \equiv \epsilon_{a+1,k} - \epsilon_{a,k}$ are then normalized so that the average level spacing across the entire spectrum is equal to one for all values of k:

$$\bar{s}_{ak} = \sum_{a=1}^{N-1} s_{ak}/(N-1) = 1. \tag{65}$$

Figure S6**a** shows the level spacing distribution of the unfolded spectrum, $p(s) \equiv N_s^{-1} \sum_{i=1}^{N_s} \delta(s - s_i)$, where N_s is the number of independent spacings (for K matrices, $N_s = K(N-1)$). The level statistics of the full matrix, Figure S6**a**, is significantly different than the level statistics obtained after all the links of distance larger than 170 km have been omitted (Figure S6**b**). Clearly, the empirical distribution $p(s)$ interpolates between the Wigner-Dyson distribution of the GOE ensemble (more precisely, the Wigner surmise [41]) that exhibits linear level repulsion typical of extended correlated states:

$$p(s) = (s\pi/2)e^{-s^2\pi/4} \qquad (s \geq 0), \tag{66}$$

and the Poisson distribution describing independent localized states:

$$p(s) = \exp(-s) \quad (s \geq 0) \tag{67}$$

In both equations (66) and (67), the average level spacing is normalized to unity: $\langle s \rangle = \int_0^\infty \mathrm{d}s \, s p(s) = 1$.

6.2 KL divergence and the level spacing entropy

To quantify the difference between Figure S6**a** and Figure S6**b**, we have calculated the Kullback-Leibler divergence between the empirical spacing distribution, $p(s)$, and the Poisson distribution

$$D(\text{SafeGraph}||\text{Poisson}) = 1 + \sum_{i=1}^{N_s} p(s_i) \log[p(s_i)] = 1 - S(\text{SafeGraph}) \tag{68}$$

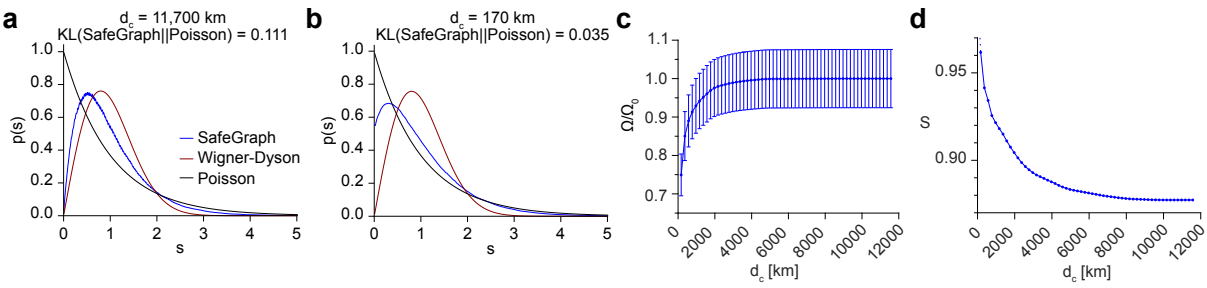

Figure S6 – The unfolded level spacing $s_i = \varepsilon_{i+1} - \varepsilon_i$. **(a)** The level spacing distribution of the full matrix, *i.e.* without removing any links. **(b)** The level spacing of the truncated matrix $d_c \leq 170$ km. Here $\varepsilon_i = \bar{G}(E_i)$ $\{i = 1, 2, \ldots, N\}$, where $\bar{G}(E)$ is the smoothed energy staircase function. The GOE level spacing (equation (66)) is plotted in red and the Poisson (equation (67)) in black. **(c)** The relative epidemic size Ω/Ω_0 as a function of d_c. The estimated relative error is $\Delta\Omega/\Omega_0 = 7\,\%$. Ω_0 denotes the unrestricted epidemic size. **(d)** The entropy S(SafeGraph) as a function of the cutoff distance d_c in the range 170-12000 km. The maximal value of d_c corresponds to a full flux matrix.

where D is the KL divergence and S is the entropy of $p(s)$. The KL divergence of the full matrix is then $D \simeq 0.11$. The KL divergence of the truncated matrix, that shows significantly less level repulsion, and is closer to Poisson statistics, is smaller than that of the full matrix by a factor of three: $D \simeq 0.04$. The KL divergence between the GOE and the Poisson distributions is

$$D(\text{GOE}||\text{Poisson}) = \frac{\pi}{2}\int_0^\infty ds\, s \exp(-s^2\pi/4)\left[\log(s\pi/2) - \pi s^2/4 + s\right] = (\log\pi - \gamma_E)/2 = 0.283057 \equiv D_0 \quad (69)$$

where $\gamma_E = 0.57721\ldots$ is the Euler constant. Therefore, in terms relative to D_0 we obtain: $D/D_0 = \{0.39, 0.12\}$ for the full and the truncated matrices respectively. Thus, D(SafeGraph||Poisson) is smaller than D_0 and approaches zero as links are being removed. Equivalently, since (i) the entropy of the Poisson distribution is equal to its average $\langle s\rangle = 1$ and (ii) the entropy of Wigner-Dyson distribution is $1 - D_0 \simeq 0.72$, the level spacing entropy varies in the range $1 - D_0 < S(\text{SafeGraph}) \leq 1$ and it approaches unity as links are removed and the states get localized.

The entropy as a function of the cutoff distance d_c is shown in Figure S6**d**. The distribution $p(s)$ is found by (i) removing all links of distance larger than d_c (ii) unfolding the spectra of the ensemble of K truncated matrices and (iii) repeating the procedure for each value of d_c. The corresponding size of the epidemic $\Omega(\infty)$ and its error $\Delta\Omega$ are estimated by solving equations (25) and (26) for each one of the truncated matrices and then calculating the mean and standard deviation over K samples. In the numerical computation we assumed a recovery rate $\gamma = 0.125[\text{day}^{-1}]$ and initial fraction of infections $\epsilon = 10^{-4}$ uniform for all the communities. The epidemic size Ω/Ω_0 as a function of d_c, relative to the value of the unrestricted epidemic size Ω_0 (*i.e.* for the case of a full untruncated matrix), is shown in Figure S6**c**. Thus, a cutoff distance of $d_c \simeq 500$ km leads to a 10 % reduction in the epidemic size. Similarly, $d_c \simeq 200$ km reduces the epidemic size by 25 %. These results are in accordance with the time evolution of the epidemic as is presented in Figure 3**b** of the main text.

6.3 Connectedness and epidemic size

The crossover from Wigner-Dyson to Poisson distribution as links between communities are being successively removed, suggests a policy of isolation which can lead to an effective reduction of the epidemic size. Edges connecting vertices a and b can be removed in several ways. For example, (i) according to the geographical distance $d(a, b)$ as has already been done in section 6.2, (ii) according to the "nominal" distance $|a - b|$, or (iii) by using the edge-betweenness centrality index, as proposed by Girvan and Newman [43]. Note, that in each one of these cases, one can as well consider a moderated mitigation policy: instead of eliminating links completely, one can impose constraints on the flux of people that are allowed to commute via central pre-determined links.

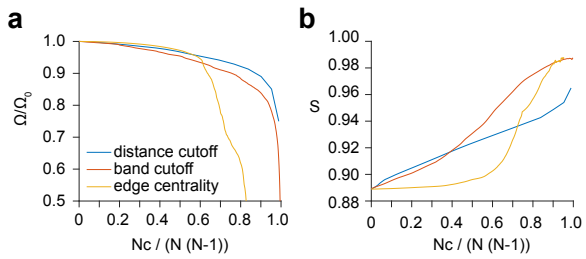

Figure S7 – **(a)** The epidemic size Ω/Ω_0 versus the relative number of cuts $N_c/N/(N-1)$ for the 3 mitigation methods: distance cutoff, bandwidth, and edge-betweenness centrality. Ω_0 is the epidemic size for $N_c = 0$. **(b)** A comparison of the entropy versus the relative number of cuts for the three mitigation method.

To study mitigation according to the nominal distance $|a - b|$, we consider the ensemble of banded matrices F_{ab} with half-bandwidth $B \equiv \max_{F_{ab}\neq 0} |a - b|$, such that the number of non-zero diagonals is $2B + 1$ irrespective of the underlying physical distance. The increase of Ω/Ω_0 as a function of B, along with a corresponding decrease of the entropy S(SafeGraph), is shown in Figure 5**d** of the main text. Comparing to Figure 5**c** of the main text, a reduction of 10 % in Ω/Ω_0 amounts to entropy values of $S = \{0.94, 0.98\}$ for the geographical and nominal cutoffs, respectively. One may argue that mitigation based on "nominal" distance is meaningless, because such a distance changes under the permutations of sites. Indeed, a certain link may either be kept as is, or be omitted, according to an arbitrary re-ordering of communities. However, performing such permutation only reshuffles the matrices of the random ensemble among themselves. Consequently, for $K, N \gg 1$, the level statistics and the resulting epidemic size are hardly affected.

Centrality indices have been used for detecting the modular structure of social and biological networks [43, 44]. The edge-betweenness centrality of edge (a, b) is defined as [45]: $C(a, b) = \sum_{ts} \sigma_{ts}(a, b)/\sigma_{ts}$, where (t, s) stands for "target" and "source" vertices, respectively. Here, σ_{ts} are shortest paths going $t \leftarrow s$ and $\sigma_{ts}(a, b)$ are such paths which, in addition, go along edge (a, b). The entries σ_{ts} are weighted by the flux that they can carry (otherwise, unweighted paths take only binary values $\sigma_{ts} = 0, 1$). The summation is carried over all (t, s) pairs that are different from (a, b). As observed in [43], edges running between loosely-connected communities should have high values of C.

The scheme proposed in [43] for identifying communities in a network is as follows: (i) calculate C for all edges of the network (ii) remove the edge with highest C (iii) re-calculate C for all edges affected by the removal. (iv) Repeat from step (ii) until no edges remain. Here, we are using the same procedure for mitigating the epidemic. There are few practical modifications though:

- Edge-betweenness is calculated for the mean matrix $H_{ab} = \sum_k F_{ab}^{(k)}/K$. The removal of

links is then applied to each matrix separately.

- For a faster computation of step (ii), edges are removed in chunks rather than one-by-one. This is done with the help of the efficient shareware package MatlabBGL that is able to rank all the edges of H in a single instance. The initial size of a chunk is 10000 (about 1 % of the total number of links in H) and it then decreases adaptively.

- The computation terminates when the resulting KL divergence is sufficiently close to zero. Our modified procedure is clearly sub-optimal. We found, however, that compared to the exhaustive search (i)-(iii), the results almost don't change because the variations of the KL divergence at the initial stages are relatively small.

The entropy $S(\text{SafeGraph})$ and the size of epidemic Ω/Ω_0 as a function of the number of removed links, N_c (*i.e.* number of cuts), are shown in Figure S7. The size of epidemic drops abruptly when the relative number of cuts $c \equiv N_c/N/(N-1)$ exceeds the value $c = 0.6$ (from a relative size of 95 % at $c = 0.6$ to 75 % at $c = 0.7$). This drop is accompanied by the increase of entropy around the same value of c. Thus, mitigation based on edge-betweenness performs better than either the distance or the nominal cutoffs, none of which exhibits a similar abrupt crossover. The betweenness index outperforms the other two methods mainly because it is sensitive to the flux that is flowing through links. The "transition" at $c = 0.6$ is expected to become much sharper as $N \gg 1$ approaches the thermodynamic limit.

Supplementary References

4. 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. *Applied Mathematics and Computation* **236,** 184–194 (2014).

18. Postnikov, E. B. & Sokolov, I. M. Continuum description of a contact infection spread in a SIR model. *Mathematical biosciences* **208,** 205–215 (2007).

21. Te Vrugt, M., Bickmann, J. & Wittkowski, R. Effects of social distancing and isolation on epidemic spreading modeled via dynamical density functional theory. *Nature communications* **11,** 1–11 (2020).

38. Naether, U., Postnikov, E. & Sokolov, I. Infection fronts in contact disease spread. *The European Physical Journal B* **65,** 353–359 (2008).

41. Mehta, M. *Random Matrices* 3rd. Sec. 1.5 (Elsevier, 2004).

43. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proceedings of the national academy of sciences* **99,** 7821–7826 (2002).

44. Luo, F., Zhong, J., Yang, Y., Scheuermann, R. H. & Zhou, J. Application of random matrix theory to biological networks. *Physics Letters A* **357,** 420–423 (2006).

45. DeMeo, P., Ferrara, E., Fiumara, G. & Ricciardello, A. A novel measure of edge centrality in social networks. *Jour. Knowledge-Based Systems* **30,** 136–150 (2012).

46. Brody, T. A., Flores, J., French, J. B., Mello, P., Pandey, A. & Wong, S. S. Random-matrix physics: spectrum and strength fluctuations. *Reviews of Modern Physics* **53.** p. 391, 385–479 (1981).

47. Guhr, T., Müeller-Groeling, A. & Weidenmüller, H. A. Random matrix theories in quantum physics: common concepts. *Phys. Rep.* **299.** p. 228, 189–425 (1998).

48. Hakke, F. *Quantum signatures of Chaos* 3rd. Sec. 4.5 (Springer, 2010).