





## Interplay between population density and mobility in determining the spread of epidemics in cities

Surendra Hazarie<sup>1,6</sup>, David Soriano-Paños <sup>2,6</sup>, Alex Arenas <sup>3</sup>, Jesús Gómez-Gardeñes <sup>2,4</sup>✉ & Gourab Ghoshal <sup>1,5</sup>✉

The increasing agglomeration of people in dense urban areas coupled with the existence of efficient modes of transportation connecting such centers, make cities particularly vulnerable to the spread of epidemics. Here we develop a data-driven approach combines with a meta-population modeling to capture the interplay between population density, mobility and epidemic spreading. We study 163 cities, chosen from four different continents, and report a global trend where the epidemic risk induced by human mobility increases consistently in those cities where mobility flows are predominantly between high population density centers. We apply our framework to the spread of SARS-CoV-2 in the United States, providing a plausible explanation for the observed heterogeneity in the spreading process across cities. Based on this insight, we propose realistic mitigation strategies (less severe than lockdowns), based on modifying the mobility in cities. Our results suggest that an optimal control strategy involves an asymmetric policy that restricts flows entering the most vulnerable areas but allowing residents to continue their usual mobility patterns.

<sup>1</sup>Department of Physics & Astronomy, University of Rochester, Rochester, NY, USA. <sup>2</sup>GOTHAM Lab—Department of Condensed Matter Physics and Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain. <sup>3</sup>Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain. <sup>4</sup>Center for Computational Social Science (CCSS), Kobe University, Kobe, Japan. <sup>5</sup>Department of Computer Science, University of Rochester, Rochester, NY, USA. <sup>6</sup>These authors contributed equally: Surendra Hazarie, David Soriano-Paños. ✉email: [gardenes@unizar.es](mailto:gardenes@unizar.es); [gghoshal@pas.rochester.edu](mailto:gghoshal@pas.rochester.edu)

During the last century, humankind has rapidly evolved into an interconnected society driven by the existence of a vast mobility network connecting different areas around the globe. In particular, the striking growth experienced by the international mobility network<sup>1</sup> has helped to bridge socio-cultural<sup>2–4</sup> and economic gaps<sup>5</sup>. Accompanying this is the phenomenon of urbanization, whereby a majority of the world's population resides in densely packed urban centers, with the trend only accelerating<sup>6,7</sup>, given the socioeconomic advantages that cities afford<sup>8,9</sup>. Allayed against these benefits, the increase in mobility has removed the main bottleneck limiting the spatial diffusion of confined epidemic outbreaks. Once the disease spreads to different regions, it takes advantage of the high population density and infrastructure networks in cities<sup>10</sup> to rapidly spread through the local population. As a consequence, over the past few years, several contagious disease outbreaks have emerged, notable among them H1N1 in 2009<sup>11</sup>, Ebola in 2014<sup>12</sup>, Zika virus in 2015<sup>13</sup>, and of course more recently SARS-CoV-2<sup>14,15</sup>. Indeed, the frequency with which these pandemics occur is troublingly increasing<sup>16</sup>.

Despite the different nature of the pathogens, their spreading, both globally and locally, is primarily explained by the sequential combination of case importation from contagion sources, followed by local community transmission converting initially unaffected regions into new endemic areas. Different flavors of mobility play a role in each process: on the one hand, long-haul flights<sup>17</sup> are usually the drivers facilitating the entry of pathogens into a given country, to the extent that the airport mobility network has proved to be a reliable proxy to determine pathogen arrival times<sup>18</sup>, international infection routes<sup>19</sup>, and to complement phylogeographic inference of emerging pathogens such as SARS-CoV-2<sup>20</sup>. On the other hand, once index cases are found within a given region, a complex combination of the local mobility<sup>21</sup> and socioeconomic features of the population<sup>22–24</sup> determines the speed of epidemic spread and the extent of its outbreak.

Quantifying the impact of local mobility on the global diffusion of a pandemic constitutes a challenging task. In this sense, several examples addressing the impact of daily recurrent mobility patterns on the spread of contagious diseases can be found in the literature<sup>25–32</sup>. The majority of these, however, are theoretical frameworks analyzing the features of synthetic mobility networks, and the influence of the total volume of travelers on the course of the epidemic. Nonetheless, recent advances made in data-gathering techniques allow for obtaining accurate representations of daily urban rhythms constructed from mobile phone traces<sup>33</sup>, geolocalized tweets<sup>34</sup>, location-based social networks<sup>35</sup>, or extensive census surveys. These data sets enable the extension of the theoretical machinery to address real epidemic scenarios. Indeed, recurrent mobility patterns have been already useful for identifying the most exposed areas in some epidemic scenarios<sup>36</sup> as well as reproducing the infection routes of H1N1 influenza<sup>37</sup>, Malaria<sup>38</sup>, and more recently SARS-CoV-2<sup>39–42</sup>.

While much attention has been spent on reconstructing past infections, or epidemic forecasting in the case of extant pandemics, an important question that immediately arises, is what makes regions—in particular urban agglomerations where most people reside—vulnerable to the spread of pathogens in the first place? While factors such as population density, levels of healthcare, quality of infrastructure, and socioeconomic disparities play a major role<sup>43</sup>, vulnerability to spread is a complex interplay between these features that is, in general, difficult to disentangle. For instance, the role of population density is an open question with evidence both for and against its influence on epidemic spreading<sup>44,45</sup>. Indeed, merely the density of contacts, while relevant at a neighborhood level, is not enough to explain the mechanisms of spread; one would also need to consider the

mobility network of flows that govern the exchange of people between the regions. In such a setting, the spatial distribution of the population densities and the strength of interaction between the regions become especially relevant.

In other words, it is reasonable to assume that the morphology of the city in terms of how its residents are distributed and how they navigate the city plays a crucial role in their susceptibility to pandemics. Recent studies have shown that the spatial patterns of how residents utilize transportation infrastructure are a strong indicator of that regions' levels of social inclusion, quality of infrastructure, and wealth creation<sup>46</sup>. Bassolas et al.<sup>47</sup> propose a measure of a city's dynamical organization based on mobility hotspots<sup>48</sup> to classify them in a spectrum between compact-hierarchical and sprawled layouts. They find that the extent to which cities are compact or sprawled serves as a low-dimensional proxy for various urban indicators related to the quality of life, health, and pollution.

In this work, we connect the dots between the morphology of human activity in cities, in terms of its associated mobility flows and the distribution of resident populations, and its effect on shaping the transmission of infectious diseases and their associated epidemic outbreaks. We collect data from 163 cities across four continents, on their population density at the zip-code level, and intra-urban mobility flows for the first half of 2020. Using this we extract population density hotspots (i.e., those areas with the highest concentration of residents) and measure the extent to which flows between hotspots dominate the total flows in the city. To capture epidemic spreading, we generalize a movement-interaction-return epidemic model<sup>28</sup> that captures the interplay between recurrent mobility flows and the distributions of resident populations. We derive the epidemic threshold, representing the minimum infectivity per contact required to instigate an epidemic outbreak and connect it with the distribution of flows among population density hotspots. In particular, we show that, despite their ostensible differences in terms of spatial layout, evolutionary history, or levels of infrastructure, all considered cities lie on a common curve capturing an inverse relationship between the epidemic threshold and the extent to which mobility flows are localized between hotspots.

The results suggest an increased susceptibility to epidemics as a function of flows being concentrated between high population centers. As a proof-of-concept, we analyze the current SARS-CoV-2 pandemic by quantifying the epidemic growth from the initial infection curves as an empirical proxy for city vulnerability and plotting it against our calculated epidemic threshold. The empirical trends match our theoretical formalism where cities with mobility concentrated primarily between hotspots are more vulnerable as compared to those with more egalitarian mobility distributions. Based on this observation, we propose a realistic mitigation policy that, being much less severe than draconian lockdowns, lowers the susceptibility of cities that lie in the vulnerable spectrum.

## Results

**Data.** The population density for cities at zip-code resolution was collected from national census bureaus<sup>49–51</sup> and from high-resolution population density estimates recently published by Facebook<sup>52</sup>. Each of these cities corresponds to the largest in their respective countries in terms of population size. From this data, we extract  $H_k$  population density hotspots for each city  $k$  (varying from city to city) by applying a nonparametric method based on the derivative of the Lorenz curve<sup>47,48</sup> (for details of the calculation see Methods).

The mobility flows within each city are sourced from the Google SARS-CoV-2 Aggregated Mobility Research Dataset, and

contain anonymized flows mobility flows aggregated over users who have turned on the Location History setting, which is off by default. The flows are between cells of approximately 5 km<sup>2</sup> for the period ranging November 3, 2019 to February 29, 2020. For the purposes of our analysis, we only consider the period before any mobility mitigation measures were initiated as a response to the SARS-CoV-2 pandemic (see Supplementary Note 1 for further details). The flows are encoded in a matrix  $\mathbf{T}$  whose elements  $T_{ij}$  correspond to the population out-flows from  $i$  to  $j$  and whose diagonal elements correspond to self-flows. For each city  $k$ , given  $H_k$  hotspots, we calculate the hotspot concentration,  $\kappa_k$ , defined as the fraction of total flows in the city that occur only between hotspots thus,

$$\kappa_k = \frac{\sum_{i,j \in H_k} T_{ij}}{\sum_{i,j} T_{ij}}. \quad (1)$$

This metric lies in the range  $0 \leq \kappa_k \leq 1$ , with the limiting cases corresponding to flows exclusively between hotspots or only between hotspot and non-hotspot areas. The list of cities for each country, the administrative unit, and the hotspot concentration is provided in Supplementary Data 1. The results show a wide spectrum of values ( $0.05 \leq \kappa_k \leq 0.79$ ) both within and between countries, indicating significant variability in cities in terms of the morphology of human flows between population centers.

In Fig. 1, we plot the spatial layout of the hotspots and the mobile network for six representative cities in Australia and the United States. The cities are organized in descending order of  $\kappa$ , and it is apparent that in those cities with high  $\kappa$ , flows are mainly confined around hotspots (marked in red), whereas they are increasingly more distributed with decreasing  $\kappa$ . An additional feature is that hotspots are more spatially concentrated in those cities with lower  $\kappa$  and dispersed in those with higher  $\kappa$ , indicating a more heterogeneous distribution of population density in the former as compared to the latter.

**Model.** To characterize a city's vulnerability to disease spread, we calculate the epidemic threshold by generalizing the formalism introduced in<sup>28</sup>. In what is to follow, we incorporate susceptible

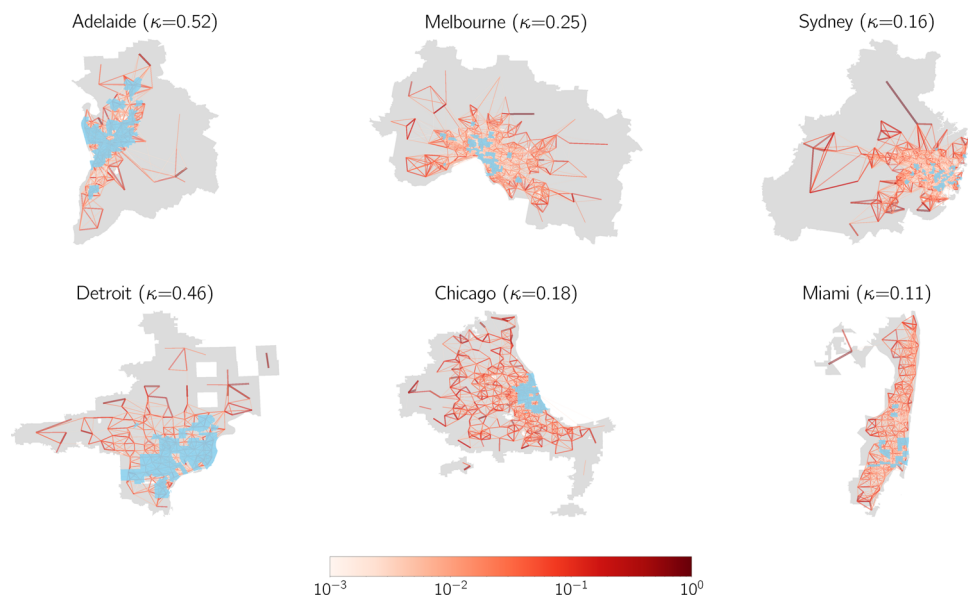
(S), infectious (I), or recovered (R) dynamics, however, we note that the formalism is easily generalizable to more elaborate compartmental schemes. An infectious individual transmits the pathogen to healthy counterparts via direct interaction at a rate  $\lambda$ . In turn, infectious individuals enter the compartment  $R$  at a rate  $\mu$ , which typically encodes the inverse of the expected contagious period. The mixing among healthy and infectious individuals is governed by the spatial distribution of the population and their mobility patterns, which are accommodated in the formalism by using metapopulations. A metapopulation is a complex network where nodes (patches) represent locations where agents are placed and movements between patches are represented by links connecting them.

In our case, the different zip-codes of the city are represented as patches  $i$ , which are initially populated by  $n_i$  residents. The activity of the residents is considered on a daily basis and split into three following stages: Movement, Interaction, and Return. First, agents decide whether or not to move with a probability  $p$ . If moving, they choose their destination according to the origin-destination matrix  $\mathbf{R}$ , being the elements  $R_{ij}$  the probability of moving from patch  $i$  to patch  $j$ . We construct this matrix from mobility data, encoded in  $\mathbf{T}$ , as

$$R_{ij} = \frac{T_{ij}}{\sum_l T_{il}}. \quad (2)$$

After all movements (or lack thereof) have been completed, interactions occur within patches according to a mean-field assumption where every individual makes the same number of contacts proportional to the population density via a function  $f_i$ . Finally, as we want to reflect the most commuting nature of human mobility, we force all the individuals to return to their residence and repeat the same process for a new time step (day).

Under these assumptions, the dynamics are totally characterized by a set of  $2 \times N$  coupled discrete equations governing the temporal evolution of the fraction of infected and recovered individuals with residence in each patch. In particular, the fraction of infected,  $\rho_i^I(t+1)$ , and recovered individuals,



**Fig. 1 Spatial representation of the urban mobility network for selected cities in Australia and the United States.** Areas marked blue correspond to population density hotspots. The cities are organized in descending order according to the extent to which mobility flows are concentrated between areas of high population density,  $\kappa$  (Eq. (1)). Line color encodes the number of inhabitants following each route normalized by the highest flow observed within each city.

$\rho_i^R(t+1)$ , associated to patch  $i$  at time  $t+1$ , read

$$\rho_i^I(t+1) = (1-\mu)\rho_i^I(t) + (1-\rho_i^I(t) - \rho_i^R(t))\Pi_i(t), \quad (3)$$

$$\rho_i^R(t+1) = \rho_i^R(t) + \mu\rho_i^I(t). \quad (4)$$

Eq. (4) determines the evolution of recovered patients. As the SIR model assumes that the compartment  $R$  constitutes the final epidemiological state, this evolution is just given by the number of infected individuals overcoming the disease. Regarding the evolution of infectious individuals, the r.h.s. of Eq. (3) corresponds to those infected overcoming the disease and the second one involves contagions of susceptible individuals. In this sense, the probability that a susceptible individual living inside  $i$  contracts the disease at time  $t$ ,  $\Pi_i(t)$ , can be expressed as

$$\Pi_i(t) = (1-p)P_i(t) + p \sum_{j=1}^N R_{ij}P_j(t). \quad (5)$$

The first term identifies those contagions occurring inside the residence patch whereas the second term contains those taking place inside neighboring areas. Likewise, the probability of contracting the disease inside a given node  $i$  at time  $t$ ,  $P_i(t)$ , is given by

$$P_i(t) = 1 - \left(1 - \lambda \frac{I_i^{\text{eff}}(t)}{n_i^{\text{eff}}}\right)^{f_i}, \quad (6)$$

where  $f_i$  determines the number of contacts per day of individuals inside  $i$ . The explicit forms assumed for the function  $f$  are explained in the Methods sections. Finally, the terms  $n_i^{\text{eff}}$  and  $I_i^{\text{eff}}(t)$ , which denote the effective population and effective number of infected individuals inside patch  $i$  at time  $t$  after population movements, read

$$n_i^{\text{eff}} = \sum_{j=1}^N n_j (\delta_{ij}(1-p) + R_{ji}), \quad (7)$$

$$I_i^{\text{eff}} = \sum_{j=1}^N n_j \rho_j^I(t) (\delta_{ij}(1-p) + R_{ji}). \quad (8)$$

Eq. (8) implies that infected individuals move analogously to the susceptible population regardless of their epidemiological condition. To account for possible mobility restrictions of symptomatic agents with respect to asymptomatic or mildly ill patients, one must introduce more complex compartmental schemes such as the ones tailored for specific diseases<sup>14,36</sup>. Nevertheless, the results presented here on the interplay between city morphology and epidemic spreading are general enough to remain valid when considering more elaborate models.

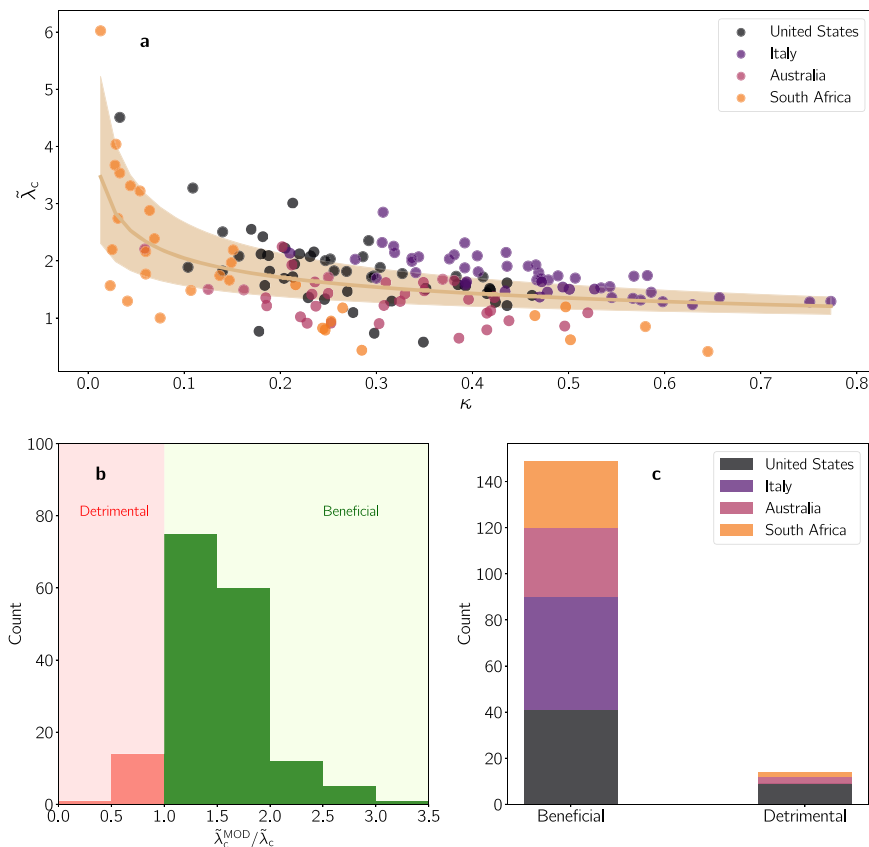
**Connecting the epidemic threshold to hotspot flows.** Given the quantitative description of human flows composing the mobility backbone of each city, we now focus on determining the effect that their morphology has on the epidemic threshold, defined as the minimum infectivity required to observe a macroscopic outbreak. Specifically, to get a fair comparison of the effects of mobility in cities with disparate average population densities, we use the normalized threshold  $\tilde{\lambda}_c = \lambda_c(p=1)/\lambda_c(p=0)$ , as detailed in the “Methods” section. A complete analysis of the dependence of the normalized threshold for mobility values different from  $p=1$  is included in Supplementary Note 2. In Fig. 2a we plot  $\tilde{\lambda}_c$  as a function of  $\kappa$  for all the cities in our dataset, represented as filled circles colored according to their respective countries. We find a clear trend, whereby all cities fall into a curve marking an inverse relationship between the vulnerability of the cities and the extent to which flows are concentrated between

hotspots. A fit to the data yields a relation  $\tilde{\lambda}_c \sim \kappa^\beta$  with  $\beta \approx -0.25$ . This exponent may vary when studying each country individually because of the different spatial resolutions here analyzed. We characterize this phenomenon in Supplementary Note 3 and Supplementary Table 1. For the sake of completeness, in Supplementary Fig. 2, we show the equivalent of Fig. 2a but now split by country finding the same trend attesting to the robustness of the inverse dependence. In particular, for cities where there is a more egalitarian distribution of flows (lower  $\kappa$ ), the epidemic threshold is higher for a mobile population as compared to a static population, indicating that movement between regions lowers the risk of an epidemic outbreak. Conversely, in those cities where the population moves primarily between hotspots, there is a little-to-no difference in risk in terms of whether residents stay in their patches, or whether they move to different ones. As a comparison to other population-related measures, we analyze the relation between different metrics such as  $\kappa$ , the average population density, the Lloyd mean crowding<sup>53</sup> and  $\tilde{\lambda}_c$  for the 50 American cities in Supplementary Note 4. In particular, we represent in Supplementary Fig. 3, the Spearman correlation between each pair of variables, proving that the concentration of flows between areas of high population density,  $\kappa$ , is a much better predictor of the vulnerability of cities than the other two measures.

To further characterize the relation existing between cities’ vulnerability and the concentration of mobility between density hotspots, we next analyze the impact of reshuffling the flows at a local scale for each of the analyzed cities. In particular, we preserve the total amount of flows, by removing all links connecting hotspots (thus setting  $\kappa=0$ ) and redistributing them across non-hotspot locations, respecting the relative proportion for the flows from hotspots to non-hotspots areas, as dictated by the original  $\mathbf{R}$  matrix.

To quantify the effect of such intervention, we represent in Fig. 2b the distribution of the values for the ratio  $\tilde{\lambda}_c^{\text{MOD}}/\tilde{\lambda}_c$ , where the numerator is the re-calculated threshold after the intervention. As the figure indicates for the vast majority of cities (irrespective of their original value of  $\kappa$ ), the effect of switching off the flows between hotspots leads to an increase in the epidemic threshold, in turn lowering their vulnerability to epidemic spread. In particular, we split the cities into two categories as a function of the outcome of the reshuffling process: *beneficial*, those where the threshold is increased greater and *detrimental*, those where the threshold is instead lowered. The number of cities belonging to each category is plotted as a bar-chart in Fig. 2c, indicating that around 91% of cities experience a beneficial effect while the remaining 9% experience an increased susceptibility (interestingly this category is dominated by American cities). For this small number of cities, listed in Supplementary Table 2, it is likely that there are other more complex features at play not considered in this analysis.

Nevertheless, these results provide evidence that, in most cases, the concentration of human mobility between densely populated areas is a feature that enhances disease spreading and makes such cities vulnerable to epidemics. Moreover, the beneficial effect caused by the reorientation of intra-hotspot flows toward less densely populated areas seems to be rooted in a homogenization of the distribution of the underlying density, which enforces infected individuals to stay away from the contagion focus, thus reducing their infection power<sup>29,32</sup>. In turn, this homogenizing flow structure appears naturally in cities with low  $\kappa$  and, as suggested by the empirical trends in Fig. 2a, characterizes the most resilient cities. Therefore, a lower  $\kappa$  translates into a greater mix of populations, between high and low population density centers, where they can actually take advantage of mobility between city sub-regions to prevent outbreaks.



**Fig. 2 Connecting morphology to vulnerability.** **a** Normalized epidemic threshold  $\tilde{\lambda}_c$  vs. population density hotspot concentration  $\kappa$  for cities in each of the chosen four countries (color code). The Spearman correlation coefficient is  $r_s = -0.45$  ( $p$ -value  $< 10^{-9}$ ), showing a moderate but statistically significant correlation. The solid line shows the result of fitting the data to a power-law function  $\tilde{\lambda}_c \sim \kappa^\beta$ , obtaining  $\beta \approx -0.25$  via least squares regression. The shadowed region covers 68.2% confidence intervals for the estimated parameters. **b** Histogram of the impact of reshuffling the flows connecting hotspots among their non-hotspots neighboring areas on the epidemic threshold, measured by the ratio  $\tilde{\lambda}_c^{MOD} / \tilde{\lambda}_c$ , where the numerator represents the recomputed threshold after reshuffling. The green-colored (red-colored) zone corresponds with a beneficial (detrimental) effect of such intervention translated into an increase (decrease) of the epidemic threshold. **c** Bar plot containing the count of cities within each category according to the outcome of the removal of flows connecting hotspots for each country here analyzed.

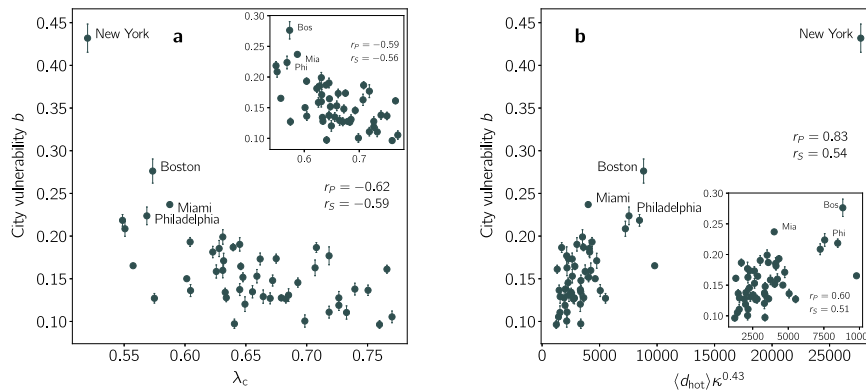
**Application to real pandemic settings.** The formalism proposed here can be readily applied to assess the exposure of cities to actual outbreaks. To illustrate this, we next focus on the spread of SARS-CoV-2 in the 50 most populated core-based-statistical-areas (CBSA) in the United States, chosen due to the appropriate spatial resolution in terms of infection data. Note that, although thus far, we have focused on the SIR model, the following will illustrate the generality of the results, in the context of the spread of SARS-CoV-2, which has been recently analyzed with more elaborate compartmental models<sup>39,53–56</sup>.

As a proxy for a city’s vulnerability to epidemic spread, we make use of the number of confirmed infected cases at the county level collected from the New York Times (<https://github.com/nytimes/covid-19-data>) and USAFacts (<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>). Given the inherent noise due to reporting artifacts, and assuming an exponential growth,  $I_k(t) \sim \exp(b_k t)$  during early onset, we apply a smoothing procedure to extract the growth rate  $b_k$  of the number of infected cases and use that as a proxy for a city  $k$ ’s susceptibility to disease spread. To remove any effects due to non-pharmaceutical interventions and behavioral changes in the population, we focus on the period before mitigation measures. The full details of the procedure explained in the Methods section, and the temporal infection plots for each city along with the fits are shown in Supplementary Fig. 4.

In order to properly connect with the growth rate, we need to reintroduce the effect of the population distribution and, therefore, the relevant variable is the epidemic threshold  $\lambda_c(p = 1)$ . Note that, by choosing  $p = 1$ , we force all the inhabitants within a city to follow the flow matrix  $T$ , but not all of them leave their residential area due to the existence of self-loops in this matrix. Indeed, according to the available data for the cities analyzed here, around 36% of the population remains on average inside their residential administrative unit. In Fig. 3a we plot the empirically extracted growth rate  $b$  as a function of the epidemic threshold,  $\lambda_c$ , finding once again an inverse trend, confirming the role of the threshold as a proxy for vulnerability. Those cities which experienced a faster epidemic growth during the early onset of the pandemic indeed had a lower threshold according to our formalism.

Next we connect, the localization of flows to hotspots to the empirical vulnerability in each of the cities. Since  $\kappa$  only takes into account flows between hotspots but does not account for their population density, the variable that captures the effective interaction between the residents in hotspots areas should be a combination of both factors. Specifically, we have chosen  $\langle d_{hot} \rangle \kappa^{0.43}$ , where the first term corresponds to the average population density within hotspots and the second term reflects the scaling obtained for the normalized epidemic threshold in the individual case of CBSA from the United States (see



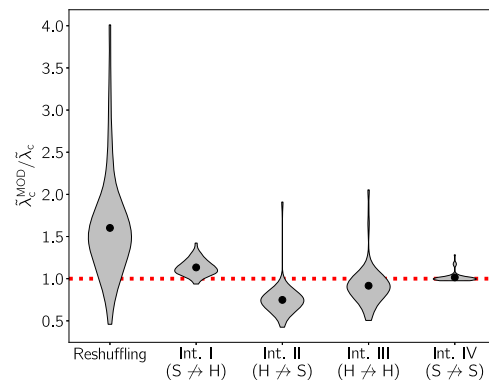


**Fig. 3 Validation of the model in the United States.** **a** Epidemic growth rate  $b$  as a function of the epidemic threshold  $\lambda_c$  computed according to Eq. (17). In the inset, we show the same plot after removing NYC. **b** Epidemic growth rate  $b$  (estimated from fitting to empirical infection data, see Methods) as a function of the average population density within hotspots  $\langle d_{hot} \rangle$ , measured in inhabitants per square mile, and the hotspot concentration  $\kappa$ . The exponent 0.43 is introduced to reflect the dependence obtained when fitting the normalized epidemic threshold to  $\kappa$  for the United States (Table S2). In both panels,  $r_p$  and  $r_s$  denote the Pearson and Spearman rank correlation coefficients among the represented quantities. Error bars correspond to the 95% confidence interval.

Supplementary Note 3 for further details). In Fig. 3b we plot  $b$  as a function of this quantity finding a clear monotonically increasing trend.

We note that the cities with the highest values of  $b$  are large transportation hubs with major airports (New York, Miami, and Boston for instance). The higher infection rates do not necessarily reflect the fact that these cities were exposed to international (domestic) travelers that seeded the infection. To check to what extent  $b$  is explained by  $\kappa$  controlling for the large inflows in such cities, we conduct a PCA analysis of  $\kappa$ ,  $b$ , the inflow per-capita, defined as the total number of international flows arriving in a given city, as well as the population density. In Supplementary Fig. 5, we plot the two most important components accounting for 63% (PC1) and 24% (PC2) of the variance in the data. We see that the quantity  $\kappa$  is roughly orthogonal to the inflow per capita, and contains unique information about the epidemic spread, quite apart from inter-city travel. Thus taken together, the results of Fig. 3 indicate that the empirical trends mirror our theoretical formalism, whereby cities that experience strong early growth of the epidemic have a lower threshold, a phenomenon for which one of the main causative mechanisms is that the movement in cities occurs primarily between hotspots.

**Potential mitigation measures.** The observations thus far, immediately suggest the possibility of effective mitigation measures that may shore up the robustness of vulnerable cities to the onset of epidemic spread. Before the availability of proven vaccines (and the continued limited efficacy of proven therapeutics) for SARS-CoV-2, the predominant global strategy was the adoption of non-pharmaceutical interventions of which a key ingredient has been aggressive lock-downs. While ostensibly being very effective in mitigating an active epidemic, significant disruption to the socio-economic fabric is one of the unfortunate consequences<sup>57,58</sup>. Having demonstrated the key role played by the interactions between population density hotspots, we next investigate some targeted interventions, or even preemptions, that are milder than completely restricting mobility city-wide and assess their efficacy in reducing vulnerability. The strategy we pursue is to modify flows between different types of locations in the city without the need to isolate individuals at home. In the first intervention (Intervention I), an asymmetrical strategy involves restricting flows from non-hotspot (heretofore referred to as suburbs) towards hotspot areas and converting them to self-loops, while keeping all other flows the same. Intervention II



**Fig. 4 Impact of mobility interventions on the epidemic threshold for cities in the United States.** The impact is quantified with violin plots representing the distribution for the ratio between the normalized threshold after each intervention,  $\tilde{\lambda}_c^{MOD}$  and the normalized threshold of the original mobility network  $\tilde{\lambda}_c$ . The dots inside each violin plot shows the mean of each distribution whereas the red dotted line shows the expected value of this ratio in absence of intervention. Here, S and H refer to suburbs and hotspots respectively, and  $S \nrightarrow H$  represents the restriction of flow from suburbs to hotspots while keeping all other flows the same.

corresponds to the reverse situation where flows from hotspots to suburbs are converted to self-loops. Finally, Interventions III and IV reduce the connections among hotspots and suburbs respectively.

To quantify the impact of each intervention, we represent in Fig. 4 the distribution of the ratios between the normalized epidemic threshold after the intervention  $\tilde{\lambda}_c^{MOD}$  and the original normalized threshold  $\tilde{\lambda}_c$  for each of the cities in the US. As a control, we also include the (unrealistic) case of reshuffling all flows according to the procedure described in Fig. 2. We find that Intervention I has the effect of increasing the epidemic threshold, thus enhancing robustness to spread. Given that a potential epidemic has a high probability of being seeded and correspondingly spreading extensively in high population density areas, preventing the residents in suburbs from visiting these locations protects them from being exposed to the disease. Conversely, restricting residents in hotspots to travel to suburbs has the opposite effect in further decreasing the threshold. This

counterproductive effect emerges due to a phenomenon discussed in<sup>29</sup>. When there is an asymmetry between the population density in cities, mobility from hotspots to suburbs leads to an increase in the threshold due to the dilution of the effective population in the hotspots thus reducing the number of contacts, as well as diverting potentially infectious individuals to lower-density regions where their impact is mitigated. Removing this route makes the situation significantly worse. Therefore, the spread of diseases across a given city is explained by a combination of the creation of contagion sources due to the movement of individuals between hotspots (governed by  $\kappa$ ) with the spatial dissemination of cases prompted by those moving from less populated areas to these sources. Finally, limiting both the mobility between hotspots or between suburbs has a mostly neutral effect, although the trends are noisier given that its effectiveness is closely related to the underlying population density distribution inside them. The results suggest that, among the scenarios outlined, the most beneficial policy is to restrict residents in suburbs from visiting hotspots, while at the same time allowing residents in hotspots to continue with their regular mobility behavior. Note, that several other combinations are possible, for instance, a combination of interventions I and III which is likely to have even more of a beneficial effect.

## Discussion

Similar to how a virus enters the human body and replicates from cell to cell, the spread of pathogens in susceptible populations is influenced by the interaction between its hosts. Thus, it is the social interactions, mediated by behavioral and mixing patterns, that shape the spread of disease in human populations. Among these aspects, human mobility is a key factor underlying the unfolding patterns of epidemic outbreaks. Understanding how human mobility shapes the spatiotemporal unfolding of contagious diseases is essential for the design of efficient containment policies to ameliorate their impact. In this paper, we investigate the interplay between the population density and the spatial distribution of flows in urban areas, and its impact on determining exposure to epidemic outbreaks. We report a clear trend worldwide: the existence of a high volume of individuals commuting among the population density centers of a given city makes it more vulnerable to the spread of epidemics. The extent of a city's vulnerability as determined from our formalism, allows us to shed light on a real epidemic scenario: the spread of SARS-CoV-2 across the United States. In particular, the epidemic threshold determined by the population density profiles and urban mobility patterns provides one of the potential mechanisms behind the different levels of infection observed across cities in the United States. Of course, many other factors such as the socioeconomic differences between regions, political affiliations, and behavioral traits play a key role, although these factors are outside the scope of this study.

Controlling for these other factors, as an application to potential epidemic waves, our indicator allows for identifying those cities that are likely to become epidemic centers once the first imported cases arrive there. This is of importance, given that it can guide authorities to identify places where timely containment policies can be locally implemented to avoid large outbreaks caused by massive community transmission. It is precisely the lack of anticipation of the SARS-CoV-2 pandemic, that has led countries to enforce aggressive containment measures aimed at ameliorating the impact of the disease. The predominant strategy has been the implementation of lockdown policies, forcing a large fraction of the population to stay isolated at home, thus reducing considerably their number of interactions. While there is consensus on the effectiveness of these interventions to mitigate an

ongoing outbreak, the collateral socio-economic damage caused by lockdowns requires a change of direction toward less aggressive containment measures.

In the case of SARS-CoV-2, the strict individual isolation characterizing the first interventions has given rise to more relaxed lockdown scenarios combined with efficient Test-Track and Isolate policies<sup>59-61</sup>. Along these lines, we present different scenarios based on modifying mobility habits to actively avoid the emergence of large areas of contagion. Our analysis suggest that an effective policy involves an asymmetric closure of neuralgic centers of the cities, restricting movement to population density hotspots from residents of other areas, while allowing those living in hotspots to commute, in order to dilute the number of contacts in the most vulnerable areas. Teleworking and effective distribution of key services in a city are practical manifestations of such interventions. Of course, one must take care in executing such a policy, keeping in mind for instance, that while it might advantage wealthier populations that typically live in suburbs, it might inadvertently leave relatively less affluent (the urban poor) at a disadvantage.

**Limitations.** These results should be interpreted in light of several important limitations. First, the Google mobility data are limited to smartphone users who have opted into the Google Location History feature, which is off by default. These data may not be representative of the population as a whole, and furthermore, their representativeness may vary by location. Importantly, these limited data are only viewed through the lens of differential privacy algorithms, specifically designed to protect user anonymity and obscure fine detail. Moreover, comparisons across rather than within locations are only descriptive since these regions can differ in substantial ways.

Despite the generality of these findings, the dimensionality reduction here proposed by coarse-graining the complex nature of the mobility flows into a single indicator  $\kappa$  sometimes fails in capturing the behavior of the epidemic threshold. The clearest example is the set of few cities for which removing the connections between hotspots accelerates epidemic spreading. To improve our results, a more exhaustive formalism taking into account, not only the differences between hotspots and suburbs but also the hierarchies existing within each category, as defined in Bassolas et al.<sup>47</sup>, can be considered. Likewise, these results rely on the assumption that population density centers are much more vulnerable to contagious diseases than scarcely dense areas. While this is arguably a logical assumption, the beneficial effect of mobility from hotspots to suburbs could be reversed for diseases with large reproduction numbers  $R_0$ . In this scenario, suburbs would also have the potential to develop large outbreaks so the existence of infection routes across the city would lead to an acceleration of the propagation of the epidemic front.

Nevertheless, an advantage of our formalism is its relative simplicity, paving the way to extend the results to more general scenarios. For example, the homogeneous assumption concerning the contacts of individuals within each subpopulation can be improved to account for the heterogeneous nature of human contacts<sup>62,63</sup>. In addition, we have restricted our study to the morphology of urban flows by neglecting the contribution coming from movements between different urban areas. Therefore, another obvious extension is to consider movement at different geographical scales<sup>64,65</sup> along with the typically longer time scales associated with them<sup>66</sup>. Accounting for movement between cities, for instance, could provide valuable information to coordinate joint efforts among different regions to modify both inter- and intra-urban flows in service of reducing the impact of a pandemic. Needless to say, pandemics are complex processes

involving a multitude of spatial and socioeconomic factors. The results presented here provide one of the building blocks for policy planners in devising effective preventive and mitigation measures for future crises.

**Methods**

**Hotspot classification.** Hotspots are identified by setting a threshold on the population densities of cells within a city. The threshold for hotspots is assigned by applying a non-parametric method, the LouBar method<sup>45,46</sup>, based on the derivative of the Lorenz curve. The Lorenz curve is the sorted cumulative distribution of population densities and is obtained by plotting, in ascending order, the normalized cumulative number of nodes vs. the normalized cumulative population density. The threshold is then obtained by taking the derivative of the Lorenz curve at (1, 1) and extrapolating it to the point at which it intersects the x-axis. We classify hotspots in cities according to this LouBar method applied to population densities, in agreement with the reliance of the model on effective densities. A cell  $i$  is considered a hotspot of city  $k$ ,  $H_k$  if it satisfies:

$$d_{i,k} > d_k^{\text{Lou}} \tag{9}$$

where  $d_i$  is the population density of cell  $i$  in city  $k$  and  $d_k^{\text{Lou}}$  is the threshold determined by performing the LouBar method on the population densities of all cells in city  $k$ . This allows us to place emphasis on zones within cities that encourage the most relative interaction, as opposed to sharp biasing due to population magnitude. We then examine the hotspots flow concentration in each city  $k$ ,  $\kappa_k$ , defined as the fraction of total flows in the city system that exist between these population density hotspots of city. Therefore,  $\kappa_k$  is given by

$$\kappa_k = \frac{\sum_{i,j \in H_k} T_{ij}}{\sum_{ij} T_{ij}}, \tag{10}$$

where  $T_{ij}$  denotes the flow of individuals going from patch  $i$  to patch  $j$  according to the mobility data.

**Estimating theoretical city vulnerability.** Eqs. (3)–(8) can be used to estimating the vulnerability to spread, for a given city. For this purpose, we study the epidemic threshold defined as the minimum infectivity per contact needed to observe an epidemic outbreak. Therefore, the lower the epidemic threshold is, the easier an epidemic wave propagates, thus reflecting a higher city vulnerability to epidemic outbreaks. To estimate the epidemic threshold, we recall that in the SIR model, a super-critical outbreak grows exponentially following the initial infectious seed(s), eventually covering a macroscopic fraction of the population. In contrast, in the sub-critical regime, the number of infected individuals decreases in time. Here, we must consider the case in which the number of initially infected individuals remains constant. Considering the size of the initial infectious seed to be negligible compared to the overall population size, we can assume that

$\vec{\rho}^I(1) = \vec{\rho}^I(0) = \vec{e} \ll \vec{1}$ . Furthermore, at the beginning of the outbreak, we assume that there are no individuals belonging to compartment  $R$  by setting  $\vec{\rho}^R(t) = \vec{0}$ . Both assumptions allow us to linearize Eq. (6) which now reads

$$P_i \simeq \lambda f_i \frac{I_i^{\text{eff}}}{n_i^{\text{eff}}}. \tag{11}$$

After introducing Eqs. (5)–(11) into Eq. (3) and taking into account the stationary regime, we obtain

$$\frac{\mu}{\lambda} \epsilon_i = \sum_{j=1}^N \underbrace{\left[ (1-p)^2 \delta_{ij} \frac{f_j}{n_j^{\text{eff}}} + p(1-p) \left( R_{ij} \frac{f_j}{n_j^{\text{eff}}} + R_{ji} \frac{f_i}{n_i^{\text{eff}}} \right) + p^2 \sum_l R_{il} R_{jl} \frac{f_l}{n_l^{\text{eff}}} \right]}_{M_{ij}} n_j \epsilon_j. \tag{12}$$

The former expression holds if  $\frac{\mu}{\lambda}$  corresponds to an eigenvalue of matrix  $\mathbf{M}$ . As our goal is obtaining the minimum  $\lambda$  value triggering epidemic outbreaks, the epidemic threshold  $\lambda_c$  is given by

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M})}, \tag{13}$$

with  $\Lambda_{\max}(\mathbf{M})$  denoting the spectral radius of matrix  $\mathbf{M}$ .

If  $p = 0$ , then the threshold would correspond with that associated with a static population that never moves, whereas if  $p = 1$  then this accounts for a fully active population which moves according to the commuting patterns encoded in the matrix  $R$ . Note that, in general, the elements of the matrix  $M$  determining the threshold are a combination of the average population density inside a city and how the population moves across the city. To get a fair comparison about the contribution of mobility in two cities with disparate population densities, we define the normalized epidemic threshold  $\tilde{\lambda}_c = \lambda_c(p = 1) / \lambda_c(p = 0)$ , which focuses on how mobility alters the population distribution with respect to the static case. In this sense, unlike the epidemic threshold, the dependence of the normalized

epidemic threshold on mobility is equal in two cities with the same OD matrices and density profiles, despite possible differences in the average population density in both cases.

**Estimating empirical city vulnerability.** We quantify the extent to which COVID-19 is able to spread in a US city by examining the time series of confirmed cases per county, from January 23 2020 to April 16 2020 (before truncating.) We aggregate this data to the level of CBSAs by summing across each CBSAs component counties. Given the noise in the data (due to collection and reporting artifacts, etc) we perform preprocessing on the curves. We use a Savitzky–Golay Filter<sup>67</sup> to smooth the data by fitting intermediate windows<sup>68</sup> with low-order polynomials. We then truncate our data to a window of two weeks after 100 cases were confirmed in each county. This allows our window of observation to capture the regime where COVID-19 awareness encouraged active testing, but before intervention methods influence how the disease propagates within cities. This way, we capture the disease behavior specific to the city structure, and not external suppression. To estimate the vulnerability of each city, we fit the filtered cumulative number of infections to an exponential function

$$I(t) = ae^{bt} \tag{14}$$

and extract the growth rate  $b$ . Although more sophisticated approaches have been proposed in the literature based on the estimation of the effective reproductive number  $R_e$ , our procedure is simple but effective to capture the vulnerability of each city at the early stage of the outbreak.

**Function governing contacts.** Let us discuss the form of the function  $f$ , determining how the number of contacts that each individual makes inside each patch depends on its density. To shed light on the role of mobility inside each city, we follow a nonparametric approach by assuming that these contacts are linearly proportional to the density inside each patch. This way, the results shown in Figs. 2 and 4 are obtained by assuming

$$f_i = \frac{n_i^{\text{eff}}}{a_i}, \tag{15}$$

being  $a_i$  the area of patch  $i$ .

Although useful for illustrating the role of mobility in each city, the nonparametric linear relation does not correspond with a realistic scenario due to the large difference in terms of contacts existing among zones with disparate densities. To make a more fair comparison of the expansion of COVID-19 over different cities, we choose a more complex function governing the number of contacts. Following<sup>69</sup>, we assume that

$$f_i = 2 - e^{-\xi \frac{n_i^{\text{eff}}}{a_i}}, \tag{16}$$

which is bounded such that  $f_i \in [1, 2)$ . Here the parameter  $\xi$  is estimated by maximizing the correlation among the theoretical and the observed vulnerabilities, yielding  $\xi = 2 \times 10^{-4}$  square miles per individual. By including this function, we estimate the epidemic threshold for each city,  $\lambda_c$ , as

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M}|_{p=1})}, \tag{17}$$

where we set  $\mu = 1$  for the sake of simplicity. Note that the latter parameter does not have any influence on the cities' ranking since it is inherent to the disease features and does not depend on human interactions.

To check the robustness of the results presented in the manuscript, We show in Supplementary Fig. 6 that the more realistic function provided by Eq. (16) yields a better estimation for the vulnerability of the city while not losing the inverse trend relating to the normalized epidemic threshold  $\tilde{\lambda}_c$  and the flow between population hotspots  $\kappa$ .

**Data availability**

The data for this work was taken from the following sources: the mobility flow data are taken from the Google SARS-CoV-2 Aggregated Mobility Research Dataset, the population density data are taken from national census bureaus (<https://blog.splitwise.com/2014/01/06/free-us-population-density-and-unemployment-rate-by-zip-code/>, <https://www.abs.gov.au/>, <http://www.statssa.gov.za/>) and estimates recently published by Facebook (<https://data.humdata.org/organization/facebook/>.) The empirical COVID-19 infection data was taken from data collected by the New York Times (<https://github.com/nytimes/covid-19-data>) and USAFacts (<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>.) The rest of the data may be made available, upon request to the authors.

**Code availability**

The code for the analysis was programmed using standard libraries in C and Python. All the calculations can be reproduced with the equations provided in the main text or the Supplementary Information. Even so, the code used here is available upon request to the authors.



Received: 3 May 2021; Accepted: 16 July 2021;

Published online: 23 August 2021

## References

- The world of air transport in 2018. Technical Report, International Civil Aviation Organization (2018).
- Varghese, N. *Globalization of Higher Education and Cross-Border Student Mobility* (Citeseer, 2008).
- Filatotchev, I., Liu, X., Lu, J. & Wright, M. Knowledge spillovers through human mobility across national borders: evidence from Zhongguancun Science Park in China. *Res. Policy* **40**, 453–462 (2011).
- Williams, A. M. & Baláz, V. What human capital, which migrants? returned skilled migration to Slovakia from the UK 1. *Int. Migr. Rev.* **39**, 439–468 (2005).
- Boubtane, E., Dumont, J.-C. & Rault, C. Immigration and economic growth in the OECD countries 1986–2006. *Oxf. Econ. Pap.* **68**, 340–360 (2016).
- United Nations, Department of Economic and Social Affairs, Population Division (2019). (World Urbanization Prospects 2018: Highlights, ST/ESA/SER.A/421, accessed 30 January 2019). <https://population.un.org/wup/Publications/Files/WUP2018-Highlights.pdf>.
- Le Néchet, F. Urban spatial structure, daily mobility and energy consumption: a study of 34 European cities. *Cybergeo* **580** <https://journals.openedition.org/cybergeo/24966> (2012).
- Bettencourt, L. M. The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
- Pan, W., Ghoshal, G., Krumme, C., Cebrian, M. & Pentland, A. Urban characteristics attributable to density-driven tie formation. *Nat. Commun.* **4**, 1961 (2013).
- Kirkley, A., Barbosa, H., Barthelemy, M. & Ghoshal, G. From the betweenness centrality in street networks to structural invariants in random planar graphs. *Nat. Commun.* **9**, 2501 (2018).
- Van Kerkhove, M. D. et al. Epidemiologic and virologic assessment of the 2009 influenza A (H1N1) pandemic on selected temperate countries in the southern hemisphere: Argentina, Australia, Chile, New Zealand and South Africa. *Influenza Other Respir. Viruses* **5**, e487–e498 (2011).
- Gomes, M. F. et al. Assessing the international spreading risk associated with the 2014 West African ebola outbreak. *PLoS Currents* **6** <https://ecurrents.outbreaks.cd818f63d40e24aef769dda7df9e0da5> (2014).
- Zhang, Q. et al. Spread of zika virus in the Americas. *Proc. Natl Acad. Sci. USA* **114**, E4334–E4343 (2017).
- Estrada, E. Covid-19 and SARS-Cov-2. modeling the present, looking at the future. *Phys. Rep.* **869**, 1–51 (2020).
- Organization, W. H. et al. Coronavirus Disease 2019 (Covid-19): Situation Report, 72. Technical Report (2020).
- Saunders-Hastings, P. R. & Krewski, D. Reviewing the history of pandemic influenza: understanding patterns of emergence and transmission. *Pathogens* **5**, 66 (2016).
- Bowen Jr, J. T. & Laroe, C. Airline networks and the international diffusion of severe acute respiratory syndrome (SARS). *Geographical J.* **172**, 130–144 (2006).
- Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337 LP–1342 (2013).
- Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study. *BMC Med.* **5**, 34 (2007).
- Lemey, P. et al. Accommodating individual travel history and unsampled diversity in bayesian phylogeographic inference of SARS-Cov-2. *Nat. Commun.* **11**, 5110 (2020).
- Barbosa, H. et al. Human mobility: models and applications. *Phys. Rep.* **734**, 1–74 (2018).
- Bonaccorsi, G. et al. Economic and social consequences of human mobility restrictions under covid-19. *Proc. Natl Acad. Sci. USA* **117**, 15530–15535 (2020).
- Ahmed, F., Ahmed, N., Pissarides, C. & Stiglitz, J. Why inequality could spread covid-19. *Lancet Public Health* **5**, e240 (2020).
- Barbosa, H. et al. Uncovering the socioeconomic facets of human mobility. *Sci. Rep.* **11**, 8616 (2021).
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
- Belik, V., Geisel, T. & Brockmann, D. Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* **1**, 011001 (2011).
- Balcan, D. & Vespignani, A. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nat. Phys.* **7**, 581–586 (2011).
- Gómez-Gardeñes, J., Soriano-Paños, D. & Arenas, A. Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks. *Nat. Phys.* **14**, 391–395 (2018).
- Soriano-Paños, D., Lotero, L., Arenas, A. & Gómez-Gardeñes, J. Spreading processes in multiplex metapopulations containing different mobility networks. *Phys. Rev. X* **8**, 031039 (2018).
- Moss, R., Naghizade, E., Tomko, M. & Geard, N. What can urban mobility data reveal about the spatial distribution of infection in a single city? *BMC Public Health* **19**, 656 (2019).
- Granel, C. & Mucha, P. J. Epidemic spreading in localized environments with recurrent mobility patterns. *Phys. Rev. E* **97**, 052302 (2018).
- Soriano-Paños, D., Ghoshal, G., Arenas, A. & Gómez-Gardeñes, J. Impact of temporal scales and recurrent mobility patterns on the unfolding of epidemics. *J. Stat. Mech.* **2020**, 024006 (2020).
- Tizzoni, M. et al. On the use of human mobility proxies for modeling epidemics. *PLoS Comput. Biol.* **10**, 1–15 (2014).
- Mazzoli, M. et al. Field theory for recurrent mobility. *Nat. Commun.* **10**, 3895 (2019).
- Barbosa, H., de Lima-Neto, F. B., Evsukoff, A. & Menezes, R. The effect of recency to human mobility. *EPJ Data Sci.* **4**, 21 (2015).
- Soriano-Paños, D. et al. Vector-borne epidemics driven by human mobility. *Phys. Rev. Res.* **2**, 013312 (2020).
- Tizzoni, M. et al. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Med.* **10**, 165 (2012).
- Wesolowski, A. et al. Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270 (2012).
- Arenas, A. et al. Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions. *Phys. Rev. X* **10**, 041055 (2020).
- Costa, G. S., Cota, W. & Ferreira, S. C. Outbreak diversity in epidemic waves propagating through distinct geographical scales. *Phys. Rev. Res. APS* **2**, 043306 (2020).
- Badr, H. S. et al. Association between mobility patterns and covid-19 transmission in the USA: a mathematical modelling study. *Lancet Infect. Dis.* **20**, P1247–1254 (2020).
- Bertuzzo, E. et al. The geography of covid-19 spread in Italy and implications for the relaxation of confinement measures. *Nat. Commun.* **11**, 4264 (2020).
- Alirol, E., Getaz, L., Stoll, B., Chappuis, F. & Loutan, L. Urbanisation and infectious diseases in a globalised world. *Lancet Infect. Dis.* **11**, 131–141 (2011).
- Kraemer, M. U. G. et al. Big city, small world: density, contact rates, and transmission of dengue across Pakistan. *J. R. Soc. Interface* **12**, 20150468–20150468 (2015).
- Li, R., Richmond, P. & Roehner, B. M. Effect of population density on epidemics. *J. Phys. A* **510**, 713–724 (2018).
- Lee, M., Barbosa, H., Youn, H., Holme, P. & Ghoshal, G. Morphology of travel routes and the organization of cities. *Nat. Commun.* **8**, 2229 (2017).
- Bassolas, A. et al. Hierarchical organization of urban mobility and its connection with city livability. *Nat. Commun.* **10**, 4817 (2019).
- Louail, T. et al. From mobile phone data to the spatial structure of cities. *Sci. Rep.* **4**, 5276 (2014).
- Free US Population Density And Unemployment Rate By Zip Code* (accessed 27 August 2020). <https://blog.splitwise.com/2014/01/06/free-us-population-density-and-unemployment-rate-by-zip-code/>.
- Australian Bureau of Statistics. <https://www.abs.gov.au/>.
- Statistics South Africa. <http://www.statssa.gov.za>.
- Facebook* (accessed 27 August 2020). <https://data.humdata.org/organization/facebook>.
- Rader, B. et al. Crowding and the shape of covid-19 epidemics. *Nat. Med.* **26**, 1829–1834 (2020).
- Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Impact of lockdown on covid-19 epidemic in île-de-France and possible exit strategies. *BMC Med.* **18**, 240 (2020).
- Prem, K. et al. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, E261–E270 (2020).
- Gatto, M. et al. Spread and dynamics of the covid-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl Acad. Sci. USA* **117**, 10484–10491 (2020).
- Gdp and employment flash estimates for the second quarter of 2020. Technical Report, Eurostat (2020).
- Gross domestic product, second quarter 2020 (advance estimate) and annual update. Technical Report, Bureau of Economic Analysis (2020).
- Hellewell, J. et al. Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, E488–E496 (2020).
- Ferretti, L. et al. Quantifying SARS-Cov-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).

61. Salathé, M. et al. Covid-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss Med. Wkly.* **150**, w20225 (2020).
62. Parino, F., Zino, L., Porfiri, M. & Rizzo, A. Modelling and predicting the effect of social distancing and travel restrictions on covid-19 spreading. *J. R. Soc. Interface* **18**, 20200875 (2021).
63. Cota, W., Soriano-Paños, D., Arenas, A., Ferreira, S. C. & Gómez-Gardeñes, J. Infectious disease dynamics in metapopulations with heterogeneous transmission and recurrent mobility. *New J. Phys.* **23**, 073019 (2021).
64. Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl Acad. Sci. USA* **102**, 11157–11162 (2005).
65. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl Acad. Sci. USA* **106**, 21484–21489 (2009).
66. Poletto, C., Tizzoni, M. & Colizza, V. Heterogeneous length of stay of host movements and spatial epidemic spread. *Sci. Rep.* **2**, 1–11 (2012).
67. Savitzky, A. & Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
68. *Smoothing in Python* (accessed 17 August 2020). <https://plotly.com/python/smoothing/>.
69. Hu, H., Nigmatulina, K. & Eckhoff, P. The scaling of contact rates with population density for the infectious disease models. *Math. Biosci.* **244**, 125–134 (2013).

## Acknowledgements

G.G. and S.H. acknowledge support from the U.S. Army Research Office under grant number W911NF-18-1-0421 and National Science Foundation under Grant No 2029095. A.A. acknowledges financial support from Spanish Ministerio de Ciencia e Innovación (grant PGC2018-094754-BC21), Generalitat de Catalunya (grant No. 2017SGR-896 and 2020PAND00098), and Universitat Rovira i Virgili (grant No. 2019PFR-URVB2-41). A.A. also acknowledges support from Generalitat de Catalunya ICREA Academia, and the James S. McDonnell Foundation (grant 220020325). D.S.P. and J.G.G. acknowledge financial support from Spanish Ministerio de Ciencia e Innovación (projects FIS2017-87519-P and PID2020-113582GB-I00), from the Departamento de Industria e Innovación del Gobierno de Aragón y Fondo Social Europeo (FENOL group E-19), and from Fundación Ibercaja and Universidad de Zaragoza (grant 224220).

## Author contributions

G.G., J.G.G., and A.A. conceived the study. D.S.P. and S.H. analyzed the data and perform the analytical derivations. All the authors analyzed the results and wrote the paper.

## 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/s42005-021-00679-0>.

**Correspondence** and requests for materials should be addressed to J.G.-G. or G.G.

**Peer review information** *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021