Supporting Information

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

Surendra Hazarie, David Soriano-Paños, Alex Arenas, Jesús Gómez-Gardeñes, and Gourab Ghoshal

Supplementary Note 1. Data

Mobility Data

The Google COVID-19 Aggregated Mobility Research Dataset contains anonymized mobility flows aggregated over users who have turned on the Location History setting, which is off by default. This is similar to the data used to show how busy certain types of places are in Google Maps — helping identify when a local business tends to be the most crowded. The dataset aggregates flows of people from region to region.

To produce this dataset, machine learning is applied to logs data to automatically segment it into semantic trips [1]. To provide strong privacy guarantees, all trips were anonymized and aggregated using a differentially private mechanism [2] to aggregate flows over time (see `https://policies.google.com/technologies/anonymization`). This research is done on the resulting heavily aggregated and differentially private data. No individual user data was ever manually inspected, only heavily aggregated flows of large populations were handled.

All anonymized trips are processed in aggregate to extract their origin and destination location and time. For example, if users traveled from location a to location b within time interval t, the corresponding cell (a, b, t) in the tensor would be $n \pm$ err, where err is Laplacian noise. The automated Laplace mechanism adds random noise drawn from a zero mean Laplace distribution and yields (ϵ, δ)–differential privacy guarantee of $\epsilon = 0.66$ and $\delta = 2.1 \times 10^{-29}$ per metric. Specifically, for each week W and each location pair (a, b), we compute the number of unique users who took a trip from location a to location b during week W. To each of these metrics, we add Laplace noise from a zero-mean distribution of scale $1/0.66$. We then remove all metrics for which the noisy number of users is lower than 100, following the process described in [2], and publish the rest. This yields that each metric we publish satisfies (ϵ, δ)–differential privacy with values defined above. The parameter ϵ controls the noise intensity in terms of its variance, while δ represents the deviation from pure ϵ–privacy. The closer they are to zero, the stronger the privacy guarantees.

We use data collected weekly from November 3rd 2019 to February 29th 2020, ensuring that we capture standard movement behavior, uninfluenced by pandemic conditions. Depending on the availability of population data per country, the mobility flows between S2 cells are aggregated to patch size of comparable resolution between countries:

- United States: 50 Urban areas with zip code patches [3].
- Italy: 49 Communes with S2 cell patches [4].
- South Africa: 31 Municipalities with ward patches [5].
- Australia: 33 Statistical Areas Level 4, with Level 2 patches [6].

To aggregate the S2 cells to the corresponding alternate patch types, the centroid points of the S2 cells were spatially joined with GIS boundaries of the alternate resolution. For example, in the US, flows from a zip code X to a zip code Y are determined by summing all of the flows that start in S2 cells whose centroids lie within X and end in S2 cells whose centroids lie within Y.

Population density data

Population density data is aggregated to patch-size in a similar manner when necessary. In the case of US zip codes, Australian statistical areas, and South African wards, population information is collected and available at those respective resolutions. For patches in Italy, population is determined by aggregating the populations of the 30 meter tiles collected by Facebook [4] to S2 cells. This is done by spatially merging the centroids of the tiles to the patches, similar to the S2-to-patch mobility flow aggregation.

Supplementary Note 2. Epidemic threshold dependence on mobility

In the main text, we characterize the interplay between mobility and epidemics by studying the value of the normalized epidemic threshold $\tilde{\lambda}_{\rm c}$ which is defined as the ratio between the threshold corresponding with the fully mobile population ($p = 1$) and the one associated with the static case ($p = 0$). To get a broader picture of the effect of the mobility, governed by the parameter p, let us redefine here the normalized epidemic threshold so that

$$\tilde{\lambda}_c(p) = \frac{\lambda_c(p)}{\lambda_c(p=0)} \ . \tag{S1}$$

Note that the values represented in the main text correspond with $\tilde{\lambda}_{\rm c}(1)$. The theoretical analysis presented in [7] predicts the existence of three different types of metapopulations as a function of the qualitative behavior of the epidemic threshold:

1. Type I: The epidemic threshold always decreases when promoting mobility. This type of cities is restricted to those metapopulations where the residential population is uniformly distributed across the system.

2. Type II: The epidemic threshold initially increases with the mobility until reaching a peak located at some intermediate value p^*. Beyond that value, promoting movements reduces the epidemic threshold, favoring epidemic spreading.

3. Type III: The epidemic threshold displays a monotonous positive trend with the mobility.

All the cities studied fall either into type II or type III metapopulations. To illustrate the differences between both cities, we explicitly represent in Supplementary Figure 1**a** the dependence of the threshold on the mobility for type II and type III cities. As shown in this figure, the location of the peak p^* allows classifying the cities. In this sense, we represent in Supplementary Figure 1**b**, a histogram of the p^* values observed for each of the 163 cities in our dataset, yielding 109 type II cities and 54 type III cities.

Supplementary Note 3. Examining the impact of hotspot concentration at the country level

In the main text, we reveal that there exists an inverse dependence of the normalized epidemic threshold $\tilde{\lambda}_{\mathrm{c}}$ on the flow concentration within hotspots κ. Nonetheless, despite the robustness of the results, the spatial resolution of the basic units composing the metapopulations is limited by data availability for each country. Therefore, different spatial resolutions are mixed in Figure 2. of the main text, which partially hinders the close relation between $\tilde{\lambda}_{\mathrm{c}}$ and κ_c.

To characterize the relevance of the flows connecting hotspots for epidemic spreading, we split the universal curve presented in the main text and represent in Supplementary Figure 2 the individual curves for each of the four countries analyzed here. The high Pearson and Spearman correlation coefficients among $\tilde{\lambda}_{\mathrm{c}}$ and κ strengthen our message about the close connection between the normalized epidemic threshold and the hotspot concentration. Note, however, that the influence of the connection between hotspots is quantitatively different in the four countries. In this sense, to quantify such different, we have also fitted the data to a power law decay function $\tilde{\lambda}_{\mathrm{c}} \sim \kappa^{\beta}$ via least-squares regression for each of the countries. Supplementary table 1 contains the estimated values for the exponent of the power-law function along with its uncertainty. Despite the high uncertainties observed in the fit, the value of the exponents suggests that the flow among hotspots is more relevant in Italy and South Africa than in United States and Australia which may be related with the different spatial definition of cities.

Supplementary Note 4. Connecting Indicators and Population Measures

We examine the relationship between normalized epidemic threshold $\tilde{\lambda}_c$, population density hotspot concentration κ and two other common population measures: average population density and Lloyd' mean crowding for cities in the US in Supplementary Figure 3.

Average Population Density

The average population density of a city k, $\langle d \rangle_k$ is calculated by:

$$\langle d \rangle_k = \frac{\sum\limits_i n_{i,k}}{\sum\limits_i a_{i,k}} \,, \tag{S2}$$

where i identifies the patches inside k.

Lloyd Mean Crowding

Lloyd mean crowding [8] for each city k was calculated according to

$$\frac{\sum\limits_i (n_{i,k} - 1) n_{i,k}}{\sum\limits_i n_{i,k}} \tag{S3}$$

for patch populations $n_{i,k}$. Lloyd mean crowding measures the number of unique contacts possible for members of patches in a city. We note that unlike κ and average population density, this measure is independent of patch size and therefore doesn't specifically address proximity.

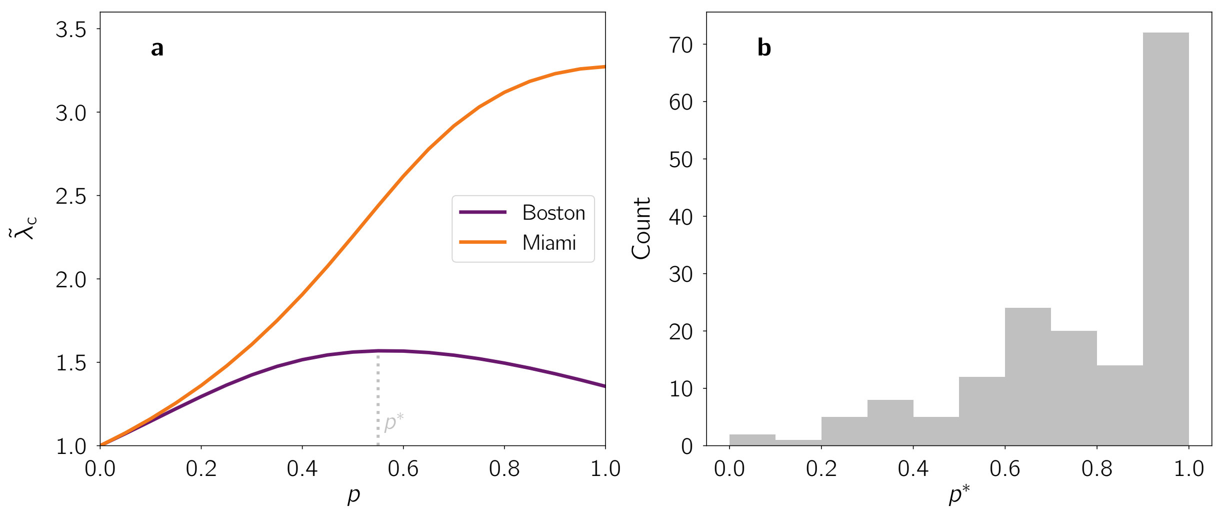
We find that our chosen pair of $\tilde{\lambda}_c$ and κ contains the strongest correlations. Of the three population-related metrics κ, average density, and Lloyd's Mean crowding, κ is the only one that contains mobility information, and like average population density, differentiates between patches of different spatial size, supporting that mobility and density are vital inclusions to properly understanding the epidemic situation of a city.

Supplementary Table 1: Exponent of the power law function $\tilde{\lambda}_c = A\kappa^{\beta}$ governing the dependence of the normalized epidemic threshold on the hotspot concentration.

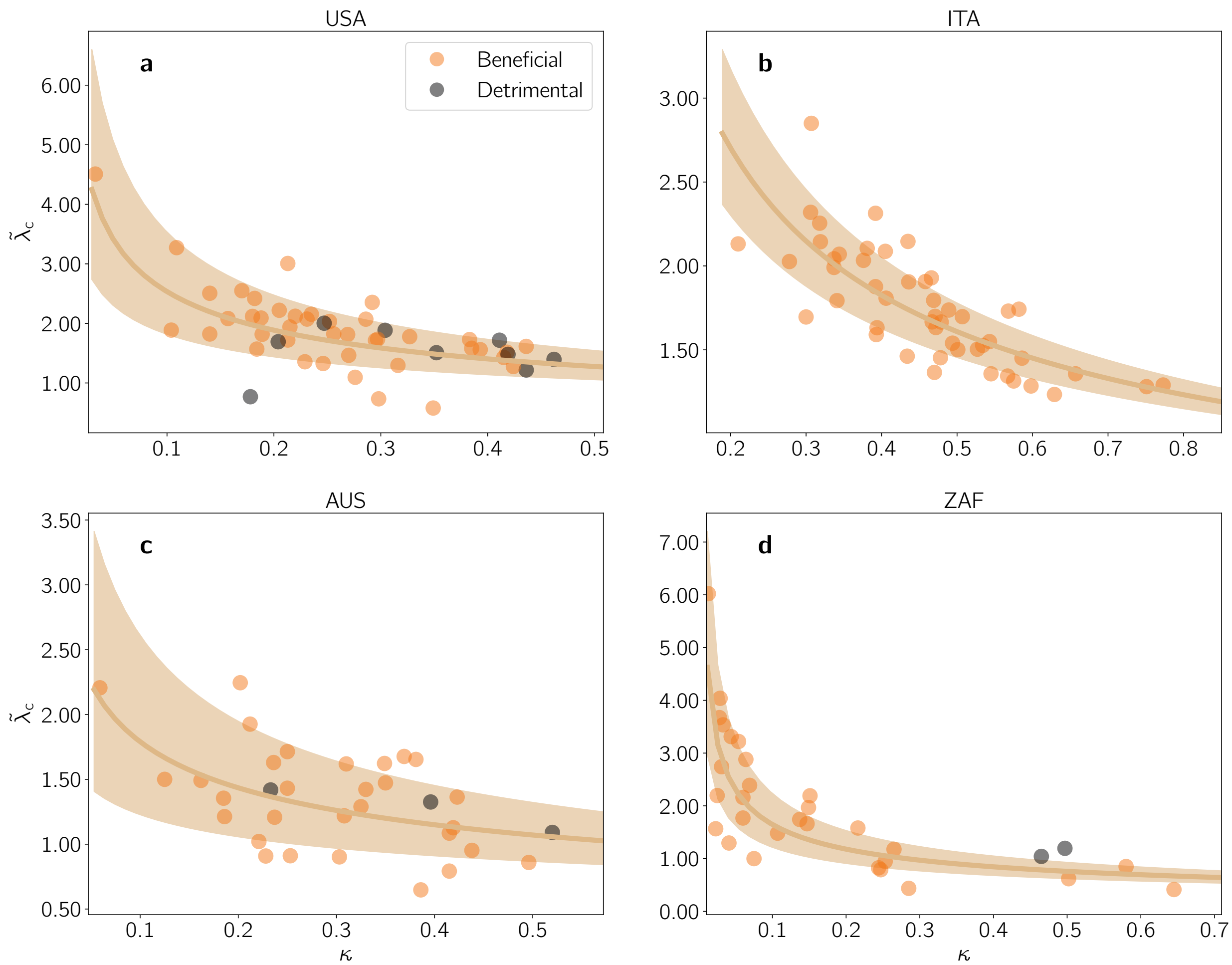
Country	**Decay rate** β
USA	-0.43 ± 0.09
ITA	-0.57 ± 0.07
AUS	-0.32 ± 0.1
ZAF	-0.48 ± 0.06

Supplementary Table 2: Cities for which removing the connections between hotspots is not beneficial to contain an ongoing outbreak. The effect of such intervention on the epidemic threshold is quantified by the ratio $\tilde{\lambda}_c^{\mathrm{MOD}}/\tilde{\lambda}_c$, being $\tilde{\lambda}_c^{\mathrm{MOD}}$ and $\tilde{\lambda}_c$ the normalized threshold after and before the reshuffling intervention respectively.

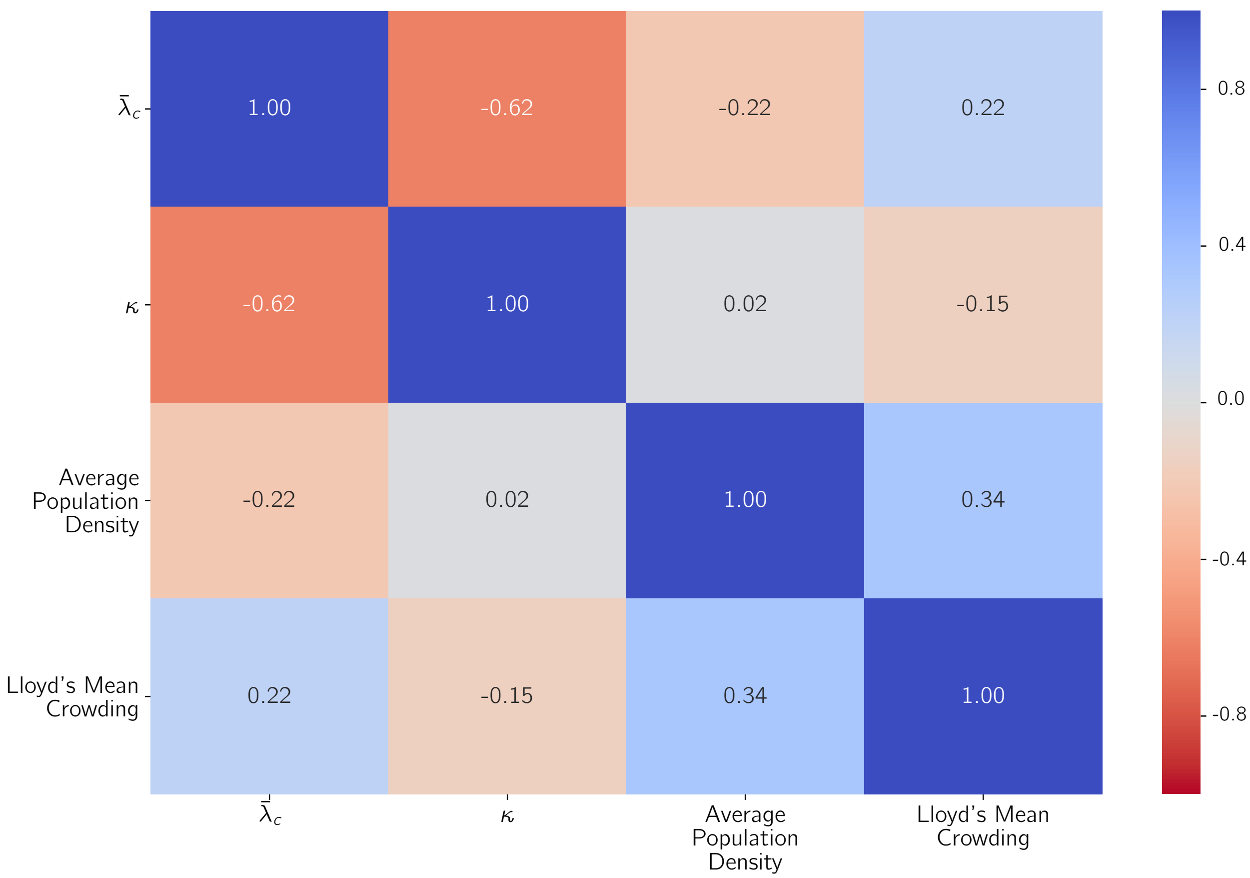
City	**Country**	$\tilde{\lambda}_{\mathbf{c}}^{\mathbf{MOD}}/\tilde{\lambda}_{\mathbf{c}}$
Ipswich	AUS	0.903
Adelaide	AUS	0.986
Moreton Bay	AUS	0.988
Tampa	USA	0.434
Detroit	USA	0.478
Pittsburgh	USA	0.537
Sacramento	USA	0.877
Dallas	USA	0.928
Indianapolis	USA	0.935
Cleveland	USA	0.946
San Diego	USA	0.951
Salt Lake City	USA	0.951
Dr JS Moroka	ZAF	0.881
Chief Albert Luthuli	ZAF	0.960



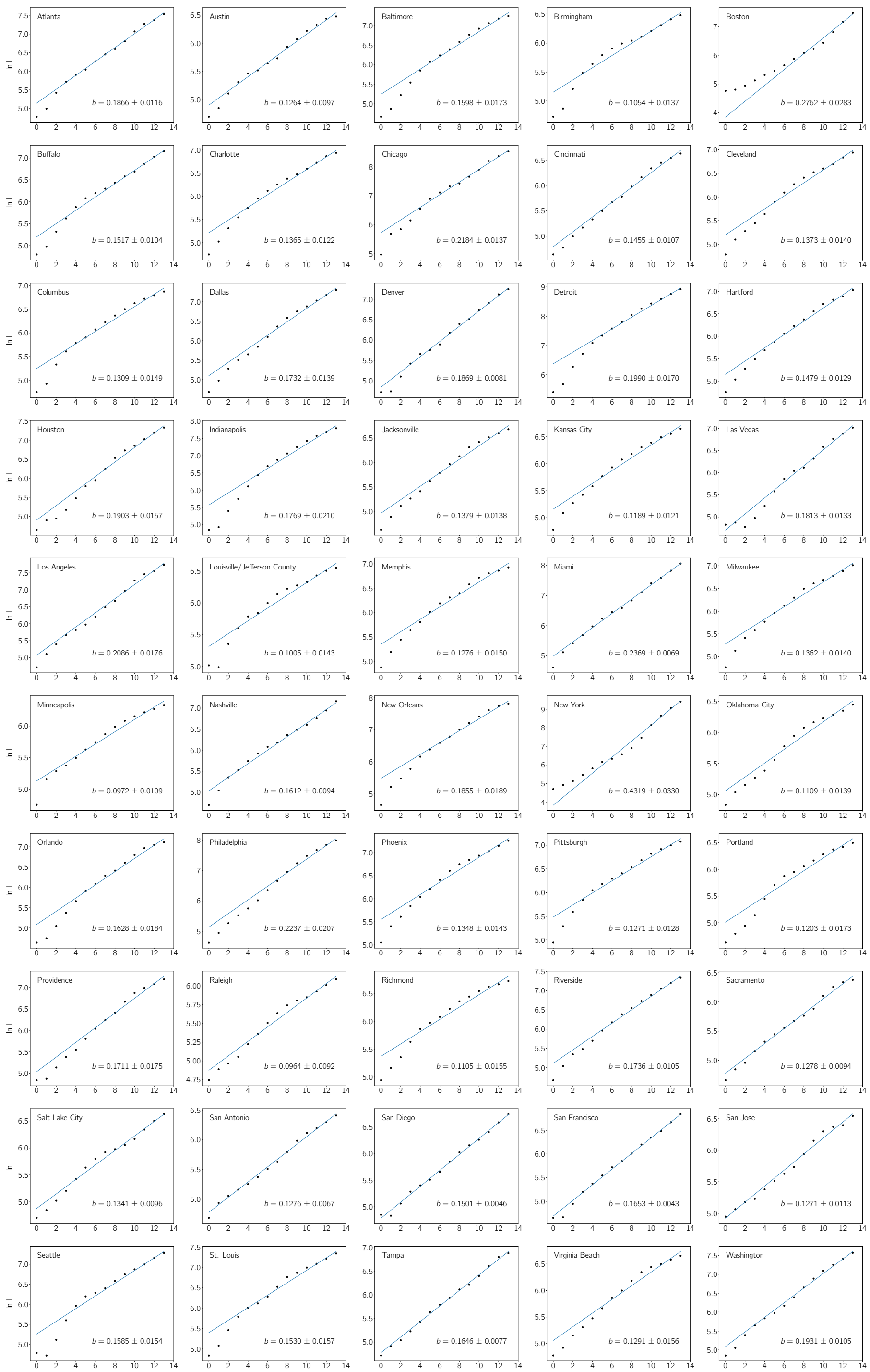
Supplementary Figure 1: **Mobility and epidemic threshold**. **a** Evolution of the normalized epidemic threshold $\tilde{\lambda}_c$ as a function of the mobility p for Boston (type II city) and Miami (type III city). The mobility value for which the epidemic threshold reaches its peak, p^*, is pinpointed with a dashed grey line. **b** Histogram of the p^* values for each of the 163 cities studied in this manuscript.



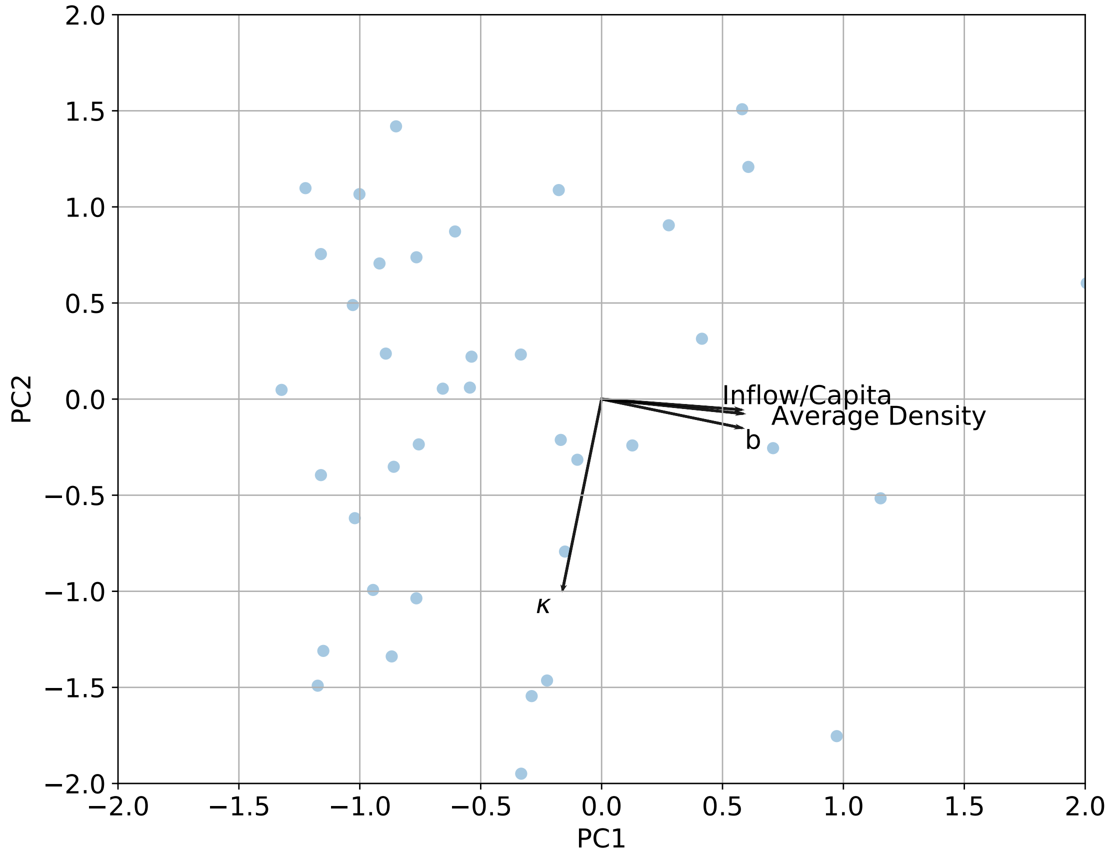
Supplementary Figure 2: **Connecting morphology to vulnerability at the country level.** Normalized epidemic threshold versus population density hotspot concentration for **a** the United States (Spearman: -0.62, Pearson = -0.59), **b** Italy (Spearman = -0.81, Pearson = -0.77), **c** South Africa (Spearman = -0.79, Pearson = -0.57), and **d** Australia (Spearman = -0.45, Pearson = -0.47) , with each city colored according to the outcome of removing flows connecting hotspots and distribute them evenly among non-hotspots neighboring areas (Fig. 2b). Solid line shows the curve computed with the mean value of the parameters obtained after fitting the data to a power-law function $\tilde{\lambda}_c = A\kappa^{\beta}$ via least-squares regression whereas the shadowed regions cover those curves generated by plugging the estimated parameters with 68.2% confidence interval.



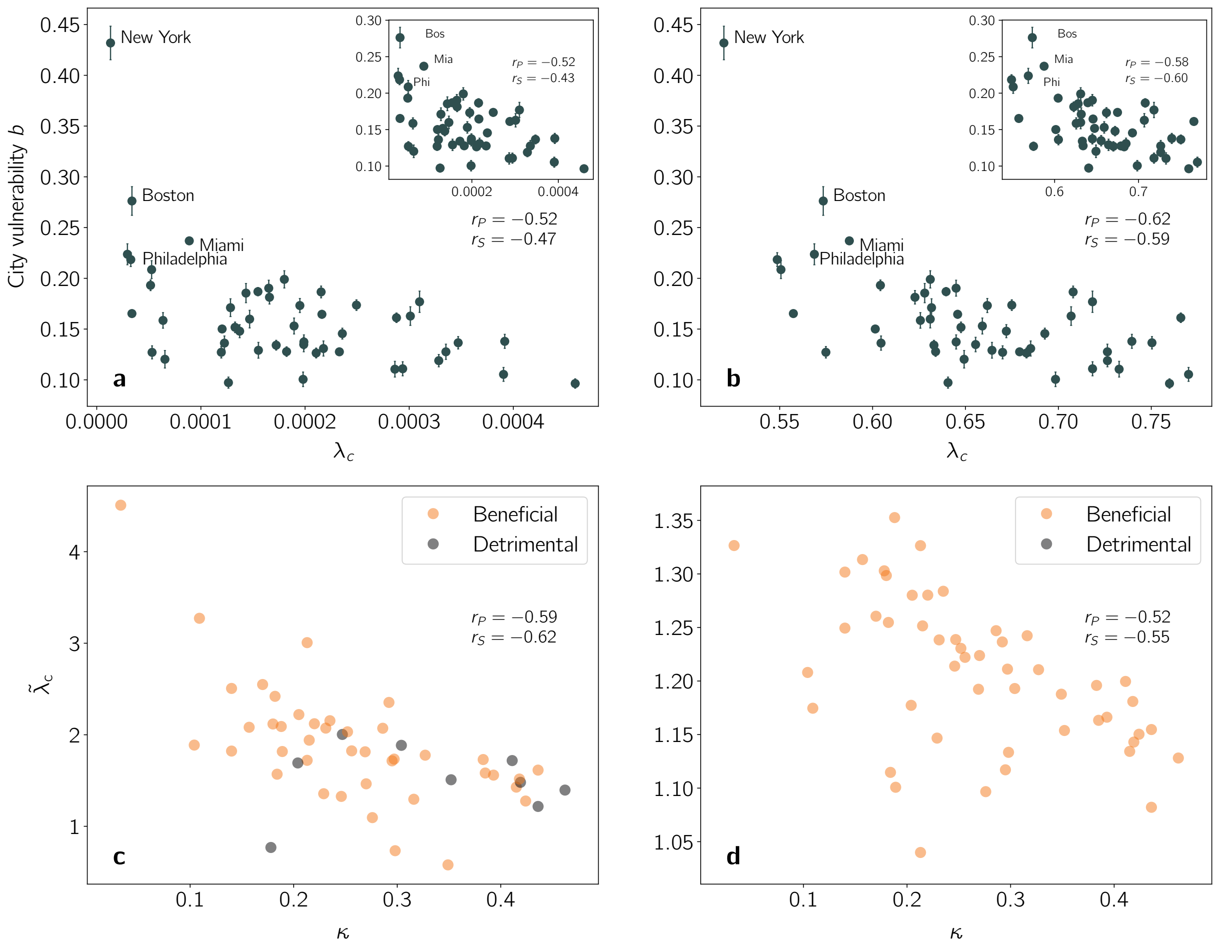
Supplementary Figure 3: **Connecting the normalized threshold with standard metrics**. Spearman Correlations between κ, the average population density in the city, and the Lloyd mean crowding with the normalized epidemic threshold $\tilde{\lambda}_c$.



Supplementary Figure 4: **Empirical estimation for cities' vulnerability.** Cumulative number of COVID-19 reported 14 days after the first 100 reported cases for the 50 cities in the United States. Dots show real data, smoothed to remove the inherent noisy nature of case reports by using a Savitzky-Golay filter. The line shows the exponential fit of the data via least-squares method to $I(t) = ae^{bt}$.



Supplementary Figure 5: **Connection between empirical vulnerability and theoretical metrics.** Principal component analysis biplot of 38 United States cities considering a space of four features: international inflow per capita, hotspot-hotspot flows (κ), epidemic growth exponent (b), and average city density.



Supplementary Figure 6: **Impact of the function governing contacts**. The number of contacts inside each area is assumed to depend linearly on the effective population density x via $f(x) = x$ (left panels) and $f(x) = 2 - e^{-\xi x}$, with $\xi = 2 \cdot 10^4$ square miles. **a-b** represent the city vulnerability b as a function of the epidemic threshold λ_c. **c-d** depict the evolution of the normalized epidemic threshold $\tilde{\lambda}_c$ as a function of the flow between hotspots encapsulated in the κ parameter. The dot color represents the effect of reshuffling the links between hotspots on the propagation of epidemics, being beneficial (orange) for those cities where the threshold is increases and detrimental (black) otherwise. r_P and r_S are the Pearson and Spearman correlation coefficients of the pair of variables represented in each panel.

Supplementary References

[1] Bassolas, A. *et al.* Hierarchical organization of urban mobility and its connection with city livability. *Nature Communications* **10**, 4817 (2019). URL `https://doi.org/10.1038/s41467-019-12809-y`.

[2] Wilson, R. J. *et al.* Differentially private sql with bounded user contribution. *Proceedings on privacy enhancing technologies* **2020**, 230–250 (2020).

[3] *Free US Population Density And Unemployment Rate By Zip Code* (accessed 2020- 08-27). URL `https://blog.splitwise.com/2014/01/06/free-us-population-density-and-unemployment-rate-by-zip-code/`.

[4] *Facebook* (accessed 2020- 08-27). URL `https://data.humdata.org/organization/facebook`.

[5] Statistics South Africa. `http://www.statssa.gov.za`.

[6] Australian Bureau of Statistics. `https://www.abs.gov.au/`.

[7] Gómez-Gardeñes, J., Soriano-Paños, D. & Arenas, A. Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks. *Nature Physics* **14**, 391–395 (2018). URL `https://doi.org/10.1038/s41567-017-0022-7`.

[8] Rader, B. *et al.* Crowding and the shape of covid-19 epidemics. *Nature Medicine* (2020).