Supporting Information

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

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

S1 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 [45,46], 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:

$$d_{i,k} > d_k^{\text{Lou}} \tag{S1}$$

where d_i is the population density of cell i in city k and d_k^{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, κ_k, defined as fraction of total flows in the city system that exist between these population density hotspots of city. Therefore, κ_k is given by:

$$\kappa_k = \frac{\sum\limits_{i,j \epsilon H_k} T_{ij}}{\sum\limits_{i,j} T_{ij}} \ , \tag{S2}$$

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

S2 Data

Table S1: Hotspot flow concentration κ for the different cities analyzed in the manuscript. The resolution column contain the two geographical divisions used inside each country to construct the metapopulations. For example, "Zip codes within CBSA" in the case of the USA implies that each city (metapopulation) corresponds to a CBSA and the entities composing each city (patches) correspond to zip codes.

City	**Country**	**Resolution**	κ
Bunbury	AUS	SA2 within SA4	0.059

City	**Country**	**Resolution**	κ
Capital Region	AUS	SA2 within SA4	0.125
Sydney	AUS	SA2 within SA4	0.162
Darwin	AUS	SA2 within SA4	0.185
Cairns	AUS	SA2 within SA4	0.186
Richmond	AUS	SA2 within SA4	0.202
Gold Coast	AUS	SA2 within SA4	0.212
South Australia	AUS	SA2 within SA4	0.221
Hume	AUS	SA2 within SA4	0.228
Moreton Bay	AUS	SA2 within SA4	0.233
Central Queensland	AUS	SA2 within SA4	0.236
Wide Bay	AUS	SA2 within SA4	0.237
Melbourne	AUS	SA2 within SA4	0.250
Townsville	AUS	SA2 within SA4	0.250
Australian Capital Territory	AUS	SA2 within SA4	0.253
Launceston and North East	AUS	SA2 within SA4	0.303
Sunshine Coast	AUS	SA2 within SA4	0.308
Central Coast	AUS	SA2 within SA4	0.310
Illawarra	AUS	SA2 within SA4	0.325
Perth	AUS	SA2 within SA4	0.330
Logan	AUS	SA2 within SA4	0.349
Brisbane	AUS	SA2 within SA4	0.350
Hunter Valley exc Newcastle	AUS	SA2 within SA4	0.369
Newcastle and Lake Macquarie	AUS	SA2 within SA4	0.381
Western Australia	AUS	SA2 within SA4	0.386
Ipswich	AUS	SA2 within SA4	0.396
West and North West	AUS	SA2 within SA4	0.415
Mackay	AUS	SA2 within SA4	0.415
Latrobe	AUS	SA2 within SA4	0.419
North West	AUS	SA2 within SA4	0.423
Hobart	AUS	SA2 within SA4	0.438
Central West	AUS	SA2 within SA4	0.496
Adelaide	AUS	SA2 within SA4	0.520
Bolzano	ITA	S2 cells within communes	0.210
Sassari	ITA	S2 cells within communes	0.278
Catania	ITA	S2 cells within communes	0.300
Livorno	ITA	S2 cells within communes	0.306
Taranto	ITA	S2 cells within communes	0.307
Foggia	ITA	S2 cells within communes	0.318
Perugia	ITA	S2 cells within communes	0.319
Bari	ITA	S2 cells within communes	0.337
Giugliano In Campania	ITA	S2 cells within communes	0.337
Terni	ITA	S2 cells within communes	0.341
Piacenza	ITA	S2 cells within communes	0.344
Genova	ITA	S2 cells within communes	0.376

City	Country	Resolution	κ
Ravenna	ITA	S2 cells within communes	0.381
Ferrara	ITA	S2 cells within communes	0.392
Ancona	ITA	S2 cells within communes	0.392
Lecce	ITA	S2 cells within communes	0.393
Cesena	ITA	S2 cells within communes	0.394
Pesaro	ITA	S2 cells within communes	0.405
Parma	ITA	S2 cells within communes	0.406
Pescara	ITA	S2 cells within communes	0.434
Roma	ITA	S2 cells within communes	0.435
Venezia	ITA	S2 cells within communes	0.436
Modena	ITA	S2 cells within communes	0.458
Trieste	ITA	S2 cells within communes	0.466
Brescia	ITA	S2 cells within communes	0.467
Trento	ITA	S2 cells within communes	0.469
Siracusa	ITA	S2 cells within communes	0.470
Napoli	ITA	S2 cells within communes	0.471
Palermo	ITA	S2 cells within communes	0.472
Milano	ITA	S2 cells within communes	0.478
Rimini	ITA	S2 cells within communes	0.479
Prato	ITA	S2 cells within communes	0.489
Verona	ITA	S2 cells within communes	0.494
Salerno	ITA	S2 cells within communes	0.501
Torino	ITA	S2 cells within communes	0.507
Arezzo	ITA	S2 cells within communes	0.527
Reggio Di Calabria	ITA	S2 cells within communes	0.534
Latina	ITA	S2 cells within communes	0.543
Udine	ITA	S2 cells within communes	0.545
Bologna	ITA	S2 cells within communes	0.567
Vicenza	ITA	S2 cells within communes	0.568
Andria	ITA	S2 cells within communes	0.575
Novara	ITA	S2 cells within communes	0.582
Padova	ITA	S2 cells within communes	0.586
Bergamo	ITA	S2 cells within communes	0.598
Firenze	ITA	S2 cells within communes	0.629
Messina	ITA	S2 cells within communes	0.657
Cagliari	ITA	S2 cells within communes	0.751
Monza	ITA	S2 cells within communes	0.773
Virginia Beach	USA	Zip codes within CBSA	0.033
Washington	USA	Zip codes within CBSA	0.104
Miami	USA	Zip codes within CBSA	0.109
Columbus	USA	Zip codes within CBSA	0.140
Seattle	USA	Zip codes within CBSA	0.140
Atlanta	USA	Zip codes within CBSA	0.157
Houston	USA	Zip codes within CBSA	0.170

City	**Country**	**Resolution**	κ
Pittsburgh	USA	Zip codes within CBSA	0.178
Richmond	USA	Zip codes within CBSA	0.180
Raleigh	USA	Zip codes within CBSA	0.182
San Francisco	USA	Zip codes within CBSA	0.184
Nashville	USA	Zip codes within CBSA	0.188
Los Angeles	USA	Zip codes within CBSA	0.189
Salt Lake City	USA	Zip codes within CBSA	0.204
Cincinnati	USA	Zip codes within CBSA	0.205
New York	USA	Zip codes within CBSA	0.213
Austin	USA	Zip codes within CBSA	0.213
Minneapolis	USA	Zip codes within CBSA	0.215
Birmingham	USA	Zip codes within CBSA	0.220
Boston	USA	Zip codes within CBSA	0.229
Providence	USA	Zip codes within CBSA	0.231
Charlotte	USA	Zip codes within CBSA	0.235
Louisville/Jefferson County	USA	Zip codes within CBSA	0.246
Dallas	USA	Zip codes within CBSA	0.247
Orlando	USA	Zip codes within CBSA	0.252
St. Louis	USA	Zip codes within CBSA	0.256
Riverside	USA	Zip codes within CBSA	0.269
Buffalo	USA	Zip codes within CBSA	0.270
Chicago	USA	Zip codes within CBSA	0.276
Kansas City	USA	Zip codes within CBSA	0.286
Baltimore	USA	Zip codes within CBSA	0.292
San Jose	USA	Zip codes within CBSA	0.295
Denver	USA	Zip codes within CBSA	0.297
Philadelphia	USA	Zip codes within CBSA	0.298
Cleveland	USA	Zip codes within CBSA	0.304
Hartford	USA	Zip codes within CBSA	0.316
San Antonio	USA	Zip codes within CBSA	0.327
Portland	USA	Zip codes within CBSA	0.349
San Diego	USA	Zip codes within CBSA	0.352
Phoenix	USA	Zip codes within CBSA	0.383
Milwaukee	USA	Zip codes within CBSA	0.385
Memphis	USA	Zip codes within CBSA	0.393
Indianapolis	USA	Zip codes within CBSA	0.411
Las Vegas	USA	Zip codes within CBSA	0.415
Oklahoma City	USA	Zip codes within CBSA	0.418
Sacramento	USA	Zip codes within CBSA	0.419
New Orleans	USA	Zip codes within CBSA	0.424
Jacksonville	USA	Zip codes within CBSA	0.436
Tampa	USA	Zip codes within CBSA	0.436
Detroit	USA	Zip codes within CBSA	0.462
Govan Mbeki	ZAF	Wards within municipalities	0.013

City	Country	Resolution	κ
Hibiscus Coast	ZAF	Wards within municipalities	0.023
City of Matlosana	ZAF	Wards within municipalities	0.025
Nelson Mandela Bay	ZAF	Wards within municipalities	0.028
City of Cape Town	ZAF	Wards within municipalities	0.029
Emalahleni	ZAF	Wards within municipalities	0.031
City of Tshwane	ZAF	Wards within municipalities	0.033
Local Municipality of Madibeng	ZAF	Wards within municipalities	0.041
Ekurhuleni	ZAF	Wards within municipalities	0.044
City of Johannesburg	ZAF	Wards within municipalities	0.054
Matjhabeng	ZAF	Wards within municipalities	0.060
Nkomazi	ZAF	Wards within municipalities	0.060
Greater Letaba	ZAF	Wards within municipalities	0.064
Mangaung	ZAF	Wards within municipalities	0.069
Makhado	ZAF	Wards within municipalities	0.075
Emfuleni	ZAF	Wards within municipalities	0.107
Greater Tzaneen	ZAF	Wards within municipalities	0.137
eThekwini	ZAF	Wards within municipalities	0.147
Mogalakwena	ZAF	Wards within municipalities	0.149
Lepele	ZAF	Wards within municipalities	0.151
Buffalo City	ZAF	Wards within municipalities	0.216
Polokwane	ZAF	Wards within municipalities	0.244
The Msunduzi	ZAF	Wards within municipalities	0.247
Rustenburg	ZAF	Wards within municipalities	0.253
Makhuduthamaga	ZAF	Wards within municipalities	0.265
Thembisile	ZAF	Wards within municipalities	0.285
Chief Albert Luthuli	ZAF	Wards within municipalities	0.465
Dr JS Moroka	ZAF	Wards within municipalities	0.497
Maluti a Phofung	ZAF	Wards within municipalities	0.502
Bushbuckridge	ZAF	Wards within municipalities	0.580
Thulamela	ZAF	Wards within municipalities	0.645

S2.1 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

[1] https://www.nature.com/articles/s41467-019-12809-y
[2] https://research.google/pubs/pub48778/

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 $n \pm \text{err}$, 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 [3], 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 [47].
- Italy: 49 Communes with S2 cell patches [50].
- South Africa: 31 Municipalities with ward patches [49].
- Australia: 33 Statistical Areas Level 4, with Level 2 patches [48].

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.

S2.2 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 [50] 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.

S3 Epidemic model

S3.1 Dynamical equations

The model used for estimating the vulnerability of a given city is a generalized version of the formalism included in [27]. From an epidemiological point of view, the model incorporates a SIR

[3] https://research.google/pubs/pub48778/

dynamics where individuals can be susceptible of contracting the disease (S), infectious (I) or recovered (R). An infectious individual transmits the pathogen to healthy counterparts via direct interaction at a rate λ. In turn, infectious individuals enter the compartment R at a rate μ, 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 composed by a set of N patches which represent places gathering agents. Those individuals can move across the metapopulation, being these movements determined by the mobility patterns usually encoded in origin-destination matrices.

In our model, each individual has an associated node which is identified as her residence. Therefore, each patch i is initially populated by n_i agents. We split each day into three stages: Movement, Interaction and Return. First, agents decide whether or not moving 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

$$R_{ij} = \frac{T_{ij}}{\sum_l T_{il}} \ . \tag{S3}$$

After all the movements have been completed, interaction among agents sharing the same current location occurs. At this point, we make a mean-field assumption within each patch, so every agent inside a given location makes the same number of contacts. We also assume that the number of contacts is proportional to the density of each patch via a function f which allows for including different ways of weighting the relevance of population density for the number of interactions inside a given patch. Finally, as we want to reflect the mostly 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 is 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, $\rho_i^R(t+1)$, associated to patch i at time $t+1$, read:

$$\rho_i^I(t+1) = (1-\mu)\rho_i^I(t) + (1-\rho_i^I(t)-\rho_i^R(t))\Pi_i(t) \ , \tag{S4}$$

$$\rho_i^R(t+1) = \rho_i^R(t) + \mu\rho_i^I(t) \ . \tag{S5}$$

Eq. (S5) 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. (S4) 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:

$$\Pi_i(t) = (1-p)P_i(t) + p\sum_{j=1}^{N} R_{ij}P_j(t) \ . \tag{S6}$$

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:

$$P_i(t) = 1 - \left(1 - \lambda \frac{I_i^{eff}(t)}{n_i^{eff}}\right)^{f_i} \ , \tag{S7}$$

where f_i determines the number of contacts per day of individuals inside i. Finally, the terms n_i^{eff} and $I_i^{eff}(t)$, which denote the effective population and effective number of infected individuals inside patch i at time t after population movements, read:

$$n_i^{eff} = \sum_{j=1}^{N} n_j \left(\delta_{ij}(1-p) + R_{ji}\right) \ , \tag{S8}$$

$$I_i^{eff} = \sum_{j=1}^{N} n_j \rho_j^I(t) \left(\delta_{ij}(1-p) + R_{ji}\right) \ . \tag{S9}$$

S3.1.1 Estimating cities' vulnerability

The former equations offer the possibility of estimating the vulnerability for each 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 assume that the disease has reached a stationary equilibrium and that the epidemic size is negligible compared with the population size. Mathematically, this imply that $\vec{\rho}^I(t+1) = \vec{\rho}(t) = \vec{\epsilon} \ll \vec{1}$. In addition, we neglect the individuals belonging to the compartment R by setting $\vec{\rho}^R(t+1) = \vec{\rho}^R(t) = \vec{0}$. Both assumptions allow us to linearize Eq. (S7) which now reads:

$$P_i \simeq \lambda f_i \frac{I_i^{eff}}{n_i^{eff}} \ . \tag{S10}$$

After introducing Eqs.(S6-S10) into Eq.(S4) and taking into account the stationary regime, we obtain:

$$\frac{\mu}{\lambda}\epsilon_i = \sum_{j=1}^{N} \underbrace{\left[(1-p)^2\delta_{ij}\frac{f_i}{n_i^{eff}} + p(1-p)\left(R_{ij}\frac{f_j}{n_j^{eff}} + R_{ji}\frac{f_i}{n_i^{eff}}\right) + p^2\sum_l R_{il}R_{jl}\frac{f_l}{n_l^{eff}}\right]}_{M_{ij}} n_j\,\epsilon_j \ . \tag{S11}$$

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

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M})} \ , \tag{S12}$$

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

S3.2 Function governing contacts

At this point, it is necessary to specify 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 non-parametric approach by assuming that these contacts are linearly proportional to the density inside each patch. This way, the results shown in Figs. 2 & 4 are obtained by assuming:

$$f_i = \frac{n_i^{eff}}{a_i} \ , \tag{S13}$$

being a_i the area of patch i.

If one is interested in isolating the role played by mobility in shaping cities' vulnerability, we must remove the dependence of the overall population density for each case. To do so, we decide to compute the normalized epidemic threshold $\tilde{\lambda}_c$ which is defined as the ratio between the actual epidemic threshold, computed by accounting for human mobility, and the threshold corresponding to a static scenario. Therefore, this quantity reads:

$$\tilde{\lambda}_c = \frac{\Lambda_{\max}(\mathbf{M})(p=0)}{\Lambda_{\max}(\mathbf{M})(p=1)} \tag{S14}$$

Although useful for illustrating the role of mobility in each city, the non-parametric 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 on the expansion of COVID-19 over different cities, we choose a more complex function governing the number of contacts. Following [**?**], we assume that

$$f_i = 2 - e^{-\xi \frac{n_i^{eff}}{a_i}} \ , \tag{S15}$$

which is bounded such that $f_i \in [1, 2)$. Here the parameter ξ is estimated by maximizing the correlation among the theoretical and the observed vulnerabilities, yielding $\xi = 2 \cdot 10^{-5}$ square miles. By including this function, we estimate the epidemic threshold for each city, λ_c, as:

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M})(p=1)} \ , \tag{S16}$$

where we set $\mu = 1$ for the sake of simplicity. Note that this 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.

S4 Examining the impact of hotspot concentration at the country level

In the main text, we reveal that all the cities analyzed here, regardless of their associated country, fall in an universal curve governing the dependence of the normalized epidemic threshold $\tilde{\lambda}_c$ on the flow concentration within hotspots κ. Moreover, we check that this universal curve is well-represented by a sub-linear power law decay function. 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 Fig. 2 of the main text, which partially hinders the close relation between $\tilde{\lambda}$ 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 Fig. S1 the individual curves for each

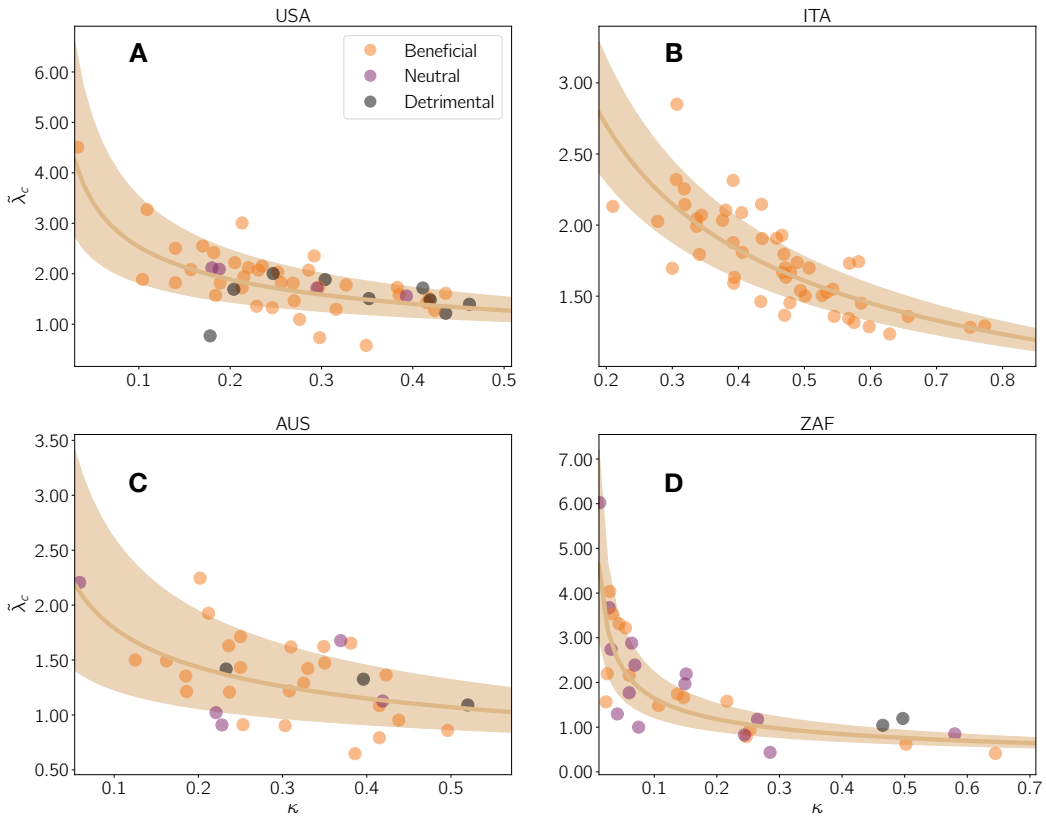

Figure S1: Normalized epidemic threshold versus population density hotspot concentration for the (**A**) United States (Spearman: -.62, Pearson = -.59), (**B**) Italy (Spearman = -.81, Pearson = -.77), (**C**) South Africa (Spearman = -.79, Pearson = -.57), and (**D**) Australia (Spearman = -.45, Pearson = -.47) , with each city colored according to the outcome of removing flows connecting hotspots and distribute them evenly among non-hotspots neighboring areas (Fig. 2**B**). Solid line shows the power-law fitting $\tilde{\lambda}_c = A\kappa^{\beta}$ and shadowed regions cover 68% confidence interval.

of the four countries analyzed here. The high Pearson and Spearman correlation coefficients among $\tilde{\lambda}$ and κ along with the goodness of the power law fitting for each individual country strengthen our message about the close connection between the normalized epidemic threshold and the hotspot concentration. The analysis at the country level indicates that, while the function governing the relationship is universal, the parameters are dependent on both the spatial resolution at which the data is available and other differences between cities, not considered in this analysis.

S5 Connecting Indicators and Population Measures

We examine the relationship between normalized epidemic threshold $\bar{\lambda}$, population density hotspot concentration κ and two other common population measures: average population density and Lloyd' mean crowding for cities in the US in Fig. S2.

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

Table S2: Exponent of the power law function $\tilde{\lambda}_c = A\kappa^{\beta}$ governing the dependence of the normalized epidemic threshold on the hotspot concentration.

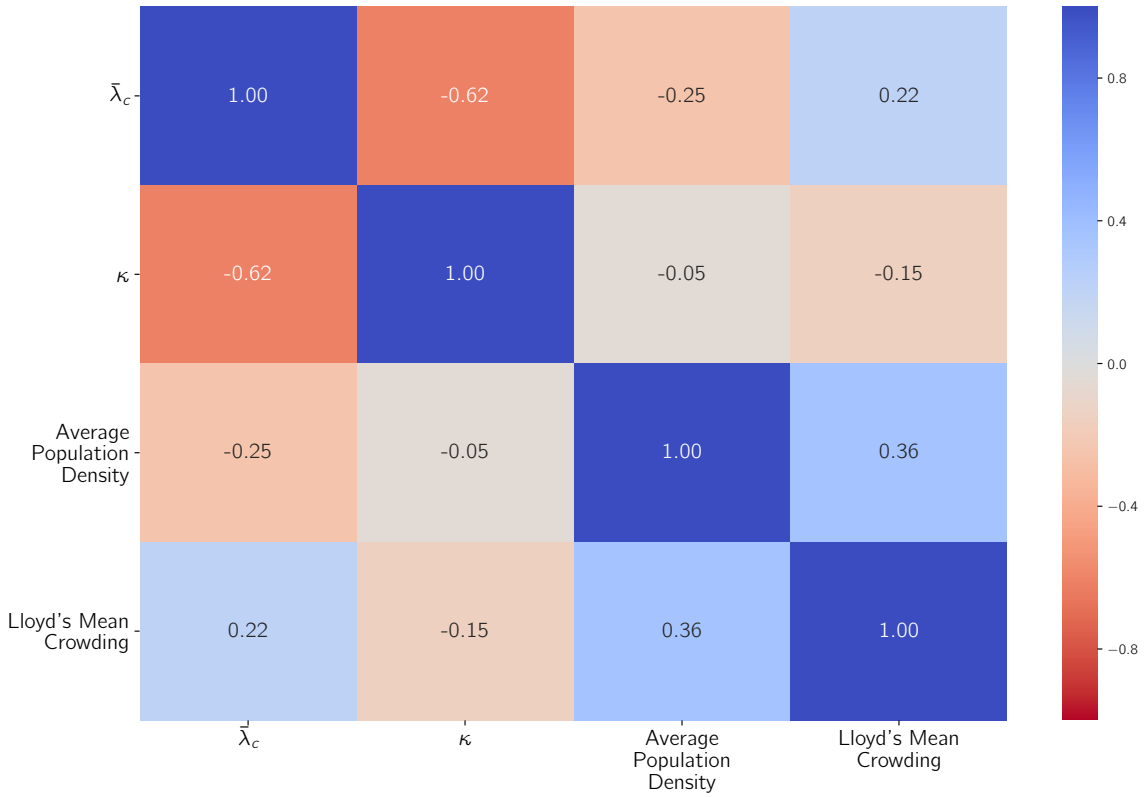

Figure S2: Spearman Correlations between κ, the average population density in the city, and the Lloyd mean crowding [51] with the normalized epidemic threshold $\tilde{\lambda}_c$

S5.1 Average Population Density

The average was calculated by averaging the population densities $d_{i,k}$ of all i_{tot} zip codes i within each city k:

$$\frac{\sum d_{i,k}}{i_{tot}} \tag{S17}$$

and describes the relative concentration of people within a city.

S5.2 Lloyd Mean Crowding

Lloyd mean crowding [**?**] was calculated according to

$$\frac{\sum (q_i - 1)q_i}{\sum q_i} \tag{S18}$$

for patch populations q_i. 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 $\bar{\lambda}$ 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.

S6 Empirical Epidemic-Indicator Correlations

We quantify the extent to which COVID-19 is able to spread in a US city by examining the timeseries of confirmed cases per county [55,56], 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 [**?**] to smooth the data by fitting intermediate windows [**?**] 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

$$I(t) = ae^{bt} \tag{S19}$$

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. The smoothed curves along with the exponential fits are presented in Fig. S3.

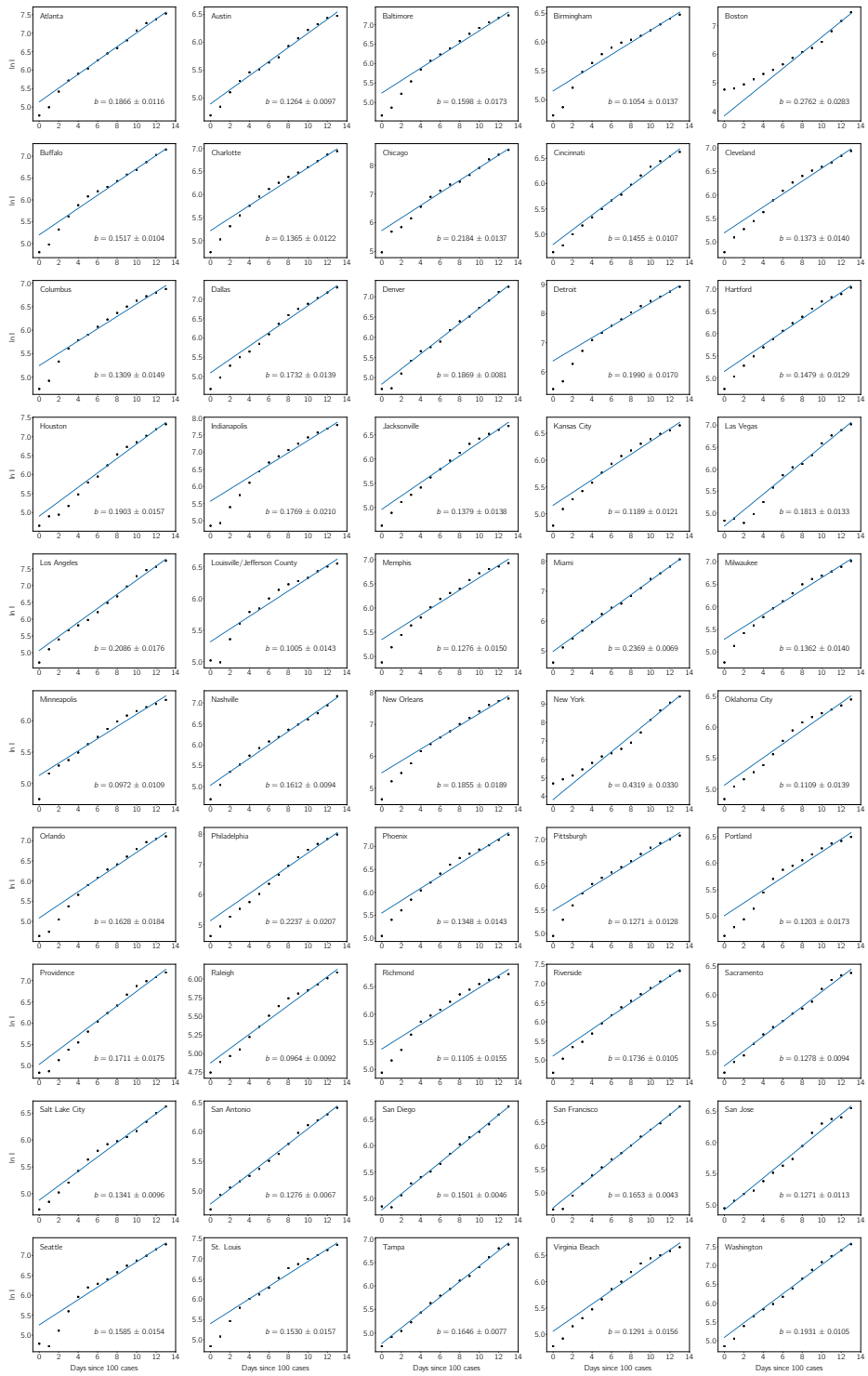

Figure S3: Cumulative number of COVID-19 reported 14 days after the first 100 reported cases for the 50 cities in the United States. Dots show real data, smoothed to remove the inherent noisy nature of case reports by using a Savitzky-Golay filter. The line shows the exponential fit of the data via least-squares method to $I(t) = ae^{bt}$.