

Universal scaling law for COVID-19 propagation in urban centers

Ben-Hur Francisco Cardoso^{a,1} and Sebastián Gonçalves^{a,b,2}

^aInstituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil
^bURPP Social Networks, University of Zürich, Zürich, Switzerland

Due to the COVID-19 pandemic, there is a high demand for Susceptible-Infective-Recovered (SIR) models to adjust and predict the number of cases in urban areas. Forecasting, however, is a difficult task, because the change in people's behavior reflects in a continuous change in the parameters of the model. An important question is what we can use from one city to another; if what happened in Madrid could have been applied to New York and then, if what we have learned from this city would be useful for São Paulo. To answer this question, we present an analysis of the transmission rate of COVID-19 as a function of population density and population size for US counties, cities of Brazil, German, and Portugal. Contrary to the common hypothesis in epidemics modeling, we observe a higher disease transmissibility for higher city's population density/size –with the latter showing more predicting power. We present a contact rate scaling theory that explain the results, predicting that the basic reproductive number R_0 of epidemics scales as the logarithm of the city size.

Urban Scaling | COVID-19 | SIR model | Epidemics

Correspondence: ¹ ben-hur.cardoso@ufrgs.br, ² sgonc@ufrgs.br

Introduction

The epidemic of COVID-19 that started in the Chinese city of Wuhan in December of 2019, was declared a pandemic on March 11th, 2020 by the World Health Organization (WHO). While in many European countries (1) the epidemic is slowing down after months of restrictions, the US is still struggling with the highest numbers of active cases and deaths. Presently, however, the epicenter is in Brazil, where the epidemic has the potential to hit even worse than in the US.

The COVID-19 is a human contact driven disease, therefore it is necessary to work at the municipal level (2, 3) to get precise predictions. Such an approach, in addition to providing more accurate forecasts, it would allow us to apply the experience of one city to others. With this idea in mind, we focus on the *basic reproductive number* R_0 , a key concept in the mathematical description of epidemics, that measure the average number of secondary cases generated by each infected subject.

For a better understating, we factorize the basic reproductive number as $R_0 = \tau p C$, where τ is the infectious period, C is the *per capita* contact rate and p the transmission probability. The probability of infection p ¹ and the infectious period τ are characteristic of the disease itself. Therefore, to make the

¹It can be drastically attenuated or even suppressed by the use of masks, for example.

comparison between cities, it is necessary to understand how the contact rate is related to demographics.

In epidemiological modeling, the *mass action* hypothesis (4) is generally followed: the *per capita* contact rate is assumed invariant of city size. Our analysis shows, however, that this assumption cannot explain the available COVID-19 data. Alternatively, we propose a scaling theory that explains the dependence between contact rate and urban size or density, which then explain how the epidemic growth rate (*i.e.* R_0) depends on size or density.

Population size and Population density

Consider a city with population size N and land area A . There are two main competing hypothesis that try to explain how the *per capita* contact rate C varies with N and A : the population **size driven** contact rate, where $C = C(N)$; and the population **density driven** contact rate, stating that $C = C(\rho)$, where $\rho = N/A$. While the first approach assumes that the social mobility network grows in larger areas, allowing more distant people to interact (5), the second one assumes that the length traveled by the individuals is invariant of the size of the city (6).

Intriguingly, based on data of disease transmission in the United States, both approaches appear to be valid (6–8). The reason for this is the quasi-linear relation between the density and size population of the US counties, as shown in Fig. 1. Indeed, we have found that the best fit is $\rho \propto N^\lambda$, with $\lambda \approx 1.03$. Assuming a linear relation instead, $\rho = kN$ gives an equally valid fit, with $k = 0.00059 \text{ Km}^{-2}$ (this gives a typical US county diameter of 46.5 Km). An almost constant density across counties, or no correlation between ρ and N , cannot explain the data.

In addition to the US, we study the COVID-19 transmission in cities of Portugal, Brazil and Germany. While the same quasi-linear scaling $\lambda = 1.03$ was found for Portugal's municipalities, in the other two countries the population density of cities does not correlate well with their size. The linear fit is strong for Portugal, weak for Brazil and almost nonexistent for Germany. Constant density also cannot explain the data for these countries (see Figs. S1, S3 and S2 in Supplementary Information Appendix).

Since the urban size and density are not well correlated in Brazil and Germany, we use these countries to verify which of the approaches, size-driven or density-driven, is the valid one. The result may be useful during the present COVID-19 pandemic and for futures ones.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

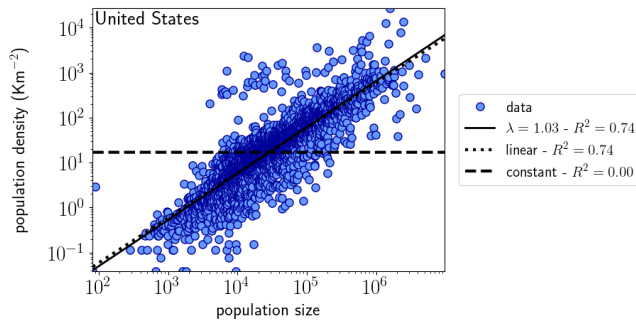


Fig. 1. Population density (ρ) vs population size (N) for US counties, represented by blue dots (data). The solid line is a fit to a power law, $\rho \propto N^\lambda$, while the dotted line is a fit to the $\rho = kN$ relation, with $k = 0.00059 \text{ Km}^{-2}$. Both fits have equal correlation coefficient, $R = 0.86$. The dashed line represents the average density. See Materials and Methods section for data source and fit methodology.

Contact rate scaling theory

Consider $N \gg 1$ individuals –represented as nodes– distributed uniformly in a two-dimensional space with land area A . As introduced by Noulas *et al* (9), we can expect an individual i to form a link with individual j with probability

$$P_{ij} = \frac{1}{\text{rank}_i(j)}, \quad (1)$$

where $\text{rank}_i(j)$ is the number of neighbors closer to i than j . Assuming that the distance between these two individuals is r , we then have

$$\text{rank}_i(j) = \rho \pi r^2 \Rightarrow P_{ij} = P(r) = \frac{1}{\rho \pi r^2}, \quad (2)$$

with $\rho = N/A$ as the population density. First, since $0 \leq P \leq 1$, we must impose a bottom cutoff radius r_0 such that

$$P(r \leq r_0) \equiv 1 \Rightarrow r_0 = \frac{1}{\sqrt{\pi \rho}}.$$

Secondly, it is natural to assume an upper cutoff radius r_1 for long distances such that $P(r > r_1) = 0$. Then, similarly to Pan *et al* (6), the mean degree of the social interaction network is

$$\begin{aligned} k &= \frac{1}{2} \int_{r_0}^{r_1} dr (2\pi r \rho) P(r) = \frac{1}{2} \left[1 + 2 \int_{r_0}^{r_1} \frac{dr}{r} \right] = \\ &= \frac{1}{2} \left[1 + \ln \left(\frac{r_1^2}{r_0^2} \right) \right] \Rightarrow k = \frac{1}{2} \ln(a_1 \rho), \end{aligned} \quad (3)$$

where $a_1 = \pi r_1^2$ is the coverage area of individual mobility and the $1/2$ factor eliminates the double counting.

It is well reasonable to expect the *per capita* contact rate C , and so R_0 , to be proportional to the mean degree k . Therefore, following the population density driven approach, which states that a_1 is invariant, from Eq. 3 we arrive at

$$R_0 \propto C \propto \ln(a_1 \rho). \quad (4)$$

On the other hand, with the population size-driven approach, the mobility increases with A ($a_1 = \kappa A$), then from Eq. 3 we now have

$$R_0 \propto C \propto \ln(\kappa N). \quad (5)$$

Results

In this Section we check on the validity of the scaling model (Eqs. 4 and 5) in comparison to the traditional *mass action* hypothesis. In the Materials and Methods section, we explain how we estimate the basic reproductive number R_0 of 1931 counties of the United States, 546 cities of Brazil, 401 cities of Germany and, 143 municipalities of Portugal, based on the COVID-19 related data. Finally, we give a plausible explanation for the marginal deviation of data from the scaling based on available data of mobility.

Population density driven contact rate. In Fig. 2 we show the relation between R_0 and the population density for different counties of the United States. We see that the proposed model (density driven, Eq. 4) provides a good fit, while the null (mass action) hypothesis cannot explain the data. A similar behaviour was found for Portugal (see Fig. S4 in Supplementary Information Appendix).

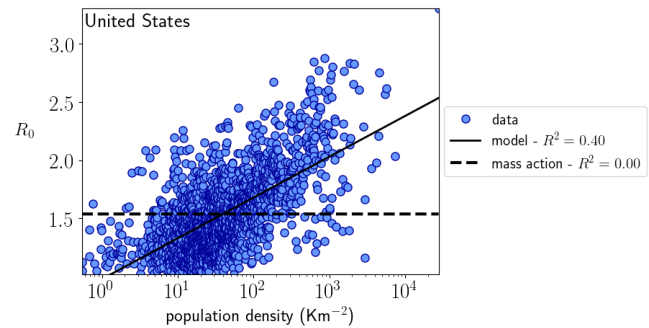


Fig. 2. R_0 vs the population density for different counties of the United States (see Materials and Methods section for fit methodology).

However, this behavior seems not to be universal. In the cases of Brazil and Germany the density driven model has almost the same predictability that the mass action hypothesis (see Figs. S5 and S6 in Supplementary Information Appendix). Yet, the *per capita* contact rate may scale with the population size instead of the population density as we see in the next Subsection.

Population size driven contact rate. In Fig. 3, we show the relation between R_0 and population size for urban regions of the United States, as well as for Brazil, Germany, and Portugal. We can see that the model (Eq. 5) is the one that provides a reasonable universal fit, while again the mass action hypothesis fails to explain the results (see Figs. S7, S9 and S10 in Supplementary Information Appendix for country-separated plots).

How can we explain that the COVID-19 data for the US and Portugal can be described by either one of the two hypothesis, the density driven and the size driven? This apparent ambiguity relies in the linear relation between population size and population density ($\rho \propto N$) observed in the urban centers for these two countries (Section *Population size and Population density*). However, for Brazil and Germany centers, only the size driven approach appears to explain the scaling; thus, we conclude that the *size* driven approach has a great explanatory power than the *density* driven hypothesis.

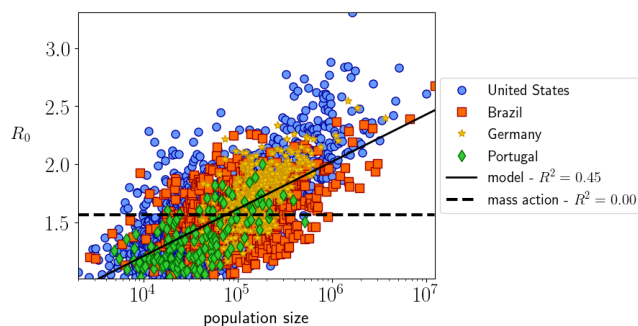


Fig. 3. R_0 vs the population size for urban regions of four different countries: the US, Brazil, Germany, and Portugal (see Materials and Methods section for fit methodology).

Marginal deviation and mobility reduction. The dispersion of points above or below the average – given by Eq. 5 – corresponds to areas where the epidemic spreads faster or slower than the average. We see in Fig. 3 that for the same population size we have municipalities with higher or lower spread, which besides the different procedures in reporting data, it could be telling something about the differences in dealing with the epidemic in different places.

Thus, we compare the marginal deviation of R_0 (the difference between each empirical value and the corresponding to the best fit, according to the theory) with the mobility reduction in each county. We estimate the mobility reduction for each county according to the data provided by *Google mobility* for the US –the only one of the four countries with municipal-level data available– on the percentage variation of time that people stay at home (see Materials and Methods section for data source and methodology).

In Fig. 4 we present the results as an histogram, where we can see a clear correlation between the reduced mobility and the the marginal deviation of R_0 . The higher the reduced mobility, the more negative the marginal deviation. In other words, the less the people move, the lower the value of R_0 relative to the expected one for a given city size.

More generally, the marginal deviations may evidence the differences in social and political aspects regarding the COVID-19 problem in different countries. In any case, our results are a warning for large cities to act promptly, applying policies of social distancing and orders to stay home.

R_0 and population size

We have found that the size driven hypothesis, besides explaining many scaling demographic characteristic of urban centers (6), it also makes sense of the observed data regarding the COVID-19 epidemics. Moreover, after the fit of Eq. 5 for the four countries (Fig. 3), we obtain a universal scaling relation for the COVID-19 basic reproductive number ²

$$R_0 = 0.18 \ln(0.1N) = 0.42 \log_{10}(0.1N). \quad (6)$$

This simple logarithmic relation between R_0 and the city size condenses both the theory and the observations regard-

²Our estimation method for R_0 is valid for urban center bigger than $N \approx 3000$, as it requires the number of cumulative cases to be small compared to the population size. For smaller towns, this is not always the case.

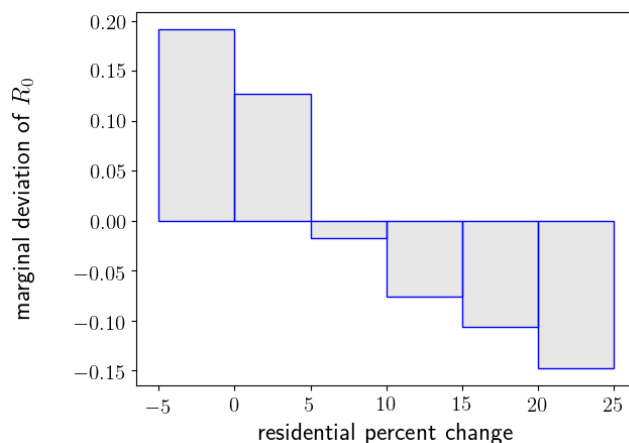


Fig. 4. Average marginal deviation of R_0 vs average mobility reduction for counties of the United States.

ing the spread of the COVID-19 pandemic in urban centers. Despite being simple, several conclusions can be made from it. It states that the mass action hypothesis –Anderson and May (10) call it proportionate mixing– is not generally valid, so the pattern of transmission depends on the city size. Moreover, Eq. 6 tell us quantitatively how its population size would impact a city in case of an epidemic. Indeed, it indicates that a multiplicative increase in population size results in the additive increase in the basic reproductive number. Therefore, the mass action hypothesis (R_0 independent of N) overestimates the epidemic impact in small cities, while it underestimates for larger ones.

With respect to the non pharmacological measurements to fight the epidemic, like the stay at home recommendation or lockouts, they are normally based on city-independent basic reproductive number –implicitly following the mass action hypothesis. According to our findings, that would make some political actions overdone for some counties and inefficient for others. A more accurate strategy, however, should consider the present contribution scaling trend. And the same applies to reopening strategies.

We remark that the values of R_0 obtained from our analysis and displayed in Fig. 3, including obviously the best fit of Eq. 6, are all within the range of the reported values in previous works (11–17).

In the following subsections we present three important conclusions that follow from the city population scaling of R_0 .

Time dependent scaling. At the early stages of the epidemic, during the exponential phase, the time evolution of number of cases \mathcal{C} obeys (see Materials and Methods section for a complete theoretical analysis):

$$\ln \mathcal{C}(t) - \ln \mathcal{C}(t_0) = \frac{R_0 - 1}{\tau} (t - t_0) \quad (7)$$

which, using Eq. 6 for R_0 and $\tau = 8$ days (18, 19), results in:

$$\ln \mathcal{C}(t) - \ln \mathcal{C}(t_0) = 0.023(t - t_0) \ln(N) - 0.177(t - t_0), \quad (8)$$

That is a power-law relation for the number of cases in term of the population size: $\mathcal{C}(t) \propto N^\alpha$, where the scaling factor α

($\alpha = 0.023t$ from Eq. 8) increases with time t . Figure 5 confirms this scaling for the US counties, while a recent empirical study shows a similar behavior for Brazilian cities (20). This is another evidence of the mass action hypothesis failure, since within it, $\alpha = 0$ would be expected. In contrast, the exponent that increases over time provides further evidence in favor of the model, diverging from the mass action predictions as time increases.

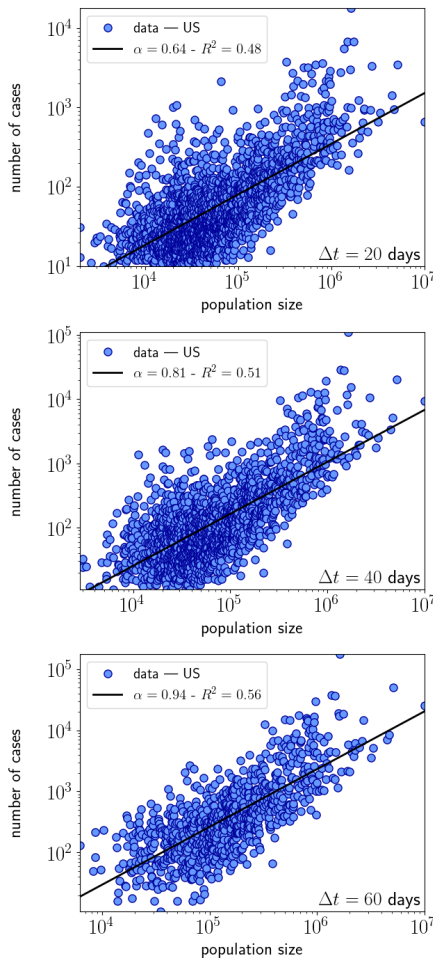


Fig. 5. Relation between the number of confirmed cases and the population size of the US counties, at three different Δt days (20, 40, and 60) after the 10th confirmed case.

Epidemic peak. The so-called *peak* of an epidemic denotes the time and the value at that time when the number of active cases, I_p , is at its maximum. Since a fraction of it requires hospitalization, a good regional/national health resources management should consider the way that number varies with the city size. According to our findings, the planning for a proportional distribution of ICU beds would lead to an underestimation of the necessary resources, with the danger of overloading the city’s health system.

The fraction of active cases at the epidemic peak, I_p/N , is related to the basic reproductive number through the relation in Eq. 11 which, using Eq. 6, can be related to the city size, as we display in Fig. 6(a), for the theoretical expected fraction of active cases at the peak as a function of population size. Ev-

idently, larger cities have relatively more active cases. More precisely, the *per capita* active cases grow approximately – fitting the exact expression– with $\ln(0.89N)$. It is worth mentioning that the theoretically predicted epidemic peak is hardly reached, because the non-pharmacological measures usually succeed in lowering the effective R_0 before that.

Community immunity. When a certain fraction of the population is immune –i.e. recovered from the infection–, the epidemic can not continue to evolve. At that point, any further outbreak would fade away. That condition is called *community immunization* and it is expressed mathematically in Eq. 12. Once more, by means of Eq. 6, we can express the fraction of people for community immunization, C^*/N , in terms of the city size, as shown in Fig. 6(b).

As the necessary fraction of the population increases fast with R_0 and this in turn increases with size, it ends up being another epidemic quantity which scales with population size – in a rather more complex relation, as can be seen in Fig. 6. In a nutshell, larger cities would require more people immunized to achieve the desired *community immunization*. This not only affects the easing of restrictions eventually made based on sample draw of antibody prevalence, but it would have to be taken into account, once a vaccine is available. The economic downside of this aspect would be eventually counterbalanced by the city income, which scales with the population too (6).

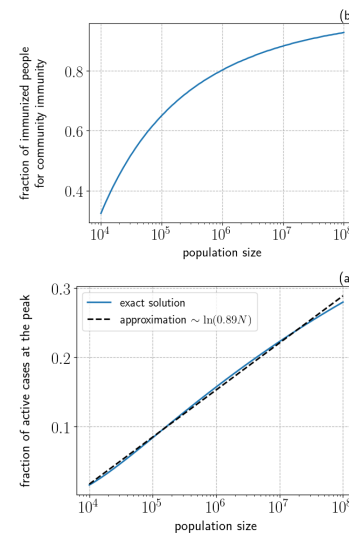


Fig. 6. Theoretical expected fraction of (a) active cases at the epidemic peak and (b) people necessary for community immunity as a function of city population size.

Materials and Methods

Demographics. With municipal-level data of population density and population size for the United States (21, 22), Brazil (23, 24), Germany (25), and Portugal (26, 27), we fit a power-law function $\rho \propto N^\lambda$ using a linear regression of the log-transformed variables, in the form $\ln(\rho) = k_1 + \lambda \ln(N)$, where k_1 is a constant. We compare the result against the linear function $\rho \propto N$, $\ln(\rho) = k_2 + \ln(N)$, and the constant function $\rho \propto 1$, $\ln(\rho) = k_3$, where k_2 and k_3 are constants.

We quantify the goodness of fit by computing the R^2 for these three models, using log-transformed variables.

Epidemic Model. To conduct our proposed analysis, which is to check how the COVID-19 transmissibility scale with the density/size of the urban population, we use the standard SIR model at the municipal level. We consider a population of size N divided in the following epidemiological compartments: susceptible S , infected I , and removed \mathcal{R} (recovered, immunized, or dead individuals). The dynamics of these compartments is driven by the following system of differential equations (4):

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta}{N}IS \\ \frac{dI}{dt} &= \frac{\beta}{N}IS - \frac{I}{\tau} \\ \frac{d\mathcal{R}}{dt} &= \frac{I}{\tau}\end{aligned}\quad (9)$$

where $\beta = pC$ is the transmission rate and τ is the infectious period. The N factor in the denominator makes β a disease only parameter, supposedly independent of the size or any other characteristic of the population. Called *mass action hypothesis* (4), it assumes that the *per capita* number of contacts is independent of the population size, resulting in similar patterns of transmission, whether it is a town or a large city (10). Now, defining the cumulative number of cases $\mathcal{C} \equiv N - S$, from Eq. 9 we got

$$\frac{d\mathcal{C}}{dt} = \frac{\beta}{N}I(N - \mathcal{C})$$

and

$$\frac{dI}{dt} = \frac{\beta}{N}I(N - \mathcal{C}) - \frac{I}{\tau}.$$

From the two above equations we have that

$$\begin{aligned}\frac{dI}{d\mathcal{C}} &= \frac{dI}{dt} \Big/ \frac{d\mathcal{C}}{dt} = 1 - \frac{N}{\beta\tau(N - \mathcal{C})} \Rightarrow \\ \Rightarrow I(\mathcal{C}) &= \mathcal{C} + \frac{N}{\beta\tau} \ln\left(\frac{N - \mathcal{C}}{N - I_0}\right)\end{aligned}\quad (10)$$

And from there, we get for the time derivative of \mathcal{C}

$$\frac{d\mathcal{C}}{dt} = \beta \frac{(N - \mathcal{C})}{N} \left[\mathcal{C} + \frac{N}{\beta\tau} \ln\left(\frac{N - \mathcal{C}}{N - I_0}\right) \right]$$

with the initial condition $I(t = 0) = \mathcal{C}(t = 0) = I_0$. At the early stages of the epidemic, when $I_0 < \mathcal{C} \ll N$, we can use a first-order approximation for the expression above to get

$$\frac{d\mathcal{C}}{dt} \approx \left(\frac{R_0 - 1}{\tau}\right)\mathcal{C} \Rightarrow \ln\mathcal{C}(t) = \ln I_0 + \left(\frac{R_0 - 1}{\tau}\right)t,$$

where $R_0 = \beta\tau$ is the *basic reproductive number*. We use it here to estimate the value of R_0 for each municipality as

$$R_0 = 1 + \tau \left[\frac{\ln\mathcal{C}(t_0 + \Delta t) - \ln\mathcal{C}(t_0)}{\Delta t} \right],$$

where t_0 is the first day since the 10^{th} confirmed case, $\Delta t = 30$ days to avoid any possible within-month measurement bias, and the infectious period is $\tau = 8$ days (18, 19). For that purpose, we use the municipal level time series data of confirmed cases for the United States (28), Brazil (29), Germany (30), and Portugal (31). As for the mobility reduction for the US counties, we measure it during the epidemic in relation with the pre-pandemic standards. To do so, we estimate it with the average of *residential percent change from baseline* variable of *Google mobility reports* (32) over the period $[t_0, t_0 + \Delta t]$.

Fit methodology. As explained in *Contact rate scaling theory* Section, we expect that $R_0 = k_1 + k_2 \ln x$, where x may be the population size or the population density (see *Demographics* for data source) and k_1 and k_2 are constants. We fit the relation using a linear regression with the log-transformed variable x and compare the result against the constant function $R_0 = k_3$, where k_3 is a constant. We quantify the goodness of fit by computing the R^2 for the two models.

Some predictions. The peak of an epidemic is when the number of active cases I is maximum. So, in that situation, we have

$$\frac{dI}{dt} \Big|_p = \frac{I}{\tau} \left(\frac{R_0 S}{N} - 1 \right) \Big|_p = 0 \Rightarrow S_p = \frac{N}{R_0} \Rightarrow \mathcal{C}_p = N - \frac{N}{R_0}.$$

Inserting this result for \mathcal{C}_p in the Eq. 10 for $I(\mathcal{C})$, the fraction of active cases during the peak is given by

$$\begin{aligned}\frac{I_p}{N} &= \frac{\mathcal{C}_p}{N} + \frac{1}{R_0} \ln\left(\frac{1 - \mathcal{C}_p/N}{1 - I_0/N}\right) \Rightarrow \\ \Rightarrow \frac{I_p}{N} &= 1 - \frac{1}{R_0} - \frac{1}{R_0} \ln(R_0),\end{aligned}\quad (11)$$

since $I_0 \ll N$.

In the equilibrium (community immunity), the cumulative number of cases is constant. So, the final fraction of infected persons is given the the solution of the transcendental equation

$$\begin{aligned}\frac{d\mathcal{C}}{dt} \Big|_* &= \frac{N}{\tau} \left(1 - \frac{\mathcal{C}^*}{N} \right) \left[R_0 \frac{\mathcal{C}^*}{N} + \ln\left(\frac{N - \mathcal{C}^*}{N - I_0}\right) \right] = 0 \Rightarrow \\ \Rightarrow R_0 \frac{\mathcal{C}^*}{N} &+ \ln\left(1 - \frac{\mathcal{C}^*}{N}\right) = 0,\end{aligned}\quad (12)$$

since $I_0 \ll N$.

ACKNOWLEDGEMENTS

SG acknowledges CAPES for support, under fellowship #003/2019 - PROPG - PRINT/UFRGS, and URPP Social Networks of University of Zürich and Prof. Claudio Tessone for hospitality. S.G. also acknowledges Dr. Eduardo N. Warman for valuable comments.

Bibliography

1. Albertine Weber, Flavio Ianelli, and Sebastian Goncalves. Trend analysis of the covid-19 pandemic in china and the rest of the world. *arXiv preprint arXiv:2003.09032*, 2020.
2. Alex Arenas, Wesley Cota, Jesús Gómez-Gardenes, Sergio Gómez, Clara Granell, Joan T Matamalas, David Soriano-Panos, and Benjamin Steinegger. A mathematical model for the spatiotemporal epidemic spreading of covid19. *MedRxiv*, 2020.

- Ben-Hur Francisco Cardoso and Sebastián Gonçalves. Urban scaling of covid-19 epidemics. *arXiv preprint arXiv:2005.07791*, 2020.
- Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.
- Luis MA Bettencourt. The origins of scaling in cities. *science*, 340(6139):1438–1441, 2013.
- W. Pan, G. Ghoshal, C. Krummel, M. Cebrian, and A. Pentland. Urban characteristics attributable to density-driven tie formation. *Nature Communications*, 4:1961, 2013.
- Hao Hu, Karima Nigmatulina, and Philip Eckhoff. The scaling of contact rates with population density for the infectious disease models. *Mathematical biosciences*, 244(2):125–134, 2013.
- Ottar N Bjørnstad, Bärbel F Finkenstädt, and Bryan T Grenfell. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series sir model. *Ecological monographs*, 72(2):169–184, 2002.
- Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PLoS one*, 7(5), 2012.
- R.M. Anderson and R.M. May. *Infectious Diseases of Humans*. Oxford University Press, 1991.
- Joseph T Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, 395(10225):689–697, 2020.
- Zhanwei Du, Lin Wang, Simon Cauchemez, Xiaoke Xu, Xianwen Wang, Benjamin J Cowling, and Lauren Ancel Meyers. Risk for transportation of coronavirus disease from wuhan to other cities in china. *Emerging infectious diseases*, 26(5):1049, 2020.
- Julien Riou and Christian L Althaus. Pattern of early human-to-human transmission of wuhan 2019 novel coronavirus (2019-ncov), december 2019 to january 2020. *Eurosurveillance*, 25(4):2000058, 2020.
- Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368(6490):489–493, 2020.
- Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The lancet infectious diseases*, 2020.
- Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, 2020.
- Shi Zhao, Qianyin Lin, Jinjun Ran, Salihu S Musa, Guangpu Yang, Weiming Wang, Yijun Lou, Daozhou Gao, Lin Yang, Daihai He, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International journal of infectious diseases*, 92: 214–217, 2020.
- Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.
- Roman Wölfel, Victor M Corman, Wolfgang Guggemos, Michael Seilmaier, Sabine Zange, Marcel A Müller, Daniela Niemeyer, Terry C Jones, Patrick Vollmar, Camilla Rothe, et al. Virological assessment of hospitalized patients with covid-2019. *Nature*, 581(7809):465–469, 2020.
- Haroldo V Ribeiro, Andre S Sunahara, Jack Sutton, Matjaz Perc, and Quentin S Hanley. City size and the spreading of covid-19 in brazil. *arXiv preprint arXiv:2005.13892*, 2020.
- County Population Totals: 2010-2019, available at www.census.gov/data/datasets/timeseries/demo/popst/2010s-counties-total.html. Accessed: 2020-06-03.
- 2010 Census Urban and Rural Classification and Urban Area Criteria, available at www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html. Accessed: 2020-06-03.
- Estimativas da População, available at www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html. Accessed: 2020-06-03.
- Área Territorial 2018, available at www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html. Accessed: 2020-06-03.
- Kreisfreie Städte und Landkreise nach Fläche, available at www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/04-kreise.html. Accessed: 2020-06-03.
- População residente por local de residência, available at www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contexto=pi&indOcorrCod=0008273. Accessed: 2020-06-03.
- Densidade Populacional por local de residência, available at www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008337. Accessed: 2020-06-03.
- CSSEGISandData / COVID-19, available at github.com/CSSEGISandData/COVID-19. Accessed: 2020-06-03.
- Brasil IO / COVID-19, available at brasil.io/covid19/. Accessed: 06-03.
- COVID-19 Germany GAE, available at github.com/jgehrcke/covid-19-germany-gae/. Accessed: 2020-06-03.
- Dados relativos à pandemia COVID-19 em Portugal, available at github.com/dssg-pt/covid19pt-data. Accessed: 2020-06-03.
- COVID-19 Community Mobility Reports, available at www.google.com/covid19/mobility/. Accessed: 2020-06-03.

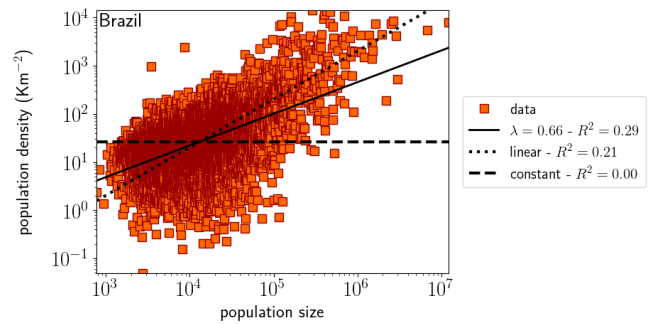


Fig. S1. Population density (ρ) vs population size (N) at the municipal level for Brazil, represented by red squares (data). The solid line is a fit to a power law, $\rho \propto N^\lambda$, while the dotted line is a fit to a linear relation, $\rho = kN$, with $k = 0.002 \text{ Km}^{-2}$. The dashed line corresponds to the average density. See Supplementary Materials for data source and fitting methodology.

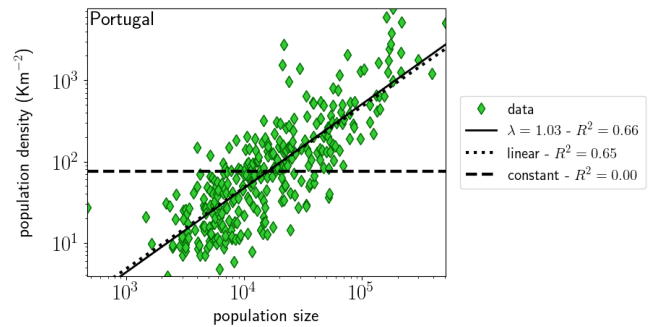


Fig. S2. Population density (ρ) vs population size (N) for Portuguese municipalities, represented by green diamonds (data). The solid line is a power law fit, $\rho \propto N^\lambda$, while the dotted line is a fit to the $\rho = kN$ relation with $k = 0.00047 \text{ km}^{-2}$. Both fits have almost equal correlation coefficient, $R = 0.81$. The dashed line represents the average density.

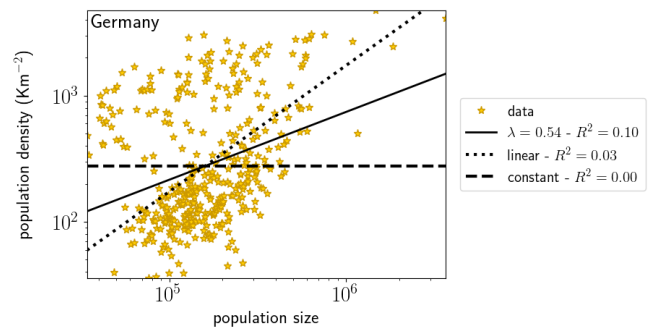


Fig. S3. Population density (ρ) vs population size (N) at the municipal level for Germany, represented by yellow stars (data). The solid line is a power law fit, $\rho \propto N^\lambda$, while the dotted line are linear fit $\rho = kN$, with $k = 0.0017 \text{ km}^{-2}$. The dashed line corresponds to the average density.

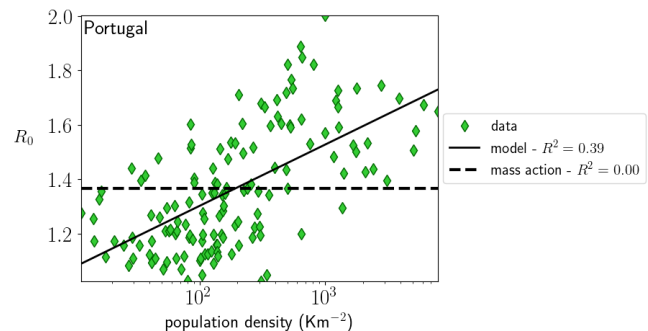


Fig. S4. R_0 vs the population density for different municipalities of Portugal.

Supplementary Information Appendix

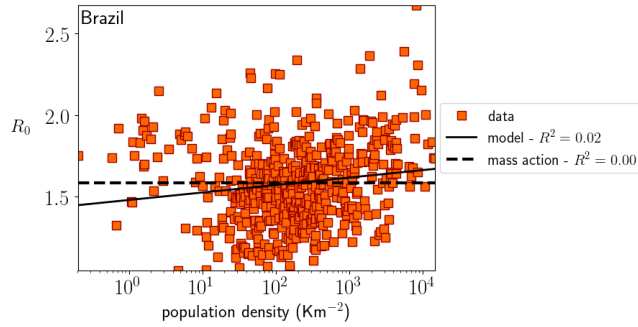


Fig. S5. R_0 vs the population density for different cities of Brazil.

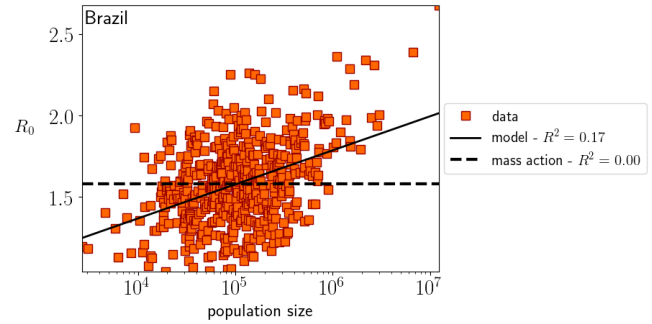


Fig. S9. R_0 vs the population size for different cities of Brazil.

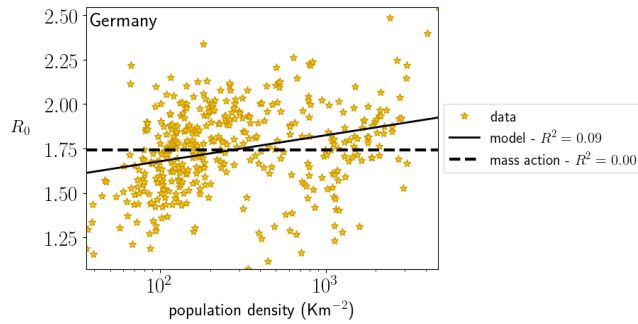


Fig. S6. Comparison between R_0 and the population density for different cities of Germany.

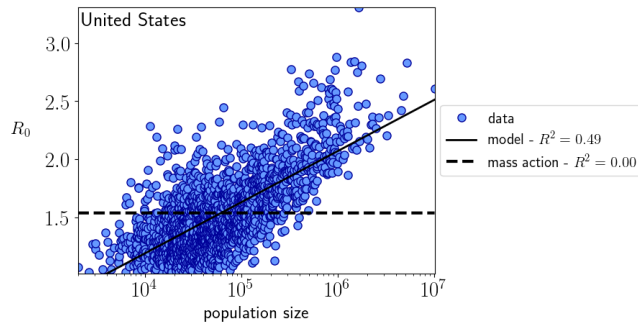


Fig. S7. R_0 vs the population size for different counties of the United States.

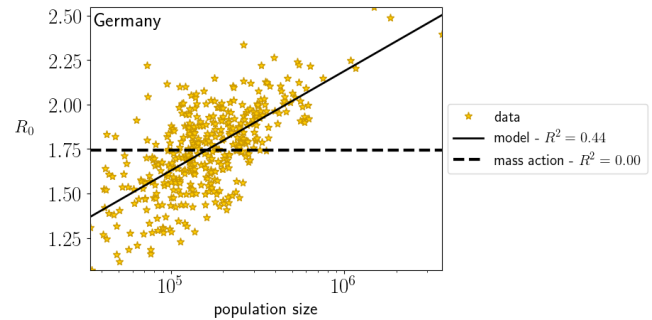


Fig. S10. R_0 vs the population size for different cities of Germany.

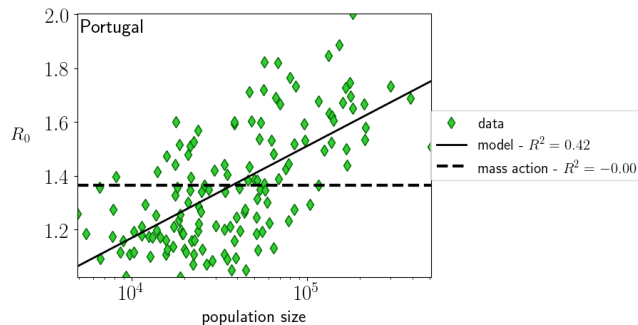


Fig. S8. R_0 vs the population size for different municipalities of Portugal.