Urban Scaling of COVID-19 Epidemics

A Preprint

Ben-Hur Francisco Cardoso  
Instituto de Física - UFRGS  
ben-hur.cardoso@ufrgs.br

Sebastián Gonçalves  
Instituto de Física - UFRGS  
sebastiangoncalves@gmail.com

May 19, 2020

Abstract

Susceptible-Infected-Recovered (SIR) mathematical models are in high demand due to the COVID-19 pandemic. They are used in their standard formulation, or through the many variants, trying to fit and hopefully predict the number of new cases for the next days or weeks, in any place, city, or country. Such is key knowledge for the authorities to prepare for the health systems demand or to apply restrictions to slow down the infectious curve. Even when the model can be easily solved —by the use of specialized software or by programming the numerical solution of the differential equations that represent the model—, the prediction is a non-easy task, because the behavioral change of people is reflected in a continuous change of the parameters. A relevant question is what we can use of 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 of use for São Paulo. With this idea in mind, we present an analysis of a spreading-rate related measure of COVID-19 as a function of population density and population size for all US counties, as long as for Brazilian cities and German cities. Contrary to what is the common hypothesis in epidemics modeling, we observe a higher per-capita contact rate for higher city’s population density and population size. Also, we find that the population size has a more explanatory effect than the population density. A contact rate scaling theory is proposed to explain the results.

Keywords COVID-19, Epidemics, Urban scaling, SIR model

1 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). Presently, the epicenter is in the US, while cases still are growing in many European countries. Soon, it will shift its center of gravity to Russia or Brazil, where the epidemic has the potential to hit even worse than in the US.

Many groups still struggle to get precise median and long term predictions of the number of expected cases, especially hospitalized or ICU ones. Such is because the epidemic parameters are continuously changing as the population changes its behavior, with or without government interventions. However, some general properties can already be identified across aggregated data, specifically related to demographic characteristics.

To give ground to our proposed analysis, which is to compare empirical municipal-level data in some countries, we use a version of the SIR model, which we call the SIRD model, because of the fourth D compartment. In this simple model, each municipal-level region (city or county) with population size N and land area A is composed of the following epidemiological compartments: susceptible (S), infected (I), recovered (R), and dead (D). Considering only within-county transmission, the dynamics of these compartments is driven by the following system of differential equations:
\[
\frac{dS}{dt} = -\frac{\beta}{N} IS \\
\frac{dI}{dt} = \frac{\beta}{N} IS - \gamma I \\
\frac{dR}{dt} = (1 - \phi)\gamma I \\
\frac{dD}{dt} = \phi\gamma I,
\]
(1)

where \( \beta \) is the transmission rate, \( \gamma \) is the removal rate, and \( \phi \) is the case fatality rate. The \( N \) factor in the denominator makes \( \beta \) a disease only parameter, supposedly independent of the size or other characteristic of the population. Indeed, in the book of Keeling and Rohani\(^\text{[2]}\) this formulation is referred to as frequency-dependent (or mass action) transmission. Called it proportionate mixing by Anderson and May\(^\text{[3]}\), it assumes that the number of contacts is independent of the population size, resulting in similar patterns of transmission, whether it is a town or a large city. However, in the unprecedented evidence that we are collecting from the ongoing COVID-19 pandemic, that common intuition seems not to be generally valid. On the opposite side, there is the pseudo-mass action formulation\(^\text{[2]}\), in which the infection rate is directly proportional to the population size—which is not usually applied to human infectious diseases. Our analysis shows that none of these extreme formulations can satisfactorily explain the available COVID-19 data. The best fit corresponds to a formulation that is somehow in between those ones, and which can be explained in terms of a contact rate scaling theory.

### 2 Population size and Population density

The time evolution of these compartments is governed by the three parameters, \( \phi \), \( \gamma \), and \( \beta \). The last one can be factorized as \( \beta = pC \), where \( p \) is the probability of transmission and \( C \) is the per capita contact rate\(^\text{[4]}\). The probability of infection \( p \) is a characteristic of the disease, most likely universal, and a key to epidemics because if it is too low, we would probably not have an outbreak.

\( C \), on the other side, condenses all the human factors that give rise to different epidemic patterns in different places, countries, or cultures. It is the only parameter that non-pharmaceutical interventions, like activity restrictions or lock-downs, can modify. Yet, we will restrict ourselves here to its urban dependency. There are two main competing hypothesis that try to explain how \( C \) 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\(^\text{[5]}\), the second one assumes that the length traveled by the individuals is invariant of the city’s size\(^\text{[6]}\).

Intriguingly, based on data of disease transmission in the United States, both approaches appear to be valid\(^\text{[4, 6, 7]}\). The reason for this is the quasi-linear correlation between density and size population of the US’s counties, as shown in Fig.\(^\text{[1]}\). 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, where \( k = A^{-1} = 0.00059km^{-2} \). An almost constant density across counties, or no correlation between \( \rho \) and \( N \) cannot explain the data. Note that from the value of the constant \( k \) we can obtain a typical county diameter in the United States of 46.5km.

In addition to the US, we study the COVID-19 transmission in Brazil’s and Germany’s cities. In these two countries, the city’s population size does not correlate well with their population density (see Fig.\(^\text{[2]}\)). The linear fitting, \( \rho = kN \), is weak for the Brazilian cities and almost nonexistent for Germany. The constant case also cannot explain the data. Since there is no correlation between \( \rho \) and \( N \), we can use these two countries to check the validity of the population size-driven or the population density-driven approaches. The results can be useful during the present COVID-19 pandemic and for futures ones.

### 3 Contact rate scaling theory

Let assume individuals distributed uniformly in a two-dimensional space according to a density \( \rho \). As introduced by Noulas \textit{et al}\(^\text{[8]}\), we can expect that the individual \( j \) interacts with the individual \( i \) with probability

\[
P_{ji} = \left( \frac{1}{\text{rank}_i(j)} \right)^{1-\alpha},
\]

\(^\text{1}\)It can be drastically attenuated or even suppressed by the use of masks, for example.
Figure 1: Population density ($\rho$) vs population size ($N$) for US counties, represented by blue dots (data). The black line is a power law fitting $\rho \propto N^\lambda$, while the red line is a linear fitting $\rho = kN$. Both have equal correlation coefficient, $R = 0.86$. The green line represents the average density. See Section 4 for the data source.

Figure 2: Population density ($\rho$) vs population size ($N$) at the municipal level for Brazil (left) and Germany (right), represented by blue dots (data). Red lines are linear fitting $\rho = kN$, with $k = 0.002$ for Brazil and $k = 0.0017$ for Germany. The green lines corresponds to the average density of each country. See Section 4 for the data source.

where rank$_{i}(j)$ is the number of neighbors closer to $i$ than $j$ and $0 \leq \alpha \leq 1$ is a scaling factor. Assuming that the distance between these two individuals is $r$, we have that

$$\text{rank}_i(j) = \rho \pi r^2 \Rightarrow P_{ji} = P(r) = \left(\frac{1}{\rho \pi r^2}\right)^{1-\alpha}.$$ 

First, since $0 \leq P \leq 1$, we must impose a bottom cutoff radius $r_0$ such that

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

Secondly, it is natural to assume an upper cutoff radius $r_1$ for long distances such that $P(r > r_1) = 0$. So, the per capita contact rate is given by

$$C_\alpha = \frac{1}{2} \int dr \ (2\pi r \rho) P(r) = \frac{1}{2} \left[ \pi r_0^2 + \int_{r_0}^{r_1} dr \ (2\pi r \rho) P(r) \right] \Rightarrow C_\alpha = \frac{(A_1 \rho)^\alpha + \alpha - 1}{2\alpha},$$ 

where $A_1 \equiv \pi r_1^2$ is the coverage area of individual mobility and the $1/2$ factor eliminate the double counting. This result generalizes the $\alpha = 0$ case deduced by Krumme et al [6], where

$$C_0 = \lim_{\alpha \to 0} \left[ \frac{(A_1 \rho)^\alpha + \alpha - 1}{2\alpha} \right] = \frac{1 + \ln(A_1 \rho)}{2}.$$
According to the population size-driven approach, the mobility increases with \( A \); \( A_1 = \kappa A \), then from Eq. 2 we have
\[
C_\alpha = C_\alpha(N) = \frac{(\kappa N)^\alpha + \alpha - 1}{2\alpha}.
\]
On the other hand, the population density driven approach states that \( A_1 \) is invariant, thus from the same Eq. 2 we get
\[
C_\alpha = C_\alpha(\rho) = \frac{(A_1 \rho)^\alpha + \alpha - 1}{2\alpha}.
\]

4 Data

We use the municipal-level time series of confirmed cases and deaths for United Sates [9], Brazil [10] and Germany [11]. Also, we use the municipal-level population size and land area for United Sates [12, 13], Brazil [14, 15] and Germany [?].

5 Methodology

Due to social distancing measures, it is expected that the value of \( \beta \) varies in time, but we can consider it as a constant for a sufficient short interval. So, let be \([t, t+\Delta t]\) such that \( S(t) < S(t+\Delta t) \) and \( D(t) > D(t+\Delta t) \). Assuming that \( \beta(t) \) is constant in this interval, we get from Eq. 1
\[
\frac{dS}{dD} = \frac{dS}{dt} \frac{dD}{dt} = -\left( \frac{\beta}{N\phi\gamma} \right) S \Rightarrow S(D) = Be^{-D/D_0}, \quad D_0 = N\phi\gamma/\beta,
\]
where \( B \) is the integration constant. Aiming to cancel some day of week seasonality bias, we choice \( \Delta t = 1 \) week. Now, noting the \( S \) is the population size minus the confirmed cases, we can construct a weekly time-series of \( D_0 \) (see Fig. 3 for an example) such that
\[
\frac{S(t)}{S(t+\Delta t)} = e^{D(t+\Delta t) - D(t)} \Rightarrow D_0(t) = \frac{D(t+\Delta t) - D(t)}{\ln S(t) - \ln S(t+\Delta t)}.
\]

Figure 3: Weekly evolution of \( D_0 \) of the New York county of US. We can note the maximum value in the 5th week, at the beginning of the PAUSE order [16].

Assuming that the social distancing measures are sufficiently distributed among cities and time, we can expect that \( D_0 \) varies in time around a value proportional to the theoretical one.

With the population density-driven contact rate hypothesis, we relate the population density of each city with its \( \langle D_0 \rangle / A \) value, where \( \langle D_0 \rangle \) is the time average of \( D_0(t) \) for this location. So, in this framework, we have
\[
\frac{\langle D_0 \rangle}{A} \propto \left( \frac{\phi\gamma}{p} \right) \frac{\rho}{C(\rho)} \propto \frac{\alpha\rho}{(A_1 \rho)^\alpha + \alpha - 1}.
\] (3)
Considering the population size-driven contact rate hypothesis, we plot the population size of each city and its \( \langle D_0 \rangle \), where \( \langle D_0 \rangle \) is the time average of \( D_0(t) \) for this location. In this approach, we expect

\[
\langle D_0 \rangle \propto \left( \frac{\phi \gamma}{p} \right) \frac{N}{C(N)} \propto \frac{\alpha N}{(\kappa N)^{\alpha + \alpha - 1}}. \tag{4}
\]

### 6 Results

#### 6.1 Population density driven contact rate

If the mass action (constant contact rate) is valid, we expect that \( \langle D_0 \rangle / A \) scales linearly with \( \rho \). In the another extreme, if the pseudo-mass action (\( C \propto N \)) is valid, we expect that \( \langle D_0 \rangle / A \) is a constant.

In Figs. 4, 5 and 6 we show, respectively, the comparison between \( \langle D_0 \rangle / A \) and the population density for different counties of United States, cities of Brazil, and cities of Germany. We can note that the model (in Eq 3) provides a good fit for United States, better than the mass action hypothesis. However, this not happens in Brazil and Germany, where the model (in Eq 3) have almost the same predictability that this hypothesis. The pseudo-mass-action cannot explain these results. This is also true in a more general scope, as shown in Fig. 7.

This result can indicate that the contact rate is, in fact, related with population size and not with the population density. The case of United states can be explained by the linear scaling between their county’s population size and population density, as show before in Fig. 1.

![Figure 4: Comparison between \( \langle D_0 \rangle / A \) and the population density for different counties of United States.](image)

#### 6.2 Population size driven contact rate

If the mass action (constant contact rate) is valid, we expect that \( \langle D_0 \rangle \) scales linearly with \( N \). In the another extreme, if the pseudo-mass action (\( C \propto N \)) is valid, we expect that \( \langle D_0 \rangle \) is a constant.

In Figs. 8, 9 and 10 we show, respectively, the comparison between \( \langle D_0 \rangle \) and the population size for different counties of United States, cities of Brazil, and cities of Germany. Now, we can note that the model (in Eq 3) provides a good fit for the three countries and are better than the mass action hypothesis. The pseudo-mass action cannot explain these results. This can be viewed in a more general scope in Fig. 11, where we found universally that \( \alpha \approx 1/3 \).

### 7 Scaling and Scale

Our main hypothesis is that closer people interact more frequently. So, we expect that if we increase the geographical scale (counties \( \rightarrow \) metropolitan areas \( \rightarrow \) states), we reduce the dependence between contact rate and the population...
size. In Fig. 12 and 13 we respectively show the relation between $\langle D_0 \rangle$ and $N$ for Metropolitan Areas and States of United States. As expected, the scaling dependence is higher for geographical scales with more granularity.

The scaling $\alpha \approx 0.2$ for Metropolitan Areas is very close to the power-law scaling found in a previous work [18]. To do so, using the approximation $S \approx N$ for short-times, they measure the growth rate of confirmed cases of COVID-19 by an exponential fit between March 16th and March 25th. Here we follow the approach described in the methodology section, since it not involves approximations, uses all available data (both confirmed cases and deaths) and allows the weekly variation of $\beta$, since now we have access of a longer period.

We aggregate the county-level to the Metropolitan Statistical Areas using delineation files of US Office of Budget and Management [17].
8 Conclusions

The epidemic dynamics are traditionally explained by two hypotheses: the mass action and the pseudo-mass action. Here we shown empirically that neither is good to describe the data. Also, we develop a theory to explain the found relation.

Our analysis and results give support to the validity of the population size driven contact rate for the COVID-19 pandemic. This result can also explain the super-linear scaling of criminality in Brazil [19], Japan [5] and United States [20]. Such is the the downside of leaving in large urban centers.

From our analysis, it is clear that the scaling is valid at the municipal, county, or city level. If we make it broader at regions, province or state level, it is washed out by the different scales averaged over such large regions.

This conclusion can provide useful insight regarding the urgent problem that cities, and the world in general, are facing. As others authors [18] already pointed out, larger cities require more strict social distancing policies. On the other side, smaller cities may relax controls before larger cities.
Figure 9: Comparison between $\langle D_0 \rangle$ and the population size for different cities of Brazil.

Figure 10: Comparison between $\langle D_0 \rangle$ and the population size for different cities of Germany.

References

[1] Albertine Weber, Flavio Iannelli, and Sebastian Gonçalves. Trend analysis of the covid-19 pandemic in china and the rest of the world. medRxiv 2020.03.19.20037192, 2020.

[2] Matt J Keeling and Pejman Rohani. Modeling infectious diseases in humans and animals. Princeton University Press, 2011.

[3] R.M. Anderson and R.M. May. Infectious Diseases of Humans. Oxford University Press, 1991.

[4] 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.

[5] Luís MA Bettencourt. The origins of scaling in cities. science, 340(6139):1438–1441, 2013.

[6] C. Krumme M. Cebrian W. Pan, G. Ghoshal and A. Pentland. Urban characteristics attributable to density-driven tie formation. Nature Communications, 4:1961, 2013.

[7] 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.
Figure 11: Comparison between $\langle D_0 \rangle$ and the population size for different cities around the world.

Figure 12: Comparison between $\langle D_0 \rangle$ and the population size for different Metropolitan Areas of United States.

[8] 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.

[9] Cssegi sand data / covid-19. [https://github.com/CSSEGISandData/COVID-19](https://github.com/CSSEGISandData/COVID-19) Accessed: 2020-05-15.

[10] Brasil io / covid-19. [https://brasil.io/covid19/](https://brasil.io/covid19/) Accessed: 2020-05-15.

[11] Covid-19 germany gae. [https://github.com/jgehrcke/covid-19-germany-gae/](https://github.com/jgehrcke/covid-19-germany-gae/) Accessed: 2020-05-15.

[12] County population totals: 2010-2019. [https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html](https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html) Accessed: 2020-05-15.

[13] 2010 census urban and rural classification and urban area criteria. [https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html](https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html) Accessed: 2020-05-15.

[14] Estimativas da população. [https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html](https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html) Accessed: 2020-05-15.

[15] Área territorial 2018. [https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html](https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html) Accessed: 2020-05-15.
Figure 13: Comparison between $\langle D_0 \rangle$ and the population size for different States of United States.

[16] No. 202: Declaring a disaster emergency in the state of new york. https://www.governor.ny.gov/news/no-202-declaring-disaster-emergency-state-new-york. Accessed: 2020-05-15.

[17] Metropolitan and micropolitan. https://www.census.gov/programs-surveys/metro-micro/about/delineation-files.html. Accessed: 2020-05-15.

[18] Andrew Stier, Marc Berman, and Luis Bettencourt. Covid-19 attack rate increases with city size. Mansueto Institute for Urban Innovation Research Paper Forthcoming, 2020.

[19] Andres Gomez-Lievano, HyeJin Youn, and Luis MA Bettencourt. The statistics of urban scaling and their connection to zipf’s law. PloS one, 7(7), 2012.

[20] Marcos Oliveira, Carmelo Bastos-Filho, and Ronaldo Menezes. The scaling of crime concentration in cities. PloS one, 12(8), 2017.