Applied Mathematical Modelling 121 (2023) 166–184

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Scale-free dynamics of COVID-19 in a Brazilian city

J.M.P. Policarpo[a], A.A.G.F. Ramos[a], C. Dye[b], N.R. Faria[b,c,g], F.E. Leal[d], O.J.S. Moraes[a], K.V. Parag[e], P.S. Peixoto[f], L. Buss[g], E.C. Sabino[g], V.H. Nascimento[h], A. Deppman[a,*]

[a] Instituto de Física - Universidade de São Paulo, Brazil
[b] Department of Biology, University of Oxford, UK
[c] Imperial Coll London, MRC Ctr Global Infect Dis Anal, Sch Publ Helth, London, England, UK
[d] Universidade de São Caetano do Sul, São Caetano do Sul and Programa de Oncovirologia - Instituto Nacional de Câncer, Rio de Janeiro, Brazil
[e] MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London W2 1PG, UK
[f] Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil
[g] Faculdade de Medicina - Universidade de São Paulo, Brazil
[h] Escola Politécnica - Universidade de São Paulo, Brazil

ARTICLE INFO

Article history:
Received 19 July 2022
Revised 13 March 2023
Accepted 29 March 2023
Available online 21 April 2023

Keywords:
Epidemic dynamics
Scale-free network
Fractal
Tsallis statistics

ABSTRACT

A common basis to address the dynamics of directly transmitted infectious diseases, such as COVID-19, are compartmental (or SIR) models. SIR models typically assume homogenous population mixing, a simplification that is convenient but unrealistic. Here we validate an existing model of a scale-free fractal infection process using high-resolution data on COVID-19 spread in São Caetano, Brazil. We find that transmission can be described by a network in which each infectious individual has a small number of susceptible contacts, of the order of 2–5. This model parameter correlated tightly with physical distancing measured by mobile phone data, such that in periods of greater distancing the model recovered a lower average number of contacts, and vice versa. We show that the SIR model is a special case of our scale-free fractal process model in which the parameter that reflects population structure is set at unity, indicating homogeneous mixing. Our more general framework better explained the dynamics of COVID-19 in São Caetano, used fewer parameters than a standard SIR model and accounted for geographically localized clusters of disease. Our model requires further validation in other locations and with other directly transmitted infectious agents.

1. Introduction

The pandemic of SARS-CoV-2 has had a huge impact on our way of life [1,2]. Non-pharmaceutical interventions were adopted and in many cases, they were successful in controlling virus transmission avoiding a potential collapse of health systems [3–6].

* Corresponding author.
E-mail addresses: josue.policarpo@usp.br (J.M.P. Policarpo), nuno.faria@zoo.ox.ac.uk (N.R. Faria), fabio.leal@online.uscs.edu.br (F.E. Leal), osmar.moraes@usp.br (O.J.S. Moraes), k.parag@imperial.ac.uk (K.V. Parag), ppeixoto@usp.br (P.S. Peixoto), vitnasci@usp.br (V.H. Nascimento), deppman@usp.br (A. Deppman).

https://doi.org/10.1016/j.apm.2023.03.039
0307-904X/

The approach used in epidemiological models is vast, but a common starting point is the Susceptible-Infected-Recovered (SIR) models or their variations [7–11], including a simpler formulation the Susceptible-Infected (SI) model [12]. The basic assumption is that one infected (I) individual can transmit, with a given probability, the virus to any other susceptible (S) individual in the population. The result is an exponential growth in the number of infected individuals at the beginning of an outbreak. However, analyses of previous epidemics [13] including COVID-19, show that the early growth of infected cases tends to be sub exponential, following a power-law [14–18].

New approaches to describing the spread of epidemics usually start by modifying the ordinary differential equations in an SIR model. Such modifications usually introduce new processes that influence the spread of infection. One example is the awareness of the susceptible individuals as an epidemic advances [19], which may result in self-imposed social distancing. The inclusion of void nodes [20] is a method for increasing the complexity of the network, obtaining more realistic descriptions of the infection process. A more comprehensive description of social networks can be found in Tanimoto [21]. In the present work, we derive the properties of epidemic spread by assuming a fractal structure of the social network and by studying, at first, transitions in a small network. The fractal structure allows the infection process to operate at scale in a larger population without the introduction of more adjustable parameters, resulting in a more testable and reliable model. Recent studies suggest that Complex Networks, where agents are the nodes of the network and the contacts among agents are represented by edges, may be useful to describe the spread of infection [22] due to the simple mechanisms that govern their growth [23]. The simple assumption that new nodes will connect to the existing nodes by a preferential attachment rule generates a scale-free or scale-invariant network. The attachment rule is a non-uniform probability distribution to attach a new node to an existing node i. To form scale-free networks, preferential attachment must be given by a power-law distribution such that the probability of a new node attaching to an existing node i in a network is

$$p(k_i) \propto \frac{1}{k_i^a}\,. \tag{1}$$

where k_i is the number of edges connecting the ith node and a is a constant. Several aspects of scale-free networks follow power laws, and scale-free networks can be obtained under a variety of attachment rules [24].

Scale-free networks obtained by preferential attachment mechanisms have the property of self-similarity [25], which is a predominant feature of fractals. Self-similarity means that a complex system, with an internal structure formed by an arbitrary number of components, is such that the components are identical to the main system when viewed on any scale. When a network is self-similar, cutting any branch of the network reduces the size of the network but leaves the structure of the network unchanged.

The relations between fractal dynamics and power-law distributions were first identified by Kopelman in a study of chemical reaction dynamics [26], where the reaction rate followed a power-law dependence on time. The theory of information spreading in a fractal network is described in Deppman and Segundo [27]. Complex networks are useful tools to investigate epidemic processes too [28–30], and the fractal dynamics approach has already been used to study COVID-19 [31], where the fractality was expressed by a power-law in the time dependence of one of the parameters of the SIR model. The model proposed in Abbasi et al. [14] follows a different approach. It assumes fractal dynamics to extend the mechanisms of infection in small groups to larger groups, using the self-similarity typical of fractals. In the present work, we show that such an assumption will lead to a power-law distribution of the parameters of the model.

In modelling the spread of diseases, a major question is how to formulate the mechanism of contact between infectious individuals and others in the population. Some of the key issues are: if one person is infected with the disease, how many other people will be at risk? How does the number of contacts differ from person to person? And how does this number change in face of social-distancing measures? In this paper, we describe a model for disease spread in which each infectious individual has a small number of contacts that is independent of the stage of the epidemic and of population size. We find that a “scale-free” or “fractal” model of contagion describes detailed data obtained from a city in Brazil better than traditional models of infectious diseases.

Roughly speaking, the fractal approach assumes the scale invariance of the complex structures of the epidemic dynamics, with different groups of infection in a given population. With such an assumption, the dynamics of the epidemic spreading over a large population can be related to that of a number of small contact groups, and the number of parameters necessary to describe the complex dynamics is reduced, simplifying the task of model fitting. In the model used here, “agents” designate the groups of close contacts into which the population is divided. Each agent is formed by other agents at smaller scales until one reaches the smallest agent, which is a group of individuals who have the most frequent and close contact. The fractal aspect is introduced by considering that any relevant property of the epidemic process in a large agent can be described by the same dynamic that happens in a smaller agent after the quantities are scaled to the larger population. In other words, the agents are self-similar.

With the self-similar assumption, the number of parameters of the model is reduced, and the modes of contagion in a large population can be investigated by observing what happens in a small group. By “modes of contagion” we mean the different ways in which an initially infected agent may infect a second agent in the same contact group (larger agent): Considering the group of four agents in Fig. 1, the agent (red) can be infected by the agents (a) directly, as shown in the left; (b) indirectly, through a third agent, as shown in the middle; (c) indirectly, through two other agents, as shown in the right. Note that there are three possible chains of direct infection: L-C, B-C, R-C, three chains of secondary infection: L-B-C, B-R-C, L-R-C (considering only the subgroups of infected agents, and disregarding the order of infection), and only one chain

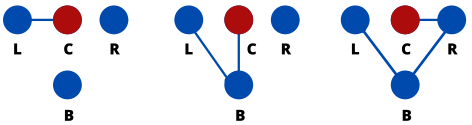

Fig. 1. Modes of contagion in a small group of four individuals ((L)eft, (C)enter, (R)ight and (B)ottom). The figures show schematically the different modes, direct (left panel) or indirect (central and right panels), that one susceptible individual can be infected in the group if one individual is initially infected.

of third-order infection L-B-R-C. We call these possibilities "modes of infection". The counting can be done by combining the number of elements in each mode. Each agent cannot infect itself, so for a group with N agents, the number of agents contributing to the number of modes is $\alpha = (N-1)$. Note that in Fig. 1 we have $\alpha = 3$. For direct contamination, we have α modes (L-C, B-C, R-C), for indirect contamination in two stage $\frac{\alpha!}{(\alpha-2)!2!}$ modes, for indirect contamination in three stage $\frac{\alpha!}{(\alpha-3)!3!}$ modes and so on. If τ is the fraction of infected agents in the group per contact and assuming that $\tau \ll 1$, we have to add powers of the τ term that correspond to secondary transmissions (power 2) and tertiary transmission (power 3), and we get

$$\nu = 1 + 3\tau + 3\tau^2 + \tau^3 . \tag{2}$$

The first term represents the first infected. The multiplicative coefficients correspond to the number of ways the contamination of each of the modes can take place. To generalise to any value of N we use the Newton Binomial expansion

$$\nu = \sum_{k=0}^{\alpha} \frac{\alpha!}{(\alpha-k)!k!}\tau^k . \tag{3}$$

Mathematically, it can be shown that the number of infected agents, ν, during the period of infection in the group is given by Abbasi et al. [14], Deppman and Segundo [27]

$$\nu = (1+\tau)^{N-1} \tag{4}$$

At small scales, our fractal approach effectively models small group (e.g., household) transmission with τ defining the probability that a person has adequate contact with an infected person for transmission to occur. This model of spreading, which also considers the various possible sequences of contact by which the infection can be transmitted across the individuals in that small group, is analogous to the chain binomial class of models (e.g., the Reed-Frost model [32]). Chain binomial models are among the most commonly used approaches for characterising spread at small scales. At larger scales, the fractal structure converges to a generalisation of the compartmental models, which are commonly used to describe transmission at population levels. In this sense, the fractal approach provides a minimalistic mechanism for bridging small and large-scale transmission dynamics.

This paper is organized as follows: in Section 2 we describe the mathematical aspects of the model and show how it generalized the SIR model; in Section 3 we describe the data and the method of analysis, and in Section 4 we discuss the results. In Section 5 we present our conclusions.

2. Mathematical description of the fractal spreading model

An epidemiological model to describe the spreading dynamics of COVID-19 was proposed in Abbasi et al. [14], where the fractal aspects of the virus transmission process observed in the analysis of real data on the transmission of the disease were included. Rather than an initial exponential increase in the number of infections, as would be expected by the SIR model, a polynomial behaviour is obtained, which means that the spreading is not as fast as one could predict by using standard SIR models. The result is in fact similar to the modified model proposed in Viboud et al. [13]. It is important to emphasize that the polynomial growth model, which was postulated in Viboud et al. [13] to approximate observed disease dynamics in Ebola outbreaks, is obtained from first principles as a consequence of the fractal model. The non-exponential behaviour is important to be considered in evaluations of the impact of different policies to combat the epidemic spreading on the rate of contamination. Roughly speaking, the fractal approach assumes the scale invariance of some complex structures of the epidemic dynamics, dividing the infected population into different groups at different scales. With such an assumption, the dynamics of the epidemic spread over a large population can be related to the process in a small group. In Fig. 2, the fractal structure of the spreading dynamics is schematically shown. At the left panel, the scaling of the agents is depicted from the smaller scale, where the agents is formed by a number of individuals, to the larger scale, where the agents are formed by a number of agents on a smaller scale. The basic idea of the fractal dynamic is that no matter the scale, the dynamic of the epidemic is the same inside any agent of the network. At large scales, the resulting network shows the self-similarity property observed in many socioeconomic networks. With the assumption of fractal dynamics, the number of parameters necessary to describe the complex epidemic dynamics is reduced (Table 1).

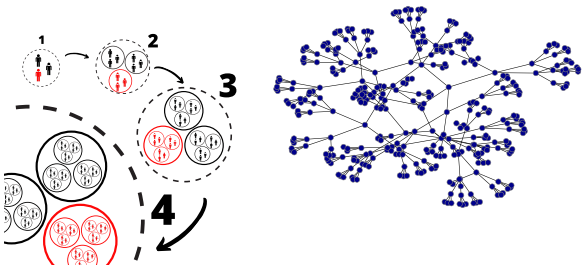

Fig. 2. Illustration of fractal approach at different scales (left). The number of people that make up each agent depends on the scale. In balloon 1, each agent in the group (dashed circle) is composed of a single persona and $N = 2$ average number of close contact, while in balloon 4, each agent is composed of 27 people and $N = 2$ average number of close contact. Self-Similarity Network example (right).

The function that describes the number of cases at the moment t is $i(t)$ and the variation of this function in the fractal model can be written as Abbasi et al. [14].

$$\frac{di}{ds}(\tau, N; s) = \left(1 + \frac{\tau s}{N}\right)^{N-1}, \tag{5}$$

where s is the susceptible population, τ (assumed small) is the transmission probability and N is the average number of close contacts of the individuals in that population. To extend the spreading mechanisms to larger populations, we assume that the population is formed by a set of agents, identified by the index j, each with approximately the same size, N. Following Ref. [14], we adopt the parameter q such that $1 - q = N^{-1}$, which leads to the differential equations of the fractal model, given by (for $t > t_{oj}$ = start of transmission)

$$\begin{cases} \dfrac{ds_j(t)}{dt} = -\kappa_j i_j^q(t)\, s_j(t) \\ \dfrac{di_j(t)}{dt} = \kappa_j i_j^q(t)\, s_j(t) + \kappa_j i_j^q(t)(t - t_{oj})\dot{s}_j(t)\,, \\ \dfrac{dr_j(t)}{dt} = -\kappa_j i_j^q(t)(t - t_{oj})\dot{s}(t) \end{cases} \tag{6}$$

The first equation corresponds to the ratio of infection in the susceptible population. The second equation has two terms on the right-hand side: the first one corresponds to the increase of the infected population due to infection of the susceptible, and is the opposite of the ratio at which the susceptible population decreases; the second term corresponds to the ratio of decrease of the infected population due to recovering. The third equation gives the ratio of increase in the recovered population, and the term at the right-hand side corresponds to the opposite of the second term in the equation for the infected population.

We assume that the initial susceptible population of the agent j is given by $s_j(t_{oj}) = s_{oj}$, to be determined in the data analysis, and that the rate of transmission, κ_j, may vary from one agent to the other. The time of introduction of the virus in agent j is t_{oj}, corresponding to the instant when the first individual of the group is infected. The removed population r_j corresponds to the individuals that are placed out of the spreading dynamics because they can no longer transmit the virus either because they have recovered or died, or because they no longer have susceptible individuals in their neighbourhood.

The expression for $i(t)$ is known in the non-extensive statistics proposed by Tsallis [33] as the q-exponential distribution. In this statistics, mean values are determined by expressions as

$$\langle a \rangle = \frac{\int a p^q(a) da}{\int p^q(a) da}, \tag{7}$$

where q is called the entropic index, and plays a role similar to the parameter q in the fractal model, and $p(a)$ is the probability density. The fact that we have q-exponential distributions and that the probability to find an infected individual appears in powers of q, shows the close connections between the Tsallis statistics and the spreading dynamics in a fractal structure.

Aside from the evident agent-based approach of the model described by the equations above, a few additional remarks are necessary. The number of parameters used for each agent is large, with q, κ_j, s_{oj} and t_{oj} being undetermined. In a general agent-based network, each agent will have independent parameters, therefore the number of parameters is proportional to the number of agents in the network. However, in the fractal network, due to the scaling invariance (or self-similarity), the number of parameters is independent of the number of agents. The parameter N, related to the agent size, can be evaluated by other means as a general property of the network or can be considered as a quantity that may vary from one agent to the other.

Observe that the term corresponding to the rate with which the removed population increases, that appears in the first and in the third equations, can be written as

$$\frac{dr_j(t)}{dt} = -[\kappa_j i_j^q(t) s_j(t)]\kappa_j(t-t_{oj}) i_j^q(t)\,. \tag{8}$$

when the second equation is used. This term is very different from its correspondent in the SIR model. Here, the removed population is determined by the epidemic process inside each agent and takes into account the fact that after some time of the first infection, all individuals in the group will be already infected through one of the possible modes of infection. Another aspect that differentiates the fractal model from the usual SIR model is the explicit time dependence of the epidemics, introduced by the used of the term $\tau_j = \kappa_j(t-t_{oj})$. Within the fractal model, the instant when the virus is transmitted to other individuals in the agent is aleatory.

We can simplify the expression above by assuming an average contamination rate for the term between brackets, multiplied by $\kappa_j(t-t_{oj})$. Taking the average of this term we have

$$\kappa_j(t-t_{oj})\kappa_j i_j^q(t) s_j(t) = -\kappa_j(t-t_{oj})\dot{s}_j(t) \sim -\langle \kappa_j(t-t_{oj})\dot{s}_j\rangle, \tag{9}$$

which corresponds to a linear approximation (because $\dot{s}(t)$ is substituted by a constant) of the behaviour of the susceptible population. Considering the period during which the virus is being transmitted inside the agent, $\Delta t_j = t_{fj} - t_{oj}$, with t_{fj} being the time when the new cases of infection in the agent cease, we have, for the first two terms at the right-hand side,

$$\langle \kappa_j(t-t_{oj})\rangle = \langle \kappa_j \Delta t_j\rangle \frac{\langle t-t_{oj}\rangle}{\Delta t_j}\,. \tag{10}$$

But we assume $\kappa_j \Delta t_j = \tau$ is constant by hypothesis, which will be discussed in Section 4.1, and $\langle t-t_{oj}\rangle \sim \Delta t_j/2$, since from observations we see that the distribution is just slightly asymmetric. Therefore, with the approximations above, we obtain

$$\frac{dr_j(t)}{dt} = \kappa_j^{\dagger} i_j^q(t)\,. \tag{11}$$

where

$$\kappa_j^{\dagger} = \frac{\tau}{2}|\langle \dot{s}_j\rangle|, \tag{12}$$

is completely determined by the parameters of the epidemic dynamics, with $\tau = \kappa_j\,\Delta t_j$, where Δt_j is the time during which the virus is circulating in the agent j. We noticed in simulations that this model results in all the susceptible populations in each agent becoming infected.

The considerations made above show that the fractal model has a structure that is similar, in some aspects, with the SIR model. However, contrary to the SIR model, in the fractal model the parameter of the removed population, $\kappa^{\dagger}$, is completely determined by the other parameters in the fractal dynamics. This happens because we are considering that the total population, s_o, will be infected, which is an unrealistic situation since, in practice, it is observed that part of the susceptible population will not be infected in the period the virus is circulating in the group. We can improve the fractal model in this aspect by including a new parameter that refers to those individuals that will be removed from the infected population before all the individuals will be infected. Defining

$$\kappa_j^{*} = (1+\gamma)\kappa^{\dagger}, \tag{13}$$

we introduce the new parameter $\gamma \geq 0$ that corresponds to the removed population, which allows us to obtain a fractal model similar to the SIR model when we substitute $\kappa^{\dagger}$ by κ^{*} in Eq. (11). Accordingly, the infected population equation will be

$$\frac{di_j(t)}{dt} = \kappa_j i_j^q(t-t_{oj})\, s(t-t_{oj}) - (1+\gamma)\kappa_{qj}^{\dagger} i_j^q(t-t_{oj})\,. \tag{14}$$

Observe that the equation above can be obtained by considering that the susceptible population in the group is $s_j(t)$ in the expression for $i(t)$. If $\gamma > 0$, the susceptible population will be larger than the total infected population, as it usually happens in the SIR model. The set of equations that describes the fractal model equivalent to the SIR model is

$$\begin{cases} \dfrac{ds_j(t)}{dt} = -\kappa_j i_j^q\, s_j \\[2ex] \dfrac{di_j(t)}{dt} = \kappa_j i_j^q\, s_j - \kappa_{qj}^{*} i_j^q\,. \\[2ex] \dfrac{dr_j(t)}{dt} = \kappa_{qj}^{*} i_j^q, \end{cases} \tag{15}$$

for $t \geq t_{oj}$. For $q = 1$, the equations above reduce to the SIR model equations. In this case, the usual notation is obtained by using the parameter β_j such that $\kappa_j = \beta_j/s_{oj}$. Equations (18) and (15) were proposed in a different and independent approach in Refs. [34,35].

From the Eq. (15), it is possible to understand another difference between the fractal model and the SIR model. While in the latter the number of close contacts increases as the population size increases, in the fractal model the two numbers are uncorrelated. This means that the number of close contacts of an individual does not increase as the population increases, reflecting the fact that the number of close contacts that an average person has does not vary significantly if that person is in a metropolis or a small city. This model describes many features and details of the COVID-19 dynamics [14].

While the transmission probability depends on the features exhibited by the different virus variants and is mostly out of our control, the number of close contacts can be reduced by social distancing and has an effective impact on the number of new cases. Therefore social distancing will be effective to reduce the number of new cases even for the new variants of SARS-CoV-2 with higher transmission probability.

It is important to stress that a clear separation between transmission probability and social distancing is never possible. For instance, the use of protective devices, such as masks, is to be considered as a reduction in the transmission probability or as social distancing? This is a question that is not even posed in the SIR model, as it does not allow a separation between the two contributions. The fractal model, however, does allow such a separation of the two contributions, as we will show below. Therefore, the question above becomes relevant and has to be addressed in some way. Our approach, in this work, is to consider as transmission probability only those features that are intrinsically related to the virus characteristics, and attribute to social distancing all features of the spreading dynamics that can be controlled by human action. Of course, this definition is arbitrary and still may not solve completely the problem of separating transmission probability from social distancing. An example is the vaccination process, which represents a human action that interferes with the virus's capability to infect individuals. However, to the present work, this aspect is irrelevant, since no vaccination took place for the entire period of analysis. The agents in the smallest groups have size $\lambda_o = 1$, i.e., they correspond to individuals. Equation (4) describes the disease transmission inside these smallest group of individuals. The scaling is made by assuming that a larger group, with population s_λ, is formed by N_λ agents of size of the order of λ, so that $s_\lambda \sim N_\lambda \lambda$, and modifying Eq. (4) to

$$\nu(s_\lambda) = \left(1 + \tau \frac{s_\lambda}{N_\lambda \lambda}\right)^{N_\lambda - 1}. \tag{16}$$

The q-exponential is usually written in the form

$$\nu = \left[1 + (1-q)\frac{\tau s}{\lambda}\right]^{\frac{1}{1-q}-1}, \tag{17}$$

with the identification of the parameter q as $1 - q = 1/N$. In the epidemic process, this is the contribution of one single agent, so the total number of infected individuals is obtained by integrating the expression above over the entire population. The number of infected agents, therefore, is of the form

$$i = \left[1 + (1-q)\frac{\tau s}{\lambda_o}\right]^{\frac{1}{1-q}} \tag{18}$$

In this fractal approach to the epidemic dynamics, it is assumed that the average number of contacts that an individual has is not strongly dependent on the population of the region in which this individual is living, that is, if she lives in a large metropolis with around 10^7 inhabitants, or if she lives in a small city with 10^3 inhabitants, her number of contacts will always be around a few tens of individuals. Of course, the meaning of close contact may vary depending on the aspect of social life one is interested in, and since here we analyse the virus spread, by close contact we mean those individuals with frequent and in-person contact, with the possibility of virus transmission. In this work, we initially consider an average number of contacts for the whole population (i.e., constant q_j), but later we investigate how small variations, relative to the total population, of the number of contacts of different agents can be observed, and we will identify some super-spreader agents.

The q-exponential function in Eq. (18) is typical of the non-extensive statistics developed by Tsallis [33], a generalization of the Boltzmann–Gibbs statistics that has found applications in many fields, including fractals [36]. The fact that it appears here is an indication of the adequacy of the non-extensive statistics to describe the epidemic dynamics. The relations between fractal structures and Tsallis statistics have been investigated in other works [36,37]. Define the rate of contagion of the agent, κ_j, such that

$$\tau = \kappa_j \, \Delta t_j, \tag{19}$$

and obtain the time-dependent function of the number of infected individuals in agent j,

$$i_j(t) = \left[1 + (1-q_j)\frac{\kappa_j s_j(t)(t - t_{oj})}{\lambda_o}\right]^{\frac{1}{1-q_j}}, \tag{20}$$

where the susceptible population is considered to be a function of time, $s_j(t)$, since those individuals that were infected are not considered susceptible anymore. There are two forms for the increase of the number of infected individuals within an agent: by the time evolution of the spreading, with a rate that is described by the derivative of $i_j(t)$; or by merging with another agent when their individuals cannot be clearly separated. In this case, the mathematical description is done by using so-called q-algebra, which is associated with Tsallis statistics, and which keeps a fixed value for q while summing the populations of the merging agents.

The infected population, as a function of time, is

$$i(t) = \sum_j i_j(t) + c, \tag{21}$$

where c is a constant representing the average daily number of new cases that come from outside and seed new clusters of infection in the city.

Here there is an important difference between the fractal model and the traditional models: rather than an exponential increase in the number of new cases, a slower increase, following the q-exponential function, is found whenever $q \neq 1$. Only in the case of $q = 1$ the SIR and SI equations are recovered.

3. Methodology

In this section, we describe the methods used to collect the data and for the analysis. The latter is divided into four parts: a test on the model consistency; an investigation on the properties of the spreading dynamics; a cross-check of the results of our analysis with additional information; and a comparison with the SIR model.

3.1. The data

The data were collected in São Caetano do Sul, a city in the metropolitan region of São Paulo, Brazil, with an estimated population of 161,957 inhabitants [38] and one of the highest Human Development Index values in the country. In April 2020, a primary care initiative offering COVID-19 care to all residents was created. Residents of the municipality aged 12 years and older with suspected COVID-19 symptoms were encouraged to contact the dedicated Corona São Caetano platform via a website (access at https://coronasaocaetano.org/) or by phone. They completed an initial screening questionnaire that included clinical and socio-demographic data. Patients meeting the suspected COVID-19 case definition were further evaluated by a health care professional and offered a home visit for sample collection for PCR testing [39]. The São Caetano platform routinely collects the patient addresses and transforms them into latitude and longitude values using an API from the Here Company (Chicago, USA). For this study, we received the lat/long values with a decreased precision of 2 decimal digits that corresponds to an area of 1.2 km^2. We also received the exact date of symptom initiation for each patient. The study was approved by the local ethics committee (CAAE 32424720.8.0000.0068). The committee waived the need for informed consent and allowed the development of an analytical dataset with no personal identification for analysis.

3.1.1. Social distancing

Social distancing adherence was investigated in a validation step of this study, and considered two different datasets. Both datasets provide a relative measure of the number of people that stayed home during a reference day, resulting in what we call the Social Isolation Index. The government of São Paulo issues an official calculation of the index based on radio signals of mobiles from network stations of the four main providers in Brazil (Oi, Vivo, Claro, Tim). This leads to an estimate of the user's location within the order of meters, depending on the number of nearby base stations. The home place is decided based on the location the user has spent the night (from 22 h to 2 h), and movements outside of the 200 m range from the house are triggered as a breach of isolation. This dataset is publicly available on a city level at the official government website (https://www.saopaulo.sp.gov.br/coronavirus/isolamento/). The second dataset uses geolocation data collected from a technology company called InLoco, which in 2020 collected positions of around one-fourth of the Brazilian population [40]. Home location is then decided in a similar way as for the network providers' data, but the breach in isolation is decided according to a change in position of around 460 m, which explains the higher percentages of isolation verified in Fig. 6. It is provided in the same way as the government data, as a Social Isolation Index aggregated by cities. During this pandemic, several studies have been using this dataset as it provides a good metric of social distancing [41–45].

3.2. The analysis

Step 1: Initially, we group the number of infected individuals in bins of increasing width, starting from bins 4-day large, until the reduced χ^2 reaches a minimum value when described by a single agent. We found that it happens for bins 30-day large.

The fitting gives the parameters of Eq. (20) s_o, κ_j and t_j with q fixed (see Section 4.1). The next step is to divide the bin width by two, double the number of agents, and repeat the process of fitting, getting, as a result, the parameters for each of the two agents, as well as the global parameter, c for this bin width. The process is repeated, doubling the number of agents and fitting the parameters until a good fit is obtained with a bin width of four days. At the end of this part, we have fitted in total 12 agents and obtained 36 best-fit parameters for those agents. The analysis of the behaviour of the parameters κ gives a power-law distribution, as we show in the next section.

Step 2: In this step we adopt, in Eq. (20), the power-law behaviour of $\kappa(s_{oj})$, that is, $\kappa(s_{oj}) = \kappa_o s_{oj}^{\beta}$, with the value of β found in Step 1. The parameter κ_o is global, being the same for all agents with a given scale. However, the number of free parameters per agent in this step remains the same because now we include q as a free parameter in the fitting process.

With this, we can infer from the data the average number of close contacts that have those individuals who are transmitting the virus.

The procedure adopted here is similar to that of Step 1, except that now, for each peak, the free parameters in the fitting procedure are s_{oj}, t_j and q_j, while κ_o and c are global parameters, being the same for all agents for the same bin width. This analysis allows us to observe how the value of q changes from one group to the other and as the epidemic process evolves.

Step 3: In this part we perform further analyses of the data, aiming to verify the robustness of the fractal model used in this work. One of the assumptions in our model, and its main difference with respect to the SIR model, is that transmission occurs only among individuals in groups of close contact, no matter the scale of the agent.

A consequence of this assumption is that the transmission of the virus may present a strong correlation with the geographical region where each infected agent lives. As the data we use have the location of residence of the infected individuals, we can investigate the correlation between the time of infection and localization.

A sequence of 2D plots indicates, by using different symbols, the groups of individuals already infected and the newly infected each day. To make a clear visualization, we just investigate the initial period of the infection in São Caetano do Sul. We calculate the least distance between newly infected and infected individuals in the city at different periods of time for the first wave of contagion. The histograms of the least distance evidence the spatial reach of the spreading dynamics.

Finally, we created a method to identify clusters of infection among the infection cases. The clustering method uses time and space correlations to identify the sequence of contagion in the different agents. The method consists in associating to the geographical localization the time elapsed between successive cases of infection, by a clustering method.

Consider that at day t we have a set of clusters of infections, C_t. To each cluster, we associate a vector position $\vec{r}_c$, given by the average position of the individuals in that cluster. For each newly infected individual on day t we attribute a probability to belong to a cluster in the set given by

$$p_i(c) = A_i d_{ci}^{-\alpha}\,(t - t_c)^{-\omega}, \tag{22}$$

where α and ω are parameters of the clustering method that are adjusted to reproduce the least-distance distributions and the time of contagion for the different agents. Here we use $\alpha = 3.5$ and $\omega = 2$. A_i is a normalization constant to ensure that the probabilities add up to one.

We also define a threshold value, L, for the acceptance in a cluster. If all values $p_i(c) < L$, then the individual forms a new cluster, which is included in the set C_{t+1} with all other clusters already in the set C_t. If there is one or more values $p_i(c) > L$, with $c \in C_t$, then the individual is included in the cluster with the highest probability. When a new individual is included in a cluster $c \in C_t$, the cluster position is recalculated as the average of the positions of all individuals in that cluster, and the time t_c is set to t.

Step 4: The comparison of the fractal model and the SIR and SI models is done by adopting Eq. (15), and fitting the data on new cases by fitting $-ds_j/dt$ to the series of new cases. This is an important difference with respect to the procedure used in Part 1 and Part 2, and is demanded by the necessity of using the SIR and SI methods of analysis. In this case, instead of fitting the q-exponential function to the data, its derivative is used in the fitting process. Since the derivative of a q-exponential is itself a q-exponential, this will just result in a slight different value for q and κ. In this part of the analysis, we investigate the role played by the free parameter, κ^*, its relation with the population size, and how all these parameters depend on the value for q. We use the chi-square as a measure of the quality of the fitting and to discuss which of the models best describes the data. Since SI and SIR are not agent-based models, we conduct the analysis using one single agent to describe the entire population.

3.3. Calculation of the distance between new cases and clusters

This analysis is performed by making 2-dimensional plots of the residence location of the infected individual over time. The residence location is obtained from the longitude, *Lo*, and latitude, *La* (in radians), of residence by defining the vector

$$\vec{r}(x, y) = R_T Lo\,\hat{i} + R_T La\cos(Lo)\,\hat{j}, \tag{23}$$

with $\hat{i}$ indicating the North-South direction and $\hat{j}$ the East-West and R_T is the Earth radius at the Equator line. The sum of all cases identified up to the day t are collected in a set I_t, where t is the day when the individual m reported the first symptoms. The distance to the previous cases is calculated by

$$d_{mn} = \sqrt{(\vec{r}_m - \vec{r}_n)^2}, \tag{24}$$

where n indicates each of the individuals in the set I_t. The least distance, d_m, is the minimum value in the set of distances for the new case, c, that is,

$$d_m = \min_{\substack{n\in I_t \\ m\neq n}}\{d_{mn}\}. \tag{25}$$

At the beginning of the analysis, it is possible to detect six cases that are very distant from each other, followed by new closer cases. The clusters were defined from the coordinates of the first six cases and a probability was assigned that each new case belonged to a cluster. The approximation performed is that the probability of a case belonging to a certain cluster depends only on the time since the first contamination and the distance to the cluster, as Eq. (26).

$$p \propto d_{ci}^{-\alpha}\,(t - t_c)^{-\omega}, \tag{26}$$

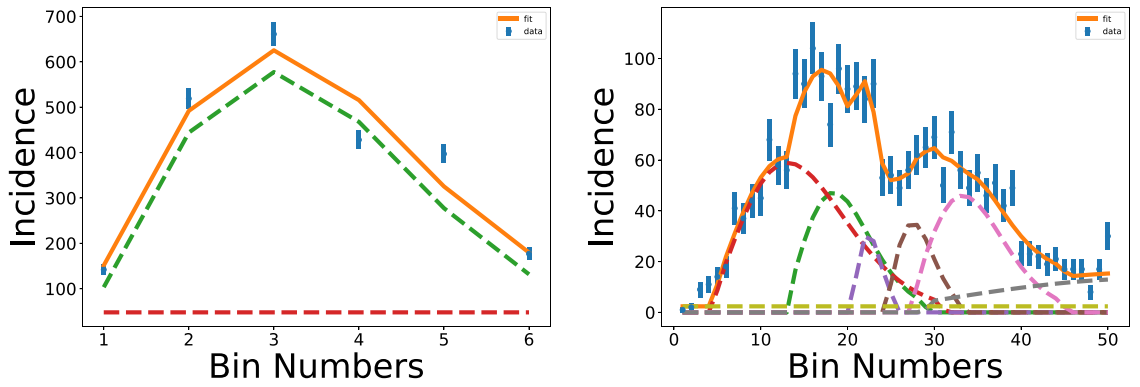

Fig. 3. The results of the fitting procedure for (left) 30-day bin and (right) 4-day bin analysis. The total number of infected (continuous line) and the contributions of each agent (dashed lines) are shown, as well as the background contribution indicated by the horizontal line. The parameters for each peak are shown in Table 2 ordered by the initial time of spreading in each agent.

were t_c is the time of the first case in cluster, α and ω are cluster parameters and d_{ci} is the distance between new case and cluster.

4. Results and discussion

We now fit the fractal model Eq. (15) to detailed data from the COVID-19 epidemic in the city of São Caetano do Sul, Brazil, in which residents with suspected COVID-19 symptoms were encouraged to contact a dedicated group that offered a home visit for a PCR test (see Section 3.1). We perform an analysis in four parts:

1. We estimate the number of agents and the parameters s_j, κ_j and Δt_j, assuming the fractal model with κ_j^* determined by the other parameters, following Eq. (13). Since previous analyses of COVID-19 showed that the number of infected individuals depends on the population as a power-law with exponent 1.15 [14], we use $q = 0.13$, which results in an exponent $1/(1-q) = 1.15$ in Eq. (20). This exponent of the power law is compatible with the observations of several quantities related to life in cities [46–48]. The goal of the analysis here is to determine the rates with which parameters κ and Δt change as the population grows, which allows us to evaluate the dynamics of the epidemic and estimate the value of q in Section 4.2 (see also the discussion at the end of Section 4.1). The analysis uses four different choices of time resolution, gathering the data in bins of widths 4, 7, 15 and 30 days. When the number of new cases is grouped into larger bins, the data is associated with larger groups of infection, and fewer details of the evolution of the epidemic can be observed in the data set. In the context of fractals, the use of different bin sizes is associated with a change in the scale of the system, so larger bins correspond to larger groups of infection. This results in models in which the number of susceptibles is approximately equal to the number of infected observed in the data (which is $N_{\text{infected}} = 2573$).
2. Having estimated the parameter κ_j for larger agents in Part 1, we now fit a different value of q_j for each agent in each scale. We obtain an estimate for social distancing and compare our results to an independent social isolation index obtained from cellular phone data.
3. In this part we use the geographical information about the approximate residence location of each individual in the data set to create clusters of infection in space and time.
4. In the last part of the analysis we compare the ability of the fractal model and the SIR model ($q \to 1$) to fit the data.

4.1. Results for Part 1

The use of different time scales to investigate the properties of the epidemic process is not new and was used to make predictions of the process at time t based on the time series up to $t-1$ [49]. Here the different scales (as mediated by the bins) are used to identify the groups throughout the outbreak that have similar properties of infection, such as the reproduction number. A lower time resolution (larger time bins) results in a smaller number of larger agents. We show in Fig. 3 the best-fit results, considering 30-day and 4-day bins, with the total number of infected individuals and the contribution of each agent and the background. Table 2 presents the best-fit parameters to all the bin sizes analyzed in the present work (we use s_{oj} to denote $s_j(t_{0j})$). We observe a good agreement between calculation and data for all cases, as evidenced by the χ^2 obtained.

The complexity shown by the data increases as the bin width is reduced, and for the 4-day bin analysis, the complex pattern of fluctuations represents a challenge for the fitting procedure, with several agents needed to obtain a good description. Nevertheless, the model successfully describes the data with six agents and a background contribution, which corresponds to the rate of people that are infected out of the city and seeds a new group of infections in the city. At the end of this part of the analysis, we obtained a set of 12 values for the agent size, transmission rate and infection period for each agent.

Table 1
Symbol table .

Symbol	Description
$s_j(t)$	Susceptible population of agent j in time t.
$i_j(t)$	Infectious population of agent j in time t.
$r_j(t)$	Recovered population of agent j in time t.
κ_j	Rate of contagion of agent j.
κ_j^*	Rate of removal for agent j.
q	Entropic index.
λ	Scale parameter determining the size of the agent.
t_0	Contamination start time of agent j.
Δt_j	Contamination period of agent j.

Table 2
Parameters of Eq. (20) adjusted for the data set separated in 4, 7, 15 and 30-day bins. The dates refer to the beginning of each group between the dates 2020-03-18 and 2020-09-19 as shown in Fig. 3. The background $\boldsymbol{c}$ is a constant adjusted to the whole period in each case.

Bin	s_{oj}	κ_j	Δt_j
	852(191)	6(1)E−04	103
	418(80)	1.6(4)E−03	60
4-day	100(16)	1.2(2)E−02	18
$c = 2 \pm 1$	183(40)	4.7(8)E−03	32
$s_{total} = 2615(233)$	459(36)	1.2(1)E−03	68
	603(92)	6(1)E−05	125
7-day	772(180)	2.5(6)E−04	149
$c = 2 \pm 1$	985(152)	3.7(8)E−04	130
$s_{total} = 2310(241)$	553(50)	4.5(5)E−04	109
15-day	671(44)	7.0(4)E−04	92
$c = 18 \pm 1$	1453(110)	2.1(2)E−04	174
$s_{total} = 2124(118)$			
30-day	2054(79)	1.8(1)E−04	196
$c = 48 \pm 11$			

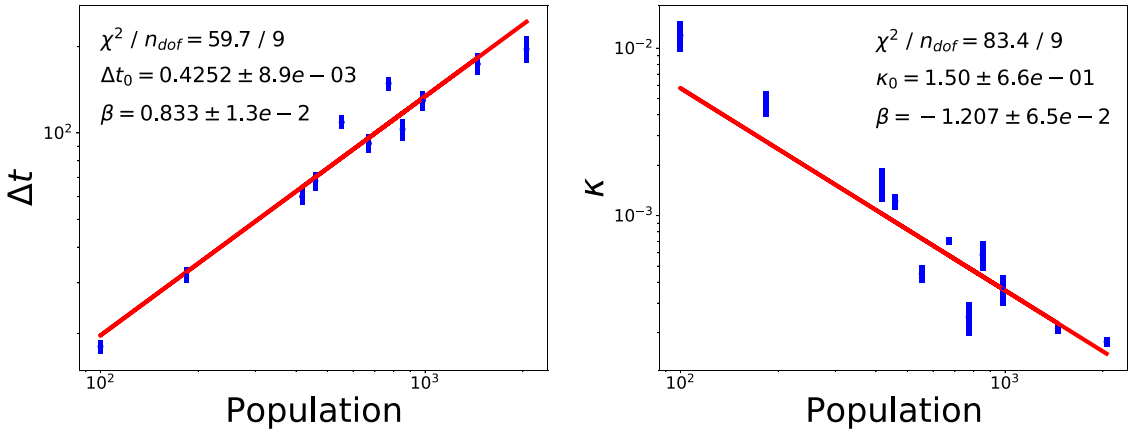

Fig. 4. Plots of the parameters Δt and κ as a function of the agent population, as obtained in the fitting procedure.

With the data for the infection period, Δt_j, κ_j and s_{oj} for each agent, we can analyse the behaviour of those parameters and verify if the model is consistent. In fact, as discussed in the Introduction, the variables Δt_j and κ_j must follow a power-law distribution due to the fractal dynamics, with $\Delta t(s_{oj}) = \Delta t_o s_{oj}^{\beta_t}$ and $\kappa(s_{oj}) = \kappa_o s_{oj}^{-\beta_\kappa}$. The plots in Fig. 4 show the data distributed along a straight line in the log-log plot, confirming that the results obtained are consistent with the assumption made in the model.

With the results obtained for κ and Δt, the parameter $\tau = \kappa \Delta t$ depends on the population size, although it was initially supposed to be constant. This dependence of τ on the population size indicates that the value chosen for q is not correct for describing the COVID-19 evolution in the city studied here. In the next section, we show that the value for q is different from the value adopted in this part of the analysis. To obtain this result, we will use the fact that κ must behave as a power law, and we adopt the exponent $\beta = -0.833$, which keeps the value of τ constant.

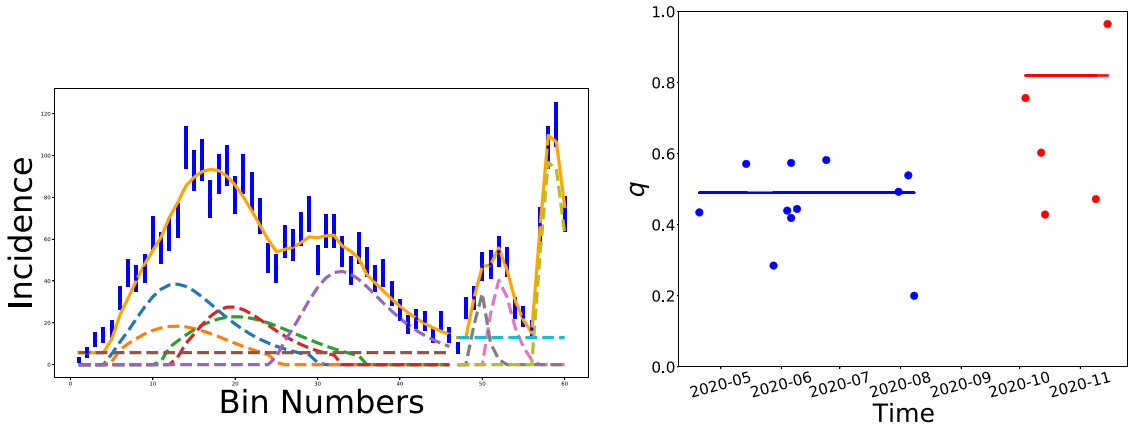

Fig. 5. The data grouped in bins of 4 days, with q free to adjust the data and using the power-law behaviour for κ in the first and second waves between 2020-03-18 and 2020-09-19 (top). The value for q as a function of time using the power-law $\kappa = \kappa_0 s^{-0.833}$ (bottom). In blue are the parameters found in the first wave and in red for the second wave. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

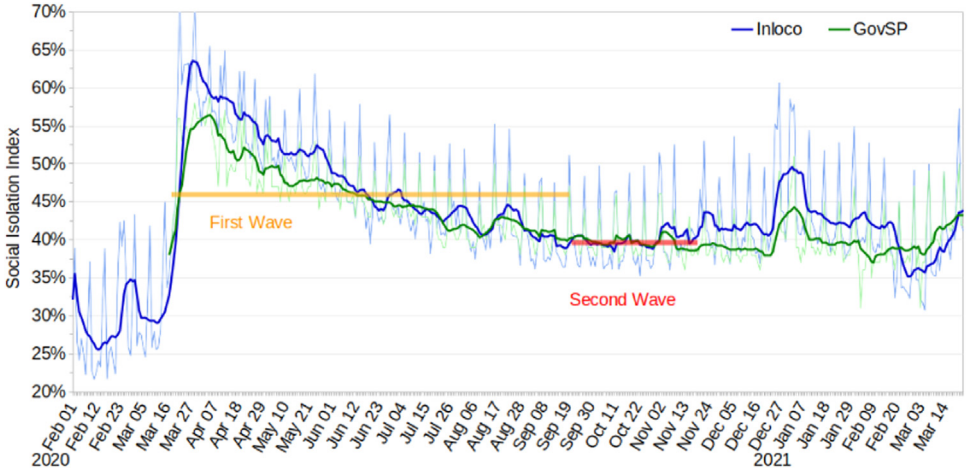

Fig. 6. Social Isolation Metrics for the city of São Caetano do Sul indicating the percentage of people that stayed home during the respective day. Two indexes are shown: in blue, the index from the Inloco company and in green the official index from the state government of São Paulo. The daily index is shown in light colours and the seven-day moving average is shown in dark colours, for each source. Horizontal bars indicate the average social isolation index, using the official government dataset, of the first and second waves. The figure shows how adherence to social distancing policies was drastically reduced in the second wave compared to the first wave. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Results for Part 2

In the second part of the present study, the power-law for $\kappa(s_{oj}) = \kappa_0 s_{oj}^{-\beta}$ is adopted with $\beta = -0.833$, while the parameter q is free to adjust the data. We also analysed the data with $\beta = -1.2$ and found that the main conclusions would be the same as obtained in the case adopted here. For instance, the fact that the social distancing observed, as described below, will decrease between the first and the second wave, stands with either value for β. Observe that we expect τ to be approximately independent of the agent considered since this parameter should capture the characteristics of the virus, while q is supposed to reflect the characteristics of social behaviour. As discussed above, virus characteristics and social behaviour cannot be completely separated, so variations on the value for τ may be expected. Therefore, we do not expect a well-defined behaviour of the parameter with the agent size, as would result from choosing $\beta = -1.2$ for κ_{oj}, which would lead to $\tau = \kappa_j \Delta t_j \sim s_o^{-0.4}$. We attribute this result to the value chosen for q, which might be slightly different from the value corresponding to the average number of contacts the population has, as we investigate below. Therefore, we adopt for κ a behaviour that corresponds to the inverse of that observed for Δt and fixes its exponent to the value mentioned. We choose this value because the interval Δt_j results from the good fitting to the data, and is determined with better precision that the values for κ_j. The results of the fitting procedure are shown in Fig. 5a, where we see the first wave, used in the previous analysis, and the peaks corresponding to the second wave of infection in the city. We notice that the second wave presents

Table 3
Best-fit parameters of Eq. (20) in the first wave obtained with bins of 4, 7, 15 and 30 days, and for the second wave between the dates 2020-09-19 and 2020-11-19 for bins of 4 and 7 days. The background $\boldsymbol{c}$ is a constant adjusted to the whole period in each case.

Global parameters	s_{oj}	q_j	Δt_j	τ_j
		1st wave		
4-day bin	554(1)	0.5712(4)	82(1)	0.0297(2)
$\kappa_0 = 0.06998(49)$	163(1)	0.435(2)	50(1)	0.0501(3)
$c = 4 \pm 1$	485(1)	0.57(1)	78(1)	0.0314(2)
$\chi^2/n_{\rm dof}$= 36/29	320(1)	0.4444(2)	69(1)	0.0395(3)
$s_{total} = 2084(2)$	652(1)	0.5388(3)	89(1)	0.0282(2)
7-day bin	569(35)	0.19(1)	76(3)	0.0398(9)
$\kappa_0 = 0.08723(43)$	1170(1)	0.2850(4)	90(1)	0.0281(3)
$c = 9 \pm 2$	402(27)	0.44(7)	72(1)	0.0401(1)
$\chi^2/n_{\rm dof}$= 35/15				
$s_{total} = 2141(44)$				
15-day bin	1701(355)	0.42(4)	134(1)	0.020(4)
$\kappa_0 = 0.07(1)$	342(150)	0.5(2)	67(1)	0.04(1)
$c = 23 \pm 1$				
$\chi^2/n_{\rm dof}$= 13/4				
$s_{total} = 2043(385)$				
30-day bin	2255(173)	0.6(2)	190(1)	0.012(6)
$\kappa_0 = 0.04(2)$				
$c = 33 \pm 29$				
$\chi^2/n_{\rm dof}$= 34/1				
		2nd wave		
4-day bin	114(11)	0.429(4)	15(1)	0.14(1)
$\kappa_0 = 0.48071(15)$	67(1)	0.756(8)	9(1)	0.1370(7)
$c = 13 \pm 1$	376(1)	0.472(1)	24(1)	0.0819(2)
$\chi^2/n_{\rm dof}$= 14/3				
$s_{total} = 557(11)$				
7-day bin	217(2)	0.602(1)	33(1)	0.0645(9)
$\kappa_0 = 0.172520(13)$	408(1)	0.9647(7)	34(1)	0.03922(4)
$c = 23 \pm 1$				
$\chi^2/n_{\rm dof}$= 6/1				
$s_{total} = 625(2)$				

peaks that are narrower than those in the first wave, indicating that the process, in this case, is somewhat different from the previous wave. The best-fit parameters are presented in Table 3.

In the first part of the analysis, we checked the consistency of the fractal hypothesis by verifying that the parameters κ and Δt, vary with the agent population according to power laws. After finding that the hypothesis of fractal dynamics is consistent with the data we now allow q_j to be a free variable to adjust the data. We observe that also in this case the model can describe the data in detail. The parameters q_j and s_{oj} are plotted in Fig. 5b, where we can observe the data of the first wave (blue online) and of the second wave (red online). The data for the first wave is concentrated in the range $0 < q_j < 0.6$ The second wave, however, has values that are above $q = 0.6$. The parameter q is related to the number of close contacts of the agents in each group, given by $N = (1 - q)^{-1}$. So these results show that our model is capable of evaluating social distancing based exclusively on the data of the number of infected individuals. For the first wave, we obtain $q \sim 0.5$ and $N \sim 2$ average number of close contacts of the individuals, while for the second wave, we obtain $q \sim 0.8$ and $N \sim 5$, indicating that the population in São Caetano relaxed social distancing during the second wave. A clear consequence is a fast increase observed in the number of infected in the second wave (Fig. 5).

We observe that χ^2 decreases as the number of agents increases showing, as expected, that the model provides a better description of the epidemic dynamics when there are assumed to be more agents. The fact that q increases from the first to the second wave shows a change in the social behaviour of the population. In each period, there are groups with small values for q, indicating small social contact, and others with higher values for q. In the second wave, as we observe in Fig. 5, some agents present a very high number of contacts and can be identified as super-spreaders.

We emphasize that the number of close contacts associated with the parameter q is not the total number of contacts made by a person; it refers only to contacts with individuals in the closest group. The average number of contacts an individual has varies according to the features of the network and of the observed process of information spreading and can be larger than the values obtained here [50]. The total number of contacts per individual is mediated by different agents at the several hierarchic levels of the fractal structure.

From the results obtained here, we observe that for $(1 - q)^{-1} \sim 10$ the curves given by the fractal model are similar to those obtained with the SIR model. This is in agreement with the historical fact that the SIR model has been successfully applied in epidemics studies for over a century. The model presented here performs as well as the SIR model, but better represents circumstances where infection is transmitted only through close contacts between individuals. This, might result

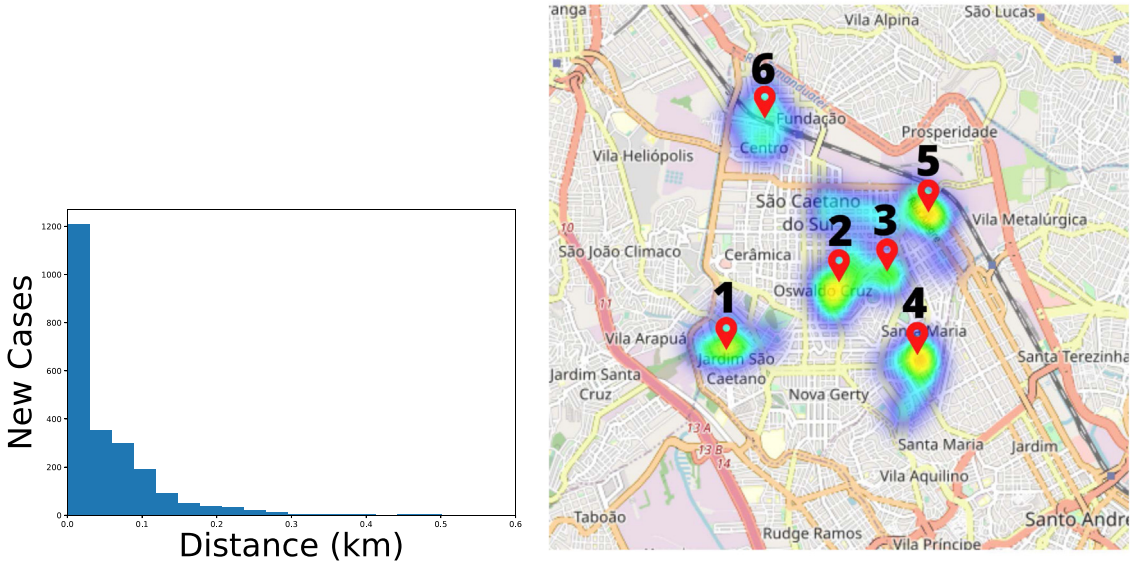

Fig. 7. Least-distance Distribution (left) between new cases and cases from previous days. Each new case is compared with all the cases from the previous days and the shortest distance between them is chosen. The analysis was carried out over the period between 2020-03-22 and 2020-09-30 corresponding to the entire first wave. Distribution of new cases (right) using the clustering method. Numbered markers represent infection groups where it is possible to associate locations of establishments with a concentration of people: Technical School - near Cluster 1, Park - near Cluster 4, Municipal Stadium - near Clusters 2 and 3, Hospital - near Cluster 5, Train station and Municipal University of São Caetano - near Cluster 6, GM Factory - near Clusters 5 and 6.

from awareness of infection spreading among individuals,leading to implement social-distancing [19], arising spontaneouslys, or by social distancing policies.

The number of contacts is a key parameter in many models of social behaviour. Dunbar's number [50,51], for instance, is considered the maximum number of contacts a person has in their social life. For the spread of information, the number of close contacts can be considerably different from Dunbar's number [52–55]. For infectious disease transmission, it has been investigated in several works, with values ranging from 3 to 8 [56,57]. The number of close contacts associated with the SARS CoV-2 transmission has been investigated in Ref. [58] for different age groups and results are in good agreement with the number of contacts obtained here. Similar variations of the number of contacts during the COVID-19 epidemic period were reported by a study of the UK transmission [59], which is between 2.4 and 4.5.

The value of τ varies very little for all agents in the first wave, with $\langle\tau\rangle = 0.03625(7)$. For the second wave, the value for τ is not as uniform, showing that the fractal model is not able to completely disentangle the properties of the virus infection contained in τ, from the social behaviour, contained in q. This difficulty of the model to separate the effects of social behaviour on the variable τ may be due to the fast change in the behaviour of the population in that period, which might be fast even compared to the infectious period, Δt, of the agents found for the 4-day bins. Another possibility is the appearance of new variants around the time of the second wave — for example, the P.2 variant seems to have appeared around mid-2020 [42], and started being reported in São Paulo state around Sept. 2020 [60].

One of the ways of verifying the intensity of the interactions between the first and second waves is by observing the degree of social isolation throughout the year and comparing it with the results obtained numerically.

In Fig. 6 we show how the Social Isolation Index, indicating the percentage of people that stayed home during the respective day, varied during 2020 in the city of São Caetano. Both datasets qualitatively agree with respect to the mobility pattern of the population. Social isolation during the first wave of the pandemic was significantly higher in the city compared to the beginning of the second wave.

4.3. Results of Part 3

The results of the third step of our analysis, where the least distance between newly infected individuals and the previous cases are computed, are presented in Fig. 7a. We observe a concentration of cases at distances around 250 m of the previous cases even at the beginning of the infection process. This correlation in the location of residence of the infected individuals is an indication of the existence of clusters of infection with geographic localization. It might be an effect of the social distancing observed at the initial phase of the first wave. The histogram for a larger period for the first wave, presented in Fig. 7, shows that the short distance contagion process remains during the spreading of the disease.

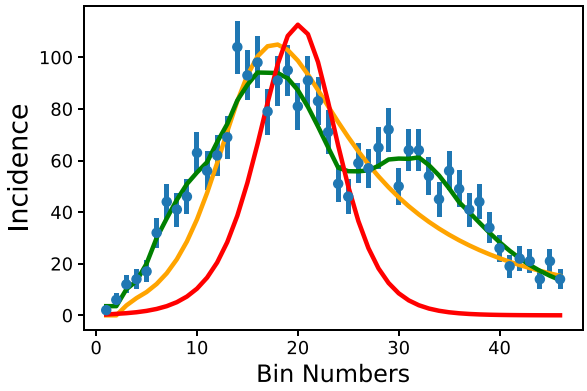

Fig. 8. Comparison between some variations of the model. In blue are the data separated in 4-day bins and the green curve represents the fit with Eq. (20). The orange curve represents the behaviour of Eq. (13) when $q \to 1$. The red curve is fit with differential Eq. (11) when $q = 1$ and $k^* = 0$ which corresponds to the traditional SI model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

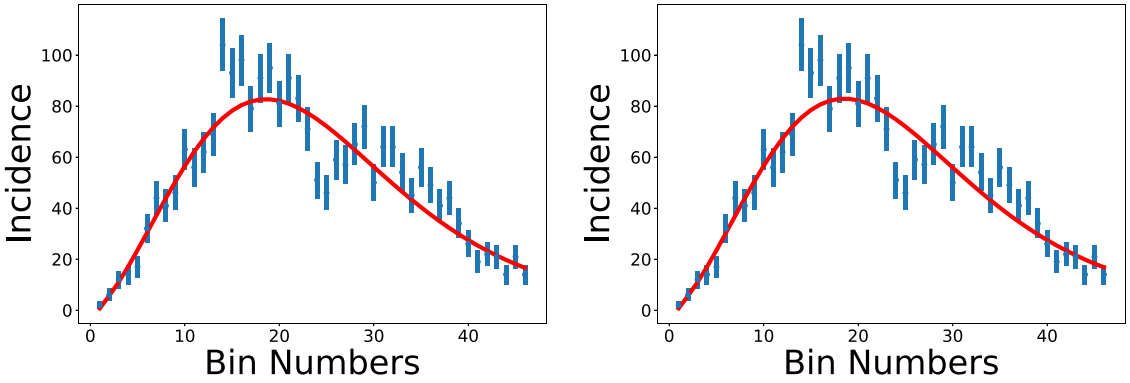

Fig. 9. Fit for the data set separated in 4 day bins with: (left panel) $k^* = 0$ and the other parameters left free; (right panel) all parameters left free.

With the location of residence of the infected individuals, we can visualize more easily the correlations between the position of new cases and previous cases. It is possible to observe a strong correlation between the new locations and the previous ones, reinforcing the hypothesis of close contact transmission.

Finally, with the clustering method described in the Materials and Methods section (Step 3, at the end of this document), we attribute each new case to a preexisting cluster, or to a new cluster of infection. In Fig. 7b we show the result of the clustering method for cases up to day 2020-04-20. The individuals in each cluster are presented with the same colour, each colour representing a different cluster. We clearly observe that the clusters are well-localized, in accordance with the least-distance findings. But in this case, the correlation is not only spatial but takes into account the time evolution of infection. The geographic localization of the clusters across the city is shown in Fig. 7b. At the beginning of the infection process, the clusters are localized, and their position in the city coincides with concentrations of people, such as universities, schools and industries. In this regard, we note that the social distancing measures adopted by the government of São Paulo started on March 23, corresponding to the initial date of our analysis and that the results reported in Fig. 7 are for the first 30 days after the date when schools and universities were closed, so the initial infections observed in the data correspond to those individuals infected before the non-pharmaceutical intervention (NPI) was put in place. Therefore, the figures show the images of the disease spreading induced by the cases of infection in the period before the closure of those institutions.

4.4. Results of Part 4

With the equations above we can calculate the epidemic process according to different scenarios. Imposing $q = 1$ results in the SIR model, and if additionally $\kappa^* = 0$ we get the SI model. For $q \neq 1$ we obtain the fractal model equivalent to SIR ($\kappa^* \neq 0$) or SI ($\kappa^* = 0$). Below we compare and discuss the results obtained with each model.

In the following, we analyse the differences and similarities of the fractal method in comparison with the SIR model or the SI model, following the method exposed in the previous section. Initially, we fix $\kappa^* = 0$ and fit the derivatives of the infectious population, di/dt, to the data corresponding to the first wave. The result is displayed in Fig. 9 and shows that a reasonable fitting is obtained with $q = 0.56 \pm 0.03$. The best-fit values for all parameters used in the different scenarios analyzed here are shown in Table 4. Observe that when all parameters are free to be adjusted, the result is compatible with

Table 4
Best-fit parameters of Eq. (15) for: fixed $\kappa^* = 0$ (first line); all parameters adjustable (second line); fixed population of **5.0E4**, **5.0E5** and **7.5E5** (lines 3 to 5); fixed $q = 1$ (last line).

s	κ	κ^*	q	χ^2/n_{dof}
2430(55)	3.0(4)E−4	0	0.56(2)	81/42
2446(77)	3.0(5)E−4	0.05± 0.14	0.57(3)	81/41
5.0E4	8.22(6)E−5	4.02(3)	0.605(7)	138/42
5.0E5	1.731(1)E−5	8.635(7)	0.999(2)	194/42
7.5E5	1.2026(7)E−5	9.01(6)	0.999(2)	318/42
1.56837(1)E5	2.936(6)E−5	4.574(9)	1	162/42

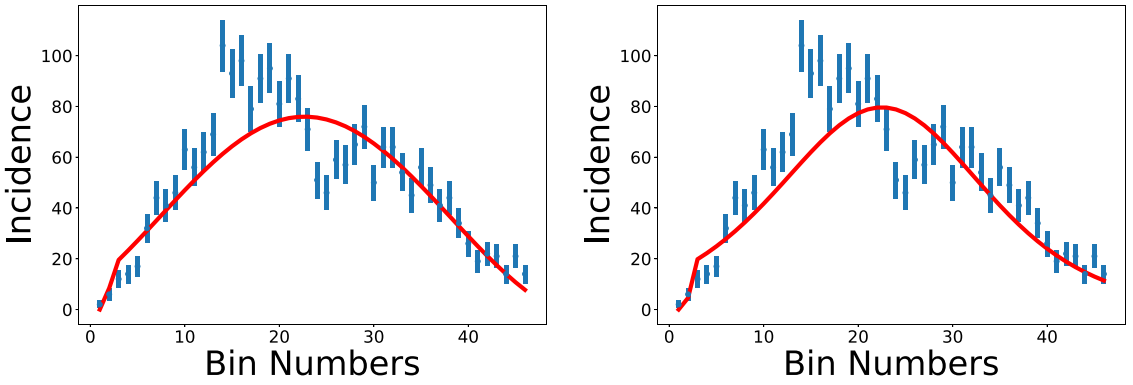

Fig. 10. Data fit considering the data set separated in 4-day bins. In the left panel the value population is fixed to 5.0E4. In the right panel we show the result for fixed $q = 1$ and free population.

$\kappa^* = 0$. The result of the best fit is shown in Fig. 9, and can be explained by the fact leaving all parameters free results in a total population that is similar to the total number of infected individuals. This happens because the set of data, alone, conveys no information about the number of individuals that are not infected.

The fitting of data can be improved, of course, by including the agent-based approach instead of using only one agent for the entire population. The fact that $q \neq 1$ is already an indication that the number of contacts is limited, and the hypothesis of the SIR model fails.

To obtain a value for κ^* different from zero we need to include additional information, as an estimate of the total susceptible population, including those that will not be infected during the epidemic episode. This is not easy to do in the context of an agent-based model, but can be done if one considers the entire city as a single group and the susceptible population as the entire population in the city. This hypothesis has some degree of arbitrariness since the city population fluctuates along the day and along the week (São Caetano is in the Metropolitan Region of São Paulo, and there is a large amount of movement between neighbouring cities), therefore the meaning of the parameter κ^* has to be interpreted with care.

In our analysis, we study the behaviour of the model for populations fixed at different values, up to those similar to the city population, estimated at $s_o = 161,957$ inhabitants [38]. The results are shown in Fig. 10. The best-fit parameters are presented in Table 4, with the results obtained showing values for κ^* that are not null, but that depends on the population size used. We observe that the parameter q also depends on the population. The fact that $q \to 1$ as we increase the population is due to the fact that we use one single agent even for large populations, and this result shows that the SIR is a special case of the fractal method, corresponding to the situation of a large agent. When we fix the parameter $q = 1$, the model still fits the data, but with a chi-square considerably higher than those obtained with other values of the same parameter. It is remarkable, however, that the population obtained for the best fit, in this case, is similar to the total population in the city. In Table 4 we observe that κ^* is dependent on the population size, therefore it can be determined only with the help of additional information, beyond that contained in the data on the new case time series: either fixing the population size or determining κ^* by other means, as the average time in which an infected individual can transmit the virus to other individuals. In the case of COVID-19, this time is around 5 days, so one expects $\kappa^* \sim 0.2\, d^{-1}$.

The expected value for κ^* is based on the average period in which one infected individual is infecting others. Including this extra information and fixing $\kappa^* = 0.2$ still leads to a reasonable description of the data, as observed in Fig. 11. The best-fit parameters are shown in Table 6. The result is obtained with only three agents, and when new agents are added, the fitting to the data worsens. In all cases, $q \neq 1$, showing the importance of taking into account the limited number of contacts. Using one single agent, with $q = 1$, which reproduces the SIR model, does not result in a good reproduction of the data, as shown in Fig. 11.

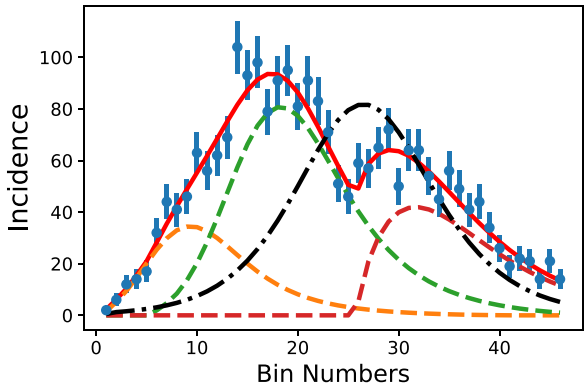

Fig. 11. Best fit with 4-day bins setting the value $k^* = 0.2$, 3 agents and other parameters free with background. The dashed lines represent the individual agent contributions, the continuous red line represents the total rate of infection, and the dot-dashed line represents the SIR result, with $q = 1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Best fit parameters of Eq. (15) obtained with bins of 4 days for orange curve and red curve of Fig. 8. The first line refers to Eq. (20) when ($\boldsymbol{q \to 1}$). The second line represents the behavior of the differential Eq. (15) when $\boldsymbol{q = 1}$ and $\boldsymbol{\kappa^* = 0}$. The parameter $\kappa = \kappa_o s^{-\beta}$, with $\beta = 0.833$ was used.

κ_0	s_o	q	Δt	τ	χ^2/n_{dof}
0.0209(5)	2319(52)	0.9999	267(1)	0.009(2)	196/43
8.1(2)E−5	1179(33)	1	128(1)	0.0103808(8)	1091/43

Table 6
First three rows: Best-fit parameters of Eq. (15) with $\boldsymbol{k^* = 0.2}$, 3 agents and other parameters free. Last row: single-agent, $\boldsymbol{q = 1}$.

s_j	κ_j	q_j	R_{0j}	χ^2/n_{dof}
461(90)	0.00136(40)	0.66(9)	3(1)	
1366(91)	0.000377(32)	0.74(3)	2.5(3)	36/33
658(62)	0.00324(79)	0.38(6)	10(3)	
3555(93)	0.000072(2)	1	1.29(5)	771/43

The best fit, with all parameters adjustable, leads to $\kappa^* \sim 0$. According to Eq. (15), if $\kappa^* = 0$ the derivative di/dt is proportional to i^q, which is itself a q-exponential. In fact, the derivative of

$$i_q = [1 + (1-q)\kappa s t]^{\frac{1}{1-q}} \tag{27}$$

is of the form

$$i_q = [1 + (1-q)\kappa s t]^{\frac{q}{1-q}}, \tag{28}$$

which can be written is a form similar to Eq. (27), if we substitute q by q' such that

$$\begin{cases} q' = \frac{1}{2-q} \\ \kappa' = \kappa \frac{1-q}{1-q'} \end{cases}. \tag{29}$$

Clearly, fitting the data with a function i_q or with or its derivative $i_{q'}^{q'}$ will lead to different values for the parameters q and q', but it is possible to find a relation between the values for q and κ found in Part 1, with q' and κ' found in Part 4, using $\kappa^* = 0$. The fact that the q-exponential function gives good fits to the data is an indication that the mechanism derived for the small group can, indeed, be applied to large groups. This is the fundamental hypothesis adopted in the fractal dynamic model, so the results obtained here corroborate the fractal assumption (Table 5).

5. Conclusion

In this study we used original data on the number of newly infected individuals in the city of São Caetano do Sul, in the metropolitan area of São Paulo. The data reports the number of new cases on a daily basis and contains the dates of initial symptoms and the geographical localization of an individual's residence.

The results obtained in the analysis show that the inclusion in the model of a mechanism to limit the number of contacts is important to accurately describe the observed data. The fractal model can be used to investigate the scale-free aspects

of the epidemic dynamics, with the possibility of including new agents (i.e., groups of close contacts) in the process, while keeping the total number of parameters small enough to obtain information on the epidemic process from the data analysis.

The data allowed us to check the fractal model of the epidemic dynamics by different approaches, and at the same time collect information on the epidemic process. We showed that the fractal dynamic model gives consistent results, such as the expected power-law distributions of the variables as a function of the population size, as obtained for the infection rate, κ and for the period of infection, Δt. When the power laws are incorporated into the model, we can investigate the effects of social distancing by analysing the parameter q, which is related to the average number of contacts. Our results show that in the first wave, the average number of contacts was approximately 2, while in the second wave, it increased to approximately 5. This was done for the first and second waves, and we observed that the number of contacts in the virus transmission process presents a significant increase in the comparison between the first and the second waves. The result obtained by using the model, and based solely on the time distribution of the newly infected individuals is corroborated by an independent analysis of the social distancing, based on the flux of the individuals inside the city. For each period, there is a variation of the number of contacts for each agent, with super-spreader agents being observed in the second wave, as can be observed in Fig. 5, with a number of contacts larger than 10, corresponding to values of $q > 0.9$ (see also Table 3).

The data on the geographic localization of the individual's residence allowed us to check the main conceptual aspect of the model, which is the transmission among groups of close contact. This assumption of the model is in agreement with the observed clustering of cases in the geographic localization. The analysis of the least distance among successive cases indicates that the average distance between the newly infected individuals and the previously infected is approximately 200 m. The clustering model allows us to identify the small groups of contagion in the epidemic process in the city, and verify that the location of the clusters coincides with places of an intense flux of people in the city. Thus, the model gives a fair description of the epidemic evolution evidencing the details of the spreading dynamics at large and small scales.

The comparison of the fractal dynamics model with the SIR and SI models was made using the fact that, as $q \to 1$, the fractal differential equations reduce to those of the SIR model (with $\kappa^* \neq 0$) and SI model (with $\kappa^* = 0$). The best description of the data, obtained in Part 4 of our analysis when all the parameters of Eq. (15) are free to adjust the data, results in $q \sim 0.5$ and $\kappa^* = 0$, so that the estimated number of close contacts is ~ 2. This result is an indication that the distribution of new cases is best described by a q-exponential function, and thus the best solution found in Parts 1 or 2 and in Part 4 are equivalent.

We conclude that the new cases are usually generated at small distances from previous cases, and that the spreading dynamics is based on highly localized clusters of infection, with the infection based on small groups. The use of complex networks to simulate the dynamical evolution of an epidemic process allows a more detailed description of the process, and confirms that those tools are useful to analyse socioeconomic aspects of modern life, as already observed in other works [61–64], and in particular, scale-free networks, or fractal networks, have been identified in several domains [47,48,65–67]. Thus, the characteristics of the model show that it can be an auxiliary method for deciding about lockdown or social distancing in specific regions, contributing to optimising the size of the population affected by those decisions and the time necessary for the sanitary measure to be effective.

The close connections between the epidemic dynamics and complex networks open several new perspectives. The resilience of networks has been investigated in applications to other fields [68], and its application to the epidemic process can help determine the best measures to reduce the rate of infection in different areas. The agent-based approach can indicate the best way to reduce or interrupt the flux of people between different areas with different stages of infection, and the length of time during which those measures must be held to optimize their results.

The model is a useful tool to evaluate and guide the public policies used to attenuate the pandemic process. It allows the identification of the infection process at small scales and offers a way to represent infection at large scales and vice-versa. As a hypothetical example, consider the problem of deciding to lockdown an entire region to mitigate the epidemic process. One of the main parameters to be determined is the duration of the lockdown, and the model described here can be used to determine the optimal duration of a lockdown in a given region. Let us suppose that, in that region, the average household is formed by four individuals. What is the optimal duration of the lockdown? According to our findings, the infection period goes as $\Delta t(s_o) = \Delta t_o s_o^{0.83}$, where s_o is the average household size. According to the power-law behaviour obtained in our analysis, the period of infection in an agent of size s_o is given by $\Delta t = s_o^{\beta}$, with $\beta = 0.833$. If t_o is the time during which one individual can infect others, for a group with $s_o = 4$ we have $\Delta t \sim 3\Delta t_o$. If the period in which one individual can infect others is, say, 10 days, then the optimal duration of lockdown is approximately 30 days.

In conclusion, we showed that the fractal dynamic model for the epidemic process of COVID-19 gives consistent results on the spread of infection. The use of the model as a tool to investigate epidemic evolution shows that it can reproduce with accuracy the data at different scales of population size. An additional, scale-free feature of the model, the independence of the average number of contacts and population size, allows us to evaluate the effects of social distancing, based only on the time series of newly infected individuals. As we have shown, the possibility of separating the transmission probability, associated with characteristics of the virus, from the average number of contacts made by infectious individuals is a benefit of the fractal dynamic model, which is shared by other models of social behaviour based on agents. In the present case, it is expressed by the independence of the parameters q and τ from the population size. Note that q varies according to the number of contacts in each group, but the distribution of values of q does not depend on the sizes λ of the agents but does depend on time, as shown in Fig. 5.

Finally, we note that the model presented here differs from the standard SIR model in those cases where the transmission of the virus depends on a small number of close contacts. When this number is around or above 10, the results obtained with the fractal model approach those of the SIR model. This does not diminish the potential of the present approach, since not only the number of contacts per individual may change during the evolution of an epidemic due to social distancing, but also there are more aspects, aside from the number of contacts, that can be explored by using the fractal model. The model may be used in the future to establish better social-distance methods, which could depend on the geographical evolution of the clusters of infection and could minimize the social burden associated with social distancing.

Data availability

Data will be made available on request.

Acknowledgments

A. D. is partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil), grant 304244/2018-0, by Project INCT-FNA Proc. No. 464 898/2014-5 and by FAPESP grant 2016/17612-7. PSP is supported by FAPESP grant 16/18445-7. V H N is supported by CNPq, grant 304714/2018-6. This work was supported by a Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0) (https://caddecentre.org). This work was supported by a Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0) (https://caddecentre.org); Wellcome Trust and Royal Society (N.R.F.: Sir Henry Dale Fellowship: 204311/Z/16/Z).

References

[1] N. Vabret, G.J. Britton, C. Gruber, S. Hegde, J. Kim, M. Kuksin, R. Levantovsky, L. Malle, A. Moreira, M.D. Park, L. Pia, E. Risson, M. Saffern, B. Salome, M.E. Selvan, M.P. Spindler, J. Tan, V. van der Heide, J.K. Gregory, K. Alexandropoulos, N. Bhardwaj, B.D. Brown, B. Greenbaum, Z.H. Gumus, D. Homann, A. Horowitz, A.O. Kamphorst, M.A.C. de Lafaille, S. Mehandru, M. Merad, R.M. Samstein, S.I.R. Project, Immunology of COVID-19: current state of the science, Immunity 52 (6) (2020) 910–941, doi:10.1016/j.immuni.2020.05.002.
[2] J.W. Goodell, COVID-19 and finance: agendas for future research, Finance Res. Lett. 35 (2020), doi:10.1016/j.frl.2020.101512.
[3] A. Sharifi, A.R. Khavarian-Garmsir, The COVID-19 pandemic: impacts on cities and major lessons for urban planning, design, and management, Sci. Total Environ. 749 (2020) 142391, doi:10.1016/j.scitotenv.2020.142391.
[4] C.H.d.B. Cruz, Social distancing in São Paulo State: demonstrating the reduction in cases using time series analysis of deaths due to COVID-19, Rev. Bras. Epidemiol. 23 (2020) e200056, doi:10.1590/1980-549720200056.
[5] S. Flaxman, S. Mishra, A. Gandy, H.J.T. Unwin, T.A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J.W. Eaton, M. Monod, A.C. Ghani, C.A. Donnelly, S. Riley, M.A.C. Vollmer, N.M. Ferguson, L.C. Okell, S. Bhatt, Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, Nature 584 (7820) (2020) 257–261, doi:10.1038/s41586-020-2405-7.
[6] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, Lancet Infect. Dis. 20 (5) (2020) 533–534, doi:10.1016/S1473-3099(20)30120-1.
[7] O. Diekmann, H. Heesterbeek, T. Britton, Mathematical tools for understanding infectious diseases dynamics, Princeton Series in Theoretical and Computational Biology, Princeton University Press, 2013.
[8] H.W. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (4) (2000) 599–653, doi:10.1137/S0036144500371907.
[9] S. He, Y. Peng, K. Sun, SEIR modeling of the COVID-19 and its dynamics, Nonlinear Dyn. 101 (3, SI) (2020) 1667–1680, doi:10.1007/s11071-020-05743-y.
[10] S. Bhatt, N. Ferguson, S. Flaxman, A. Gandy, S. Mishra, J.A. Scott, Semi-mechanistic Bayesian modeling of COVID-19 with renewal processes, 2020, arXiv:2012.00394
[11] D. Champredon, J. Dushoff, D.J.D. Earn, Equivalence of the erlang-distributed SEIR epidemic model and the renewal equation, SIAM J. Appl. Math. 78 (6) (2018) 3258–3278, doi:10.1137/18M1186411.
[12] N.T.J. Bailey, The Mathematical Theory of Epidemics, Charles Griffin & Co., Ltd., London, 1957.
[13] C. Viboud, L. Simonsen, G. Chowell, A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks, Epidemics 15 (2016) 27–37, doi:10.1016/j.epidem.2016.01.002.
[14] M. Abbasi, A. Bollini, J. Castillo, A. Deppman, J. Guidio, P. Matuoka, A. Meirelles, J. Policarpo, A. Ramos, S. Simionatto, A. Varona, E. Andrade-II, H. Panjeh, L. Trevisan, Fractal signatures of the COVID-19 spread, Chaos, Solitons Fractals 140 (2020) 110119, doi:10.1016/j.chaos.2020.110119.
[15] H. Jahanshahi, J.M. Munoz-Pacheco, S. Bekiros, N.D. Alotaibi, A fractional-order SIRD model with time-dependent memory indexes for encompassing the multi-fractional characteristics of the COVID-19, Chaos, Solitons Fractals 143 (2021), doi:10.1016/j.chaos.2020.110632.
[16] G. Vasconcelos, A. Macêdo, G. Duarte-Filho, et al., Power law behaviour in the saturation regime of fatality curves of the COVID-19 pandemic, Sci. Rep. 11 (2021) 4619, doi:10.1038/s41598-021-84165-1.
[17] S. Padhy, V.P. Dimri, Apparent scaling of virus surface roughness—An example from the pandemic SARS-nCoV, Physica D 414 (2020) 132704, doi:10.1016/j.physd.2020.132704.
[18] S. Thurner, P. Klimek, R. Hanel, A network-based explanation of why most COVID-19 infection curves are linear, Proc. Natl. Acad. Sci. 117 (37) (2020) 22684–22689, doi:10.1073/pnas.2010398117.
[19] K.A. Kabir, K. Kuga, J. Tanimoto, Analysis of sir epidemic model with information spreading of awareness, Chaos, Solitons Fractals 119 (2019) 118–125, doi:10.1016/j.chaos.2018.12.017.
[20] K. Kuga, J. Tanimoto, Effects of void nodes on epidemic spreads in networks, Sci. Rep. 12 (1) (2022) 3957, doi:10.1038/s41598-022-07985-9.
[21] J. Tanimoto, Sociophysics Approach to Epidemics, vol. 23, Springer, 2021.
[22] M.E. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[23] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512, doi:10.1126/science.286.5439.509.
[24] R. Albert, A.L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (2002) 47–97, doi:10.1103/RevModPhys.74.47.
[25] C. Song, S. Havlin, H.A. Makse, Self-similarity of complex networks, Nature 433 (2005) 392, doi:10.1038/nature03248.
[26] R. Kopelman, Fractal reaction kinetics, Science 241 (4873) (1988) 1620–1626, doi:10.1126/science.241.4873.1620.
[27] A. Deppman, E.O.A. Segundo, Flux of information in scale-free networks, 2021, arXiv:2106.08959
[28] M.J. Keeling, K.T.D. Eames, Networks and epidemic models, J. R. Soc. Interface 2 (2005) 295–307, doi:10.1098/rsif.2005.0051.
[29] R. Pastor-Satorras, C. Castellano, P.V. Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. 87 (2015) 925–979, doi:10.1103/RevModPhys.87.925.
[30] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200–3203, doi:10.1103/PhysRevLett.86.3200.

[31] K. Kosmidis, P. Macheras, A fractal kinetics SI model can explain the dynamics of COVID-19 epidemics, PLoS One 15 (2020), doi:10.1371/journal.pone.0237304.
[32] H. Abbey, An examination of the reed-frost theory of epidemics, Hum. Biol. 24 (3) (1952) 201.
[33] C. Tsallis, Possible generalization of Boltzmann–Gibbs statistics, J. Stat. Phys. 52 (1988) 479–487.
[34] C. Tsallis, U. Tirnakli, Predicting COVID-19 peaks around the world, Front. Phys. 8 (2020) 217.
[35] U. Tirnakli, C. Tsallis, Epidemiological model with anomalous kinetics: early stages of the COVID-19 pandemic, Front. Phys. 8 (2020) 613168.
[36] A. Deppman, Thermodynamics with fractal structure, Tsallis statistics, and hadrons, Phys. Rev. D 93 (5) (2016), doi:10.1103/PhysRevD.93.054001.
[37] A. Deppman, T. Frederico, E. Megias, D.P. Menezes, Fractal structure and non-extensive statistics, Entropy 20 (9) (2018), doi:10.3390/e20090633.
[38] IBGE | Cidades@ | São Paulo | São Caetano do Sul | Panorama, 2021, https://cidades.ibge.gov.br/brasil/sp/sao-caetano-do-sul/panorama. Accessed Jul. 18, 2021.
[39] F.E. Leal, M.C. Mendes-Correa, L.F. Buss, S.F. Costa, J.C. Bizario, S.R. De Souza, O. Thomaz, T.R. Tozetto-Mendoza, L.S. Villas-Boas, L.C. De Oliveira-Da Silva, et al., Clinical features and natural history of the first 2073 suspected COVID-19 cases in the corona São Caetano primary care programme: a prospective cohort study, BMJ Open 11 (1) (2021) e042745.
[40] P.S. Peixoto, D. Marcondes, C. Peixoto, S.M. Oliva, Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in brazil, PLoS One 15 (7) (2020) e0235732, doi:10.1371/journal.pone.0235732.
[41] D.S. Candido, I.M. Claro, J.G. de Jesus, W.M. Souza, F.R. Moreira, S. Dellicour, T.A. Mellan, L.D. Plessis, R.H. Pereira, F.C. Sales, et al., Evolution and epidemic spread of SARS-CoV-2 in Brazil, Science 369 (6508) (2020) 1255–1260.
[42] N.R. Faria, T.A. Mellan, C. Whittaker, I.M. Claro, D.da S. Candido, S. Mishra, M.A. Crispim, F.C. Sales, I. Hawryluk, J.T. McCrone, et al., Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil, Science 372 (6544) (2021) 815–821.
[43] L.F. Buss, C.A. Prete Jr., C.M. Abrahim, A. Mendrone Jr., T. Salomon, C. de Almeida-Neto, R.F. França, M.C. Belotti, M.P. Carvalho, A.G. Costa, et al., Three-quarters attack rate of SARS-CoV-2 in the Brazilian amazon during a largely unmitigated epidemic, Science 371 (6526) (2021) 288–292.
[44] F.L. Dias, M. Assumpção, P.S. Peixoto, M.B. Bianchi, B. Collaço, J. Calhau, Using seismic noise levels to monitor social isolation: an example from Rio de Janeiro, Brazil, Geophys. Res. Lett. 47 (16) (2020). e2020GL088748
[45] E.C. Sabino, L.F. Buss, M.P. Carvalho, C.A. Prete, M.A. Crispim, N.A. Fraiji, R.H. Pereira, K.V. Parag, P.d.S. Peixoto, M.U. Kraemer, et al., Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence, Lancet 397 (10273) (2021) 452–455.
[46] L.M.A. Bettencourt, The origins of scaling in cities, Science 340 (6139) (2013) 1438–1441, doi:10.1126/science.1235823.
[47] L.M.A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G.B. West, Growth, innovation, scaling, and the pace of life in cities, Proc. Natl. Acad. Sci. 104 (17) (2007) 7301–7306, doi:10.1073/pnas.0610172104.
[48] M. Batty, A theory of city size, Science 340 (6139) (2013) 1418–1419, doi:10.1126/science.1239870.
[49] K. Parag, C. Donnelly, Using information theory to optimise epidemic models for real-time prediction and estimation, PLoS Comput. Biol. 16 (2020) e1007990, doi:10.1371/journal.pcbi.1007990.
[50] B. Gonçalves, N. Perra, A. Vespignani, Modeling users' activity on twitter networks: validation of Dunbar's number, PLoS One 6 (2011).
[51] Z. Adrian, M. Tymon, A new extraordinary means of appeal in the Polish criminal procedure: the basic principles of a fair trial and a complaint against a cassatory judgment, AJEE 6 (2023) 1–18.
[52] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, S. Vigna, Four degrees of separation, in: Proceedings of the 3rd Annual ACM Web Science Conference, 2012, pp. 33–42.
[53] Z. Wang, Z. Yang, H. Shan, J. Shi, A measurement analysis of information propagation in online social network, in: 2013 Third International Conference on Instrumentation, Measurement, Computer, Communication and Control, 2013, pp. 23–27, doi:10.1109/IMCCC.2013.13.
[54] J. Leskovec, K.J. Lang, A. Dasgupta, M.W. Mahoney, Statistical properties of community structure in large social and information networks, in: Proceedings of the 17th International Conference on World Wide Web, ACM, New York, NY, USA, 2008.
[55] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? in: Proceedings of the 19th International Conference on World Wide Web, ACM, New York, NY, USA, 2010.
[56] K. Leung, M. Jit, E.H.Y. Lau, J.T. Wu, Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong, Sci. Rep. 7 (2017) 7974.
[57] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G.S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, W.J. Edmunds, Social contacts and mixing patterns relevant to the spread of infectious diseases, PLoS Med. 5 (3) (2008) e74.
[58] C.G. McAloon, P. Wall, F. Butler, M. Codd, E. Gormley, C. Walsh, J. Duggan, T.B. Murphy, P. Nolan, B. Smyth, K. O'Brien, C. Teljeur, M.J. Green, L. O'Grady, K. Culhane, C. Buckley, C. Carroll, S. Doyle, J. Martin, S.J. More, Numbers of close contacts of individuals infected with SARS-CoV-2 and their association with government intervention strategies, BMC Public Health 21 (1) (2021) 2238, doi:10.1186/s12889-021-12318-y.
[59] A. Gimma, J.D. Munday, K.L.M. Wong, P. Coletti, K. van Zandvoort, K. Prem, CMMID COVID-19 working group, in: P. Klepac, G.J. Rubin, S. Funk, W.J. Edmunds, C.I. Jarvis (Eds.), Changes in Social Contacts in England During the COVID-19 Pandemic Between March 2020 and March 2021 as Measured by the CoMix Survey: A Repeated Cross-Sectional Study, PLoS Med., vol. 19, 2022, p. e1003907.
[60] Presença das linhagens por estado, 2021, http://www.genomahcov.fiocruz.br/presenca-das-linhagens-por-estado/. Accessed Jul. 26, 2021.
[61] R. Albert, A.L. Barabási, Topology of evolving networks: local events and universality, Phys. Rev. Lett. 85 (2000) 5234–5237, doi:10.1103/PhysRevLett.85.5234.
[62] H. Gong, C. Guo, Y. Liu, Measuring network rationality and simulating information diffusion based on network structure, Physica A 564 (2021) 125501, doi:10.1016/j.physa.2020.125501.
[63] M.J. Hamilton, B.T. Milne, R.S. Walker, O. Burger, J.H. Brown, The complex structure of hunter-gatherer social networks, Proc. R. Soc. B. 274 (2007) 2195–2203, doi:10.1098/rspb.2007.0564.
[64] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. 99 (12) (2002) 7821–7826, doi:10.1073/pnas.122653799.
[65] H. Youn, L.M.A. Bettencourt, J. Lobo, D. Strumsky, H. Samaniego, G.B. West, Scaling and universality in urban economic diversification, J. R. Soc. Interface 13 (114) (2016) 20150937, doi:10.1098/rsif.2015.0937.
[66] D. Brockmann, F. Theis, Money circulation, trackable items, and the emergence of universal human mobility patterns, IEEE Pervasive Comput. 7 (4) (2008) 28–35, doi:10.1109/MPRV.2008.77.
[67] G.B. West, J.H. Brown, B.J. Enquist, The fourth dimension of life: fractal geometry and allometric scaling of organisms, Science 284 (5420) (1999) 1677–1679, doi:10.1126/science.284.5420.1677.
[68] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Resilience of the internet to random breakdowns, Phys. Rev. Lett. 85 (2000) 4626–4628, doi:10.1103/PhysRevLett.85.4626.