



**Cite this article:** Ribeiro FL, Meirelles J, Ferreira FF, Neto CR. 2017 A model of urban scaling laws based on distance-dependent interactions. *R. Soc. open sci.* **4**: 160926. <http://dx.doi.org/10.1098/rsos.160926>

Received: 24 November 2016

Accepted: 21 February 2017

**Subject Category:**

Physics

**Subject Areas:**

computer modelling and simulation/mathematical modelling/complexity

**Keywords:**

agglomeration effects, complex systems, power laws, scaling laws, urban indicators, public policies

**Author for correspondence:**

Fabiano L. Ribeiro

e-mail: [fribeiro@dfi.ufla.br](mailto:fribeiro@dfi.ufla.br)


# A model of urban scaling laws based on distance-dependent interactions

Fabiano L. Ribeiro<sup>1</sup>, Joao Meirelles<sup>2</sup>, Fernando F. Ferreira<sup>3</sup> and Camilo Rodrigues Neto<sup>3</sup>

<sup>1</sup>Departamento de Física (DFI), Universidade Federal de Lavras (UFLA), Caixa Postal 3037, 37200-000 Lavras, Minas Gerais, Brazil

<sup>2</sup>Laboratory on Human-Environment Relations in Urban Systems—HERUS, École polytechnique fédérale de Lausanne (EPFL) Station 2, 1015 Lausanne, Switzerland

<sup>3</sup>EACH—Universidade de São Paulo (USP), Av. Arlindo Bettio, 1000 (Vila Guaraciaba), 03828-000 São Paulo, SP, Brazil

 CRN, 0000-0001-6783-6695

Socio-economic related properties of a city grow faster than a linear relationship with the population, in a log–log plot, the so-called *superlinear scaling*. Conversely, the larger a city, the more efficient it is in the use of its infrastructure, leading to a *sublinear scaling* on these variables. In this work, we addressed a simple explanation for those scaling laws in cities based on the interaction range between the citizens and on the fractal properties of the cities. To this purpose, we introduced a measure of social potential which captured the influence of social interaction on the economic performance and the benefits of amenities in the case of infrastructure offered by the city. We assumed that the population density depends on the fractal dimension and on the distance-dependent interactions between individuals. The model suggests that when the city interacts as a whole, and not just as a set of isolated parts, there is improvement of the socio-economic indicators. Moreover, the bigger the interaction range between citizens and amenities, the bigger the improvement of the socio-economic indicators and the lower the infrastructure costs of the city. We addressed how public policies could take advantage of these properties to improve cities development, minimizing negative effects. Furthermore, the model predicts that the sum of the scaling exponents of social-economic and infrastructure variables are 2, as observed in the literature. Simulations with an agent-based model are confronted with the theoretical approach and they are compatible with the empirical evidences.

## 1. Introduction

When humanity built the first cities, they brought together individuals previously separated by space, increasing social and

economic interactions, through the shared infrastructure of the cities, in a more constant and efficient way. Since then, the city as a social organization of humanity, has become an important place to create interactions in space and time between individuals [1].

Following Warren Weaver's ideas on organized complexity [2], Jane Jacobs first proposed the idea of the city as an integrator of individuals in her seminal work, *The Death and Life of the Great American City* [3]. In her work, Jacobs saw cities as complex entities, organized in a bottom-up manner, by the behaviour of its inhabitants. In the book's concluding chapter, Jacobs defines the city as a problem of organized complexity. The ideas of complexity were young then, and Jacobs quickly realized the importance of understanding cities from this conceptual framework. Thereafter, this complexity approach to cities have yielded some important insights into urban science [4–7].

In the same vein, the hypothesis raised by Bettencourt & Lobo [8] proposes a new explanation based fundamentally on network effects: the mere fact that individuals are spatially close to each other increases the number of potential encounters between them. Theoretically, the probability of encounters and interaction grows as the individuals get closer and this process lowers the transaction and communication costs.

Recent research made it possible to empirically support these propositions [1,9,10], thanks to the increasing availability of urban data, from official statistics to private databases. Taking advantage of this large volume of data, covering thousands of cities in the world, researchers have apparently found statistical regularities on how concentrations of people affect economic activities, infrastructure and social vitality [11].

The main findings from this new field of study, integrating urban planning, geography and physics of complex systems, can be summarized by the assumption that all cities produce spatial scale economies as they grow and at the same time, achieve gains in their socio-economic productivity. In general, when different cities within the same urban system (i.e. in the same country) are compared, the largest cities are denser than the smallest towns, and therefore the total amount of infrastructure per inhabitant is smaller. At the same time, as a general rule, the largest city is richer, costlier and more culturally and technologically productive when the metrics *per capita* are analysed.

If the socio-economic variables of a city grow faster than the population, we have an effect called *superlinear scaling* [10], characterized by a scaling exponent greater than one. The data indicate that the urban infrastructure follows a similar law: the larger a city, the more efficient it is in the use of its infrastructure, leading to significant economies of space. For example, when the population of a city doubles its material infrastructure grows less than that, from the number of gas stations to the total length of water pipes, streets and electric cables [9]. If the infrastructure grows less than the population, we have something called *sublinear scaling*, characterized by a scaling exponent smaller than one. Last, there is a third class of variables which presents *linear scaling* with the population size. According to empirical evidence, those variables are related to individual needs, as consumption of water and electrical energy, number of employment, and so on [9].

The most remarkable properties of the urban scaling hypothesis are (i) the concentration of people in space and time and (ii) greater intensive use of urban infrastructure. Together, (i) and (ii) promote matching and social coordination, increasing social indicators such as wealth, innovation and crime, enabling better use of infrastructure such as the street and transportation networks, electrical and communication cables and many others as a city grows [10]. Some evidence, however, indicates a considerable sensitivity of the scaling exponent in relation to the adopted definition of city, challenging the universality hypothesis of these scaling laws [12–15]. More empirical evidence is clearly needed in this discussion but theoretical understanding, as presented in this work, can contribute to this dispute.

In order to build a quantitative theory of cities, we must take into account the city's geometry to calculate aggregate amounts that generate social and infrastructure indicators. According to the statistical mechanics approach, to explain the scaling law observed in many complex systems we can ignore several microscopic details and focus on the important features. More specifically, we can focus on proposing a spatial interaction model involving the individuals and how they make choices about their destinations or how they are influenced by the social network.

In the work herein, we look for a simple explanation for the scaling laws in cities based on the fractal properties of the cities, as well as the behaviour of individuals. For that, we introduce a measure of social potential that captures the influence of social interaction on the economic performance and the benefits of amenities in the case of infrastructure offered by the city. We assume that the populational density depends on the fractal properties and that the individual interaction intensity decays with distance. As a result we obtain a power law scaling for social indicators and infrastructure. The scaling exponents we

found are coherent with the empirical data. An agent-based model is proposed to compute experiments and to test the hypotheses.

The paper is organized in five sections, including this introduction. Section 2 presents the model based on the hypothesis just exposed above. Section 3 presents the model simulation results and §4 discusses our results. We conclude in §5 with suggestions of further research.

## 2. The model

### 2.1. Scaling laws in the cities: socio-economic production

Individuals with limited information can take advantage from another person's opinions, behaviours and abilities in order to fulfil their needs and solve their problems. Through a positive feedback mechanism, those interactions allow individuals to explore the information and resources available in the city to enhance creativity and socio-economic production. Each individual in the city is stimulated by the others. In this context, we use the word *stimulus* to mean all the social characteristics that influence the individual, such as cultural attitude, dialect and demand of products, just to cite a few.

The distance between individuals in a city is an important factor. For instance, the chance of two individuals meeting each other through common friends depends on the distance they live one from each other. Moreover, the impact or influence which one person has on another also may depend on the distance between them. Based on this, let us consider the stimulus strength of the individual  $j$  over the individual  $i$  be represented by the function  $f(r_{ij})$ , where  $r_{ij}$  is the distance between them. So, based on recent empirical evidence [16,17], it is quite plausible to assume that the stimulus between two citizens decays with the distance, according to:

$$f(r) = \begin{cases} \frac{1}{r^\gamma} & \text{if } r > 2r_0 \\ \frac{1}{(2r_0)^\gamma} & \text{otherwise,} \end{cases} \quad (2.1)$$

where  $\gamma$  is the decay exponent. This function was first used to model the interaction of living cells in a competitive [18] and cooperative [19–21] environment, and recently it was used to model the interaction between tumour cells [22]. The distance  $2r_0$  is the minimal distance between the agents (the distance of a house, for instance). For convenience purposes, we assume that  $r_0 = \frac{1}{2}$ . We find empirical support for the hypothesis expressed by equation (2.1), with experimental values of  $\gamma$  ranging in the interval:  $1 \leq \gamma \leq 1.5$  [17]. For example, the distribution of physical distances of Facebook contacts are  $\gamma = 1.03$  [16] and  $\gamma = 1.12$  [23]; the distribution of e-mail distances:  $\gamma = 1.0$  [16] and  $\gamma = 1.20$  [23]; and the frequency of cell phone call with the distance is found in the interval:  $1 \leq \gamma \leq 1.5$  [17].

The influence between individuals separated by longer distances may be small, but if many people share the same opinions, their collective strength may be meaningful and affect them. The extent of the social tissue, measured by the size of the social network, contributed to the amount of social stimulus or influence to which one individual is exposed. For simplicity, we proposed an additive model to measure the total effective influence or stimulus. The individual stimulus of a single individual  $i$  from his/her interaction with all the other ( $N$ ) citizens of the city (influences), can be written as

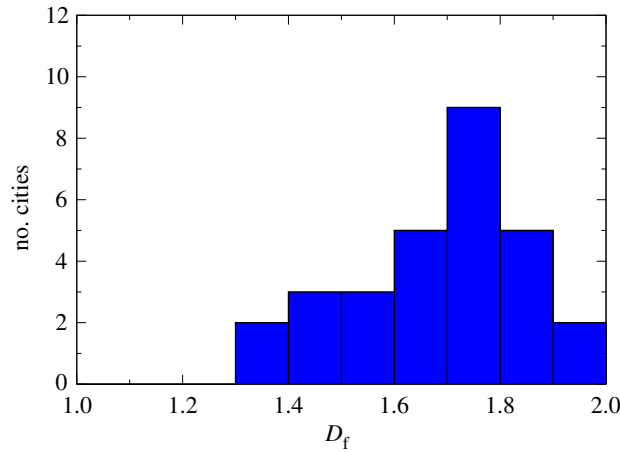
$$I_i = \sum_{j=1}^N f(r_{ij}), \quad (2.2)$$

while the socio-economic stimulus of the city is given by

$$I_{se} = \sum_{i=1}^N I_i. \quad (2.3)$$

We are not modelling the mechanism behind how an individual may influence others. We are just considering that the influence strength decays over the social network through distance in a pairwise interaction framework at the micro level. Moreover, we are hypothesizing that the aggregate information emerging from individual interaction is the mechanism for the power law observed at the macro level, as presented in the following.

Considering that the individuals are continually distributed along the city area with density  $\rho(\mathbf{r})$ , where  $\mathbf{r}$  is a position vector, then the sum in the equation (2.2) can be approximated by an integral. That



**Figure 1.** Histogram of the fractal dimension of many cities around the world. Data collected from [24, p. 242]. Most of the cities studied have their fractal dimension in the interval  $D_f = [1.6, 1.8]$ , with average  $\bar{D}_f = 1.7$ .

is, the total stimulus felt by  $i$  can be computed by

$$I_i = \int_{\text{all space}} \rho(\mathbf{r}) f(r) d^D \mathbf{r}, \quad (2.4)$$

where  $D$  is the Euclidean dimension of the space in which the population is embedded. In the case of a city, which lies in a plane, we can assume that  $D = 2$ . However, following the approach adopted in many studies, the spatial distribution of a city obeys fractal behaviours [24,25]. Figure 1 presents a histogram built from a table of the fractal dimension  $D_f$  of many cities around the world presented in [24, p. 242]. The *box-counting method* was used to compute the fractal dimension of the cities [26,27]. According to the data, most of the cities have their fractal dimension in the interval  $D_f = [1.6, 1.8]$ , with average  $\bar{D}_f = 1.7$ , as shown in figure 1. There are other references that suggest this is the average value of the fractal dimension of the cities [28].

Cities usually present fractal behaviour, thus we assume that  $r$  is the linear length, and then the number of individuals must scale as  $r^{D_f}$ , while the area that embedded the population scales as  $r^D$ . Thus, we write the population density equation as

$$\rho(\mathbf{r}) = \frac{\text{number of individuals}}{\text{Area}} = \rho_0 \frac{r^{D_f}}{r^D} = \rho_0 r^{D_f - D}, \quad (2.5)$$

where  $\rho_0$  is constant (for details, see [27]). This hypothesis suggests that the density is a radially symmetric function. With this hypothesis, we solve equation (2.4) using

$$d^D \mathbf{r} = r^{D-1} dr d\Omega_D \quad (2.6)$$

(transforming from Cartesian coordinates to hyperspherical coordinates). Using the periodic boundary condition, we have

$$I_i(N) = \frac{\omega_D}{D_f} \frac{1}{1 - \gamma/D_f} \left[ \left( \frac{\omega_D}{D_f} N \right)^{1 - \gamma/D_f} - 1 \right] + \frac{\omega_D}{D_f}, \quad (2.7)$$

where  $\omega_D$  depends only on the Euclidean dimension (for details, see [19–21]). The right side of the above equation does not depend on the index  $i$ , which means that, given the hypothesis presented, all the individuals of the population felt the same stimulus intensity, which depends on the population size of the city. It is due to the self-similarity of the fractal structure formed by the population [27]. In this way, the individual stimulus scales with  $N$  by

$$I_i(N) = c_1 N^{1 - \gamma/D_f} + c_2, \quad (2.8)$$

where  $c_1$  and  $c_2$  are constants. The ratio  $\gamma/D_f$  plays an important role in this context. In fact, this ratio is responsible for the range of interaction between each pair of individuals in the city. If  $(\gamma/D_f) > 1$ , the individual stimulus given by (2.4) does not depend on the population size (given a population sufficiently large). Let us call this situation the *short-range interaction regime* since the individual interacts only with its closer neighbours. In this case, the stimulus felt by a single individual converges to a

**Table 1.** Empirical results of the scaling exponent for two particular indicators of the cities in three different countries: one correlated to an infrastructure indicator (gas station) and other related to a socio-economic one (GDP). The data came from [8,29], but similar results were presented in other studies [9,10,29]. The number of gas stations scales sublinearly ( $\beta \approx \frac{5}{6} < 1$ ), while the GDP scales superlinearly ( $\beta \approx \frac{7}{6} > 1$ ) with the population size of the cities. Different countries present similar scale exponent values, and the evidence suggests a correlation between the exponents, given that the sum, no matter the country, is around 2.

country	$\beta_{\text{infra}}$ (gas stations)	$\beta_{\text{se}}$ (GDP)	$\beta_{\text{infra}} + \beta_{\text{se}}$
France	0.90 [0.8, 1.0]	1.20 [1.15, 1.26]	2.1 [1.95, 2.2]
Spain	0.75 [0.65, 0.85]	1.13 [0.97, 1.30]	1.88 [1.62, 2.15]
Germany	0.80 [0.75, 0.85]	1.17 [1.06, 1.28]	1.97 [1.81, 2.13]

constant when the population size is large enough. On the other hand, if  $(\gamma/D_f) < 1$ , the individual stimulus given by (2.4) diverges when  $N \rightarrow \infty$  as a consequence of the fact that the individuals interact effectively with everybody, despite being less intense with people who are further away. We will call this case the *long-range interaction regime*.

From equation (2.4), the total socio-economic stimulus produced in the city, that is  $I_{\text{se}} = NI_i$ , is given by

$$I_{\text{se}} = c_1 N^{\beta_{\text{se}}} + c_2 N \quad (2.9)$$

where

$$\beta_{\text{se}} \equiv 2 - \frac{\gamma}{D_f}, \quad (2.10)$$

is the scaling exponent associated with the socio-economic activities. So the total stimulus in the city is a power law of the population size.

In the short-range interaction regime,  $\gamma/D_f > 1$ , the total stimulus of the city scales as  $I_{\text{se}} \sim N$  (linear behaviour), which is not compatible with the empirical data presented in table 1. It means that if the individuals of the city interact only with their neighbours and not with the distant ones, the intellectual and socio-economic activities of the city will not scale superlinearly, but only linearly with the population size. Conversely, in the long-range interaction regime, that is  $\gamma/D_f < 1$ , the total stimulus scales as

$$I_{\text{se}} \sim N^{\beta_{\text{se}}}, \quad (2.11)$$

with  $\beta_{\text{se}} > 1$ . This result agrees qualitatively with the empirical data presented in table 1.

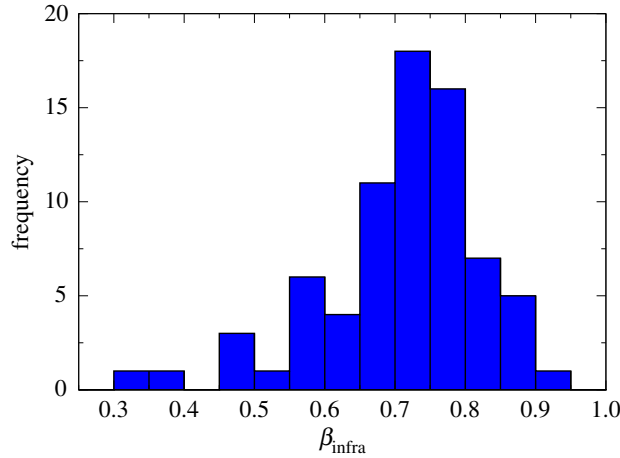
This means that, according to the theory presented here, if a city presents superlinear behaviour of their socio-economic indicators, then the individuals (or the regions) of this city must be able to interact with all other individuals (or other regions) of this city. In other words, the city will only enhance their socio-economic indicators if the city behaves as a whole. On the contrary, that is, if the city is just a collection of isolated regions, then the socio-economic indicators will be only linearly dependent on the population size. As we presented, the last case is not supported by empirical data.

Medellin, Colombia, is a example of such a situation. This city had a strong increase in development indicators when it began to integrate its less developed areas, previously isolated in the hills, with the rest of the city. This integration process, made by the implementation of the aerial lift (or metrocable), generated new revenue for the inhabitants of these places and promoted the development of the city as a whole [30].

## 2.2. Scaling laws in the cities: infrastructure

Let us now focus on the infrastructure of the cities. In the present work, the focus is directed to the number of amenities a city has to offer. They are related to the infrastructure sector, and they usually present sublinear behaviour, as shown by figure 2. We built this histogram using 74 kinds of amenities (e.g. bakery ( $\beta_{\text{infra}} = 0.847$ ), beauty salon ( $= 0.745$ ), gas station ( $= 0.652$ ) and so on) across 47 US cities. The data were collected directly from the references [31,32], but other studies found similar results [9,29]. Although the model we present below relates to the number of amenities in a city, the framework must be valid to other infrastructure sectors of the city.

Suppose  $U$  to be the total consumption of an individual human need quantity, and then  $U \sim N$  [9]. This consumption is supplied by the amenities of the infrastructure sector in consideration. For instance,  $U$  can be the total consumption of bread in the city, and the amenities that supply this product are the bakeries. The average *per capita* demand of this product is  $\langle u \rangle = U/N$ .



**Figure 2.** Histogram showing the number of amenities with a particular value of scaling exponent ( $\beta_{\text{infra}}$ ). This histogram was built with 74 amenities (for instance, bakery ( $\beta_{\text{infra}} = 0.847$ ), beauty salon ( $= 0.745$ ), gas station ( $= 0.652$ ) and so on), all of them presenting scaling law with the population size, across 47 US cities. It is possible to note an evident sublinear scaling for all the amenities presented. The data used were collected directly from the references [31,32].

Considering  $u_i$  as the consumption of the product (following the last example, the consumption of bread) by a single citizen  $i$ , we have  $U = \sum_{i=1}^N u_i$ . The citizen can acquire this product in  $P$  amenities distributed around the city, and the choice of what amenity through which he/she will get the product depends on many factors, such as the quality of services, conservation, overcrowding, price, distance, etc. However, we will consider an ideal situation that all the amenities are completely similar, and the only relevant variable for the citizen is the distance. In this way, a rational citizen will choose the amenity that minimizes the transport cost. In other words, we can use the function  $f(r_{ik})$ , given by equation (2.1), to represent the total supply of the amenity  $k$  for citizen  $i$ , and they are separated by the distance  $r_{ik}$ . Then the total demand—the provision—of citizen  $i$  can be computed by

$$u_i = \sum_{k=1}^P f(r_{ik}). \quad (2.12)$$

Here, we assume that the citizen is more influenced, or more supplied in this context, by closest amenities.

In order to compute the number of amenities of a given infrastructure sector, let us to compute first the total provision (of a particular product) of the city. It can be reached by the sum

$$U \equiv \sum_{i=1}^N u_i = \sum_{i=1}^N \sum_{k=1}^P f(r_{ik}), \quad (2.13)$$

which can also be written as

$$U = \sum_{k=1}^P \left( \sum_{i=1}^N f(r_{ik}) \right). \quad (2.14)$$

The sum inside the parenthesis can be computed in a similar way that was done in the previous section, which conducts us to  $\sum_{i=1}^N f(r_{ik}) \sim N^{1-\gamma/D_f}$ , and then (from equation (2.14))

$$U \sim PN^{1-\gamma/D_f}. \quad (2.15)$$

As  $U \sim N$ , the above equation yields to

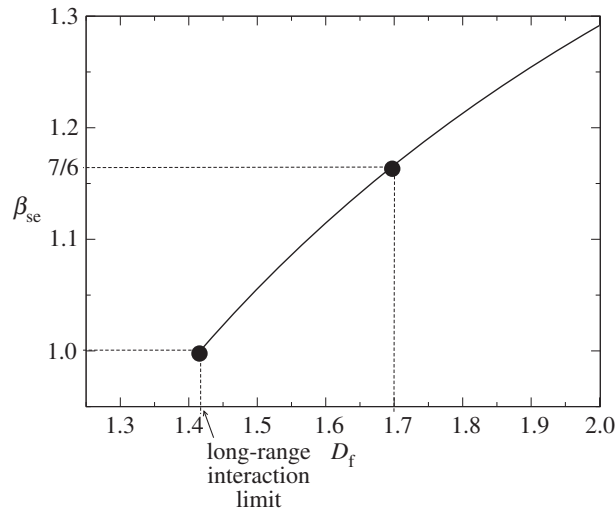
$$P \sim N^{\beta_{\text{infra}}}, \quad (2.16)$$

where

$$\beta_{\text{infra}} \equiv \frac{\gamma}{D_f} \quad (2.17)$$

is the scaling exponent associated with the infrastructure.





**Figure 3.** The socio-economic scale exponent as a function of the fractal dimension of the city, according to the equation (2.10), for  $\gamma = 1.41666 \dots$ . The socio-economic scale exponent is a monotonically increasing function of the fractal dimension of the city. When one considers, by approximation, that the city is a compacted structure, that is  $D_f = D = 2$ , it implies an overestimation of the scale exponent. The extension of the solid line ( $D_f > \gamma = 1.41666 \dots$ ) represents the long-range interaction regime. For  $D_f < \gamma$ , that is the short-range interaction regime, the socio-economic variables scale linearly with the population size.

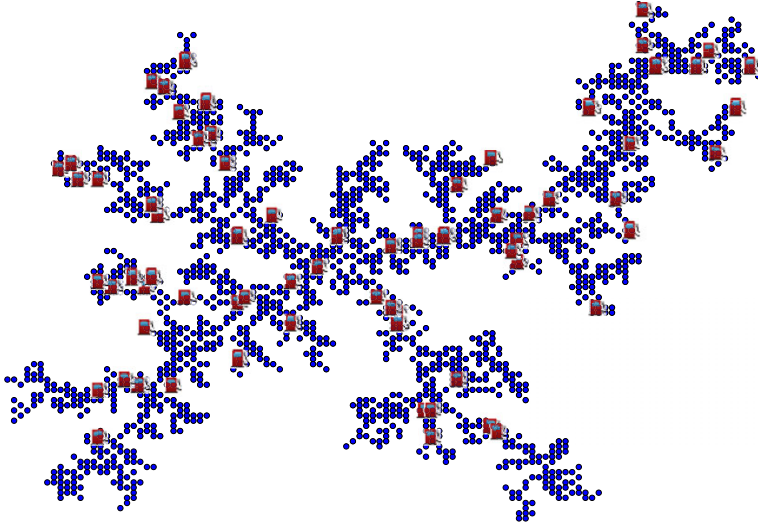
If we are in the *long-range interaction regime*, that is  $(\gamma/D_f) < 1$ , the number of amenities behaves sublinearly in relation to  $N$ . Since the empirical data suggest a sublinear behaviour, the theory predicts that the individuals search for gas stations or other infrastructure variables in an interaction of long range. That is, there will be scale economies in infrastructure variables (sublinear behaviour) only if the city interacts as a whole, as a single organism. If the citizens use only the amenities that are in their blocks, it means that the citizens do not use effectively the city, but only a part of it. In this case, we observe a linear relation between the number of amenities (infrastructure) and the population size, which is not consistent with the empirical evidence.

In conclusion, and connecting with what we described in the previous section, when the city interacts as a whole, and not just as a set of isolated parts, then there is improvement of the socio-economic indicators (superlinear behaviour). Consequently, the infrastructure variables show scale economies (sublinear behaviour). Moreover, the bigger the interaction range between citizens and amenities (i.e. the smaller  $\gamma$  is), the bigger the improvement of the socio-economic indicators and the lower the infrastructure costs of the city are.

### 2.3. Relation between the scale exponent and the fractality of the city

In the long-range interaction regime the scale exponent  $\beta_{se}$  is a monotonically increasing function of the fractal dimension of the city, as suggested by equation (2.10) and shown in figure 3. It means that more compact cities ( $D_f \approx 2$ ) tend to have bigger scale exponents than fractal cities ( $D_f < 2$ ). That is because, according to our theoretical results, the more compact the city is, the bigger the cultural and social stimuli are. This assertion is in accordance with Bettencourt’s discussion [10]. In his urban mobility-based model, assuming that the population completely fills the area, the scale exponent predicted by his theory is bigger than the one observed experimentally. According to Bettencourt, this is related to the fact that the population does not occupy all the space, but only a fragment of that. The model presented here and consequently the result (2.10), based on social stimulus, also follows the same interpretation.

Given that the three parameters of the result (2.10) are accessible experimentally, the theory can be tested. For instance, if  $\beta_{se} = \frac{7}{6}$  and  $D_f = 1.7$ , then  $\gamma = 1.41666 \dots$ , which coincides with the interval of numeric values that the experimental results suggest (between 1 and 1.5). Moreover, this numeric value strengthens the evidence that cities present a long-range interaction regime because  $(\gamma/D_f) = (1.41666 \dots / 1.7) < 1$ .



**Figure 4.** Spatial distribution at equilibrium of a given simulated city. The citizens,  $N = 2000$  fixed agents represented by the small circles, were distributed according to the DLA algorithm, forming a structure with fractal dimension  $D_f \approx 1.7$ . The amenities, represented by a fuel dispenser, were generated by the rules described in the §3.

### 3. Model simulations

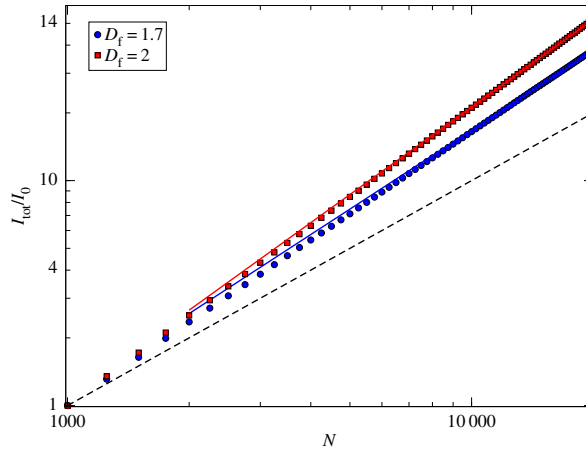
In this section, we present computational simulations in order to test the proposed theory. The algorithm used to simulate a virtual city is described as follows. Consider  $N$  fixed individuals spatially distributed, whose stimulus between them and the total stimulus of the city are computed by equations (2.2) and (2.3), respectively. We will consider two types of spatial configurations in the simulations: compacted and fractal landscape. In the first one, the individuals are fixed and randomly distributed in a square. In this case, the population (compacted) presents dimension  $D_f = D = 2$ . The other type of spatial distribution is fractal and generated by a *diffusion-limited aggregation* (DLA) algorithm [27,33]. One particular configuration generated by this algorithm is the one presented in figure 4 (small circles). In this case, the spatial distribution of the population has dimension  $D_f \approx 1.7$  [27]. In fact, this artificial structure has been used to model growth of cities due to the numerical similarity between its fractal dimension and the dimension of real cities [24,34]. We will use these two kinds of spatial distribution (compacted and fractal) to see, computationally, how the socio-economic and infrastructure scale exponents behave according to the dimension of the spatial structure formed by the population.

The list below summarizes the algorithm used to simulate a city and to compute the socio-economic variables:

1. First, it generates a city with  $N$  individuals: using DLA algorithm ( $D_f = 1.7$ ) or distributing the individuals randomly in a square ( $D_f = 2$ ).
2. It computes the distance  $r_{ij}$  between any two individuals.
3. It computes the stimulus of all individuals of the population ( $\{I_i\}_{i=1\dots N}$ ), in which  $I_i = \sum_{j=1}^N f(r_{ij})$ , and  $f(r_{ij})$  is given by equation (2.1). The parameter  $\gamma$  is kept fixed throughout the process.
4. Lastly, it computes the total stimulus of the city by  $I_{se} = \sum_{i=1}^N I_i$ .

Figure 5 presents the average of the total stimulus of the two kinds of spatial structure as a function of the population size when  $\gamma < D_f$ . It is evident, by the simulation results that the socio-economic exponent follows a superlinear behaviour ( $\beta_{se} > 1$ ). That is, the total stimulus of the city is boosted with increases in population size. In relation to the quantitative analysis of the social exponent, there is also a good agreement between the simulation and the analytic prediction given by (2.11). The comparison between the theoretical predictions of the socio-economic exponent and the simulation results is better for a larger population size, due to the minimization of the finite size effects. As was expected, a compact structure presents a greater socio-economic exponent than the fractal one. In other words, the more compact the population is, the more efficient the city will be in the socio-economic aspects.





**Figure 5.** Superlinear behaviour between the total stimulus of the city and the population size, presented by the simulation of the model. The parameters of simulations were:  $\gamma = 1.41666 \dots$ ,  $D_f = 2$  (homogeneous distribution) and  $D_f = 1.7$  (DLA algorithm). Points (circles and squares) represent averages over 50 independent samples of numerical simulations, where each simulation is performed keeping  $N$  fixed. The error bars are smaller than the size of the points. The continuous lines are theoretical predictions (equation (2.10)) where  $\beta_{se} = \frac{7}{6}$  (blue line) and  $\beta_{se} = 1.29166$  (red line). Dashed line represents the linear scaling.

To simulate the dynamics of the amenities, which represent an infrastructure sector, we used the following algorithm. Given a city generated by the previous algorithm, we have  $P = P_0$  amenities (initial condition) that are randomly distributed in the space of the virtual city. Then the provision function of all the individuals of the population is computed via equation (2.12). The average individual provision can be computed by  $\bar{u} = (1/N) \sum_{i=1}^N u_i$ . If  $\bar{u} < c$  (supply shortages), where  $c$  is a constant, then one new amenity is introduced in a random position in the city. If  $\bar{u} \geq c$ , then the less efficient amenity is deleted (bankruptcy) from the city. As the time evolves, the number of amenities converges to an equilibrium (optimal) value. Figure 4 shows a particular city generated using DLA algorithm and the distribution (at equilibrium) of the amenities generated by the algorithm described above.

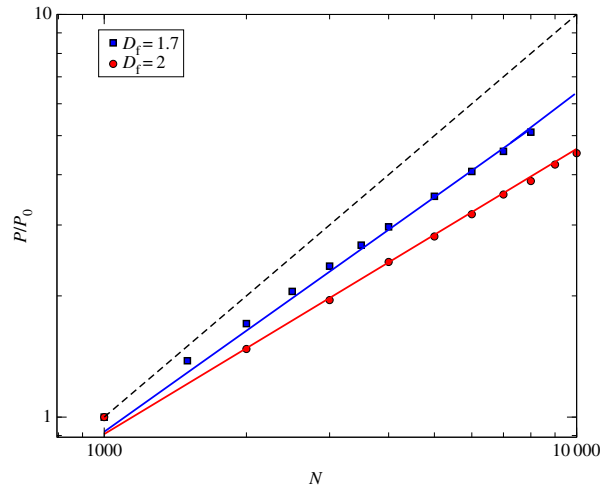
The algorithm used to simulate the amenities and the variables associated with infrastructure is summarized below:

1. It generates a city with  $N$  individuals by the same procedure described above.
2. It generates  $P_0$  amenities, each of them is allocated randomly in some point of the city.
3.  $P := P_0$ .
4. It computes the distance from any amenity to all the other citizens to obtain  $\{r_{ik}\}_{i=1, \dots, N, k=1, \dots, P}$ .
5. It calculates the supply function of all individuals, that is the set  $\{u_i\}_{i=1 \dots N}$ , and the average individual-used provisions, by  $u_i = \sum_{k=1}^P f(r_{ik})$  and  $\bar{u} = (1/N) \sum_{i=1}^N u_i$ , respectively.
6. If  $\bar{u} > c$ , where  $c$  is a constant, then the amenity that produces less provision to the individuals is deleted. Then do  $P := P - 1$ . Otherwise ( $\bar{u} \leq c$ ), a new amenity is generated and allocated randomly in some point of the city. Then  $P := P + 1$ .
7. Back to the item 4.
8. The simulation stops when the number of amenities converges.

Figure 6 shows us that the equilibrium quantity of amenities scales sublinearly with the population size (given  $\gamma < D_f$ , i.e. long-range interaction regime). That means we have scale economies, and therefore, greater cities need less amenities *per capita*. These scale economies (sublinear behaviour), according to the model, are a direct consequence of the long-range interaction regime. The opposite situation, that is the short-range interaction regime, must conduct to a linear behaviour, which does not correspond to the empirical evidence.

## 4. Discussions

In our model, the ratio  $\gamma/D_f$  plays an important role. Being  $\gamma/D_f > 1$  implies a short-range interaction between the population, that will tend to form isolated groups, and then the city would not behave as a unit. In this situation, as shown by equation (2.9) in this limit, the socio-economic



**Figure 6.** Sublinear behaviour between the number of amenities at equilibrium (normalized by the division by  $P_0 \equiv P(N = 1000)$ ) and the population size, presented by the simulation of the model. The parameters of simulations are:  $\gamma = 1.41666 \dots$ ,  $D_f = 2$  (homogeneous distribution) and  $D_f = 1.7$  (DLA algorithm). Points (circles and squares) represent averages over 30 independent samples of numerical simulations, where each simulation is performed keeping  $N$  fixed. The error bars are smaller than the size of the points. Continuous lines are theoretical predictions (equation (2.16)) where  $\beta_{\text{infra}} = \frac{5}{6}$  (blue line) and  $\beta_{\text{infra}} = 0.708$  (red line). Dashed line represents the linear scaling.

variables will scale linearly with the population size, which is not supported by the empirical data. However,  $\gamma/D_f < 1$  implies a long-range interaction between the population and the socio-economic variables will exhibit superlinear behaviour, as shown by equation (2.11), in accordance with empirical evidence.

In the context of this model, because we observe the superlinear behaviour, we expect that the individuals interact in a long-range manner in the cities. So, the definition of what is a city and how to measure its properties must be done with the following requirement in mind: it is not the political borders nor the arbitrary geographical borders that matters. What defines the cities is the geographical areas with interacting population, the so-called *functional cities* [10].

In the following, we discuss the main contributions of the present theory that can shed light into the study of cities.

#### 4.1. The explicit dependence of the scaling exponents

One of the main contributions of this work is to obtain an explicit dependence of the scaling exponents  $\beta$ , equations (2.10) and (2.17), with ratio  $\gamma/D_f$ . As defined in this work, the decay exponent of interaction  $\gamma$  and the fractal dimension of the city  $D_f$  are, in principle, both measurable.

As stated by equation (2.10) and shown in figure 3, the scaling exponent associated with the socio-economic activities  $\beta_{\text{se}}$  is a monotonically increasing function of the fractal dimension of the city  $D_f$ . As a consequence, for given  $\gamma$ , the denser the city, the greater the socio-economic scaling exponent, more intense the social interaction and so the city development, in accordance with [10]. As stated by equation (2.17), a similar reasoning is valid for the scaling exponent associated with the infrastructure  $\beta_{\text{infra}}$ .

#### 4.2. Individual productive capacity

One interesting aspect of the model proposed concerns the *individual productive capacity*, called  $G$  by Bettencourt [10]. He has shown that this quantity can be expressed as the product of the socio-economic production *per capita* and the infrastructure *per capita*, which reveals to be scale invariant:  $G \sim N^0$ . In the particular case of Bettencourt's empirical studies, it was shown that these properties of  $G$  can be observed when we use variables related to area (for instance, road surface area, circumscribing land area, etc.) as the infrastructure variable. In this work, one can show that this property of  $G$  can also be observed when we use the number of amenities (*per capita*) as the infrastructure variable. That is, according to the results

given by equations (2.10) and (2.17),

$$G = \left(\frac{I_{se}}{N}\right) \cdot \left(\frac{P}{N}\right) \sim N^{2-\gamma/D_f-1} N^{\gamma/D_f-1} = N^0, \quad (4.1)$$

which means  $G$  is a constant in relation to  $N$  (in the absence of a scaling law), regardless of the numeric values of  $D_f$  or  $\gamma$ . Thus, the result of the current work is compatible with the fact that  $G$ , being related to individual effort and thus limited to physical constraints, should not depend on the size of the city.

### 4.3. Relation between socio-economic and the infrastructure scaling exponents

Another important feature that emerges from the model is the relation between the socio-economic and infrastructure scaling exponents. For instance, the sum of these exponents,

$$\beta_{se} + \beta_{infra} = \left(2 - \frac{\gamma}{D_f}\right) + \left(\frac{\gamma}{D_f}\right) = 2 \quad (4.2)$$

is a number ( $= 2$ ), which does not depend on the parameters of the model ( $\gamma$  or  $D_f$ ). Moreover, this result corroborates the same relation found in [7] and it is in accordance with the empirical data, presented in table 1.

### 4.4. Different decay exponents

In our model, we are supposing that the decay exponent  $\gamma$  is the same for both socio-economic and infrastructure variables. However, the way citizens behave in relation to them can be different. Therefore, if we build a  $\gamma$  for each variable as:  $\beta_{infra} = \gamma_{infra}/D_f$  and  $\beta_{se} = 2 - \gamma_{se}/D_f$ , the sum of the scaling exponents will be

$$\beta_{se} + \beta_{infra} = 2 + \frac{\gamma_{infra} - \gamma_{se}}{D_f}. \quad (4.3)$$

Note that this sum can be below or above 2. Empirically, as we showed in table 1, the sum varies around 2. There are two alternative explanations for this: (i) this fluctuation is due to some measurement error or (ii) it is due to the difference between  $\gamma_{se}$  and  $\gamma_{infra}$ . More empirical evidence and, consequently, a more detailed study are necessary to clarify this issue.

### 4.5. Different urban variables with different scaling exponents

Now, we consider the possibility that different infrastructure sectors present different scaling exponents. For instance, let's take two different kinds of amenities in the city: bakeries and gas stations. According to empirical data presented by [31,32],  $\beta_{bakery} = 0.847$ , while  $\beta_{gas\ station} = 0.652$  in the USA. The model sheds some light on the difference between these values. If  $D_f$  is the same for both, the model says that if  $\beta_{bakery} > \beta_{gas\ station}$ , then  $\gamma_{bakery} > \gamma_{gas\ station}$ , which means that bakeries are more restricted, that is, they have a smaller range of interaction than gas stations. That makes sense, because people usually go to bakeries that are in the neighbourhood. However, it is more common for people to fuel their cars further away from their homes. The same idea can be applied in the context of the socio-economic variables.

### 4.6. The fluctuations in real data

There is one more interesting possibility allowed by the approach presented in this work: it can also give some insights about the fluctuations of the urban variables that are observed in empirical allometries. Empirical data from recent works [35,36] suggest that the residuals surrounding the scaling laws are log-normally distributed. The present model can explain such properties as a consequence of fluctuation of the interaction range of the citizen, given by particularities of the cities.

To understand that, consider the hypothesis that the exponent decay  $\gamma$  is Gaussian distributed in an ensemble of cities of the same population size. This implies that the scaling exponent  $\beta$  will be also normally distributed, since  $\beta \propto \gamma$ , which is consistent with the empirical data presented in figure 2. From the allometric equation  $Y \propto N^\beta$ , being  $\beta$  normally distributed and  $\log Y \propto \beta \log N$ , then  $\log Y$  will be also normally distributed. Then the urban indicator  $Y$ , which can represent both a socio-economic  $I_{se}$  or an infrastructure  $P$  variable, must be lognormally distributed, in the same way that it is observed in empirical data. In [35], the authors consider Gaussian noise in the multiplicative constant of the scaling equation to explain the lognormal distribution observed in real data. While they assumed  $\beta$  fixed without noise, we consider  $\beta$  is noise.

In conclusion, the model proposed suggests that the fluctuation in the interaction range of the individuals conducts to the lognormal distribution of the allometric metrics, in accordance with the empirical facts. In other words, fluctuations in urban variables would be due to the diversity and peculiarities of each city in promoting the spatial integration of individuals.

#### 4.7. Practical applications for urban development

Based on the previous considerations, we now address the role of the parameters  $\gamma$  and  $D_f$  in practical applications. In general, it is usual to say that the development of a city for good is associated with the increase of the socio-economic scaling exponents, equation (2.10), and the decrease of the infrastructure scaling exponents, equation (2.17). For both exponents, the smaller the ratio  $\gamma/D_f$ , the better the city development.

Creating public policies able to decrease the ratio  $\gamma/D_f$  should be the goal of cities' administration. This could be done (i) by increasing the density of the cities, or what is the same, the fractal dimension  $D_f$ ; (ii) by decreasing the decay exponent  $\gamma$ , which means increasing the interaction range between the citizens, by creating public policies capable of promoting distant individuals to interact by decreasing the cost of transportation, or improving and diversifying urban mobility, for instance; and (iii) by a combination of both, since those variables might be dependent or correlated one to the other.

However, these strategies might be bounded by correlations between what is generally seen as prosperity variables (such as GDP, income, employment, etc.) and what is generally seen as disadvantage variables (such as criminality and CO<sub>2</sub> emission levels). While the increase of the former variables are welcome, the increase of the latter are not. We need first to know to what extent those variables are independent or correlated before discussing specific policies. Empirical evidence regarding the values of  $\gamma$  for the different urban variables and their correlations is still missing and needs to be further studied before discussing specific policies.

### 5. Conclusion

This work sheds light on the understanding of scaling laws which emerge in cities. We argued that the cities' scaling laws observed in empirical data can be explained considering distance-dependent interactions at individual level. We developed a microscopic model approach, based on pair interaction of individuals with distance, characterized by a decay exponent  $\gamma$  that accounts for the superlinear and sublinear scaling, respectively, of the socio-economic and infrastructure variables. To obtain scaling exponents compatible with empirical data, the long-range interaction must be prevalent. We built an agent-based model to test the hypotheses behind the presented assumptions. The main idea is that the model can be built without taking some details into consideration. The scaling laws obtained from simulations were compatible with empirical data.

The proposed model presents an explicit relation for the scaling exponents  $\beta$  in terms of the decay exponent  $\gamma$ , and the fractal dimension  $D_f$  of the city. The model predicts that the sum of the scaling exponents of socio-economic and infrastructure variables are 2, as observed in empirical data, independent of the parameters of the model. The scaling exponents  $\beta$  can be different for different socio-economic and infrastructure variables, and we addressed how public policies could take advantage of these properties to improve cities' development, minimizing negative effects.

The model also presents an alternative interpretation for the lognormal distribution of the scaling variables. It proposes that the fluctuation in the interaction range of the individuals conducts to the lognormal distribution of the allometric metrics, in accordance with the empirical facts. In other words, fluctuations in urban variables would be due to the diversity and peculiarities of each city in promoting the spatial integration of individuals.

This framework elegantly allows researchers to make microscopic measurements of individual interactions and to integrate the outcome with the city geometrical properties to estimate the underlying scaling exponents. This study relies on the fact that cities have to be seen as a unit as their emergent scaling laws depend on the fact that cities work as integrated systems. Further research in this line is warranted.

**Data accessibility.** All data used are publicly available and properly cited.

**Authors' contributions.** The authors contributed equally to this study.

**Competing interests.** We declare we have no competing interests.

**Funding.** The authors thank the Brazilian funding agencies CAPES and CNPq for financial support. F.F.F. thanks the São Paulo Research Foundation (FAPESP) for supporting his research.

Acknowledgements. We acknowledge the useful insights of the anonymous referees, in particular, those that ultimately led the final version of subsections ‘The fluctuations in real data’ and ‘Practical applications for urban development’ in the Discussion section.

## References

- Batty M. 2013 *The new science of cities*. Cambridge, MA: MIT Press.
- Weaver W. 1948 Science and complexity. *Am. Sci.* **36**, 536–544.
- Jacobs J. 1961 *The death and life of great American cities*. New York, NY: Random House.
- Gabaix X. 1999 Zipf’s law for cities: an explanation. *Q. J. Econ.* **114**, 739–767. (doi:10.1162/003355399556133)
- Beckmann MJ. 1958 City hierarchies and the distribution of city size. *Econ. Dev. Cult. Change* **6**, 243–248. (doi:10.1086/449769)
- Henderson JV. 1974 The sizes and types of cities. *Am. Econ. Rev.* **64**, 640–656.
- Rozenfeld HD, Rybski D, Gabaix X, Makse HA. 2011 The area and population of cities: new insights from a different perspective on cities. *Am. Econ. Rev.* **101**, 2205–2225. (doi:10.1257/aer.101.5.2205)
- Bettencourt LMA, Lobo J. 2016 Urban scaling in Europe. *J. R. Soc. Interface* **13**, 20160005. (doi:10.1098/rsif.2016.0005)
- Bettencourt LMA, Lobo J, Helbing D, Kühnert C, West GB. 2007 Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306. (doi:10.1073/pnas.0610172104)
- Bettencourt LMA. 2013 The origins of scaling in cities. *Science* **340**, 1438–1441. (doi:10.1126/science.1235823)
- Bettencourt LMA, West G. 2010 A unified theory of urban living. *Nature* **467**, 912–913. (doi:10.1038/467912a)
- Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A, Batty M. 2015 Constructing cities, deconstructing scaling laws. *J. R. Soc. Interface* **12**, 20140745. (doi:10.1098/rsif.2014.0745)
- Cottineau C, Finance O, Hatna E, Arcaute E, Batty M. 2016 Defining urban agglomerations to detect agglomeration economies. (<http://arxiv.org/abs/1601.05664>)
- Arcaute E, Molinero C, Hatna E, Murcio R, Vargas-Ruiz C, Masucci AP, Batty M. 2016 Cities and regions in Britain through hierarchical percolation. *R. Soc. open sci.* **3**, 150691. (doi:10.1098/rsos.150691)
- Oliveira EA, Andrade Jr JS, Makse HA. 2014 Large cities are less green. *Sci. Rep.* **4**, 04235. (doi:10.1038/srep04235)
- Goldenberg J, Levy M. 2009 Distance is not dead: social interaction and geographical distance in the internet era. (<http://arxiv.org/abs/0906.3202>)
- Herrera-Yagüe C, Schneider CM, Couronne T, Smoreda Z, Benito RM, Zufiria PJ, MC. 2015 The anatomy of urban social networks and its implications in the searchability problem. *Sci. Rep.* **5**, 10265. (doi:10.1038/srep10265)
- Mombach JC, Lemke N, Bodmann BE, Idiart MAP. 2002 A mean-field theory of cellular growth. *Europhys. Lett.* **59**, 923–928. (doi:10.1209/epl/i2002-00244-6)
- dos Santos RV, Ribeiro FL, Martinez AS. 2015 Models for Allee effect based on physical principles. *J. Theor. Biol.* **385**, 143–152. (doi:10.1016/j.jtbi.2015.08.018)
- Ribeiro FL. 2015 A non-phenomenological model of competition and cooperation to explain population growth behaviors. *Bull. Math. Biol.* **77**, 409–433. (doi:10.1007/s11538-014-0059-z)
- Ribeiro FL, Ribeiro KN. 2015 A one dimensional model of population growth. *Phys. A* **434**, 201–210. (doi:10.1016/j.physa.2015.03.021)
- Ribeiro FL, Santos RVD, Mata AS. 2016 Fractal dimension and universality in avascular tumor growth. (<http://arxiv.org/abs/1610.05789>)
- Newman ME. 2005 Power laws, Pareto distributions and Zipf’s law. *Contemp. Phys.* **46**, 323–351. (doi:10.1080/00107510500052444)
- Batty M, Longley PA. 1994 *Fractal cities: a geometry of form and function*. New York, NY: Academic Press.
- Caglioni M, Giovanni R. 2004 Contribution to fractal analysis of cities: a study of metropolitan area of Milan. *Cybergeo* (online), Dossiers, document 269. (doi:10.4000/cybergeo.3634)
- Baish JW, Jain RK. 2000 Fractals and cancer. *Cancer Res.* **60**, 3683–3688.
- Falconer K. 2004 *Fractal geometry: mathematical foundations and applications*. New York, NY: John Wiley & Sons.
- Chen Y. 2010 Characterizing growth and form of fractal cities with allometric scaling exponents. *Discrete Dynamics in Nature and Society* **2010**, 194715. (doi:10.1155/2010/194715)
- Kühnert C, Helbing D, West GB. 2006 Scaling laws in urban supply networks. *Phys. A* **363**, 96–103. (doi:10.1016/j.physa.2006.01.058)
- Brand P, Dávila JD. 2011 Mobility innovation at the urban margins. *City* **15**, 647–661. (doi:10.1080/13604813.2011.609007)
- Hidalgo CA, Castañer EE. 2015 The amenity space and the evolution of neighborhoods. (<http://arxiv.org/abs/1509.02868>)
- Ensenat EC. 2015 Beyond city size: characterizing and predicting the location of urban amenities, PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Vicsek M. 1991 *Fractal growth: demonstrations of fractal aggregation models*, Dskt edition. Singapore: World Scientific.
- Murcio R, Masucci AP, Arcaute E, Batty M. 2015 Multifractal to monofractal evolution of the London street network. *Phys. Rev. E* **92**, 062130. (doi:10.1103/PhysRevE.92.062130)
- Alves LGA, Ribeiro HV, Lenzi EK, Mendes RS. 2013 Distance to the scaling law: a useful approach for unveiling relationships between crime and urban metrics. *PLoS ONE* **8**, e69580. (doi:10.1371/journal.pone.0069580)
- Leitão JC, Miotto JM, Gerlach M, Altmann EG. 2016 Is this scaling nonlinear? *R. Soc. open sci.* **3**, 150649. (doi:10.1098/rsos.150649)