
POTSDAM-INSTITUT FÜR KLIMAFOLGENFORSCHUNG

# Highlights

## Mathematical models to explain the origin of urban scaling laws

Fabiano L. Ribeiro,Diego Rybski

- We review the main mathematical models present in the literature that aim at explaining the origin and emergence of urban scaling.
- We identify similarities and connections between these models.
- The models treated in this paper explains urban scaling from different premises: from gravity ideas, passing through densification ideas and cites' geometry, to a hierarchical organization and socio-network properties.
- Regarding the gravity idea, we propose a general framework that includes all gravity models analyzed as particular cases.

# Mathematical models to explain the origin of urban scaling laws

Fabiano L. Ribeiro[a,*,1], Diego Rybski[b,c,d,*,2]

[a] Universidade Federal de Lavras (UFLA), Aquenta Sol, Lavras, 37200-900, MG, Brazil
[b] Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 601203, Potsdam, 14412, , Germany
[c] Department of Environmental Science Policy and Management, University of California Berkeley, 130 Mulford Hall #3114, Berkeley, 94720, CA, USA
[d] Complexity Science Hub Vienna, Josefstädterstrasse 39, Vienna, A-1090, CA, Austria

ARTICLE INFO

*Keywords*:
urban scaling
cities
complex systems

ABSTRACT

The quest for a theory of cities that could offer a quantitative and systematic approach to managing cities represents a top priority. If such a theory is feasible, then its formulation must be in a mathematical way. As a contribution to organizing the mathematical ideas that deal with such a systematic way of understanding urban phenomena, we review the main theoretical models present in the literature that aim at explaining the origin and emergence of urban scaling. We intend to present the models, identify similarities and connections between them, and find situations in which different models lead to the same output. In addition, we report situations where some ideas initially introduced in a particular model can also be introduced in another one, generating more diversification and increasing the scope of the original works. The models treated in this paper explain urban scaling from different premises, i.e. from gravity ideas, densification and cites' geometry to a hierarchical organization and social network properties. We also investigate scenarios in which these different fundamental ideas could be interpreted as similar – where the similarity is likely but not obvious. Furthermore, concerning the gravity model, we propose a general framework that includes all analyzed models as particular cases. We conclude the paper by discussing perspectives of this field and how future research designs and schools of thought can build on the ideas treated here.

## CRediT authorship contribution statement

**Fabiano L. Ribeiro:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Paper writing. **Diego Rybski:** Conceptualization, Funding acquisition, Methodology, Project administration, Paper writing.

fribeiro@ufla.br (F.L. Ribeiro); diego.rybski@pik-potsdam.de (D. Rybski)
ORCID(s): 0000-0002-2719-6061 (F.L. Ribeiro); 0000-0001-6125-7705 (D. Rybski)
https://twitter.com/@fabianorib77 (F.L. Ribeiro), https://twitter.com/@DiegoRybski (D. Rybski)
1

## Contents

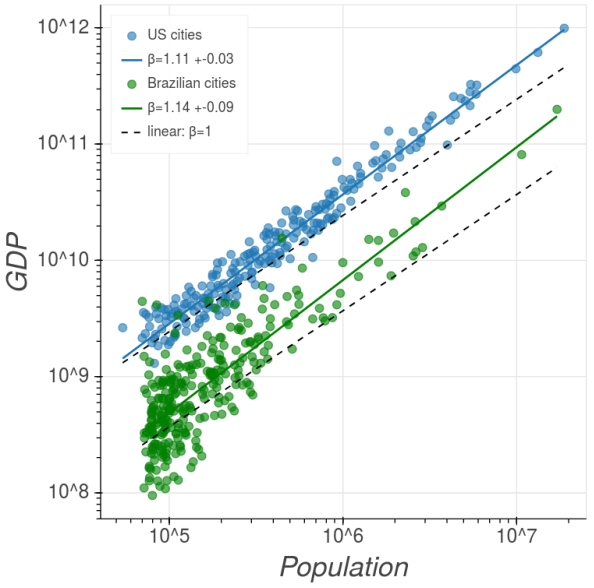


**Figure 1:** Log-log plot of the gross domestic product (GDP) as a function of population $N$, for USA and Brazilian cities (blue and green, respectively). The solid lines represent regressions to the points and reveal power law properties. Both countries have similar scaling exponents $\beta > 1$ (super-linear) despite the socio-economic differences between these two countries. The dashed lines indicate linearity ($\beta = 1$) and serve as guides to the eye.

## 1. Introduction

For the first time in human history, the urbanised population surpasses the rural population, and United Nations estimate that until 2050 more than 70% of the people around the world will live in cities[1]. To deal with all the problems that come with this urban intensification, like extreme density, traffic, infrastructure saturation, it is urgent to develop a quantitative theory in order to understand the urban phenomena and to govern our cities systematically [1, 2]. This theory will involve an interdisciplinary effort and could better predict scenarios to be explored by decision-makers and suggest new observations about the cities' growth and their organisation [3]. This theory, if successful, can point out where the data is missing and what we need to measure to obtain a deeper understanding of this phenomenon. Besides, if we expect this theory gives a quantitative description of cities, it must be formulated mathematically. With the purpose of providing a perspective, we organize and present mathematical ideas that aim at understanding cities systematically, concentrating on one aspect that is central to the *new science of cities* [3]: *urban scaling*.

Urban scaling analysis proposes that some quantity, say $Y$, grows free-of-scale and non-linearly with the population size $N$ of a city, following the form

$$Y = Y_0 N^{\beta}, \tag{1}$$

where $Y_0$ is a constant and $\beta$ is the *scaling exponent* [4]. To a great extent, empirical evidence reveals three distinct scaling regimes.

Variables related to socio-economic activities (e.g. GDP, Patents, AIDS cases) scale in a super-linear manner with the population size ($\beta > 1$). Empirical evidence for economically and culturally different countries and also for different urban metrics suggest a numerical value of the scaling exponent around $\beta = 1.15$ for socio-economic variables [4, 5]. It means the per-capita quantity of these socio-economic variables tends to increase with the size of a city – the so-called increasing returns to scale [6]. Fig. (1) presents an example of super-linear scaling of the GDP with the city population

[1] https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

size. Typically, one large city generates more wealth than two cities of half the size together. We can intuitively say that *the bigger the city is, the more wealth it generates* [7, 8].

On the other hand, variables associated with basic individual services (e.g. number of houses, water consumption) scale linearly with city population ($\beta = 1$). And infrastructure-related variables (e.g. electrical cables, number of gas stations) scale in a sub-linear manner ($\beta < 1$). Empirical evidence suggests a numerical value of the scaling exponent around $\beta = 0.85$ for infrastructure variables [4, 9, 5]. This means that bigger cities demand less infrastructure per-capita [10], which allows to say that: *bigger cities do more with less* [7, 8]. There is also evidence for some kind of constraint on the numerical value of the scaling exponents, such that these super- and sub-linear exponents add up to $\approx 2$ [11, 12].

A special form of urban scaling is the scaling relation between city area and population

$$A \sim N^{\beta_{\mathrm{FA}}} . \tag{2}$$

It takes a fundamental role because both – area size $A$ and population size $N$ – are measures of city size. As such, it relates to population density which is also an important characteristic of cities [13, 14, e.g.]. Therefore, we call it "*fundamental allometry*" [15]. Early results go back to [16] and [17], reporting $\beta_{\mathrm{FA}} \simeq 3/4$ and $\beta_{\mathrm{FA}} \simeq 2/3$, respectively. There is a set of papers empirically analyzing the fundamental allometry and [18] collected the exponents estimated in many such studies. Mostly $\beta_{\mathrm{FA}} < 1$ is reported, i.e. large cities exhibit higher population density. Plotting $\beta_{\mathrm{FA}}$ as a function of the year, a gradual shift can be seen [15, Fig.3], which is likely due to altering city definitions over the years. The topic gained new interest in the context of archaeology [19, 20, e.g.]. In a recent study, it was reported that 18 out of 38 countries (47 %) support $\beta_{\mathrm{FA}} = 5/6$ while for 17 out of 38 the scaling is indistinguishable from linearity [21].

Batty [3, p.41] argues, following [17], that in cities also the third dimension is being used, i.e. population in space $N \sim r^3$, where $r$ is the length scale. The area size of the city is in a plane, i.e. $A \sim r^2$. Eliminating $r$ we obtain $A \sim N^{2/3}$, corresponding to $\beta_{\mathrm{FA}} = 2/3$. Coffey [22, footnote 7] even argues that $\beta_{\mathrm{FA}} = 2/3$ represents the isometric reference value. Of course this is a very simple consideration and in particular, the vertical dimension is very different from the horizontal ones. But it indicates that we can expect $\beta_{\mathrm{FA}} < 1$ and higher densities (population per area) in large cities.

But why does urban scaling emerge at all? This paper tries to answer this question and focuses on works explaining non-linear urban scaling by some sort of model that goes beyond empirical characterization. Many (but not all) of the models discussed here are built on the idea that urban scaling results from a multiplicative combination of population, density, geometry, and hierarchical organization that enhance or disfavour the interaction between people. The premise is that interaction, and consequently the exchange of knowledge, generates ideas that result in innovation, economic growth, increasing returns, and economies of scale. In some models, the geometrical and network properties of the cities also play an essential role to explain the observed scaling laws, given that human interactions depend strongly on the city's spatial structure. Natural factors, such as rugged relief or the presence of physical barriers (mountains, rivers, lakes, etc.), promote or intensify the isolation of certain parts of the city. In addition, artificial factors, e.g. the geometry of the street networks or the city's shape, should also affect such human interactions.

We are aware that an enormous number of papers have been dedicated to presenting empirical and theoretical evidence about urban scaling in the last few years. Of course, it will not be possible to organize in a single paper all the results and ideas contained in those works. We opt to present only models that explain or derive urban scaling properties as an emergent phenomenon, giving special attention to those that derive it mathematically.

With the purpose of gaining insights from relating and comparing the models, we present them in a more straightforward manner than in the original publications. The intention is to focus only on what is strictly essential to explain urban scaling quantitatively. Some models are rewritten using a different notation from the original to get homogenization and coherence among the models. Most of the models' mathematical deductions are presented in a self-contained manner in this paper. However, in some cases, we opt to omit very extensive mathematical passages to preserve the text's dynamic and flux.

The paper also aims at synergies by relating all those models. That is, what emerges from the interconnection between different models to explain urban scaling? Is it possible to see some common properties in different approaches? In which direction could future research develop? Also, writing the models in a standard notation allows us to identify what hypotheses and results they have in common. Therefore, we organize the models in a taxonomy, identifying groups of models that share the same fundamental ideas, as organized in Figs. (2) and (3), and Tab. 1. For instance, we find that a set of models differs only in how the interactions between people are estimated, i.e. how the

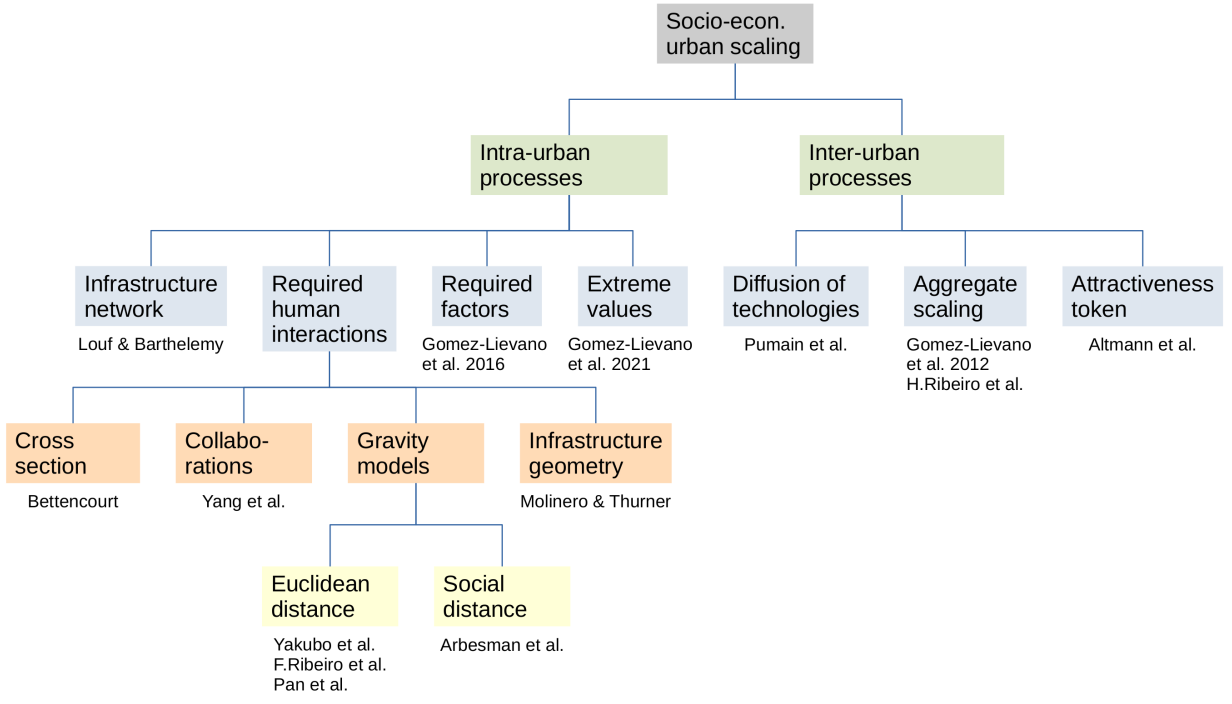


**Figure 2:** Taxonomy of models explaining socio-economic urban scaling. Whether processes take place within or between cities is the first distinguishing factor. Most models are based on required human interactions within cities.Only a small number of models are based on inter-city processes.

probability of interaction is derived. In the case of models based on gravity processes, this organization allows us to formulate a general framework with these models as particular cases.

The models that are presented here are divided into two categories: intra- and inter-city models. The intra-city models, which can be found in Sec. (I), refer to models that consider only city internal factors to explain the scaling laws. In these models, the interaction between people, and how geometry and the city spatial distribution affect it, is the main component to explain the laws of scales. We also propose general formulations from which some models could be derived as particular cases. In this category, all gravity models represent special cases of a general formalization.

The second category is about inter-city models, and it can be found in Sec. II. They consider the exchange of some kind of information between cities to explain the scaling laws. Not all of the models are based on derivations as a backbone and not all lead to the emergence of urban scaling, but they have been included to better represent the somewhat less developed group of inter-city models. The main mechanisms this category of models proposed to explain the scaling laws are: Zipf's law, hierarchical organization, and interaction among people of different cities.

We conclude with a discussion of perspectives for future research. Without anticipating too much of Sec. 7, we see a clear need for more sophisticated inter-city models – and ultimately for unification with their intra-city counterpart. At the same time, there is a strong presence of gravity models and, while being well linked to urban scaling, the connection to other gravity contexts, such as population flows or morphological growth, remains unknown territory. These and further perspectives are discussed in more detail at the end of the paper.

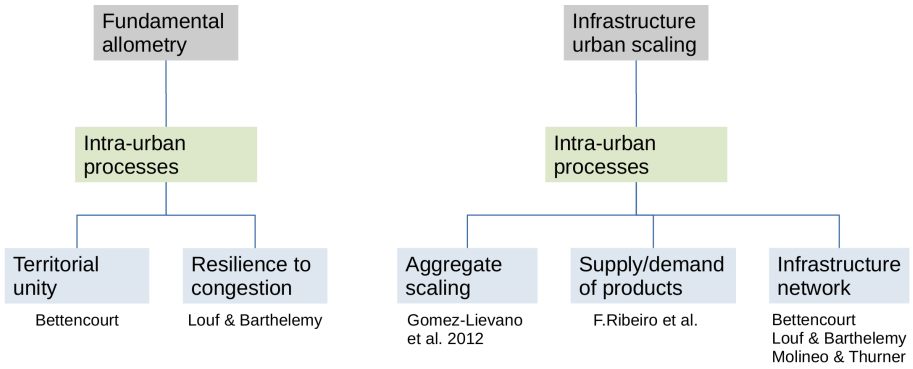


**Figure 3:** Taxonomy of models explaining fundamental allometry and infrastructure urban scaling.

# Part I
# Intra-city models

In the last years, we have seen a large number of works that consider (with empirical justifications) that urban variables are dependent on the population size $N$, without the necessity of incorporating other variables or even information from neighbouring cities. Indeed, the main idea behind it is that the outcome of an urban variable is a consequence of only the city's internal processes. However, we know that cities are in constant interaction with each other, and the dependence of urban variables solely on the city size may be a manifestation of a successful first-order approximation. This section is dedicated to presenting the ideas of the mathematical models that explain urban scaling and deriving an interpretation for the scaling exponent, using only endogenous factors.

## 2. Required Human interaction

Most publications on the topic consider human interaction as the primary mechanism to explain the origin of urban scaling. This fact obviously reflects in a larger space dedicated to this approach in the paper in hand. We begin the description by introducing some quantities common to the models that belong to this category. In addition, we present which properties a model based on interaction must have to be compatible with the empirical data.

### 2.1. General framework of human interaction models

Consider that the city is composed of $N$ individuals (population size of the city) that live in an area $A$ of the considered city. When two individuals meet in the city they generate ideas that correspond to a quantity $g$ of socio-economic activity. For instance, $g$ could represent the number of patents, the amount of wealth, etc., that this encounter contributes to. If each individual, say $i$, meets with $k_i$ persons, then the total wealth generated by these meetings is $g \cdot k_i$. The number $k_i$ can represent, depending on the model and without loss of generality, the *number of contacts* (friends, colleagues, or random encounters) of this individual, or the number of interactions that he/she has in a specific period of time, or even the node degree in a complex network.

What distinguishes the models presented below, is the way they propose to compute/determine the average number of contacts of the individuals, and consequently the production of socio-economic wealth generated by these contacts. The considered models essentially obey the following description. The city-wide total outcome $Y$ of a specific socio-economic variable can be understood as the sum of all individual socio-economic output. In turn, the *individual socio-economic production*, namely $y$, is the result of the socio-economic output generated by a single interaction, multiplied by the total number of interactions of each individual, $y = g \cdot k_i$. This idea is summarized in the diagram presented in Fig. (4) and yields the relation

$$Y = gN^2 n_c \,, \tag{3}$$

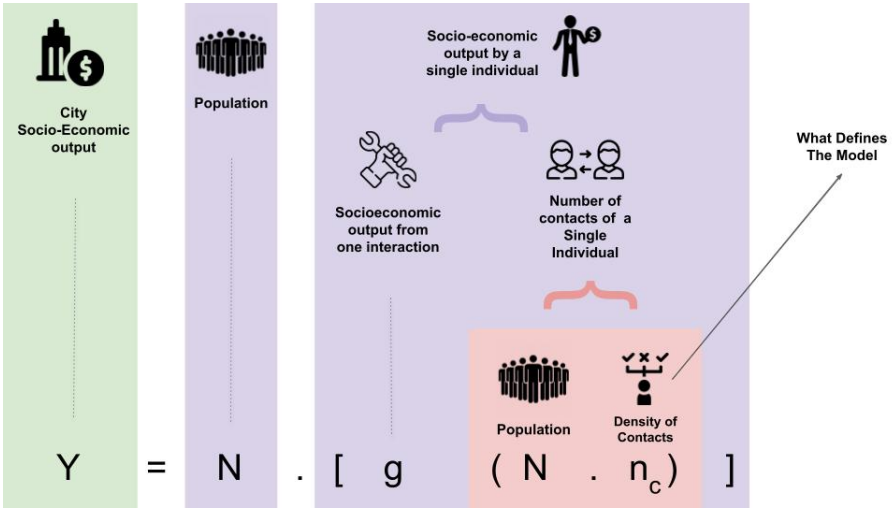


**Figure 4:** Diagram illustrating the models that consider urban scaling as a result of knowledge exchange via human interactions. The idea is that the total outcome $Y$ of a specific socio-economic variable can be understood as the sum of all individual socio-economic outputs of the city, that is $Y = Ny$. In turn, the individual socio-economic production $y$ is the result of the socio-economic output generated by a single interaction ($g$), multiplied by the total number of interactions of each individual ($k_i$). This idea yields the relation $Y = gN\langle k_i\rangle = gN^2 n_c$, where $n_c$, the average density of contacts, is what defines the models.

where

$$n_c \equiv \langle k_i \rangle / N \tag{4}$$

is the *average density of contacts*, also related to the probability of interaction. Essentially, the models differ in the way the authors propose to estimate this density.

If we consider that $g$ is scale-independent (i.e. $g \sim N^0$, as suggested by [11]), one obtains

$$Y \sim N^2 n_c \,. \tag{5}$$

This means we are looking for $n_c$ that leads from Eq. (5) to the empirical evidence

$$Y \sim N^{\hat{\beta}_{\text{super}}} \,, \tag{6}$$

where $\hat{\beta}_{\text{super}} \equiv 1.15$ is (approximately) the *empirical value* of the socio-economic scaling exponent. Equaling Eqs. (5) and (6), implies that the average density of contacts must follow the power-law

$$n_c \sim N^{\hat{\beta}_{\text{super}}-2} \tag{7}$$

in order to be compatible with the empirical findings.

Let us also use $\hat{\beta}_{\text{sub}} \equiv 0.85$ to represent (approximately) the *empirical value* of the infrastructure scaling exponent. As it was suggested in [11, 12], there is some kind of complementarity between these two scaling exponents that can be expressed by the constraint

$$\hat{\beta}_{\text{super}} + \hat{\beta}_{\text{sub}} = 2 \,. \tag{8}$$

It suggests that the exponent in Eq. (7) is, in fact, $\hat{\beta}_{\text{sub}}$, which allows us to write

$$n_c \sim N^{-\hat{\beta}_{\text{sub}}} \,. \tag{9}$$

This consideration is speculative at this point, but the following sections will provide some justification. The expressions Eqs. (7) and (9) represent a "rule of thumb" that the models based on interactions need to comply when quantitatively explaining scaling laws. The take-home message here is that the density of contacts and consequently the probability of interaction of any proposed model based on interaction must result in Eqs. (7) and (9) in order to be empirically consistent.

Table 1 summarizes the models that will be presented in the next sections, organizing their main mechanism, as well as the scaling exponents predicted by them. Specific to the models based on interaction, all of them consist of computing the average number of contacts of the people; and this quantity, in turn, will depend on the parameters used in each model. The following sections are dedicated to presenting these mathematical models in more detail.

| Kind of model | model | model mechanism | $n_c = \langle k_i \rangle / N$ | Interaction Probability | $\beta_{subb}$ | $\beta_{super}$ | $\beta_{FA}$ | obs. | Urban Variable |
|---|---|---|---|---|---|---|---|---|---|
| Human Interaction | Bettencourt (2013) | travel cost proportional to the soc.-econ. output $c \sim y$ | $n_c = a/A_n$ | $p_{intt} \propto n_c$ | $\frac{D_f+1/2}{D_f+1}$ | $2-\frac{(D_f+1/2)}{D_f+1}$ | $\frac{D_f}{D_f+1}$ | - | Urbanized area, infrastructure network, and socio-economic activities |
| Human Interaction | Yang et al. (2019) | It is necessary $q$ partners to do a socio-economic activity | - | - | - | $1+q\frac{\Delta \log \langle k_i \rangle}{\Delta \log (N)}$ | - | similar to other models when $q=1$ | socio-economic activities |
| Gravity model | Ribeiro et al. (2017) | - interaction among indiv. decay with the distance; - city as a fractal structure | $n_c \sim N^{-\frac{\gamma}{D_p}}$ | $p_{intt}(r) \sim \frac{1}{r^{\gamma}}$ | - | $2-\frac{\gamma}{D_p}$ | - | $\langle k_i \rangle = c_1 N^{1-\frac{\gamma}{D_p}} + c_2$ | socio-economic activities |
| Gravity model | Actrativess model Yakubo et al. (2014) | Actractivess of the indiv. is power-law distributed | $n_c \sim N^{-\frac{m(\alpha-1)}{D_p}}$ | $p_{intt}(r) \sim \frac{1}{r^{m(\alpha-1)}}$ | - | $2-\frac{m}{D_p}(\alpha-1)+\frac{\eta}{D_p}$ | - | When one considers the average degree is scaling invariant, then the model predicts Eq. (51) | socio-economic activities |
| Gravity model | Social Network Arbesman et al. (2009) | Population in a tree-shaped social network | $n_c \sim N^{-\phi+\lambda}$ | $p_{intt}(d) \sim b^{-\phi d}$ | - | $2-\phi+\lambda$ | - | - | socio-economic activities |
| Gravity model | Amenities distribution Ribeiro et al. (2017) | People buy products in closer amenities | - | $f(r_{ik}) \sim \frac{1}{r_{ik}^{\gamma}}$ | $\frac{\gamma}{D_p}$ | - | - | - | amenities |
| Gravity model | Pan e al. (2013) | Rank model and population uniformly distributed | $n_c \sim 1$ | $p_{intt}(r) \sim \frac{1}{r^2}$ | - | - | - | Conducts to $Y \sim N \ln(N)$ | socio-economic activities |
| Human Interaction and Infrastructure network | Molinero and Thurner (2019) | Relation between population distribution and street newtwork | $n_c \sim N^{-\frac{D_{infra}}{D_p}}$ | $p_{intt}(r) \sim \frac{1}{r^{D_{infra}}}$ | $\frac{D_{infra}}{D_p}$ | $2-\frac{D_{infra}}{D_p}$ | - | $\langle k_i \rangle \sim N^{1-\frac{D_{infra}}{D_p}}$ | Urban Infrastructure and socio-economic activities |
| Infrastructure network | Bettencourt (2013) | Hierarchical infrastructure network properties | - | - | $\alpha$ | - | - | $\alpha$ is a constant related to the way the width of the hyper-roads changes from one hierarchical level to the next. | Infrastructure network Area |
| Infrastructure network | Louf and Barthelemy (2013/2014) | Tra c congestion and city's activity centres | - | - | $\frac{\mu}{2\mu+1}+\frac{1}{2}$ | $\frac{\mu}{2\mu+1}+1$ | $\frac{2\mu}{2\mu+1}$ | $\mu$: resilience of the transport network to congestion | City's Area and activity centres number |
| Required Factors | Gomez-Lievano et al. (2016) | urban socio-economic phenomenon occurs when a number of necessary complementary factors are available in the city. | - | - | - | $1+Mbq$ | - | It is necessary that the probability that a given factor be provided by the city be logarithmically dependent on the population size | socio-economic activities |
| Extreme Values | Gomez-Lievano et al. (2021) | Selection process acting on independent randon variables | - | - | - | $\frac{\sigma}{\sqrt{2 \ln N}}$ | - | Superlinear scaling only for small N | socio-economic activities |
| Inter-Urban process | Pumain et al. (2006) | Hierarchical di usion process of innovations | - | - | - | - | - | - | socio-economic activities |
| Inter-Urban process | Gomez-Lievano et al. (2012) | urban scaling and city size distributions | - | - | $\frac{\alpha}{\alpha_Y}$ | $\frac{\alpha}{\alpha_Y}$ | - | predicted $\beta$ only represents an upper limit | Infrastructure and socio-economic activities |
| Inter-urban process | H.Ribeiro et al. (2021) | country-wide urban scaling and city size distributions | - | - | - | Eq. (157) | - | - | socio-economic activities |
| Human Interaction | Altmann et al. (2020) | Tokens are randomly assigned to the people | - | - | - | - | - | This model introduce ideas about integrating intra- and inter-city aspects | socio-economic activities |

**Table 1**
Overview of the main features of models that explain the non-linear urban scaling. This table also presents each model's mechanisms and their predicted scaling exponents.

## 2.2. Bettencourt model – human interaction as cross section

Probably the most influential model proposed to explain the origin of urban scaling is the one by Bettencourt [11]. The model shares similarities with the concept of the *cross section* as used in physics. It considers that each individual moves throughout the city prescribing, with an interaction radius $l_0$, a trajectory of length $l$. It implies that this individual accesses in his/her trajectory an area $a = l_0 \cdot l$ of the city.

The density of contacts can be considered to be the ratio

$$n_c = \frac{\text{Area accessible to the individual}}{\text{Area of the city}} = \frac{a}{A}\,. \tag{10}$$

Using Eq. (5) and keeping in mind that the area accessible $a$ is an intrinsic property of individuals, and therefore scale-independent ($a \sim N^0$), one obtains the relation

$$Y \sim \frac{N^2}{A}\,. \tag{11}$$

Before continuing, lets to focus initially on the fundamental allometry ($A$ as a function of $N$). To do that, Bettencourt assumes that each individual has a cost $c$ to move around in the city, which is proportional to the transversal length $L$ of the city, that is $c \sim L$. Then the total transport cost of the city is $T = Nc \sim NL$. Considering that area is a fractal structure with dimension $D_f$, and therefore it scales with the transversal length as $A \sim L^{D_f}$, then the total transport cost can be written as

$$T \sim N A^{\frac{1}{D_f}}\,. \tag{12}$$

Bettencourt also explores the hypothesis that the individual socio-economic production $y$ must be sufficient for each person to travel through the city. That is, $y$ must be sufficient to pay the transportation cost, which means $y \sim c$, ensuring a territorial unity of the city. As a consequence of this hypothesis, and of $y = Y/N$, one has $T \sim Y$. This result, together with Eqs. (11) and (12), yields

$$A \sim N^{\frac{D_f}{D_f+1}}\,, \tag{13}$$

and consequently, one has the fundamental allometry (FA) scaling exponent

$$\beta_{\text{FA}} = \frac{D_f}{D_f+1}\,, \tag{14}$$

revealing a sub-linear regime between area and city population, as supported by empirical evidence [23], where the city fractal dimension determines the scaling.

As a next step, Bettencourt also used those ideas to find the scaling ($\beta_{\text{sub}}$) of the infrastructure network area, namely $A_n$. To do this, he suggests considering the average distance between individuals, namely $\lambda$, and the density of individuals, say $\rho$. These two quantities are related to each other via $\rho = N/A = 1/\lambda^2$, which means that, on average, we have one single individual inside a square of size $\lambda$. It implies that $\lambda = \sqrt{A/N}$, and given the result (13), we obtain

$$\lambda \sim N^{-\frac{1}{2(D_f+1)}}\,, \tag{15}$$

and

$$\rho \sim N^{\frac{1}{D_f+1}}\,. \tag{16}$$

That is, the average distance between individuals decreases with the increase in city size. In other words, bigger cities are denser than smaller ones, as empirical studies suggest [23, 24].

Finally, one can write that $A_n \sim N \cdot \lambda$ (given that $\lambda$ is a kind of infrastructure per capita), and consequently, from Eq. (15),

$$A_n \sim N^{\frac{D_f+\frac{1}{2}}{D_f+1}}\,, \tag{17}$$

implying in a smaller than 1 scaling exponent

$$\beta_{\text{sub}} = \frac{D_f + \frac{1}{2}}{D_f + 1} \tag{18}$$

for the infrastructure network area. The particular case $D_f = 2$ leads to $\beta_{\text{sub}} = 5/6 \approx 0.83$, very close to the empirical value.

In relation to the socio-economic scaling ($\beta_{\text{super}}$), it can be determined by Eq. (11) and the fundamental allometry (13), that implies

$$Y \sim N^{2-\left(\frac{D_f}{D_f+1}\right)} . \tag{19}$$

That is

$$\beta_{\text{super}} = 2 - \left(\frac{D_f}{D_f + 1}\right) , \tag{20}$$

which shows that the socio-economic scaling exponent is greater than 1 (super-linear) and is governed by the city's fractal dimension. However, for this result to be compatible with the empirical finds $\hat{\beta}_{\text{super}} \approx 7/6 \approx 1.16$, it is necessary that $D_f = 5$, which obviously is incompatible. Bettencourt, in order to solve this problem, proposes that Eq. (11) must be written in terms of the infrastructure network area, that is

$$Y \sim \frac{N^2}{A_n} . \tag{21}$$

It means the density of contacts ($n_c$) depends on the ratio between the area accessible to an individual and the infrastructure network area: $n_c = a/A_n$. Then, from Eq. (17), one gets

$$Y \sim N^{2-\left(\frac{D_f+1/2}{D_f+1}\right)} \tag{22}$$

leading to an alternative value for the scaling:

$$\beta_{\text{super}} = 2 - \left(\frac{D_f + 1/2}{D_f + 1}\right) . \tag{23}$$

Note that if $D_f = 2$ then $\beta_{\text{super}} = 7/6$, which is the empirical value [4, 5]. For more components and details about this model and further development, we refer to [11, 25, 26, 27].

In conclusion, the result Eqs. (18) and (22) shows that Bettencourt's considerations predict the empirical scaling exponents quantitatively. This is remarkable, given that the model is based on rather specific hypotheses. For instance, the model considers only the area as an infra-structure variable and does not take into account the number of amenities, that also scales sub-linearly with the population size [28, 9, 12]. In addition, buildings are not taken into account by the model, which would expand the interaction range from a two-dimensional area to a three-dimensional volume. The following models shed some light on this discussion.

## 2.3. Yang et al. model – required collaboration

Yang et al. [29] explain super-linear scaling as the result of the likelihood of finding the required collaboration in the city, necessary for an undertaking. Consider that the development of a certain prototype requires $q + 1$ experts. The case $q = 0$ means that one single person can accomplish all the processes necessary for such activity. In the other mathematical models presented here, as the Bettencourt model discussed in the last section and the gravity models presented in the next, the activity and the productivity (that we call $g$) demand two persons, that is $q = 1$.

The authors introduce $p_q\left(k_i\right)$ as the probability that an individual $i$ finds all $q$ collaborators required for an undertaking, among $k_i$ contacts. The socio-economic production is then given by

$$Y \sim N p_q\left(\langle k_i \rangle\right) . \tag{24}$$

where $\langle k_i \rangle$ is the average number of unique contacts for a person living in the city. They show that $p_q\left(\langle k_i \rangle\right) \sim \langle k_i \rangle^q$ and consequently

$$Y \sim N \langle k_i \rangle^q . \tag{25}$$

Let us first define

$$\Delta \log(N) \equiv \log N - \log N' , \tag{26}$$

where $N$ and $N'$ are the populations of two different cities. The same definition can be applied to $\Delta \log(Y)$ and $\Delta \log\langle k_i \rangle$ in an analogous way. With this definition and given that $Y = Y_0 N^{\beta_{\text{super}}}$, where $Y_0$ is a constant (the intercept), one gets

$$\begin{aligned} \Delta \log(Y) &\equiv \log Y - \log Y' \\ &= \beta_{\text{super}} \log N + \log Y_0 - \beta_{\text{super}} \log N' - \log Y_0 , \end{aligned} \tag{27}$$

and consequently

$$\beta_{\text{super}} = \frac{\Delta \log(Y)}{\Delta \log(N)} . \tag{28}$$

Returning to Eq. (25), applying the logarithm, and using the definition above yields

$$\Delta \log Y \sim \Delta \log N + q \Delta \log\langle k_i \rangle , \tag{29}$$

in which dividing all the terms by $\Delta \log N$ and identifying Eq. (28), one obtains the following expression for the scaling exponent

$$\beta_{\text{super}} \sim 1 + q \frac{\Delta \log\langle k_i \rangle}{\Delta \log N} . \tag{30}$$

Some implications can be inferred from this result. First of all, a necessary condition for super-linear scaling is that the number of contacts is greater in larger cities, that is $\frac{\Delta \log\langle k_i \rangle}{\Delta \log N} \geq 0$, which is supported by empirical data [30]. The second necessary condition for the super-linearity is that more than 1 person is necessary (that is $q > 0$) to implement the undertaking. In this sense, an undertaking that can be done alone ($q = 0$) generates a liner scaling ($\beta_{\text{super}} \sim 1$) which falls into the category of individual needs variables. Eq. (30) also implies that if one does not need connections to realize something, the aggregate volume of this activity will scale linearly, without increasing returns to scale. In this sense, increasing returns can be attributed to enterprises that exceed $q = 0$. In contrast, Eq. (30) suggests that urban outputs requiring more participants should lead to more pronounced super-linear scaling.

## 2.4. Gravity Models

This section presents a set of models that explain urban scaling laws using the idea that the interaction between any pair of individuals within a city decays with the distance separating them – similar to the Newtonian gravity of two massive bodies. For a review of the application of gravity ideas in urban phenomena see [31, 32, 33, 34]. Before presenting these models in more detail, let's define some quantities and introduce some concepts that are common to them.

First of all, we denote $dN(\mathbf{r})$ as the number of people inside a *hyper-volume element* $d\mathbf{r}$ embedded in a $D$ dimensional space. For instance, this hyper-volume element is an area if $D = 2$; or a volume if $D = 3$. Moreover, $\mathbf{r}$ is a vector directed from one particular individual, say $i$, to the individuals that are in the hyper-volume element $d\mathbf{r}$. The vector $\mathbf{r}$ can be interpreted in two ways, (i) as *position vector* in $D$ dimensions, and consequently its modulus is the Euclidean distance $r$, conform illustrated in Fig. (5); or (ii) as a non-Euclidean distance that separates any two

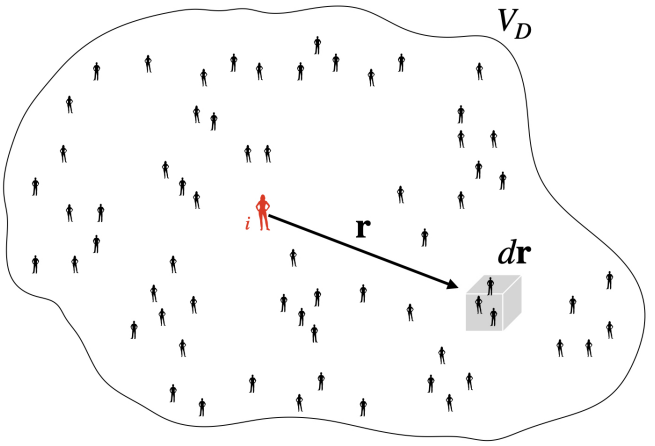


**Figure 5:** Illustration of the particular case that the vector $\mathbf{r}$ represents the Euclidean distance between the individual $i$ and the *hyper-volume element* $d\mathbf{r}$ (as used in the context of gravity models). All individuals inside this hyper-volume element ($d\mathbf{r}$) are at distance $r = |\mathbf{r}|$ from $i$. The population (of the entire city) is completely embedded in the hyper-volume $V_D$.

individuals inside a complex network. With the definition of theses quantities, one can compute, for instance, the total number of individuals in the city as $N = \int dN(\mathbf{r})d\mathbf{r}$ or the *density of individuals* at $\mathbf{r}$ as $\rho(\mathbf{r}) = dN(\mathbf{r})/d\mathbf{r}$, which characterize how the population arranges itself in space.

The gravity idea enters when we consider the probability, say $p_{\text{int}}(\mathbf{r})$, of the $i$-th individual interacting with – or to be a contact of – someone who is at $\mathbf{r}$. Then $p_{\text{int}}(\mathbf{r})dN(\mathbf{r})$ is the (average) number of contacts that this individual has in $\mathbf{r}$, and consequently the total number of contacts of this individual, say $k_i$, will be given by the integral

$$k_i = \int p_{\text{int}}(\mathbf{r})dN(\mathbf{r}) = \int p_{\text{int}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}\,, \tag{31}$$

where the integral cover all the space, that is, the hyper-volume $V_D$ where the city is embedded (see Fig. (5)).

Moreover, let's denote $g(\mathbf{r})$ as the socio-economic production generated by the interaction between $i$ and another individual located at $\mathbf{r}$. It is plausible to assume that the interaction between two more distant individuals can be more productive than the interaction between closer individuals ("The strength of weak ties" [35, 36, 37]), since distant individuals are exposed to different experiences. With these considerations, it makes sense to interpret $p_{\text{int}}(\mathbf{r})dN(\mathbf{r})\cdot g(\mathbf{r})$ as the socio-economic production generated by the interaction between $i$ and all the other individuals at $\mathbf{r}$. Then the total socio-economic production of this individual is

$$y_i = \int p_{\text{int}}(\mathbf{r})\rho(\mathbf{r})g(\mathbf{r})d\mathbf{r}\,. \tag{32}$$

Finally, the total socio-economic production of the city, that is $Y = Ny_i$, can be written as

$$Y = N\int p_{\text{int}}(\mathbf{r})\rho(\mathbf{r})g(\mathbf{r})d\mathbf{r}\,. \tag{33}$$

This is the generic formulation of the models based on the gravity idea. What differentiates these models is the way the authors propose to define:

1. the metric associated with the vector $\mathbf{r}$;
2. the probability of interaction $p_{\text{int}}(\mathbf{r})$ and its dependence (decay) with $\mathbf{r}$;
3. the spatial distribution of the population, characterized by $\rho(\mathbf{r})$;
4. and the socio-economic production $g(\mathbf{r})$ generated per encounter/contact.

In the following, we present the gravity models, from the simplest version to more complex ones. Table 1 summarizes the main findings of these models.

### *2.4.1. F.Ribeiro et al. model – simple gravity*

The model studied by Ribeiro et al. [12] considers that the probability of interaction between two individuals decays with the *Euclidean distance* $r$ that separates them according to a power-law

$$p_{\text{int}}(r) = \frac{1}{r^{\gamma}}\,, \tag{34}$$

where $\gamma$, the *decay exponent*, is a parameter of the model which measures the *interaction range*. Some empirical evidence that support the hypothesis Eq. (34) can be found in [38, 39, 40].

The model also considers that the *population space distribution* forms a fractal structure with fractal dimension $D_p$ embedded in a hyper-volume with Euclidean dimension $D$ (holding $D_p \leq D$). This assumption allows writing the density of individuals at $\mathbf{r}$ as

$$\rho(\mathbf{r}) = \frac{\text{number of individuals}}{\text{hyper-volume}} = \rho_0 \frac{r^{D_p}}{r^{D}} = \rho_0 r^{D_p - D}\,, \tag{35}$$

where $\rho_0$ is a constant. Finally, they consider that all interactions have the same socio-economic production, that is

$$g(\mathbf{r}) = \text{const}\,. \tag{36}$$

Combining all assumptions [Eqs. (34), (35) and (36)], the total socio-economic production of the city can be determined by solving the integral Eq. (33). This can be done by transforming the integration element from Cartesian to hyperspherical coordinates, that is using

$$d\mathbf{r} = r^{D-1} dr d\Omega\,, \tag{37}$$

where $d\Omega$ is the *solid-angle*, which leads to

$$Y = c_1 N^{2-\frac{\gamma}{D_p}} + c_2 N\,, \tag{38}$$

where $c_1$ and $c_2$ are constants. A similar calculus using polar coordinates is presented in [41]. Using similar ideas one can calculate the average number of contacts using Eq. (31), which yields $\langle k_i \rangle = c_1 N^{1-\frac{\gamma}{D_p}} + c_2$, and the density of contacts using Eq. (4), which yields $n_c \sim N^{-\frac{\gamma}{D_p}}$.

This result allows the following interpretation. The case $\gamma > D_p$ characterizes a *short-range interaction regime* where the linear term in Eq. (38) dominates for sufficiently large $N$; that is, the system converges to $Y \sim N$, i.e. a linear relation between urban metrics and population. The case $\gamma < D_p$ characterizes a *long-range interaction regime* where the non-linear term in Eq. (38) dominates for sufficiently large $N$. That is, the system behaves in a super-linear way, characterized by $Y \sim N^{2-\frac{\gamma}{D_p}}$. This means when there are long-range interactions between the individuals, then the super-linear scaling exponent emerges, and is given by

$$\beta_{\text{super}} = 2 - \frac{\gamma}{D_p}\,. \tag{39}$$

This result also suggests that the scaling exponent is determined by the ratio of two geometrical parameters, the decay exponent and the population fractal dimension. Some recent works [42, 43] applied the same approach adopted here to study tumour growth, reaching similar results. It allows some kind of analogy between urban systems and the dynamics of cancer cells.

According to this model, the super-linearity of the socioeconomic variables, expressed by Eq. (39), is a consequence of the integrity of the city, in the sense that the super-linear behaviour of the socio-economic activity should only appear when there are interactions within the entire city (long-range regime). Otherwise, if the city is formed by isolated regions (short-range regime), the number of interactions and consequently the socio-economic metrics will depend linearly on the population size, without increasing returns to scale. It is interesting to note that this result is in accordance with an argument used by Bettencourt (see Sec. 2.2), namely that the per-capita socio-economic production must be sufficient to pay the transportation cost ($y \sim c$), ensuring a territorial unity of the city. Remarkably, two different

approaches use the same argument to explain urban scaling, i.e. the interconnection between the parts that constitute the city.

In addition, the result Eq. (39) also allows us to conclude that the smaller the $\gamma$, that is, the larger the access of the people to more distant parts of the city, the more pronounced is the socio-economic scaling. That is, the larger the region of people's access, e.g. due to an efficient transport system, the better the city's socio-economic metrics and the more pronounced are the increasing return to scale[2]. Obviously, an efficient transport system only represents a necessary condition to increasing return to scale. Further conditions – as the presence of influential people integrating distant parts of the city or the interaction between socially distant people – are discussed in the following sections (other gravity models). The decay exponent $\gamma$ rather represents a compound value of the influence of distance [44, 45]. This leads to the question about the role of information and communication technology (ICT). With the establishment of the Internet and e.g. video conference systems, physical vis-à-vis meetings might become less important, which would reduce the influence of the distance [45, 46]. However, the temporal evolution of $\gamma$ remains to be proven empirically. With the following models and the introduction of further concepts, we present some insights and possible interpretations for this parameter $\gamma$ once it is only an arbitrary parameter at this point.

### *2.4.2. Yakubo et al. model of individual attractiveness*

Yakubo et al. [47, 48] consider an additional ingredient in the gravity approach, namely that people exhibit different attractiveness to one another depending on how influential an individual is. To model this aspect, the authors consider a set of random variables $\{x_i\}_{i=1..N}$, each one associated with a given individual and following a power-law distribution (a pdf)

$$s(x) \sim x^{-\alpha}\,, \tag{40}$$

where $\alpha$ is a parameter of the model. The higher the value of $x$ the more attractive the individual is.

Any two individuals, say $i$ and $j$, are connected to each other if

$$\frac{x_i x_j}{r_{ij}^m} > \Theta\,, \tag{41}$$

where $\Theta$ is a threshold constant, $r_{ij}$ is the Euclidean distance between them, and the exponent $m$ is a parameter of the model. The probability that the $i$-th individual is connected to another individual at a distance $r$ can be computed by

$$p_{\text{int}}(r) = \begin{cases} 1, & \text{if } r \le \xi \\ \int_{x>\Theta r^m/x_i} s(x)dx, & \text{if } r > \xi \end{cases}\,, \tag{42}$$

where $x = \Theta r^m/x_i$ is the lower limit of the $x$ values necessary for an individual at $r$ to interact with $i$, and

$$\xi \equiv \left(\frac{x_{\text{min}}^2}{\Theta}\right)^{\frac{1}{m}} \tag{43}$$

is a distance below which any two individuals are connected regardless of $x$. Here $x_{\text{min}}$ is the smallest value assumed by $x$, i.e. $x_{\text{min}} \equiv \min\{x_i\}$. A condition for convergence of the integral in Eq. (42) is $\alpha > 1$, and if this is the case, then the solution of this integral, using the distribution (40), is

$$p_{\text{int}}(r) \sim \frac{1}{r^{m(\alpha-1)}}\,, \tag{44}$$

for $r > \xi$. Note that if we identify

$$\gamma = m(\alpha - 1)\,, \tag{45}$$

then we recover Eq. (34), as used in Ribeiro et al. model (Sec. 2.4.1).

[2]It is worth mentioning that a larger $\beta_{\text{super}}$ does not automatically imply a wealthier urban system. It can also be a result of an economic imbalance between smaller and larger cities.

The authors also consider a more generic shape for the socio-economic production generated by each interaction, namely

$$g(r) \sim r^{\eta}\,, \tag{46}$$

where $\eta$ is a parameter, in the sense that the productivity increases with the distance when $\eta > 0$ and decreases with the distance when $\eta < 0$; $\eta = 0$ means that all connections have the same socio-economic contribution as considered in the previous model. The authors also consider that the *population spatial distribution* is a fractal structure with dimension $D_p$ and therefore Eq. (35) is also valid in this context.

Combining all ingredients of the model, i.e. using Eqs. (44), (46), and (35), it is possible to compute the average degree (average number of contacts), namely $\langle k_i \rangle$, via

$$\langle k_i \rangle = \int \int \rho(r) p_{\text{int}}(r) s(x) dx dr\,, \tag{47}$$

(from (31) and considering the average from the distribution $s(x)$) and the socio-economic output with Eq. (33). It results, respectively, in

$$\langle k_i \rangle = c_1 \xi^{D_P} + c_2 \xi^{m(\alpha-1)} N^{1-\frac{m(\alpha-1)}{D_P}} \tag{48}$$

and

$$Y = c_3 \xi^{\eta+D_P} N + c_4 \xi^{m(\alpha-1)} N^{2+\frac{\eta-m(\alpha-1)}{D_P}}\,, \tag{49}$$

where $c_1, c_2, c_3$ and $c_4$ are constants.

For the interpretation of this result, it is necessary to distinguish two situations concerning the distance $\xi$ (or $\Theta$ parameter in Eq. (43)) and its scaling properties. If $\xi$ is scale-invariant (i.e. $\xi \sim N^0$), and $N$ sufficiently large, then

1. $Y \sim N$ when $m > (D_p + \eta)/(\alpha - 1)$, i.e. the first right-hand term in Eq. (49) dominates, which characterize a short-range interaction regime (see Sec. 2.4.1); and
2. $Y \sim N^{2+\frac{\eta-m(\alpha-1)}{D_p}}$ when $m < (D_p+\eta)/(\alpha-1)$, i.e. the second right-hand term in Eq. (49) dominates, characterizing a long-range interaction.

This means the super-linear scaling behaviour happens in a long-range kind regime, with exponent

$$\beta_{\text{super}} = 2 - \frac{m}{D_p}(\alpha - 1) + \frac{\eta}{D_p}\,. \tag{50}$$

According to the model, the super-linearity of the socio-economic scaling exponent can occur even when the productivity generated by the interaction is independent of the distance (i.e. when $\eta = 0$). In fact, the main factor that controls this super-linearity is the ratio between the interaction range (expressed by $\gamma = m(\alpha - 1)$) and the fractal dimension of the city ($D_p$), as it was already suggested in the previous section. Indeed the Yakubo et al. model and the gravity model studied by Ribeiro et al. are equivalent when $\eta = 0$ and $\alpha = 2$.

In addition, the result Eq. (45) gives some insights for the parameter $\gamma$. It suggests that this parameter, which controls the interaction range, depends not only on the geometric properties – expressed by the parameter $m$ – but also on the degree of influence of the people who compose the city, expressed by the parameter $\alpha$. Moreover, when the parameter $\alpha$ is sufficiently large, representing the situation where the influence is distributed around a typical value, a short-range interaction regime is observed. It corresponds to a more homogeneous population in terms of influence. A larger $\alpha$ – which can also be thought of as an absence of a concentration of influence – leads to a smaller $\beta_{\text{super}}$, i.e. it reduces the increasing returns to scale. Conversely, suppose $\alpha$ is sufficiently small, representing the situation where some people exert a considerable influence on the population. In that case, a long-range interaction regime is observed, where the city behaves in a more integrated way. To sum up, the result Eq. (50) suggests that to improve the urban socio-economic metrics (larger $\beta_{\text{super}}$), it is essential not only to provide good access to other parts of the city – as it was discussed in the previous subsections – but also to have influencers in the population who can establish interaction between distant parts of a city.

The result described above holds for $\langle k_i \rangle$ scaling with $N$ according to Eq. (48). However, the result of the scaling exponent changes drastically when we consider an average degree that is scale-invariant, i.e. $\langle k_i \rangle \sim N^0$, as proposed originally [48]. According to Eq. (48), this implies that $\xi$ scales with the population size $\xi \sim \langle k_i \rangle^{\frac{1}{D_p}}$ when $m > D_p/(\alpha - 1)$ and $\xi \sim N^{\frac{1}{D_p} - \frac{1}{m(\alpha-1)}}$. Inserting such result in Eq. (49) yields the following scaling exponents:

$$\beta = \begin{cases} 1, & \text{if } m(\alpha-1) \leq D_p \\ & \text{and } m(\alpha-1) \leq D_p + \eta \\ 2 - \frac{m(\alpha-1)-\eta}{D_p}, & \text{if } D_p \leq m(\alpha-1) \leq D_P + \eta \\ 2 + \frac{\eta}{D_p} - \frac{D_p+\eta}{m(\alpha-1)}, & \text{if } m(\alpha-1) < D_p \\ & \text{and } m(\alpha-1) > D_p + \eta \\ 1 + \frac{\eta}{D_p}, & \text{if } m(\alpha-1) < D_p \\ & \text{and } m(\alpha-1) < D_p + \eta \end{cases} . \tag{51}$$

Figure (6) synthesizes these results, revealing many possibilities of regimes (sub-linear, super-linear, and linear) for the scaling exponents according to the parameters of the model. It is important to note the role of the parameter $\eta$ in this context. The non-linearity (super- or sub-linear regimes) only happens for $\eta \neq 0$. The value of this parameter can also change the regimes, from sub-linear ($\eta < 1$) to super-linear ($\eta > 1$). The parameter $\eta$ also changes the region of parameters that we are interpreting as a long-range interaction (blue part of Fig. 6) and the short-range interaction (light-red part of Fig. 6). The main results of this model are summarized in Tab. (1).

### *2.4.3. Arbesman et al. model – Tree-shaped social network*

Up to now, the models that we presented are based on the idea that the interaction between people needs to overcome geographical distance. However, the work developed by Abersman et al. [36] shows that urban super-linear scaling can also emerge when a hierarchically organized social network is considered.

The authors propose that the population is organized in a tree-shaped social network, as the one sketched in Fig. (7). The $N$ individuals of the population are in the $N$ leaves of this tree, and every branch in this hierarchical network splits into $b$ new branches. The social distance $d$ between two individuals is defined as the height of their lowest common ancestor. In relation to the general gravitational framework that was presented at the beginning of this section, the distance vector becomes a scalar, that is $\mathbf{r} \to d$, and it does not represent the physical distance, but rather the social distance.

One can demonstrate that the *number of leaves* that are at social distance $d$ from a given individual, say $N(d)$, grows exponentially

$$N(d) = b^d \, . \tag{52}$$

The exponential structure of this relation motivates us to model the other quantities necessary to compute Eq. (33) with other exponential functions. For instance, the authors considered that the probability of interaction between two nodes drops off exponentially with the network distance as

$$p_{\text{int}}(d) \sim b^{-\phi d} \, , \tag{53}$$

where $\phi$ is the parameter that controls the range of interaction inside this network. Finally, the authors propose that the social productivity of the interaction between two individuals also depends exponentially on the social distance between them

$$g(d) = b^{\lambda d} \, . \tag{54}$$

Here $\lambda$ is a parameter, in the sense that the productivity increases with the social distance when $\lambda > 0$ and decreases with the social distance when $\lambda < 0$. All interactions have the same productivity when $\lambda = 0$. This parameter is similar to the parameter $\eta$ in the context of the Yakubo et al. model of individual attractiveness (see Sec. 2.4.2).

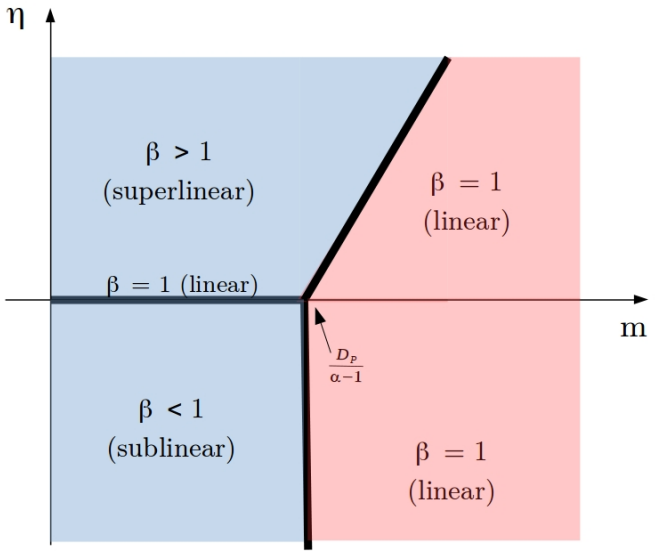


**Figure 6:** Phase diagram of the possible regimes (super-linear, sub-linear, and linear) according to the Yakubo et al. model of individual attractiveness, Eq. (51), and the model parameters $\eta$, $m$, $D_p$ and $\alpha$, assuming $\langle k_i \rangle \sim N^0$ (scale-invariant). In this case, the non-linearity (sub or super-linear) only happens when $\eta \neq 0$. The diagram also presents the parameter configuration that yields a long-range interaction regime (blue filling) and a short-range interaction regime (light-red filling). The super-linearity only occurs in the presence of long-range interaction (sufficiently small $m$) and with $\eta$ positive; that is, when the productivity among pairs increases with the distance. Source: modified after Yakubo et al. [48].

We determine the total productivity inserting these tree relations Eqs. (52), (53), and (54) in the general relation Eq. (33). Given that $d$ is a discrete variable, implying $d\mathbf{r} \to \Delta d = 1$, the integral in Eq. (33) becomes the sum

$$Y = N \sum_{d=1}^{\log_b N} b^{-\phi d} b^d b^{\lambda d} , \tag{55}$$

where $\log_b N$ is the maximum social distance in the network. The sum in Eq. (55) is in fact a geometric progression which can be solved analytically, yielding $Y \sim N^{2-\phi+\lambda}$, from which

$$\beta_{\text{super}} = 2 - \phi + \lambda \tag{56}$$

follows.

This result shows that the socio-economic scaling exponent is larger when socially distant people interact (characterized by smaller values of $\phi$ and positive values of $\lambda$). This means the city improves its socio-economy when the interaction among socially different people is possible – in a similar way as it was discussed previously in the context of geographic distance. As argued in [36], "rich interconnectivity between communities creates better cities", implying that socially distant ties can be a socio-economic force.

For the sake of completeness, given the probability of interaction Eq. (53), together with Eq. (52), one can calculate the average number of contacts of this model using Eq. (31), which yields $\langle k_i \rangle = N^{1-\phi+\lambda}$, and the density of contacts using Eq. (4), which yields $n_c \sim N^{-\phi+\lambda}$. These results are summarised in the Tab. (1).

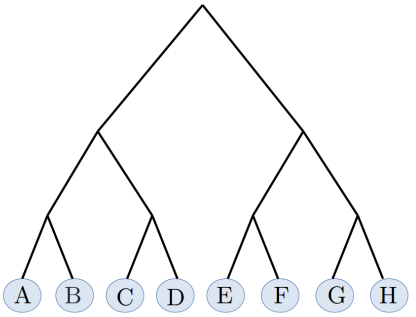


**Figure 7:** Tree-shaped social-network – as used by the Arbesman et al. model – composed by $N = 8$ individuals, named $A, B, ..., H$, disposed in $N$ leaves. In this particular case, all branches are split into $b = 2$ other branches. The distance $d$ between any two individuals in this network is the height of their lowest common ancestor. For instance, the social distance between $A$ and $B$ ($AB$) is $d = 1$; $AC$ and $AD$ are $d = 2$; $AE$, $AF$, $AG$ and $AH$ are $d = 3$. Source: [36].

### *2.4.4. Connection between Euclidean and social distance*

We have discussed the gravity ideas in two versions, considering the physical distance and social distance. However, we would also like to show that under certain circumstances these two versions can be understood as equivalent. First of all, if we consider that the two models are compatible, then the probability of interaction between any two individuals must be the same, that is

$$p_{\text{int}} \sim b^{-\phi d} \sim r^{-\gamma} \tag{57}$$

if we combine Eqs. (34) and (53). Moreover, if both model-versions are compatible, then also the scaling exponent must be the same, i.e. combining Eqs. (39) and (56), and considering $\lambda = 0$ without loss of generality leads to $\beta_{\text{super}} = 2 - \phi = 2 - \gamma/D_p$, which implies

$$\phi = \frac{\gamma}{D_p}\,. \tag{58}$$

Inserting Eq. (58) in Eq. (57) one can conclude that the two approaches are similar – i.e. they lead to the same results – when the Euclidean and social distances ($r$ and $d$, respectively) are related by $d \sim D_p \log_b(r)$ or

$$r \sim b^{\frac{d}{D_p}}\,. \tag{59}$$

This result indicates that according to the models, Euclidean and social distances should be correlated, which is plausible since we live close to people we know [38]. Moreover, if Eq. (59) holds, then the two approaches, in fact, represent the same urban system.

If we solve Eq. (59) for the fractal dimension $D_p$, one gets

$$D_p \sim \frac{d}{\log_b(r)}\,. \tag{60}$$

The network distance $d$ must be proportional to the logarithm of a "mass" since otherwise Eq. (60) would not comply with the definition of the fractal dimension [49]. Interestingly, in the social-network model the number of nodes and the distance are related via $N(d) = b^d$, Eq. (52), and consequently $d \sim \log_b(N)$; that is, $d$ is proportional to the logarithm of a "mass". Inserting in Eq. (60) yields $D_p \sim \log(N)/\log(r)$, which makes sense in terms of fractal geometry.

This means if the fractal dimension $D_p$ relates population and space – where population takes the role of mass and Euclidean distance the role of scale – then the gravity models in both versions are equivalent. This is remarkable since the various authors [12, 48, 36] developed their models independently employing different ideas and approaches. Under this condition, the population is spatially located in a fractal manner and follows a hierarchical social network. From Eq. (58) we conclude that the decay exponent $\gamma$ also relates to social ties, expressed by the parameter $\phi$, which gives one more insight about the $\gamma$ parameter and consequently how the interaction between people behave. The following subsection presents additional alternative interpretations for the $\gamma$ parameter.

### *2.4.5. F.Ribeiro et al. supply-demand model*

The infrastructure scaling exponent $\beta_{\text{sub}}$ can also be deduced using the gravity approach. However, different from the Bettencourt model (Sec. 2.2), which focuses on the area to explain the sub-linear urban scaling, Ribeiro et al. [12] focus on the number of amenities in the city necessary to satisfy the people needs. They use the idea that people tend to choose close places to buy products.

The model considers that the individual $i$ consumes $u_i$ quantities of a given *individual need* product (e.g. bread) per period of time and, consequently, the city, as a whole, consumes $U = \sum_{i=1}^{N} u_i$ units of this product during this period. It is assumed that the demand is always fully provided by the city. Suppose that the consumers can buy this product in $P$ amenities (e.g. bakeries, if the product is bread) distributed throughout the city, but these consumers choose, preferentially, the amenities that are close. In order to model this fact and using the same idea discussed around Eq. (34), the total supply (per time) by the $k$-th amenity for the person $i$ will be given by

$$f(r_{ik}) \propto \frac{1}{r_{ik}^{\gamma}}\,, \tag{61}$$

where $r_{ik}$ is the Euclidean distance between them. Equation (61) can also be understood as the number of products the individual $i$ bought in the $k$-th amenity during a period of time. As $\gamma > 0$, this person buys more products in closer amenities. The total demand of $i$ can then be computed by

$$u_i = \sum_{k=1}^{P} f(r_{ik})\,, \tag{62}$$

and consequently the total provision of the city is

$$U \equiv \sum_{i=1}^{N} u_i = \sum_{i=1}^{N}\sum_{k=1}^{P} f(r_{ik})\,, \tag{63}$$

which can also be written as

$$U = \sum_{k=1}^{P}\left(\sum_{i=1}^{N} f(r_{ik})\right). \tag{64}$$

If we consider that the population is homogeneously distributed in a fractal structure, as considered before (in Sec. 2.4.1), then the inner sum of Eq. (64) can be transformed into an integral that can be solved as before using Eqs. (35), (37), and (61)

$$\sum_{i=1}^{N} f(r_{ik}) \sim \int f(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \sim N^{1-\frac{\gamma}{D_p}}\,. \tag{65}$$

It reveals that this sum, on average, is the same for all amenities, because the right-hand side of this proportionality does not depend on the index $k$. Inserting this result in Eq. (64), the second sum transforms into a multiplication by the total number of amenities ($P$), leading to

$$U \sim P N^{1-\frac{\gamma}{D_p}}\,. \tag{66}$$

As we are dealing with individual need products, the empirical evidence suggests that the total consumption must *scale linearly* with the population size, that is $U \sim N$. Under this condition, Eq. (66) can be rewritten as

$$P \sim N^{\frac{\gamma}{D_p}}, \tag{67}$$

and therefore one gets the scaling exponent

$$\beta_{\text{sub}} = \frac{\gamma}{D_p}, \tag{68}$$

which governs the scaling properties of the number of amenities – an infrastructure variable – as a function of the population size. It is worth mentioning that Eq. (68) together with Eq. (39) fulfill the rule of thumb $\beta_{\text{sub}} + \beta_{\text{sub}} = 2$ (see around Eq. (8).

In conclusion, this model suggests that sub-linear exponent of the infrastructure occurs only if $\gamma < D_p$. This $\gamma$-range in fact characterizes the long-range interaction regime, which is a consequence of the city acting as an entire *coupled system*, as mentioned in Sec. 2.4.1 in the context of the gravity model and comparison to the Bettencourt model. In addition, the bigger the interaction range is, the bigger the economy of scales and consequently the lower the infrastructure costs.

### *2.4.6. Alternative interpretations of $\gamma$ parameter*

Previously [see Yakubo et al. (2.4.2) and Abersman et al. (2.4.3) models], we presented some possible interpretation for the gravity model parameter $\gamma$ [introduced in Eq. (34)]. However, are there other possible explanations, or better, is there a more fundamental way to explain the origin of this exponent? Following this idea, and given that the previous models suggest the dependence between the scaling and gravity model exponents, we present in this sub-section some physical interpretations for the numeric value of $\gamma$.

We start this discussion from the so-called *Rank model* [50], which considers that the probability of one person choosing another person to interact with depends on the number of people closer to $i$ than $j$ is to $i$. The main idea is that, as illustrated in Fig. (8-a), everybody that lies inside the circle of radius $r_{ij}$, that separates $i$ from $j$, is closer to $i$ than $j$ is to $i$. In this sense, and if the probability of interaction $p_{ij}$ between two individuals is a distant-depend function, then any person inside the circle of Fig (8-a) must have a probability of interaction with $i$ greater than $p_{ij}$. That is, $p_{ik} \geq p_{ij}$ if $r_{ik} \leq r_{ij}$ or, in a more general way, one can say that $p_{i1} \leq p_{i2} \leq \cdots \leq p_{i,j-1} \leq p_{ij}$, if $1, 2, \cdots, j-1$ are individuals inside the circle of Fig. (8-a). One way to capture this idea is to consider that $p_{ij}$ obeys a rank rule of the type

$$p_{ij} = \frac{1}{\text{Rank}_i(j)}\,, \tag{69}$$

where $\text{Rank}_i(j)$ is the rank position of $j$, in terms of distance, in relation to $i$. The function $\text{Rank}_i(j)$ is numerically identical to the number of individuals that lie inside the circle centered in $i$ and with radius $r_{ij}$; that is,

$$\text{Rank}_i(j) = N(r_{ij})\,. \tag{70}$$

This Rank idea found applications in migrations [51] – with the so-called *intervening opportunities model* –, urban mobility [52], and social-network friendship [50].

If we consider that the population is distributed spatially as a fractal object with dimension $D_p$, then the number of people inside a circle of radius $r$ will be given by $N(r) = N_0 r^{D_P}$, and consequently $p_{ij}$ can be written, from Eq. (69) and (70), as

$$p_{ij} \sim \frac{1}{r^{D_p}}\,. \tag{71}$$

From the rank population idea, in comparison with Eq. (34), it follows $\gamma = D_p$ and that $\gamma$ can be interpreted as the fractal dimension of the population distribution.

However, it may be more natural to consider that $i$ accesses $j$ not according to the closer individuals but instead according to places/infrastructure that are closer to $i$ than the place that $j$ is. In this case, the rank function $\text{Rank}_i(j)$ is indeed relative to the number of places/infrastructure inside the circle centered in $i$ and with radius $r_{ij}$, as illustrated in Fig. (8-b). In other words, $i$ has more chance (probability) to interact with people living in places closer than the place where $j$ lives. In this situation, $\text{Rank}_i(j)$ is then the distance rank position of the place/infrastructure where $j$ is in relation to place/infrastructure where $i$ is. That is, the function $\text{Rank}_i(j)$ is numerically identical to the number of places/infrastructure inside the circle centered in $i$ and with radius $r_{ij}$. If the infrastructure/place has a fractal shape whose number of units scales as $\sim r^{D_{\text{infra}}}$ then, analogous to the idea above, one gets

$$p_{ij} \sim \frac{1}{r^{D_{\text{infra}}}}\,. \tag{72}$$

The $\gamma$ parameter is now interpreted as the fractal dimension of the city's infrastructure. We will explore more about this interpretation when we present the Molinero & Thurner model [53] in sec (2.5).

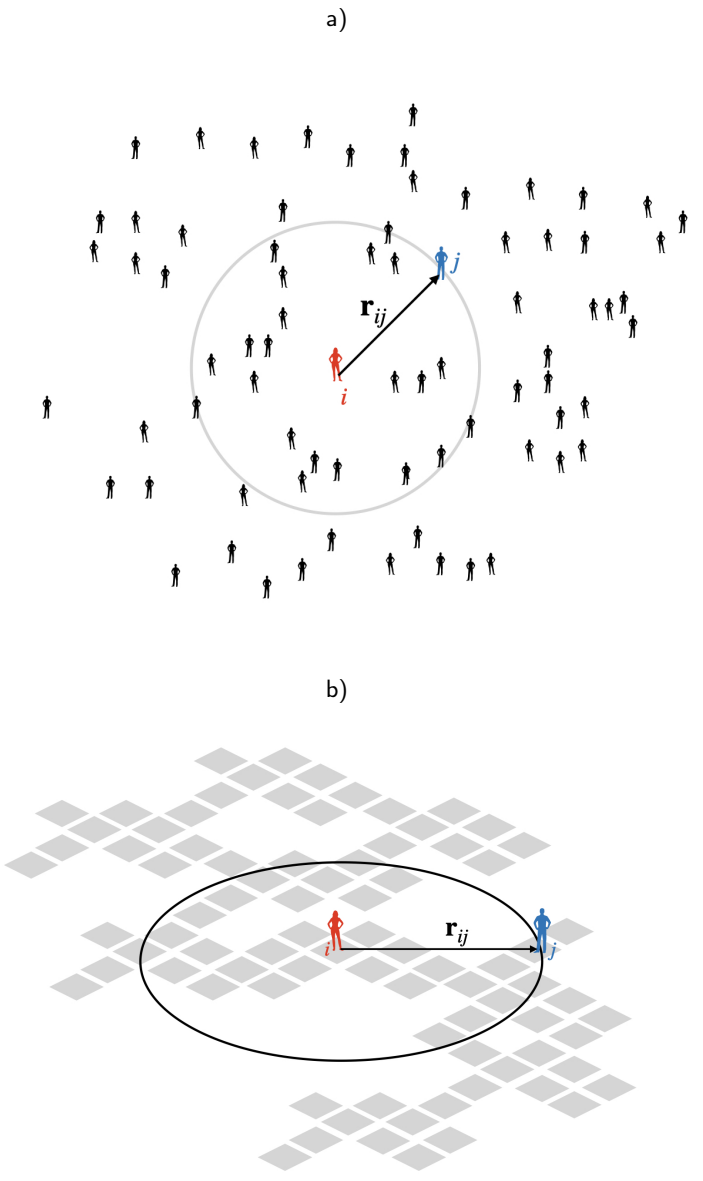


**Figure 8:** Illustration of the spatial distribution of population and infrastructure places. Panel (a) shows the spatial population distribution. All people inside the circle centred in $i$ and with radius $r_{ij}$ (distance between $i$ and $j$) are closer to $i$ than $j$ is to $i$. Panel (b) represents the spatial distribution of spaces/infrastructure. All places/infrastructure inside the circle are closer to $i$ than the place where $j$ is.

Alternatively to the Rank model, Simini et al. [54] proposed the *radiation model*. The name comes from the fact that this model is inspired by the radiation processes in physics. Suppose that the individual/location $i$ emits particles that can be absorbed by other individuals/locations, according to the following rule. Every particle emitted from $i$ has a random number which represents its absorption threshold. Any other individual/location, for instance $j$, also has a random number that represents its absorbance. The particle will be absorbed by the closest individual/location where the absorbance (receiver) is greater than the absorption threshold (sender). Some examples of radiation model applications include [55, 56, 57].

Analytical derivation of this model [54, supplementary material] shows that the probability of a particle emitted from $i$ to be absorbed by $j$ is given by

$$p_{ij} = \frac{1}{(1+s_{ij})(2+s_{ij})} = \frac{1}{s_{ij}^2 + 3s_{ij} + 2}\,, \tag{73}$$

where we consider that every place is occupied by just one person. Here $s_{ij}$ is the population or place/infrastructure inside a circle centered in $i$, and with radius $r_{ij}$; that is, $s_{ij}$ is the very rank function described above, that can be interpreted in two scenarios, in a similar way that was described above. In the first scenario, $s_{ij}$ represents the population that is closer to $i$ than $j$ is (Fig. (8-a); in the second scenario, $s_{ij}$ represents the places that are closer to the place where $i$ is than the place where $j$ is (Fig. (8-b).

In the first scenario we have $s_{ij} = N(r_{ij}) \sim r_{ij}^{D_p}$, then the radiation model [Eq. (73)] is analogous to a form of gravity model

$$p_{ij} \sim \frac{1}{r^{2D_p}}\,. \tag{74}$$

Comparing this result to with Eq. (34), leads to $\gamma = 2D_P$. Similar to the Rank model, the gravity exponent can be interpreted as a quantity related to the fractal dimension of the spatial population distribution $D_P$. In the particular example presented by Simini et al. [54], a two-dimensional uniformly distributed population, i.e. $D_p = 2$, yields $\gamma = 4$.

Analogously, for the second scenario we obtain $s_{ij} \sim r_{ij}^{D_{\text{infra}}}$ and the radiation model relates to the gravity model via

$$p_{ij} \sim \frac{1}{r^{2D_{\text{infra}}}}\,, \tag{75}$$

and $\gamma = 2D_{\text{infra}}$. In this scenario, the gravity exponent is related to the fractal dimension of the space/infrastructure of the cities $D_{\text{infra}}$.

Another work that can give insights about the interpretation and determination of the $\gamma$ exponent is the empirical study published by Dong et al. [58]. The authors analyze empirically urban scaling *within* a city ("mesoscale" in their terminology). Therefore, they divided a considered city into $n$ cells, for which they have geographic position and population data at cell resolution. The authors report that socio-economic activity scales super-linearly and infrastructure volume scales sub-linearly for various cities in China. Conceptually, they explore the set $\{N_k\}_k$, where $N_k$ is the population in the $k$-th cell, and the set $\{r_{k,l}\}_{k,l}$, where $r_{k,l}$ is the distance between any two cells $k$ and $l$. With this empirical data, they compute the quantity

$$q_{kl} = \text{cte} \cdot \frac{N_k N_l}{r_{kl}^{\gamma}}\,, \tag{76}$$

where $q_{kl}$ is interpreted as related to the number of interactions between the cells $k$ and $l$. From this quantity it is possible to numerically get the "total interaction" of the cell $k$ via the sum $Q_k = \sum_{l\neq k}^{n} q_{kl}$, i.e.

$$Q_k = \text{cte} \cdot N_k \left( \sum_{l\neq k}^{n} \frac{N_l}{r_{kl}^{\gamma}} \right). \tag{77}$$

Then, given a specific value of $\gamma$ and the empirical data ($\{N_k\}_k$ and $\{r_{kl}\}_l$) it is possible to compute numerically $Q_k$ by Eq. (77).

As a next step, and for a specific $\gamma$ value, the authors plot $Q_k$ as a function of $N_k$ for all the cells that compose a specific city, trying to fit the data using the power law hypothesis $Q_k \sim N_k^{\beta}$. They show that this power law fits the data very well for many Chinese cities, and then it is possible to get $\beta$ as a function of the gravity exponent $\gamma$, i.e. $\beta(\gamma)$. Numerically, they find a linear dependence between $\beta$ and $\gamma$ when $\gamma \approx 1$, and saturation for $\beta$ when $\gamma$ is sufficiently large. More precisely, for $\gamma \sim 1$ they report

$$\beta(\gamma) \approx a + b(\gamma - 1)\,, \tag{78}$$

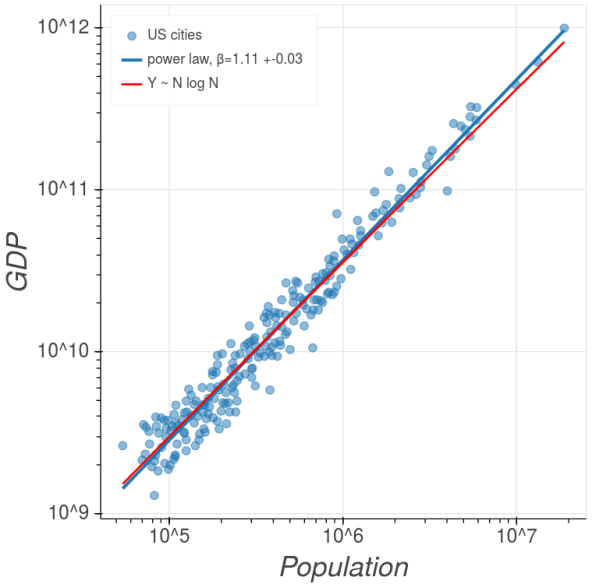


**Figure 9:** Comparison of $Y \sim N \ln N$ and the power law $Y \sim N^{\beta}$ to fit USA cities' data (GDP as a function of the population). Note that there is no apparent difference between these two regressions to describe the data.

where $a = 1.153$ and $b = 0.186$ [58]. Note that when $\gamma = 1$, $\beta$ approaches to the empirical value of the scaling exponent ($\gamma = 1 \implies \beta = 1.153$).

This numeric approach provides a link between the gravity exponent $\gamma$ the scaling exponent $\beta$. In addition, it emphasizes the importance of tree effects on urban scaling: (i) The spatial structure of the city (expressed by the distance between cells), (ii) the population size, and (iii) how these two mechanisms convert into interaction.

To summarize, we present here some alternative explanations and interpretations for the gravity model decay exponent $\gamma$. In the context of the Yakubo et al. model, this exponent is related to the degree of influence of the people (parameter $\alpha$), while, in the context of the Arbesman et al. model, it is related to the social ties in the social network. However, in the context discussed in this subsection, we present some insights suggesting that $\gamma$ could also be related to the fractal configuration of the population distribution in space or of the spatial infrastructure distribution, both being characterized by their respective fractal dimensions.

### *2.4.7. Pan et al. model*

Pan et al. [59] proposed an alternative to the power Eq. (1) to describe the relationship between socio-economic variables and city population size. They suggest that a function of the form $Y \sim N \log N$ could describe the empirical data. Figure (9) presents data from the US and a comparative plot using the power-law equation and the alternative equation proposed. There is no apparent difference between the fit of these two equations.

The authors justify this alternative equation using the *rank model* (see Sec. 2.4.6) to define the probability of interaction between the people. They also considered, for simplicity, that the population is uniformly distributed in a two-dimensional space in such a way that the density $\rho = N/A$ is always constant. In this way, the number of individuals inside the circle of radius $r_{ij}$, that is the rank function, will be simply

$$Rank_i(j) \propto r_{ij}^2 \,. \tag{79}$$

Using the idea that the probability of interaction $p_{ij}$ between two persons, $i$ and $j$, is given by the inverse of the rank function (see details in Sec. 2.4.6), one has $p_{ij} \sim 1/r_{ij}^2$, or simply

$$p(r) \sim \frac{1}{r^2} \,. \tag{80}$$

This consideration, i.e. the rank rule and uniformly distributed population, yields the gravity model with $\gamma = 2$.

Proceeding with the general Eq. (33) using (i) $g = 1$ (same weight for all interactions), (ii) $\rho = \text{cte}$, (iii) transformation from Cartesian to polar coordinates $d\mathbf{r} \sim r^2 dr$ (in two-dimensional space), and (iv) the probability of interaction given by Eq. (80), yields

$$Y \sim N \ln N \,. \tag{81}$$

This function is steeper than linear scaling once the number of contacts per individual ($y = Y/N = \ln N$) grows with $N$ (logarithmically). This result challenges the power law equation as a candidate to describe increasing returns. In fact, in [60] other functions are presented that can also describe the super-linear urban scaling as well as the power law equation (e.g. logistic scaling relationship). It is worth mentioning that Eq. (81) does not have any scaling parameter. The only parameter is the pre-factor controlling the $Y$ in absolute terms, i.e. vertical adjustment in Fig. (9).

#### *2.4.8. Findings and conclusions from gravity models*

At least three groups of authors independently employ gravity ideas to explain urban scaling – here we show that they lead to consistent results. This observation emphasizes the importance of the gravity approach to understanding urban phenomena, as it was already suggested qualitatively by Tobler with his first law of geography[3] [61].

A novelty of the present work is the organization of these ideas in a single and general framework that permits the identification of similarities and common results. The various gravity models are equivalent in special cases, outside this overlap they represent variants. However, from consolidating the models we conclude that *urban scaling is essentially a consequence of spatial distribution imposed by the geometry and the social ties that enhance or reduce interactions*. Specifically, we make the following interpretations.

- *Good access to all parts of the city.* Increasing returns to scale require a geographically well-connected city, allowing interactions within the entire city and permitting integrity of the city. In practice, this can be achieved by an efficient transport system.
- *Influencers reaching distant parts of the city.* The presence of outstanding influential people can integrate distant parts of the city and promote interconnectivity, resulting in a more pronounced urban scaling.
- *Interaction between socially distant people.* The socio-economic scaling exponent is larger when socially distant people can interact better. We have demonstrated that under certain circumstances social and geographic distance are related.

### 2.5. Molinero & Thurner model – infrastructure geometry

This section presents the model proposed by Molinero and Thurner (M&T) [53] which, as some of the models presented in the previous section, also employs geometrical considerations as the main factor responsible for urban scaling. The authors introduce new ingredients to the discussion, like the city verticalization and the distinction between the fractal dimension of the population ($D_p$) and the fractal dimension of the infrastructure ($D_{\text{infra}}$), represented by the street network. M&T argue that the distinction between population and infrastructure fractal dimensions is essential to the scaling laws observed across cities.

As the street network rests on the two-dimensional earth surface, its dimension is constrained by $D_{\text{infra}} \leq 2$ [62]. The population is located in houses and buildings, which are situated along the streets. The authors argue, that if we neglect the vertical extent, then the fractal dimension of the population would be very similar to the fractal dimension of the street structure. However, as they argue, the cities have a vertical component that cannot be disregarded, which constrains $D_p$ to the interval $D_{\text{infra}} \leq D_p \leq D_{\text{infra}} + D_h$, where $D_h$ is the dimension associated with the city building height. If one considers that people fulfil all three-dimensional embedded components of the infrastructure, then

$$D_p = D_{\text{infra}} + D_h \,. \tag{82}$$

The population fractal dimension $D_p$ is defined by the power-law relation between the population and a linear metric $r$, that is

$$N \sim r^{D_p} \,. \tag{83}$$

[3] "Everything is related to everything else, but near things are more related than distant things."

In the same way, the street network fractal dimension $D_{\text{infra}}$ is defined by the power-law relation between the *street network total length* $L_{\text{tot}}$ and a linear metric

$$L_{\text{tot}} \sim r^{D_{\text{infra}}} . \tag{84}$$

In addition, as proposed by the authors, the number of individuals can be written using these two dimensions

$$N \sim C_c \cdot r^{D_p} \tag{85}$$

and

$$N \sim C_{\text{infra}} \cdot r^{D_{\text{infra}}} . \tag{86}$$

In Eq. (85), $C_c$ can be understood as the number of people living in a *cube* of size 1 (in any units). This means $r^{D_p}$ is the number of non-empty cubes of size 1 in the city. Such a cube can be, for instance, a house, an apartment, or a floor. Theoretically, $C_c$ must be scale-invariant, that is, $C_c \sim N^0$, because of the physical limit to accommodate a maximum number of people in a house/apartment/floor. The authors verified that $C_c$ increases with the city size for cities smaller than 100.000 inhabitants, but it indeed saturates and stabilises for cities larger than that. Similarly, the other quantity, $C_{\text{infra}}$ in Eq. (86), can be understood as the number of people living in a *square* of size 1. This means $r^{D_{\text{infra}}}$ is related to the number of non-empty squares of size 1 (in any units). Indeed $C_{\text{infra}}$ represents the projection of the three-dimension space population into a two-dimensional plane (vertical projection). In contrast to $C_c$ that is constant for a sufficiently large population, $C_{\text{infra}}$ grows with the population size obeying a power-law relation $C_{\text{infra}} \sim N^{0.09}$ in the UK and the authors observe similar results in other countries. It reveals an absence of a typical size value for $C_{\text{infra}}$.

The relation between $C_{\text{infra}}$ and $C_c$ can be obtained equalling Eqs. (85) and (86) to get

$$C_{\text{infra}} = C_c r^{D_p - D_{\text{infra}}}. \tag{87}$$

Using Eq. (83) in this relation and considering that $C_c \sim N^0$, leads to

$$C_{\text{infra}} \sim N^{1-\frac{D_{\text{infra}}}{D_P}} . \tag{88}$$

The saturation of $C_c$ for sufficiently large cities and the power-law relation between $C_{\text{infra}}$ and $N$ implies that the densification of the "cube" of size 1 happens until the city population reaches a limit (around $N = 100.0000$ inhabitants). For cities larger than this, the number of people in this cube is stabilized, but to increase the number of people per square meter (that is, to continue increasing $C_{\text{infra}}$ with $N$), the city starts to grow vertically.

Defining such quantities, one can derive the urban scaling exponents. Using Eqs. (84) and (85) one can show that $L_{\text{tot}} \sim N^{\frac{D_{\text{infra}}}{D_p}}$, i.e.

$$\beta_{\text{sub}} = \frac{D_{\text{infra}}}{D_p} . \tag{89}$$

This result implies that the urban scaling exponent is the result of the relationship between the two fractal structures, namely the population and infrastructure (street) network. In addition, the emergence of non-linearity ($\beta \neq 1$) happens because of the difference of the fractal dimensions of these structures, and the sub-linearity ($\beta < 1$) is due to $D_{\text{infra}} \leq D_p$.

The authors also consider that the socio-economic variable must be dependent on the number of interactions in the city. To estimate this number they consider, by hypothesis, that the number of interactions inside a square of size 1 will be proportional to the maximum number of interactions $C_{\text{infra}}(C_{\text{infra}} - 1)/2 \sim C_{\text{infra}}^2$. With this consideration, the total number of interactions in the city ($N_{\text{int}}$) is given by the number of interactions inside a square of size 1 multiplied by the number of squares of this size, that is

$$N_{\text{int}} \sim C_{\text{infra}}^2 r^{D_{\text{infra}}} . \tag{90}$$

Using Eqs. (87) and (85) one can show that $N_{\text{int}} \sim N^{2-\frac{D_{\text{infra}}}{D_p}}$, and therefore, considering that $Y \sim N_{\text{int}}$,

$$\beta_{\text{super}} = 2 - \frac{D_{\text{infra}}}{D_p} \,. \tag{91}$$

It shows, which role fractal structures (population and street networks) play in urban scaling. Urban scaling emerges as a result of an imbalance between where people live and the structure on which they move. It is also important to stress that, as demonstrated by the authors, while the population and infrastructure fractal dimensions ($D_p$ and $D_{\text{infra}}$) vary largely for the individual cities, the empirical ratio $D_{\text{infra}}/D_p$ is remarkably robust (for thousands of cities) and around $D_{\text{infra}}/D_p \approx 0.86$ [53, Fig. 2]. The closeness of this numeric value to the empirical value of the infrastructure scaling exponent ($\beta_{\text{sub}} \approx 0.85$) is remarkable and puts this theory (conform Eq. (89)) as one of the most successful in explaining urban scaling in the context of infrastructure.

***Conection with the gravity models***

Apparently Eqs. (39), $\beta_{\text{super}} = 2 - \frac{\gamma}{D_p}$, and Eq. (91), $\beta_{\text{super}} = 2 - \frac{D_{\text{infra}}}{D_p}$, have a very similar form and only differ in the numerator, implying $\gamma = D_{\text{infra}}$ [63]. Indeed, comparing these two independent results, together with what was discussed in sec. (2.4.6), provides an additional interpretation of the gravity exponent $\gamma$ and suggests a more fundamental explanation for the M&T result.

In the gravity model context – Section (2.4) – the super-linear scaling exponent in Eq. (91) emerges when the probability of interaction between two individuals separated by the Euclidean distance $r$ is given by

$$p_{\text{int}}(r) \sim \frac{1}{r^{D_{\text{infra}}}} \,, \tag{92}$$

where we have replaced $\gamma$ with $D_{\text{infra}}$ in Eq. (34). This result indicates how space – or the street network structure, here represented by its fractal dimension $D_{\text{infra}}$, – affects the connection between the people. If Eq. (92) holds, then the more compact the street network is (larger $D_{\text{infra}}$), the smaller the interaction range. In other words, with increasing density fewer parts of the city can be accessed by individuals. Remains to interpret what $r^{D_{\text{infra}}}$ means or what quantity it represents. There are at least two candidate quantities that scale as $r^{D_{\text{infra}}}$, and that could be responsible for this impedance on the individuals' interaction. First, the *total length* inside a circle of radius $r$, which according to the definition Eq. (84) scales as $\sim r^{D_{\text{infra}}}$. Second, the *number of sites/places/houses* inside a circle of radius $r$; as the houses are coupled to the streets, this number must also scale as $r^{D_{\text{infra}}}$). Indeed, conform presented in the sec. (2.4.6), the rank and radiation model can explain how $\gamma$ and the infrastructure fractal dimension are related one each other (see around eq. (72) and (74)). In this sense, Eq. (92) suggests a quantitative way to understand how the structure of the streets affects the interaction of the people and, consequently, how it reverberates on urban scaling.

## 3. Bettencourt infrastructure network model

In the second part of his 2013 paper [11], Bettencourt proposes a more detailed model to explain the scaling laws of urban infrastructure. The model was inspired by West, Brown, and Enquist (WBE)'s theory [64, 65, 66, 67], which explains the allometric scaling in biology based on the hierarchical network properties of the distribution and allocation of resources.

Bettencourt proposes that the urban network infrastructure system, for instance, the road network of a city, is composed of $h$ hierarchical levels. In addition, every unit of specific hierarchical level branches into another $b$ units of the following level. That is if $N_k$ is the number of infrastructure units of the hierarchical level $k$, then

$$N_{k+1} = bN_k \,. \tag{93}$$

From this, one can show that

$$N_k = N_0 b^k \,, \tag{94}$$

where $N_0$ is the number of units in the lowest level ($k = 0$). In the case of the road network system, $k = 0$ (the lowest hierarchical level) represents the highways, $k = 1$ represents roads, and so on until $k = h$ (highest hierarchical level),

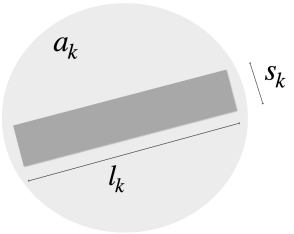


**Figure 10:** The geometry of a *hyper-road* (a highway/road/path at the $k$-th level), with length $l_k$ and width $s_k$, while $a_k$ is the area that this hyper-road serves, as used in the Bettencourt infrastructure network model.

which represents, for instance, the smallest paths. The number of units in a hierarchical level will always be larger than the number of units in the previous hierarchical level, then $b > 1$. In the case of a road network, $N_0 = 1$ (number of highways) and the number of small paths is equal (or of the order) to the population size $N_h = N$ – both by definition. These considerations can be expressed by these inequalities

$$\underbrace{N_0}_{\text{Numb. of Highways}} < \underbrace{N_1}_{\text{Numb. of Roads}} < \cdots < \underbrace{N_h}_{\text{Numb. of paths}} \equiv \underbrace{N}_{\text{Population size}} . \tag{95}$$

In Fig. 10 the geometry of a *hyper-road* (a highway/road/path at the $k$-th level) is shown, which has length $l_k$ and width $s_k$, while $a_k$ is the area that this hyper-road serves. It must be valid that $s_k > s_{k+1}$, that is

$$\underbrace{s_0}_{\text{Highways width}} > \underbrace{s_1}_{\text{Roads width}} > \cdots > \underbrace{s_h}_{\text{paths width}} \equiv s_* , \tag{96}$$

where $s_*$ is the width of the smallest network unit, considered to be scale-invariant ($s_* \sim N^0$) – inspired by the WBE theory. Analogous to Eq. (93), the author assumes (by hypotheses) that

$$s_k = b^\alpha s_{k+1} , \tag{97}$$

where $\alpha$ is a parameter representing the shrinkage/expansion of the hyper-roads width from one level to the next. If $\alpha = 1$, the increase of hyper- road's width from one level to the next happens in the same quantitative way as the decrease in the number of units of hyper-roads from one level to the next. Note that, in order to obey that $b > 1$ and $s_{k+1} < s_k$, it is required that $\alpha > 0$. From these definitions, one can show that $s_0 = b^\alpha s_1 = b^{2\alpha} s_2 = \cdots$, and $s_0 = s_* b^{\alpha h}$, then

$$s_k = s_* b^{\alpha(h-k)} . \tag{98}$$

Let us try now to infer the area $a_k$ that a hyper-road at level $k$ serves. For that, Bettencourt assumes that the fundamental allometry $A(N) = aN^{\beta_{\text{FA}}}$, where $a$ is a constant, must be valid for all hierarchical levels in such a way that

$$A_k = aN_k^{\beta_{\text{FA}}} , \tag{99}$$

where $A_k$ is the total area of all the hyper-roads of the level $k$. By definition, $A_h = A$ (area of the city). In addition, $a_k$ (area served by a hyper-road at level $k$), which can also be understood as the area of the infrastructure level per unit of infrastructure, can be written as

$$a_k = \frac{A_k}{N_k} . \tag{100}$$

Consequently, from Eqs (94) and (99), one has

$$a_k = ab^{k(\beta_{\text{FA}}-1)} , \tag{101}$$

when $N_0 = 1$. For instance, $a_h$ (for $k = h$) can be understood as the area occupied by a single person ($a_h = A_h/N_h = A/N$).

Now, one needs to determine the hyper-road length $l_k$, and for this Bettencourt suggests that

$$l_k = l a_k \tag{102}$$

where $l$ is a constant (scale-independent) defined as

$$l_h = \frac{a_h}{l} = \frac{1}{l}\frac{A}{N} = \frac{a}{l} N^{\beta_{\text{FA}}-1} . \tag{103}$$

To summarize what we have so far are,

$$\begin{cases} N_k = b^k \\ a_k = a b^{k(\beta_{\text{FA}}-1)} \\ l_k = l a_k \\ s_k = s_* b^{\alpha(h-k)} . \end{cases}$$

With these quantities, it is possible to compute the *total length of the road network* by

$$L_{tot} = \sum_{k=0}^{h} l_k N_k ; \tag{104}$$

and the *total area of the road network* by

$$A_n = \sum_{k=0}^{h} (s_k l_k) N_k . \tag{105}$$

Using P.G. sum, one can show that $L_{tot} \sim N^{\beta_{\text{FA}}}$, i.e. the total road length scales with the area of the city (i.e. according to the fundamental allometry). It is also possible to show that the area of the total roads is $A_n \sim N^{\alpha}$, where $\alpha$ is the exponent defined in Eq. (97). Accordingly, the model explains the scaling of the area of the road network from the way the width of the hyper-roads changes from one hierarchical level to the next.

## 4. Louf & Barthelemy Model

Louf & Barthelemy (L&B)[68, 69] present a model that aims to show how traffic congestion relates to the number of city's activity centers (and vice versa) and how it affects urban scaling, especially the fundamental allometry. The model explains multi-centres' emergence and the condition for mono-centric cities to exist or be maintained.

The starting point is the classical model by Fujita and Ogawa in spatial economics [70, 34], which states that the individuals that live in place $i$ choose to work at a sub-center $j$ which maximizes some wage and location function. L&B use the function

$$Z_{ij} = W_j - C_{ij} \tag{106}$$

where $W_j$ is the average wage paid by businesses located at the sub-center $j$, and $C_{ij}$ is the commuting cost between $i$ and $j$.

The authors propose that $W_j$ is a random variable, given by $W_j = s\eta_j$, where $s$ is a parameter which defines the maximum wage and $\eta_j \in [0, 1]$ is a uniform random variable. Concerning the mobility cost, they assume that $C_{ij}$ depends on (i) the distance $r_{ij}$ between $i$ and $j$ and (ii) the traffic between these two places. More specifically, they consider

$$C_{ij} = \tau r_{ij} \left[ 1 + \left( \frac{T_{ij}}{c} \right)^{\mu} \right] \tag{107}$$

where $\tau$ is the transport cost, $T_{ij}$ is the traffic between $i$ and $j$, $c$ (constant) is the typical capacity of a road, and $\mu$ is a parameter representing the resilience of the transport network to congestion. The smaller the parameter $\mu$ is, the better the transport network system is.

To treat the model analytically, they propose some simplification to Eq. (107). Firstly, they consider that there are no correlations between the workplace spatial distribution, so that

$$r_{ij} \sim L \sim \sqrt{A}\,, \tag{108}$$

where $L$ and $A$ are the diagonal extent and area of the city, respectively. Secondly, they only take the incoming traffic to the workplace $j$ into account, which makes another simplification possible

$$T_{ij} \to T_j\,. \tag{109}$$

Then Eq. (106) becomes

$$Z_{ij} = s\eta_j - \text{cte}\cdot\tau\sqrt{A}\left[1+\left(\frac{T_j}{c}\right)^{\mu}\right]. \tag{110}$$

As we are interested in $j$ that maximize $Z_{ij}$, we can rewrite $Z_{ij}/s \to Z_{ij}$ without loss of generality. It allows us to write

$$Z_{ij} = \eta_j - \text{cte}\cdot\frac{\sqrt{A}}{l}\left[1+\left(\frac{T_j}{c}\right)^{\mu}\right], \tag{111}$$

where $l \equiv s/\tau$. It is interesting to note that the unit of $s$ (maximum wage) is *money*, and the unit of $\tau$ (cost to travel) is *money per distance*. Then, $l$, which has the unit of distance, can be interpreted as the maximum commuting distance people can travel. Note also that if $l$ is too small, the system will always be in a poly-centric regime and never in a mono-centric one. Here we consider only the situation that $l$ is sufficiently large, which guarantees the city's integrity, similarly to the previous models' considerations, as the Bettencourt (sec. (2.2)) and Ribeiro et al (sec. (2.4.1)) models.

When the population increases, the traffic also increases, allowing new effective sub-centers to arise. For instance, the second effective sub-center appears when, for a new individual $i$, we have

$$Z_{i2} > Z_{i1}\,, \tag{112}$$

where $Z_{i1}$ is given by Eq. (111) in relation to the most attractive sub-center and $Z_{i2}$ to the second most attractive sub-center.

At this stage (mono-centric regime), we have $T_1 = N$ (all previous individuals work at the sub-center 1) and $T_j = 0$ for all the other $N_c - 1$ (potential) sub-centers. However, when Eq. (112) is valid, then

$$\underbrace{\eta_2 - \text{cte}\frac{\sqrt{A}}{l}}_{Z_{i2}} > \underbrace{\eta_1 - \text{cte}\frac{\sqrt{A}}{l}\left[1+\left(\frac{N}{c}\right)^{\mu}\right]}_{Z_{i1}}. \tag{113}$$

It implies that the critical population size $N^*$ giving rise to the second sub-center is determined by

$$\eta_2 - \eta_1 = -\text{cte}\frac{\sqrt{A}}{l}\left(\frac{N^*}{c}\right)^{\mu}. \tag{114}$$

Given that the set $\{\eta_j\}_j$ is generated from a uniform distribution, it is valid that $\overline{\eta_1 - \eta_1} \sim \frac{1}{N_c}$, where $N_c$ is the number of potential sub-centers, and consequently

$$N^* \sim c\left(\frac{l}{\sqrt{A}N_c}\right)^{\mu}. \tag{115}$$

This population size represents the transition from a mono-centric to a poly-centric (two effective sub-centers) organization. Note that small $\mu$ or large $c$ (representing cities with good transportation infrastructure) can absorb large traffic and maintain a mono-centric regime for a larger population.

A natural question which comes out is: how many effective sub-centers are needed for a city of a given population size? Or putting this question in another way, can the model predict the population $N$ for which the $P$-th effective sub-center appears? To answer those questions, suppose that $P-1$ effective sub-centers (from a total of $N_c$ potential sub-centers) have emerged in $t-1$ time steps, and then it is valid that

$$\eta_1 \geq \eta_2 \geq \cdots \geq \eta_{P-1} \,. \tag{116}$$

In the next time step ($t$) the new worker $i$ will choose a new emergent effective sub-center $P$, with random variable $\eta_P$, if

$$Z_{iP} > \max_{j\in[1,P-1]} \{Z_{ij}\} \,. \tag{117}$$

Assuming that all sub-centers have approximately the same number of commuters, then $T_j \sim N/(P-1)$, and that the new sub-center has minimal traffic in comparison to the other sub-centers, we get

$$\eta_P - \eta_1 > -\text{cte}\frac{\sqrt{A}}{l}\frac{1}{c^\mu}\frac{N^\mu}{(P-1)^\mu} \,. \tag{118}$$

This is the condition for the $P$-th sub-center to emerge, given a city with $N$ individuals.

Given a uniform random distribution of $\{\eta_j\}$, it is possible to prove that

$$\overline{(\eta_1 - \eta_P)} \sim \frac{P-1}{N_c+1} \,. \tag{119}$$

Consequently, the critical condition for the emergence of the $P$-th sub-center (from the above equations) is

$$P \sim A^{\frac{1}{2(1+\mu)}} N^{\frac{\mu}{1+\mu}} = f(A,N) \,. \tag{120}$$

This result suggests that the total number of effective sub-centers is an interplay between area $A$ and population $N$. One can also draw parallels between Eq. (120) and the Cobb-Douglas production function (see Sec. 4.2). This means area and population can substitute each other, e.g. a lack of area can be compensated by population to achieve the same number of sub-centers. Moreover, the city's resilience to absorb transport congestion (represented by $\mu$) determines the scaling exponents. In other words, according to this model, the capacity of the city to facilitate the flux of people is the main responsible for the scaling economy.

The authors also argue that the model can predict fundamental allometry. In fact, they suggest that the total area $A$ of the city is directly related to the number of activity centers. This means, if $A_1$ is the typical attraction area of one sub-center, then

$$A \sim PA_1 \,. \tag{121}$$

In addition, $A_1$ is related to the distance that the individuals can travel daily, namely $l_c$, and consequently one can say that

$$l_c = \frac{L_{\text{travel}}}{N} \sim \sqrt{A_1} \tag{122}$$

where $L_{\text{travel}}$ is the total daily distance travel of the population. Empirical findings [69] show that $L_{\text{travel}}$ scales linearly with the population size, and consequently $l_c$ and $A_1$ are scale-independent [4]. It implies that $P \sim A$, and inserting such result in Eq. (120) it is possible to write $A \sim A^{\frac{1}{2(1+\mu)}} N^{\frac{\mu}{1+\mu}}$, which yields $A \sim N^{\frac{2\mu}{2\mu+1}}$, and then

$$\beta_{\text{FA}} = \frac{2\mu}{2\mu+1} \,. \tag{123}$$

[4] For instance, the authors have found that $l_c \approx 23$ miles.

That is, the model predicts a sub-linear behaviour between area and population size, and the scaling (fundamental allometry) is explained as a consequence of the city's resilience to absorb traffic congestion (given by $\mu$). Note that a large $\mu$ represents a city with a bad congestion resilience system, yields $A \sim N$, i.e. the linear regime. The scaling economy is more pronounced in cities that operate a better transport system, which is consistent with the other models treated in this paper.

Based on this result, the authors also show how other urban metrics scale with $N$. For instance, they show (see details in [69]) that the total street length scales as $N^{\frac{\mu}{2\mu+1}+\frac{1}{2}}$ (sub-linear regime), and then

$$\beta_{\text{sub}} = \frac{\mu}{2\mu+1} + \frac{1}{2}\,, \tag{124}$$

and also that the total carbon emissions scale as $N^{\frac{\mu}{2\mu+1}+1}$ (super-linear regime, as a consequence of the congestion), and then

$$\beta_{\text{super}} = \frac{\mu}{2\mu+1} + 1\,. \tag{125}$$

As in Sec. (2.4.4), where we discuss the connection between Euclidean and social distance, or in Sec. (2.5), where we discuss the connection between Molinero & Thurner model with the gravity models, next we discuss similarities of the Louf & Barthelemy expressions with those of another model.

## 4.1. Connection between the Louf & Barthelemy model and the Bettencourt models

If we compare the Louf & Barthelemy expressions with those of the Bettencourt 2013 models (Sec. (2.2) and (3), and Tab. (1)), then it becomes apparent, that the expressions for $\beta_{FA}$, i.e. Eqs. (14) and (123), are identical if

$$2\mu = D_f \tag{126}$$

holds true. If $2\mu = D_f$ the expressions are also identical for $\beta_{\text{sub}}$, i.e. Eqs. (18) and (124). However, for $\beta_{\text{super}}$ this equivalence does not work. While in the Bettencourt model $\beta_{\text{super}} = 2 - \beta_{\text{sub}}$, in the Louf & Barthelemy model it is $\beta_{\text{super}} = \beta_{\text{sub}} + 1/2$. The disagreement can be attributed to the different nature of what the $\beta_{\text{super}}$ exponent describes in both models. In the case of the Bettencourt model, it is socio-economic efficiency (e.g. GDP) with increasing city size, while in the case of the Louf & Barthelemy model, it is increasing carbon emissions with city size, i.e. an inefficiency. Thus, the two $\beta_{\text{super}}$ are not really comparable.

Nevertheless, it is remarkable that for two out of three predictions both models agree – given Eq. (126) – despite their derivations stemming from very different arguments. Relating these two models we find a (not obvious) relation between the fractal dimension $D_f$, which characterizes the shape and structure of the city, and $\mu$, which represents the city's transport network efficiency. The comparison of these two models, thus, indicates a relation between these seemingly unrelated quantities.

In front of this background, the Louf & Barthelemy model represents an interesting link between morphology and mobility. It is consistent with other models, including the gravity approaches and the Bettencourt models, which suggests that urban scaling is more pronounced the better the city is connected as a whole, i.e. when each individual has potential access to any other. In other words, the facilitation of mobility and access is identified as an essential ingredient to promoting a scaling economy.

## 4.2. Cobb-Douglas form generalizing urban scaling

The derivation by Louf & Barthelemy includes an interesting interim result. Equation (120) involves *population size* and *area size*. Ribeiro et al. [71] investigate this extension of urban scaling. Specifically, instead of $Y \sim N^{\beta}$, they propose

$$Y \sim N^{\beta_N} A^{\beta_A}\,, \tag{127}$$

where $N$ is the population size, $A$ is the area size; $\beta_N, \beta_A$ are exponents and in general different from $\beta$ [in [71] $Y$ are urban $CO_2$ emissions]. Equation (127) represents a form of Cobb-Douglas production function [72, 73], where

the population and area are in analogy to labour and capital, respectively. An interesting property of Eq. (127) is substitution, i.e. moving along so-called isoquants of constant $Y$, population and area can substitute each other. Equation (127) implicitly also involves the density. Cities of high population and low area and cities of low population and high area can exhibit the same $Y$ but have completely different densities.

Including the relation between population and area, Eq. (1) and Eq. (2) in Eq. (127), leads to

$$\beta = \beta_N + \beta_A \beta_{\mathrm{FA}} \tag{128}$$

and equivalently $\beta_N = \beta - \beta_A \beta_{\mathrm{FA}}$ and $\beta_A = \frac{1}{\beta_{\mathrm{FA}}}(\beta - \beta_N)$. An additional equation is necessary to separate $\beta_N$ and $\beta_A$. E.g. with *constant returns to scale* ($\beta_N + \beta_A = 1$) we obtain

$$\beta_N = \frac{\beta_{\mathrm{FA}} - \beta}{\beta_{\mathrm{FA}} - 1} \quad \text{and} \quad \beta_A = \frac{\beta - 1}{\beta_{\mathrm{FA}} - 1} \tag{129}$$

for $\beta_{\mathrm{FA}} \neq 1$ [71]. Accordingly, without the constraint of constant returns to scale, Eq. (127) represents a generalization of $Y \sim N^{\beta}$ even under the constraint of the fundamental allometry Eq. (2) [71]. For $\beta_{\mathrm{FA}} = 1$ the original urban scaling is recovered.

Following econometrics, one can extend Eq. (127) by including more independent variables $X_i$, i.e. $Y \sim f(X_1, X_2, X_3, \dots)$. This might lead to a better fitting when empirical data is analyzed e.g. by means of multi-linear regression. However, the urban scaling paradigm relates city characteristics to size – while population and area both represent measures of size, other independent variables might not.

## 5. Gomez-Lievano et al. – model of required factors

This section presents the Gomez-Lievano et al. 2016 model of required factors [74], which belongs to the intra-city model category, but differs from the other models presented so far. While other intra-city models are based on *human interaction* premises, this one is rather based on required factors within the city. Consequently, the general framework introduced in Sec. (2.1) does not apply here. The model employs concepts of economic complexity and cultural evolution to explain the origin of urban scaling. The main idea of the model is that an urban socio-economic phenomenon occurs when a number of necessary complementary factors are available in the city.

Consider $M$ as the number of possible factors required for a particular socio-economic activity. For instance, if $Y$ is the total number of patents in a city, $M$ can be the number of different skills and capabilities needed by an individual to develop a patent. That is, $M$ is a measure of the "sophistication" (complexity, difficulty, etc.) of the phenomenon in question [75]. Suppose that each one of these factors is provided by the respective city with probability $z$. Then, if these factors are independent, the probability that the city provides $m$ of these $M$ factors, say $p(m)$, will follow the binomial distribution

$$p(m) = \frac{M!}{m!(M-m)!} z^m (1-z)^{M-m} . \tag{130}$$

According to the authors, the parameter $z$ can be interpreted as a measure of urban diversity.

One particular individual $i$ will only succeed to implement/develop this socio-economic activity if she/he only needs factors that the city can provide. The authors introduce the binary random variable $w_i$ that is $w_i = 1$ when the individual succeeds or $w_i = 0$ when the individual fails to implement this socio-economic activity. An analogous proposition, but without scaling analysis, was studied in [76].

Let us call $p(w_i = 1|m)$ the probability of success given that the city provides $m$ factors. This probability is identical to the probability that this individual *does not need* the $(M - m)$ factors that the city *does not provide*. If we denote $q$ as the probability that $i$ needs any given factor – i.e. it is a measure of the ability of the individuals [75] – then we can write

$$p(w_i = 1|m) = (1-q)^{M-m} . \tag{131}$$

The total socio-economic activity in the city will be given by the expected value of the aggregate output $Y = \sum_{i=1}^{N} w_i$, that is

$$\langle Y \rangle = N \sum_{m=0}^{M} p(w_i = 1, m)\,, \tag{132}$$

where $p(w_i = 1, m)$ is the joint probability also given by

$$p(w_i = 1, m) = p(w_i = 1|m)p(m)\,. \tag{133}$$

Then, using the probability distributions Eqs. (130) and (131), the expected value of the totality of this socio-economic activity (from Eq. (132) and using binomial properties) will be

$$\langle Y \rangle = N\Big(1 - q(1 - z)\Big)^{M}\,. \tag{134}$$

The term $q(1 - z)$ represents the probability that a factor is neither possessed by the individual nor by the city, which is likely to be very low. This can happen for sufficiently small $q$, e.g. if the number of skills per individual is large; but also for $z \to 1$, e.g. cities tend to be very diverse places, in the sense that there are very few factors that cannot be hired/found/bought therein. Then, for $q(1 - z)$ sufficiently low Eq. (134) yields

$$\langle Y \rangle \approx N e^{-Mq(1-z)}\,. \tag{135}$$

As argued by the authors, this expression corresponds to a power law if $z$ is a logarithmic function of the population size,

$$z(N) = a + b\ln(N)\,, \tag{136}$$

where $a$ and $b$ are constants. It turns Eq. (135) into the power-law $\langle Y \rangle = Y_0 N^{\beta_{\text{super}}}$, where

$$\beta_{\text{super}} = 1 + Mbq\,. \tag{137}$$

The authors argue that Eq. (136) can be explained by considering the way cities accumulate factors as they grow in size. According to them, this relationship emerges if factors are added to cities as they increase their size and a selection process occurs in which only the best or most useful factors survive. Such cumulative evolutionary processes have been analyzed in the cultural evolution literature [77], and they give rise to factors accumulating with the logarithm of population size.

To sum up, the model predicts that the super-linearity of the urban scaling exponent is due to (i) the number of factors a given socio-economic activity requires to happen (expressed by $M$), (ii) the capacity of the city to provide the necessary complementary factors ($b$); and (iii) the dependence of individuals to get factors from their urban environment ($q$). The larger these quantities are, the more pronounced the super-linearity.

## 6. Gomez-Lievano et al. – extreme value model

Gomez-Lievano et al. 2021 [78] argue that the super-linear scaling may not be a consequence of increasing returns to scale, as it is usually assumed. In this work, the authors propose a hypothetical situation where the non-linearity of the urban scaling could emerge even without an interaction process between the agents. They show that non-linearity can emerge by a selection process acting on independent random variables. In this sense, urban scaling would rather represent an artefact.

To demonstrate this argument, the authors propose the following model. Assume that a given individual has a productivity $w$, which is an independent random variable, and consequently, it does not depend on the size of the city he/she lives in. Moreover, this productivity is log-normally distributed, following

$$p_w(w|x_0, \sigma^2) = \frac{1}{w\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln w - \ln w_0)^2}{2\sigma^2}}\,, \tag{138}$$

where $w_0$ and $\sigma$ are positive parameters, such that $\ln w_0 = \langle \ln w \rangle$ is the expectation value of $\ln w$, $\sigma^2 = \mathrm{VAR}[\ln w]$ is the variance, and the productivity expectation value is given by $\mu = \langle w \rangle = w_0 e^{\frac{\sigma^2}{2}}$. The choice of a log-normal probability density function (pdf) is justified by some empirical evidence suggesting that productivity across workers, measured indirectly by wages, follows such distribution [79, 80].

The authors model the total socio-economic production of the $c$-th city as the sum of the productivity of all citizens of the city, that is

$$Y_c(N_c) = \sum_{i=1}^{N_c} w_i^c \tag{139}$$

where $N_c$ is the population of the $c$-th city. The expectation value of this production can be computed as $\langle Y(N) \rangle = \sum_{i=1}^{N} \langle w_i \rangle$ which implies $\langle Y(N) \rangle = N \langle w_i \rangle$. Similarly, for any other city with population $\lambda N$, where $\lambda$ is an increase factor, one gets $\langle Y(\lambda N) \rangle = \lambda N \langle w_i \rangle$. Consequently, it is possible to identify

$$\frac{\langle Y(\lambda N) \rangle}{\lambda N} = \frac{\langle Y(N) \rangle}{N}\,, \tag{140}$$

which means the per-capita socio-economic output remains unchanged in the face of an increase in population, which shows that a basic situation of independent random variables yields $\beta = 1$ (linear scaling).

However, scaling properties can appear if one considers the production $Y$ as proportional to the maximum value of the productivity in the city, i.e.

$$Y(N) \equiv \max\{\mathrm{w}_1, \cdots, \mathrm{w}_\mathrm{N}\}\,. \tag{141}$$

This represents the case where productivity is dominated by the most productive individuals.

For convenience, let's rewrite $w_i$ as $w_i = e^{\sigma z_i + \ln w_0}$, where $z_i$ is an independent random variable sampled from a normal distribution with mean 0 and variance 1. The parameter $w_0$ is choose to be $\ln w_0 = -\frac{\sigma^2}{2}$, by convenience only, to promotes $\langle w_i \rangle = e^{\ln w_0 + \frac{\sigma^2}{2}} = 1$. Then Eq. (141) can be rewritten as

$$Y(N) \sim e^{\sigma Z(N) - \frac{\sigma^2}{2}} \tag{142}$$

for $\sigma \gg \sqrt{\ln N}$. Here $Z(N) \equiv \max\{\mathrm{z}_1, \cdots, \mathrm{z}_\mathrm{N}\}$, and it is well known that it converges to the *Gumbel distribution* [81, 82], leading to

$$Z(N) \sim \sqrt{2 \ln N}\,. \tag{143}$$

Hence, from Eq. (142) one obtains

$$Y(N) \sim e^{\sigma \sqrt{2 \ln N} - \frac{\sigma^2}{2}}\,. \tag{144}$$

Considering the power law of the form $Y \sim N^\beta$ holds, the scaling exponent is given by

$$\beta = \frac{\frac{d}{dN} \ln Y}{\frac{d}{dN} \ln N}\,. \tag{145}$$

Applying such derivative to Eq. (144) yields

$$\beta = \frac{\sigma}{\sqrt{2 \ln N}}\,, \tag{146}$$

which is valid for $N \ll \sigma$, implying super-linearity. Otherwise prevails $\beta = 1$, as discussed previously.

The result of this model has at least two consequences. First, increasing super-linear scaling, and increasing returns to scale, can be a mere artefact because it can happen from the selection process of independent random variables. Second, it allows different regimes, linear and super-linear, depending on the relation between the size of the city and the variance in the distribution of workers' productivity. Small cities ($N \ll \sigma$) would exhibit super-linear scaling, while larger cities ($N \gg \sigma$) would exhibit a linear regime.

# Part II
# Inter-city Models

The models presented so far are based on city internal (intra-city) aspects. However, cities are not closed or isolated objects and, indeed, cities are in constant interaction with each other, be it by relations among individuals/firms from different cities [83], be it by the very migration flows between cities [84]. Consequently, processes taking place between cities must, in some way, interfere with the productivity and the use of infrastructure. It is natural also to elaborate how such interactions between cities, that is *inter-city* processes, could explain urban scaling or interfere with it, even if only in a second-order approximation. This section presents models that can be assigned to the intra-city category. This line of research is less advanced, as suggested by a smaller number of models and a wider range of concepts, and some of them only comprise a qualitative approach.

## 6.1. Pumain et al. model of technological diffusion

Pumain et al. [85] argue that non-linearities are due to interactions within the *system* of cities. More specifically, they propose that non-linearity emerges through a hierarchical diffusion process of innovations, from the largest to the smaller cities. According to this proposition, the innovation process is disproportionately higher in larger cities. Super-linear scaling of economic indicators represents a stage of the emergence of new technologies, which take place in larger cities; linear scaling represents the diffusion stage, from larger cities to towns and small cities of the system, and sub-linear scaling represents the mature stage of technologies, also characterised by decay or substitution processes. The merit of this theory is that it brings the interconnection between cities to the forefront, evidenced, for instance, by Zipf's law which reveals some kind of hierarchical structure in urban systems [86, 87, 88, 89, 90, 91, 92, 93, 94].

## 6.2. Gomez-Lievano et al. relation

Gomez-Lievano et al. [95] propose a statistical framework to characterize urban scaling and city size distributions. In simple terms, they derive the scenario under which the city population sizes $N$ follow Zipf's law [94, 96, 34], i.e. a power-law distribution according to

$$P(N) \sim N^{-\alpha}\,, \tag{147}$$

with $\alpha \approx 1$. Here, $P(N)$ represents the *(complementary) cumulative distribution function* (ccdf). In addition, the authors consider, based on empirical evidence, that the ccdf of a given urban indicator $Y$, namely $P(Y)$, also follows a power-law

$$P(Y) \sim Y^{-\alpha_Y}\,, \tag{148}$$

where $\alpha_Y$ is usually different from 1 for socio-economic and infrastructure urban variables.

While the authors use a probabilistic characterization of urban scaling, in the following we present a back-of-an-envelop derivation. If $p(N)$ is the *probability density function* (pdf), in the sense that $P(N) = \int p(N)dN$, and similarly for $p(Y)$, then one distribution can be transformed into the other obeying the density transformation

$$p(N)dN = p(Y)dY\,. \tag{149}$$

If we use the distributions Eq. (147) and (148) then we can write the integrals

$$\int N^{-\alpha-1}dN \sim \int Y^{-\alpha_Y-1}dY\,, \tag{150}$$

which leads to $N^{-\alpha} \sim Y^{-\alpha_Y}$ and

$$Y \sim N^{\frac{\alpha}{\alpha_Y}}\,. \tag{151}$$

Finally, comparison with Eq. (6) provides

$$\beta = \frac{\alpha}{\alpha_Y}\,. \tag{152}$$

This means that the scaling exponent is directly related to the Zipf exponent and vice versa; Zipf's law and urban scaling are connected phenomena. Recently, various relations were proposed explaining the connection between the Zipf and scaling exponents quantitatively [97] and qualitatively [98]. More specifically, urban scaling transforms the Zipf distribution ($\alpha \approx 1$) into another power-law distribution with another exponent $\alpha_Y$ that differs from $\alpha$ if $\beta \neq 1$.

However, Ribeiro et al. [97] argue that Eq. (152) only represents an upper limit. By permuting the values $N$ and $Y$ from different cities, the association is destroyed so that correlations vanish ($\beta \approx 0$) – but the distributions and the exponents $\alpha$, $\alpha_Y$ remain unaffected. This thought experiment leads to a situation where Eq. (152) is violated. However, different degrees of correlations permit $\beta$-values up to $\frac{\alpha}{\alpha_Y}$. In other words, Eq. (147) and $Y \sim N^\beta$ imply Eq. (148) but Eqs. (147) and (148) do not imply $Y \sim N^\beta$ with $\beta = \frac{\alpha}{\alpha_Y}$.

### 6.3. H.Ribeiro et al. model – country scaling

Ribeiro et al. [97] empirically relate Zipf's law for cities and urban scaling. Based on data from many countries, they find correlations between the Zipf-exponent $\alpha$ and the urban scaling exponent $\beta$ for GDP. In order to explain these correlations, the authors argue that for a given total urban population the country-wide urban GDP is fixed and different values of $\alpha$ require an adjustment of $\beta$ so that the country-wide aggregate is preserved. The same argument is used vice versa, i.e. the authors make no statement about the direction of a possible causality.

Combining Zipf's law and urban scaling, the total country-wide output of a considered (additive) socio-economic urban metric is given by

$$Y^* = \sum_{N=N_{\min}}^{N_{\max}} Y(N)h(N) \sim \int_{N=N_{\min}}^{N_{\max}} N^\beta h_1 N^{-\alpha-1} dN \,. \tag{153}$$

where $h(N)$ is the frequency function, which according to Zipf's law Eq. (147) is $h(N) = h_1 N^{-\alpha-1}$; and $h_1$ is the normalization constant from the city size distribution. The constant $h_1$ can be obtained from the total urban population

$$N^* = \sum_{N=N_{\min}}^{N_{\max}} N h(N) \approx \int_{N_{\min}}^{N_{\max}} N h_1 N^{-\alpha-1} dN \,. \tag{154}$$

These equations are solved by considering that the sizes of the largest and smallest cities in a country depend on the total population of that country, following power-laws,

$$N_{\max} = b(N^*)^\theta \qquad \text{and} \qquad N_{\min} = a(N^*)^\delta \,, \tag{155}$$

where $a, b, \theta, \delta$ are constants. Introducing these relations in Eq. (153), the authors obtain

$$Y^* \sim \frac{(\alpha-1)N^*}{\alpha-\beta} \left( \frac{a^\beta b^\alpha (N^*)^{\delta\beta+\theta\alpha} - a^\alpha b^\beta (N^*)^{\theta\beta+\delta\alpha}}{ab^\alpha (N^*)^{\delta+\theta\alpha} - a^\alpha b (N^*)^{\theta+\delta\alpha}} \right) . \tag{156}$$

Country populations are large ($N^* \gg 1$) and in this limit the total aggregate urban metric becomes

$$Y^* = Y_0^* (N^*)^\gamma \,, \tag{157}$$

where $Y_0^*$ and $\gamma$ are constant. Which of the four terms in the parenthesis of Eq. (156) dominates in the limit, depends on the values of the exponents $\alpha, \beta, \delta, \theta$. Considering Eq. (156) in the limit $N^* \gg 1$ and comparing with Eq. (157), the exponents can be solved for $\beta$ yielding

$$\beta = \begin{cases} 1 + \frac{\gamma-1}{\theta} & 0 < \alpha \leq 1 \\ \frac{\gamma+\delta-1}{\theta} + \left(1 - \frac{\delta}{\theta}\right)\alpha & 1 < \alpha < 1 + \frac{\gamma-1}{\delta} \\ 1 + \frac{\gamma-1}{\delta} & \alpha \geq 1 + \frac{\gamma-1}{\delta} \end{cases} , \tag{158}$$

for $\gamma > 1$ and $\delta < \theta$ (the authors provide similar expressions for other conditions). It is a step-wise function (see Fig. 11), which in the middle regime exhibits a linear relation between $\beta$ and $\alpha$. Solving for $\beta$ does not mean that there is a causality from $\alpha$ on $\beta$.

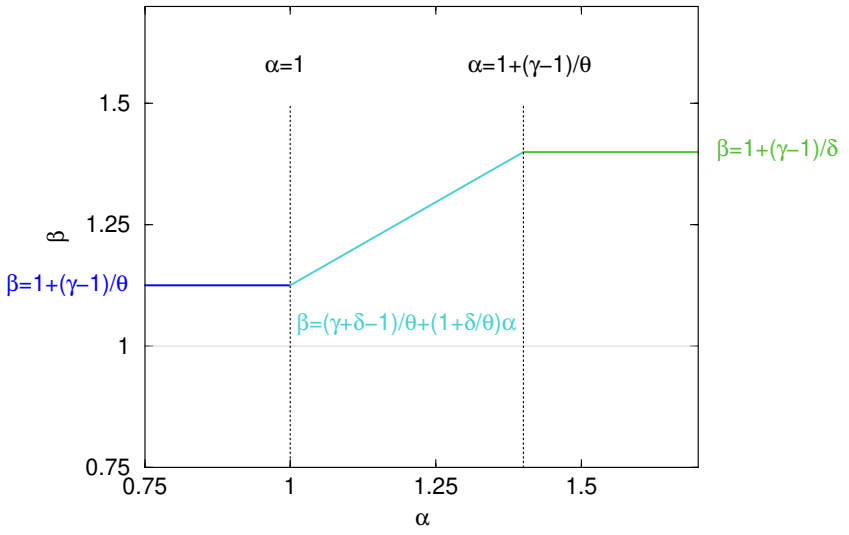


**Figure 11:** Illustration of Eq. (158) relating the urban scaling exponent $\beta$ and the Zipf exponent $\alpha$. Here the parameters $\gamma = 1.1$, $\delta = 0.25$, and $\theta = 0.8$ have been used. Source: adapted after [97].

Last, it needs to be mentioned that $\alpha$ and $\beta$ are exponents across cities and there is one value each for a country. The other exponents $\gamma, \delta, \theta$ are country scaling exponents across countries. This means, considering a country or a similar confined region as it is generally done in this paper, the country scaling exponents are constants. E.g. for a given country population $N^*$ Eq. (157) with $\gamma$ implies a fixed country aggregate $Y^*$.

Comparison of Eq. (152) and the middle regime of Eq. (158) leads to $\frac{\gamma+\delta-1}{\theta} = 0$ and $(1 - \frac{\delta}{\theta}) = \frac{1}{\alpha_Y}$. It supports the compatibility of both views. Fitting the country scaling relationships Eqs. (155) and (157), Ribeiro et al. [97] report $\gamma \approx 1.31$, $\delta \approx 0.20$, and $\theta \approx 0.79$, approximately confirming the first relation and leading to $\alpha_Y \approx 1.34$. However, as the country scaling relations only describe the mean field, the value $\alpha_Y \approx 1.34$ needs to be interpreted in the same sense and one needs to keep in mind some spread around it.

### 6.4. Altmann et al. model – attractiveness token

The last model to be discussed is the one developed by Altmann [99] based on an approach initially proposed by Leitão et al. [100]. This model differs from the other models presented here in the sense that it does not lead to the scaling exponent as an emergent phenomenon. However, it is essential to mention it because the model builds on interactions between individuals of different cities, introducing ideas about integrating intra- and inter-city aspects. Moreover, this model suggests ideas that could be incorporated into the intra-city models, in order to expand them to an inter-city approach.

The model considers an urban system composed of $N_{\text{cit}}$ cities, where $N_c$ is the population of the $c$-th city, in such a way that the total population of the system is $N^* = \sum_{c=1}^{N_{\text{cit}}} N_c$. There are also $Y^*$ tokens that are randomly assigned to the people of the system. This token can be, for example, a patent or a socio-economic output. Let's denote $p(i)$ the probability that one token is attributed to the individual $i$ out of $N^*$ individuals. Altmann proposes that this distribution is a function of the attractiveness $x_i$ of this individual [99]. The attractiveness would be related to the ability of the individual to attract one of the tokens – for example, charisma, leadership, or professional training. In other words, token attractiveness can also be understood in an active sense, i.e. related to the skills of an individual or to which extent an individual is able to convince others. This quantity can be thought of as the result of the connection between this individual and all the others, in the form

$$x_i \propto \sum_{j=1}^{N^*} p_{\text{int}}(r_{ij}) \tag{159}$$

where $p_{\text{int}}(r_{ij})$ is a measure of the interaction between the individuals $i$ and $j$, that are at the distance $r_{ij}$ one each other. Here $r_{ij}$ can be within or across the boundaries of a city, since $i$ and $j$ can belong to the same or to different cities.

Note that the attractiveness $x_i$ also appears in the Yakubo et al. model (see Sec. 2.4.2), and $p_{\text{int}}(r_{ij})$ is the probability of interaction in the context of gravity models (see Sec. 2.4). This way, $p_{\text{int}}(r_{ij})$ can, for instance, obey the power-law form given by Eq. (34). Moreover, Eq. (159) can also recover a strict intra-city process when $p_{\text{int}}(r_{ij}) = 0$ for $i$ and $j$ residing in different cities. There is a subtle but essential difference between the attractiveness considered by Yakubo et al. and the one considered by Altmann. In the former, the attractiveness *implies* the probability of interaction (as given by Eq (42)); while in the latter the attractiveness is a *consequence* of the probability of interaction (as given by Eq. (159)).

Returning to the model, the probability that a token is allocated to a particular city $c$ is $p_{\text{tok}}(c) = \sum_{i \in c} p(i)$, and the expected number of tokens at this city can be computed using this probability, i.e.

$$Y_c = Y^* \cdot p_{\text{tok}}(c)\,. \tag{160}$$

Altmann considers, by hypothesis, a non-linear efficiency of individuals, expressed by

$$p(i) = \frac{x_i^{1-\beta}}{Z(\beta)}\,, \tag{161}$$

where $Z(\beta)$ is the normalization constant given by the sum $\sum_{i=1}^{N^*} p(i) = 1$. In the particular case that $x_i$ is the same for all individuals of the same city $c$, that is $x_i = x_c$ for any $i$ belong to city $c$, the probability of this city to attract a token is

$$p_{\text{tok}}(c) = N_c \cdot \frac{x_c^{1-\beta}}{Z(\beta)}. \tag{162}$$

Then, from Eq. (160), the expected number of tokens at the $c$-th city is given by

$$Y_c = Y^* \cdot N_c \cdot \frac{x_c^{1-\beta}}{Z(\beta)}\,. \tag{163}$$

The probability to observe the set $\{Y_c\}_{c=1,\cdots,N_{\text{cit}}}$ in the set of cities of size $\{N_c\}_{c=1,\cdots,N_{\text{cit}}}$ is indeed a multinomial distribution

$$P(Y_1, \cdots, Y_{N_{\text{cit}}} | N_1, \cdots, N_{N_{\text{cit}}}) = \frac{Y^*!}{\prod_{c=1}^{N_{\text{cit}}} Y_c!} \prod_{c=1}^{N_{\text{cit}}} \left( \frac{N_c \cdot x_c^{\beta-1}}{Z(\beta)} \right)^{Y_c}\,. \tag{164}$$

It correspond to the likelihood of the data $\{Y_c\}_{c=1,\cdots,N_{\text{cit}}}$ given a fixed population $\{N_c\}_{c=1,\cdots,N_{\text{cit}}}$.

The first version of the model [100] is without spatial interactions. However, in the version presented in [99], Altmann has generalized the model to take into account spatial interactions between individuals by Eq. (159). He investigates the parameter $\beta$ that maximizes the posterior of this likelihood, varying the *typical interaction distance* $r_0$ that governs the interaction range between the individuals. For instance, in the case where the probability of interaction is governed by an exponential decay $p_{\text{int}}(r_{ij}) \sim e^{-\ln(2) r_{ij}/r_0}$, implies that the interaction becomes 0.5 for $r_{ij} = r_0$ and 1 for $r_{ij} = 0$ (intra-city case).

He finds that the value of $\beta$ that maximizes Eq. (164), say $\beta^*$, is a function of the typical distance $r_0$, i.e. $\beta^* = \beta^*(r_0)$. In addition, he finds that $\beta^*$ has a maximum value for $r_0 \neq 0$, which reflects an effect from an inter-city process. For instance, for the urban GDP of Brazilian cities, $\beta^*$ is maximized when $r_0 = 14.6\,\text{km}$, suggesting that this is the typical interaction range for this system of cities.

Altmann argues that urban scaling is a consequence of the non-linearity of the individuals' efficiency that depends on the size of the city they live in, modelled by Eq. (161), which is in line with what other authors argue [11]. Another distinguishing aspect of Altmann's work is a systematic way – via the likelihood (164) – to test different models, with their parameters being extracted and tested from data. It involves, according to Altmann, separating the analysis into two steps. First "explaining" the emergence of scaling (the model) and second, making a fit to the data (by maximum likelihood approach).

For future work, it could be interesting to incorporate into this model some ideas from the models presented previously. For example, to incorporate the power-law distribution of attractiveness – as done by Yakubo (see section 2.4.2) – or to use a range of interaction as captured by the parameter $\gamma$ discussed in the gravity models context (section 2.4.1).

## 7. Discussion and future directions

The purpose of this work goes beyond simply presenting the mathematical models that aim at explaining the urban scaling phenomenon. Instead, here we explore what one can find out when all such models are organized and compared systematically. Which similarities do they share and what distinguishes a specific model? And most importantly, what can be inferred from the junction of all these models for future research designs and schools of thought? These perspectives will be discussed in this section.

In order to facilitate the organization, comparison, and identification of research gaps, we propose the taxonomy depicted in Figs. (2) and (3). The first distinguishing property is that the models can be divided into two types, the ones based on processes or interactions within a city and the ones based on processes between cities. In the former, intra-city type, most models consider that urban scaling emerges from human interactions, i.e. intentionally or accidentally meeting people, in the considered city. In the following subsections, we synthesize the particularities in more detail (in arbitrary order).

### 7.1. Intra- versus inter-city process

A point that has been little explored by the theoretical research is the role of exogenous factors in the scaling of cities. The imbalance between models of intra- and inter-city processes (as also visible in Fig. 2) may be attributed to early successes of publications employing human interactions within cities to explain urban scaling. However, external factors – including inter-city processes – should be better investigated in future work, either to show that they represent negligible perturbations compared to the internal factors, responsible for urban scaling, or the opposite, that external factors can significantly affect the scaling laws discussed here.

Intra-city models are based on reasonable assumptions and prove to be successful in describing urban scaling. Nevertheless, at the same time, cities are not isolated objects, and there are important interactions between them. An example includes commuting [101, 102] – quantified by the number of people who live in certain cities but work in others – and how it can increase urban productivity. The models based on processes between cities, in addition to being few, are less explicit, and the emergence of urban scaling is barely derived but, e.g. somewhat related to other scaling laws.

Furthermore, it will be interesting to study how digital means of interaction (e.g. video conferencing) change the role of distance and endogenous/exogenous interactions and how this will have short-, medium-, and long-term repercussions in socio-economic urban productivity. Conducting experiments to quantify these issues is essential for developing a systematic understanding of cities.

### 7.2. Probability of interaction

We find that models employing human interactions within a city are all based on a given probability of interaction. The main idea behind this kind of models is that the exchange of knowledge and experiences generates ideas and innovations, which results in increasing returns and economy scale. The super-linearity that the number of contacts that people have represents strong evidence and argument favouring the idea that the connections between people are a crucial mechanism that causes increasing returns to scale. It is supported by empirical findings and the theoretical background presented here.

The models that are based on the interaction within cities differ one from the others only in the reasoning behind how the interaction probability is estimated theoretically. Exploring this similarity, we unify this conceptual overlap in a framework that formalizes this probability (see Fig. 4). Supporting this framework with empirical analysis would represent a great step forward, e.g. quantifying the frequency of encounters between personal profiles. Moreover, the type of interaction plays an important role but is neglected by all models. Overall, the spatial organization of social interactions on urban and pan-urban scales is of high relevance. Recent works that are going empirically in this direction include [103, 104, 105]. Growing data, as collected by mobile phones and GPS, might provide a better measure of these human interactions in space.

## 7.3. Gravity and city integrity

The analogy to gravity in physics has a long history and has been studied in a broader context, including population flows [54], like commuters [56], or spatial explicit modelling [106]. Reviewing the gravity models used to explain urban scaling, we find that three groups of authors independently employ different ideas that lead to equivalent results (within given parameter configurations). We interpret this observation as strong support for the gravity idea, i.e. that some sort of interaction decays with some sort of distance, interfering in some way with urban scaling. A closer inspection of the variants permits us to draw conclusions, that might serve as guidelines to public managers and as theoretical justifications for implementations of social integration policies.

Different approaches of gravity show the consistency of the presented models in terms of (i) good access to all parts of the city, (ii) influencers reaching distant parts of the city, and (iii) interaction between socially distant people. It was also possible to see via these models how geometry can interfere with people's interactions and how it affects the socio-economic development of cities; and finally, how it could be responsible for generating urban scaling. One point that must be highlighted is the axiom of the city's integrity as one necessary condition to explain scaling properties to emerge. Different models (Ribeiro et al. and Yakubo et al. models in the context of gravity idea, and also the Bettencourt model) show that the increasing returns to scale are only possible if the city behaves as a whole, with a good integration of its different geographical parts. This aspect deserves more attention, especially when it concerns the design of empirical research to understand the integrity and connection more profoundly. One work that went in this direction is [104], which using mobile phone data and analyzing the trajectory of thousands of people found that the density of visitors of a place decays with the inverse of the square of both distance and frequency of visits. However, as important as the design of experiments to corroborate or contradict these models' outcomes is to design urban policies and to use such findings to improve the cities' efficiency. That is, to understand this connection force as a guiding principle to generate an assisted socio-economic improvement of cities.

One example of improved socio-economic output due to increased inter-connectivity is the city of Medellin (Columbia) and its cable car, which connects previously isolated areas of the city [107]. It had a significant and persistent influence on the socio-economic index of the city, showing that integrating people is a vital ingredient to enhancing the development of the cities. For future work, it remains to systematically identify such examples and to investigate quantitatively the effect of integration, i.e. how it reverberates quantitatively to the city development. For instance, how the improvement of transport efficiency – for example, implementing links in a subway network or increasing the number of bus lines – changed the city's GDP. Ex-post analysis of developing cities could provide us with an excellent opportunity to quantify these relationships between integration and productivity. This includes also the validation of models presented here – and perhaps the creation of more general and more robust ones.

Gravity approaches appear in the urban literature in at least three domains: (i) geo-spatial interactions [56, 84, e.g.], (ii) urban form and growth [108, 109, e.g.], and (iii) urban scaling [48, 12, e.g.]. They all resemble Newton's law of gravity, with a decreasing distance power-law function controlled by the exponent $\gamma$.

In the context of (i) *geo-spatial interactions*, the urban dynamics – such as population flows (e.g. commuters) – is empirically modelled by a gravity form consisting of the product of origin and destination mass (usually populations), and the exponent $\gamma$ is usually obtained by regression. In the context of (ii) *urban form and growth*, the spatial distribution of the population, or the built-up area of a city, can be simulated on a grid where the probability of urbanization is modelled by a decreasing power-law of the distance to already urban locations. The exponent $\gamma$ of this power-law is a parameter to be chosen. In the context of (iii) *urban scaling*, the gravity form is used to model the probability of interaction between people, which is controlled by the $\gamma$ parameter.

It will be highly relevant to study and understand the relations between these three forms of gravity approaches, and if it is possible to unify them. It could lead to an explanation of the $\gamma$-exponent itself, that is, the parameter that controls the space impedance. In other words, it is desirable to derive the exponent $\gamma$ instead of treating it as an inexplicable parameter. Solving this problem will also help to understand the role of morphology in urban scaling, specifically the interaction between urban scaling, fractality, and gravity.

## 7.4. Urban morphology and geometry

Geometry plays a fundamental role in describing urban scaling, and it was evidenced by gravity models and the Molinero & Thurner model, both presented here. It is important to stress that the findings of these models are not merely naive analogies. Of course, we know that the presence of lakes or mountains (components of the geometry/geography of the city) interferes with the flux and, consequently, the interaction between people. However, the models' implications go beyond these basic premises. Indeed we found quantitative relations between urban scaling

and the fractal dimensions of the spatial structure formed by the people and the fractal dimension of the street network. This means, the models and the empirical findings reveal, to some extent, how the shape of the city facilitates or obstructs human contact, which, in turn, reverberates in economic productivity and scaling. In other words, through the models, we can observe a quantitative relationship between the shape (geometry) and scaling (socio-economic properties).

Given these insights, it remains to investigate these properties more deeply, especially in other countries, given that Molinero & Thurner exclusively analyzed European cities. They found that the ratio between the population and infrastructure fractal dimensions is similar to those (European) cities. However, is this ratio universal? Can the same value of this ratio be observed across a wider set of countries? One point that calls attention in this context is that the entropy of the spatial distribution of urban settlements strongly depends on the degree of development of the country/city [110, 111]. Given that entropy is related to the fractal dimension [112] and that entropy depends on the country's development, it suggests that maybe the ratio between fractal dimensions found by Molinero & Thurner could be different for non-developed countries and thus it could not be universal. Of course, that is speculative, and only empirical analysis in a broader set of countries can bring some light to this issue.

## 7.5. The influence of the city definition

Although this paper has a theoretical focus, we cannot leave without discussing an empirical issue. When we see a city, we immediately recognize it as such, but where does it start and where does it end? The question of how to define cities is probably as old as cities research itself. It is an ongoing problem if and how it affects scaling exponents – see [96, Sec. 3.1] for a discussion of the "units of observation" in the context of Zipf's law for cities.

Arcaute et al. [113] define urban units via commuting and population density thresholds. In many cases, subsequent urban scaling analysis leads to vanishing non-linearity or non-universal exponents. Similarly, Cottineau et al. [114] report that "different scaling regimes can be encountered for the same territory, time and attribute, depending on the criteria used to delineate cities". Strikingly, [115] analyze urban $CO_2$ emissions in the USA and obtain super-linear scaling ($\beta = 1.37$) for "urban areas" and sub-linear scaling ($\beta = 0.95$) for "combined statistical areas".

In contrast, Dong et al. [58] analyze urban scaling on a within-city scale and confirm super-linear scaling for socio-economic activity. In the opposite direction, Ribeiro et al. [97] find that (urban) GDP also follows a power-law on the country scale, i.e. when each point in the scaling-plot represents a country instead of a city. Both findings support the robustness of (urban) scaling.

On the one hand, these works add to the problem of "city definition", and it would be interesting to contextualize the findings with the city's integrity condition as discussed in the context of gravity models. On the other hand, the sensitivity to the choice of the units of observation represents an interesting scientific problem. Researchers do mention the Modifiable Areal Unit Problem (MAUP) [114, e.g.] but to our best knowledge there is no theoretical work solving the problem in the urban scaling context.

A starting point could be the ordinary least squares (OLS) slope given by $\beta_{\mathrm{OLS}} = \rho(\sigma_{\ln Y}/\sigma_{\ln N})$ [116, e.g.], where $\rho$ is the correlation coefficient between $\ln Y$ and $\ln N$, and $\sigma_{\ln Y}, \sigma_{\ln N}$ are the respective standard deviations. Changing the aggregational scale, the standard deviations obey $\sigma \sim s^{h-1}$, where $s$ is the number of smallest units in the aggregated values, and $h$ is a fluctuation exponent. In the absence of correlations, $h = 1/2$ and $h - 1 < 0$ so that the standard deviation decreases with increasing aggregational scale following $s^{-1/2}$. Changes of $\beta$ with the aggregational scale can occur when population and urban indicators exhibit different spatial correlations ($h_{\ln Y} \neq h_{\ln N}$). Moreover, the correlation coefficient $\rho$ is affected by the aggregational scale (ecological fallacy [117, e.g.]).

## 7.6. Longitudinal and transversal scaling

We also need to mention that here we exclusively consider urban scaling cross-sectionally. The equivalence of cross-sectional (transversal) and temporal scaling (longitudinal) is tempting but requires some formal considerations. Recently, two groups independently and consistently provide the theoretical description [118, 119]. Essentially, these works suggest that temporal and cross-sectional scaling are the same, given that cities have sufficiently large growth rates and given that exogenous factors can be neglected.

Remains for future works to analyze the connection between the models presented in this paper in the context of individual cities' growth. In most cases, models of urban scaling leave it open if they are only valid cross-sectionally or also temporally. While longitudinal and transversal scaling is basically understood from the empirical perspective, the modelling is not explored at all. This is best visible by the fact that none of the models includes the time variable in any sense. Accordingly, they implicitly operate at a constant time where the time $\Delta t$ needed by the processes is negligible

to the time that changes of parameters and constraints take. This, however, might be an unjustified assumption that requires further investigation.

### 7.7. Zipf's law and urban scaling

If large cities are wealthier than smaller ones, then they should also be more attractive. In this way, a systematic population flow among cities of different sizes should lead to a change in the city size distribution. Empirically, such a temporal process has not been observed. However, Ribeiro et al. [97] do find an association between urban scaling and city size distribution. Their finding not only suggests inter-city processes but also that there must be some truth to the intuition. Nevertheless, it is also plausible that there must be some sort of counteracting force holding people back, otherwise everyone would migrate to the largest city. In a general sense, the dynamics can be formulated by a balance equation that combines intrinsic population growth and migration between cities [33]. Significant work has been done in solving the balance equation [120], but more effort is needed to understand the connection between urban scaling and Zipf's law for cities.

A better understanding of inter-city processes (and the association between Zipf's law and urban scaling) will require a better understanding of the spatial organization of cities in a country or region. Certainly, such attempts go back to Christaller, who formulated the Central Places Theory (CPT) in 1933 [121]. While Zipf's law implicitly gives a hierarchical organization of cities, an explicit manifestation is given by CPT [122]. Accordingly, we suggest reviewing CPT in view of today's understanding of urban systems, particularly in front of urban scaling. E.g. according to CPT, cities produce goods and services that exceed the needs of their own population and serve a "basin of attraction", which includes smaller cities and settlements. Therefore, the larger the centrality of the city is, the bigger it will be. Moreover, Christaller quantifies the centrality by the number of telephones relative to population – similar to [123] but assuming $\beta = 1$ [121, Sec.B1, p.152]. However, to what extent the various components of CPT hold true represents an additional question and requires up-to-date empirical analysis.

### 7.8. Forms of density scaling

In the context of the fundamental allometry, Eq. (2), we had mentioned the population density. Some colleagues explore the population density as the alternative independent variable, i.e. instead of the plain population.

E.g. Newman & Kenworthy [13] plot the annual gasoline use per capita as a function of population density and find for a set of metropolises a strong dependence with low density being associated with high gasoline use. More recently, a similar relation was found for urban carbon emissions in the USA which was characterized by means of power laws [124]. The Cobb-Douglas form discussed in Sec. 4.2 can to some extent reconcile this form of density scaling with the conventional urban scaling [71].

A different form of density scaling is studied by Gastner & Newman [125], i.e. facilities (hospitals, airports, or malls) per area as a function of population per area. The authors find a power law with the exponent 0.66. This value agrees with their theoretical prediction $2/3$ which is based on the global minimization of total travel distance between people and facilities [125]. Um et al. [126] expand the understanding by distinguishing public and commercial facilities. They argue that public facilities follow the minimization of travel distance but commercial facilities aim at maximizing the number of people in their respective Voronoi cells (defined by the spatial organization of other facilities of the same type). Then, in the case of commercial facilities, their derivation leads to an exponent 1 between facility density and population density [126]. Although this form of density scaling is well described, it remains open how it relates to conventional urban scaling. In this context, it could also be of interest to relate density scaling to Reilly's law of retail gravitation [127].

### 7.9. Urban scaling and the New Science of Cities

Given all the challenges cities are facing, such as spatial segregation, accessibility, mobility, pollution, energy demand, congestion, and crime, it is an urgent need to develop an operational science of the city. That is, a systematic and quantitative way to govern our cities using a predictive theory that is empirically grounded.

One way to approach this problem is via the analyses (by machine learning) of a massive amount of data that is nowadays available due to the omnipresence of sensors (e.g. mobile phones, GPS, digital office). Another way is by performing computer simulations to predict scenarios and opportunities. However, such approaches – while being fashionable and tempting – have their own pitfalls. While big data and machine learning usually cannot reveal the processes of the system, the computational simulations are often based on a large set of parameters which mostly impedes any general interpretation. However, minimal mathematical models, like the ones presented here, are the

way to understand more profoundly a complex system as the urban phenomenon. This approach allows identifying the predominant mechanisms – the *essence* – of the phenomenon without getting lost in the less important and non-fundamental details.

Finally, we have to see how these models contribute to the New Science of Cities [3, 1, 2, 128]. Here we discuss only one facet of the urban phenomenon, namely urban scaling. Nevertheless, the various emergent properties observed in cities and urban systems, such as fractality, segregation, and mobility patterns, are mainly studied and understood as independent features. We argue in favour of a holistic view, where the findings presented here only represent a particular case of a more general theory. A beginning could be models that cover both intra- and inter-city processes parsimoniously. Specifically, such models could explain the association between urban scaling and city size distribution as recently found [97]. As such models would include two important urban regularities, they could represent an important step towards a Unified Urban Theory (UUT) – a quantitative theory that combines few premises and derives many different observed urban patterns as particular cases [98].

## 8. Final Remarks

The research of urban scaling attained a mature state where initial indication in data has been expanded to a solid empirical finding – and to a range of theoretical models mathematically deriving the empirical exponents. We discuss and contextualize a set of modelling approaches, whereas most of the considered models (a) aim at explaining the emergence of non-linear urban scaling and (b) consist of a formal derivation. On the one hand, our work summarizes the models in a comprehensible and coherent manner and, on the other hand, compares and relates them in order to identify similarities and dissimilarities. Consequently, the purpose of this work is to infer perspectives for the field.

To finalize, we would also like to mention works that challenge the urban scaling framework (in addition to the problem of how to define cities, see Sec. 7.5). Leitão et al. [100] find that empirically different estimations of the exponent $\beta$ can be obtained depending on the assumptions made about the fluctuations around the urban scaling relation. Gomez-Lievano et al. [78] show that non-linear urban scaling can emerge as an artefact when extreme value theory is employed. Despite overwhelming evidence favouring urban scaling, these works and similar publications need to be taken more seriously as they can also point towards weak spots in the theoretical models.

Overall, we hope this paper disentangles the plethora of urban scaling models and thereby synthesizes new understanding. Substantial progress has been made by the community, but we also think that the chapter of urban scaling models cannot be closed yet. We also hope that approaches and concepts developed in related disciplines – including social sciences, economics, and geography – can be translated into mathematical models in order to obtain a broader and more general understanding of cities and urban systems.

## Acknowledgments

We wrote this paper in memoriam of João Meirelles, who unexpectedly passed away in 2020 and who helped with the first insights that gave rise to this work. We would like to thank the various authors who supported us by verifying the description of their models and approaches. F. L. Ribeiro thanks CAPES (grant number 88881.119533/2016-01), CNPq (grant numbers 403139/2021-0 and 424686/2021-0), and Fapemig (grant number APQ-00829-21). D. Rybski thanks the Alexander von Humboldt Foundation for financial support under the Feodor Lynen Fellowship. D. Rybski is grateful to the Leibniz Association (project IMPETUS) for financially supporting this research.

## References

[1] J. Lobo, M. Alberti, M. Allen-Dumas, E. Arcaute, M. Barthelemy, L. A. Bojorquez Tapia, S. Brail, L. Bettencourt, A. Beukes, W. Chen, R. Florida, M. Gonzalez, N. Grimm, M. Hamilton, C. Kempes, C. E. Kontokosta, C. Mellander, Z. P. Neal, S. Ortman, D. Pfeiffer, M. Price, A. Revi, C. Rozenblat, D. Rybski, M. Siemiatycki, S. T. Shutters, M. E. Smith, E. C. Stokes, D. Strumsky, G. West, D. White, J. Wu, V. C. Yang, A. York, and H. Youn, "Urban Science: Integrated Theory from the First Cities to Sustainable Metropolises," *SSRN Electronic Journal*, 2020.

[2] L. Bettencourt and G. West, "A unified theory of urban living.," *Nature*, vol. 467, pp. 912–3, oct 2010.

[3] M. Batty, *The new science of cities*. MIT Press,, 2013.

[4] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, "Growth, innovation, scaling, and the pace of life in cities.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 7301–6, apr 2007.

[5] J. Meirelles, C. R. Neto, F. F. Ferreira, F. L. Ribeiro, and C. R. Binder, "Evolution of urban scaling: Evidence from Brazil," *PLOS ONE*, vol. 13, p. e0204574, oct 2018.

[6] L. Sveikauskas, "The Productivity of Cities," *The Quarterly Journal of Economics*, vol. 89, no. 3, 1975.

[7] L. M. Bettencourt and G. B. West, "Bigger cities do more with less.," *Scientific American*, vol. 305, no. 3, pp. 52–53, 2011.

[8] J. Meirelles, F. Ribeiro, G. Cury, and C. Binder, "More from Less? Environmental Rebound Effects of City Size," *arXiv preprint arXiv:2001.09968*, pp. 1–21, 2020.

[9] C. Kühnert, D. Helbing, and G. B. West, "Scaling laws in urban supply networks," *Physica A: Statistical Mechanics and its Applications*, vol. 363, pp. 96–103, apr 2006.

[10] J. Norman, H. L. MacLean, and C. A. Kennedy, "Comparing high and low residential density: life-cycle analysis of energy use and greenhouse gas emissions," *Journal of urban planning and development*, vol. 132, no. 1, pp. 10–21, 2006.

[11] L. M. A. Bettencourt, "The origins of scaling in cities," *Science*, vol. 340, no. 6139, pp. 1438–41, 2013.

[12] F. L. Ribeiro, Joao Meirelles, F. F. Ferreira, and C. R. Neto, "A model of urban scaling laws based on distance-dependent interactions," *Royal Society Open Science*, vol. 4, no. 160926, 2017.

[13] P. W. G. Newman and J. R. Kenworthy, "Gasoline consumption and cities: a comparison of us cities with a global survey," *J. Am. Plann. Assoc.*, vol. 55, no. 1, pp. 24–37, 1989.

[14] R. Gudipudi, T. Fluschnik, A. G. C. Ros, C. Walther, and J. P. Kropp, "City density and co2 efficiency," *Energ. Policy*, vol. 91, pp. 352–361, 2016.

[15] D. Rybski, D. E. Reusser, A. L. Winz, C. Fichtner, T. Sterzel, and J. P. Kropp, "Cities as nuclei of sustainability?," *Environ. Plan. B*, vol. 44, no. 3, pp. 425–440, 2017.

[16] J. Q. Stewart, "Suggested principles of "social physics"," *Science*, vol. 106, no. 2748, pp. 179–180, 1947.

[17] S. Nordbeck, "Urban allometric growth," *Geogr. Ann. B*, vol. 53, no. 1, pp. 54–67, 1971.

[18] M. Batty and P. Ferguson, "Defining city size," *Environ. Plan. B*, vol. 38, no. 5, pp. 753–756, 2011.

[19] S. G. Ortman, A. H. F. Cabaniss, J. O. Sturm, and L. M. A. Bettencourt, "The pre-history of urban scaling," *PLoS One*, vol. 9, no. 2, p. e87902, 2014.

[20] M. J. Hamilton, B. T. Milne, R. S. Walker, and J. H. Brown, "Nonlinear scaling of space use in human hunter–gatherers," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 111, pp. 4765–4769, 2007.

[21] J. R. Burger, J. G. Okie, I. Hatton, V. P. Weinberger, M. Shrestha, K. J. Liedtke, T. Be, A. R. Cruz, X. Feng, C. Hinojo-Hinojo, A. S. M. G. Kibria, K. C. Ernst, and B. J. Enquist, "Global city densities: re-examining urban scaling theory," *arXiv e-Print*, vol. arXiv:2210.08067 [physics.soc-ph], 2022.

[22] W. J. Coffey, "Allometric growth in urban and regional social-economic systems," *Canadian Journal of Regional Science*, vol. 11, no. 1, pp. 49–65, 1979.

[23] S. Nordbeck, "Urban Allometric Growth," *Geografiska Annaler. Series B, Human Geography*, vol. 53, no. 1, p. 54, 1971.

[24] M. Batty and P. Ferguson, "Defining city size," *Environment and Planning B: Planning and Design*, vol. 38, no. 5, pp. 753–756, 2011.

[25] L. M. A. Bettencourt, *Introduction to Urban Science: Evidence and Theory of Cities as Complex Systems*. Cambridge, MA: The MIT Press, 2021.

[26] L. M. Bettencourt, "Towards a statistical mechanics of cities," *Comptes Rendus Physique*, vol. 20, no. 4, pp. 308–318, 2019.

[27] L. M. A. Bettencourt, "Urban Growth and the Emergent Statistics of Cities," *Science Advances*, no. August, p. 20, 2020.

[28] C. A. Hidalgo and E. E. Castañer, "The Amenity Space and The Evolution of Neighborhoods," *arXiv:1509.02868 [physics.soc-ph]*, pp. 1–17, 2015.

[29] V. C. Yang, A. V. Papachristos, and D. M. Abrams, "Modeling the origin of urban-output scaling laws," *Physical Review E*, vol. 100, no. 3, p. 32306, 2019.

[30] M. Schläpfer, L. M. a. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West, and C. Ratti, "The scaling of human interactions with city size.," *Journal of the Royal Society, Interface / the Royal Society*, vol. 11, no. 98, pp. 20130789–, 2014.

[31] A. T. Philbrick, "Short History of the Development of the Gravity Model.," *Aust Road Res*, vol. 5, no. 4, pp. 40–54, 1973.

[32] K. E. Haynes and A. S. Fotheringham, *Gravity and Spatial Interaction Models*. Morgantown: Regional Research Institute, West Virginia University, 1985.

[33] M. Barthelemy, "The statistical physics of cities," *Nature Reviews Physics*, vol. 1, no. 6, pp. 406–415, 2019.

[34] M. BARTHELEMY, *THE STRUCTURE AND DYNAMICS OF CITIES*. Cambridge Univ. Press,, 2016.

[35] M. Granovetter, "The Strength Of Weak Ties," *Am. J. Sociol.*, vol. 78, no. 1360, 1973.

[36] S. Arbesman, J. M. Kleinberg, and S. H. Strogatz, "Superlinear scaling for innovation in cities," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 79, no. 1, pp. 1–5, 2009.

[37] K. Rajkumar, G. Saint-Jacques, I. Bojinov, E. Brynjolfsson, and S. Aral, "A causal test of the strength of weak ties," *Science (New York, N.Y.)*, vol. 377, no. 6612, pp. 1304–1310, 2022.

[38] J. Goldenberg and M. Levy, "Distance Is Not Dead: Social Interaction and Geographical Distance in the Internet Era," *ArXiv*, 2009.

[39] C. Herrera-Yagüe, C. M. Schneider, T. Couronné, Z. Smoreda, R. M. Benito, P. J. Zufiria, and M. C. González, "The anatomy of urban social networks and its implications in the searchability problem," *Scientific Reports*, pp. 1–27, 2015.

[40] S. Scellato, R. Lambiotte, and C. Mascolo, "Socio-spatial Properties of Online Location-based Social Networks," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2011.

[41] R. Li, L. Dong, J. Zhang, X. Wang, W. X. Wang, Z. Di, and H. E. Stanley, "Simple spatial scaling rules behind complex cities," *Nature Communications*, vol. 8, no. 1, pp. 1–7, 2017.

[42] V. M. Pérez-García, G. F. Calvo, J. J. Bosque, O. León-Triana, J. Jiménez, J. Pérez-Beteta, J. Belmonte-Beitia, M. Valiente, L. Zhu, P. García-Gómez, P. Sánchez-Gómez, E. Hernández-San Miguel, R. Hortigüela, Y. Azimzade, D. Molina-García, Á. Martínez, Á. Acosta Rojas, A. Ortiz de Mendivil, F. Vallette, P. Schucht, M. Murek, M. Pérez-Cano, D. Albillo, A. F. Honguero Martínez, G. A. Jiménez Londoño, E. Arana, and A. M. García Vicente, "Universal scaling laws rule explosive growth in human cancers," *Nature Physics*, vol. 16, no. 12, pp. 1232–1237, 2020.

[43] F. L. Ribeiro, R. V. Dos Santos, and A. S. Mata, "Fractal dimension and universality in avascular tumor growth," *Physical Review E*, vol. 95, no. 4, pp. 1–9, 2017.

[44] A. Cliff, R. Martin, and J. Ord, "Evaluating the friction of distance parameter in gravity models," *Regional Studies*, vol. 8, no. 3-4, pp. 281–286, 1974.

[45] H. Couclelis, "Editorial," *Environment and Planning B: Planning and Design*, vol. 23, pp. 387–389, 1996.

[46] B. Lengyel, A. Varga, B. S??gv??ri, k. Jakobi, and J. Kert??sz, "Geographies of an online social network," *PLoS ONE*, vol. 10, no. 9, pp. 1–13, 2015.

[47] K. Yakubo and D. Koroak, "Scale-free networks embedded in fractal space," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 83, no. 6, pp. 1–11, 2011.

[48] K. Yakubo and Y. Saijo, "Superlinear and sublinear urban scaling in geographical networks modeling cities," *Phys. Rev. E*, vol. 022803, pp. 1–10, 2014.

[49] A. Bunde and S. Havlin, *Fractals in Science*, ch. 1, pp. 1–25. Berlin: Springer-Verlag, 1995.

[50] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins, "Geographic routing in social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 33, pp. 11623–11628, 2005.

[51] S. A. Stouffer, "INTERVENING OPPORTUNITIES: A THEORY RELATING MOBILITY AND DISTANCE," *American Sociological Review*, vol. 5, no. 6, pp. 845–867, 1940.

[52] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: Universal patterns in human urban mobility," *PLoS ONE*, vol. 7, no. 5, 2012.

[53] C. Molinero and S. Thurner, "How the geometry of cities explains urban scaling laws and determines their exponents," *Interface*, vol. 18, 2021.

[54] F. Simini, M. C. González, A. Maritan, and A. L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, no. 7392, pp. 96–100, 2012.

[55] I. Hong, W. S. Jung, and H. H. Jo, "Gravity model explained by the radiation model on a population landscape," *PLoS ONE*, vol. 14, no. 6, pp. 1–13, 2019.

[56] G. Spadon, A. C. de Carvalho, J. F. Rodrigues-Jr, and L. G. Alves, "Reconstructing commuters network using machine learning and urban indicators," *Scientific Reports*, vol. 9, no. 1, pp. 1–13, 2019.

[57] M. Schläpfer, "The hidden universality of movement in cities," *arXiv preprint*, 2020.

[58] L. Dong, Z. Huang, J. Zhang, and Y. Liu, "Understanding the mesoscopic scaling patterns within cities," *Scientific Reports*, vol. 10, pp. 1–12, 2020.

[59] W. Pan, G. Ghoshal, C. Krumme, M. Cebrian, and A. Pentland, "Urban characteristics attributable to density-driven tie formation.," *Nature communications*, vol. 4, p. 1961, 2013.

[60] C. R. Shalizi, "Scaling and Hierarchy in Urban Economies," *arXiv preprint*, vol. I, no. 1, pp. 1–15, 2011.

[61] W. R. Tobler, "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, vol. 46, no. 2, pp. 234–240, 1970.

[62] Z. Lu, H. Zhang, F. Southworth, and J. Crittenden, "Fractal dimensions of metropolitan area road networks and the impacts on the urban built environment," *Ecological Indicators*, vol. 70, no. November, pp. 285–296, 2016.

[63] A. Carbone, P. Murialdo, A. Pieroni, and C. Toxqui-Quitl, "Atlas of urban scaling laws," *Journal of Physics: Complexity*, vol. 3, no. 2, 2022.

[64] B. J. E. Geoffrey B. West, James H. Brown, "A General Model for the Origin of Allometric Scaling Laws in Biology," *Science*, vol. 276, pp. 122–126, apr 1997.

[65] G. B. West, "The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms," *Science*, vol. 284, pp. 1677–1679, jun 1999.

[66] G. B. West and J. H. Brown, "Life's Universal Scaling Laws," *Physics Today*, no. September, 2004.

[67] F. L. Ribeiro and W. R. Pereira, "a Gentle Introduction To Scaling Relations in Biological Systems," *Revista Brasileira de Ensino de Fisica*, vol. 44, 2022.

[68] R. Louf and M. Barthelemy, "Modeling the polycentric transition of cities," *Physical Review Letters*, vol. 111, no. 19, 2013.

[69] R. Louf and M. Barthelemy, "How congestion shapes cities: From mobility patterns to scaling," *Scientific Reports*, vol. 4, pp. 1–9, 2014.

[70] M. Fujita and H. Ogawa, "Multiple equilibria and structural transition of non-monocentric urban configurations," *Regional Science and Urban Economics*, vol. 12, no. 2, pp. 161–196, 1982.

[71] H. V. Ribeiro, D. Rybski, and J. P. Kropp, "Effects of changing population or density on urban carbon dioxide emissions," *Nat. Commun.*, vol. 10, p. 3204, 2019.

[72] C. W. Cobb and P. H. Douglas, "A theory of production," *American Economic Review*, vol. 18, no. 1, pp. 139–165, 1928.

[73] S. W. David F. Heathfield, *An Introduction to Cost and Production Functions*. London: Macmillan, 1987.

[74] A. Gomez-Lievano, O. Patterson-Lomba, and R. Hausmann, "Explaining the Prevalence, Scaling and Variance of Urban Phenomena," *Nature Human Behaviour*, vol. 1, no. 0012, pp. 1–6, 2016.

[75] A. Gomez-Lievano and O. Patterson-Lomba, "The drivers of urban economic complexity and their connection to urban economic performance," *ArXiv*, 2018.

[76] R. Hausmann and C. A. Hidalgo, "The network structure of economic output," *Journal of Economic Growth*, vol. 16, no. 4, pp. 309–342, 2011.

[77] J. Henrich, "Demography and Cultural Evolution: How Adaptive Cultural Processes can Produce Maladaptive Losses: The Tasmanian Case," *American Antiquity*, vol. 69, no. 2, pp. 197–214, 2004.

[78] A. Gómez-Liévano, V. Vysotsky, and J. Lobo, "Artificial increasing returns to scale and the problem of sampling from lognormals," *Environment and Planning B: Urban Analytics and City Science*, vol. 48, no. 6, pp. 1574–1590, 2021.

[79] J. Eeckhout, R. Pinheiro, and K. Schmidheiny, "Spatial sorting," *Journal of Political Economy*, vol. 122, no. 3, pp. 554–620, 2014.

[80] P. Combes and et al. G Duranton , Gobillon L, "The productivity advantages of large cities: Distinguishing agglomeration from firm selection.," *Econometrica*, vol. 80, no. 6, pp. 2543–2594, 2012.

[81] S. Coles, *An Introduction to Statistical Modeling of Extreme Values, Ser. in Stat.* London: Springer, 2001.

[82] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer, 1983.

[83] A. S. Alderson, J. Beckfield, and J. Sprague-Jones, "Intercity relations and globalisation: The evolution of the global urban hierarchy, 1981—2007," *Urban Stud.*, vol. 47, no. 9, pp. 1899–1923, 2010.

[84] R. P. Curiel, L. Pappalardo, L. Gabrielli, and S. R. Bishop, "Gravity and scaling laws of city to city migration," *PLoS ONE*, vol. 13, no. 7, pp. 1–19, 2018.

[85] L. J. Pumain D, Paulus F, Vacchiani-Marcuzzo C, "An evolutionary theory for interpreting urban scaling laws," *Cybergeo: European Journal of Geography;*, 2006.

[86] F. Auerbach, "Das Gesetz der Bevölkerungskonzentration," *Petermanns Geogr. Mitteilungen*, vol. 59, no. 74, pp. 73–76, 1913.

[87] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Manfield Centre, CT: Martino Publishing, 2012. (Reprint of 1949 Edition).

[88] B. J. L. Berry and A. Okulicz-Kozaryn, "The city size distribution debate: Resolution for US urban regions and megalopolitan areas," *Cities*, vol. 29, no. SI1, pp. S17–S23, 2012.

[89] D. Rybski, "Auerbach's legacy," *Environ. Plan. A*, vol. 45, no. 6, pp. 1266–1268, 2013.

[90] V. Nitsch, "Zipf zipped," *J. Urban. Econ.*, vol. 57, no. 1, pp. 86–100, 2005.

[91] K. T. Soo, "Zipf's law for cities: a cross-country investigation," *Reg. Sci. Urban. Econ.*, vol. 35, no. 3, pp. 239–263, 2005.

[92] C. Cottineau, "MetaZipf. A dynamic meta-analysis of city size distributions," *PLoS One*, vol. 12, no. 8, p. e0183919, 2017.

[93] H. D. Rozenfeld, D. Rybski, X. Gabaix, and H. A. Makse, "The area and population of cities: New insights from a different perspective on cities," *Am. Econ. Rev.*, vol. 101, no. 5, pp. 2205–2225, 2011.

[94] X. Gabaix, "Zipf's Law for Cities: An explanation," *The Quarterly Journal of Economics*, vol. 114, no. 3, pp. 739–767, 2009.

[95] A. Gomez-Lievano, H. J. Youn, and L. M. Bettencourt, "The statistics of urban scaling and their connection to Zipf's law," *PLoS ONE*, vol. 7, no. 7, 2012.

[96] B. J. L. Berry and A. Okulicz-Kozaryn, "The city size distribution debate: Resolution for US urban regions and megalopolitan areas," *Cities*, vol. 29, no. SUPPL. 1, pp. S17–S23, 2012.

[97] H. V. Ribeiro, M. Oehlers, A. I. Moreno-monroy, P. Kropp, and D. Rybski, "Effects of population distribution on urban scaling," *PLoS ONE*, vol. 16, no. 1, 2021.

[98] F. L. Ribeiro, J. Lobo, and D. Rybski, "Zipf's law and urban scaling: Hypotheses towards a Unified Urban Theory," *arXiv preprint*, pp. 1–3, 2021.

[99] E. G. Altmann, "Spatial interactions in urban scaling laws," *PloS one*, vol. 15, no. 12, pp. 1–12, 2020.

[100] J. C. Leitão, J. M. Miotto, M. Gerlach, and E. G. Altmann, "Is this scaling nonlinear?," *Royal Society Open Science*, vol. 3, 2016.

[101] G. D. Nelson and A. Rae, "An economic geography of the United States: From commutes to megaregions," *PLoS ONE*, vol. 11, no. 11, pp. 1–23, 2016.

[102] L. G. Alves, D. Rybski, and H. V. Ribeiro, "Commuting network effect on urban wealth scaling," *Scientific Reports*, vol. 11, no. 1, pp. 1–10, 2021.

[103] L. Alessandretti, P. Sapiezynski, S. Lehmann, and A. Baronchelli, "Evidence for a Conserved Quantity in Human Mobility," *Nature Human Behaviour*, vol. 2, pp. 485–491, 2018.

[104] M. Schläpfer, L. Dong, K. O'Keeffe, P. Santi, M. Szell, H. Salat, S. Anklesaria, M. Vazifeh, C. Ratti, and G. B. West, "The universal visitation law of human mobility," *Nature*, vol. 593, no. 7860, pp. 522–527, 2021.

[105] L. Alessandretti, "What human mobility data tell us about COVID-19 spread," *Nature Reviews Physics*, 2021.

[106] D. Rybski, A. G. C. Ros, and J. P. Kropp, "Distance-weighted city growth," *Phys. Rev. E*, vol. 87, 2013.

[107] Peter Brand and J. D. Dávila, "Mobility innovation at the urban margins Medellín's Metrocables," *City*, vol. 15, no. 6, 2011.

[108] D. Rybski, A. G. C. Ros, and J. P. Kropp, "Distance weighted city growth," *arXiv preprint*, 2012.

[109] Y. Li, D. Rybski, and J. P. Kropp, "Singularity cities," *Environment and Planning B: Urban Analytics and City Science*, vol. 0, no. 0, pp. 1–17, 2019.

[110] V. M. Netto, E. Brigatti, and C. Cacholas, "From urban form to information: Cellular configurations in different spatial cultures," *Environment and Planning B: Urban Analytics and City Science*, vol. 50, no. 1, pp. 146–161, 2023.

[111] E. Brigatti, V. M. Netto, F. N. De Sousa Filho, and C. Cacholas, "Entropy and hierarchical clustering: Characterizing the morphology of the urban fabric in different spatial cultures," *Chaos*, vol. 31, no. 11, 2021.

[112] S. Encarnação, M. Gaudiano, F. C. Santos, J. A. Tenedório, and J. M. Pacheco, "Fractal cartography of urban areas," *Scientific Reports*, vol. 2, pp. 1–5, 2012.

[113] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty, "Constructing cities, deconstructing scaling laws," *Journal of The Royal Society Interface*, vol. 12, no. 102, p. 20140745, 2015.

[114] C. Cottineau, E. Hatna, E. Arcaute, and M. Batty, "Diverse cities or the systematic paradox of urban scaling laws," *Computers, environment and urban systems*, vol. 63, pp. 80–94, 2017.

[115] R. Louf and M. Barthelemy, "Scaling: lost in the smog," *Environ. Plan. B*, vol. 41, no. 5, pp. 767–769, 2014.

[116] G. J. Babu and E. D. Feigelson, "Analytical and monte carlo comparisons of six different linear least squares fits," *Communications in Statistics-Simulation and Computation*, vol. 21, no. 2, pp. 533–549, 1992.

[117] S. Openshaw and P. J. Taylor, *Statistical applications in spatial sciences*, ch. A million or so correlation coefficients: three experiments on the modifiable areal unit problem, pp. 127–144. London: Pion, 1979.

[118] L. M. Bettencourt, V. C. Yang, J. Lobo, C. P. Kempes, D. Rybski, and M. J. Hamilton, "The interpretation of urban scaling analysis in time," *Journal of the Royal Society, Interface*, vol. 17, no. 163, p. 20190846, 2020.
[119] F. L. Ribeiro, J. Meirelles, V. M. Netto, C. R. Neto, and A. Baronchelli, "On the relation between Transversal and Longitudinal Scaling in Cities," *PLoS ONE*, pp. 1–20, 2020.
[120] V. Verbavatz and M. Barthelemy, "The growth equation of cities," *Nature*, vol. 587, no. 7834, pp. 397–401, 2020.
[121] W. Christaller, *Central Places in Southern Germany*. London UK: Prentice-Hall International, Inc., 1966.
[122] B. J. L. Berry and W. L. Garrison, "Alternate explanations of urban rank-size relationships," *Annals of the Association of American Geographers*, vol. 48, no. 1, pp. 83–90, 1958.
[123] L. M. A. Bettencourt, J. Lobo, D. Strumsky, and G. B. West, "Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities," *PLoS ONE*, vol. 5, no. 11, pp. 20–22, 2010.
[124] R. Gudipudi, T. Fluschnik, A. G. C. Ros, C. Walther, and J. P. Kropp, "City density and CO2 efficiency," *Energy Policy*, vol. 91, pp. 352–361, 2016.
[125] O. design of spatial distribution networks, "Gastner, m. t. and newman, m. e. j.," *Phys. Rev. E*, vol. 74, no. 1, p. 016117, 2006.
[126] J. Um, S. W. Son, S. I. Lee, H. Jeong, and J. K. Beom, "Scaling laws between population and facility densities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 34, pp. 14236–14240, 2009.
[127] W. J. Reilly, *The Law of Retail Gravitation*. 1931.
[128] F. L. Ribeiro, "Física das Cidades," *Revista de Morfologia Urbana*, vol. 8, no. 1, p. e00159, 2020.