

Distance-weighted city growth

Diego Rybski* and Anselmo García Cantú Ros

Potsdam Institute for Climate Impact Research–14412 Potsdam, Germany, EU

Jürgen P. Kropp

*Potsdam Institute for Climate Impact Research–14412 Potsdam, Germany, EU and**Institute of Earth and Environmental Science, University of Potsdam–14476 Potsdam, Germany, EU*

(Received 17 September 2012; revised manuscript received 7 February 2013; published 16 April 2013)

Urban agglomerations exhibit complex emergent features of which Zipf's law, i.e., a power-law size distribution, and fractality may be regarded as the most prominent ones. We propose a simplistic model for the generation of citylike structures which is solely based on the assumption that growth is more likely to take place close to inhabited space. The model involves one parameter which is an exponent determining how strongly the attraction decays with the distance. In addition, the model is run iteratively so that existing clusters can grow (together) and new ones can emerge. The model is capable of reproducing the size distribution and the fractality of the boundary of the largest cluster. Although the power-law distribution depends on both, the imposed exponent and the iteration, the fractality seems to be independent of the former and only depends on the latter. Analyzing land-cover data, we estimate the parameter-value $\gamma \approx 2.5$ for Paris and its surroundings.

DOI: [10.1103/PhysRevE.87.042114](https://doi.org/10.1103/PhysRevE.87.042114)

PACS number(s): 64.60.al, 61.43.-j, 89.65.Lm

I. INTRODUCTION

Cities or urban agglomerations exhibit signatures of complex phenomena, such as broad size distributions [1–5] and fractal structure [6,7] (and references therein). The last decades have witnessed a strong interest within the scientific community in characterizing the worldwide urbanization phenomenon. This line of research has strongly benefited from accessibility of demographic databases and from application of tools originated in statistical physics enabling the identification and analysis of universal aspects of urban forms and scaling features [8]. Beyond the descriptive level, various attempts to obtain insight into mechanisms that underly the complex features of cities have been proposed.

(i) Multiplicative models [9–12] have explored the connection between random city growth and city size distributions. In particular, building on discrete random walk theory, multiplicative models have proved successful at reproducing Zipf's law (i.e., power-law city size distribution with exponent close to 2). Furthermore, some of these models have proposed plausible explanations for the origin of these mechanisms, based on spatial economics theory [10]. Notwithstanding this fact, multiplicative models are space independent and, thus, are unable to address other important features of city structures, such as self-similarity. (ii) Approaches based on cellular automata have been used to model spatial structure of urban land use over time [13] reproducing fractal properties. (iii) The correlated percolation model (CPM) [14,15] assumes that an urban built environment is shaped by spatial correlations where the occupation probabilities of two sites are more similar the closer they are. The model involves the empirical findings on the radial decay of density around a city center. For certain ranges in the space of parameters, the CPM reproduces basic features of real urban aggregates, such as broad size distributions in urban clusters and the fractal scaling of the

perimeter. (iv) Reaction diffusion models [16–18] have been introduced in order to explore the role of intermittency in creating spatial inhomogeneities in agreement with Zipf's law. (v) Spatial explicit preferential attachment has been shown to be capable of reproducing Zipf's law [19]. Here, the probability that a city grows is essentially assumed to be proportional to the size of the city. (vi) Agent based modeling has been employed to simulate urban growth [20], reproducing the formation of new clusters as well as the merging of neighboring ones. (vii) A random group formation is presented in Ref. [21] from which a Bayesian estimate is obtained based on minimal information. It represents a general approach for power-law distributed group sizes.

Although the term *demographic gravitation* was coined by Stewart [22], in geographical economics, gravitational models have been investigated for many decades. Carrothers [23] provides a review of gravity and potential concepts of human interaction. The so-called Reilly's law of retail gravitation describes the breaking point of the boundary of equal attraction [24]. Similarly, Huff's law of shopper attraction [25] provides the probability of an agent at a given site to travel to a particular facility. Last but not least, the volume of trade between countries has been described from the point of view of gravity analogy [26]. In contrast, limitation of gravitational models has been pointed out in the context of mobility and migration [27].

Following the first law of geography “Everything is related to everything else, but near things are more related than distant things” [28], we elaborate on the role of gravity effects in shaping the most salient universal features of cities, namely, size distribution and fractality. To this end, we introduce a model where individual lattice sites of a grid are more likely to be occupied the closer they are to already occupied sites. We find that the cluster sizes follow Zipf's law except for the largest cluster which outgrows Zipf's law, i.e., the largest cluster is too big and can be considered as the central business district [14]. Applying box counting [29], we find self-similarity of the largest cluster boundary, whereas, the fractal

*ca-dr@rybski.de

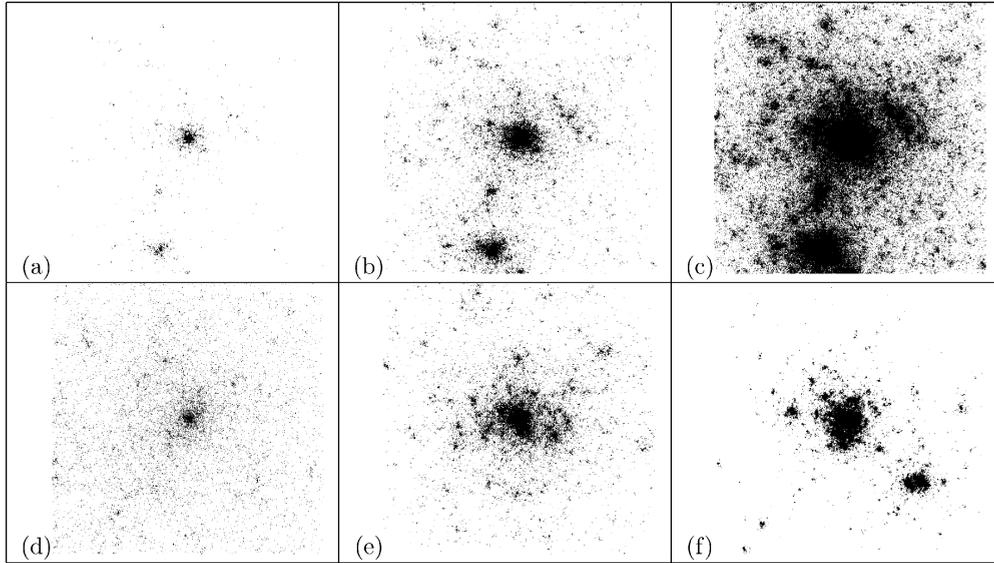


FIG. 1. Illustrative examples of model realizations for different iterations i and different exponents γ . (a)–(c) Different iterations of the model $i = 6, 10, 14$ ($\gamma = 2.5$, $N = 630$). Growth takes place close to occupied sites. (d)–(f) Realizations with different exponents $\gamma = 2.0, 2.5, 3.0$ (the occupation probability is $p \approx 0.04$, $N = 630$). The smaller γ , the more scattered the emerging structures are, the larger γ , the more compact they are.

exponent seems to be independent of the chosen exponent. Despite being very simple, our model intrinsically features radial symmetry, as in (iii), and preferential attachment, as in (v).

II. MODEL

We consider a two dimensional square lattice of size $N \times N$ whose sites w_j with coordinates $j \in \{(1 \cdots N, 1 \cdots N)\}$ can be either empty or occupied. We start with an empty grid ($w_j = 0$ for all j) and, without loss of generality, set the single central site as occupied ($w_j = 1$, $j = (N/2, N/2)$ for even N , $j = [(N+1)/2, (N+1)/2]$ for odd N). Then, the probability that the sites will be occupied is

$$q_j = C \frac{\sum_{k \neq j} w_k d_{j,k}^{-\gamma}}{\sum_{k \neq j} d_{j,k}^{-\gamma}}, \quad (1)$$

where $d_{j,k}$ is the Euclidean distance between sites j and k . The proportionality constant C is determined by normalization, i.e., $C = 1/\max(q_j)$ so that the maximum probability is 1. The exponent $\gamma > 0$ is a free parameter that determines how strong the influence of occupied sites decays with the distance. This model is inspired by Ref. [25] where the probability that a site will be occupied is solely determined by the distance to already occupied sites.

It is apparent that only sites within close proximity of the initially occupied site are likely to be occupied, whereas, distant sites mostly remain empty. The procedure is then iterated by repeating the process, involving recalculation of Eq. (1) for each step. Note that a different choice of C would only influence how many iterations are needed to completely fill the lattice.

III. ANALYSIS

The model output depends on a set of factors. Beyond the exponent γ , the system size $N \times N$ needs to be chosen. As the model works iteratively, the emerging structures can be investigated at different iterations i . Moreover, we run the model for M realizations in order to obtain better statistics.

Figure 1 shows examples of model realizations. Visually, the emerging structures could be associated with urban space. Figures 1(a)–1(c) show three iterations of a single realization. For high values of γ , the spatiotemporal evolution is strongly influenced by the sites which are occupied early, see Figs. 1(d)–1(f). Such a path dependency is also reflected in the reduction of rotational symmetry observed for increasing values of γ . In particular, the larger γ is chosen, the more compact and less scattered the obtained structures are. Large γ also leads to slower filling of the lattice.

A. Cluster size distribution

We begin our analysis by studying the cluster size distribution. We employ the city clustering algorithm [4,30] and find that the largest cluster is markedly larger than the remaining ones [Fig. 2(a)], i.e., larger than expected from Zipf's law. The presence of such anomalous extremes in size distributions is denoted as *Dragon Kings* and can be a signature of strongly cooperative dynamics [31]. A similar effect has been found in another model [19], where—in order to avoid their appearance—the domination of the largest cluster is inhibited by excluding it from proportionate growth. Exclusion is not feasible in our model, and we omit it when studying the cluster size distribution [14,15].

Figure 2(a) shows examples of the probability density $P(S)$ of the cluster size S disregarding the largest cluster of each realization. We find approximate power laws according to

$$P(S) \sim S^{-\zeta}, \quad (2)$$

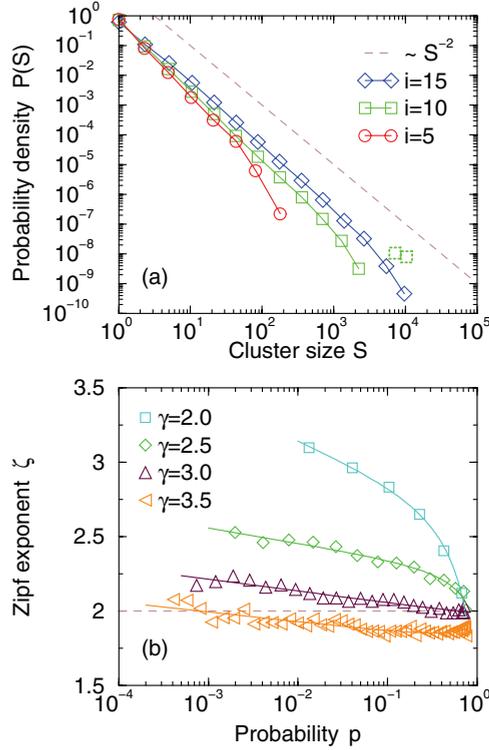


FIG. 2. (Color online) Cluster size distribution and dependence of the Zipf exponent on the occupation probability. (a) Examples of probability density distributions $P(S)$ of the cluster size S disregarding the largest cluster of each realization: $i = 5, 10, 15$ (from top to bottom, $p \approx 0.004, 0.078, 0.573$), $\gamma = 2.5$, $N = 630$, and $M = 100$ (solid lines connect symbols). The two green dotted squares represent the contribution of the largest cluster ($i = 10$). (b) The obtained Zipf exponents ζ as a function of the occupation probability p : $\gamma = 2, 2.5, 3, 3.5$ (from top to bottom), $N = 630$, $M = 100$. Solid lines are fits according to Eq. (3). Dashed lines indicate $\zeta = 2$.

where ζ is the Zipf exponent. In Fig. 2(a), one can see deviations from Eq. (2) in the form of too few large clusters. Naturally, for late iterations, Eq. (2) extends over more decades of cluster size than for early iterations. As can be seen, the Zipf exponent is close to 2. To be more precise, ζ is smaller for large iteration i than for small i . Accepting minor deviations from $\zeta = 2$, the model produces cluster size distributions compatible with Zipf's law.

In order to better understand how ζ relates to the model parameters, we express the iteration i in terms of the overall occupation probability p , which for a given i , is defined by the number of occupied sites divided by the total number of sites, i.e., $N \times N$. The probability p increases monotonically with the iteration i . In Fig. 2(b), ζ is plotted as a function of p . As can be seen, it decreases monotonically with increasing probability and strongly depends on the model exponent γ . Whereas, for $\gamma < 3$, convex $\zeta(p)$'s are found with overall $\zeta > 2$, for $\gamma = 3$, an almost logarithmic form can be identified with $\zeta > 2$ and $\zeta \rightarrow 2$ for $p \rightarrow 1$. In contrast, for $\gamma > 3$, we see a slightly concave relation and $\zeta < 2$ (except for small p). Accordingly, the Zipf exponent depends strongly on both, the model exponent γ and the iteration of model i .

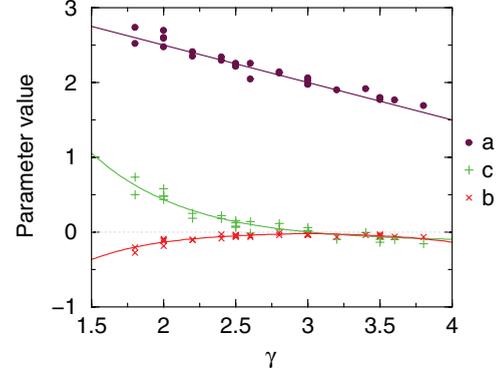


FIG. 3. (Color online) Parameter values a , b , and c as a function of the exponent γ from fitting the Zipf exponent versus the occupation probability [Fig. 2(b)]. The dots represent the parameters obtained from fitting Eq. (3) to the numerical values, the solid lines are given by $a = -k_1\gamma + k_2$, $b_{\gamma \leq 3} = -e^{-k_3\gamma + k_4}$, $b_{\gamma > 3} = -e^{k_3\gamma - k_5}$, and $c = e^{-k_6\gamma + k_7} - e^{-k_8 + k_7}$ (k_1-k_8 are fitting parameters). In order to have enough values, we do not separate the different system sizes.

Moreover, we find that the dependence of ζ on p can be well approximated by

$$\zeta(p) = a + b \ln(p) + c \ln(1 - p), \quad (3)$$

which also is related to the logarithm of a β distribution. The solid lines in Fig. 2(b) are nonlinear fits to the numerical model results, providing the fit parameters a , b , and c . In Fig. 3, the obtained values of these parameters are plotted against γ .

B. Fractality

Next, we analyze fractal properties of the *urban envelope* of the largest cluster. Therefore, we first extract the boundary of the cluster. This is performed by identifying those largest cluster sites which have, at least, one empty neighboring cell which connects to the system border via a nearest neighbor path of empty sites (the latter condition is necessary to exclude inclusions). Thus, here, the boundary is defined as the occupied neighbors of the external perimeter [32]. Then, we apply box counting, i.e., perform coarse graining and count how many sites are occupied. Thus, we regularly group $m \times m$ sites and accordingly reduce the system size to $(N \times N)/(m \times m)$. Finally, we count the number of occupied sites N_B for a chosen box size m .

Examples of $N_B(m)$ are displayed in Fig. 4(a). Apart from minor deviations for small and large m , straight lines are found in the log-log representation, corresponding to

$$N_B(m) \sim m^{-d_B}, \quad (4)$$

where d_B is a measure of the fractal dimension of the cluster boundary. In the displayed examples, we approximately find $d_B \approx 1.25$.

Figure 4(b) shows d_B as a function of the occupation probability p . Qualitatively, we find a logarithmic dependence, implying $d_B \rightarrow \text{const.}$ for $p \rightarrow 1$, which seems to be independent of γ .

Overall, the values are below those expected from uncorrelated percolation slightly above or below the percolation transition [33]. The fractal dimension for the perimeter in

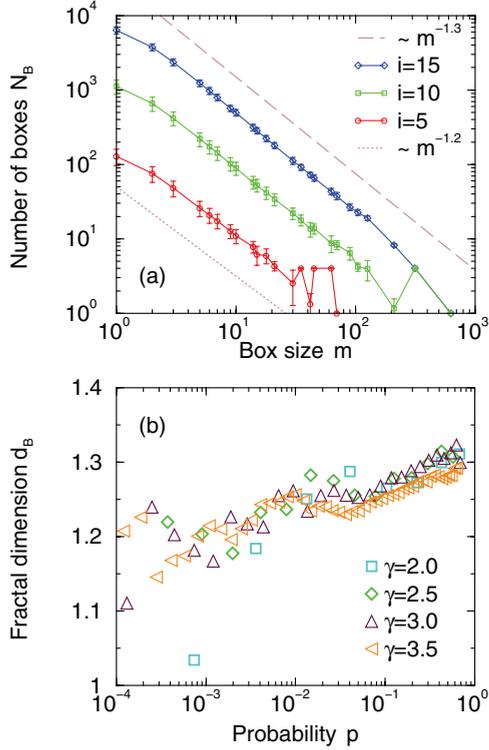


FIG. 4. (Color online) Self-similar scaling of the boundary of the largest cluster. (a) Number of boxes necessary to cover the boundary as a function of the size of the boxes: $\gamma = 2.5$, $N = 630$, $i = 5, 10, 15$ ($p \approx 0.004, 0.078, 0.573$), and $M = 100$. The results follow Eq. (4) indicating the fractal property of the boundary with a fractal dimension $1 < d_B < 2$. The error bars represent standard deviations among the realizations. (b) Fractal dimension as a function of the occupation probability for $\gamma = 2.0, 2.5, 3.0, 3.5$ ($N = 630$ and $M = 100$). d_B increases approximately logarithmically, independent of γ .

uncorrelated percolation is $4/3$. This difference could be due to inherent correlations in our model. Nevertheless, the model generates self-similar (fractal) largest clusters. The evolving fractal dimension is, at least, qualitatively consistent with urban areas, see, e.g., Ref. [34].

Last, we would like to note that the definition of the boundary has a substantial influence on the fractal dimension [32]. Moreover, box-counting results can differ from those obtained with other techniques, such as the equipaced polygon method [35]. Further analysis is required to shed light on these aspects.

C. Percolation transition

One may argue that, at a certain iteration, the system might undergo a percolation transition. Thus, finally, we characterize the percolation threshold of the model. Therefore, we calculate the average cluster size disregarding the largest component $\langle S \rangle$ as a function of the occupation probability p . At the percolation transition p_c , the average cluster size exhibits a maximum [36]. Figure 5 depicts $\langle S \rangle(p)$ for some values of γ . A distinct peak can be found which moves to larger p with decreasing γ . For $\gamma < 3$, the maximum becomes less clear, and we cannot determine p_c .

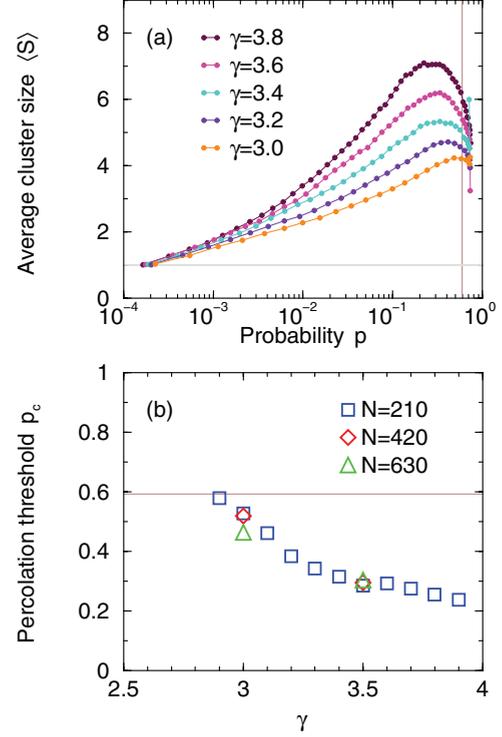


FIG. 5. (Color online) Percolation transition of the model. (a) Average size of the *finite* clusters: $\gamma = 3.8, 3.6, 3.4, 3.2, 3.0$ (from top to bottom), $N = 210$, and $M = 1000$. The vertical line represents the percolation transition of uncorrelated percolation $p_c^* \simeq 0.593$ (site percolation in the square lattice, [36]). The maximum is located at the percolation transition. (b) Percolation transition p_c as a function of γ : $N = 210, 420, 630$ ($M = 1000, 400, 100$). For small γ , the percolation transition is close to p_c^* as indicated by the horizontal line.

The obtained percolation thresholds are plotted versus γ in Fig. 5(b). The transition decreases monotonically with increasing γ . For $\gamma \approx 3$, the value is close to the transition of uncorrelated site percolation in the square lattice ($p_c^* \approx 0.593$, [36]). For $\gamma \approx 4$, we find $p_c \approx 0.2$.

We would like to note that the results of Zipf and fractality analysis seem to be independent from percolation transition, i.e., there is no change in the behavior below or above p_c . Accordingly, scaling in the form of Zipf's law and fractality is reproduced even away from criticality.

IV. ANALYZING REAL DATA

Finally, it is of interest which γ -value real city growth exhibits. In order to address this question, we consider Paris and its surroundings. We analyze CORINE [37] land-cover data in 250 m resolution and only distinguish between urban and nonurban land grid cells. For the years 2000 and 2006, we extract a window of 1000×1000 grid points [Fig. 6(a)] and study the land-cover change. Since our model only includes growth, we disregard those cells which change from urban to nonurban.

First, we calculate the probabilities q_j according to Eq. (1) for the year 2000 with $w_k = 1$ for urban and $w_k = 0$ for nonurban cells. Then, we determine the logarithmic likelihood

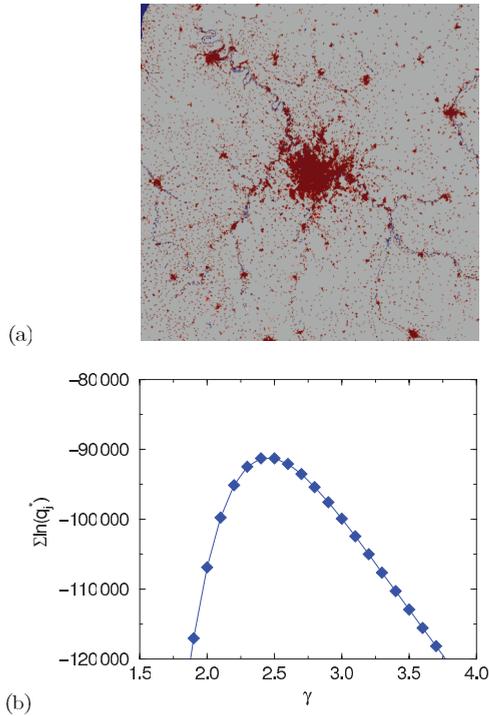


FIG. 6. (Color online) Urban growth and the estimation of γ for Paris and its surroundings in the years 2000 and 2006. (a) Considered land-cover data in a window of 1000×1000 grid points (250 m resolution). The panel distinguishes the following sites: dark red: populated in 2000; gray: unpopulated in 2000; blue: water; light red: nonurban \rightarrow urban; and yellow: urban \rightarrow nonurban. (b) Logarithmic likelihood of urbanization. First, we calculate the probabilities q_j according to Eq. (1) for the year 2000, whereas ($w_k = 1$ for urban and $w_k = 0$ for nonurban cells). Then, we determine Q according to Eq. (5) over all nonurban cells in 2000 (the change urban to nonurban is disregarded). We find a maximum at $\gamma_{\text{Paris}} \approx 2.5$.

by summing over all nonurban cells in 2000,

$$Q = \sum_{j \in A} \ln q_j + \sum_{j \in B} \ln(1 - q_j), \quad (5)$$

where A is the set of cells changing from nonurban to urban and B is the set of cells remaining nonurban. The quantity Q can be understood as the logarithm of the joint probability of obtaining the observed urbanization, i.e., the product of probabilities that the cells become either urbanized or not, respectively.

Varying γ , we can identify the value for which Q is maximized, i.e., for which the probabilities calculated with Eq. (1) best represent the nonurban to urban land-cover change in the real data. As can be seen in Fig. 6(b), the maximum is located at $\gamma_{\text{Paris}} \approx 2.5$. The qualitative similarity between

Figs. 6(a) and 1(e) supports this quantitative result, but the comparison also shows that the real example is richer in structure.

One may ask if the period 2000–2006 could be two or more steps and if this would lead to another exponent γ . However, assuming time independence of γ , the rate of urbanization is determined by the proportionality constant C in Eq. (1). Accordingly, the exponent should not depend on the period between the snapshots. Although the analysis does not provide sufficient evidence to support our model, it leads to the value of γ for which the model best fits the growth of Paris.

V. DISCUSSION

We also find that the growth rate of clusters between two iterations is independent of the cluster size (not shown). This implies proportionate growth, a characteristic which is also featured by preferential attachment [38,39]. We would like to note that, in the proposed model, such a mechanism emerges and is not included explicitly. We further find that the standard deviation in the growth rate decays as a power law with exponent $1/2$ (not shown), which indicates uncorrelated growth [30,40]. Again, analyzing the growth, we have disregarded the largest cluster.

Although the work in hand briefly introduces our model, more research is necessary to characterize it. This includes: (i) an analytical description of the model, (ii) further numerical analysis, in particular, refining the fractal characterization (as mentioned in Sec. III B) or other features, such as the area-perimeter scaling [41], and (iii) relating our model to other physical approaches, such as Refs. [41–43].

The analogy of gravitation has a long history in geography and spatial economy. However, the early papers were limited by scientific background from statistical physics as well as computational power. Here, we reexamine the concept of *gravity cities* by proposing a simple statistical model which generates citylike structures. The emergent complex structures are similar to urban space. On one hand, we find that the largest cluster, which can be considered as the central business district [14], exhibits fractality, consistent with measured urban area. On the other hand, clusters around the largest one can be considered as towns surrounding a large city [14]. Their cluster size distribution is compatible with Zipf's law.

ACKNOWLEDGMENTS

We would like to thank B.F. Prah, B. Zhou, L. Kaack, E. Giese, X. Gabaix, H. D. Rozenfeld, and N. Schwarz for useful discussions and comments. We appreciate financial support by BaltCICA (partly financed by the EU Baltic Sea Region Programme No. 2007-2013).

[1] F. Auerbach, *Petermanns Geogr. Mitteilungen* **59**, 73 (1913).
 [2] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology (Reprint of 1949 ed.)* (Martino, Manfield Centre, CT, 2012).

[3] A. Saichev, Y. Malevergne, and D. Sornette, *Theory of Zipf's Law and Beyond* (Springer, Berlin, 2010).
 [4] H. D. Rozenfeld, D. Rybski, X. Gabaix, and H. A. Makse, *Am. Econ. Rev.* **101**, 560 (2011).

- [5] B. J. L. Berry and A. Okulicz-Kozaryn, *Cities* **29**, S17 (2012).
- [6] M. Batty and P. Longley, *Fractal Cities: A Geometry of Form and Function* (Academic, San Diego/London, 1994).
- [7] M. Batty and P. A. Longley, *Area* **19**, 215 (1987).
- [8] L. Bettencourt and G. West, *Nature (London)* **467**, 912 (2010).
- [9] M. Levy and S. Solomon, *Int. J. Mod. Phys. C* **7**, 595 (1996).
- [10] X. Gabaix, *Q. J. Econ.* **114**, 739 (1999).
- [11] O. Malcai, O. Biham, and S. Solomon, *Phys. Rev. E* **60**, 1299 (1999).
- [12] X. Gabaix, *Annu. Rev. Econ.* **1**, 255 (2009).
- [13] R. White and G. Engelen, *Environ. Plan. A* **25**, 1175 (1993).
- [14] H. A. Makse, H. S., and H. E. Stanley, *Nature (London)* **377**, 608 (1995).
- [15] H. A. Makse, J. S. Andrade, M. Batty, S. Havlin, and H. E. Stanley, *Phys. Rev. E* **58**, 7054 (1998).
- [16] D. H. Zanette and S. C. Manrubia, *Phys. Rev. Lett.* **79**, 523 (1997).
- [17] M. Marsili, S. Maslov, and Y. C. Zhang, *Phys. Rev. Lett.* **80**, 4830 (1998).
- [18] D. H. Zanette and S. C. Manrubia, *Phys. Rev. Lett.* **80**, 4831 (1998).
- [19] F. Schweitzer and J. Steinbrink, *Appl. Geogr.* **18**, 69 (1998).
- [20] F. Schweitzer, *Brownian Agents and Active Particles* (Springer, Berlin, 2003), p. 295.
- [21] S. K. Baek, S. Bernhardsson, and P. Minnhagen, *New J. Phys.* **13**, 043004 (2011).
- [22] J. Q. Stewart, *Sociometry* **11**, 31 (1948).
- [23] G. A. P. Carrothers, *J. Am. I. Planners* **22**, 94 (1956).
- [24] W. J. Reilly, *The Law of Retail Gravitation* (Reilly, New York, 1931).
- [25] D. L. Huff, *Land Econ.* **39**, 81 (1963).
- [26] P. Pöyhönen, *Weltwirtschaftliches Archiv* Bd. **90**, 93 (1963).
- [27] F. Simini, M. C. Gonzalez, A. Maritan, and A.-L. Barabási, *Nature (London)* **484**, 96 (2012).
- [28] W. R. Tobler, *Econ. Geogr.* **46**, 234 (1970).
- [29] A. Bunde and S. Havlin, *Fractals in Science* (Springer, Berlin, 1994), p. 1.
- [30] H. D. Rozenfeld, D. Rybski, J. S. Andrade, Jr., M. Batty, H. E. Stanley, and H. A. Makse, *Proc. Natl. Acad. Sci. USA* **105**, 18702 (2008).
- [31] V. F. Pisarenko and D. Sornette, *Eur. Phys. J.-Spec. Top.* **205**, 95 (2012).
- [32] T. Grossman and A. Aharony, *J. Phys. A* **20**, L1193 (1987).
- [33] R. F. Voss, *J. Phys. A* **17**, L373 (1984).
- [34] G. Shen, *Int. J. Geograph. Inf. Sci.* **16**, 419 (2002).
- [35] B. H. Kaye and G. G. Clark, *Part. Part. Syst. Charact.* **2**, 143 (1985).
- [36] A. Bunde and S. Havlin, *Fractal and Disordered Systems* (Springer, Berlin, 1991), p. 51.
- [37] M. Bossard, J. Feranec, and J. Otahel, European Environment Agency, Technical Report No. 40, 2000 (unpublished).
- [38] H. A. Simon, *Biometrika* **42**, 425 (1955).
- [39] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [40] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, *Proc. Natl. Acad. Sci. USA* **106**, 12640 (2009).
- [41] C. Andersson, K. Lindgren, S. Rasmussen, and R. White, *Phys. Rev. E* **66**, 026204 (2002).
- [42] Y. P. Jeon and B. J. McCoy, *Phys. Rev. E* **72**, 037104 (2005).
- [43] S. Ree, *Phys. Rev. E* **73**, 026115 (2006).