Title	Scale-free networks embedded in fractal space
Author(s)	Yakubo, K.; Korošak, D.
Citation	Physical Review E, 83(6), 066111 https://doi.org/10.1103/PhysRevE.83.066111
Issue Date	2011-06
Doc URL	http://hdl.handle.net/2115/46767
Rights	© 2011 American Physical Society
Type	article
File Information	PRE83-6_066111.pdf

Instructions for use

Scale-free networks embedded in fractal space

K. Yakubo[1,*] and D. Korošak[2,3,†]

[1]*Department of Applied Physics, Hokkaido University, Sapporo 060-8628, Japan*
[2]*University of Maribor, Institute of Physiology, Faculty of Medicine, Maribor SI-2000, Slovenia*
[3]*University of Maribor, Faculty of Civil Engineering, Maribor SI-2000, Slovenia*

The impact of an inhomogeneous arrangement of nodes in space on a network organization cannot be neglected in most real-world scale-free networks. Here we propose a model for a geographical network with nodes embedded in a fractal space in which we can tune the network heterogeneity by varying the strength of the spatial embedding. When the nodes in such networks have power-law distributed intrinsic weights, the networks are scale-free with the degree distribution exponent decreasing with increasing fractal dimension if the spatial embedding is strong enough, while the weakly embedded networks are still scale-free but the degree exponent is equal to $\gamma = 2$ regardless of the fractal dimension. We show that this phenomenon is related to the transition from a noncompact to compact phase of the network and that this transition accompanies a drastic change of the network efficiency. We test our analytically derived predictions on the real-world example of networks describing the soil porous architecture.

I. INTRODUCTION

Scale-free organization of networks [1–3] seems to be the underlying common principle of many complex systems. In real-world networks such as the Internet, social networks, or communication networks [4–9], the inhomogeneous arrangement of nodes in space has a strong impact on the network organization, so the linking rules must include the dependence on the distances between nodes [10–14]. In addition to actual Euclidean distances between nodes in real spaces, it has been also pointed out that the existence of underlying *hidden* metric spaces is crucial to understanding many properties of real-world complex networks, such as the scale-free property, the strong clustering nature, the self-similarity, and the navigability of (ultra) small-world networks [15–19]. However, most of the network models studied so far considered either randomly distributed nodes in metric space [20–23] or nodes placed on a lattice [24–28]. The importance of inhomogeneous spatial positions of nodes was emphasized in [4,29] where it was shown that the fractal or self-similar property of node sets is crucial to constrain the Internet models describing the Internet's large-scale topology and its observed scale-free character [30–32] at the router and autonomous system levels. While preferential attachment [2,3] seems to be the main underlying mechanism structuring the Internet, the original form of the preferential attachment [2] should be altered [4,31,33] to account for the observed spatial and/or functional heterogeneity of the nodes.

Here we study the structure of networks formed by a geographical network model in which the nodes with power-law distributed intrinsic weights (i.e., fitnesses [34–36]) are embedded in a fractal space. We show analytically that the networks produced by such a model are scale-free with the degree exponent influenced by the fractal dimension of the embedding space if the spatial embedding is strong

enough. By explicitly deriving the degree and the edge-length distribution functions, we classify these networks into noncompact phase with infinite average degree and average edge length and compact phase with finite average degree and average edge length, separated by an intermediate phase characterized by finite average degree and infinite average edge length. It is also shown that the transition between these phases accompanies a drastic change of the network efficiency. Finally, we use our findings in the analysis of the networks describing the soil porous architecture as an example.

II. MODEL

An inhomogeneous distribution of nodes is realized by placing nodes homogeneously in a fractal space S_D with the fractal dimension D. If D is noninteger, a homogeneous node distribution in S_D implies an inhomogeneous arrangement of nodes in a space S_d with the Euclidean (integer) dimension d ($> D$) embedding S_D. This situation can be easily understood by considering a highly inhomogeneous distribution of nodes in the incipient infinite cluster of a percolation system at the percolation threshold though the node set takes a *homogeneous* fractal structure with a noninteger dimension, as shown by nodes in Fig. 1 where $D = 1.89$ and $d = 2$. The fractal space S_D is assumed to be large but finite with linear size L. In our analytical arguments we assume furthermore that the space S_D is isotropic from every point in S_D to eliminate irrelevant boundary effects. The number of nodes is then given by $N = \rho\Omega \int_0^L l^{D-1}\, dl$, where ρ is the density of nodes and Ω is the surface area of the D-dimensional unit sphere (the D-dimensional total solid angle), that is, $\Omega = 2\pi^{D/2}/\Gamma(D/2)$ with the gamma function $\Gamma(x)$.

Each node has a real and continuous fitness x randomly assigned according to the probability distribution function $s(x)$. If a node pair (i-j) satisfies

$$\frac{F(x_i, x_j)}{R_{ij}^m} > \Theta, \qquad (1)$$
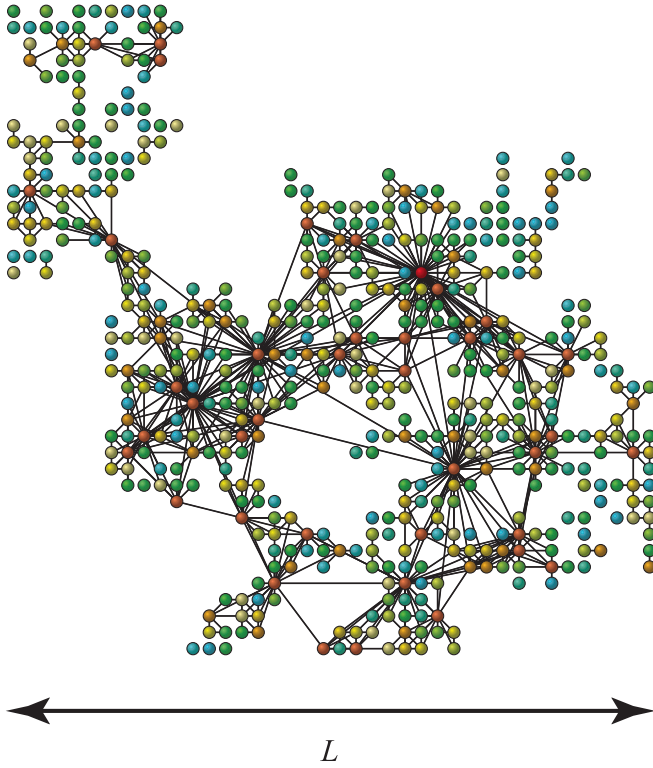
*yakubo@eng.hokudai.ac.jp
†dean.korosak@uni-mb.si

FIG. 1. (Color online) Geographical network on a fractal node set. Nodes are located on sites in the largest cluster of a two-dimensional site-percolation system at the percolation threshold $p_c = 0.593$. The node set takes a fractal structure with the fractal dimension $D = 1.89$ and is embedded in a two-dimensional space (S_2). The inhomogeneous layout of nodes is regarded as a uniform node distribution in $S_{1.89}$ of linear size L. Although the fractal space reflecting the node arrangement in this figure is not isotropic, we assume in analytical treatments that S_D is isotropic from every point in S_D. Parameters to form the network are $\alpha = 2.0$, $m = 3.0$, and $x_{\min} = 1.0$. The threshold Θ is chosen to realize $\langle k \rangle = 4.0$.

then these two nodes are connected by an edge, where R_{ij} is the Euclidean distance between the nodes i and j, m ($\geqslant 0$) is a real parameter quantifying the strength of the spatial embedding (namely, the strength of the geographical effect), Θ is a threshold value, and $F(x,y)$ is a function of two fitnesses. If $m = 0$, we have a conventional threshold model which provides a scale-free network for a variety of combinations of forms of $F(x,y)$ and $s(x)$ [34,35]. If $F(x,y)$ is chosen to be constant and $m = 1$, our network model is reduced to the model proposed by [21]. Here we concentrate on the case of

$$F(x,y) = xy \qquad (2)$$

and

$$s(x) = s_0 x^{-\alpha}, \qquad (3)$$

where x (and y) are in the range of $[x_{\min}, \infty)$ and $\alpha > 1$. From the normalization condition of $s(x)$, we have

$$s_0 = (\alpha - 1)x_{\min}^{\alpha-1}. \qquad (4)$$

The network shown in Fig. 1 is formed by this rule with $\alpha = 2.0$, $m = 3.0$, and $x_{\min} = 1.0$. It should be noted that the fractal

space S_D of this example is not isotropic from every point in S_D because the isotropic condition cannot be satisfied in a finite-size numerical realization.

Similar geographical network models have been studied under the assumption that nodes are homogeneously distributed in metric spaces [15–19,22,23], while here we investigate how the inhomogeneity (fractality) of spatial distribution of nodes affects the properties of the network. Another important difference from most of previous work is the introduction of the parameter m. Preceding studies [15–19,23] fixed the value of m as $m = D$, so that the fitness x is just the expected degree of the node. This choice of m is surely quite reasonable because the degree is straightforwardly reflected by the fitness. It is, however, not obvious how important the fitness is for obtaining edges from other nodes if the fitness is defined as an intrinsic attractiveness of a node independently of the degree. In an acquaintance network, for example, a very attractive person (large fitness) living in a depopulated area has sometimes only a few friends (like large fitness nodes in the top-left peninsula in Fig. 1), while a less-attractive person (small fitness) in a densely populated city may have a lot of acquaintances. Since our purpose is to elucidate how a network embedded in a fractal space changes its statistical or structural properties by varying the strength of the geographical effect, we introduce the additional parameter m controlling the relative importance of the geographical effect against the role of the fitness. This additional degree of freedom m has been considered by [22] as well.

III. DEGREE DISTRIBUTION FUNCTION

First, we calculate the degree distribution function of the network formed by the above geographical algorithm. Let $k_i(l)dl$ be the number of nodes connected to the node i and included in a thin spherical shell of the radius l (width dl) centered at the position of the node i. Since the distance between the node i and a node in the shell is l, the connectivity condition Eq. (1) tells us that nodes with $x_j > \Theta l^m / x_i$ can connect to the node i. Thus, the number of connected nodes $k_i(l)dl$ in an average sense is

$$k_i(l)dl = n(l)dl \cdot \int_{\Theta l^m / x_i}^{\infty} s(x)\,dx, \qquad (5)$$

where $n(l)dl = \rho \Omega l^{D-1} dl$ is the number of nodes in this spherical shell. In this expression we assume that $\Theta l^m / x_i > x_{\min}$, which is equivalent to $l > l_{\min}(x_i)$ where

$$l_{\min}(x_i) = \left(\frac{x_{\min} x_i}{\Theta}\right)^{1/m}. \qquad (6)$$

All nodes within distance $l_{\min}(x_i)$ from the node i must be connected to i. Thus, using Eq. (4) we have

$$k_i(l) = \begin{cases} \rho \Omega \left(\frac{x_{\min}}{\Theta}\right)^{\alpha-1} x_i^{\alpha-1} l^{D-1-m(\alpha-1)}, & l > l_{\min}(x_i), \\ \rho \Omega l^{D-1}, & l \leqslant l_{\min}(x_i). \end{cases} \qquad (7)$$

Integrating $k_i(l)$ with respect to l over $(0,L]$ we obtain the total number of nodes connected to the node i, or the degree k_i of

the node i given by

$$k_i = \frac{\rho \Omega x_{\min}^{\alpha-1} L^{D-m(\alpha-1)}}{[D-m(\alpha-1)]\Theta^{\alpha-1}} x_i^{\alpha-1}$$

$$+ \frac{\rho \Omega x_{\min}^{D/m}}{\Theta^{D/m}} \left[\frac{1}{D} - \frac{1}{D-m(\alpha-1)} \right] x_i^{D/m}. \quad (8)$$

It should be noted that geometrical coefficients in two terms of Eq. (8) coming from the volume integration are not very meaningful for realistic systems because of our artificial isotropic conditions. In the case of $D - m(\alpha - 1) > 0$, the first term of Eq. (8) dominates the second term when L is sufficiently large and we have $k_i \propto x_i^{\alpha-1}$. The degree distribution function $P(k)$ calculated from the relation $|P(k)dk| = |s(x)dx|$ is then proportional to k^{-2} independently of m, α, and D. If $D - m(\alpha - 1) < 0$, however, the second term of Eq. (8) becomes much larger than the first term. The relation $k_i \propto x_i^{D/m}$ in this case leads to $P(k) \propto k^{-m(\alpha-1)/D-1}$. Just at $D - m(\alpha - 1) = 0$, the first and second terms of Eq. (8) evenly contribute to k_i. However, these two terms are both proportional to $x_i^{\alpha-1}$, and lead $P(k) \propto k^{-2}$. Therefore, the degree distribution function of the present geographical network model is given by

$$P(k) \propto \begin{cases} k^{-2}, & m \leqslant m_{c0}, \quad (9a) \\ k^{-m(\alpha-1)/D-1}, & m > m_{c0}, \quad (9b) \end{cases}$$

where

$$m_{c0} = \frac{D}{\alpha - 1}. \quad (10)$$

The distribution function $P(k)$ obeys power-law forms in both cases of $m \leqslant m_{c0}$ and $m > m_{c0}$, and the degree exponent γ is

$$\gamma = \begin{cases} 2, & m \leqslant m_{c0}, \\ \dfrac{m}{D}(\alpha - 1) + 1, & m > m_{c0}. \end{cases} \quad (11)$$

The geographical inhomogeneity of the node distribution does not affect the degree distribution of the network when the geographical effect is weak (small m), while γ depends on D when this effect is strong (large m).

Let us consider the above results from a viewpoint of previous work on similar network models with uniformly distributed nodes. First, we note that the condition $m > m_{c0}$ is always satisfied if $m = D$ and $\alpha > 2$ as assumed in [15–19]. The degree k_i of the node i with the fitness x_i is then given by $k_i \propto x_i^{D/m} \propto x_i$. This is consistent with the argument in [15,23] where the degree is essentially equivalent to the fitness by setting m as $m = D$. In this case the degree exponent γ must be identical to the exponent α of the fitness distribution, which can be confirmed by Eq. (11). The gravity model with the Pareto distribution of fitnesses studied by Masuda *et al.* [22] is the same as our model except for the uniform distribution of nodes in Euclidean space. In this model the exponent m is independent of the spatial dimension as treated in this paper. They calculated k_i and $P(k)$ only for $m > m_{c0}$ and obtained the same results as Eqs. (8) and (11) with the integer dimension d instead of D.

In the derivation of $P(k)$, we assumed that the system size L is always larger than $l_{\min}(x_i)$ for any x_i because we are interested in the thermodynamic case ($L \to \infty$). This is, however, not obvious, because x_i [and then $l_{\min}(x_i)$] can also

diverge in the thermodynamic limit under a constant density ρ. Let us consider carefully this condition $L > l_{\min}(x_i)$. In a finite system with $N = \rho \Omega L^D / D$ nodes, the fitness x is truncated at a finite value. The maximum fitness $x_{\max}$ is given by $N \int_{x_{\max}}^{\infty} s(x)\, dx = 1$. Using the fitness distribution Eq. (3) with Eq. (4), the quantity $x_{\max}$ is given by

$$x_{\max} = N^{1/(\alpha-1)} x_{\min}. \quad (12)$$

Thus, the length $l_{\min}(x_i)$ can be as large as $l_{\min}(x_{\max})$, where

$$l_{\min}(x_{\max}) = \left[\frac{(\rho \Omega)^{1/(\alpha-1)} L^{D/(\alpha-1)} x_{\min}^2}{\Theta D^{1/(\alpha-1)}} \right]^{1/m}. \quad (13)$$

From the above expression, the condition $l_{\min}(x_{\max}) < L$ is equivalent to $\Theta > \Theta_0$, where

$$\Theta_0 = \left(\frac{\rho \Omega}{D} \right)^{\frac{1}{\alpha-1}} x_{\min}^2 L^{\frac{D}{\alpha-1}-m}. \quad (14)$$

If $m > m_{c0}$, the quantity Θ_0 goes to zero in the thermodynamic limit and any finite Θ satisfies $\Theta > \Theta_0$, namely, $l_{\min}(x_{\max}) < L$. Thus, the degree distribution function $P(k)$ is given by Eq. (9b) in the thermodynamic limit with $m > m_{c0}$. On the contrary, if $m < m_{c0}$, Θ is always less than Θ_0 because Θ_0 diverges as $L \to \infty$. In this case there must be nodes satisfying $l_{\min}(x_i) > L$ for which $x_i > \Theta L^m / x_{\min}$. Since the condition $x_i > \Theta L^m / x_{\min}$ implies that any node in the whole system can connect to the node i, the degree of such a node is $N - 1$ independent of x_i. Thus, these nodes give an additional δ-functional contribution $\delta(k - N + 1)$ to the degree distribution $P(k)$ given by Eq. (9a) for nodes with $x_i < \Theta L^m / x_{\min}$. Let us estimate the magnitude of this δ-functional contribution. It is proportional to the number of nodes n_0 having $x_i > \Theta L^m / x_{\min}$. The quantity n_0 is given by $N \int_{\Theta L^m / x_{\min}}^{\infty} s(x)\, dx$ and can be written as

$$n_0 = N \left(\frac{L}{\xi} \right)^{-m(\alpha-1)}, \quad (15)$$

where ξ is the node-pair distance defined by

$$\xi = l_{\min}(x_{\min}) = \left(\frac{x_{\min}^2}{\Theta} \right)^{1/m}, \quad (16)$$

below which the two nodes are connected independently of the fitness. The properly normalized δ-functional part of $P(k)$ is then presented by $(L/\xi)^{-m(\alpha-1)} \delta(k - N + 1)$. Since $m(\alpha - 1)$ is always positive, the δ-functional contribution vanishes in the thermodynamic limit. In the case of $m = m_{c0}$, Θ_0 defined by Eq. (14) is finite independently of L, which implies that Θ can be either larger or smaller than Θ_0 even in the thermodynamic limit. If Θ is larger than Θ_0, Eq. (9) providing $P(k) \propto k^{-2}$ at $m = m_{c0}$ is valid. For $\Theta \leqslant \Theta_0$, the δ-functional correction to $P(k)$ should be considered. The magnitude n_0 of this correction is, however, infinitesimal because $n_0 = N(L/\xi)^{-D}$ goes to zero for $L \to \infty$. Therefore, Eq. (9) is valid both for the cases of $m \leqslant m_{c0}$ and $m > m_{c0}$ in an infinite system. For a finite L, however, Θ can be chosen to be less than Θ_0 independently of m and the δ-functional contribution remains finite. Thus, the degree distribution function of a finite system with $\Theta < \Theta_0$ must have the δ-functional correction term, that is,

$$P(k) = p_0 k^{-\gamma} + \left(\frac{L}{\xi} \right)^{-m(\alpha-1)} \delta(k - N + 1) \quad (17)$$

for both $m \leqslant m_{c0}$ and $m > m_{c0}$. Here p_0 is a normalization constant and the exponent γ is given by Eq. (11). It should be noted that Eq. (17) holds for a finite but large L because the first term of Eq. (17) is valid in the large L limit.

The quantity Θ_0 is a characteristic value of Θ peculiar to a finite system with a fixed density ρ. There are two other characteristic values of Θ for a finite L. One is $\Theta_{\min}$ below which every node can connect to all other $N - 1$ nodes. Network constructed under $\Theta < \Theta_{\min}$ becomes the complete graph. Obviously, $\Theta_{\min}$ is given by

$$\Theta_{\min} = \frac{x_{\min}^2}{L^m}. \tag{18}$$

Another characteristic Θ is $\Theta_{\max}$ above which no node can connect to any other nodes. Network with $\Theta > \Theta_{\max}$ is a set of isolated nodes. We have

$$\Theta_{\max} = \frac{x_{\max}^2}{\Delta l^m}, \tag{19}$$

where $\Delta l = (\rho \Omega / D)^{-1/D}$ is the minimum edge length. Using Eq. (12), $\Theta_{\max}$ is written as

$$\Theta_{\max} = \left(\frac{\rho \Omega}{D} \right)^{\frac{2}{\alpha-1} + \frac{m}{D}} x_{\min}^2 L^{\frac{2D}{\alpha-1}}. \tag{20}$$

In a finite system we always assume that Θ satisfies the condition $\Theta_{\min} \ll \Theta \ll \Theta_{\max}$ leading to nontrivial networks. It is not necessary to consider this condition in an infinite system because $\Theta_{\min}$ and $\Theta_{\max}$ vanishes and diverges, respectively, in the thermodynamic limit.

IV. RELATION BETWEEN $\langle k \rangle$ AND Θ

From Eq. (9) it is clear that the average degree $\langle k \rangle$ diverges for $m \leqslant m_{c0}$ because of $\gamma = 2$ and it remains finite for $m > m_{c0}$ in the thermodynamic limit. The average degree $\langle k \rangle$ in a finite system is, however, always finite and depends on Θ. A large Θ restricts a connection of a node pair and leads a small $\langle k \rangle$. It is important to know the relation between Θ and $\langle k \rangle$ for a finite but large system. We calculate the average degree by

$$\langle k \rangle = \int_{x_{\min}}^{x_{\max}} k_i s(x) \, dx, \tag{21}$$

instead of $\langle k \rangle = \int k P(k) dk$, using the derived expression (not asymptotic form) of k_i for a finite system.

Let us consider Eq. (21) separately for $\Theta < \Theta_0$ and for $\Theta \geqslant \Theta_0$. In the case of $\Theta < \Theta_0$, the length $l_{\min}(x_i)$ can be larger than L, which implies $\Theta L^m / x_{\min} < x_{\max}$. Thus, the integral of Eq. (21) is separated into two regions

$$\langle k \rangle = \int_{x_{\min}}^{\Theta L^m / x_{\min}} k(x) s(x) \, dx + (N-1) \int_{\Theta L^m / x_{\min}}^{x_{\max}} s(x) \, dx, \tag{22}$$

where $k(x)$ is given by Eq. (8) regarding x_i as a continuous variable x. The coefficient $N - 1$ in the second term comes from the fact that nodes satisfying $x_i > \Theta L^m / x_{\min}$ connect to all $N - 1$ nodes. Using Eq. (8) and approximating $\int_{\Theta L^m / x_{\min}}^{x_{\max}} dx$

by $\int_{\Theta L^m / x_{\min}}^{\infty} dx$, we have

$$\langle k \rangle = X \rho \left(\frac{x_{\min}^2}{\Theta} \right)^{\alpha-1} L^{D-m(\alpha-1)} \ln \left(\frac{Y \Theta L^m}{x_{\min^2}} \right)$$
$$+ Z \rho \left(\frac{x_{\min}^2}{\Theta} \right)^{D/m}, \tag{23}$$

where

$$X = \frac{\Omega(\alpha - 1)}{D - m(\alpha - 1)}, \tag{24}$$

$$Y = \exp \left\{ \frac{d - 2m(\alpha - 1)}{(\alpha - 1)[d - m(\alpha - 1)]} \right\}, \tag{25}$$

$$Z = \frac{\Omega m^2 (\alpha - 1)^2}{D[D - m(\alpha - 1)]^2}. \tag{26}$$

It should be again emphasized that these geometrical quantities X, Y, and Z resulting from the volume integration depend strongly on the boundary conditions and are not very meaningful. We can evaluate the asymptotic behavior of $\langle k \rangle$ for large L by using Eq. (23). In the case of $m < m_{c0}$, the first term of Eq. (23) obviously dominates the second term. Then, ignoring unimportant geographical coefficient, $\langle k \rangle$ behaves asymptotically as

$$\langle k \rangle \sim \Theta^{1-\alpha} (\ln \Theta + c). \tag{27}$$

where c is a constant depending on the boundary condition. On the other hand, a careful treatment is required for $m > m_{c0}$. It seems that the second term of Eq. (23) dominates the first term for $m > m_{c0}$ in the thermodynamic limit. However, we should note that Θ must be infinitesimal to satisfy the condition $\Theta < \Theta_0$ in this calculation because Θ_0 for $L \to \infty$ goes to zero for $m > m_{c0}$. So, it is not obvious which term is dominating in Eq. (23). In order to find the dominant term, we evaluate the lower bounds of these terms by replacing Θ by Θ_0 and using Eq. (14). Then, the lower bounds of the first and the second terms are proportional to $\ln[Y(\rho \Omega / D)^{1/(\alpha-1)} L^{D/(\alpha-1)}]$ and $L^{[1 - \frac{D}{m(\alpha-1)}]}$, respectively. This suggests that the second term dominates the first term for large L because the exponent $1 - \frac{D}{m(\alpha-1)}$ is positive for $m > m_{c0}$. We have then

$$\langle k \rangle \sim \Theta^{-D/m} \tag{28}$$

for $\Theta < \Theta_0$ and $m > m_{c0}$. At $m = m_{c0}$, both terms in Eq. (23) should be considered.

Next, we treat the case of $\Theta \geqslant \Theta_0$. Here $l_{\min}(x_i)$ is always less than L, and the degree k_i is given by Eq. (8) for any x_i in the range of $x_{\min} \leqslant x_i \leqslant x_{\max}$. The average degree $\langle k \rangle$ is then simply presented by

$$\langle k \rangle = \int_{x_{\min}}^{x_{\max}} k(x) s(x) \, dx$$
$$= \frac{\rho X}{\alpha - 1} \left(\frac{x_{\min}^2}{\Theta} \right)^{\alpha-1} L^{D-m(\alpha-1)} \ln \left(W L^D \right)$$
$$- \rho Z [D - m(\alpha - 1)] \left(\frac{x_{\min}^2}{\Theta} \right)^{D/m} \left(W L^D \right)^{\frac{D}{m(\alpha-1)} - 1}$$
$$+ \rho Z [D - m(\alpha - 1)] \left(\frac{x_{\min}^2}{\Theta} \right)^{D/m}, \tag{29}$$

where $W = \rho\Omega/D$ and we used Eq. (8) for $k(x)$ and Eq. (12). For $m > m_{c0}$, it is easy to understand that the third term dominates other two terms for large L. So we have

$$\langle k \rangle \sim \Theta^{-D/m}. \tag{30}$$

In the case of $m < m_{c0}$, the infinitely large Θ must be considered when we find the dominant term of Eq. (29), because Θ is larger than Θ_0 and Θ_0 diverges for $m < m_{c0}$ in the thermodynamic limit. As in the case of $\Theta < \Theta_0$ and $m > m_{c0}$, replacing Θ in Eq. (29) by Θ_0, the L dependence of the upper bounds of these terms points to the dominant term. Since the upper bounds of the first, second, and third terms are proportional to $\ln(\rho\Omega L^D/D)$, L^0, and $L^{D[1-\frac{1}{m(\alpha-1)}]}$, respectively, the first term dominates the second and third terms because of $m < m_{c0}$. Therefore, the average degree $\langle k \rangle$ is asymptotically given by

$$\langle k \rangle \sim \Theta^{1-\alpha}. \tag{31}$$

This relation differs from Eq. (27) by a logarithmic correction.

In summary, the relation between Θ and $\langle k \rangle$ is given by

$$\langle k \rangle \sim \begin{cases} \Theta^{1-\alpha}(\ln\Theta + c), & m < m_{c0}, \\ \Theta^{-D/m}, & m > m_{c0} \end{cases} \tag{32}$$

for $\Theta < \Theta_0$ and

$$\langle k \rangle \sim \begin{cases} \Theta^{1-\alpha}, & m < m_{c0}, \\ \Theta^{-D/m}, & m > m_{c0} \end{cases} \tag{33}$$

for $\Theta \geqslant \Theta_0$. At $m = m_{c0}$, $\langle k \rangle$ is related to Θ through Eq. (23) for $\Theta < \Theta_0$ and through Eq. (29) for $\Theta \geqslant \Theta_0$, because every term contributes equally to $\langle k \rangle$ even in the thermodynamic limit. A similar result to Eq. (33) has been obtained by [23] where the nodes were uniformly distributed in a d-dimensional space with the L-max norm and m is fixed at $m = D(= d)$. The asymptotic L dependence of $\langle k \rangle$ for a fixed Θ can be also evaluated from Eqs. (23) and (29). In the case of $m < m_{c0}$ and enough large L (then, $\Theta < \Theta_0$), the dominant first term of Eq. (23) gives $\langle k \rangle \propto L^{D-m(\alpha-1)} \ln L$. For $m > m_{c0}$ (and then $\Theta > \Theta_0$), we have $\langle k \rangle \propto L^0$. These L dependences are consistent with those calculated by $\langle k \rangle = \int_1^{N-1} k P(k) dk$ by using Eqs. (17) and (9) for $m < m_{c0}$ and $m > m_{c0}$, respectively, and taking into account the L dependence of p_0 in Eq. (17).

V. EDGE-LENGTH DISTRIBUTION FUNCTION

In this section, we derive the edge-length distribution function $R(l)$ of our geographical networks embedded in a fractal space. To this end, we regard $k_i(l)$ given by Eq. (7) as a continuous function $k(x,l)$ of the fitness x and the edge length l. The average number of edges, $k(l)dl$, of length $[l, l+dl]$ from a given node is obtained by averaging $k(x,l)$ over the fitness x, that is,

$$k(l) = \int_{x_{\min}}^{x_{\max}} s(x) k(x,l) dx. \tag{34}$$

Equation (7) expresses the forms of $k_i(l)$ by separating two cases $l > l_{\min}(x_i)$ and $l \leqslant l_{\min}(x_i)$ for a fixed x_i. Corresponding to this classification, $k(x,l)$ in Eq. (34) for a fixed l has different

forms for $x_{\min} \leqslant x < \Theta l^m/x_{\min}$ and $\Theta l^m/x_{\min} \leqslant x \leqslant x_{\max}$, respectively. Thus, the integral of Eq. (34) is calculated as

$$k(l) = \rho\Omega\left(\frac{x_{\min}}{\Theta}\right)^{\alpha-1} l^{D-1-m(\alpha-1)} \int_{x_{\min}}^{x_l} s(x) x^{\alpha-1} dx$$
$$+ \rho\Omega l^{D-1} \int_{x_l}^{\infty} s(x) dx, \tag{35}$$

where $x_l = \Theta l^m/x_{\min}$. Here l is assumed to be larger than ξ (namely $x_{\min} < x_l$) as argued in [22] and the upper cutoff of the integral is, as an approximation, extended to infinity. Using Eqs. (3), (4), (7), and (16), $k(l)$ is expressed by

$$k(l) = \rho\Omega l^{D-1}\left(\frac{l}{\xi}\right)^{-m(\alpha-1)}\left[1 + m(\alpha-1)\ln\left(\frac{l}{\xi}\right)\right]. \tag{36}$$

The probability distribution function $R(l)$ is given by

$$R(l) = \frac{k(l)}{\displaystyle\int_L k(l')dl'}, \tag{37}$$

where the integration in the denominator is done over the whole range of l. Neglecting the normalization constant, the edge-length distribution is

$$R(l) \propto l^{D-1}\left(\frac{l}{\xi}\right)^{-m(\alpha-1)}\left[1 + m(\alpha-1)\ln\left(\frac{l}{\xi}\right)\right]. \tag{38}$$

We should remark that $R(l)$ for $m \leqslant m_{c0}$ goes to infinity as $l \to \infty$. Since the length l does not exceed the size L in a finite system, the distribution $R(l)$ is actually truncated at $l = L$. In the thermodynamic limit, however, we must consider the infinitesimal normalization constant coming from the denominator of Eq. (37).

In the case of $l \leqslant \xi$ (namely $x_{\min} \geqslant x_l$), $k(x,l)$ in Eq. (34) is given by Eq. (7) for any x in the integration range $[x_{\min}, x_{\max}]$. Thus, $k(l)$ is given by $\rho\Omega l^{D-1} \int_{x_{\min}}^{\infty} s(x)dx$, namely

$$k(l) = \rho\Omega l^{D-1} \tag{39}$$

and then

$$R(l) \propto l^{D-1}. \tag{40}$$

It should be noted that the distribution $R(l)$ depends on the fractal dimension D independently of m while the degree distribution $P(k)$ does not depend on D for $m \leqslant m_{c0}$ (weak geographical effect region).

Two expressions Eqs. (38) and (40) of $R(l)$ for $l > \xi$ and $l \leqslant \xi$ must coincide at $l = \xi$. This condition concludes that the proportionality coefficients for Eqs. (38) and (40) are identical. We can derive the common coefficient C for an infinite system from the normalization condition of $R(l)$ given by

$$1 = C \int_0^{\xi} l^{D-1} dl$$
$$+ C \int_{\xi}^{\infty} l^{D-1}\left(\frac{l}{\xi}\right)^{-\beta}\left[1 + \beta\ln\left(\frac{l}{\xi}\right)\right] dl, \tag{41}$$

where $\beta = m(\alpha - 1)$. When $m > m_{c0}$, this equation provides a finite coefficient expressed by

$$C = \frac{D}{\xi^D}\left(1 - \frac{m_{c0}}{m}\right)^2, \tag{42}$$

while C is infinitesimal for $m \leqslant m_{c0}$ as mentioned above. It should be noted that the coefficient C does not depend on the boundary condition because the boundary-condition dependent factors in the numerator and the denominator in Eq. (37) are canceled out.

From the above argument, we can immediately derive the probability $g(l)$ of two nodes with distance l to be connected. This probability is given by the ratio of the number of connected nodes $k(l)dl$ to $n(l)dl$ nodes located at distances $[l,l + dl]$ from a given node, namely $g(l) = k(l)/n(l)$. Since $n(l) = \rho\Omega l^{D-1}$ as mentioned below Eq. (5) and $k(l)$ is given by Eq. (36) or (39) for $l > \xi$ or $l \leqslant \xi$, respectively, we have

$$g(l) = \begin{cases} \left(\frac{l}{\xi}\right)^{-m(\alpha-1)}\left[1 + m(\alpha-1)\ln\left(\frac{l}{\xi}\right)\right], & l > \xi, \\ 1, & l \leqslant \xi. \end{cases} \tag{43}$$

This expression can be alternatively derived directly from the meaning of $g(l)$,

$$g(l) = \int_{x_{\min}}^{\infty} s(x)dx \int_{xy/l^m > \Theta} s(y)dy. \tag{44}$$

Considering that two nodes are always connected if the fitness of one node exceeds $x_l(= \Theta l^m/x_{\min})$, we can separate the above integration into two parts as

$$g(l) = \int_{x_l}^{\infty} s(x)dx + \int_{x_{\min}}^{x_l} s(x)dx \int_{\Theta l^m/x}^{\infty} s(y)dy \tag{45}$$

for $x_l > x_{\min}$ ($l > \xi$). For $x_l \leqslant x_{\min}$, the integral range of the second integral in Eq. (44) is extended over the whole region of y, then $g(l) = \int_{x_{\min}}^{\infty}\int_{x_{\min}}^{\infty} s(x)s(y)dxdy$. These equations again lead Eq. (43) if we use Eq. (3). We should note that the probability $g(l)$ given by Eq. (43) does not depends on the fractal dimension D for any value of m.

VI. NUMERICAL CONFIRMATIONS

In this section the above analytical results are numerically verified. First, we confirm the behavior of the degree distribution function $P(k)$ for geographical networks on the Sierpinski node set (Sierpinski geographical networks). Nodes are located on vertices of the Sierpinski gasket with the fractal dimension $D = \ln 3/\ln 2 \approx 1.585$ and connected by edges according to the condition Eq. (1) with Eqs. (2) and (3). In our numerical calculations in this work, the lower cut-off $x_{\min}$ is set to be unity. A typical network on the sixth generation Sierpinski gasket ($N = 366$) is depicted in Fig. 2. There exist nodes (hubs) possessing a large number of edges. Numerically calculated degree distribution functions $P(k)$ for two networks with different m and α are presented in Fig. 3. The scale-free property of networks formed by our algorithm is clearly shown in this figure. The scale-free exponents γ calculated numerically for many Sierpinski geographical networks with different combinations of α and m are plotted in Fig. 4 as a
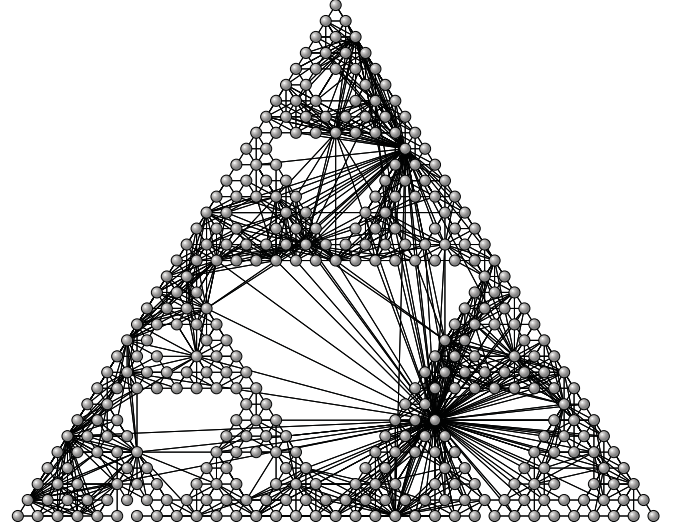

FIG. 2. Typical Sierpinski geographical network. Nodes are located on vertices of the Sierpinski gasket in the sixth generation ($N = 366$). Parameters to form the network are $\alpha = 2.0$ and $m = 3.0$. The threshold Θ is chosen to satisfy $\langle k \rangle = 10.0$.

function of $m(\alpha - 1)/D$. In Fig. 4 values of γ for networks with nodes distributed homogeneously in two-dimensional Euclidean space (2D-geographical networks) are also plotted. We should remark that our theoretical arguments are valid for a geographical network with homogeneously distributed nodes which is a special case with the Euclidean dimension d instead of the fractal dimension D. In fact, Eq. (11) with $D = d$ reproduces the results by [15–19,22,23] as mentioned in Sec. III. We see that all numerical results in Fig. 4 collapse onto the theoretical line given by Eq. (11). These results strongly support the theoretical predictions on $P(k)$ presented in Sec. III.
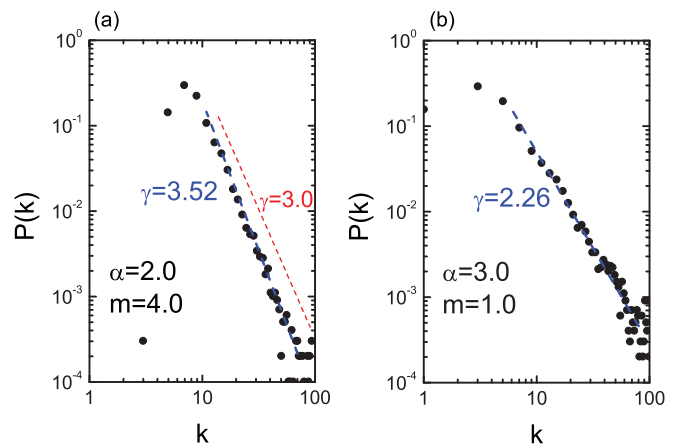

FIG. 3. (Color online) Degree distribution of Sierpinski geographical networks in the ninth generation ($N = 9,843$). The threshold Θ is chosen to satisfy $\langle k \rangle = 10.0$. Parameters α and m are set as (a) $\alpha = 2.0$ and $m = 4.0$ and (b) $\alpha = 3.0$ and $m = 1.0$. Thick dashed lines through dots indicate the slopes γ predicted by Eq. (11). Thin dashed line in (a) represents the slope calculated by using the Euclidean dimension $d = 2$ instead of the fractal dimension $D = 1.585$.
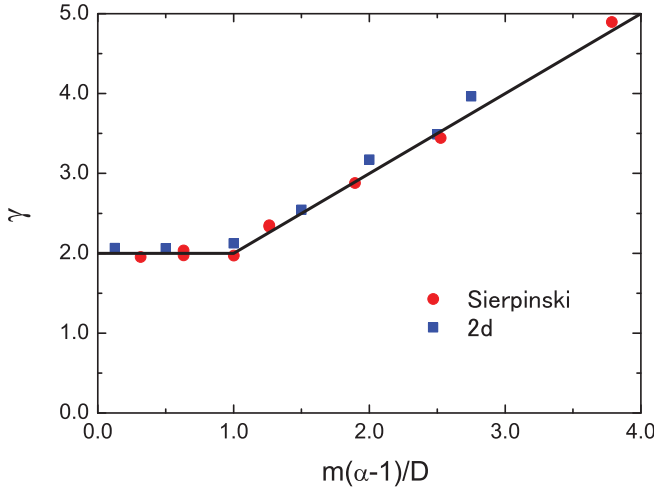
FIG. 4. (Color online) Degree exponent γ as a function of $m(\alpha - 1)/D$. Solid line shows the theoretical prediction Eq. (11). Filled circles and squares represent numerically obtained γ for Sierpinski geographical networks (ninth generation) and 2D-geographical networks ($N = 1000$) with several combinations of α and m. Values of γ are evaluated by the least-squares fit.

Next, we confirm numerically the relation between $\langle k \rangle$ and Θ presented by Eqs. (32) and (33). Since the Θ dependence of $\langle k \rangle$ is complicated as classified into four cases depending on values of m and Θ, we checked this relation for 2D-geographical networks to avoid additional complications due to the noninteger dimension D. It is necessary to evaluate Θ_0 defined by Eq. (14) to distinguish the four regions with respect to m and Θ. In addition, Θ must satisfy the condition $\Theta_{min} \ll \Theta \ll \Theta_{max}$, where Θ_{min} and Θ_{max} are given by Eqs. (18) and (20), respectively. Rewriting Eqs. (14) and (20) by $\Theta_0 = N^{1/(\alpha-1)}x_{min}^2 L^{-m}$ and $\Theta_{max} = N^{[2D+m(\alpha-1)]/[D(\alpha-1)]}x_{min}^2 L^{-m}$, we can estimate Θ_0 and Θ_{max} without treating the boundary-condition dependent geometrical factor $\rho\Omega/D$. Figures 5(a) and 5(b) show numerically calculated $\langle k \rangle$ as a function of Θ for $m < m_{c0}$. In these calculations, the exponent m characterizing the strength of the geographical effect are chosen to be 1.0, and $N = 1000$ nodes are distributed in a square of size $L = 100.0$. The fitness x is allocated to each node according to the probability distribution function $s(x)$ given by Eq. (3) with $\alpha = 1.5$. The square system has the periodic boundary conditions in the x and y directions. Since $m_{c0} = 4.0$, $\Theta_0 = 10^4$, $\Theta_{min} = 0.01$, and $\Theta_{max} = 3.2 \times 10^{11}$ from these parameters, the conditions $\Theta_{min} \ll \Theta < \Theta_0 \ll \Theta_{max}$ and $\Theta_{min} \ll \Theta_0 < \Theta \ll \Theta_{max}$ are satisfied for Figs. 5(a) and 5(b), respectively, and $m < m_{c0}$ for both figures. Numerically calculated $\langle k \rangle$s are well described by solid lines representing Eqs. (32) and (33), where the constant c in Eq. (32) and prefactors are suitably chosen. Results for $m > m_{c0}$ are presented in Figs. 5(c) and 5(d) which show the $\langle k \rangle$-Θ relation for $\Theta < \Theta_0$ and $\Theta > \Theta_0$, respectively. As in the cases of Figs. 5(a) and 5(b), nodes are distributed in a square of size $L = 100.0$. Parameters α and m are chosen as $\alpha = 2.0$ and $m = 3.0$ so that the condition $m > m_{c0}(= 2.0)$ is satisfied. In order to realize the condition $\Theta_{min} \ll \Theta < \Theta_0 \ll \Theta_{max}$ in Fig. 5(c), we treated a large network with $N = 100\,000$, which has $\Theta_{min} = 10^{-6}$, $\Theta_{max} = 3.2 \times 10^{11}$, and $\Theta_0 = 0.1$. On the

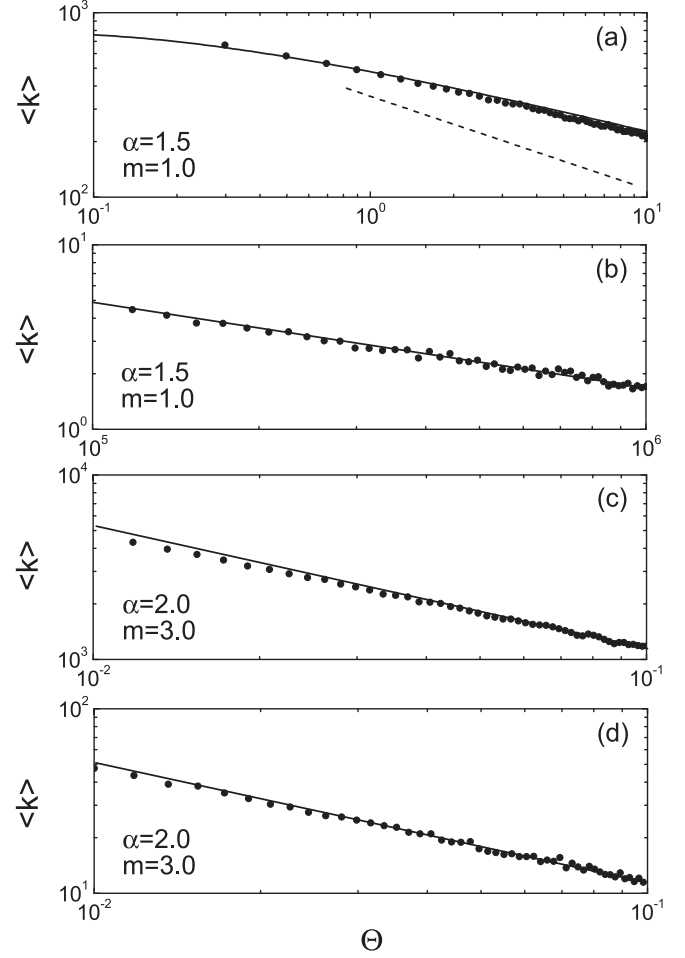FIG. 5. Relation between the average degree $\langle k \rangle$ and the threshold value Θ for (a) $m < m_{c0}$ and $\Theta < \Theta_0$, (b) $m < m_{c0}$ and $\Theta > \Theta_0$, (c) $m > m_{c0}$ and $\Theta < \Theta_0$, and (d) $m > m_{c0}$ and $\Theta > \Theta_0$. Dots represent results calculated numerically for 2D-geographical networks. Details of parameters and conditions are given in the text. Solid curves show the theoretical predictions given by Eqs. (32) and (33). Dashed line in (a) indicates the slope given by Eq. (32) without the logarithmic term.

contrary, the number of nodes for Fig. 5(d) is $N = 1000$, in which $\Theta_{min} \ll \Theta_0 < \Theta \ll \Theta_{max}$ is satisfied with $\Theta_{min} = 10^{-6}$, $\Theta_{max} = 3.2 \times 10^4$, and $\Theta_0 = 0.001$. Similarly to Figs. 5(a) and 4(b), numerical results agree well with the theoretical results shown by solid lines.

The behavior of the edge-length distribution function $R(l)$ is also examined. If we employ Sierpinski geographical networks to compute $R(l)$, the distance l between nodes becomes a discrete variable, which makes it difficult to compare $R(l)$ to the analytically obtained result with continuous l. Thus we calculate numerically $R(l)$ for 2D-geographical networks again. Results (dots) shown in Fig. 6 are calculated in the condition of $m > m_{c0}$, for which $R(l)$ can be properly normalized even in the thermodynamic limit. Solid and dashed curves represent the theoretical prediction given by Eqs. (38) and (40) with the common prefactor presented by Eq. (42). It should be emphasized that there is no fitting parameter to obtain the theoretical curve. Numerical results agree quite well with the theoretical prediction. The peak structure and the
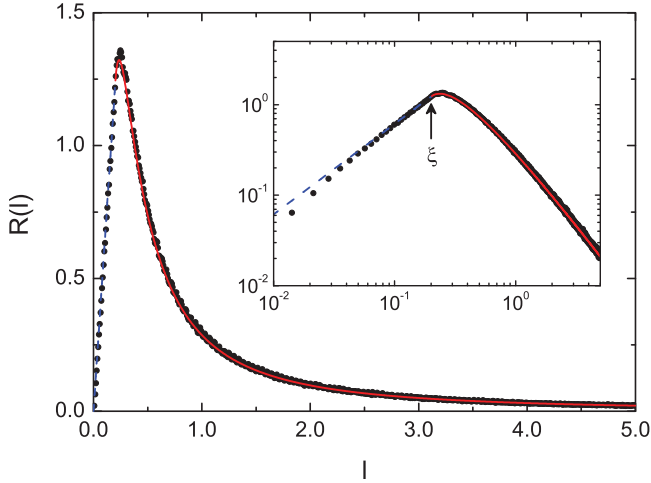
FIG. 6. (Color online) Edge-length distribution function $R(l)$ for 2D-geographical networks in squares of size $L = 10.0$. In order to obtain the numerical results (dots), we employ $\alpha = 2.0$ ($m_{c0} = 2.0$), $m = 3.0$, and $N = 1000$. The threshold value Θ is chosen as $\Theta = 125.2$ so that $\langle k \rangle$ becomes equal to 10.0. The results are averaged over 1000 realizations. Solid and dashed lines represent the theoretical results given by Eqs. (38) and (40) with the common prefactor presented by Eq. (42). The inset shows the same result in a logarithmic scale. The length $\xi (= 0.20)$ defined by Eq. (16) is indicated by the arrow.

tail profile slightly deviating from a power law (see the inset of Fig. 6) result from the logarithmic term of Eq. (38). The reason of the slight deviation between numerical results and the theoretical line for $l \lesssim 0.05$ (see the inset) is due to a finite Δl defined below Eq. (19).

Finally, we address the small-world and fractal property of our geographical networks. For $m = 0$, a network belongs to the threshold model [34,35], in which the network structure depends only on the fitness x regardless of spatial positions of nodes. In this case, the node-pair (topological) distance in a connected graph is bounded by 2 independently of the network size N, because the node with the largest fitness connects to all other nodes. This implies that the network is "smaller" than a ultra-small-world network (diameter$\propto \ln \ln N$). In the opposite limit, that is, $m \to \infty$, with a finite $\langle k \rangle$, nodes are connected only to neighboring nodes. Thus, the network must take a fractal structure with the same fractal dimension D as that of the node set [37]. It is plausible that a network with a finite m exhibits an intermediate morphology between these limiting structures. Figure 7 shows the small-world and fractal properties of Sierpinski geographical networks with $\alpha = 2.0$ and $\langle k \rangle = 4.0$. Since the exclusive relation between fractality and small worldness does not depend on how we characterize networks [38,39], we employ two methods to study the structural feature of networks depending on diameters of networks. For networks with small diameters (small m), the diameter d_{max} of the largest singly connected graph (cluster) is calculated as a function of the number of nodes N_{max} in the cluster (cluster-growing method), as shown in Fig. 7(a). The diameter d_{max} is almost constant for m close to zero, and seems to become proportional to $\ln \ln N_{max}$ or $\ln N_{max}$ as increasing m. Further increasing m, the N_{max} dependence of d_{max} becomes
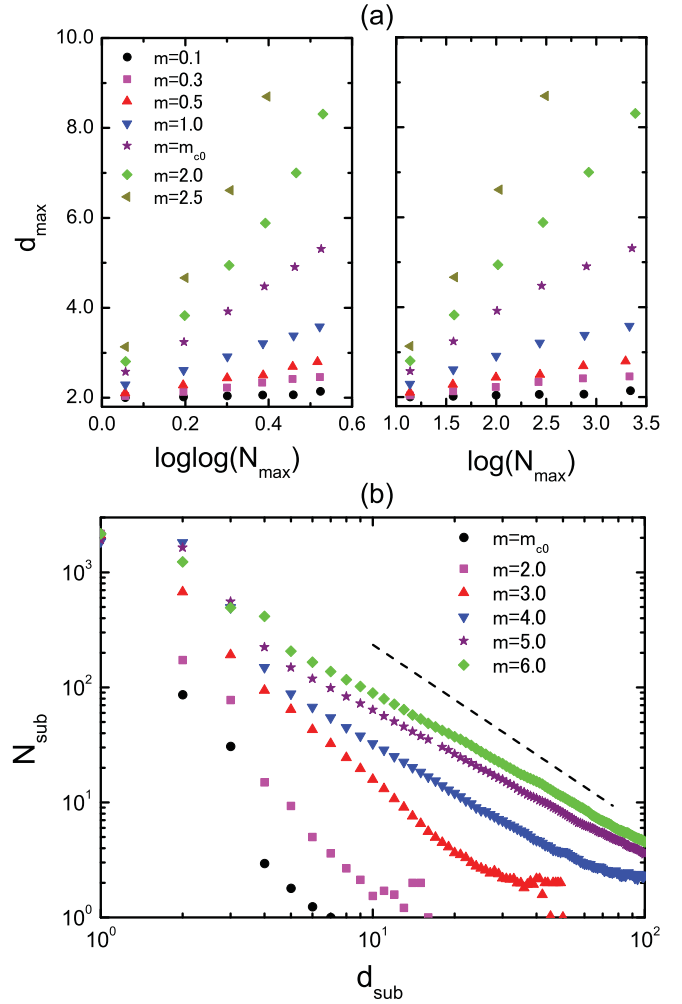

FIG. 7. (Color online) Small-world and fractal properties of Sierpinski geographical networks with $\alpha = 2.0$ and $\langle k \rangle = 4.0$. (a) Upper two panels represent the diameter of the largest connected cluster in Sierpinski geographical networks in the third to eighth generations with one-generation increments as a function of the number of nodes in the cluster. Different symbols indicate results for different values of m. The value of $m_{c0} = D/(\alpha - 1)$ is 1.585. Results are averaged over 10 000 (third generation) to 200 (eighth generation) samples depending on the generation. (b) Lower panel shows the number of subgraphs covering minimally the largest cluster in eighth-generation Sierpinski geographical networks as a function of the diameter of subgraphs. Results are averaged over 200 samples. To achieve the minimum covering, we employed the compact-box-burning algorithm [40]. Dashed line represents the slope of the fractal dimension $D = \ln 3 / \ln 2$.

faster than a logarithmic form. For networks formed by $m \geqslant m_{c0} = \ln 3 / \ln 2$ with relatively large diameters, we compute the number of subgraphs (N_{sub}) covering the largest cluster as a function of the subgraph diameter d_{sub} (box-covering method). When m is large enough compared to m_{c0}, the subgraph number N_{sub} seems to be proportional to $d_{sub}^{-D_{net}}$, which implies that the network is fractal with the fractal dimension D_{net} [37]. Such a power-law dependence has been also observed in geographical networks formed by uniformly

distributed nodes with the L-max norm [23]. It is interesting that the network fractal dimension D_{net} is not the same with the fractal dimension D of the Sierpinski node set for intermediate values of m ($> m_{c0}$) though D_{net} approaches to D as increasing m to infinity. In conclusion, the geographical network changes its structure from a constant-diameter network to ultra-small-world, small-world, and finally to fractal network as increasing the strength of the geographical effect m. However, these numerical results do not present the transition (or crossover in a finite system) values of m giving these structural changes. Furthermore, we need to study the small-world and fractal properties of geographical networks by taking into account the percolation transition caused by varying the threshold value Θ [39].

VII. COMPACTNESS AND EFFICIENCY

From the functional forms of $P(k)$ and $R(l)$, we can immediately find the convergence property of the average degree $\langle k \rangle$ and the average edge length $\langle l \rangle$ in the thermodynamic limit. As argued at the end of Sec. IV, the quantity $\langle k \rangle$ for $m \leqslant m_{c0}$ diverges as $L^{D-m(\alpha-1)} \ln L$ and it converges for $m > m_{c0}$. On the other hand, the convergence of $\langle l \rangle$ is governed by Eq. (38) describing $R(l)$ in the asymptotic l regime. From Eq. (38), the average $\langle l \rangle$ diverges for $m \leqslant m_{c1}$ and remains finite for $m > m_{c1}$, where

$$m_{c1} = \frac{D+1}{\alpha-1}. \tag{46}$$

The behavior of $\langle k \rangle$ and $\langle l \rangle$ suggests that the impact of the geographical effect can be classified into three regions. For $0 \leqslant m \leqslant m_{c0}$, both quantities $\langle k \rangle$ and $\langle l \rangle$ diverge in the thermodynamic limit. Since a node can connect with a huge number of nodes far away, we call this region the *noncompact phase*. In the same sense, the region of $m > m_{c1}$ is termed the *compact phase* where both quantities $\langle k \rangle$ and $\langle l \rangle$ remain finite and a node connects with only a small number of nodes in its vicinity. In the *intermediate phase*, that is, $m_{c0} < m \leqslant m_{c1}$, the average degree converges but $\langle l \rangle$ diverges. These regions are summarize in Table I.

Let us consider the relation between the compactness of a network and the geographical efficiency e defined by

$$e = \frac{2}{N(N-1)} \sum_{i>j} \frac{1}{r_{ij}}, \tag{47}$$

where r_{ij} is the minimum value of Euclidean distances along any possible network paths connecting the node i to j, that is,

$$r_{ij} = \min_{\substack{1 \leqslant k_1,k_2,\dots,k_m \leqslant N \\ m=1,2,\dots}} (l_{ik_1} + l_{k_1 k_2} + \cdots + l_{k_m j}), \tag{48}$$

TABLE I. Three phases of geographical networks embedded in a fractal space.

Phase	Region of m	$\langle k \rangle$	$\langle l \rangle$
Noncompact	$m \leqslant m_{c0}$	infinite	infinite
Intermediate	$m_{c0} < m \leqslant m_{c1}$	finite	infinite
Compact	$m > m_{c1}$	finite	finite

where l_{ij} is the Euclidean length of the edge $(i$-$j)$ if the edge exists and infinity otherwise. This quantity is a natural extension of the global efficiency E defined by [41]

$$E = \frac{2}{N(N-1)} \sum_{i>j} \frac{1}{d_{ij}}, \tag{49}$$

where d_{ij} is the shortest network distance between two nodes i and j. Both quantities e and E characterize the efficiency of the information exchange or the flow in the network. If the efficiency is governed mainly by the Euclidean distance along the path rather than the number of steps, the geographical efficiency e is more meaningful than E, and vice versa.

We calculated numerically these quantities and examined how the compactness is related to the efficiencies e and E. Solid lines in Fig. 8 represent the efficiencies for Sierpinski geographical networks in three different generations (having the same linear size $L = 1.0$), while dashed lines show those for 2D-geographical networks with different numbers of nodes in squares of size $L = 10.0$. In order to eliminate the L and N dependence of the maximum and minimum values of the efficiencies, we plot the rescaled quantities $\tilde{e}$ and $\tilde{E}$ defined by $\tilde{e} = (e - e_a)/(e_b - e_a)$ and $\tilde{E} = (E - E_a)/(E_b - E_a)$, where e_a (E_a) and e_b (E_b) are $e(E)$ at $m/m_{c0} = 10^{-2}$ and 10^2, respectively. The rescaled geographical (global) efficiency $\tilde{e}$ ($\tilde{E}$) rapidly increases (decreases) near $m = m_{c0}$ as increasing m. (We see that $\tilde{e}$ forms a small hump near $m/m_{c0} \sim 1$ and exceeds unity when approaching the thermodynamic limit for Sierpinski geographical networks. The rescaled efficiency larger than unity is allowed because the rescaling factor is not the maximum value of the efficiency.) From the fact that enlarging the system the slope near $m = m_{c0}$ becomes steeper and the point at the intersection of $\tilde{e}$ and $\tilde{E}$ approaches $m = m_{c0}$, the efficiencies e and E in the thermodynamic limit are supposed to show step-like forms at $m = m_{c0}$. This behavior can be interpreted as follows. In the noncompact
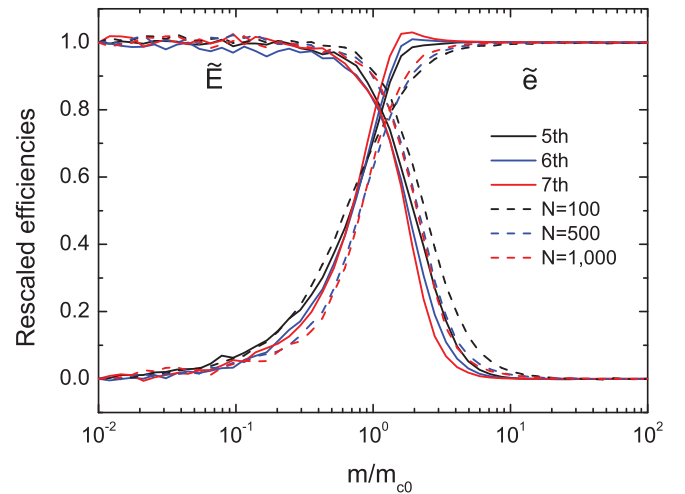

FIG. 8. (Color online) Rescaled geographical and global efficiencies as a function of m/m_{c0} for Sierpinski (solid lines) and 2D-(dashed lines) geographical networks. For both network systems, we employ $\alpha = 2.0$ and Θ giving $\langle k \rangle = 10.0$, and all results are averaged over 1000 realizations. Numbers of nodes in Sierpinski geographical networks in the fifth, sixth, and seventh generations are 123, 336, and 1,095, respectively.
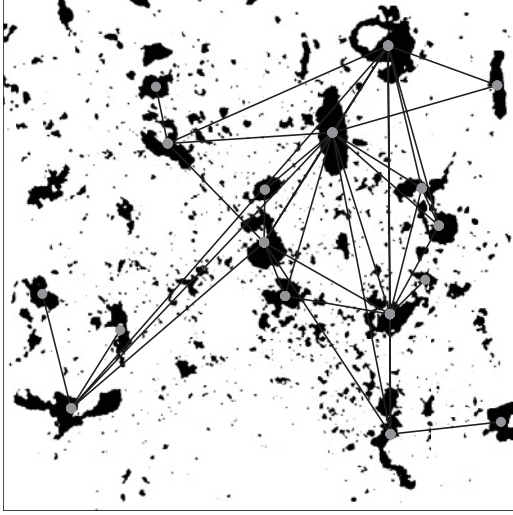
FIG. 9. 2D x-ray CT image of a soil structure with network edges connecting pores (black patches) according to Eq. (1) with $m = 5$ and $\Theta = 0.01$. From [42].

phase ($m < m_{c0}$), existing short-cut edge (in the topological sense) connects a starting node and a target node by a small number of edges with going back-and-forth around the target node. This back-and-forth motion, however, requires an extra Euclidean distance and leads the low geographical efficiency. On the other hand, in the compact phase, the network structure resembles the structure of a regular lattice (or regular Sierpinski gasket). Although we need lots of edges to connect two distant nodes, the total Euclidean length along the path can be minimized as the geodetic distance between two nodes. Thus, the geographical efficiency e becomes large in this region. For $m \gg m_{c0}$, the shortest Euclidean path often coincides with the shortest topological path though r_{ij} is typically smaller than that for $m < m_{c0}$ but d_{ij} is larger than that for $m < m_{c0}$. The above consideration supports that the global efficiency E measuring the number of edges to connect
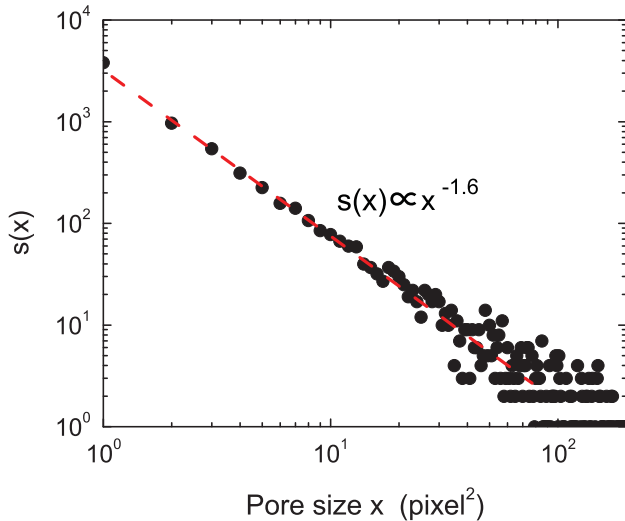

FIG. 10. (Color online) Pore-size distribution obtained from the image analysis of the soil structure shown in Fig. 9. Dashed line shows the power-law fit $s(x) \sim x^{-\alpha}$ to the distribution with $\alpha = 1.6$.
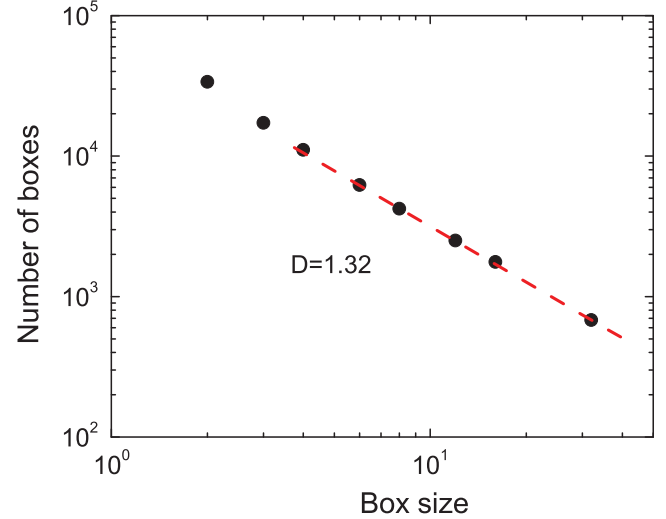

FIG. 11. (Color online) Number of boxes required to cover the soil-pore image shown in Fig. 9 as a function of the box size. The slope obtained by the least-squares fit (dashed line) indicates that the fractal dimension of the soil-pore structure is 1.32.

nodes is large for $m < m_{c0}$ and small for $m > m_{c0}$. It should be noted that the abrupt change in e or E occurs at $m = m_{c0}$ but not at $m = m_{c1}$ though the transition point m_{c1} is related to the edge length. Since e and E behave oppositely, it is crucial to clarify which efficiency is more relevant to a given problem by considering how strongly the cost of the flow is influenced by the Euclidean distance. It is also interesting that both efficiencies e and E are relatively high in the network at $m = m_{c0}$ at which the competition between order and disorder in the geometrical sense is balanced.

VIII. EXAMPLE

We will demonstrate the above described approach using soil-pore networks as an example of a real-world spatially embedded network.

Two approaches to build complex network models of soil-pore organization have recently been developed [42–44]. Here we will concentrate on the networks presented in [42] formed by geographical threshold algorithm as described in Sec. II. We consider a set of N pores representing the nodes of the network. The nodes of the network are located at the centers of the pores and the edges between nodes are drawn according to Eq. (1) with Eqs. (2) and (3), where in this case the continuous fitness variable x is the size of the pore. Pore sizes and their relative positions of actual soil specimens are obtained from the image analysis of 2D soil X-CT scans [42]. The edges between pores do not have any physical meaning in this network model. However, characterizing a soil by using the network concept provides a new method to understand properties of soils and gives a novel description of the soil pore structure.

An example of the soil-pore network overlain on the 2D soil porous structure image is presented in Fig. 9 (from [42]). The pore-size distribution of the soil sample shown in Fig. 9 is analyzed to be of the form $s(x) \sim x^{-\alpha}$ with $\alpha = 1.6$ (Fig. 10). The fractal dimension $D = 1.32$ of the soil-pore structure is also determined using box-counting method as
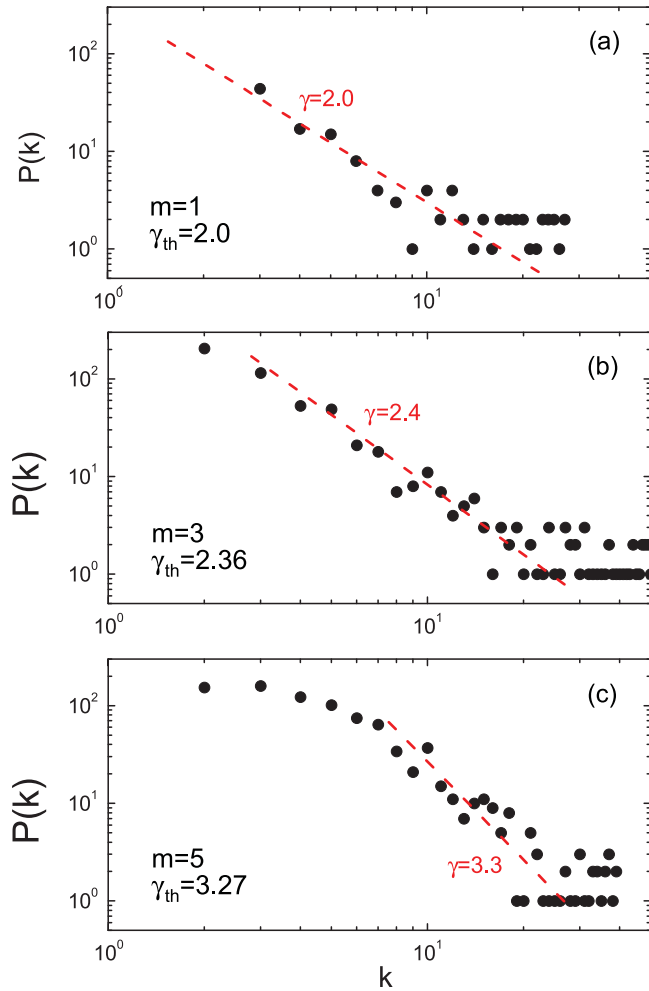
FIG. 12. (Color online) Degree distributions for three soil-pore networks formed by (a) $m = 1$, (b) $m = 3$, and (c) $m = 5$. Dashed lines through dots represent the power-law fits with the slopes indicated beside the lines. γ_{th} in each panel gives the degree exponent predicted by Eq. (11) with $D = 1.32$ and $\alpha = 1.6$.

shown in Fig. 11. Our theory predicts that the soil-pore network has a power-law degree distribution function and the degree exponent γ depends on m above $m_{c0} = 2.2$. Figure 12 shows the degree distributions for three networks based on soil image data constructed with different values of the parameter m ($m = 1$, 3, and 5). In Fig. 12 the power-law fits (dashed lines) and calculated exponents of the degree distributions are shown together with theoretically predicted scale-free exponents from Eq. (11). The results clearly demonstrate the agreement of the empirically obtained scale-free exponents and the theoretically predicted ones. This also shows how the fractality of pore

spatial distribution is reflected in the network organization. Using networks we get an independent way to measure spatial (fractal) structure of the soil. By measuring the scaling exponent γ of the degree distribution for any $m > m_{c0}$ and the exponent α of the pore-size distribution, the fractal dimension D can be determined. A possible further application of this network model might help to understand the self-organization of the soil-microbe complex [45] and the relation between the biological function and the soil-pore structure which changes between the more open (to enhance the rate of oxygen supply) and the more closed one (to protect the soil biological function) depending on the microbial activity.

IX. CONCLUSIONS

We have proposed a geographical scale-free network model with the nodes embedded in a fractal space and analytically and numerically studied several network properties. The fractal dimension D of the embedding space was found to influence the scale-free exponent as $\gamma = m(\alpha - 1)/D + 1$ only if the spatial embedding is strong enough (i.e., when $m > m_{c0}$) otherwise $\gamma = 2$. The analyses of the average degree and average edge length revealed that this type of network can exist either in the noncompact, compact, or intermediate phase depending on the importance of the spatial arrangement of nodes. We derived the edge-length distribution functions for our network model and showed that it has a peak-like structure similar to the profile of the shortest-path-distance distribution observed in a large-scale structure of the Internet [31,46]. It is interesting to apply our approach to modeling the Internet at the autonomous systems level considering the observed long-tailed distribution of autonomous systems sizes [47]. The measured degree distribution exponent of the Internet is slightly larger than $\gamma = 2$ ($\gamma = 2.1$–2.2) [31,32] and seems to be decreasing with time [32]. In our network model this would be an indication of the evolution of the Internet toward the noncompact phase ($m \to m_{c0}$).

We hope that our work will help in advancing the understanding of the complex systems in which the heterogeneity of intrinsic properties and the spatial arrangement of the elements play an important role.

[1] A.-L. Barabasi, Science **325**, 412 (2009).

[2] A.-L. Barabasi and R. Albert, Science **286**, 509 (1999).

[3] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).

[4] S.-H. Yook, H. Jeong, and A.-L. Barabàsi, Proc. Natl. Acad. Sci. U.S.A. **99**, 13382 (2002).

[5] S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature (London) **429**, 180 (2004).

[6] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins, Proc. Natl. Acad. Sci. U.S.A. **102**, 11623 (2005).

[7] M. T. Gastner and M. E. J. Newman, Eur. Phys. J. B **49**, 247 (2006).

[8] R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Doorena, Physica A **387**, 5317 (2008).

[9] G. Bianconi, P. Pin, and M. Marsili, Proc. Natl. Acad. Sci. U.S.A. **106**, 11433 (2009).

[10] M. Barthélemy, Phys. Rep. **499**, 1 (2011).

[11] R. Xulvi-Brunet and I. M. Sokolov, Phys. Rev. E **66**, 026118 (2002).

[12] M. Barthélemy, Europhys. Lett. **63**, 915 (2003).

[13] M. Kaiser and C. C. Hilgetag, Phys. Rev. E **69**, 036103 (2004).

[14] R. Xulvi-Brunet and I. M. Sokolov, Phys. Rev. E **75**, 046117 (2007).

[15] M. A. Serrano, D. Krioukov, and M. Boguñá, Phys. Rev. Lett. **100**, 078701 (2008).

[16] M. Boguñá, D. Krioukov, and K. C. Claffy, Nat. Phys. **5**, 74 (2009).

[17] M. Boguñá and D. Krioukov, Phys. Rev. Lett. **102**, 058701 (2009).

[18] D. Krioukov, F. Papadopoulos, A. Vahdat, and M. Boguñá, Phys. Rev. E **80**, 035101 (2009).

[19] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Phys. Rev. E **82**, 036106 (2010).

[20] J. Dall and M. Christensen, Phys. Rev. E **66**, 016121 (2002).

[21] C. Herrmann, M. Barthélemy, and P. Provero, Phys. Rev. E **68**, 26128 (2003).

[22] N. Masuda, H. Miwa, and N. Konno, Phys. Rev. E **71**, 036108 (2005).

[23] S. Morita, Phys. Rev. E **73**, 035104 (2006).

[24] J. M. Kleinberg, Nature (London) **406**, 845 (2000).

[25] C. P. Warren, L. M. Sander, and I. M. Sokolov, Phys. Rev. E **66**, 056105 (2002).

[26] A. F. Rozenfeld, R. Cohen, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **89**, 218701 (2002).

[27] D. ben-Avraham, A. F. Rozenfeld, R. Cohen, and S. Havlin, Physica A **330**, 107 (2003).

[28] K. Kosmidis, S. Havlin, and A. Bunde, Europhys. Lett. **82**, 48005 (2008).

[29] Y. Hayashi and Y. Ono, Phys. Rev. E **82**, 016108 (2010).

[30] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).

[31] A. Vazquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).

[32] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos, IEEE/ACM Trans. Networking **11**, 514 (2003).

[33] S. Shakkottai, M. Fomenkov, R. Koga, D. Krioukov, and K. C. Claffy, Eur. Phys. J. B **74**, 271 (2010).

[34] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz, Phys. Rev. Lett. **89**, 258702 (2002).

[35] N. Masuda, H. Miwa, and N. Konno, Phys. Rev. E **70**, 36124 (2004).

[36] D. Garlaschelli, A. Capocci, and G. Caldarelli, Nat. Phys. **3**, 813 (2007).

[37] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).

[38] F. Kawasaki and K. Yakubo, Phys. Rev. E **82**, 036113 (2010).

[39] H. D. Rozenfeld, C. Song, and H. A. Makse, Phys. Rev. Lett. **104**, 025701 (2010).

[40] C. Song, L. K. Gallos, S. Havlin, and H. A. Makse, J. Stat. Mech. Theory Exp. (2007) P03006.

[41] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).

[42] S. J. Mooney and D. Korošak, Soil Sci. Soc. Am. J. **73**, 1094 (2009).

[43] A. Santiago, J. P. Cardenas, J. C. Losada, R. M. Benito, A. M. Tarquis, and F. Borondo, Nonlin. Processes Geophys. **15**, 893 (2008).

[44] J. P. Cardenas, A. Santiago, A. M. Tarquis, J. C. Losada, F. Borondo, and R. M. Benito, Geoderma **160**, 13 (2010).

[45] I. M. Young and J. W. Crawford, Science **304**, 1634 (2004).

[46] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K. Claffy, and A. Vahdat, Comput. Commun. Rev. **36**, 17 (2006).

[47] A. Lakhina, J. W. Byers, M. Crovella, and I. Matta, IEEE J. Sel. Areas Commun. **21**, 934 (2003).