Adaptive walk on complex networks

Paulo R. A. Campos*
Departamento de Física e Matemática, Universidade Federal Rural de Pernambuco, Dois Irmãos 52171-900, Recife-Pernambuco, Brazil

F. G. Brady Moreira†
Departamento de Física, Universidade Federal de Pernambuco, Cidade Universitária 50670-901, Recife-Pernambuco, Brazil
(Received 17 January 2005; revised manuscript received 28 February 2005; published 28 June 2005)

We investigate the properties of adaptive walks on an uncorrelated fitness landscape which is established in sequence spaces of complex structure. In particular, we perform numerical simulations of adaptive walks on random graphs and scale-free networks. For the former, we also derive some analytical approximations for the density of local optima of the fitness landscape and the mean length walk. We compare our results with those obtained for regular lattices. We obtain that the density of local optima decreases as $1/z$, where z is the mean connectivity, for all networks we have investigated. In random graphs, the mean length walk $\bar{L}$ reaches the asymptotic value $e-1$ for large z, which corresponds to the result for regular networks. Although we could not find an exact estimate, we derive an underestimated value for $\bar{L}$. Unlike random graphs, scale-free networks show an upper asymptotic value of $\bar{L}$.

DOI: 10.1103/PhysRevE.71.061921 PACS number(s): 87.23.Kg, 89.75.Hc, 87.15.Aa

I. INTRODUCTION

The investigation of dynamical properties of adaptive evolution performed by populations of self-replicating entities on rugged landscapes has attracted increasing interest in the evolutionary biology community [1–6]. It is well established that species evolve by increasing their adaptation to the environment where they live [7,8]. In this sense, Wright created the metaphor of an uphill climb to the Darwinian evolution [9]. In the context of rugged landscapes and making use of Wright's metaphor evolutionary optimization is a permanent search for better local optima of the fitness landscape. Long-term evolution is thus best pictured as an adaptive walk in a discrete space of sequences. In an adaptive walk, the optimization process proceeds from an initial entity toward a local or global optimum of the fitness landscape through fitter neighbors.

All previous investigations regarding adaptive walks have been restricted to sequence spaces which are represented as hypercubes [10–12]. In this formulation, each entity is connected to its one-mutant neighbors. In a binary representation, each entity has D neighbors, where D denotes the sequence size. In the current work, we aim to study the dynamical properties of adaptive walks in more complex topologies. The main motivation of our approach regards the fact that not all sequences in genotype space are viable. Actually, a great amount of mutations are lethal and so a considerable fraction of organisms do not remain viable after suffering a single point mutation [13,14]. For instance, recent analysis of experiments in silico shows that up to 93% of sequences are not viable [13], and investigations in RNA viruses show that up to 39.6% of single point mutations have a lethal effect [15]. Thus, it seems more realistic to ponder sequence spaces where the node's connectivity is not the same for every node, as it is in hypercubes. In that context, we wish to examine how the topological properties of sequence spaces can affect the evolutionary optimization process. An interesting question is to know the strength of the influence of the distribution of connectivities, as well as the mean connectivity of the network, on the mean length walk performed by single populations before reaching local optima of fitness landscapes. For this purpose, we investigate the adaptive process on two distinct topologies: random graphs [16,17] and scale-free networks [19]. The study of random graphs is especially interesting because they have small distance between nodes and low clustering, which is exactly opposed to those features of regular networks. On the other hand, scale-free networks interpolate between an ordered finite-dimensional lattice and completely random graphs. We compare our results with those obtained for regular networks. We restrict our study to an uncorrelated fitness landscape, where we ascribe at random the fitness of each node in the sequence space. We focus our analysis primarily on the special case where the population always jumps to the most favorable neighbor. The aforementioned algorithm is referred to as gradient adaptation [2,5,10].

The paper is organized as follows. In Sec. II we describe the model together with the different topologies considered in this work. Also, we derive some important analytical results for the density of local optima of the fitness landscape and for the mean length walk. In Sec. III we present the simulation data and compare the results with our theoretical predictions.

II. THE MODEL

We consider different network topologies with fixed number of nodes N and mean coordination number z. The number of neighbors of a given node i is denoted by k_i. In the case of a regular network, as in a D-dimensional hypercube, the connectivity is exactly the same for all nodes, $k_i=D$, $i=1,\ldots,N$. In order to define a fitness landscape we ascribe to

*Electronic address: prac@ufrpe.br
†Electronic address: brady@df.ufpe.br

each node i a fitness value $f_i \in (0,1]$ which is taken from a uniform distribution. The set of values f_i, $i=1,\ldots,N$ establishes the fitness landscape. In the simulations, we perform averages over distinct sets $\{f_i,\ i=1,\ldots,N\}$ in order to estimate the relevant quantities.

Since the fitnesses f_i are uniformly distributed random variables, the fitness landscape is completely uncorrelated and holds a high degree of ruggedness. A given node i is a local optimum when its fitness value f_i is greater than the fitness values of all its neighboring nodes, i.e., $f_i > f_j$, $j=1,\ldots,k_i$. As the fitnesses f_i are uniformly distributed in the interval (0, 1], the likelihood that a given node i is a local optimum of the fitness landscape equals $f_i^{k_i}$. We can therefore calculate the density of local optima of the fitness landscape, given by

$$\eta = \sum_k \int_0^1 x^k P(k) dx, \tag{1}$$

where $P(k)$ denotes the probability distribution of connectivities.

We initiate the adaptive walk in a randomly chosen sequence (node), and then natural selection will move the population from this wild-type sequence to the fittest neighboring sequence [2,5,12]. This process is repeated until the population finds a locally optimum node. Besides the density of local optima η given by Eq. (1), we also consider in our investigation two other relevant statistical quantities: the mean length walk $\bar{L}$, which corresponds to the mean number of steps taken before reaching a local optimum, and the difference in fitness of the values reached by the population and the global optimum of the fitness landscape, which we denote by Δf_G. Since it is extremely difficult to find exact analytical expressions for the mean number of steps, here we derive an underestimated value for $\bar{L}$ in the case of random graphs. Below we describe the different topologies considered in this work.

A. Regular networks

The genotype space of sequences of size D is most easily represented as a hypercube where each node in the underlying space represents one entity, and its neighbors are those entities differing from the first by one mutation [20]. The number of links is exactly the same for all nodes. In this situation, the probability distribution is a simple function of k, $P(k)=\delta_{k,D}$, and the density of local optima becomes

$$\eta = \sum_k \int_0^1 x^k \delta_{k,D} dx = \frac{1}{D+1}. \tag{2}$$

We see that the density of states decreases as $1/(D+1)$, and for large D, $\eta \simeq 1/D$.

The expected number of steps taken in gradient adaptive walks on regular networks has been estimated by Orr [10]. Orr has demonstrated that the probability P_n that a population takes exactly n steps is

$$P_n = \frac{n}{(n+1)!}, \tag{3}$$

and the mean number of steps taken during gradient adaptation $\bar{L}$ is

$$\bar{L} = \sum_{n=1}^{\infty} n P_n = e - 1 \approx 1.72. \tag{4}$$

B. Random graphs

The random graphs were introduced by Erdös and Rényi about 40 years ago [16]. In this model, each node is connected to any other node with probability p. Such procedure produces a statistically homogeneous network, in which the connectivity distribution is Poissonian

$$P_{random}(k) = \frac{e^{-z} z^k}{k!}, \tag{5}$$

where $z=p(N-1)$ is the mean coordinate number and corresponds to the relevant parameter of the model.

By substituting Eq. (5) in Eq. (1), we obtain the following expression for the density of local optima:

$$\eta = \sum_k \int_0^1 x^k \frac{e^{-z} z^k}{k!} dx = \int_0^1 \sum_k \frac{(xz)^k}{k!} e^{-z} dx,$$

$$\eta = \frac{1}{z}(1 - e^{-z}). \tag{6}$$

For intermediate to large values of z, we should expect that the density η decreases as $1/z$.

In regular networks, one can use symmetry arguments to determine the mean length walk $\bar{L}$ [10]. For random graphs, we cannot make an exact formulation to obtain $\bar{L}$ because the node connectivities are not the same any more. Nevertheless, we can underestimate $\bar{L}$ by making the assumption that since a given node has a neighbor which is a local optimum it moves to that node. This formulation would correspond to the gradient walk procedure when this local optimum is in fact the fittest neighbor node. However, this conjecture is not always true since the fittest node is not necessarily a local optimum. The above assumption is even less robust for network with large mean connectivity z.

Henceforth, we assume that the starting point of an adaptive walk is not a local optimum of the fitness landscape. In the present underestimation of $\bar{L}$ we also suppose that $e^{-z} \ll 1$. In that case, the likelihood that the adaptive walk stops in the first step is given by

$$P_1 = 1 \times \left\{ \sum_{k=0}^{\infty} [1 - (1-\eta)^k] \times \frac{e^{-z} z^k}{k!} \right\},$$

$$P_1 = 1 - e^{-\eta z}. \tag{7}$$

The term $[1-(1-\eta)^k]$ is the probability that at least one of the neighboring nodes of the starting point S_0 is a local optimum. The sum is taken over all possible values of connectivity k of node S_0. Now, the probability that the adaptive walk stops in the second step is obtained through

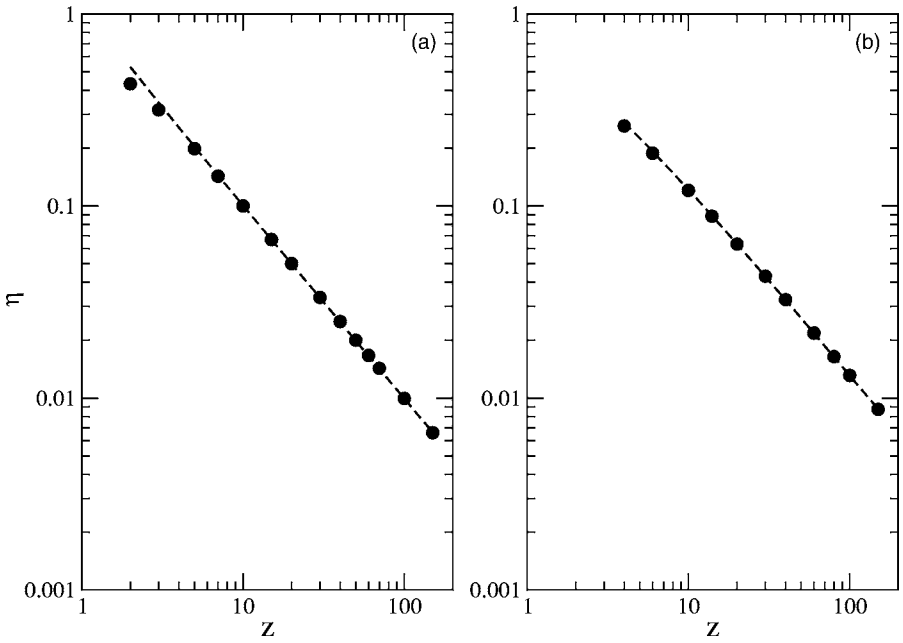

FIG. 1. Density of local optima as a function of the mean connectivity z. In (a) we have the results for random graphs and (b) for scale-free networks. The data points are the simulation data, whereas the dashed lines correspond to the fit according to $\eta \propto 1/z$. The simulation data are averages taken over 100 distinct sets of fitness values. The error bars are smaller than the symbols.

$$P_2 = 1 \times \left\{ \sum_{k=0}^{\infty} (1-\eta)^k \times \frac{e^{-z} z^k}{k!} \right\} \times \left\{ \sum_{k=0}^{\infty} [1-(1-\eta)^k] \frac{e^{-z} z^k}{k!} \right\},$$

$$P_2 = e^{-\eta z} \times (1 - e^{-\eta z}) = e^{-\eta z} - e^{-2\eta z}. \tag{8}$$

The term $e^{-\eta z}$ is just the probability that none of the neighbors of node S_0 is a local optimum. The second term $(1 - e^{-\eta z})$ is the chance that the second visited node S_1 has at least one local optimum in its neighborhood. In general, the probability that adaptation stops at step n (stops at node S_n) is simply given by

$$P_n = [e^{-\eta z}]^{n-1} \times (1 - e^{-\eta z}) = e^{-(n-1)\eta z} - e^{-n\eta z}. \tag{9}$$

We estimate the mean walk length as

$$\bar{L}_u = \sum_{n=0}^{\infty} n P_n,$$

$$\bar{L}_u = \sum_{n=0}^{\infty} n (e^{-(n-1)\eta z} - e^{-n\eta z}),$$

$$\bar{L}_u = \sum_{n=0}^{\infty} e^{-n} = \frac{e}{e-1}. \tag{10}$$

To obtain $\bar{L}_u$, we have used $\eta = 1/z$, which is the limiting value of η for large z [see Eq. (6)].

C. Scale-free networks

Recent investigations show that such distinct systems as the world-wide web [21] and scientific [22] and biochemical networks [23,24] self-organize into a scale-free state [19]. All those investigations claim that the probability P_k that a given node has k edges follows a power-law like

$$P_k \propto k^{-\gamma}. \tag{11}$$

Such systems are referred to as scale-free networks. The two mechanisms which are responsible for the emergence of scale-free patterns are the growth and preferential attachment [19,21,25]. The former mechanism tell us that scale-free networks are created in a dynamical way. The second mechanism means that news nodes are most likely attached to vertices with higher connectivities. In the context of adaptive walks, the study of scale-free networks is especially interesting because of their appearance in natural networks [26,27]. Recently, it has been verified that preferential attachment governs the protein network evolution [26]. Doye has found a scale-free character in the potential energy landscape [27], but in contrast to other scale-free networks where the topology results from the dynamics of growth, the potential energy landscape is a static entity [28–31]. Actually a straightforward relation among the guiding forces responsible for the appearance of scale-free patterns and the already witnessed evolutionary forces in a biological context emerges if we interpret natural selection as a preferential attachment mechanism [32]. In our simulations, we generate scale-free networks by means of the growth and preferential attachment mechanisms [19,21]. These mechanisms generate scale-free networks with exponent $\gamma = 3$ when the number of nodes N becomes infinitely large [21].

III. NUMERICAL RESULTS AND DISCUSSION

In Fig. 1 we show the density of local optima η as a function of the mean connectivity z. In part (a) we plot the simulation data for random graphs with fixed number of nodes $N = 32\,768$. The straight line corresponds to the theoretical prediction of Eq. (6) for large z. We have observed a

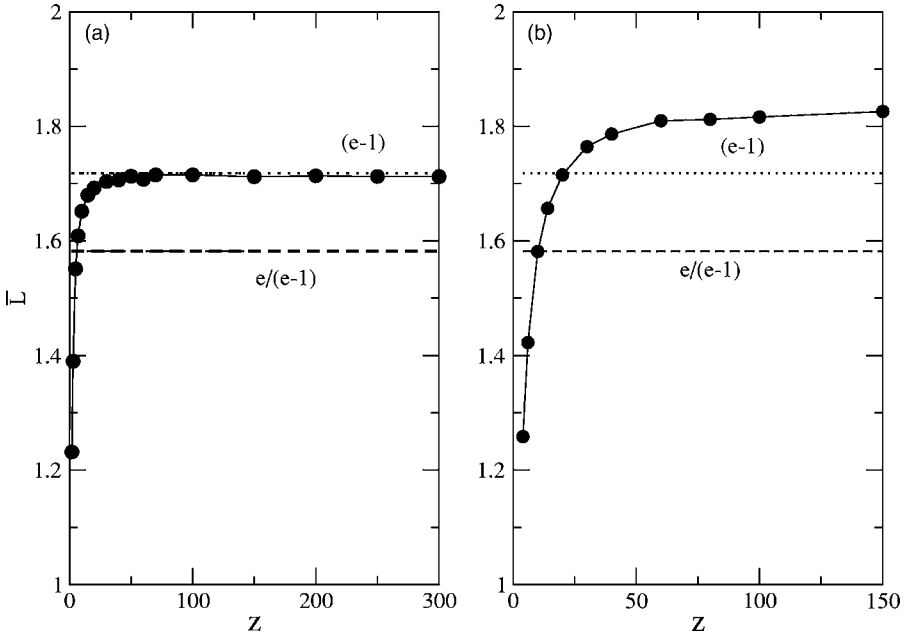

FIG. 2. Mean length walk $\bar{L}$ as a function of the mean connectivity z. The data points correspond to the results from computer simulations. The dotted line represents $\bar{L}=e-1$, whereas the dashed line corresponds to the estimate in Eq. (10). In the figure, we have considered networks with $N=32\,768$ nodes. In (a) we have the results for random graphs and in (b) for scale-free networks. The data points correspond to averages over 100 000 replicates. In (a) for $z\geqslant 200$ averages were taken over 500 000 replicates. The error bars are smaller than the symbols.

good agreement between theory and numerical simulations. Part (b) of Fig. 1 displays the density of local optima as a function of the mean connectivity $z=2m$ (where m is the minimum number of links of a given node) for scale-free networks. We also observe that the simulation data are also well described by $\eta\propto 1/z$. These results suggest that the density of local optima is insensitive to the particular distribution of connectivities of the networks, but depends on their mean connectivities only.

Figure 2 shows the mean length walk $\bar{L}$ as a function of z. For small z, we notice an abrupt increase of $\bar{L}$ as z grows. For random graphs [Fig. 2(a)], for one further augment z, $\bar{L}$ approaches the asymptotic value $e-1$ which is represented by the dotted line. In the same figure, the dashed line corresponds to the underestimated mean length walk $\bar{L}_u=e/(e-1)$ [Eq. (10)]. We check that for $z=2$ and 3, the simulation results lie below the estimated value $\bar{L}_u$, which is actually expected since we have assumed in the analytical calculation of $\bar{L}_u$ that $e^{-z}\ll 1$. We remark that for small z, the gradient adaptive walk is performed by considering the percolating cluster [17]. For instance, when $z=2$ only 79% of the nodes belong to the giant cluster [18], and so the effective size of the network is smaller than N. When $z\geqslant 5$, the simulation data surpass $\bar{L}_u$ and the disagreement $\bar{L}-\bar{L}_u$ increases with z, up to $z\approx 50$. Right away, $\bar{L}_u$ underestimates the mean length walk by about 9%. Interestingly, the asymptotic value of $\bar{L}$ coincides exactly with Orr's prediction for $\bar{L}$ in hypercubes, where $\bar{L}=e-1\approx 1.72$ steps. In part (b) of Fig. 2 we plot the mean length walk $\bar{L}$ for scale-free networks. Although we notice that $\bar{L}$ behaves in a similar way as in part (a), the asymptotic value of $\bar{L}$ is higher than $(e-1)$. A possible explanation for this fact is that in scale-free networks, the nodes are most likely connected to *hubs* (highly connected nodes), and so the probability that a hub is a locally optimum node decreases exponentially with increasing connectivity.

Figure 3 displays $\Delta f_G=f_{global}-f_{local}$, where f_{global} denotes the expected maximum value of a set of N random numbers, and f_{local} means the expected value of fitness of local optima reached by the population. For uniformly distributed random variable f_{global} is simply $N/(N+1)$. Like the density of local optima, Δf_G also decreases as $1/z$ and is not influenced by the network topology, as we can infer from collapse of the data points.

In summary, we have investigated the influence of the topology of sequence space on the dynamical properties of adaptive walks performed by single populations. We have considered an uncorrelated fitness landscape and compare our results for random graphs and scale-free networks with those for regular lattice. For random graphs we present ana-

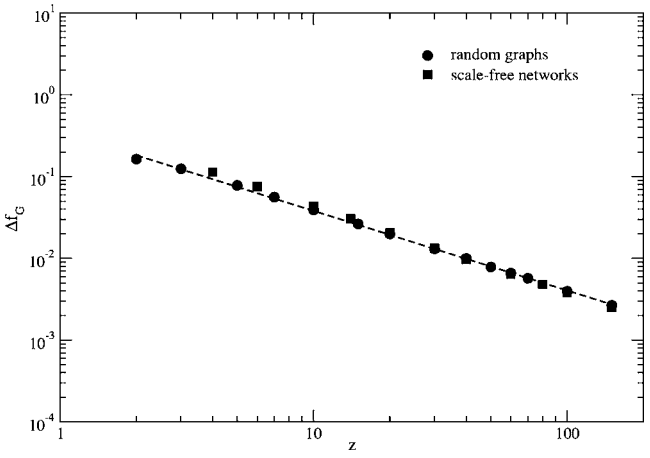

FIG. 3. The quantity Δf_G as a function of the mean connectivity z. The circles denote the results for random graphs, and the squares denote data for scale-free networks. The dashed line represents the best fitting according to $\Delta f_G\propto 1/z$. The data points correspond to averages over 100 000 replicates. The error bars are smaller than the symbols.

lytical expressions for the density of local optima and mean length walk. We expect that these expressions together with the results for hypercubes [10] will be useful for comparison in future works on adaptive processes. Remarkably, we find that the density of local optima does not depend on the particular form of the distribution of connectivities. Indeed, the density of local optima decreases as $1/z$, where z is the mean connectivity of the network. For random graphs we have determined exactly the dependence of the density η with z and have also obtained an underestimated value for the mean length walk. Interestingly, we have ascertained that the asymptotic value of $\bar{L}$, which corresponds to large z, is the same as that in regular networks. However, we observe an upper limit when dealing with scale-free networks.

ACKNOWLEDGMENT

The authors acknowledge the financial support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

[1] W. Fontana, W. Schnabl, and P. Schuster, Phys. Rev. A **40**, 3301 (1989).
[2] W. Fontana, D. A. Konings, P. F. Stadler, and P. F. Schuster, Biopolymers **33**, 1389 (1993).
[3] P. R. A. Campos, C. Adami, and C. O. Wilke, Physica A **304**, 295 (2002).
[4] W. Fontana, P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster, Phys. Rev. E **47**, 2083 (1993).
[5] E. D. Weinberger, Phys. Rev. A **44**, 6399 (1991).
[6] P. R. A. Campos and V. M. de Oliveira, Int. J. Mod. Phys. C **13**, 1003 (2002).
[7] P. R. A. Campos and V. M. de Oliveira, Evolution (Lawrence, Kans.) **58**, 932 (2004).
[8] P. R. A. Campos, V. M. de Oliveira, R. Giro, and D. S. Galvão, Phys. Rev. Lett. **93**, 098107 (2004).
[9] S. Wright, Genetics **16**, 97 (1931).
[10] H. A. Orr, J. Theor. Biol. **220**, 241 (2003).
[11] S. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).
[12] S. A. Kauffman, *The Origins of Order* (Oxford University Press, Oxford, 1992).
[13] R. E. Lenski, C. Ofria, T. C. Collier, and C. Adami, Nature (London) **400**, 661 (1999).
[14] S. F. Elena and R. E. Lenski, Nature (London) **390**, 395 (1997).
[15] R. Sanjuán, A. Moya, and S. F. Elena, Proc. Natl. Acad. Sci. U.S.A. **101**, 8396 (2004).
[16] P. Erdos and A. Rényi, Publ. Math., Inst. Hung. Acad. Sci. **5**, 17 (1960).
[17] L. F. C. Pereira and F. G. B. Moreira, Phys. Rev. E **71**, 016123 (2005).
[18] J. Dall and M. Christensen, Phys. Rev. E **66**, 016121 (2002).
[19] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[20] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[21] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[22] M. E. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[23] H. Jeong *et al.*, Nature (London) **407**, 651 (2000).
[24] H. Jeong, S. Mason, A. L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[25] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[26] E. Eisenberg and E. Y. Levanon, Phys. Rev. Lett. **91**, 138701 (2003).
[27] J. P. K. Doye, Phys. Rev. Lett. **88**, 238701 (2002).
[28] D. Garlaschelli and M. I. Loffredo, Phys. Rev. Lett. **93**, 188701 (2004).
[29] M. Baiesi and S. S. Manna, Phys. Rev. E **68**, 047103 (2003).
[30] J. Berg and M. Lässig, Phys. Rev. Lett. **89**, 228701 (2002).
[31] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[32] P. R. A. Campos and V. M. de Oliveira, Physica A **325**, 570 (2003).