PHYSICAL REVIEW E **70**, 061914 (2004)

Model ecosystems with random nonlinear interspecies interactions

Danielle O. C. Santos and José F. Fontanari
Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos, São Paulo, Brazil

(Received 29 June 2004; published 29 December 2004)

The principle of competitive exclusion in ecology establishes that two species living together cannot occupy the same ecological niche. Here we present a model ecosystem in which the species are described by a series of phenotypic characters and the strength of the competition between two species is given by a nondecreasing (modulating) function of the number of common characters. Using analytical tools of statistical mechanics we find that the ecosystem diversity, defined as the fraction of species that coexist at equilibrium, decreases as the complexity (i.e., number of characters) of the species increases, regardless of the modulating function. By considering both selective and random elimination of the links in the community web, we show that ecosystems composed of simple species are more robust than those composed of complex species. In addition, we show that the puzzling result that there exists either rich or poor ecosystems for a linear modulating function is not typical of communities in which the interspecies interactions are determined by a complementarity rule.

DOI: 10.1103/PhysRevE.70.061914 PACS number(s): 87.23.Cc, 75.10.Nr

I. INTRODUCTION

Assemblies of biological species—ecosystems—can be viewed as systems of elements that influence each other through competition and cooperation [1]. Mathematical models advanced to describe the dynamics of these systems are usually couched in terms of nonlinear rate equations which are then integrated numerically, especially in the case of many interacting species [2]. For large ecosystems, however, an analytical statistical approach is more insightful since it is practically impossible to explore fully the space of parameters associated with the rate constants and interaction strengths solely through numerical means. Unfortunately, the methods of statistical mechanics can be applied only in very particular situations—namely, when the interactions are symmetric so that the dynamics maximize a quadratic Lyapunov function [3] and when the interactions are antisymmetric so that a constant of motion can be identified [4]. This contribution focuses on the former case, since it can be thoroughly investigated using tools of equilibrium statistical mechanics.

The uncertainty and complexity of the interspecies interactions in nature grounds the suggestion that those interactions should be considered as random variables. This approach may be viewed as a null hypothesis that provides a standard to which the behavior pattern of realistic ecosystems should be compared. In particular, the measures of how the encounter between ith and jth species affects the growth of species i, denoted by J_{ij} for $i,j=1,\ldots,N$, are assumed to be independent, Gaussian-distributed random variables [5]. This setting, however, fails to take into account the possibility of an underlying, nonrandom structure in the interspecies interactions, which might explain cooperative behavior as cross feeding and symbioses as well as the competitive behavior that results from the exploitation of the same niche by different species. Here we explore an alternative model ecosystem in which the pairwise species interactions are regulated by the degree of complementarity between species—i.e., by the number of features or characters that distinguish the interacting species [6]. Since the assignment of the characters to each species is done randomly, the model retains the randomness ingredient of the Gaussian model, while accounting for an explicit, biologically motivated structure for the interspecies interactions.

Borrowing from models for molecular recognition used in the study of the immune [7] and olfactory [8] systems, it is assumed that each species is characterized by a set of p phenotypic characters, $\mu=1,\ldots,p$. The resulting interaction between a pair of species—say, i and j—depends on the presence or not of the same character in both species according to a complementarity principle. Explicitly, we assume that the interactions are functions of the Hebb rule

$$J_{ij} = \frac{\sqrt{p}}{N} F\left(\frac{1}{\sqrt{p}} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \right), \quad i \neq j, \tag{1}$$

where the ξ_i^μ are quenched, independent random variables that take on the values ±1 with equal probability. If species i exhibits character μ, then ξ_i^μ is set to 1; otherwise, it is set to -1. In what follows we will assume that the number of characters is extensive—i.e., $p=\alpha N$. Strictly, taxonomists define species by searching for a reliable group of characters—i.e., characters that are possessed by all members of the species to be defined and not by members of other species [9]. Here we take a slightly different approach and define species by a list of p identifiable morphological characters that a particular species may or may not possess. Although we can guarantee that each species is assigned a unique set of characters (the probability that two or more species are assigned the same set of characters vanishes as $2^{-\alpha N}N^2$ in the limit of large N), the odds of finding a reliable character to describe a species in the taxonomic sense are negligible (it vanishes as 2^{-N} for large N). However, since for some closely related species no reliable characters have been identified, taxonomists have to resort to a global approach akin to ours [9].

The modulating function $F(x)$ in Eq. (1) stands for any *odd* nondecreasing function of x. Hence, $\langle J_{ij}\rangle=0$ and

$$\langle J_{ij}^2 \rangle = \frac{\alpha}{N} \int_{-\infty}^{\infty} Dz F^2(z), \tag{2}$$

where $Dz=\exp(-z^2/2)dz/\sqrt{2\pi}$ is the Gaussian measure. Up to now only the linear case [i.e., $F(x)=x$] was considered in the literature [6,12]. Invoking Gause's principle of competitive exclusion [13]—two species living together cannot occupy the same ecological niche—we assume that the larger the number of features shared by a pair of species (matches), the stronger the competition between them, so that $J_{ij}>0$ corresponds to pairs of competing species whereas $J_{ij}<0$ to pairs of cooperating species. The nonlinear Hebb rule was proposed and studied within the context of attractor neural networks [10,11], as a more realistic, from the biological viewpoint, generalization of the standard (linear) Hebb rule (see [14] for a review). This generalization seems also appropriate in the present context of model ecosystems. For example, one could envisage a situation in which the interaction becomes effective (nonzero) provided the net number of matches exceeds a threshold value or the interaction could simply saturate after a certain number of matches, the extreme situation being the clipped interaction $F(x)=\mathrm{sgn}(x)$.

In this contribution we use the replicator dynamics framework to model the evolution of an isolated community of interacting species [15]. As pointed out, in the case of symmetric interactions $J_{ij}=J_{ji}$, such as the nonlinear Hebb rule, this dynamics maximizes a quadratic fitness functional so that standard tools of the statistical mechanics of disordered systems can be used to characterize its stationary states. Actually, the statistical characterization of these states is possible thanks to the remarkable result derived by Sompolinsky that in the limit of extensive p the model with nonlinear interactions is equivalent to the model with linear interactions plus a Gaussian static noise, the variance of which is determined by the nonlinear function [10]. In what follows we will use that finding together with the replica method to study how the choice of the modulating function affects the statistical properties of the stationary states of the replicator dynamics.

II. MODEL

We assume that the abundance of individuals of species $i=1,\ldots,N$ in the ecosystem, $x_i\in[0,\infty)$, is governed by the nonlinear system of equations, so-called replicator equations,

$$\frac{dx_i}{dt}=x_i\left(\mathcal{F}_i-\frac{\phi}{N}\right), \qquad (3)$$

where $\mathcal{F}_i=-\Sigma_j J_{ij}x_j$ can be identified with the fitness of species i and the term $\phi=\Sigma_i x_i\mathcal{F}_i$ ensures that the total concentration

$$\sum_{i=1}^{N}x_i=N \qquad (4)$$

is kept fixed.

Apparently, this framework seems to be suitable only to study chemostat cultures of microrganisms where the total concentration is regulated to a constant level via some flux control mechanism. However, since it was demonstrated that the replicator dynamics for N species is equivalent to the Lotka-Volterra equation for $N-1$ species, this formulation can actually be applied to describe a much broader class of ecosystems [15]. Finally, to prevent the unbounded growth of any single species in the thermodynamic limit $N\to\infty$, we introduce a competition term between individuals of the same species by setting $J_{ii}=u>0$ for all i. Hence the parameter u will be referred to as the intraspecies interaction.

In the case of symmetric interactions the asymptotic regime of Eq. (3) can be characterized by looking at the maxima of the fitness functional

$$\mathcal{F}(\{x_i\})=-\sum_{i,j}J_{ij}x_ix_j, \qquad (5)$$

and so it can be shown that the only stationary states are fixed points [15]. Following Sompolinsky, in the case that the interactions J_{ij} are given by the nonlinear Hebb rule (1) and the number of characters is extensive, we write the fitness functional as [10]

$$\mathcal{F}(\{x_i\})=-\sum_{i\neq j}\left(\frac{J}{N}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu+\delta_{ij}\right)x_ix_j-u\sum_i x_i^2, \qquad (6)$$

where δ_{ij} is a random Gaussian noise of zero mean and variance δ^2/N with $\delta^2=\alpha(\hat{J}-J^2)$. Here,

$$J=\int_{-\infty}^{\infty}DzzF(z), \qquad (7)$$

$$\hat{J}=\int_{-\infty}^{\infty}DzF^2(z). \qquad (8)$$

Note that $\hat{J}$ is proportional to the variance of the interspecies interactions J_{ij} given in Eq. (2). It must be stressed that the equivalence between the model with nonlinear interactions and the model in which a Gaussian noise is added to the linear interactions is valid only if $F(x)$ is odd, p is large, and the expected value of random variables ξ_i^μ is zero.

Before embarking on the statistical mechanics analysis of our ecosystem model, it is instructive to compare it with a related model recently proposed in the literature—the tangled nature model [16,17]. Inspired in models of adaptive walks on complex fitness landscapes [18] (see [19] for a formulation of stochastic walks which is more pertinent for this discussion) this model identifies each individual in a population of variable size by a binary string of fixed length: the individual's genome. The fitness of a given individual is proportional to the overlap between its genome and the average genome of rest of the population, weighted by a random interaction term that depends on the genomes of the interacting individuals. The main effect of this random term is to break the symmetry of the interactions. The individual's genome is akin to the feature vector $\xi_i=(\xi_i^1,\ldots,\xi_i^p)$ and the dependence of the fitness on the overlap between genomes is similar to that exhibited in Eq. (6). A crucial difference between the models, aside from the symmetry of the interspecies interactions, is that in the tangled nature model the genomes can change via mutations, leading to a very rich dynamics characterized by long stasis periods and very sharp transitions. Moreover, the notion of species appears as an emergent structure in the distribution of the population in genome space [16,17]. As expected, the complexity of the tangled nature model precludes an analytical approach to study its dynamics and the investigations of that model were limited so far to computer simulations. It is interesting to note, however, that it is possible to study analytically within the equilibrium statistical mechanics framework the coupled dynamics of species and features by exploring the physical interpretation of the finite number of replicas proposed by Coolen *et al.* [20]. We shall pursue this research line in a future contribution.

III. REPLICA APPROACH

The maxima of the fitness functional can be obtained within the statistical mechanics framework in the limit of infinite N (and hence p). Since the presence or absence of a character in a given species is decided randomly, we must resort to the replica approach in order to obtain the average free-energy density from which the relevant parameters of the model can immediately be derived. We begin by defining the average free-energy density f as

$$-\beta f = \lim_{N\to\infty}\frac{1}{N}\langle \ln Z\rangle, \tag{9}$$

where

$$Z = \int_0^\infty \prod_i \mathrm{d}x_i\, \delta\!\left(N - \sum_i x_i\right)\exp[\beta\mathcal{F}(\{x_i\})] \tag{10}$$

is the partition function and $\beta=1/T$ is the inverse temperature. The notation $\langle\cdots\rangle$ stands for the average over the probability distribution of the quenched random variables ξ_i^μ and the Gaussian noise δ_{ij}. Taking the limit $T\to 0$ in Eq. (10) ensures that only the states that maximize $\mathcal{F}$ will contribute to Z. As usual, the evaluation of the quenched average can be carried out through the replica method, which consists of calculating $\langle Z^n\rangle$ for *integer* n—i.e., $Z^n=\prod_{a=1}^n Z^a$—and then using the identity

$$\langle \ln Z\rangle = \lim_{n\to 0}\frac{1}{n}\ln\langle Z^n\rangle, \tag{11}$$

in which the analytical continuation to $n=0$ is implicit.

After standard algebric manipulations that include saddle-point integrations in the limit of large N, we find

$$\begin{aligned} f = -\lim_{n\to 0}\operatorname{extr}\frac{1}{n}\Bigg\{ & \beta\sum_{a<b} q_{ab}(2\delta^2 q_{ab} - \hat q_{ab}) + \sum_a r_a \\ & + \sum_a Q_a[\beta(\hat Q_a + \delta^2 Q_a) - u + \alpha J] + \frac{1}{\beta}\ln G_0(\hat q_{ab},\hat Q_a, r_a) \\ & + \frac{\alpha}{\beta}\ln G_1(q_{ab},Q_a)\Bigg\}, \end{aligned} \tag{12}$$

where

$$G_0 = \int_0^\infty \prod_a dx_a \exp\left[-\beta\sum_a x_a(r_a + \beta\hat Q_a x_a) + \beta^2\sum_{a<b}\hat q_{ab}x_a x_b\right] \tag{13}$$

and

$$\begin{aligned} G_1 = \int_{-\infty}^\infty \prod_a \frac{dy_a}{\sqrt{2\pi}}\exp\Bigg[& -\frac{1}{2}\sum_a y_a^2(1+2\beta J Q_a) \\ & - 2\beta J\sum_{a<b} q_{ab}y_a y_b\Bigg]. \end{aligned} \tag{14}$$

The relevant physical order parameters are

$$q_{ab} = \frac{1}{N}\sum_i \langle\langle x_{ia}x_{ib}\rangle_T\rangle, \quad a<b, \tag{15}$$

and

$$Q_a = \frac{1}{N}\sum_i \langle\langle x_{ia}^2\rangle_T\rangle, \tag{16}$$

which measure the overlap between a pair of stationary states labeled by the replica indices a and b and the overlap between the stationary state labeled by a with itself. Here $\langle\cdots\rangle_T$ stands for a thermal average taken with the Gibbs distribution

$$\mathcal{W}(\{x_i\}) = \frac{1}{Z}\delta\!\left(N - \sum_i x_i\right)\exp[\beta\mathcal{F}(\{x_i\})]. \tag{17}$$

To proceed further and evaluate the extremum in Eq. (12) we must make some simplifying assumption about the structure of the saddle-point parameters.

A. Replica-symmetric solution

The simplest guess is that the saddle-point parameters are symmetric under permutations of the replica indices—i.e., $q_{ab}=q$, $\hat q_{ab}=\hat q$, $Q_a=Q$, $\hat Q_a=\hat Q$, and $r_a=r$. This prescription facilitates greatly the evaluation of the integrals in Eqs. (13) and (14), resulting in the following replica-symmetric free energy density

$$\begin{aligned} f_T = {} & \frac{\beta}{2}q(2\delta^2 q - \hat q) + \frac{\alpha J q}{1+2\beta J(Q-q)} - Q[\beta(\hat Q + \delta^2 Q) \\ & - u + \alpha J] - r + \frac{1}{2\beta}\ln(\hat Q + \hat q/2) - \frac{1}{2\beta}\ln(\pi/4\beta^2) \\ & + \frac{\alpha}{2\beta}\ln[1+2\beta J(Q-q)] - \frac{1}{\beta}\int Dz\,\ln[\mathrm{e}^{\Xi_z^2}\operatorname{erfc}(\Xi_z)], \end{aligned} \tag{18}$$

where

$$\Xi_z = \frac{r - \hat q^{1/2}z}{2(\hat Q + \hat q/2)^{1/2}}. \tag{19}$$

In this framework the definitions (15) and (16) become

$$q = \frac{1}{N}\sum_i \langle\langle x_i\rangle_T^2\rangle, \tag{20}$$

$$Q = \frac{1}{N}\sum_i \langle\langle x_i^2\rangle_T\rangle, \tag{21}$$

where the thermal average is now calculated using the replica-symmetry prescription. Hence $Q-q>0$ is the average variance of the species concentration at equilibrium. The saddle-point parameters q, Q, r, $\hat{q}$, and $\hat{Q}$ can be obtained by solving the five coupled nonlinear equations that result from extremizing the free energy with respect to each of them. Since our aim is to single out the states that maximize the fitness functional, we must consider the zero-temperature limit $\beta\to\infty$ of the saddle-point equations. In this limit it is convenient to introduce the auxiliary parameters $v=2\beta(Q-q)$, $\Delta=r/\hat{q}^{1/2}$ and $\lambda=2\beta(\hat{Q}+\hat{q}/2)$ so that the saddle-point equations are written as

$$v = \frac{1}{\lambda}\int_\Delta^\infty Dz\, z(z-\Delta), \tag{22}$$

$$1 = \frac{\hat{q}^{1/2}}{\lambda}\int_\Delta^\infty Dz(z-\Delta), \tag{23}$$

$$\lambda = u - \delta^2 v - \frac{\alpha v J^2}{1+vJ}, \tag{24}$$

$$Q = \frac{\hat{q}^{1/2}}{\lambda}(\hat{q}^{1/2}v - \Delta), \tag{25}$$

$$\hat{q} = Q\left[\delta^2 + \frac{\alpha J^2}{(1+vJ)^2}\right]. \tag{26}$$

Using Eqs. (22) and (23) we can rewrite Eq. (25) as

$$Q = \frac{\int_\Delta^\infty Dz(z-\Delta)^2}{\left[\int_\Delta^\infty Dz(z-\Delta)\right]^2} \tag{27}$$

in order to make clear that Q is a positive quantity. Finally, taking the zero-temperature limit in Eq. (18) and using these saddle-point equations to simplify and rearrange the terms yields the expression

$$f_0 = \frac{Q}{2}(\lambda - \delta^2 v) - \frac{\alpha J^2 Q v}{2(1+vJ)^2} \tag{28}$$

for the average fitness of the population. The results for the standard random replicator model are recovered by setting $J=0$ and $\delta^2=1$ [5] and those for the linear Hebb rule by setting $J=1$ and $\delta^2=0$ [12]. Except for Q, which is proportional to the probability that two randomly selected individuals belong to the same species, a measure known as Simpson's index [21], and the suceptibility v that measures the fluctuations around the average species concentration, knowledge of the saddle-point parameters is not very illuminating. However, they are necessary to the evaluation of more informative quantities such as the values that the species concentrations x_i can take on. This can be done by calculating explicitly the cumulative distribution that the concentration of a given species—say, x_k—assumes a value smaller than x, defined by

$$C_k(x) = \lim_{\beta\to\infty}\int_0^\infty \prod_j dx_j \Theta(x-x_k)\mathcal{W}(\{x_i\}), \tag{29}$$

where $\Theta(x)=1$ if $x\geqslant 0$ and 0 otherwise, and $\mathcal{W}(\{x_i\})$ is given by Eq. (17). Since all species concentrations are equivalent, we can write $C_k(x)=C(x)\,\forall\, k$ and evaluate Eq. (29) by adding the field term $h\Sigma_i\Theta(x-x_i)$ to Eq. (6). Taking the derivatives of the resulting free energy with respect to h and then the limit $h\to 0$ yields

$$C(x) = 1 - \frac{1}{2}\,\mathrm{erfc}\left[\frac{\Delta\hat{q}^{1/2} + \lambda x}{(2\hat{q})^{1/2}}\right]. \tag{30}$$

An interesting outcome of the random replicator model is that a fraction $C(0)$ of the species are extinct in the stationary regime [5]. Since the mechanism of extinction is clearly outcompetition, the extinction sizes must be determined by the form of the interspecies interactions. To investigate quantitatively this effect we define the diversity index,

$$d = \frac{1}{2}\,\mathrm{erfc}(\Delta/\sqrt{2}), \tag{31}$$

as the fraction of surviving species at equilibrium. We note that there seems to exist a remarkable correlation between the reciprocal of the parameter Q and the diversity index [22]. In fact, the larger the fraction of surviving species, the smaller the probability that two randomly chosen individuals are of the same species.

B. Stability analysis

Before we begin the presentation and discussion of the numerical solution of the saddle-point equations it is important to check that the replica-symmetric solution is in fact locally stable. An instability of the solution is determined by a sign change in (at least) one of the eigenvalues of the matrix of quadratic fluctuations around the replica-symmetric solution. Following the standard stability analysis [23] it can be shown that the stability is determined by the eigenvalues of the matrix

$$\begin{pmatrix} \partial^2 G_0 & -1 \\ -1 & 4\beta^2\delta^2 + \alpha\partial^2 G_1 \end{pmatrix}, \tag{32}$$

where $\partial^2 G_0$ is the $[(1/2)n(n+3)]$-dimensional matrix of second derivatives of G_0 $(\hat{q}_{ab},\hat{Q}_a,r_a)$ with respect to its arguments. Similarly, $\partial^2 G_1$ is the $[(1/2)n(n+1)]$-dimensional matrix of second derivatives of G_1 with respect to q_{ab} and Q_a. Requiring all the eigenvalues of this matrix to be positive leads to the following condition (see [24] for a similar calculation)

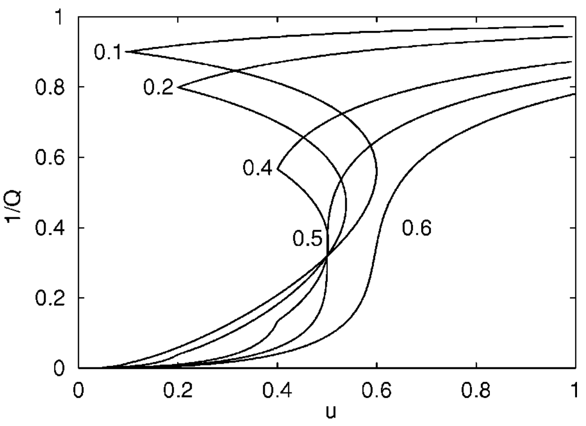

FIG. 1. Reciprocal of the saddle-point parameter Q as function of u for linear interactions and α as indicated in the figure. For $u > \alpha$ and $\alpha < 1/2$ only the higher branch (larger $1/Q$) is physical.

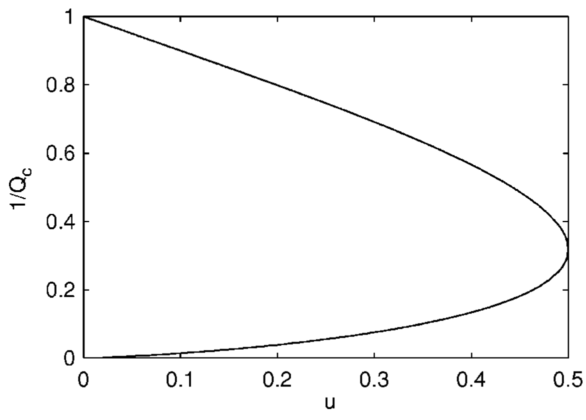

FIG. 2. The reciprocal of the saddle-point parameter Q at $u = \alpha \leqslant 0.5$ where the suceptibility v diverges in the case of linear interactions.

$$\Lambda = \left[\delta^2 + \frac{\alpha J^2}{(1+vJ)^2}\right]\frac{d}{\lambda^2} < 1. \tag{33}$$

We note that the factor d defined by Eq. (31) is missing in an analogous formula presented in Ref. [12]. This factor is necessary to recover the results for the Gaussian model [25].

IV. RESULTS

To solve numerically the saddle-point equations (22)–(26) one has to fix the control parameters u and α as well as the function $F(x)$. To set a standard for comparing the effects of different choices of the function $F(x)$ we discuss first the linear case—i.e., $F(x)=x$ (see [12] for a preliminary analysis of this limit).

A. Linear interactions

This case is obtained by setting $\delta^2=0$ and $J=1$ in the saddle-point equations. The trick to solve those equations is to consider Δ as a fixed, given parameter and u as unknown. Hence by varying Δ we can find the saddle-point solutions for different values of u. This procedure is effective because in the linear case we can write all sadddle-point parameters as explicit functions of Δ as in the case of Q [see Eq. (27)] and v,

$$v = \frac{[\alpha Q(\Delta)]^{1/2}}{[\alpha Q(\Delta)]^{1/2} - R(\Delta)}, \tag{34}$$

where $R(\Delta) = \int_\Delta^\infty Dz z(z-\Delta) / \int_\Delta^\infty Dz(z-\Delta)$ and we have made explicit the dependence of Q on Δ through Eq. (27). The final result is summarized in Fig. 1 where we show the dependence of the reciprocal of Q on the parameter u. The unusual feature is the appearance of three solutions in the regime $u \geqslant \alpha$ and $\alpha < 1/2$. However, only the solution corresponding to the larger value of $1/Q$ (upper branch) is physical. The other two solutions correspond to negative values of the suceptibility v and so they must be discarded. In fact, the regime of multiple solutions occurs when the denominator of Eq. (34) is negative and the onset of this regime is determined by the vanishing of this denominator and the consequent divergence of v. As it is clear from Eq. (22), $v \to \infty$ is solution provided that $\lambda \to 0$, which, according to Eq. (24), implies $u=\alpha$. The precise values of Δ (and hence of $1/Q$) at which this divergence takes place are obtained by solving $\alpha Q(\Delta) - R^2(\Delta) = 0$ and the solutions, in terms of $1/Q$, are shown in Fig. 2 as function of $u=\alpha$. This equation does not have solutions for $\alpha > 1/2$ and hence there is no divergence of the susceptibility in this regime. In particular, we find $Q_c = \pi$ at $\alpha = u = 1/2$.

The discontinuity of the saddle-point parameter Δ at $u = \alpha < 1/2$ resulting from the dropping of the unphysical solution is, of course, reflected on the diversity measure. Figure 3 illustrates this point and highlights the remarkable result that for certain choices of the intraspecies competition parameter $(u < 1/2)$ there are either rich or poor ecosystems: intermediate diversity values are simply forbidden. We have no physical or biological intuition to explain this behavior pattern, especially because it seems to be unique to the linear case. In fact, from the mathematical viewpoint, the saddle-point equations admit a solution with diverging suceptibility v because of the simultaneous vanishing of λ such that the product λv goes to a finite constant according to Eq. (22). Inspection of Eq. (24) shows that such a solution is not ad-

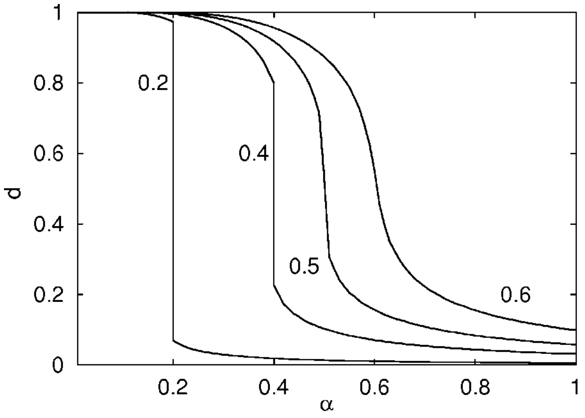

FIG. 3. Diversity d as function of α for linear interactions and u as indicated in the figure.

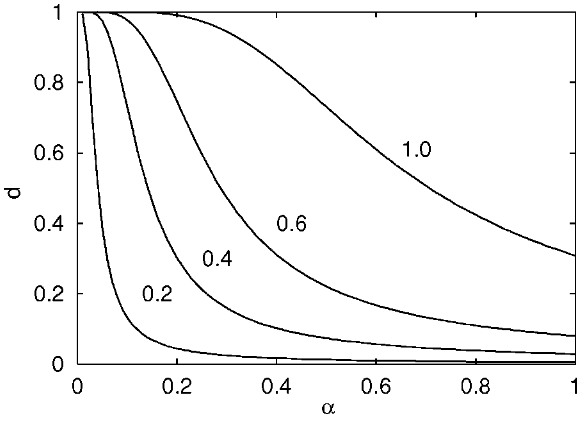

FIG. 4. Diversity as function of α for clipped interactions ($\theta=0$) and u as indicated in the figure.

missible in the case that $\delta^2>0$. We conclude then that the threshold phenomenon illustrated in Fig. 3 is not robust to the presence of Gaussian noise, even for a vanishingly small variance or, equivalently, to any distortion of the linear Hebb rule.

Although it is clear that increasing the intraspecies competitive interaction u leads to an increase of diversity, it is more difficult to understand why the increase of the number of features of each species results in a reduction of diversity. This can be explained by noting that, whatever the choice of $F(x)$, the variance of the interspecies interaction strengths grows linearly with α [see Eq. (2)]. This means that the larger α, the greater the odds of producing pairs of strongly cooperating species such that the interactions among them overshadow the self-restraint parameter u. The diversity reduction observed in Fig. 3 is then a consequence of the unrestricted growth of small groups of species that ends up to drive less favorable species to extinction in order to fulfill constraint (4). A word is in order about the stability of the replica-symmetric solution. For $\alpha<1/2$ we find that this solution is stable for $u<\alpha$, while for $\alpha>1/2$ it is stable for u slightly smaller than α (see Fig. 6).

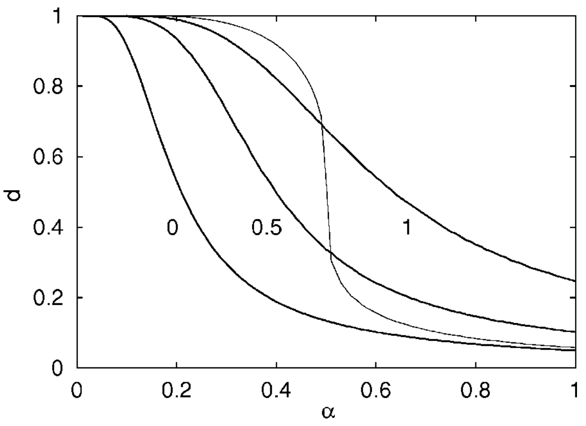

FIG. 5. Diversity as function of α for three-level interactions (values of θ indicated in the figure) and $u=0.5$. The thin line is the result for the linear case.

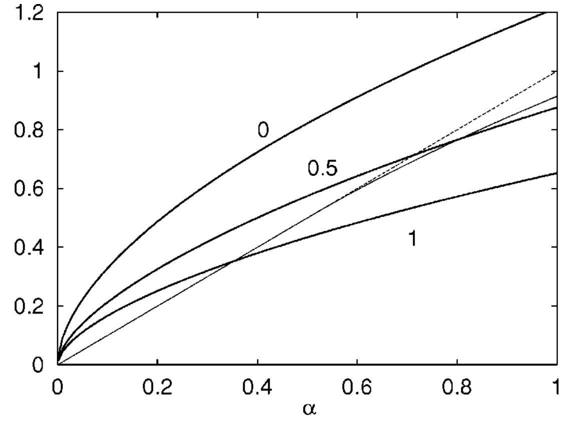

FIG. 6. Almeida-Thouless lines in the plane (α,u) for the three-level interactions model (values of the cutoff θ as indicated). The thin solid line is the result for the linear case and the dashed straight line is the diagonal $\alpha=u$.

B. Three-level interactions

The choice of the modulating function $F(x)$ determines the magnitudes of the interspecies interactions. However, it is often the case that no quantitative information is available, and the best we can infer are the signs of the interactions [1]. To model this scenario we set $F(x)=\mathrm{sgn}(x)$ if $|x|>\theta$ and 0 otherwise. Hence the interactions become effective only if the net number of matches (or mismatches) between homologous characters exceeds the cutoff value $\theta\geqslant 0$. Clipped interactions are obtained by setting $\theta=0$. Evaluation of Eqs. (7) and (8) yields $J=(2/\pi)^{1/2}\exp(-\theta^2/2)$ and $\hat{J}=\mathrm{erfc}(\theta/\sqrt{2})$, respectively. Hence the variance δ^2 is a monotonically decreasing function of θ. Figure 4 shows the dependence of the diversity on α in the case of clipped interactions $\theta=0$. As pointed out before, there is no threshold phenomenon in case of nonlinear interactions ($\delta^2>0$); i.e., all saddle-point parameters are smooth functions of the control parameters u and α. We note a substantial reduction of diversity as compared with the linear case. This point is reinforced in Fig. 5 where the effect of varying the cutoff θ is presented together with the result for the linear case. For fixed α the diversity increases monotonically with increasing θ, approaching its maximum value $d=1$ in the limit $\theta\rightarrow\infty$. This behavior is expected for large values of the cutoff (i.e., $\theta>\alpha^{1/2}$) since then most of the off-diagonal interactions and, consequently, of the randomness are removed from the fitness functional. However, the effect of a small cutoff is more subtle and, as we shall see, depends on the particular choice of the modulating function.

The regions in the space (α,u) where the stability condition (33) is satisfied are shown in Fig. 6. The equation $\Lambda=1$ defines the so-called Almeida-Thouless (AT) line that delimits the stable (above the AT line) and the unstable (below the AT line) regions. The nonlinearity makes the replica-symmetric prescription less reliable for small α, which is expected since there is an additional source of (Gaussian) randomness in this case. For large α, however, this prescription is more reliable since for $\theta>0$ the variance and hence the typical magnitude of the off-diagonal interactions $\langle J_{ij}^2\rangle\propto\alpha\hat{J}$ is smaller than that of the linear model.

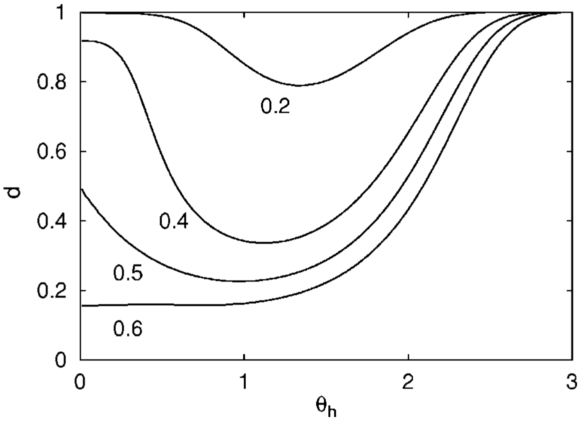

FIG. 7. Diversity as function of the cutoff θ_h for the high-pass filter with $u=0.5$ and α as indicated in the figure.

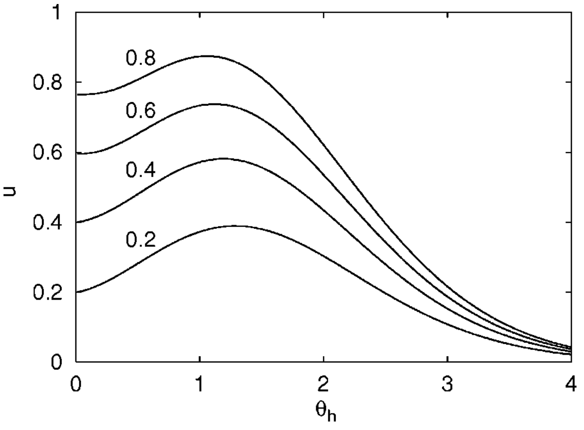

FIG. 8. Almeida-Thouless lines in the plane (θ_h, u) for the high-pass filter and α as indicated in the figure.

C. High- and low-pass filters

The main advantage of studying a model ecosystem is, no doubt, the option to carry out any type of experiment to clarify the behavior of the observables of interest. In the present context, an interesting possibility is to carry out a selective elimination of interactions, depending on their magnitudes. In the case of the high-pass filter, $F(x)=x$ if $|x|>\theta_h$ and 0 otherwise; this may be justified by identifying the cutoff $\theta_h \geqslant 0$ with the precision with which the interaction strengths are measured, so that magnitudes below the cutoff value are simply not detectable. Alternatively, by knocking out the small interactions we can infer their influence on the ecosystem diversity. Evaluation of Eqs. (7) and (8) now yields $\hat{J}=\hat{J}_h$ and $J=J_h$ with

$$J_h = (2/\pi)^{1/2}\theta_h \exp(-\theta_h^2/2) + \mathrm{erfc}(\theta_h/\sqrt{2}) \qquad (35)$$

and $\hat{J}_h=J_h$, so that $\delta^2=\alpha J_h(1-J_h)$. We note that δ^2 reaches its maximum value $\alpha/4$ at $\theta_h^* \approx 1.538$ and the linear case is recovered by setting $\theta_h=0$. Figure 7 illustrates the nontrivial role played by the weak interactions—depending on the value of α, their removal results in a great reduction of diversity, in stark contrast to our previous finding for the three-level model. As pointed out before, the increase of d for large cutoff values is an expected effect of the elimination of the random off-diagonal interactions. The diversity minimum displayed in this figure for α not too large is probably a consequence of the existence of a maximum in the variance δ^2 of the Gaussian noise term. We note, however, that while the value of the cutoff corresponding to the diversity minimum decreases and ultimately disappears for increasing α, the maximum of δ^2 occurs at θ_h^*, regardless of α. Naturally, one expects the replica-symmetric saddle-point to be less reliable in this situation of maximum randomness. This intuition is actually confirmed by the Almeida-Thouless lines exhibited in Fig. 8, although the points of maxima in these lines do not coincide exactly with the cutoff that maximizes δ^2.

This analysis was complemented by the inspection of the low-pass filter, $F(x)=x$ if $|x|<\theta_l$ and 0 otherwise, with $\theta_l \geqslant 0$. The linear case is recovered in the limit $\theta_l \to \infty$. Using the notation $\hat{J}=\hat{J}_l$ and $J=J_l$, we find $\hat{J}_l=J_l=1-J_h$, with J_h given by Eq. (35). Hence the variance δ^2 coincides with the variance of the high-pass filter and, in particular, it reaches a maximum value at $\theta_l=\theta_l^* \approx 1.538$. This explains the striking similarity between Fig. 9, which shows the diversity as function of the cutoff for the low-pass filter, and its counterpart for the high-pass filter, Fig. 7. These two figures indicate that the removal of either very weak or very strong interactions has little or none effect on the diversity of the ecosystem, except for α close to u, when, in the case of the high-pass filter, the diversity decreases linearly with increasing θ_h for small values of this cutoff. We note that the qualitative dependence of d on the cutoff exhibited in Figs. 7 and 9 is not altered by changing the value of the intraspecies interactions u. Moreover, the Almeida-Thouless lines for the low-pass filter exhibit maxima for all values of α, similarly to what we have found in the analysis of the high-pass filter.

D. General remarks

The high- and low-pass filters discussed above can be viewed as a kind of discriminating dilution or disconnection

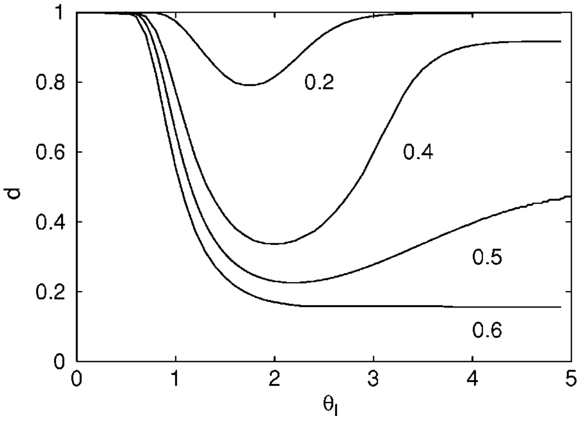

FIG. 9. Diversity as function of the cutoff θ_l for the the low-pass filter with $u=0.5$ and α as indicated in the figure.

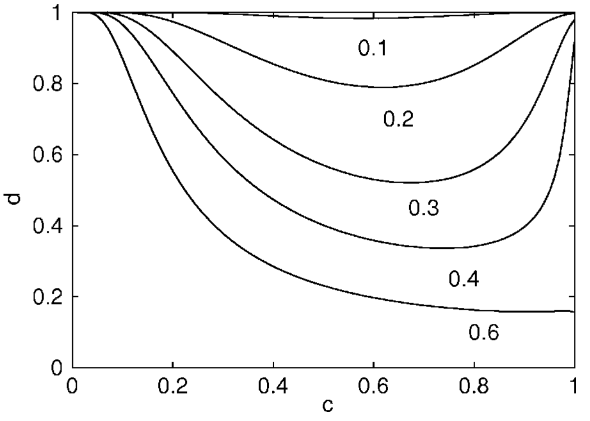

FIG. 10. Diversity as function of the average web connectivity c in the case of random dilution with $u/c=0.5$ and α as indicated in the figure.

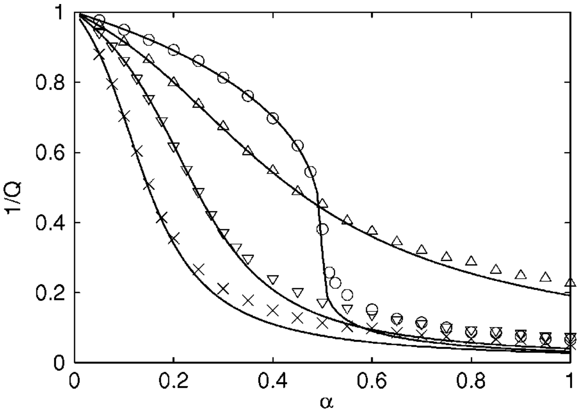

FIG. 11. Reciprocal of the saddle-point parameter Q as function of α for the linear (◯), clipped (×), high-pass filter (▽), and low-pass filter (△) modulating functions. The parameters are $u=0.5$, $N=500$, and $\theta_h=\theta_l=1$. The lines are the analytical predictions.

of (bidirectional) links in the community web. Interestingly, the very same set of saddle-point equations, Eqs. (22)–(26), can be used to describe the situation in which the species are disconnected at random. In fact, Sompolinsky has shown that if the links are cut randomly so that the average number of links for each species is Nc with $c\in(0,1]$, then the effect of dilution will be equivalent to the addition of an independent Gaussian noise of variance $\delta^2=\alpha(1-c)/c$ to the Hebb rule ($J=1$) [10]. This result depends on rescaling the mutual interactions, Eq. (1), by c and so the self-interaction u must be rescaled accordingly. The results for the fully connected, linear case are recovered for $c=1$. In Fig. 10 we illustrate the effect of random dilution on the ecosystem diversity. The results are very similar to those of selective dilution depicted in Figs. 7 and 9, although in this case there is no direct correlation between dilution and the magnitude of the interactions, except perhaps for a negligible sampling effect.

A traditional approach to study the structure of ecosystems is to look at the number of species that a habitat can sustain and at the relative abundance of each species in the community. In particular, field ecologists had long observed that most species of plants in a secluded community were relatively rare, while a few species were fairly common [26]. Those data are satisfactorily fitted by geometric distributions, which, in turn, are derived using continuous-time branching process to model the birth and death of species [27,28]. In the present model, the distribution of species abundance x is given by the derivative of the cumulative distribution, Eq. (30), which yields a Gaussian term centered at $x=-\Delta\hat{q}^{1/2}/\lambda$ plus a Dirac δ at the origin $x=0$. Hence most of the surviving species occurs at a well defined abundance value, so that both rare and common species are very infrequent. Since the particular form of the cumulative distribution (30) results from the mean-field, quadratic (or higher-order [29]) nature of the interactions between species, it is unlikely that a geometric (power-law) distribution for the species abundance is derived in the random replicator framework.

To conclude this section we present in Fig. 11 a comparison between the analytical predictions based on the replica-symmetric prescription and the numerical results of the brute-force solution of Eq. (3) for several choices of the modulating function $F(x)$. Here we focus on the reciprocal of the order parameter Q, in order to highlight its similarity with the diversity measure privileged in the previous figures. The agreement is excellent in the regions where the replica-symmetric saddle point is stable, which correspond to the regime of small α, but even outside this region the agreement is not bad. When the variance of the off-diagonal interactions, which is proportional to $\alpha\hat{J}$ [see Eqs. (2) and (8)], is very small compared to the self-interaction term u, we find a regime characterized by the coexistence of a macroscopic number of species (i.e., $x_i\approx 1$ for all species i) and hence $1/Q\to 1$. In the other extreme, when $\alpha\hat{J}\gg u$, the stationary solution is characterized by the dominance of a few species only (i.e., $x_i\approx N$ for a finite number of species), implying that $1/Q\propto 1/N$. In this sense, for large α the discrepancy between the theoretical prediction derived in the thermodynamic limit and the simulations carried out for finite N was expected already. The result for the low-pass filter tends to zero much slower than for the other modulating functions because the factor $\hat{J}$ in the variance of the off-diagonal interactions is much smaller in that case.

V. CONCLUSION

Ecology was one of the first disciplines of biology to embrace mathematical thinking and modeling as a guide to attain a unified theoretical structure on the facts of nature. It is thus not a surprise that the trade-off between precision, realism, and generality involved in the formulation of a mathematical model had been so passionately debated by the ecological community (see [30] for an account of the mathematical thinking in population ecology). It was in this historical context that MacArthur introduced a minimum principle in ecology, akin to the principle of least action of mechanics, by showing that a special kind of competition equation—namely, Eq. (3) with symmetric interspecies interactions—minimizes a quadratic expression. That finding was then explored to interpret species packing and competitive equilibria [3]. Recently this idea was taken up and considerably refined by the application of statistical mechanics techniques that enabled not only the study of very large communities but also the investigation of communities in which the strengths of the interactions between species are assigned randomly [5,6].

The distinctive feature of ecosystems with (nonstructured) interspecies random interactions is the enormous number (exponential in the number of species) of stable states or distinct equilibria [5]. The lack of features to identify the species, however, precludes the study of basic principles of ecology. In this contribution we resolve the identity issue by assigning to each species a series of characters [6]. Explicitly, species i is characterized by a binary string $(\xi_i^1, \xi_i^2, \ldots, \xi_i^p)$ where $\xi_i^\mu = 1$ or -1 indicates that species i displays or lacks character μ. Then it is possible to describe 2^p distinct species in this framework. According to Gause's principle of competitive exclusion, the strength of the competition between a pair of interacting species should increase with the number of features they have in common. In that sense, the generalized form of the Hebb rule explicited in Eq. (1) provides a very broad framework to investigate the consequences of Gause's principle to the competitive equilibrium for many species.

In this contribution we have focused mainly on the ecosystem diversity, defined as the fraction of the N species that survive and hence coexist at equilibrium. A very robust result that holds for any choice of the modulating function is the reduction of diversity that accompanies the increase of the number of features p needed to specify the species. It is natural to associate this number or, equivalently, the ratio $\alpha = p/N$ with the overall complexity of the species that compose the ecosystem and hence with the complexity of the ecosystem itself. Accordingly, our findings imply that the number of species that can coexist decreases with the complexity of the species. In fact, increase of α leads to the appearance of pairs or small groups of strongly complementary species, in the sense that the magnitudes of their interactions exceed those of their self-restraint parameters. The resulting unrestricted growth of this group of species culminates then with the extinction of the remaining species of the ecosystem. Moreover, the analysis of selective as well as random deletion of links between species shows that ecosystems composed of simple species (in the above sense) are very robust against dilution. We stress that this apparently intuitive result is actually a mathematical consequence of the complementarity principle used to set the interspecies interactions. In addition, we find that the puzzling result that there exists either rich or poor ecosystems in the case of a linear modulating function is not robust to the presence of static noise as well as to nonlinear deviations of the Hebb rule. Hence, the discontinuity of the diversity index illustrated in Fig. 3 is not typical of communities of species whose interactions are determined by a complementarity principle.

To conclude, a few remarks relating our results to the long-standing diversity-stability debate in ecology are in order (see [31,32] for reviews). The central issue in that argument is the insurance hypothesis of biodiversity which asserts that the robustness against perturbations (resistance) as well as the capacity to return to the original state after perturbation (resilience) increases with diversity, simply because a greater number of species is likely to exhibit a larger repertoire of responses to environmental perturbation [33] (see, however, [34]). This intuitive idea was confronted by May's remark that greater diversity lead to lower local stability of multispecies equilibria in a Lotka-Volterra competition model [1]. This result follows from the proof that systems of equations became less stable as the number of equations increases [35]. As in the present analysis, that result was derived under the condition that both the connectance of the community web and the strengths of the interactions be random variables (see [36] for the generalization of May's analysis to random hierarchical webs). The adequacy of this scenario to describe real community webs is, of course, arguable. In addition, field experiments designed to assess the insurance hypothesis often fail to distinguish the effect of diversity from that of species composition, so the relationship between diversity and stability remains controversial despite (or because of) its importance to management of biological resources [31,32]. In a sense our conclusions parallel those of May [1]. In fact, viewing the random mutual interactions as a perturbation (the relative strength of which is measured by the ratio $\alpha\hat{J}/u$) of the ideal ecosystem in which all N species coexist, our results show how the gradual introduction of mutual interactions destabilizes the regime of perfect coexistence $(d=1)$ among the species. This instability is reflected by the extinction of a fraction of the initial species. However, Figs. 7, 9, and 10 illustrate the complexity of the situation: the diversity depends not only on the connectance of the web but also on the strengths of the mutual interactions that are being added. Compensatory effects do occur so that the diversity can actually increase as new links are incorporated in the community web. These effects are probably due to species composition since they depend strongly on the species complexity and disappear for $\alpha > 0.5$. We must note, however, that our approach differs from the standard ecological modeling which, as a rule, takes the number of species (diversity) as the control parameter or independent variable and measures the stability in terms of some appropriate response variable, usually total biomass production. Here we consider diversity as a dynamical variable and concentrate on its dependence on the nature of the species composing the system as well as on the connectance of the community web.

Statistical mechanics can contribute uniquely to the study of large ecosystems with random interspecies interactions. This kind of randomness stirs up difficult issues such as the lack of self-averageness of nonextensive quantities, the nontrivial ergodicity breaking associated with the instability of the replica-symmetric solution, and the very slow relaxation (aging) to equilibrium which are now fairly well understood [37]. Interestingly, a phenomenon akin to aging in spin glasses was observed in the evolutionary dynamics of the tangled nature model as the average duration of the periods of stasis was reported to increase with time [16]. In this contribution we ally statistical mechanics know-how to a biologically motivated prescription for setting the magnitude of the interspecies interactions in a form suitable to investigate ecologically relevant issues such as the link between species composition and ecosystem diversity.

ACKNOWLEDGMENTS

The work of J.F.F. was supported in part by CNPq and FAPESP, Project No. 99/09644-9. D.O.C.S. was supported by CAPES.

[1] R. M. May, *Stability and Complexity in Model Ecosystems* (Princeton University Press, Princeton, 1974).
[2] J. D. Murray, *Mathematical Biology* (Springer-Verlag, Berlin, 1989).
[3] R. MacArthur, Theor Popul. Biol. **1**, 1 (1970).
[4] N. S. Goel, S. C. Maitra, and E. W. Montroll, Rev. Mod. Phys. **43**, 231 (1971).
[5] S. Diederich and M. Opper, Phys. Rev. A **39**, 4333 (1989).
[6] V. M. de Oliveira and J. F. Fontanari, Phys. Rev. Lett. **89**, 148101 (2002).
[7] J. D. Farmer, N. H. Packard, and A. Perelson, Physica D **22**, 187 (1986).
[8] D. Lancet, E. Sadovsky, and E. Seidemann, Proc. Natl. Acad. Sci. U.S.A. **90**, 3715 (1993).
[9] M. Ridley, *Evolution* (Blackwell Science, Cambridge, England, 1996).
[10] H. Sompolinsky, in *Heidelberg Colloquium on Glassy Dynamics*, edited by J. L. van Hemmen and I. Morgenstern, Lecture Notes in Physics Vol. 275 (Springer, Berlin, 1987), pp. 485–527.
[11] J. L. van Hemmen and R. Kuhn, Phys. Rev. Lett. **57**, 913 (1986).
[12] V. M. de Oliveira, Eur. Phys. J. B **31**, 259 (2003).
[13] G. F. Gause, *The Struggle for Existence* (Dover, New York, 1971).
[14] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, Cambridge, England, 1989).
[15] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, England, 1998).
[16] K. Christensen, S. A. di Collobiano, M. Hall, and H. J. Jensen, J. Theor. Biol. **216**, 73 (2002).
[17] M. Hall, K. Christensen, S. A. di Collobiano, and H. J. Jensen, Phys. Rev. E **66**, 011904 (2002).
[18] S. A. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).
[19] C. Amitrano, L. Peliti, and M. Saber, J. Mol. Evol. **29**, 513 (1989).
[20] A. C. C. Coolen, R. W. Penney, and D. Sherrington, Phys. Rev. B **48**, 16116 (1993).
[21] E. H. Simpson, Nature (London) **163**, 688 (1949).
[22] V. M. de Oliveira and J. F. Fontanari, Phys. Rev. E **64**, 051911 (2001).
[23] J. R. L. de Almeida and D. J. Thouless, J. Phys. A **11**, 983 (1978).
[24] E. Gardner and B. Derrida, J. Phys. A **21**, 271 (1988).
[25] P. Biscari and G. Parisi, J. Phys. A **28**, 4697 (1995).
[26] N. G. Hairston, Ecology **40**, 404 (1959).
[27] G. U. Yule, Proc. R. Soc., London, Ser. B **213**, 21 (1925).
[28] J. Chu and C. Adami, Proc. Natl. Acad. Sci. U.S.A. **96**, 15 017 (1999).
[29] V. M. de Oliveira and J. F. Fontanari, Phys. Rev. Lett. **85**, 4984 (2000).
[30] S. E. Kingsland, *Modeling Nature* (University of Chicago Press, Chicago, 1985).
[31] D. Tilman, Ecology **80**, 1455 (1999).
[32] M. Loreau, S. Naeem, P. Inchausti, J. Bengtsson, J. P. Grime, A. Hector, D. U. Hooper, M. A. Huston, D. Raffaelli, B. Schmid, D. Tilman, and D. A. Wardle, Science **294**, 804 (2001).
[33] S. Yachi and M. Loreau, Proc. Natl. Acad. Sci. U.S.A. **96**, 1463 (1999).
[34] A. B. Pfisterer and B. Schmid, Nature (London) **416**, 84 (2002).
[35] M. R. Gardner and W. R. Ashby, Nature (London) **228**, 784 (1970).
[36] T. Hogg, B. A. Huberman, and J. M. McGlade, Proc. R. Soc. London, Ser. B **237**, 43 (1989).
[37] K. H. Fischer and J. A. Hertz, *Spin Glasses* (Cambridge University Press, Cambridge, England, 1991).