CrossMark

Study of the influence of the phylogenetic distance on the interaction network of mutualistic ecosystems

Roberto P.J. Perazzo [a], Laura Hernández [b,*], Horacio Ceva [c], Enrique Burgos [c,d], José Ignacio Alvarez-Hamelin [a]

[a] *Departamento de Investigación y Desarrollo, Instituto Tecnológico de Buenos Aires, Avenida E. Madero 399, Buenos Aires, Argentina*

[b] *Laboratoire de Physique Théorique et Modélisation, UMR CNRS, Université de Cergy-Pontoise, 2 Avenue Adolphe Chauvin, 95302, Cergy-Pontoise Cedex, France*

[c] *Departamento de Física, Comisión Nacional de Energía Atómica, Avenida del Libertador, 8250, 1429 Buenos Aires, Argentina*

[d] *Consejo Nacional de Investigaciones Científicas y Técnicas, Avenida Rivadavia 1917, C1033AAJ, Buenos Aires, Argentina*

H I G H L I G H T S

- We study the influence of phylogenetic distance on mutualistic ecosystems' networks.
- Instead of a statistical study, we reformulated it as an optimization problem.
- We generalize the self-organizing network model (SNM) to handle phylogenetic data.
- Reducing the distance among the counterparts of each species destroys nestedness.

A R T I C L E   I N F O

A B S T R A C T

We investigate how the phylogenetic relationship between the species of each interacting guild in a mutualistic ecosystem influences its network of contacts. We develop a dynamical self organized model that reallocates contacts between mutualists, according to a contact preference rule (CPR) that takes into account phylogenetic distances. We conclude that a CPR that promotes phylogenetic proximity among the counterparts of the species of each guild leads to highly unrealistic contact patterns. We find that nestedness can instead be attributed to a general rule by which species tend to behave as generalists holding contacts with counterparts that already have a large number of contacts.

1. Introduction

Mutualistic ecosystems usually involve species of animals and plants that interact to fulfill essential biological functions such as feeding or reproduction. Typical examples are *pollination networks*, where insects feeding from the nectar of flowers contribute to the pollination process and *seed dispersal networks*, where animals (usually birds) feed from the fruits while dispersing the seeds contained in them. A large amount of research has been devoted to studying mutualism [1,2]. In traditional studies, the interaction of all active plant and animal species is recorded within a restricted geographical extension [3,4]. Collecting these data is a difficult work and several observed ecosystems are not large enough so as to obtain reasonable statistical results. Interestingly, some large ecosystems have been recorded like for example [5–7].

* Corresponding author. Tel.: +33 134257521; fax: +33 134257500.
E-mail addresses: rperazzo@itba.edu.ar (R.P.J. Perazzo), laura.hernandez@u-cergy.fr, laura.hernandez1@sfr.fr (L. Hernández), ceva@cnea.gov.ar (H. Ceva), burgos@cnea.gov.ar (E. Burgos), ihameli@cnet.fi.uba.ar (J.I. Alvarez-Hamelin).

The interaction among the mutualists may be described in terms of a complex network which is said to be *bipartite* as there are two different kind of vertices, those representing plant species and those representing animal species, while the considered interactions only happen between vertices of different kinds.

The corresponding adjacency matrices of natural mutualistic ecosystems strongly indicate that they are not a random collection of interacting species, but that they display instead a high degree of internal organization. A pervading feature that is generally found is that such adjacency matrices have a *nested* pattern of interactions, in which both generalists (species holding many interactions) and specialists (holding few interactions) tend to interact with generalists whereas specialist-to-specialist interactions are infrequent [8]. In other words, if species are ordered by decreasing number of contacts, then the contacts of a given species constitute a subset of the contacts of the preceding species in the list [9].

This nested structure has been attributed to a number of different causes and the controversy about the ultimate reasons making this pattern so frequently observed is still open. It is fairly obvious that a detailed explanation of the interaction behavior of individual species can be of little help to understand such a generalized pattern found in ecological systems of very different sizes and types, and involving plants of different nature and animals that range from insects to birds.

Nestedness has been shown to offer some advantage for the robustness of the whole system; thus suggesting that systems that are currently observed are those that have survived less disturbed thanks to its nested structure [8,10].

Interestingly, the same nested structure has been recently found in a completely different context: industrial structure of the countries and the import–export world trade network. In these cases the two different kinds of nodes represent the products and the countries that produce or buy them. It has been shown that the techniques used to measure the degree of order of mutualistic ecosystems seem to be adapted to a description of these economical systems and that they are more precise than the magnitudes habitually used to address these problems such as, for instance, the global amount of products bought/sold by each country [11–14].

Different hypotheses aiming to explain the origin of this widespread pattern of interactions in mutualist ecosystems have been proposed. In Ref. [15] nestedness has been attributed to phenotypic affinity between species of different guilds while in Refs. [16,17] an extensive analysis is performed, concluding that phylogenetic proximity could explain the nested organization of contacts of some cases of mutualistic systems. Ref. [18] considers the contrary, that the modest percentage of correlations between phylogenetic relatedness and ecological similarity found in Ref. [16] indicates that phylogenetic relationships do not have a marked effect.

Most of the works that study the possible mechanisms that lead to this nested pattern search for a positive statistical correlation between the nested pattern of contacts and its supposed cause. However, the sole fact that in a part of the empirical observations two elements appear to be statistically correlated should not be taken to mean that one is the cause of the other. Such a correlation may rather indicate instead, that both elements are not incompatible, i.e., they do not mutually exclude each other or they stem from a third, common cause.

One example of this analysis is given by the strong positive correlation found between the species' abundance and hence the frequency of interactions, with the pattern of contacts of some species [19]. It has been suggested that locally abundant species are prone to accumulate interactions and conversely rare species are prone to lose them [20], as also suggested by neutral theories [21].

We propose an alternative way to identify the important features that may lead to nestedness, by studying the states reached by a system that follows a dynamical process based on some assumed interaction mechanism between the mutualists, thus comforting or falsifying hypotheses concerning such interaction mechanisms between the species. This complementary approach is based on the Self Organizing Network Model (SNM) that we have proposed and developed in Refs. [22,23] and that we will briefly recall in Section 2.1.

In the present paper we generalize the SNM in order to integrate external data that is not contained in the evolving network but has to be taken from the phylogenetic trees of the interacting mutualists. Mathematically, phylogenetic proximity is accounted for through a matrix of distances separating any two species of each guild. The distance matrix is directly obtained from the topology of the phylogenetic tree.

We complete our study with the comparison of the statistical properties of observed networks with those of two extreme idealized benchmarks.

2. Description of the numerical study

2.1. The SNM algorithm

Mutualistic systems can be analyzed as bipartite graphs [24]. The interaction pattern is usually coded into a (rectangular) adjacency matrix in which rows and columns are labeled respectively by the plant and animal species. Its elements $K_{p,a} \in \{0, 1\}$ represent respectively the absence or presence of an interaction (contact) between the plant species p and the animal species a. The number of contacts of each species is the degree of the corresponding node in the bipartite graph, $G_{a,(p)}^{A,(P)}$, for animals (plants).

Several reasons have been given to explain the pattern of interactions between the two guilds of a mutualistic network. They have been usually based on positive statistical significance of correlations. One alternative way to elucidate a possible *causal* link between some hypothetical interaction mechanisms between mutualistic species and the pattern of contacts, is to use a dynamical model.

The basic idea behind this strategy is to verify the *consistency* of the empirically observed contact pattern, and some hypothetical interaction rule that may favor or hamper the contact between mutualistic species. Such interaction can, in principle, be based on phenotypic complementarity, phylogenetic affinity, degree, or any other possibility. We call the interaction mechanism the *contact preference rule* (CPR) in the sense that it is assumed that species verifying this rule tend to hold contacts among each other.

The dynamics of the SNM is artificial and should not be interpreted as representing the complex evolutionary processes that have led to the species themselves as observed today and to their interactions. Instead, from a purely theoretical point of view, this setup is equivalent to considering that the observed pattern of interactions corresponds to an *optimal* assignment of the contacts between both guilds, with two constraints. The first constraint is the fulfillment of an assumed CPR, and the second is the given (constant) total number of contacts between the two mutualistic guilds. This number might be considered as an indicator of the energy invested by all the species of the ecological system in their exchange of nourishment. In other words one may attempt to describe the observed pattern of contacts as the result of a (combinatorial) optimization problem by which contacts in the adjacency matrix are placed in such a way so as to reach an extreme of an utility function that corresponds to an optimal fulfillment of some prevailing CPR.

Comparing with the real system, the pattern that emerges when running the SNM with the chosen CPR, starting from an initial random adjacency matrix, one can easily rule out the wrong hypothesis when the reached stationary states differ from the real system considered. This can also be done by analyzing the stability of real adjacency matrices evolved with the studied CPR.

On the other hand an emerging pattern that is consistent with the organization of real systems suggests that the studied CPR may be considered as playing an important role in such organization.

In the SNM discussed in Refs. [8,22] the topology of that network emerges as the result of a self-organization process where links are redefined gradually, alternating plants and animals. For instance, a plant is first chosen at random and one of its contacts is redefined by spotting a mutualistic counterpart that has already more contacts that the original one (called SNM-I in the previously cited references). Next an animal is chosen and the same procedure is accomplished. The iteration of these steps provides a simple heuristic, leading to a good approximation of the optimal assignment of contacts mentioned above.

Unlike in the preferential attachment algorithm [25], in the SNM the topology of a *non-growing* network with a fixed number of nodes is progressively reshaped: at each iteration a connection between two nodes of different kinds is rewired to favor a contact with the node having the highest degree. It is worthwhile noting that in this sense, the CPR of the SNM is *local*: it does not take into account the whole probability distribution, but only the degrees of the two randomly chosen species.

In the above references we show that such a CPR always leads to nested networks with degree distributions that closely resemble the ones reported from the observation of real mutualistic systems.

The SNM has to be modified to take into consideration the phylogenetic structure of both guilds. Therefore, it is necessary to have a simple quantitative measure of the phylogenetic structure. We now describe how this is made.

2.2. The phylogenetic-SNM

To study the influence of phylogenetic proximity on the pattern formation of the adjacency matrix, one has to modify the SNM by properly defining a CPR based on some measure of the phylogenetic distance between any two species of the same guild.

The classification of species according to their similarities has been a major endeavor since the origins of biology as a natural science. This classification may be summarized in a symmetric $N \times N$ distance matrix with vanishing diagonal elements providing a measure of similarities and differences between any pair of species. Due to the central role of evolution, these classifications are depicted by phylogenetic trees that are determined using several sources of information.

Comparative studies of phenotypic traits are also widely used. The resemblance of species is measured through a phylogenetic signal that is quantitatively estimated through statistical analyses [26,27] of the distribution of the values of different traits. These studies may also be supplemented whenever possible with fossil records.

While the tips of the tree correspond to presently observed species, the remaining nodes are associated to their presumed ancestors. A hierarchical organization of all living species is therefore provided and those that closely resemble each other are neighboring tips of the tree.

The phylogenetic classification of a group of species gathers them in taxa within taxa of an ever increasing generality. Based on the topology of the tree it is possible to define a *distance* that provides a quantitative estimation of resemblances and differences between them. In the appendix we describe the simple procedure to extract a distance matrix directly from the topology of the phylogenetic tree.

A distance matrix $d(k, k')$ constructed in this way is not only fully consistent from the start with the results of statistical analyses but also fully agrees with what can be expected from an intuitive picture: small values of $d(k, k')$ remain associated to species that share the same branching sequence and a common evolutionary history while large values correspond to species that have followed a different evolutionary process because they have been separated at earlier stages. With these conventions the closest possible distance between any two species is 1 and, if all species are at a distance 1 they belong to a *star phylogeny*.

We are now in the position to define the self organization process that casts into a modified SNM a CPR based on the distance between species of each guild. Certainly, different CPRs based on the phylogenetic distance could be imagined, it is then interesting to study the consequences of two extreme opposite limits: minimizing (MIN-CPR) or maximizing (MAX-CPR) the phylogenetic distance among the counterparts of a given species.

At each iteration of the MIN-CPR algorithm, the following actions are performed:

- Two elements $K_{p,a_1} = 1$ and $K_{p,a_0} = 0$ are chosen at random in the same row p, corresponding to connected and disconnected species respectively.
- Two total distances Γ_0 and Γ_1 are calculated between the plant p and the other plants p' which hold a contact with a_0 and a_1 respectively.

$$\Gamma_0 = \sum_{p'} d(p, p') K_{p',a_0} \tag{1}$$

$$\Gamma_1 = \sum_{p'} d(p, p') K_{p',a_1}. \tag{2}$$

Notice that Γ_0 and Γ_1 are the total distances separating the randomly chosen plant from all other plants holding contacts with the same animal. Γ_1 corresponds to the current contact and Γ_0 is calculated for an alternative location of the contact.

- A swap between these two elements corresponding to the redefinition of the link, i.e., $K_{p,a_0}^{\text{new}} = 1$ and $K_{p,a_1}^{\text{new}} = 0$, is proposed and it will be accepted if the following two conditions are satisfied:
  - (i) neither animal species a_1 nor a_0 remain without contacts due to the swapping,
  - (ii) $\Gamma_1 \geq \Gamma_0$
    In this case a_0 is a better mutualistic counterpart of p; it belongs to a group whose members are closer phylogenetic relatives than the original group of contacts of a_1.
    Whenever the conditions (i) or (ii) are not met, the proposed swapping is rejected.

These steps are repeated, alternatively inverting the role of rows and columns, until the algorithm converges (no more changes are possible).

The extreme opposite rule, which we call MAX-CPR, consists of replacing the condition (ii) by $\Gamma_1 \leq \Gamma_0$ to accept the swapping.

Since the MIN-CPR and MAX-CPR conditions are mutually exclusive we perform independent runs of the SNM algorithm.

Within the MIN-CPR possibility, the animal counterpart that will finally be selected for the plant p, is such that the set of plants p' having contact with the animal a_0 are phylogenetically closer to p than those in contact with a_1. This is so because the sums Γ_0 and Γ_1 involve all distances between the plant p (that has been selected at random) and all other plants that hold contacts with the two animals a_0 and a_1. With this algorithm the configuration of contacts is progressively dominated by phylogenetic proximity as measured by the corresponding distance matrices. Within this strategy species of one guild are assumed to interact in the same fashion as all other species of the same guild belonging to their phylogenetic neighborhood.

Within the MAX-CPR possibility, the animal counterpart that will be selected for the plant p is such that all other plants p' that have some contact with it are phylogenetically further from p. In this way the set of species that share contacts with p tend to have a greater phylogenetic diversity. This is so because the pattern of contacts is progressively dominated by greater phylogenetic distances as measured by the corresponding distance matrices. In this alternative all species tend to be as generalist as possible in what refers their phylogenetic grouping.

We introduce an order parameter that accounts for the *effective distance between interacting species*:

$$D^{A,P} = \frac{1}{\langle d_{A,P} \rangle} \frac{\sum_{k,k'} d_{A,P}(k, k') \tilde{W}_{k,k'}^{A,P}}{\sum_{k,k'} \tilde{W}_{k,k'}^{A,P}}. \tag{3}$$

In Eq. (3) A and P respectively represent animals and plants, and $\tilde{W}$ represents the unweighted adjacency matrix of the projected graphs for animals or plants that are defined as (with K^T being the transposed of K)

$$W_{p,p'}^{P} = \sum_{a} K_{p,a} K_{a,p'}^{T} (1 - \delta_{p,p'}) \tag{4}$$

$$W_{a,a'}^{A} = \sum_{p} K_{a,p}^{T} K_{p,a'} (1 - \delta_{a,a'}) \tag{5}$$

and $\tilde{W}_{k,k'} = 1$ if $W(k, k') \neq 0$ or 0 if $W(k, k') = 0$, i.e., two species share or do not share mutualistic counterparts; $d(k, k')$ is the distance matrix obtained from the real phylogenetic tree. This equation provides different results for plants or animals and should therefore be evaluated separately for the two guilds.
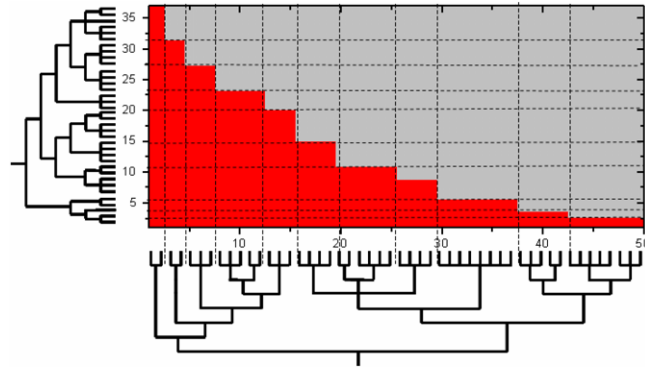
Fig. 1. (*Color online*) Adjacency matrix of the benchmark PERF system in which a perfectly nested contact pattern is tailored to follow the (arbitrary) phylogenetic tree of animals and plants. Dark (light) gray means $K_{pa} = 1(0)$. Dotted lines are drawn to help the eye in the correspondence of contacts with phylogenetic groups.

The sum in the denominator of Eq. (3) is just the number of terms appearing in the numerator, therefore D represents the average distance between species of the same guild that share at least one counterpart of the other guild. D is measured in units of the average distance $\langle d \rangle$ between *all* plant (animal) species of the system, namely

$$\langle d_{A,P} \rangle = \frac{\sum\limits_{k,k'} d_{A,P}(k, k')}{N_{A,P}(N_{A,P} - 1)} \tag{6}$$

where $N_{A,P}$ is the number of animal or plant species of the system.

A value $D^{A,P}$ significantly less than unity indicates that phylogenetic proximity is a dominant effect, because species that share at least one mutualistic counterpart are closer than the average separation of species of that guild. If species are ordered as the tips of a phylogenetic tree, $\tilde{W}_{k,k'}$ has most of its non vanishing elements close to the diagonal. On the other hand $D^{A,P} \simeq 1$ is instead a signature that phylogenetic proximity is not relevant. This value is compatible with both the random case and the situation where generalists are present. Correspondingly $\tilde{W}_{k,k'}$ tends to display non vanishing elements far from the diagonal. The value of $D^{A,P}$ is therefore a good order parameter characterizing the whole pattern of contacts that can be used to check if the ordering process implied in the SNM converges to stable configuration.

2.3. Compared statistical analysis of real and benchmark networks

In Ref. [16] it is stated that if phylogenetically close species share some common feeding or pollenization strategy, they should also be expected to have similar degrees. It is then interesting to compare the correlation observed between the nested structure and some magnitude measuring the phylogenetic proximity of the species of each guild, in the real networks and both a null model on one hand and an ideal network tailored so as to have the maximum possible correlation, in the sense of Ref. [16].

This idealized benchmark system, called PERF, is built in such a way that its adjacency matrix is perfectly nested and contacts faithfully follow some assumed (arbitrary) phylogenetic ordering. The size of the system (36 plants and 50 animals) was chosen to compare to the natural network discussed in Ref. [16]. The adjacency matrix together with the (arbitrary) phylogenetic trees of a realization of the PERF system are shown in Fig. 1.

We also consider null models consisting of systems in which the adjacency matrix and the topology of phylogenetic trees of both guilds are the same as the observed ones but where the species labeling the tips are randomly permuted. We refer to these as the "randomized systems". The null model is constructed considering the average over many realizations of such random sorting. This is done to provide statistically significant results of a situation in which the correlation between the structure of the network and the prevailing phylogenetic order has been destroyed.

We consider the ecological similarity as defined in Ref. [16]:

$$S_{k,k'} = \frac{W_{k,k'}^{A,(P)}}{G_k + G_{k'}}. \tag{7}$$

Again, the matrix $W_{k,k'}^{A,(P)}$ corresponds to the *projected graphs*. Then $S_{k,k'}$ measures the number of mutualistic counterparts shared by the two species as compared to the total number of counterparts of both; in this way its maximum value is 1/2. In order to consider the dependence with the distance we calculate the average $\langle S(\delta) \rangle$ defined as:

$$\langle S(\delta) \rangle = \frac{1}{N_\delta} \sum_{(k,k')|d(k,k')=\delta} S_{k,k'} \tag{8}$$

where N_δ is the number of pairs of species of the same guild that are separated by a phylogenetic distance δ.
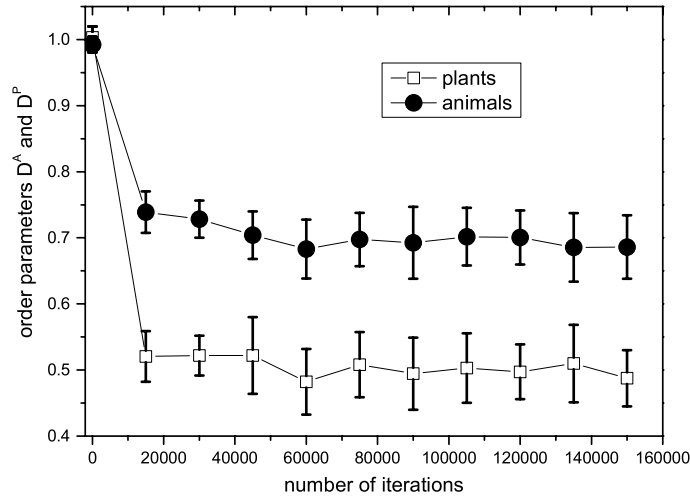
Fig. 2. Order parameters D^A and D^P (effective distance between species sharing the same mutualistic counterparts), as a function of the number of iteration steps of the SNM algorithm, using the MIN-CPR alternative. The curves are the average over several realizations with random initial conditions, always using the same number of rows, columns, and contacts as the system NCOR. Error bars are the corresponding standard deviations.

We also study the dispersion of the degrees of pairs of species as a function of their phylogenetic distance. We define:

$$\Delta G^{A(P)}(\delta) = \sqrt{\frac{1}{N_\delta} \sum_{(k,k')|d(k,k')=\delta} (\Delta G_{k,k'}^{A,(P)})^2} \tag{9}$$

with,

$$\Delta G_{k,k'}^{A,(P)} = G_k^{A,(P)} - G_{k'}^{A,(P)} \tag{10}$$

and where G_k is the degree of the species k. The sum is extended over all pairs of species (k, k') that are separated by the phylogenetic distance δ.

3. Results

3.1. The phylogenetic SNM

We have studied the following observed systems taken from Ref. [16] (NCOR, OLAU, ARR1 and ARR2). For the sake of simplicity, we illustrate most of the following discussion using the case of the ecosystem Nava de las Correhuelas (NCOR) as it has been presented as a prototypical example where phylogenetic proximity explains the contact pattern between mutualistic guilds. Anyway, our results hold for all the studied systems.

When considering the organization of a system under a given CPR, the order parameter $D^{A(P)}$ is a way to test for the convergence of the algorithm. In Figs. 2 and 3 we plot the values of the parameters D^A and D^P as a function of the number of iteration steps of the SNM for animals and plants for the two extreme CPRs considered above.

The initial conditions are always adjacency matrices of the same number of rows and columns and with the same number of contacts than the observed system under study, except for the fact that all contacts are randomly distributed. The CPR is built using the phylogenetic distance matrix that corresponds to the real system.

The curves shown are issued from averages over a statistically significant set of realizations of such random initial conditions. The error bars are the standard deviation of the results obtained over such set. These error bars provide also a measure of the convergence of the SNM.

As the number of SNM iterations grows, both CPR produce values of D^A and D^P that reach asymptotic constant values. This indicates that both CPR succeed in driving the system to a stable pattern of contacts. These are however different. While for the MIN-CPR case it is found that the asymptotic value of the order parameters D^A and D^P stabilizes at values that are significantly smaller than unity, for the MAX-CPR alternative it is found instead, that they reach a stable value of 1.

As expected, for the first iterations, $D^A \simeq D^P \simeq 1$ in both Figs. 2 and 3. The system is very near the initial condition, where the contacts have been randomly distributed with no relation whatsoever to the phylogenetic tree of either guild.

On the other hand, values significantly lower than 1 as those seen in Fig. 2 after many iterations of the SNM, indicate that species that share the same counterparts are close phylogenetic neighbors.

The ordered pattern that prevails asymptotically for the MAX-CPR case is a single, perfectly nested pattern. After some initial fluctuations $D^{A,P}$ stabilize very closely to 1. The ordered pattern turns out to be essentially unique except for a random
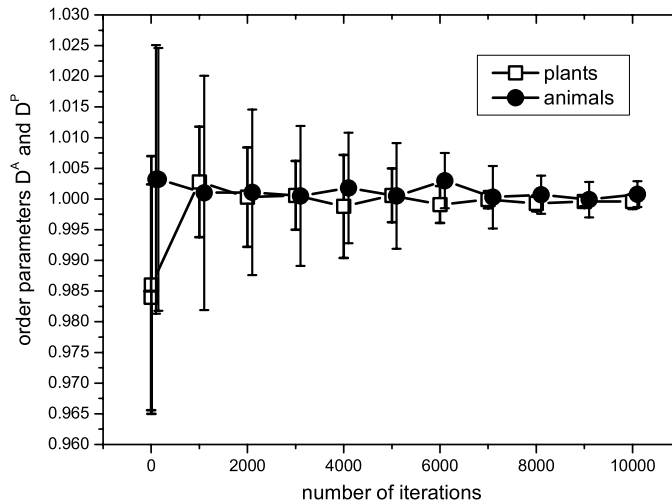
Fig. 3. Value of the order parameters D^A and D^P as a function of the number of iteration steps of the SNM algorithm using the MAX-CPR alternative. Notice that the vertical scale is strongly expanded as compared to that of Fig. 2. Conventions are the same as in that figure. Data for plants and animals have been slightly horizontally displaced, to facilitate its observation.

permutation of phylogenetic labels and therefore the error bars tend to diminish as the number of iterations of the SNM increases.

This is not the case for the MIN-CPR in which error bars do not diminish in the same way. The reason is that there are several possible distinct asymptotically ordered patterns, all of them corresponding to modular ecosystems and all having slightly different values of D^A and D^P. The phylogenetic SNM converges to any of these ordered patterns thus producing a somewhat larger dispersion in the values of these parameters. Nevertheless, the limiting constant values that are reached are many standard deviations away from the value $D^{A(P)} = 1$.

Moreover, the stability analysis of a given pattern of contacts under the SNM can be nicely illustrated using the projected graphs of each guild W^P and W^A.

In Fig. 4, panel (A) shows the distance matrix of plants of NCOR. Species are ordered as in the corresponding phylogenetic tree. In such order phylogenetically close species fall near the diagonal. In panel (B) the corresponding projected graph of plants is shown. It can be observed that species near the diagonal do not share more contacts than the others. In panel (C) we show the matrix W^P, after 140,000 iterations of the CPR-MIN, taking the real system (corresponding to the projected graph in panel (B)) as initial condition. Comparing with panel (A), it is clear that species that are nearest neighbors in the phylogenetic tree (near diagonal elements), share the same amount of contacts, thus showing that the CPR-MIN algorithm converges to the expected state. Nevertheless, the comparison with the real system (panel (B)) show that this state does not correspond to the one observed in nature.

In Fig. 5(A) we show the observed adjacency matrix of NCOR ordered in the same way as the phylogenetic tree. In panels (B) and (C) the asymptotic contact patterns obtained when applying the MIN-CPR and the MAX-CPR respectively are shown. These matrices correspond to configurations in which the parameters $D^{A,P}$ have reached a stationary value.

The phylogenetic trees of plants and animals, taken from Ref. [16], are included to guide the eye.

The MIN-CPR is dominated by phylogenetic proximity. To better understand the emerging contact pattern shown in Panel (B), one has to bear in mind that both animals and plants are considered on an equal footing. This generates an adjacency matrix that breaks into disconnected blocks in which phylogenetically close species of one guild interact with a similar group of the other guild. Such a situation excludes a nested scheme, since the species tend to specialize its contacts and generalists are ruled out of the system. The contact pattern of the NCOR system used as an initial condition, becomes therefore severely disturbed putting in evidence that it is unstable under the presence of the MIN-CPR in which phylogenetic proximity is the dominant rule.

A similar analysis for the MAX-CPR rule, shows the opposite behavior (panel (C)). The SNM causes few changes, reinforcing instead the presence of generalists and keeping the real matrix almost unchanged. The observed adjacency matrix must therefore be considered stable under such CPR. This run of the SNM also provides additional information. The NCOR system hosts a group of animals that are phylogenetically close and that are all fairly good generalists (e.g. the *turdus* group). Such a correlation between degree and phylogenetic proximity is not destroyed by the perturbations introduced by the SNM, *if the prevailing CPR is of the MAX type* (compare panels (A) and (C)).

The MAX-CPR rule leads to an asymptotically stable state which is almost perfectly nested. In panel (D) we show the adjacency matrices of panels (A) and (C) but reordered according to decreasing degree; in this way we can compare the empirically observed nested structure of the NCOR system with an asymptotically nearly perfect nested pattern.

It is worthwhile noticing that the CPR-MAX combines two effects. Contacts are reallocated according to the larger value between Γ_1 or Γ_0. These quantities may become large not only by involving species that are more phylogenetically distant
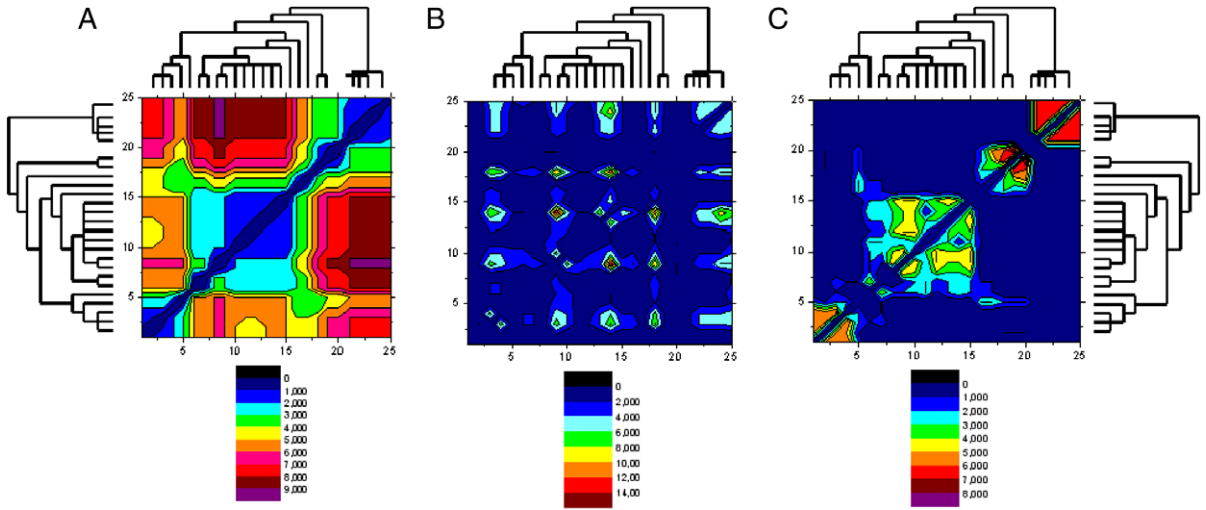
Fig. 4. (Color online) Panel (A): distance matrix for plants, NCOR system with species ordered according to the phylogenetic tree (shown along both margins of the matrix); panel (B):projected matrix for plants of the real system NCOR; panel (C):projected matrix for plants by the SNM MIN-CPR, after 140,000 iterations using as input the empirical matrix shown in panel (B). The NCOR matrix is not stable under this CPR. The axes order is the same for the three matrices.
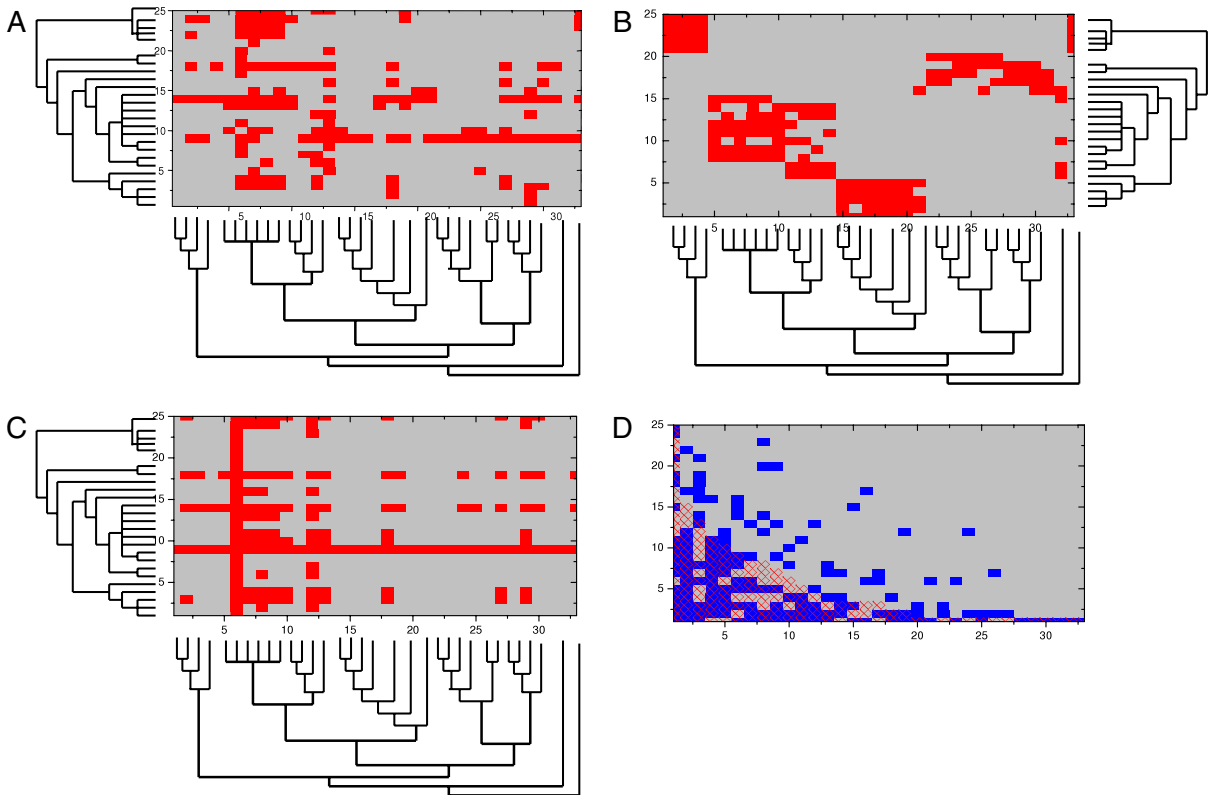

Fig. 5. (Color online) Adjacency matrices related to the study of the NCOR ecosystem. Panel (A): the empirical contact pattern with species ordered according to the phylogenetic tree (shown along both margins of the matrix); panel (B): contact pattern produced by the SNM after 80,000 iterations using as input the empirical matrix shown in panel (A) and the MIN-CPR alternative, species in the same order as in panel (A); panel (C): contact pattern produced by the SNM after 5000 iterations, using MAX-CPR, species in the same order as in panel (A); panel (D) same contact pattern as panel (A) (dark pixels) and (C) (hatched pixels) but species are ordered by their degree. Dark pixels correspond to observed contacts (panel (A)) while a slanted pattern corresponds to the theoretical results (panel (C)). Notice that while in panel (C) there are generalists and specialists, they do not show up in panel (B).

but also by involving a *greater number* of counterparts. In order to separate both effects we have made an alternative test dividing both sums by the number of counterparts that are found in both rows. In this way contacts are placed in those
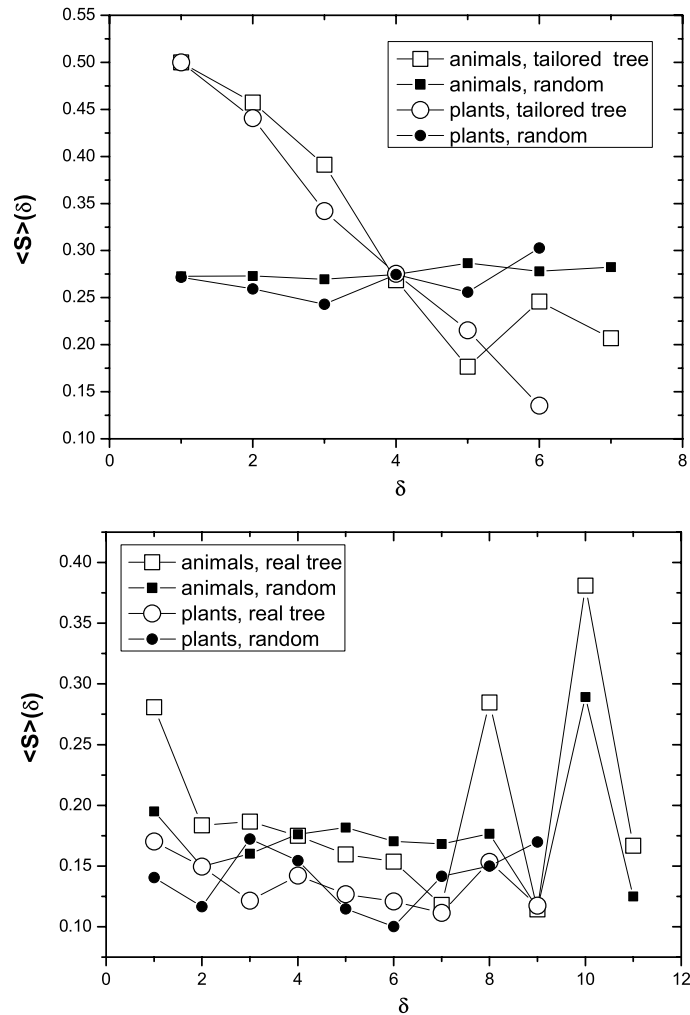
Fig. 6. Average ecological similarity as a function of the phylogenetic distance for plants and animals. We show in the same graph the results corresponding to both the ordered and the randomized phylogenetic tree. Upper panel: PERF-system. Bottom panel: NCOR matrix.

positions that correspond to a greater *average* phylogenetic distance. This test completely separates phylogenetic influence from any other. The results of these calculations point in the direction that a stationary stable contact pattern is never reached thus confirming what was said in Ref. [22] namely that the chief effect leading to a nested configuration is that all species tend to place their contacts with already crowded counterparts.

The above arguments have a greater reach because they hold for *any* distance matrix or similarity measure. We have made separate tests by checking the stationary contact patterns that are obtained by introducing alternative phylogenetic trees. We have considered trees that display a uniform branching rate and another in which all species successively stem from a single branch. Notice that in the case of a star phylogeny the self organization algorithm becomes identical to that of Ref. [22]. In all cases the MIN-CPR strategy leads to modular ecosystems in which contacts gather in nearly disconnected groups. As long as it is imposed that contacts have to take place between species that are phylogenetically close to each other, according to some criterion, some kind of specialization is favored and nestedness turns out to be hampered.

3.2. Comparative statistical properties: real system vs. benchmark and null models

In Fig. 6 we plot the average ecological similarity $\langle S(\delta) \rangle$ defined in Eq. (8) as a function of the distance δ separating any pair of species of the NCOR and PERF systems. It can be seen that for the PERF-system $\langle S(\delta) \rangle$ decreases monotonically with the phylogenetic distance for plants and animals. This monotonous trend is destroyed in the corresponding randomized system. The same plots made for the NCOR-system show no qualitative difference between the real and the randomized case.

In Fig. 7 we show the plots of $\Delta G^{A(P)}(\delta)$ as defined in Eq. (9) for all the systems (NCOR,OLAU, ARR1, ARR2) together with the one corresponding to the PERF benchmark system. If the number of contacts were dominated by the phylogenetic
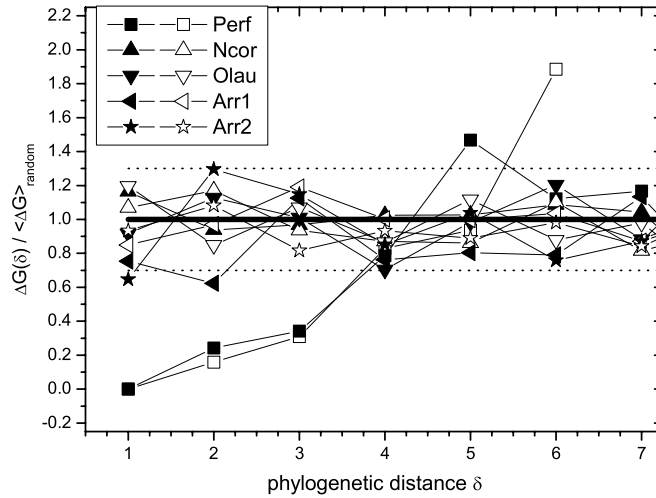
Fig. 7. ΔG as a function of the phylogenetic distance δ for $\delta \leq 7$. Filled (empty) symbols represent animals (plants). In every case, the values are normalized by dividing them by the average obtained from the null models for the PERF, NCOR, OLAU, ARR1 and ARR2 systems $\langle \Delta G \rangle_{random}$, therefore the horizontal full line at 1 corresponds to the random case.

proximity, the smallest value of $\Delta G(\delta)$ should be found at $\delta = 1$ because phylogenetically close species must have essentially the same degree. Since for a greater phylogenetic distance there is also a greater diversity of species, $\Delta G(\delta)$ must also be an increasing function of δ. The eventual decrease of the curves for large phylogenetic distances is a finite size effect due to the scarcity of pairs of species that are far away in a finite (and rather small) phylogenetic tree.

Those two features are clearly displayed in the PERF system. Both disappear when the phylogenetic trees are randomized. The curves corresponding to the observed adjacency matrices of the NCOR, OLAU, ARR1 and ARR2 natural systems are compared in Fig. 7 to an average of the null models for all systems obtained after 100 randomizations of each phylogenetic tree. They appear as random positive or negative fluctuations around the randomized average values thus not showing any appreciable trend making a difference between the observed and the randomized systems.

4. Conclusions

In the present paper we have developed a generalization of the SNM algorithm introduced in Ref. [22]. This aims at developing an alternative tool to test the validity of the various hypotheses that can be made to explain the internal organization of a mutualistic network. Usually, a hypothesis is made about the importance of some plausible behavioral trait on the pattern of contacts between the species of the system, and positive statistical correlations between them are interpreted as a plausibility argument in favor of the tested hypothesis. Instead, the present generalization introduces a dynamical algorithm that leads to the self organization of the system through a CPR that uses *external data*, in the sense that data are not obtained from the topology of the evolving network but obtained from some property of the system. Here, by introducing data from the phylogenies of the mutualists, we can test the conjecture, claimed in Ref. [16] that the pattern of contacts between species is dominated by the phylogenetic proximity between the species of the guild.

As many CPRs taking into account different degrees of influence of the phylogenetic trees of the mutualists may be considered, we chose to study in this article two extreme opposite situations. It is certainly possible to build a CPR that combines, for example, phylogenetic influence with some other property of the system (see for instance Ref. [28]), it is then essential to know the behavior in the limiting situations studied here. A straightforward modification of the phylogenetic SNM introduced here allows for testing different possible degrees of influence of the phylogenies of the interacting mutualists.

The general conclusion that stems from the SNM is that an interaction between species dominated by a CPR that favors *phylogenetic proximity* between mutualists of each guild cannot give rise to a nested contact pattern. Such a CPR tends to destroy a nested pattern of contacts leading to adjacency matrices with a clear tendency to break down into separate, nearly independent components, in which groups of phylogenetically close neighbors of both guilds hold contacts among each other but not with the rest of the species of the ecosystem.

We have also shown that an alternative interaction pattern dominated by phylogenetic diversity reproduces much better the observed order in mutualistic ecosystems. This interaction mechanism is one in which species hold contacts with the greatest possible diversity of mutualistic counterparts that are already visited by a greater number of species. It is worth stressing that this conclusion is fully consistent with those of Refs. [22,8] that are based upon a quantitative comparison with data observed in several representative and reasonably large mutualistic systems.

If a group of phylogenetically close species happens to have similar contact patterns, this turns out to be stable under such a maximal diversity interaction rule. In fact, it may happen that among species at distance $d = 1$, a subgroup of generalists
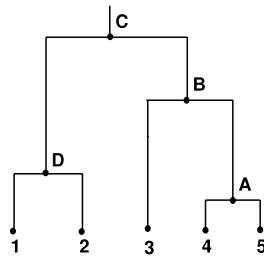
Fig. 8. An example of a simple phylogenetic tree is shown. The matrix of distances is: $d(1, 2) = d(4, 5) = 1$ $d(1, 3) = d(1, 4) = d(1, 5) = d(2, 3) = d(2, 4) = d(2, 5) = 3$ $d(3, 4) = d(3, 5) = 2$.

(as is the case of the turdus group in the NCOR system) might be found. In this case such a subgroup will remain stable under the organization algorithm of the SNM-MAX. This does not mean that *all* species at distance $d = 1$, will have a similar degree.

These results place serious doubts on the fact of considering the correlations between degree distributions and phylogenetic proximity as a sign of causation. The few circumstances in which they have been found to be statistically significant [16,18], point to the direction of considering that these are largely accidental, i.e. due to the existence of some phylogenetically related generalists.

The comparison of statistical studies performed on the real networks with those performed on a benchmark ideal model and on a null model, are compatible with the results issued from the phylogenetic SNM.

A dominant cause of the generalized nestedness found in mutualistic ecosystems lies perhaps on the simple fact that species that we observe in real systems today are those that tend to put the least possible restrictions on their mutualistic counterparts.

Acknowledgment

Appendix. The distance matrix

If a tree-like diagram is provided it is possible to extract from it a square matrix $d(k, k')$ of all distances between any two living species k and k'. One biologically plausible way to define such distance is to extract it from their evolutionary history. This amounts to considering that two species are "separated by a distance" that is measured by the time elapsed since they were differentiated in the course of evolution.

The evolutionary time can be represented by the length of the branches of the tree.

Resemblances and differences measured by this distance could be considered to involve a compound effect of all the traits that where considered in the analyses that led to the phylogenetic tree.

In order to obtain the distance matrix with all the distances one has to provide a uniform time scale for all branches, i.e., to provide a time order for all the branching points of the phylogenetic tree. The most parsimonious way of doing this is by defining that all branches that stem from a common ancestor and reach the tips of the tree must have the same length, counting lengths by starting from the tips and climbing upwards. This assumption is consistent with the constancy of an evolutionary clock [29].

With this procedure the distance matrix can directly be read from the topology of the phylogenetic tree. We exemplify this procedure in Fig. 8. We define that the two branches that lead to the species labeled (4) and (5) having a common ancestor in node (A) have a length equal to 1. By the same rule, the branch starting at species (3) that has a common ancestor with (4) and (5) in the branchpoint (B) has a length equal to 2. Moreover, the total length of the branches that have to be climbed starting from (1) or (2) to reach a common ancestor to all species in (C) must then have a total length of 3. In all these cases the lengths are defined except for an overall multiplicative scale factor. This ambiguity is however not relevant for the present analysis.

References

[1] C.M. Herrera, O. Pellmyr (Eds.), Plant-animal Interactions. An Evolutionary Approach, Blackwell Science, Oxford, 2002.
[2] N. Waser, J. Ollerton (Eds.), Plant–pollinator Interactions. From Specialization to Generalization, University of Chicago Press, Chicago, 2006.
[3] J. Memmott, N.M. Waser, M.V. Price, Tolerance of pollination networks to species extinctions, Proc. R. Soc. Lond. B (271) (2004) 2605–2611.
[4] D. Medan, N.H. Montaldo, M. Devoto, A. Mantese, V. Vasellati, G.G. Roitman, N.H. Bartoloni, Plant–pollinator relationships at two altitudes in the Andes of Mendoza, Argentina. Arctic, Antarctic and Alpine Research (ISSN: 1523-0430) 34 (2002) 233–241.
[5] R.F. Clements, F.L. Long, Experimental Pollination. An Outline of the Ecology of Flowers and Insects, Carnegie Institute of Washington, Washington, 1923.
[6] C. Robertson, Flowers and Insects: Lists of Visitors to Four Hundred and Fifty-Three Flowers, C. Robertson, Carlinville, IL, 1929.

[7] M. Kato, T. Makutani, T. Inoue, T. Itino, Insect-flower relationship in the primary beech forest of Ashu, Kyoto: an overview of the flowering phenology and seasonal pattern of insect visits, Contr. Biol. Lab. Kyoto Univ. 27 (1990) 309–375.

[8] E. Burgos, H. Ceva, R.P.J. Perazzo, M. Devoto, D. Medan, M. Zimmermann, A.M. Delbue, J. Theoret. Biol. 249 (2007) 307.

[9] J. Bascompte, P. Jordano, C.J. Melián, J.M. Olesen, Proc. Natl. Acad. Sci. USA 100 (2003) 9383–9387;
P. Jordano, J. Bascompte, J.M. Olesen, in: N. Waser, J. Ollerton (Eds.), Plant-Pollinator Interactions : From Specialization to Generalization, University of Chicago Press, 2006, pp. 173–199.

[10] J. Memmott, N.M. Waser, M.V. Price, Proc. R. Soc. B 271 (2004) 2605;
J. Memmott, D. Alonso, E.L. Berlow, A. Dobson, J.A. Dunne, R. Solé, J. Weitz, J. Pascual, M. Dunne, J.A. (Eds.), Ecological Networks: Linking Structure to Dynamics in Food Webs, Oxford University Press, New York, 2005, p. 325.

[11] César Hidalgo, Ricardo Hausmann, Proc. Natl. Acad. Sci. USA 106 (2009) 10570–10575.

[12] Sebastián Bustos, Charles Gomez, Ricardo Hausmann, César Hidalgo, PLOS ONE 7 (2012) e49393.

[13] Guido Caldarelli, Matthieu Cristelli, Andrea Gabrielli, Luciano Pietronero, Antonio Scala, Andrea Tacchella, PLOS ONE 7 (2012) e47278.

[14] L. Ermann, D.L. Shepelyansky, Phys. Lett. A 377 (2013) 250–256.

[15] P. Jordano, J. Bascompte, J.M. Olesen, Ecology Lett. 6 (2003) 69.

[16] E.L. Rezende, J.E. Lavabre, P.R. Guimaraes, Jr., P. Jordano, J. Bascompte, Nature 448 (2007) 925.

[17] E.L. Rezende, P. Jordano, J. Bascompte, Oikos 116 (2007) 1919.

[18] Susanne S. Renner, Nature 448 (2007) 877.

[19] D.P. Vázquez, M.A. Aizen, in: N. Waser, J. Ollerton (Eds.), Plant-Pollinator Interactions : From Specialization to Generalization, University of Chicago Press, Chicago, 2006, p. 200.

[20] M. Stang, P.G.L. Klinkhamer, E.van der Meijden, visitor web Oikos 112 (2006) 111–121.

[21] D.P. Vazquez, N.P. Chacoff, L. Cagnolo, Ecology 90 (2009) 2039.

[22] D. Medan, R.P.J. Perazzo, M. Devoto, E. Burgos, M. Zimmermann, H. Ceva, A.M. Delbue, J. Theoret. Biol. 246 (2007) 510.

[23] E. Burgos, H. Ceva, L. Hernández, R.P.J. Perazzo, M. Devoto, D. Medan, Phys. Rev. E78 (2008) 046113.

[24] M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. E 64 (2001) 026128.

[25] A.L. Barabási, R. Albert, Science 286 (1999) 500.

[26] J. Felsestein, Amer. Nat. 125 (1985) 1.

[27] S.P. Blomberg, T. Garland, A.R. Ives, Evolution **57** (4) (2003) 717.

[28] A.R. Ives, H.C.J. Godfray, Amer. Nat. 168 (2006) E1.

[29] M. Kimura M, New Sci. 107 (N1464) (1985) 41.