Article

Using Expectation Maximization and Resource Overlap Techniques to Classify Species According to Their Niche Similarities in Mutualistic Networks

Hugo Fort [1,†,*] and Muhittin Mungan [2,†]

[1] Complex Systems Group, Institute of Physics, Facultad de Ciencias, Universidad de la República, 11400 Montevideo, Uruguay

[2] Physics Department, Boğaziçi University, Bebek, 34342 Istanbul, Turkey;
  E-Mail: mmungan@boun.edu.tr

[†] These authors contributed equally to this work.

[*] Author to whom correspondence should be addressed; E-Mail: hugo@fisica.edu.uy;
  Tel.: +598-2-525-8624.

Abstract: Mutualistic networks in nature are widespread and play a key role in generating the diversity of life on Earth. They constitute an interdisciplinary field where physicists, biologists and computer scientists work together. Plant-pollinator mutualisms in particular form complex networks of interdependence between often hundreds of species. Understanding the architecture of these networks is of paramount importance for assessing the robustness of the corresponding communities to global change and management strategies. Advances in this problem are currently limited mainly due to the lack of methodological tools to deal with the intrinsic complexity of mutualisms, as well as the scarcity and incompleteness of available empirical data. One way to uncover the structure underlying complex networks is to employ information theoretical statistical inference methods, such as the expectation maximization (EM) algorithm. In particular, such an approach can be used to cluster the nodes of a network based on the similarity of their node neighborhoods. Here, we show how to connect network theory with the classical ecological niche theory for mutualistic plant-pollinator webs by using the EM algorithm. We apply EM to classify the nodes of an extensive collection of mutualistic plant-pollinator networks according to their connection similarity. We find that EM recovers largely the

same clustering of the species as an alternative recently proposed method based on resource overlap, where one considers each party as a consuming resource for the other party (plants providing food to animals, while animals assist the reproduction of plants). Furthermore, using the EM algorithm, we can obtain a sequence of successfully-refined classifications that enables us to identify the fine-structure of the ecological network and understand better the niche distribution both for plants and animals. This is an example of how information theoretical methods help to systematize and unify work in ecology.

---

1. Introduction

Mutualisms, *i.e.*, interactions between two species in which both benefit from the association, have been essential in the evolutionary diversification of life and are fundamental to preserve nature's biodiversity [1,2]. Without mutualism, the biosphere would be entirely different. For instance, approximately 90% of flowering plants are pollinated by animals [3], mainly by pollinator insects. In this mutualistic relationship, the pollinator obtains food (nectar), inadvertently picks up pollen and moves on to the next flower, allowing plants to reproduce.

However, the study of mutualisms as a non-trivial biological phenomena is a recent development in biology relative to the mainstream or Newtonian ecology [1], based on antagonistic interspecific interactions (e.g., competition, predation). Antagonistic interactions have received much more attention than mutualistic ones, because mutualistic interactions were initially viewed merely as insignificant oddities that in no way challenge the Darwinian natural selection paradigm based on competition. Still, even Darwin recognized that cooperative relationships might provide organisms with a competitive advantage; however, it was assumed that such phenomena were relatively rare. For these reasons, single trophic-level ecological communities have virtually been synonymous with the study of interspecific competition [4,5] and approached via the Lotka–Volterra equations of competition [6,7], and similarly, the dominant interspecific interaction in food webs involving different trophic levels is predator-prey [5]. However, during the past three decades, a dramatic change has occurred, and there has been an upsurge of interest in mutualism among ecologists and, in particular, to the application of cooperative versions of the well-known Lotka–Volterra equations [5,8]. In fact, mutualism still lacks a formal conceptual treatment at a depth comparable to that of antagonistic interactions [8] in such a way as to include all in a single comprehensive ecological framework.

In the case of plant-pollinator mutualism, interactions among plants and pollinators in general form large and complex mutualistic networks. These networks are bipartite, *i.e.*, plants and pollinators are the nodes, and the pollination interactions form the links between these nodes, including only links connecting nodes from the plant community to nodes from the animal community. Several authors have analyzed pollination networks [9–18] with respect to their nestedness (the extent to which specialist species interact with subsets of the species with which generalist species interact) and modularity (defined as the tendency of species to interact within and not between groups). The main findings are that

these networks tend to exhibit high nestedness (see also [17] for a measure of nestedness that takes into account qualitative network properties, as well, resulting in low nestedness) and a correlation between modularity and nestedness that depends on the size and connectance of the network [15,16]. Here, connectance is the the ratio of actual links over the total number of possible links in the network.

An important limitation in the study of mutualistic plant-pollinator networks is the scarcity and incompleteness of empirical available data. For most of these networks, there are only qualitative data in the form of qualitative interaction matrices (also called adjacency matrices) I_{ap} with binary elements: 1(0) if the pair formed for animal in row a and plant in column p interacts or not. Less coarse information, like quantitative interaction matrices q_{ap} specifying the number of observed visits of each animal species a to each plant species p, only exist for a small subset of pollination networks.

From the adjacency matrices, it is possible to obtain useful insights into the architecture of the ecological communities for developing quantitative methods for mutualistic networks. Quantitative methods in turn are crucial to assess the degree to which mutualisms determine the relative species abundances, recognized as a fundamental question that, if answered, would substantially advance population and community ecology [19]. For example, a method was recently proposed to approach bi-partite plant-pollinator communities by considering each party as a resource for the other party [20]: plants provide food to animals, while animals facilitate the reproduction of plants. Since different species of pollinators often visit the same species, there is resource overlapping. In a similar manner, resource overlap also occurs between plants, since different species of plants share the same pollinator resource. Resource overlapping implies that plants and pollinators compete for resources (pollinators and plants, respectively). For interspecific competition, the classical community ecology theoretical approach is niche theory [5]. The concept of resource utilization niche, introduced by MacArthur and Levins [21], focuses on how species use consumable resources and is a central element of the classical theory of community ecology. In fact, niche theory was essentially a group of theoretical models designed to address the problem of how many and how similar coexisting species could be within a given community [21,22]. According to niche theory, differences among species in their niches are important in determining the outcome of species interactions, as might be revealed in their distributions and/or abundances, as well as in their biodiversity and their functional role in ecosystems. The relative utilization of resources along a resource spectrum or niche axis can be described as a frequency distribution. Species are thus characterized in terms of their similarity in resource use or their niche overlap. According to niche theory, there is a strong correspondence between the degree of niche overlap between two species and the intensity of their competition for shared resources [21]. An important property of the species closely connected with their niche overlap is the degree of generalization D [14]. For a species s of one of the two mutualistic network partitions, D_s is defined as the number of species of the other network partition with which this species interacts, *i.e.*, for an animal (plant) species, D_a (D_p) is given by the number of ones in a row (column) of I_{ap}. The lower (higher) the degree D is, the more specialist (generalist) the species is and the narrower (wider) its niche. It turns out that for a large set of plant-pollinator networks, covering a wide geographical range, the resource overlap intensity for pollinator species exhibits a pattern of clusters of generalists separated by clusters of specialists [20]. A similar relationship between D and niche position x emerges from the application of the resource overlap (RO) method when applied to plant species (see Supplementary Material in [20]).

The purpose of this paper is to connect network theory with classical ecological niche theory for mutualistic plant-pollinator webs by using the expectation maximization (EM) algorithm, a versatile statistical inference method [23–26]. One feature that makes EM particularly useful is that it can uncover a very broad range of types of structures in networks, using only the patterns of connections between the nodes (the information contained in I_{ap}) and without prior knowledge of what one is looking for [24]. Here, we consider a set of 20 plant-pollinator networks obtained from the supplementary dataset of Rezende *et al.* [27] and the on-line interaction web database (IWDB) [28]. We apply the RO and EM methods to each partition of these bi-partite networks, *i.e.*, plants and pollinators. We find that the clustering of nodes, representing species of plants or animals, classified according to their connection similarity by EM, is highly compatible with the ordering of these species on a linear niche axis produced by the RO method used in [20]. The classes uncovered by the EM method are a collection of nodes with a similar connection pattern. From an ecological point of view, we can therefore regard these classes as collections of species that play a similar role in the network. The identification of such classes/roles is useful, since it allows for a description of the ecological network at the level of aggregated nodes, *i.e.*, compartments [29,30].

This article is organized as follows: in the Methods Section, we describe our dataset and review the RO and EM methods. We present our results in Section 3, followed by a discussion of our results in Section 4 and provide conclusions in Section 5.

2. Methods

In this section, we present the dataset that we analyzed and next review the RO and EM methods, which are both based on qualitative data about the mutualistic interactions. The reader is referred to the references provided in the subsections for further details. Network figures were drawn using the graphics tool Pajek [31]. We apply the RO and EM methods on the unipartite pollinator-pollinator and plant-plant networks. These networks are obtained from the bi-partite plant pollinator networks by a one-mode projection as follows: given any two pollinators i and j, we draw an edge between them, if there exists at least one plant species that both pollinate. The plant-plant network is obtained in a similar manner by drawing an edge between any pair of plants if they are pollinated by at least one common pollinator.

2.1. Dataset

Table 1 shows the plant-pollinator networks that we have analyzed. For each plant-pollinator network, we also provide the source reference. Four-letter network abbreviations follow the naming convention used in the supplement of main source [27].

Table 1. The 20 plant-pollinator networks forming the dataset of this study.

Network Abbreviation	Data Reference	Plants	Pollinators	Interactions
ARR3	[27]	36	25	81
DISH	[27]	16	36	85
DUPO	[27]	11	38	106
EOLZ	[27]	31	76	456
ESKI	[27]	14	13	52
HOCK	[27]	29	81	179
LLAO	[32,33]	9	27	39
MED2	[27]	23	72	125
MEMM	[27]	25	79	2183
MOMA	[27]	11	18	38
MULL	[27]	105	54	204
OFLO	[27]	10	12	30
OFST	[27]	8	42	79
OLAU	[27]	29	55	145
OLLE	[27]	9	56	103
PERC	[27]	61	36	178
PRCA	[27]	41	139	374
SAAG	[34]	51	25	190
SCHE	[35]	7	32	59
STAN	[36]	25	111	231

2.2. Construction of the Niche Axis from Resource Overlap

As remarked before, the idea of the RO method is to regard plants (animals) as a one-dimensional discrete resource spectrum [22] for animals (plants), and the position of species along the niche axis will be obtained from their niche overlap [21]. Our choice of a one-dimensional niche-axis is for simplicity, *i.e.*, following the principle of parsimony, we introduce the minimal and sufficient complexity to achieve the desired goal. Furthermore, higher dimensional niches do not introduce dramatic changes (see, for example, [37]). We choose integer niche variables rather than real variables, because we do not have information on relevant niche variables. For example, had we data on trait sizes for pollinator species, we could use these as real-valued niche variables. Ours is therefore a reverse engineering problem: we determine the RO between all pairs of species (i, j), where i and j are arbitrary integers, and we sort these in such a way that all species with large RO become grouped together.

We quantify the resource overlap between species i and j by the Jaccard similarity index [38], which can be readily obtained from I_{ap} and is given by [38]:

$$J_{ij} = \frac{P_{ij}^{11}}{P_{ij}^{01} + P_{ij}^{10} + P_{ij}^{11}} \tag{1}$$

where P_{ij}^{11} represents the number of species of one party (plants or animals) that interacts with both species, i and j, of the other party (animals or plants, respectively) and P_{ij}^{10} (P_{ij}^{01}) the number of mutualistic species that interact only with i (j). Notice that if the two species share all of the mutualistic species, *i.e.*, when $P_{ij}^{10} = P_{ij}^{01} = 0$), then $J_{ij} = 1$. This situation of full overlap in general occurs between

maximum specialists with $D = 1$ and is less likely to occur between species with intermediate values of D. To construct the niche axis for each party, it is necessary to group together all species, either plants or animals, that exhibit a large resource overlap, reordering them from degree-order to niche-order, by assigning to each species an integer x between 1 and N (N_A animal species or N_P plant species) that denotes its niche "position". To accomplish this, we use a "simulated annealing" algorithm [39]. This algorithm favors the proximity of interacting species by minimizing a cost function or "energy" E, assigned to each ordering, given by:

$$E = \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} \left\{ |J_{ij} - J_{i+1j}| + |J_{ij} - J_{i-1j}| + |J_{ij} - J_{ij+1}| + |J_{ij} - J_{ij-1}| \right\} \tag{2}$$

where i and j are integers, $|\dots|$ stands for the positive value of the difference of the Jaccard similarity indices for neighboring positions and $d_{ij} = |i - j|$ is the distance from the "site" (i, j) to the diagonal. First, the absolute values of the differences between the Jaccard coefficient of a focal species J_{ij} and each of its four nearest neighbors penalizes large differences between neighbor species and attempts to group together all species with large RO. Second, diagonal terms always have maximum RO, *i.e.*, $J_{ij} = 1$ (because the RO of one species with itself is by definition = 1). Weighting terms proportional to the distance d_{ij} accelerate the convergence of the algorithm, but this is not required. Starting with a decreasing D-ordered list of species and its corresponding interaction matrix with energy E_0, the algorithm proceeds by randomly choosing a pair of species and swapping their positions on the ordering. A new Jaccard matrix is produced for this new ordering, and its cost is recalculated using Equation (2). If the cost decreases, the change is accepted. If the cost is increased by ΔE, the change is accepted with probability $\exp(-\Delta E/T)$, where T is a virtual temperature. We started with $T = 5E_0$, and every 50 of these simulated annealing steps (SAS), the temperature is lowered to 5% of its previous value. A SAS comprises a number of random choices equal to N. When changes are not accepted, they are discarded, and a new pair of species is chosen to repeat the process. This procedure is repeated until the calculated value of cost is stabilized for a configuration in which sites (i, j) far from the diagonal, or equivalently very different positions along the niche axis, correspond to pairs of species that have a small overlap.

2.3. Applying Statistical Inference Methods to Extract the Structure Underlying A Network

It is plausible to assume that similar pollinators pollinate similar plants. In terms of the network of plant-pollinator interactions, such an assumption implies that the node neighborhoods of two such pollinators will be similar, so that the set of (plant) nodes that each of these pollinators connect to has a strong overlap. Nodes whose node-neighborhood is identical are called structurally equivalent [40]. Thus, the extent to which two nodes are structurally equivalent can provide information for assessing their similarity. We therefore ask: given a network and making the assumption that the connection between nodes arise as a result of some inherent, not necessarily readily-observable property of the nodes, can we cluster the nodes based on the similarity with respect to such a property? It turns out that a statistical inference method based on expectation maximization (EM), introduced for networks by Newman and Leicht [24], is highly suitable to extract such information [25]. In this approach, it is assumed that each node belongs to one of a finite number N_c of distinct classes (so that nodes belonging to the same class have the same properties). In the EM implementation of Newman and Leicht, the

class membership of each node is treated as unobserved, or missing information [23], meaning that this information is not *a priori* available to us, and we use the EM method to infer this information *a posteriori* from the network connection information and an underlying probability model, by maximizing the corresponding likelihood function. In the following, we briefly describe the EM implementation and refer the reader to [24–26] for further details.

Consider a network specified in terms of the adjacency matrix a_{ij}, where $a_{ij} = 1$ if nodes i and j are connected by an (undirected) edge, and $a_{ij} = 0$ otherwise. In the probability model, we assume that each node i has an intrinsic property g_i that can take as many different values as there are number of classes N_c. The probability of a connection between two nodes i and j, given that they belong to classes g_i and g_j, is defined as:

$$\text{Prob(edge between } i \text{ and } j | g_i, g_j) = \theta_{g_i, j} \theta_{g_j, i} \tag{3}$$

where θ_{rj} is the probability of having a connection from a node belonging to class r to a certain node j. Given the probabilities θ_{rj} and a class assignment of nodes $\{g_i\}$, it is assumed that edges are observed independently according to Equation (3) above. With the corresponding log-likelihood given as:

$$\mathcal{L} = \sum_{ir} \gamma_{ir} \left[\ln \pi_r + \sum_j a_{ij} \ln \theta_{rj} \right] \tag{4}$$

the EM implementation then results in the following set of self-consistent equations (EM equations):

$$\pi_r = \frac{1}{N} \sum_i \gamma_{ir}, \tag{5}$$

$$\theta_{rj} = \frac{\sum_i a_{ij} \gamma_{ir}}{\sum_i k_i \gamma_{ir}}, \tag{6}$$

$$\gamma_{ir} = \frac{\pi_r \prod_j \theta_{rj}^{a_{ij}}}{\sum_s \pi_s \prod_j \theta_{sj}^{a_{ij}}} \tag{7}$$

Here, N is the number of nodes, k_i is the degree of node i (number of nodes adjacent to i), π_r is the probability that a randomly-selected node belongs to class r, while γ_{ir} is the posterior probability that node i belongs to class r. The solution to the EM equations is obtained by treating them as a mapping and iterating until a fixed point is reached: we start with randomly-chosen distributions γ_{ir} and evaluate the left-hand sides of the EM equations, Equations (5) to (7). Next, these values are substituted into the right-hand sides of the equations to obtain updated values, and this is repeated until convergence is reached.

As the network a_{ij}, we use the pollinator-pollinator network, *i.e.*, the one-mode projection of the bi-partite plant-pollinator network. Recall that this means that we consider as nodes the set of pollinators only and set $a_{ij} = 1$, if there is at least one plant species that both pollinators i and j pollinate (otherwise, we set $a_{ij} = 0$). Note that the resulting network is not necessarily transitive, meaning that the existence of edges between species i and j, as well as j and k does not imply an edge between species i and k, since the common sets of plants giving rise to each of the two edges could be disjoint. On the other hand, if in the bi-partite network, n pollinators $i_1, i_2, i_3, \ldots i_n$ are structurally equivalent, then in the pollinator-pollinator network, they will form a complete subgraph, meaning that all possible edges between those nodes will be present. plant-plant networks can be constructed in a similar manner.

For each network, and a given number of classes N_c, we generated 100 different sets of random initial distributions γ_{ir} and iterated the EM equations to convergence (to machine precision). From the resulting classifications, we chose the one with the largest log-likelihood Equation (4) [24,25].

We have attempted classifications of the pollinator-pollinator networks with up to 16 classes and found that, owing to the number of pollinators in these networks, classifications into $N_c > 8$ classes rarely yield genuine classifications. By this, we mean that the resulting classifications contain classes to which no pollinators have been assigned.

The situation is more pronounced in the case of the plant-plant networks, where the number of plant species is generally less than that of the pollinator species and often of the order of the number of classes N_c, as is evident from Table 1. In particular, the plant-plant networks of DUPO, EOLZ, ESKI, OFST and OLLE are almost complete graphs, meaning that all plants are connected to almost all other plants. The connection patterns of plants being highly similar means that the connectivity of individual plant nodes does not provide useful information to distinguish these plants. As a result, the EM classifications assigns all of these plants to the same class. To give an example, the plant-plant network of EOLZ, containing 31 plant species, has 14 species having connections to all other species. Only 7 species connect to less than 27 other species, and the minimum number of connections to other species in the network is 23.

For most of the networks we considered and particularly for the classifications in to 2–5 classes, despite starting from 100 different initial conditions, only a few distinct classifications were obtained; thus, it was not deemed necessary to perform extended searches for classifications with the highest achievable log-likelihood, as was done for instance in [25].

Let us emphasize that we are not attempting to fit for the optimal number of classes, *i.e.*, we are not doing model selection. Rather, our aim here is to use the EM method as an exploratory tool. We therefore treat the number of classes as a parameter in order to track how the classifications change as we vary the number of classes. Alternatively, if our goal were to determine the optimal number of classes, this could be achieved by using the Akaike information criterion [41], as for example done by Allesina and Pasqual for a similar statistical clustering approach to ecological networks [29].

3. Results

3.1. EM Classifications and Network Structure

Figure 1 shows the four-fold EM classification of the unipartite pollinator-pollinator network SCHE of Schemske *et al.* [35]. The color of each node i codes the class membership, namely the class r for which γ_{ir} is maximum. Notice:

(i) clique-like properties: nodes assigned to the same class tend to connect to each other;

(ii) distinct connection-preferences: for example, "red" pollinators connect predominantly to "red" and "green" pollinators; likewise, "blue" pollinators connect to "blue" and "green" pollinators, *etc.*

These two properties are generic features of almost all networks that we have analyzed. Thus, the EM algorithm when applied to plant-pollinator networks reveals useful information with respect to connection similarity and connection preferences, thereby validating the suitability of the method.

Figure 1. Four-fold expectation maximization (EM) classification of the unipartite pollinator-pollinator network SCHE. The color of each node i codes the class membership, namely the class r for which γ_{ir} is maximum. Since the network contains pollinators with similar pollination preferences, the resulting network structure contains cliques. The EM algorithm uncovers this clique structure, as is evident from the fact that the node colors coincide with the cliques.

3.2. Crispness of EM Classification

The class assignment probabilities γ_{ir} also allow us to assess the crispness of a classification. To this end, we consider the classification entropy [25],

$$S = \frac{1}{N} \sum_{ir} \gamma_{ir} \ln \gamma_{ir} \tag{8}$$

where N is the number of nodes in the network and the indices i and r run over the nodes and classes, respectively. Hence, $0 \leq S \leq \ln N_c$, with N_c being the number of classes chosen for the classification. A crisp classification will imply $S = 0$, *i.e.*, all nodes are classified as belonging to unique classes. Thus, the parameter S assesses the crispness of the classification found.

Tables 2 and 3 provide the classification entropy for the 2–5 fold-classifications of the networks that we have considered. For the pollinator networks in Table 2 and the plant-networks in Table 3, we see that for most of the classification entropies S, as given by Equation (8), we obtain values $S < 0.1$. Thus, the classifications resulting from the EM algorithm are rather crisp.

Table 2. Classification entropy S and niche alignment quality Q for the 20 pollinator-pollinator networks. Note that a five-fold classification for the network ESKI could not be found.

| | pollinator-pollinator | | | | | | | |
| | 2 Class | | 3 Class | | 4 Class | | 5 Class | |
Network	S	Q	S	Q	S	Q	S	Q
ARR3	0.000	**1.000**	0.020	**1.000**	0.015	0.840	0.044	0.960
DISH	0.039	0.556	0.013	0.722	0.043	0.889	0.055	0.889
DUPO	0.042	0.658	0.079	0.526	0.116	0.579	0.105	0.579
EOLZ	0.024	0.684	0.030	0.684	0.029	0.671	0.021	0.658
ESKI	0.040	**1.000**	0.174	**1.000**	0.148	**1.000**	-	-
HOCK	0.016	0.741	0.003	0.617	0.004	0.617	0.005	0.716
LLAO	0.000	0.778	0.000	0.926	0.028	0.815	0.032	0.889
MED2	0.015	0.222	0.015	0.222	0.007	0.181	0.011	0.139
MEMM	0.017	0.165	0.008	0.190	0.012	0.177	0.011	0.228
MOMA	0.000	**1.000**	0.007	**1.000**	0.019	**1.000**	0.133	0.889
MULL	0.026	0.833	0.038	0.926	0.028	0.778	0.017	0.852
OFLO	0.022	**1.000**	0.258	**1.000**	0.193	**1.000**	0.222	**1.000**
OFST	0.013	0.714	0.025	0.929	0.006	**1.000**	0.016	0.833
OLAU	0.020	0.600	0.018	0.509	0.016	0.636	0.011	0.727
OLLE	0.040	0.786	0.026	0.929	0.031	0.964	0.025	0.875
PERC	0.013	0.861	0.035	0.861	0.034	0.639	0.059	0.667
PRCA	0.001	0.813	0.008	0.763	0.009	0.863	0.007	0.806
SAAG	0.055	0.640	0.199	0.680	0.202	0.800	0.258	0.840
SCHE	0.000	**1.000**	0.027	0.938	0.006	0.969	0.003	0.969
STAN	0.000	**1.000**	0.000	**1.000**	0.001	0.901	0.009	0.775

Table 3. Classification entropy S and niche alignment quality Q for plant-plant networks. The networks for DUPO, EOLZ, ESKI, OFST and OLLE are nearly complete graphs, and classification into 2 or more classes was not found (see the end of Section 2.3 for details).

| | plant-plant | | | | | | | |
| | 2 Class | | 3 Class | | 4 Class | | 5 Class | |
Network	S	Q	S	Q	S	Q	S	Q
ARR3	0.000	**1.000**	0.006	0.833	0.003	0.750	0.007	0.694
DISH	0.011	**1.000**	0.063	**1.000**	0.216	0.875	0.236	0.875
LLAO	0.000	**1.000**	0.000	**1.000**	-	-	-	-
HOCK	0.000	0.828	0.024	0.828	0.047	0.828	0.026	0.724
MED2	0.011	0.435	0.005	0.348	0.021	0.391	0.025	0.435

Table 3 . *Cont.*

| | plant-plant | | | | | | | |
| | 2 Class | | 3 Class | | 4 Class | | 5 Class | |
Network	S	Q	S	Q	S	Q	S	Q
MEMM	0.239	0.760	0.481	0.560	-	-	-	-
MOMA	0.000	**1.000**	0.000	**1.000**	-	-	-	-
MULL	0.008	0.848	0.017	0.914	0.013	0.867	0.030	0.867
OFLO	0.016	**1.000**	0.365	**1.000**	0.327	**1.000**	0.407	0.800
OLAU	0.001	**1.000**	0.056	0.966	0.027	0.862	0.043	0.724
PERC	0.023	0.639	0.011	0.639	0.021	0.672	0.020	0.803
PRCA	0.109	0.927	0.089	0.463	0.095	0.561	0.074	0.659
SAAG	0.014	0.745	0.011	0.804	0.026	0.706	0.050	0.686
SCHE	0.136	0.571	0.411	0.857	-	-	-	-
STAN	0.040	**1.000**	0.085	0.920	0.143	0.920	0.160	0.920

3.3. Compatibility of RO-Based Niche-Ordering of Species with Their EM Classifications

The RO approach of Subsection 2.2 results in an ordering of the species of the unipartite pollinator-pollinator and plant-plant networks along a one-dimensional circular niche axis. We ask next how compatible such an ordering is with the classification of the species as obtained from the EM algorithm. In order to measure the extent to which the EM classification is compatible with the inferred niche-ordering, we proceed as follows: Given a niche ordering, for each EM class r, we identify the regions of contiguous sites on the niche-axis that belong to the same class. Among these, we pick the largest region and denote its size by L_r. If the classification is fully compatible with the niche order, then for each class r, there is only a single contiguous region, and the sum of the sizes of these regions L_r is N. In all other cases, this sum is less than N. We therefore define the quality index Q as:

$$Q = \frac{1}{N} \sum_r L_r \tag{9}$$

so that $0 < Q \leq 1$, and full compatibility of EM classification with RO niche ordering implies $Q = 1$.

To illustrate this, consider the 2-, 3- and 4-fold classifications of the SCHE pollinator-pollinator network [35] shown in Figure 2. This network consists of 32 pollinators. In the two-fold classification, the "blue" and "red" groups each form single contiguous domains of 24 and eight pollinators, respectively. Thus, $L_{\text{blue}} = 24$, and $L_{\text{red}} = 8$, so that $Q = 1$. Now, consider the three-fold classification, as shown in the middle panel of the figure. Comparing to the two-fold classification in the top panel of Figure 2, we see that the "blue" group has split into two, forming a "blue" and "green" group. However, the new "blue" group is not contiguous anymore; it has a second domain on the niche axis of size two (the pollinators with IDs 13 and 22). The total number of pollinators in the largest contiguous domains is therefore $32 - 2 = 30$, and hence, $Q = 30/32 = 0.9375$, as indicated in Table 2. Comparing to the four-fold classification, we see that the boundary of the "blue" and "red" domains turns into a new group,

"yellow", so that there is now only one pollinator that is not part of a contiguous domain, and therefore, $Q = 31/32 = 0.969$.
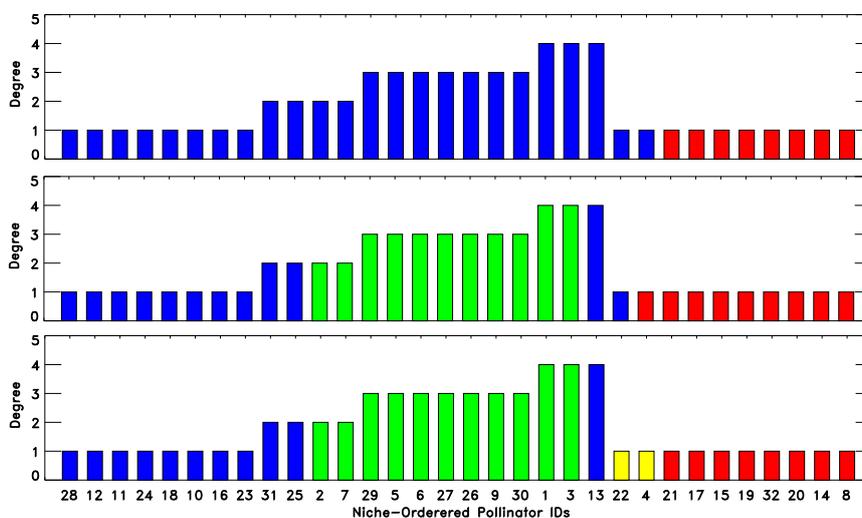
Figure 2. (Top to bottom) 2-, 3- and 4-fold expectation maximization (EM) classification of the SCHE [35] unipartite pollinator-pollinator network. Each figure is a plot of degree *vs.* pollinator ID (shown on the x-axis of the bottom figure). The pollinators have been placed according to their ordering on the niche axis, as inferred by the resource overlap (RO) method of Subsection 2.2. The color of the bars denotes the class that the pollinator has been assigned to by the EM algorithm. The degree refers to the number of different plant species that the pollinator pollinates. A large (low) value of the degree implies that the pollinator is a generalist (specialist).

In Table 2, we list the Q values for the pollinator-pollinator networks. Quite a number of classifications have a niche alignment quality of $Q = 1$, as indicated in bold in the table. This implies that the EM classification of the pollinators is in perfect agreement with their ordering on a niche axis, as obtained by the RO method. However, there are a few cases where this alignment is very poor, as for example in the MED2and MEMMnetworks. Table 3 shows similar findings for the plant-plant networks: close to crisp classifications and high niche alignment quality Q.

4. Discussion

The RO method produces two clear patterns. Firstly, the Jaccard matrices display a block structure along the diagonal that corresponds to the niche axis (dark red blocks of a side greater than one in the right panel of Figure 3). Such a structure is typical of Jaccard matrices computed by using MacArthur and Levins' niche overlap [20]. These blocks correspond to sectors of strong resource overlap. The blocks of maximum overlap, $J_{ij} = 1$, occur for sets of specialists ($D_i = 1$ or 2) that all share one or at most two plants as their resource.

Secondly, the degree of generalism *versus* the niche position, $D(x)$, exhibits a pattern of "hills" separated by shallow "valleys" corresponding to alternating clusters of generalists and specialists (upper panel of Figures 2 and 3). This undulated landscape can be understood, because as generalist species

(plants or pollinators) have wider niches than specialists, the niche overlap between a generalist and a specialist is in general smaller than the niche overlap between pairs of generalists or pairs of specialists that share some resource.

In turn, these clusters of generalists or specialists are often associated with classes obtained when applying the EM method. Note in particular how in Figure 3, the order and size of the contiguous EM classes in the right top panel closely follows the block structure along the diagonal of the resource overlaps depicted in the bottom right of the figure.
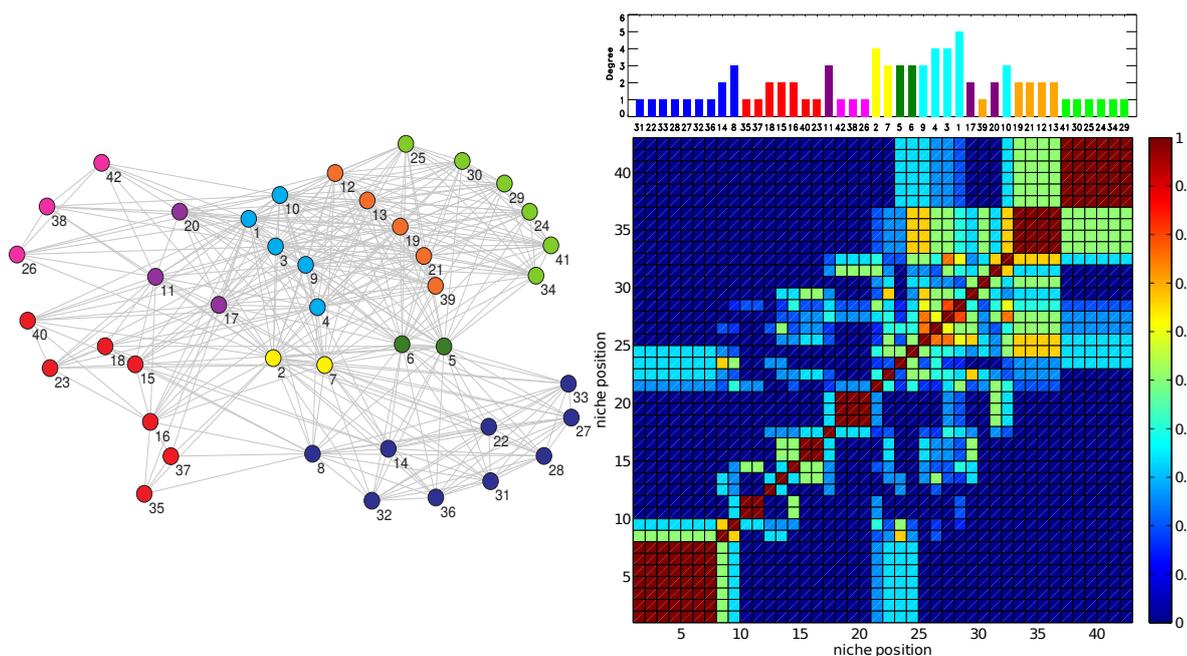
Figure 3. (Left) The OFST pollinator-pollinator network with the nine-fold EM classification. The color of each node denotes the class in which the pollinator has been classified by the EM algorithm. (Right, top) The corresponding plot of degree D *vs.* inferred relative location on a one-dimensional niche axis (legend as in Figure 2). The numbering of the pollinators corresponds to the IDs given in the data source [27] and agrees with those in the left panel. (Right, bottom) The Jaccard J_{ij} matrix measuring the resource overlap of species i and j. $J_{ij} = 0$ (dark blue) corresponds to no resource overlap, while $J_{ij} = 1$ (dark red) corresponds to complete resource overlap, in general only possible between two specialists with $D = 1$ that have the same resource.

Therefore, our analysis sheds light on how those groups of ecologically-similar species self-organize into niches based on the similarity of some of their biological traits and their abundance [20]. Indeed, the niche concept is an important element in many aspects of ecological thinking, across levels of organization, from individuals (concerning their behavior, morphology and physiology) to ecosystems (to evaluate how species participate in ecosystem functioning) [42]. The niche framework helps to devise each species place and role in the community to which it belong, as well as it provides a better understanding of species coexistence, new insights and interpretations about ecological patterns and processes in ecology and how fluctuating environments might regulate population dynamics and species interactions.

As we have seen, there are classes that emerge from a classification into a low number of classes N_c and persist while N_c increases. At the same time, there are classes that are refined as the number of classes increases. Both types of behavior can reveal meaningful information about the network. To illustrate these points, we consider the classifications of the DISH pollinator-pollinator network into $N_c = 3, 4, 6, 7$ and 9 classes, as depicted in Figure 4. Consider for example the set of pollinators with IDs $\{9, 11, 13, 14\}$. These pollinators are classified as belonging to the same "green" group for all of the classifications shown. In fact, this turned out to be true for all classifications attempted ($N_c \leq 12$).
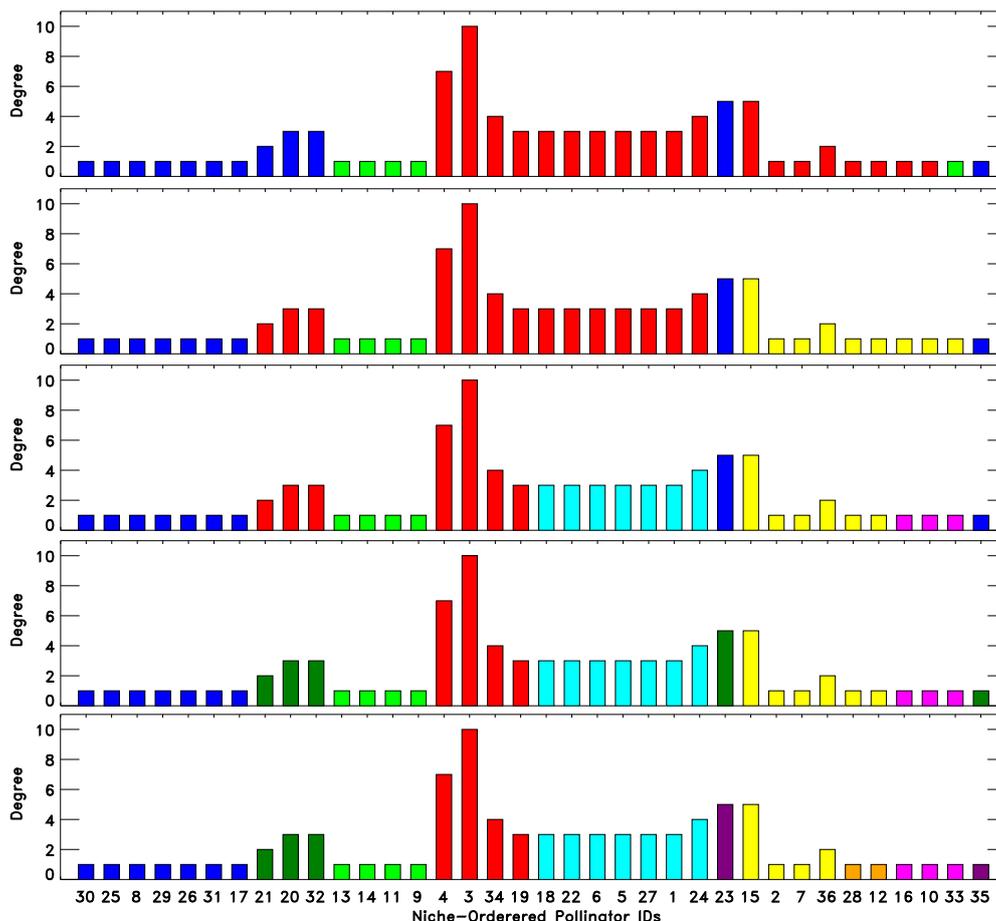
Figure 4. (Top to bottom) 3-, 4-, 6-, 7- and 9-fold EM classification of the DISH unipartite pollinator-pollinator network. As the number of classes increase, we see both robustness, namely certain species being classified together, as well as the emergence of finer structure, as classes are split up (legend as in Figure 2).

On the other hand, most of the other classes, as expected, split up into smaller classes when N_c is increased. In this case, an increase of N_c often reveals the internal structure of a cluster of species. Again, for DISH, Figure 4 considers the first ten species on the niche-axis, $\{30, 25, 8, 29, 26, 31, 17, 21, 20, 32\}$. These 10 species constitute almost the whole blue class for the $N_c = 3$ classification, while when $N_c = 4$, they split up into two classes, Species 21, 20 and 32 becoming part of the "red" class and then forming their own class when $N_c = 7$. Likewise, comparing the three- and four-fold classifications, we see that the "red" class splits into two, forming a new class containing the pollinator

IDs $\{4, 3, 34, 19, 18, 22, 6, 5, 27, 1, 24\}$, which coincides with the set of generalists (large degrees), at the center of the niche axis.

5. Conclusions

Ecological networks are an interdisciplinary research field, where physicists, biologists and applied mathematicians have been working together in order to understand the structure of ecological networks, one of the outstanding challenges in the study of complex systems.

The case of mutualistic interactions is of paramount importance; for example, we have seen that the mutualism between plants and pollinators is vital for the maintenance of most natural systems. Moreover, pollinators play a key role in communities, and their decline is leading to a crisis that is affecting both native and cultivated plant species [43–45]. A better understanding of the structure of plant-pollinator communities and how this relate to their dynamics is key to developing models capable of explaining the relative species abundance. These quantitative models are useful for several purposes, like predicting the consequences of the loss of pollinators on crop yield [46].

A recently-developed model for pollination, in terms of mutualistic interactions between plants and pollinators aggregated into classes with similar traits, was able to predict the relative abundances of plants and pollinators of a Mediterranean network [30]. The problem is that only for a handful of these networks are there public data about species traits, like the nectar depth for plants and the proboscis length for animals, which seem to play a crucial role in determining the set of mutualistic interactions occurring between plants and pollinators [36,47–51]. The general procedure we proposed here, based on the combination of EM and RO, offers a possible alternative to overcome this lack of data on traits by classifying plants and pollinators into classes associated with clusters of species with overlapping niches (*i.e.*, grouped on a niche axis).

From quantitative interaction matrices, q_{ap}, it is possible to compute using a generalization of Jaccard similarity indices, basically internal products, more refined overlap matrices and the corresponding species niche distributions [20]. Something worth addressing in future work would be to extend the EM method in such a way as to incorporate quantitative information on mutualistic interactions. For example, we have formed the unipartite pollinator-pollinator networks by including a link between two pollinators when they have at least one common plant resource. However, it would be natural to consider weighted links, where the weight of a link is the quantitative amount of resource overlap for the two pollinators. Likewise, the number of visits of a pollinator to a plant could also be used as a quantitative input into the EM method.

Author Contributions

Both authors substantially contributed to the reported work. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Boucher, D.H. *The Biology of Mutualism: Ecology and Evolution*; Oxford University Press: Oxford, UK, 1988.
2. Stachowicz, J.J. Mutualism, facilitation, and the structure of ecological communities. *BioScience* **2001**, *51*, 235–246.
3. Regan, E.C.; Santini, L.; Ingwall-King, L.; Hoffmann, M.; Rondinini, C.; Symes, A.; Taylor, J.; Butchart, S.H.M. Global Trends in the Status of Bird and Mammal Pollinators. *Conserv. Lett.* **2015**, doi:10.1111/conl.12162.
4. May, R.M. *Stability and Complexity in Model Ecosystems*; Princeton University Press: Princeton, NJ, USA, 1973.
5. Morin, P.J. *Community Ecology*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2011.
6. Levins, R. *Evolution in Changing Environments*; Princeton University Press: Princeton, NJ, USA, 1968.
7. Rose, M.R. *Quantitative Ecological Theory*; Johns Hopkins University Press: Baltimore, MD, USA, 1987.
8. Bascompte, J.; Jordano, P. *Mutualistic Networks*; Princeton University Press: Princeton, NJ, USA, 2013.
9. Memmott, J. The structure of a plant-pollinator food web. *Ecol. Lett.* **1999**, *2*, 276–280.
10. Memmott, J.; Waser, N.M. Integration of alien plants into a native flower-pollinator visitation web. *Proc. R. Soc. Lond. B* **2002**, *269*, 2395–2399.
11. Olesen, J.M.; Jordano, P. Geographic patterns in plant-pollinator mutualistic networks. *Ecology* **2002**, *83*, 2416–2424.
12. Ollerton, J.; Johnson, S.D.; Cranmer, L.; Kellie, S. The pollination ecology of an assemblage of grassland asclepiads in South Africa. *Ann. Bot.* **2003**. *92*, 807–834.
13. Bascompte, J.; Jordano, P.; Melián, C.J.; Olesen, J.M. The nested assembly of plant–animal mutualistic networks. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9383–9387.
14. Jordano, P.; Bascompte, J.; Olesen, J.M. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.* **2003**, *6*, 69–81.
15. Olesen, J.M.; Bascompte, J.; Dupont, Y.L.; Jordano, P. The modularity of pollination networks. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19891–19896.
16. Fortuna, M.A.; Stouffer, D.B.; Olesen, J.M.; Jordano, P.; Mouillot, D.; Krasnov, B.R.; Poulin, R.; Bascompte, J. Nestedness versus modularity in ecological networks: Two sides of the same coin? *J. Anim. Ecol.* **2010**, *79*, 811–817.

17. Staniczenko, P.P.A.; Kopp, J.C.; Allesina, S. The ghost of nestedness in ecological networks. *Nat. Commun.* **2013**, *4*, doi:10.1038/ncomms2422.

18. Jonhson, S.; Domínguez-García, V.; Muñoz, M.A. Factors Determining Nestedness in Complex Networks. *PLoS ONE* **2013**, *8*, e74025.

19. Agrawal, A.A.; Ackerly, D.D.; Adler, F.; Arnold, A.E.; Càceres, C.; Doak, D.F.; Post, E.; Hudson, P.J.; Maron, J.; Mooney, K.A.; *et al*. Filling key gaps in population and community ecology. *Front. Ecol. Environ.* **2007**, *5*, 145–152.

20. Fort, H. Quantitative predictions of pollinators' abundances from qualitative data on their interactions with plants and evidences of emergent neutrality. *Oikos* **2014**, *123*, 1469–1478.

21. MacArthur, R.; Levins, R. The limiting similarity, convergence and divergence of coexisting species. *Am. Nat.* **1967**, *101*, 377–385.

22. May, R.M.; MacArthur, R.H. Niche overlap as a function of environmental variability. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 1109–1113.

23. McLachlan, G.J.; Krishnan, T. *The EM Algorithm and Extensions*; Wiley: Hoboken, NY, USA, 1996.

24. Newman, M.E.J.; Leicht, E.A. Mixture models and exploratory analysis in networks. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9564–9569.

25. Ramasco, J.J.; Mungan, M. Inversion method for content-based networks. *Phys. Rev. E* **2008**, *77*, 036122.

26. Mungan, M.; Ramasco, J.J. Stability of maximum likelihood-based clustering methods: Exploring the backbone of classifications. *J. Stat. Mech.* **2010**, *2010*, P04028.

27. Rezende, E.L.; Lavabre, J.E.; Guimarães, P.R.; Jordano, P.; Bascompte, J. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **2007**, *448*, 925–928.

28. Interaction Web DataBase. Available online: https://www.nceas.ucsb.edu/interactionweb/resources.html (accessed on 12 November 2015).

29. Allesina, S.; Pascual, M. Food web models: A plea for groups. *Ecol. Lett.* **2009**, *12*, 652–662.

30. Fort, H.; Mungan, M. Predicting abundances of plants and pollinators using a simple compartmental mutualistic model. *Proc. R. Soc. Lond. B* **2015**, *282*, doi:10.1098/rspb.2015.0592.

31. Batagelj, V.; Mrvar, A. Pajek—Program for Large Network Analysis. *Connections* **1998**, *21*, 47–57.

32. Vázquez, D.P. Interactions among Introduced Ungulates, Plants, and Pollinators: A Field Study in the Temperate Forest of the Southern Andes. Ph.D. Thesis, University of Tennessee, Knoxville, TN, USA, 2002.

33. Vázquez, D.P.; Simberloff, D. Ecological specialization and susceptibility to disturbance: Conjectures and refutations. *Am. Nat.* **2002**, *159*, 606–623.

34. Santos, G.M.M.; Aguiar, C.M.L.; Mello, M.A.R. Flower-visiting guild associated with the Caatinga flora: Trophic interaction networks formed by social bees and social wasps with plants. *Apidologie* **2010**, *41*, 466–475.

35. Schemske, D.W.; Willson, M.F.; Melampy, M.N.; Miller, L.J.; Verner, L.; Schemske, K.M.; Best, L.B. Flowering Ecology of Some Spring Woodland Herbs. *Ecology* **1978**, *59*, 351–366.

36. Stang, M. The Structure of Flower Visitation Webs: How Morphology and Abundance Affect Interation Patterns between Flowers and Flower Visitors. Ph.D. Thesis, Leiden University, Leiden, The Netherlands, 2007.

37. Fort, H.; Scheffer, M.; van Nes, E. The clumping transition in niche competition: A robust critical phenomenon. *J. Stat. Mech.* **2010**, *2010*, P05005.

38. Jaccard, P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles* **1901**, *37*, 547–579. (In French)

39. Kirkpatrick, S.; Gelatt, C.D., Jr.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.

40. Doreian, P.; Batagelj, V.; Ferligoj, A. *Generalized Blockmodeling*; Cambridge University Press: Cambridge, UK, 2004.

41. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.

42. Chase, J.M.; Leibold, M.A. *Ecological Niches*; University of Chicago Press: Chicago, IL, USA, 2003.

43. Kearns, C.A.; Inouye, D.W.; Waser, N.M. Endangered mutualisms: The conservation of plant-pollinator interactions. *Ann. Rev. Ecol. Syst.* **1998**, *29*, 83–112.

44. Ghazoul, J. Buzziness as usual? Questioning the global pollination crisis. *Trends Ecol. Evol.* **2005**, *20*, 367–373.

45. Biesmeijer, J.C.; Roberts, S.P.M.; Reemer, M.; Ohlemüeller, R.; Edwards, M.; Peeters, T.; Schaffers, A.P.; Potts, S.G.; Kleukers, R.; Thomas, C.D.; *et al.* Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **2006**, *313*, 351–354.

46. Aizen, M.A.; Garibaldi, L.A.; Cunningham, S.A.; Klein, A.M. How much does agriculture depend on pollinators? Lessons from long-term trends in crop production. *Ann. Bot.* **2009**, *103*, 1579–1588.

47. Ranta, E.; Lundberg, H. Resource partitioning in bumblebees: The significance of differences in proboscis length. *Oikos* **1980**, *35*, 298–302.

48. Harder, L.D. Morphology as a predictor of flower choice by bumblebees. *Ecology* **1985**, *66*, 198–210.

49. Herrera, C.M. Pollinator abundance, morphology, and flower visitation rate: Analysis of the quantity component in a plant-pollinator system. *Oecologia* **1989**, *80*, 241–248.

50. Herrera, C.M. Floral traits and plant adaptation to insect pollinators: A devil's advocate approach. In *Floral Biology*; Lloyd, D.G., Barrett, S.C.H., Eds.; Springer: Berlin/Heidelberg, Germany, 1996; pp. 65–87.

51. Stang, M.; Klinkhamer, P.G.L.; Waser, N.M.; Stang, I.; van der Meijden, E. Size-specific interation patterns and size matching in a plant-pollinator interaction web. *Ann. Bot.* **2009**, *103*, 1459–1469.