

# Sampling networks of ecological interactions

Pedro Jordano<sup>\*a</sup>

<sup>a</sup>Integrative Ecology Group, Estación Biológica de Doñana, Consejo  
Superior de Investigaciones Científicas (EBD-CSIC), Avenida  
Americo Vesputio s/n, E-41092 Sevilla, Spain

In review, Functional Ecology, July 19, 2016

## Summary

1. Sampling ecological interactions presents similar challenges, problems, potential biases, and constraints as sampling individuals and species in biodiversity inventories. Robust estimates of the actual number of interactions (links) within diversified ecological networks require adequate sampling effort that needs to be explicitly gauged. Yet we still lack a sampling theory explicitly focusing on ecological interactions.
2. While the complete inventory of interactions is likely impossible, a robust characterization of its main patterns and metrics is probably realistic. We

---

<sup>\*</sup>jordano@ebd.csic.es

must acknowledge that a sizeable fraction of the maximum number of interactions  $I_{max}$  among, say,  $A$  animal species and  $P$  plant species (i.e.,  $I_{max} = AP$ ) is impossible to record due to forbidden links, i.e., life-history restrictions. Thus, the number of observed interactions  $I$  in robustly sampled networks is typically  $I \ll I_{max}$ , resulting in sparse interaction matrices with low connectance.

3. Here I provide a review and outline a conceptual framework for interaction sampling by building an explicit analogue to individuals and species sampling, thus extending diversity-monitoring approaches to the characterization of complex networks of ecological interactions. Contrary to species inventories, a sizable fraction of non-observed pairwise interactions cannot be sampled, due to biological constraints that forbid their occurrence.
4. Reasons for forbidden links are multiple but mainly stem from spatial and temporal uncoupling, size mismatches, and intrinsically low probabilities of interspecific encounter for most potential interactions of partner species. Adequately assessing the completeness of a network of ecological interactions thus needs knowledge of the natural history details embedded, so that forbidden links can be accounted for when addressing sampling effort.
5. Recent implementations of inference methods for unobserved species or for individual-based data can be combined with the assessment of forbidden links. This can help in estimating their relative importance, simply by the difference between the asymptotic estimate of interaction richness *in a robustly-sampled* assemblage and the maximum richness  $I_{max}$  of interactions. This is crucial to assess the rapid and devastating effects of defauna-

driven loss of key ecological interactions and the services they provide and the analogous losses related to interaction gains due to invasive species and biotic homogenization.

## Keywords

complex networks, food webs, frugivory, mutualism, plant-animal interactions, pollination, seed dispersal

## Introduction

Biodiversity sampling is a labour-intensive activity, and sampling is often not sufficient to detect all or even most of the species present in an assemblage. Gotelli & Colwell (2011).

1 Biodiversity species assessment aims at sampling individuals in collections and  
2 determining the number of species represented. Given that, by definition, samples  
3 are incomplete, these collections do not enumerate the species actually present.  
4 The ecological literature dealing with robust estimators of species richness and di-  
5 versity in collections of individuals is immense, and a number of useful approaches  
6 have been used to obtain such estimates (Magurran, 1988; Gotelli & Colwell, 2001;  
7 Colwell *et al.*, 2004; Hortal *et al.*, 2006; Colwell, 2009; Gotelli & Colwell, 2011; Chao  
8 *et al.*, 2014). Recent effort has been also focused at defining essential biodiversity  
9 variables (EBV) (Pereira *et al.*, 2013) that can be sampled and measured repeat-  
10 edly to complement biodiversity estimates. Yet sampling species or taxa-specific

EBVs is just probing a single component of biodiversity; interactions among species are another fundamental component, one that supports the existence, but in some cases also the extinction, of species. For example, the extinction of interactions represents a dramatic loss of biodiversity because it entails the loss of fundamental ecological functions (Valiente-Banuet *et al.*, 2014). This missed component of biodiversity loss, the extinction of ecological interactions, very often accompanies, or even precedes, species disappearance. Interactions among species are a key component of biodiversity and here I aim to show that most problems associated with sampling interactions in natural communities relate to, and are even worse than, problems associated with sampling species diversity. I consider pairwise interactions among species at the habitat level, in the context of alpha diversity and the estimation of local interaction richness from sampling data (Chao *et al.*, 2014). In the first part I provide a succinct overview of previous work addressing sampling issues for ecological interaction networks. In the second part, I discuss specific rationales for sampling the biodiversity of ecological interactions. Finally, I provide a short overview of asymptotic diversity estimates (Gotelli & Colwell, 2001), and a discussion of its application to interaction sampling. Most of the examples come from the analysis of plant-animal interaction networks, yet are applicable to other types of interspecific interactions.

Interactions can be a much better indicator of the richness and diversity of ecosystem functions than a simple list of taxa and their abundances and/or related biodiversity indicator variables (EBVs) (Mommott *et al.*, 2006; Valiente-Banuet *et al.*, 2014). Thus, sampling interactions should be a central issue when identifying and diagnosing ecosystem services (e.g., pollination, seeding by frugivores, etc.). Fortunately, the whole battery of biodiversity-related tools used by ecolo-

gists to sample biodiversity (species, *sensu stricto*) can be extended and applied to the sampling of interactions (see Table 2 in Colwell *et al.*, 2004). Monitoring interactions is a type of biodiversity sampling and is subject to similar methodological shortcomings, especially under-sampling (Jordano, 1987; Jordano *et al.*, 2009; Vázquez *et al.*, 2009; Dorado *et al.*, 2011; Rivera-Hutinel *et al.*, 2012). We are interested in having a complete list of all the pairwise interactions among species (e.g., all the distinct, species-species interactions, or links, among the pollinators and flowering plants) that do actually exist in a given community. Sampling these interactions thus entails exactly the same problems, limitations, constraints, and potential biases as sampling individual organisms and species diversity. As Mao & Colwell (2005) put it, these are the workings of Preston’s demon, the moving “veil line” (Preston, 1948) between the detected and the undetected interactions as sample size increases.

Early efforts to recognize and solve sampling problems in analyses of interactions stem from research on food webs and to determine how undersampling biases food web metrics (Martinez, 1991; Cohen *et al.*, 1993; Wells & O’Hara, 2012, among others). In addition, the myriad of classic natural history studies documenting animal diets, host-pathogen infection records, plant herbivory records, etc., represent efforts to document interactions occurring in nature. All of them share the problem of sampling incompleteness influencing the patterns and metrics reported. Yet, despite the early recognition that incomplete sampling may seriously bias the analysis of ecological networks (Jordano, 1987), only recent studies have explicitly acknowledged it and attempted to determine its influence (Ollerton & Cranmer, 2002; Nielsen & Bascompte, 2007; Vázquez *et al.*, 2009; Gibson *et al.*, 2011; Olesen *et al.*, 2011; Chacoff *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito &

61 Fox, 2014; Bascompte & Jordano, 2014; Vizentin-Bugoni *et al.*, 2014, 2016; Frund  
 62 *et al.*, 2015). The sampling approaches have been extended to predict patterns of  
 63 coextinctions in interaction assemblages (e.g., hosts-parasites) (Colwell *et al.*, 2012).  
 64 Most empirical studies provide no indication of sampling effort, implicitly assuming  
 65 that the reported network patterns and metrics are robust. Yet recent evidences  
 66 point out that number of partner species detected, number of actual links, and  
 67 some aggregate statistics describing network patterns, are prone to sampling bias  
 68 (Nielsen & Bascompte, 2007; Dorado *et al.*, 2011; Olesen *et al.*, 2011; Chacoff *et al.*,  
 69 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox, 2014; Frund *et al.*, 2015). Most of  
 70 these evidences, however, come either from simulation studies (Frund *et al.*, 2015)  
 71 or from relatively species-poor assemblages. Most certainly, sampling limitations  
 72 pervade biodiversity inventories and we might rightly expect that frequent inter-  
 73 actions may be over-represented and rare interactions may be missed entirely in  
 74 studies of mega-diverse assemblages (Bascompte & Jordano, 2014); but, to what  
 75 extent?

## 76 Sampling interactions: methods

77 When we sample interactions in the field we record the presence of two species that  
 78 interact in some way. For example, Snow and Snow(1988) recorded an interaction  
 79 whenever they saw a bird “touching” a fruit on a plant. We observe and record  
 80 feeding observations, visitation, occupancy, presence in pollen loads or in fecal  
 81 samples, etc., of *individual* animals or plants and accumulate pairwise interactions,  
 82 i.e., lists of species partners and the frequencies with which we observe them. We  
 83 assume that the matrix (species numbers) is predefined (i.e., all species interacting

are well documented).

Most types of ecological interactions can be illustrated with bipartite graphs, with two or more distinct groups of interacting partners (Bascompte & Jordano, 2014); for illustration purposes I'll focus more specifically on plant-animal interactions. Sampling interactions requires filling the cells of an interaction matrix with data. The matrix,  $\Delta = AP$  (the adjacency matrix for the graph representation of the network), is a 2D inventory of the interactions among, say,  $A$  animal species (rows) and  $P$  plant species (columns) (Jordano, 1987; Bascompte & Jordano, 2014). The matrix entries illustrate the values of the pairwise interactions visualized in the  $\Delta$  matrix, and can be 0 or 1, for presence-absence of a given pairwise interaction, or take a quantitative weight  $w_{ji}$  to represent the interaction intensity or unidirectional effect of species  $j$  on species  $i$  (Bascompte & Jordano, 2014; Vázquez *et al.*, 2015). The outcomes of most ecological interactions are dependent on frequency of encounters (e.g., visit rate of pollinators, number of records of ant defenders, frequency of seeds in fecal samples). Thus, a frequently used proxy for interaction intensities  $w_{ji}$  is just how frequent new interspecific encounters are, whether or not appropriately weighted to estimate interaction effectiveness (Vázquez *et al.*, 2005).

We need to define two basic steps in the sampling of interactions: 1) which type of interactions we sample; and 2) which type of record we get to document the existence of an interaction. In step #1 we need to take into account whether we are sampling the whole community of interactor species (all the animals, all the plants) or just a subset of them, i.e., a sub matrix  $\Delta_{m,n}$  of  $m < A$  animal species and  $n < P$  plant species of the adjacency matrix  $\Delta_{AP}$  (i.e., the matrix representation of interactions among the partner species). Subsets can be: a) all

109 the potential plants interacting with a subset of the animals (Fig. 1a); b) all the  
 110 potential animal species interacting with a subset of the plant species (Fig. 1b);  
 111 c) a subset of all the potential animal species interacting with a subset of all the  
 112 plant species (Fig. 1c). While some discussion has considered how to establish  
 113 the limits of what represents a network (Strogatz, 2001) (in analogy to discussion  
 114 on food-web limits; Cohen, 1978), it must be noted that situations a-c in Fig. 1  
 115 do not represent complete interaction networks. Subnet sampling is generalized  
 116 in studies of biological networks (e.g., protein interactions, gene regulation), yet  
 117 it is important to recognize that most properties of subnetworks (even random  
 118 subsamples) do not represent properties of whole networks (Stumpf *et al.*, 2005).

119 In step #2 above we face the problem of the type of record we take to sample  
 120 interactions. This is important because it defines whether we approach the problem  
 121 of filling up the interaction matrix in a “zoo-centric” way or in a “phyto-centric”  
 122 way. Zoo-centric studies directly sample animal activity and document the plants  
 123 ‘touched’ by the animal. For example, analysis of pollen samples recovered from the  
 124 body of pollinators, analysis of fecal samples of frugivores, radio-tracking data, etc.  
 125 Phyto-centric studies take samples of focal individual plant species and document  
 126 which animals ‘arrive’ or ‘touch’ the plants. Examples include focal watches of  
 127 fruiting or flowering plants to record visitation by animals, raising insect herbivores  
 128 from seed samples, identifying herbivory marks in samples of leaves, etc.

129 Most recent analyses of plant-animal interaction networks are phyto-centric;  
 130 just 3.5% of available plant-pollinator ( $N = 58$ ) or 36.6% plant-frugivore ( $N = 22$ )  
 131 interaction datasets are zoo-centric (see Schleuning *et al.*, 2012). Moreover, most  
 132 available datasets on host-parasite (parasitoid) or plant-herbivore interactions are  
 133 “host-centric” or phyto-centric (e.g., Thébault & Fontaine, 2010; Morris *et al.*,



2013; Eklöf *et al.*, 2013). This may be related to a variety of causes, like preferred methodologies by researchers working with a particular group or system, logistic limitations, or inherent taxonomic focus of the research questions. A likely result of phyto-centric sampling would be adjacency matrices with large  $A : P$  ratios. In contrast, zoo-centric samplings might be prone to detect plants from outside the habitat, complicating the definition of network boundaries. In any case we don't have a clear view of the potential biases that taxa-focused sampling may generate in observed network patterns, for example by generating consistently asymmetric interaction matrices (Dormann *et al.*, 2009).

Reasonably complete analyses of interaction networks can be obtained when combining both phyto-centric and zoo-centric sampling. For example, Bosch *et al.* (2009) showed that the addition of pollen load data on top of focal-plant sampling of pollinators unveiled a significant number of interactions, resulting in important network structural changes. Olesen *et al.* (2011) identified pollen loads on sampled insects and added the new links to an observation-based visitation matrix, with an extra 5% of links representing the estimated number of missing links in the pollination network. The overlap between observational and pollen-load recorded links was only 33%, underscoring the value of combining methodological approaches. Zoo-centric sampling has recently been extended with the use of DNA-barcoding, for example with plant-herbivore (Jurado-Rivera *et al.*, 2009), host-parasitoid (Wirta *et al.*, 2014), and plant-frugivore interactions (González-Varo *et al.*, 2014). For mutualistic networks we would expect that zoo-centric sampling could help unveiling interactions of the animals with rare plant species or for relatively common plants species which are difficult to sample by direct observation. Future methodological work may provide significant advances showing

159 how mixing different sampling strategies strengthens the completeness of network  
 160 data. These mixed strategies may combine, for instance, timed watches at focal  
 161 plants, spot censuses along walked transects, pollen load or seed contents analy-  
 162 ses, monitoring with camera traps, and DNA barcoding records. However, there  
 163 are no tested protocols and/or sampling designs for ecological interaction studies  
 164 to suggest an optimum combination of approaches. Ideally, pilot studies would  
 165 provide adequate information for each specific study setting.

## 166 Sampling interactions: rationale

167 The number of distinct pairwise interactions that we can record in a landscape (an  
 168 area of relatively homogeneous vegetation) is equivalent to the number of distinct  
 169 classes in which we can classify the recorded encounters among *individuals* of  
 170 two different species. Yet, individual-based interaction networks have been only  
 171 recently studied (Dupont *et al.*, 2011; Wells & O’Hara, 2012). The most usual  
 172 approach has been to pool individual-based interaction data into species-based  
 173 summaries, an approach that ignores the fact that only a fraction of individuals  
 174 may actually interact given a per capita interaction effect (Wells & O’Hara, 2012).  
 175 Wells & O’Hara (2012) illustrate the pros and cons of the approach. We walk in  
 176 the forest and see a blackbird  $Tm$  picking an ivy  $Hh$  fruit and ingesting it: we  
 177 have a record for  $Tm - Hh$  interaction. We keep advancing and record again a  
 178 blackbird feeding on hawthorn  $Cm$  fruits so we record a  $Tm - Cm$  interaction;  
 179 as we advance we encounter another ivy plant and record a blackcap swallowing a  
 180 fruit so we now have a new  $Sa - Hh$  interaction, and so on. At the end we have  
 181 a series of classes (e.g.,  $Sa - Hh$ ,  $Tm - Hh$ ,  $Tm - Cm$ , etc.), along with their

observed frequencies.

We get a vector  $c = [c_1 \dots c_n]'$  where  $c_j$  is the number of classes represented  $j$  times in our sampling:  $c_1$  is the number of singletons (interactions recorded once),  $c_2$  is the number of twin pairs (interactions with just two records),  $c_3$  the number of triplets, etc. The problem thus turns to be estimating the number of distinct classes  $C$  from the vector of  $c_j$  values and the frequency of unobserved interactions (see “The real missing links” below).

More specifically, we usually obtain a type of reference sample (Chao *et al.*, 2014) for interactions: a series of repeated samples (e.g., observation days, 1h watches, etc.) with quantitative information, i.e., recording the number of instances of each interaction type on each day. This replicated abundance data, can be treated in three ways: 1) Abundance data within replicates: the counts of interactions, separately for each day; 2) Pooled abundance data: the counts of interactions, summed over all days (the most usual approach); and 3) Replicated incidence data: the number of days on which we recorded each interaction. Assuming a reasonable number of replicates, replicated incidence data is considered to be the most robust statistically, as it takes account of heterogeneity among days (Colwell *et al.*, 2004, 2012; Chao *et al.*, 2014). Thus, both presence-absence and weighted information on interactions can be accommodated for this purpose.

## The species assemblage

When we consider an observed and recorded sample of interactions on a particular assemblage of  $A_{obs}$  and  $P_{obs}$  species (or a set of replicated samples) as a reference sample (Chao *et al.*, 2014) we may have three sources of undersampling error.

205 These sources are ignored if we treat the reference sample as a true representation  
 206 of the interactions in a well-defined assemblage: 1) some animal species are actually  
 207 present but not observed (zero abundance or incidence in the interactions in the  
 208 reference sample),  $A_0$ ; 2) some plant species are actually present but not observed  
 209 (zero abundance or incidence in the interactions in the reference sample),  $P_0$ ; 3)  
 210 some unobserved links (the zeroes in the adjacency matrix,  $UL$ ) may actually  
 211 occur but not recorded. Thus a first problem is determining if  $A_{obs}$  and  $P_{obs}$  truly  
 212 represent the actual species richness interacting in the assemblage. To this end  
 213 we might use the replicated reference samples to estimate the true number of  
 214 interacting animal  $A_{est}$  and plant  $P_{est}$  species as in traditional diversity estimation  
 215 analysis (Chao *et al.*, 2014). If there are no uniques (species seen on only one day),  
 216 then  $A_0$  and  $P_0$  will be zero (based on the Chao2 formula), and we have  $A_{obs}$  and  
 217  $P_{obs}$  as robust estimates of the actual species richness of the assemblage. If  $A_0$   
 218 and  $P_0$  are not zero they estimate the minimum number of undetected animal and  
 219 plant species that can be expected with a sufficiently large number of replicates,  
 220 taken from the same assemblage/locality by the same methods in the same time  
 221 period. We can use extrapolation methods (Colwell *et al.*, 2012) to estimate how  
 222 many additional replicate surveys it would take to reach a specified proportion  $g$   
 223 of  $A_{est}$  and  $P_{est}$ .

## 224 The interactions

225 We are then faced with assessing the sampling of interactions  $I$ . Table 1 summa-  
 226 rizes the main components and targets for estimation of interaction richness. In  
 227 contrast with traditional species diversity estimates, sampling networks has the

paradox that despite the potentially interacting species being present in the sampled assemblage (i.e., included in the  $A_{obs}$  and  $P_{obs}$  species lists), some of their pairwise interactions are impossible to record. The reason is forbidden links. Independently of whether we sample full communities or subset communities we face a problem: some of the interactions that we can visualize in the empty adjacency matrix  $\Delta$  will simply not occur. With a total of  $A_{obs}P_{obs}$  “potential” interactions (eventually augmented to  $A_{est}P_{est}$  in case we have undetected species), a fraction of them are impossible to record, because they are forbidden (Jordano *et al.*, 2003; Olesen *et al.*, 2011).

Our goal is to estimate the true number of non-null  $AP$  interactions, including interactions that actually occur but have not been observed ( $I_0$ ) from the replicated incidence frequencies of interaction types:  $I_{est} = I_{obs} + I_0$ . Note that  $I_0$  estimates the minimum number of undetected plant-animal interactions that can be expected with a sufficiently large number of replicates, taken from the same assemblage/locality by the same methods in the same time period. Therefore we have two types of non-observed links:  $UL^*$  and  $UL$ , corresponding to the real assemblage species richness and to the observed assemblage species richness, respectively (Table 1).

Forbidden links are non-occurrences of pairwise interactions that can be accounted for by biological constraints, such as spatio-temporal uncoupling (Jordano, 1987), size or reward mismatching, foraging constraints (e.g., accessibility) (Moré *et al.*, 2012), and physiological-biochemical constraints (Jordano, 1987). We still have very little information about the frequency of forbidden links in natural communities (Jordano *et al.*, 2003; Stang *et al.*, 2009; Vázquez *et al.*, 2009; Olesen *et al.*, 2011; Ibanez, 2012; Maruyama *et al.*, 2014; Vizentin-Bugoni *et al.*, 2014)

(Table 1). Forbidden links are thus represented as structural zeroes in the interaction matrix, i.e., matrix cells that cannot get a non-zero value. Therefore, we need to account for the frequency of these structural zeros in our matrix before proceeding.

Our main problem then turns to estimate the number of true missed links, i.e., those that can't be accounted for by biological constraints and that might suggest undersampling. Thus, the sampling of interactions in nature, as the sampling of species, is a cumulative process. In our analysis, we are not re-sampling individuals, but interactions, so we built interaction-based accumulation curves. If an interaction-based curve suggests a robust sampling, it does mean that no new interactions are likely to be recorded, irrespectively of the species, as it is a whole-network sampling approach. We add new, distinct, interactions recorded as we increase sampling effort (Fig. 2). We can obtain an Interaction Accumulation Curve (*IAC*) analogous to a Species Curve (*SAC*) (see Supporting Information in the online data availability repository): the observed number of distinct pairwise interactions in a survey or collection as a function of the accumulated number of observations or samples (Colwell, 2009).

## Empirical data on Forbidden Links

Adjacency matrices are frequently sparse, i.e., they are densely populated with zeroes, with a fraction of them being structural (unobservable interactions) (Bascompte & Jordano, 2014). Thus, it would be a serious interpretation error to attribute the sparseness of adjacency matrices for bipartite networks to just the result of undersampling. The actual typology of link types in ecological interac-

tion networks is thus more complex than just the two categories of observed and  
 unobserved interactions (Table 1). Unobserved interactions are represented by  
 zeroes and belong to two categories. Missing interactions may actually exist but  
 require additional sampling or a variety of methods to be observed. Forbidden  
 links, on the other hand, arise due to biological constraints limiting interactions  
 and remain unobservable in nature, irrespectively of sampling effort (Table 1).  
 Forbidden links  $FL$  may actually account for a relatively large fraction of unob-  
 served interactions  $UL$  when sampling taxonomically-restricted subnetworks (e.g.,  
 plant-hummingbird pollination networks) (Table 1). Phenological uncoupling is  
 also prevalent in most networks, and may add up to explain ca. 25-40% of the for-  
 bidden links, especially in highly seasonal habitats, and up to 20% when estimated  
 relative to the total number of unobserved interactions (Table 2). In any case, we  
 might expect that a fraction of the missing links  $ML$  would be eventually explained  
 by further biological reasons, depending on the knowledge of natural details of the  
 particular systems. Our goal as naturalists would be to reduce the fraction of  $UL$   
 which remain as missing links; to this end we might search for additional biological  
 constraints or increase sampling effort. For instance, habitat use patterns by hum-  
 mingbirds in the Arima Valley network (Table 2; Snow & Snow, 1972) impose a  
 marked pattern of microhabitat mismatches causing up to 44.5% of the forbidden  
 links. A myriad of biological causes beyond those included as  $FL$  in Table 1 may  
 contribute explanations for  $UL$ : limits of color perception, presence of secondary  
 metabolites in fruit pulp and leaves, toxins and combinations of monosaccharides  
 in nectar, etc. For example, aside from  $FL$ , some pairwise interactions may sim-  
 ply have an asymptotically-zero probability of interspecific encounter between the  
 partner species, if they are very rare. However, it is surprising that just the limited

set of forbidden link types considered in Table 1 explain between 24.6-77.2% of the unobserved links. Notably, the Arima Valley, Santa Virgínia, and Hato Ratón networks have  $> 60\%$  of the unobserved links explained, which might be related to the fact that they are subnetworks (Arima Valley, Santa Virgínia) or relatively small networks (Hato Ratón). All this means that empirical networks may have sizable fractions of structural zeroes. Ignoring this biological fact may contribute to wrongly inferring undersampling of interactions in real-world assemblages.

To sum up, two elements of inference are required in the analysis of unobserved interactions in ecological interaction networks: first, detailed natural history information on the participant species that allows the inference of biological constraints imposing forbidden links, so that structural zeroes can be identified in the adjacency matrix. Second, a critical analysis of sampling robustness and a robust estimate of the actual fraction of missing links,  $M$ , resulting in a robust estimate of  $I$ . In the next sections we explore these elements of inference, using *IACs* as analogs to *SACs* to assess the robustness of interaction sampling.

## Assessing sampling effort when recording interactions: asymptotic diversity estimates

A plot of the cumulative number of species recorded,  $S_n$ , as a function of some measure of sampling effort (say,  $n$  samples taken) yields the species accumulation curve (SAC) or collector's curve (Colwell & Coddington, 1994). Similarly, interaction accumulation curves (IAC), analogous to SACs (Gotelli & Colwell, 2001; Hortal *et al.*, 2006; Chao, 2005; Colwell, 2013), can be used to assess the robust-



ness of interactions sampling for plant-animal community datasets (Jordano, 1987; Jordano *et al.*, 2009; Olesen *et al.*, 2011).

The basic method to estimate sampling effort and explicitly show the analogues with rarefaction analysis in biodiversity research is to vectorize the interaction matrix  $AP$  so that we get a vector of all the potential pairwise interactions ( $I_{max}$ , Table 1) that can occur in the observed assemblage with  $A_{obs}$  animal species and  $P_{obs}$  plant species. The new “species” we aim to sample are the pairwise interactions (Table 3), as previously discussed. In general, if we have  $A = 1...i$ , animal species and  $P = 1...j$  plant species (assuming a complete list of species in the assemblage), we’ll have a vector of “new” species to sample:  $A_1P_1, A_1P_2, ...A_2P_1, A_2P_2, ...A_iP_j$ . We can represent the successive samples where we can potentially get records of these interactions in a matrix with the vectorized interaction matrix and columns representing the successive samples we take (Table 3). This is simply a vectorized version of the interaction matrix. This is analogous to a biodiversity sampling matrix with species as rows and sampling units (e.g., quadrats) as columns (Jordano *et al.*, 2009). The package *EstimateS* (Colwell, 2013) includes a complete set of functions for estimating the mean IAC and its unconditional standard deviation from random permutations of the data, or subsampling without replacement (Gotelli & Colwell, 2001); it further reports asymptotic estimators for the expected number of distinct pairwise interactions included in a given reference sample of interaction records (see also the `specaccum` function in library `vegan` of the R Package)(R Development Core Team, 2010; Jordano *et al.*, 2009; Olesen *et al.*, 2011). In particular, we may take advantage of replicated incidence data, as it takes account of heterogeneity among samples (days, censuses, etc.; R.K Colwell, pers. comm.) (see also Colwell *et al.*, 2004, 2012; Chao *et al.*, 2014). Future

theoretical work will be needed to formally assess the similarities and differences between the species vs. interactions sampling approaches and developing biologically meaningful null models of expected interaction richness with added sampling effort.

Diversity-accumulation analysis (Magurran, 1988; Hortal *et al.*, 2006) comes up immediately with this type of dataset. This procedure plots the accumulation curve for the expected number of distinct pairwise interactions recorded with increasing sampling effort (Jordano *et al.*, 2009; Olesen *et al.*, 2011). Asymptotic estimates of interaction richness and its associated standard errors and confidence intervals can thus be obtained (Hortal *et al.*, 2006) (see Table 4 and Supplementary Online Material). The characteristic feature of interaction datasets is that, due to forbidden links, a number of pairwise interactions among the  $I_{max}$  number specified in the  $\Delta$  adjacency matrix cannot be recorded, irrespective of sampling effort.

We may expect undersampling specially in moderate to large sized networks with multiple modules (i.e., species subsets requiring different sampling strategies) (Jordano, 1987; Olesen *et al.*, 2011; Chacoff *et al.*, 2012); adequate sampling may be feasible when interaction subwebs are studied (Olesen *et al.*, 2011; Vizentin-Bugoni *et al.*, 2014), typically with more homogeneous subsets of species (e.g., bumblebee-pollinated flowers).

Mixture models incorporating detectabilities have been proposed to effectively account for rare species (Mao & Colwell, 2005). In an analogous line, mixture models could be extended to samples of pairwise interactions, also with specific detectability values. These detection rate/odds could be variable among groups of interactions, depending on their specific detectability. For example, detectability of

373 flower-pollinator interactions involving bumblebees could have a higher detectabil-  
 374 ity than flower-pollinator pairwise interactions involving, say, nitidulid beetles.  
 375 These more homogeneous groupings of pairwise interactions within a network de-  
 376 fine modules (Bascompte & Jordano, 2014), so we might expect that interactions  
 377 of a given module (e.g., plants and their hummingbird pollinators; Fig. 1a) may  
 378 share similar detectability values, in an analogous way to species groups receiving  
 379 homogeneous detectability values in mixture models (Mao & Colwell, 2005). In  
 380 its simplest form, this would result in a sample with multiple pairwise interactions  
 381 detected, in which the number of interaction events recorded for each distinct  
 382 interaction found in the sample is recorded (i.e., a column vector in Table 3, cor-  
 383 responding to, say, a sampling day). The number of interactions recorded for the  
 384  $i_{th}$  pairwise interaction (i.e.,  $A_i P_j$  in Table 3),  $Y_i$  could be treated as a Poisson  
 385 random variable with a mean parameter  $\lambda_i$ , its detection rate. Mixture models  
 386 (Mao & Colwell, 2005) include estimates for abundance-based data (their analogs  
 387 in interaction sampling would be weighted data), where  $Y_i$  is a Poisson random  
 388 variable with detection rate  $\lambda_i$ . This is combined with the incidence-based model,  
 389 where  $Y_i$  is a binomial random variable (their analogous in interaction sampling  
 390 would be presence/absence records of interactions) with detection odds  $\lambda_i$ . Let  
 391  $T$  be the number of samples in an incidence-based data set. A Poisson/binomial  
 392 density can be written as (Mao & Colwell, 2005):

$$g(y; \lambda) = \begin{cases} \frac{\lambda^y}{y!e^\lambda} & [1] \\ \binom{T}{y} \frac{\lambda^y}{(1+\lambda)^T} & [2] \end{cases}$$

393 where [1] corresponds to a weighted network, and [2] to a qualitative network.

394 The detection rates  $\lambda_i$  depend on the relative abundances  $\phi_i$  of the interactions,  
 395 the probability of a pairwise interaction being detected when it is present, and the  
 396 sample size (the number of interactions recorded), which, in turn, is a function  
 397 of the sampling effort. Unfortunately, no specific sampling model has been devel-  
 398 oped along these lines for species interactions and their characteristic features. For  
 399 example, a complication factor might be that interaction abundances,  $\phi_i$ , in real  
 400 assemblages are a function of the abundances of interacting species that determine  
 401 interspecific encounter rates; yet they also depend on biological factors that ulti-  
 402 mately determine if the interaction occurs when the partner species are present.  
 403 For example,  $\lambda_i$  should be set to zero for all  $FL$ . In its simplest form,  $\phi_i$  could be  
 404 estimated from just the product of partner species abundances, an approach re-  
 405 cently used as a null model to assess the role of biological constraints in generating  
 406 forbidden links and explaining interaction patterns (Vizentin-Bugoni *et al.*, 2014).  
 407 Yet more complex models (e.g., Wells & O'hara 2012; ) should incorporate not  
 408 only interspecific encounter probabilities, but also interaction detectabilities, phe-  
 409 notypic matching and incidence of forbidden links. Mixture models are certainly  
 410 complex and for most situations of evaluating sampling effort better alternatives  
 411 include the simpler incidence-based rarefaction and extrapolation (Colwell *et al.*,  
 412 2012; Chao *et al.*, 2014).

## 413 The *real* missing links

414 Given that a fraction of unobserved interactions can be accounted for by forbidden  
 415 links, what about the remaining missing interactions? We have already discussed  
 416 that some of these could still be related to unaccounted constraints, and still others

would be certainly attributable to insufficient sampling. Would this always be the case? A crucial ecological aspect limiting interactions within multispecific assemblages of distinct taxonomic relatedness (Fig. 2) is the probability of interspecific encounter, i.e., the probability that two individuals of the partner species actually encounter each other in nature.

Given log-normally distributed abundances of the two species groups, the expected probabilities of interspecific encounter ( $PIE$ ) would be simply the product of the two lognormal distributions. Thus, we might expect that for very low  $PIE$  values, pairwise interactions would be either extremely difficult to sample, or simply do not occur in nature. Consider the Nava de las Correhuelas interaction web (NCH, Table 2, 4), with  $A = 36$ ,  $P = 25$ ,  $I = 181$ , and almost half of the unobserved interactions not accounted for by forbidden links, thus  $M = 53.1\%$  (Jordano *et al.*, 2009). A sizable fraction of these possible but missing links would be simply not occurring in nature, most likely due to extremely low  $PIE$ , in fact asymptotically zero. Given the vectorized list of pairwise interactions for NCH, I computed the  $PIE$  values for each one by multiplying element-wise the two species abundance distributions. The  $PIE_{max} = 0.0597$ , being a neutral estimate, based on the assumption that interactions occur in proportion to the species-specific local abundances. With  $PIE_{median} < 1.4 \cdot 10^{-4}$  we may safely expect (note the quantile estimate  $Q_{75\%} = 3.27 \cdot 10^{-4}$ ) that a sizable fraction of these missing interactions may not occur according to this neutral expectation (Jordano, 1987; Olesen *et al.*, 2011) (neutral forbidden links, *sensu* Canard *et al.*, 2012).

When we consider the vectorized interaction matrix, enumerating all pairwise interactions for the  $AP$  combinations, the expected probabilities of finding a given interaction can be estimated with a Good-Turing approximation (Good, 1953).

442 The technique, developed by Alan Turing and I.J. Good with applications to lin-  
 443 guistics and word analysis (Gale & Sampson, 1995) has been recently extended in  
 444 novel ways for ecological analyses (Chao *et al.*, 2015). In our present context it  
 445 estimates the probability of recording an interaction of a hitherto unseen pair of  
 446 partners, given a set of past records of interactions between other species pairs.  
 447 Let a sample of  $N$  interactions so that  $n_r$  distinct pairwise interactions have ex-  
 448 actly  $r$  records. All Good-Turing estimators obtain the underlying frequencies of  
 449 events as:

$$P(X) = \frac{(N_X + 1)}{T} \left(1 - \frac{E(1)}{T}\right) \quad (1)$$

450 where  $X$  is the pairwise interaction,  $N_X$  is the number of times interaction  $X$   
 451 is recorded,  $T$  is the sample size (number of distinct interactions recorded) and  
 452  $E(1)$  is an estimate of how many different interactions were recorded exactly once.  
 453 Strictly speaking Equation (1) gives the probability that the next interaction type  
 454 recorded will be  $X$ , after sampling a given assemblage of interacting species. In  
 455 other words, we scale down the maximum-likelihood estimator  $\frac{n}{T}$  by a factor of  
 456  $\frac{1-E(1)}{T}$ . This reduces all the probabilities for interactions we have recorded, and  
 457 makes room for interactions we haven't seen. If we sum over the interactions we  
 458 have seen, then the sum of  $P(X)$  is  $1 - \frac{1-E(1)}{T}$ . Because probabilities sum to one,  
 459 we have the left-over probability of  $P_{new} = \frac{E(1)}{T}$  of seeing something new, where  
 460 new means that we sample a new pairwise interaction.

## Discussion

Recent work has inferred that most data available for interaction networks are incomplete due to undersampling, resulting in a variety of biased parameters and network patterns (Chacoff *et al.*, 2012). It is important to note, however, that in practice, most surveyed networks to date have been subnets of much larger networks. This is also true for protein interaction, gene regulation, and metabolic networks, where only a subset of the molecular entities in a cell have been sampled (Stumpf *et al.*, 2005). Despite recent attempts to document whole ecosystem meta-networks (Pocock *et al.*, 2012), it is likely that most ecological interaction networks will illustrate just major ecosystem compartments. Due to their high generalization, high temporal and spatial turnover, and high complexity of association patterns, adequate sampling of ecological interaction networks is challenging and requires extremely large sampling effort. Undersampling of ecological networks may originate from the analysis of assemblage subsets (e.g., taxonomically or functionally defined), and/or from logistically-limited sampling effort. It is extremely hard to robustly sample the set of biotic interactions even for relatively simple, species-poor assemblages; thus, we need to assess how robust is the characterization of the adjacency matrix  $\Delta$ . Concluding that an ecological network dataset is undersampled just by its sparseness would be unrealistic. The reason stems from a biological fact: a sizeable fraction of the maximum, potential links that can be recorded among two distinct sets of species is simply unobservable, irrespective of sampling effort (Jordano, 1987). In addition, sampling effort needs to be explicitly gauged because of its potential influence on parameter estimates for the network.

Missing links are a characteristic feature of all plant-animal interaction networks, and likely pervade other ecological interactions. Important natural history details explain a fraction of them, resulting in unrealizable interactions (i.e., forbidden interactions) that define structural zeroes in the interaction matrices and contribute to their extreme sparseness. Sampling interactions is a way to monitor biodiversity beyond the simple enumeration of component species and to develop efficient and robust inventories of functional interactions. Yet no sampling theory for interactions is available. Focusing just on the realized interactions or treating missing interactions as the expected unique result of sampling bias would miss important components to understand how all sorts of interactions coevolve within complex webs of interdependence among species.

Contrary to species inventories, a sizable fraction of non-observed pairwise interactions cannot be sampled, due to biological constraints that forbid their occurrence. Moreover, recent implementations of inference methods for unobserved species (Chao *et al.*, 2015) or for individual-based data (Wells & O'Hara, 2012) can be combined with the forbidden link approach. They do not account either for the existence of these ecological constraints, but can help in estimating their relative importance, simply by the difference between the asymptotic estimate of interaction richness *in a robustly-sampled* assemblage and the maximum richness  $I_{max}$  of interactions.

Ecological interactions provide the wireframe supporting the lives of species, and they also embed crucial ecosystem functions which are fundamental for supporting the Earth system. We still have a limited knowledge of the biodiversity of ecological interactions, and they are being lost (extinct) at a very fast pace, frequently preceding species extinctions (Valiente-Banuet *et al.*, 2014). We ur-



510 gently need robust techniques to assess the completeness of ecological interactions  
511 networks because this knowledge will allow the identification of the minimal com-  
512 ponents of their ecological complexity that need to be restored to rebuild functional  
513 ecosystems after perturbations.

## 514 Acknowledgements

515 I am indebted to Jens M. Olesen, Alfredo Valido, Jordi Bascompte, Thomas  
516 Lewinshon, John N. Thompson, Nick Gotelli, Carsten Dormann, and Paulo R.  
517 Guimaraes Jr. for useful and thoughtful discussion at different stages of this  
518 manuscript. Jeferson Vizentin-Bugoni kindly helped with the Sta Virginia data.  
519 Jens M. Olesen kindly made available the Grundvad dataset; together with Robert  
520 K. Colwell, Néstor Pérez-Méndez, JuanPe González-Varo, and Paco Rodríguez pro-  
521 vided most useful comments to a final version of the ms. Robert Colwell shared  
522 a number of crucial suggestions that clarified my vision of sampling ecological in-  
523 teractions, and the final manuscript was greatly improved with comments from  
524 three anonymous reviewers. The study was supported by a Junta de Andalucía  
525 Excellence Grant (RNM-5731), as well as a Severo Ochoa Excellence Award from  
526 the Ministerio de Economía y Competitividad (SEV-2012-0262). The Agencia  
527 de Medio Ambiente, Junta de Andalucía, provided generous facilities that made  
528 possible my long-term field work in different natural parks.

## Data accessibility

This review does not use new raw data, but includes some re-analyses of previously published material. All the original data supporting the paper, R code, supplementary figures, and summaries of analytical protocols is available at the author's GitHub repository ([https://github.com/pedroj/MS\\_Network-Sampling](https://github.com/pedroj/MS_Network-Sampling)), with DOI: 10.5281/zenodo.29437.

## References

- Bartomeus, I., D., G., J., T., M., A., I., D. & M., B.V. (????) .
- Bascompte, J. & Jordano, P. (2014) *Mutualistic networks*. Monographs in Population Biology, No. 53, Princeton University Press, Princeton, NJ.
- Bosch, J., Martín González, A.M., Rodrigo, A. & Navarro, D. (2009) Plant-pollinator networks: adding the pollinator's perspective. *Ecology Letters* **12**, 409–419.
- Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Moullot, D. (2012) Emergence of structural patterns in neutral trophic networks. *PLoS ONE* **7**, e38295.
- Chacoff, N.P., Vázquez, D.P., Lomáscolo, S.B., Stevani, E.L., Dorado, J. & Padrón, B. (2012) Evaluating sampling completeness in a desert plant-pollinator network. *Journal of Animal Ecology* **81**, 190–200.
- Chao, A. (2005) Species richness estimation. *Encyclopedia of Statistical Sciences*, pp. 7909–7916, Oxford University Press, New York, USA.

- 550 Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Elli-  
551 son, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework  
552 for sampling and estimation in species diversity studies. *Ecological Monographs*  
553 **84**, 45–67.
- 554 Chao, A., Hsieh, T.C., Chazdon, R.L., Colwell, R.K. & Gotelli, N.J. (2015) Un-  
555 veiling the species-rank abundance distribution by generalizing the Good-Turing  
556 sample coverage theory. *Ecology* **96**, 1189–1201.
- 557 Cohen, J.E. (1978) *Food webs and niche space*. Princeton University Press, Prince-  
558 ton, New Jersey, US.
- 559 Cohen, J.E., Beaver, R.A., Cousins, S.H., DeAngelis, D.L., Goldwasser, L., Heong,  
560 K.L., Holt, R.D., Kohn, A.J., Lawton, J.H., Martinez, N., O'Malley, R., Page,  
561 L.M., Patten, B.C., Pimm, S.L., Polis, G., Rejmanek, M., Schoener, T.W.,  
562 Schoenly, K., Sprules, W.G., Teal, J.M., Ulanowicz, R., Warren, P.H., Wilbur,  
563 H.M. & Yodzis, P. (1993) Improving Food Webs. *Ecology* **74**, 252–258.
- 564 Colwell, R. & Coddington, J. (1994) Estimating terrestrial biodiversity through ex-  
565 trapolation. *Philosophical Transactions Of The Royal Society Of London Series*  
566 *B-Biological Sciences* **345**, 101–118.
- 567 Colwell, R.K. (2009) Biodiversity: concepts, patterns, and measurement. *The*  
568 *Princeton Guide to Ecology* (ed. S.A. Levin), pp. 257–263, Princeton Univer-  
569 sity Press, Princeton.
- 570 Colwell, R.K. (2013) EstimateS: Biodiversity Estimation. - pp. 1–33.
- 571 Colwell, R.K., Dunn, R.R. & Harris, N.C. (2012) Coextinction and persistence of

- 572 dependent species in a changing world. *Annual Review of Ecology Evolution and*  
573 *Systematics* **43**, 183–203.
- 574 Colwell, R.K., Mao, C.X. & Chang, J. (2004) Interpolating, extrapolating, and  
575 comparing incidence-based species accumulation curves. *Ecology* **85**, 2717–2727.
- 576 Dorado, J., Vázquez, D.P., Stevani, E.L. & Chacoff, N.P. (2011) Rareness and  
577 specialization in plant-pollinator networks. *Ecology* **92**, 19–25.
- 578 Dormann, C.F., Frund, J., Bluthgen, N. & Gruber, B. (2009) Indices, graphs and  
579 null models: Analyzing bipartite ecological networks. *Open Ecology Journal* **2**,  
580 7–24.
- 581 Dupont, Y.L., Trøjelsgaard, K., Hagen, M., Henriksen, M.V., Olesen, J.M., Ped-  
582 ersen, N.M.E. & Kissling, W.D. (2014) Spatial structure of an individual-based  
583 plant-pollinator network. *Oikos* **123**, 1301–1310.
- 584 Dupont, Y.L., Trøjelsgaard, K. & Olesen, J.M. (2011) Scaling down from species  
585 to individuals: a flower–visitation network between individual honeybees and  
586 thistle plants. *Oikos* **120**, 170–177.
- 587 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P.,  
588 Dalsgaard, B., de Sassi, C., Galetti, M., Guimaraes, P.R., Lomáscolo, S.B.,  
589 Martín González, A.M., Pizo, M.A., Rader, R., Rodrigo, A., Tylianakis, J.M.,  
590 Vázquez, D.P. & Allesina, S. (2013) The dimensionality of ecological networks.  
591 *Ecology Letters* **16**, 577–583.
- 592 Frund, J., McCann, K.S. & Williams, N.M. (2015) Sampling bias is a challenge

- 593 for quantifying specialization and network structure: lessons from a quantitative  
594 niche model. *Oikos* pp. n/a–n/a.
- 595 Gale, W.A. & Sampson, G. (1995) Good-Turing frequency estimation without  
596 tears. *Journal of Quantitative Linguistics* **2**, 217–237.
- 597 Gibson, R.H., Knott, B., Eberlein, T. & Memmott, J. (2011) Sampling method  
598 influences the structure of plant–pollinator networks. *Oikos* **120**, 822–831.
- 599 González-Varo, J.P., Arroyo, J.M. & Jordano, P. (2014) Who dispersed the seeds?  
600 The use of DNA barcoding in frugivory and seed dispersal studies. *Methods in*  
601 *Ecology and Evolution* **5**, 806–814.
- 602 Good, I.J. (1953) The population frequencies of species and the estimation of  
603 population parameters. *Biometrika* **40**, 237–264.
- 604 Gotelli, N. & Colwell, R. (2001) Quantifying biodiversity: procedures and pitfalls  
605 in the measurement and comparison of species richness. *Ecology Letters* **4**, 379–  
606 391.
- 607 Gotelli, N.J. & Colwell, R.K. (2011) Estimating species richness. *Biological Di-*  
608 *versity Frontiers in Measurement and Assessment* (eds. A.E. Magurran & B.J.  
609 McGill), pp. 39–54, Oxford University Press, Oxford, UK.
- 610 Hortal, J., Borges, P. & Gaspar, C. (2006) Evaluating the performance of species  
611 richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*  
612 **75**, 274–287.
- 613 Ibanez, S. (2012) Optimizing size thresholds in a plant–pollinator interaction web:

- 614 towards a mechanistic understanding of ecological networks. *Oecologia* **170**, 233–  
615 242.
- 616 Jordano, P. (1987) Patterns of mutualistic interactions in pollination and seed dis-  
617 persal: connectance, dependence asymmetries, and coevolution. *The American*  
618 *Naturalist* **129**, 657–677.
- 619 Jordano, P., Bascompte, J. & Olesen, J. (2003) Invariant properties in coevolu-  
620 tionary networks of plant-animal interactions. *Ecology Letters* **6**, 69–81.
- 621 Jordano, P., Vázquez, D. & Bascompte, J. (2009) Redes complejas de interac-  
622 ciones planta—animal. *Ecología y evolución de interacciones planta-animal* (eds.  
623 R. Medel, R. Dirzo & R. Zamora), pp. 17–41, Editorial Universitaria, Santiago,  
624 Chile.
- 625 Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., Petitpierre, E. & Gomez-Zurita,  
626 J. (2009) DNA barcoding insect-host plant associations. *Proceedings Of The*  
627 *Royal Society B-Biological Sciences* **276**, 639–648.
- 628 Magurran, A. (1988) *Ecological diversity and its measurement*. Princeton Univer-  
629 sity Press, Princeton, US.
- 630 Mao, C. & Colwell, R.K. (2005) Estimation of species richness: mixture models,  
631 the role of rare species, and inferential challenges. *Ecology* **86**, 1143–1153.
- 632 Martinez, N. (1991) Artifacts or attributes? Effects of resolution on food-web  
633 patterns in Little Rock Lake food web. *Ecological Monographs* **61**, 367–392.
- 634 Maruyama, P.K., Vizentin-Bugoni, J., Oliveira, G.M., Oliveira, P.E. & Dalsgaard,

- 635 B. (2014) Morphological and spatio-temporal mismatches shape a neotropical  
636 savanna plant-hummingbird network. *Biotropica* **46**, 740–747.
- 637 Memmott, J., Alonso, D., Berlow, E., Dobson, A., Dunne, J.A., Solé, R.V. &  
638 Weitz, J.S. (2006) Biodiversity loss and ecological network structure. *Ecological*  
639 *Networks Linking Structure to Dynamics in Food Webs* (eds. M. Pascual & J.A.  
640 Dunne), pp. 325–347, Oxford University Press, Oxford, UK.
- 641 Moré, M., Amorim, F.W., Benitez-Vieyra, S., Medina, A.M., Sazima, M. &  
642 Cocucci, A.A. (2012) Armament Imbalances: Match and Mismatch in Plant-  
643 Pollinator Traits of Highly Specialized Long-Spurred Orchids. *PLoS ONE* **7**,  
644 e41878.
- 645 Morris, R.J., Gripenberg, S., Lewis, O.T. & Roslin, T. (2013) Antagonistic inter-  
646 action networks are structured independently of latitude and host guild. *Ecology*  
647 *Letters* **17**, 340–349.
- 648 Nielsen, A. & Bascompte, J. (2007) Ecological networks, nestedness and sampling  
649 effort. *Journal of Ecology* **95**, 1134–1141–1141.
- 650 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H. & Jordano, P. (2011)  
651 Missing and forbidden links in mutualistic networks. *Proceedings Of The Royal*  
652 *Society B-Biological Sciences* **278**, 725–732.
- 653 Olito, C. & Fox, J.W. (2014) Species traits and abundances predict metrics of  
654 plant-pollinator network structure, but not pairwise interactions. *Oikos* **124**,  
655 428–436.

- 656 Ollerton, J. & Cranmer, L. (2002) Latitudinal trends in plant-pollinator interac-  
657 tions: are tropical plants more specialised? *Oikos* **98**, 340–350.
- 658 Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes,  
659 R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops,  
660 N., Dulloo, E., Faith, D., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurtt,  
661 G., Jetz, W., Karp, D.S., Mcgeoch, M., Obura, D., Onoda, Y., Pettorelli, N.,  
662 Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S., Turak, E., Walpole, M.  
663 & Wegmann, M. (2013) Essential biodiversity variables. *Science* **339**, 277–278.
- 664 Pocock, M.J.O., Evans, D.M. & Memmott, J. (2012) The Robustness and Restora-  
665 tion of a Network of Ecological Networks. *Science* **335**, 973–977.
- 666 Preston, F. (1948) The commonness, and rarity, of species. *Ecology* **29**, 254–283.
- 667 R Development Core Team (2010) *R: A language and environment for statis-*  
668 *tical computing*. R Foundation for Statistical Computing. Vienna, Austria.  
669 <http://www.R-project.org>, Vienna, Austria.
- 670 Rivera-Hutinel, A., Bustamante, R.O., Marín, V.H. & Medel, R. (2012) Effects  
671 of sampling completeness on the structure of plant-pollinator networks. *Ecology*  
672 **93**, 1593–1603.
- 673 Schleuning, M., Frund, J., Klein, A.M., Abrahamczyk, S., Alarcón, R., Albrecht,  
674 M., Andersson, G.K.S., Bazarian, S., Böhning-Gaese, K., Bommarco, R., Dals-  
675 gaard, B., Dehling, D.M., Gotlieb, A., Hagen, M., Hickler, T., Holzschuh, A.,  
676 Kaiser-Bunbury, C.N., Kreft, H., Morris, R.J., Sandel, B., Sutherland, W.J.,  
677 Svenning, J.C., Tschardtke, T., Watts, S., Weiner, C.N., Werner, M., Williams,



- 678 N.M., Winqvist, C., Dormann, C.F. & Blüthgen, N. (2012) Specialization of  
679 mutualistic interaction networks decreases toward tropical latitudes. *Current*  
680 *Biology* **22**, 1925–1931.
- 681 Snow, B. & Snow, D. (1972) Feeding niches of hummingbirds in a Trinidad valley.  
682 *Journal of Animal Ecology* **41**, 471–485.
- 683 Snow, B. & Snow, D. (1988) *Birds and berries*. Poyser, Calton, UK.
- 684 Stang, M., Klinkhamer, P., Waser, N.M., Stang, I. & van der Meijden, E. (2009)  
685 Size-specific interaction patterns and size matching in a plant-pollinator inter-  
686 action web. *Annals Of Botany* **103**, 1459–1469.
- 687 Strogatz, S. (2001) Exploring complex networks. *Nature* **410**, 268–276.
- 688 Stumpf, M.P.H., Wiuf, C. & May, R.M. (2005) Subnets of scale-free networks  
689 are not scale-free: Sampling properties of networks. *Proceedings of the National*  
690 *Academy of Sciences USA* **102**, 4221–4224.
- 691 Thébault, E. & Fontaine, C. (2010) Stability of ecological communities and the  
692 architecture of mutualistic and trophic networks. *Science* **329**, 853–856.
- 693 Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A.,  
694 Galetti, M., García, M.B., García, D., Gomez, J.M., Jordano, P., Medel, R.,  
695 Navarro, L., Obeso, J.R., Oviedo, R., Ramírez, N., Rey, P.J., Traveset, A.,  
696 Verdú, M. & Zamora, R. (2014) Beyond species loss: the extinction of ecological  
697 interactions in a changing world. *Functional Ecology* **29**, 299–307.
- 698 Vázquez, D., Morris, W. & Jordano, P. (2005) Interaction frequency as a surrogate  
699 for the total effect of animal mutualists on plants. *Ecology Letters* **8**, 1088–1094.

- 700 Vázquez, D.P., Chacoff, N.P. & Cagnolo, L. (2009) Evaluating multiple determi-  
 701 nants of the structure of plant-animal mutualistic networks. *Ecology* **90**, 2039–  
 702 2046.
- 703 Vázquez, D.P., Ramos-Jiliberto, R., Urbani, P. & Valdovinos, F.S. (2015) A con-  
 704 ceptual framework for studying the strength of plant-animal mutualistic inter-  
 705 actions. *Ecology Letters* **18**, 385–400.
- 706 Vizentin-Bugoni, J., Maruyama, P.K., Debastiani, V.J., Duarte, L.d.S., Dalsgaard,  
 707 B. & Sazima, M. (2016) Influences of sampling effort on detected patterns and  
 708 structuring processes of a neotropical plant–hummingbird network. *Journal of*  
 709 *Animal Ecology* **85**, 262–272.
- 710 Vizentin-Bugoni, J., Maruyama, P.K. & Sazima, M. (2014) Processes entangling  
 711 interactions in communities: forbidden links are more important than abundance  
 712 in a hummingbird-plant network. *Proceedings Of The Royal Society B-Biological*  
 713 *Sciences* **281**, 20132397–20132397.
- 714 Wells, K. & O’Hara, R.B. (2012) Species interactions: estimating per-individual  
 715 interaction strength and covariates before simplifying data into per-species eco-  
 716 logical networks. *Methods in Ecology and Evolution* **4**, 1–8.
- 717 Wirta, H.K., Hebert, P.D.N., Kaartinen, R., Prosser, S.W., Várkonyi, G. & Roslin,  
 718 T. (2014) Complementary molecular information changes our perception of food  
 719 web structure. *Proceedings of the National Academy of Sciences USA* **111**, 1885–  
 720 1890.

## Figure captions

**Figure 1.** Sampling ecological interaction networks (e.g., plant-animal interactions) usually focus on different types of subsampling the full network, yielding submatrices  $\Delta[m, n]$  of the full interaction matrix  $\Delta$  with  $A$  and  $P$  animal and plant species. a) all the potential plants interacting with a subset of the animals (e.g., studying just the hummingbird-pollinated flower species in a community); b) all the potential animal species interacting with a subset of the plant species (e.g., studying the frugivore species feeding on figs *Ficus* in a community); and c) sampling a subset of all the potential animal species interacting with a subset of all the plant species (e.g., studying the plant-frugivore interactions of the rainforest understory).

**Figure 2.** Sampling species interactions in natural communities. Suppose an assemblage with  $A = 3$  animal species (red, species 1–3 with three, two, and 1 individuals, respectively) and  $P = 3$  plant species (green, species a-c with three individuals each) (colored balls), sampled with increasing effort in steps 1 to 6 (panels). In Step 1 we record animal species 1 and plant species 1 and 2 with a total of three interactions (black lines) represented as two distinct interactions: 1 – a and 1 – b. As we advance our sampling (panels 1 to 6, illustrating e.g., additional sampling days) we record new distinct interactions. Note that we actually sample and record interactions among individuals, yet we pool the data across species to get a species by species interaction matrix. Few network analyses have been carried out on individual data (Dupont *et al.*, 2014). Above and below each panel are the cumulative number of distinct species and interactions sampled, so

745 that panel 6 illustrates the final network.

746

747 **Figures**

Figure 1:

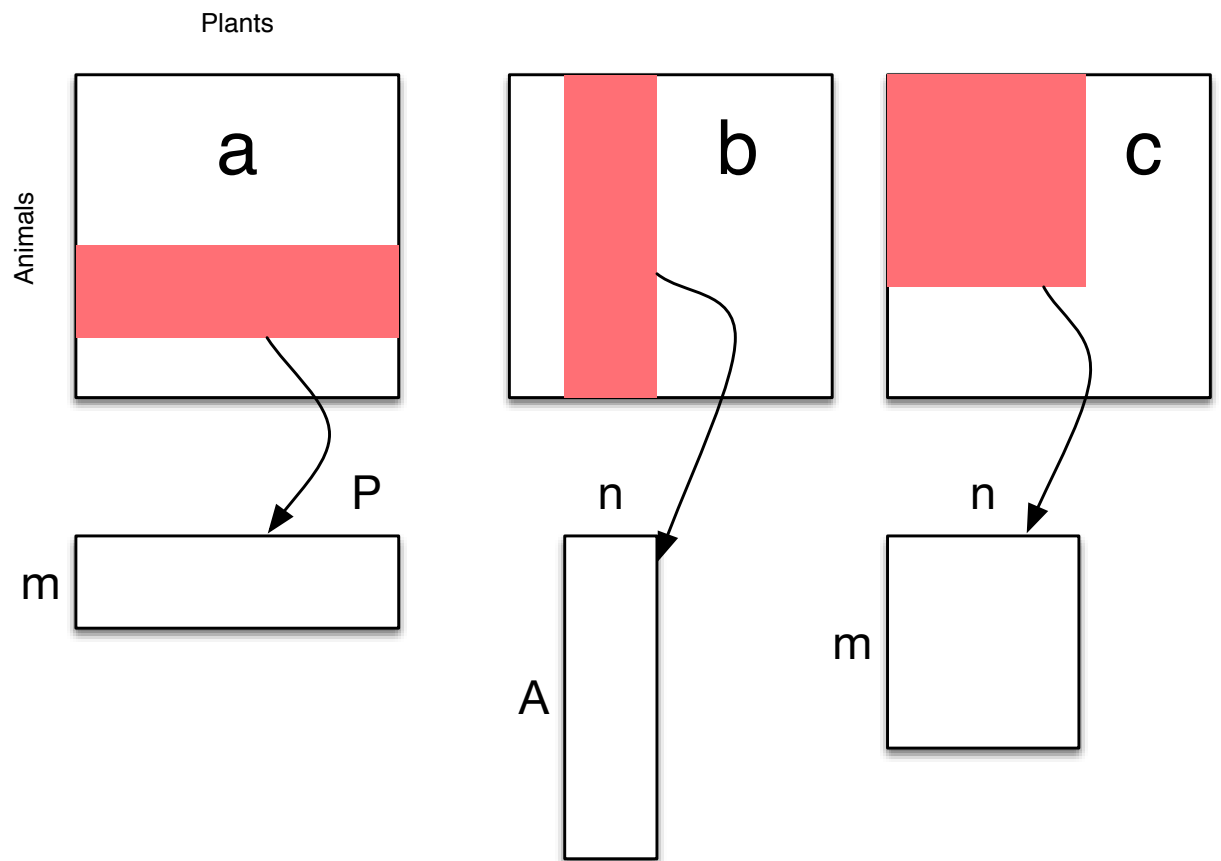
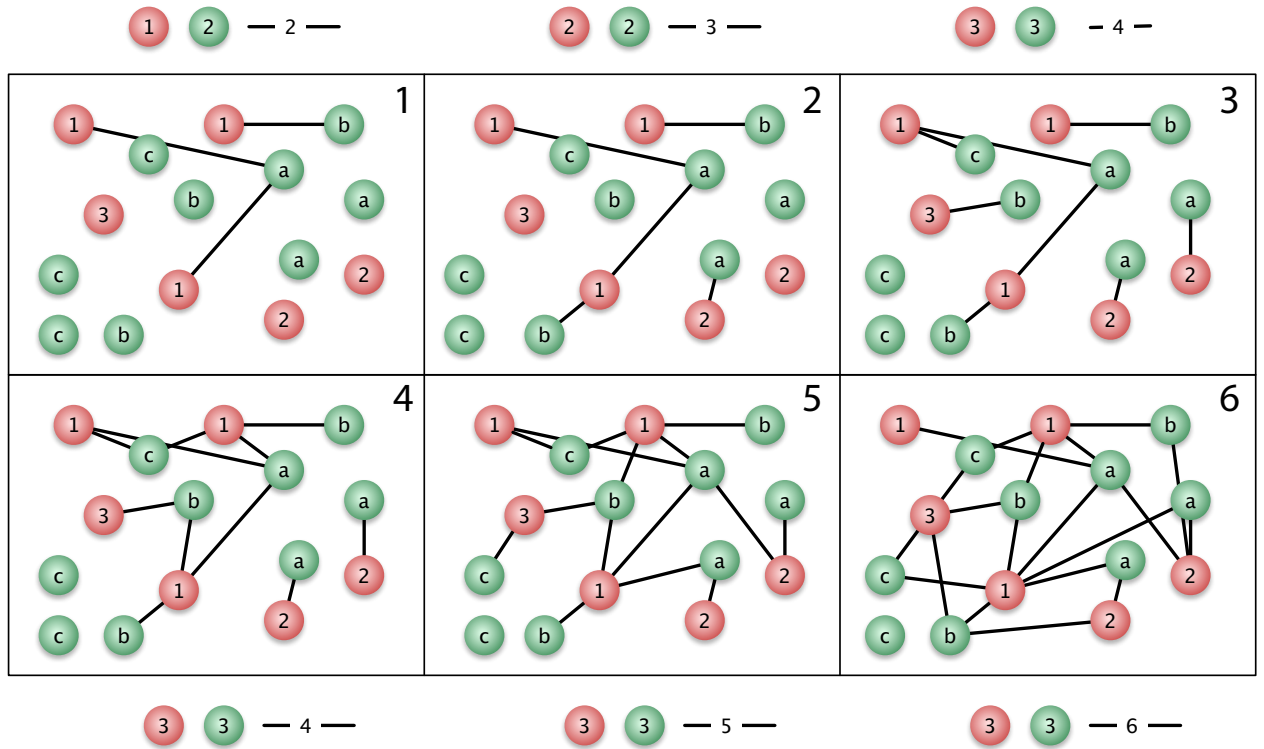


Figure 2:



Jordano - Figure 1

## Table captions

**Table 1.** A taxonomy of link types for ecological interactions (Olesen *et al.* 2011).

$A$ , number of animal species;  $P$ , number of plant species;  $I$ , number of observed links;  $C = 100I/(AP)$ , connectance;  $FL$ , number of forbidden links; and  $ML$ , number of missing links. As natural scientists, our ultimate goal is to eliminate  $ML$  from the equation  $FL = AP - I - ML$ , which probably is not feasible given logistic sampling limitations. When we, during our study, estimate  $ML$  to be negligible, we cease observing and estimate  $I$  and  $FL$ .

**Table 2.** Frequencies of different type of forbidden links in natural plant-animal interaction assemblages.  $AP$ , maximum potential links,  $I_{max}$ ;  $I$ , number of observed links;  $UL$ , number of unobserved links;  $FL$ , number of forbidden links;  $FL_P$ , phenology;  $FL_S$ , size restrictions;  $FL_A$ , accessibility;  $FL_O$ , other types of restrictions;  $ML$ , unknown causes (missing links). Relative frequencies (in parentheses) calculated over  $I_{max} = AP$  for  $I$ ,  $ML$ , and  $FL$ ; for all forbidden links types, calculated over  $FL$ . References, from left to right: Olesen *et al.* 2008; Olesen & Myrthue unpubl.; Snow & Snow 1972 and Jordano *et al.* 2006; Vizentin-Bugoni *et al.* 2014; Jordano *et al.* 2009; Olesen *et al.* 2011.

**Table 3.** A vectorized interaction matrix.

**Table 4.** Sampling statistics for three plant-animal interaction networks (Olesen *et al.* 2011). Symbols as in Table 1;  $N$ , number of records;  $Chao1$  and  $ACE$  are asymptotic estimators for the number of distinct pairwise interactions  $I$  (Hortal

772 *et al.* 2006), and their standard errors;  $C$ , sample coverage for rare interactions  
773 (Chao & Jost 2012). Scaled asymptotic estimators and their confidence intervals  
774 ( $CI$ ) were calculated by weighting *Chao1* and *ACE* with the observed frequencies  
775 of forbidden links.

776

## 777 Tables



Table 1:

Link type	Formulation	Definition
Potential links	$I_{max} = A_{obs}P_{obs}$	Size of observed network matrix, i.e. maximum number of potentially observable interactions; $A_{obs}$ and $P_{obs}$ , numbers of interacting animal and plant species, respectively. These might be below the real numbers of animal and plant species, $A_{est}$ and $P_{est}$ .
Observed links	$I_{obs}$	Total number of observed links in the network given a sufficient sampling effort. Number of ones in the adjacency matrix.
True links	$I_{est}$	Total number of links in the network given a sufficient sampling effort; expected for the augmented $A_{est}P_{est}$ matrix.
Unobserved links	$UL = I_{max} - I_{obs}$	Number of zeroes in the adjacency matrix.
True unobserved links	$UL* = I_{max} - I_{obs}$	Number of zeroes in the augmented adjacency matrix that, eventually, includes unobserved species.
Forbidden links	$FL$	Number of links, which remain unobserved because of linkage constraints, irrespectively of sufficient sampling effort.
Observed Missing links	$ML = A_{obs}P_{obs} - I_{obs} - FL$	Number of links, which may exist in nature but need more sampling effort and/or additional sampling methods to be observed.
True Missing links	$ML* = A_{est}P_{est} - I_{est} - FL$	Number of links, which may exist in nature but need more sampling effort and/or additional sampling methods to be observed. Augments $ML$ for the $A_{est}P_{est}$ matrix.

Table 2:

Link type	Pollination			Seed dispersal		
	Zackenber	Grundvad	Arima Valley	Sta. Virginia	Hato Ratón	Nava Correhuelas
$I_{max}$	1891	646	522	423	272	825
$I$	268 (0.1417)	212 (0.3282)	185 (0.3544)	86 (0.1042)	151 (0.4719)	181 (0.2194)
$UL$	1507 (0.7969)	434 (0.6718)	337 (0.6456)	337 (0.4085)	169 (0.5281)	644 (0.7806)
$FL$	530 (0.3517)	107 (0.2465)	218 (0.6469)	260 (0.7715)	118 (0.6982)	302 (0.4689)
$FL_P$	530 (1.0000)	94 (0.2166)	0 (0.0000)	120 (0.1624)	67 (0.3964)	195 (0.3028)
$FL_S$	... (...)	8 (0.0184)	30 (0.0890)	140 (0.1894)	31 (0.1834)	46 (0.0714)
$FL_A$	... (...)	5 (0.0115)	150 (0.445) <sup>a</sup>	... (...)	20 (0.1183)	61 (0.0947)
$FL_O$	... (...)	... (...)	38 (0.1128) <sup>b</sup>	... (...)	... (...)	363 (0.5637)
$ML$	977 (0.6483)	327 (0.7535)	119 (0.3531)	77 (0.1042)	51 (0.3018)	342 (0.5311)

<sup>a</sup>, Lack of accessibility due to habitat uncoupling, i.e., canopy-foraging species vs. understory species.

<sup>b</sup>, Colour restrictions, and reward per flower too small relative to the size of the bird.

Dots indicate no data available for the  $FL$  type.

Table 3:

Interaction	Sample 1	Sample 2	Sample 3	...	Sample $i$
A1 - P1	12	2	0	...	6
A1 - P2	0	0	0	...	1
...	...	...	...	...	...
A5 - P3	5	0	1	...	18
A5 - P4	1	0	1	...	3
...	...	...	...	...	...
A <sub>i</sub> - P <sub>i</sub>	1	0	1	...	2

Table 4:

	Hato Ratón	Nava Correhuelas	Zackenberg
$A$	17	33	65
$P$	16	25	31
$I_{max}$	272	825	1891
$N$	3340	8378	1245
$I$	151	181	268
$C$	0.917	0.886	0.707
$Chao1$	$263.1 \pm 70.9$	$231.4 \pm 14.2$	$509.6 \pm 54.7$
$ACE$	$240.3 \pm 8.9$	$241.3 \pm 7.9$	$566.1 \pm 14.8$
% <i>unobserved</i> <sup>a</sup>	8.33	15.38	47.80

<sup>a</sup>, estimated with library Jade (R Core Development Team 2010, Chao *et al.* 2015)