Sampling networks of ecological interactions

Pedro Jordano*[a]

[a]Integrative Ecology Group, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas (EBD-CSIC), Avenida Americo Vespucio s/n, E–41092 Sevilla, Spain

*Corresponding author: jordano@ebd.csic.es

Summary

1. Sampling ecological interactions presents similar challenges, problems, potential biases, and constraints as sampling individuals and species in biodiversity inventories. Interactions are just pairwise relationships among individuals of two unrelated species, such as those among plants and their seed dispersers in frugivory interactions or those among plants and their pollinators. Sampling interactions is a fundamental step to build robustly estimated interaction networks, yet few analyses have attempted a formal approach to their sampling protocols.

*jordano@ebd.csic.es

2. 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.

3. While the complete inventory of interactions is likely impossible, a robust characterization of its main patterns and metrics is probably realistic. We must acknowledge that a sizable 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, the restrictions imposed by the organisms life-histories. Thus, the number of observed interactions I in robustly sampled networks is typically $I << I_{max}$, resulting in extremely sparse interaction matrices with low connectance.

4. Reasons for forbidden links are multiple but mainly stem from spatial and temporal uncoupling of partner species encounters and from intrinsically low probabilities of interspecific encounter for many of the potential pairwise interactions. Adequately assessing the completeness of a network of ecological interactions thus needs a deep knowledge of the natural history details embedded, so that forbidden links can be "discounted" when addressing sampling effort.

5. 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. This is crucial to assess the fast-paced and devastating effects of defaunation-driven loss of key ecological

interactions and the services they provide.

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).

Biodiversity assessment aims at sampling individuals in collections and determining the number of species represented. Given that, by definition, samples are incomplete, these collections enumerate a lower number of the species actually present. The ecological literature dealing with robust estimators of species richness and diversity in collections of individuals is immense, and a number of useful approaches have been used to obtain such estimates (Magurran, 1988; Gotelli & Colwell, 2001; Hortal, Borges & Gaspar, 2006; Colwell, 2009; Gotelli & Colwell, 2011). Recent effort has been also focused at defining essential biodiversity variables (EBV) (Pereira *et al.*, 2013) that can be sampled and measured repeatedly 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, the one that supports the existence of species

(Memmott *et al.*, 2006). 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 to sampling interactions in natural communities have to do with problems associated to 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 (Mao & Colwell, 2005). 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.

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). Thus, sampling interactions should be a central issue when identifying and diagnosing ecosystem services (e.g., pollination, natural seeding by frugivores, etc.). Fortunately, all the whole battery of biodiversity-related tools used by ecologists to sample biodiversity (species, *sensu stricto*) can be extended and applied to the sampling of interactions. Analogs are evident between these approaches (Colwell, Dunn & Harris, 2012). Monitoring interactions is analogous to any biodiversity sampling [i.e., a species inventory Jordano (1987); Jordano, Vázquez & Bascompte (2009)] and is subject to similar methodological shortcomings, especially under-sampling (Coddington *et al.*, 2009; Vazquez, Chacoff & Cagnolo, 2009; Dorado *et al.*, 2011; Rivera-Hutinel *et al.*, 2012). For example, when we study mutualistic networks, our goal is to make an inventory of the distinct pairwise interactions that made up the network. 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 can 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 (Mao & Colwell, 2005) put it, these are the workings of Preston's demon, the moving "veil line" between detected and the undetected interactions as sample size increases (Preston, 1948).

Early efforts to recognize and solve sampling problems in analyses of interactions stem from researchers interested in food web analyses and in determining the biases of undersampled food web metrics (Martinez, 1991; Cohen *et al.*, 1993; Martinez, 1993; Bersier, Banasek-Richter & Cattin, 2002; Brose, Martinez & Williams, 2003; Banasek-Richter, Cattin & Bersier, 2004). 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; Vazquez, Chacoff & Cagnolo, 2009; Gibson *et al.*, 2011; Olesen *et al.*, 2011; Chacoff *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox, 2014; Bascompte & Jordano, 2014; Vizentin-Bugoni, Maruyama & Sazima, 2014; Frund, McCann &

Williams, 2015). The sampling approaches have been extended to predict patterns of coextintions in interaction assemblages (e.g., hosts-parasites) (Colwell, Dunn & Harris, 2012). Most empirical studies provide no estimate of sampling effort, implicitly assuming that the reported network patterns and metrics are robust. Yet recent evidences point out that number of partner species detected, number of actual links, and some aggregate statistics describing network patterns, are prone to sampling bias (Nielsen & Bascompte, 2007; Dorado *et al.*, 2011; Olesen *et al.*, 2011; Chacoff *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox, 2014; Frund, McCann & Williams, 2015). Most of these evidences, however, come from either theoretical, simulation, studies (Frund, McCann & Williams, 2015) or from relatively species-poor assemblages. Even for species-rich, tropical assemblages it might be erroneous to conclude that network data routinely come from insufficiently sampled datasets (Ollerton & Cranmer, 2002; Chacoff *et al.*, 2012), given the extremely sparse nature of these interaction matrices because of the prevalence of forbidden links (which, by definition, cannot be documented despite extensive sampling effort). However, most certainly, sampling limitations pervade biodiversity inventories in tropical areas (Coddington *et al.*, 2009) and we might rightly expect that frequent interactions may be over-represented and rare interactions may be missed entirely in studies of mega-diverse assemblages (Bascompte & Jordano, 2014); but, to what extent?

Sampling interactions: methods

When we sample interactions in the field we record the presence of two species that interact in some way. For example, Snow and Snow(Snow & Snow, 1988) recorded an interaction whenever they saw a bird "touching" a fruit on a plant. In a similar way, interactions between pollinators and plants are tallied by recording any visit of a pollinator entering a flower and touching the reproductive parts. We observe and record feeding observations, visitation, occupancy, presence in pollen loads or in fecal samples, etc., of *individual* animals or plants and accumulate pairwise interactions, i.e., lists of species partners and the frequencies with which we observe them. Therefore, estimating the sampling completeness of pairwise interactions for a whole network, requires estimating the number (richness) of distinct pairwise interactions accumulated as sampling effort is increased, pooling the data for all partner species. Most, if not all, types of ecological interactions can be illustrated by 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, $\Delta = AP$, is a 2D representation of the interactions among, say, A animal species (rows) and P plant species (columns) (Jordano, 1987; Bascompte & Jordano, 2014). An interaction matrix Δ consists of an array of zeroes or ones, or an array of numeric values (including zeroes)- if the data (interaction frequencies) are quantified. The matrix entries illustrate the values of the pairwise interactions visualized in the Δ 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; Vazquez *et al.*, 2015). Given that 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), a frequently used proxy for interaction intensities w_{ji} is just how frequent are new interspecific encounters, whether or not appropriately weighted to estimate interaction effectiveness (Vazquez, Morris & Jordano, 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 there are two considerations we need to take into account. First is whether we are sampling the whole community of interactor species (all the animals, all the plants) or we sample 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 Δ_{AP}. Subsets can be: a) all the potential plants interacting with a subset of the animals (Fig. 1a); b) all the potential animal species interacting with a subset of the plant species (Fig. 1b); c) a subset of all the potential animal species interacting with a subset of all the plant species (Fig. 1c). While some discussion has considered how to establish the limits of what represents a network (Strogatz, 2001) (in analogy to discussion on food-web limits (Cohen, 1978)), it must be noted that situations a-c in Fig. 1 do not represent complete interaction networks. As vividly stated by Cohen et al. (1993): "*As more comprehensive, more detailed, more explicit webs become available, smaller, highly aggregated, incompletely described webs may progressively be dropped from analyses of web structure (though such webs may remain useful for other purposes, such as pedagogy)*". Subnet sampling is generalized in studies of biological networks (e.g., protein interactions, gene regulation), yet it is important to recognize that most properties of subnetworks (even random subsamples) do not represent properties of whole networks (Stumpf, Wiuf & May, 2005).

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

Most recent analyses of plant-animal interaction networks are phyto-centric; just 3.5% of available plant-pollinator ($N = 58$) or 36.6% plant-frugivore ($N = 22$) interaction datasets are zoo-centric (see (Schleuning *et al.*, 2012)). Moreover, most available datasets on host-parasite or plant-herbivore interactions are "host-centric" or phyto-centric (e.g., (Thébault & Fontaine, 2010; Eklöf *et al.*, 2013)). This maybe 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 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). System symmetry has been suggested to influence estimations of generalization levels in plants and animals when measured as I_A and I_P (Elberling & Olesen, 1999); thus, differences in I_A and I_P between networks may arise from different $A : P$ ratios rather than other ecological factors (Olesen & Jordano, 2002).

Interestingly enough, quite complete analyses of interaction networks can be obtained when combining both phyto-centric and zoo-centric sampling. For example, Bosch et al. (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. Connectance increased 1.43-fold, mean plant connectivity went from 18.5 to 26.4, and mean pollinator connectivity from 2.9 to 4.1; moreover, extreme specialist pollinator species (singletons in the adjacency matrix) decreased 0.6-fold. Zoo-centric sampling has recently been extended with the use of DNA-barcoding, for example with plant-herbivore (Jurado-Rivera *et al.*, 2009) and plant-frugivore interactions (González-Varo, Arroyo & Jordano, 2014). For mutualistic networks we would expect that zoo-centric sampling could help unveiling interactions for rare species or for relatively common species which are difficult to sample by direct observation. Future methodological work may provide significant advances showing how mixing different sampling strategies strengthens the completeness of network data. These mixed strategies may combine, for instance, focal analyses, pollen load or seed contents, camera traps, and DNA barcoding records. We might expect increased power of these mixed sampling approaches when combining different methods from both phyto- and zoo-centric perspectives (Bosch *et al.*, 2009; Bluthgen, 2010).

Sampling interactions: rationale

The number of distinct pairwise interactions that we can record in a landscape (an area of relatively homogeneous vegetation, analogous to the one we would use to monitor species diversity) is equivalent to the number of distinct classes in which we can classify the recorded encounters among individuals of two different species. Yet, individual-based plant-animal interaction networks have been only recently studied (Dupont, Trøjelsgaard & Olesen, 2011). We walk in the forest and see a blackbird Tm picking an ivy Hh fruit and ingesting it: we have a record for $Tm - Hh$ interaction. We keep advancing and record again a blackbird feeding on hawthorn Cm fruits so we record a $Tm - Cm$ interaction; as we advance we encounter another ivy plant and record a blackcap swallowing a fruit so we now have a new $Sa - Hh$ interaction, and so on. At the end we have a series of classes (e.g., $Sa - Hh$, $Tm - Hh$, $Tm - Cm$, etc.), along with their observed frequencies. Bunge & Fitzpatrick (Bunge & Fitzpatrick, 1993) review the main aspects and approaches to estimate the number of distinct classes C in a sample of observations. 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 made interaction-based accumulation curves. If an interaction-based curve points towards 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 (N. Gotelli, pers. com.). 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 cumulating Curve (SAC): 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).

Our sampling above would have resulted in a vector $n = [n_1 ... n_C]'$ where n_i is the number of records in the i^{th} class. As stressed by Bunge & Fitzpatrick (Bunge & Fitzpatrick, 1993), however, the i^{th} class would appear in the sample if and only if $n_i > 0$, and we don't know *a priori* which n_i are zero. So, n is not observable. Rather, what we get is a vector $c = [c_1...c_n]'$ where c_j is the number of classes represented j times in our sampling: c_1 is the number of singletons, c_2 is the number of twin pairs, 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.

Estimating the number of interactions with resulting robust estimates of network parameters is a central issue in the study of ecological interaction networks (Jordano, 1987; Bascompte & Jordano, 2014). In 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 and P species lists), some of their pairwise interactions are impossible to be recorded. The reason is forbidden links. Independently of whether we sample full communities of subset communities we face a problem: some of the interactions that we can visualize in the empty adjacency matrix Δ with size AP will simply not occur. Thus, independently of the sampling effort we put, we'll never document these pairwise interactions. With a total of AP "potential" interactions , a fraction of them are impossible to record, because they are forbidden (Jordano, Bascompte & Olesen, 2003; Olesen *et al.*, 2011). Forbidden links are constraints for the establishment of new links, and mainly arise from the biological attributes of the species: no link can be established between a plant and an animal mutualist differing in phenology, i.e. the seeds of a winter-ripening plant cannot be dispersed by a frugivore that is a summer stopover migrant (Jordano, 1987). Or, for instance, short-tongued pollinators cannot successfully reach the nectar in long-corolla flowers and pollinate them efficiently (Moré *et al.*, 2012). Forbidden links are thus represented as structural zeroes in the interaction matrix, i.e., matrix cells that cannot get a non-zero value. So, we need to account for the frequency of these structural zeros in our matrix before proceeding. For example, most measurements of connectance ($C = I/(AP)$) implicitly ignore the fact that by taking the full product AP in the denominator they are underestimating the actual connectance value, i.e., the fraction of actual interactions I relative to the *biologically possible* ones, not to the total maximum $I_{max} = AP$.

Adjacency matrices are frequently sparse, i.e., they are densely populated with zeroes, with a fraction of them being structural (i.e., unobservable interactions) (Bascompte & Jordano, 2014). It would be thus a serious interpretation error to attribute the sparseness of adjacency matrices for bipartite networks to under-sampling. The actual typology of link types in ecological interaction 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 (Jordano, Bascompte & Olesen, 2003; Olesen *et al.*, 2011). Forbidden links are non-occurrences of pairwise interactions that can be accounted for by biological constraints, such as spatio-temporal uncoupling, size or reward mismatching, foraging constraints (e.g., accessibility), and physiological-biochemical constraints (Jordano, 1987). We still have extremely reduced information about the frequency of forbidden links in natural communities (Jordano, Bascompte & Olesen, 2003; Stang *et al.*, 2009; Vazquez, Chacoff & Cagnolo, 2009; Olesen *et al.*, 2011; Ibanez, 2012; Maruyama *et al.*, 2014; Vizentin-Bugoni, Maruyama & Sazima, 2014) (Table 1). Forbidden links FL may actually account for a relatively large fraction of unobserved interactions UL when sampling taxonomically-restricted subnetworks (e.g., plant-hummingbird pollination networks) (Table 1). Phenological unmatching is also prevalent in most networks, and may add up to explain ca. 25–40% of the forbidden 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 added sampling effort. For instance, habitat use patterns by hummingbirds in the Aroma Valley network (Table 2; (Snow & Snow, 1972)) impose a marked pattern of microhabitat mismatches causing up to 44.5% of the forbidden links. There are a myriad of biological causes beyond those included as FL in Table 2 that may contribute explanations for UL: limits of color perception and or partial preferences, presence of secondary metabolites in fruit pulp and leaves, toxins and combinations of monosaccharides in nectar, etc. However, it is surprising that just the limited set of forbidden link types in Table 1 explain between 24.6–77.2% of the unobserved links. Notably, the Arima Valley, Santa Virgńia, 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 infer 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 by identified in the adjacency matrix; second, a critical analysis of sampling robustness a robust estimate of the actual fraction of missing links, M, and thus, a robust estimate of I.

Asymptotic diversity estimates

Let's assume a sampling of the diversity in a specific locality, over relatively homogeneous landscape where we aim at determining the number of species present for a particular group of organisms. To do that we carry out transects or plot samplings across the landscape, adequately replicated so we obtain a number of samples. Briefly, S_{obs} is the total number of species observed in a sample, or in a set of samples. S_{est} is the estimated number of species in the community represented by the sample, or by the set of samples, where *est* indicates an estimator. With abundance data, let S_k be the number of species each represented by exactly k individuals in a single sample. Thus, S_0 is the number of undetected species (species present in the community but not included in the sample), S_1 is the number of singleton species, S_2 is the number of doubleton species, etc. The total number of individuals in the sample would be:

$$n = \sum_{k=1}^{S_{obs}} S_k$$

A frequently used asymptotic, bias corrected, non-parametric estimator is S_{Chao} (Hortal, Borges & Gaspar, 2006; Chao, 2005; Colwell, 2013):

$$S_{Chao} = S_{obs} + \frac{S_1(S_1 - 1)}{2(S_2 + 1)}$$

Another frequently used alternative is the Chao2 estimator, S_{Chao2} (Gotelli & Colwell, 2001), which has been reported to have a limited bias for small sample sizes (Colwell & Coddington, 1994; Chao, 2005):

$$S_{Chao2} = S_{obs} + \frac{S_1^2}{2S_2}$$

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). Such a curve eventually reaches an asymptote converging with S_{est}. In an analogous way, interaction accumulation curves (IAC), analogous to SACs, can be used to assess the robustness of interactions sampling for plant-animal community datasets (Jordano, 1987; Jordano, Vázquez & Bascompte, 2009; Olesen *et al.*, 2011). For instance, a random accumulator function (e.g., library vegan in the R Package (R Development Core Team, 2010)) which finds the mean IAC and its standard deviation from random permutations of the data, or subsampling without replacement (Gotelli & Colwell, 2001) can be used to estimate the expected number of distinct pairwise interactions included in a given sampling of records (Jordano, Vázquez & Bascompte, 2009; Olesen *et al.*, 2011). We start with a vectorized interaction matrix representing the pairwise interactions (rows) recorded during a cumulative number of censuses or sampling periods (columns) (Table 3) , in a way analogous to a biodiversity sampling matrix with species as rows and sampling units (e.g., quadrats) as columns (Jordano, Vázquez & Bascompte, 2009). In this way we effectively extend sampling theory developed for species diversity to the sampling of ecological interactions. Yet future theoretical work will be needed to formally assess the similarities and differences in the two approaches and developing biologically meaningful null models of expected interaction richness with added sampling effort.

Assessing sampling effort when recording interactions

The basic method we can propose 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 a community of A animal species and P plant species. The new "species" we aim to sample are the pairwise interactions (Table 3). So, if we have in our community *Turdus merula* (Tm) and *Rosa canina* (Rc) and *Prunus mahaleb* (Pm), our problem will be to sample 2 new "species": $Tm - Rc$ and $Tm - Pm$. In general, if we have $A = 1...i$, animal species and $P = 1...j$ plant species, 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.

For example, 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 flower-pollinator interactions involving bumblebees could have a higher detectability than flower-pollinator pairwise interactions involving, say, nitidulid beetles. These more homogeneous groupings of pairwise interactions within a network define modules (Bascompte & Jordano, 2014), so we might expect that interactions of a given module (e.g., plants and their hummingbird pollinators; Fig. 1a) may share similar detectability values, in an analogous way to species groups receiving homogeneous detectability values in mixture models (Mao & Colwell, 2005). Such sampling, in its simplest form, would result in a sample with multiple pairwise interactions detected, in which the number of interaction events recorded for each distinct interaction found in the sample is recorded (i.e., a column vector in Table 3, corresponding to, say, a sampling day). The number of interactions recorded for the i_{th} pairwise interaction (i.e., $A_i P_j$ in Table 3), Y_i could be treated as a Poisson random variable with a mean parameter λ_i, its detection rate. Mixture models (Mao & Colwell, 2005) include estimates for abundance-based data (their analogous in interaction sampling would be weighted data), where Y_i is a Poisson random variable with detection rate λ_i. This is combined with the incidence-based model, where Y_i is a binomial random variable (their analogous in interaction sampling would be presence/absence records of interactions) with detection odds λ_i. Let T be the number of samples in an incidence-based data set. A Poisson/binomial density can be written as (Mao & Colwell, 2005):

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

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

The detection rates λ_i depend on the relative abundances ϕ_i of the interactions, the probability of a pairwise interaction being detected when it is present, and the sample size (the number of interactions recorded), which, in turn, is a function of the sampling effort. Unfortunately, no specific sampling model has been developed along these lines for species interactions and their characteristic features. For example, a complication factor might be that interaction abundances, ϕ_i, in real assemblages are a function of the abundances of interacting species, that determine interspecific encounter rates; yet they also depend on biological factors that ultimately determine if the interaction occurs when the partner species are present. It its simplest form, ϕ_i could be estimated from just the product of partner species abundances, an approach recently used as a null model to assess the role of biological constraints in generating forbidden links and explaining interaction patterns (Vizentin-Bugoni, Maruyama & Sazima, 2014). Yet more complex models should incorporate not only interspecific encounter probabilities, but also phenotypic matching and incidence of forbidden links.

Rarefaction analysis and diversity-accumulation analysis (Magurran, 1988; Hortal, Borges & Gaspar, 2006) come 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, Vázquez & Bascompte, 2009; Olesen *et al.*, 2011). Asymptotic estimates of interaction richness and its associated standard errors and confidence intervals can thus be obtained (Hortal, Borges & Gaspar, 2006). It should be noted that the asymptotic estimate of interaction richness implicitly ignores the fact that, due to forbidden links, a number of pairwise interactions among the I_{max} number specified in the adjacency matrix Δ cannot be recorded, irrespective of sampling effort. Therefore, the asymptotic value most likely is an overestimate of the actual maximum number of links that can be present in an assemblage. If forbidden links are taken into account, the asymptotic estimate should be lower. Yet, to the best of my knowledge, there is no theory developed to estimate this "biologically real" asymptotic value. Not unexpectedly, most recent analyses of sampling effort in ecological network studies found evidences of undersampling (Chacoff *et al.*, 2012). This needs not to be true, especially when interaction subwebs are studied (Olesen *et al.*, 2011; Vizentin-Bugoni, Maruyama & Sazima, 2014), and once the issue of structural zeroes in the interaction matrices is effectively incorporated in the estimates.

The *real* missing links

Given that a fraction of unobserved interactions can be accounted for by forbidden links, what about the remaining missing interactions? We have already discussed 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? Multispecific assemblages of distinct taxonomic relatedness, whose interactions can be represented as bipartite networks (e.g., host-parasite, plant-animal mutualisms, plant-herbivore interactions- with two distinct sets of unrelated higher taxa), are shaped by interspecific encounters among individuals of the partners (Fig. 2). A crucial ecological aspect limiting these interactions 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 "neutral" probabilities of interspecific encounter (PIE) would be simply the product of the two lognormal distributions. Thus, we might expect that for low PIE values, pairwise interactions would be either extremely difficult to sample, or just simply non-occurring in nature. Consider the Nava de las Correhuelas interaction web (NCH, Table 2), with $A = nnn$, $P = nnn$, $I = nnn$, and almost half of the unobserved interactions not accounted for by forbidden links missing links, $M = 53.1\%$. Given the robust sampling of this network (Jordano, Vázquez & Bascompte, 2009), a sizable fraction of these possible but missing links would be simply not occurring in nature, most likely by 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 $PIE_{median} < 1.4\ 10^{-4}$ we may safely expect (note the quantile estimate $Q_{75\%}$ =3.27 10^{-4}) that a sizable fraction of these missing interactions may simply not occur according to this neutral expectation (Jordano, 1987) (Olesen *et al.*, 2011) (neutral forbidden links, *sensu* (Canard *et al.*, 2012)). Which is the expected frequency for pairwise interactions? and, which is the expected probability for unobserved interactions? More specifically, which is the probability of missing interactions, M (i.e., the unobserved ones that cannot be accounted for as forbidden links)?

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). The technique, developed by Alan Turing and I.J. Good with applications to linguistics and word analysis (Gale & Sampson, 1995) has been recently applied in ecology (Chao *et al.*, 2015), estimates the probability of recording an interaction of a hitherto unseen pair of partners, given a set of past records of interactions between other species pairs. Let a sample of N interactions so that n_r distinct pairwise interactions have exactly r records. All Good-Turing estimators obtain the underlying frequencies of events as:

$$P(X) = \frac{(N_X + 1)}{T} \, (1 - \frac{E(1)}{T}) \tag{1}$$

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

Note, however, that Good-Turing estimators, as the traditional asymptotic estimators, do not account in our case for the forbidden interactions. To account for these FL I re-scaled the asymptotic estimates, so that a more meaningful estimate could be obtained (Table 4). The scaling was calculated as $Chao1 * (I + ML)/AP$, just correcting for the FL frequency, given that $I + ML$ represent the total *feasible* interactions when discounting the forbidden links (Table 1). After scaling, observed I values (Table 2) are within the $Chao1$ and ACE asymptotic estimates but below the ACE estimates for Hato Ratón and Zackenberg (Table 4). Thus, even after re-scaling for FL, it is likely that adequate characterization of most interaction networks will require intensive sampling effort.

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, many surveyed networks to date have been subnets of much larger networks. This is true for protein interaction, gene regulation, and metabolic networks, where only a subset of the molecular entities in a cell have been sampled (Stumpf, Wiuf & May, 2005). Despite recent attempts to document whole ecosystem meta-networks (Pocock, Evans & Memmott, 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 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; yet, concluding that all ecological network datasets are undersampled would be unrealistic. The reason stems form 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).

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 unobservable 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. Some key components of this sampling are analogous to species sampling and traditional biodiversity inventories; however, there are important differences. 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 mutualisms 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. A re-scaling of traditional asymptotic estimates for interaction richness can be applied whenever the knowledge of natural history details about the study system is sufficient to estimate at least the main causes of forbidden links. Moreover, recent implementations of inference methods for unobserved species (Chao *et al.*, 2015) can be combined with the forbidden link approach, yet they do not account either for the existence of these ecological constraints.

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. Yet we still have a limited knowledge of the biodiversity of ecological interactions, but they are being lost (extinct) at a very fast pace, frequently preceding species extinctions (Valiente-Banuet *et al.*, 2014). We urgently need robust techniques to assess the completeness of ecological interactions networks because this knowledge will allow the identification of the minimal components of ecological complexity that need to be restored after perturbations to rebuild functional ecosystems.

Acknowledgements

I am indebted to Jens M. Olesen, Alfredo Valido, Jordi Bascompte, Thomas Lewinshon, John N. Thompson, Nick Gotelli, Carsten Dormann, and Paulo R. Guimaraẽs Jr. for useful and thoughtful comments and discussion at different stages of this manuscript. The study was supported by a Junta de Andalucía Excellence Grant (RNM–5731), as well as a Severo Ochoa Excellence Award from the Ministerio de Economía y Competitividad (SEV–2012–0262). The Agencia de Medio Ambiente, Junta de Andalucía, provided generous facilities that made possible my long-term field work in different natural parks.

Data accessiblity

Please state where you have deposited the raw data underlying your analyses. It will need to include the name of the repository (e.g. Dryad, figshare, GenBank etc.) and location of the data (i.e DOI). For authors archiving at Dryad, we can facilitate the process when your paper is accepted.

References

Banasek-Richter, C., Cattin, M. & Bersier, L. (2004) Sampling effects and the robustness of quantitative and qualitative food-web descriptors. *Journal of Theoretical Biology*, **226**, 23–32.

Bascompte, J. & Jordano, P. (2014) *Mutualistic networks*. Monographs in Population Biology, No. 53. Princeton University Press, Princeton, NJ.

Bersier, L., Banasek-Richter, C. & Cattin, M. (2002) Quantitative descriptors of food-web matrices. *Ecology*, **83**, 2394–2407.

Bluthgen, N. (2010) Why network analysis is often disconnected from community ecology: A critique and an ecologist's guide. *Basic And Applied Ecology*, **11**, 185–195.

Bosch, J., Martín González, A.M., Rodrigo, A. & Navarro, D. (2009) Plant-pollinator networks: adding the pollinator's perspective. *Ecology Letters*, **12**, 409–419.

Brose, U., Martinez, N. & Williams, R. (2003) Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*, **84**, 2364–2377.

Bunge, J. & Fitzpatrick, M. (1993) Estimating the number of species: a review. *Journal of the American Statistical Association*, **88**, 364–373.

Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012) Emergence of structural patterns in neutral trophic networks. *PLoS ONE*, **7**, e38295.

Chacoff, N.P., Vazquez, D.P., Lomascolo, 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.

Chao, A., Hsieh, T.C., Chazdon, R.L., Colwell, R.K. & Gotelli, N.J. (2015) Unveiling the species-rank abundance distribution by generalizing the Good-Turing sample coverage theory. *Ecology*, **96**, 1189–1201.

Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009) Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology*, **78**, 573–584.

Cohen, J.E. (1978) *Food webs and niche space.* Princeton University Press, Princeton, New Jersey, US.

Cohen, J.E., Beaver, R.A., Cousins, S.H., DeAngelis, D.L., Goldwasser, L., Heong, K.L., Holt, R.D., Kohn, A.J., Lawton, J.H., Martinez, N., O'Malley, R., Page, L.M., Patten, B.C., Pimm, S.L., Polis, G., Rejmanek, M., Schoener, T.W., Schenly, K., Sprules, W.G., Teal, J.M., Ulanowicz, R., Warren, P.H., Wilbur, H.M. & Yodis, P. (1993) Improving food webs. *Ecology*, **74**, 252–258.

Colwell, R. & Coddington, J. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences*, **345**, 101–118.

Colwell, R.K. (2009) Biodiversity: concepts, patterns, and measurement. *The Princeton Guide to Ecology* (ed. S.A. Levin), pp. 257–263. Princeton University Press, Princeton.

Colwell, R.K. (2013) EstimateS: Biodiversity Estimation. -, pp. 1–33.

Colwell, R.K., Dunn, R.R. & Harris, N.C. (2012) Coextinction and persistence of dependent species in a changing world. *Annual Review of Ecology Evolution and Systematics*, **43**, 183–203.

Dorado, J., Vazquez, D.P., Stevani, E.L. & Chacoff, N.P. (2011) Rareness and specialization in plant-pollinator networks. *Ecology*, **92**, 19–25.

Dormann, C.F., Frund, J., Bluthgen, N. & Gruber, B. (2009) Indices, graphs and null models: Analyzing bipartite ecological networks. *Open Ecology Journal*, **2**, 7–24.

Dupont, Y.L., Trøjelsgaard, K. & Olesen, J.M. (2011) Scaling down from species to individuals: a flower–visitation network between individual honeybees and thistle plants. *Oikos*, **120**, 170–177.

Dupont, Y.L., Trøjelsgaard, K., Hagen, M., Henriksen, M.V., Olesen, J.M., Pedersen, N.M.E. & Kissling, W.D. (2014) Spatial structure of an individual-based plant-pollinator network. *Oikos*, **123**, 1301–1310.

Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P., Dalsgaard, B., de Sassi, C., Galetti, M., Guimaraes, P.R., Lomáscolo, S.B., Martín González, A.M., Pizo, M.A., Rader, R., Rodrigo, A., Tylianakis, J.M., Vazquez, D.P. & Allesina, S. (2013) The dimensionality of ecological networks. *Ecology Letters*, **16**, 577–583.

Elberling, H. & Olesen, J.M. (1999) The structure of a high latitude plant-flower visitor system: the dominance of flies. *Ecography*, **22**, 314–323.

Frund, J., McCann, K.S. & Williams, N.M. (2015) Sampling bias is a challenge for quantifying specialization and network structure: lessons from a quantitative niche model. *Oikos*, pp. n/a–n/a.

Gale, W.A. & Sampson, G. (1995) Good-Turing frequency estimation without tears. *Journal of Quantitative Linguistics*, **2**, 217–237.

Gibson, R.H., Knott, B., Eberlein, T. & Memmott, J. (2011) Sampling method influences the structure of plant–pollinator networks. *Oikos*, **120**, 822–831.

González-Varo, J.P., Arroyo, J.M. & Jordano, P. (2014) Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. *Methods in Ecology and Evolution*, **5**, 806–814.

Good, I.J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237–264.

Gotelli, N.J. & Colwell, R.K. (2011) Estimating species richness. *Biological Diversity Frontiers in Measurement and Assessment* (eds. A.E. Magurran & B.J. McGill), pp. 39–54. Oxford University Press, Oxford, UK.

Gotelli, N. & Colwell, R. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Hortal, J., Borges, P. & Gaspar, C. (2006) Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*, **75**, 274–287.

Ibanez, S. (2012) Optimizing size thresholds in a plant–pollinator interaction web: towards a mechanistic understanding of ecological networks. *Oecologia*, **170**, 233–242.

Jordano, P. (1987) Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *The American Naturalist*, **129**, 657–677.

Jordano, P., Bascompte, J. & Olesen, J. (2003) Invariant properties in coevolutionary networks of plant-animal interactions. *Ecology Letters*, **6**, 69–81.

Jordano, P., Vázquez, D. & Bascompte, J. (2009) Redes complejas de interacciones planta—animal. *Ecología y evolución de interacciones planta-animal* (eds.

R. Medel, R. Dirzo & R. Zamora), pp. 17–41. Editorial Universitaria, Santiago, Chile.

Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., Petitpierre, E. & Gomez-Zurita, J. (2009) DNA barcoding insect-host plant associations. *Proceedings Of The Royal Society B-Biological Sciences*, **276**, 639–648.

Magurran, A. (1988) *Ecological diversity and its measurement.* Princeton University Press, Princeton, US.

Mao, C. & Colwell, R.K. (2005) Estimation of species richness: mixture models, the role of rare species, and inferential challenges. *Ecology*, **86**, 1143–1153.

Martinez, N.D. (1993) Effects of resolution on food web structure. *Oikos*, **66**, 403–412.

Martinez, N. (1991) Artifacts or attributes? Effects of resolution on food-web patterns in Little Rock Lake food web. *Ecological Monographs*, **61**, 367–392.

Maruyama, P.K., Vizentin-Bugoni, J., Oliveira, G.M., Oliveira, P.E. & Dalsgaard, B. (2014) Morphological and spatio-temporal mismatches shape a neotropical savanna plant-hummingbird network. *Biotropica*, **46**, 740–747.

Memmott, J., Alonso, D., Berlow, E., Dobson, A., Dunne, J.A., Solé, R.V. & Weitz, J.S. (2006) Biodiversity loss and ecological network structure. *Ecological Networks Linking Structure to Dynamics in Food Webs* (eds. M. Pascual & J.A. Dunne), pp. 325–347. Oxford University Press, Oxford, UK.

Moré, M., Amorim, F.W., Benitez-Vieyra, S., Medina, A.M., Sazima, M. & Cocucci, A.A. (2012) Armament Imbalances: Match and Mismatch in Plant-

Pollinator Traits of Highly Specialized Long-Spurred Orchids. *PLoS ONE*, **7**, e41878.

Nielsen, A. & Bascompte, J. (2007) Ecological networks, nestedness and sampling effort. *Journal of Ecology*, **95**, 1134–1141–1141.

Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H. & Jordano, P. (2011) Missing and forbidden links in mutualistic networks. *Proceedings Of The Royal Society B-Biological Sciences*, **278**, 725–732.

Olesen, J. & Jordano, P. (2002) Geographic patterns in plant-pollinator mutualistic networks. *Ecology*, **83**, 2416–2424.

Olito, C. & Fox, J.W. (2014) Species traits and abundances predict metrics of plant-pollinator network structure, but not pairwise interactions. *Oikos*, **124**, 428–436.

Ollerton, J. & Cranmer, L. (2002) Latitudinal trends in plant-pollinator interactions: are tropical plants more specialised? *Oikos*, **98**, 340–350.

Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N., Dulloo, E., Faith, D., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurtt, G., Jetz, W., Karp, D.S., Mcgeoch, M., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S., Turak, E., Walpole, M. & Wegmann, M. (2013) Essential biodiversity variables. *Science*, **339**, 277–278.

Pocock, M.J.O., Evans, D.M. & Memmott, J. (2012) The Robustness and Restoration of a Network of Ecological Networks. *Science*, **335**, 973–977.

Preston, F. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.

R Development Core Team (2010) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org, Vienna, Austria.

Rivera-Hutinel, A., Bustamante, R.O., Marín, V.H. & Medel, R. (2012) Effects of sampling completeness on the structure of plant-pollinator networks. *Ecology*, **93**, 1593–1603.

Schleuning, M., Frund, J., Klein, A.M., Abrahamczyk, S., Alarcón, R., Albrecht, M., Andersson, G.K.S., Bazarian, S., Böhning-Gaese, K., Bommarco, R., Dalsgaard, B., Dehling, D.M., Gotlieb, A., Hagen, M., Hickler, T., Holzschuh, A., Kaiser-Bunbury, C.N., Kreft, H., Morris, R.J., Sandel, B., Sutherland, W.J., Svenning, J.C., Tscharntke, T., Watts, S., Weiner, C.N., Werner, M., Williams, N.M., Winqvist, C., Dormann, C.F. & Blüthgen, N. (2012) Specialization of mutualistic interaction networks decreases toward tropical latitudes. *Current Biology*, **22**, 1925–1931.

Snow, B. & Snow, D. (1972) Feeding niches of hummingbirds in a Trinidad valley. *Journal of Animal Ecology*, **41**, 471–485.

Snow, B. & Snow, D. (1988) *Birds and berries.* Poyser, Calton, UK.

Stang, M., Klinkhamer, P., Waser, N.M., Stang, I. & van der Meijden, E. (2009) Size-specific interaction patterns and size matching in a plant-pollinator interaction web. *Annals Of Botany*, **103**, 1459–1469.

Strogatz, S. (2001) Exploring complex networks. *Nature*, **410**, 268–276.

Stumpf, M.P.H., Wiuf, C. & May, R.M. (2005) Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences USA*, **102**, 4221–4224.

Thébault, E. & Fontaine, C. (2010) Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, **329**, 853–856.

Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A., Galetti, M., García, M.B., García, D., Gomez, J.M., Jordano, P., Medel, R., Navarro, L., Obeso, J.R., Oviedo, R., Ramírez, N., Rey, P.J., Traveset, A., Verdú, M. & Zamora, R. (2014) Beyond species loss: the extinction of ecological interactions in a changing world. *Functional Ecology*, **29**, 299–307.

Vazquez, D.P., Chacoff, N.P. & Cagnolo, L. (2009) Evaluating multiple determinants of the structure of plant-animal mutualistic networks. *Ecology*, **90**, 2039–2046.

Vazquez, D.P., Ramos-Jiliberto, R., Urbani, P. & Valdovinos, F.S. (2015) A conceptual framework for studying the strength of plant-animal mutualistic interactions. *Ecology Letters*, **18**, 385–400.

Vazquez, D., Morris, W. & Jordano, P. (2005) Interaction frequency as a surrogate for the total effect of animal mutualists on plants. *Ecology Letters*, **8**, 1088–1094.

Vizentin-Bugoni, J., Maruyama, P.K. & Sazima, M. (2014) Processes entangling interactions in communities: forbidden links are more important than abundance in a hummingbird-plant network. *Proceedings Of The Royal Society B-Biological Sciences*, **281**, 20132397–20132397.

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

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

Table 1:

Link type	Formulation	Definition
Potential links	$I_{max} = AP$	Size of network matrix, i.e. maximum number of potentially observable interactions; A and P, numbers of interacting animal and plant species, respectively.
Observed links	I	Total number of observed links in the network given a sufficient sampling effort. Number of ones in the adjacency matrix.
Unobserved links	$UL = I_{max} - I$	Number of zeroes in the adjacency matrix.
Forbidden links	FL	Number of links, which remain unobserved because of linkage constraints, irrespectively of sufficient sampling effort.
Missing links	$ML = AP - I - FL$	Number of links, which may exist in nature but need more sampling effort and/or additional sampling methods to be observed.

Table 2:

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

[a], Lack of accessibility due to habitat uncoupling, i.e., canopy-foraging species vs. understory spec

[b], Colour restrictions, and reward per flower too small relative to the size of the bird.

Table 3:

Interaction	Sample 1	Sample 2	Sample 3	...	Sample i
A1 - P2	12	2	0	...	6
A1 - P2	0	0	0	...	1
...	...	...	...	...	...
A5 - P3	5	0	1	...	18
A5 - P4	1	0	1	...	3
...	...	...	...	...	...
A_i - P_i	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 ± 70.9	231.4 ± 14.2	509.6 ± 54.7
ACE	240.3 ± 8.9	241.3 ± 7.9	566.1 ± 14.8
Scaled Chao	195.4	162.7	308.4
CI	[124.5–266.3]	[148.5–176.9]	[253.6–363.1]
Scaled ACE	178.5	169.7	342.6
CI	[169.5–187.4]	[161.8–177.6]	[327.8–357.4]
% *unobserved*[a]	8.33	15.38	47.8

[a], estimated with library Jade (R Core Development Team 2010,Chao et al. 2015)

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 Δ 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).

Figure 1:

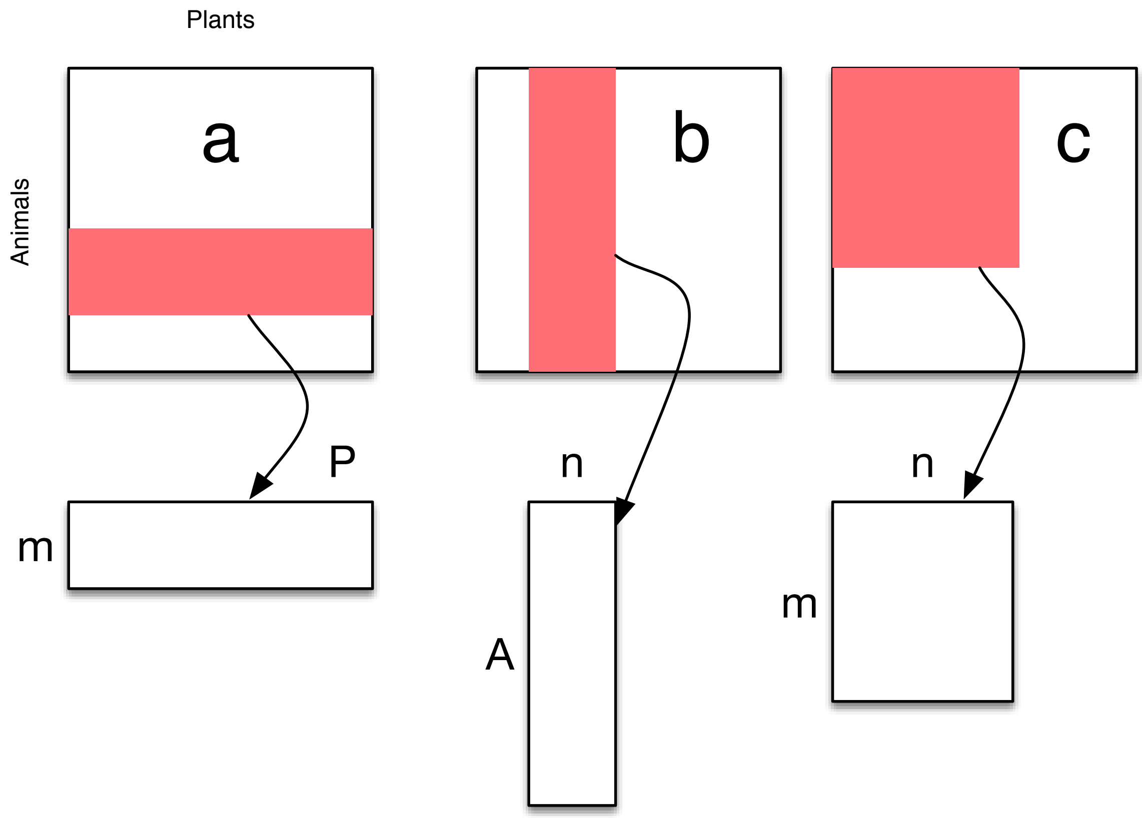

Figure 2: