

# 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

Sevilla, September 4, 2015

## 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 different 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.

---

<sup>\*</sup>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, i.e., life-history restrictions. Thus, the number of observed interactions  $I$  in robustly sampled networks is typically  $I \ll I_{max}$ , resulting in extremely sparse interaction matrices with low connectance.
4. Reasons for forbidden links are multiple but mainly stem from spatial and temporal uncoupling, 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 “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 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, Mao & Chang, 2004; Hortal, Borges & Gaspar, 2006; Colwell, 2009; Gotelli  
8 & Colwell, 2011; Chao *et al.*, 2014). Recent effort has been also focused at defining  
9 essential biodiversity variables (EBV) (Pereira *et al.*, 2013) that can be sampled  
10 and measured repeatedly to complement biodiversity estimates. Yet sampling  
11 species or taxa-specific EBVs is just probing a single component of biodiversity;  
12 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 problems associated with sampling species diversity, even worse. We 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 we provide a succinct overview of previous work addressing sampling issues for ecological interaction networks. In the second part, after a short overview of asymptotic diversity estimates (Gotelli & Colwell, 2001), we discuss specific rationales for sampling the biodiversity of ecological interactions. Most of my examples come from the analysis of plant-animal interaction networks, yet are applicable to other types of species-species 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, 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 (see Table 2 in 2004; 2009). Monitoring interactions is a biodiversity sampling and is subject to similar methodological shortcomings,

38 especially under-sampling (Jordano, 1987; Jordano, Vázquez & Bascompte, 2009;  
 39 Coddington *et al.*, 2009; Vazquez, Chacoff & Cagnolo, 2009; Dorado *et al.*, 2011;  
 40 Rivera-Hutinel *et al.*, 2012). For example, when we study mutualistic networks,  
 41 our goal is to make an inventory of the distinct pairwise interactions that made  
 42 up the network. We are interested in having a complete list of all the pairwise  
 43 interactions among species (e.g., all the distinct, species-species interactions, or  
 44 links, among the pollinators and flowering plants) that do actually exist in a given  
 45 community. Sampling these interactions thus entails exactly the same problems,  
 46 limitations, constraints, and potential biases as sampling individual organisms  
 47 and species diversity. As Mao & Colwell (2005) put it, these are the workings of  
 48 Preston’s demon, the moving “veil line” (Preston, 1948) between the detected and  
 49 the undetected interactions as sample size increases.

50 Early efforts to recognize and solve sampling problems in analyses of interac-  
 51 tions stem from research on food webs and to determine how undersampling biases  
 52 food web metrics (Martinez, 1991; Cohen *et al.*, 1993; Martinez, 1993; Bersier,  
 53 Banasek-Richter & Cattin, 2002; Brose, Martinez & Williams, 2003; Banasek-  
 54 Richter, Cattin & Bersier, 2004; Wells & O’Hara, 2012). In addition, the myriad  
 55 of classic natural history studies documenting animal diets, host-pathogen infection  
 56 records, plant herbivory records, etc., represent efforts to document interactions  
 57 occurring in nature. All of them share the problem of sampling incompleteness in-  
 58 fluencing the patterns and metrics reported. Yet, despite the early recognition that  
 59 incomplete sampling may seriously bias the analysis of ecological networks (Jor-  
 60 dano, 1987), only recent studies have explicitly acknowledged it and attempted to  
 61 determine its influence (Ollerton & Cranmer, 2002; Nielsen & Bascompte, 2007;  
 62 Vazquez, Chacoff & Cagnolo, 2009; Gibson *et al.*, 2011; Olesen *et al.*, 2011; Chacoff

63 *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox, 2014; Bascompte & Jordano,  
 64 2014; Vizentin-Bugoni, Maruyama & Sazima, 2014; Frund, McCann & Williams,  
 65 2015). The sampling approaches have been extended to predict patterns of coex-  
 66 tintions in interaction assemblages (e.g., hosts-parasites) (Colwell, Dunn & Harris,  
 67 2012). Most empirical studies provide no estimate of sampling effort, implicitly  
 68 assuming that the reported network patterns and metrics are robust. Yet recent ev-  
 69 idences point out that number of partner species detected, number of actual links,  
 70 and some aggregate statistics describing network patterns, are prone to sampling  
 71 bias (Nielsen & Bascompte, 2007; Dorado *et al.*, 2011; Olesen *et al.*, 2011; Chacoff  
 72 *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox, 2014; Frund, McCann &  
 73 Williams, 2015). Most of these evidences, however, come either from simulation  
 74 studies (Frund, McCann & Williams, 2015) or from relatively species-poor assem-  
 75 blages. Most certainly, sampling limitations pervade biodiversity inventories in  
 76 tropical areas (Coddington *et al.*, 2009) and we might rightly expect that frequent  
 77 interactions may be over-represented and rare interactions may be missed entirely  
 78 in studies of mega-diverse assemblages (Bascompte & Jordano, 2014); but, to what  
 79 extent?

## 80 Sampling interactions: methods

81 When we sample interactions in the field we record the presence of two species that  
 82 interact in some way. For example, Snow and Snow(1988) recorded an interaction  
 83 whenever they saw a bird “touching” a fruit on a plant. We observe and record  
 84 feeding observations, visitation, occupancy, presence in pollen loads or in fecal  
 85 samples, etc., of *individual* animals or plants and accumulate pairwise interactions,

186 i.e., lists of species partners and the frequencies with which we observe them.  
 187 Therefore, estimating the sampling completeness of pairwise interactions for a  
 188 whole network, requires some gauging of the sampling completeness (i.e., how the  
 189 number (richness) of distinct pairwise interactions accumulates as sampling effort  
 190 is increased) and/or estimating the uncertainty around the missed links (Wells &  
 191 O’Hara, 2012).

192 Most types of ecological interactions can be illustrated with bipartite graphs,  
 193 with two or more distinct groups of interacting partners (Bascompte & Jordano,  
 194 2014); for illustration purposes I’ll focus more specifically on plant-animal inter-  
 195 actions. Sampling interactions requires filling the cells of an interaction matrix  
 196 with data. The matrix,  $\Delta = AP$  (the adjacency matrix for the graph representa-  
 197 tion of the network), is a 2D inventory of the interactions among, say,  $A$  animal  
 198 species (rows) and  $P$  plant species (columns) (Jordano, 1987; Bascompte & Jor-  
 199 dano, 2014). The matrix entries illustrate the values of the pairwise interactions  
 200 visualized in the  $\Delta$  matrix, and can be 0 or 1, for presence-absence of a given  
 201 pairwise interaction, or take a quantitative weight  $w_{ji}$  to represent the interaction  
 202 intensity or unidirectional effect of species  $j$  on species  $i$  (Bascompte & Jordano,  
 203 2014; Vazquez *et al.*, 2015). The outcomes of most ecological interactions are  
 204 dependent on frequency of encounters (e.g., visit rate of pollinators, number of  
 205 records of ant defenders, frequency of seeds in fecal samples). Thus, a frequently  
 206 used proxy for interaction intensities  $w_{ji}$  is just how frequent new interspecific  
 207 encounters are, whether or not appropriately weighted to estimate interaction ef-  
 208 fectiveness (Vazquez, Morris & Jordano, 2005).

209 We need to define two basic steps in the sampling of interactions: 1) which  
 210 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 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 (parasitoid) or plant-herbivore interactions are “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 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).

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. Connectance increased 1.43-fold, mean plant connec-

161 tivity went from 18.5 to 26.4, and mean pollinator connectivity from 2.9 to 4.1;  
 162 moreover, extreme specialist pollinator species (singletons in the adjacency matrix)  
 163 decreased 0.6-fold. (Olesen *et al.* 2011) identified pollen loads on sampled insects  
 164 and added the new links to an observation-based visitation matrix, with an extra  
 165 5% of links representing the estimated number of missing links in the pollination  
 166 network. The overlap between observational and pollen-load recorded links was  
 167 only 33%, underscoring the value of combining methodological approaches. Zoo-  
 168 centric sampling has recently been extended with the use of DNA-barcoding, for  
 169 example with plant-herbivore (Jurado-Rivera *et al.*, 2009), host-parasitoid (Wirta  
 170 *et al.*, 2014), and plant-frugivore interactions (González-Varo, Arroyo & Jordano,  
 171 2014). For mutualistic networks we would expect that zoo-centric sampling could  
 172 help unveiling interactions of the animals with rare plant species or for relatively  
 173 common plants species which are difficult to sample by direct observation. Fu-  
 174 ture methodological work may provide significant advances showing how mixing  
 175 different sampling strategies strengthens the completeness of network data. These  
 176 mixed strategies may combine, for instance, timed watches at focal plants, spot  
 177 censuses along walked transects, pollen load or seed contents analyses, monitoring  
 178 with camera traps, and DNA barcoding records. We might expect increased power  
 179 of these mixed sampling approaches when combining different methods from both  
 180 phyto- and zoo-centric perspectives (Bosch *et al.*, 2009; Bluthgen, 2010). Note also  
 181 that the different methods could be applied in different combinations to the two  
 182 distinct sets of species. However, there are no tested protocols and/or sampling  
 183 designs for ecological interaction studies to suggest an optimum combination of  
 184 approaches. Ideally, pilot studies would provide adequate information for each  
 185 specific study setting.

## 186 Sampling interactions: rationale

187 The number of distinct pairwise interactions that we can record in a landscape  
 188 (an area of relatively homogeneous vegetation, analogous to the one we would  
 189 use to monitor species diversity) is equivalent to the number of distinct classes in  
 190 which we can classify the recorded encounters among *individuals* of two different  
 191 species. Yet, individual-based interaction networks have been only recently studied  
 192 (Dupont, Trøjelsgaard & Olesen, 2011; Wells & O'Hara, 2012). The most usual  
 193 approach has been to pool individual-based interaction data into species-based  
 194 summaries, an approach that ignores the fact that only a fraction of individuals  
 195 may actually interact given a per capita interaction effect (Wells & O'Hara, 2012).  
 196 Wells & O'Hara (2012) illustrate the pros and cons of the approach. We walk in  
 197 the forest and see a blackbird  $Tm$  picking an ivy  $Hh$  fruit and ingesting it: we  
 198 have a record for  $Tm - Hh$  interaction. We keep advancing and record again a  
 199 blackbird feeding on hawthorn  $Cm$  fruits so we record a  $Tm - Cm$  interaction;  
 200 as we advance we encounter another ivy plant and record a blackcap swallowing a  
 201 fruit so we now have a new  $Sa - Hh$  interaction, and so on. At the end we have  
 202 a series of classes (e.g.,  $Sa - Hh$ ,  $Tm - Hh$ ,  $Tm - Cm$ , etc.), along with their  
 203 observed frequencies. Bunge & Fitzpatrick (1993) provide an early review of the  
 204 main aspects and approaches to estimate the number of distinct classes  $C$  in a  
 205 sample of observations.

206 Our sampling above would have resulted in a vector  $n = [n_1 \dots n_C]'$  where  $n_i$  is  
 207 the number of records in the  $i^{th}$  class. As stressed by Bunge & Fitzpatrick (1993),  
 208 however, the  $i^{th}$  class would appear in the sample if and only if  $n_i > 0$ , and we  
 209 don't know *a priori* which  $n_i$  are zero. So,  $n$  is not observable. Rather, what we

get is a vector  $c = [c_1 \dots c_n]'$  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 replicated 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 the most robust statistically, as it takes account of heterogeneity among days (Colwell, Mao & Chang, 2004; Colwell, Dunn & Harris, 2012; Chao *et al.*, 2014). Thus, both presence-absence and weighted information on interactions can be accommodated for this purpose.

## The species assemblage

Estimating the number of interactions with resulting robust estimates of network parameters is a central issue in the study of ecological networks (Jordano, 1987; Bascompte & Jordano, 2014). When we consider an observed and recorded sample

233 of interactions on a particular assemblage of  $A_{obs}$  and  $P_{obs}$  species (or a set of  
 234 replicated samples) as a reference sample (Chao *et al.*, 2014) we may have three  
 235 sources of undersampling error that are ignored by treating a reference sample as a  
 236 true representation of the interactions in well-defined assemblage: 1) some animal  
 237 species are actually present but not observed (zero abundance or incidence in the  
 238 interactions in the reference sample),  $A_0$ ; 2) some plant species are actually present  
 239 but not observed (zero abundance or incidence in the interactions in the reference  
 240 sample),  $P_0$ ; 3) some unobserved links (the zeroes in the adjacency matrix,  $UL$ )  
 241 may actually occur but not recorded. Thus a first problem is determining if  $A_{obs}$   
 242 and  $P_{obs}$  truly represent the actual species richness interacting in the assemblage.  
 243 To this end we might use the replicated reference samples to estimate the true  
 244 number of interacting animal  $A_{est}$  and plant  $P_{est}$  species as in traditional diversity  
 245 estimation analysis (Chao *et al.*, 2014). If there are no uniques (species seen on  
 246 only one day), then  $A_0$  and  $P_0$  will be zero, and we have  $A_{obs}$  and  $P_{obs}$  as robust  
 247 estimates of the actual species richness of the assemblage. If  $A_0$  and  $P_0$  are not  
 248 zero they estimate the minimum number of undetected animal and plant species  
 249 that can be expected with a sufficiently large number of replicates, taken from the  
 250 same assemblage/locality by the same methods in the same time period. We can  
 251 use extrapolation methods (Colwell, Dunn & Harris, 2012) to estimate how many  
 252 additional replicate surveys it would take to reach a specified proportion  $g$  of  $A_{est}$   
 253 and  $P_{est}$ .

## 254 The interactions

255 We are then faced with assessing the sampling of interactions  $I$ . Table 1 sum-  
 256 marizes the main components and targets for estimation of interaction richness.  
 257 In contrast with traditional species diversity estimates, sampling networks has  
 258 the paradox that despite the potentially interacting species being present in the  
 259 sampled assemblage (i.e., included in the  $A_{obs}$  and  $P_{obs}$  species lists), some of  
 260 their pairwise interactions are impossible to be recorded. The reason is forbidden  
 261 links. Independently of whether we sample full communities or subset communi-  
 262 ties we face a problem: some of the interactions that we can visualize in the empty  
 263 adjacency matrix  $\Delta$  will simply not occur. With a total of  $A_{obs}P_{obs}$  “potential” in-  
 264 teractions (eventually augmented to  $A_{est}P_{est}$  in case we have undetected species),  
 265 a fraction of them are impossible to record, because they are forbidden (Jordano,  
 266 Bascompte & Olesen, 2003; Olesen *et al.*, 2011).

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

274 Forbidden links are non-occurrences of pairwise interactions that can be ac-  
 275 counted for by biological constraints, such as spatio-temporal uncoupling (Jordano,  
 276 1987), size or reward mismatching, foraging constraints (e.g., accessibility) (Moré  
 277 *et al.*, 2012), and physiological-biochemical constraints (Jordano, 1987). We still

278 have extremely reduced information about the frequency of forbidden links in natu-  
 279 ral communities (Jordano, Bascompte & Olesen, 2003; Stang *et al.*, 2009; Vazquez,  
 280 Chacoff & Cagnolo, 2009; Olesen *et al.*, 2011; Ibanez, 2012; Maruyama *et al.*, 2014;  
 281 Vizentin-Bugoni, Maruyama & Sazima, 2014) (Table 1). Forbidden links are thus  
 282 represented as structural zeroes in the interaction matrix, i.e., matrix cells that  
 283 cannot get a non-zero value. [ROB] UL\*: unobserved links between the [Aest =  
 284 Aobs + A0)] \* [Pest = Pobs + P0] interacting species. The total number of cells in  
 285 the augmented adjacency matrix is thus Aest \* Pest. (2) Thus, UL\* is a mixture  
 286 of unobserved links between A and P (your UL), and unobserved links that involve  
 287 unobserved A or unobserved P (or both). [ROB]

288 We might expect different types of  $FL$  to occupy different parts of the  $\Delta$  ma-  
 289 trix, with missing cells due to phenological uncoupling,  $FL_P$ , largely distributed  
 290 in the lower-right half  $\Delta$  matrix and actually missed links  $ML$  distributed in its  
 291 central part (Olesen *et al.*, 2010). Yet, most of these aspects remain understud-  
 292 ied. Therefore, we need to account for the frequency of these structural zeros in  
 293 our matrix before proceeding. For example, most measurements of connectance  
 294  $C = I/(AP)$  implicitly ignore the fact that by taking the full product  $AP$  in the  
 295 denominator they are underestimating the actual connectance value, i.e., the frac-  
 296 tion of actual interactions  $I$  relative to the *biologically possible* ones, not to the  
 297 total maximum  $I_{max} = AP$ .

298 Our main problem then turns to estimate the number of true missed links,  
 299 i.e., those that can't be accounted for by biological constraints and that might  
 300 suggest undersampling. Thus, the sampling of interactions in nature, as the sam-  
 301 pling of species, is a cumulative process. In our analysis, we are not re-sampling  
 302 individuals, but interactions, so we made interaction-based accumulation curves.

303 If an interaction-based curve suggests a robust sampling, it does mean that no  
 304 new interactions are likely to be recorded, irrespectively of the species, as it is  
 305 a whole-network sampling approach (N. Gotelli, pers. com.). We add new, dis-  
 306 tinct, interactions recorded as we increase sampling effort (Fig. 2). We can obtain  
 307 an Interaction Accumulation Curve (*IAC*) analogous to a Species Curve (*SAC*)  
 308 (see Supplementary Online Material): the observed number of distinct pairwise  
 309 interactions in a survey or collection as a function of the accumulated number of  
 310 observations or samples (Colwell, 2009).

## 311 0.1 Empirical data on Forbidden Links

312 Adjacency matrices are frequently sparse, i.e., they are densely populated with  
 313 zeroes, with a fraction of them being structural (unobservable interactions) (Bas-  
 314 compte & Jordano, 2014). Thus, it would be a serious interpretation error to  
 315 attribute the sparseness of adjacency matrices for bipartite networks to undersam-  
 316 pling. The actual typology of link types in ecological interaction networks is thus  
 317 more complex than just the two categories of observed and unobserved interactions  
 318 (Table 1). Unobserved interactions are represented by zeroes and belong to two  
 319 categories. Missing interactions may actually exist but require additional sampling  
 320 or a variety of methods to be observed. Forbidden links, on the other hand, arise  
 321 due to biological constraints limiting interactions and remain unobservable in na-  
 322 ture, irrespectively of sampling effort (Table 1). Forbidden links *FL* may actually  
 323 account for a relatively large fraction of unobserved interactions *UL* when sam-  
 324 pling taxonomically-restricted subnetworks (e.g., plant-hummingbird pollination  
 325 networks) (Table 1). Phenological uncoupling 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 increase sampling effort. For instance, habitat use patterns by hummingbirds 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 2 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. For example, aside from  $FL$ , some pairwise interactions may simply 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 infor-

351 mation on the participant species that allows the inference of biological constraints  
 352 imposing forbidden links, so that structural zeroes can be identified in the adjacency  
 353 matrix. Second, a critical analysis of sampling robustness and a robust  
 354 estimate of the actual fraction of missing links,  $M$ , resulting in a robust estimate  
 355 of  $I$ . In the next sections I explore these elements of inference, using *IACs* to  
 356 assess the robustness of interaction sampling.

## 357 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 or use any other type of direct or indirect recording method, 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 (represented by just one individual),  $S_2$  is the number of doubleton species (species with two individuals), etc. The total number of individuals in the sample would be:

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

358

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

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

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

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

364 [ROB] No. The unbiased form of Chao2 is identical in formation to Chao1 (above),  
 365 but instead of counts it uses incidence frequencies among samples (usually Q in-  
 366 stead of S). [ROB]

367 A plot of the cumulative number of species recorded,  $S_n$ , as a function of some  
 368 measure of sampling effort (say,  $n$  samples taken) yields the species accumulation  
 369 curve (SAC) or collector's curve (Colwell & Coddington, 1994). Similarly, inter-  
 370 action accumulation curves (IAC), analogous to SACs, can be used to assess the  
 371 robustness of interactions sampling for plant-animal community datasets (Jordano,  
 372 1987; Jordano, Vázquez & Bascompte, 2009; Olesen *et al.*, 2011), as discussed in  
 373 the next section.

## 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 (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, Vázquez & Bascompte, 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) and the 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, 2010) (Jordano, Vázquez & Bascompte, 2009; Olesen *et al.*, 2011).

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.

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) (see Supplementary Online Material). It should be noted that the asymptotic estimate of interaction richness explicitly ignores the fact that, due to forbidden links, a number of pairwise interactions among the  $I_{max}$  number specified in the adjacency matrix  $\Delta$  cannot be recorded, irrespective of sampling effort. We may expect undersampling specially in moderate to large size 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, Maruyama & Sazima, 2014), typically with more homogeneous subsets of species (e.g., bumblebee-pollinated flowers). In any case the sparseness of the  $\Delta$  matrix is by no means an indication of undersampling the issue of structural zeroes in the interaction matrices is effectively incorporated in the estimates.

For example, mixture models incorporating detectabilities have been proposed to effectively account for rare species (Mao & Colwell, 2005). In an analogous line,

423 mixture models could be extended to samples of pairwise interactions, also with  
 424 specific detectability values. These detection rate/odds could be variable among  
 425 groups of interactions, depending on their specific detectability. For example,  
 426 detectability of flower-pollinator interactions involving bumblebees could have a  
 427 higher detectability than flower-pollinator pairwise interactions involving, say, ni-  
 428 tidulid beetles. These more homogeneous groupings of pairwise interactions within  
 429 a network define modules (Bascompte & Jordano, 2014), so we might expect that  
 430 interactions of a given module (e.g., plants and their hummingbird pollinators; Fig.  
 431 1a) may share similar detectability values, in an analogous way to species groups  
 432 receiving homogeneous detectability values in mixture models (Mao & Colwell,  
 433 2005). In its simplest form, this would result in a sample with multiple pairwise  
 434 interactions detected, in which the number of interaction events recorded for each  
 435 distinct interaction found in the sample is recorded (i.e., a column vector in Table  
 436 3, corresponding to, say, a sampling day). The number of interactions recorded for  
 437 the  $i_{th}$  pairwise interaction (i.e.,  $A_i P_j$  in Table 3),  $Y_i$  could be treated as a Poisson  
 438 random variable with a mean parameter  $\lambda_i$ , its detection rate. Mixture models  
 439 (Mao & Colwell, 2005) include estimates for abundance-based data (their analogs  
 440 in interaction sampling would be weighted data), where  $Y_i$  is a Poisson random  
 441 variable with detection rate  $\lambda_i$ . This is combined with the incidence-based model,  
 442 where  $Y_i$  is a binomial random variable (their analogous in interaction sampling  
 443 would be presence/absence records of interactions) with detection odds  $\lambda_i$ . Let  
 444  $T$  be the number of samples in an incidence-based data set. A Poisson/binomial  
 445 density can be written as (Mao & Colwell, 2005):

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

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

447 The detection rates  $\lambda_i$  depend on the relative abundances  $\phi_i$  of the interactions,  
 448 the probability of a pairwise interaction being detected when it is present, and the  
 449 sample size (the number of interactions recorded), which, in turn, is a function  
 450 of the sampling effort. Unfortunately, no specific sampling model has been de-  
 451 veloped along these lines for species interactions and their characteristic features.  
 452 For example, a complication factor might be that interaction abundances,  $\phi_i$ , in  
 453 real assemblages are a function of the abundances of interacting species that de-  
 454 termine interspecific encounter rates; yet they also depend on biological factors  
 455 that ultimately determine if the interaction occurs when the partner species are  
 456 present. For example,  $\lambda_i$  should be set to zero for all  $FL$ . In its simplest form,  $\phi_i$   
 457 could be estimated from just the product of partner species abundances, an ap-  
 458 proach recently used as a null model to assess the role of biological constraints in  
 459 generating forbidden links and explaining interaction patterns (Vizentin-Bugoni,  
 460 Maruyama & Sazima, 2014). Yet more complex models (e.g., Wells & O'hara  
 461 2012) should incorporate not only interspecific encounter probabilities, but also  
 462 interaction detectabilities, phenotypic matching and incidence of forbidden links.  
 463 Mixture models are certainly complex and for most situations of evaluating sam-  
 464 pling effort better alternatives include the simpler incidence-based rarefaction and  
 465 extrapolation (Colwell, Dunn & Harris, 2012; Chao *et al.*, 2014).

## 466 The *real* missing links

467 Given that a fraction of unobserved interactions can be accounted for by for-  
 468 bidden links, what about the remaining missing interactions? We have already  
 469 discussed that some of these could still be related to unaccounted constraints, and  
 470 still others would be certainly attributable to insufficient sampling. Would this  
 471 always be the case? Multispecific assemblages of distinct taxonomic relatedness,  
 472 whose interactions can be represented as bipartite networks (e.g., host-parasite,  
 473 plant-animal mutualisms, plant-herbivore interactions- with two distinct sets of  
 474 unrelated higher taxa), are shaped by interspecific encounters among individuals  
 475 of the partner species (Fig. 2). A crucial ecological aspect limiting these inter-  
 476 actions is the probability of interspecific encounter, i.e., the probability that two  
 477 individuals of the partner species actually encounter each other in nature.

478 Given log-normally distributed abundances of the two species groups, the ex-  
 479 pected probabilities of interspecific encounter (*PIE*) would be simply the product  
 480 of the two lognormal distributions. Thus, we might expect that for low *PIE* val-  
 481 ues, pairwise interactions would be either extremely difficult to sample, or just  
 482 simply not occurring in nature. Consider the Nava de las Correhuelas interaction  
 483 web (NCH, Table 2), with  $A = 36$ ,  $P = 25$ ,  $I = 181$ , and almost half of the unob-  
 484 served interactions not accounted for by forbidden links, thus  $M = 53.1\%$ . Given  
 485 the robust sampling of this network (Jordano, Vázquez & Bascompte, 2009), a  
 486 sizable fraction of these possible but missing links would be simply not occurring  
 487 in nature, most likely by extremely low *PIE*, in fact asymptotically zero. Given  
 488 the vectorized list of pairwise interactions for NCH, I computed the *PIE* values for  
 489 each one by multiplying element-wise the two species abundance distributions. The



490  $PIE_{max} = 0.0597$ , being a neutral estimate, based on the assumption that interac-  
 491 tions occur in proportion to the species-specific local abundances. With  $PIE_{median}$   
 492  $< 1.4 \cdot 10^{-4}$  we may safely expect (note the quantile estimate  $Q_{75\%} = 3.27 \cdot 10^{-4}$ )  
 493 that a sizable fraction of these missing interactions may not occur according to  
 494 this neutral expectation (Jordano, 1987; Olesen *et al.*, 2011) (neutral forbidden  
 495 links, *sensu* Canard *et al.*, 2012).

496 When we consider the vectorized interaction matrix, enumerating all pairwise  
 497 interactions for the  $AP$  combinations, the expected probabilities of finding a given  
 498 interaction can be estimated with a Good-Turing approximation (Good, 1953).  
 499 The technique, developed by Alan Turing and I.J. Good with applications to lin-  
 500 guistics and word analysis (Gale & Sampson, 1995) has been recently extended in  
 501 novel ways for ecological analyses (Chao *et al.*, 2015). It estimates the probability  
 502 of recording an interaction of a hitherto unseen pair of partners, given a set of past  
 503 records of interactions between other species pairs. Let a sample of  $N$  interactions  
 504 so that  $n_r$  distinct pairwise interactions have exactly  $r$  records. All Good-Turing  
 505 estimators obtain the underlying frequencies of events as:

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

506 where  $X$  is the pairwise interaction,  $N_X$  is the number of times interaction  $X$   
 507 is recorded,  $T$  is the sample size (number of distinct interactions recorded) and  
 508  $E(1)$  is an estimate of how many different interactions were recorded exactly once.  
 509 Strictly speaking Equation (1) gives the probability that the next interaction type  
 510 recorded will be  $X$ , after sampling a given assemblage of interacting species. In  
 511 other words, we scale down the maximum-likelihood estimator  $\frac{n}{T}$  by a factor of

512  $\frac{1-E(1)}{T}$ . This reduces all the probabilities for interactions we have recorded, and  
 513 makes room for interactions we haven't seen. If we sum over the interactions we  
 514 have seen, then the sum of  $P(X)$  is  $1 - \frac{1-E(1)}{T}$ . Because probabilities sum to one,  
 515 we have the left-over probability of  $P_{new} = \frac{E(1)}{T}$  of seeing something new, where  
 516 new means that we sample a new pairwise interaction. Note, however, that Good-  
 517 Turing estimators, the traditional asymptotic estimators, do not account in our  
 518 case for the forbidden interactions.

## 519 Discussion

520 Recent work has inferred that most data available for interaction networks are  
 521 incomplete due to undersampling, resulting in a variety of biased parameters and  
 522 network patterns (Chacoff *et al.*, 2012). It is important to note, however, that  
 523 in practice, many surveyed networks to date have been subnets of much larger  
 524 networks. This is true for protein interaction, gene regulation, and metabolic  
 525 networks, where only a subset of the molecular entities in a cell have been sam-  
 526 pled (Stumpf, Wiuf & May, 2005). Despite recent attempts to document whole  
 527 ecosystem meta-networks (Pocock, Evans & Memmott, 2012), it is likely that most  
 528 ecological interaction networks will illustrate just major ecosystem compartments.  
 529 Due to their high generalization, high temporal and spatial turnover, and high  
 530 complexity of association patterns, adequate sampling of ecological interaction  
 531 networks requires extremely large sampling effort. Undersampling of ecological  
 532 networks may originate from the analysis of assemblage subsets (e.g., taxonomi-  
 533 cally or functionally defined), and/or from logistically-limited sampling effort. It is  
 534 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. 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. 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 urgently need robust techniques to assess the completeness of ecological interactions networks because this knowledge will allow the identification of the minimal components of their ecological complexity that need to be restored to rebuild functional ecosystems after perturbations.

## Acknowledgements

I am indebted to Jens M. Olesen, Alfredo Valido, Jordi Bascompte, Thomas Lewinshon, John N. Thompson, Nick Gotelli, Carsten Dormann, and Paulo R. Guimarães Jr. for useful and thoughtful discussion at different stages of this manuscript. Jeferson Vizentin-Bugoni kindly helped with the Sta Virgínia data. Jens M. Olesen kindly made available the Grundvad dataset; together with Robert K. Colwell, Néstor Pérez-Méndez, JuanPe González-Varo, and Paco Rodríguez provided most useful comments to a final version of the ms. Robert Colwell shared a number of crucial suggestions that clarified my vision of sampling ecological in-

teractions. 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 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/pedrojm/MS\\_Network-Sampling](https://github.com/pedrojm/MS_Network-Sampling)), with DOI: 10.5281/zenodo.29437.

## 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: Sen-  
sitivity 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., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Elli-  
son, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework  
for sampling and estimation in species diversity studies. *Ecological Monographs*,  
**84**, 45–67.

- 625 Chao, A., Hsieh, T.C., Chazdon, R.L., Colwell, R.K. & Gotelli, N.J. (2015) Un-  
626 veiling the species-rank abundance distribution by generalizing the Good-Turing  
627 sample coverage theory. *Ecology*, **96**, 1189–1201.
- 628 Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009)  
629 Undersampling bias: the null hypothesis for singleton species in tropical arthro-  
630 pod surveys. *Journal of Animal Ecology*, **78**, 573–584.
- 631 Cohen, J.E. (1978) *Food webs and niche space*. Princeton University Press, Prince-  
632 ton, New Jersey, US.
- 633 Cohen, J.E., Beaver, R.A., Cousins, S.H., DeAngelis, D.L., Goldwasser, L., Heong,  
634 K.L., Holt, R.D., Kohn, A.J., Lawton, J.H., Martinez, N., O'Malley, R., Page,  
635 L.M., Patten, B.C., Pimm, S.L., Polis, G., Rejmanek, M., Schoener, T.W.,  
636 Schenly, K., Sprules, W.G., Teal, J.M., Ulanowicz, R., Warren, P.H., Wilbur,  
637 H.M. & Yodis, P. (1993) Improving food webs. *Ecology*, **74**, 252–258.
- 638 Colwell, R. & Coddington, J. (1994) Estimating terrestrial biodiversity through ex-  
639 trapolation. *Philosophical Transactions Of The Royal Society Of London Series*  
640 *B-Biological Sciences*, **345**, 101–118.
- 641 Colwell, R.K. (2009) Biodiversity: concepts, patterns, and measurement. *The*  
642 *Princeton Guide to Ecology* (ed. S.A. Levin), pp. 257–263. Princeton University  
643 Press, Princeton.
- 644 Colwell, R.K. (2013) EstimateS: Biodiversity Estimation. -, pp. 1–33.
- 645 Colwell, R.K., Dunn, R.R. & Harris, N.C. (2012) Coextinction and persistence of

- 646 dependent species in a changing world. *Annual Review of Ecology Evolution and*  
647 *Systematics*, **43**, 183–203.
- 648 Colwell, R.K., Mao, C.X. & Chang, J. (2004) Interpolating, extrapolating, and  
649 comparing incidence-based species accumulation curves. *Ecology*, **85**, 2717–2727.
- 650 Dorado, J., Vazquez, D.P., Stevani, E.L. & Chacoff, N.P. (2011) Rareness and  
651 specialization in plant-pollinator networks. *Ecology*, **92**, 19–25.
- 652 Dormann, C.F., Frund, J., Bluthgen, N. & Gruber, B. (2009) Indices, graphs and  
653 null models: Analyzing bipartite ecological networks. *Open Ecology Journal*, **2**,  
654 7–24.
- 655 Dupont, Y.L., Trøjelsgaard, K. & Olesen, J.M. (2011) Scaling down from species  
656 to individuals: a flower-visitation network between individual honeybees and  
657 thistle plants. *Oikos*, **120**, 170–177.
- 658 Dupont, Y.L., Trøjelsgaard, K., Hagen, M., Henriksen, M.V., Olesen, J.M., Ped-  
659 ersen, N.M.E. & Kissling, W.D. (2014) Spatial structure of an individual-based  
660 plant-pollinator network. *Oikos*, **123**, 1301–1310.
- 661 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P.,  
662 Dalsgaard, B., de Sassi, C., Galetti, M., Guimaraes, P.R., Lomáscolo, S.B.,  
663 Martín González, A.M., Pizo, M.A., Rader, R., Rodrigo, A., Tylianakis, J.M.,  
664 Vazquez, D.P. & Allesina, S. (2013) The dimensionality of ecological networks.  
665 *Ecology Letters*, **16**, 577–583.
- 666 Elberling, H. & Olesen, J.M. (1999) The structure of a high latitude plant-flower  
667 visitor system: the dominance of flies. *Ecography*, **22**, 314–323.



- 668 Frund, J., McCann, K.S. & Williams, N.M. (2015) Sampling bias is a challenge  
 669 for quantifying specialization and network structure: lessons from a quantitative  
 670 niche model. *Oikos*, pp. n/a–n/a.
- 671 Gale, W.A. & Sampson, G. (1995) Good-Turing frequency estimation without  
 672 tears. *Journal of Quantitative Linguistics*, **2**, 217–237.
- 673 Gibson, R.H., Knott, B., Eberlein, T. & Memmott, J. (2011) Sampling method  
 674 influences the structure of plant–pollinator networks. *Oikos*, **120**, 822–831.
- 675 González-Varo, J.P., Arroyo, J.M. & Jordano, P. (2014) Who dispersed the seeds?  
 676 The use of DNA barcoding in frugivory and seed dispersal studies. *Methods in*  
 677 *Ecology and Evolution*, **5**, 806–814.
- 678 Good, I.J. (1953) The population frequencies of species and the estimation of  
 679 population parameters. *Biometrika*, **40**, 237–264.
- 680 Gotelli, N.J. & Colwell, R.K. (2011) Estimating species richness. *Biological Di-*  
 681 *versity Frontiers in Measurement and Assessment* (eds. A.E. Magurran & B.J.  
 682 McGill), pp. 39–54. Oxford University Press, Oxford, UK.
- 683 Gotelli, N. & Colwell, R. (2001) Quantifying biodiversity: procedures and pitfalls  
 684 in the measurement and comparison of species richness. *Ecology Letters*, **4**,  
 685 379–391.
- 686 Hortal, J., Borges, P. & Gaspar, C. (2006) Evaluating the performance of species  
 687 richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*,  
 688 **75**, 274–287.

- 689 Ibanez, S. (2012) Optimizing size thresholds in a plant–pollinator interaction web:  
690 towards a mechanistic understanding of ecological networks. *Oecologia*, **170**,  
691 233–242.
- 692 Jordano, P. (1987) Patterns of mutualistic interactions in pollination and seed dis-  
693 persal: connectance, dependence asymmetries, and coevolution. *The American*  
694 *Naturalist*, **129**, 657–677.
- 695 Jordano, P., Bascompte, J. & Olesen, J. (2003) Invariant properties in coevolu-  
696 tionary networks of plant-animal interactions. *Ecology Letters*, **6**, 69–81.
- 697 Jordano, P., Vázquez, D. & Bascompte, J. (2009) Redes complejas de interac-  
698 ciones planta—animal. *Ecología y evolución de interacciones planta-animal* (eds.  
699 R. Medel, R. Dirzo & R. Zamora), pp. 17–41. Editorial Universitaria, Santiago,  
700 Chile.
- 701 Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., Petitpierre, E. & Gomez-Zurita,  
702 J. (2009) DNA barcoding insect-host plant associations. *Proceedings Of The*  
703 *Royal Society B-Biological Sciences*, **276**, 639–648.
- 704 Magurran, A. (1988) *Ecological diversity and its measurement*. Princeton Univer-  
705 sity Press, Princeton, US.
- 706 Mao, C. & Colwell, R.K. (2005) Estimation of species richness: mixture models,  
707 the role of rare species, and inferential challenges. *Ecology*, **86**, 1143–1153.
- 708 Martinez, N.D. (1993) Effects of resolution on food web structure. *Oikos*, **66**,  
709 403–412.

- 710 Martinez, N. (1991) Artifacts or attributes? Effects of resolution on food-web  
711 patterns in Little Rock Lake food web. *Ecological Monographs*, **61**, 367–392.
- 712 Maruyama, P.K., Vizentin-Bugoni, J., Oliveira, G.M., Oliveira, P.E. & Dalsgaard,  
713 B. (2014) Morphological and spatio-temporal mismatches shape a neotropical  
714 savanna plant-hummingbird network. *Biotropica*, **46**, 740–747.
- 715 Moré, M., Amorim, F.W., Benitez-Vieyra, S., Medina, A.M., Sazima, M. &  
716 Cocucci, A.A. (2012) Armament Imbalances: Match and Mismatch in Plant-  
717 Pollinator Traits of Highly Specialized Long-Spurred Orchids. *PLoS ONE*, **7**,  
718 e41878.
- 719 Morris, R.J., Gripenberg, S., Lewis, O.T. & Roslin, T. (2013) Antagonistic inter-  
720 action networks are structured independently of latitude and host guild. *Ecology*  
721 *Letters*, **17**, 340–349.
- 722 Nielsen, A. & Bascompte, J. (2007) Ecological networks, nestedness and sampling  
723 effort. *Journal of Ecology*, **95**, 1134–1141–1141.
- 724 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H. & Jordano, P. (2011)  
725 Missing and forbidden links in mutualistic networks. *Proceedings Of The Royal*  
726 *Society B-Biological Sciences*, **278**, 725–732.
- 727 Olesen, J.M., Dupont, Y.L., O’gorman, E., Ings, T.C., Layer, K., Melin, C.J.,  
728 Trjelsgaard, K., Pichler, D.E., Rasmussen, C. & Woodward, G. (2010) From  
729 Broadstone to Zackenberg. *Advances in Ecological Research*, **42**, 1–69.
- 730 Olesen, J. & Jordano, P. (2002) Geographic patterns in plant-pollinator mutualistic  
731 networks. *Ecology*, **83**, 2416–2424.

- 732 Olito, C. & Fox, J.W. (2014) Species traits and abundances predict metrics of  
733 plant-pollinator network structure, but not pairwise interactions. *Oikos*, **124**,  
734 428–436.
- 735 Ollerton, J. & Cranmer, L. (2002) Latitudinal trends in plant-pollinator interac-  
736 tions: are tropical plants more specialised? *Oikos*, **98**, 340–350.
- 737 Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes,  
738 R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops,  
739 N., Dulloo, E., Faith, D., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurtt,  
740 G., Jetz, W., Karp, D.S., Mcgeoch, M., Obura, D., Onoda, Y., Pettorelli, N.,  
741 Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S., Turak, E., Walpole, M.  
742 & Wegmann, M. (2013) Essential biodiversity variables. *Science*, **339**, 277–278.
- 743 Pocock, M.J.O., Evans, D.M. & Memmott, J. (2012) The Robustness and Restora-  
744 tion of a Network of Ecological Networks. *Science*, **335**, 973–977.
- 745 Preston, F. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.
- 746 R Development Core Team (2010) *R: A language and environment for statis-*  
747 *tical computing*. R Foundation for Statistical Computing. Vienna, Austria.  
748 <http://www.R-project.org>, Vienna, Austria.
- 749 Rivera-Hutinel, A., Bustamante, R.O., Marín, V.H. & Medel, R. (2012) Effects of  
750 sampling completeness on the structure of plant-pollinator networks. *Ecology*,  
751 **93**, 1593–1603.
- 752 Schleuning, M., Frund, J., Klein, A.M., Abrahamczyk, S., Alarcón, R., Albrecht,  
753 M., Andersson, G.K.S., Bazarian, S., Böhning-Gaese, K., Bommarco, R., Dals-

- 754 gaard, B., Dehling, D.M., Gotlieb, A., Hagen, M., Hickler, T., Holzschuh, A.,  
755 Kaiser-Bunbury, C.N., Kreft, H., Morris, R.J., Sandel, B., Sutherland, W.J.,  
756 Svenning, J.C., Tcharntke, T., Watts, S., Weiner, C.N., Werner, M., Williams,  
757 N.M., Winqvist, C., Dormann, C.F. & Blüthgen, N. (2012) Specialization of  
758 mutualistic interaction networks decreases toward tropical latitudes. *Current*  
759 *Biology*, **22**, 1925–1931.
- 760 Snow, B. & Snow, D. (1972) Feeding niches of hummingbirds in a Trinidad valley.  
761 *Journal of Animal Ecology*, **41**, 471–485.
- 762 Snow, B. & Snow, D. (1988) *Birds and berries*. Poyser, Calton, UK.
- 763 Stang, M., Klinkhamer, P., Waser, N.M., Stang, I. & van der Meijden, E. (2009)  
764 Size-specific interaction patterns and size matching in a plant-pollinator inter-  
765 action web. *Annals Of Botany*, **103**, 1459–1469.
- 766 Strogatz, S. (2001) Exploring complex networks. *Nature*, **410**, 268–276.
- 767 Stumpf, M.P.H., Wiuf, C. & May, R.M. (2005) Subnets of scale-free networks are  
768 not scale-free: Sampling properties of networks. *Proceedings of the National*  
769 *Academy of Sciences USA*, **102**, 4221–4224.
- 770 Thébault, E. & Fontaine, C. (2010) Stability of ecological communities and the  
771 architecture of mutualistic and trophic networks. *Science*, **329**, 853–856.
- 772 Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A.,  
773 Galetti, M., García, M.B., García, D., Gomez, J.M., Jordano, P., Medel, R.,  
774 Navarro, L., Obeso, J.R., Oviedo, R., Ramírez, N., Rey, P.J., Traveset, A.,

- 775 Verdú, M. & Zamora, R. (2014) Beyond species loss: the extinction of ecological  
776 interactions in a changing world. *Functional Ecology*, **29**, 299–307.
- 777 Vazquez, D.P., Chacoff, N.P. & Cagnolo, L. (2009) Evaluating multiple deter-  
778 minants of the structure of plant-animal mutualistic networks. *Ecology*, **90**,  
779 2039–2046.
- 780 Vazquez, D.P., Ramos-Jiliberto, R., Urbani, P. & Valdovinos, F.S. (2015) A con-  
781 ceptual framework for studying the strength of plant-animal mutualistic inter-  
782 actions. *Ecology Letters*, **18**, 385–400.
- 783 Vazquez, D., Morris, W. & Jordano, P. (2005) Interaction frequency as a surrogate  
784 for the total effect of animal mutualists on plants. *Ecology Letters*, **8**, 1088–1094.
- 785 Vizentin-Bugoni, J., Maruyama, P.K. & Sazima, M. (2014) Processes entangling  
786 interactions in communities: forbidden links are more important than abundance  
787 in a hummingbird-plant network. *Proceedings Of The Royal Society B-Biological*  
788 *Sciences*, **281**, 20132397–20132397.
- 789 Wells, K. & O’Hara, R.B. (2012) Species interactions: estimating per-individual  
790 interaction strength and covariates before simplifying data into per-species eco-  
791 logical networks. *Methods in Ecology and Evolution*, **4**, 1–8.
- 792 Wirta, H.K., Hebert, P.D.N., Kaartinen, R., Prosser, S.W., Várkonyi, G. & Roslin,  
793 T. (2014) Complementary molecular information changes our perception of food  
794 web structure. *Proceedings of the National Academy of Sciences USA*, **111**,  
795 1885–1890.

## 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).

## 820 Figures

Figure 1:

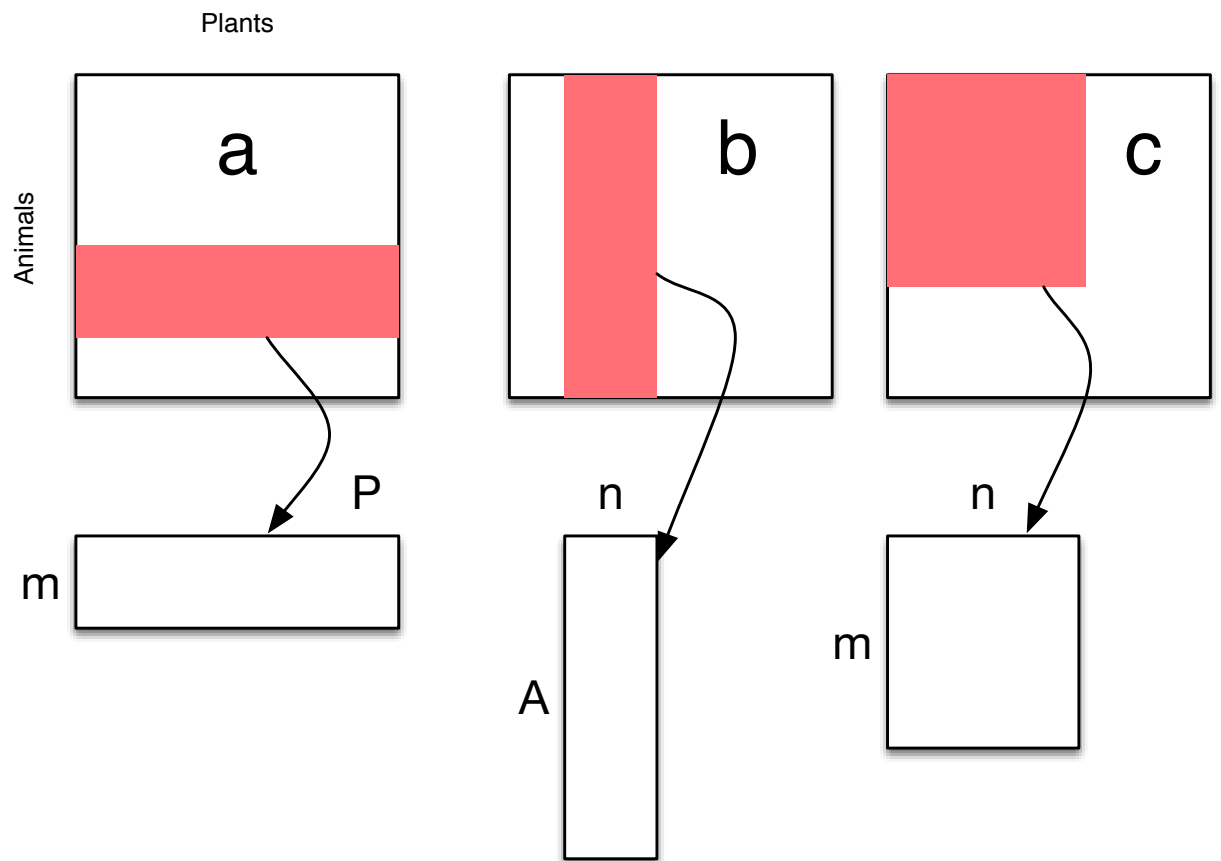
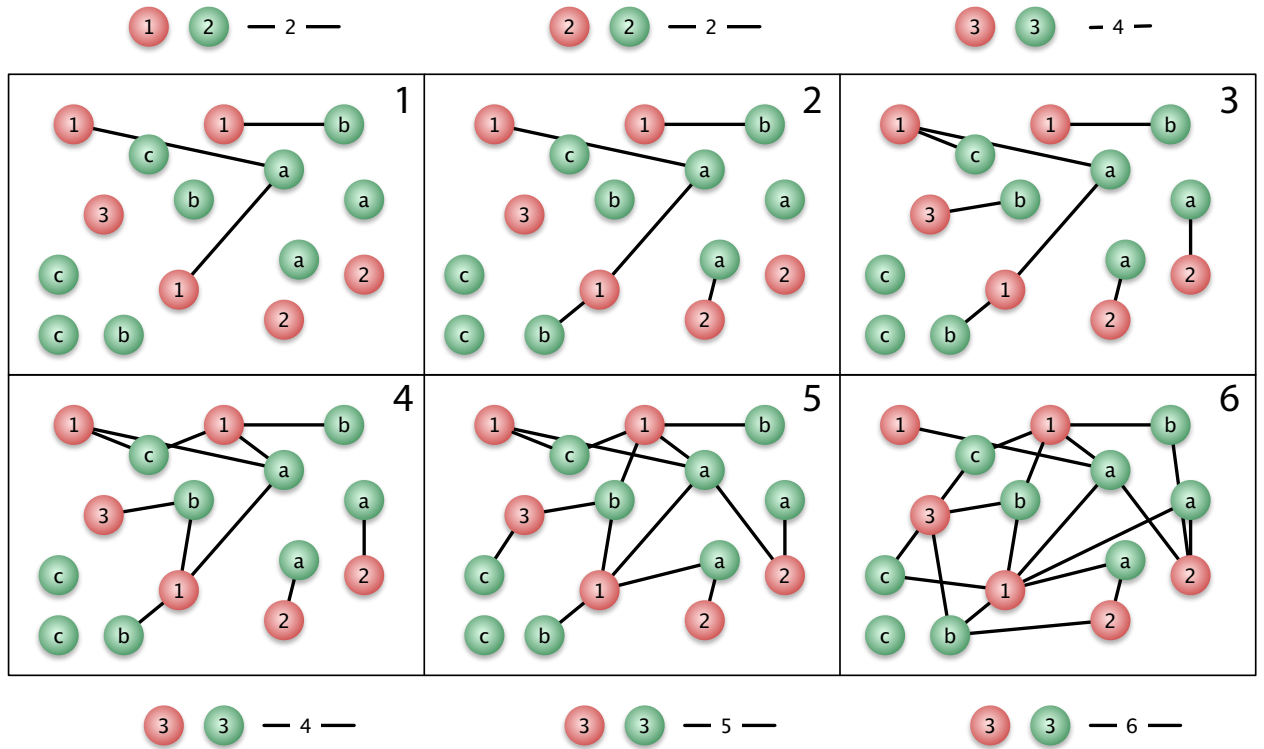




Figure 2:



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

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

849

850 **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)
$FL_S$	$\dots(\dots)$	8 (0.0184)	30 (0.0890)	140 (0.1894)	31 (0.1834)	46 (0.0714)
$FL_A$	$\dots(\dots)$	5 (0.0115)	150 (0.445) <sup>a</sup>	$\dots(\dots)$	20 (0.1183)	61 (0.0947)
$FL_O$	$\dots(\dots)$	$\dots(\dots)$	38 (0.1128) <sup>b</sup>	$\dots(\dots)$	$\dots(\dots)$	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.

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 <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
$Chao1$	$263.1 \pm 70.9$	$231.4 \pm 14.2$	$509.6 \pm 54.7$
$ACE$	$240.3 \pm 8.9$	$241.3 \pm 7.9$	$566.1 \pm 14.8$
$CI$	[169.5–187.4]	[161.8–177.6]	[327.8–357.4]
% <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)