

# 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 Vespucio s/n, E 41092 Sevilla, Spain

Sevilla, September 1, 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 unrelated [delete unrelated? All species are somewhat related, right?] 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, 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).

1 Biodiversity assessment aims at sampling individuals in collections and deter-  
2 mining the number of species represented. Given that, by definition, samples are  
3 incomplete, these collections enumerate a lower number of the species actually  
4 present. The ecological literature dealing with robust estimators of species rich-  
5 ness and diversity in collections of individuals is immense, and a number of useful  
6 approaches have been used to obtain such estimates (Magurran, 1988; Gotelli &  
7 Colwell, 2001; Hortal, Borges & Gaspar, 2006; Colwell, 2009; Gotelli & Colwell,  
8 2011). Recent effort has been also focused at defining essential biodiversity vari-  
9 ables (EBV) (Pereira *et al.*, 2013) that can be sampled and measured repeatedly  
10 to complement biodiversity estimates. Yet sampling species or taxa-specific EBVs  
11 is just probing a single component of biodiversity; interactions among species are  
12 another fundamental component, the one that supports the existence of species.

13 For example, the extinction of interactions represents a dramatic loss of biodiver-  
14 sity because it entails the loss of fundamental ecological functions (Valiente-Banuet  
15 *et al.*, 2014). This missed component of biodiversity loss, the extinction of ecolog-  
16 ical interactions, very often accompanies, or even precedes, species disappearance.  
17 Interactions among species are a key component of biodiversity and here I aim to  
18 show that most problems associated to sampling interactions in natural communi-  
19 ties have to do with problems associated to sampling species diversity. I consider  
20 pairwise interactions among species at the habitat level, in the context of alpha  
21 diversity and the estimation of local interaction richness from sampling data (Mao  
22 & Colwell, 2005). In the first part I provide a succinct overview of previous work  
23 addressing sampling issues for ecological interaction networks. In the second part  
24 I discuss specific rationales for sampling the biodiversity of ecological interactions.

25 Interactions can be a much better indicator of the richness and diversity of  
26 ecosystem functions than a simple list of taxa and their abundances and/or re-  
27 lated biodiversity indicator variables (EBVs). Thus, sampling interactions should  
28 be a central issue when identifying and diagnosing ecosystem services (e.g., pollin-  
29 nation, natural seeding by frugivores, etc.). Fortunately, all the whole battery of  
30 biodiversity-related tools used by ecologists to sample biodiversity (species, *sensu*  
31 *stricto*) can be extended and applied to the sampling of interactions. Analogs  
32 are evident between these approaches (Colwell, Dunn & Harris, 2012). Monitor-  
33 ing interactions is analogous to any biodiversity sampling [i.e., a species inventory  
34 Jordano (1987); Jordano, Vázquez & Bascompte (2009)] and is subject to similar  
35 methodological shortcomings, especially under-sampling (Coddington *et al.*, 2009;  
36 Vazquez, Chacoff & Cagnolo, 2009; Dorado *et al.*, 2011; Rivera-Hutinel *et al.*,  
37 2012). For example, when we study mutualistic networks, our goal is to make an

38 inventory of the distinct pairwise interactions that made up the network. We are  
39 interested in having a complete list of all the pairwise interactions among species  
40 (e.g., all the distinct, species-species interactions, or links, among the pollinators  
41 and flowering plants) that can exist in a given community. Sampling these in-  
42 teractions thus entails exactly the same problems, limitations, constraints, and  
43 potential biases as sampling individual organisms and species diversity. As Mao &  
44 Colwell (2005) put it, these are the workings of Preston's demon, the moving "veil  
45 line" between detected and the undetected interactions as sample size increases  
46 (Preston, 1948).

47 Early efforts to recognize and solve sampling problems in analyses of interac-  
48 tions stem from researchers interested in food web analyses and in determining the  
49 biases of undersampled food web metrics (Martinez, 1991; Cohen *et al.*, 1993; Mar-  
50 tinez, 1993; Bersier, Banasek-Richter & Cattin, 2002; Brose, Martinez & Williams,  
51 2003; Banasek-Richter, Cattin & Bersier, 2004). In addition, the myriad of classic  
52 natural history studies documenting animal diets, host-pathogen infection records,  
53 plant herbivory records, etc., represent efforts to document interactions occurring  
54 in nature. All of them share the problem of sampling incompleteness influenc-  
55 ing the patterns and metrics reported. Yet, despite the early recognition that  
56 incomplete sampling may seriously bias the analysis of ecological networks (Jor-  
57 dano, 1987), only recent studies have explicitly acknowledged it and attempted to  
58 determine its influence (Ollerton & Cranmer, 2002; Nielsen & Bascompte, 2007;  
59 Vazquez, Chacoff & Cagnolo, 2009; Gibson *et al.*, 2011; Olesen *et al.*, 2011; Cha-  
60 coff *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox, 2014; Bascompte &  
61 Jordano, 2014; Vizentin-Bugoni, Maruyama & Sazima, 2014; Frund, McCann &  
62 Williams, 2015) [What about Wells O'hara 2013 MEE? It's not cited in the ms, but

63 looks quite relevant to me... ([http://onlinelibrary.wiley.com/doi/10.1111/j.2041-  
64 210x.2012.00249.x/full](http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210x.2012.00249.x/full)]). The sampling approaches have been extended to predict  
65 patterns of coextinctions in interaction assemblages (e.g., hosts-parasites) (Colwell,  
66 Dunn & Harris, 2012). Most empirical studies provide no estimate of sampling  
67 effort, implicitly assuming that the reported network patterns and metrics are  
68 robust. Yet recent evidences point out that number of partner species detected,  
69 number of actual links, and some aggregate statistics describing network patterns,  
70 are prone to sampling bias (Nielsen & Bascompte, 2007; Dorado *et al.*, 2011; Ole-  
71 sen *et al.*, 2011; Chacoff *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox,  
72 2014; Frund, McCann & Williams, 2015). Most of these evidences, however, come  
73 from either theoretical, simulation, studies (Frund, McCann & Williams, 2015)  
74 or from relatively species-poor assemblages. Even for species-rich, tropical assem-  
75 blages it might be erroneous to conclude that network data routinely come from  
76 insufficiently sampled datasets (Ollerton & Cranmer, 2002; Chacoff *et al.*, 2012),  
77 given the extremely sparse nature of these interaction matrices because of the  
78 prevalence of forbidden links (which, by definition, cannot be documented despite  
79 extensive sampling effort). However, most certainly, sampling limitations pervade  
80 biodiversity inventories in tropical areas (Coddington *et al.*, 2009) and we might  
81 rightly expect that frequent interactions may be over-represented and rare inter-  
82 actions may be missed entirely in studies of mega-diverse assemblages (Bascompte  
83 & Jordano, 2014); but, to what extent?  
84 [awesome introduction! :)]

## 85 Sampling interactions: methods

86 When we sample interactions in the field we record the presence of two species  
87 that interact [engage? To avoid repeating interact...] in some way. For example,  
88 Snow and Snow(1988) recorded an interaction whenever they saw a bird “touch-  
89 ing” a fruit on a plant. We observe and record feeding observations, visitation,  
90 occupancy, presence in pollen loads or in fecal samples, etc., of *individual* animals  
91 or plants and accumulate pairwise interactions, i.e., lists of species partners and  
92 the frequencies with which we observe them. Therefore, estimating the sampling  
93 completeness of pairwise interactions for a whole network, requires estimating the  
94 number (richness) of distinct pairwise interactions accumulated as sampling effort  
95 is increased, pooling the data for all partner species. [While I agree with most of  
96 last sentence, I'm not completely sure that estimating network completeness \*re-  
97 quires\* comparing the number of interactions in relation to increasing sampling  
98 effort. I mean, that can be a fruitful way to assess it, but probably not the only one  
99 (thus not a \*requirement\*). For instance, we might consider a network quite com-  
100 plete if we know (nearly) all missing links are forbidden (thus not requiring extra  
101 sampling effort). Or alternatively, we could assess the uncertainty around missing  
102 links (i.e. how certain we are that the link is actually absent vs unobserved) by  
103 accounting for detectability (i.e. false negatives), again not requiring increasing  
104 sampling effort. The latter relates to the well-developed field of occupancy mod-  
105 elling (e.g. Royle Link Ecology 2006) which I think could be very useful in this  
106 context. The problem is very similar: inferring the presence of species/interactions  
107 when we have imperfect observations with false negatives. Let me know If you'd  
108 like to discuss this further! Also, Wells O'hara AFAIK didn't require varying

109 sampling efforts to estimate the true number of interactions.]

110 Most, if not all, types of ecological interactions can be illustrated with bipar-  
111 tite graphs, with two or more distinct groups of interacting partners (Bascompte &  
112 Jordano, 2014); for illustration purposes I'll focus more specifically on plant-animal  
113 interactions. Sampling interactions requires filling the cells of an interaction ma-  
114 trix with data. The matrix,  $\Delta = AP$ , is a 2D representation of the interactions  
115 among, say,  $A$  animal species (rows) and  $P$  plant species (columns) (Jordano, 1987;  
116 Bascompte & Jordano, 2014). [The next two sentences sound repetitive; I'd edit  
117 them or simply leave the second one, more detailed.] An interaction matrix  $\Delta$  con-  
118 sists of an array of zeroes or ones, or an array of numeric values (including zeroes)-  
119 if the data (interaction frequencies) are quantified. The matrix entries illustrate  
120 the values of the pairwise interactions visualized in the  $\Delta$  matrix, and can be 0  
121 or 1, for presence-absence of a given pairwise interaction, or take a quantitative  
122 weight  $w_{ji}$  to represent the interaction intensity or unidirectional effect of species  
123  $j$  on species  $i$  (Bascompte & Jordano, 2014; Vazquez *et al.*, 2015). Given that the  
124 outcomes of most ecological interactions are dependent on frequency of encounters  
125 (e.g., visit rate of pollinators, number of records of ant defenders, frequency of  
126 seeds in fecal samples), a frequently used proxy for interaction intensities  $w_{ji}$  is  
127 just how frequent are new interspecific encounters, whether or not appropriately  
128 weighted to estimate interaction effectiveness (Vazquez, Morris & Jordano, 2005).

129 We need to define two basic steps in the sampling of interactions: 1) which  
130 type of interactions we sample; and 2) which type of record we get to document  
131 the existence of an interaction. In step #1 we need to take into account whether  
132 we are sampling the whole community of interactors (all the animals, all  
133 the plants) or just a subset of them, i.e., a sub matrix  $\Delta_{m,n}$  of  $m < A$  animal

---

Jordano - Sampling networks

134 species and  $n < P$  plant species of the adjacency matrix  $\Delta_{AP}$ . Subsets can be:  
 135 a) all the potential plants interacting with a subset of the animals (Fig. 1a); b)  
 136 all the potential animal species interacting with a subset of the plant species (Fig.  
 137 1b); c) a subset of all the potential animal species interacting with a subset of all  
 138 the plant species (Fig. 1c). While some discussion has considered how to establish  
 139 the limits of what represents a network (Strogatz, 2001) (in analogy to discussion  
 140 on food-web limits; Cohen, 1978), it must be noted that situations a-c in Fig.  
 141 1 do not represent complete interaction networks. As vividly stated by Cohen  
 142 et al. (1993): “*As more comprehensive, more detailed, more explicit webs become  
 143 available, smaller, highly aggregated, incompletely described webs may progressively  
 144 be dropped from analyses of web structure (though such webs may remain useful for  
 145 other purposes, such as pedagogy)*”. Subnet sampling is generalized in studies of  
 146 biological networks (e.g., protein interactions, gene regulation), yet it is important  
 147 to recognize that most properties of subnetworks (even random subsamples) do  
 148 not represent properties of whole networks (Stumpf, Wiuf & May, 2005).

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

---

Jordano - Sampling networks

159        Most recent analyses of plant-animal interaction networks are phyto-centric;  
160        just 3.5% of available plant-pollinator ( $N= 58$ ) or 36.6% plant-frugivore ( $N= 22$ )  
161        interaction datasets are zoo-centric (see Schleuning *et al.*, 2012). Moreover, most  
162        available datasets on host-parasite or plant-herbivore interactions are “host-centric”  
163        or phyto-centric (e.g., Thébault & Fontaine, 2010; Eklöf *et al.*, 2013). This maybe  
164        related to a variety of causes, like preferred methodologies by researchers working  
165        with a particular group or system, logistic limitations, or inherent taxonomic focus  
166        of the research questions. A likely result of phyto-centric sampling would be adja-  
167        cency matrices with large  $A : P$  ratios. In any case we don’t have a clear view of  
168        the potential biases that taxa-focused sampling may generate in observed network  
169        patterns, for example by generating consistently asymmetric interaction matrices  
170        (Dormann *et al.*, 2009). System symmetry has been suggested to influence esti-  
171        mations of generalization levels in plants and animals when measured as  $I_A$  and  
172         $I_P$  (Elberling & Olesen, 1999); thus, differences in  $I_A$  and  $I_P$  between networks  
173        may arise from different  $A : P$  ratios rather than other ecological factors (Olesen  
174        & Jordano, 2002).

175        Interestingly enough, quite complete analyses of interaction networks can be  
176        obtained when combining both phyto-centric and zoo-centric sampling. For ex-  
177        ample, Bosch et al. (Bosch *et al.*, 2009) showed that the addition of pollen load  
178        data on top of focal-plant sampling of pollinators unveiled a significant number  
179        of interactions, resulting in important network structural changes. Connectance  
180        increased 1.43-fold, mean plant connectivity went from 18.5 to 26.4, and mean  
181        pollinator connectivity from 2.9 to 4.1; moreover, extreme specialist pollinator  
182        species (singletons in the adjacency matrix) decreased 0.6-fold. Zoo-centric sam-

183 pling has recently been extended with the use of DNA-barcoding, for example

---

Jordano - Sampling networks

184 with plant-herbivore (Jurado-Rivera *et al.*, 2009) and plant-frugivore interactions  
185 (González-Varo, Arroyo & Jordano, 2014). For mutualistic networks we would ex-  
186 pect that zoo-centric sampling could help unveiling interactions for rare species or  
187 for relatively common species which are difficult to sample by direct observation.  
188 Future methodological work may provide significant advances showing how mixing  
189 different sampling strategies strengthens the completeness of network data. These  
190 mixed strategies may combine, for instance, focal analyses, pollen load or seed  
191 contents, camera traps, and DNA barcoding records. We might expect increased  
192 power of these mixed sampling approaches when combining different methods from  
193 both phyto- and zoo-centric perspectives (Bosch *et al.*, 2009; Bluthgen, 2010).

194 [Very nice and clear section]

## 195 Sampling interactions: rationale

196 The number of distinct pairwise interactions that we can record in a landscape (an  
197 area of relatively homogeneous vegetation, analogous to the one we would use to  
198 monitor species diversity) is equivalent to the number of distinct classes in which  
199 we can classify the recorded encounters among individuals of two different species.  
200 Yet, individual-based plant-animal interaction networks have been only recently  
201 studied (Dupont, Trøjelsgaard & Olesen, 2011). We walk in the forest and see  
202 a blackbird *Tm* picking an ivy *Hh* fruit and ingesting it: we have a record for  
203 *Tm – Hh* interaction. We keep advancing and record again a blackbird feeding  
204 on hawthorn *Cm* fruits so we record a *Tm – Cm* interaction; as we advance we  
205 encounter another ivy plant and record a blackcap swallowing a fruit so we now

206 have a new  $Sa - Hh$  interaction, and so on. At the end we have a series of classes

Jordano - Sampling networks

207 (e.g.,  $Sa - Hh$ ,  $Tm - Hh$ ,  $Tm - Cm$ , etc.), along with their observed frequencies.

208 Bunge & Fitzpatrick (1993) review the main aspects and approaches to estimate  
209 the number of distinct classes  $C$  in a sample of observations. [Not sure to have fol-  
210 lowed the reasoning: isn't the maximum number of interactions directly defined by  
211 the numbers of animal and plant species in the site? That is, `expand.grid(animals,`  
212 `plants)`? Why do we need to determine a (reduced) number of classes within that?

213 We have a list of all possible interactions and each one would have a 0 or 1 depend-  
214 ing on having being observed, right? To me the problem looks more analogous to  
215 rarefaction analyses, or discerning true zeros (i.e. forbidden links) from unobserved  
216 (but existent) interactions, as you expand later.] The sampling of interactions in  
217 nature, as the sampling of species, is a cumulative process. In our analysis, we are  
218 not re-sampling individuals, but interactions, so we made interaction-based accu-  
219 mulation curves. If an interaction-based curve points towards a robust sampling,  
220 it does mean that no new interactions are likely to be recorded, irrespectively of  
221 the species, as it is a whole-network sampling approach (N. Gotelli, pers. com.).

222 We add new, distinct, interactions recorded as we increase sampling effort (Fig. 2).

223 We can obtain an Interaction Accumulation Curve (*IAC*) analogous to a Species  
224 cumulating Curve (*SAC*): the observed number of distinct pairwise interactions  
225 in a survey or collection as a function of the accumulated number of observations  
226 or samples (Colwell, 2009). [I think more details are needed here; I feel the IAC  
227 has been introduced too fast (e.g. as compared to the description in Github). Or  
228 alternatively, just leave here a gist of the idea, as it is expanded later in the next  
229 section. Also, I think it would be \*\*great\*\* to include such figure (the IAC) in

230 the ms, along with a short vignette (or even package) showing how to do these  
231 analyses with real networks.]

Jordano - Sampling networks

232 Our sampling above would have resulted in a vector  $n = [n_1 \dots n_C]'$  where  $n_i$   
233 is the number of records in the  $i^{th}$  class. As stressed by Bunge & Fitzpatrick  
234 (1993), however, the  $i^{th}$  class would appear in the sample if and only if  $n_i > 0$ ,  
235 and we don't know *a priori* which  $n_i$  are zero. [Or, alternatively, we do know  
236 all possible interactions, but can't distinguish true zeros (forbidden links) from  
237 unobserved links (due to insufficient sampling effort)] So,  $n$  is not observable [or  
238 is observed with error (which we could account for)]. Rather, what we get is a  
239 vector  $c = [c_1 \dots c_n]'$  where  $c_j$  is the number of classes represented  $j$  times in our  
240 sampling:  $c_1$  is the number of singletons [should we assume readers know what this  
241 mean? First time the term appears],  $c_2$  is the number of twin pairs,  $c_3$  the number  
242 of triplets, etc. The problem thus turns to be estimating the number of distinct  
243 classes  $C$  from the vector of  $c_j$  values. [This paragraph is harder to follow... Also,  
244 are you using this approach later in the ms? I can't find its connection to the rest  
245 of the approaches in the ms...]

246 Estimating the number of interactions with resulting robust estimates of net-  
247 work parameters is a central issue in the study of ecological interaction networks  
248 (Jordano, 1987; Bascompte & Jordano, 2014). In contrast with traditional species  
249 diversity estimates, sampling networks has the paradox that despite the poten-  
250 tially interacting species being present in the sampled assemblage (i.e., included in  
251 the  $A$  and  $P$  species lists), some of their pairwise interactions are impossible to be  
252 recorded. The reason is forbidden links. Independently of whether we sample full  
253 communities or subset communities we face a problem: some of the interactions

254 that we can visualize in the empty adjacency matrix  $\Delta$  will simply not occur.  
255 Thus, independently of the sampling effort we put, we'll never document these  
256 pairwise interactions. With a total of  $AP$  “potential” interactions, a fraction of

257 them are impossible to record, because they are forbidden (Jordano, Bascompte  
258 & Olesen, 2003; Olesen *et al.*, 2011) [mention Table 1 here too?]. Forbidden links  
259 are constraints for the establishment of new links, and mainly arise from the bio-  
260 logical attributes of the species: no link can be established between a plant and an  
261 animal mutualist differing in phenology, i.e. the seeds of a winter-ripening plant  
262 cannot be dispersed by a frugivore that is a summer stopover migrant (Jordano,  
263 1987). Or, for instance, short-tongued pollinators cannot successfully reach the  
264 nectar in long-corolla flowers and pollinate them efficiently (Moré *et al.*, 2012).  
265 Forbidden links are thus represented as structural zeroes in the interaction matrix,  
266 i.e., matrix cells that cannot get a non-zero value. So, we need to account for the  
267 frequency of these structural zeros in our matrix before proceeding. For example,  
268 most measurements of connectance  $C = I/(AP)$  implicitly ignore the fact that  
269 by taking the full product  $AP$  in the denominator they are underestimating the  
270 actual connectance value, i.e., the fraction of actual interactions  $I$  relative to the  
271 *biologically possible* ones, not to the total maximum  $I_{max} = AP$ .

272 Adjacency matrices are frequently sparse, i.e., they are densely populated with  
273 zeroes, with a fraction of them being structural (i.e., unobservable interactions)  
274 (Bascompte & Jordano, 2014). It would be thus a serious interpretation error to  
275 attribute the sparseness of adjacency matrices for bipartite networks to undersam-  
276 pling. The actual typology of link types in ecological interaction networks is thus  
277 more complex than just the two categories of observed and unobserved interactions

278 (Table 1). Unobserved interactions are represented by zeroes and belong to two  
279 categories. Missing interactions may actually exist but require additional sam-  
280 pling or a variety of methods to be observed. Forbidden links, on the other hand,  
281 arise due to biological constraints limiting interactions and remain unobservable

Jordano - Sampling networks

282 in nature, irrespectively of sampling effort (Jordano, Bascompte & Olesen, 2003;  
283 Olesen *et al.*, 2011). Forbidden links are non-occurrences of pairwise interactions  
284 that can be accounted for by biological constraints, such as spatio-temporal un-  
285 coupling, size or reward mismatching, foraging constraints (e.g., accessibility), and  
286 physiological-biochemical constraints (Jordano, 1987). [These sentences sound a  
287 bit repetitive with the last half of the former paragraph, where forbidden links  
288 have already been introduced and some examples given] We still have extremely  
289 reduced information about the frequency of forbidden links in natural communi-  
290 ties (Jordano, Bascompte & Olesen, 2003; Stang *et al.*, 2009; Vazquez, Chacoff &  
291 Cagnolo, 2009; Olesen *et al.*, 2011; Ibanez, 2012; Maruyama *et al.*, 2014; Vizentin-  
292 Bugoni, Maruyama & Sazima, 2014) (Table 1). Forbidden links *FL* may actually  
293 account for a relatively large fraction of unobserved interactions *UL* when sam-  
294 pling taxonomically-restricted subnetworks (e.g., plant-hummingbird pollination  
295 networks) (Table 2). Phenological unmatching is also prevalent in most networks,  
296 and may add up to explain ca. 25–40% of the forbidden links, especially in highly  
297 seasonal habitats, and up to 20% when estimated relative to the total number  
298 of unobserved interactions (Table 2). In any case, we might expect that a frac-  
299 tion of the missing links *ML* would be eventually explained by further biological  
300 reasons, depending on the knowledge of natural details of the particular systems.  
301 Our goal as naturalists would be to reduce the fraction of *UL* which remain as  
302 missing links; to this end we might search for additional biological constraints or

303 increase sampling effort. For instance, habitat use patterns by hummingbirds in  
304 the Arima Valley network (Table 2; Snow & Snow, 1972) impose a marked pattern  
305 of microhabitat mismatches causing up to 44.5% of the forbidden links. A myr-  
306 iad of biological causes beyond those included as *FL* in Table 2 may contribute

Jordano - Sampling networks

307 explanations for *UL*: limits of color perception and or partial preferences, pres-  
308 ence of secondary metabolites in fruit pulp and leaves, toxins and combinations  
309 of monosaccharides in nectar, etc. However, it is surprising that just the limited  
310 set of forbidden link types considered in Table 2 explain between 24.6 77.2% of  
311 the unobserved links. Notably, the Arima Valley, Santa Virginia, and Hato Ratón  
312 networks have > 60% of the unobserved links explained, which might be related  
313 to the fact that they are subnetworks (Arima Valley, Santa Virgínia) or relatively  
314 small networks (Hato Ratón). All this means that empirical networks may have  
315 sizable fractions of structural zeroes. Ignoring this biological fact may contribute  
316 to wrongly infer undersampling of interactions in real-world assemblages.

317 To sum up, two elements of inference are required in the analysis of unobserved  
318 interactions in ecological interaction networks: first, detailed natural history infor-  
319 mation on the participant species that allows the inference of biological constraints  
320 imposing forbidden links, so that structural zeroes can be identified in the adj-  
321 cency matrix; second, a critical analysis of sampling robustness, including a robust  
322 estimate of the actual fraction of missing links,  $M$ , and thus, a robust estimate  
323 of  $I$ . [I agree, but considering that information to infer forbidden links is always  
324 going to be limited, could we use your IACs to infer the number of total links  
325 in the network (as the asymptote of the rarefaction curve, hence accounting for  
326 sampling effort)? Then the number of forbidden links could be estimated as  $I_{max}$

327 minus that asymptote from the IAC, right? If you can't explain the forbidden  
328 links because you lack natural history information, at least you can tell how many  
329 of the unobserved links are forbidden and how many missing due to insufficient  
330 sampling. Does this makes sense? (Surely not)]

### 331 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 (true) 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. [These concepts (singleton, doubleton, etc) have not been explained yet] The total number of individuals in the sample would be:

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

332

333 A frequently used asymptotic, bias corrected, non-parametric estimator is  $S_{Chao}$

<sup>334</sup> (Hortal, Borges & Gaspar, 2006; Chao, 2005; Colwell, 2013):

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

<sup>335</sup> Another frequently used alternative is the Chao2 estimator,  $S_{Chao2}$  (Gotelli &  
<sup>336</sup> Colwell, 2001), which has been reported to have a limited bias for small sample

Jordano - Sampling networks

<sup>337</sup> sizes (Colwell & Coddington, 1994; Chao, 2005):

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

<sup>338</sup> A plot of the cumulative number of species recorded,  $S_n$ , as a function of  
<sup>339</sup> some measure of sampling effort (say,  $n$  samples taken) yields the species accu-  
<sup>340</sup> mulation curve (SAC) or collector's curve (Colwell & Coddington, 1994). Such a  
<sup>341</sup> curve eventually reaches an asymptote converging with  $S_{est}$ . In an analogous way,  
<sup>342</sup> interaction accumulation curves (IAC), analogous to SACs, can be used to as-  
<sup>343</sup> sess the robustness of interactions sampling for plant-animal community datasets  
<sup>344</sup> (Jordano, 1987; Jordano, Vázquez & Bascompte, 2009; Olesen *et al.*, 2011). For  
<sup>345</sup> instance, a random accumulator function (e.g., library vegan in the R Package, R  
<sup>346</sup> Development Core Team, 2010) which finds the mean IAC and its standard devia-  
<sup>347</sup> tion from random permutations of the data, or subsampling without replacement  
<sup>348</sup> (Gotelli & Colwell, 2001) can be used to estimate the expected number of distinct  
<sup>349</sup> pairwise interactions included in a given sampling of records (Jordano, Vázquez  
<sup>350</sup> & Bascompte, 2009; Olesen *et al.*, 2011). We start with a vectorized interaction  
<sup>351</sup> matrix representing the pairwise interactions (rows) recorded during a cumulative  
<sup>352</sup> number of censuses or sampling periods (columns) (Table 3), in a way analogous

353 to a biodiversity sampling matrix with species as rows and sampling units (e.g.,  
354 quadrats) as columns (Jordano, Vázquez & Bascompte, 2009). [From the previous  
355 sentence (We start...), it looks like you are going to explain the method, at least  
356 briefly. But that happens more in the next section and paragraph. I suggest to  
357 move these details about data formatting and possible algorithms there, to avoid  
358 sounding repetitive] In this way we effectively extend sampling theory developed

Jordano - Sampling networks

359 for species diversity to the sampling of ecological interactions. Yet future theoret-  
360 ical work will be needed to formally assess the similarities and differences in the  
361 two approaches and developing biologically meaningful null models of expected  
362 interaction richness with added sampling effort.

## 363 Assessing sampling effort when recording interac- 364 tions

365 The basic method we can propose to estimate sampling effort and explicitly show  
366 the analogues with rarefaction analysis in biodiversity research is to vectorize the  
367 interaction matrix  $AP$  so that we get a vector of all the potential pairwise inter-  
368 actions ( $I_{max}$ , Table 1) that can occur in a community of  $A$  animal species and  
369  $P$  plant species. The new “species” we aim to sample are the pairwise interac-  
370 tions (Table 3). So, if we have in our community *Turdus merula* ( $Tm$ ) and *Rosa*  
371 *canina* ( $Rc$ ) and *Prunus mahaleb* ( $Pm$ ), our problem will be to sample 2 new  
372 “species”:  $Tm - Rc$  and  $Tm - Pm$ . In general, if we have  $A = 1...i$ , animal  
373 species and  $P = 1...j$  plant species, we’ll have a vector of “new” species to sample:  
374  $A \cdot P$ ,  $A \cdot P_s$ ,  $A_s P$ ,  $A_s P_s$ ,  $A \cdot P_r$ . We can represent the successive samples where

374  $\alpha_{11}, \alpha_{12}, \dots, \alpha_{1j}, \dots, \alpha_{21}, \alpha_{22}, \dots, \alpha_{2j}$ . You can represent the successive samples where  
375 we can potentially get records of these interactions in a matrix with the vectorized  
376 interaction matrix and columns representing the successive samples we take (Table  
377 3). This is simply a vectorized version of the interaction matrix.

378 [I think the next logical paragraph is this one about rarefaction analysis, so I  
379 suggest to move it here. Also, I think it might be good to give more details about  
380 the procedure (maybe as Suppl Mat), as these IACs look like a main contribution

381 of the ms. Or at least refer to the instructions in the Github repo!]

382 Rarefaction analysis and diversity-accumulation analysis (Magurran, 1988; Hor-  
383 tal, Borges & Gaspar, 2006) come up immediately with this type of dataset. This  
384 procedure plots the accumulation curve for the expected number of distinct pair-  
385 wise interactions recorded with increasing sampling effort (Jordano, Vázquez &  
386 Bascompte, 2009; Olesen *et al.*, 2011). Asymptotic estimates of interaction rich-  
387 ness and its associated standard errors and confidence intervals can thus be ob-  
388 tained (Hortal, Borges & Gaspar, 2006) [([include a figure of an IAC as in Github!.  
389 It should be noted that the asymptotic estimate of interaction richness implicitly  
390 ignores the fact that, due to forbidden links, a number of pairwise interactions  
391 among the  $I_{max}$  number specified in the adjacency matrix  $\Delta$  cannot be recorded,  
392 irrespective of sampling effort. Therefore, the asymptotic value most likely is an  
393 overestimate of the actual maximum number of links that can be present in an  
394 assemblage. [Mmmm, I know very little about these rarefaction analyses, but I  
395 can't see why they should overestimate the number of real interactions... They  
396 just analyse the rate of accumulation of interactions with increasing sampling ef-  
397 fort and calculate the asymptote, right? So, as far as I can see, forbidden links

398 are left out of the asymptote, which only increases if new interactions are added  
399 with increased sampling effort. If this is right, it would be nice because we could  
400 then calculate how many of the unobserved links are due to insufficient sampling  
401 error and then calculate the number of forbidden links by deduction (as in Table  
402 1). Maybe you can expand a bit about why the asymptote overestimates the num-  
403 ber of links?] If forbidden links are taken into account, the asymptotic estimate  
404 should be lower. Yet, to the best of my knowledge, there is no theory developed  
405 to estimate this “biologically real” asymptotic value. Not unexpectedly, most re-

20

---

Jordano - Sampling networks

406 cent analyses of sampling effort in ecological network studies found evidences of  
407 undersampling (Chacoff *et al.*, 2012). This needs not to be true, especially when  
408 interaction subwebs are studied (Olesen *et al.*, 2011; Vizentin-Bugoni, Maruyama  
409 & Sazima, 2014), and once the issue of structural zeroes in the interaction matrices  
410 is effectively incorporated in the estimates.

411 [I may be completely lost by now, but I think these models below are of a  
412 different class, and could perhaps go in a different subsection. Or leave them  
413 here, but I'd put them after the simpler rarefaction analyses. Also, they match  
414 better with the next section more focussed on accounting for detectability, species  
415 abundances...]

416 For example, mixture models incorporating detectabilities have been proposed  
417 to effectively account for rare species (Mao & Colwell, 2005). In an analogous line,  
418 mixture models could be extended to samples of pairwise interactions, also with  
419 specific detectability values. These detection rate/odds could be variable among  
420 groups of interactions, depending on their specific detectability. For example,  
421 detectability of flower-pollinator interactions involving bumblebees could have a

422 higher detectability than flower-pollinator pairwise interactions involving, say, mi-  
 423 tidulid beetles. These more homogeneous groupings of pairwise interactions within  
 424 a network define modules (Bascompte & Jordano, 2014), so we might expect that  
 425 interactions of a given module (e.g., plants and their hummingbird pollinators; Fig.  
 426 1a) may share similar detectability values, in an analogous way to species groups  
 427 receiving homogeneous detectability values in mixture models (Mao & Colwell,  
 428 2005). Such sampling, in its simplest form, would result in a sample with multiple  
 429 pairwise interactions detected, in which the number of interaction events recorded  
 430 for each distinct interaction found in the sample is recorded (i.e., a column vector

#### Jordano - Sampling networks

431 in Table 3, corresponding to, say, a sampling day). The number of interactions  
 432 recorded for the  $i_{th}$  pairwise interaction (i.e.,  $A_i P_j$  in Table 3),  $Y_i$  could be treated  
 433 as a Poisson random variable with a mean parameter  $\lambda_i$ , its detection rate. Mix-  
 434 ture models (Mao & Colwell, 2005) include estimates for abundance-based data  
 435 (their analogous in interaction sampling would be weighted data), where  $Y_i$  is  
 436 a Poisson random variable with detection rate  $\lambda_i$ . This is combined with the  
 437 incidence-based model, where  $Y_i$  is a binomial random variable (their analogous  
 438 in interaction sampling would be presence/absence records of interactions) with  
 439 detection odds  $\lambda_i$ . Let  $T$  be the number of samples in an incidence-based data set.  
 440 A Poisson/binomial density can be written as (Mao & Colwell, 2005):

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

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

442 I would have a look at Wells O'hara 2013 MEE (<http://onlinelibrary.wiley.com/doi/10.1111/j.20>)

444 The detection rates  $\lambda_i$  depend on the relative abundances  $\phi_i$  of the interactions,  
 445 the probability of a pairwise interaction being detected when it is present, and the  
 446 sample size (the number of interactions recorded), which, in turn, is a function  
 447 of the sampling effort. Unfortunately, no specific sampling model has been de-  
 448 veloped along these lines for species interactions and their characteristic features.  
 449 For example, a complication factor might be that interaction abundances,  $\phi_i$ , in  
 450 real assemblages are a function of the abundances of interacting species, that de-  
 451 termine interspecific encounter rates; yet they also depend on biological factors

452 that ultimately determine if the interaction occurs when the partner species are  
 453 present. In its simplest form,  $\phi_i$  could be estimated from just the product of part-  
 454 ner species abundances, an approach recently used as a null model to assess the  
 455 role of biological constraints in generating forbidden links and explaining interac-  
 456 tion patterns (Vizentin-Bugoni, Maruyama & Sazima, 2014). Yet more complex  
 457 models should incorporate not only interspecific encounter probabilities, but also  
 458 phenotypic matching and incidence of forbidden links.

## 459 The *real* missing links

460 Given that a fraction of unobserved interactions can be accounted for by for-  
 461 bidden links, what about the remaining missing interactions? We have already  
 462 discussed that some of these could still be related to unaccounted constraints, and  
 463 still others would be certainly attributable to insufficient sampling. Would this  
 464 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 partner species (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 sam-

ple, or just simply non-occurring in nature. Consider the Nava de las Correhuelas interaction web (NCH, Table 2), with  $A = 36$ ,  $P = 25$ ,  $I = 181$ , and almost half of the unobserved interactions not accounted for by forbidden links missing links [missing or forbidden? I guess the latter],  $M = 53.1\%$ . 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  $PIE_{median} < 1.4 \cdot 10^{-4}$  we may safely expect (note the quantile estimate  $Q_{75\%} = 3.27 \cdot 10^{-4}$ ) that a sizable fraction of these missing interactions may simply not occur according to this neutral expectation (Jordano, 1987; Olesen *et al.*, 2011) (neutral forbidden links, *sensu*

489 Canard *et al.*, 2012). Which is the expected frequency for pairwise interactions?  
490 And, which is the expected probability for unobserved interactions? More specific-  
491 ally, which is the probability of missing interactions,  $M$  (i.e., the unobserved ones  
492 that cannot be accounted for as forbidden links)?

493 When we consider the vectorized interaction matrix, enumerating all pairwise  
494 interactions for the  $AP$  combinations, the expected probabilities of finding a given  
495 interaction can be estimated with a Good-Turing approximation (Good, 1953).  
496 The technique, developed by Alan Turing and I.J. Good with applications to lin-  
497 guistics and word analysis (Gale & Sampson, 1995) and recently applied in ecology  
498 (Chao *et al.*, 2015), estimates the probability of recording an interaction of a hith-  
499 erto unseen pair of partners, given a set of past records of interactions between

500 other species pairs. Let a sample of  $N$  interactions so that  $n_r$  distinct pairwise  
501 interactions have exactly  $r$  records. All Good-Turing estimators obtain the under-  
502 lying frequencies of events as:

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

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

511 have seen, then the sum of  $P(X)$  is  $1 - \frac{1-E(1)}{T}$ . Because probabilities sum to one,  
512 we have the left-over probability of  $P_{new} = \frac{E(1)}{T}$  of seeing something new, where  
513 new means that we sample a new pairwise interaction. [very cool! :)]

514 Note, however, that Good-Turing estimators, as the traditional asymptotic  
515 estimators, do not account in our case for the forbidden interactions. To account for  
516 these *FL* I re-scaled the asymptotic estimates, so that a more meaningful estimate  
517 could be obtained (Table 4). The scaling was calculated as [*Chao1\*(I+ML)*]/*AP*,  
518 just correcting for the *FL* frequency [I got lost here...], given that *I+ML* represent  
519 the total *feasible* interactions when discounting the forbidden links (Table 1). After  
520 scaling, observed *I* values (Table 2) are within the *Chao1* and *ACE* [What is  
521 ACE? Define here?] asymptotic estimates but below the *ACE* estimates for Hato

Jordano - Sampling networks

522 Ratón and Zackenberg (Table 4). Thus, even after re-scaling for *FL*, it is likely  
523 that adequate characterization of most interaction networks will require intensive  
524 sampling effort.

525 [This section is very cool. I wonder if you could strengthen the connection with  
526 the previous sections by including a wrap-up (either here or in the Discussion) on  
527 how to connect the different analyses. I can't see well the big picture of how the  
528 different approaches you have explained in the ms can be used and related - but  
529 I'm newbie in the topic, so it can surely be my fault!]

## 530 Discussion

531 Recent work has inferred that most data available for interaction networks are  
532 incomplete due to undersampling, resulting in a variety of biased parameters and

533 network patterns (Chacoff *et al.*, 2012). It is important to note, however, that  
534 in practice, many surveyed networks to date have been subnets of much larger  
535 networks. This is true for protein interaction, gene regulation, and metabolic  
536 networks, where only a subset of the molecular entities in a cell have been sam-  
537 pled (Stumpf, Wiuf & May, 2005). Despite recent attempts to document whole  
538 ecosystem meta-networks (Pocock, Evans & Memmott, 2012), it is likely that most  
539 ecological interaction networks will illustrate just major ecosystem compartments.  
540 Due to their high generalization, high temporal and spatial turnover, and high  
541 complexity of association patterns, adequate sampling of ecological interaction  
542 networks requires extremely large sampling effort. Undersampling of ecological  
543 networks may originate from the analysis of assemblage subsets (e.g., taxonomi-  
544 cally or functionally defined), and/or from logistically-limited sampling effort. It

Jordano - Sampling networks

545 is extremely hard to robustly sample the set of biotic interactions even for rela-  
546 tively simple, species-poor assemblages; yet, concluding that all ecological network  
547 datasets are undersampled would be unrealistic. The reason stems from a biologi-  
548 cal fact: a sizeable fraction of the maximum, potential links that can be recorded  
549 among two distinct sets of species is simply unobservable, irrespective of sampling  
550 effort (Jordano, 1987).

551 Missing links are a characteristic feature of all plant-animal interaction net-  
552 works, and likely pervade other ecological interactions. Important natural history  
553 details explain a fraction of them, resulting in unobservable [unrealized? Unob-  
554 servable suggests to me that the interaction may exist but we'll never be able to  
555 observe it] interactions (i.e., forbidden interactions) that define structural zeroes  
556 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.

565 Contrary to species inventories, a sizable fraction of non-observed pairwise  
566 interactions cannot be sampled, due to biological constraints that forbid their oc-  
567 currence. A re-scaling of traditional asymptotic estimates for interaction richness  
568 can be applied whenever the knowledge of natural history details about the study  
569 system is sufficient to estimate at least the main causes of forbidden links. More-

over, 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 their ecological complexity that need to be restored after perturbations

581 to rebuild functional ecosystems.

## 582 Acknowledgements

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

## 591 Data accessibility

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

## 597 References

598 Banasek-Richter, C., Cattin, M. & Bersier, L. (2004) Sampling effects and the ro-

- 599 robustness of quantitative and qualitative food-web descriptors. *Journal of Theoretical Biology*, **226**, 23–32.
- 601 Bascompte, J. & Jordano, P. (2014) *Mutualistic networks*. Monographs in Population Biology, No. 53. Princeton University Press, Princeton, NJ.
- 603 Bersier, L., Banasek-Richter, C. & Cattin, M. (2002) Quantitative descriptors of food-web matrices. *Ecology*, **83**, 2394–2407.
- 605 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.
- 608 Bosch, J., Martín González, A.M., Rodrigo, A. & Navarro, D. (2009) Plant-pollinator networks: adding the pollinator's perspective. *Ecology Letters*, **12**, 409–419.

---

Jordano - Sampling networks

- 611 Brose, U., Martinez, N. & Williams, R. (2003) Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*, **84**, 2364–2377.
- 614 Bunge, J. & Fitzpatrick, M. (1993) Estimating the number of species: a review. *Journal of the American Statistical Association*, **88**, 364–373.
- 616 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.
- 619 Chacoff, N.P., Vazquez, D.P., Lomascolo, S.B., Stevani, E.L., Dorado, J. & Padrón,

- 620 B. (2012) Evaluating sampling completeness in a desert plant-pollinator network.  
621 *Journal of Animal Ecology*, **81**, 190–200.
- 622 Chao, A. (2005) Species richness estimation. *Encyclopedia of Statistical Sciences*,  
623 pp. 7909–7916. Oxford University Press, New York, USA.
- 624 Chao, A., Hsieh, T.C., Chazdon, R.L., Colwell, R.K. & Gotelli, N.J. (2015) Un-  
625 veiling the species-rank abundance distribution by generalizing the Good-Turing  
626 sample coverage theory. *Ecology*, **96**, 1189–1201.
- 627 Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009)  
628 Undersampling bias: the null hypothesis for singleton species in tropical arthro-  
629 pod surveys. *Journal of Animal Ecology*, **78**, 573–584.
- 630 Cohen, J.E. (1978) *Food webs and niche space*. Princeton University Press, Prince-  
631 ton, New Jersey, US.

30

---

Jordano - Sampling networks

- 632 Cohen, J.E., Beaver, R.A., Cousins, S.H., DeAngelis, D.L., Goldwasser, L., Heong,  
633 K.L., Holt, R.D., Kohn, A.J., Lawton, J.H., Martinez, N., O’Malley, R., Page,  
634 L.M., Patten, B.C., Pimm, S.L., Polis, G., Rejmanek, M., Schoener, T.W.,  
635 Schenly, K., Sprules, W.G., Teal, J.M., Ulanowicz, R., Warren, P.H., Wilbur,  
636 H.M. & Yodis, P. (1993) Improving food webs. *Ecology*, **74**, 252–258.
- 637 Colwell, R. & Coddington, J. (1994) Estimating terrestrial biodiversity through ex-  
638 trapolation. *Philosophical Transactions Of The Royal Society Of London Series*  
639 *B-Biological Sciences*, **345**, 101–118.

- 640 Colwell, R.K. (2009) Biodiversity: concepts, patterns, and measurement. *The  
641 Princeton Guide to Ecology* (ed. S.A. Levin), pp. 257–263. Princeton University  
642 Press, Princeton.
- 643 Colwell, R.K. (2013) EstimateS: Biodiversity Estimation. -, pp. 1–33.
- 644 Colwell, R.K., Dunn, R.R. & Harris, N.C. (2012) Coextinction and persistence of  
645 dependent species in a changing world. *Annual Review of Ecology Evolution and  
646 Systematics*, **43**, 183–203.
- 647 Dorado, J., Vazquez, D.P., Stevani, E.L. & Chacoff, N.P. (2011) Rareness and  
648 specialization in plant-pollinator networks. *Ecology*, **92**, 19–25.
- 649 Dormann, C.F., Frund, J., Bluthgen, N. & Gruber, B. (2009) Indices, graphs and  
650 null models: Analyzing bipartite ecological networks. *Open Ecology Journal*, **2**,  
651 7–24.
- 652 Dupont, Y.L., Trøjelsgaard, K. & Olesen, J.M. (2011) Scaling down from species

---

Jordano - Sampling networks

- 653 to individuals: a flower visitation network between individual honeybees and  
654 thistle plants. *Oikos*, **120**, 170–177.
- 655 Dupont, Y.L., Trøjelsgaard, K., Hagen, M., Henriksen, M.V., Olesen, J.M., Ped-  
656 ersen, N.M.E. & Kissling, W.D. (2014) Spatial structure of an individual-based  
657 plant-pollinator network. *Oikos*, **123**, 1301–1310.
- 658 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P.,  
659 Dalsgaard, B., de Sassi, C., Galetti, M., Guimaraes, P.R., Lomáscolo, S.B.,  
660 Martín González, A.M., Pizo, M.A., Rader, R., Rodrigo, A., Tylianakis, J.M.,

- 661 Vazquez, D.P. & Allesina, S. (2013) The dimensionality of ecological networks.  
662 *Ecology Letters*, **16**, 577–583.
- 663 Elberling, H. & Olesen, J.M. (1999) The structure of a high latitude plant-flower  
664 visitor system: the dominance of flies. *Ecography*, **22**, 314–323.
- 665 Frund, J., McCann, K.S. & Williams, N.M. (2015) Sampling bias is a challenge  
666 for quantifying specialization and network structure: lessons from a quantitative  
667 niche model. *Oikos*, pp. n/a–n/a.
- 668 Gale, W.A. & Sampson, G. (1995) Good-Turing frequency estimation without  
669 tears. *Journal of Quantitative Linguistics*, **2**, 217–237.
- 670 Gibson, R.H., Knott, B., Eberlein, T. & Memmott, J. (2011) Sampling method  
671 influences the structure of plant-pollinator networks. *Oikos*, **120**, 822–831.
- 672 González-Varo, J.P., Arroyo, J.M. & Jordano, P. (2014) Who dispersed the seeds?  
673 The use of DNA barcoding in frugivory and seed dispersal studies. *Methods in  
674 Ecology and Evolution*, **5**, 806–814.

- 675 Good, I.J. (1953) The population frequencies of species and the estimation of  
676 population parameters. *Biometrika*, **40**, 237–264.
- 677 Gotelli, N.J. & Colwell, R.K. (2011) Estimating species richness. *Biological Di-  
678 versity Frontiers in Measurement and Assessment* (eds. A.E. Magurran & B.J.  
679 McGill), pp. 39–54. Oxford University Press, Oxford, UK.
- 680 Gotelli, N. & Colwell, R. (2001) Quantifying biodiversity: procedures and pitfalls  
681 in the measurement and comparison of species richness. *Ecology Letters*, **4**,

- 683 Hortal, J., Borges, P. & Gaspar, C. (2006) Evaluating the performance of species  
 684 richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*,  
 685 **75**, 274–287.
- 686 Ibanez, S. (2012) Optimizing size thresholds in a plant–pollinator interaction web:  
 687 towards a mechanistic understanding of ecological networks. *Oecologia*, **170**,  
 688 233–242.
- 689 Jordano, P. (1987) Patterns of mutualistic interactions in pollination and seed dis-  
 690 persal: connectance, dependence asymmetries, and coevolution. *The American  
 691 Naturalist*, **129**, 657–677.
- 692 Jordano, P., Bascompte, J. & Olesen, J. (2003) Invariant properties in coevolu-  
 693 tionary networks of plant-animal interactions. *Ecology Letters*, **6**, 69–81.
- 694 Jordano, P., Vázquez, D. & Bascompte, J. (2009) Redes complejas de interac-  
 695 ciones planta—animal. *Ecología y evolución de interacciones planta-animal* (eds.

- 
- 696 R. Medel, R. Dirzo & R. Zamora), pp. 17–41. Editorial Universitaria, Santiago,  
 697 Chile.
- 698 Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., Petitpierre, E. & Gomez-Zurita,  
 699 J. (2009) DNA barcoding insect-host plant associations. *Proceedings Of The  
 700 Royal Society B-Biological Sciences*, **276**, 639–648.
- 701 Magurran, A. (1988) *Ecological diversity and its measurement*. Princeton Univer-

- 702 sity Press, Princeton, US.
- 703 Mao, C. & Colwell, R.K. (2005) Estimation of species richness: mixture models,  
704 the role of rare species, and inferential challenges. *Ecology*, **86**, 1143–1153.
- 705 Martinez, N.D. (1993) Effects of resolution on food web structure. *Oikos*, **66**,  
706 403–412.
- 707 Martinez, N. (1991) Artifacts or attributes? Effects of resolution on food-web  
708 patterns in Little Rock Lake food web. *Ecological Monographs*, **61**, 367–392.
- 709 Maruyama, P.K., Vizentin-Bugoni, J., Oliveira, G.M., Oliveira, P.E. & Dalsgaard,  
710 B. (2014) Morphological and spatio-temporal mismatches shape a neotropical  
711 savanna plant-hummingbird network. *Biotropica*, **46**, 740–747.
- 712 Moré, M., Amorim, F.W., Benitez-Vieyra, S., Medina, A.M., Sazima, M. &  
713 Cocucci, A.A. (2012) Armament Imbalances: Match and Mismatch in Plant-  
714 Pollinator Traits of Highly Specialized Long-Spurred Orchids. *PLoS ONE*, **7**,  
715 e41878.
- 716 Nielsen, A. & Bascompte, J. (2007) Ecological networks, nestedness and sampling  
717 effort. *Journal of Ecology*, **95**, 1134–1141–1141.

---

Jordano - Sampling networks

- 718 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H. & Jordano, P. (2011)  
719 Missing and forbidden links in mutualistic networks. *Proceedings Of The Royal  
720 Society B-Biological Sciences*, **278**, 725–732.
- 721 Olesen, J. & Jordano, P. (2002) Geographic patterns in plant-pollinator mutualistic  
722 networks. *Ecology*, **83**, 2416–2424.
- 723 Olito, C. & Fox, I.W. (2014) Species traits and abundances predict metrics of

- 723 Zhou, C. & Liu, S.Y. (2011) Species status and abundance predict metrics of  
724 plant-pollinator network structure, but not pairwise interactions. *Oikos*, **124**,  
725 428–436.
- 726 Ollerton, J. & Cranmer, L. (2002) Latitudinal trends in plant-pollinator interac-  
727 tions: are tropical plants more specialised? *Oikos*, **98**, 340–350.
- 728 Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes,  
729 R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops,  
730 N., Dulloo, E., Faith, D., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurttt,  
731 G., Jetz, W., Karp, D.S., Mcgeoch, M., Obura, D., Onoda, Y., Pettorelli, N.,  
732 Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S., Turak, E., Walpole, M.  
733 & Wegmann, M. (2013) Essential biodiversity variables. *Science*, **339**, 277–278.
- 734 Pocock, M.J.O., Evans, D.M. & Memmott, J. (2012) The Robustness and Restora-  
735 tion of a Network of Ecological Networks. *Science*, **335**, 973–977.
- 736 Preston, F. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254–283.
- 737 R Development Core Team (2010) *R: A language and environment for statis-  
738 tical computing*. R Foundation for Statistical Computing. Vienna, Austria.  
739 <http://www.R-project.org>, Vienna, Austria.

- 761 Thébault, E. & Fontaine, C. (2010) Stability of ecological communities and the  
762 architecture of mutualistic and trophic networks. *Science*, **329**, 853–856.
- 763 Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A.,  
764 Galetti, M., García, M.B., García, D., Gomez, J.M., Jordano, P., Medel, R.,  
765 Navarro, L., Obeso, J.R., Oviedo, R., Ramírez, N., Rey, P.J., Traveset, A.,

766 Verdú, M. & Zamora, R. (2014) Beyond species loss: the extinction of ecological  
767 interactions in a changing world. *Functional Ecology*, **29**, 299–307.

768 Vazquez, D.P., Chacoff, N.P. & Cagnolo, L. (2009) Evaluating multiple deter-  
769 minants of the structure of plant-animal mutualistic networks. *Ecology*, **90**,  
770 2039–2046.

771 Vazquez, D.P., Ramos-Jiliberto, R., Urbani, P. & Valdovinos, F.S. (2015) A con-  
772 ceptual framework for studying the strength of plant-animal mutualistic inter-  
773 actions. *Ecology Letters*, **18**, 385–400.

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

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

780 **Figure captions**

781 **Figure 1.** Sampling ecological interaction networks (e.g., plant-animal interac-  
782 tions) usually focus on different types of subsampling the full network, yielding  
783 submatrices  $\Delta[m, n]$  of the full interaction matrix  $\Delta$  with  $A$  and  $P$  animal and

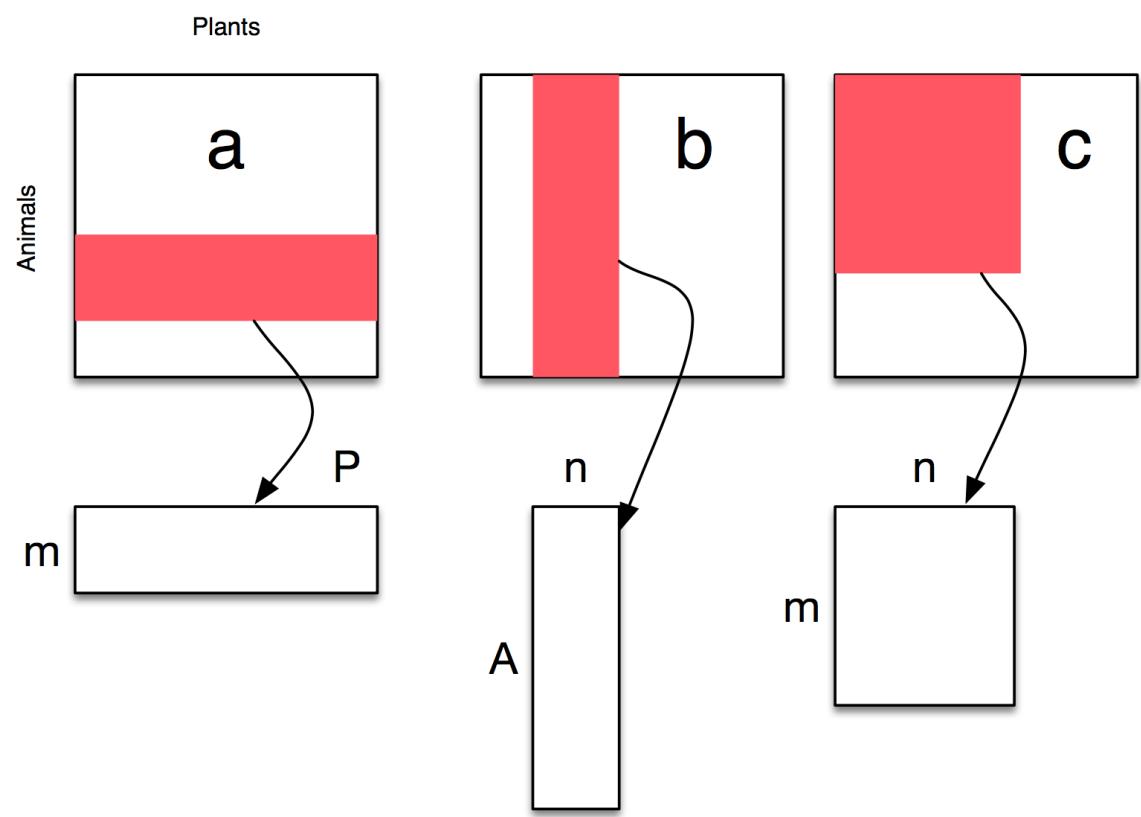
784 plant species, a) all the potential plants interacting with a subset of the animals  
785 (e.g., studying just the hummingbird-pollinated flower species in a community);  
786 b) all the potential animal species interacting with a subset of the plant species  
787 (e.g., studying the frugivore species feeding on figs *Ficus* in a community); and c)  
788 sampling a subset of all the potential animal species interacting with a subset of all  
789 the plant species (e.g., studying the plant-frugivore interactions of the rainforest  
790 understory).

791

792 **Figure 2.** Sampling species interactions in natural communities. Suppose an  
793 assemblage with  $A = 3$  animal species (red, species 1 3 with three, two, and 1  
794 individuals, respectively) and  $P = 3$  plant species (green, species a-c with three  
795 individuals each) (colored balls), sampled with increasing effort in steps 1 to 6  
796 (panels). In Step 1 we record animal species 1 and plant species 1 and 2 with  
797 a total of three interactions (black lines) represented as two distinct interactions:  
798 1 – a and 1 – b. As we advance our sampling (panels 1 to 6, illustrating e.g.,  
799 additional sampling days) we record new distinct interactions. Note that we actu-  
800 ally sample and record interactions among individuals, yet we pool the data across  
801 species to get a species by species interaction matrix. Few network analyses have  
802 been carried out on individual data(Dupont *et al.*, 2014).

803

804 **Figures**



39

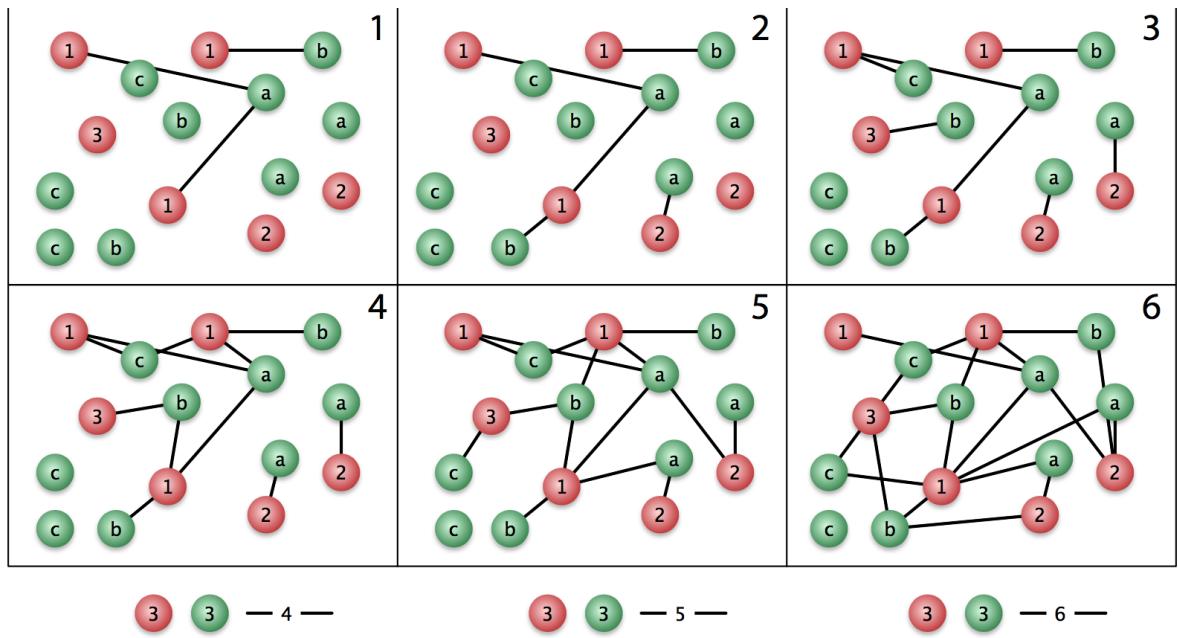
Jordano - Sampling networks

Figure 2:

1 2 — 2 —

2 2 — 2 —

3 3 — 4 —



Jordano – Figure 1

40

Jordano - Sampling networks

- 740 Rivera-Hutinel, A., Bustamante, R.O., Marín, V.H. & Medel, R. (2012) Effects of  
 741 sampling completeness on the structure of plant-pollinator networks. *Ecology*,  
 742 **93**, 1593–1603.

- 743 Schleuning, M., Frund, J., Klein, A.M., Abrahamszyk, S., Alarcón, R., Albrecht,  
744 M., Andersson, G.K.S., Bazarian, S., Böhning-Gaese, K., Bommarco, R., Dals-  
745 gaard, B., Dehling, D.M., Gotlieb, A., Hagen, M., Hickler, T., Holzschuh, A.,  
746 Kaiser-Bunbury, C.N., Kreft, H., Morris, R.J., Sandel, B., Sutherland, W.J.,  
747 Svenning, J.C., Tscharntke, T., Watts, S., Weiner, C.N., Werner, M., Williams,  
748 N.M., Winqvist, C., Dormann, C.F. & Blüthgen, N. (2012) Specialization of  
749 mutualistic interaction networks decreases toward tropical latitudes. *Current*  
750 *Biology*, **22**, 1925–1931.
- 751 Snow, B. & Snow, D. (1972) Feeding niches of hummingbirds in a Trinidad valley.  
752 *Journal of Animal Ecology*, **41**, 471–485.
- 753 Snow, B. & Snow, D. (1988) *Birds and berries*. Poyser, Calton, UK.
- 754 Stang, M., Klinkhamer, P., Waser, N.M., Stang, I. & van der Meijden, E. (2009)  
755 Size-specific interaction patterns and size matching in a plant-pollinator inter-  
756 action web. *Annals Of Botany*, **103**, 1459–1469.
- 757 Strogatz, S. (2001) Exploring complex networks. *Nature*, **410**, 268–276.
- 758 Stumpf, M.P.H., Wiuf, C. & May, R.M. (2005) Subnets of scale-free networks are  
759 not scale-free: Sampling properties of networks. *Proceedings of the National*  
760 *Academy of Sciences USA*, **102**, 4221–4224.

805 **Table captions**

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

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

813

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

823

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

825

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

830 (Chao & Jost 2012). Scaled asymptotic estimators and their confidence intervals  
831 (*CI*) were calculated by weighting *Chao1* and *ACE* with the observed frequencies  
832 of forbidden links.

833

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 | Pollination           |                           |                             | Seed dispersal  |                 |                  |
|-----------|-----------------------|---------------------------|-----------------------------|-----------------|-----------------|------------------|
|           | Zackenberg            | Grundvad                  | Arima Valley                | Sta. Virginia   | Hato Ratón      | Nava Correhuelas |
| $I_{max}$ | 1891                  | 646                       | 522                         | 423             | 272             | 825              |
| $I$       | 268<br>(0.1417)       | 212<br>(0.3282)           | 185<br>(0.3544)             | 86<br>(0.1042)  | 151<br>(0.4719) | 181<br>(0.2194)  |
| $UL$      | 1507<br>(0.7969)      | 434<br>(0.6718)           | 337<br>(0.6456)             | 337<br>(0.4085) | 169<br>(0.5281) | 644<br>(0.7806)  |
| $FL$      | 530<br>(0.3517)       | 107<br>(0.2465)           | 218<br>(0.6469)             | 260<br>(0.7715) | 118<br>(0.6982) | 302<br>(0.4689)  |
| $FL_P$    | 530<br>(1.0000)       | 94<br>(0.2166)            | 0<br>(0.0000)               | 120<br>(0.1624) | 67<br>(0.3964)  | 195<br>(0.3028)  |
| $FL_S$    | ... (...)<br>(0.0184) | 8<br>(0.0184)             | 30<br>(0.0890)              | 140<br>(0.1894) | 31<br>(0.1834)  | 46 (0.0714)      |
| $FL_A$    | ... (...)<br>(0.0115) | 5<br>(0.445) <sup>a</sup> | 150<br>(0.445) <sup>a</sup> | ... (...)       | 20<br>(0.1183)  | 61 (0.0947)      |
| $FL_O$    | ... (...)             | ... (...)                 | 38<br>(0.1128) <sup>b</sup> | ... (...)       | ... (...)       | 363<br>(0.5637)  |
| $ML$      | 977<br>(0.6483)       | 327<br>(0.7535)           | 119<br>(0.3531)             | 77<br>(0.1042)  | 51<br>(0.3018)  | 342<br>(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>i</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               |
| ...         | ...      | ...      | ...      | ... | ...             |
| Ai - Pi     | 1        | 0        | 1        | ... | 2               |

Table 4:

|                           | Hato Ratón       | Nava Correhuelas | Zackenberg       |
|---------------------------|------------------|------------------|------------------|
| $A$                       | 17               | 33               | 65               |
| $P$                       | 16               | 25               | 31               |
| $I_{max}$                 | 272              | 825              | 1891             |
| $N$                       | 3340             | 8378             | 1245             |
| $I$                       | 151              | 181              | 268              |
| $C$                       | 0.917            | 0.886            | 0.707            |
| $Chao1$                   | $263.1 \pm 70.9$ | $231.4 \pm 14.2$ | $509.6 \pm 54.7$ |
| $ACE$                     | $240.3 \pm 8.9$  | $241.3 \pm 7.9$  | $566.1 \pm 14.8$ |
| <i>Scaled Chao</i>        | 195.4            | 162.7            | 308.4            |
| $CI$                      | [124.5–266.3]    | [148.5–176.9]    | [253.6–363.1]    |
| <i>Scaled ACE</i>         | 178.5            | 169.7            | 342.6            |
| $CI$                      | [169.5–187.4]    | [161.8–177.6]    | [327.8–357.4]    |
| % unobserved <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)