Sampling networks of ecological interactions

Pedro Jordano\_a

aIntegrative Ecology Group, Estación Biológica de Doñana, Consejo

Superior de Investigaciones Científicas (EBD-CSIC), Avenida

Americo Vespucio s/n, E–41092 Sevilla, Spain

Sevilla, August 28, 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 species, such as those among plants and their seed dispersers in

frugivory interactions or those among plants and their pollinators. Sampling

interactions is a fundamental step to build robustly estimated interaction

networks, yet few analyses have attempted a formal approach to their sampling

protocols.

\_jordano@ebd.csic.es

1

Jordano - Sampling networks

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

Imax among, say, A animal species and P plant species (i.e., Imax = 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 << Imax, 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

2

Jordano - Sampling networks

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 species assessment aims at sampling individuals in collections and deter2

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 rich5

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 vari9

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, but in some cases, also the extinction of species.

3

Jordano - Sampling networks

13 For example, the extinction of interactions represents a dramatic loss of biodiver14

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 ecolog16

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 communi19

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 re27

lated biodiversity indicator variables (EBVs). Thus, sampling interactions should

28 be a central issue when identifying and diagnosing ecosystem services (e.g., polli29

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). Monitor33

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 Vázquez, 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

4

Jordano - Sampling networks

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 in42

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 the detected and undetected interactions as sample size increases

46 (Preston, 1948).

47 Early efforts to recognize and solve sampling problems in analyses of interac48

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; Mar50

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 influenc55

ing the patterns and metrics reported. Yet, despite the early recognition that

56 incomplete sampling may seriously bias the analysis of ecological networks (Jor57

dano, 1987), only recent studies have explicitly acknowledged it and attempted to

58 determine its influence (Ollerton & Cranmer, 2002; Nielsen & Bascompte, 2007;

59 Vázquez, Chacoff & Cagnolo, 2009; Gibson et al., 2011; Olesen et al., 2011; Cha60

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). The sampling approaches have been extended to predict patterns

5

Jordano - Sampling networks

63 of coextintions in interaction assemblages (e.g., hosts-parasites) (Colwell, Dunn &

64 Harris, 2012). Most empirical studies provide no estimate of sampling effort, im65

plicitly assuming that the reported network patterns and metrics are robust. Yet

66 recent evidences point out that number of partner species detected, number of

67 actual links, and some aggregate statistics describing network patterns, are prone

68 to sampling bias (Nielsen & Bascompte, 2007; Dorado et al., 2011; Olesen et al.,

69 2011; Chacoff et al., 2012; Rivera-Hutinel et al., 2012; Olito & Fox, 2014; Frund,

70 McCann & Williams, 2015). Most of these evidences, however, come from either

71 simulation studies (Frund, McCann & Williams, 2015) or from rel72

atively species-poor assemblages. Even for species-rich, tropical assemblages it

73 might be erroneous to conclude that network data routinely come from insuffi74

ciently sampled datasets (Ollerton & Cranmer, 2002; Chacoff et al., 2012), given

75 the extremely sparse nature of these interaction matrices because of the prevalence

76 of forbidden links (which, by definition, cannot be documented despite extensive

77 sampling effort). However, most certainly, sampling limitations pervade biodiver78

sity inventories in tropical areas (Coddington et al., 2009) and we might rightly

79 expect that frequent interactions may be over-represented and rare interactions

80 may be missed entirely in studies of mega-diverse assemblages (Bascompte & Jor81

dano, 2014); but, to what extent?

82 Sampling interactions: methods

83 When we sample interactions in the field we record the presence of two species

84 that interact in some way. For example, Snow and Snow(1988) recorded an inter85

action whenever they saw a bird “touching” a fruit on a plant. We observe and

6

Jordano - Sampling networks

86 record feeding observations, visitation, occupancy, presence in pollen loads or in

87 fecal samples, etc., of individual animals or plants and accumulate pairwise inter88

actions, i.e., lists of species partners and the frequencies with which we observe

89 them. Therefore, estimating the sampling completeness of pairwise interactions

90 for a whole network, requires estimating the number (richness) of distinct pairwise

91 interactions accumulated as sampling effort is increased, pooling the data for all

92 partner species.

93 Most, if not all, types of ecological interactions can be illustrated with bipar94

tite graphs, with two or more distinct groups of interacting partners (Bascompte &

95 Jordano, 2014); for illustration purposes I’ll focus more specifically on plant-animal

96 interactions. Sampling interactions requires filling the cells of an interaction ma97

trix with data. The matrix, \_ = AP, is a 2D representation of the interactions

98 among, say, A animal species (rows) and P plant species (columns) (Jordano,

99 1987; Bascompte & Jordano, 2014). An interaction matrix \_ consists of an array

100 of zeroes or ones, or an array of numeric values (including zeroes)- if the data

101 (interaction frequencies) are quantified. The matrix entries illustrate the values

102 of the pairwise interactions visualized in the \_ matrix, and can be 0 or 1, for

103 presence-absence of a given pairwise interaction, or take a quantitative weight wji

104 to represent the interaction intensity or unidirectional effect of species j on species

105 i (Bascompte & Jordano, 2014; Vázquez et al., 2015). Given that the outcomes of

106 most ecological interactions are dependent on frequency of encounters (e.g., visit

107 rate of pollinators, number of records of ant defenders, frequency of seeds in fecal

108 samples), a frequently used proxy for interaction intensities wji is just how fre109

quent new interspecific encounters are, whether or not appropriately weighted to

110 estimate interaction effectiveness (Vázquez, Morris & Jordano, 2005).

7

Jordano - Sampling networks

111 We need to define two basic steps in the sampling of interactions: 1) which

112 type of interactions we sample; and 2) which type of record we get to document

113 the existence of an interaction. In step #1 we need to take into account whether

114 we are sampling the whole community of interactor species (all the animals, all

115 the plants) or just a subset of them, i.e., a sub matrix \_m;n of m < A animal

116 species and n < P plant species of the adjacency matrix \_AP . Subsets can be:

117 a) all the potential plants interacting with a subset of the animals (Fig. 1a); b)

118 all the potential animal species interacting with a subset of the plant species (Fig.

119 1b); c) a subset of all the potential animal species interacting with a subset of all

120 the plant species (Fig. 1c). While some discussion has considered how to establish

121 the limits of what represents a network (Strogatz, 2001) (in analogy to discussion

122 on food-web limits; Cohen, 1978), it must be noted that situations a-c in Fig.

123 1 do not represent complete interaction networks. As vividly stated by Cohen

124 et al. (1993): “As more comprehensive, more detailed, more explicit webs become

125 available, smaller, highly aggregated, incompletely described webs may progressively

126 be dropped from analyses of web structure (though such webs may remain useful for

127 other purposes, such as pedagogy)”. Subnet sampling is generalized in studies of

128 biological networks (e.g., protein interactions, gene regulation), yet it is important

129 to recognize that most properties of subnetworks (even random subsamples) do

130 not represent properties of whole networks (Stumpf, Wiuf & May, 2005).

131 In step #2 above we face the problem of the type of record we take to sample

132 interactions. This is important because it defines whether we approach the problem

133 of filling up the interaction matrix in a “zoo-centric” way or in a “phyto-centric”

134 way. Zoo-centric studies directly sample animal activity and document the plants

135 ‘touched’ by the animal. For example, analysis of pollen samples recovered from the

8

Jordano - Sampling networks

136 body of pollinators, analysis of fecal samples of frugivores, radio-tracking data, etc.

137 Phyto-centric studies take samples of focal individual plant species and document

138 which animals ‘arrive’ or ‘touch’ the plants. Examples include focal watches of

139 fruiting or flowering plants to record visitation by animals, raising insect herbivores

140 from seed samples, identifying herbivory marks in samples of leaves, etc.

141 Most recent analyses of plant-animal interaction networks are phyto-centric;

142 just 3.5% of available plant-pollinator (N= 58) or 36.6% plant-frugivore (N= 22)

143 interaction datasets are zoo-centric (see Schleuning et al., 2012). Moreover, most

144 available datasets on host-parasite or plant-herbivore interactions are “host-centric”

145 or phyto-centric (e.g., Thébault & Fontaine, 2010; Eklöf et al., 2013). This may be

146 related to a variety of causes, like preferred methodologies by researchers working

147 with a particular group or system, logistic limitations, or inherent taxonomic focus

148 of the research questions. A likely result of phyto-centric sampling would be adja149

cency matrices with large A : P ratios. In any case we don’t have a clear view of

150 the potential biases that taxa-focused sampling may generate in observed network

151 patterns, for example by generating consistently asymmetric interaction matrices

152 (Dormann et al., 2009). System symmetry has been suggested to influence esti153

mations of generalization levels in plants and animals when measured as IA and

154 IP (Elberling & Olesen, 1999); thus, differences in IA and IP between networks

155 may arise from different A : P ratios rather than other ecological factors (Olesen

156 & Jordano, 2002).

157 Interestingly enough, quite complete analyses of interaction networks can be

158 obtained when combining both phyto-centric and zoo-centric sampling. For ex159

ample, Bosch et al. (Bosch et al., 2009) showed that the addition of pollen load

160 data on top of focal-plant sampling of pollinators unveiled a significant number

9

Jordano - Sampling networks

161 of interactions, resulting in important network structural changes. Connectance

162 increased 1.43-fold, mean plant connectivity went from 18.5 to 26.4, and mean

163 pollinator connectivity from 2.9 to 4.1; moreover, extreme specialist pollinator

164 species (singletons in the adjacency matrix) decreased 0.6-fold. Zoo-centric sam165

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

166 with plant-herbivore (Jurado-Rivera et al., 2009) and plant-frugivore interactions

167 (González-Varo, Arroyo & Jordano, 2014). For mutualistic networks we would ex168

pect that zoo-centric sampling could help unveiling interactions for rare species or

169 for relatively common species which are difficult to sample by direct observation.

170 Future methodological work may provide significant advances showing how mixing

171 different sampling strategies strengthens the completeness of network data. These

172 mixed strategies may combine, for instance, focal analyses, pollen load or seed

173 contents, camera traps, and DNA barcoding records. We might expect increased

174 power of these mixed sampling approaches when combining different methods from

175 both phyto- and zoo-centric perspectives (Bosch et al., 2009; Blüthgen, 2010).

176 Sampling interactions: rationale

177 The number of distinct pairwise interactions that we can record in a landscape (an

178 area of relatively homogeneous vegetation, analogous to the one we would use to

179 monitor species diversity) is equivalent to the number of distinct classes in which

180 we can classify the recorded encounters among individuals of two different species.

181 Yet, individual-based plant-animal interaction networks have been only recently

182 studied (Dupont, Trøjelsgaard & Olesen, 2011). We walk in the forest and see

183 a blackbird Tm picking an ivy Hh fruit and ingesting it: we have a record for

10

Jordano - Sampling networks

184 Tm 􀀀 Hh interaction. We keep advancing and record again a blackbird feeding

185 on hawthorn Cm fruits so we record a Tm 􀀀 Cm interaction; as we advance we

186 encounter another ivy plant and record a blackcap swallowing a fruit so we now

187 have a new Sa 􀀀 Hh interaction, and so on. At the end we have a series of

188 classes (e.g., Sa 􀀀 Hh, Tm 􀀀 Hh, Tm 􀀀 Cm, etc.), along with their observed

189 frequencies. Bunge & Fitzpatrick (1993) review the main aspects and approaches

190 to estimate the number of distinct classes C in a sample of observations. The

191 sampling of interactions in nature, as the sampling of species, is a cumulative

192 process. In our analysis, we are not re-sampling individuals, but interactions, so we

193 made interaction-based accumulation curves. If an interaction-based curve points

194 towards a robust sampling, it does mean that no new interactions are likely to be

195 recorded, irrespectively of the species, as it is a whole-network sampling approach

196 (N. Gotelli, pers. com.). We add new, distinct, interactions recorded as we increase

197 sampling effort (Fig. 2). We can obtain an Interaction Accumulation Curve (IAC)

198 analogous to a Species cumulating Curve (SAC): the observed number of distinct

199 pairwise interactions in a survey or collection as a function of the accumulated

200 number of observations or samples (Colwell, 2009).

201 Our sampling above would have resulted in a vector n = [n1:::nC]0 where ni is

202 the number of records in the ith class. As stressed by Bunge & Fitzpatrick (1993),

203 however, the ith class would appear in the sample if and only if ni > 0, and we

204 don’t know a priori which ni are zero. So, n is not observable. Rather, what we

205 get is a vector c = [c1:::cn]0 where cj is the number of classes represented j times

206 in our sampling: c1 is the number of singletons, c2 is the number of twin pairs, c3

207 the number of triplets, etc. The problem thus turns to be estimating the number

208 of distinct classes C from the vector of cj values.

11

Jordano - Sampling networks

209 Estimating the number of interactions with resulting robust estimates of net210

work parameters is a central issue in the study of ecological interaction networks

211 (Jordano, 1987; Bascompte & Jordano, 2014). In contrast with traditional species

212 diversity estimates, sampling networks has the paradox that despite the poten213

tially interacting species being present in the sampled assemblage (i.e., included in

214 the A and P species lists), some of their pairwise interactions are impossible to be

215 recorded. The reason is forbidden links. Independently of whether we sample full

216 communities or subset communities we face a problem: some of the interactions

217 that we can visualize in the empty adjacency matrix \_ will simply not occur. Thus,

218 independently of the sampling effort we put, we’ll never document these pairwise

219 interactions. With a total of AP “potential” interactions , a fraction of them are

220 impossible to record, because they are forbidden (Jordano, Bascompte & Olesen,

221 2003; Olesen et al., 2011). Forbidden links are constraints for the establishment of

222 new links, and mainly arise from the biological attributes of the species: no link

223 can be established between a plant and an animal mutualist differing in phenology,

224 i.e. the seeds of a winter-ripening plant cannot be dispersed by a frugivore that

225 is a summer stopover migrant (Jordano, 1987). Or, for instance, short-tongued

226 pollinators cannot successfully reach the nectar in long-corolla flowers and pol227

linate them efficiently (Moré et al., 2012). Forbidden links are thus represented

228 as structural zeroes in the interaction matrix, i.e., matrix cells that cannot get a

229 non-zero value. So, we need to account for the frequency of these structural zeros

230 in our matrix before proceeding. For example, most measurements of connectance

231 C = I=(AP) implicitly ignore the fact that by taking the full product AP in the

232 denominator they are underestimating the actual connectance value, i.e., the frac233

tion of actual interactions I relative to the biologically possible ones, not to the

12

Jordano - Sampling networks

234 total maximum Imax = AP.

235 Adjacency matrices are frequently sparse, i.e., they are densely populated with

236 zeroes, with a fraction of them being structural (i.e., unobservable interactions)

237 (Bascompte & Jordano, 2014). Thus it would be a serious interpretation error to

238 attribute the sparseness of adjacency matrices for bipartite networks to under239

sampling. The actual typology of link types in ecological interaction networks is

240 thus more complex than just the two categories of observed and unobserved inter241

actions (Table 1). Unobserved interactions are represented by zeroes and belong

242 to two categories. Missing interactions may actually exist but require additional

243 sampling or a variety of methods to be observed. Forbidden links, on the other

244 hand, arise due to biological constraints limiting interactions and remain unobserv245

able in nature, irrespectively of sampling effort (Jordano, Bascompte & Olesen,

246 2003; Olesen et al., 2011). Forbidden links are non-occurrences of pairwise interac247

tions that can be accounted for by biological constraints, such as spatio-temporal

248 uncoupling, size or reward mismatching, foraging constraints (e.g., accessibility),

249 and physiological-biochemical constraints (Jordano, 1987). We still have extremely

250 reduced information about the frequency of forbidden links in natural communi251

ties (Jordano, Bascompte & Olesen, 2003; Stang et al., 2009; Vázquez, Chacoff &

252 Cagnolo, 2009; Olesen et al., 2011; Ibanez, 2012; Maruyama et al., 2014; Vizentin-

253 Bugoni, Maruyama & Sazima, 2014) (Table 1). Forbidden links FL may actually

254 account for a relatively large fraction of unobserved interactions UL when sam255

pling taxonomically-restricted subnetworks (e.g., plant-hummingbird pollination

256 networks) (Table 1). Phenological unmatching is also prevalent in most networks,

257 and may add up to explain ca. 25–40% of the forbidden links, especially in highly

258 seasonal habitats, and up to 20% when estimated relative to the total number

13

Jordano - Sampling networks

of unobserved 259 interactions (Table 2). In any case, we might expect that a frac260

tion of the missing links ML would be eventually explained by further biological

261 reasons, depending on the knowledge of natural details of the particular systems.

262 Our goal as naturalists would be to reduce the fraction of UL which remain as

263 missing links; to this end we might search for additional biological constraints or

264 increase sampling effort. For instance, habitat use patterns by hummingbirds in

265 the Arima Valley network (Table 2; Snow & Snow, 1972) impose a marked pattern

266 of microhabitat mismatches causing up to 44.5% of the forbidden links. A myr267

iad of biological causes beyond those included as FL in Table 2 may contribute

268 explanations for UL: limits of color perception and or partial preferences, pres269

ence of secondary metabolites in fruit pulp and leaves, toxins and combinations

270 of monosaccharides in nectar, etc. However, it is surprising that just the limited

271 set of forbidden link types considered in Table 1 explain between 24.6–77.2% of

272 the unobserved links. Notably, the Arima Valley, Santa Virgnia, and Hato Ratón

273 networks have > 60% of the unobserved links explained, which might be related

274 to the fact that they are subnetworks (Arima Valley, Santa Virgínia) or relatively

275 small networks (Hato Ratón). All this means that empirical networks may have

276 sizable fractions of structural zeroes. Ignoring this biological fact may contribute

277 to wrongly infering undersampling of interactions in real-world assemblages.

278 To sum up, two elements of inference are required in the analysis of unobserved

279 interactions in ecological interaction networks: first, detailed natural history infor280

mation on the participant species that allows the inference of biological constraints

281 imposing forbidden links, so that structural zeroes can by identified in the adja282

cency matrix; second, a critical analysis of sampling robustness a robust estimate

283 of the actual fraction of missing links, M, and thus, a robust estimate of I.

14

Jordano - Sampling networks

284 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, Sobs is the total number of species observed in a sample, or in a set

of samples. Sest 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 Sk be the number of species each represented by exactly k individuals

in a single sample. Thus, S0 is the number of undetected species (species

present in the community but not included in the sample), S1 is the number of

singleton species, S2 is the number of doubleton species, etc. The total number of

individuals in the sample would be:

n =

SXobs

k=1

Sk

285

286 A frequently used asymptotic, bias corrected, non-parametric estimator is SChao

287 (Hortal, Borges & Gaspar, 2006; Chao, 2005; Colwell, 2013):

SChao = Sobs +

S1(S1 􀀀 q \_\_\_\_\_\_\_\_1)

2(S2 + 1)

288 Another frequently used alternative is the Chao2 estimator, SChao2 (Gotelli &

289 Colwell, 2001), which has been reported to have a limited bias for small sample

290 sizes (Colwell & Coddington, 1994; Chao, 2005):

15

Jordano - Sampling networks

SChao2 = Sobs +

S2

1

2S2

A plot of the 291 cumulative number of species recorded, Sn, as a function of

292 some measure of sampling effort (say, n samples taken) yields the species accu293

mulation curve (SAC) or collector’s curve (Colwell & Coddington, 1994). Such a

294 curve eventually reaches an asymptote converging with Sest. In an analogous way,

295 interaction accumulation curves (IAC), analogous to SACs, can be used to as296

sess the robustness of interactions sampling for plant-animal community datasets

297 (Jordano, 1987; Jordano, Vázquez & Bascompte, 2009; Olesen et al., 2011). For

298 instance, a random accumulator function (e.g., library vegan in the R Package, R

299 Development Core Team, 2010) which finds the mean IAC and its standard devia300

tion from random permutations of the data, or subsampling without replacement

301 (Gotelli & Colwell, 2001) can be used to estimate the expected number of distinct

302 pairwise interactions included in a given sampling of records (Jordano, Vázquez

303 & Bascompte, 2009; Olesen et al., 2011). We start with a vectorized interaction

304 matrix representing the pairwise interactions (rows) recorded during a cumulative

305 number of censuses or sampling periods (columns) (Table 3) , in a way analogous

306 to a biodiversity sampling matrix with species as rows and sampling units (e.g.,

307 quadrats) as columns (Jordano, Vázquez & Bascompte, 2009). In this way we

308 effectively extend sampling theory developed for species diversity to the sampling

309 of ecological interactions. Yet future theoretical work will be needed to formally

310 assess the similarities and differences in the two approaches and developing biolog311

ically meaningful null models of expected interaction richness with added sampling

312 effort.

16

Jordano - Sampling networks

313 Assessing sampling effort when recording interac314

tions

315 The basic method we can propose to estimate sampling effort and explicitly show

316 the analogues with rarefaction analysis in biodiversity research is to vectorize the

317 interaction matrix AP so that we get a vector of all the potential pairwise inter318

actions (Imax, Table 1) that can occur in a community of A animal species and

319 P plant species. The new “species” we aim to sample are the pairwise interac320

tions (Table 3). So, if we have in our community Turdus merula (Tm) and Rosa

321 canina (Rc) and Prunus mahaleb (Pm), our problem will be to sample 2 new

322 “species”: Tm 􀀀 Rc and Tm 􀀀 Pm. In general, if we have A = 1:::i , animal

323 species and P = 1:::j plant species, we’ll have a vector of “new” species to sample:

324 A1P1;A1P2; :::A2P1;A2P2; :::AiPj . We can represent the successive samples where

325 we can potentially get records of these interactions in a matrix with the vectorized

326 interaction matrix and columns representing the successive samples we take (Table

327 3). This is simply a vectorized version of the interaction matrix.

328 For example, mixture models incorporating detectabilities have been proposed

329 to effectively account for rare species (Mao & Colwell, 2005). In an analogous line,

330 mixture models could be extended to samples of pairwise interactions, also with

331 specific detectability values. These detection rate/odds could be variable among

332 groups of interactions, depending on their specific detectability. For example,

333 detectability of flower-pollinator interactions involving bumblebees could have a

334 higher detectability than flower-pollinator pairwise interactions involving, say, ni335

tidulid beetles. These more homogeneous groupings of pairwise interactions within

336 a network define modules (Bascompte & Jordano, 2014), so we might expect that

17

Jordano - Sampling networks

interactions 337 of a given module (e.g., plants and their hummingbird pollinators; Fig.

338 1a) may share similar detectability values, in an analogous way to species groups

339 receiving homogeneous detectability values in mixture models (Mao & Colwell,

340 2005). Such sampling, in its simplest form, would result in a sample with multiple

341 pairwise interactions detected, in which the number of interaction events recorded

342 for each distinct interaction found in the sample is recorded (i.e., a column vector

343 in Table 3, corresponding to, say, a sampling day). The number of interactions

344 recorded for the ith pairwise interaction (i.e., AiPj in Table 3), Yi could be treated

345 as a Poisson random variable with a mean parameter \_i, its detection rate. Mix346

ture models (Mao & Colwell, 2005) include estimates for abundance-based data

347 (their analogous in interaction sampling would be weighted data), where Yi is

348 a Poisson random variable with detection rate \_i. This is combined with the

349 incidence-based model, where Yi is a binomial random variable (their analogous

350 in interaction sampling would be presence/absence records of interactions) with

351 detection odds \_i. Let T be the number of samples in an incidence-based data set.

352 A Poisson/binomial density can be written as (Mao & Colwell, 2005):

g(y; \_) =

8>><

>>:

\_y

y!e\_ [1]

\_

T

y

\_

\_y

(1+\_)T [2]

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

354 The detection rates \_i depend on the relative abundances \_i of the interactions,

355 the probability of a pairwise interaction being detected when it is present, and the

356 sample size (the number of interactions recorded), which, in turn, is a function

357 of the sampling effort. Unfortunately, no specific sampling model has been de-

18

Jordano - Sampling networks

358 veloped along these lines for species interactions and their characteristic features.

359 For example, a complication factor might be that interaction abundances, \_i, in

360 real assemblages are a function of the abundances of interacting species, that de361

termine interspecific encounter rates; yet they also depend on biological factors

362 that ultimately determine if the interaction occurs when the partner species are

363 present. In its simplest form, \_i could be estimated from just the product of part364

ner species abundances, an approach recently used as a null model to assess the

365 role of biological constraints in generating forbidden links and explaining interac366

tion patterns (Vizentin-Bugoni, Maruyama & Sazima, 2014). Yet more complex

367 models should incorporate not only interspecific encounter probabilities, but also

368 phenotypic matching and incidence of forbidden links.

369 Rarefaction analysis and diversity-accumulation analysis (Magurran, 1988; Hor370

tal, Borges & Gaspar, 2006) come up immediately with this type of dataset. This

371 procedure plots the accumulation curve for the expected number of distinct pair372

wise interactions recorded with increasing sampling effort (Jordano, Vázquez &

373 Bascompte, 2009; Olesen et al., 2011). Asymptotic estimates of interaction rich374

ness and its associated standard errors and confidence intervals can thus be ob375

tained (Hortal, Borges & Gaspar, 2006). It should be noted that the asymptotic

376 estimate of interaction richness implicitly ignores the fact that, due to forbidden

377 links, a number of pairwise interactions among the Imax number specified in the

378 adjacency matrix \_ cannot be recorded, irrespective of sampling effort. Therefore,

379 the asymptotic value most likely is an overestimate of the actual maximum number

380 of links that can be present in an assemblage. If forbidden links are taken into ac381

count, the asymptotic estimate should be lower. Yet, to the best of my knowledge,

382 there is no theory developed to estimate this “biologically real” asymptotic value.

19

Jordano - Sampling networks

383 Not unexpectedly, most recent analyses of sampling effort in ecological network

384 studies found evidences of undersampling (Chacoff et al., 2012). This needs not

385 to be true, especially when interaction subwebs are studied (Olesen et al., 2011;

386 Vizentin-Bugoni, Maruyama & Sazima, 2014), and once the issue of structural

387 zeroes in the interaction matrices is effectively incorporated in the estimates.

388 The real missing links

389 Given that a fraction of unobserved interactions can be accounted for by for390

bidden links, what about the remaining missing interactions? We have already

391 discussed that some of these could still be related to unaccounted constraints, and

392 still others would be certainly attributable to insufficient sampling. Would this

393 always be the case? Multispecific assemblages of distinct taxonomic relatedness,

394 whose interactions can be represented as bipartite networks (e.g., host-parasite,

395 plant-animal mutualisms, plant-herbivore interactions- with two distinct sets of

396 unrelated higher taxa), are shaped by interspecific encounters among individuals

397 of the partner species (Fig. 2). A crucial ecological aspect limiting these inter398

actions is the probability of interspecific encounter, i.e., the probability that two

399 individuals of the partner species actually encounter each other in nature.

400 Given log-normally distributed abundances of the two species groups, the ex401

pected “neutral” probabilities of interspecific encounter (PIE) would be simply

402 the product of the two lognormal distributions. Thus, we might expect that for

403 low PIE values, pairwise interactions would be either extremely difficult to sam404

ple, or just simply non-occurring in nature. Consider the Nava de las Correhuelas

405 interaction web (NCH, Table 2), with A = 36, P = 25, I = 181, and almost

20

Jordano - Sampling networks

406 half of the unobserved interactions not accounted for by forbidden links missing

407 links, M = 53.1%. Given the robust sampling of this network (Jordano, Vázquez

408 & Bascompte, 2009), a sizable fraction of these possible but missing links would

409 be simply not occurring in nature, most likely by extremely low PIE, in fact

410 asymptotically zero. Given the vectorized list of pairwise interactions for NCH, I

411 computed the PIE values for each one by multiplying element wise the two species

412 abundance distributions. The PIEmax = 0.0597, being a neutral estimate, based

413 on the assumption that interactions occur in proportion to the species-specific local

414 abundances. With PIEmedian < 1:4 10􀀀4 we may safely expect (note the quantile

415 estimate Q75% =3:27 10􀀀4) that a sizable fraction of these missing interactions

416 may simply not occur according to this neutral expectation (Jordano, 1987; Ole417

sen et al., 2011) (neutral forbidden links, sensu Canard et al., 2012). Which is the

418 expected frequency for pairwise interactions? and, which is the expected proba419

bility for unobserved interactions? More specifically, which is the probability of

420 missing interactions, M (i.e., the unobserved ones that cannot be accounted for as

421 forbidden links)?

422 When we consider the vectorized interaction matrix, enumerating all pairwise

423 interactions for the AP combinations, the expected probabilities of finding a given

424 interaction can be estimated with a Good-Turing approximation (Good, 1953).

425 The technique, developed by Alan Turing and I.J. Good with applications to lin426

guistics and word analysis (Gale & Sampson, 1995) has been recently applied in

427 ecology (Chao et al., 2015), estimates the probability of recording an interaction

428 of a hitherto unseen pair of partners, given a set of past records of interactions

429 between other species pairs. Let a sample of N interactions so that nr distinct

430 pairwise interactions have exactly r records. All Good-Turing estimators obtain

21

Jordano - Sampling networks

431 the underlying frequencies of events as:

P(X) =

(NX + 1)

T

(1 􀀀

E(1)

T

) (1)

432 where X is the pairwise interaction, NX is the number of times interaction X

433 is recorded, T is the sample size (number of distinct interactions recorded) and

434 E(1) is an estimate of how many different interactions were recorded exactly once.

435 Strictly speaking Equation (1) gives the probability that the next interaction type

436 recorded will be X, after sampling a given assemblage of interacting species. In

other words, we scale down the maximum-likelihood estimator n

T 437 by a factor of

1􀀀E(1)

T 438 . This reduces all the probabilities for interactions we have recorded, and

439 makes room for interactions we haven’t seen. If we sum over the interactions we

have seen, then the sum of P(X) is 1 􀀀 1􀀀E(1)

T 440 . Because probabilities sum to one,

we have the left-over probability of Pnew = E(1)

T 441 of seeing something new, where

442 new means that we sample a new pairwise interaction.

443 Note, however, that Good-Turing estimators, as the traditional asymptotic

444 estimators, do not account in our case for the forbidden interactions. To account for

445 these FL I re-scaled the asymptotic estimates, so that a more meaningful estimate

446 could be obtained (Table 4). The scaling was calculated as [Chao1\_(I+ML)]=AP,

447 just correcting for the FL frequency, given that I+ML represent the total feasible

448 interactions when discounting the forbidden links (Table 1). After scaling, observed

449 I values (Table 2) are within the Chao1 and ACE asymptotic estimates but below

450 the ACE estimates for Hato Ratón and Zackenberg (Table 4). Thus, even after

451 re-scaling for FL, it is likely that adequate characterization of most interaction

452 networks will require intensive sampling effort.

22

Jordano - Sampling networks

453 Discussion

454 Recent work has inferred that most data available for interaction networks are

455 incomplete due to undersampling, resulting in a variety of biased parameters and

456 network patterns (Chacoff et al., 2012). It is important to note, however, that

457 in practice, many surveyed networks to date have been subnets of much larger

458 networks. This is true for protein interaction, gene regulation, and metabolic

459 networks, where only a subset of the molecular entities in a cell have been sam460

pled (Stumpf, Wiuf & May, 2005). Despite recent attempts to document whole

461 ecosystem meta-networks (Pocock, Evans & Memmott, 2012), it is likely that most

462 ecological interaction networks will illustrate just major ecosystem compartments.

463 Due to their high generalization, high temporal and spatial turnover, and high

464 complexity of association patterns, adequate sampling of ecological interaction

465 networks requires extremely large sampling effort. Undersampling of ecological

466 networks may originate from the analysis of assemblage subsets (e.g., taxonomi467

cally or functionally defined), and/or from logistically-limited sampling effort. It

468 is extremely hard to robustly sample the set of biotic interactions even for rela469

tively simple, species-poor assemblages; yet, concluding that all ecological network

470 datasets are undersampled would be unrealistic. The reason stems from a biologi471

cal fact: a sizeable fraction of the maximum, potential links that can be recorded

472 among two distinct sets of species is simply unobservable, irrespective of sampling

473 effort (Jordano, 1987).

474 Missing links are a characteristic feature of all plant-animal interaction net475

works, and likely pervade other ecological interactions. Important natural history

476 details explain a fraction of them, resulting in unobservable interactions (i.e., for-

23

Jordano - Sampling networks

477 bidden interactions) that define structural zeroes in the interaction matrices and

478 contribute to their extreme sparseness. Sampling interactions is a way to monitor

479 biodiversity beyond the simple enumeration of component species and to develop

480 efficient and robust inventories of functional interactions. Yet no sampling theory

481 for interactions is available. Some key components of this sampling are analo482

gous to species sampling and traditional biodiversity inventories; however, there

483 are important differences. Focusing just on the realized interactions or treating

484 missing interactions as the expected unique result of sampling bias would miss

485 important components to understand how mutualisms coevolve within complex

486 webs of interdependence among species.

487 Contrary to species inventories, a sizable fraction of non-observed pairwise

488 interactions cannot be sampled, due to biological constraints that forbid their oc489

currence. A re-scaling of traditional asymptotic estimates for interaction richness

490 can be applied whenever the knowledge of natural history details about the study

491 system is sufficient to estimate at least the main causes of forbidden links. More492

over, recent implementations of inference methods for unobserved species (Chao

493 et al., 2015) can be combined with the forbidden link approach, yet they do not

494 account either for the existence of these ecological constraints.

495 Ecological interactions provide the wireframe supporting the lives of species,

496 and they also embed crucial ecosystem functions which are fundamental for sup497

porting the Earth system. Yet we still have a limited knowledge of the biodiversity

498 of ecological interactions, but they are being lost (extinct) at a very fast pace, fre499

quently preceding species extinctions (Valiente-Banuet et al., 2014). We urgently

500 need robust techniques to assess the completeness of ecological interactions net501

works because this knowledge will allow the identification of the minimal compo-

24

Jordano - Sampling networks

502 nents of their ecological complexity that need to be restored after perturbations

503 to rebuild functional ecosystems.

504 Acknowledgements

505 I am indebted to Robert Colwell, Jens M. Olesen, Alfredo Valido, Jordi Bas506

compte, Thomas Lewinshon, John N. Thompson, Nick Gotelli, Carsten Dormann,

507 and Paulo R. Guimara˜es Jr. for useful and thoughtful comments and discussion

508 at different stages of this manuscript. The study was supported by a Junta de

509 Andalucía Excellence Grant (RNM–5731), as well as a Severo Ochoa Excellence

510 Award from the Ministerio de Economía y Competitividad (SEV–2012–0262). The

511 Agencia de Medio Ambiente, Junta de Andalucía, provided generous facilities that

512 made possible my long-term field work in different natural parks.

513 Data accessiblity

514 This review does not use new raw data, but includes some re-analyses of previously

515 published material. All the original data supporting the paper, R code, supple516

mentary figures, and summaries of analytical protocols is available at the author’s

517 GitHub repository (https://github.com/pedroj/MS\_Network-Sampling), with

518 DOI: 10.5281/zenodo.29437.

25

Jordano - Sampling networks

519 References

520 Banasek-Richter, C., Cattin, M. & Bersier, L. (2004) Sampling effects and the ro521

bustness of quantitative and qualitative food-web descriptors. Journal of Theo522

retical Biology, 226, 23–32.

523 Bascompte, J. & Jordano, P. (2014) Mutualistic networks. Monographs in Popu524

lation Biology, No. 53. Princeton University Press, Princeton, NJ.

525 Bersier, L., Banasek-Richter, C. & Cattin, M. (2002) Quantitative descriptors of

526 food-web matrices. Ecology, 83, 2394–2407.

527 Blüthgen, N. (2010) Why network analysis is often disconnected from community

528 ecology: A critique and an ecologist’s guide. Basic And Applied Ecology, 11,

529 185–195.

530 Bosch, J., Martín González, A.M., Rodrigo, A. & Navarro, D. (2009) Plant531

pollinator networks: adding the pollinator’s perspective. Ecology Letters, 12,

532 409–419.

533 Brose, U., Martinez, N. & Williams, R. (2003) Estimating species richness: Sen534

sitivity to sample coverage and insensitivity to spatial patterns. Ecology, 84,

535 2364–2377.

536 Bunge, J. & Fitzpatrick, M. (1993) Estimating the number of species: a review.

537 Journal of the American Statistical Association, 88, 364–373.

538 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot,

539 D. (2012) Emergence of structural patterns in neutral trophic networks. PLoS

540 ONE, 7, e38295.

26

Jordano - Sampling networks

541 Chacoff, N.P., Vázquez, D.P., Lomascolo, S.B., Stevani, E.L., Dorado, J. & Padrón,

542 B. (2012) Evaluating sampling completeness in a desert plant-pollinator network.

543 Journal of Animal Ecology, 81, 190–200.

544 Chao, A. (2005) Species richness estimation. Encyclopedia of Statistical Sciences,

545 pp. 7909–7916. Oxford University Press, New York, USA.

546 Chao, A., Hsieh, T.C., Chazdon, R.L., Colwell, R.K. & Gotelli, N.J. (2015) Un547

veiling the species-rank abundance distribution by generalizing the Good-Turing

548 sample coverage theory. Ecology, 96, 1189–1201.

549 Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009)

550 Undersampling bias: the null hypothesis for singleton species in tropical arthro551

pod surveys. Journal of Animal Ecology, 78, 573–584.

552 Cohen, J.E. (1978) Food webs and niche space. Princeton University Press, Prince553

ton, New Jersey, US.

554 Cohen, J.E., Beaver, R.A., Cousins, S.H., DeAngelis, D.L., Goldwasser, L., Heong,

555 K.L., Holt, R.D., Kohn, A.J., Lawton, J.H., Martinez, N., O’Malley, R., Page,

556 L.M., Patten, B.C., Pimm, S.L., Polis, G., Rejmanek, M., Schoener, T.W.,

557 Schenly, K., Sprules, W.G., Teal, J.M., Ulanowicz, R., Warren, P.H., Wilbur,

558 H.M. & Yodis, P. (1993) Improving food webs. Ecology, 74, 252–258.

559 Colwell, R. & Coddington, J. (1994) Estimating terrestrial biodiversity through ex560

trapolation. Philosophical Transactions Of The Royal Society Of London Series

561 B-Biological Sciences, 345, 101–118.

562 Colwell, R.K. (2009) Biodiversity: concepts, patterns, and measurement. The

27

Jordano - Sampling networks

563 Princeton Guide to Ecology (ed. S.A. Levin), pp. 257–263. Princeton University

564 Press, Princeton.

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

566 Colwell, R.K., Dunn, R.R. & Harris, N.C. (2012) Coextinction and persistence of

567 dependent species in a changing world. Annual Review of Ecology Evolution and

568 Systematics, 43, 183–203.

569 Dorado, J., Vázquez, D.P., Stevani, E.L. & Chacoff, N.P. (2011) Rareness and

570 specialization in plant-pollinator networks. Ecology, 92, 19–25.

571 Dormann, C.F., Frund, J., Blüthgen, N. & Gruber, B. (2009) Indices, graphs and

572 null models: Analyzing bipartite ecological networks. Open Ecology Journal, 2,

573 7–24.

574 Dupont, Y.L., Trøjelsgaard, K. & Olesen, J.M. (2011) Scaling down from species

575 to individuals: a flower–visitation network between individual honeybees and

576 thistle plants. Oikos, 120, 170–177.

577 Dupont, Y.L., Trøjelsgaard, K., Hagen, M., Henriksen, M.V., Olesen, J.M., Ped578

ersen, N.M.E. & Kissling, W.D. (2014) Spatial structure of an individual-based

579 plant-pollinator network. Oikos, 123, 1301–1310.

580 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P.,

581 Dalsgaard, B., de Sassi, C., Galetti, M., Guimaraes, P.R., Lomáscolo, S.B.,

582 Martín González, A.M., Pizo, M.A., Rader, R., Rodrigo, A., Tylianakis, J.M.,

583 Vázquez, D.P. & Allesina, S. (2013) The dimensionality of ecological networks.

584 Ecology Letters, 16, 577–583.

28

Jordano - Sampling networks

585 Elberling, H. & Olesen, J.M. (1999) The structure of a high latitude plant-flower

586 visitor system: the dominance of flies. Ecography, 22, 314–323.

587 Frund, J., McCann, K.S. & Williams, N.M. (2015) Sampling bias is a challenge

588 for quantifying specialization and network structure: lessons from a quantitative

589 niche model. Oikos, pp. n/a–n/a.

590 Gale, W.A. & Sampson, G. (1995) Good-Turing frequency estimation without

591 tears. Journal of Quantitative Linguistics, 2, 217–237.

592 Gibson, R.H., Knott, B., Eberlein, T. & Memmott, J. (2011) Sampling method

593 influences the structure of plant–pollinator networks. Oikos, 120, 822–831.

594 González-Varo, J.P., Arroyo, J.M. & Jordano, P. (2014) Who dispersed the seeds?

595 The use of DNA barcoding in frugivory and seed dispersal studies. Methods in

596 Ecology and Evolution, 5, 806–814.

597 Good, I.J. (1953) The population frequencies of species and the estimation of

598 population parameters. Biometrika, 40, 237–264.

599 Gotelli, N.J. & Colwell, R.K. (2011) Estimating species richness. Biological Di600

versity Frontiers in Measurement and Assessment (eds. A.E. Magurran & B.J.

601 McGill), pp. 39–54. Oxford University Press, Oxford, UK.

602 Gotelli, N. & Colwell, R. (2001) Quantifying biodiversity: procedures and pitfalls

603 in the measurement and comparison of species richness. Ecology Letters, 4,

604 379–391.

605 Hortal, J., Borges, P. & Gaspar, C. (2006) Evaluating the performance of species

29

Jordano - Sampling networks

606 richness estimators: sensitivity to sample grain size. Journal of Animal Ecology,

607 75, 274–287.

608 Ibanez, S. (2012) Optimizing size thresholds in a plant–pollinator interaction web:

609 towards a mechanistic understanding of ecological networks. Oecologia, 170,

610 233–242.

611 Jordano, P. (1987) Patterns of mutualistic interactions in pollination and seed dis612

persal: connectance, dependence asymmetries, and coevolution. The American

613 Naturalist, 129, 657–677.

614 Jordano, P., Bascompte, J. & Olesen, J. (2003) Invariant properties in coevolu615

tionary networks of plant-animal interactions. Ecology Letters, 6, 69–81.

616 Jordano, P., Vázquez, D. & Bascompte, J. (2009) Redes complejas de interac617

ciones planta—animal. Ecología y evolución de interacciones planta-animal (eds.

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

619 Chile.

620 Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., Petitpierre, E. & Gomez-Zurita,

621 J. (2009) DNA barcoding insect-host plant associations. Proceedings of the Royal

622 Society B-Biological Sciences, 276, 639–648.

623 Magurran, A. (1988) Ecological diversity and its measurement. Princeton Univer624

sity Press, Princeton, US.

625 Mao, C. & Colwell, R.K. (2005) Estimation of species richness: mixture models,

626 the role of rare species, and inferential challenges. Ecology, 86, 1143–1153.

30

Jordano - Sampling networks

627 Martinez, N.D. (1993) Effects of resolution on food web structure. Oikos, 66,

628 403–412.

629 Martinez, N. (1991) Artifacts or attributes? Effects of resolution on food-web

630 patterns in Little Rock Lake food web. Ecological Monographs, 61, 367–392.

631 Maruyama, P.K., Vizentin-Bugoni, J., Oliveira, G.M., Oliveira, P.E. & Dalsgaard,

632 B. (2014) Morphological and spatio-temporal mismatches shape a neotropical

633 savanna plant-hummingbird network. Biotropica, 46, 740–747.

634 Moré, M., Amorim, F.W., Benitez-Vieyra, S., Medina, A.M., Sazima, M. &

635 Cocucci, A.A. (2012) Armament Imbalances: Match and Mismatch in Plant-

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

637 e41878.

638 Nielsen, A. & Bascompte, J. (2007) Ecological networks, nestedness and sampling

639 effort. Journal of Ecology, 95, 1134–1141–1141.

640 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H. & Jordano, P. (2011)

641 Missing and forbidden links in mutualistic networks. Proceedings of the Royal

642 Society B-Biological Sciences, 278, 725–732.

643 Olesen, J. & Jordano, P. (2002) Geographic patterns in plant-pollinator mutualistic

644 networks. Ecology, 83, 2416–2424.

645 Olito, C. & Fox, J.W. (2014) Species traits and abundances predict metrics of

646 plant-pollinator network structure, but not pairwise interactions. Oikos, 124,

647 428–436.

31

Jordano - Sampling networks

648 Ollerton, J. & Cranmer, L. (2002) Latitudinal trends in plant-pollinator interac649

tions: are tropical plants more specialised? Oikos, 98, 340–350.

650 Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes,

651 R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops,

652 N., Dulloo, E., Faith, D., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurtt,

653 G., Jetz, W., Karp, D.S., Mcgeoch, M., Obura, D., Onoda, Y., Pettorelli, N.,

654 Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S., Turak, E., Walpole, M.

655 & Wegmann, M. (2013) Essential biodiversity variables. Science, 339, 277–278.

656 Pocock, M.J.O., Evans, D.M. & Memmott, J. (2012) The Robustness and Restora657

tion of a Network of Ecological Networks. Science, 335, 973–977.

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

659 R Development Core Team (2010) R: A language and environment for statis660

tical computing. R Foundation for Statistical Computing. Vienna, Austria.

661 http://www.R-project.org, Vienna, Austria.

662 Rivera-Hutinel, A., Bustamante, R.O., Marín, V.H. & Medel, R. (2012) Effects of

663 sampling completeness on the structure of plant-pollinator networks. Ecology,

664 93, 1593–1603.

665 Schleuning, M., Frund, J., Klein, A.M., Abrahamczyk, S., Alarcón, R., Albrecht,

666 M., Andersson, G.K.S., Bazarian, S., Böhning-Gaese, K., Bommarco, R., Dals667

gaard, B., Dehling, D.M., Gotlieb, A., Hagen, M., Hickler, T., Holzschuh, A.,

668 Kaiser-Bunbury, C.N., Kreft, H., Morris, R.J., Sandel, B., Sutherland, W.J.,

669 Svenning, J.C., Tscharntke, T., Watts, S., Weiner, C.N., Werner, M., Williams,

32

Jordano - Sampling networks

670 N.M., Winqvist, C., Dormann, C.F. & Blüthgen, N. (2012) Specialization of

671 mutualistic interaction networks decreases toward tropical latitudes. Current

672 Biology, 22, 1925–1931.

673 Snow, B. & Snow, D. (1972) Feeding niches of hummingbirds in a Trinidad valley.

674 Journal of Animal Ecology, 41, 471–485.

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

676 Stang, M., Klinkhamer, P., Waser, N.M., Stang, I. & van der Meijden, E. (2009)

677 Size-specific interaction patterns and size matching in a plant-pollinator inter678

action web. Annals Of Botany, 103, 1459–1469.

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

680 Stumpf, M.P.H., Wiuf, C. & May, R.M. (2005) Subnets of scale-free networks are

681 not scale-free: Sampling properties of networks. Proceedings of the National

682 Academy of Sciences USA, 102, 4221–4224.

683 Thébault, E. & Fontaine, C. (2010) Stability of ecological communities and the

684 architecture of mutualistic and trophic networks. Science, 329, 853–856.

685 Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A.,

686 Galetti, M., García, M.B., García, D., Gomez, J.M., Jordano, P., Medel, R.,

687 Navarro, L., Obeso, J.R., Oviedo, R., Ramírez, N., Rey, P.J., Traveset, A.,

688 Verdú, M. & Zamora, R. (2014) Beyond species loss: the extinction of ecological

689 interactions in a changing world. Functional Ecology, 29, 299–307.

690 Vázquez, D.P., Chacoff, N.P. & Cagnolo, L. (2009) Evaluating multiple deter-

33

Jordano - Sampling networks

691 minants of the structure of plant-animal mutualistic networks. Ecology, 90,

692 2039–2046.

693 Vázquez, D.P., Ramos-Jiliberto, R., Urbani, P. & Valdovinos, F.S. (2015) A con694

ceptual framework for studying the strength of plant-animal mutualistic inter695

actions. Ecology Letters, 18, 385–400.

696 Vázquez, D., Morris, W. & Jordano, P. (2005) Interaction frequency as a surrogate

697 for the total effect of animal mutualists on plants. Ecology Letters, 8, 1088–1094.

698 Vizentin-Bugoni, J., Maruyama, P.K. & Sazima, M. (2014) Processes entangling

699 interactions in communities: forbidden links are more important than abundance

700 in a hummingbird-plant network. Proceedings of the Royal Society B-Biological

701 Sciences, 281, 20132397.

34

Jordano - Sampling networks

702 Figure captions

703 Figure 1. Sampling ecological interaction networks (e.g., plant-animal interac704

tions) usually focus on different types of subsampling the full network, yielding

705 submatrices \_[m; n] of the full interaction matrix \_ with A and P animal and

706 plant species. a) all the potential plants interacting with a subset of the animals

707 (e.g., studying just the hummingbird-pollinated flower species in a community);

708 b) all the potential animal species interacting with a subset of the plant species

709 (e.g., studying the frugivore species feeding on figs Ficus in a community); and c)

710 sampling a subset of all the potential animal species interacting with a subset of all

711 the plant species (e.g., studying the plant-frugivore interactions of the rainforest

712 understory).

713

714 Figure 2. Sampling species interactions in natural communities. Suppose an

715 assemblage with A = 3 animal species (red, species 1–3 with three, two, and 1

716 individuals, respectively) and P = 3 plant species (green, species a-c with three

717 individuals each) (colored balls), sampled with increasing effort in steps 1 to 6

718 (panels). In Step 1 we record animal species 1 and plant species 1 and 2 with

719 a total of three interactions (black lines) represented as two distinct interactions:

720 1 􀀀 a and 1 􀀀 b. As we advance our sampling (panels 1 to 6, illustrating e.g.,

721 additional sampling days) we record new distinct interactions. Note that we actu722

ally sample and record interactions among individuals, yet we pool the data across

723 species to get a species by species interaction matrix. Few network analyses have

724 been carried out on individual data(Dupont et al., 2014).

725

35

Jordano - Sampling networks

726 Figures

Figure 1:

Animals

Plants

a b c

m

P

A

n n

m

36

Jordano - Sampling networks

Figure 2:

Jordano - Figure 1

b

c

a

a

a

b

b

c

b

c

a

a

a

b

b

c

b

c

a

a

a

b

b

c

b

c

a

a

a

b

b

c

b

c

a

a

a

b

b

c

b

c

a

a

a

b

b

c

1 1

3

1

2

2

1 1

3

1

2

2

1 1

3

1

2

2

1 1

3

1

2

2

1 1

3

1

2

2

1 1

3

1

2

2

1 2 3

4 5 6

1 2 2 2 2 2 3 3 4

3 3 4 3 3 5 3 3 6

c

c c c

c c

37

Jordano - Sampling networks

727 Table captions

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

729 A, number of animal species; P, number of plant species; I, number of observed

730 links; C = 100I=(AP), connectance; FL, number of forbidden links; and ML,

731 number of missing links. As natural scientists, our ultimate goal is to eliminate

732 ML from the equation FL = AP 􀀀 I 􀀀ML, which probably is not feasible given

733 logistic sampling limitations. When we, during our study, estimate ML to be

734 negligible, we cease observing and estimate I and FL.

735

736 Table 2. Frequencies of different type of forbidden links in natural plant-animal

737 interaction assemblages. AP, maximum potential links, Imax; I, number of ob738

served links; UL, number of unobserved links; FL, number of forbidden links;

739 FLP , phenology; FLS, size restrictions; FLA, accessibility; FLO, other types of

740 restrictions; ML, unknown causes (missing links). Relative frequencies (in paren741

theses) calculated over Imax = AP for I, ML, and FL; for all forbidden links types,

742 calculated over FL. References, from left to right: Olesen et al. 2008; Olesen &

743 Myrthue unpubl.; Snow & Snow 1972 and Jordano et al. 2006; Vizentin-Bugoni

744 et al. 2014; Jordano et al. 2009; Olesen et al. 2011.

745

746 Table 3. A vectorized interaction matrix.

747

748 Table 4. Sampling statistics for three plant-animal interaction networks (Olesen

749 et al. 2011). Symbols as in Table 1; N, number of records; Chao1 and ACE are

750 asymptotic estimators for the number of distinct pairwise interactions I (Hortal

38

Jordano - Sampling networks

751 et al. 2006), and their standard errors; C, sample coverage for rare interactions

752 (Chao & Jost 2012). Scaled asymptotic estimators and their confidence intervals

753 (CI) were calculated by weighting Chao1 and ACE with the observed frequencies

754 of forbidden links.

755

39

Jordano - Sampling networks

756 Tables

Table 1:

Link type Formulation Definition

Potential links Imax = 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 = Imax 􀀀 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.

40

Jordano - Sampling networks

Table 2:

Pollination Seed

dispersal

Link

type

Zackenberg Grundvad Arima

Valley

Sta.

Virginia

Hato

Ratón

Nava

Correhuelas

Imax 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)

FLP 530

(1.0000)

94

(0.2166)

0

(0.0000)

120

(0.1624)

67

(0.3964)

195

(0.3028)

FLS \_ \_ \_ (\_ \_ \_) 8

(0.0184)

30

(0.0890)

140

(0.1894)

31

(0.1834)

46 (0.0714)

FLA \_ \_ \_ (\_ \_ \_) 5

(0.0115)

150

(0.445)a

\_ \_ \_ (\_ \_ \_) 20

(0.1183)

61 (0.0947)

FLO \_ \_ \_ (\_ \_ \_) \_ \_ \_ (\_ \_ \_) 38

(0.1128)b

\_ \_ \_ (\_ \_ \_) \_ \_ \_ (\_ \_ \_) 363

(0.5637)

ML 977

(0.6483)

327

(0.7535)

119

(0.3531)

77

(0.1042)

51

(0.3018)

342

(0.5311)

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

understory species.

b, Colour restrictions, and reward per flower too small relative to the size of the

bird.

41

Jordano - Sampling networks

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

. . . . . . . . . . . . . . . . . .

Ai - Pi 1 0 1 . . . 2

42

Jordano - Sampling networks

Table 4:

Hato Ratón Nava Correhuelas Zackenberg

A 17 33 65

P 16 25 31

Imax 272 825 1891

N 3340 8378 1245

I 151 181 268

C 0.917 0.886 0.707

Chao1 263:1 \_ 70:9 231:4 \_ 14:2 509:6 \_ 54:7

ACE 240:3 \_ 8:9 241:3 \_ 7:9 566:1 \_ 14:8

Scaled Chao 195.4 162.7 308.4

CI [124.5–266.3] [148.5–176.9] [253.6–363.1]

Scaled ACE 178.5 169.7 342.6

CI [169.5–187.4] [161.8–177.6] [327.8–357.4]

% unobserveda 8.33 15.38 47.80

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

43