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, September 5, 2015

Summary

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

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

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

5. Here I provide a review and outline a conceptual framework for interaction sampling by building an explicit analogue to individuals and species sampling, thus extending diversity-monitoring approaches to the characterization of complex networks of ecological interactions. This is crucial to assess the fast-paced and devastating effects of defaunation-driven loss of key ecological interactions and the services they provide and the analogous losses related
to interaction gains due to invasive species and biotic homogenization.

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

Introduction

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

Biodiversity species assessment aims at sampling individuals in collections and determining the number of species represented. Given that, by definition, samples are incomplete, these collections do not enumerate the species actually present. The ecological literature dealing with robust estimators of species richness and diversity in collections of individuals is immense, and a number of useful approaches have been used to obtain such estimates (Magurran, 1988; Gotelli & Colwell, 2001; Colwell, Mao & Chang, 2004; Hortal, Borges & Gaspar, 2006; Colwell, 2009; Gotelli & Colwell, 2011; Chao et al., 2014). Recent effort has been also focused at defining essential biodiversity variables (EBV) (Pereira et al., 2013) that can be sampled and measured repeatedly to complement biodiversity estimates. Yet sampling species or taxa-specific EBVs is just probing a single component of biodiversity; interactions among species are another fundamental component, one that supports
the existence, but in some cases also the extinction, of species. For example, the ex-
tinction of interactions represents a dramatic loss of biodiversity because it entails
the loss of fundamental ecological functions (Valiente-Banuet et al., 2014). This
missed component of biodiversity loss, the extinction of ecological interactions,
very often accompanies, or even precedes, species disappearance. Interactions
among species are a key component of biodiversity and here we aim to show that
most problems associated with sampling interactions in natural communities relate
to problems associated with sampling species diversity, even worse. We consider
pairwise interactions among species at the habitat level, in the context of alpha di-
versity and the estimation of local interaction richness from sampling data (Chao
et al., 2014). In the first part we provide a succinct overview of previous work
addressing sampling issues for ecological interaction networks. In the second part,
after a short overview of asymptotic diversity estimates (Gotelli & Colwell, 2001),
we discuss specific rationales for sampling the biodiversity of ecological interac-
tions. Most of the examples come from the analysis of plant-animal interaction
networks, yet are applicable to other types of species-species interactions.

Interactions can be a much better indicator of the richness and diversity of
ecosystem functions than a simple list of taxa and their abundances and/or related
biodiversity indicator variables (EBVs). Thus, sampling interactions should be a
central issue when identifying and diagnosing ecosystem services (e.g., pollination,
natural seeding by frugivores, etc.). Fortunately, the whole battery of biodiversity-
related tools used by ecologists to sample biodiversity (species, sensu stricto) can
be extended and applied to the sampling of interactions. Analogs are evident
between these approaches (see Table 2 in Colwell, Mao & Chang, 2004). Monito-
ing interactions is a biodiversity sampling and is subject to similar methodological
shortcomings, especially under-sampling (Jordano, 1987; Jordano, Vázquez & Bascompte, 2009; Coddington et al., 2009; Vázquez, Chacoff & Cagnolo, 2009; Dorado et al., 2011; Rivera-Hutinel et al., 2012). For example, when we study mutualistic networks, our goal is to make an inventory of the distinct pairwise interactions that made up the network. We are interested in having a complete list of all the pairwise interactions among species (e.g., all the distinct, species-species interactions, or links, among the pollinators and flowering plants) that do actually exist in a given community. Sampling these interactions thus entails exactly the same problems, limitations, constraints, and potential biases as sampling individual organisms and species diversity. As Mao & Colwell (2005) put it, these are the workings of Preston’s demon, the moving “veil line” (Preston, 1948) between the detected and the undetected interactions as sample size increases.

Early efforts to recognize and solve sampling problems in analyses of interactions stem from research on food webs and to determine how undersampling biases food web metrics (Martinez, 1991; Cohen et al., 1993; Martinez, 1993; Bersier, Banasek-Richter & Cattin, 2002; Brose, Martinez & Williams, 2003; Banasek-Richter, Cattin & Bersier, 2004; Wells & O’Hara, 2012). In addition, the myriad of classic natural history studies documenting animal diets, host-pathogen infection records, plant herbivory records, etc., represent efforts to document interactions occurring in nature. All of them share the problem of sampling incompleteness influencing the patterns and metrics reported. Yet, despite the early recognition that incomplete sampling may seriously bias the analysis of ecological networks (Jordano, 1987), only recent studies have explicitly acknowledged it and attempted to determine its influence (Ollerton & Cranmer, 2002; Nielsen & Bascompte, 2007; Vázquez, Chacoff & Cagnolo, 2009; Gibson et al., 2011; Olesen et al., 2011; Chacoff
et al., 2012; Rivera-Hutinel et al., 2012; Olito & Fox, 2014; Bascompte & Jordano, 2014; Vizentin-Bugoni, Maruyama & Sazima, 2014; Frund, McCann & Williams, 2015). The sampling approaches have been extended to predict patterns of coextinctions in interaction assemblages (e.g., hosts-parasites) (Colwell, Dunn & Harris, 2012). Most empirical studies provide no estimate of sampling effort, implicitly assuming that the reported network patterns and metrics are robust. Yet recent evidences point out that number of partner species detected, number of actual links, and some aggregate statistics describing network patterns, are prone to sampling bias (Nielsen & Bascompte, 2007; Dorado et al., 2011; Olesen et al., 2011; Chacoff et al., 2012; Rivera-Hutinel et al., 2012; Olito & Fox, 2014; Frund, McCann & Williams, 2015). Most of these evidences, however, come either from simulation studies (Frund, McCann & Williams, 2015) or from relatively species-poor assemblages. Most certainly, sampling limitations pervade biodiversity inventories in tropical areas (Coddington et al., 2009) and we might rightly expect that frequent interactions may be over-represented and rare interactions may be missed entirely in studies of mega-diverse assemblages (Bascompte & Jordano, 2014); but, to what extent?

Sampling interactions: methods

When we sample interactions in the field we record the presence of two species that interact in some way. For example, Snow and Snow(1988) recorded an interaction whenever they saw a bird “touching” a fruit on a plant. We observe and record feeding observations, visitation, occupancy, presence in pollen loads or in fecal samples, etc., of individual animals or plants and accumulate pairwise inter
actions, i.e., lists of species partners and the frequencies with which we observe them. Therefore, estimating the sampling completeness of pairwise interactions for a whole network, requires some gauging of how the number (richness) of distinct pairwise interactions accumulates as sampling effort is increased) and/or estimating the uncertainty around the missed links (Wells & O’Hara, 2012).

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

We need to define two basic steps in the sampling of interactions: 1) which type of interactions we sample; and 2) which type of record we get to document the existence of an interaction. In step #1 we need to take into account whether
we are sampling the whole community of interactor species (all the animals, all the plants) or just a subset of them, i.e., a sub matrix \( \Delta_{m,n} \) of \( m < A \) animal species and \( n < P \) plant species of the adjacency matrix \( \Delta_{AP} \) (i.e., the matrix representation of interactions among the partner species). Subsets can be: a) all the potential plants interacting with a subset of the animals (Fig. 1a); b) all the potential animal species interacting with a subset of the plant species (Fig. 1b); c) a subset of all the potential animal species interacting with a subset of all the plant species (Fig. 1c). While some discussion has considered how to establish the limits of what represents a network (Strogatz, 2001) (in analogy to discussion on food-web limits; Cohen, 1978), it must be noted that situations a-c in Fig. 1 do not represent complete interaction networks. As vividly stated by Cohen et al. (1993): “As more comprehensive, more detailed, more explicit webs become available, smaller, highly aggregated, incompletely described webs may progressively be dropped from analyses of web structure (though such webs may remain useful for other purposes, such as pedagogy)”. Subnet sampling is generalized in studies of biological networks (e.g., protein interactions, gene regulation), yet it is important to recognize that most properties of subnetworks (even random subsamples) do not represent properties of whole networks (Stumpf, Wiuf & May, 2005).

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

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

Reasonably complete analyses of interaction networks can be obtained when
combining both phyto-centric and zoo-centric sampling. For example, Bosch et al.
(2009) showed that the addition of pollen load data on top of focal-plant sampling
of pollinators unveiled a significant number of interactions, resulting in important
network structural changes. Connectance increased 1.43-fold, mean plant connec-
tivity went from 18.5 to 26.4, and mean pollinator connectivity from 2.9 to 4.1;
moreover, extreme specialist pollinator species (singletons in the adjacency matrix) decreased 0.6-fold. Olesen et al. (2011) identified pollen loads on sampled insects and added the new links to an observation-based visitation matrix, with an extra 5% of links representing the estimated number of missing links in the pollination network. The overlap between observational and pollen-load recorded links was only 33%, underscoring the value of combining methodological approaches. Zoo-centric sampling has recently been extended with the use of DNA-barcoding, for example with plant-herbivore (Jurado-Rivera et al., 2009), host-parasitoid (Wirta et al., 2014), and plant-frugivore interactions (González-Varo, Arroyo & Jordano, 2014). For mutualistic networks we would expect that zoo-centric sampling could help unveiling interactions of the animals with rare plant species or for relatively common plants species which are difficult to sample by direct observation. Future methodological work may provide significant advances showing how mixing different sampling strategies strengthens the completeness of network data. These mixed strategies may combine, for instance, timed watches at focal plants, spot censuses along walked transects, pollen load or seed contents analyses, monitoring with camera traps, and DNA barcoding records. We might expect increased power of these mixed sampling approaches when combining different methods from both phyto- and zoo-centric perspectives (Bosch et al., 2009; Blüthgen, 2010). Note also that the different methods could be applied in different combinations to the two distinct sets of species. However, there are no tested protocols and/or sampling designs for ecological interaction studies to suggest an optimum combination of approaches. Ideally, pilot studies would provide adequate information for each specific study setting.
**Sampling interactions: rationale**

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

Our sampling above would have resulted in a vector \( n = [n_1 \ldots n_C]' \) where \( n_i \) is the number of records in the \( i^{th} \) class. As stressed by Bunge & Fitzpatrick (1993), however, the \( i^{th} \) class would appear in the sample if and only if \( n_i > 0 \), and we don’t know *a priori* which \( n_i \) are zero. So, \( n \) is not observable. Rather, what we
get is a vector \( c = [c_1...c_n]' \) where \( c_j \) is the number of classes represented \( j \) times in our sampling: \( c_1 \) is the number of singletons (interactions recorded once), \( c_2 \) is the number of twin pairs (interactions with just two records), \( c_3 \) the number of triplets, etc. The problem thus turns to be estimating the number of distinct classes \( C \) from the vector of \( c_j \) values and the frequency of unobserved interactions (see “The real missing links” below).

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

The species assemblage

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

The interactions

We are then faced with assessing the sampling of interactions $I$. Table 1 summarizes the main components and targets for estimation of interaction richness. In contrast with traditional species diversity estimates, sampling networks has the paradox that despite the potentially interacting species being present in the sam-
pled assemblage (i.e., included in the $A_{obs}$ and $P_{obs}$ species lists), some of their pairwise interactions are impossible to be recorded. The reason is forbidden links.

Independently of whether we sample full communities or subset communities we face a problem: some of the interactions that we can visualize in the empty adjacency matrix $\Delta$ will simply not occur. With a total of $A_{obs}P_{obs}$ “potential” interactions (eventually augmented to $A_{est}P_{est}$ in case we have undetected species), a fraction of them are impossible to record, because they are forbidden (Jordano, Bascompte & Olesen, 2003; Olesen et al., 2011).

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

Forbidden links are non-occurrences of pairwise interactions that can be accounted for by biological constraints, such as spatio-temporal uncoupling (Jordano, 1987), size or reward mismatching, foraging constraints (e.g., accessibility) (Moré et al., 2012), and physiological-biochemical constraints (Jordano, 1987). We still have extremely reduced information about the frequency of forbidden links in natural communities (Jordano, Bascompte & Olesen, 2003; Stang et al., 2009; Vázquez, Chacoff & Cagnolo, 2009; Olesen et al., 2011; Ibanez, 2012; Maruyama et al., 2014; Vizentin-Bugoni, Maruyama & Sazima, 2014) (Table 1). Forbidden links are thus
represented as structural zeroes in the interaction matrix, i.e., matrix cells that cannot get a non-zero value.

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

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

Adjacency matrices are frequently sparse, i.e., they are densely populated with zeroes, with a fraction of them being structural (unobservable interactions) (Bascompte & Jordano, 2014). Thus, it would be a serious interpretation error to attribute the sparseness of adjacency matrices for bipartite networks to undersampling. The actual typology of link types in ecological interaction networks is thus more complex than just the two categories of observed and unobserved interactions (Table 1). Unobserved interactions are represented by zeroes and belong to two categories. Missing interactions may actually exist but require additional sampling or a variety of methods to be observed. Forbidden links, on the other hand, arise due to biological constraints limiting interactions and remain unobservable in nature, irrespectively of sampling effort (Table 1). Forbidden links $FL$ may actually account for a relatively large fraction of unobserved interactions $UL$ when sampling taxonomically-restricted subnetworks (e.g., plant-hummingbird pollination networks) (Table 1). Phenological uncoupling is also prevalent in most networks, and may add up to explain ca. 25–40% of the forbidden links, especially in highly seasonal habitats, and up to 20% when estimated relative to the total number of unobserved interactions (Table 2). In any case, we might expect that a fraction of the missing links $ML$ would be eventually explained by further biological reasons, depending on the knowledge of natural details of the particular systems. Our goal as naturalists would be to reduce the fraction of $UL$ which remain as missing links; to this end we might search for additional biological constraints or increase sampling effort. For instance, habitat use patterns by hummingbirds in the Arima Valley network (Table 2; Snow & Snow, 1972) impose a marked pattern of microhabitat
mismatches causing up to 44.5% of the forbidden links. A myriad of biological causes beyond those included as FL in Table 2 may contribute explanations for UL: limits of color perception and or partial preferences, presence of secondary metabolites in fruit pulp and leaves, toxins and combinations of monosaccharides in nectar, etc. For example, aside from FL, some pairwise interactions may simply have an asymptotically-zero probability of interspecific encounter between the partner species, if they are very rare. However, it is surprising that just the limited set of forbidden link types considered in Table 1 explain between 24.6–77.2% of the unobserved links. Notably, the Arima Valley, Santa Virgínia, and Hato Ratón networks have > 60% of the unobserved links explained, which might be related to the fact that they are subnetworks (Arima Valley, Santa Virgínia) or relatively small networks (Hato Ratón). All this means that empirical networks may have sizable fractions of structural zeroes. Ignoring this biological fact may contribute to wrongly inferring undersampling of interactions in real-world assemblages.

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

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

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

A frequently used asymptotic, bias corrected, non-parametric estimator is $S_{Chao1}$ (Hortal, Borges & Gaspar, 2006; Chao, 2005; Colwell, 2013):

$$S_{Chao1} = S_{obs} + \frac{S_1(S_1 - 1)}{2(S_2 + 1)}$$

Another frequently used alternative is the Chao2 estimator, $S_{Chao2}$ (Gotelli &
Colwell, 2001), which has been reported to have a limited bias for small sample sizes (Colwell & Coddington, 1994; Chao, 2005). Instead of using counts it uses incidence frequencies \( Q_k \) among samples (number of species present in just one sample, in two samples, etc.):

\[
S_{\text{Chao2}} = S_{\text{obs}} + \frac{Q_1(Q_1 - 1)}{2(Q_2 + 1)}
\]

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

**Assessing sampling effort when recording interactions**

The basic method we can propose to estimate sampling effort and explicitly show the analogues with rarefaction analysis in biodiversity research is to vectorize the interaction matrix \( AP \) so that we get a vector of all the potential pairwise interactions \( I_{\text{max}} \) (Table 1) that can occur in the observed assemblage with \( A_{\text{obs}} \) animal species and \( P_{\text{obs}} \) plant species. The new “species” we aim to sample are the pairwise interactions (Table 3). So, if we have in our community *Turdus merula* (Tm) and *Rosa canina* (Rc) and *Prunus mahaleb* (Pm), our problem will be to sample 2 new
“species”: $Tm – Rc$ and $Tm – Pm$. In general, if we have $A = 1 ... i$, animal species and $P = 1 ... j$ plant species (assuming a complete list of species in the assemblage), we’ll have a vector of “new” species to sample: $A_1P_1, A_1P_2, ... A_2P_1, A_2P_2, ... A_iP_j$.

We can represent the successive samples where we can potentially get records of these interactions in a matrix with the vectorized interaction matrix and columns representing the successive samples we take (Table 3). This is simply a vectorized version of the interaction matrix. This is analogous to a biodiversity sampling matrix with species as rows and sampling units (e.g., quadrats) as columns (Jordano, Vázquez & Bascompte, 2009). The package EstimateS (Colwell, 2013) includes a complete set of functions for estimating the mean IAC and its unconditional standard deviation from random permutations of the data, or subsampling without replacement (Gotelli & Colwell, 2001) and the asymptotic estimators for the expected number of distinct pairwise interactions included in a given reference sample of interaction records (see also the specaccum function in library vegan of the R Package)(R Development Core Team, 2010; Jordano, Vázquez & Bascompte, 2009; Olesen et al., 2011). In particular, we may take advantage of replicated incidence data, as it takes account of heterogeneity among samples (days, censuses, etc.; R.K Colwell, pers. comm.) (see also Colwell, Mao & Chang, 2004; Colwell, Dunn & Harris, 2012; Chao et al., 2014).

In this way we effectively extend sampling theory developed for species diversity to the sampling of ecological interactions. Yet future theoretical work will be needed to formally assess the similarities and differences in the two approaches and developing biologically meaningful null models of expected interaction richness with added sampling effort.

Diversity-accumulation analysis (Magurran, 1988; Hortal, Borges & Gaspar,
2006) comes up immediately with this type of dataset. This procedure plots the accumulation curve for the expected number of distinct pairwise interactions recorded with increasing sampling effort (Jordano, Vázquez & Bascompte, 2009; Olesen et al., 2011). Asymptotic estimates of interaction richness and its associated standard errors and confidence intervals can thus be obtained (Hortal, Borges & Gaspar, 2006) (see Supplementary Online Material). It should be noted that the asymptotic estimate of interaction richness explicitly ignores the fact that, due to forbidden links, a number of pairwise interactions among the $I_{max}$ number specified in the adjacency matrix $\Delta$ cannot be recorded, irrespective of sampling effort.

We may expect undersampling specially in moderate to large sized networks with multiple modules (i.e., species subsets requiring different sampling strategies) (Jordano, 1987; Olesen et al., 2011; Chacoff et al., 2012); adequate sampling may be feasible when interaction subwebs are studied (Olesen et al., 2011; Vizentin-Bugoni, Maruyama & Sazima, 2014), typically with more homogeneous subsets of species (e.g., bumblebee-pollinated flowers). In any case the sparseness of the $\Delta$ matrix is by no means an indication of undersampling whenever the issue of structural zeroes in the interaction matrices is effectively incorporated in the estimates.

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

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

where [1] corresponds to a weighted network, and [2] to a qualitative network. The detection rates \( \lambda_i \) depend on the relative abundances \( \phi_i \) of the interactions, the probability of a pairwise interaction being detected when it is present, and the
sample size (the number of interactions recorded), which, in turn, is a function of the sampling effort. Unfortunately, no specific sampling model has been developed along these lines for species interactions and their characteristic features. For example, a complication factor might be that interaction abundances, $\phi_i$, in real assemblages are a function of the abundances of interacting species that determine interspecific encounter rates; yet they also depend on biological factors that ultimately determine if the interaction occurs when the partner species are present. For example, $\lambda_i$ should be set to zero for all $FL$. It its simplest form, $\phi_i$ could be estimated from just the product of partner species abundances, an approach recently used as a null model to assess the role of biological constraints in generating forbidden links and explaining interaction patterns (Vizentin-Bugoni, Maruyama & Sazima, 2014). Yet more complex models (e.g., Wells & O’hara 2012) should incorporate not only interspecific encounter probabilities, but also interaction detectabilities, phenotypic matching and incidence of forbidden links. Mixture models are certainly complex and for most situations of evaluating sampling effort better alternatives include the simpler incidence-based rarefaction and extrapolation (Colwell, Dunn & Harris, 2012; Chao et al., 2014).

The real missing links

Given that a fraction of unobserved interactions can be accounted for by forbidden links, what about the remaining missing interactions? We have already discussed that some of these could still be related to unaccounted constraints, and still others would be certainly attributable to insufficient sampling. Would this always be the case? Multispecific assemblages of distinct taxonomic relatedness,
whose interactions can be represented as bipartite networks (e.g., host-parasite, plant-animal mutualisms, plant-herbivore interactions- with two distinct sets of unrelated higher taxa), are shaped by interspecific encounters among individuals of the 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 probabilities of interspecific encounter \((PIE)\) would be simply the product of the two lognormal distributions. Thus, we might expect that for low \(PIE\) values, pairwise interactions would be either extremely difficult to sample, or just simply not 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, thus \(M = 53.1\%\). 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 \(PIE_{median} < 1.4 \times 10^{-4}\) we may safely expect (note the quantile estimate \(Q_{75\%} = 3.27 \times 10^{-4}\)) that a sizable fraction of these missing interactions may not occur according to this neutral expectation (Jordano, 1987; Olesen et al., 2011) (neutral forbidden links, sensu Canard et al., 2012).

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

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

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

Recent work has inferred that most data available for interaction networks are incomplete due to undersampling, resulting in a variety of biased parameters and network patterns (Chacoff et al., 2012). It is important to note, however, that in practice, many surveyed networks to date have been subnets of much larger networks. This is also true for protein interaction, gene regulation, and metabolic networks, where only a subset of the molecular entities in a cell have been sampled (Stumpf, Wiuf & May, 2005). Despite recent attempts to document whole ecosystem meta-networks (Pocock, Evans & Memmott, 2012), it is likely that most ecological interaction networks will illustrate just major ecosystem compartments. Due to their high generalization, high temporal and spatial turnover, and high complexity of association patterns, adequate sampling of ecological interaction networks is challenging and requires extremely large sampling effort. Undersampling of ecological networks may originate from the analysis of assemblage subsets (e.g., taxonomically or functionally defined), and/or from logistically-limited sampling effort. It is extremely hard to robustly sample the set of biotic interactions even for relatively simple, species-poor assemblages; thus, we need to assess how robust is the characterization of the adjacency matrix $\Delta$. Concluding that an ecological network dataset is undersampled just by its sparseness would be unrealistic. The reason stems from a biological fact: a sizeable fraction of the maximum, potential links that can be recorded among two distinct sets of species is simply unobservable, irrespective of sampling effort (Jordano, 1987). In addition, sampling effort needs to be explicitly gauged because of its potential influence on parameter estimates for the network.
Missing links are a characteristic feature of all plant-animal interaction networks, and likely pervade other ecological interactions. Important natural history details explain a fraction of them, resulting in unrealizable interactions (i.e., forbidden interactions) that define structural zeroes in the interaction matrices and contribute to their extreme sparseness. Sampling interactions is a way to monitor biodiversity beyond the simple enumeration of component species and to develop efficient and robust inventories of functional interactions. Yet no sampling theory for interactions is available. Some key components of this sampling are analogous to species sampling and traditional biodiversity inventories; however, there are important differences. Focusing just on the realized interactions or treating missing interactions as the expected unique result of sampling bias would miss important components to understand how mutualisms coevolve within complex webs of interdependence among species.

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

Ecological interactions provide the wireframe supporting the lives of species, and they also embed crucial ecosystem functions which are fundamental for supporting the Earth system. We still have a limited knowledge of the biodiversity
of ecological interactions, and they are being lost (extinct) at a very fast pace, frequently preceding species extinctions (Valiente-Banuet et al., 2014). We urgently need robust techniques to assess the completeness of ecological interactions networks because this knowledge will allow the identification of the minimal components of their ecological complexity that need to be restored to rebuild functional ecosystems after perturbations.

Acknowledgements

I am indebted to Jens M. Olesen, Alfredo Valido, Jordi Bascompte, Thomas Lewinshon, John N. Thompson, Nick Gotelli, Carsten Dormann, and Paulo R. Guimaraes Jr. for useful and thoughtful discussion at different stages of this manuscript. Jeferson Vizentin-Bugoni kindly helped with the Sta Virgina data. Jens M. Olesen kindly made available the Grundvad dataset; together with Robert K. Colwell, Nestor Perez-Mendez, Juan Pe Gonzalez-Varo, and Paco Rodriguez provided most useful comments to a final version of the ms. Robert Colwell shared a number of crucial suggestions that clarified my vision of sampling ecological interactions. The study was supported by a Junta de Andalucía Excellence Grant (RNM–5731), as well as a Severo Ochoa Excellence Award from the Ministerio de Economía y Competitividad (SEV–2012–0262). The Agencia de Medio Ambiente, Junta de Andalucia, provided generous facilities that made possible my long-term field work in different natural parks.
Data accessibility

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

References


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


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


Figure captions

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

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

Figure 1:
Figure 2:
Table captions

Table 1. A taxonomy of link types for ecological interactions (Olesen et al. 2011).
A, number of animal species; P, number of plant species; I, number of observed links; C = 100I/(AP), connectance; FL, number of forbidden links; and ML, number of missing links. As natural scientists, our ultimate goal is to eliminate ML from the equation FL = AP − I − ML, which probably is not feasible given logistic sampling limitations. When we, during our study, estimate ML to be negligible, we cease observing and estimate I and FL.

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

Table 3. A vectorized interaction matrix.

Table 4. Sampling statistics for three plant-animal interaction networks (Olesen et al. 2011). Symbols as in Table 1; N, number of records; Chao1 and ACE are asymptotic estimators for the number of distinct pairwise interactions I (Hortal
et al. 2006), and their standard errors; $C$, sample coverage for rare interactions (Chao & Jost 2012). Scaled asymptotic estimators and their confidence intervals ($CI$) were calculated by weighting $Chao_1$ and $ACE$ with the observed frequencies of forbidden links.

Tables
<table>
<thead>
<tr>
<th>Link type</th>
<th>Formulation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential links</td>
<td>$I_{\text{max}} = A_{\text{obs}} P_{\text{obs}}$</td>
<td>Size of observed network matrix, i.e. maximum number of potentially observable interactions; $A_{\text{obs}}$ and $P_{\text{obs}}$, numbers of interacting animal and plant species, respectively. These might be below the real numbers of animal and plant species, $A_{\text{est}}$ and $P_{\text{est}}$.</td>
</tr>
<tr>
<td>Observed links</td>
<td>$I_{\text{obs}}$</td>
<td>Total number of observed links in the network given a sufficient sampling effort. Number of ones in the adjacency matrix.</td>
</tr>
<tr>
<td>True links</td>
<td>$I_{\text{est}}$</td>
<td>Total number of links in the network given a sufficient sampling effort; expected for the augmented $A_{\text{est}} P_{\text{est}}$ matrix.</td>
</tr>
<tr>
<td>Unobserved links</td>
<td>$UL = I_{\text{max}} - I_{\text{obs}}$</td>
<td>Number of zeroes in the adjacency matrix.</td>
</tr>
<tr>
<td>True unobserved links</td>
<td>$UL* = I_{\text{max}} - I_{\text{obs}}$</td>
<td>Number of zeroes in the augmented adjacency matrix that, eventually, includes unobserved species.</td>
</tr>
<tr>
<td>Forbidden links</td>
<td>$FL$</td>
<td>Number of links, which remain unobserved because of linkage constraints, irrespectively of sufficient sampling effort.</td>
</tr>
<tr>
<td>Observed Missing links</td>
<td>$ML = A_{\text{obs}} P_{\text{obs}} - I_{\text{obs}} - FL$</td>
<td>Number of links, which may exist in nature but need more sampling effort and/or additional sampling methods to be observed.</td>
</tr>
<tr>
<td>True Missing links</td>
<td>$ML* = A_{\text{est}} P_{\text{est}} - I_{\text{est}} - FL$</td>
<td>Number of links, which may exist in nature but need more sampling effort and/or additional sampling methods to be observed. Augments $ML$ for the $A_{\text{est}} P_{\text{est}}$ matrix.</td>
</tr>
</tbody>
</table>
Table 2:

<table>
<thead>
<tr>
<th>Link type</th>
<th>Pollination</th>
<th>Seed dispersal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zackenberg Grundvad</td>
<td>Arima Valley</td>
</tr>
<tr>
<td>$I_{\text{max}}$</td>
<td>1891</td>
<td>646</td>
</tr>
<tr>
<td>$I$</td>
<td>268 (0.1417)</td>
<td>212 (0.3282)</td>
</tr>
<tr>
<td>$UL$</td>
<td>1507 (0.7969)</td>
<td>434 (0.6718)</td>
</tr>
<tr>
<td>$FL$</td>
<td>530 (0.3517)</td>
<td>107 (0.2465)</td>
</tr>
<tr>
<td>$FL_P$</td>
<td>530 (1.0000)</td>
<td>94 (0.2166)</td>
</tr>
<tr>
<td>$FL_S$</td>
<td>···(···)</td>
<td>8 (0.0184)</td>
</tr>
<tr>
<td>$FL_A$</td>
<td>···(···)</td>
<td>5 (0.0115)</td>
</tr>
<tr>
<td>$FL_O$</td>
<td>···(···)</td>
<td>···(···)</td>
</tr>
<tr>
<td>$ML$</td>
<td>977 (0.6483)</td>
<td>327 (0.7535)</td>
</tr>
</tbody>
</table>

*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.
Jordano - Sampling networks

Table 3:

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
<th>...</th>
<th>Sample i</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 - P2</td>
<td>12</td>
<td>2</td>
<td>0</td>
<td>...</td>
<td>6</td>
</tr>
<tr>
<td>A1 - P2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>A5 - P3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>18</td>
</tr>
<tr>
<td>A5 - P4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>A_i - P_i</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4:

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

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