

**Sampling bias is a challenge for quantifying specialization and network structure:  
lessons from a quantitative niche model**

Jochen Fründ<sup>1,2</sup>, Kevin S. McCann<sup>1</sup> and Neal M. Williams<sup>2</sup>

<sup>1</sup>*Integrative Biology, Univ. of Guelph, Guelph, ON, Canada.*

<sup>2</sup>*Entomology and Nematology, Univ. of California, Davis, CA, USA.*

**Corresponding author:** Jochen Fründ, Integrative Biology, Univ. of Guelph, Guelph, ON, Canada. E-mail: [jfruent@uoguelph.ca](mailto:jfruent@uoguelph.ca)

**Decision date:** 07-May-2015

---

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/oik.02256].

**Abstract**

Network approaches have become a popular tool for understanding ecological complexity in a changing world. Many network descriptors relate directly or indirectly to specialization, which is a central concept in ecology and measured in different ways. Unfortunately, quantification of specialization and network structure using field data can suffer from sampling effects. Previous studies evaluating such sampling effects either used field data where the true network structure is unknown, or they simulated sampling based on completely generalized interactions. Here, we used a quantitative niche model to generate bipartite networks representing a wide range of specialization and evaluated potential sampling biases for a large set of specialization and network metrics for different network sizes. We show that with sample sizes realistic for species-rich networks, all metrics are biased towards overestimating specialization (and underestimating generalization and niche overlap). Importantly, this sampling bias depends on the true degree of specialization and is strongest for generalized networks. As a result, null models simulating generalized interactions may misrepresent sampling bias. We show that other methods used for empirical data may be biased in the opposite direction and strongly overestimate sampling completeness. Some network metrics are barely related between small and large sub-samples of the same network and thus may often not be meaningful. Small samples also overestimate interspecific variation of specialization within generalized networks. While new approaches to deal with these challenges have to be developed, we also identify metrics that are relatively unbiased and fairly consistent across sampling intensities and we identify a provisional rule of thumb for the number of observations required for accurate estimates. Our quantitative niche model can help understand variation in network structure capturing both sampling effects and biological meaning. This is needed to connect network science to fundamental ecological theory and to give robust quantitative answers for applied ecological problems.

**Introduction**

Network analysis has become an important tool in community ecology (Ings et al. 2009) that helps to reveal the organization and functioning of complex systems of interacting species (Rohr et al. 2014) and how they may respond to ongoing global environmental change (Tylianakis et al. 2007). Here we focus on bipartite networks, such as those representing the interactions between plants and their pollinators, or parasitoids and their hosts (Vázquez et al. 2009a, Morris et al. 2014). These networks are essentially consumer-resource matrices (Ings et al. 2009), which were considered in traditional studies on specialization of ecological niches (Colwell and Futuyma 1971) using terms such as diet breadth, niche width and niche overlap. As such, our conclusions are applicable for measuring specialization also without a network perspective. Most network properties and corresponding metrics for bipartite networks describe the magnitude or distribution of specialization (or specificity) among species, putting different emphasis on neutral and niche effects (Vázquez et al. 2009a). Neutral effects describe the consequences of how many species are present and in what abundance, without considering interspecific differences (Krishna et al. 2008). Niche effects describe the outcome of species having different biotic niches and functional traits determining which species interact preferentially with each other (Junker et al. 2013, Eklöf et al. 2013).

A ubiquitous challenge for network studies is that the compilation of interaction data is prone to be incomplete even after intensive sampling (Chacoff et al. 2012). Incomplete sampling likely produces sampling bias, as recognized for resource use specialization (Colwell and Futuyma 1971, Ricklefs and Lau 1980, Devictor et al. 2010) and related food web metrics (Banašek-Richter et al. 2004). Network metrics that isolate niche effects from neutral effects have been advocated (Blüthgen 2010), because neutral effects may often be confounded with the number of observations (Vázquez et al. 2007, Blüthgen et al. 2008). Nevertheless, new developments in network research do not always account for sampling effects. For example, theory shows ways in which network structure influences network stability (Memmott et al. 2004, Rohr et al. 2014), but when theoretical results are compared to empirical networks, sampling effects are rarely considered. Recent trends in empirical network studies to assess the variation of interactions in space (Schleuning et al. 2012), time (Petanidou et al. 2008), among individuals (Tur et al. 2014) and in response to human impact (Tylianakis et al. 2007, Aizen et al. 2012) further increase the relevance of incomplete observations. In all of these cases, available sampling effort must be split up, making them

especially vulnerable to bias. Therefore, a careful assessment of how sampling intensity interacts with other network characteristics to affect measurement of specialization is timely.

Previous studies investigating sampling effects have used mainly two approaches. Some studies have varied sampling effort (mostly duration) empirically, finding that number of links and specialization are strongly underestimated with low sampling effort, whereas other network property metrics are relatively insensitive (Nielsen and Bascompte 2007, Petanidou et al. 2008, Hegland et al. 2010, Chacoff et al. 2012, Rivera-Hutinel et al. 2012). However, the true network is unknown in these cases. Increasing sampling duration simultaneously increases the temporal heterogeneity over which interactions are pooled to form a network. Therefore, the number of species in the network increases with sampling, both because more species are observed and because more species enter the network. Sampling completeness of existing networks has been estimated to about 60 % (Chacoff et al. 2012), but because the true network is unknown the accuracy of this estimate is unclear and it does not tell how results may be influenced. An alternative approach has been to use null models to estimate the potential influence of sampling effects (Vázquez and Aizen 2003). Commonly used null models randomize only binary network structure (Bascompte et al. 2003), or they simulate neutral (fully generalized) interactions, as in quantitative null models (Blüthgen et al. 2008, Dormann et al. 2009). Modifications exist that constrain connectance (Vázquez et al. 2007), but only the neutral quantitative null models simulate an ecological process from which network patterns emerge. These null models suggest high sampling bias in most network metrics (Blüthgen et al. 2008) and often successfully predict changes in network metrics resulting from community change (Vázquez et al. 2009a). However, their usefulness for correcting network metrics and effect size (as advocated in Dormann et al. 2009) is unproven, because ecological networks are known to be significantly different from the assumption of full generalization (Blüthgen et al. 2007).

In this study, we use a mechanistic model to generate a set of networks with a wide range of specialization. From these we simulate network sampling and quantify the effect of sampling intensity on specialization and network metrics. In contrast to most previous studies on sampling bias (Blüthgen et al. 2008), we also consider networks of different sizes (number of consumer and resource species). Comparing values of 48 metrics of specialization and network structure between small and large samples, we show that

- 1) sampling bias is smaller for more specialized networks;
- 2) networks with more resource species require more observations to overcome bias; and

- 3) the level and consistency of bias varies among metrics

We explore solutions for the challenges of sampling bias in network metrics: we give an estimate of how many observations may be needed for accurate measurements and identify best available metrics for different situations. Our results show that neutral null models may overestimate sampling bias and richness estimators may overestimate sampling completeness. We discuss how advanced methods may overcome these limitations. We hope that our approach will help bring theory and data more closely together and make network ecology more quantitatively predictive.

## Methods

### Setting up theoretical webs with a quantitative niche model

All simulations were performed in R 3.02 (R Development Core Team 2014). We used a quantitative niche model to generate quantitative bipartite networks with variable degrees of specialization. Network structure emerged from consumer preferences for resource species, which were determined by a simple trait matching approach (Fig. 1). This approach is similar to the “niche model” for binary food web structure (Williams et al. 2010) and similar models for binary bipartite networks (Joppa and Williams 2013), but produces quantitative networks. A trait value between 0 and 1 was chosen from a uniform distribution for both consumer and resource species ( $t_j$  and  $t_i$ , respectively). Equal trait values indicate a perfect match between consumer and resource trait. Consumer  $j$ 's preference (raw preference,  $P_{i,j}$ ) for a resource species  $i$  was determined by the consumer “niche” as a function of the pairwise difference in trait values between consumer and resource as follows:

$$P_{i,j} = \exp(- (t_i - t_j)^2 * s^2)$$

This represents a Gaussian niche shape, in which specialization was controlled by a parameter  $s = 1 /$  standard deviation. As levels for  $s$  we used 1, 2, 3, 4.1, 5.5, 8, 11, 14.5, 19, 30 and 55, realizing an evenly spread gradient from almost equal preference for all resources to specialization effectively on a single resource species.  $P_{i,j}$  was rescaled per consumer in relation to the resource species present in a network, yielding  $P'_{i,j}$  that sum to 1 across resources for each consumer. Thus,  $P'_{i,j}$  can be interpreted as a per-visit probability (for example, which plant species is visited by a bee species in one interaction event). The “true web”  $W_{i,j}$  is a matrix of interaction probabilities ( $P'_{i,j}$ , consumers' relative preferences for resources) multiplied by abundance vectors (additional details for methods in Supplementary Material Appendix

A1). Thus, our quantitative networks are consistent with a foraging process and suitable for sampling simulations. Both interaction frequency and observation probability are proportional to *true webs*,  $W_{i,j}$ . The same value of  $s$  was used for all consumers in a network of a given specialization level, but variation in specialization within webs arose from the random draws of trait values. We generated a new set of resource trait values for each consumer species (i.e., new  $t_i$  were drawn for each  $t_j$ ), reflecting that different consumer species respond to different combinations of resource traits (Junker et al. 2013). Thus, traits  $t_i$  and  $t_j$  reflect the degree of matching in a multidimensional trait space (Eklöf et al. 2013). This approach differs from the classic niche model that uses only one trait dimension, but our model produced similar networks whether using a single trait dimension or high trait dimensionality (see Supplementary Material Fig. S1). Traits as used here are hypothetical, but the method could be adapted for use with empirical trait data.

#### **Set of networks analyzed**

With this quantitative niche model, we generated a set of networks for which we evaluated network metrics under simulated sampling. This set was a factorial combination of eleven levels of specialization (see above) and seven different web sizes (“4 on 4” = 4 consumer species on 4 resource species, “4 on 10”, “10 on 4”, “10 on 10”, “10 on 40”, “40 on 10” and “40 on 40”). For each specialization level and web size, we generated 100 replicate networks (i.e. a total of 7700 webs).

Abundance distributions in ecological communities are typically strongly skewed and as a consequence, the number of observations strongly varies among species within networks in many field studies (Blüthgen et al. 2008). These skewed distributions can have a strong influence on defining and measuring specialization and network structure (Blüthgen et al. 2008). To represent skewed abundance distributions, we used log-normally distributed species frequencies of both consumers and resources, with a mean of  $\log(10)$  and an s.d. of 1.5 (both on the log-scale), which is similar to plant-pollinator networks in the NCEAS *interaction web database* (<http://www.nceas.ucsb.edu/interactionweb/>). Abundance vectors that ideally fit the log-normal distribution were generated deterministically (Supplement Appendix A1). For each network, we simulated sampling from the true network to construct quantitative networks with different sampling intensities (numbers of observations), resulting in resource (rows)  $\times$  consumer (columns) matrices with cell values equaling interaction frequencies (“observed webs”). For a given total number of observations (number of consumers multiplied by mean number of observations), we drew

interactions with a probability relative to expected interaction frequency in the *true webs*  $W_{i,j}$ . We fixed the number of observations per network, so that the number of observations per species varied in relation to species abundance, similar to network sampling with transect walks.

For most analyses, we evaluated sampling effects by comparing two sampling intensities for each *true web*. By these pairwise comparisons, we isolated true sampling effects from the “neutral effects” (influence of abundance on interaction probability) they are often confused with. “Small sample webs” with a mean of 5 observations per consumer species are comparable to sampling effort in many published networks (such as those in the *interaction web database*). “Large sample webs” were defined by a mean of 1000 observations per consumer species, which is more than in any empirical network dataset that we are aware of. We used *large sample webs* rather than *true webs*, because the entries of *true webs* would conflict with metrics defined for binary networks (all interaction probabilities are non-zero in our framework) or with quantitative metrics defined only for integer values. Our sampling mimics that of discrete interaction events, for example, counting flower visitors. We expect similar sampling effects for sampling of pollen loads or stomach contents, but with those methods it is more difficult to define the number of independent observations as they can reveal several links per consumer individual. Note that ‘consumer’ here refers to the focal guild for which specialization is being quantified; this could as easily be plants and their specialization on pollinators. We also explored how metric values change with increasing number of observations using a rarefaction simulation (described below), which supported our choice of the number of observations (Fig. S2). We are confident that *large sample webs* well approximate fully known networks of discrete interaction events.

We also generated a second set of networks, *even webs*, for which all species (both consumers and resources) within a network had equal abundance and the number of observations was fixed per consumer (Appendix A1). We always refer to the main network set with uneven abundances and web-based sampling (termed “*uneven webs*” if necessary for distinction) if not mentioned otherwise, and show results for even webs in the Supplement. The sampling approach used for *even webs* has the advantage of equal weighting for all species and is easier to interpret, but is not the most typical scenario of network sampling. However, datasets collected with the aim of balanced sampling effort for a set of species may look similar to our *even webs* set. Furthermore, many theoretical and conceptual studies do not incorporate the influence of uneven abundances. Therefore, the two simulated network datasets (*even*

*webs* and *uneven webs*) together help integrate conceptual with empirical studies and work on consumer-resource niches with work on networks.

### Network metrics

For each observed network, we calculated a large set of network metrics (or indices, see list in Tab. 1 and Tab. S1) with the R package *bipartite* 2.04 (Dormann et al. 2009). Eighteen of these metrics measure different aspects of specialization, some of them being developed for resource use data without an explicit network focus. Other metrics look at other network properties, which are nevertheless often indirectly related to specialization and its distribution among species. We decided to integrate as many metrics as possible, as variation in metric choice often complicates comparisons among studies and many supposedly improved metrics have not been thoroughly evaluated. In all calculations, we used option *empty = TRUE*, which excludes species without observed interactions from the network (a common practice in network analysis). We differentiate among metrics based on the type of network data they consider: quantitative metrics are designed for quantitative networks and use information about the frequency or strength of each link, whereas metrics designed for qualitative (or binary) networks only use information about the presence or absence of each link and here are referred to as qualitative metrics. The different metrics quantify properties at the (consumer) species-level, for a community of consumers or for the network as a whole. All network metrics (except modularity) were calculated with the functions *specieslevel* and *networklevel* in the *bipartite* package. For species-level metrics, we calculated the average for the consumer community, allowing us to compare them in the same analytical framework as network-level metrics. Weighted means (weight = number of observations per consumer) were used where applicable (see Tab. S1), which has been recommended to alleviate sampling effects (Blüthgen et al. 2007).

Modularity was calculated with the quantitative algorithm *QuanBiMo* (Dormann and Strauss 2014) available in package *bipartite* (setting *steps = 10e5*, see also Fig. S3) and with the binary algorithm “fast greedy” in package *igraph* 0.7 (Csardi and Nepusz 2006). Some network metrics were not defined for all networks (for example,  $H_2'$  for webs with only one resource, *nestedness* for fully connected webs and *compartment diversity* for non-compartmented webs, see Tab. 1 for proportions). Extreme outliers in *Fisher alpha* (two values  $> 10^9$ ) and *weighted nestedness (wine)* (4 values  $< -1$ ) were excluded from



analyses and figures. Note that we refer to qualitative generality (number of links or number of resource species used) as *degree* here, while *generality* always refers to the quantitative version.

Bias was calculated as the difference between metric values of *large sample webs* and *small sample webs*, in relation to the range of the metric across all *large sample webs* of the same web size:  $\% \text{ Bias}_k = 100 \cdot (x_{k,s} - x_{k,l}) / (\max(x_{.,l}) - \min(x_{.,l}))$ , where  $x_k$  = metric value in web  $k$ ,  $s$  = small sample,  $l$  = large sample, and “.” indicates all webs of the same size and type (i.e., even or uneven webs separated). Thus, negative bias means underestimation and positive bias overestimation of the “true” metric value. Smoothing curves per web size were fitted with function *loess* in R package *stats* and shown wherever this was possible.

### Rarefaction analyses

We explored how the accuracy of specialization metrics is improved by increasing sampling intensity (the number of observations per consumer species,  $N_{obs}$ ). For this analysis, we used networks with equal  $N_{obs}$  for all consumers (the *even webs* set, Appendix A1), thus avoiding the dropping of consumer species from the network and more clearly defining  $N_{obs}$  per consumer. Beginning with *large sample webs*, rarefaction sequences were constructed by stepwise reducing  $N_{obs}$  by one, which gave an *observed web* for each  $N_{obs}$  between 1 and 100. In each step, the probability of reducing the frequency of a link by 1 was equal to this link’s interaction frequency in the *observed web* of the step before. This gave a continuous sequence of networks reflecting stepwise decreases in  $N_{obs}$ , which equivalently reflects how metrics change along a sampling process.

From these sequences, we estimated the number of observations  $N_{obs}$  needed for a reliable metric estimate. We calculated the difference  $dif_N$  between the metric value at  $N_{obs}$  and the metric value at  $N_{obs} = 100$ . The threshold for what is an accurate estimate was set to a  $dif_N$  of not more than 5% of the full range of the metric across all *large sample webs* of the same web size. Note that we focused on accuracy in this analysis and do not differentiate between bias and noise. By calculating  $dif_N$  along one rarefaction sequence per web, we found the highest  $N_{obs}$  ( $N_{obs,max}$ ) with a difference above the threshold, which then gave the *minimum reliable sample size* as  $N_{obs,max} + 1$ . Because it was computationally intensive and we were only interested in an overall mean trend, we calculated *minimum reliable sample size* for only 10 of the 100 replicate webs per web size and specialization level in the *even webs* set (a total sample of 440 webs). The choice of one sequence per web and the maximum of 100 observations are unlikely to influence our results (Fig. S2).

### **Estimators of sampling completeness**

We tested the accuracy of a method that others have used to assess sampling completeness in empirical pollination networks (Chacoff et al. 2012). We used a species richness estimator on *small sample webs* to estimate the true number of links and compared this estimate to the number of links in *large sample webs* (calculating % of links observed for both methods). The % of *large sample web* links observed in a *small sample web* is a conservative estimate of truth because even in the *large sample webs* not all rare links are observed. We used the species richness estimator *Chao1* (calculated with *estimateR* in R package *vegan* 2.0 (Oksanen et al. 2013)), because it is commonly used and considered to be relatively unbiased (Oksanen et al. 2013) and because the similar *Chao2* is not applicable here where sampling intensity is defined by number of observations without distinct sets of samples.

### **Interspecific variation of specialization within webs**

To explore the variation of specialization within our networks, we selected six metrics that characterize specialization or generalization on the level of a (consumer) species: degree, generality, proportional generality,  $d'$ , species specificity index (SSI) and paired difference index (PDI). For each metric, we calculated the standard deviation of metric values among species in a network, separately for small sample and large sample webs. All species were weighted equally in this analysis.

### **Results**

Our simulated networks (*large sample webs*) spanned almost the whole potential range of most metrics of specialization (or generalization), except for node specialization (NSI) and paired difference index (PDI).

All metrics overestimated specialization (underestimated generalization) when based on *small sample webs* compared to *large sample webs* (Fig. 2). For most of these metrics the mean bias was approximately between 10 and 30 % of the full metric range for a given web size (Tab. 1). Notable exceptions were *functional complementarity* (mean bias 58 %) and *weighted connectance* (no average bias, positive and negative biases canceled each other out).

### Specialization bias in different networks: influence of true specialization, web size and sampling scheme

Importantly, the specialization bias resulting from limited sampling generally increased with the true degree of specialization in the network as represented by the metric value from the *large sample web* (Fig. 2). For example, niche overlap in the most generalized networks should be close to 1.0 (0.94 in *large sample webs*), but was only 0.24 for the 40 x 40 species *small sample webs*. This represents a bias of -78 %, compared to a mean bias of -32 % for this web size. For truly specialized interactions, bias was much weaker or absent for all specialization and generalization metrics, except for NSI (bias unrelated to the large sample value) and PDI (all values suggesting specialization and small samples biased towards generalization). Among the least biased metrics were  $d'$ , SSI,  $H_2'$  and PSI (except for webs with fewer consumer than resource species). For specialized networks with true values above approximately 0.5, these four metrics tended to be unbiased on average.

Bias was influenced by network size for most metrics. The underestimation of generalization generally increased with the number of resource species. The number of consumer species influenced only some metrics. Several metrics are normalized by the number or diversity of (resource) species in the network (for example, connectance and proportional generality). Although the normalization did not reduce bias overall, these metrics tended to be less biased for webs with fewer consumers (at a given number of resources), because resources were more likely to be unobserved and thus not included in the normalization. In contrast, PSI and  $d'$  were more biased in webs with fewer consumers than resources. Overall, *even webs* (8 to 20 species) with a fixed number of observations per consumer showed similar patterns as the *uneven webs* (8 to 80 species) with web-based sampling (Tab. S2, Figs. S4, S5, S6). There were some relevant differences: many metrics, particularly those derived from graph theory performing very poorly in *uneven webs* (such as NSI and centrality metrics), tended to perform slightly better in *even webs*. Some metrics derived from community ecology performed much worse in *even webs* with very low correlations or very high bias (such as V-ratio or Fisher alpha, respectively). As expected, the difference between qualitative and quantitative metrics was less for *even webs*. For example, degree was more biased than generality in *uneven webs*, but these metrics showed similar patterns in *even webs*. Considerable specialization heterogeneity among species within a web was generated by our model despite  $s$  being the same for all species within a web (Fig. S7). Similar to the pattern for average

specialization bias, within-web standard deviation of species-level specialization was overestimated in small sample webs, but only for generalized networks (approximately  $s < 5$ ).

### **Differences between metrics: metric consistency; other network structures**

Sampling causes noise in addition to bias, but most specialization metrics calculated from *small sample webs* are still correlated to those calculated from the *large sample webs*. These correlations were often reasonably high, which means that although values do not match precisely, small sample values can still reflect real differences in specialization. The strength of these correlations differed greatly among metrics of specialization or generalization (correlation coefficients in Tab. 1, clouds of points in Fig. 2). Some metrics, such as degree and generality, showed a fairly consistent bias, so that correlations were high ( $r > 0.92$ ) despite the strong bias. Several other metrics showed strong correlations ( $r > 0.93$ ) and little bias, for example, SSI and  $H_2'$ . Other metrics, such as proportional generality and especially connectance and NSI, showed weaker correlations ( $r < 0.8$ ).

Network metrics focusing on aspects of network structure other than specialization strongly varied in bias and consistency between *large sample webs* and *small sample webs* (Figs. 3 and S6, Tab. 1). The realized number of species (with observed interactions) was often lower in *small sample webs* than in *large sample webs* (Fig. 3). For highly specialized webs with fewer consumers than resources, some resources were without interactions also in *large sample webs*, leading to high correlations between number of resource species in *small sample webs* and *large sample webs*, but also to a negative bias of as much as 50%. Among metrics that consider the distribution of specialization, specialization asymmetry was strongly correlated among small and large samples. In contrast, nestedness was weakly correlated among small and large samples (all nestedness metrics,  $r$  between 0.03 and 0.53). In particular, webs with zero nestedness in *large sample webs* often had high nestedness scores in *small sample webs*. Many network metrics showed patterns similar to specialization metrics. For example, quantitative modularity (Fig. 3) was more strongly overestimated in less modular webs. Similarly, cluster coefficients, metrics of robustness to secondary extinction, interaction diversity metrics and linkage density are correlated to generalization and thus showed underestimating bias that tended to be stronger with higher true metric value (Tab. 1, Fig. S6). However, Shannon interaction evenness was unbiased, while Alatalo interaction evenness tended to overestimate the true value and showed considerable noise. Some other metrics showed irregular patterns or little variation within web size (Fig. S6).

### **Possible rule of thumb for required number of observations**

Metrics roughly fell into four groups based on the number of observations at which they approached the value of a “fully sampled” web with 5% accuracy (based on simulations with the *even webs* set, Tab. S2): most quantitative metrics saturated on average between 15-24 observations per consumer and most qualitative metrics approached the final value on average between 23-37 observations per consumer. Some metrics approached the final value on average with 10 observations or less (these were mostly invariant to our specialization parameter  $s$ ), while a fourth group did not saturate until well above 40 observations on average (various metrics with strong noise).

### **Estimating sampling completeness**

Sampling completeness was overestimated for most networks by the *Chao1* estimator (Figs. 4, see S8 for *even webs*). Among *even webs*, many *small sample webs* had between 10 and 30 % (rarely over 50 %) of the links of *large sample webs*, but *Chao1* estimated between 30 and 80 % (up to 100 %) sampling completeness. On the other hand, published quantitative null models might often overestimate sampling bias (Figs. S9, S10).

### **Discussion**

We evaluate the consistency of specialization and network metrics across different sampling intensities for differently sized networks representing a large gradient of specialization generated by a quantitative niche model. Our simulated sub-sampling from full networks with known specialization properties shows that all metrics are biased towards overestimating specialization at least under some conditions of limited sampling. This bias increases with true generalization, which complicates correction methods. Apart from these overall patterns, the magnitude of sampling bias strongly varied among metrics. Our results are important for plant-mutualist, plant-herbivore and host-parasitoid networks, but applicable also to many other types of bipartite networks and consumer-resource matrices. For example, columns could be re-interpreted as consumer individuals instead of species (individual specialization, Tur et al. 2014) or rows could be interpreted as habitats instead of resource species (habitat specialization, Devictor et al. 2010). We identify a rule of thumb for how many observations may be needed for reliable metric values, indicating that unfortunately most species in published network datasets are likely undersampled and most specialization estimates are thus substantially biased. Solutions for dealing with limited accuracy of

specialization estimates depend on the severity of sampling artefacts in relation to effect size in a study. For studies dealing with large effect sizes, choosing best metrics may be enough and we identify several metrics that appear relatively consistent across sample sizes. However, studies that need to quantify relatively small changes in specialization or that need to estimate the absolute value, such as for comparison to theoretical models, face a challenge in the complexity of sampling effects and will need to explore other solutions, several of which we discuss below.

### **Quantitative niche model**

Our new niche-based model generates quantitative bipartite webs with variable degree of specialization and as such has substantial advantages over previous methods. The established “niche model” (Williams et al. 2010, Joppa and Williams 2013) only generates binary network structure (presence or absence of links), but quantitative data are often collected by empirical studies and are important for network dynamics, functioning and impacts (Blüthgen 2010). By simulating interactions as foraging of freely moving consumers with preferences, the network structure in our model emerges from fundamental processes consistent with niche theory and animal behavior, rather than being restricted to abundance-based (neutral) null models (Dormann et al. 2009), potentially constrained by binary empirical overlap matrices (Vázquez et al. 2009b). Because probability of observation in typical network sampling is proportional to interaction frequency, our model also readily allows sampling effects to be studied without being confined to incomplete empirical observations or unrealistic null models. This makes it a useful tool for exploring relationships between network processes and patterns and for connecting theory to data.

The model could contribute to developing a more practical niche concept and improve the search for determinants of network structure. Here, we used virtual traits (as in the niche model), but it could also work with real biological traits that appear to be important for the structure of bipartite networks (Junker et al. 2013, Eklöf et al. 2013). Such a trait-based approach would require a better understanding of the shape of niches or preference distributions in relation to continuous resource traits. Here we used a symmetric niche shape as is often assumed (Devictor et al. 2010), but a skewed shape may be more realistic for some traits (Schneider et al. 2012). For the purpose of this study, we simply required plausible networks differing in specialization.

Defining niches by the distribution of preference (interaction probability) in relation to traits means that we focused on realized (Eltonian) niches (Devictor et al. 2010). In the context of networks, the fundamental niche is more concerned with the set of resource species a consumer species is able to interact with (Devictor et al. 2010). A tension exists in network research between the two niche concepts: most data sets represent samples of realized niches, but many analyses and theoretical concepts rather reflect fundamental niches (e.g. robustness, Memmott et al. 2004). Whereas the fundamental niche may be well defined by suitable binary data, the realized niche is more usefully measured by quantitative data. Modeling approaches such as ours may help bridge the gap between theoreticians and empiricists and put network research on a more solid basis.

### **Specialization bias in different networks**

Here, we confirmed the high potential for specialization bias (Winemiller et al. 2001, Blüthgen 2010, Morris et al. 2014) and show that it occurs to some extent with all metrics. The degree of bias depends on the underlying true degree of specialization: for a highly specialized species, a few observations may reveal all the resources it ever interacts with. However, to detect full generalization at least as many observations as the number of potential interaction partners (resources) are needed. To account for sampling processes and detect less preferred resources, at least two to four times as many observations as the number of potential resources for each consumer may be needed to minimize the influence of sampling effects (or even more: Linton et al. 1981). Only a few long-term studies have exceeded this threshold (Winemiller et al. 2001) and generally support our conclusions from simulated data, but most published network datasets fall below this threshold (Vázquez et al. 2009a). The potential for sampling artifacts should more commonly be explored and reported, especially for metrics based on few observations of likely highly generalized interactions such as plant-seed disperser or ant-nectar associations (Blüthgen et al. 2007).

Network size (species richness) strongly influences sampling bias, which means that observed differences between webs of different size may merely be sampling artefacts. The influence of network size on sampling bias, or inversely, on the required sampling intensity for unbiased specialization estimates, is best separately considered for number of resources and consumers. Unsurprisingly, more observations are required for more resource species (potential interaction partners). The effect of number of consumer species is less obvious in our sampling scheme and differs more strongly among metrics. All else being

equal, more species-rich networks can be expected to have stronger specialization bias. However, increasing the number of species and simultaneously the total number of observations may also increase precision for certain metrics and analyses. We analyzed networks in the lower size range of published networks and therefore may have underestimated average specialization bias of published datasets. Specialization and network structure may scale with species richness in different ways depending on the underlying causes for richness gradients (Winemiller et al. 2001, Banašek-Richter et al. 2009). In our model, the diversity but not the outer bounds of resource traits increased with resource richness. For some comparisons across ecological gradients, higher species richness might be driven by higher heterogeneity or wider niche space giving more opportunity for niche partitioning (MacArthur 1972), but in practice, the potential niche space and how it scales with web size is virtually impossible to know. Network size is an important, but not a linear predictor of sampling bias.

The comparison to the *even webs* set representing a different sampling scenario showed that strong sampling bias is not restricted to a scenario in which species (and their interactions) are observed in proportion to their highly uneven abundances (Blüthgen et al. 2008). This makes our results useful for a wide range of network types and sampling approaches. While uneven webs would often result from field sampling of networks, *even webs* represent standardized sampling intensity per consumer, as would be achieved by rarefaction or species-based approaches to measuring specialization. The tendency for less bias in *even webs* may be partly due to smaller networks in this set. Notably, higher sampling bias in *uneven webs* than *even webs* was mostly found for qualitative, but only for few quantitative metrics. Quantitative structure of uneven networks may be simpler than qualitative structure (Banašek-Richter et al. 2009). Nevertheless, whenever the interspecific distribution of specialization is concerned (for nestedness or interspecific comparisons) sampling effects are likely more severe when observations are distributed unevenly. Eventually, the effect size in comparative studies must be judged in relation to potential differences in sampling bias. For example, a 10% latitudinal difference in herbivore overlap (Dyer et al. 2007) might also be caused by differences in sampling bias, which we show can be more than 30 % of the potential range of specialization (or generalization) metrics. Available data may not yet allow for solid conclusions about latitudinal differences in the specialization of interactions (Schleuning et al. 2012, Morris et al. 2014). In light of the multiple factors contributing to it, properly accounting for sampling bias will often be a challenge.



Our quantitative niche model also generated interspecific variation among species in a network, potentially comparable to real networks. In addition to average specialization, heterogeneity in specialization can also be overestimated by low sampling effort. This is most severe for highly generalized networks, which also have the highest sampling bias. For more specialized networks, specialization heterogeneity may be of similar magnitude in small samples as in very large samples, despite more noise in the small samples.

### **Differences between metrics**

Despite a general trend for specialization bias in observed networks, there were large differences among metrics.

Quantitative metrics were often less sensitive than qualitative metrics (Banašek-Richter et al. 2004, Blüthgen 2010). The quantitative metrics SSI,  $d'$  and  $H_2'$  were among the best performing specialization metrics and quantitative generality and niche overlap were less sensitive to sampling effects than qualitative analogues degree and number of shared partners. The strong bias in qualitative connectance could affect studies using empirical connectance as a constraint (Eklöf et al. 2013). Using the additional information contained in interaction frequency seems to be advised (Dormann and Strauss 2014), although it does not fully safeguard against sampling effects.

We followed Blüthgen et al. (2006) and estimated both consumer abundances and resource availability from marginal totals (species interaction frequencies) when these were required for normalized metrics. This estimation may often be the best available option, but it shifts the focus to measuring how a species differs from other species. Increasing the number of consumer species may increase accuracy in these cases. Similarly, with more consumers, more resource species are observed in interactions. The number of species is commonly used for defining potential interactions and standardizing connectance and other metrics. We did not regard unobserved species as part of the network, as this is usually not possible in empirical studies and many metrics are not defined for species without interactions. The exclusion of non-interacting species may limit sampling bias in empirical evaluations (Nielsen and Bascompte 2007). In this case, opposing patterns of sampling effects may be found depending on whether increased sampling detects many new species or whether sampling only increases the number of interactions among existing species. Connectance may even decrease with higher sampling effort when it means including more species (Martinez et al. 1999), but once most species are detected, connectance will increase with

sampling as new interactions among detected species are observed. In some applications the inclusion of non-interacting species is useful (if metrics are defined for this case). For example, it has been advised to include unparasitized hosts when calculating metrics for host-parasitoid webs (Morris et al. 2014), which may change the interpretation and potentially the sensitivity to sampling effects.

Although our analysis focused on the overall magnitude of specialization, we also evaluated other aspects of network structure that may be more closely related to the distribution of specialization. Metrics focusing on other network properties varied enormously in how they respond to the number of observations. Some metrics are closely correlated to the degree of specialization (modularity, interaction diversity, robustness) and show similar bias as specialization metrics. Metrics that were accurately estimated with very few observations either barely varied among our simulated networks (for example, node specialization index) or almost completely reflect properties of the participating communities irrespective of interactions (for example, interaction strength asymmetry: Blüthgen 2010).

Some network metrics performed poorly and had very low similarity between values of small and large samples. These were mostly not developed for species interaction networks specifically, but derived from graph theory, such as betweenness (Martín González et al. 2010), or community ecology, such as all nestedness metrics (Bascompte et al. 2003). Calculating these metrics from networks with limited observations may not give meaningful results. Based on low sampling bias, nestedness has been suggested for monitoring (Hegland et al. 2010, Rivera-Hutinel et al. 2012), but it might not be as useful if it is dominated by noise or does not change in response to human impact. In contrast, commonly used metrics with fairly consistent bias (for example, degree and generality) may meaningfully reflect underlying gradients or ecological differences, as long as the absolute metric value (effect size) is not important and sampling bias is consistent across comparisons. However, such metrics are prone to showing rarity-specialization relationships as an artefact of variable numbers of observations (Dorado et al. 2011).

### **Solutions for the challenge**

We first consider the level of sampling intensity that would help ameliorate bias in metrics and then discuss different strategies to minimize such biases. Our study suggests that a useful rule of thumb for sampling effort may be at least 20 observations per consumer for accurate estimates of most quantitative metrics, although this rule needs to be explored for more different networks. More observations would be

needed for qualitative metrics, for highly generalized interactions and for higher numbers of available resources. This level of sampling completeness will often not be possible to achieve for many rarer species, especially when the study question does not allow observations to be aggregated over long time periods. Common richness estimators may overestimate sampling completeness and thus hide the extent of the problem. Estimators such as *Chao1* only provide a lower bound for asymptotic richness and may not be reliable for very small samples (Colwell et al. 2012). Additionally, in our assessment we assumed that observations are independent (similar to quantitative null models); whether the number of independent observations is well approximated by the number of individuals is an open question – if not, sampling problems would be even more severe. Network patterns with typical sampling effort should often be interpreted with caution and the potential of sampling effects to influence key results should be considered.

The challenges of sampling artefacts in the measurement of specialization and network structure may be solved differently in different situations. In some cases, choosing metrics with minimal sampling bias, rarefying to a common number of observations per species (Rivera-Hutinel et al. 2012) or using weighted means (Blüthgen et al. 2007) may be simple and good solutions. However, these approaches have limitations: robust metrics may not reflect the meaning of interest and they also have some sampling bias at least when interactions are generalized. Rarefaction is only as good as the minimum sample size, and thus not an option if some species are represented by very few observations. Using weighted means (as we did here) lowers the influence of undersampled species, but this means results are dominated by abundant species.

When these simple solutions are not enough, null models can help explore the possible influence of sampling effects (Dormann et al. 2009). Qualitative null models do not simulate a sampling process and may give misleading results (Blüthgen et al. 2008). Existing (neutral) quantitative null models may overestimate sampling bias, because they simulate unrealistically generalized interactions. They are still useful for estimating the upper bounds of sampling effects, but their applicability for correcting effect sizes may be limited. The same is true for correction methods that use sampling-related covariates in linear models (Aizen et al. 2012, Morris et al. 2014): as the amount of sampling bias depends in complex ways on network size and the number and distribution of observations, “corrected” results are unlikely to represent true effect sizes. We recommend these two methods for more exploratory analyses, but caution against their use for quantification or prediction.

Advanced techniques should be developed for overcoming the limitations of current methods. Promising hierarchical and Bayesian methods for estimating detectability and true network structure have been proposed (Sørensen et al. 2011, Wells et al. 2013, Bartomeus 2013), but are not widely used or sufficiently evaluated (but see Wells et al. 2013). The quantitative niche model presented here could prove useful for evaluating performance and sampling bias of newly proposed network methods, as it can generate realistic network structure that doesn't suffer the unavoidable incompleteness of empirical data. The model may also be used for developing unbiased estimators. Observed patterns could be compared to patterns generated by this model instead of a neutral quantitative null model, thereby more accurately representing the structure of the data when the null hypothesis is "no change in specialization" rather than "no specialization". For example, null patterns based on the observed number of observations per species and network could be generated for a range of specialization values, or the most likely average specialization (specialization parameter  $s$ ) for the study system might be estimated from the network, from natural history or from behavior to generate the most appropriate null pattern.

A different approach to alleviate sampling bias is to use additional information. Here, we used only interaction frequency, which is typically available for quantitative networks (for example, the number of flower visitors or the number of parasitized hosts). Independent estimates of abundance are commonly available for resource species (flower cover or number of unparasitized hosts), but rarely also for consumers (Vázquez et al. 2007). These can be used to improve estimates of resource availability or of a species' overall involvement in the network, but they do not solve the problem of few observations. Pollen or fecal samples can be used to complement visitation data and effectively detect more links, especially if these are collected independent of the primary sampling method (Dorado et al. 2011). Some metrics use additional information about species traits or phylogeny to define specialization (Junker et al. 2013, Jorge et al. 2014), which may make specialization definitions more meaningful. Molecular methods (Wirta et al. 2014) can also improve our ability to detect species-specific interactions and overcome sampling issues. Finally, catalogues, databases and expert knowledge may be useful to better capture all potential interactions for applications more interested in specialization of fundamental niches (Pearse and Altermatt 2013). However, in addition to potentially changing the biological interpretation, all complementary information will likely only improve data quality, but not completely avoid sampling effects. Specialization bias and artefacts can arise from incomplete sampling for all methods.

Full consideration of sampling artefacts may alter current beliefs about the structure of real networks. It will be important for developing ecological network science towards being more predictive and meaningful and for advancing new frontiers such as spatiotemporal dynamics and the impact of global change.

#### **Acknowledgments**

#### **Authorship:**

JF and NMW conceived and designed the study, JF performed analyses and JF, NMW and KSM discussed results and wrote the paper. We thank C. F. Dormann for maintaining the bipartite package, which allowed us to cover so many network metrics. JF was supported by a research fellowship by the DFG (German Research Foundation) and by University of California, Davis, College of Agriculture and Natural Resources.

**References**

- Aizen, M. A. et al. 2012. Specialization and rarity predict nonrandom loss of interactions from mutualist networks. - *Science* 335: 1486–1489.
- Banašek-Richter, C. et al. 2004. Sampling effects and the robustness of quantitative and qualitative food-web descriptors. - *J. Theor. Biol.* 226: 23–32.
- Banašek-Richter, C. et al. 2009. Complexity in quantitative food webs. - *Ecology* 90: 1470–1477.
- Bartomeus, I. 2013. Understanding linkage rules in plant-pollinator networks by using hierarchical models that incorporate pollinator detectability and plant traits. - *PLoS ONE* 8: e69200.
- Bascompte, J. et al. 2003. The nested assembly of plant-animal mutualistic networks. - *Proc. Natl. Acad. Sci. U. S. A.* 100: 9383–9387.
- Blüthgen, N. 2010. Why network analysis is often disconnected from community ecology: a critique and an ecologist's guide. - *Basic Appl. Ecol.* 11: 185–195.
- Blüthgen, N. et al. 2006. Measuring specialization in species interaction networks. - *BMC Ecol.* 6: 9.
- Blüthgen, N. et al. 2007. Specialization, constraints, and conflicting interests in mutualistic networks. - *Curr. Biol.* 17: 341–346.
- Blüthgen, N. et al. 2008. What do interaction network metrics tell us about specialization and biological traits? - *Ecology* 89: 3387–3399.
- Chacoff, N. P. et al. 2012. Evaluating sampling completeness in a desert plant-pollinator network: Sampling a plant-pollinator network. - *J. Anim. Ecol.* 81: 190–200.
- Colwell, R. K. and Futuyma, D. J. 1971. On the measurement of niche breadth and overlap. - *Ecology* 52: 567.
- Colwell, R. K. et al. 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. - *J. Plant Ecol.* 5: 3–21.
- Csardi, G. and Nepusz, T. 2006. The igraph software package for complex network research. - *InterJournal Complex Syst.* 1695: -.
- Devictor, V. et al. 2010. Defining and measuring ecological specialization. - *J. Appl. Ecol.* 47: 15–25.
- Dorado, J. et al. 2011. Rareness and specialization in plant-pollinator networks. - *Ecology* 92: 19–25.
- Dormann, C. F. and Strauss, R. 2014. A method for detecting modules in quantitative bipartite networks. - *Methods Ecol. Evol.* 5: 90–98.
- Dormann, C. F. et al. 2009. Indices, graphs and null models: analyzing bipartite ecological networks. - *Open Ecol. J.* 2: 7–24.
- Dyer, L. A. et al. 2007. Host specificity of Lepidoptera in tropical and temperate forests. - *Nature* 448: 696–699.
- Eklöf, A. et al. 2013. The dimensionality of ecological networks. - *Ecol. Lett.* 16: 577–583.
- Hegland, S. J. et al. 2010. How to monitor ecological communities cost-efficiently: The example of plant-pollinator networks. - *Biol. Conserv.* 143: 2092–2101.
- Ings, T. C. et al. 2009. Review: Ecological networks - beyond food webs. - *J. Anim. Ecol.* 78: 253–269.
- Joppa, L. N. and Williams, R. 2013. Modeling the building blocks of biodiversity. - *PLoS ONE* 8: e56277.

- Jorge, L. R. et al. 2014. An integrated framework to improve the concept of resource specialisation. - *Ecol. Lett.* 17: 1341–1350.
- Junker, R. R. et al. 2013. Specialization on traits as basis for the niche-breadth of flower visitors and as structuring mechanism of ecological networks. - *Funct. Ecol.* 27: 329–341.
- Krishna, A. et al. 2008. A neutral-niche theory of nestedness in mutualistic networks. - *Oikos* 117: 1609–1618.
- Linton, L. R. et al. 1981. Resource Utilization Indices: An Assessment. - *J. Anim. Ecol.* 50: 283–292.
- MacArthur, R. H. 1972. *Geographical ecology: patterns in the distribution of species.* - Princeton University Press.
- Martinez, N. D. et al. 1999. Effects of sampling effort on characterization of food-web structure. - *Ecology* 80: 1044–1055.
- Martín González, A. M. et al. 2010. Centrality measures and the importance of generalist species in pollination networks. - *Ecol. Complex.* 7: 36–43.
- Memmott, J. et al. 2004. Tolerance of pollination networks to species extinctions. - *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* 271: 2605–2611.
- Morris, R. J. et al. 2014. Antagonistic interaction networks are structured independently of latitude and host guild. - *Ecol. Lett.* 17: 340–349.
- Nielsen, A. and Bascompte, J. 2007. Ecological networks, nestedness and sampling effort. - *J. Ecol.* 95: 1134–1141.
- Oksanen, J. et al. 2013. *vegan: community ecology package.* R package version 2.0-10.
- Pearse, I. S. and Altermatt, F. 2013. Extinction cascades partially estimate herbivore losses in a complete Lepidoptera-plant food web. - *Ecology* 94: 1785–1794.
- Petanidou, T. et al. 2008. Long-term observation of a pollination network: fluctuation in species and interactions, relative invariance of network structure and implications for estimates of specialization. - *Ecol. Lett.* 11: 564–575.
- R Development Core Team 2014. *R: A language and environment for statistical computing.* - R Foundation for Statistical Computing.
- Ricklefs, R. E. and Lau, M. 1980. Bias and dispersion of overlap indices: results of some Monte Carlo simulations. - *Ecology* 61: 1019–1024.
- Rivera-Hutinel, A. et al. 2012. Effects of sampling completeness on the structure of plant–pollinator networks. - *Ecology* 93: 1593–1603.
- Rohr, R. P. et al. 2014. On the structural stability of mutualistic systems. - *Science* 345: 1253497.
- Schleuning, M. et al. 2012. Specialization of mutualistic interaction networks decreases toward tropical latitudes. - *Curr. Biol.* 22: 1925–1931.
- Schneider, F. D. et al. 2012. Body mass constraints on feeding rates determine the consequences of predator loss. - *Ecol. Lett.* 15: 436–443.
- Sørensen, P. B. et al. 2011. A method for under-sampled ecological network data analysis: plant-pollination as case study. - *J. Pollinat. Ecol.* 6: 129–139.
- Tur, C. et al. 2014. Downscaling pollen-transport networks to the level of individuals. - *J. Anim. Ecol.* 83: 306–317.

- Tylianakis, J. M. et al. 2007. Habitat modification alters the structure of tropical host-parasitoid food webs. - *Nature* 445: 202–205.
- Vázquez, D. P. and Aizen, M. A. 2003. Null model analyses of specialization in plant-pollinator interactions. - *Ecology* 84: 2493–2501.
- Vázquez, D. P. et al. 2007. Species abundance and asymmetric interaction strength in ecological networks. - *Oikos* 116: 1120–1127.
- Vázquez, D. P. et al. 2009a. Uniting pattern and process in plant–animal mutualistic networks: a review. - *Ann. Bot.* 103: 1445–1457.
- Vázquez, D. P. et al. 2009b. Evaluating multiple determinants of the structure of plant-animal mutualistic networks. - *Ecology* 90: 2039–2046.
- Wells, K. et al. 2013. Inferring host specificity and network formation through agent-based models: tick–mammal interactions in Borneo. - *Oecologia* 172: 307–316.
- Williams, R. J. et al. 2010. The probabilistic niche model reveals the niche structure and role of body size in a complex food web. - *PLoS ONE* 5: e12092.
- Winemiller, K. O. et al. 2001. Food web laws or niche theory? Six independent empirical tests. - *Am. Nat.* 158: 193–199.
- Wirta, H. K. et al. 2014. Complementary molecular information changes our perception of food web structure. - *Proc. Natl. Acad. Sci. U. S. A.* 111: 1885–1890.

**Supplementary Material**

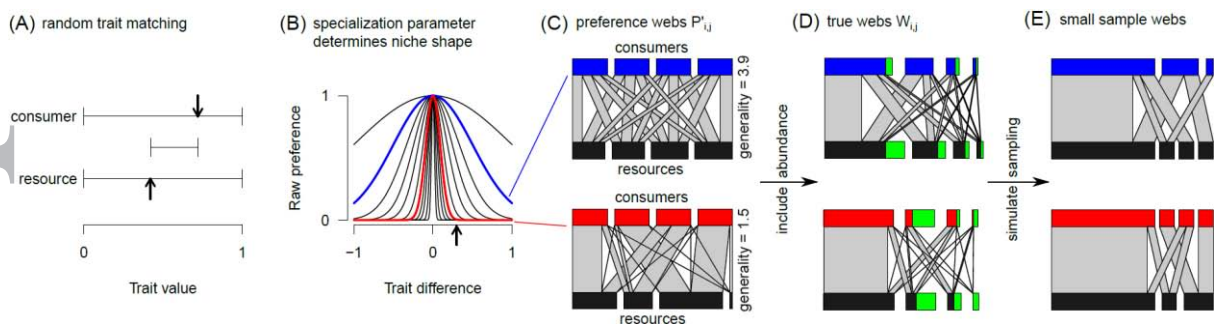
Supplementary material (Appendix oik.XXXXX at <[www.oikosjournal.org/readers/appendix](http://www.oikosjournal.org/readers/appendix)>). Tab.

S1-S2, Figs. S1-S10, Appendix A1-A3.

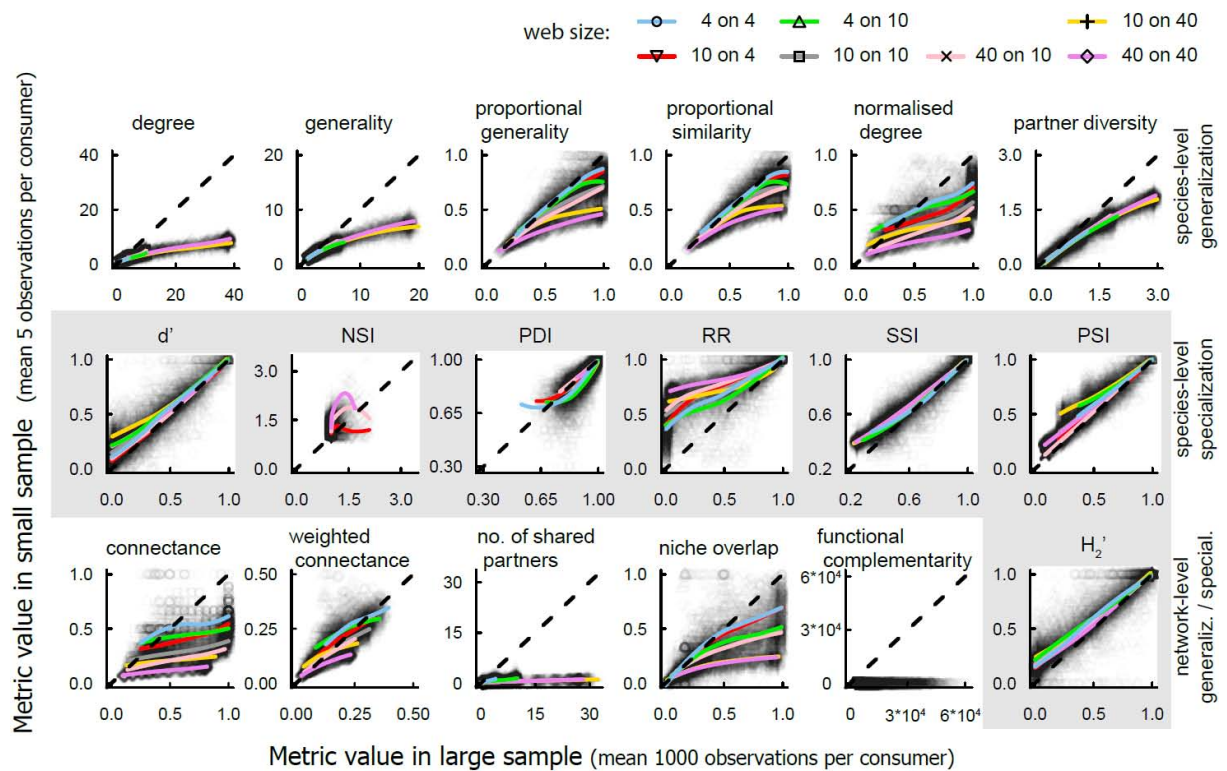


**Figure Legends**

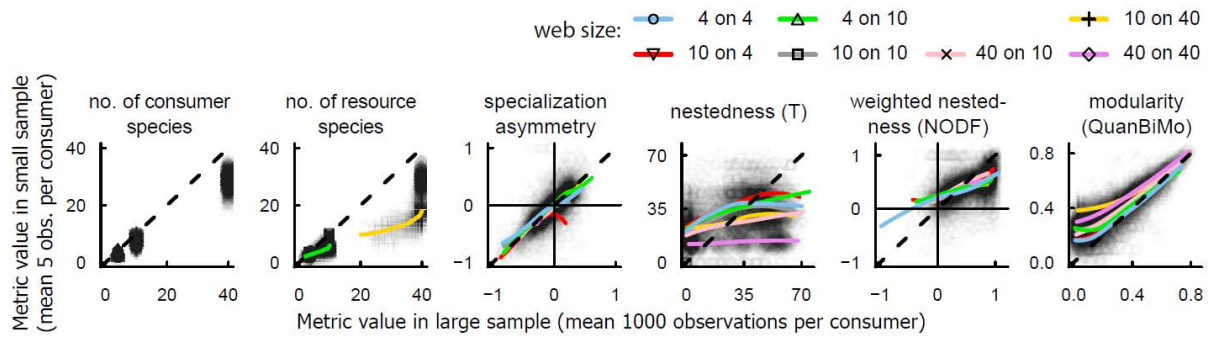
**Figure 1.** Illustration of the “quantitative niche model” used to generate webs with different specialization. Consumer X resource preference matrices were generated in five steps following a trait matching concept. (A) A random trait value was drawn for each bee and plant species. The distance between trait values reflects the degree of matching. Arrows indicate example values. (B) Gaussian ‘niche’ curves were used to determine the ‘raw’ preference ( $P_{i,j}$ ) of a consumer for a resource in relation to the trait distance (arrow corresponds to the example in A). We set up a gradient of 11 levels of specialization by varying the standard deviation of niche curves. A relatively generalized curve is highlighted in blue and a relatively specialized curve in red. Two example webs generated by this approach are shown in C-E, with colors corresponding to curves in B. (C) Standardized preference webs  $P'_{i,j}$  reflect the quantitative network structure solely due to trait matching. (D) True quantitative webs ( $W_{i,j}$ ) are the product of preference and abundance. Width of light grey connecting bars is proportional to expected interaction frequency. Abundance per consumer and resource species (approximately log-normally distributed) is reflected by width of horizontal bars (including light green portions). A large light green portion indicates that a species interacts less frequently than expected from relative abundance, which follows from specialization on rare partners and the influence of both consumer and resource abundance on interaction frequency. (E) To obtain networks comparable to empirical network datasets, discrete sampling events are simulated with observation probability proportional to true webs. Here, the width of species’ horizontal bars reflects observation frequency (species without observations not shown), with a network total of 20.



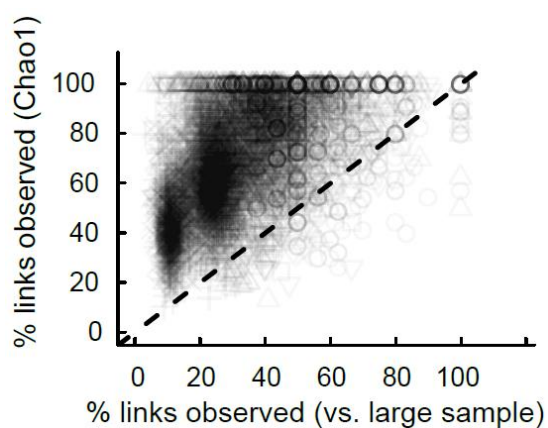
**Figure 2.** Effects of sampling intensity on specialization metrics. Biplots compare the metric value in a small sample of the network to the value in a very intensive sample of the same network for all our simulated networks. For an ideal metric (one independent of sampling intensity) all values would be on the dashed line of identity. Sampling bias overestimating specialization is indicated by values below the line for metrics of generalization (white background) and by values above the line for metrics of specialization (grey background). Web size (number of consumer species “on” number of resource species) is indicated by different symbols and colors of fitted smoothing curves.



**Figure 3.** Sampling sensitivity of network metrics that do not aim to quantify the degree of specialization, but other network properties. See legend of Fig. 2 for further explanations. See Supplementary Material Fig. S6 for all other metrics listed in Tab. 1.



**Figure 4.** Exploring the accuracy of sampling completeness estimated by the widely used Chao1 estimator. Each dot is one of our simulated networks. The X-axis represents a conservative estimate of the true % links observed in a small sample (number of links in *small sample web* divided by number of links in *large sample web*). The Y-axis represents % of links observed as estimated by Chao1 based on *small sample webs*. Most values are above the line of identity, showing that estimators may commonly underestimate the true number of links.



**Table**

**Table 1:** Summary of sampling sensitivity statistics for metrics of specialization and network structure; roughly sorted by metric type (similar metrics together). See Table S1 for descriptions and details for each metric.

metric name	mean bias (%)	r(small, large)	% NA	data type <sup>(c)</sup>	Figure
<i>Generalization / specialization metrics</i>					
degree	-40	<b>0.92</b>	0	bin.	2
generality	-18	<b>0.95</b>	0	quan.	2
proportional generality	-16	0.79	0	quan.	2
proportional similarity	-13	0.81	0	quan.	2
normalised degree	-29	0.67	1	bin.	2
partner diversity	-17	<b>0.97</b>	0	quan.	2
d <sup>*</sup> (species specialization)	<b>+9</b>	<b>0.91</b>	1	quan.	2
NSI (node specialization index)	+15	0.35	39	bin.	2
PDI (paired difference index)	<b>-9</b>	0.79	1	quan.	2
RR (resource range)	+32	0.77	1	bin.	2
SSI (species specificity index)	<b>+9</b>	<b>0.95</b>	1	quan.	2
PSI (pollination service index)	+13	<b>0.93</b>	1	quan.	2
connectance	-38	0.52	0	bin.	2
weighted connectance	<b>0</b>	0.82	0	quan.	2
number of shared partners	-41	0.41	0	bin.	2
niche overlap (Horn)	-20	0.64	0	quan.	2
functional complementarity	-58	0.97	0	quan.	2
H <sub>2</sub> <sup>*</sup> (network specialization)	<b>+10</b>	<b>0.93</b>	1	quan.	2
<i>Other network metrics</i>					
no. of consumer species	-22 <sup>(b)</sup>	<b>0.99</b>	0	bin.	3
no. of resource species	-127	<b>0.91</b>	0	bin.	3
specialization asymmetry	<b>0</b>	0.83	1	quan.	3
interaction strength asymmetry	<b>+8</b>	0.72	0	quan.	S6
species strength	-15	<b>0.90</b>	1	quan.	S6
interaction push pull	<b>-4</b>	0.87	1	quan.	S6
cluster coefficient (group)	-28	0.64	0	bin.	S6
weighted cluster coefficient	-60	0.46	20	quan.	S6
extinction slope	-26	0.43	6	bin.	S6
robustness	-59	0.79	1	bin.	S6
togetherness	<b>-5</b>	-0.04	15	bin.	S6
C score	+27	0.71	0	bin.	S6
V ratio	-10	0.82	11	quan.	S6
discrepancy	-18	0.86	0	bin.	S6
betweenness (centrality)	+45	-0.01	1	bin.	S6
closeness (centrality)	+85	<b>0.90</b>	4	bin.	S6
web asymmetry	-26	<b>0.94</b>	0	bin.	S6
no. of compartments	+35	0.29	0	bin.	S6
compartment diversity	+873	0.12	92	quan.	S6
linkage density	-19	<b>0.93</b>	0	quan.	S6
Shannon diversity	+22	<b>0.96</b>	0	quan.	S6
Fisher alpha	-14	<b>0.98</b>	0	quan.	S6
interaction evenness	<b>-2</b>	0.87	0	quan.	S6
Alatalo interaction evenness	+28	0.60	0	quan.	S6
nestedness (T)	<b>-3</b>	0.33	11	bin.	3
weighted nestedness (NODF)	<b>0</b>	0.20	1	quan.	3
nestedness (NODF)	-11	0.03	1	bin.	S6
weighted nestedness (wine) <sup>(a)</sup>	<b>-9</b>	0.53	26	quan.	S6
cluster coefficient (web)	-40	0.42	0	bin.	S6
modularity (QuanBiMo)	+12	0.85	4	quan.	3
modularity (FG)	+38	0.72	0	bin.	S6

**Table footnotes:**

**mean bias %:** overall average of bias per web, defined as  $100 * (\text{value}_{\text{small sample}} - \text{value}_{\text{large sample}}) / (\text{max. value}_{\text{within size}} - \text{min. value}_{\text{within size}})$ ; best performing ( $\leq 10$ ) in bold

**r(small, large):** overall correlation coefficient (Pearson) of values from large sample webs with values from small sample webs; best performing ( $\geq 0.9$ ) in bold

**% NA:** percent of webs for which metric was not defined (either in small or large sample)

a: extreme outliers removed

b: bias not defined, because there was no variation among full webs within (some) web sizes, which would lead to division by 0; an alternative measure of bias was calculated (small - large sample difference divided by maximum)

c: Abbreviations: bin. = binary / qualitative networks; quan. = quantitative / weighted networks