



Finding missing links in interaction networks

J. CHRISTOPHER D. TERRY ¹ AND OWEN T. LEWIS

Department of Zoology, University of Oxford, Oxford OX1 3PS United Kingdom

Citation: Terry, J. C. D., and O. T. Lewis. 2020. Finding missing links in interaction networks. *Ecology* 101 (7):e03047. 10.1002/ecy.3047

Abstract. Documenting which species interact within ecological communities is challenging and labor intensive. As a result, many interactions remain unrecorded, potentially distorting our understanding of network structure and dynamics. We test the utility of four structural models and a new coverage-deficit model for predicting missing links in both simulated and empirical bipartite networks. We find they can perform well, although the predictive power of structural models varies with the underlying network structure. The accuracy of predictions can be improved by ensembling multiple models. Augmenting observed networks with most-likely missing links improves estimates of qualitative network metrics. Tools to identify likely missing links can be simple to implement, allowing the prioritization of research effort and more robust assessment of network properties.

Key words: *bipartite network; ecological network; missing links; network metrics; prediction; sampling completeness.*

INTRODUCTION

Networks documenting trophic, competitive, or mutualistic interactions among species are an essential tool for understanding the structure and dynamics of ecological communities (Memmott 2009, Poisot et al. 2016b, Dormann et al. 2017, Delmas et al. 2019). Interaction networks can be a better indicator of the resilience of communities than species counts (Valiente-Banuet et al. 2015) and are increasingly applied in conservation (Tylianakis et al. 2010) and biomonitoring (Gray et al. 2014).

Documenting the species present in a system can be difficult enough (Gotelli and Colwell 2010); accurately enumerating their interactions is even more challenging. As a result, ecological interaction networks are routinely under-sampled, with consequences for our understanding of communities (Goldwasser and Roughgarden 1997, Blüthgen 2010, Chacoff et al. 2012, Jordano 2016). Ecological networks are sparse and hence any unobserved interactions could be either “missing links” (interactions that would be observed with complete sampling), or true negatives. Differences or deficiencies in sampling can lead to divergent conclusions about underlying mechanisms (Lee and Guénard 2019). Furthermore, under-sampling will skew the apparent prevalence of interactions toward frequent interactions and may miss infrequent interactions entirely (Dormann et al. 2017).

Common solutions to this problem have included rarefying networks to a consistent level of sampling (Morris

et al. 2014), sampling until metric values stabilize (Riviera-Hutinel et al. 2012, Costa et al. 2016), explicit observation models (Weinstein and Graham 2017), and the use of network metrics that are comparatively robust to under sampling (Vizentin-Bugoni et al. 2016). There has been an extensive literature testing these different approaches (Nielsen and Bascompte 2007, Blüthgen et al. 2008, Poisot et al. 2012b, Fründ et al. 2016, de Aguiar et al. 2019, Henriksen et al. 2019), often arguing for an emphasis on quantitative rather than qualitative metrics, which reduce the influence of missed infrequent interactions.

Nonetheless, infrequent interactions may play an important role in ecological communities. Since networks are dynamic, there is value in distinguishing infrequent interactions from those that are genuinely absent. Rarely observed links have the potential to grow in importance with changes in biotic (Terry et al. 2017) or abiotic conditions (Staniczenko et al. 2017). At equilibrium, interactions may occur at low frequency, yet be important for the structure of the community. For instance, while interactions involving a species suppressed to a low level by a consumer may be observed infrequently, removal of the consumer can allow prey abundance to increase rapidly (Paine 1980). At a network level, the prevalence and distribution of apparently weak or infrequent interactions can have marked consequences for network stability and dynamics (McCann et al. 1998, Berlow 1999, Gellner and McCann 2016, Jacquet et al. 2016, Kadoya et al. 2018).

Ecological networks are manifestly non-random (Dunne and Pascual 2006, Montoya et al. 2006, Vázquez et al. 2009, Bascompte and Jordano 2013). In principle, it should be possible to use this underlying structure to help infer where links are most likely to have been missed

(Bartomeus 2013, Morales-Castilla et al. 2015, Valdovinos 2019). To this end, a number of missing link inference approaches have been tested on ecological networks. These have included hierarchical structuring models (Clauset et al. 2008), linear filtering (Stock et al. 2017), matching-centrality models (Rohr et al. 2016), stochastic block models (Guimera and Sales-Pardo 2009), and k -nearest neighbor recommenders (Desjardins-Proulx et al. 2017). These models have shown that missing links can be inferred from partially observed structured networks. However, although ecological networks are used as test sets, often the scenarios are not orientated toward ecological needs and to date there has been limited uptake of these methods from the wider ecological research community.

Furthermore, it is common in empirical ecological networks for species to have been sampled to markedly different extents (but see Novotny et al. 2010). This has the potential to give further information about the potential location of missing links, since missing links are more likely for poorly sampled species (Blüthgen et al. 2008).

Here, we explore the potential to identify missing links in a diverse collection of ecological bipartite networks with a range of models, including a new species-level coverage-deficit approach that incorporates information on the completeness of sampling for individual species. We show that inferential models can perform well, especially when used in ensemble. We demonstrate how they could inform empirical work to document networks and help evaluate hypotheses.

METHODS

We first present a set of predictive models to detect missing links in interaction networks. To test their capacity to identify missing links and improve the estimation of network-level metrics, we use a suite of simulated and empirical networks that we subsample to create known missing link targets.

Predictive models

Our predictive models seek to identify the true missing links within an observed bipartite adjacency matrix \mathbf{O} composed of observation counts o_{ij} . This has rows representing each focal layer species i and columns detailing each potential interaction partner j . We assume that this set of observations is derived from a set of true interaction frequencies, where the relative rate of interaction between two species is given by a_{ij} . If the species do not interact $a_{ij} = 0$. We use our statistical models to assign to each unobserved interaction in a network a relative probability of being a “missing link” (false negative), rather than a true negative: $P(a_{ij} > 0 | \mathbf{O})$ for the cases where $\mathbf{O}_{ij} = 0$. We assume throughout that the species observed (the dimensions of \mathbf{O}) are a complete list; we do not attempt to predict the identity of interactions with unobserved species. Here we introduce and detail a new “coverage-deficit”

model and four structural models described in the literature. Visualizations of the predictions each of the core models makes based on an example partially observed network are shown in Fig. 1. All models were fit using the cassandra R package (Terry 2019).

Sample size and coverage-deficit models.—Interactions are more likely to have been missed between species that are relatively poorly sampled. When ecological bipartite networks are constructed, it is common that individuals of one level of a network are selected and then the interaction partners in the other level observed (Jordano 2016). Examples include gathering insect hosts and rearing parasitoids from them or observing pollinators that visit a set of focal plant species. This sampling may be opportunistic, matched to the local abundance, or standardized in some other manner. The result is a set of observed interaction counts with each focal layer species. Although interaction counts may be later rescaled to better represent local abundances, here we assume the raw observation data are available of discretizable relationships where individual interaction events can be observed. This includes many interactions such as parasitism, pollination visits and feeding observations, but excludes diffuse facilitative or competitive interactions. A simple approach assigns a relative probability of missed interactions based solely on the number of observations of each focal layer species.

$$P(a_{ij} > 0 | \mathbf{O}) \sim \frac{1}{\sum_j \mathbf{O}_{ij}}.$$

We refer to this as our “sample-size” model. This model captures in a simple way the intuition that interactions are more likely to have been missed where they involve poorly observed species.

This approach can be refined using techniques from the extensive literature developed to assess the sample completeness of species inventories (Chao and Jost 2012). Rooted in information theory originally derived for cryptanalysis by Turing, these approaches can be applied to any case of estimating sample coverage, including interaction sample coverage (Chacoff et al. 2012, Jordano 2016). The core of our coverage-deficit model estimates the interaction partner sample completeness for each focal layer species, based on the observed occurrence frequencies and the Chao1 estimator (Chao 1984, Chao and Jost, 2012). Where the data is sufficient, this coverage-deficit estimator (\hat{C}_{def}) is an estimate of the probability that the next interaction partner to be sampled would not already have been observed interacting with that species. It is defined as follows:

$$\hat{C}_{\text{def}} = \frac{f_1}{n} \frac{f_1(n-1)}{f_1(n-1) + 2f_2}$$

where f_1 is the number of interactions observed only once, f_2 is the number of interactions observed exactly

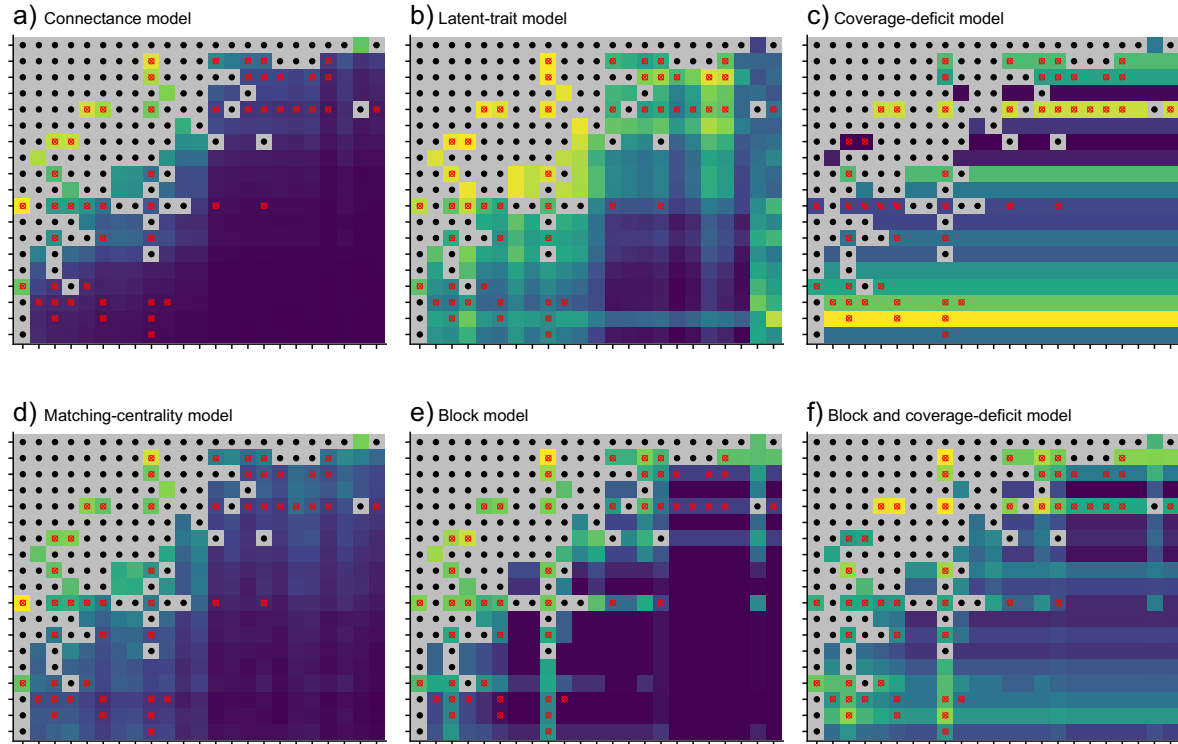


FIG. 1. Demonstration of the fit of predictive models to partially observed networks. The underlying network in each case is a simulated observed bipartite network, generated using the model described in Appendix S1. Brighter colors indicate a higher probability of an unobserved interaction being present according to that model. Black dots indicate observed interactions and red targets indicate unobserved missing links that we wish to identify. Area under receiver operating characteristic curve (AUC) values for this example were as follows: connectance 0.80, latent-trait 0.76, matching-centrality 0.81, block 0.88, coverage-deficit 0.7, averaging the block and the coverage-deficit model 0.92.

twice, and n is the total observed interactions involving that focal species (Chao and Jost, 2012). For each focal layer species i , we use a \hat{C}_{def_i} estimate to assign a relative probability to each potential missing interaction as follows:

$$P(a_{ij} > 0 | \mathbf{O}) \sim \frac{\hat{C}_{\text{def}_i}}{\sum_j (\mathbf{O}_{ij} = 0)}.$$

Unfortunately, ecological interaction networks typically contain a number of focal layer species that are extremely poorly observed, resulting in situations where the Chao1 estimator for coverage deficit is undefined or inappropriate. For very poorly sampled species, the distribution of interactions cannot itself provide much information (Colwell and Coddington 1994). Where observations are exclusively singletons ($f_1 = n$) the index will estimate zero coverage. At the other end of the scale, if a focal species is observed with no singleton interactions the index will estimate that there is zero sample coverage deficit.

In these cases, we use a simple binomial model to estimate \hat{C}_{def} , the probability that the next interactor drawn will not yet have been observed interacting with the focal species. The likelihood of having still not observed any

of the missing links after n samples can be defined as follows: $L = (1 - \hat{C}_{\text{def}})^n$. While the maximum likelihood estimate of \hat{C}_{def} will be 0, with a Bayesian approach and assuming a flat prior to the posterior mean can be found directly as follows: $\mathbb{E}(\hat{C}_{\text{def}} | n) = 1 - (1/[n + 2])$. We use this alternative estimate for \hat{C}_{def} when $n \leq 5$ (cut-off selected via accuracy tests on simulated samples), all the observations are singletons ($f_1 = n$), or if there are no singletons or doubletons ($f_1 + f_2 = 0$).

Latent-trait model.—Species do not interact at random with each other; similar species tend to interact with a similar set of interaction partners. It should be possible to use this information to guide our expectations of missing links through a trait-based framework (Morales-Castilla et al. 2015, Bartomeus et al. 2016, Laigle et al. 2018). These similarities can be in directly measured traits, such as corolla and tongue length (Vizentin-Bugoni et al. 2014, Klumpers et al. 2019, Pichler et al. 2020) or through inferred “latent” traits that are not directly measurable. Latent traits can be identified whether the community is structured by phylogeny, microhabitat, or physiological differences. A survey of published networks found that the number of latent traits required may be quite low (Eklöf et al. 2013).

Intraspecific trait variation caused by spatial, temporal, or ontogenetic differences can pose considerable challenges (González-Varo and Traveset 2016). Nonetheless, such approaches have had successes and a suite of approaches has been developed including matching models (Rohr et al. 2016), Dirichlet-multinomial regression (Crea et al. 2016), matrix factorization (Seo and Hutchinson 2018), and fourth-corner analyses (Spitz et al. 2014). While it is possible to fit any number of latent traits, here we explore use of a simple single latent trait model to reduce the potential of overfitting (Rohr et al. 2016).

In this model, each focal layer species i is assigned a trait value m_i and each partner-layer species j a value m_j . The probability of the interaction existing between each i and j is determined from the difference in trait values through a logistic model

$$\text{logit}(P(a_{ij} > 0|\mathbf{O})) = -\lambda(m_i - m_j)^2 + k_M$$

where λ , k_M , and the vector of trait parameters \mathbf{m} are found through maximum likelihood optimization (details given in Appendix S3). We constrain λ to be positive, such that greater trait differences between potential interactors reduce the probability of an interaction. the parameter k_M is an intercept parameter to capture the average probability of an interaction between two species with perfectly matching traits. We penalize strongly divergent traits by introducing a Cauchy-distributed prior with mean 0 on \mathbf{m} .

Multiple trait distributions may have very similar likelihoods, especially with several specialized interactions or where species bridge network components. To account for this, we optimize 10 differently initialized models and average the probabilities assigned to unobserved interactions by the five models with the greatest likelihood. This model averaging approach is conceptually similar to, but computationally far cheaper than, sampling from posterior distributions of trait vectors.

Connectance model.—It is readily observable that some species have many interaction partners while other species interact with very few. This manifests itself in skewed degree distributions within ecological networks (Jordano et al. 2003). All else being equal, it would be less surprising to discover a missed interaction involving a known generalist species than one with few observed interactions. Many bipartite interaction networks, especially mutualistic networks, are observed to be “nested”: the range of specialists is a subset of the generalists (Bascompte et al. 2003, Ulrich et al. 2009). By assigning each species a connectivity value, simple models can fit such binary nestedness and preferential patterns well. Although nestedness can be generated by a variety of processes (Song et al. 2017) and patterns observable in binary networks may not be reflected in the quantitative networks (Staniczenko et al. 2013), our principle objective here is to identify missing links.

To fit a connectance model, we assign each species a connectance term, c , and determine the probability of an interaction between focal species i and interactor j as follows:

$$\text{logit}(P(a_{ij} > 0|\mathbf{O})) = c_i + c_j + k_C$$

where k_C is a constant intercept term and all parameters are found by maximum likelihood optimization. When both species have high degrees (many interaction partners), their connectance parameters will be high, and the probability of those species interacting will be high.

Matching-centrality model.—The matching-centrality approach (Rohr et al. 2016, 2018) combines a latent-trait and a connectance model and is fit in a similar manner. This dual approach has been shown to have a high capacity to fit tightly to diverse network structures (Rohr et al. 2016). The probability of each interaction existing is found by

$$\text{logit}(P(a_{ij} > 0|\mathbf{O})) = c_i + c_j - \lambda(m_i - m_j)^2 + k_B$$

As in the latent-trait model, λ is defined to be positive, k_B is a constant intercept term and all parameters are found by maximum likelihood optimization. As for the latent-trait model, we again apply a weak Cauchy prior centered at 0 onto the latent trait terms \mathbf{m} and fit 10 models, averaging assigned probabilities from the best five models.

Block model.—Ecological networks, especially antagonistic networks, frequently show comparatively discrete compartments or modularization, where subsets of species interact more strongly within their group than with outsiders (Olesen et al. 2007, 2008, Allesina and Pascual 2009, Schleuning et al. 2014). This may reflect fundamental incompatibilities in physiological traits, temporal mismatch (such as nocturnal/diurnal partitioning), or spatial segregation.

This grouping can be represented by stochastic block models (SBMs), which have been shown to perform well on ecological data in comparison to other clustering algorithms (Leger et al. 2015) and are used increasingly in ecology (Sander et al. 2015, Kéfi et al. 2016). In SBMs, each species is assigned to a group, G_x , in a defined set: $x \in \{1, \dots, g\}$. The relative probability of interaction between two species is determined based on their group membership

$$P(a_{ij} > 0|\mathbf{O}) = \omega_{G_i G_j}.$$

The elements of ω are the between-group interaction probabilities. These are directly specified as the fraction of observed interactions between each of the groups. For example, if 4 of a possible 8 interactions between group A and group B had been observed, $\omega_{G_A G_B} = 0.5$. We find optimal group assignments and fit the model using

a degree-corrected, bipartite-SBM-specific algorithm (Larremore et al. 2014; detailed in Appendix S3). To capture inherent indeterminacy in underlying group structure, we fit 10 differently initialized models and use the average prediction of the five models with the highest likelihood.

Combinations of models.—Individually the above models each capture discrete pieces of information about the identity of missing links. However, the structure of ecological networks is the product of many separate drivers. To capture this diversity, we combine the predictions of multiple models into ensembles. We test combining the matching-centrality model with the block model, and each of the structural models with the coverage-deficit model.

We test two ensembling approaches: multiplication and averaging. Multiplying the relative probabilities assigned to each putative missing link, $P_{\text{ens}}(a_{ij} > 0) \sim P_1(a_{ij} > 0) \times P_2(a_{ij} > 0)$, emphasizes the extreme probabilities of the constituent models. Averaging the relative probabilities, $P_{\text{ens}}(a_{ij} > 0) \sim P_1(a_{ij} > 0) \times P_2(a_{ij} > 0)$, highlights possible interactions that are consistently identified by multiple models. Before combining models, we standardize each set of probabilities assigned to unobserved interactions to sum to 1.

Data sets

Testing the efficacy of extrapolations requires knowledge of the “true” network. We take two complimentary approaches to defining our true networks. First, we generate a set of simulated networks that vary over a wide range of network properties. Second, we use a large and diverse set of empirical networks from literature sources. In both cases, “observed” networks were generated by taking draws from a multinomial distribution parameterized by the true interaction frequency. This replicates sampling processes in ecological studies where the most frequent interactions are highly likely to have been observed and the rarest missed.

We initially generated 2,000 simulated networks from a probabilistic two-trait niche model (modified from Fründ et al. 2016) described in full in Appendix S1. Key network properties were calculated using the bipartite R package (Dormann et al. 2008): connectance, weighted nestedness (Galeano et al. 2009), cluster coefficient (Watts and Strogatz 1998), distribution of interaction frequency (Shannon diversity; Bersier et al. 2002), and network-level specificity $H2'$ (Blüthgen et al. 2006). Distributions of these metrics are shown in Appendix S4: Fig. S1 and the (low) correlations between them are depicted in Appendix S1: Fig. S1.

Our objective was to generate interaction matrices that represent a wide range of plausible bipartite networks. However, models used to generate plausible interaction networks have many similarities to the predictive models described above. This is for a good reason: models are

chosen because they are thought to represent ecological networks parsimoniously. To reduce circularity, our generative model is considerably more complex than our predictive models and we examine how predictive model performance changes in response to network properties.

Using the true networks, we took 300–2,000 samples per network to generate our “observed” networks. We discarded occasional cases where all true interactions were observed. This resulted in a low mean true sample coverage deficit of 2.4%, corresponding to an average network completeness of 64% due to skewed interaction strength distributions. Our observed networks therefore represent reasonably well-sampled systems that are nonetheless missing a large proportion of the interactions present.

We collated a diverse set of 113 empirical networks representing antagonistic, mutualistic, and commensalistic interaction types (Appendix S2). We collated quantitative single-class bipartite networks from the *Web of Life* repository with 30–200 species and 40–431 observed interactions (data available online).² This included 48 plant–pollinator (19 sources), 23 mammal–flea (one source, but across a wide geographic range), 2 plant–ant mutualisms (two sources), and 10 seed-disperser (7 sources) networks. We supplemented this with 25 mammal–dung-beetle interaction networks (Frank et al. 2018, selecting those with ≥ 30 total species, ≥ 40 observed interactions, and ≥ 5 mammal species) and five host–parasitoid networks (Tylianakis et al. 2007).

These empirical networks are not exhaustively sampled; the Chao1 estimator indicates that the observed interactions represent on average only 76% of the underlying network (Appendix S2, Fig. S1). Nonetheless, we took the empirical interaction frequency to be the true network to define a multinomial distribution. We subsampled each empirical network 20 times, each time taking a different fraction between 0.2 and 1.0 of the original sample size, taken to be the sum of the observations. We excluded cases where the observed web included more than 95% of the true interactions.

Model testing

Identifying missing links.—Using each model or combination of models we assign each unobserved interaction in an observed network a relative probability of being a missing link. We assume that all observations are true positives (i.e., there have been no misidentifications) although in principle similar methods could be used to identify candidates for reevaluation. Where a species in the true network was not included in the observed network, we did not attempt to predict its interactions and reduced the size of the true network correspondingly for all purposes.

We assess model performance at identifying missing links in a particular network with the area under receiver

² www.web-of-life.es

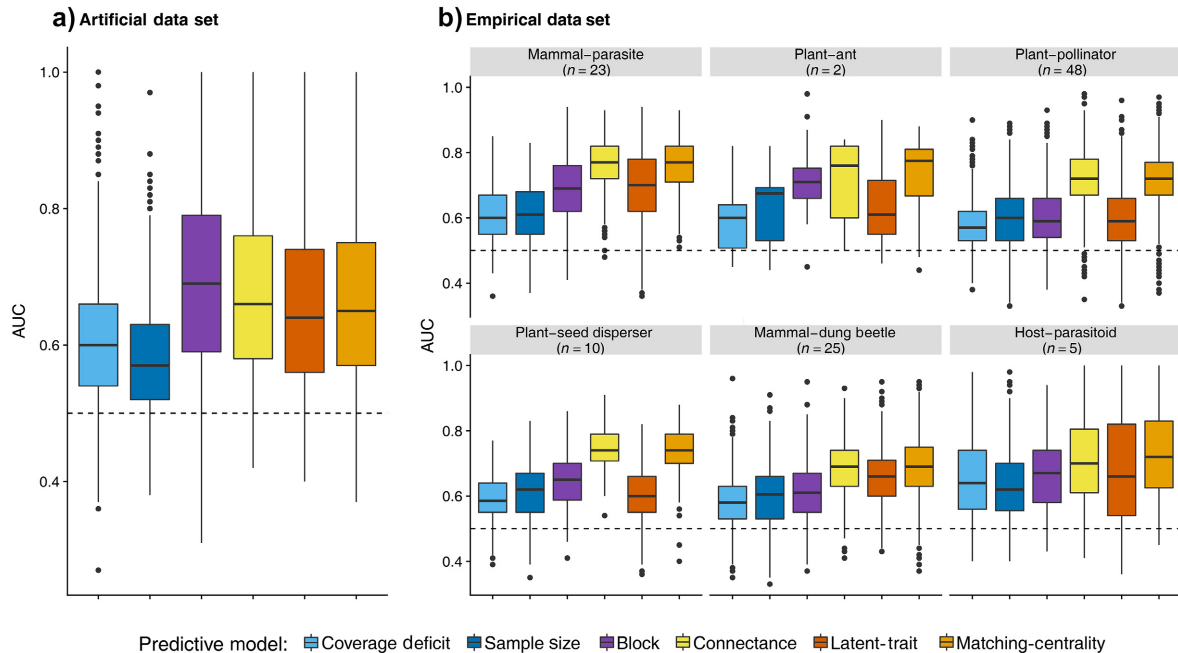


FIG. 2. (a) Model performance across 1985 simulated networks. Mean differences between the statistical models were small, although given the large sample size all pairwise differences were statistically significant (Holm-corrected t tests, paired by network ID, $P < 0.0001$). (b) Model performance on the empirical data sets, subdivided by network class. The original number of networks in each subset of the empirical data set is given as n . Each empirical network was subsampled 20 times to different fractions of the original sample size. Models perform differently on different classes of data (mModel comparison, accounting for data set ID as a random effect, shows model \times class interaction term is significant: $P < 0.0001$, $n_{\text{obs}} = 12,672$). In all cases, boxes show first and third quartiles and the dotted line at AUC = 0.5 shows the expected accuracy from random guessing.

operating characteristic curve (AUC) metric to assess the information content of a signal, using the pROC R package (Robin et al. 2011). AUC can be considered the chance that a false negative (missing link) was assigned a higher relative probability than a true negative. A value of 0.5 indicates no useful signal while a value of 1 indicates perfect discrimination. To determine if $P(a_{ij} > 0)$ is reflective of true relative interaction frequency, we calculated the mean Spearman rank correlation between these values for the unobserved interactions.

Improving estimates of network metrics.—We next tested whether inferring missing links can improve estimates of network-level metrics. We did this by assessing whether network-level properties of an “augmented” network (including inferred missing links) were closer to the true network than the originally observed network. For this analysis, we used the simulated data sets and the ensemble model identified as best able to identify missing links (averaging the block, matching-centrality, and coverage-deficit models). To determine the number of putative missing links with which to augment the observed network, we estimated the true number of interactions with the bias-corrected Chao1 index (using estimateR in the vegan R package; Oksanen et al. 2018). We augmented the observed interaction networks by adding this number of the most probable missing links to the network.

Each selected missing link was taken to be observed once.

Network-level metrics were calculated for the observed, augmented and true networks. We tested three qualitative network metrics (connectance, nestedness, and niche overlap) identified as being susceptible to misestimation in under-sampled networks (Fründ et al. 2016), alongside two metrics considered to be comparatively resistant to under-sampling: specialization ($H2'$) and nestedness (weighted NODF).

RESULTS

Identifying missing links: Simulated networks

In general, the network models showed impressive performance at identifying likely missing links (Fig. 2a). The best-performing individual model overall was the block model, which had a mean AUC of 0.69, well above the accuracy expected from random guessing. This was closely followed by the other three network models (mean AUC values: latent-trait, 0.66; connectance, 0.68; and matching-centrality, 0.67). The coverage-deficit model lagged behind the structural models, but was more informative than the sample-size model. Combining the predictions of multiple models tended to improve performance relative to individual models (Table 1). The

TABLE 1. Mean area under receiver operating characteristic curve (AUC) values using different predictive models or combinations of models across the whole simulated ($n = 1,985$) and empirical (113 networks, 2112 subsamples) network sets.

Model	Simulated data set			Empirical data set		
	Structural model only	Ensembled with coverage-deficit model		Structural model only	Ensembled with coverage-deficit model	
		(\times)	($+$)		(\times)	($+$)
Focal layer sampling						
Coverage-deficit	0.600			0.589 ^{abc}		
Observation count	0.580			0.603 ^{de}		
Network models						
Latent-trait	0.660 ^a	0.649	0.654 ^a	0.631 ^f	0.588 ^{abc}	0.590 ^{abc}
Connectance	0.676	0.708 ^{efg}	0.683 ^b	0.721 ⁱ	0.651 ^h	0.663
Matching-centrality	0.671	0.702 ^{de}	0.684 ^b	0.718 ⁱ	0.642 ^g	0.658
Block	0.691 ^{bcd}	0.689 ^{bc}	0.695 ^{cd}	0.630 ^f	0.584 ^a	0.590 ^c
Combined models						
Block and matching-centrality (multiplied)	0.702 ^{de}	0.705 ^{ef}	0.711 ^f	0.647 ^{gh}	0.594 ^c	0.604 ^{de}
Block and matching-centrality (averaged)	0.705 ^{ef}	0.709 ^f	0.713 ^g	0.648 ^{gh}	0.600 ^d	0.603 ^d

Notes: Within each data set, results are shown with just the network structure-based model (left column), ensembled with the coverage-deficit model, by multiplying the probabilities (\times , center column of each pair) or by averaging the assigned probabilities ($+$, right column of each pair). Incorporating coverage deficit tends to lead to modest improvements in the simulated networks but not in the empirical networks. Sets of results found not to differ significantly (Holm-corrected t tests, paired by network ID) are indicated by the same lowercase letter superscripts, applied separately for simulated and empirical results.

overall best-performing model averaged the predictions of the block model, the matching-centrality model, and the coverage-deficit model. However, the additional improvements from combining models were often small, around 0.01 AUC units.

Model performance and the relative hierarchy of models were affected by the underlying true network structure (Fig. 3). On average, missing links were identified best in nested, unclustered networks where interaction frequencies were more evenly distributed. The coverage-deficit and sample-size models were the least responsive to network structure, since our simulated network generator did not result in a pronounced abundance–generality correlation. Average model performance did not change appreciably with increased sample size, overall interaction coverage deficit, or samples per host (Appendix S4: Fig. S2b). However, when a greater fraction of interactions had been observed the structural models increased their performance, while performance of the coverage-deficit model remained unchanged (Appendix S4: Fig. S2b).

Identifying missing links- Empirical networks

The performance of the different predictive models varied considerably between the different categories of empirical network (Fig. 2b). As expected, the connectance model performed best for mutualistic networks, and the latent-trait model for antagonistic networks. In all six network categories, the matching-centrality model had the best overall performance. The ‘true network’ properties (Appendix S4: Fig. S1) or the degree of

sampling rarefaction (Appendix S4: Fig. S2a) had relatively little effect on predictive performance, except for networks that show “anti-nestedness.” In these cases, the predictive capacity of the connectance model was low, although the matching-centrality model maintained its advantage. The coverage-deficit model was very poor at predicting missing links in the subsampled empirical data sets and combining it with the network structure models led to a *reduction* in overall predictive power (Table 1).

Identifying interaction frequency of missing links

The probabilities assigned to missing links $P(a_{ij} > 0 | \mathbf{O})$ could be related to their true interaction frequency. The correlations were stronger in the simulated data set than the empirical data set (Appendix S4: Table S1). Of the structural models, the connectance model showed the strongest positive correlation ($\rho_{\text{Sim}} = 0.51$, $\rho_{\text{Emp}} = 0.15$), while the block model showed the weakest ($\rho_{\text{Sim}} = 0.28$, $\rho_{\text{Emp}} = 0.06$). The coverage-deficit model was negatively correlated with true frequency ($\rho_{\text{Sim}} = -0.35$, $\rho_{\text{Emp}} = -0.05$), i.e., missing links successfully identified via this model tended to be infrequent.

Improving estimates of network metrics

Integrating inferred interactions into observed networks could improve estimates of key metrics. Qualitative metrics of the augmented networks showed increased correlation, reduced bias, and reduced error with the true network values from the original observed

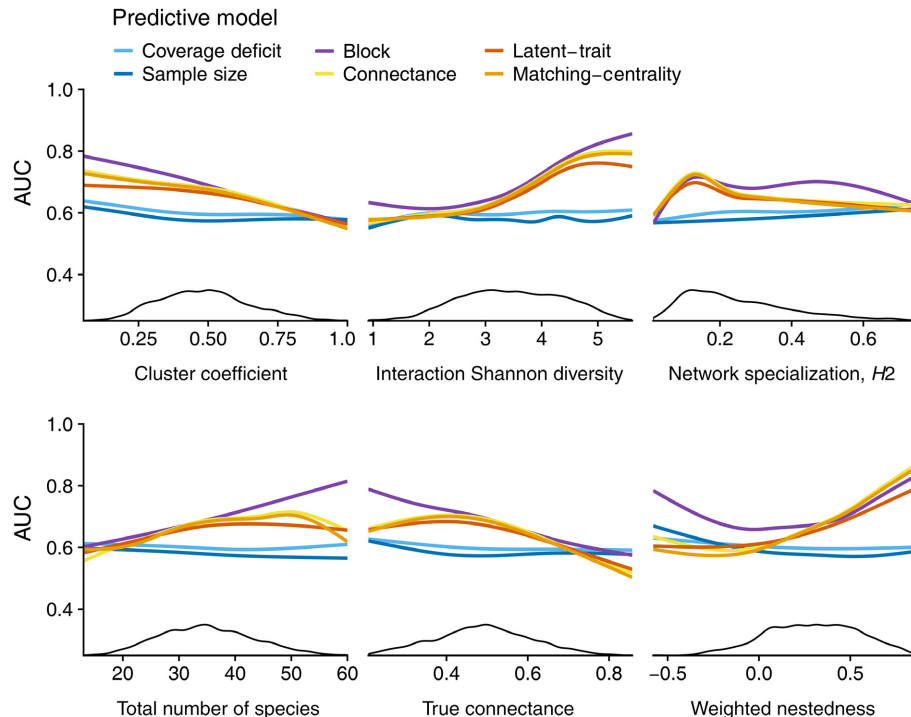


FIG. 3. Response of model performance to “true” network structure. Smoothed line shows a GAM fit for each predictive model. Lower black line shows the local density distribution of that property across the simulated test data set.

networks (Fig. 4a). By contrast, the quantitative network metrics ($H2'$ and weighted NODEF, Fig. 4b) were already well-estimated by the under-sampled network, and the inclusion of additional interactions made little difference. The Chao1 estimator for the number of missing interactions appeared to perform well for the empirical networks but consistently underestimated the true interaction count and coverage deficit in the simulated data (Appendix S4; Fig.S3).

DISCUSSION

We have shown that it is possible to infer missing links in a range of empirical networks with a reasonable degree of accuracy and this approach can improve estimates of network metrics. The challenge now is to put these tools into practice to leverage additional insight into ecological processes.

Choosing between predictive models

Ideally, a model used to infer missed interactions would capture the true ecological mechanisms that structure the network, but this is not essential for all purposes. Given the diversity of ecological networks, there will never be a single “best” model and victors in comparisons will depend on the data set used. In our simulated data, our block model appears to perform best, while in our empirical data sets, the matching-centrality

model performed consistently well (as found by Rohr et al. 2016). However, we have reason to distrust the empirical data sets (discussed below). The pronounced, likely artefactual, abundance–generality relationship will favor connectance models.

Nonetheless, splitting hairs over the best-performing model is not necessarily a productive route. Networks are structured by multiple processes and in our simulated data sets the very best performance comes from combining different models, which identify different missing interactions. Structure-based models pick out the more frequent interactions while coverage-deficit models highlight comparatively infrequent interactions that would be the hardest to determine through further undirected sampling. Future progress will come from operationalizing estimated missing links, rather than from further incremental model refinements.

Distinctions between empirical and simulated networks

The empirical and simulated data sets overlapped substantially in key network metrics. A key difference between these data sets is the stronger correlation in empirical networks between species’ marginal totals and their generality (mean Spearman $\rho_{\text{Sim}} = 0.43$, $\rho_{\text{Emp}} = 0.76$). This can account for the poor performance of the coverage-deficit model in these cases. It is challenging to disentangle the extent to which apparent specialism of rarely observed species is a sampling

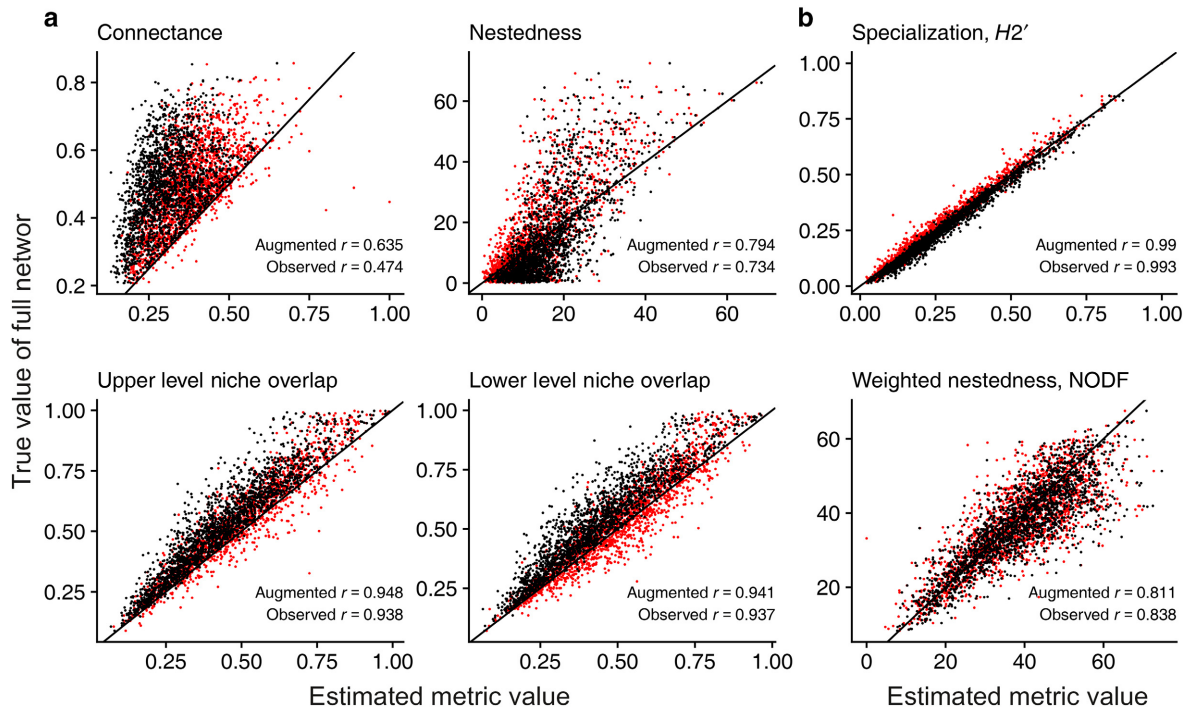


FIG. 4. Relationship between network metrics calculated on the observed (black) and augmented (red) interaction networks to the true network value in the simulated networks. The line shows the 1:1 relationship. Pearson's correlation coefficient (r) is shown for each metric. (a) Qualitative metrics show notable improvements in correlation with the true values when augmented, while (b) quantitative metrics are largely unchanged.

artifact (Blüthgen et al. 2006), as there are ecological explanations for such a relationship (Fort et al. 2016, Dormann et al. 2017, Simmons et al. 2019). For our purposes here, this likely bias in the structure of empirical networks has the consequence that predictive models may identify missing links introduced both by our sub-sampling procedure, and due to gaps in the original data set. We therefore place more weight on the results from the simulated data, while noting that, despite the obstacles, the structural models are still able to perform reasonably well on the sparse empirical data.

Estimating the quality of sampling and the number of missing interactions

Sample coverage estimators offer promise for estimating the number of missed interactions (Chacoff et al. 2012, Traveset et al. 2015, Jordano 2016, MacGregor et al. 2017). The distinction between sample coverage and network completeness is worth emphasizing. Apparently high values of sample coverage alone should not be used as a simple proxy for network completeness, although they are related to the reliability of certain quantitative network metrics (Henriksen et al. 2019). We found that the whole-network-level Chao1 estimate of missed interactions was strongly correlated with the true number of missing links in our simulated networks, but notably underestimated this number, a pattern also

observed by Fründ et al. (2016). Although there appeared to be no systematic bias in estimated network completeness in the empirical data sets, these true networks are themselves highly incomplete, suggesting an overall underestimation also in these cases. Our focus in this paper is on the location rather than the number of missing links. Future work should investigate how long tails of infrequent interactions relate to this underestimation, and consider whether species rank abundance distribution estimators (Chao et al. 2014) can be usefully generalized to assign a frequency distribution to estimated missing links.

The way forward

Most of the models used here, and indeed our fitting approaches, likely have further room for refinement, and there are many further classes of predictive models that we did not test, such as nearest neighbor algorithms (Desjardins-Proulx et al. 2017). Furthermore, we based our models entirely on the information contained in the network itself. In most cases, ecologists constructing a network have access to (sometimes considerable) additional information and expertise on the study system. Introducing preexisting knowledge about likely or forbidden interactions, or pooled knowledge from multiple sample sites offers considerable opportunities to maximize the information in a network (Gray et al. 2015,

Pomeranz et al. 2018). External information would be particularly useful in separating “forbidden links” (Olesen et al. 2011) that never occur due to fundamental incompatibilities from those that occur elsewhere but not within those spatiotemporal boundaries. For instance, species may not be observed to interact because of a fundamental mismatch in body sizes (a forbidden link) or because they interact at night but data were gathered only during the day.

Because the naive models used here do not use external information, or the judgment of the ecologists constructing the network, the quality of the estimation very much represents a lower bound of what is possible with the approach in practice. As a result of the underlying idiosyncrasies of ecological systems, often euphemistically labelled “species identity effects,” there will always be an upper limit to the predictive capacity of any statistical model. We suggest that, rather than developing new models to gain marginal improvements in predictive capacity, a more productive focus would be to develop frameworks exploiting this information to test the robustness of conclusions derived from ecological networks.

Complete network inventories are not a realistic or necessarily useful goal (Jordano 2016). Enormous sample sizes and strictly delimited boundaries are required to exhaustively sample all interactions (Martinez et al. 1999). There is undoubtedly a long tail of extremely infrequent interactions that will never be observed in a realistic sampling regime. Furthermore, there is rarely a clear line defining network boundaries (Poisot et al. 2015). Ecological networks are a constantly moving target (Rasmussen et al. 2013, Caradonna et al. 2017) and there may be little value in including certain rare, occasional visitors and unusual interactions. It can be challenging to determine if infrequent interactions are simply a curiosity (e.g., Mercer 1966) or have significant consequences (e.g., Dudley et al. 2016). The utility of increasing the accuracy with which one type of interaction is understood will plateau as the impact of other interaction types and ecological guilds predominate (Lafferty et al. 2008). Despite the challenges, we see two principal uses for inferred missing links.

First, inferring missing links to direct further sampling where the goal is a descriptive network. In many cases, the topography of the network is the focus, given the potential for interaction strengths to vary through time and in response to perturbation. For example, when tracing the relationships of species that may be the target for eradication measures, such as disease-vector mosquitoes (Collins et al. 2019), a critical first step is an inventory of interactions in the community. Since in most cases investment in sampling is a finite resource, directing sampling toward high-likelihood interactions could more efficiently increase the overall network completeness. However, we caution that directed sampling will introduce biases in the generated network (Gibson et al.

2011, Falcão et al. 2016) and may hinder cross-community comparability (Poisot et al. 2012a). In certain cases, independently sampled networks (even if less complete) may be more useful than comparisons between augmented networks. The use of information from other networks may lead them to appear more homogenous than they really are and lead to underestimates of interaction β -diversity through time and space (Poisot et al. 2015, Graham and Weinstein 2018).

Second, the identification of probable missing links can improve tests of the reliability of conclusions drawn from networks, potentially through explicitly probabilistic networks (Poisot et al. 2016a). Statistical resampling and rarefaction procedures that can only work backward or by adding random interactions may both introduce biases. While particular quantitative metrics are relatively robust to under-sampling (Blüthgen et al. 2008, Fründ et al. 2016, Henriksen et al. 2019), these metrics may not align with the goals of the study. If the conclusions are robust to the inclusion of probable missing links, then confidence in the results would be increased. Conversely, if including most-likely missing links alters conclusions, this would signal that these conclusions are unlikely to be robust. In these cases, care should be taken when selecting a predictive model to avoid a clash with the properties under investigation.

CONCLUSION

The identification of likely missing links is a powerful tool to include in the arsenal of network ecologists. We have shown that models to identify missing links can perform well across a wide range of network types. Importantly, inferential models need not be technically challenging to implement; for example, the *cassandRa* package allows all the models discussed here to be fit with a single function call. We hope that such approaches become more widely used in guiding ecological sampling and testing the robustness of conclusions drawn from bipartite networks.

ACKNOWLEDGMENTS

J. C. D. Terry initiated the research, in discussion with O. T. Lewis. J. C. D. Terry conducted all analyses and wrote the first draft of the manuscript. Both authors contributed to revisions. We thank the CERO group for comments on this project, and two anonymous reviewers for their constructive comments. Funding to both authors was provided by NERC Standard Grant NE/N010221/1.

LITERATURE CITED

- Allesina, S., and M. Pascual. 2009. Food web models: a plea for groups. *Ecology Letters* 12:652–662.
- Bartomeus, I. 2013. Understanding linkage rules in plant-pollinator networks by using hierarchical models that incorporate pollinator detectability and plant traits. *PLoS ONE* 8:1–8.
- Bartomeus, I., D. Gravel, J. M. Tylianakis, M. A. Aizen, I. A. Dickie, and M. Bernard-Verdier. 2016. A common framework

- for identifying linkage rules across different types of interactions. *Functional Ecology* 30:1894–1903.
- Bascompte, J., and P. Jordano. 2013. *Mutualistic networks*. Princeton University Press, Princeton, New Jersey, USA.
- Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences USA* 100:9383–9387.
- Berlow, E. L. 1999. Strong effects of weak interactions in ecological communities. *Nature* 398:330–334.
- Bersier, L.-F., C. Banašek-Richter, and M.-F. Cattin. 2002. Quantitative descriptors of food web matrices. *Ecology* 83:2917–2926.
- Blüthgen, N. 2010. Why network analysis is often disconnected from community ecology: A critique and an ecologist's guide. *Basic and Applied Ecology* 11:185–195.
- Blüthgen, N., J. Fründ, D. P. Vázquez, and F. Menzel. 2008. What do interaction network metrics tell us about specialization and biological traits. *Ecology* 89:3387–3399.
- Blüthgen, N., F. Menzel, and N. Blüthgen. 2006. Measuring specialization in species interaction networks. *BMC Ecology* 6:9.
- CaraDonna, P. J., W. K. Petry, R. M. Brennan, J. L. Cunningham, J. L. Bronstein, N. M. Waser, and N. J. Sanders. 2017. Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecology Letters* 20:385–394.
- Chacoff, N. P., D. P. Vázquez, S. B. Lomáscolo, E. L. Stevani, J. Dorado, and B. Padrón. 2012. Evaluating sampling completeness in a desert plant–pollinator network. *Journal of Animal Ecology* 81:190–200.
- Chao, A. 1984. Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* 11:265–270.
- Chao, A., T. C. Hsieh, R. L. Chazdon, and R. K. Colwell. 2014. Unveiling the species–rank abundance distribution by generalizing good–turing sample coverage theory. *Ecology* 96:1189–1201.
- Chao, A., and L. Jost. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93:2533–2547.
- Clauset, A., C. Moore, and M. E. J. Newman. 2008. Hierarchical structure and the prediction of missing links in networks. *Nature* 453:98–101.
- Collins, C. M., J. A. S. Bonds, M. M. Quinlan, and J. D. Mumford. 2019. Effects of the removal or reduction in density of the malaria mosquito, *Anopheles gambiae* s.l., on interacting predators and competitors in local ecosystems. *Medical and Veterinary Entomology* 33:1–15.
- Colwell, R. K., and J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society B* 345:101–118.
- Costa, J. M., L. P. da Silva, J. A. Ramos, and R. H. Heleno. 2016. Sampling completeness in seed dispersal networks: When enough is enough. *Basic and Applied Ecology* 17:155–164.
- Crea, C., R. A. Ali, and R. Rader. 2016. A new model for ecological networks using species-level traits. *Methods in Ecology and Evolution* 7:232–241.
- de Aguiar, M. A. M. et al 2019. Revealing biases in the sampling of ecological interaction networks. *PeerJ* 7:e7566.
- Delmas, E. et al 2019. Analysing ecological networks of species interactions. *Biological Reviews* 94:16–36.
- Desjardins-Proulx, P., I. Laigle, T. Poisot, and D. Gravel. 2017. Ecological interactions and the Netflix problem. *PeerJ* 5:e3644.
- Dormann, C. F., J. Fründ, and H. M. Schaefer. 2017. Identifying causes of patterns in ecological networks: opportunities and limitations. *Annual Review of Ecology, Evolution, and Systematics* 48:559–584.
- Dormann, C. F., B. Gruber, and J. Fründ. 2008. Introducing the bipartite package: Analysing ecological networks. *R News* 8:8–11.
- Dudley, J. P., B. M. Hang'Ombe, F. H. Leendertz, L. J. Dorward, J. de Castro, A. L. Subalusky, and M. Clauss. 2016. Carnivory in the common hippopotamus *Hippopotamus amphibius*: implications for the ecology and epidemiology of anthrax in African landscapes. *Mammal Review* 46:191–203.
- Dunne, J. A., and M. Pascual. 2006. *Ecological networks: linking structure to dynamics in food webs*. Oxford University Press, Oxford, UK.
- Eklöf, A. et al 2013. The dimensionality of ecological networks. *Ecology Letters* 16:577–583.
- Falcão, J. C. F., W. Dáttilo, and V. Rico-Gray. 2016. Sampling effort differences can lead to biased conclusions on the architecture of ant–plant interaction networks. *Ecological Complexity* 25:44–52.
- Fort, H., D. P. Vázquez, and B. L. Lan. 2016. Abundance and generalisation in mutualistic networks: Solving the chicken-and-egg dilemma. *Ecology Letters* 19:4–11.
- Frank, K., F.-T. Krell, E. M. Slade, E. H. Raine, L. Y. Chiew, T. Schmitt, C. S. Vairappan, P. Walter, and N. Blüthgen. 2018. Global dung webs: high trophic generalism of dung beetles along the latitudinal diversity gradient. *Ecology Letters* 21:1229–1236.
- Fründ, J., K. S. McCann, and N. M. Williams. 2016. Sampling bias is a challenge for quantifying specialization and network structure: Lessons from a quantitative niche model. *Oikos* 125:502–513.
- Galeano, J., J. M. Pastor, and J. M. Iriando. 2009. Weighted-Interaction Nestedness Estimator (WINE): A new estimator to calculate over frequency matrices. *Environmental Modelling & Software* 24:1342–1346.
- Gellner, G., and K. S. McCann. 2016. Consistent role of weak and strong interactions in high- and low-diversity trophic food webs. *Nature Communications* 7:11180.
- Gibson, R. H., B. Knott, T. Eberlein, and J. Memmott. 2011. Sampling method influences the structure of plant–pollinator networks. *Oikos* 120:822–831.
- Goldwasser, L., and J. Roughgarden. 1997. Sampling effects and the estimation of food-web properties. *Ecology* 78:41–54.
- González-Varo, J. P., and A. Traveset. 2016. The labile limits of forbidden interactions. *Trends in Ecology and Evolution* 31:700–710.
- Gotelli, N. J., and R. K. Colwell. 2010. Estimating species richness. Pages 416–422 in *A. Magurran and McGill, B. J., editors. Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford, NY.*
- Graham, C. H., and B. G. Weinstein. 2018. Towards a predictive model of species interaction beta diversity. *Ecology Letters* 21:1299–1310.
- Gray, C. et al 2014. Ecological networks: the missing links in biomonitoring science. *Journal of Applied Ecology* 51:1444–1449.
- Gray, C., D. H. Figueroa, L. N. Hudson, A. Ma, D. Perkins, and G. Woodward. 2015. Joining the dots: An automated method for constructing food webs from compendia of published interactions. *Food Webs* 5:11–20.
- Guimera, R., and M. Sales-Pardo. 2009. Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences USA* 106:22073–22078.
- Henriksen, M. V., D. G. Chapple, S. L. Chown, and M. A. McGeoch. 2019. The effect of network size and sampling

- completeness in depauperate networks. *Journal of Animal Ecology* 88:211–222.
- Jacquet, C., C. Moritz, L. Morissette, P. Legagneux, F. Massol, P. Archambault, and D. Gravel. 2016. No complexity–stability relationship in empirical ecosystems. *Nature Communications* 7:12573.
- Jordano, P. 2016. Sampling networks of ecological interactions. *Functional Ecology* 30:1883–1893.
- Jordano, P., J. Bascompte, and J. M. Olesen. 2003. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters* 6:69–81.
- Kadoya, T., G. Gellner, and K. S. McCann. 2018. Potential oscillators and keystone modules in food webs. *Ecology Letters* 21:1330–1340.
- Kéfi, S., V. Miele, E. A. Wieters, S. A. Navarrete, and E. L. Berlow. 2016. How structured is the entangled bank? The surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience. *PLOS Biology* 14:e1002527.
- Klumpers, S. G. T., M. Stang, and P. G. L. Klinkhamer. 2019. Foraging efficiency and size matching in a plant–pollinator community: the importance of sugar content and tongue length. *Ecology Letters* 22:469–479.
- Lafferty, K. D. et al 2008. Parasites in food webs: the ultimate missing links. *Ecology Letters* 11:533–546.
- Laigle, I., I. Aubin, C. Digel, U. Brose, I. Boulangeat, and D. Gravel. 2018. Species traits as drivers of food web structure. *Oikos* 127:316–326.
- Larremore, D. B., A. Clauset, and A. Z. Jacobs. 2014. Efficiently inferring community structure in bipartite networks. *Physical Review E* 90:012805.
- Lee, R. H., and B. Guénard. 2019. Choices of sampling method bias functional components estimation and ability to discriminate assembly mechanisms. *Methods in Ecology and Evolution* 10:867–878.
- Leger, J.-B., J.-J. Daudin, and C. Vacher. 2015. Clustering methods differ in their ability to detect patterns in ecological networks. *Methods in Ecology and Evolution* 6:474–481.
- MacGregor, C. J., D. M. Evans, and M. J. O. Pocock. 2017. Estimating sampling completeness of interactions in quantitative bipartite ecological networks: incorporating variation in species specialisation. *bioRxiv*:195917.
- Martinez, N. D., B. A. Hawkins, H. A. Dawah, and B. P. Feifarek. 1999. Effects of sampling effort on characterization of food-web structure. *Ecology* 80:1044–1055.
- McCann, K., A. Hastings, and G. R. Huxel. 1998. Weak trophic interactions and the balance of nature. *Nature* 395:794–798.
- Memmott, J. 2009. Food webs: a ladder for picking strawberries or a practical tool for practical problems? *Philosophical Transactions of the Royal Society B* 364:1693–1699.
- Mercer, A. J. 1966. Turnstones feeding on human corpse. *British Birds* 59:307.
- Montoya, J. M., S. L. Pimm, and R. V. Solé. 2006. Ecological networks and their fragility. *Nature* 442:259–264.
- Morales-Castilla, I., M. G. Matias, D. Gravel, and M. B. Araújo. 2015. Inferring biotic interactions from proxies. *Trends in Ecology & Evolution* 30:347–356.
- Morris, R. J., S. Gripenberg, O. T. Lewis, and T. Roslin. 2014. Antagonistic interaction networks are structured independently of latitude and host guild. *Ecology Letters* 17:340–349.
- Nielsen, A., and J. Bascompte. 2007. Ecological networks, nestedness and sampling effort. *Journal of Ecology* 95:1134–1141.
- Novotny, V. et al 2010. Guild-specific patterns of species richness and host specialization in plant–herbivore food webs from a tropical forest. *Journal of Animal Ecology* 79:1193–1203.
- Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs, and H. Wagner. 2018. *vegan: Community Ecology Package*. <https://CRAN.R-project.org/package=vegan>
- Olesen, J. M., J. Bascompte, Y. L. Dupont, H. Elberling, C. Rasmussen, and P. Jordano. 2011. Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B* 278:725–732.
- Olesen, J. M., J. Bascompte, Y. L. Dupont, and P. Jordano. 2007. The modularity of pollination networks. *Proceedings of the National Academy of Sciences USA* 104:19891–19896.
- Olesen, J. M., J. Bascompte, H. Elberling, and P. Jordano. 2008. Temporal dynamics in a pollination network. *Ecology* 89:1573–1582.
- Paine, R. T. 1980. Food Webs: Linkage, interaction strength and community infrastructure. *Journal of Animal Ecology* 49:666–685.
- Pichler, M., V. Boreux, A. M. Klein, M. Schleuning, and F. Hartig. 2020. Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology and Evolution* 11:281–293.
- Poisot, T., E. Canard, D. Mouillot, N. Mouquet, and D. Gravel. 2012a. The dissimilarity of species interaction networks. *Ecology Letters* 15:1353–1361.
- Poisot, T., E. Canard, N. Mouquet, and M. E. Hochberg. 2012b. A comparative study of ecological specialization estimators. *Methods in Ecology and Evolution* 3:537–544.
- Poisot, T., A. R. Cirtwill, K. Cazelles, D. Gravel, M.-J. Fortin, and D. B. Stouffer. 2016a. The structure of probabilistic networks. *Methods in Ecology and Evolution* 7:303–312.
- Poisot, T., D. B. Stouffer, and D. Gravel. 2015. Beyond species: why ecological interaction networks vary through space and time. *Oikos* 124:243–251.
- Poisot, T., D. B. Stouffer, and S. Kéfi. 2016b. Describe, understand and predict: why do we need networks in ecology? *Functional Ecology* 30:1878–1882.
- Pomeranz, J. P. F., R. M. Thompson, T. Poisot, and J. S. Harding. 2018. Inferring predator–prey interactions in food webs. *Methods in Ecology and Evolution* 2018:1–12.
- Rasmussen, C., Y. L. Dupont, J. B. Mosbacher, K. Trøjelsgaard, and J. M. Olesen. 2013. Strong impact of temporal resolution on the structure of an ecological network. *PLoS ONE* 8:e81694.
- Rivera-Hutinel, A., R. O. Bustamante, V. H. Marín, and R. Medel. 2012. Effects of sampling completeness on the structure of plant–pollinator networks. *Ecology* 93:1593–1603.
- Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77.
- Rohr, R. P., R. E. Naisbit, C. Mazza, and L.-F. Bersier. 2016. Matching–centrality decomposition and the forecasting of new links in networks. *Proceedings of the Royal Society B* 283:20152702.
- Rohr, R. P., R. E. Naisbit, C. Mazza, and L.-F. Bersier. 2018. Statistical approaches for inferring and predicting food-web architecture. Pages 178–192 in *J. C. Moore, de Ruiter, P. C., McCann, K. S. and Wolters, V., editors. Adaptive food webs, stability and transitions of real and model ecosystems. Cambridge University Press, Cambridge, UK.*
- Sander, E. L., J. T. Wootton, and S. Allesina. 2015. What can interaction webs tell us about species roles? *PLoS Computational Biology* 11:e1004330.

- Schleuning, M. et al 2014. Ecological, historical and evolutionary determinants of modularity in weighted seed-dispersal networks. *Ecology Letters* 17:454–463.
- Seo, E., and R. A. Hutchinson. 2018. Predicting links in plant-pollinator interaction networks using latent factor models with implicit feedback. Pages 808–815 *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*.
- Simmons, B. I. et al 2019. Abundance drives broad patterns of generalisation in plant–hummingbird pollination networks. *Oikos* 128: 1287–1295.
- Song, C., R. P. Rohr, and S. Saavedra. 2017. Why are some plant–pollinator networks more nested than others? *Journal of Animal Ecology* 86:1417–1424.
- Spitz, J., V. Ridoux, and A. Brind’Amour. 2014. Let’s go beyond taxonomy in diet description: Testing a trait-based approach to prey–predator relationships. *Journal of Animal Ecology* 83:1137–1148.
- Staniczenko, P. P. A., J. C. Kopp, and S. Allesina. 2013. The ghost of nestedness in ecological networks. *Nature Communications* 4:1391.
- Staniczenko, P. P. A., O. T. Lewis, J. M. Tylianakis, M. Albrecht, V. Coudrain, A.-M. Klein, and F. Reed-Tsochas. 2017. Predicting the effect of habitat modification on networks of interacting species. *Nature Communications* 8:792.
- Stock, M., T. Poisot, W. Waegeman, and B. De Baets. 2017. Linear filtering reveals false negatives in species interaction data. *Scientific Reports* 7:45908.
- Terry, J. C. D. 2019. cassandRa: Finds Missing Links and Metric Confidence Intervals in Ecological Bipartite Networks. R package version 0.1.0. <https://CRAN.R-project.org/package=cassandRa>
- Terry, J. C. D., R. J. Morris, and M. B. Bonsall. 2017. Trophic interaction modifications: an empirical and theoretical framework. *Ecology Letters* 20:1219–1230.
- Traveset, A., J. M. Olesen, M. Nogales, P. Vargas, P. Jaramillo, E. Antolín, M. M. Trigo, and R. Heleno. 2015. Bird–flower visitation networks in the Galápagos unveil a widespread interaction release. *Nature Communications* 6:6376.
- Tylianakis, J. M., E. Laliberté, A. Nielsen, and J. Bascompte. 2010. Conservation of species interaction networks. *Biological Conservation* 143:2270–2279.
- Tylianakis, J. M., T. Tschamtkke, and O. T. Lewis. 2007. Habitat modification alters the structure of tropical host–parasitoid food webs. *Nature* 445:202–205.
- Ulrich, W., M. Almeida-Neto, and N. J. Gotelli. 2009. A consumer’s guide to nestedness analysis. *Oikos* 118:3–17.
- Valdovinos, F. S. 2019. Mutualistic networks: moving closer to a predictive theory. *Ecology Letters* 22:1517–1534.
- Valiente-Banuet, A. et al 2015. Beyond species loss: The extinction of ecological interactions in a changing world. *Functional Ecology* 29:299–307.
- Vázquez, D. P., N. Bluthgen, L. Cagnolo, and N. P. Chacoff. 2009. Uniting pattern and process in plant–animal mutualistic networks: A review. *Annals of Botany* 103:1445–1457.
- Vizentin-Bugoni, J., P. K. Maruyama, V. J. Debastiani, and L. S. Duarte, B. Dalsgaard, and M. Sazima. 2016. Influences of sampling effort on detected patterns and structuring processes of a Neotropical plant–hummingbird network. *Journal of Animal Ecology* 85:262–272.
- Vizentin-Bugoni, J., P. K. Maruyama, and M. Sazima. 2014. Processes entangling interactions in communities: forbidden links are more important than abundance in a hummingbird–plant network. *Proceedings of the Royal Society B* 281:20132397.
- Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of “small-world” networks. *Nature* 393:440–442.
- Weinstein, B. G., and C. H. Graham. 2017. On comparing traits and abundance for predicting species interactions with imperfect detection. *Food Webs* 11:17–25.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1002/ecy.3047/supinfo>

DATA AVAILABILITY

All code and necessary data to generate simulated data sets and carry out analyses are archived on Zenodo: <https://doi.org/10.5281/zenodo.3686770>.